Estimation of Large Volatility Matrices with Low-Rank Signal Plus Sparse Noise Structures

Runyu Dai, Yasumasa Matsuda

Graduate School of Economics and Management, Tohoku University

June 26, 2023

Abstract

In this paper, we propose a parsimonious model to estimate large volatility matrices by combining DCC-GARCH, sparsity-induced weak factors (sWFs) and POET framework in Fan et al. (2013). We call this method the DCC and sWFs extended POET (DCC-ePOET). Built on the mixed factor structures, we estimate volatility matrices through the univariate volatilities of observable factors and weak latent factors with a linear transformation. We further include a sparse noise covariance estimator obtained by an adaptive threshold method proposed in POET to address the singularity issue when the cross-sectional dimension N is larger than the sample size T, and capture the weak correlations in the factor models' idiosyncratic terms. Simulation studies show that our proposed method achieves good finite-sample performance. Empirical studies demonstrate that the developed method is superior to several candidates in the analysis of out-of-sample minimum variance portfolio allocations.

Keywords: Volatility matrix; multivariate GARCH; factor models; thresholding; high-dimensional data; ePOET.

1 Introduction

Much existing research focuses on estimating large static covariance matrix. However, over long time horizons, the covariance matrices of economic or financial time series variables often show structural changes such as volatility clustering and structural breaks. Therefore, there has been a growing interest in estimating a wide variety of large dynamic covariance matrices in recent years. see, e.g., Guo et al. (2017), Chen et al. (2019), Dendramis et al. (2021), Wang et al. (2021), and Ke et al. (2022). See more in a recent survey by Li (2021). In this paper, we study a typical type of large dynamic covariance matrices, volatility matrices, which are used for modelling the co-volatile structures of many financial time series.

There are various ways to model volatility matrices. Among them, the multivariate (generalized) autoregressive conditional heteroskedasticity model, a multivariate extension of famous univariate ARCH (Engle, 1982) and GARCH (Bollerslev, 1986) model, is a popular strand. Multivariate ARCH model is first considered by Engle et al. (1984) and then a general version, multivariate GARCH (MGARCH), is introduced by Bollerslev et al. (1988) in the form of VECH-GARCH. Later, Engle and Kroner (1995) extend VECH-GARCH to famous BEKK model.

It is well-known that conventional MGARCH models usually suffer from the curse of dimensionality that the number of parameters will be significantly increasing as the cross-sectional dimension (or number of series) N increases. To address this problem, Bollerslev (1990) propose the constant-conditional-correlated GARCH (CCC-GARCH) model, which assume the MGARCH model comprises N univariate GARCH processes that are related through with a constant conditional correlation matrix. In the early 2000s, Engle (2002) extended the CCC-GARCH to DCC-GARCH to allows for time-varying conditional correlation. However, in large or high dimension, estimating the targeting correlation matrix of CCC or DCC-GARCH is quite challenging. Engle et al. (2019) recently propose DCC-NL-GARCH model by employing nonlinear shrinkage (NL) method of Ledoit and Wolf (2012) to estimate the correlation targeting matrix of DCC in large dimension.

On the other hand, Ding (1994) first proposes the principal component (PC) GARCH to capture the co-volatility by several PCs. PC-GARCH provides an effective way to parameterize the MGARCH model and has been extended to many variants, such as O-GARCH and GO-GARCH models. See Alexander (2000) and Van der Weide (2002). Note that PCs also can be treated as some latent factors that drive the variation of the data, so PC-GARCH is often denoted as a type of Factor-GARCH model. Factor-GARCH family mainly assumes that the data is composed of a series of common factors, whatever real and latent factors. Besides PC-GARCH, many other types of Factor-GARCH have been extensively studied in the past two decades. Vrontos et al. (2003) propose the Full-Factor GARCH model, assuming the number of conditionally uncorrelated factors are the same as N. Hafner and Preminger (2009) prove the asymptotic properties of the QMLE for general Factor-GARCH models. In Zhang and Chan (2009), they further improve the Factor-GARCH model by embedding DCC to capture the dynamic structures in the factors. The aforementioned methods are based on so-called static factor models, namely factor loadings are assumed to be time invariant. More flexible and complex dynamic factor models are also utilised to estimate MGARCH in the past decade. See typical examples such as Santos and Moura (2014) and Barigozzi and Hallin (2017).

In the high dimensional space, Fan et al. (2008), Fan et al. (2011) and Fan et al. (2013) consecutively propose the structure of common factors plus noise covariance matrix to model the static covariance matrix. The first two assume the factors are known such as famous Famma-French 3 and 5 factors, whereas the third one captures the unobservable latent factors by PCs. In the same paper, they further assume that the error terms after extracting latent factors are weakly correlated so that the error covariance has conditional sparsity structure. Combined PCs and sparse error matrix, they propose the famous the Principal Orthogonal complEment Thresholding (POET) framework. Inspired from Fan's consecutive contributions, Li et al. (2022) use observable factors equipped with CCC-GARCH structure plus the diagonal matrix of sample error covariance to estimate large volatility matrix. This method is denoted by Factor-CCC-Diag hereafter. Following Fan et al. (2011) and Fan et al. (2013), Li et al. (2022) extends the diagonal error covariance matrix estimate of Factor-CCC-Diag to a sparse one, and we denote this modification as CCC-POET. However, using only observable factors to estimate covariance matrix arises a typical concern of omitted variables and factors. For instance, see omitted factors issues in asset pricing by Giglio and Xiu (2021) and the research of factor zoos by Feng et al. (2020). A natural and intuitive remedy is to extract more latent factors in the residuals. Shi et al. (2022) augment Fama-French models by introducing latent factors to extract information from Fama-French model's residuals for estimating large covariance matrices. Their approach inherently assumes that the latent factors in the residuals (LFR) are strong factors, namely LFR diverges with the same rate of N. However, as pointed out in the extended POET (ePOET) method by Dai et al. (2022), the latent factors in the residuals of Fama-French models are usually relatively weak in that each latent factor may diverge slower than N. In particular, the scree plot in that article shows that the sample eigenvalues of residuals decreases in a relatively smooth manner, which violates the strong factor assumption that a big gap exists among the eigenvalues. Following these discussions, based on ePOET, we propose a novel large volatility matrix estimator by utilising the benefits of Factor-GARCH and weak latent factors in the residuals. The weak latent factors can be estimated by the sparsity-induced weak factor (sWF) framework developed in Uematsu and Yamagata (2022). Moreover, to further capture the dynamic structure of observable factors, following DCC-Factor strategy proposed by Zhang and Chan (2009), we assume the observable factors have DCC-GARCH structure and we call our method DCC-ePOET.

