Multivariate Rotated ARCH Models

Diaa Noureldin
Department of Economics, University of Oxford,
& Oxford-Man Institute,
Eagle House, Walton Well Road, Oxford OX2 6ED, UK,
diaa.noureldin@economics.ox.ac.uk

Neil Shephard
Nuffield College, New Road, Oxford OX1 1NF, UK,
Department of Economics, University of Oxford,
& Oxford-Man Institute, University of Oxford,
neil.shephard@economics.ox.ac.uk

Kevin Sheppard
Department of Economics, University of Oxford,
& Oxford-Man Institute,
Eagle House, Walton Well Road, Oxford OX2 6ED, UK,
kevin.sheppard@economics.ox.ac.uk

February 16, 2012

Abstract

This paper introduces a new class of multivariate volatility models which is easy to estimate using covariance targeting, even with rich dynamics. We call them rotated ARCH (RARCH) models. The basic structure is to rotate the returns and then to fit them using a BEKK-type parameterization of the time-varying covariance whose long-run covariance is the identity matrix. The extension to DCC-type parameterizations is given, introducing the rotated conditional correlation (RCC) model. Inference for these models is computationally attractive, and the asymptotics are standard. The techniques are illustrated using data on some DJIA stocks.

Keywords: RARCH; RCC; multivariate volatility; covariance targeting; common persistence; empirical Bayes; predictive likelihood.

JEL classification: C32; C52; C58.

1   Introduction

Estimating multivariate volatility models with flexible dynamics that are feasible in moderately large dimensions remains a challenging problem. Modelling and forecasting multivariate volatility is not only crucial for asset pricing and optimal portfolio allocation, but also to characterize the systemic risk profile of individual firms.

The recent financial crisis forcefully demonstrated the need for more robust models to project financial risk, especially to capture correlation dynamics. However, in practice, new, richly parameterized models

face the 'curse of dimensionality' in reference to the - often exponential - increase in the number of model parameters as the number of assets under study grows. Reviews of the multivariate generalized autoregressive conditional heteroskedasticity (GARCH) literature are given by, for example, Bauwens et al. (2006), Engle (2009a), Silvennoinen and Teräsvirta (2009) and Francq and Zakoian (2010, Ch. 11).

The main idea in this paper is to undertake a transformation (in particular, a rotation) of the raw returns, which enables us to easily extend the idea of variance targeting (Engle and Mezrich, 1996) to covariance targeting in multivariate models of any dimension. We call this model the rotated ARCH (RARCH) model. The proposed transformation is not novel, but our model is. The transformation is related to work on the orthogonal GARCH (OGARCH) model of Alexander and Chibumba (1997) and Alexander (2001), and its extensions in van der Weide (2002), Lanne and Saikkonen (2007), Fan et al. (2008) and Boswijk and van der Weide (2011). The interest in these papers is to find orthogonal or unconditionally uncorrelated components in the raw returns which can then be modelled individually through univariate volatility models.[1] In contrast, we utilize a closely related transformation enabling us to fit flexible multivariate models to the rotated returns using covariance targeting. Unlike the majority of the existing literature, we do not assume that the *unconditional* rotation produces returns which are also conditionally uncorrelated.

The RARCH model utilizes the popular BEKK parameterization introduced in Engle and Kroner (1995) using covariance targeting where the long-run covariance is the identity matrix. We also apply the same rotation technique in the context of the dynamic conditional correlation (DCC) model of Engle (2002) to introduce the rotated conditional correlation (RCC) model. The RARCH and RCC models are particularly attractive in terms of estimation and inference, and offer computational advantages compared to existing models.

Our attention will be devoted throughout to diagonal specifications. For d assets, the diagonal RARCH or diagonal RCC model will have a number of dynamic parameters equal to $2d$.[2] In relation to diagonal specifications, we propose a novel parameterization that offers flexibility in modelling both the volatilities and correlations while economizing on the number of parameters, hence it may be attractive in large dimensions. We call it the common persistence (CP) specification as it imposes common persistence on all elements of the conditional covariance/correlation matrix and has only $d + 1$ dynamic parameters.

---

[1]The model of Fan et al. (2008) differs in that the estimated components are also conditionally uncorrelated, or the least conditionally correlated if conditionally uncorrelated components do not exist. We discuss the relation of our model to OGARCH models in Section 2.5.

