Efficient Factor GARCH Models and Factor-DCC Models

Kun Zhang KZHANG@CSE.CUHK.EDU.HK

Laiwan Chan LWCHAN@CSE.CUHK.EDU.HK

Department of Computer Science and Engineering

The Chinese University of Hong Kong

Hong Kong

Abstract

We reveal that in the estimation of univariate GARCH or multivariate generalized orthogonal GARCH (GO-GARCH) models, maximizing the likelihood is equivalent to making the standardized residuals as independent as possible. Based on that, we propose three factor GARCH models in the framework of GO-GARCH: independent-factor GARCH exploits factors that are statistically as independent as possible; factors in best-factor GARCH have the largest autocorrelation in their squared values such that their volatilities could be forecasted well by univariate GARCH; factors in conditional-decorrelation GARCH are conditionally as uncorrelated as possible. A two-step method for estimating these models is introduced, and it gives easy and reliable estimation. Since the extracted factors may still have weak conditional correlations, we further propose factor-DCC models, as an extension to the above factor GARCH models with dynamic conditional correlation (DCC) modelling the remaining conditional correlations between factors. Experimental results on the Hong Kong stock market shows that conditional-decorrelation GARCH and independent-factor GARCH have better generalization performance than the original GO-GARCH, and that conditional-decorrelation GARCH (among factor GARCH models) and its extension with DCC embedded (among factor-DCC models) behave best.

Keywords: Factor GARCH model, Mutual information, Mutual independence, Conditional uncorrelatedness, Autocorrelation, Dynamic conditional correlation

1. Introduction

It has been over 20 years since the autoregressive conditional heteroscedasticity (ARCH) model was proposed (Engle, 1982). A natural and powerful extension of the ARCH model is the generalized ARCH (GARCH) model (Bollerslev, 1986). These models have been shown to be very useful in modelling and forecasting the volatility of financial return series. Furthermore, financial assets may be inter-correlated, and their correlations play an important role in financial management, such as portfolio construction. It was then natural to generalize the GARCH models from modelling the conditional variance of a univariate series to modelling the conditional covariance matrix of multivariate series (Bollerslev et al., 1988); for a recent survey on multivariate GARCH model, see Bauwens et al. (2006).

Conventionally univariate and multivariate GARCH models are estimated by maximum likelihood (ML) estimation, or quasi maximum likelihood (QML) estimation, which simply assumes that the returns are conditionally normally distributed. The maximum likelihood of GARCH models is actually closely related to statistical dependence in standardized residuals.[i] In the literature, after estimating univariate GARCH models, some statistical tests, such as the BDS test (Brock et al., 1996), can be used to test the independence of standardized residuals, so that we can compare the behavior of different formulations for the GARCH model. In this work, we reveal the relationship between the mutual information of standardized residuals and the likelihood value attained by GARCH models. We show that in the estimation of GARCH models, maximizing the likelihood is equivalent to minimizing the mutual information of standardized residuals. Actually, when we use different GARCH models to fit the given data, the likelihood is a direct indicator of how independent the standardized residuals are.

Usually multivariate GARCH models involve quite a lot of parameters, and they are difficult to estimate for high-dimensional data (Bauwens et al., 2006). Factor GARCH models (e.g., Engle et al., 1990; Alexander, 2000; van der Weide, 2002) provide one way to efficiently parameterize the GARCH model. In this paper, by considering the generalized orthogonal GARCH (GO-GARCH) model (van der Weide, 2002) from an information-theoretic point of view, we propose to use independent component analysis (ICA) and two other statistical methods to construct factor GARCH models in the framework of GO-GARCH. The proposed models can all be estimated easily in two separate steps. In the first step the factors are estimated according to some statistical criteria. In the second step we estimate the univariate GARCH model for each factor and

i. In the GARCH literature, the innovation normalized by its time-varying standard deviation is defined as the standardized residual.

construct the conditional covariance matrix of return series. Due to the simplicity and efficiency of parameter estimation, these models are suitable for high-dimensional data. Since the extracted factors may still have weak conditional correlations, we show that the estimate of conditional correlations between returns can be further improved by modelling the remaining time-varying conditional correlations between factors with the dynamic conditional correlation (DCC) model. This leads to factor-DCC models.

This paper is organized as follows. In Section 2 we show the relationship between statistical dependence in standardized residuals and the data likelihood of the univariate GARCH models. Section 3 reviews multivariate GARCH models and discusses the estimation the GO-GARCH model by the minimization of mutual information. Three forms of factor GARCH models, which exploit ICA and other techniques for factor extraction, are proposed and discussed in Section 4. In Section 5 we further embed the DCC model into the factor models to improve the forecasting performance. 10 stocks selected from Hong Kong stock market are used to compare the performance of our proposed factor GARCH models, the orthogonal GARCH , GO-GARCH, the DCC model, and the factor-DCC models in Section 6. Section 7 concludes this paper.

2. Estimation of Univariate GARCH Model Based on Mutual Information

2.1 Univariate GARCH Models

It is well known that financial return series are almost serially uncorrelated, but the squares of the returns are often strongly correlated, as indicated by the phenomenon of volatility clustering. Based on these facts, one can see that the return series are not serially independent, and that the dependency in the return series is demonstrated by the autocorrelation of the squared values. Consequently, the return magnitude is predictable to some extent.

Inspired by volatility clustering, the ARCH-type models (Engle, 1982; Bollerslev et al., 1992; Bera & Higgins, 1993; Poon & Granger, 2003) were proposed to model the volatility of the return series. A popular one is the GARCH(p,q) model (Bollerslev, 1986), which is described by the following system,

$$r_t = u_t + \epsilon_t, \tag{1}$$

$$\epsilon_t = \sqrt{h_t} z_t, \tag{2}$$

$$h_t = w + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}, \tag{3}$$

where r_t denotes the return at time t, u_t denotes the conditional mean of r_t conditional on information at time $t-1$ and can be described by any type of regression models, ϵ_t is the innovation, h_t

is the conditional variance of r_t based on information up through time $t-1$, z_t is i.i.d. with mean zero and variance one, and restrictions $w > 0$, $\alpha_i \geq 0$, and $\beta_i \geq 0$ are imposed to ensure that h_t is positive (actually the restrictions on parameters may be weaker, see Nelson & Cao, 1992). In Engle's original ARCH model (Engle, 1982), z_t is assumed to be Gaussian distributed. The main approach to parameter estimation in GARCH models is based on ML estimation. Although there is a lot of empirical evidence that the standardized residual z_t does not follow the Gaussian distribution, the normality of z_t is often assumed and produces the QML estimates (Bollerslev & Wooldridge, 1992).

2.2 Estimation of Univariate GARCH Model by Mutual Information Minimization

Parameters in the GARCH models are adjusted to capture the autocorrelation in squared returns, which reflects the dependency in the return series. Consequently, after eliminating the effect the time-varying volatility modeled by GARCH models, the standardized residuals, z_t, should be much more serially independent. Therefore, we can estimate the parameters in GARCH models by making z_t serially as independent as possible. The statistical dependence can be measured by mutual information.

Mutual information is a natural and canonical measure of statistical dependence. In information theory, the mutual information between n random variables $y_1, ..., y_n$ is defined as

$$I(Y_1, ..., Y_n) = \sum_{i=1}^{n} H(Y_i) - H(\mathbf{Y}), \tag{4}$$

where $\mathbf{Y} = (Y_1, ..., Y_n)^T$, and $H(\cdot)$ denotes the (differential) entropy, defined as $H(X) = -\int p(x) \log p(x) dx$ (Cover & Thomas, 1991). $I(Y_1, ..., Y_n)$ is always non-negative, and is zero if and only if y_i are mutually independent. Different from the coefficient of linear correlation, which measures the linear dependency between two variables, mutual information can capture both linear and nonlinear dependence, with no need to specify any form of dependency. Mutual information has been applied to detect the dependency or to examine the predictability of financial time series (Dionísio et al., 2003; Darbellay & Wuertz, 2000; Maasoumi & Racine, 2002).

In the GARCH(p,q) model (Eq. 1–3), we can see that the standardized residual z_t is a function of $\epsilon_t, \epsilon_{t-1}, \cdots$, denoted by

$$z_t = f(\epsilon_t, \epsilon_{t-1}, ...; \boldsymbol{\theta}), \tag{5}$$

where $\boldsymbol{\theta}$ is the parameter set containing the parameters in the GARCH formulation. In particular, for the GARCH(p,q) model (Eq. 3), $\boldsymbol{\theta} = \{w, \alpha_1, ..., \alpha_q, \beta_1, .., \beta_p\}$. Bollerslev (1986) showed that the necessary and sufficient condition for z_t to be covariance stationary is $\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i < 1$. Under this condition, the effect of ϵ_{t-k} on h_t (and z_t) diminishes as k increases.

Denote by $\mathbf{z}^*$ the vector consisting of z_t at different time indices, i.e. $\mathbf{z}^* = (z_t, z_{t-1}, ..., z_{t-N})^T$, and similarly for $\mathbf{r}^*$ and $\boldsymbol{\epsilon}^*$. The dependence in the process $\mathcal{Z} \triangleq \{z_t\}$ can be measured by the mutual information rate (Taleb et al., 2001), which is defined as

$$\mathcal{I}(\mathcal{Z}) = H(z_t) - \mathcal{H}(\mathcal{Z}), \tag{6}$$

where $\mathcal{H}(\mathcal{Z}) = \lim_{N\to\infty} \frac{H(\mathbf{z}^*)}{N+1}$ is the entropy rate of the process $\mathcal{Z}$ (Cover & Thomas, 1991). The mutual information rate of the process $\mathcal{Z}$ can also be written as $\mathcal{I}(\mathcal{Z}) = \lim_{N\to\infty} \frac{1}{N+1} \cdot I(\mathbf{z}^*)$. When N is sufficiently large, we can neglect the effect of $\epsilon_{t-N-1}, \epsilon_{t-N-2}, \cdots$, as they are very early, and write Eq. 5 in matrix form:

$$\mathbf{z}^* = \mathcal{F}(\boldsymbol{\epsilon}^*; \boldsymbol{\theta}), \tag{7}$$

where $\mathcal{F}$ denotes the mapping from $\boldsymbol{\epsilon}^*$ to $\mathbf{z}^*$. Similarly, we denote the mapping from $\mathbf{r}^*$ to $\boldsymbol{\epsilon}^*$ by $\mathcal{G}$, i.e.

$$\boldsymbol{\epsilon}^* = \mathcal{G}(\mathbf{r}^*; \boldsymbol{\psi}), \tag{8}$$

where $\boldsymbol{\psi}$ contains the parameters in the regression model for u_t. The overall mapping from $\mathbf{r}^*$ to $\mathbf{z}^*$ is then $\mathcal{F} \circ \mathcal{G}$. We can estimate the parameters in the GARCH models by minimizing $I(\mathbf{z}^*)$, the mutual information between components of $\mathbf{z}^*$.

The Jacobian matrix of the mapping $\mathcal{F}$ is

$$\mathbf{J}_{\mathcal{F}} = \begin{bmatrix} \frac{\partial z_t}{\partial r_t} & \frac{\partial z_t}{\partial r_{t-1}} & \cdots & \frac{\partial z_t}{\partial r_{t-N}} \\ \frac{\partial z_{t-1}}{\partial r_t} & \frac{\partial z_{t-1}}{\partial r_{t-1}} & \cdots & \frac{\partial z_{t-1}}{\partial r_{t-N}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial z_{t-N}}{\partial r_t} & \frac{\partial z_{t-N}}{\partial r_{t-1}} & \cdots & \frac{\partial z_{t-N}}{\partial r_{t-N}} \end{bmatrix} = \begin{bmatrix} h_t^{-1/2} & \frac{\partial z_t}{\partial r_{t-1}} & \cdots & \frac{\partial z_t}{\partial r_{t-N}} \\ 0 & h_{t-1}^{-1/2} & \cdots & \frac{\partial z_{t-1}}{\partial r_{t-N}} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h_{t-N}^{-1/2} \end{bmatrix}. \tag{9}$$

Denote the determinant of $\mathbf{J}_{\mathcal{F}}$ by $|\mathbf{J}_{\mathcal{F}}|$. Obviously $|\mathbf{J}_{\mathcal{F}}| = \prod_{i=0}^{N} h_{t-i}^{-1/2}$. No matter $u_t = 0$ or is described by a regression model, the Jacobian matrix $\mathbf{J}_{\mathcal{G}}$ is always a upper triangular matrix with the entries on its diagonal being 1, so $|\mathbf{J}_{\mathcal{G}}| = 1$. The determinant of the Jacobian matrix associated with the overall transformation $\mathcal{F} \circ \mathcal{G}$ is $|\mathbf{J}| = |\mathbf{J}_{\mathcal{F}\circ\mathcal{G}}| = |\mathbf{J}_{\mathcal{F}}| \cdot |\mathbf{J}_{\mathcal{G}}| = |\mathbf{J}_{\mathcal{F}}| = \prod_{i=0}^{N} h_{t-i}^{-1/2}$. We then have the following relationship between the joint density $p_{\mathbf{z}^*}(\mathbf{z}^*)$ and $p_{\mathbf{r}^*}(\mathbf{r}^*)$:

$$p_{\mathbf{z}^*}(\mathbf{z}^*) = \frac{p_{\mathbf{r}^*}(\mathbf{r}^*)}{|\mathbf{J}|} = \frac{p_{\mathbf{r}^*}(\mathbf{r}^*)}{\prod_{i=0}^{N} h_{t-i}^{-1/2}}. \tag{10}$$

As $H(\mathbf{z}^*) = -E\{\log p_{\mathbf{z}^*}(\mathbf{z}^*)\}$, we further have

$$H(\mathbf{z}^*) = H(\mathbf{r}^*) + E\Big\{\log \prod_{i=0}^{N} h_{t-i}^{-1/2}\Big\} = H(\mathbf{r}^*) - \frac{1}{2}\sum_{i=0}^{N} E \log h_{t-i}.$$

Consequently, the mutual information rate of the process $\{z_t\}$ is

$$\begin{aligned}\mathcal{I}(\mathcal{Z}) &= H(z_t) - \mathcal{H}(\mathcal{Z}) = H(z_t) - \lim_{N\to\infty} \frac{H(\mathbf{z}^*)}{N+1} \\ &= H(z_t) + \lim_{N\to\infty} \cdot \frac{1}{N+1} \cdot \frac{1}{2} \sum_{i=0}^{N} E \log h_{t-i} - \lim_{N\to\infty} \frac{H(\mathbf{r}^*)}{N+1} \\ &= H(z_t) + \frac{1}{2} E \log h_t - \mathcal{H}(\mathcal{R}), \qquad (11)\end{aligned}$$

where $\mathcal{R}$ denotes the process $\{r_t\}$. As $\mathcal{H}(\mathcal{R})$ does not depend on the parameters $\boldsymbol{\theta}$, minimizing $\mathcal{I}(\mathcal{Z})$ is equivalent to maximizing the following function (T denotes the number of observations)

$$\begin{aligned}l_T(\boldsymbol{\theta}) &= -T \cdot \left[H(z_t) + \frac{1}{2} E \log h_t \right] \qquad (12) \\ &= \sum_{t=1}^{T} \log[p_z(z_t)] - \frac{1}{2} \sum_{t=1}^{T} \log h_t, \qquad (13)\end{aligned}$$

which is exactly the log-likelihood function for the observations $r_1, ..., r_T$. Note that z in Eq. 13 denotes the variable associated with z_t.[ii]

This reveals the fact that in the estimation of univariate GARCH models, maximizing the data likelihood is equivalent to minimizing the statistical dependence in standardized residuals. Although these two approaches have the same objective function, we should address their distinction. In ML estimation, the density of the standardized residuals, p_z, must be correctly specified in advance. However, in practice we do not have such information exactly. QML estimation simply assumes p_z to be Gaussian. In the approach based on the minimization of mutual information, p_z is not necessarily known as prior information, and it may be adaptively estimated from data in the maximization of Eq. 13, as the so-called "semiparametric ARCH models" does in Engle and González-rivera (1991). Consequently this method may produce a larger likelihood value (or more independent standardized residuals).