1.1 Notations

Throughout the paper, $\mathbf{I}_N$ is a $N \times N$ identity matrix and $\mathbf{0}$ is a zero matrix. For any square matrix $\mathbf{A}$, we denote the kth, the largest and the smallest largest eigenvalues by $\lambda_k(\mathbf{A})$, $\lambda_{\max}(\mathbf{A})$ and $\lambda_{\min}(\mathbf{A})$, respectively. $|\mathbf{A}|$ means the determinant of $\mathbf{A}$. For any matrix $\mathbf{M} = (m_{ti}) \in \mathbb{R}^{T\times N}$, we define l_2 (spectral) norm, the Frobenius norm, the entrywise maximum norm and l_1 norm as $\|\mathbf{M}\|_2 = \lambda_{\max}^{1/2}(\mathbf{M}'\mathbf{M})$, $\|\mathbf{M}\|_{\mathrm{F}} = (\sum_{t,i} m_{ti}^2)^{1/2}$, $\|\mathbf{M}\|_{\max} = \max_{t,i}|m_{ti}|$, respectively. Given a $N \times N$ positive definite matrix $\mathbf{\Sigma}$, the weighted quadratic norm of an $N \times N$ matrix $\mathbf{P}$ of $\mathbf{\Sigma}$ is defined as $\|\mathbf{P}\|_{\mathbf{\Sigma}} = N^{-1/2}\|\mathbf{\Sigma}^{-1/2}\mathbf{P}\mathbf{\Sigma}^{-1/2}\|_{\mathrm{F}}$. Let $\lesssim$ and $\gtrsim$ represent $\leqslant$ and $\geqslant$ up to a positive constant factor. For two positive values x and y, let $x \wedge y =: \min\{x, y\}$ and $x \vee y =: \max\{x, y\}$. Finally, for two positive sequences a_n and b_n depending on n, $a_n \asymp b_n$ if $a_n \gtrsim b_n$ and $a_n \lesssim b_n$. The terms "errors","noise", and "idiosyncratic terms/errors" share the same meaning in the paper.

1.2 Organisation

The rest of the paper is arranged as follows. Section 2 introduces the DCC-ePOET model. The estimation methodologies are discussed in Section 3. In Section 4, we discuss the selection of the threshold constant for the DCC-ePOET estimator. Two sets of Monte Carlo experiments are presented in Section 5. In Section 6, we further compare the DCC-ePOET model with other candidate models in portfolio analysis. Section 7 concludes this paper.

2 Model descriptions

We consider the estimation of the volatility matrix of N-dimensional financial series vector $\mathbf{y}_t$, generated by the linear model

$$\mathbf{y}_t = \mathbf{A}^0\mathbf{x}_t + \mathbf{u}_t, \tag{2.1}$$

where $\mathbf{x}_t$ and $\mathbf{A}^0 = (\mathbf{a}_1^0, \dots, \mathbf{a}_r^0)$ represent the r-dimensional observable factor and its factor loadings, respectively. Furthermore, the error term $\mathbf{u}_t$ has the latent factor structure

$$\mathbf{u}_t = \mathbf{B}^0\mathbf{f}_t^0 + \mathbf{e}_t, \tag{2.2}$$

where $\mathbf{f}_t^0$ and $\mathbf{B}^0 = (\mathbf{b}_1^0, \dots, \mathbf{b}_K^0)$ represent the K-dimensional unobservable factor and its factor loadings, respectively, and $\mathbf{e}_t$ is the idiosyncratic error term. We can also rewite the models in the matrix from: $\mathbf{Y} = \mathbf{X}\mathbf{A}' + \mathbf{U}$ and $\mathbf{U} = \mathbf{F}\mathbf{B}' + \mathbf{E}$. The number of observable factors r and the number of latent factors K are assumed to be finite and small. We further suppose that $\mathbf{x}_t$ and latent parts $\mathbf{u}_t$ are uncorrelated to separately estimate $\mathbf{A}^0$ and $\mathbf{B}^0$. One may stand on a more rigorous model that considers the endogeneity problem such as Bai (2009). It is also worth mentioning that to avoid omitted variable problem, model (2.1) is allowed to include many observable factors, whereas we do not pursuit theses directions to avoid technical and computational issues. Without loss of generality, we suppose the following conditions to identify the latent factors and loadings,

$$E[\mathbf{f}_t^0\mathbf{f}_t^{0'}] = \mathbf{I}_K \quad \text{and} \quad \mathbf{B}^{0'}\mathbf{B}^0 \text{ diagonal}, \tag{2.3}$$

throughout the paper. For the unobservable weak factors in 2.2, we formally introduce the sWF model according to Uematsu and Yamagata (2022). We assume that $\lambda_{\min}(\mathbf{B}^{0'}\mathbf{B}^0)$ is bounded away from zero and $\lambda_{\max}(\mathbf{\Sigma}_e)$ is bounded from infinity. Then for $k = 1, \ldots, K$, the latent factor model in (2.2) is called the *weak factor (WF) model* if

$$\lambda_k(\mathbf{B}^0\mathbf{B}^{0'}) \asymp N^{\alpha_k}, \quad k = 1, \ldots, K \tag{2.4}$$

for some $\alpha_k \in (0, 1]$. The sWF model is realised by assuming $\mathbf{B}^0$ is sparse. That is for each $k = 1, \ldots, K$, suppose $\mathbf{b}_k^0$ has $N_k := \lfloor N^{\alpha_k} \rfloor$ nonzero elements for some constant $\alpha_k \in (0, 1]$, and this is so-called sparsity-induced WF (sWF). Notice that α_k being close to zero means a factor is extremely weak, while $\alpha_k = 1$ implies a so-called strong factor. In contrast to the sWF model, a typical strong factor model such as Fan et al. (2013) relies on

$$\lambda_k(\mathbf{B}^0\mathbf{B}^{0'}) \asymp N, \quad k = 1, \ldots, K. \tag{2.5}$$

Regarding the volatility modeling of $\mathbf{y}_t$, for each $l = 1, ..., r$ and $k = 1, ..., K$, we denote the univariate volatility $(h_{tl}^x, ..., h_{tr}^x)$ for observable factor $\mathbf{x}_l$ and $(h_{t1}^f, ..., h_{tK}^f)$ for latent factor $\mathbf{f}_k$. The observable factors $\mathbf{x}_t$ is assumed to have DCC-GARCH(p, q) structure as follows. For i.i.d $\boldsymbol{\eta}_t$ and $\mathbf{D}_t = \operatorname{diag}\left(\sqrt{h_{t1}^x}, ..., \sqrt{h_{tr}^x}\right)$,

$$\begin{aligned}
\mathbf{x}_t &= \mathbf{\Sigma}_x(t)^{1/2}\boldsymbol{\eta}_t, \quad \mathbf{\Sigma}_x(t) = \mathbf{D}_t\mathbf{R}_t\mathbf{D}_t, \\
\mathbf{R}_t &= \operatorname{diag}\mathbf{Q}_t{}^{-1/2}\mathbf{Q}_t \operatorname{diag}\mathbf{Q}_t{}^{-1/2}, \\
\mathbf{Q}_t &= (1 - a - b)\mathbf{S} + a\boldsymbol{\eta}_{t-1}^*\boldsymbol{\eta}_{t-1}^{*'} + b\mathbf{Q}_{t-1}, \\
h_{tl}^x &= \omega_l^0 + \sum_{i=1}^{q} \alpha_{il}^0 x_{t-i,l}^2 + \sum_{i=1}^{p} \beta_{il}^0 h_{t-i,l}^x,
\end{aligned} \tag{2.6}$$