[2]We use the term 'dynamic parameters' to denote the parameters of the dynamic equation for the conditional covariance matrix in the RARCH model, and for the conditional correlation matrix in the RCC model. However, covariance targeting also requires the estimation of 'static parameters' which characterize the long-run covariance of the non-standardized returns in the RARCH model, or the standardized returns in the RCC model. Estimation is typically undertaken in stages as discussed later.

This specification is motivated by the empirical observation that parameter estimates of GARCH(1,1) processes tend to show similar persistence across assets, but differ in levels of smoothness. We show that it performs quite favourably in comparison to diagonal specifications which have $2d$ dynamic parameters.

We discuss quasi-maximum likelihood (QML) estimation of the RARCH and RCC models and the asymptotic distribution of the QML estimator. In addition, we use empirical Bayes methods to improve inference for the diagonal specifications. The empirical Bayes estimator applies shrinkage to the QML parameter estimates in the spirit of James and Stein (1961), which is known to dominate the QML estimator in terms of estimation risk under quadratic loss; see Efron (2010) for details. The use of empirical Bayes methods enables us to assess the poolability of the dynamic parameters by offering a more flexible alternative to the scalar specification.

Our empirical analysis shows that both RARCH and RCC lead to statistically significant gains in the 1-step predictive joint likelihood compared to the OGARCH model. We also show that capturing the dynamics of the covariances of the rotated returns, which is missing in the OGARCH model, does improve the prediction of the conditional correlations of the unrotated returns. Given its flexibility, the RCC model performs best in the 10-dimensional case we study, and our proposed common persistence specification performs quite favourably in comparison to the diagonal specification.

The structure of the paper is as follows: Section 2 discusses the RARCH and RCC models, and their properties. Section 3 shows how to estimate the models using QML and empirical Bayes methods, and also discusses our model evaluation strategy. In Section 4, the usefulness of the rotation technique is illustrated using data on some DJIA stocks. Section 5 draws some conclusions.

2 Modelling Approach

2.1 The rotated ARCH (RARCH) Model

2.1.1 First Some Preliminaries

Let r_t, $t = 1, ..., T$, denote the d-dimensional time series of daily asset returns, and assume $\mathrm{E}[r_t|\mathcal{F}_{t-1}] = 0$ where $\mathcal{F}_{t-1}$ denotes the information set at time $t-1$. The unconditional covariance of r_t is given by

$$\mathrm{Var}[r_t] \equiv \overline{H} = P\Lambda P',$$

using the spectral decomposition in the second equality, where P is a matrix of eigenvectors, and the eigenvalue matrix, Λ, is diagonal with non-negative elements. Orthogonality of P implies $P'P = \mathrm{I}_d$ where I_d denotes the d-dimensional identity matrix. The symmetric square root of $\overline{H}$ is $\overline{H}^{1/2} = P\Lambda^{1/2}P'$.

2.1.2 Core of the Paper

We now move on to the core of the paper. Letting $r_t = \overline{H}^{1/2} e_t$, we define the rotated returns as

$$e_t = \overline{H}^{-1/2} r_t = P\Lambda^{-1/2} P' r_t, \tag{1}$$

where $\text{Var}[e_t] = \text{I}_d$. In the case that there are k zero eigenvalues in Λ, using the asymmetric square root $\overline{H}^{1/2} = P\Lambda^{1/2}$ is preferred since it delivers a reduced-dimension vector of the rotated returns of length $d - k$.[3] Without loss of generality, we will assume in the rest of the paper that all eigenvalues in Λ are strictly positive.[4]

The conditional covariance of the rotated returns is

$$\text{Var}[e_t | \mathcal{F}_{t-1}] \equiv G_t,$$

where $\text{E}[e_t | \mathcal{F}_{t-1}] = 0$ which follows from $\text{E}[r_t | \mathcal{F}_{t-1}] = 0$. Consider a covariance-targeting BEKK-type parameterization (Engle and Kroner, 1995) for G_t:

$$G_t = \left(\text{I}_d - AA' - BB'\right) + Ae_{t-1}e'_{t-1}A' + BG_{t-1}B', \quad G_0 = \text{I}_d, \tag{2}$$

where A and B are conformable parameter matrices, and we assume that

$$\left(\text{I}_d - AA' - BB'\right) \geq 0$$

in the sense of being positive semidefinite.[5] Note that by taking the unconditional expectation of (2) we have $\text{E}[G_t] - A\text{E}[G_t]A' - B\text{E}[G_t]B' = \text{I}_d - AA' - BB'$, where $\text{E}[G_t] = \text{I}_d$ is a solution to this equation implying $\text{E}[e_t e'_t] = \text{I}_d$.