3. Estimation of GO-GARCH by Mutual Information Minimization

3.1 Multivariate GARCH Models

Since the financial return series may be correlated, in addition to the time-varying volatility of each series, the time-varying correlations among them are also very useful and need to be modeled and forecasted. For example, the time-varying correlations can help us to construct a short-term

ii. In the following we drop the time index t in the subscript to denote the variable associated with a time series.

portfolio. Hence there is a need to extend the univariate GARCH models to the multivariate case. For a survey on multivariate GARCH models, see Bauwens et al. (2006); Long (2005).

Suppose we have n return series $r_{it}, i = 1, ..., n$, which form a vector $\mathbf{r}_t = (r_{1t}, ..., r_{nt})^T$. Let $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, ..., \epsilon_{nt})^T$ be the zero-mean version of $\mathbf{r}_t$, obtained by subtracting the mean from $\mathbf{r}_t$. Denote the conditional covariance matrix of $\mathbf{r}_t$ by $\mathbf{H}_t$. The basic and general form for modelling the multivariate conditional covariance matrix is the *vech* model (Bollerslev et al., 1988). Using the *vech* operator to stack the lower triangular portion of a symmetric matrix into a column vector, it models each element of $\mathbf{H}_t$ as a linear combination of the lagged squared errors, cross-products of errors, and the lagged elements of $\mathbf{H}_t$. As a special case of the *vech* model, the BEKK model was proposed in Baba et al. (1991) and Engle and Kroner (1995). However, the number of parameters in these models grows very rapidly as the data dimension increases. This causes problems in parameter estimation when the data dimension is high.

In order to efficiently parameterize multivariate GARCH, some methods have been developed based on univariate GARCH models, in two directions. The first is to construct the conditional covariance matrix by explicitly modelling both the volatility of each return series and the conditional correlation matrix. For example, the constant conditional correlation (CCC) model of Bollerslev (1990) assumes the conditional correlation to be constant and the time-varying behavior of conditional covariances is due to the time-varying conditional variances. To relax the constant conditional correlation assumption, Engle (2002), Christodoulakis and Satchell (2002), and Tse and Tsui (2002) proposed the dynamic conditional correlation (DCC) model, the Correlated ARCH (CorrARCH) model, and the variable conditional correlation (VCC) model, respectively. They generalize the CCC model to allow the conditional correlation to be time dependent.

The second direction is to exploit the idea of factor models to construct multivariate GARCH models. It is believed that there exist some common factors, which may be determined empirically or constructed by some statistical methods, driving the evolution of the return series. The idea of the multivariate factor GARCH model dates to Engle et al. (1990), in which only a small number of common factors are empirically determined and believed to underly the observed return series. In this model, the factors are not necessarily uncorrelated, and it is difficult to determine what the factors are.

3.2 Orthogonal GARCH Models

Recently the orthogonal GARCH (O-GARCH) model was proposed by exploiting the principal component analysis (PCA) technique to determine the factors (Ding, 1994; Alexander, 2000; Alexander, 2001). The O-GARCH model allows $n \times n$ covariance matrices to be generated from

m (normally $m < n$) univariate GARCH processes, each of which models a principal component of the data $\boldsymbol{\epsilon}_t$. In order to do that, we first use PCA to extract m principal components, i.e. $\mathbf{y}_t = \mathbf{P}_m^T \boldsymbol{\epsilon}_t$, where $\mathbf{y}_t = (y_{1t}, ..., y_{mt})^T$ consists of the m principal components, and $\mathbf{P}_m$ is a $n \times m$ matrix consisting of the eigenvectors associated with the m largest eigenvalues of the covariance matrix of $\boldsymbol{\epsilon}_t$. Next, we use a univariate GARCH model to estimate $h_{y_{it}}$, the conditional variance of each principal component, and construct the $m \times m$ time-varying diagonal matrix $\boldsymbol{\Sigma}_t = diag\{h_{y_{1t}}, ..., h_{y_{mt}}\}$. Finally we can obtain the conditional covariance matrix of $\boldsymbol{\epsilon}_t$ as

$$\mathbf{H}_t = \mathbf{P}_m \boldsymbol{\Sigma}_t \mathbf{P}_m^T. \tag{14}$$

In determining the factors in the O-GARCH model, we just use the unconditional covariance matrix of the return series, and the conditional information is not considered. A requirement for the validity of O-GARCH is that principal components are also conditionally uncorrelated. However, as the conditional covariance matrix of the financial return series is time-varying, the unconditional uncorrelatedness between principal components does not guarantee their conditional uncorrelatedness. Recently, GO-GARCH was proposed as a generalization of O-GARCH (van der Weide, 2002; Lanne & Saikkonen, 2005). In GO-GARCH, the orthogonality constraint of the factor loading matrix is relaxed.

3.3 Estimation of GO-GARCH Models by Mutual Information Minimization

In GO-GARCH, the error series $\boldsymbol{\epsilon}_t$ (which are the zero-mean version of the return series $\mathbf{r}_t$) are assumed to be generated from some latent uncorrelated factors $\mathbf{y}_t = (y_{1t}, ..., y_{mt})^T$ by linear transformation (van der Weide, 2002):

$$\boldsymbol{\epsilon}_t = \mathbf{A}\mathbf{y}_t, \tag{15}$$

where $m \leq n$,[iii] y_{it} have unit variance, and the conditional covariance matrix of $\mathbf{y}_t$ is assumed to be diagonal:

$$\boldsymbol{\Sigma}_t = diag\{h_{y_{1t}}, ..., h_{y_{mt}}\}. \tag{16}$$

Each factor is described as a GARCH(1,1) process:[iv]

$$h_{y_{it}} = (1 - \alpha_i - \beta_i) + \alpha_i y_{i,t-1}^2 + \beta_i h_{y_{i,t-1}}. \tag{17}$$

iii. In van der Weide (2002) and Lanne and Saikkonen (2005) it is assumed that $m = n$. Here the case that $m < n$ can also be included.

iv. The model proposed by Lanne and Saikkonen (2005) is slightly different—they assume that there exist some factors with a constant volatility, rather than the time-varying one.

The conditional covariance matrix of $\boldsymbol{\epsilon}_t$ is therefore given as

$$\mathbf{H}_t = \mathbf{A}\boldsymbol{\Sigma}_t\mathbf{A}^T. \tag{18}$$

This model was estimated by QML estimation (van der Weide, 2002) or ML estimation with standardized residuals modeled by the mixture of Gaussians (Lanne & Saikkonen, 2005). But the latter is computationally quite intensive. Here we investigate the GO-GARCH model from the information-theoretic viewpoint. This point of view explicitly relates the likelihood function to the contemporaneous and temporal statistical dependence of standardized residuals. It allows adaptive estimation of the distribution of the standardized residuals, as Lanne and Saikkonen (2005) does. Furthermore, it inspires our proposal of the factor GARCH models in the framework of GO-GARCH, which will be given in Section 4.

Denote the standardized residuals by z_{it}, i.e. $z_{it} = y_{it}h_{y_{it}}^{-1/2}$. Let $\mathbf{z}_t = (z_{1t}, ..., z_{mt})^T$. In the GO-GARCH model, the observed error series generated by $\boldsymbol{\epsilon}_t = \mathbf{A}\mathbf{y}_t = \mathbf{A}\boldsymbol{\Sigma}_t^{1/2}\mathbf{z}_t$. In order to find the factors y_{it} uniquely from $\boldsymbol{\epsilon}_t$ (the permutation of y_{it} does not matter), we assume that the mixing matrix $\mathbf{A}$ is of full column rank. The factors y_{it} can then be recovered from $\boldsymbol{\epsilon}_t$ by the following linear transformation:

$$\mathbf{y}_t = \mathbf{W}\boldsymbol{\epsilon}_t, \tag{19}$$

where $\mathbf{W}$ is a $m \times n$ matrix. $\mathbf{A}$ can be constructed based on the estimated $\mathbf{W}$. If $m = n$, $\mathbf{A} = \mathbf{W}^{-1}$; otherwise $\mathbf{A}$ is the pseudo-inverse of $\mathbf{W}$, i.e. $\mathbf{A} = \mathbf{W}^T(\mathbf{W}\mathbf{W}^T)^{-1}$.

One way of selecting a small number of factors is to do dimension reduction by applying PCA on $\boldsymbol{\epsilon}_t$, and to extract m factors in the space spanned by m principal components of $\boldsymbol{\epsilon}_t$. PCA can also be used to whiten the data. Denote the whitening matrix by $\mathbf{V}$, which can be obtained by eigenvalue decomposition (EVD) of the covariance matrix $E\{\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^T\}$: $\mathbf{V} = \mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^T$, where $\mathbf{D}$ is the diagonal matrix of the m largest eigenvalues of $E\{\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^T\}$, and the columns of the matrix $\mathbf{E}$ are the corresponding unit-norm eigenvectors. Denote the whitened errors by $\tilde{\boldsymbol{\epsilon}_t}$, i.e.

$$\tilde{\boldsymbol{\epsilon}_t} = \mathbf{V}\boldsymbol{\epsilon}_t. \tag{20}$$

Clearly $E\{\tilde{\boldsymbol{\epsilon}_t}\tilde{\boldsymbol{\epsilon}_t}^T\} = \mathbf{I}_m$. $\mathbf{W}$ in Eq. 19 can be decomposed as $\mathbf{W} = \widetilde{\mathbf{W}}\mathbf{V}$. Since $\mathbf{I}_m = E\{\mathbf{y}_t\mathbf{y}_t^T\} = \widetilde{\mathbf{W}}E\{\tilde{\boldsymbol{\epsilon}_t}\tilde{\boldsymbol{\epsilon}_t}^T\}\widetilde{\mathbf{W}}^T = \widetilde{\mathbf{W}}\widetilde{\mathbf{W}}^T$, $\widetilde{\mathbf{W}}$ is an orthogonal matrix. Now the parameters to be estimated are the orthogonal matrix $\widetilde{\mathbf{W}}$ and the parameters in univariate GARCH(1,1) models $\{\alpha_i, \beta_i\}_{i=1}^m$.

In Section 2 we have seen that univariate GARCH models actually aim to capture the temporal dependence in a univariate return series. Moreover, multivariate GARCH models aim to capture not only the temporal dependence in each return series, but also the contemporaneous dependence between different return series. That is, multivariate GARCH makes z_{it} contemporaneously and temporally as independent as possible. Let $\mathbf{z}^*$ be a collection of standardized residuals:

$$\mathbf{z}^* = (z_{1t}, \cdots, z_{m,t}, z_{1,t-1}, \cdots, z_{m,t-1}, \cdots\cdots, z_{1,t-N}, \cdots, z_{m,t-N})^T,$$

where N is sufficiently large, and similarly for $\tilde{\boldsymbol{\epsilon}}^*$, which is the collection of whitened errors defined in Eq. 20. To measure the statistical temporal and contemporaneous dependence in the vector process $\mathcal{Z} \triangleq \{\mathbf{z}_t\}$, we define the the mutual information rate of $\mathcal{Z}$ as $\mathcal{I}(\mathcal{Z}) = \lim_{N\to\infty} \frac{1}{N+1} \cdot I(\mathbf{z}^*)$. Multivariate GARCH can be estimated by minimizing $\mathcal{I}(\mathcal{Z})$.