where $\boldsymbol{\eta}_t^* = \left(x_{t1}/\sqrt{h_{t1}^x}, ..., x_{tr}/\sqrt{h_{tr}^x}\right)$ and $\mathbf{S}$ is a positive-definite unconditional covariance matrix. Note that h_{tl}^x demonstrates a standard univariate GARCH(p, q) process (Bollerslev, 1986) and the definition of h_{tk}^f is similar. Then, under the orthogonal condition between $\mathbf{x}_t$ and $\mathbf{u}_t$, the volatility matrix of $\mathbf{y}_t$ is given by

$$\mathbf{\Sigma}_y(t) = \mathbf{A}^0\mathbf{\Sigma}_x(t)\mathbf{A}^{0'} + \mathbf{B}^0\mathbf{\Sigma}_f(t)\mathbf{B}^{0'} + \mathbf{\Sigma}_e, \tag{2.7}$$

where $\mathbf{\Sigma}_f(t)$ is the volatility matrix of latent factors. In this work, we assume the latent factors are conditionally uncorrelated so that $\mathbf{\Sigma}_f(t) = \operatorname{diag}\left(h_{t1}^f, ..., h_{tK}^f\right)$. One may also assume DCC (or CCC) structure in the latent factors if necessary, while our empirical studies will show that equipping DCC to latent factors makes a very tiny difference.

After taking out the pervasive observable and (weak) latent factors from the target variables, it is reasonable to assume the remaining terms are weakly correlated. Thus, following Cai and Liu (2011) and Fan et al. (2011), we assume the noise covariance matrix $\mathbf{\Sigma}_e$ is conditionally sparse. That is $\mathbf{\Sigma}_e \in \Upsilon(m_N)$, where

$$\Upsilon(m_N) = \left\{ \mathbf{\Sigma}_e = (\sigma_{ij}^e)_{N\times N} \succ 0 : \max_{i\leqslant N} \sum_{j=1}^{N} |\sigma_{ij}^e|^q (\sigma_{ii}^e \sigma_{jj}^e)^{(1-q)/2} \leqslant m_N \right\} \tag{2.8}$$

for the sparsity measure $m_N = \max_{i\leqslant N} \sum_{j\leqslant N} |\sigma_{ij}^e|^q$ with some constant $q \in [0, 1]$.

3 Estimation

The estimation procedure first applies the ordinary least squares (OLS) to (2.1) to obtain the estimated loadings $\hat{\mathbf{A}}$ for observable factors and the initial residuals $\hat{\mathbf{U}}$. Then, we estimate latent factors with loadings from $\hat{\mathbf{U}}$ and calculate the error $\hat{\mathbf{E}}$. After these, the remaining is to estimate the volatility matrix of observable factors, the volatility matrix of latent factors and the error covariance matrix, respectively.

3.1 Estimation of the volatility matrix of observable factors

As the factors and r are known, we only have to estimate the r-dimensional DCC-GARCH model in this stage. The estimation is simply based on standard quasi-maximum likelihood estimation (QMLE) and not computationally challenging due to the small r. We follow the usual DCC-GARCH estimation method proposed in Engle (2002). The log-likelihood function is given by

$$\begin{aligned}
\mathbf{L} &= -\frac{1}{2}\sum_{t=1}^{T}[r\log(2\pi) + \log|\mathbf{\Sigma}_x(t)| + \mathbf{x}_t'\mathbf{\Sigma}_x(t)^{-1}\mathbf{x}_t] \\
&= -\frac{1}{2}\sum_{t=1}^{T}[r\log(2\pi) + \log|\mathbf{D}_t\mathbf{R}_t\mathbf{D}_t| + \mathbf{x}_t'\mathbf{D}_t^{-1}\mathbf{R}_t^{-1}\mathbf{D}_t^{-1}\mathbf{x}_t] \\
&= -\frac{1}{2}\sum_{t=1}^{T}[r\log(2\pi) + 2\log|\mathbf{D}_t| + \mathbf{x}_t'\mathbf{D}_t^{-2}\mathbf{x}_t - \boldsymbol{\eta}_t'\boldsymbol{\eta}_t + \log|\mathbf{R}_t| + \boldsymbol{\eta}_t'\mathbf{R}_t^{-1}\boldsymbol{\eta}_t].
\end{aligned}$$

We observe that $\mathbf{L}$ can be decomposed into two sub-functions, the univariate volatility one and the conditional correlation one. Denote θ_g^x as the univariate GARCH parameter for observable factors and $\phi_c = (a, b)$ as the correlation parameter. The usual stationarity conditions: $0 < \sum_{i=1}^{q} \alpha_{il} + \sum_{i=1}^{p} \beta_{il} < 1, \omega_l > 0$ for GARCH(p,q) and $a, b \geqslant 0, a+b < 1$ for DCC are satisfied. Then, we have

$$\begin{aligned}
\mathbf{L}(\theta_g^x, \phi_c) &= -\frac{1}{2}\sum_{t=1}^{T}[r\log(2\pi) + 2\log|\mathbf{D}_t| + \mathbf{x}_t'\mathbf{D}_t^{-2}\mathbf{x}_t] \\
&\quad - \frac{1}{2}\sum_{t=1}^{T}[-\boldsymbol{\eta}_t'\boldsymbol{\eta}_t + \log|\mathbf{R}_t| + \boldsymbol{\eta}_t'\mathbf{R}_t^{-1}\boldsymbol{\eta}_t] \\
&= \mathbf{L}_g(\theta_g^x) + \mathbf{L}_c(\theta_g^x, \phi_c), \qquad (3.9)
\end{aligned}$$

where $\mathbf{L}_g(\theta_g^x)$ is for estimating the univariate GARCH, and $\mathbf{L}_c(\theta_g^x, \phi_c)$ is for estimating the conditional correlation matrix. It is obvious that given θ_g^x, the maximum of $\mathbf{L}_c$ can be achieved, leading to a two-step QMLE. The final estimator of DCC-GARCH is then obtained by solving the following:

$$\hat{\theta}_g^x = \arg\min -\mathbf{L}_g(\theta_g^x), \quad \hat{\phi}_c = \max_{\phi_c} \mathbf{L}_c(\hat{\theta}_g^x, \phi_c).$$

Denote the volatility estimator and correlation estimator according to above estimations by $\hat{h}_{tl}^x$ and $\hat{\mathbf{R}}_t$, respectively. The estimator of $\boldsymbol{\Sigma}_x(t)$ is given by $\hat{\boldsymbol{\Sigma}}_x(t) = \hat{\mathbf{D}}_t\hat{\mathbf{R}}_t\hat{\mathbf{D}}_t$. Recall the model for DCC-GARCH, we also need an estimate for unconditional covariance matrix $\mathbf{S}$ to conduct the correlation estimation, and in our case we use $\hat{\mathbf{S}} = T^{-1}\sum_{t=1}^{T}\hat{\boldsymbol{\eta}}_t^*\hat{\boldsymbol{\eta}}_t^{*'}$, where $\hat{\boldsymbol{\eta}}_t^* = \left(x_{t1}/\sqrt{\hat{h}_{t1}^x}, ..., x_{tr}/\sqrt{\hat{h}_{tr}^x}\right)$. Note that in this study, we assume the number of observable factors, r, is finite and small, and T is very large. Thus, compared to the sample covariance matrix in high-dimensional cases, the sample version in our work is not a concern. If r is allowed to be large or grow comparably to N and T, another issue of estimating a large dimensional covariance matrix will arise. Furthermore, it has been pointed out in Aielli (2013) that the sample version, $\hat{\mathbf{S}}$, may lead to an inconsistent estimate of the DCC-GARCH model as $\mathbf{S} \neq E(\boldsymbol{\eta}_t^*\boldsymbol{\eta}_t^{*'})$, thus he propose a corrected or tractable DCC (cDCC) model as a remedy to this issue. For more details, please check that article. While based on our preliminary numerical studies, cDCC and DCC produce almost identical results (when r is small.). Thus we keep the usual DCC in our work.