The "RARCH model" in (1) and (2) means e_t follows a covariance-targeting BEKK-type parameterization with long-run covariance equal to I_d. The dynamics in (2) can, of course, be generalized to include a higher-order lag structure, or introduce asymmetric terms as proposed in Cappiello et al. (2006) in the context of the DCC model. In addition, high-frequency volatility estimators can be utilized to model and forecast the conditional covariance of daily returns as proposed in Noureldin et al. (2011), where this is shown to provide significant forecast improvements. This is straightforward to apply in (2) by replacing $e_{t-1}e'_{t-1}$ by a high-frequency covariance estimator which has been rotated so that it has unconditional

---

[3] By ignoring the zero eigenvalues, the dimension of Λ will be $(d - k) \times (d - k)$, while P will be $d \times (d - k)$.

[4] It is worth noting that the distinction between the symmetric and asymmetric square roots also has implications for the implied model for the unrotated returns when the rotated returns follow a model with a diagonal structure. This is discussed in more detail in Section 2.3.

[5] For the univariate GARCH model, variance targeting was introduced in Engle and Mezrich (1996). The usefulness of covariance targeting in large models is that it allows for 2-step estimation. In the first step, the $d(d+1)/2$ parameters in $\overline{H}$ are estimated using the method of moments, while the dynamic parameters, A and B, are estimated in the second step by quasi-maximum likelihood.

expectation equal to I_d. Also, a component structure which decomposes the conditional covariance matrix into long- and short-run components can be adopted as proposed in, for example, Colacito et al. (2011) for the DCC model.

2.1.3 Some Initial Comments

Let $\text{Var}[r_t|\mathcal{F}_{t-1}] \equiv H_t$. The advantage of rotation in conjunction with covariance targeting can be seen by considering the covariance-targeting BEKK model for the unrotated returns:

$$H_t = \left(\overline{H} - A\overline{H}A' - B\overline{H}B'\right) + Ar_{t-1}r'_{t-1}A' + BH_{t-1}B', \tag{3}$$

assuming $\left(\overline{H} - A\overline{H}A' - B\overline{H}B'\right)$ is positive semidefinite. It is clear from (3) that in the case of diagonal (when A and B are assumed diagonal) and full (when A and B are unrestricted) BEKK models, it is difficult to specify sufficient parameter restrictions to ensure that $\left(\overline{H} - A\overline{H}A' - B\overline{H}B'\right)$ is positive semidefinite. Fitting the model to e_t instead, as in (2), circumvents this problem and allows for diagonal and full BEKK-type specifications to be easily fitted.

The dynamic properties of the RARCH model can be studied when (2) is vectorized. It is straightforward to show that

$$vec\left(G_t\right) = \left(I_{d^2} - A \otimes A - B \otimes B\right)vec(I_d) + (A \otimes A + B \otimes B)vec\left(G_{t-1}\right) + (A \otimes A)u_{t-1},$$

where $u_t = vec\left(e_t e'_t - G_t\right)$ and noting that the vector martingale difference property $\text{E}[u_t|\mathcal{F}_{t-1}] = 0$ holds. The vector u_t is essentially a vector weak white noise sequence. Covariance stationarity follows directly from the analysis in Engle and Kroner (1995) and requires the eigenvalues of $(A \otimes A) + (B \otimes B)$ to be less than one in modulus. Thus $vec\left(G_t\right)$ has a covariance stationary vector autoregression representation while

$$
\begin{aligned}
vec(e_t e'_t) &= \left(I_{d^2} - A \otimes A - B \otimes B\right)vec(I_d) + (A \otimes A + B \otimes B)vec(e_{t-1}e'_{t-1}) \\
&\quad + u_t - (B \otimes B)u_{t-1},
\end{aligned}
$$

is a covariance stationary vector autoregressive moving average representation.