Denote the mapping from $\tilde{\boldsymbol{\epsilon}}^*$ to $\mathbf{z}^*$ by $\mathcal{F}$, i.e. $\mathbf{z}^* = \mathcal{F}(\tilde{\boldsymbol{\epsilon}}^*; \boldsymbol{\phi})$, where $\boldsymbol{\phi}$ denotes the parameter set. One can calculate the Jacobian matrix of $\mathcal{F}$:

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_t \cdot \widetilde{\mathbf{W}} & \cdots & \cdots & \cdots \\ \mathbf{0} & \mathbf{J}_{t-1} \cdot \widetilde{\mathbf{W}} & \cdots & \cdots \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{J}_{t-N} \cdot \widetilde{\mathbf{W}} \end{bmatrix},$$

where

$$\mathbf{J}_k = \begin{bmatrix} h_{y_{1k}}^{-1/2} & 0 & \cdots & 0 \\ 0 & h_{y_{2k}}^{-1/2} & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & h_{y_{mk}}^{-1/2} \end{bmatrix}.$$

Since $p_{\mathbf{z}^*}(\mathbf{z}^*) = \frac{p_{\tilde{\boldsymbol{\epsilon}}^*}(\tilde{\boldsymbol{\epsilon}}^*)}{|\mathbf{J}|}$ and $|\mathbf{J}| = \Pi_{j=1}^m \Pi_{i=0}^N h_{j,t-i}^{-1/2}$, we have

$$\begin{aligned} \mathcal{I}(\mathcal{Z}) &= \lim_{N\to\infty} \frac{I(\mathbf{z}^*)}{N+1} = \lim_{N\to\infty} \frac{1}{N+1} \Big[\sum_{j=1}^m \sum_{i=0}^N H(z_{j,t-i}) - H(\mathbf{z}^*) \Big] \\ &= \lim_{N\to\infty} \frac{1}{N+1} \Big\{ \sum_{i=0}^N \sum_{j=1}^m H(z_{j,t-i}) - [H(\tilde{\boldsymbol{\epsilon}}^*) + E \log |\mathbf{J}|] \Big\} \\ &= \sum_{j=1}^m \Big[H(z_{jt}) + \frac{1}{2} E \log h_{y_{jt}} \Big] - \lim_{N\to\infty} \frac{1}{N+1} H(\tilde{\boldsymbol{\epsilon}}^*). \end{aligned} \quad (21)$$

As the last term in the above equation does not depend on $\boldsymbol{\phi}$, minimizing $\mathcal{I}(\mathcal{Z})$ is equivalent to maximizing the following function

$$\begin{aligned} l(\boldsymbol{\phi}) &= \sum_{j=1}^m \Big[-H(z_{jt}) - \frac{1}{2} E \log h_{y_{jt}} \Big] \\ &= \sum_{j=1}^m E \Big[\log p_{z_j}(z_{jt}) - \frac{1}{2} \log h_{y_{jt}} \Big]. \end{aligned} \quad (22)$$

Below we give the likelihood function for $\tilde{\boldsymbol{\epsilon}}_t$. Since z_{it} are assumed to be mutually independent, we have $p_{\mathbf{z}}(\mathbf{z}_t) = \prod_{i=1}^{m} p_{z_i}(z_{it})$. Since $\tilde{\boldsymbol{\epsilon}}_t = \widetilde{\mathbf{W}}^T \mathbf{y}_t = \widetilde{\mathbf{W}}^T \boldsymbol{\Sigma}_t^{1/2} \mathbf{z}_t$, we have $p_{\tilde{\boldsymbol{\epsilon}}}(\tilde{\boldsymbol{\epsilon}}_t) = \frac{p_{\mathbf{z}}(\mathbf{z}_t)}{|\widetilde{\mathbf{W}}^T \boldsymbol{\Sigma}_t^{1/2}|}$. The log-likelihood function of $\tilde{\boldsymbol{\epsilon}}_t$ is therefore

$$\log p_{\tilde{\boldsymbol{\epsilon}}}(\tilde{\boldsymbol{\epsilon}}_t) = \sum_{t=1}^{T} \sum_{i=1}^{m} \Big[\log p_{z_i}(z_{it}) - \frac{1}{2} \log h_{y_{it}} \Big]. \tag{23}$$

Clearly the likelihood function (Eq. 23) is the same as Eq. 22 (with the constant factor ignored), which is derived by mutual information minimization.

In particular, if we simply adopt the standard Gaussian density for p_{z_j}, regardless of the true distribution of standardized residuals, the above objective function (Eq. 23) turns to be the quasi log-likelihood function in van der Weide (2002). When some factors have constant volatilities, i.e. $h_{j,t} = 1$ for some j, the above objective function is reduced to that in Lanne and Saikkonen (2005).

Estimation of GO-GARCH with QML is time-consuming, and can be problematic, especially when the data dimension is high, for two reasons. First, in the estimation of GO-GARCH, the update of $\mathbf{W}$ (or equivalently $\widetilde{\mathbf{W}}$) and $\{\alpha_i, \beta_i\}_{i=1}^m$ interferes with each other, which may cause estimation difficulties. Second, the quasi likelihood is quite flat in the neighborhood of its optimum and ill-conditioned for multivariate GARCH models (Jerez et al., 2000). As an illustration, let us consider the behavior of QML in a degenerate case, where all factors have a constant volatility. The objective function of GO-GARCH turns to be that of the independent component analysis (ICA) model, which is discussed in Section 4.1. This model is not identifiable by using QML, while it can be identified by using the true ML (Hyvärinen et al., 2001).

Alternatively, we can estimate the factors and the univariate GARCH models in two separate steps. We can find the statistical characteristics of the factors extracted in GO-GARCH, and determine $\mathbf{W}$ and the factors by optimizing some statistical criterion regarding the factors in the first step. In the second step we simply fit a univariate GARCH model, which may be a complex extension of the standard GARCH (Eq. 1–3), such as the exponential GARCH (EGARCH) (Nelson, 1991) and the threshold GARCH (TGARCH), to these factors. Finally the conditional covariance matrix is constructed according to Eq. 18. In this way parameters involved in both steps can be estimated reliably and fast.

4. Three Factor Models in the Framework of GO-GARCH

Now we present three factor GARCH models in the framework of GO-GARCH by analyzing the property of the factors in GO-GARCH. The proposed models are the independent-factor

GARCH (IF-GARCH) model, the best-factor GARCH (BF-GARCH) model, and the conditional-decorrelation GARCH (CD-GARCH) model. Generally speaking, these models have the following features:

- They can all be estimated in a convenient way, because $\mathbf{W}$ and parameters in univariate GARCH models are estimated separately. Hence they are more suitable for high-dimensional data. Reliability and low computation in the estimation of these models also make it possible to couple these models with others to achieve better flexibility and performance.

- Factors extracted in these models have clear statistical properties, so they may also be useful in other financial analysis scenarios. Apart from the unconditional uncorrelatedness constraint, different criteria are used to determine $\mathbf{W}$ in these models. In IF-GARCH, $\mathbf{W}$ is determined by making the factors y_{it} as statistically independent as possible; in BF-GARCH, each factor has the largest autocorrelation in its squared values such that its volatility is forecasted well by univariate GARCH; CD-GARCH produces the factors which are conditionally as uncorrelated as possible.

Below we present these models by discussing the rationale behind them, the uniqueness of solution, together with the method for parameter estimation. The second step in the estimation of the proposed factor GARCH models is simply to estimate univariate GARCH for each factor, which is very easy. So we focus on the first step, in which $\mathbf{W}$ and factors are estimated.

4.1 Independent-Factor GARCH Model

4.1.1 Rationale

If GO-GARCH is specified and estimated correctly, the standardized residuals z_{it} should be contemporaneously and temporally as independent as possible. On the other hand, if z_{it} are contemporaneously independent, the factors y_{it} must be mutually independent, since y_{it} is determined by z_{ik} $(k = t, t-1, ...)$ and does not depend on $z_{jk}, j \neq i$. And as discussed below, under weak assumptions, which usually hold for financial return series, $\mathbf{W}$ and y_{it} can be determined by making the factors y_{it} as mutually independent as possible. This is exactly the objective of the independent component analysis (ICA) technique (Hyvärinen et al., 2001).

ICA, as a generative model, is a statistical technique for revealing hidden factors that underlie the observed signals. In ICA, we only have some observable variables $\epsilon_1, ..., \epsilon_m$, which are assumed to be linear mixtures of some unknown statistically independent source variables

$s_1, ..., s_m$ with an unknown mixing matrix $\mathbf{A}$. Let $\boldsymbol{\epsilon} = (\epsilon_1, ..., \epsilon_m)^T$ and $\mathbf{s} = (s_1, ..., s_m)^T$. The latent data generation procedure is $\boldsymbol{\epsilon} = \mathbf{A}\mathbf{s}$. Under the assumption that $\mathbf{A}$ is of full column rank and that at most one of the sources s_i is normal,[v] s_i and $\mathbf{A}$ can be recovered by ICA. ICA of the data $\boldsymbol{\epsilon}$ aims at finding the linear transform $\mathbf{y} = \mathbf{W}\boldsymbol{\epsilon}$ with components of $\mathbf{y}$ as independent as possible, such that $\mathbf{y}$ provides an estimate of $\mathbf{s}$ (or equivalently $\mathbf{WA}$ is a generalized permutation matrix). Note that the scaling and permutation of y_i can be arbitrary, and in practice the variance of y_i is usually set to one.

ICA has been applied in some finance scenarios (Back, 1997; Kiviluoto & Oja, 1998; Cha & Chan, 2000). In Chan and Cha (2001), ICA is used to construct factor models in finance, and in particular, several factor selection criteria are given.

4.1.2 Parameter Estimation

Various algorithms from different points of view have been developed for ICA (Hyvärinen et al., 2001). For example, FastICA (Hyvärinen & Oja, 1997) is a fixed-point algorithm for maximizing the non-Gaussianity of y_{it}; JADE (Cardoso & Souloumiac, 1993) is based on joint diagonalization of fourth order cross-cumulant matrices; the Infomax algorithm (Bell & Sejnowski, 1995) is based on the information-maximization principle. Due to its simplicity and efficiency, we adopt the FastICA algorithm in our experiments.

4.1.3 Remark

IF-GARCH exploits ICA to enforce mutual independence between factors, since the standardized residuals are approximately contemporaneously independent. In fact, for GO-GARCH, the standardized residuals are just as independent as possible, and they may not be completely independent. Consequently, it is possible that IF-GARCH does not give the optimal estimate for conditional covariances. ICA just exploits contemporaneous information of factors, while the methods discussed below also take into consideration temporal information of the factors.

v. Note that uncorrelated and jointly normal variables are mutually independent, and that the uncorrelatedness does not change after any orthogonal transformation. Consequently, in order to recover the original sources uniquely (with scaling and permutation indeterminacies ignored) by ICA, one has to assume that at most one of the sources is normal (Hyvärinen et al., 2001).

4.2 Best-Factor GARCH Model

4.2.1 Rationale

From Eq. 22 and Eq. 12, we can see that the objective function of GO-GARCH is actually the sum of the likelihood of fitting each factor y_{it} with univariate GARCH . In order to maximize Eq. 22, $\mathbf{W}$ should maximize the likelihood of fitting univariate GARCH to each factor, which can be approximately achieved by maximizing the autocorrelation in squared values of each factor y_{it}, as shown below.

For a GARCH series, the maximum likelihood attained by univariate GARCH is related to the autocorrelation in its squared values, given that the GARCH model is specified and estimated correctly, as explained below. Looking at Eq. 11 and Eq. 12, we can see that the likelihood, $l_T(\boldsymbol{\theta})$, is given by

$$\begin{aligned} l_T(\boldsymbol{\theta}) &= T \cdot [-\mathcal{H}(\mathcal{R}) - \mathcal{I}(\mathcal{Z})] \\ &= T \cdot [\mathcal{I}(\mathcal{R}) - \mathcal{I}(\mathcal{Z}] - T \cdot H(r_t). \end{aligned} \tag{24}$$

Recall that $\mathcal{I}(\mathcal{Z})$ and $\mathcal{I}(\mathcal{R})$ denote the mutual information rate in the processes $\{z_t\}$ and $\{r_t\}$, respectively. The last term in Eq. 24 is determined by the density of the return and is approximately a constant. If the GARCH model is specified and estimated correctly, the standardized residual is almost serially independent and consequently $\mathcal{I}(\mathcal{Z}) \approx 0$. From Eq. 24 we can see that the larger the autocorrelation in squared returns, the larger $\mathcal{I}(\mathcal{R})$, and as a consequence, the likelihood $l_T(\boldsymbol{\theta})$ is higher.

Now let us investigate how the linear transformation $\mathbf{W}$ changes the autocorrelation in squared values of each series. Roughly speaking, the mixture of independent GARCH processes tends to lose the GARCH effect, according to the following theorem.

Theorem 1 *Suppose that $g_{1t}, ..., g_{kt}$ are k zero-mean independent GARCH processes with no linear time-correlations, and that they have finite kurtosis.*[vi] *Denote their variances by $\sigma_1^2, ..., \sigma_k^2$ and let $\sigma^2 = \sigma_1^2 + \cdots + \sigma_k^2$. Denote by x_t the sum of $g_{it}, i = 1, ..., k$. Under the condition that $\frac{\sigma_1^4 + \cdots + \sigma_k^4}{\sigma^4} \to 0$ when $k \to \infty$, we have:*

1. *The autocovariance in squared values of $\frac{x_t}{std(x_t)}$ tends to vanish when $k \to \infty$.*
2. *The autocorrelation in squared values of x_t tends to vanish when $k \to \infty$.*

vi. Here we use the following definitions. Kurtosis of the zero-mean variable x is defined by $\text{kurt}(x) = \frac{E\{x^4\}}{E[x^2]^2}$. The excess kurtosis of x is $\tilde{\kappa}(x) = \text{kurt}(x) - 3$.

Proof: see Appendix.

In fact the condition that $\frac{\sigma_1^4+\cdots+\sigma_k^4}{\sigma^4} \to 0$ when $k \to \infty$ is weak. It is used just to exclude the case that most of σ_i^2 are very small. In an extreme case, one can see that when all σ_i^2 are equal, $\frac{\sigma_1^4+\cdots+\sigma_k^4}{\sigma^4} = \frac{1}{k}$ and the condition is obviously satisfied. Interestingly, this theorem can be considered as a counterpart of the central limit theorem regarding temporal information. Hyvärinen (2001) presented the uniqueness of the solution to $\widetilde{\mathbf{W}}$ by maximizing the autocovariance of the squared values of each factor (of unit variance) when the latent factors have autocorrelation in squared values and are independent.