3.2 Estimation of the volatility matrix of latent factors

To estimate the sWF model, we first obtain *sparse orthogonal factor regression (SOFAR)* estimator (Uematsu et al., 2019; Uematsu and Yamagata, 2022) of $(\mathbf{B}, \mathbf{F})$ by

$$(\hat{\mathbf{B}}, \hat{\mathbf{F}}) = \underset{(\tilde{\mathbf{B}},\tilde{\mathbf{F}}) \in \mathbb{R}^{N\times \hat{K}} \times \mathbb{R}^{T\times \hat{K}}}{\arg\min} \frac{1}{2}\|\hat{\mathbf{U}} - \tilde{\mathbf{F}}\tilde{\mathbf{B}}'\|_{\mathrm{F}}^2 + \eta_{NT}\|\tilde{\mathbf{B}}\|_1 \tag{3.10}$$

$$\text{subject to} \quad \tilde{\mathbf{F}}'\tilde{\mathbf{F}}/T = \mathbf{I}_{\hat{K}} \quad \text{and} \quad \tilde{\mathbf{B}}'\tilde{\mathbf{B}} \text{ diagonal,}$$

where $\hat{K}$ is the estimated number of factors and $\eta_{NT} > 0$ is a penalty coefficient. Next, we focus on the estimation of GARCH(p, q) model using estimated factors $\hat{\mathbf{F}}$ through QMLE. For each $k = 1, ..., K$, the volatility is modelled by GARCH(p, q):

$$h_{tk}^f = h_{tk}^f(\theta_k^f, \hat{f}_{tk}) = \omega_k^f + \sum_{i=1}^q \alpha_{ik}^f \hat{f}_{t-i,k}^2 + \sum_{i=1}^p \beta_{ik}^f h_{t-i,k}^f$$

And the conditional log-likelihood function is given by

$$\mathbf{L}_k = \mathbf{L}_k(\theta_k) = \frac{1}{T}\sum_{t=1}^T \log\left(h_{tk}^f\right) + \frac{\hat{f}_{tk}^2}{h_{tk}^f},$$

The univariate GARCH parameter $\hat{\theta}_k^f$ is achieved according to

$$\hat{\theta}_k^f = \underset{\theta_k \in \Theta_k^f}{\operatorname{argmin}} \; -\mathbf{L}_k(\theta_k),$$

where Θ_k^f is a compact parametric space in which the stationarity conditions $0 < \sum_{i=1}^q \alpha_{ik}^f + \sum_{i=1}^p \beta_{ik}^f < 1$ and $\omega_k^f > 0$ are satisfied. The univariate volatility estimator is denoted by $\hat{h}_{tk}^f$ so that the estimator of $\mathbf{\Sigma}_f(t)$ is given by $\hat{\mathbf{\Sigma}}_f(t) = \operatorname{diag}\left(\sqrt{\hat{h}_{t1}^f}, ..., \sqrt{\hat{h}_{tK}^f}\right)$.

3.3 Determining the number of latent factors

In practice, we have to estimate the number of latent factors. There have been many techniques developed in the past decades, including famous Bai and Ng (2002) and Ahn and Horenstein (2013). However, these two methods are designed for the strong factor models with all the K signal eigenvalues diverging proportionally to N. For example, POET paper adopts the information criteria (IC) methods from Bai and Ng (2002), while the simulation studies of Uematsu and Yamagata (2022) demonstrates the $\hat{K}$ estimated by IC could be far away from the true K, especially when weak factors exist. See the details in that paper. In this work, we recommend to use the approach of Onatski (2010). Basically, this method is to determine the number of (weak) factors by $\hat{K} = \hat{K}(\delta)$ with

$$\hat{K}(\delta) = \{k = 1, \ldots, k_{\max} - 1 : \ \lambda_k - \lambda_{k+1} \geqslant \delta\},$$

where $\delta > 0$ is a fixed constant. Empirically, δ is predetermined on a calibration by the *edge distribution* (ED) method of Onatski (2010). The consistency of this selection has been proved in Uematsu and Yamagata (2022) for sWF models.

3.4 Noise covariance matrix via POET.

We follow Fan et al. (2013) to impose the conditional sparsity (2.8) on the noise covariance matrix. We obtain the estimate of $\mathbf{\Sigma}_e$ by employing adaptive thresholding techniques on the sample covariance matrix of the residuals $\hat{\mathbf{E}} = \hat{\mathbf{U}} - \hat{\mathbf{F}}\hat{\mathbf{B}}'$ as follows.

$$\hat{\mathbf{\Sigma}}_e^{\tau} = (\hat{\sigma}_{ij}^{\tau})_{N\times N} \ \text{ with } \ \hat{\sigma}_{ij}^{\tau} = \begin{cases} \hat{\sigma}_{ii}^{e} & \text{for } i = j, \\ s_{ij}^{\tau}(\hat{\sigma}_{ij}^{e}) & \text{for } i \neq j. \end{cases} \tag{3.11}$$

where $\hat{\sigma}_{ij}^{e}$ is the (i,j)-th entry of the sample covariance matrix of $\hat{\mathbf{e}}_t$, $s_{ij}^{\tau}(\cdot)$ is a soft thresholding function given by

$$s_{ij}^{\tau}(\hat{\sigma}_{ij}^{e}) = \operatorname{sign}(\hat{\sigma}_{ij}^{e}) \max(0, |\hat{\sigma}_{ij}^{e}| - \tau_{ij})$$

with

$$\tau_{ij} = C_\tau \omega_{NT} (\hat{\theta}_{ij})^{1/2} \ \text{ and } \ \hat{\theta}_{ij} = T^{-1} \sum_{t=1}^{T} (\hat{e}_{ti}\hat{e}_{tj} - \hat{\sigma}_{ij}^{e})^2 \tag{3.12}$$

for some sequence $\omega_{NT} > 0$ depending on N & T, and sufficiently large constant $C_\tau > 0$. Finally, the estimate of volatility matrix of $\mathbf{y}_t$ is given by

$$\begin{aligned} \hat{\mathbf{\Sigma}}_y(t) &= \hat{\mathbf{A}}\hat{\mathbf{\Sigma}}_x(t)\hat{\mathbf{A}}' + \hat{\mathbf{B}}\hat{\mathbf{\Sigma}}_f(t)\hat{\mathbf{B}}' + \hat{\mathbf{\Sigma}}_e^{\tau}, \\ &= \hat{\mathbf{A}}\hat{\mathbf{D}}_t\hat{\mathbf{R}}_t\hat{\mathbf{D}}_t\hat{\mathbf{A}}' + \hat{\mathbf{B}} \operatorname{diag}\left(\sqrt{\hat{h}_{t1}^{f}}, ..., \sqrt{\hat{h}_{tK}^{f}}\right) \hat{\mathbf{B}}' + \hat{\mathbf{\Sigma}}_e^{\tau}. \end{aligned} \tag{3.13}$$