2.2 Leading Special Cases

2.2.1 Scalar Specification

The scalar specification assumes $A = \alpha^{1/2}I_d$ and $B = \beta^{1/2}I_d$ so that all the elements of G_t have the same dynamic parameters. The (i,j)-th element of G_t is

$$g_{ij,t} = (1 - \alpha - \beta)1_{[i=j]} + \alpha e_{i,t-1}e_{j,t-1} + \beta g_{ij,t-1}, \quad i,j = 1, ...d,$$

where $1_{[\cdot]}$ is the indicator function. Here we assume $\alpha, \beta \geq 0$. Note that if $\alpha = 0$, β is unidentified and needs to be set equal to zero indicating conditional homoskedasticity in the model. Covariance stationarity requires $\alpha + \beta < 1$.

2.2.2 Diagonal Specification

In the diagonal specification, A and B are assumed to be diagonal with elements $\alpha_{ii}^{1/2}$ and $\beta_{ii}^{1/2}$, $i = 1, ...d$. This specification implies

$$g_{ij,t} = (1 - \alpha_{ii}^{1/2}\alpha_{jj}^{1/2} - \beta_{ii}^{1/2}\beta_{jj}^{1/2})1_{[i=j]} + \alpha_{ii}^{1/2}\alpha_{jj}^{1/2}e_{i,t-1}e_{j,t-1} + \beta_{ii}^{1/2}\beta_{jj}^{1/2}g_{ij,t-1}, \quad i,j = 1,...d.$$

We assume $\alpha_{ii}^{1/2}, \beta_{ii}^{1/2} \geq 0$.[6] The cross-equation parameter restrictions between the variance and covariance equations are a feature of BEKK-type parameterizations. Covariance stationarity is determined by the eigenvalues of the diagonal matrix:

$$A \otimes A + B \otimes B = \begin{pmatrix} \alpha_{11}^{1/2}A + \beta_{11}^{1/2}B & 0 & \cdots & 0 \\ 0 & \alpha_{22}^{1/2}A + \beta_{22}^{1/2}B & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \alpha_{dd}^{1/2}A + \beta_{dd}^{1/2}B \end{pmatrix}.$$

Define

$$\lambda_{ij} = \alpha_{ii}^{1/2}\alpha_{jj}^{1/2} + \beta_{ii}^{1/2}\beta_{jj}^{1/2},$$

which we call 'persistence parameters' as they control the persistence of G_t. Obviously $\lambda_{ij} = \lambda_{ji}$ so that $A \otimes A + B \otimes B$ has $d(d+1)/2$ unique elements. To ensure covariance stationarity, we impose

$$\max \lambda_{ij} < 1, \quad i,j = 1,...d. \tag{4}$$

Note that this restriction also ensures that $I_d - AA' - BB'$ is positive definite which highlights the particular advantage of the RARCH model. In practice, imposing $\lambda_{ii} = \alpha_{ii} + \beta_{ii} < 1$, by parameterizing the model in terms of α_{ii} and β_{ii}, is a necessary and sufficient condition for (4) to hold; see Engle and Kroner (1995).

It will be convenient to introduce measures of the heterogeneity of the smoothness and persistence parameters in the elements of G_t. By smoothness parameters we refer to the coefficients β_{ii} for $g_{ii,t}$ and $\beta_{ii}^{1/2}\beta_{jj}^{1/2}$ for $g_{ij,t}$, while λ_{ii} and λ_{ij} determine the persistence of $g_{ii,t}$ and $g_{ij,t}$, respectively. For ease of interpretation, we focus only on the smoothness and persistence parameters for $g_{ii,t}$ (i.e. the conditional

---

[6]It is worth noting that Engle and Kroner (1995) did not impose non-negativity restrictions on the elements of A and B, except for restricting their first elements, $\alpha_{11}^{1/2}$ and $\beta_{11}^{1/2}$, to be positive for identification. This allows for the possibilities of (i) a positive cross product of returns decreasing the conditional covariance, and (ii) an oscillatory path for the conditional covariance/correlation. By assuming $\alpha_{ii}^{1/2}, \beta_{ii}^{1/2} \geq 0$ we rule out such possibilities.