4.2.2 Parameter Estimation

Hyvärinen (2001) gives the fixed-point update rule for $\tilde{\mathbf{w}}_i^T$, the i-th row of $\widetilde{\mathbf{W}}$, with the fourth-order cumulant (Eq. 34) as the objective function under the constraint var$(y_{it}) = 1$. We prefer not to take it account the last term in Eq. 34, which is related to the autocorrelation of y_{it}. So we aim to estimate the i-th factor y_{it} by maximizing cov$(y_{it}^2, y_{i,t-\tau}^2)$, subject to var$(y_{it}) = 1$ and the uncorrelatedness among estimated factors. This is equivalent to maximizing cov$(y_{it}^2, y_{i,t-\tau}^2)$ under the conditions $||\tilde{\mathbf{w}}_i|| = 1$ and $\tilde{\mathbf{w}}_i^T \tilde{\mathbf{w}}_j = 0$ (for $j \neq i$). One can find the learning rule by deriving the gradient of cov$(y_{it}^2, y_{i,t-\tau}^2)$ w.r.t. $\tilde{\mathbf{w}}_i$:

$$\triangle \tilde{\mathbf{w}}_i \propto E\{\tilde{\boldsymbol{\epsilon}}_t \tilde{\mathbf{w}}_i^T \tilde{\boldsymbol{\epsilon}}_t (\tilde{\mathbf{w}}_i^T \tilde{\boldsymbol{\epsilon}}_{t-\tau})^2\} + E\{\tilde{\boldsymbol{\epsilon}}_{t-\tau} \tilde{\mathbf{w}}_i^T \tilde{\boldsymbol{\epsilon}}_{t-\tau} (\tilde{\mathbf{w}}_i^T \tilde{\boldsymbol{\epsilon}}_t)^2\}. \tag{25}$$

In each iteration the lag τ is randomly chosen between 1 and 7 with probability $1/3$ for 1. If $i > 1$, after each iteration of Eq. 25, $\tilde{\mathbf{w}}_i$ is made orthogonal to $\tilde{\mathbf{w}}_p, p = 1, ...i-1$, by the Gram-Schmidt like procedure:

$$\tilde{\mathbf{w}}_i \longleftarrow \tilde{\mathbf{w}}_i - \sum_{p=1}^{i-1} (\tilde{\mathbf{w}}_i^T \tilde{\mathbf{w}}_p) \tilde{\mathbf{w}}_p.$$

$\tilde{\mathbf{w}}_i$ is then normalized to unit length: $\tilde{\mathbf{w}}_i \longleftarrow \frac{\tilde{\mathbf{w}}_i}{||\tilde{\mathbf{w}}_i||}$.

Alternatively we can update all rows of $\widetilde{\mathbf{W}}$ in parallel. To do that, we need to update all $\tilde{\mathbf{w}}_i$ according to Eq. 25, followed by symmetric orthogonalization of $\widetilde{\mathbf{W}}$, until convergence. The symmetric orthogonalization of $\widetilde{\mathbf{W}}$ is accomplished by

$$\widetilde{\mathbf{W}} \longleftarrow (\widetilde{\mathbf{W}}\widetilde{\mathbf{W}}^T)^{-1/2} \widetilde{\mathbf{W}}. \tag{26}$$

where $(\widetilde{\mathbf{W}}\widetilde{\mathbf{W}}^T)^{-1/2}$ is obtained from the EVD decomposition of $\widetilde{\mathbf{W}}\widetilde{\mathbf{W}}^T$. Let the EVD decomposition of $\widetilde{\mathbf{W}}\widetilde{\mathbf{W}}^T$ be $\widetilde{\mathbf{W}}\widetilde{\mathbf{W}}^T = \mathbf{E}\mathbf{D}\mathbf{E}^T$. We have $(\widetilde{\mathbf{W}}\widetilde{\mathbf{W}}^T)^{-1/2} = \mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^T$.

4.3 Conditional-Decorrelation GARCH Model

4.3.1 Rationale

In GO-GARCH it is assumed that the factors y_{it} are conditionally uncorrelated such that the conditional covariance matrix of $\mathbf{y}_t$ is diagonal (see Eq. 16). But we just explicitly enforce unconditional uncorrelatedness between factors, and certainly unconditional uncorrelatedness can not guarantee conditional uncorrelatedness. It may cause large error in the estimated covariance matrix if we simply regard unconditional uncorrelatedness as conditional uncorrelatedness. In order to reduce the error in modelling the conditional covariance matrix, it is very natural to enforce conditional decorrelation between factors.

In Matsuoka et al. (1995) it was shown that if the latent factors are conditionally uncorrelated and their local variances fluctuate somewhat independently of each other,[vii] the factors (and the matrix $\mathbf{W}$) can be determined uniquely except for the trivial scaling and permutation indeterminacies by making y_{it} conditionally uncorrelated. So one can estimate the factors by enforcing conditional decorrelation between factors.

4.3.2 Parameter Estimation

Matsuoka et al. (1995) gave an algorithm to estimate y_{it} and $\widetilde{\mathbf{W}}$ by achieving conditional decorrelation between y_{it}. A simpler version was given in Hyvärinen et al., 2001, chap. 18. But in these algorithms, we need to re-calculate the local covariances of factors after each iteration, which causes high computational load.

In fact, in order to achieve conditional uncorrelatedness between factors, after whitening of the data, the orthogonal matrix $\widetilde{\mathbf{W}}$ should make all local covariance matrices of the whitened return series jointly as diagonal as possible. Therefore, after estimating a series of local covariance matrices, $\widetilde{\mathbf{W}}$ can be estimated by applying simultaneous diagonalization to these matrices. In our experiments, we adopt this method. We use the simultaneous diagonalization method given in Cardoso and Souloumiac (1996), and the local covariance is estimated using the exponentially weighted moving average (EWMA) with the smoothing constant $\lambda = 0.90$.

---

vii. Mathematically speaking, the condition, which states that the local variances fluctuate somewhat independently of each other, means that for every pair of the factors, the ratio of their local variances is not constant over time.

4.3.3 Remark

It should be mentioned that recently, an approach for modelling multivariate volatilities via conditional uncorrelated components (CUC's) was proposed by Fan et al. (2005). The CUC's in their approach are actually the same as the conditionally uncorrelated factors in our CD-GARCH model. In addition, as they exploit the extended GARCH model, which has a more general setting than conventional GARCH, to model the volatility of each CUC, their approach is more flexible than CD-GARCH. The consistency of the estimate of $\mathbf{W}$ is also proved in their work. In this paper we discuss CD-GARCH from a different viewpoint. It is proposed as one of the simple factor GARCH models and compared to others extensively. Moreover, in order to improve the model flexibility, one can combine the factor model with the DCC model, as shown in Section 5.

4.4 Discussion

Now we give a summary and comments on the models proposed above. The three models use different criteria to find the factors and $\mathbf{W}$. Under some conditions, we can unify these models. In fact, Independent-factor GARCH, best-factor GARCH, and conditional-decorrelation GARCH give the same result if the following conditions are satisfied:

1. The factors y_{it} are statistically independent of each other.
2. At most one of y_{it} is normally distributed.
3. All y_{it} have autocorrelation in their squared values and are temporally uncorrelated.
4. For every pair of y_{it}, the ratio of their local variances is not constant over time.

Under the above conditions, each of the three models can give a unique result for the factors (except for the permutation and scaling arbitrariness of the outputs), which is an estimate of the latent factors generating the return series. Therefore the estimated factors and the matrix $\mathbf{W}$ must be the same for the three models (with permutation indeterminacy ignored).

We should emphasize that in practice the above conditions may not exactly hold. Consequently, these models may produce different estimates for $\mathbf{W}$, since they focus on information of different aspects and exploit different objective functions. The choice of the model depends on which criterion is most consistent with the characteristics of financial data. We know that financial return series exhibit the volatility clustering phenomenon and their covariances change over time. In order to capture the nonlinear dependency between outputs, ICA algorithms use higher-order statistics explicitly or implicitly, which may be sensitive to the chosen data period of the

	IF-GARCH	BF-GARCH	CD-GARCH
Criterion for estimating factors	Factors are statistically as independent as possible (at most one factor is normal.)	Each factor has the largest autocorrelation in squared values	Factors are conditionally as uncorrelated as possible
Estimation procedure	ICA (FastICA given by Hyvärinen and Oja (1997) is used) for estimating $\mathbf{W}$; Eq. 18 for constructing $\mathbf{H}_t$	Eq. 25 for estimating $\widetilde{\mathbf{W}}$; Eq. 18 for constructing $\mathbf{H}_t$	Simultaneous diagonalization (Cardoso & Souloumiac, 1996) for estimating $\widetilde{\mathbf{W}}$; Eq. 18 for constructing $\mathbf{H}_t$
Computation	involve low computation		
Remark	provide the same result under certain conditions		

Table 1: Summary of the three multivariate factor GARCH models.

nonstationary financial data and outliers. This may cause a problem in constructing independent factor models in finance. Autocorrelation in squared values of a series is also sensitive to outliers (Alexander, 2001, chap. 4), so the result of BF-GARCH may also be sensitive to outliers. In CD-GARCH, we do not exploit the information regarding the distributions of the factors. We just utilize the second-order statistics averaged for all time instants. Hence it should be more robust to outliers. Also recall that GO-GARCH explicitly assumes that the factors are conditionally uncorrelated. We therefore conjecture that CD-GARCH should describe the multivariate financial return series best.

All these three models are estimated easily, and the estimation involves low computation. In particular, IF-GARCH and CD-GARCH (based on simultaneous diagonalization of local covariances) are computationally very efficient. Note that in the estimation of BF-GARCH, in each iteration of the learning rule Eq. 25, the time lag τ is randomly chosen, which may slower the convergence a little. Table 1 summarizes the characteristics of these models. The behavior of these models will be experimentally studied in Section 7.

In Section 3.3 we claimed that one can use PCA as preprocessing to reduce the number of factors. In fact, even if we don't use PCA, it is still possible to reduce the number of factors, by analyzing the estimated $\mathbf{A}$. Denote the i-th cloumn of $\mathbf{A}$ by $\mathbf{a}_i$. Clearly $||\mathbf{a}_i||^2$ indicates the total variance that the i-th factor contributes to all return series. We can reduce the number of factors according to the order of $||\mathbf{a}_i||^2$. As a consequence, large covariance matrices can be approximately generated by only a small number of factors. If necessary, one may resort to other more complex criteria to determine the number of factors. For instance, Chan and Cha (2001) exploited the minimum description length (MDL) principle to do that.

5. Factor-DCC Model: Factor Model coupled with DCC

5.1 The DCC Model

The DCC model, given by Engle (2002), is a generalization of CCC model (Bollerslev, 1990). The CCC model assumes that the time-varying of conditional covariances is caused by the time-varying of the conditional variance of each return series:

$$\mathbf{H}_t = \mathbf{D}_t \mathbf{R} \mathbf{D}_t, \tag{27}$$

where $\mathbf{D}_t = diag\{\sqrt{h_{it}}\}$ and h_{it} denotes the conditional variance of the i-th return.

The DCC model differs only in allowing $\mathbf{R}$ to be time-varying:

$$\mathbf{H}_t = \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t. \tag{28}$$

The conditional correlation matrix $\mathbf{R}_t$ is determined by another matrix $\mathbf{Q}_t$, which is used to model the dynamic correlation structure of the standardized residual $\mathbf{z}_t$ by the univariate GARCH(1,1) formulation:

$$\mathbf{Q}_t = (1-\alpha-\beta)\overline{\mathbf{Q}} + \alpha \mathbf{z}_{t-1}\mathbf{z}_{t-1}^T + \beta \mathbf{Q}_{t-1}, \tag{29}$$

where $\overline{\mathbf{Q}}$ is the unconditional covariance matrix of $\mathbf{z}_t$. The (i,j)-th entry of the conditional correlation matrix $\mathbf{R}_t$, denoted by $\rho_{i,j,t}$, is estimated as

$$\rho_{i,j,t} = \frac{q_{i,j,t}}{\sqrt{q_{i,i,t}q_{j,j,t}}}, \tag{30}$$

where $q_{i,j,t}$ is the (i,j)-th entry of $\mathbf{Q}_t$.

5.2 Why is DCC Embedded?

van der Weide (2002) demonstrated that GO-GARCH produces time-varying conditional correlation and that the conditional correlation is bounded. In practice, factor GARCH models may not fully capture the dynamic behavior of the conditional correlation of return series. Consequently, the factors may still have some remaining conditional correlation.

In order to model $\mathbf{H}_t$ more accurately, we no longer treat $\mathbf{\Sigma}_t$ as a diagonal matrix, as shown in Eq. 16, but model it with the DCC model. In this way the factor-DCC model is constructed, as an extension of the corresponding factor-GARCH model. Clearly DCC would provide a better estimate for $\mathbf{\Sigma}_t$ than the diagonal matrix, and as a consequence, the capacity of factor GARCH is extended by embedding DCC for modelling the conditional correlations between factors. In other words, the factor-DCC model should have better flexibility than the corresponding factor GARCH model. On the other hand, the performance of the extended model is also better than that of DCC, as explained below.

In the factor GARCH models we proposed, the factors are estimated very conveniently such that the recovery of factors can even be considered as a preprocessing step. After this step, the factors are expected to exhibit the following features.

- The conditional correlation matrix of $\mathbf{y}_t$ is close to the identity matrix and its variability is greatly reduced. This is because the factors (especially those extracted in IF-GARCH and CD-GARCH) are approximately conditionally uncorrelated.
- Compared to the original return series, the factors (especially those obtained in BF-GARCH) would exhibit higher autocorrelation in squared values and can be modeled better with univariate GARCH models.

According to these features, $\mathbf{\Sigma}_t$ should be modeled better by the DCC model (Engle, 2002) than $\mathbf{H}_t$. In other words, the performance of DCC is improved by incorporating the linear transformation stage $\mathbf{W}$.

5.3 Factor-DCC Model

Now we adopt the DCC model to describe the conditional covariance matrix between factors, i.e.

$$\mathbf{\Sigma}_t = \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t, \tag{31}$$

where $\mathbf{D}_t = diag\{\sqrt{h_{y_{it}}}\}$ and $\mathbf{R}_t$ denotes the conditional correlation matrix of $\mathbf{y}_t$. $\mathbf{R}_t$ is given by Eq. 29 and Eq. 30. The conditional covariance matrix of return series is given by Eq. 18. The structure of the factor-DCC is illustrated in Figure 1.

