Modelling Multivariate Volatilities via Latent Common Factors

Weiming Li

International School of Economics and Management, Capital University of Economics and Business, Beijing, 100070, China

Jing Gao

International School of Economics and Management, Capital University of Economics and Business, Beijing, 100070, China

Kunpeng Li

International School of Economics and Management, Capital University of Economics and Business, Beijing, 100070, China

Qiwei Yao

Department of Statistics, London School of Economics and Political Science, UK.

Abstract

Volatility, represented in the form of conditional heteroscedasticity, plays an important role in controlling and forecasting risks in various financial operations including asset pricing, portfolio allocation, and hedging futures. However, modeling and forecasting multi-dimensional conditional heteroscedasticity are technically challenging. As the volatilities of many financial assets are often driven by a few common and latent factors, we propose in this paper a dimension reduction method to model a multivariate volatility process and to estimate a lower-dimensional space, to be called the volatility space, within which the dynamics of the multivariate volatility process is confined. The new method is simple to use, as technically it boils down to an eigenanalysis for a non-negative definite matrix. Hence it is applicable to the cases when the number of assets

concerned is in the order of thousands (using an ordinary PC/laptop). On the other hand, the model has the capability to cater for complex conditional heteroscedasticity behavior for multi-dimensional processes. Some asymptotic properties for the new method are established. We further illustrate the new method using both simulated and real data examples.

Keywords: Eigenanalysis, Latent factors, Multi-dimensional volatility process, Volatility space.

1 Introduction

Volatility is a measure for the uncertainty and the risk of asset returns. It is usually defined as the conditional standard deviation of an asset return given all the available information up to the present time. The importance of correctly specifying and forecasting volatilities is reflected in almost every facet of finance. It is related to option pricing (e.g. Black-Scholes formula), risk measures (e.g. value-at-risk), risk-adjusted return (e.g. Sharpe ratio), securities regulations (e.g. capital requirement under Basel III), portfolio allocation and hedging. While a large number of statistical models as well as the associated inference methods and theory have been developed for modelling and forecasting univariate volatilities, almost all real financial applications require to specify the volatilities for multiple assets jointly. This calls for the modelling of conditional variance-covariance matrix processes. Though there is little conceptual difficulty in extending most univariate volatility models to multivariate cases, the inference and the implementation for those models are challenging. Those models are typically over-parameterized; leading to flat likelihoods which cause innate difficulties in inference. Also the high dimensionality of the volatility process (i.e. in the order of N^2 for N assets) causes operational difficulties in implementing inference methods. Therefore some extremely simple models (with a very small number of parameters) are proposed at the early stage of the development in order to make the inference feasible and the models interpretable. See the surveys in Wang and Yao (2005), Bauwens, Laurent, and Rombouts (2006), Asai, McAleer and Yu (2006), Silvemmponen and Terasvirta (2008) and the references within.

The phrase 'markets move together' could be taken as a folklore in finance. It partially reflects the fact that the prices and their volatilities across different assets are often driven by a few common factors. The celebrated CAPM and Fama-French models are cases in point. In this paper, we propose a simple method to identify the latent common factors which drive the volatilities of multiple or even a large number of assets. By assuming that those latent factors are linear combinations of the observed returns, the task is to estimate the factor loadings (on the returns). Technically our estimation method boils down to an eigenanalysis for a non-negative definite matrix. Hence it is applicable for modelling

volatility processes for thousands of assets using an ordinary PC or laptop. The method itself is generic in the sense that we do not impose any explicit parametric forms on the volatility process; see Section 2.1 below.

Our approach can be viewed as a version of volatility factor model. However it differs from the factor model of Engle and Kroner (1995) which assumes that the factors are known and observable. See also Tao et al (2011) which deals factor volatility models with some initial estimators obtained from using high-frequency data. The latent factors in our approach can also be viewed as a generalized version of principal volatility components (PVC) of Hu and Tsay (2014), although our version is only characterized by the second moments properties instead of (truncated) fourth moments adopted in Hu and Tsay (2014). Unlike Hu and Tsay (2014), we do not assume that the underlying volatility process is a vector ARCH(∞) process. Though our approach is based on the same idea as Pan et al. (2010), our implementation is radically different. While the innovation expansion method proposed in Pan et al. (2010) requires to solve a sequence of complex nonlinear optimization problems, our method boils down to a single eigenanalysis for a non-negative definite matrix, and therefore is applicable when the dimension of time series is large or much larger.

The rest of the paper is organized as follows. In Section 2, we introduce our setting and the eigenanalysis-based estimation method. The asymptotic properties of the proposed method is presented in Section 3. We illustrate the proposed method by simulation in Section 4 which also contains some numerical comparisons with the PVC method of Hu and Tsay (2014). Further illustration using two real data sets is presented in Section 5.

2 Methodology

2.1 Basic setting

Let $y_t = (y_{1t}, y_{2t}, \ldots, y_{Nt})'$ be an N-dimensional strictly stationary time series with finite first two moments. Let $\mathscr{F}_t = \sigma(y_t, y_{t-1}, \ldots,)$ denote the σ-algebra generated by $y_t, y_{t-1}, ...$ (i.e., the information available up to the time t). For expositional simplicity, we assume

$E(y_t|\mathscr{F}_{t-1}) = 0$. Our objective is to model the conditional variance and covariance process

$$\Sigma_y(t) = \text{var}(y_t|\mathcal{F}_{t-1}) = E(y_t y_t'|\mathcal{F}_{t-1}), \quad t = 0, \pm 1, \pm 2, \cdots .$$

To simplify the matter concerned, we assume that $\text{var}(y_t) = I_N$, where I_N is the $N \times N$ identity matrix. In practice, this amounts to replacing y_t by $S^{-1/2}y_t$, where S is the sample covariance matrix of y_t, and is assumed to be invertible.

Let

$$\mathcal{M}_0 = \{\, a \in R^N \;\big|\; \text{var}(a'y_t|\mathcal{F}_{t-1}) \neq \text{var}(a'y_t) \,\}.$$

Then $\mathcal{M}_0$ consists of all the directions in which y_t exhibits conditional heteroscedasticity. Let $\mathcal{M}_1$ be the linear space spanned by $\mathcal{M}_0$, which is called the volatility space of y_t. We assume that the dimension of $\mathcal{M}_1$ is an positive integer r smaller than N. Let A be an $N \times r$ matrix whose columns form an orthonormal basis of $\mathcal{M}_1$. Then such an A always exists but not uniquely. Furthermore, let B be an $N \times (N - r)$ matrix such that (A, B) forms an $N \times N$ orthogonal matrix. Now y_t can be formally written as

$$y_t = (AA' + BB')y_t = Ax_t + \varepsilon_t, \tag{2.1}$$

where $x_t = A'y_t$ can be viewed as an r-variate latent factor process which drives the conditional heteroscedasticity of y_t, and $\varepsilon_t = BB'y_t$ exhibits no conditional heteroscedasticity as $\text{var}(\varepsilon_t|\mathcal{F}_{t-1}) \equiv \text{var}(\varepsilon_t)$. It follows from (2.1) and Lemma 1 below that

$$\Sigma_y(t) = A\Sigma_x(t)A' + \Sigma_\varepsilon, \tag{2.2}$$

where $\Sigma_x(t) = \text{var}(x_t|\mathcal{F}_{t-1})$ and $\Sigma_\varepsilon = \text{var}(\varepsilon_t)$. This is the standard form of volatility factor models; see, e.g. Engle and Kroner (1995), though the factor process x_t is unobservable now.