variances), noting that the dynamic parameters of $g_{ij,t}$ are linked to those of $g_{ii,t}$ and $g_{jj,t}$ as shown above. Let

$$\mu_\alpha = d^{-1} \sum_{i=1}^{d} \alpha_{ii}$$

denote the mean estimate of α_{ii}, and

$$\sigma_\alpha = \sqrt{d^{-1} \sum_{i=1}^{d} (\alpha_{ii} - \mu_\alpha)^2}$$

be a corresponding measure of heterogeneity. We similarly define μ_β and σ_β for β_{ii}, and μ_λ and σ_λ for λ_{ii}. Note that for the scalar model, $\sigma_\alpha = \sigma_\beta = \sigma_\lambda = 0$. These measures are useful for motivating the common persistence specification, as well as introducing empirical Bayes estimation in Section 3.2.

2.2.3 Common Persistence Specification

In the diagonal case, $(A \otimes A) + (B \otimes B)$ will be a diagonal matrix with diagonal elements given by $\lambda_{ij} = \alpha_{ii}^{1/2} \alpha_{jj}^{1/2} + \beta_{ii}^{1/2} \beta_{jj}^{1/2}$. We propose a new specification in the context of diagonal structures, which we call the common persistence (CP) specification. The CP specification imposes

$$\lambda_{ij} = \lambda, \quad \forall i, j = 1, ..., d,$$

implying $A \otimes A + B \otimes B = \lambda \mathrm{I}_{d^2}$. This gives the dynamic equation

$$G_t = (1 - \lambda)\mathrm{I}_d + A(e_{t-1}e'_{t-1} - G_{t-1})A' + \lambda G_{t-1}, \tag{5}$$

where A is a diagonal matrix with elements $\alpha_{ii}^{1/2} \geq 0$, and $0 < \lambda < 1$ is a scalar parameter satisfying $\lambda \geq \max \alpha_{ii}$. This implies

$$g_{ij,t} = (1 - \lambda)1_{[i=j]} + \alpha_{ii}^{1/2} \alpha_{jj}^{1/2} e_{i,t-1} e_{j,t-1} + (\lambda - \alpha_{ii}^{1/2} \alpha_{jj}^{1/2}) g_{ij,t-1}, \quad i, j = 1, ...d,$$

The CP specification has $d + 1$ dynamic parameters compared to $2d$ dynamic parameters in the diagonal model. It imposes common persistence on all elements of G_t through a common eigenvalue, λ, for the dynamic equation for G_t.[7] The condition for covariance stationarity is simply $\lambda < 1$, which also guarantees that $(1 - \lambda)\mathrm{I}_d$ is positive definite. This specification allows the different elements of G_t to load freely on the lagged variances/covariances and the corresponding shocks allowing them to have different smoothness levels; however it restricts them to have common persistence through λ. In contrast to the diagonal specification, here we have $\sigma_\lambda = 0$, while $\sigma_\alpha \neq 0$ which also implies $\sigma_\beta \neq 0$.