The factor-DCC model represents the conditional covariance matrix of the whitened data $\tilde{\boldsymbol{\epsilon}}_t$ as $\widetilde{\mathbf{H}}_t = \widetilde{\mathbf{W}}^T \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t \widetilde{\mathbf{W}}$. The quasi log-likelihood for $\tilde{\boldsymbol{\epsilon}}_t$ is

$$\begin{aligned} l_T &= -\frac{1}{2}\sum_{t=1}^{T}[m\log(2\pi) + \log|\widetilde{\mathbf{H}}_t| + \tilde{\boldsymbol{\epsilon}}_t^T \widetilde{\mathbf{H}}_t^{-1} \tilde{\boldsymbol{\epsilon}}_t] \\ &= -\frac{1}{2}\sum_{t=1}^{T}\Big\{\Big[m\log(2\pi) + \sum_{i=1}^{m}\Big(\log h_{y_{it}} + z_{it}^2\Big)\Big] \\ &\qquad + \Big[\log|\mathbf{R}_t| + \mathbf{z}_t^T \mathbf{R}_t^{-1} \mathbf{z}_t - \mathbf{z}_t^T \mathbf{z}_t\Big]\Big\}. \end{aligned}$$

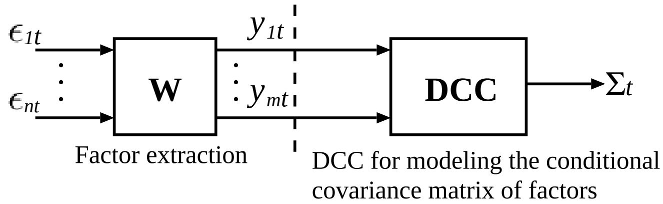

Figure 1: The factor-DCC model. The conditional covariance matrix $\mathbf{H}_t = \mathbf{A}\mathbf{\Sigma}_t\mathbf{A}^T$, where $\mathbf{\Sigma}_t$ is the conditional covariance matrix of factors y_{it} and $\mathbf{A}$ is the pseudo-inverse of $\mathbf{W}$. $\mathbf{\Sigma}_t$ is modeled by the DCC model.

Let $l_V = -\frac{1}{2}\sum_t[m\log(2\pi) + \sum_{i=1}^m(\log h_{y_{it}} + z_{it}^2)]$, and $l_C = -\frac{1}{2}\sum_t[\log|\mathbf{R}_t| + \mathbf{z}_t^T\mathbf{R}_t^{-1}\mathbf{z}_t - \mathbf{z}_t^T\mathbf{z}_t]$. The log-likelihood consists of l_V and l_C. Comparing l_V with Eq. 23, one can see that l_V is exactly the quasi log-likelihood of the GO-GARCH model. Also note that l_C is the objective function of the DCC model (Engle, 2002). Analogously to the two-step approach for estimation of the DCC model (Engle, 2002), we can also use the two-step approach to maximize l_T. In the first step it finds the parameters in the factor GARCH model which maximize l_V. With these estimated values as given, in the second step DCC models the conditional covariance between the factors, and the parameters in $\mathbf{R}_t$ are estimated by maximizing l_C.

Hence, by embedding the DCC model into our proposed factor models, we can get the IF-DCC, BD-DCC, and CD-DCC models. For the reasons given in Section 5.2, the performance of each extended model is expected to be better than that of the original factor model and that of the DCC model. Since CD-GARCH provides approximately conditionally uncorrelated factors and is expected to model the financial data well, we conjecture that CD-DCC model gives the best performance among the three extended models.

The factor-DCC model can be thought of as a hierarchical model: in the first level the strong conditional correlation between return series is expressed in terms of the weak conditional correlation between factors, using the linear transformation; in the second level, the weak conditional correlation between factors is calculated accurately by the DCC model. Consequently, the conditional correlations between original return series are modeled more accurately using this hierarchical structure.

Stock	Mean	Std.	Skewness	Excess kurtosis	First-order autocorrelation coeff. in squared returns	Box-Pierce LM (Lag p=5)
CHEUNG KONG (0001.HK)	0.00071	0.0223	0.644	7.92	0.165	48.13*
CLP HOLDINGS (0002.HK)	0.00052	0.0161	0.729	12.17	0.245	17.89*
HK & CHINA GAS (0003.HK)	0.00063	0.0178	0.388	7.47	0.233	28.64*
HSBC HOLDINGS (0005.HK)	0.00091	0.0176	0.337	11.90	0.390	24.75*
HK ELECTRIC (0006.HK)	0.00052	0.0162	0.212	9.66	0.327	25.49*
HANG LUNG GROUP (0010.HK)	0.00047	0.0214	0.161	5.98	0.191	1.26
HANG SENG BANK (0011.HK)	0.00081	0.0190	0.160	7.33	0.368	27.02*
HENDERSON LAND (0012.HK)	0.00073	0.0243	0.558	5.50	0.189	26.02*
HUTCHISON (0013.HK)	−0.00073	0.0226	0.498	6.91	0.222	25.11*
CATHAY PAC AIR (0293.HK)	0.00043	0.0240	0.168	5.38	0.128	18.53*

* Significant at 1% level.

Table 2: Return series used in the experiment and their statistics.

6. Experiment with Real Data: Empirical Study

In the experiment, 10 stock returns selected from the Hang Seng Index constitutes in the Hong Kong stock market are used. We use the daily dividend/split adjusted closing prices started from June 22th, 1990 to April 9th, 2004. Denoting the closing price by P_t, the return is calculated as

$$r_t = \frac{P_t - P_{t-1}}{P_{t-1}} \tag{32}$$

Each return series contains 3600 observations. For the few days when the price is not available, we use the simple linear interpolation to estimate the price. Table 2 gives some statistics of the daily returns of these stocks. From this table, we can see clearly that most stocks have the GARCH effect.

Here we compare 10 multivariate GARCH models, which are the O-GARCH (Alexander, 2000), DCC model (Engle, 2002), GO-GARCH with QML (van der Weide, 2002), IF-GARCH, BF-GARCH, CD-GARCH, and the factor-DCC models GO-DCC (GO-GARCH coupled with

DCC), IF-DCC, BF-DCC, CD-DCC. The DCC model is chosen for comparison for two reasons. First, other basic models, such as the the *vech* model (Bollerslev et al., 1988) and the BEKK model (Engle & Kroner, 1995), have too many parameters and thus are not suitable for modelling the conditional covariance matrix of high dimension. Second, the DCC model is a popular one and has been shown to perform well in a number of situations (Engle, 2002). All experiments in this paper are conducted using MATLAB. For the estimation of DCC, O-GARCH, and the univariate GARCH model, we use the UCSD_GARCH toolbox developed by Sheppard (2002). In the estimation of GO-GARCH, we use the MATLAB nonlinear constrained optimization toolbox to optimize the QML. Other MATLAB source codes used in experiments are available from http://www.cse.cuhk.edu.hk/∼kzhang/factor_garch.

We use 3000 points from all 3600 observations for training, and the other 600 are used for testing. We use two measures to evaluate the forecasting performance of these models. They are (log-)QML and the Value-at-Risk (VaR) performance. The (log-)QML has been used to evaluate the performance of GARCH models in Bollerslev et al. (1994):

$$\begin{aligned} \text{QML} &= \sum_{t=1}^{T} \log \mathcal{G}(\boldsymbol{\epsilon}_t; \mathbf{0}, \mathbf{H}_t) \\ &= -\frac{1}{2} \sum_{t=1}^{T} [n \log(2\pi) + \log |\mathbf{H}_t| + \boldsymbol{\epsilon}_t^T \mathbf{H}_t^{-1} \boldsymbol{\epsilon}_t]. \end{aligned}$$

We use the QML, rather than the true likelihood, because it is hard to specify the true conditional distribution of returns and the QML provides an approximate for the true maximum likelihood. Note that in Section. 3.3, we have shown the relationship between the likelihood attained and the statistical dependence in standardized residuals. For some given multivariate return series, the larger the likelihood function, the more independent the standardized residuals.

To save space, the estimated parameters for modelling the conditional covariance matrix of the 10 stocks are not reported here, as the number of parameters is too large. Later we will give the estimated parameter values together with the standard errors for 5-dimensional data. Table 3 presents the in-sample and out-of-sample QML values. Note that of course GO-GARCH by maximizing QML gives the highest in-sample QML among the factor models, since it explicitly maximizes the quasi likelihood. But it does not necessarily maximize the true likelihood. As we care about the generalization behavior of the models, we would evaluate the GARCH models by comparing their out-of-sample QML.

Among the first six GARCH models, GO-GARCH, CD-GARCH, IF-GARCH, and DCC give similar performance. If we take into account of the computational load, obviously CD-GARCH and IF-GARCH are plausible. Among the factor-DCC models, the CD-DCC model performs

Model	QML		Cpu time (seconds)
	In-sample (3000 observations)	Out-of sample (600 observations)	
O-GARCH	85,158.4	17,547.8	4.32
DCC model	85,569.6	**17,568.9**	120.8
GO-GARCH (QML)	**85,634.4**	**17,574.8**	$2 \times 10^4 *$
IF-GARCH	85,450.0	**17,571.5**	10.7
BF-GARCH	85,304.5	17,471.7	19.6
CD-GARCH	**85,597.6**	**17,566.4**	17.8
GO-DCC	**85,763.3**	*17,635.4*	$2 \times 10^4 *$
IF-DCC	85,669.2	**17,649.3**	132.9
BF-DCC	85,632.7	17,608.4	142.4
CD-DCC	**85,760.9**	**17,643.8**	140.6

Table 3: Comparison of the 10 models with 10 stock returns. Algorithms are implemented using MATLAB. *Estimation of GO-GARCH takes quite a long time, mainly for two reasons. The first is the estimation difficulties discussed in Section 3.3. The second is that we use the MATLAB nonlinear constrained optimization toolbox for parameter estimation without providing the gradient explicitly.

the best, with IF-DCC close behind. Each factor-DCC model gives better performance than the corresponding factor model as well as the DCC model. We can see that by incorporating the factor extraction procedure, the performance of DCC is greatly improved, with only slight increase in computational time. This also verifies the usefulness of factor models for constructing multivariate GARCH models. Note that the difference in QML for different models is not very large, since the quasi likelihood function is almost flat in the neighborhood of its optimum (Jerez et al., 2000).

Figure 2 shows the conditional correlation between the first and second return series estimated by O-GARCH, CD-GARCH, and the CD-DCC model, respectively. We can see that the conditional correlation estimated by CD-GARCH (Figure 2B) and that estimated by the CD-DCC model (Figure 2C) are very similar. Their trend lies in a lower level after November, 1999, while the result obtained by O-GARCH does not (Figure 2A). The conditional correlation between the CD-GARCH factors y_{1t} and y_{2t} estimated by the DCC model is also given in Figure 2 (D), from which we can see the correlation between y_{1t} and y_{2t} is comparatively small and oscillates around 0, which agree with our claims in Section 5.2.

In order to reduce the random effect and to compare these models more convincingly, we conduct five more experiments. In each experiment we randomly select five stocks from all of the 10 stocks used in the first experiment. Table 4 lists the parameter values estimated in the first experiment with five stocks. The standard errors are given in the parentheses. They are computed by the bootstrapping method described in Fan et al. (2005) (with 200 replications).[viii] We use the Amari performance index P_{err} (Cichocki & Amari, 2003) to measure the distance between two matrices $\mathbf{W}_1$ and $\mathbf{W}_2$. Let $p_{ij} = [\mathbf{W}_1\mathbf{W}_2^{-1}]_{ij}$. P_{err} is defined as

$$P_{err} = \frac{1}{m(m-1)} \sum_{i=1}^{m} \Big\{ \Big(\sum_{j=1}^{m} \frac{|p_{ij}|}{\max_k |p_{ik}|} - 1 \Big) + \Big(\sum_{j=1}^{m} \frac{|p_{ji}|}{\max_k |p_{ki}|} - 1 \Big) \Big\}. \qquad (33)$$

Particularly, this performance index measures how close $\mathbf{W}_1\mathbf{W}_2^{-1}$ is to the generalized permutation matrix. The smaller P_{err} is, the closer $\mathbf{W}_1\mathbf{W}_2^{-1}$ is to the generalized permutation matrix. Permutation and scaling of row of $\mathbf{W}_1$ and $\mathbf{W}_2$ do not affect this measure. From the parameter

viii. Fan et al. (2005) gave the bootstrapping procedure for computing standard errors (or confidence sets) of the parameters in factor GARCH models. For DCC and factor-DCC models, the procedure is quite similar. In the bootstrap sampling procedure, we just need to obtain the standardized residuals as $\mathbf{H}_t^{-1/2}\boldsymbol{\epsilon}_t$, to draw the standardized residuals by sampling with replacement, and to generate new observations according to the multivariate GARCH model under study with the estimated parameters.

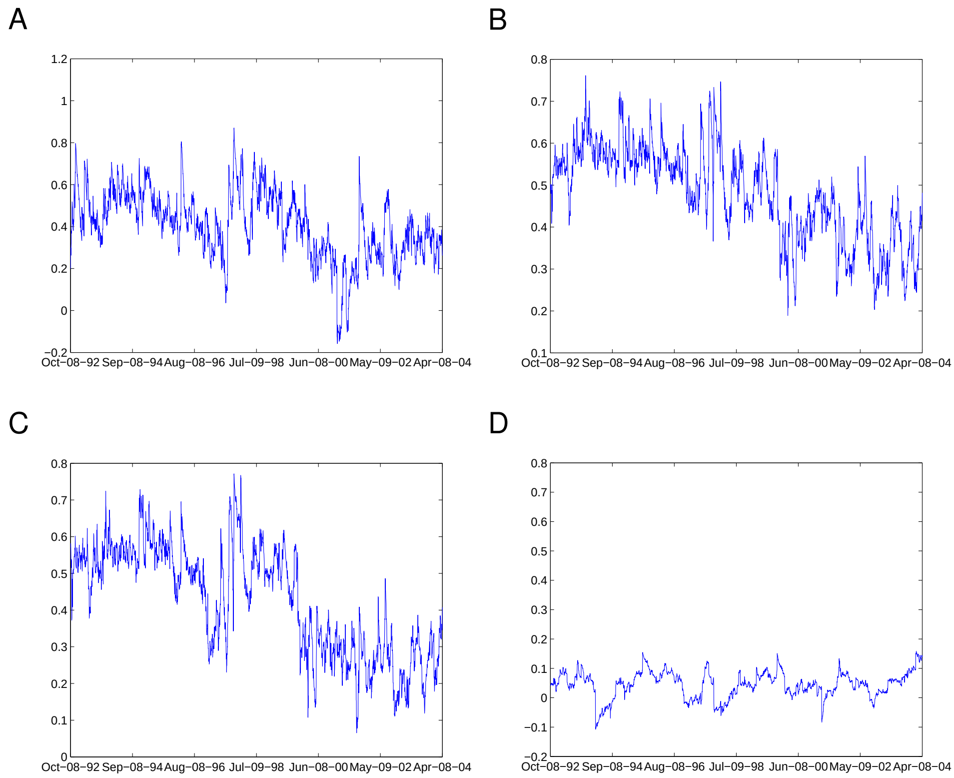

Figure 2: (A) The conditional correlation between ϵ_{1t} and ϵ_{2t} estimated by O-GARCH. (B) That estimated by CD-GARCH. (C) That estimated by the CD-DCC model. (D) The conditional correlation between the CD-GARCH factors y_{1t} and y_{2t} estimated by the DCC model.

values in Table 4, one can see that some factors may be fitted better by an integrated GARCH model. Table 5 gives the distance among $\hat{\mathbf{W}}$ in different models. We can see that $\hat{\mathbf{W}}$ in CD-GARCH is close to that in GO-GARCH and IF-GARCH, meaning that the criteria for factor extraction in these three models are consistent to a certain extent.