When $r > 0$ but small, the conditional heteroscedasticity of y_t is confined to an r-dimensional volatility space $\mathcal{M}_1$. While $\mathcal{M}_1$ is uniquely defined by y_t, (A, x_t) in (2.1) and (2.2) are not. In fact they can be replaced by $(AH, H'x_t)$ for any $r \times r$ orthogonal matrix

H.

Lemma 1 *Under the assumption* $\text{var}(y_t) = I_N$, $\text{cov}(x_t, \varepsilon_t|\mathcal{F}_{t-1}) = 0$.

Proof. Let $V = I_N - \Sigma_y(t)$. As $c'Vc = \text{var}(c'y_t) - \text{var}(c'y_t|\mathcal{F}_{t-1}) \geq 0$ for any $c \in R^N$, V is a non-negative definite matrix. By the definition of ε_t, $B'VB = 0$. Therefore $V^{1/2}B = 0$, and consequently

$$0 = VB = \{I_N - \Sigma_y(t)\}B. \tag{2.3}$$

Now

$$\text{cov}(x_t, \varepsilon_t|\mathcal{F}_{t-1}) = A'\Sigma_y(t)BB' = A'I_NBB' = A'BB' = 0.$$

2.2 Estimation

Since the factor loading matrix A in (2.1) is not unique, any A can be used as long as the linear space spanned by its columns is equal to the volatility space $\mathcal{M}_1$. Note that $\mathcal{M}_1$ is also uniquely determined by B characterized in equation (2.3). By Theorem 7.1.1 of Chow and Teicher (1997), (2.3) is equivalent to

$$E\{(y_ty_t' - I_N)I(W)\}B = 0 \quad \text{for all } \ W \in \mathcal{B}_t, \tag{2.4}$$

where $\mathcal{B}_t$ is any π-class such that the σ-algebra generated by $\mathcal{B}_t$ is $\mathcal{F}_{t-1}$, and $I(\cdot)$ is the indicator function. Also note that (2.4) can be equivalently expressed as $MB = 0$, where

$$M = \sum_{W \in \mathcal{B}_t} \big[E\{(y_ty_t' - I_N)I(W)\}\big]^2.$$

In practice we replace M by

$$M_1 = \sum_{k=1}^{m} \sum_{W \in \mathcal{B}} w(W)\big[E\{(y_ty_t' - I_N)I(y_{t-k} \in W)\}\big]^2, \tag{2.5}$$

where m is a prescribed positive integer, $w(\cdot)$ is a weight function, and

$$\mathcal{B} = \big\{(u \in R^N : \|u\| \leq \|y_t\|), \ t = 1, 2, 3...T\big\}. \tag{2.6}$$

See also Fan, Wang and Yao (2008). Note that M_1 is a non-negative definite matrix, and $M_1 B = 0$ implies that the columns of B are the eigenvectors of M_1 corresponding to the eigenvalue 0. Therefore the columns of A are the eigenvectors of M_1 corresponding to non-zero eigenvalues. Furthermore the rank of matrix M_1 is the number of latent factors r. Thus we can directly estimate A.

The matrix M_1 facilitates a natural estimator

$$\widehat{M_1} = \sum_{k=1}^{m} \sum_{W \in \mathcal{B}} w(W) \big[\frac{1}{T-k} \sum_{t=k+1}^{T} \{(y_t y_t' - I_N) I(y_{t-k} \in W)\} \big]^2. \tag{2.7}$$

Perform the eigenanalysis for $\widehat{M_1}$, and let $\widehat{\lambda}_1 \geq \cdots \geq \widehat{\lambda}_N$ be its eigenvalues, and $\widehat{\gamma}_1, \cdots, \widehat{\gamma}_N$ be the corresponding eigenvectors. Then $\widehat{A} = (\widehat{\gamma}_1, \cdots, \widehat{\gamma}_r)$ is an estimated factor loading matrix. Denoted by $\widehat{\mathcal{M}}_1$ the linear space spanned by the columns of $\widehat{A}$. Then $\widehat{\mathcal{M}}_1$ is an estimator for the volatility space $\mathcal{M}_1$. Furthermore $\widehat{x}_t = \widehat{A}' y_t$ is an estimated factor process.

In principle one might like to use a large m in (2.7). However in practice a small m is often sufficient, as the information carried in the first a few lags is enough to determine the volatility space. Furthermore our method is not so sensitive to the choice of m, as all the terms on the RHS of (2.7) are non-negative definite matrices. There is no information cancellation among different lags.

In practice we also need to estimate the number of factors r. We adopt the ratio estimator

$$\widehat{r} = \max_{1 \leq j \leq N-1} \widehat{\lambda}_j / \widehat{\lambda}_{j+1}. \tag{2.8}$$

See Lam and Yao (2012) for this method in the context of the factor modelling for means.

3 Asymptotic properties

We provide some asymptotic results when $T \to \infty$ and N fixed. Some regularity conditions are now in order.

1. The process $\{y_t\}$ is strictly stationary.

2. $\{y_t\}$ is a β-mixing process with the β mixing coefficients satisfying $\beta(\tau) = O(\tau^{-s})$ for

some $s > p$, where $p > 2$ is a constant defined in condition 3 below.

3. $E\|y_t\|^{2p} < \infty$ for some $p > 2$ and $s > p/(p-2)$, where $\|\cdot\|$ denotes the Euclidean norm.

Since both A and $\widehat{A}$ are half orthogonal matrices, i.e. $A'A = \widehat{A}'\widehat{A} = I_r$, we measure the distance between $\mathcal{M}_1$ and its estimator $\widehat{\mathcal{M}}_1$ by the measure

$$d(\widehat{\mathcal{M}}_1, \mathcal{M}_1) = \sqrt{1 - \frac{\text{trace}(\widehat{A}\widehat{A}'AA')}{r}} \tag{3.1}$$

Then $d(\widehat{\mathcal{M}}_1, \mathcal{M}_1) \in [0, 1]$, it is equal to 1 if and only if the two spaces are orthogonal with each other, and 0 if and only if the two spaces are identical.

Theorem 1 *Under conditions 1 – 3 above, it holds that*

1. $\|\widehat{M}_1 - M_1\|_2 = O_p(T^{-1/2})$,
2. $d(\widehat{\mathcal{M}}_1, \mathcal{M}_1) = O_p(T^{-1/4})$.