---

[7]More generality can be introduced in (5) by assuming $G_t = (\mathrm{I}_d - \widetilde{\lambda}\widetilde{\lambda}') + A(e_{t-1}e'_{t-1} - G_{t-1})A' + \widetilde{\lambda}G_{t-1}\widetilde{\lambda}'$, where $\widetilde{\lambda}$ is diagonal matrix with a block structure such that different groupings of assets have different persistence levels.
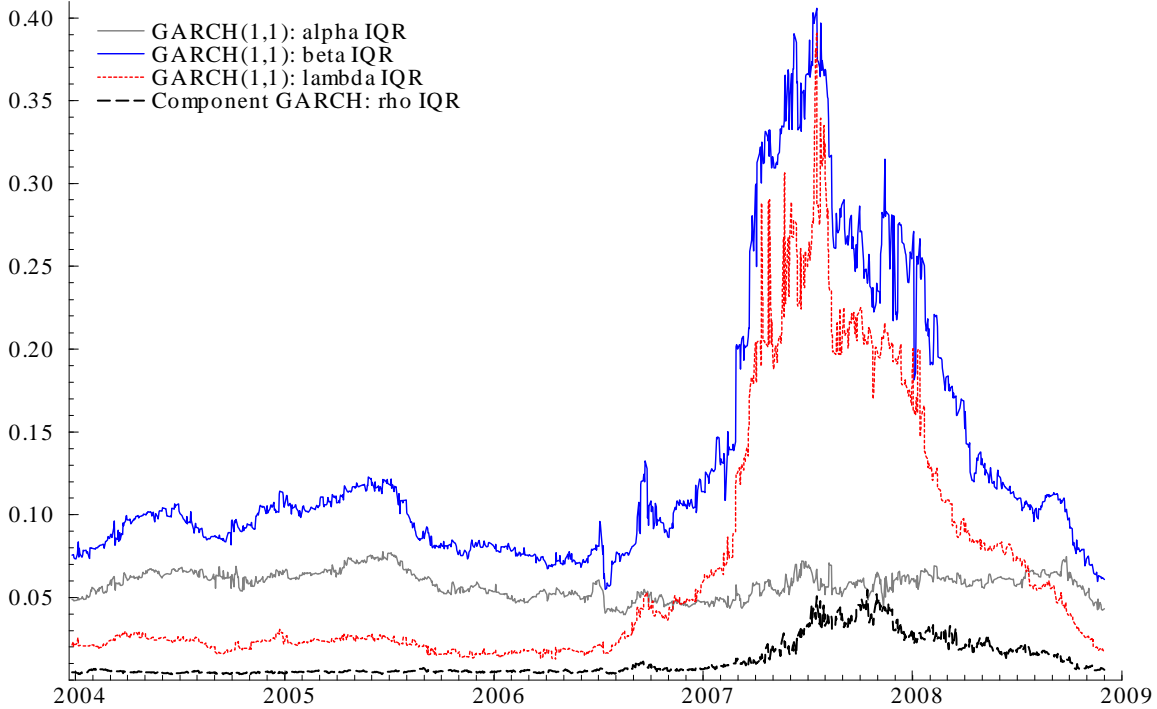
Figure 1: Heterogeneity in the smoothness and persistence parameters of the GARCH(1,1) model and the component GARCH model of Engle and Lee (1999) for the S&P 100 stocks during the period 1/4/2000-1/12/2008. The figure shows a time series of the inter-quartile range (IQR) for rolling-window estimates of the GARCH(1,1) model parameters, α_{ii} and β_{ii}, as well as the persistence, $\lambda_{ii} = \alpha_{ii} + \beta_{ii}$. The dashed black line is the IQR for rolling-window estimates of the persistence parameter in the long-run variance component, ρ_{ii}, using the component GARCH model of Engle and Lee (1999). A rolling window of 1,000 observations is used to estimate the parameters.

This specification is motivated by the empirical observation that persistence levels in the conditional variances of asset returns are less heterogeneous than their smoothness levels. Figure 1 shows a time series of the inter-quartile range (IQR) for rolling-window estimates of the GARCH(1,1) model parameters, α_{ii} and β_{ii}, as well as the implied persistence, $\lambda_{ii} = \alpha_{ii} + \beta_{ii}$, for 94 stocks of the S&P 100 index which consistently appeared on the index during the period 1/4/2000-1/12/2008.[8] Before the financial crisis, the IQR for λ_{ii} was noticeably lower than that for both α_{ii} and β_{ii}, especially compared to β_{ii}. During the recent financial crisis, more heterogeneity appears in λ_{ii} probably due to a decline in volatility persistence for some stocks in reaction to the sudden increase in their average levels of volatility during the crisis. However, it remains lower than the β_{ii} IQR, and gradually returns to its pre-crisis level by the end of the sample.