Remark 3.1 *In practice, we have to determine some rate ω_{NT} in (3.12). Note that Dai et al. (2022) prove that under some mild assumptions, $\|\hat{\mathbf{\Sigma}}_e^{\tau} - \mathbf{\Sigma}_e\| \lesssim \omega_{NT}^{1-q} m_N$ with high probability, where*

$$\omega_{NT} = T^{-1/2} \frac{N_1^{3/2}(N_1 \vee T)^{1/2}}{N_K^{3/2}(N_K \wedge T)^{1/2}} \log^{1/2}(N \vee T) \log^{1/2} T,$$

and we may use this complex rate ω_{NT} or $T^{-1/2}\log^{1/2} N$ for simplicity. They perform almost identically because the value of τ_{ij} is also adjusted by some threshold constant C_τ, which is obtained by optimal selections.

Remark 3.2 *If no factor is observed, one may estimate the volatility matrix starting from Section 3.2. Then, the volatility matrix is obtained by a linear transformation of univariate volatilities of several weak/strong latent factors plus a ePOET noise covariance matrix. It is feasible to start the estimatin from Section 3.2 even if some observable factors are known. However, based on our preliminary numerical studies, including observable factors will significantly enhance the robustness of estimation.*

Remark 3.3 *One may also consider a time-varying noise covariance matrix. For instance, we can estimate the noise volatility matrix based on DCC-NL method proposed in Engle et al. (2019), where the unconditional covariance matrix of DCC is achieved according to the non-linear shrinkage technique developed in Ledoit and Wolf (2012). However, the DCC-NL method performs unstable in some cases and the underlying intuition in economics data is not clear.*

4 Choice of the threshold constant

In practice, we have to select the tuning constant C_τ in the threshold level $\tau_{ij} = C_\tau \omega_{NT}(\theta_{ij})^{1/2}$. Following the procedures of Fan et al. (2013), we use cross-validation to select C_τ. Repeat the following Step 1-2 H times and denote each time as h.

Step 1. Obtain the residuals $\{\hat{\mathbf{e}}_t\}_{t=1}^T$ based on our two-step regressions.

Step 2. Randomly divide $\{\hat{\mathbf{e}}_t\}_{t=1}^T$ into two sets, denoted by M_1 and M_2. Let M_1 be the training set $\{\hat{\mathbf{e}}_t\}_{t\in M_1}$, and M_2 be the validation set $\{\hat{\mathbf{e}}_t\}_{t\in M_2}$. The training set has size $T(M_1)$ and the validation set has size $T(M_2)$, where we simply set $T(M_1) = \lfloor T(1-\log^{-1} T)\rfloor$ and $T(M_1) + T(M_2) = T$.

Step 3. Obtain the optimal tuning parameters C_τ^* according to the Frobenius loss:

$$C_\tau^* = \mathop{\arg\min}_{C_\tau \in [C^{\min},\ \bar{C}]} \frac{1}{H} \sum_{h=1}^{H} \left\| \hat{\boldsymbol{\Sigma}}_e^\tau(C_\tau)^{M_1,h} - \hat{\boldsymbol{\Sigma}}_e^{M_2,h} \right\|_{\mathrm{F}}^2 .$$

For each time h, $\hat{\mathbf{\Sigma}}_e^{\tau}(C_\tau)^{M_1,h}$ is the ePOET estimator of $\{\hat{\mathbf{e}}_t\}_{t\in M_1}^h$ with the threshold constant C_τ. $\hat{\mathbf{\Sigma}}_e^{M_2,h}$ is the sample covariance matrix of $\{\hat{\mathbf{e}}_t\}_{t\in M_2}^h$. Note that C_τ^* is selected from a interval $[C^{\min},\ \bar{C}]$, where $C^{\min}$ is the minimum constant to ensure the positive definiteness and $\bar{C}$ is some large constant set by users.

5 Monte Carlo simulations

In this section, we investigate the finite sample behaviour of DCC-ePOET through Monte Carlo experiments.

5.1 Design 1: data are generated from observabale factors plus latent factors

In this experiment, we consider a data generating process (DGP) from the DCC-ePOET model in Section 2. Recall the linear factor model we proposed:

$$\mathbf{y}_t = \mathbf{A}^0\mathbf{x}_t + \mathbf{B}^0\mathbf{f}_t^0 + \mathbf{e}_t.$$

We consider $\mathbf{x}_t$ follows a standard DCC-GARCH process:

$$\mathbf{x}_t = \mathbf{\Sigma}_x(t)^{1/2}\boldsymbol{\eta}_t, \quad \mathbf{\Sigma}_x(t) = \mathbf{D}\mathbf{R}_t\mathbf{D},$$

where

$$\begin{aligned}
\boldsymbol{\eta}_t &\sim \text{i.i.d. } N(0, \mathbf{I}_r),\\
\mathbf{D}_t &= \operatorname{diag}\left(\sqrt{h_{t1}^x}, ..., \sqrt{h_{tr}^x}\right),\\
\mathbf{R}_t &= \operatorname{diag}\mathbf{Q}_t{}^{-1/2}\mathbf{Q}_t \operatorname{diag}\mathbf{Q}_t{}^{-1/2},\\
\mathbf{Q}_t &= (1-a-b)\mathbf{S} + a\boldsymbol{\eta}_{t-1}^*\boldsymbol{\eta}_{t-1}^{*'} + b\mathbf{Q}_{t-1},
\end{aligned}$$

with $\sqrt{h_{tl}^x}$ following GARCH (1,1) for each $l \leqslant r$, $\boldsymbol{\eta}_t^* = \left(x_{t1}/\sqrt{h_{t1}^x}, ..., x_{tr}/\sqrt{h_{tr}^x}\right)$ and $\mathbf{S}$ is a positive-definite unconditional covariance matrix. In this experiment, we assume the true $\mathbf{S}$ is the sample covariance matrix of Fama-French 3 factors collecting from April 2, 2002 to December 29, 2017. Namely, $r = 3$. We set the univariate GARCH(1,1) parameters $\omega = 0.02$, $\alpha = 0.25$ and $\beta = 0.7$. The DCC parameters are designed to be $a = 0.1$ and $b = 0.85$. The factor loading a_{il}^0 is generated from i.i.d. $N(0, 1)$.