Proof. We introduce some notation. Let

$$C_{T,k}(W) = \frac{1}{T-k}\sum_{t=k+1}^{T}(y_t y_t' - I_N)I(y_{t-k} \in W)$$
$$C_k(W) = E\big[(y_t y_t' - I_N)I(y_{t-k} \in W)\big].$$

Then we can write M_1 and $\widehat{M}_1$ as

$$\widehat{M}_1 = \sum_{k=1}^{m}\sum_{W\in\mathcal{B}} w(W)C_{T,k}^2(W),$$
$$M_1 = \sum_{k=1}^{m}\sum_{W\in\mathcal{B}} w(W)C_k^2(W).$$

Thus,

$$\widehat{M}_1 - M_1 = \sum_{k=1}^{m}\sum_{W\in\mathcal{B}} w(W)\big[C_{T,k}^2(W) - C_k^2(W)\big]. \tag{3.2}$$

Under the conditions 1-3 given above, Theorem 1 of Arcones and Yu (1994) implies that the set of functions $\{(y_{ti}y_{tj}' - \delta_{ij})I(y_{t-k} \in W), W \in \mathcal{B}\}$ is a Donsker class, and hence

the process $\{\Delta_{T,k}(W), W \in \mathcal{B}\}$ that is indexed by $W \in \mathcal{B}$ converges weakly to a Gaussian process, where $\Delta_{T,k}(W) = T^{1/2}\{C_{T,k}(W) - C_k(W)\}$. It follows from equation (3.2) that:

$$\begin{aligned}
\widehat{M_1} - M_1 &= \sum_{k=1}^{m} \sum_{W \in \mathcal{B}} w(W)\big[C_{T,k}^2(W) - C_k^2(W)\big] \\
&= \sum_{k=1}^{m} \sum_{W \in \mathcal{B}} w(W)\big[\big(C_k(W) + T^{-1/2}\Delta_{T,k}(W)\big)^2 - C_k^2(W)\big] \\
&= \sum_{k=1}^{m} \sum_{W \in \mathcal{B}} w(W)\big[2C_k(W)T^{-1/2}\Delta_{T,k}(W) + T^{-1}\Delta_{T,k}^2(W)\big] \\
&= T^{-1/2} \sum_{k=1}^{m} \sum_{W \in \mathcal{B}} w(W)\big[2C_k(W)\Delta_{T,k}(W) + T^{-1/2}\Delta_{T,k}^2(W)\big] \\
&= T^{-1/2} O_p(1) \\
\Rightarrow \|\widehat{M_1} - M_1\|_2 &= O_p(T^{-1/2}),
\end{aligned}$$

where $\|A\|_2$ is the Euclidean norm of A.

For the proof of the second part of Theorem 1, we need to use Theorem 8.1.10 in Golub & Van Loan (1996), which is stated explicitly in Lemma 2 below. See also Johnstone & Lu (2009), Lam, Yao & Bathia (2011).

Lemma 2 *Suppose M and $M + E$ are $T \times T$ symmetric matrices and that $Q = [A, B]$, where A has size $T \times r$ and B has size $T \times (T - r)$, is an orthogonal matrix such that span(A) is an invariant subspace for M; that is, $M \times span(A) \subset A$. Partition the matrices $Q'MQ$ and $Q'EQ$ as follows:*

$$Q'MQ = \begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix},$$
$$Q'EQ = \begin{pmatrix} E_{11} & E_{21}' \\ E_{21} & E_{22} \end{pmatrix}.$$

If $sep(D_1, D_2) \stackrel{def}{=} \min_{\lambda \in \lambda(D_1), \mu \in \lambda(D_2)} |\lambda - \mu| > 0$, where $\lambda(M)$ denotes the set of eigenvalues of the matrix M, and $\|E\|_2 \leq sep(D_1, D_2)/5$, where $\|\cdot\|_2$ is the Euclidean norm of

$\cdot$, *then there exists a matrix* $P \in \mathbb{R}^{(T-r)\times r}$ *with*

$$\|P\|_2 \leq \frac{4}{sep(D_1, D_2)}\|E_{21}\|_2$$

such that the columns of $\bar{A} = (A + BP)(I + P'P)^{-1/2}$ *define an orthonormal basis for a subspace that is invariant for* $M + E$.

In our setting, $D_1 = A'M_1A = O(1)$ and $D_2 = B'M_1B = 0$. Thus $sep(D_1, D_2) = \lambda_{min}(D_1) = O(1)$, while $\|E\|_2 = \|\widehat{M}_1 - M_1\|_2 = O_p(T^{-1/2})$. Hence, for sufficient large T, we have $\|E\|_2 \leq sep(D_1, D_2)/5$. This allows us to apply Lemma 1 to conclude that there exists a matrix $P \in \mathbb{R}^{(T-r)\times r}$ such that

$$\|P\|_2 \leq \frac{4}{sep(D_1, D_2)}\|E_{21}\|_2 \leq \frac{4}{sep(D_1, D_2)}\|E\|_2,$$

and $\bar{A} = (A + BP)(I + P'P)^{-1/2}$ is an orthonormal basis for a subspace that is invariant for $\widehat{M}_1 = M_1 + E$. Then we have

$$\begin{aligned}
\|\bar{A} - A\|_2 &= \|[A\{I - (I + P'P)^{1/2}\} + BP](I + P'P)^{-1/2}\|_2 \\
&\leq \|I - (I + P'P)^{1/2}\|_2 + \|P\|_2 \\
&\leq 2\|P\|_2 \leq \frac{8}{sep(D_1, D_2)}\|E\|_2 = O_p(T^{-1/2})
\end{aligned}$$

The difference between $\bar{A}$ and A is defines as $\Delta = \bar{A} - A$ and $\|\Delta\|_2 = O_p(T^{-1/2})$.

For any estimator $\widehat{A}$ for A, which is not necessary equal to $\bar{A}$, there exist a $r \times r$ orthogonal matrix H such that $\widehat{A} = \bar{A}H$. Both of $\widehat{A}$ and $\bar{A}$ are orthonormal basis for

subspaces that are invariant for $\widehat{M}_1 = M_1 + E$. Thus,

$$\begin{aligned} d(\widehat{\mathcal{M}}_1, \mathcal{M}_1) &= \sqrt{1 - \frac{\text{trace}(\bar{A}HH'\bar{A}'AA')}{r}} \\ &= \sqrt{1 - \frac{\text{trace}(\bar{A}\bar{A}'AA')}{r}} \\ &= \sqrt{1 - \frac{\text{trace}((A+\Delta)(A+\Delta)'AA')}{r}} \\ &= \sqrt{1 - \frac{\text{trace}(AA'AA') + trace(\Delta A'AA' + A\Delta'AA' + \Delta\Delta'AA')}{r}} \\ &= \sqrt{-\frac{\text{trace}(\Delta A'AA' + A\Delta'AA' + \Delta\Delta'AA')}{r}} \\ &= \sqrt{O_p(T^{-1/2})} = O_p(T^{-1/4}). \end{aligned}$$

This completes the proof of Theorem 1.