The GARCH(1,1) model is likely to miss the volatility dynamics during the crisis due to rapidly

---

[8]Here we are studying univariate GARCH models for a panel of assets as should be clear from the context. The source of the data is Thompson Reuters Datastream, and the stocks are listed in Appendix A. We used a rolling window of 1,000 observations, about 4 years of data, to estimate the parameters.

changing volatility levels. To capture this feature, we also estimated the component GARCH model of Engle and Lee (1999) using the same data. In this model the conditional variance for asset i is modelled as $h_{ii,t} = q_{ii,t} + s_{ii,t}$, where

$$
\begin{aligned}
q_{ii,t} &= \omega_{ii} + \rho_{ii}q_{ii,t-1} + \phi_{ii}(r_{i,t-1}^2 - h_{ii,t-1}), \\
s_{ii,t} &= (\alpha_{ii}^s + \beta_{ii}^s)s_{ii,t-1} + \alpha_{ii}^s(r_{i,t-1}^2 - h_{ii,t-1}),
\end{aligned}
$$

such that $q_{ii,t}$ and $s_{ii,t}$ represent the long- and short-run variance components, respectively. The short-run component has persistence $\alpha_{ii}^s + \beta_{ii}^s$ where the superscripts are used in reference to the $s_{ii,t}$ equation, while $q_{ii,t}$ has persistence ρ_{ii} which is assumed to be larger than $\alpha_{ii}^s + \beta_{ii}^s$ for identification. This model is a restricted GARCH(2,2) model and is better able to capture changes in the long-run variance over time. We find the IQR for the cross-section estimates of ρ_{ii} to be quite low, and although it increases during the financial crisis, it remains indicative of a high degree of homogeneity in the persistence levels of the long-run component. The IQR for $\alpha_{ii}^s + \beta_{ii}^s$, not plotted in Figure 1 to improve presentation, is much larger reaching as high as 0.763 during the crisis.

Brownlees (2010) studies a large cross section of U.S. financial firms during the crisis, and finds the cross-sectional variation in λ_{ii} to be negligible, while smoothness tends to decline with the leverage of the company. In the context of the DCC model, Hafner and Franses (2009) make a related observation by arguing that heterogeneity in α_{ii} is greater than that in β_{ii}, and in one of their specifications they impose a common smoothing parameter β.[9] We conjecture that imposing a common persistence parameter, λ, is more intuitive since assets with different α_{ii}'s are also likely to display varying levels of smoothness through β_{ii}. In addition, the advantage of our specification is that a single parameter, λ, controls both covariance stationarity and positive definiteness of $(1-\lambda)\mathrm{I}_d$ regardless of the dimensionality of the system.

2.2.4 Orthogonal Parameter Matrices

Another interesting case, which we outline here but do not pursue empirically, is when A and B are made up of orthogonal vectors

$$
A = (a_1, ..., a_d), \quad B = (b_1, ..., b_d)
$$

and so

$$
\left(AA'\right)_{ij} = a_i a_j' = \alpha_{ij}1_{[i=j]}, \quad \left(BB'\right)_{ij} = b_i b_j' = \beta_{ij}1_{[i=j]}, \quad i,j = 1, 2, ..., d.
$$

Orthogonality of A and B implies that $\mathrm{I}_d - AA' - BB'$ is diagonal, which means one can easily impose restrictions on A and B to ensure $\mathrm{I}_d - AA' - BB'$ is positive semidefinite.