Table 6 shows the in-sample and out-of-sample QML for these five experiments. When summed over all five cases, we can see clearly that among the factor GARCH models and the DCC model, CD-GARCH and DCC are the best, and IF-GARCH is very close behind. For the factor-DCC models, the CD-DCC model always performs best, followed by the IF-DCC model. Note that the independent factor model is estimated very fast, so IF-GARCH and IF-DCC give comparatively good results with low computation.

Now let us compare these models by investigating the forecasting performance of the 5% VaR for portfolios with fixed weights. We consider an equally weighted portfolio and a hedge portfolio. For constructing the hedge portfolio of 10 stocks, the weight for the first five stocks is 0.2 and that for the others is -0.2. The weight vector for the hedge portfolio consisting of five stocks is $(0.5, 0.5, -0.3, -0.3, -0.4)^T$. In calculation of the VaR, the standardized residuals are assumed to follow the normal distribution.

Model	DCC	GO-GARCH	GO-DCC	IF-GARCH	IF-DCC	BF-GARCH	BF-DCC	CD-GARCH	CD-DCC
$\hat{\mathbf{W}}$	—	* (P_{err}: 0.0190)		** (P_{err}: 0.0262)		*** (P_{err}: 0.0661)		**** (P_{err}: 0.0245)	
$\hat{\alpha}_1$	0.1186 (0.0102)	0.0544 (0.0084)		0.0799 (0.0103)		0.0905 (0.0118)		0.0659 (0.0108)	
$\hat{\beta}_1$	0.8809 (0.0105)	0.9379 (0.0106)		0.9093 (0.0118)		0.9004 (0.0135)		0.9253 (0.0132)	
$\hat{\alpha}_2$	0.0851 (0.0111)	0.0437 (0.0073)		0.0312 (0.0083)		0.0293 (0.0038)		0.0294 (0.0055)	
$\hat{\beta}_2$	0.8942 (0.0162)	0.9513 (0.0094)		0.9639 (0.0125)		0.9673 (0.0047)		0.9665 (0.0075)	
$\hat{\alpha}_3$	0.1220 (0.0166)	0.0548 (0.0047)		0.0737 (0.0114)		0.0580 (0.0090)		0.0841 (0.0118)	
$\hat{\beta}_3$	0.8722 (0.0177)	0.9397 (0.0052)		0.9161 (0.0140)		0.9303 (0.0125)		0.9079 (0.0122)	
$\hat{\alpha}_4$	0.0994 (0.0095)	0.0277 (0.0042)		0.0549 (0.0069)		0.0424 (0.0046)		0.0547 (0.0050)	
$\hat{\beta}_4$	0.8902 (0.0101)	0.9691 (0.0059)		0.9360 (0.0092)		0.9529 (0.0057)		0.9394 (0.0056)	
$\hat{\alpha}_5$	0.0874 (0.0062)	0.0902 (0.0093)		0.0501 (0.0001)		0.0493 (0.0091)		0.0423 (0.0064)	
$\hat{\beta}_5$	0.8973 (0.0085)	0.9030 (0.0100)		0.9499 (0.0001)		0.9416 (0.0126)		0.9530 (0.0081)	
$\hat{\alpha}_{DCC}$	0.0130 (0.0008)	—	0.0101 (0.0011)	—	0.0129 (0.0008)	—	0.0118 (0.0014)	—	0.0113 (0.0009)
$\hat{\beta}_{DCC}$	0.9829 (0.00014)	—	0.9729 (0.0045)	—	0.9799 (0.0022)	—	0.9805 (0.0033)	—	0.9748 (0.0028)

$$*\hat{\mathbf{W}}_{GO} = \begin{bmatrix} -63.1309 & -1.3950 & 70.6358 & -3.9647 & 4.9817 \\ -7.6802 & -1.9740 & 5.8658 & 71.4124 & -74.8239 \\ 27.1660 & 13.1141 & 21.5107 & -33.2691 & -25.6074 \\ -20.3211 & 55.6060 & -14.3885 & 2.3388 & -17.4138 \\ 36.8576 & 6.8855 & 23.1277 & 7.4234 & -4.8296 \end{bmatrix} . **\hat{\mathbf{W}}_{IF} = \begin{bmatrix} -69.0825 & 2.6734 & 13.2643 & -8.0793 & 8.5702 \\ 9.2869 & -56.3526 & 12.2650 & -0.5828 & 16.7259 \\ -32.6966 & -0.9289 & 73.2649 & -0.6530 & 3.1127 \\ -12.8686 & -9.6418 & -22.2190 & -9.9944 & 61.3561 \\ -21.4619 & -6.3775 & -6.3761 & 78.2108 & -49.7823 \end{bmatrix}$$

.

$$***\hat{\mathbf{W}}_{BF} = \begin{bmatrix} 20.6899 & 8.3840 & 30.3021 & 13.3088 & -3.9912 \\ 22.8921 & -39.1274 & 26.5732 & 21.4409 & -32.8858 \\ -70.7030 & 8.6297 & 59.5367 & 16.2540 & -20.1137 \\ -23.9947 & -33.0844 & -14.6084 & 38.2933 & 20.4068 \\ 5.2096 & 23.3872 & -29.0875 & 62.5736 & -68.4641 \end{bmatrix} . ****\hat{\mathbf{W}}_{CD} = \begin{bmatrix} 41.0158 & 1.3977 & -75.2921 & 3.5068 & -4.4484 \\ -19.6151 & 53.9973 & -13.2330 & 2.8600 & -9.9846 \\ 61.9166 & 2.0958 & -4.8322 & 6.7886 & -6.0013 \\ -23.2127 & -19.8721 & -17.9987 & 30.4136 & 30.0664 \\ 10.6948 & -0.5534 & -6.0945 & -72.7414 & 74.4732 \end{bmatrix}$$

.

Table 4: Estimate of the parameters in different models for a combination of five return series. Numbers in the parentheses are the standard errors. In the DCC model, α_i and β_i are GARCH parameters for the return series, while in other models, they are GARCH parameters for the factors.

	$\hat{\mathbf{W}}_{GO}$	$\hat{\mathbf{W}}_{IF}$	$\hat{\mathbf{W}}_{BF}$	$\hat{\mathbf{W}}_{CD}$
$\hat{\mathbf{W}}_{GO}$				
$\hat{\mathbf{W}}_{IF}$	0.3531			
$\hat{\mathbf{W}}_{BF}$	0.4192	0.8347		
$\hat{\mathbf{W}}_{CD}$	**0.1818**	*0.2326*	0.5447	

Table 5: The distance among $\hat{\mathbf{W}}_{GO}$, $\hat{\mathbf{W}}_{IF}$, $\hat{\mathbf{W}}_{BF}$, and $\hat{\mathbf{W}}_{CD}$, measured by the Amari performance index P_{err}.

Two statistical tests are used to evaluate the VaR forecasting performance. They are the Dynamic quantile (DQ) test by Engle and Manganelli (2004) and the Kupiec LR test given in Kupiec (1995). The DQ test examines the independence of a HIT from past HITs, from the predicted VaR, or from any other variables, and HIT is defined as $I(r_t < \text{VaR}_\alpha) - \alpha$, where α is the VaR level. In our experiments, we use five lags for past HITs and the current VaR. The Kupiec LR test compares the empirical failure rate to its theoretical value. Tables 7 and 8 summarize the in-sample $\text{VaR}_{0.05}$ violations and the statistical test results, and Tables 9 and 10 give the out-of-sample results. The models are compared in terms of the number of rejections at 1% or 5% significant level.

One should be aware that the evaluation of GARCH models by examining the VaRs of a set of given portfolios may not be very reliable. Engle and Manganelli (2004) claimed that although GARCH might be useful for describing the evolution of volatility, it might provide an unsatisfactory approximation when applied to tail estimation. Moreover, the sensitivity of the VaR failure rates with respect to the distributional assumptions is found to be larger than that with respect to the parametric specification for the multivariate GARCH models (Rombouts & Verbeek, 2004).[ix] However, when summed over all the cases, we can see that factor-DCC models give the best $\text{VaR}_{0.05}$ prediction performance. And among the factor GARCH models, the result of O-GARCH is not as good as that of others. These findings are especially obvious in the out-of-sample case.

ix. Rombouts and Verbeek (2004) also proposed a semi-parametric model for the distribution of the standardized residuals and found that it can improve the VaR forecasting performance. However, this approach does not apply here since the data dimension is high.

Model	QML	Exp 1	Exp 2	Exp 3	Exp 4	Exp 5
O-GARCH	In-sample	39,852.0 (-177.6)	42,360.7 (-343.9)	40,944.4 (-271.8)	40,216.0 (-132.6)	41,140.4 (-568.7)
	Out-of-sample	*8,622.9 (-13.1)*	8,747.6 (-23.3)	**8,626.2 (0)**	**8,476.3 (-2.6)**	8,742.9 (-42.3)
DCC model	In-sample	**40,029.6 (0)**	**42,675.2 (-29.4)**	**41,216.2 (0)**	**40,342.9 (-5.7)**	**41,709.1 (0)**
	Out-of-sample	**8,636.0 (0)**	*8,759.7 (-11.2)*	**8,623.3 (-2.9)**	**8,476.7 (-2.2)**	*8,768.1 (-17.1)*
GO-GARCH	In-sample	**40,026.1 (-3.5)**	**42,704.6 (0)**	**41,185.8 (-30.4)**	**40,348.6 (0)**	**41,708.1 (-1.0)**
	Out-of-sample	8,616.9 (-19.1)	8,748.5 (-22.4)	8,594.8 (-31.4)	8,443.7 (-35.2)	8,759.8 (-25.4)
IF-GARCH	In-sample	*39,980.7 (-48.9)*	42,564.7 (-139.9)	41,100.1 (-116.1)	40,254.8 (-93.8)	41,602.1 (-107.0)
	Out-of-sample	**8,632.2 (-3.8)**	*8,757.8 (-13.1)*	**8,620.5 (-5.7)**	**8,478.9 (0)**	**8,785.2 (0)**
BF-GARCH	In-sample	*39,976.5 (-53.1)*	42,586.4 (-118.2)	40,987.0 (-229.2)	**40,330.9 (-17.7)**	41,503.8 (-205.3)
	Out-of-sample	8,618.9 (-17.1)	*8,757.4 (-13.5)*	8,561.7 (-64.5)	**8,472.3 (-6.6)**	8,730.6 (-54.6)
CD-GARCH	In-sample	*39,988.3 (-41.3)*	**42,647.5 (-57.1)**	*41,145.9 (-70.3)*	*40,292.8 (-55.8)*	**41,687.5 (-21.6)**
	Out-of-sample	**8,625.8 (-10.2)**	**8,770.9 (0)**	**8,623.6 (-2.6)**	**8,476.3 (-2.6)**	*8,769.9 (-15.3)*
GO-DCC	In-sample	**40,096.0 (0)**	**42,770.6 (0)**	**41,256.1 (0)**	**40,374.7 (-2.4)**	**41,801.2 (0)**
	Out-of-sample	8,649.9 (-7.1)	8,788.9 (-9.3)	*8,641.3 (-11.9)*	8,475.4 (-25.1)	**8,801.3 (-5.5)**
IF-DCC	In-sample	**40,075.0 (-21.0)**	42,705.2 (-65.4)	**41,241.6 (-14.5)**	40,363.1 (-14.0)	41,764.5 (-36.7)
	Out-of-sample	**8,657.0 (0)**	*8,791.3 (-6.9)*	**8,649.6 (-3.6)**	**8,500.5 (0)**	**8,806.8 (0)**
BF-DCC	In-sample	40,053.4 (-42.6)	*42,740.6 (-30.0)*	41,191.6 (-64.5)	**40,376.7 (-0.4)**	41,734.0 (-67.2)
	Out-of-sample	**8,653.8 (-3.2)**	**8,798.2 (0)**	8,629.1 (-24.1)	**8,498.4 (-2.1)**	8,790.5 (-16.3)
CD-DCC	In-sample	**40,070.8 (-25.2)**	**42,748.5 (-22.1)**	**41,251.1 (-5.0)**	**40,377.1 (0)**	**41,793.1 (-8.1)**
	Out-of-sample	**8,654.0 (-3.0)**	**8,797.0 (-1.2)**	**8,653.2 (0)**	**8,499.6 (-0.9)**	**8,803.5 (-3.3)**

Table 6: Quasi likelihood of the 10 models modelling the conditional covariance of five return series. The number in the parentheses is the difference between the corresponding value and the largest value in that group. Group 1 includes the first six models, and the other models are in group 2.