4 Monte Carlo Simulation

In this section, we conduct simulation to examine the finite sample performance of the proposed method. We also compare it with the PVC method by Hu and Tsay (2014). For each setting, we replicate the simulation 2000 times in Examples 1 –3, and 1000 times in Example 4. We always set the weight function $w(W) = 1/T$ in (2.7) in this section and also Section 5.

Example 1. Let y_t be a 4×1 process defined as

$$\begin{aligned} y_t &= Ax_t + \varepsilon_t, \\ x_t &= \sigma_t e_t, \quad \sigma_t^2 = 1 + 0.9x_{t-1}^2, \end{aligned} \tag{4.1}$$

where $A = (0.1, 0.7, -0.1, -0.7)'$, $\varepsilon_t = (\varepsilon_{1t}, \cdots, \varepsilon_{4t})'$, and ε_{it} and e_t are independent and $N(0,1)$ random variables.

For the time series generated from (4.1) with length $T = 250, 500$ or 1000, we apply the new method proposed in Section 2.2 with $m = 5$ in (2.7) to estimate A. We also calculate the estimate for A using Hu and Tsay's PVC method with $c = 2.5$ and $m = 5$ specified by Hu and Tsay (2014). The means and standard deviations for $d(\widehat{\mathcal{M}}_1, \mathcal{M}_1)$ defined in (3.1)

over 2000 replication are reported in Table 1. The boxplots of $d(\widehat{\mathcal{M}}_1, \mathcal{M}_1)$ are presented in Figure 1. As A is a vector now, we can also measure the goodness of the estimation by looking at the sample correlation between $\widehat{A}'y_t$ and $A'y_t$. To this end, we define a discrepancy measure

$$d(\widehat{A}, A) = 1 - \frac{\left[\sum_t (y_t - \bar{y})'\widehat{A}A'(y_t - \bar{y})\right]^2}{\left[\sum_t (y_t - \bar{y})'\widehat{A}\widehat{A}'(y_t - \bar{y})\right]\left[\sum_t (y_t - \bar{y})'AA'(y_t - \bar{y})\right]}. \tag{4.2}$$

The mean and the standard deviation of $d(\widehat{A}, A)$ are also included in Table 1. As $N = 4$ and $r = 1$ now, both the methods perform well for this simple example, although the new method provides slightly more accurate estimates. Also as expected, the estimation errors decrease when the sample size T increases.

Table 1: Means and standard deviations (in parentheses) of $d(\widehat{\mathcal{M}}_1, \mathcal{M}_1)$ defined in (3.1) and $d(\widehat{A}, A)$ defined in (4.2) in simulation with 2000 replications in Example 1.

		New method	PVC method
$T = 250$	$d(\widehat{\mathcal{M}}_1, \mathcal{M}_1)$	0.035294 (0.021559)	0.042552 (0.027078)
	$d(\widehat{A}, A)$	0.000117 (0.000198)	0.000177 (0.000325)
$T = 500$	$d(\widehat{\mathcal{M}}_1, \mathcal{M}_1)$	0.022892 (0.012991)	0.027038 (0.015871)
	$d(\widehat{A}, A)$	0.000040 (0.000059)	0.000058 (0.000090)
$T = 1000$	$d(\widehat{\mathcal{M}}_1, \mathcal{M}_1)$	0.015411 (0.008200)	0.018677 (0.010316)
	$d(\widehat{A}, A)$	0.000016 (0.000022)	0.000024 (0.000033)

For $T = 1000$, we repeat the simulation also with $m = 1$ and also $m = 10$. The results together with those with $m = 5$ are reported in Table 2 and Figure 2. Since the results are almost the same with different values of m, this indicates that the proposed method is insensitive to the choice of m in (2.7).

Example 2. Now in model (4.1) we let A be the following 4×2 matrix

$$A = \begin{bmatrix} 0 & 7/10 \\ \sqrt{2}/2 & -1/10 \\ 0 & -7/10 \\ \sqrt{2}/2 & 1/10 \end{bmatrix},$$

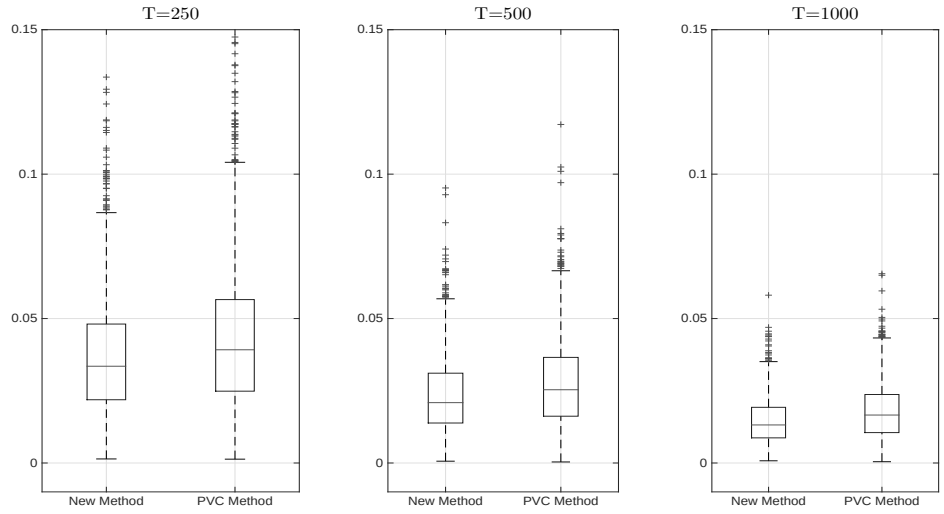

Figure 1: Boxplots of $d(\widehat{\mathcal{M}}_1, \mathcal{M}_1)$ with $T = 250, 500, 1000$ for Example 1.

Table 2: Means and standard deviations (in parentheses) of $d(\widehat{\mathcal{M}}_1, \mathcal{M}_1)$ defined in (3.1) and $d(\widehat{A}, A)$ defined in (4.2) in simulation with 2000 replications in Example 1 when $T = 1000$ and lag $m = 1, 5, 10$.