For the latent factors part, we follow a similar DGP process in Uematsu and Yamagata (2022). The factors f_{tk} and the corresponding factor loadings b_{ik} satisfy the conditions $T^{-1}\sum_{t=1}^{T} f_{tk}^{0} f_{ts}^{0} = 1\{s = k\}$ and $N^{-1}\sum_{i=1}^{N} b_{ik}^{0} b_{is}^{0} = 1\{s = k\}$. These conditions can be realized by the Gram-Schmidt orthogonalisation to f_{tk}^{*} and b_{ik}^{*}. Here, $b_{ik}^{*} \sim$ i.i.d.$N(0,1)$ for $i = 1, \ldots, N_k$ and $b_{ik}^{*} = 0$ for $N_{k+1}, \ldots, N$. $f_{tk}^{*} = \rho_{fk} f_{t-1,k}^{*} + v_{tk}$ with $|\rho_{fk}| < 1$, $f_{t-1,k}^{*} \sim$ i.i.d.$N(0,1)$, and $v_{tk} \sim$ i.i.d.$N(0, 1-\rho_{fk}^{2})$. We assume the number of weak latent factors $K = 2$ and the factor strength is $(0.8, 0.8)$. The factors follows the identical univariate GARCH(1,1) process with parameters $\omega = 0.02,\ \alpha = 0.25$ and $\beta = 0.7$.

For the idiosyncratic term, $\mathbf{e}_t = (e_{ti})_{N\times 1}$, its covariance matrix $\mathbf{\Sigma}_e$ is assumed to be block-diagonal. Each block has 10×10 non-zero elements, with the diagonal randomly generated from i.i.d. Unif(0.25,0.5). We fixed the within-block correlation to be 0.2. $\mathbf{e}_t$ is also assumed to be (weakly) serially dependent following AR(1) process:

$$\mathbf{e}_t = \phi_e \mathbf{e}_{t-1} + (1-\phi_e^2)^{1/2} \boldsymbol{\epsilon}_t, \tag{5.14}$$

where $\boldsymbol{\epsilon}_t$ for $t = 1, ..., T$ are generated independently from $N(\mathbf{0}, \mathbf{\Sigma}_\epsilon)$. Note that the covariance matrix of ϵ, $\mathbf{\Sigma}_\epsilon$, is identical to $\mathbf{\Sigma}_e$. We set the AR(1) parameter $\phi_e = 0.2$.

5.2 Design 2: data are generated from observable factors and only partial factors are known

In this case, we do not consider the latent factors in the DGP. Suppose

$$\mathbf{y}_t = \mathbf{A}^0 \mathbf{x}_t + \mathbf{e}_t,$$

where the observable factors $\mathbf{x}_t$ follows DCC-GARCH(1,1) with the same DCC and GARCH parameters as in Design 1. We further set the number of factors $\mathbf{x}_t$ is five. The true unconditional covariance matrix $\mathbf{S}$ is the sample covariance matrix of Fama-French five factors spanning from April 2, 2002 to December 29, 2017[1]. We keep the DGP process of error term $\mathbf{e}_t$ in Design 1. Most importantly, we assume that only first three factors are observable in this experiment, and we estimate the number of latent factors according to the selection method proposed in Onatski (2010).

[1]See https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

5.3 Results

We use the following standard statistical losses to evaluate the estimation accuracy.

(a) Relative error: $\|\hat{\mathbf{\Sigma}}(t) - \mathbf{\Sigma}(t)\|_{\mathbf{\Sigma}} = \|\mathbf{\Sigma}(t)^{-1/2}\hat{\mathbf{\Sigma}}(t)\mathbf{\Sigma}(t)^{-1/2} - \mathbf{I}_N\|_{\mathrm{F}}$.

(b) Maximum norm loss: $\|\hat{\mathbf{\Sigma}}(t) - \mathbf{\Sigma}(t)\|_{\max}$.

(c) Spectral loss of the inverse: $\|\hat{\mathbf{\Sigma}}(t)^{-1} - \mathbf{\Sigma}(t)^{-1}\|_2$.

Loss (a) is a relative error measure for covariance matrix estimation, which is initially proposed by Fan et al. (2011). (b) and (c) are usual criteria. We compare our proposed method with the CCC-POET approach recently developed in Li et al. (2022), where they consider the volatility matrix is composed of observable factors equipped with CCC-GARCH and POET noise covariance matrix. The simulation results are stored in Table 1 and 2.

Table 1: Performance of the DCC-ePOET and CCC-POET estimates when data generated from observable factors and latent weak factors.

T = 100	**Methods**		$\|\hat{\mathbf{\Sigma}}(t) - \mathbf{\Sigma}(t)\|_{\mathbf{\Sigma}}$	$\|\hat{\mathbf{\Sigma}}(t) - \mathbf{\Sigma}(t)\|_{\max}$	$\|\hat{\mathbf{\Sigma}}(t)^{-1} - \mathbf{\Sigma}(t)^{-1}\|_2$
N = 50	**DCC-ePOET**	**mean**	0.890	4.679	3.722
		median	0.840	3.622	3.630
	CCC-POET	**mean**	1.233	4.886	52.706
		median	1.118	3.637	28.475
N = 100	**DCC-ePOET**	**mean**	0.984	5.458	3.378
		median	0.904	4.532	3.240
	CCC-POET	**mean**	1.587	5.843	2380.258
		median	1.454	4.795	568.148
N = 300	**DCC-ePOET**	**mean**	1.256	6.708	3.004
		median	1.171	5.556	2.917
	CCC-POET	**mean**	4.225	7.989	221.062
		median	3.724	6.426	123.592

Our simulation results show that (i) The proposed DCC-ePOET method generally works well whether the observable factors are fully or partially known. The DCC-ePOET estimators are consistent under any statistical losses, whatever $N < T$ and $N \geqslant T$. (ii) In any case, our approach has beaten the recently developed CCC-POET method in Li et al.

Table 2: Performance of the DCC-ePOET and CCC-POET estimates when only partial factors are observable.

T = 100	**Methods**		$\Vert\hat{\mathbf{\Sigma}}(t) - \mathbf{\Sigma}(t)\Vert_{\mathbf{\Sigma}}$	$\Vert\hat{\mathbf{\Sigma}}(t) - \mathbf{\Sigma}(t)\Vert_{\max}$	$\Vert\hat{\mathbf{\Sigma}}(t)^{-1} - \mathbf{\Sigma}(t)^{-1}\Vert_2$
N = 50	**DCC-ePOET**	mean	0.706	3.431	3.161
		median	0.682	2.903	2.985
	CCC-POET	mean	0.985	4.251	23.764
		median	0.933	3.391	10.212
N = 100	**DCC-ePOET**	mean	0.827	4.503	2.898
		median	0.773	3.637	2.782
	CCC-POET	mean	1.262	5.277	1144.294
		median	1.215	4.218	169.251
N = 300	**DCC-ePOET**	mean	1.048	5.193	2.788
		median	0.983	4.482	2.740
	CCC-POET	mean	2.786	7.065	358.619
		median	2.434	5.623	214.293

(2022). It is clear that the superiority of our method is enhanced as N increases. CCC-POET performs acceptably in the relative error and maximum norm. However, it collapses in terms of precision (inverse of the covariance matrix) estimator mainly because it cannot capture the dynamic structure in the observable factors and sufficient correlation information in the residuals. In the following section, we will show how the collapse of the precision estimator deteriorates the performance in a real dataset.