		O-GARCH	DCC	GO-GARCH	IF-GARCH	BF-GARCH	CD-GARCH	GO-DCC	IF-DCC	BF-DCC	CD-DCC
With 10 stocks	% of HITs	5.00	4.40	4.83	45.00	4.87	4.60	5.07	4.87	5.17	5.00
	DQ test (p-value)	*0.0180*	0.1047	0.238	0.0541	**0.0009**	0.1785	0.3902	0.1436	0.3037	0.3769
	Kupiec test (p-value)	1	0.124	0.6737	0.2015	0.7365	0.3085	0.8672	0.7365	0.6769	1
Exp 1 with 5 stocks	% of HITs	4.53	4.47	4.60	4.47	4.37	4.47	4.37	4.40	4.27	4.20
	DQ test (p-value)	0.8330	0.8302	0.6919	0.8617	0.6390	0.8687	0.4305	0.5007	0.4188	0.3280
	Kupiec test (p-value)	0.2337	0.1726	0.3085	0.1726	0.1041	0.1726	0.1041	0.1240	0.0590	*0.0389*
Exp 2 with 5 stocks	% of HITs	4.67	4.37	5.03	4.63	5.33	5.23	4.93	4.53	4.80	5.03
	DQ test (p-value)	*0.0120*	**0.0015**	0.0576	**0.0078**	0.0552	*0.0210*	*0.0363*	**0.0016**	0.0975	**0.0051**
	Kupiec test (p-value)	0.3971	0.1041	0.9333	0.3511	0.4070	0.5605	0.8667	0.2337	0.6130	0.9333
Exp 3 with 5 stocks	% of HITs	4.77	4.67	4.93	4.87	4.77	4.80	4.97	4.67	4.83	4.97
	DQ test (p-value)	0.8590	0.7941	0.2424	0.6745	0.7227	0.9176	0.2643	0.8145	0.8990	0.2643
	Kupiec test (p-value)	0.5547	0.3971	0.8667	0.7365	0.5547	0.6130	0.9332	0.3971	0.6737	0.9332
Exp 4 with 5 stocks	% of HITs	5.37	4.83	5.50	5.23	5.37	5.37	5.00	5.00	5.27	5.27
	DQ test (p-value)	**0.0054**	*0.0208*	**0.0007**	*0.0264*	**0.0008**	**0.0009**	**0.0069**	0.0605	*0.0116*	**0.0030**
	Kupiec test (p-value)	0.3622	0.6737	0.2159	0.5605	0.3622	0.3622	1.0000	1.0000	0.5063	0.5063
Exp 5 with 5 stocks	% of HITs	4.07	3.60	4.13	3.90	4.07	3.83	4.17	4.13	4.30	4.17
	DQ test (p-value)	0.2004	*0.0313*	0.5562	*0.0488*	0.6882	0.1659	0.1364	0.2003	0.2460	*0.0319*
	Kupiec test (p-value)	*0.0155*	**0.0002**	*0.0249*	**0.0041**	*0.0155*	**0.0023**	*0.0313*	*0.0249*	0.0718	*0.0313*
No. of rejections at 1% level	DQ test	1	1	1	1	2	1	1	1	0	2
	Kupiec test	0	1	0	1	0	1	0	0	0	0
No. of rejections at 5% level	DQ test	3	3	1	3	2	2	2	1	1	3
	Kupiec test	1	1	1	1	1	1	1	1	0	2

Note: Coefficients significant at 5% (1%) formatted in italic (bold).

Table 7: *In-sample* $\mathrm{VaR}_{0.05}$ violations and p-values of the statistical tests for a *hedge* portfolio. QR test denotes the dynamic quantile test. For the portfolio consisting of 10 stocks, the weight for the fist five stocks is 0.2, and that for the others is -0.2. For the portfolio consisting of 5 stocks, the weight vector is $(0.5, 0.5, -0.3, -0.3, -0.4)^T$.

		O-GARCH	DCC	GO-GARCH	IF-GARCH	BF-GARCH	CD-GARCH	GO-DCC	IF-DCC	BF-DCC	CD-DCC
With 10 stocks	% of HITs	5.07	4.20	4.63	3.47	4.40	3.63	4.87	4.43	4.43	4.30
	DQ test (p-value)	0.0951	*0.0218*	**0.0028**	**0**	**0.0010**	**0**	**0.0017**	**0.0009**	**0.0017**	**0.0001**
	Kupiec LR test (p-value)	0.8672	*0.0389*	0.3511	**0**	0.124	**0.0003**	0.7365	0.1468	0.1468	0.0718
Exp 1 with 5 stocks	% of HITs	4.00	4.83	4.53	4.17	4.47	4.07	4.70	4.77	4.67	4.63
	DQ test (p-value)	**0.0000**	0.1279	**0.0085**	*0.0247*	**0.0060**	**0.0048**	**0.0060**	**0.0097**	**0.0032**	*0.0286*
	Kupiec LR test (p-value)	**0.0093**	0.6737	0.2337	*0.0313*	0.1726	*0.0155*	0.4465	0.5547	0.3971	0.3511
Exp 2 with 5 stocks	% of HITs	3.60	4.40	4.67	3.57	4.77	3.83	4.67	4.37	4.57	4.33
	DQ test (p-value)	**0.0000**	0.0707	**0.0085**	**0.0003**	0.0650	**0.0009**	*0.0240*	*0.0117*	*0.0172*	**0.0060**
	Kupiec LR test (p-value)	**0.0002**	0.1241	0.3971	**0.0002**	0.5547	**0.0023**	0.3971	0.1041	0.2694	0.0867
Exp 3 with 5 stocks	% of HITs	3.97	4.57	4.60	3.80	4.70	3.97	4.53	4.40	4.50	4.33
	DQ test (p-value)	**0.0000**	**0.0087**	**0.0092**	**0.0000**	*0.0433*	**0.0001**	**0.0071**	**0.0008**	*0.0171*	**0.0004**
	Kupiec LR test (p-value)	**0.0071**	0.2694	0.3085	**0.0017**	0.4465	**0.0071**	0.2337	0.1241	0.2015	0.0867
Exp 4 with 5 stocks	% of HITs	3.67	4.63	4.17	3.67	4.93	3.70	4.43	4.30	4.97	4.43
	DQ test (p-value)	**0.0000**	0.6719	**0.0002**	**0.0000**	0.0869	**0.0000**	**0.0048**	**0.0001**	0.0867	**0.0023**
	Kupiec LR test (p-value)	**0.0004**	0.3511	*0.0313*	**0.0004**	0.8667	**0.0006**	0.1468	0.0718	0.9332	0.1468
Exp 5 with 5 stocks	% of HITs	4.00	4.60	4.67	4.17	4.80	4.17	4.83	4.60	4.80	4.67
	DQ test (p-value)	**0.0000**	0.0758	*0.0197*	**0.0003**	*0.0467*	**0.0010**	*0.0474*	*0.0124*	0.0906	*0.0360*
	Kupiec LR test (p-value)	**0.0093**	0.3085	0.3971	*0.0313*	0.6130	*0.0313*	0.6737	0.3085	0.6130	0.3971
No. of rejections at 1% level	DQ test	5	1	5	5	2	6	4	4	2	4
	Kupiec test	5	0	0	4	0	4	0	0	0	0
No. of rejections at 5% level	DQ test	5	2	6	6	4	6	6	6	4	6
	Kupiec test	5	1	1	6	0	6	0	0	0	0

Note: Coefficients significant at 5% (1%) formatted in italic (bold).

Table 8: *In-sample* $\text{VaR}_{0.05}$ violations and corresponding statistical tests for a *equally weighted* portfolio. For the portfolio consisting of 10 stocks, the weight for each stock is 0.1. For the portfolio consisting of 5 stocks, the weight for each stock is 0.2.

		O-GARCH	DCC	GO-GARCH	IF-GARCH	BF-GARCH	CD-GARCH	GO-DCC	IF-DCC	BF-DCC	CD-DCC
With 10 stocks	% of HITs	4.00	4.33	4.33	4.00	3.17	4.17	4.50	4.33	4.50	4.33
	DQ test (p-value)	0.7082	0.3558	0.7320	0.8016	0.2898	0.9548	0.7202	0.7366	0.5832	0.7388
	Kupiec test (p-value)	0.2449	0.4437	0.4437	0.2449	*0.0276*	0.3355	0.5678	0.4437	0.5678	0.4437
Exp 1 with 5 stocks	% of HITs	3.00	4.83	3.83	3.83	4.17	4.00	4.67	4.67	4.67	4.67
	DQ test (p-value)	0.4044	0.8727	0.6036	0.6108	0.9435	0.8912	0.9635	0.9903	0.9900	0.9904
	Kupiec test (p-value)	*0.0155*	0.8506	0.1722	0.1722	0.3355	0.2449	0.7049	0.7049	0.7049	0.7049
Exp 2 with 5 stocks	% of HITs	2.00	4.33	2.00	1.83	1.67	3.00	4.83	3.83	3.67	5.17
	DQ test (p-value)	0.0587	0.9582	*0.0311*	*0.0367*	*0.0196*	0.4324	0.9867	0.6858	0.6707	0.9883
	Kupiec test (p-value)	**0.0001**	0.4437	**0.0001**	**0.0000**	**0.0000**	*0.0155*	0.8506	0.1722	0.1164	0.8522
Exp 3 with 5 stocks	% of HITs	3.67	3.50	3.67	3.67	3.67	4.00	3.33	3.67	4.00	3.83
	DQ test (p-value)	*0.0333*	**0.0007**	*0.0334*	*0.0328*	*0.0337*	**0.0073**	**0.0095**	*0.0320*	**0.0074**	**0.0037**
	Kupiec test (p-value)	0.1164	0.0754	0.1164	0.1164	0.1164	0.2449	*0.0467*	0.1164	0.2449	0.1722
Exp 4 with 5 stocks	% of HITs	5.00	3.67	5.17	4.67	3.83	5.17	3.50	3.33	2.17	3.83
	DQ test (p-value)	0.9836	0.3880	0.9777	0.9660	0.8774	0.9779	0.5514	0.5030	0.4796	0.8158
	Kupiec test (p-value)	1.0000	0.1164	0.8522	0.7049	0.1722	0.8522	0.0754	*0.0467*	*0.0276*	0.1722
Exp 5 with 5 stocks	% of HITs	2.00	2.50	3.00	2.83	2.17	3.17	3.67	3.67	3.83	3.67
	DQ test (p-value)	**0.0047**	*0.0418*	0.0775	*0.0441*	*0.0122*	0.1222	0.1446	0.1481	0.1635	0.1437
	Kupiec test (p-value)	**0.0001**	**0.0019**	*0.0155*	**0.0082**	**0.0004**	*0.0276*	0.1164	0.1164	0.1722	0.1164
No. of rejections at 1% level	DQ test	1	1	0	0	0	1	1	0	1	1
	Kupiec test	2	1	1	2	2	0	0	0	0	0
No. of rejections at 5% level	DQ test	2	2	2	3	3	1	1	1	1	1
	Kupiec test	3	1	2	2	3	2	1	1	1	0

Note: Coefficients significant at 5% (1%) formatted in italic (bold).

Table 9: *Out-of-sample* $\mathrm{VaR}_{0.05}$ violations and corresponding statistical tests for a *hedge* portfolio. For the portfolio consisting of 10 stocks, the weight for the fist five stocks is 0.2, and that for the others is -0.2. For the portfolio consisting of 5 stocks, the weight vector is $(0.5, 0.5, -0.3, -0.3, -0.4)^T$.

		O-GARCH	DCC	GO-GARCH	IF-GARCH	BF-GARCH	CD-GARCH	GO-DCC	IF-DCC	BF-DCC	CD-DCC
With 10 stocks	% of HITs	3.50	2.83	2.50	1.50	1.83	1.67	3.33	2.83	2.83	3.00
	DQ test (p-value)	**0.0067**	*0.0427*	*0.0109*	**0.0021**	**0.0065**	**0.0049**	0.117	*0.0244*	*0.0246*	**0.0030**
	Kupiec test (p-value)	0.0754	**0.0082**	**0.0019**	**0**	**0**	**0**	*0.0467*	**0.0082**	**0.0082**	*0.0155*
Exp 1 with 5 stocks	% of HITs	2.33	2.50	2.33	2.33	2.17	2.50	3.33	2.50	3.17	2.67
	DQ test (p-value)	0.0615	0.0879	**0.0038**	**0.0044**	**0.0064**	*0.0102*	0.1831	**0.0099**	0.1186	*0.0217*
	Kupiec test (p-value)	**0.0009**	**0.0019**	**0.0009**	**0.0009**	**0.0004**	**0.0019**	*0.0467*	**0.0019**	*0.0276*	**0.0041**
Exp 2 with 5 stocks	% of HITs	2.83	3.17	3.33	2.83	3.00	3.00	3.67	3.50	3.67	3.67
	DQ test (p-value)	*0.0105*	*0.0319*	0.1267	0.0538	0.1407	0.0826	0.2952	*0.0464*	0.3044	0.0788
	Kupiec test (p-value)	**0.0082**	*0.0276*	*0.0467*	**0.0082**	*0.0155*	*0.0155*	0.1164	0.0754	0.1164	0.1164
Exp 3 with 5 stocks	% of HITs	2.17	2.67	2.33	2.00	2.67	2.00	3.33	2.67	3.00	2.60
	DQ test (p-value)	*0.0359*	*0.0220*	**0.0040**	**0.0007**	*0.0107*	**0.0007**	0.1191	*0.0237*	*0.0423*	*0.0113*
	Kupiec test (p-value)	**0.0004**	**0.0041**	**0.0009**	**0.0001**	**0.0041**	**0.0001**	*0.0467*	**0.0041**	*0.0155*	**0.0019**
Exp 4 with 5 stocks	% of HITs	3.00	3.00	2.17	2.67	3.67	2.83	3.33	3.00	3.00	3.17
	DQ test (p-value)	**0.0021**	0.0906	**0.0083**	**0.0086**	*0.0294*	*0.0184*	0.0807	*0.0247*	0.0857	*0.0466*
	Kupiec test (p-value)	*0.0155*	*0.0155*	**0.0004**	**0.0041**	0.1164	**0.0082**	*0.0467*	*0.0155*	*0.0155*	*0.0276*
Exp 5 with 5 stocks	% of HITs	2.17	2.83	2.00	2.00	2.50	2.00	3.50	3.33	3.50	3.50
	DQ test (p-value)	**0.0010**	**0.0073**	**0.0025**	**0.0028**	**0.0013**	**0.0026**	*0.0423*	*0.0232*	*0.0396*	*0.0441*
	Kupiec test (p-value)	**0.0004**	**0.0082**	**0.0001**	**0.0001**	**0.0019**	**0.0001**	0.0754	*0.0467*	0.0754	0.0754
No. of rejections at 1% level	DQ test	3	1	4	5	3	3	0	1	0	1
	Kupiec test	4	4	5	6	4	5	0	3	1	2
No. of rejections at 5% level	DQ test	5	4	5	5	5	5	1	6	3	5
	Kupiec test	5	6	6	6	5	6	4	5	4	4

Note: Coefficients significant at 5% (1%) formatted in italic (bold).