		New method	PVC method
$m = 1$	$d(\widehat{\mathcal{M}}_1, \mathcal{M}_1)$	0.014209 (0.008165)	0.018012 (0.010125)
	$d(\widehat{A}, A)$	0.000014 (0.000021)	0.000022 (0.000032)
$m = 5$	$d(\widehat{\mathcal{M}}_1, \mathcal{M}_1)$	0.015411 (0.008200)	0.018677 (0.010316)
	$d(\widehat{A}, A)$	0.000016 (0.000022)	0.000024 (0.000033)
$m = 10$	$d(\widehat{\mathcal{M}}_1, \mathcal{M}_1)$	0.015672 (0.008650)	0.018993 (0.010148)
	$d(\widehat{A}, A)$	0.000017 (0.000026)	0.000024 (0.000033)

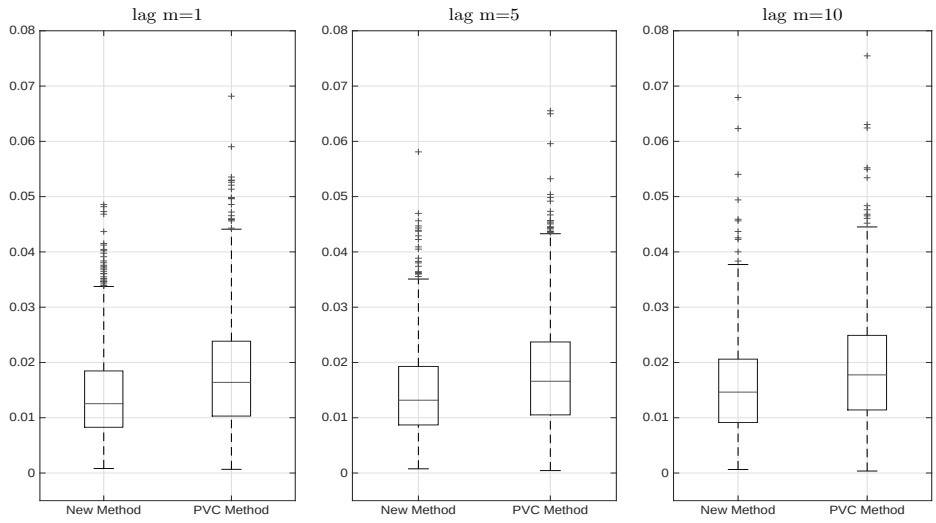

Figure 2: Boxplots of $d(\widehat{\mathcal{M}}_1, \mathcal{M}_1)$ with lag $m = 1, 5, 10$ for Example 1 when $T = 1000$.

the factor series be defined as

$$x_t = \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} \sigma_{1t} e_{1t} \\ \sigma_{2t} e_{2t} \end{pmatrix}, \quad \begin{pmatrix} \sigma_{1t}^2 \\ \sigma_{2t}^2 \end{pmatrix} = \begin{pmatrix} 1 + 0.8 x_{1,t-1}^2 \\ 2 + 0.9 x_{2,t-1}^2 \end{pmatrix},$$

and ε_{it}, e_{it} be independent and $N(0,1)$.

We conduct the simulation in the same manner as in Example 1, and the results are presented in Table 3 and Figure 3, which display the similar patterns observed in Table 1 and Figure 1.

Table 3: Means and standard deviations (in parentheses) of $d(\widehat{\mathcal{M}}_1, \mathcal{M}_1)$ defined in (3.1) in simulation with 2000 replications in Example 2.

		New method	PVC method
$T = 250$	$d(\widehat{\mathcal{M}}_1, \mathcal{M}_1)$	0.0356 (0.0196)	0.0489 (0.0470)
$T = 500$	$d(\widehat{\mathcal{M}}_1, \mathcal{M}_1))$	0.0232 (0.0136)	0.0338 (0.0412)
$T = 1000$	$d(\widehat{\mathcal{M}}_1, \mathcal{M}_1)$	0.0152 (0.0081)	0.0216 (0.0223)

Example 3. We use the same setting as in Example 2, except now ε_{it}, e_{it} are independent and $t(5)$ (instead of $N(0,1)$) random variables. The results are reported in Table 4 and Figure 4. Comparing the results for Example 2, the newly proposed method performs better for the models with heavy-tailed (i.e. $t(5)$-distributed) innovations. This may be due to the

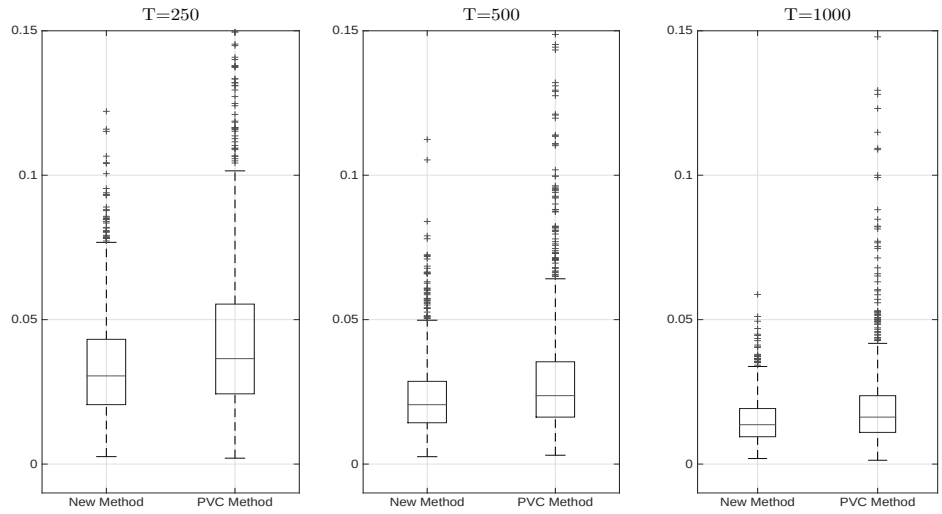

Figure 3: Boxplots of $d(\widehat{\mathcal{M}}_1, \mathcal{M}_1)$ with $T = 250, 500, 1000$ for Example 2.

fact that the sets in $\mathcal{B}$ defined in (2.6) are more spread-out with heavier tailed innovations. In contrast, the performance of the PVC method is slightly worse now as the information at (heavy) tails is truncated by the definition of the method. See Hu and Tsay (2014).

Table 4: Means and standard deviations (in parentheses) of $d(\widehat{\mathcal{M}}_1, \mathcal{M}_1)$ defined in (3.1) in simulation with 2000 replications in Example 3.

		New method	PVC method
$T = 250$	$d(\widehat{\mathcal{M}}_1, \mathcal{M}_1)$	0.0188 (0.0180)	0.0436 (0.0795)
$T = 500$	$d(\widehat{\mathcal{M}}_1, \mathcal{M}_1))$	0.0100 (0.0112)	0.0328 (0.0761)
$T = 1000$	$d(\widehat{\mathcal{M}}_1, \mathcal{M}_1)$	0.0057 (0.0057)	0.0266 (0.0763)

Example 4. Let's consider a high-dimensional case of the simulated model. Now in model (4.1), we let the loading matrix A be a 100×1 matrix, where A is normalized as: $A = \frac{\bar{A}}{\|\bar{A}\|}$ and each element of $\bar{A}$ is randomly and independently drawn from uniform distribution on interval [-1,1].

The factor series remain the same as in (4.1). Innovations ε_t are normalized as: $\varepsilon_t = \frac{\bar{\varepsilon}_t}{\sqrt{100}}$, and $\bar{\varepsilon}_{it}$ and e_t are independent and $N(0,1)$ random variables. Normalization we made here ensures that Ax_t and ε_t are of the same magnitude.