6 Empirical studies

To evaluate the goodness of the estimation of volatility matrix in practice, we carry out minimum variance portfolio (MVP) analysis of S&P 500 data based on several multivariate GARCH approaches and ePOET estimate. We follow the similar strategy and the same dataset in Dai et al. (2022), where they aim to evaluate the performance of static covariance matrix estimates. The goal of MVP is to allocate N financial assets to make portfolio risk $\mathbf{w}'\tilde{\mathbf{\Sigma}}_t\mathbf{w}$ as low as possible. Here, $\mathbf{w}$ is a vector of weights and $\tilde{\mathbf{\Sigma}}$ is a volatility matrix estimate of the given assets at time t. In detail, the MVP solves the following optimisation problem:

$$\min_{\mathbf{w}} \mathbf{w}'\tilde{\boldsymbol{\Sigma}}_t\mathbf{w} \ \text{ subject to } \ \mathbf{w}'\mathbf{1}_N = 1, \tag{6.15}$$

where $\mathbf{1}_N = (1, \dots, 1)'$. We allow short sales and ignore any transaction cost for simplicity. The optimal weight $\mathbf{w}^*$ is obtained by the quadratic problem (6.15) and the corresponding risk $\mathbf{R}_t^*$ are computed as

$$\mathbf{w}^* = \frac{\tilde{\boldsymbol{\Sigma}}_t^{-1}\mathbf{1}_N}{\mathbf{1}_N{}'\tilde{\boldsymbol{\Sigma}}_t^{-1}\mathbf{1}_N}, \qquad \mathbf{R}_t^* = \mathbf{w}^{*\prime}\tilde{\boldsymbol{\Sigma}}_t\mathbf{w}^*. \tag{6.16}$$

We obtained the S&P 500 data from Yahoo Finance, comprising 2520 daily excess returns spanning from April 2, 2002, to March 30, 2012 with full information. This period covers approximately 10 years of trading, with an average of 21 trading days per month. Alongside the time span of the S&P 500 data, we also collected the Fama-French five factors data from Kenneth R. French-Data Library.

The trading strategy is: on the first trading day of each month, we constructed an optimal portfolio using a candidate covariance estimate on the historical data from the preceding T days. We set the time dimension T to be 126, about six months of trading days. Under a rolling window scheme, the vector of optimal portfolio weights ($\mathbf{w}_t^*$) is updated monthly for constructing next month's portfolios until March 30, 2012. The cross-sectional dimension N is fixed at 395, the maximum number of stocks available in the dataset. Once obtaining all the out-of-sample portfolios, we calculate the out-of-sample variance (**Var**), the total out-of-sample excess returns (**TR**) and the mean Sharpe ratio (**SR**) according to the formulas in DeMiguel et al. (2009) and Lam (2016) as follows.

$$\begin{aligned}
\mathbf{TR} &= \sum_{i=6}^{120}\sum_{t=21i+1}^{21i+21} \mathbf{w}_i^{*\prime}\mathbf{R}_t{}^*, \quad \mathbf{TR}_i = \sum_{t=21i+1}^{21i+21} \mathbf{w}_i^{*\prime}\mathbf{R}_t^*, \\
\mathbf{Var} &= \frac{1}{2520}\sum_{i=6}^{120}\sum_{t=21i+1}^{21i+21} (\mathbf{w}_i^{*\prime}\mathbf{R}_t^* - \mathbf{TR}_i)^2, \quad \mathbf{Var}_i = \frac{1}{21}\sum_{t=21i+1}^{21i+21} (\mathbf{w}_i^{*\prime}\mathbf{R}_t^* - \mathbf{TR}_i/21)^2, \\
\mathbf{SR} &= \frac{1}{114}\sum_{i=6}^{120} \frac{\mathbf{TR}_i}{\mathbf{Var}_i}.
\end{aligned}$$

We compare the out-of-sample forecasting performance of following 7 candidate methods.

(i) **ePOET**: A static covariance matrix proposed in Dai et al. (2022) that the covariance matrix consists of observable factors, latent factors and the sparse noise covariance matrix and this estimator is the winner in that article.

(ii) **DCC-ePOET-1**: Our proposed estimator;

(iii) **DCC-ePOET-2**: This method is similar to DCC-ePOET-1, and the latent factors are also allowed to be DCC-GARCH;

(iv) **ePOET-GARCH**: This approach is similar to DCC-ePOET-1, but does not impose the DCC structure on any factor;

(v) **DCC-Factor-GARCH**: This method is DCC-ePOET-1 without latent factors, inspired from Zhang and Chan (2009) but including sparse noise covariance matrix to tackle the singularity issues arising from large N;

(vi) **CCC-POET**: This method is proposed by Li et al. (2022), where the volatility matrix is composed of observable factors with CCC-GARCH structure plus the sparse noise covariance matrix estimate according to Fan et al. (2011).

(vii) **PC-GARCH**: This one applies univariate GARCH on each latent factor in the POET model by Fan et al. (2013).

Note that all the DCC-GARCH estimations are implemented by the R package, `xdcclarge`[2]. We employ GARCH(1,1) in each univariate volatility estimation as suggested in Hansen and Lunde (2005). GARCH(1,1) is typically adequate for capturing the clustered nature of volatilities observed in the data for most scenarios. In each window, the number of latent factors is selected according to the method of Onatski (2010) and the optimal threshold tuning parameters is determined by CV.

Table 3 reports the out-of-sample MVP performance of all the candidate methods. The most important criterion is the out-of-sample variance because our primary focus is to make the risk as low as possible. The total return and mean Sharpe ratios should be placed as the secondary criteria in our evaluation. Table 3 shows that (i) As expected, considering dynamic structure (DCC) in the observable factors significantly reduces the out-of-sample risk while imposing dynamic structure on latent factors has no effect on risk-reduction and only brings some tiny improvements in TR and SR. (ii) For DCC-Factor-GARCH and CCC-POET, their performance collapsed in any aspect. This is not a surprising result

[2]See https://cran.rproject.org/web/packages/xdcclarge/index.html

Table 3: Performance of 7 methods in out-of-sample minimum variance portfolio analysis.

Criteria	ePOET	DCC-ePOET-1	DCC-ePOET-2	ePOET-GARCH
Var	0.448	**0.438**	**0.438**	0.448
TR	39.89	42.272	42.416	**43.333**
SR	1.362	1.375	1.379	**1.398**

Criteria	DCC-Factor-GARCH	CCC-POET	PC-GARCH
Var	0.803	0.805	0.462
TR	1.976	1.883	36.871
SR	1.199	1.200	1.366

because omitting latent factors in the residuals will lead to unstable precision estimations, as we also discussed in the simulation studies. (iii) Finally, even though ePOET still works acceptably among the competitors, replacing the static covariance matrix with the ePOET-based dynamic covariance estimates significantly and robustly improves the out-of-sample performance in every criterion. Generally, our proposed method works very well in every aspect.