Table 10: *Out-of-sample* $\text{VaR}_{0.05}$ violations and corresponding statistical tests for a *equally weighted* portfolio. For the portfolio consisting of 10 stocks, the weight for each stock is 0.1. For the portfolio consisting of 5 stocks, the weight for each stock is 0.2.

7. Conclusion

In this paper we consider the estimation problem of GARCH models from the information-theoretic viewpoint. We reveal the relationship between the statistical dependence in standardized residuals and the maximum likelihood when estimating univariate or multivariate GARCH models. We show that maximizing the likelihood is equivalent to minimizing the statistical dependence in standardized residuals. In particular, we focus on the framework of the generalized orthogonal GARCH and propose three efficient models in this framework. These models are independent-factor GARCH, best-factor GARCH, and conditional-decorrelation GARCH. Independent-factor GARCH exploits ICA to make factors mutually as independent as possible. Best-factor GARCH gives the factors which have the largest autocorrelation in their squared values such that their volatilities are forecasted well by univariate GARCH. And conditional-decorrelation GARCH aims at ensuring that factors are not only unconditionally uncorrelated, but also conditionally as uncorrelated as possible. These models exploit the data information from different aspects to construct factors, and the factors are all estimated fast and feasibly and they have clear statistical properties.

We further exploit the DCC model to estimate the weak conditional correlation between factors and develop factor-DCC models. We explain that such models give better performance than the original factor GARCH and DCC. Experimental results on the Hong Kong stock market with various GARCH models give the following findings: (1) among DCC and factor GARCH models, conditional-decorrelation GARCH, independent-factor GARCH, and DCC exhibit the best generalization performance; (2) each factor-DCC model has better performance than DCC and the corresponding factor GARCH model; (3) conditional-decorrelation GARCH (among factor GARCH models) and its extension embedded with the DCC model (among factor-DCC models) provide the best results. The independent-factor GARCH and its extension with DCC embedded are also recommended since their performance is very close to the best and their estimation is very fast.

APPENDIX

Proof of Thoerem 1: As $g_{it}, i = 1, ..., k$, are independent, the variance of x_t is σ^2. According to the property of cumulants (Nikias & Mendel, 1993),[x] cumulants of the sum of statistically

x. The fourth-order cross-cumulant of the zero-mean series y_{it} is given by

$$\text{cum}(y_{it}, y_{it}, y_{i,t-\tau}, y_{i,t-\tau}) = E\{y_{it}^2, y_{i,t-\tau}^2\} - E\{y_{it}^2\}E\{y_{i,t-\tau}^2\} - 2(E\{y_{it}y_{i,t-\tau}\})^2 \quad (34)$$

independent variables equal the sum of the cumulants of the individual variables. Consequently, for any τ, we have

$$\begin{aligned}
& \text{cov}\Big[\big(\frac{x_t}{\text{std}(x_t)}\big)^2, \big(\frac{x_{t-\tau}}{\text{std}(x_t)}\big)^2\Big] \\
= & \frac{1}{\sigma^4}\text{cum}(x_t, x_t, x_{t-\tau}, x_{t-\tau}) \\
= & \frac{1}{\sigma^4}\sum_{i=1}^{k}\text{cum}(g_{it}, g_{it}, g_{i,t-\tau}, g_{i,t-\tau}) \quad \text{(as } g_{1t}, ..., g_{kt} \text{ are independent)} \\
= & \frac{1}{\sigma^4}\sum_{i=1}^{k}\sigma_i^4\text{cov}\Big[\big(\frac{g_{it}}{\text{std}(g_{it})}\big)^2, \big(\frac{g_{i,t-\tau}}{\text{std}(g_{it})}\big)^2\Big] \\
\leq & \max_i\Big\{\text{cov}\Big[\big(\frac{g_{it}}{\text{std}(g_{it})}\big)^2, \big(\frac{g_{i,t-\tau}}{\text{std}(g_{it})}\big)^2\Big]\Big\} \cdot \frac{\sum_{i=1}^{k}\sigma_i^4}{\sigma^4} \to 0, \text{ when } k \to \infty.
\end{aligned}$$

Therefore (1) is true. Note that for zero-mean variable x, $\text{var}(x^2) = E\{x^4\} - E[x^2]^2 = E[x^2]^2 \cdot \tilde{\kappa}(x) + 2E[x^2]^2$. We further have

$$\begin{aligned}
\text{corr}(x_t^2, x_{t-\tau}^2) = & \frac{1}{\text{var}(x_t^2)}\text{cum}(x_t, x_t, x_{t-\tau}, x_{t-\tau}) \\
= & \frac{1}{\text{var}(x_t^2)}\sum_{i=1}^{k}\text{var}(g_{it}^2)\text{corr}(g_{it}^2, g_{i,t-\tau}^2) \\
\leq & \max_i[\text{corr}(g_{it}^2, g_{i,t-\tau}^2)] \cdot \frac{\sum_{i=1}^{k}\text{var}(g_{it}^2)}{\text{var}(x_t^2)} \\
= & \max_i[\text{corr}(g_{it}^2, g_{i,t-\tau}^2)] \cdot \frac{\sum_{i=1}^{k}\sigma_i^4\tilde{\kappa}(g_{it}) + 2\sum_{i=1}^{k}\sigma_i^4}{\sigma^4\tilde{\kappa}(x_t) + 2\sigma^4} \\
= & \max_i[\text{corr}(g_{it}^2, g_{i,t-\tau}^2)] \cdot \frac{\sum_{i=1}^{k}\sigma_i^4\tilde{\kappa}(g_{it}) + 2\sum_{i=1}^{k}\sigma_i^4}{\sum_{i=1}^{k}\sigma_i^4\tilde{\kappa}(g_{it}) + 2\sigma^4} \\
\leq & \max_i[\text{corr}(g_{it}^2, g_{i,t-\tau}^2)] \cdot \frac{\max_i\tilde{\kappa}(g_{it}) + 2}{\max_i\tilde{\kappa}(g_{it}) + 2\frac{\sigma^4}{\sum_{i=1}^{k}\sigma_i^4}} \to 0, \text{ when } k \to \infty.
\end{aligned}$$

This means that the autocorrelation of x_t^2 also tends to zero as $k \to \infty$. □

References

Alexander, C. (2000). Orthogonal methods for generating large positive semi-definite covariance matrices. In *ISMA center discussioni papers in finance 2000-06*. University of Reading, UK. available from http://www.ismacentre.rdg.ac.uk.

where τ is the time lag. If y_{it} is temporally uncorrelated, this cumulant is equal to the autocovariance $\text{cov}(y_{it}^2, y_{i,t-\tau}^2)$. As its special case, the fourth-order cumulant of y_{it} is $\kappa_4(y_i) = E\{y_i^4\} - 3E[y_i^2]^2$.

Alexander, C. (2001). *Market models: A guide to financial data analysis*. Chichester, UK: Wiley.

Baba, Y., Engle, R. F., Kraft, D. F., & Kroner, K. F. (1991). Multivariate simultaneous generalized ARCH. Unpublished manuscript, Dept. of Economics, University of California, San Diego.

Back, A. D. (1997). A first application of independent component analysis to extracting structure from stock returns. *International Journal of Neural Systems*, *8*, 473–484.

Bauwens, L., Laurent, S., & Rombouts, J. V. K. (2006). Multivariate GARCH models: A survey. *Journal of Applied Econometrics*, *21*, 79–109.

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, *7*, 1129–1159.

Bera, A., & Higgins, M. (1993). ARCH models: Properties, estimation and testing. *Journal of Economic Surveys*, *7*, 305–366.

Bollerslev, T. (1986). A generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*, 307–327.

Bollerslev, T. (1990). Modeling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH approach. *Review of Economics and Statistics*, *72*, 498–505.

Bollerslev, T., Chou, R., & Kroner, K. (1992). ARCH modeling in finance: A review of the theory and empirical evidence. *Journal of Econometrics*, *52*, 5–59.

Bollerslev, T., Engle, R. F., & Nelson, D. B. (1994). ARCH models. In R. F. Engle and D. McFadden (Eds.), *Handbook of econometrics*, vol. 4, 2959–3038. North-Holland, Amsterdam.

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

Bollerslev, T., & Wooldridge, J. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews*, *11*, 143–179.

Brock, W. A., Dechert, W. D., Sheinkman, J. A., & LeBaron, B. (1996). A test for independence based on the correlation dimension. *Econometric Reviews*, *15*, 197–235.

Cardoso, J. F., & Souloumiac, A. (1993). Blind beamforming for non-Gaussian signals. *IEE Proceeding-F*, *140*, 362–370.

Cardoso, J. F., & Souloumiac, A. (1996). Jacobi angles for simultaneous diagonalization. *SIAM J. Mat. Anal. Appl.*, *17*, 161–164.

Cha, S. M., & Chan, L. W. (2000). Applying independent component analysis to factor model in finance. *Intelligent Data Engineering and Automated Learning-IDEAL 2000, Springer*, 538–544.

Chan, L., & Cha, S. (2001). Selection of independent factor model in finance. *proceedings of 3rd International Conference on Independent Component Analysis and blind Signal Separation*. San Diego, California, USA.

Christodoulakis, G. A., & Satchell, S. E. (2002). Correlated ARCH (CorrARCH): Modelling the time-varying conditional correlation between financial asset returns. *European Journal of Operational Research*, *139*, 351–370.

Cichocki, A., & Amari, S. (2003). *Adaptive blind signal and image processing: Learning algorithms and applications*. UK: John Wiley & Sons. corrected and revisited edition edition.

Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory*. Wiley.

Darbellay, G., & Wuertz, D. (2000). The entropy as a tool for analyzing statistical dependencies in financial time series. *Physica A*, *287*, 429–439.

Ding, Z. (1994). *Time series analysis of speculative returns*. PhD dissertation, University of Califoria, San Diego.

Dionísio, A., Menezes, R., & Mendes, D. (2003). Mutual information: A dependence measure for nonlinear time series. Economics Working Paper Archive at WUSTL, 0311003.

Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of u.k. inflation. *Econometrica*, *50*, 987–1008.

Engle, R., & González-rivera, G. (1991). Semiparametric ARCH models. *Journal of Business and Economic Statistics*, *9*, 345–359.

Engle, R., & Manganelli, S. (2004). Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of Business and Economic Statistics*, *22*, 367–381.

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

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

Engle, R. F., Ng, V. K., & Rothschild, M. (1990). Asset pricing with a factor ARCH covariance structure: Empirical estimates for treasury bills. *Journal of Econometrics*, *45*, 213C238.

Fan, J., Wang, M., & Yao, Q. (2005). *Modelling multivariate volatilities via conditionally uncorrelated components*. unpublished manuscript.

Hyvärinen, A. (2001). Blind source separation by nonstationarity of variance: A cumulant based approach. *IEEE Transaction on Neural Networks*, *12*, 1471–1474.

Hyvärinen, A., Karhunen, J., & Oja, E. (2001). *Independent component analysis*. John Wiley & Sons, Inc.

Hyvärinen, A., & Oja, E. (1997). A fast fixed-point algorithm for independent component analysis. *Neural Computation*, *9*, 1483–1492.

Jerez, M., Casals, J., & Sotoca, S. (2000). The likelihood of multivariate GARCH models is ill-conditioned. Preliminary version.

Kiviluoto, K., & Oja, E. (1998). Independent component analysis for parallel financial time series. *Proc. ICONIP'98* (pp. 895–898). Tokyo, Japan.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives*, *3*, 73–84.

Lanne, M., & Saikkonen, P. (2005). A multivariate generalized orthogonal factor GARCH model. In *Discussion paper no 63, helsinki center of economic research*. ISSN 1795-0562.

Long, X. D. (2005). Semiparametric multivariate GARCH model. *FMA European Conference*.

Maasoumi, E., & Racine, J. (2002). Entropy and predictability of stock market returns. *Journal of Econometrics*, *107*, 291–312.

Matsuoka, K., Ohya, M., & Kawamoto, M. (1995). A neural net for blind separation of nonstationary signals. *Neural Networks*, *8*, 411–419.

Nelson, D. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, *59*, 347ł370.

Nelson, D., & Cao, C. (1992). Inequality constraints in the univariate GARCH model. *Journal of Business and Economic Statistics*, *10*, 229–235.

Nikias, C., & Mendel, J. (1993). Signal processing with higher-order spectra. *IEEE Signal Processing Magazine*, *10*, 10–37.

Poon, S., & Granger, C. W. (2003). Forecasting volatility in financial markets: A review. *Journal of Economic Literature*, *41*, 478–539.

Rombouts, J., & Verbeek, M. (2004). *Evaluating portfolio value-at-risk using semi-parametric garch models*. available at http://neumann.hec.ca/iea/cahiers/2004/iea0414_jrom.pdf.

Sheppard, K. (2002). Ucsd_garch toolbox, version 2.0, 01-Jan-2002. available at http://www.kevinsheppard.com/research/ucsd_garch/ucsd_garch.aspx.

Taleb, A., Solé, J., & Jutten, C. (2001). Quasi-nonparametric blind inversion of Wiener systems. *IEEE Transaction on Signal Processing*, *49*, 917–924.

Tse, Y. K., & Tsui, K. C. (2002). A multivariate GARCH model with time-varying correlations. *Journal of Business and Economic Statistics*, *20*, 351–362.

van der Weide, R. (2002). GO-GARCH: a multivariate generalized orthogonal GARCH model. *Journal of Applied Econometrics*, *17*, 549–564.