We conduct the simulation in the same manner as in Example 1. The results are presented in Table 5 and Figure 5. Similar patterns to those in Example 1 are observed

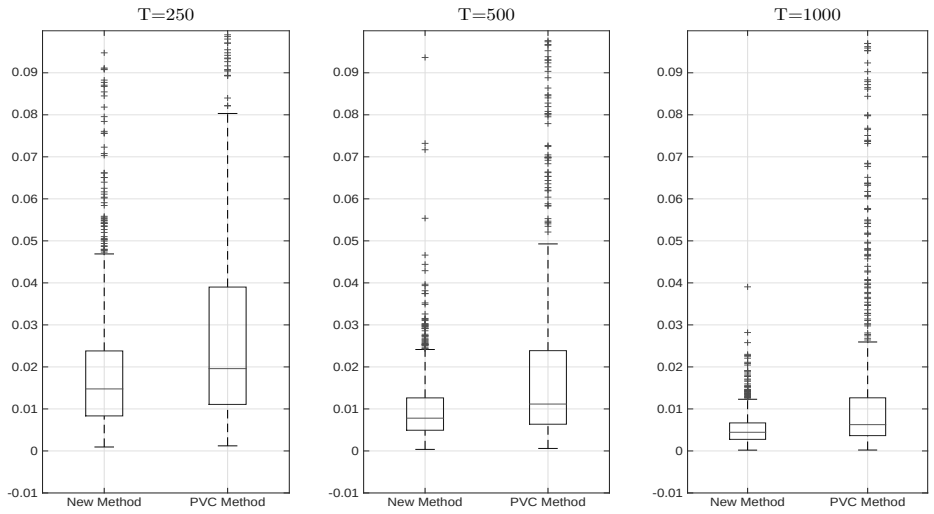

Figure 4: Boxplots of $d(\widehat{\mathcal{M}}_1, \mathcal{M}_1)$ with $T = 250, 500, 1000$ for Example 3.

here. This example indicates the potential application of the proposed method in high-dimensional cases.

Table 5: Means and standard deviations (in parentheses) of $d(\widehat{\mathcal{M}}_1, \mathcal{M}_1)$ defined in (3.1) and $d(\widehat{A}, A)$ defined in (4.2) in simulation with 1000 replications in Example 4.

		New method	PVC method
$T = 250$	$d(\widehat{\mathcal{M}}_1, \mathcal{M}_1)$	0.0424565 (0.0176764)	0.0566304 (0.0236632)
	$d(\widehat{A}, A)$	0.0000081 (0.0000110)	0.0000149 (0.0000190)
$T = 500$	$d(\widehat{\mathcal{M}}_1, \mathcal{M}_1)$	0.0276088 (0.0101171)	0.0382413 (0.0148118)
	$d(\widehat{A}, A)$	0.0000024 (0.0000027)	0.0000049 (0.0000052)
$T = 1000$	$d(\widehat{\mathcal{M}}_1, \mathcal{M}_1)$	0.0181864 (0.0056076)	0.0260939 (0.0089832)
	$d(\widehat{A}, A)$	0.0000008 (0.0000007)	0.0000018 (0.0000016)

5 Illustration with real data

We further illustrate the proposed method with two real data examples. Both data sets were downloaded from *Yahoo!Finance*.

Example 5. We consider the daily returns of the six stocks in the period of January 2, 2002 — July 10, 2008. The stocks concerned are Bank of America corporation, Dell Inc., JPMorgan Chase&Co., FedEx Corporation, McDonald's Corp. and American International Group. For this example, $T = 1642$ and $N = 6$. The returns are calculated based on the

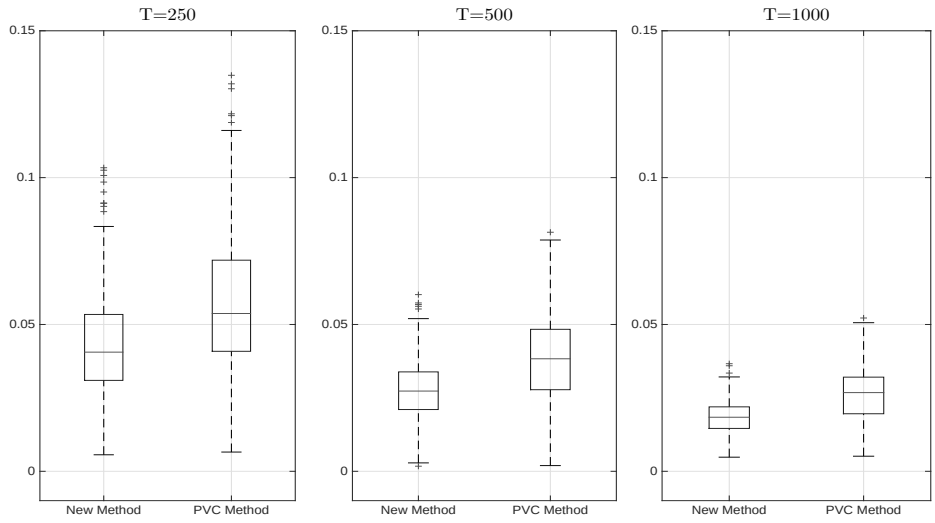

Figure 5: Boxplots of $d(\widehat{\mathcal{M}}_1, \mathcal{M}_1)$ with $T = 250, 500, 1000$ for Example 4.

daily closing prices.

We apply the proposed method with $m = 5$ in (2.7). The ratio estimator (2.8) leads to $\widehat{r} = 1$, indicating the volatility space $\mathcal{M}_1$ is one-dimensional. In fact the estimation results are hardly changed for m between 1 and 10. We also apply the PVC method of Hu and Tsay (2014), obtaining almost identical results for this example. The ratios of the eigenvalues and the factor loadings (i.e. $\widehat{A}$) for the two methods are listed in Tables 6 and 7. In fact that the sample correlation coefficient between the two loadings in Table 7 is 0.9993.

Table 6: Ratios of eigenvalues for Example 5.

	New Method	PVC method
λ_1/λ_2	36.90	25.77
λ_2/λ_3	1.22	1.22
λ_3/λ_4	1.41	2.18
λ_4/λ_5	1.52	1.59
λ_5/λ_6	3.13	1.77

The above analysis indicates that the volatilities (including co-volatilities) of the six return series are driven by a one-dimensional process. To visualize this extremely low dimensional structure, we plot the sample autocorrelation function (ACF) of the transformed squared return series $(\widehat{\gamma}_i' y_t)^2$, $i = 1, \cdots, 6$, where $\widehat{\gamma}_i$ is the eigenvector of matrix $\widehat{M}_1$, defined in (2.7), corresponding to the i-th largest eigenvalue. Figure 6 shows significant and persis-

Table 7: Factor loadings for Example 5.