7 Conclusion remarks

This paper proposes a new type of Factor-GARCH model, DCC-ePOET, to estimate large volatility matrices, especially addressing the issues of omitted variables and singularity in Factor-GARCH families under the high-dimensional space. We capture the dynamic structure in the factors by embedding DCC-GARCH into observable factors. We further capture the covariance information in the residuals by a Factor-GARCH using sparsity-induced weak factors plus a sparse noise unconditional covariance obtained from ePOET techniques. Monte Carlo simulations demonstrate that the finite sample performance of the proposed estimator is quite satisfactory even if the observable factors are just partially known. The out-of-sample MVP analysis confirms that our proposed DCC-ePOET estimator works robustly and generally outperforms the other candidate models.

There are three potential extensions for future studies. (i) As our model uses the simplest ARMA(0,0)-GARCH model in the signal part, we may consider more extensions in univariate GARCH models, such as A-GARCH, T-GARCH, and E-GARCH for some specific scenarios. (ii) We assume the factor loadings are time-invariant, while it is natural to allow the loadings of observable factors to be time-varying for more flexibility. For instance, many researchers have pointed out that structure breaks in factor models are reflected through the changes in loadings. The time-varying loading for sWF factors is still an open question for future research. (iii) Finally, we may incorporate the QMLE of GARCH Models with heavy-tailed likelihoods proposed in Fan et al. (2014) to enhance the efficiency when the factor innovation distribution exhibits heavy tails.

References

Ahn, S. C. and A. R. Horenstein (2013). Eigenvalue ratio test for the number of factors. *Econometrica 81*(3), 1203–1227.

Aielli, G. P. (2013). Dynamic conditional correlation: on properties and estimation. *Journal of Business & Economic Statistics 31*(3), 282–299.

Alexander, C. (2000). A primer on the orthogonal garch model. Manuscript ISMA Centre, University of Reading, UK, 2.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica 77*(4), 1229–1279.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*(1), 191–221.

Barigozzi, M. and M. Hallin (2017). Generalized dynamic factor models and volatilities: Estimation and forecasting. *Journal of Econometrics 201*(2), 307–321.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics 31*(3), 307–327.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized arch model. *The Review of Economics and Statistics*, 498–505.

Bollerslev, T., R. F. Engle, and J. M. Wooldridge (1988). A capital asset pricing model with time-varying covariances. *Journal of Political Economy 96*(1), 116–131.

Cai, T. and W. Liu (2011). Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association 106*(494), 672–684.

Chen, J., D. Li, and O. Linton (2019). A new semiparametric estimation approach for large dynamic covariance matrices with multiple conditioning variables. *Journal of Econometrics 212*(1), 155–176.

Dai, R., Y. Uematsu, Y. Matsuda, et al. (2022). Estimation of large covariance matrices with mixed factor structures. Technical report, Graduate School of Economics and Management, Tohoku University.

DeMiguel, V., L. Garlappi, F. J. Nogales, and R. Uppal (2009). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management science 55*(5), 798–812.

Dendramis, Y., L. Giraitis, and G. Kapetanios (2021). Estimation of time-varying covariance matrices for large datasets. *Econometric Theory 37*(6), 1100–1134.

Ding, Z. (1994). Time series analysis of speculative returns. University of California, San Diego, Department of Economics.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics 20*(3), 339–350.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the econometric society*, 987–1007.

Engle, R. F., C. W. Granger, and D. Kraft (1984). Combining competing forecasts of inflation using a bivariate arch model. *Journal of Economic Dynamics and Control 8*(2), 151–165.

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

Engle, R. F., O. Ledoit, and M. Wolf (2019). Large dynamic covariance matrices. *Journal of Business & Economic Statistics 37*(2), 363–375.

Fan, J., Y. Fan, and J. Lv (2008). High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics 147*(1), 186–197.

Fan, J., Y. Liao, and M. Mincheva (2011). High dimensional covariance matrix estimation in approximate factor models. *Annals of Statistics 39*(6), 3320.

Fan, J., Y. Liao, and M. Mincheva (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society. Series B, Statistical Methodology 75*(4).

Fan, J., L. Qi, and D. Xiu (2014). Quasi-maximum likelihood estimation of garch models with heavy-tailed likelihoods. *Journal of Business & Economic Statistics 32*(2), 178–191.

Feng, G., S. Giglio, and D. Xiu (2020). Taming the factor zoo: A test of new factors. *The Journal of Finance 75*(3), 1327–1370.

Giglio, S. and D. Xiu (2021). Asset pricing with omitted factors. *Journal of Political Economy 129*(7), 1947–1990.

Guo, S., J. L. Box, and W. Zhang (2017). A dynamic structure for high-dimensional covariance matrices and its application in portfolio allocation. *Journal of the American Statistical Association 112*(517), 235–253.

Hafner, C. M. and A. Preminger (2009). Asymptotic theory for a factor garch model. *Econometric Theory 25*(2), 336–363.

Hansen, P. R. and A. Lunde (2005). A forecast comparison of volatility models: Does anything beat a garch (1, 1)? *Journal of Applied Econometrics 20*(7), 873–889.

Ke, Y., H. Lian, and W. Zhang (2022). High-dimensional dynamic covariance matrices with homogeneous structure. *Journal of Business & Economic Statistics 40*(1), 96–110.

Lam, C. (2016). Nonparametric eigenvalue-regularized precision or covariance matrix estimator. *Annals of Statistics 44*(3), 928–953.

Ledoit, O. and M. Wolf (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices.

Li, D. (2021). Estimation of large dynamic covariance matrices: A selective review. *Econometrics and Statistics*.

Li, X., X. Zhang, and Y. Li (2022). High-dimensional conditional covariance matrices estimation using a factor-garch model. *Symmetry 14*(1), 158.

Li, X.-l., Y. Li, J.-z. Pan, and X.-f. Zhang (2022). A factor-garch model for high dimensional volatilities. *Acta Mathematicae Applicatae Sinica, English Series 38*(3), 635–663.

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics 92*(4), 1004–1016.

Santos, A. A. and G. V. Moura (2014). Dynamic factor multivariate garch model. *Computational Statistics & Data Analysis 76*, 606–617.

Shi, F., L. Shu, and X. Gu (2022). An enhanced factor model for portfolio selection in high dimensions. *Journal of Financial Econometrics*.

Uematsu, Y., Y. Fan, K. Chen, J. Lv, and W. Lin (2019). SOFAR: large-scale association network learning. *IEEE transactions on information theory 65*(8), 4924–4939.

Uematsu, Y. and T. Yamagata (2022). Estimation of sparsity-induced weak factor models. *Journal of Business & Economic Statistics 41*(1), 213–227.

Van der Weide, R. (2002). Go-garch: A multivariate generalized orthogonal garch model. *Journal of Applied Econometrics 17*(5), 549–564.

Vrontos, I. D., P. Dellaportas, and D. N. Politis (2003). A full-factor multivariate garch model. *The Econometrics Journal 6*(2), 312–334.

Wang, H., B. Peng, D. Li, and C. Leng (2021). Nonparametric estimation of large covariance matrices with conditional sparsity. *Journal of Econometrics 223*(1), 53–72.

Zhang, K. and L. Chan (2009). Efficient factor garch models and factor-dcc models. *Quantitative Finance 9*(1), 71–91.