	New Method	PVC method
Bank of America corporation	0.3922	0.3663
Dell Inc.	0.3138	0.3037
JPMorgan Chase&Co.	0.6492	0.6690
FedEx Corporation	0.2224	0.2140
McDonald's Corp.	0.1263	0.1398
American International Group	0.5107	0.5105

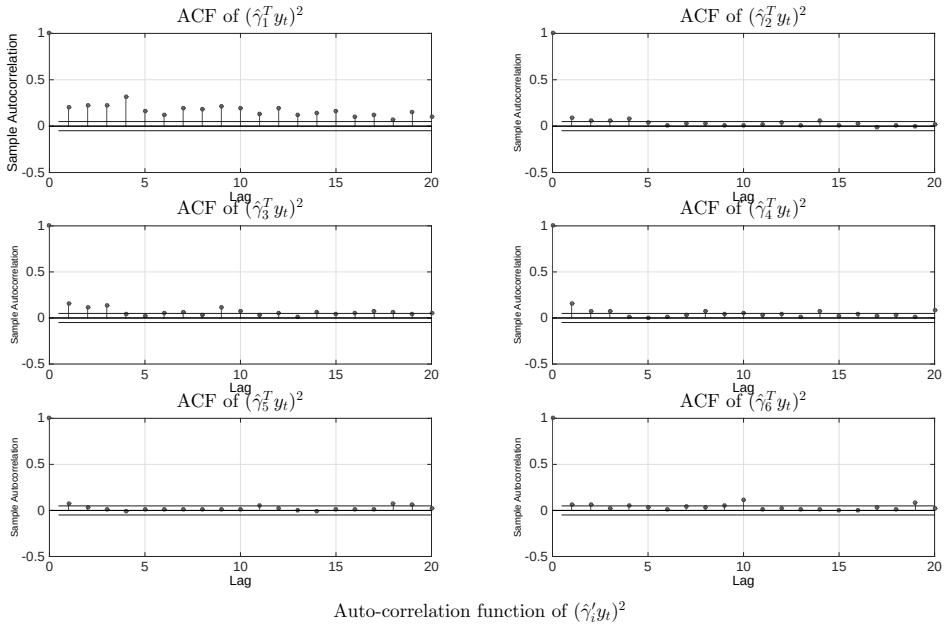

Figure 6: ACF of $(\widehat{\gamma}_i' y_t)^2, i = 1, \cdots, 6$ for Example 5

tent autocorrelations in the series $(\widehat{\gamma}_1' y_t)^2$, and this is not the case for $(\widehat{\gamma}_2' y_t)^2, \cdots, (\widehat{\gamma}_6' y_t)^2$. The displays in Figure 6 lend further support to the conclusion that the volatility space for these 6 return series is one-dimensional.

Example 6. Now we model the daily returns of 196 stocks included in S&P500 in the period of July 13, 2009 – July 11, 2014. The returns are calculated based on the daily closing prices. Now the dimension $N = 196$ and the number of the total observations is $T = 1259$.

We use the proposed method with $m = 5$ in (2.7). The ratio estimator (2.8) again leads to $\widehat{r} = 1$, suggesting that the dimension of volatility space $\mathcal{M}_1$ is one. The result hardly changed when we vary m between 1 and 10.

The PVC method of Hu and Tsay (2014) is also applied in this example, indicating

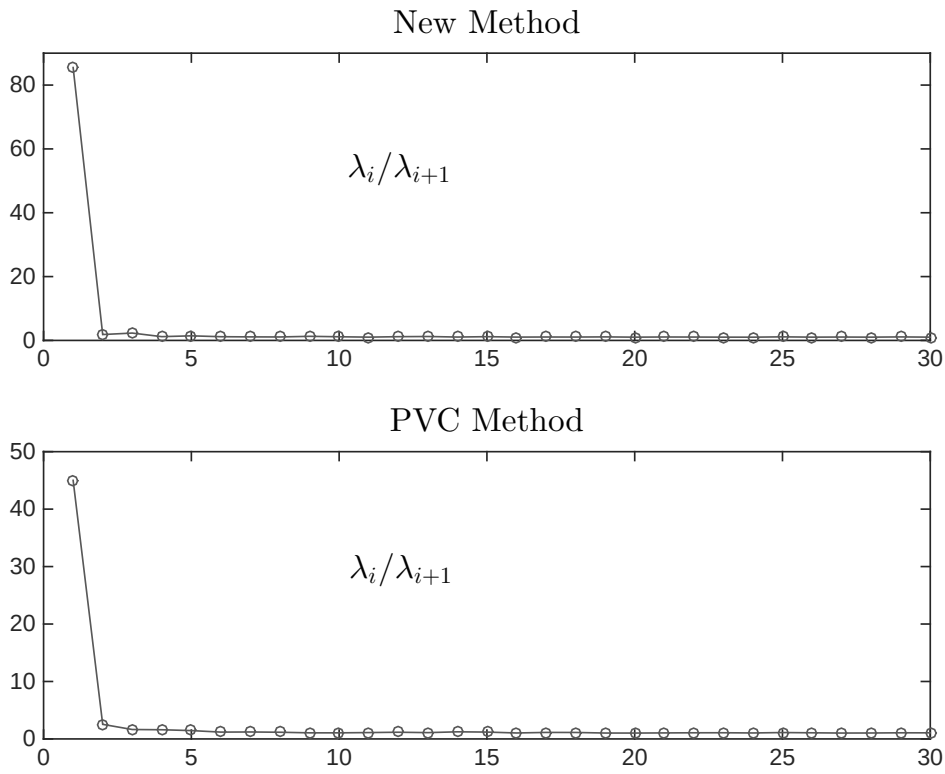

Figure 7: First 30 ratios of eigenvalues for Example 6: $\lambda_i/\lambda_{i+1}, i = 1, \cdots, 30$

almost the same result. For both the methods, the first 30 ratios of the eigenvalues are plotted in Figure 7. The remaining ratios after $i = 30$ are at most 1.17 and they are not included in the figure. The factor loadings (i.e. $\widehat{A}$) are presented in Figure 8. Moreover, the sample correlation coefficient between the two loadings in Figure 8 is 0.936.

Since $\widehat{r} = 1$, the conditional heteroscedasticity effect of the 196 time series of stock returns may be regarded as driven by one scalar process. The other 195 orthogonal combinations of the 196 returns exhibit little conditional heteroscedasticity effect. Thus this high-dimensional volatility process is reduced to a one-dimensional process, achieving the maximum dimension reduction.

To visualize the extremely low-dimension of the conditional heteroscedasticity for this data set, the sample ACFs of the first 6 transformed squared return series $(\widehat{\gamma}_i' y_t)^2$, $i = 1, \cdots, 6$ are plotted in Figure 9, where $\widehat{\gamma}_i$ is the eigenvector of matrix $\widehat{M}_1$, defined in (2.7), corresponding to the i-th largest eigenvalue. Figure 9 shows significant and persistent

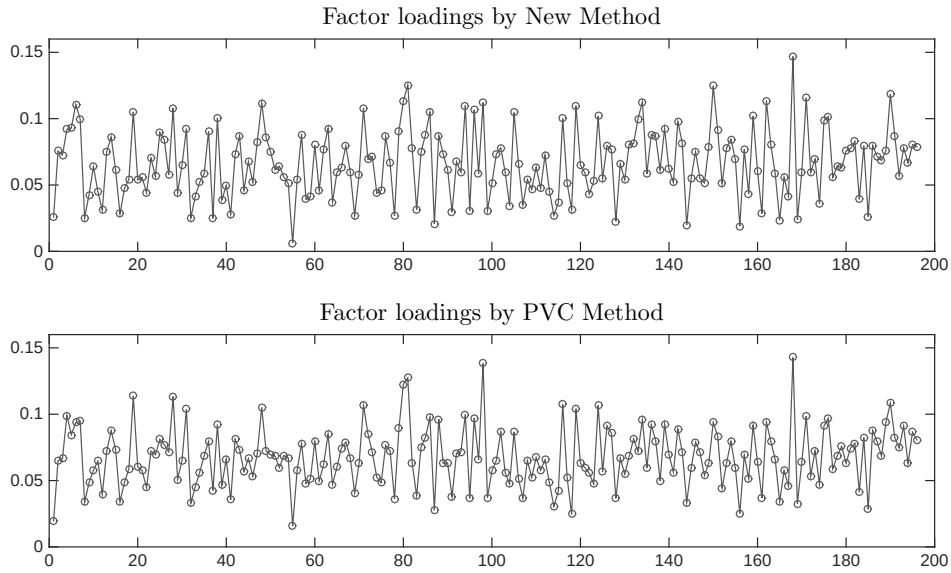

Figure 8: Factor loadings for Example 6

autocorrelations in the series $(\widehat{\gamma}_1' y_t)^2$. This is not the case for $(\widehat{\gamma}_2' y_t)^2, \cdots, (\widehat{\gamma}_6' y_t)^2$. Figure 9 provides further evidence to support the claim that the volatility space for these 196 returns series is one-dimensional.

6 Concluding remarks

For multiple volatility processes with low-dimensional dynamic structures, we propose to model the conditional variance and covariance by latent common factors. Technically the method boils down to an eigenanalysis for a non-negative definite matrix. Thus it is applicable when the dimension the time series is in the order of thousands. The method can be viewed as a generalized version of the PVC method of Hu and Tsay (2014). In addition to the computational efficiency, the new method imposes fewer moment conditions, and is more efficient in dealing with the processes with heavy tailed innovations (as shown in Example 3). The method also can be applied to high-dimensional cases (see, e.g. Examples 4 and 6), though it remains as an open problem to extend the asymptotic theory with diverging N in addition to diverging T.

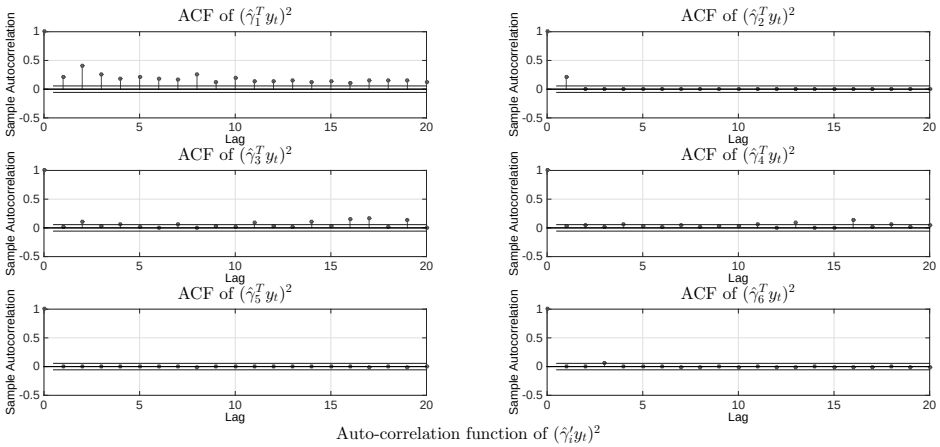

Figure 9: ACF of $(\widehat{\gamma}_i' y_t)^2, i = 1, \cdots, 6$ for example 6

Acknowledgements

We would like to thank the Associate Editor and the two referees for their critical and constructive comments that have greatly improved our paper. We would also like to thank Professor Qi Li for his helpful comments/suggestions during the writing of this paper. Kunpeng Li's research is supported by the NSFC No. 71201031 and No. 71571122. Qiwei Yao's research is partially supported by an EPSRC research grant.

References

Arcones, M. A., and Yu, B. (1994). Central limit theorems for empirical and U-processes of stationary mixing sequences. *Journal of Theoretical Probability*, **7**, 47-71.

Asai, M., McAleer, M. and Yu, J. (2006). Multivariate stochastic volatility: a review. *Econometric Reviews*, **25**, 145-175.

Bauwens, L., Laurent, S., and Rombouts, J. V. (2006). Multivariate GARCH models: a survey. *Journal of applied econometrics*, **21**, 79-109.

Chow, Y. S., and Teicher, H. (1997). *Probability Theory*. Springer, New York.

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

Fan, J., Wang, M., and Yao, Q. (2008). Modelling multivariate volatilities via conditionally uncorrelated components. *Journal of the Royal Statistical Society*, **B**, **70**, 679-702.

Golub, G. H., and van Van Loan, C. F. (1996). *Matrix computations*, 3rd ed. Johns Hopkins University Press.

Hu, P.-Y., and Tsay, R.S. (2014). Principal Volatility Component Analysis, *Journal of Business & Economic Statistics*, **32**, 153-164.

Johnstone, I. M., and Lu, A. Y. (2009). On consistency and sparsity for principal components analysis in high dimensions. *Journal of the American Statistical Association*, **104**, 682-693.

Lam, C., Yao, Q., and Bathia, N. (2011). Estimation of latent factors for high-dimensional time series. *Biometrika*, **98**, 901-918.

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

Lawley, D. N., and Maxwell, A. E. (1971). *Factor analysis as a statistical method*. American Elsevier Pub.

Pan, J., Polonik, W., and Yao, Q. (2010). Estimating factor models for multivariate volatilities: An innovation expansion method. In Lechevallier, Y. and Saporta, G., editors, *Proceedings of COMPSTAT'2010*, 305-314. Physica-Verlag HD.

Polonik, W., and Yao, Q. (2000). Conditional minimum volume predictive regions for stochastic processes. *Journal of the American Statistical Association*, **95**, 509-519.

Polonik, W., and Yao, Q. (2002). Set-indexed conditional empirical and quantile processes based on dependent data. *Journal of Multivariate Analysis*, **80**, 234-255.

Silvennoinen, A., and Terasvirta, T. (2008). Multivariate GARCH Models. Available at `http://dx.doi.org/10.2139/ssrn.1148139`.

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

Wang, M. and Yao, Q. (2005). Modelling multivariate volatilities: an ad hoc approach. In *Contemporary Multivariate Analysis and Experimental Designs – In Celebration of Professor Kai-Tai Fang's 65th Birthday* J. Fan & G. Li (edit.), 87-97. World Scientific, Singapore.