---

[9]In their paper, the parameters are those of the dynamic equation for the conditional correlations; we will discuss the DCC model in more detail in Section 2.4.
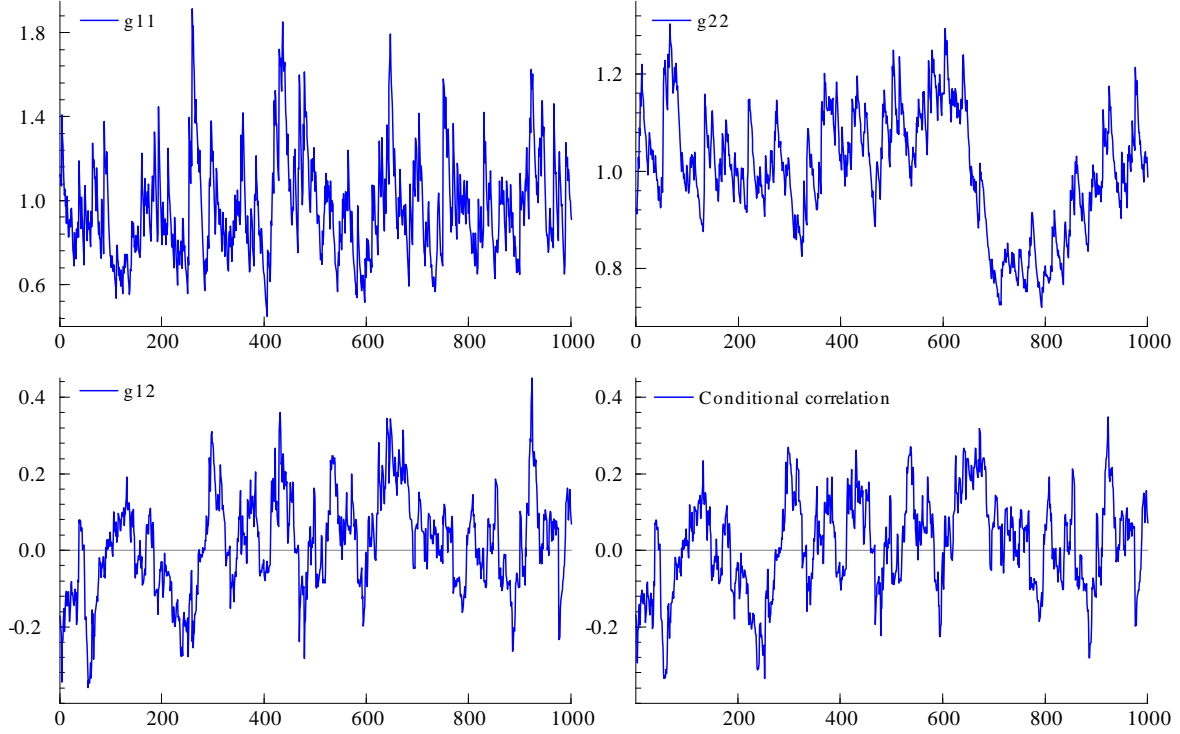
Figure 2: Sample path of G_t for 1,000 simulations.

Example 1 *Suppose $d = 2$ and we parameterize A as*

$$A = \begin{pmatrix} \alpha_{11}^{1/2} & -c\alpha_{11}^{1/2} \\ c\alpha_{22}^{1/2} & \alpha_{22}^{1/2} \end{pmatrix}, \quad Ae_{t-1} = \begin{pmatrix} \alpha_{11}^{1/2} \\ c\alpha_{22}^{1/2} \end{pmatrix} e_{1,t-1} + \begin{pmatrix} -c\alpha_{11}^{1/2} \\ \alpha_{22}^{1/2} \end{pmatrix} e_{2,t-1},$$

then A is orthogonal and AA' is diagonal with first element $\alpha_{11}\left(1 + c^2\right)$, and second element $\alpha_{22}\left(1 + c^2\right)$. When $c = 0$ then A is diagonal. In the diagonal case suppose $\alpha_{11}^{1/2} = 0.300$, $\beta_{11}^{1/2} = 0.950$, $\alpha_{22}^{1/2} = 0.150$, $\beta_{22}^{1/2} = 0.980$. Figure 2 shows the sample path of G_t for $1,000$ simulated observations assuming dynamics similar to (2). Top left is $g_{11,t}$, top right is $g_{22,t}$, bottom left is $g_{12,t}$ while bottom right is the conditional correlation.

Remark 1 *One parameterization for A and B in the orthogonal parameters version of RARCH is to use*

$$A = diag\left(\alpha_{ii}^{1/2}\right) P_A, \quad B = diag\left(\beta_{ii}^{1/2}\right) P_B,$$

where $\alpha_{ii}^{1/2}$ and $\beta_{ii}^{1/2}$ are identical to those in the diagonal specification, and P_A and P_B are orthonormal matrices.

P_A can be computed using a set of $d(d-1)/2$ parameters on $[-1,1]^{d(d-1)/2}$, by exploiting the relationship between correlation matrices and the space of orthonormal matrices. Let the $d(d-1)/2$ parameter be denoted

$$\rho_{21}, \rho_{31}, \rho_{32}, \rho_{41}, \rho_{42}, \rho_{43}, \rho_{51}, \cdots, \rho_{d(d-1)}$$

and define $C_A = [c_{ij}]$ where

$$c_{ij} = \begin{cases} 1 & i,j = 1, \\ \sqrt{1 - \sum_{k=1}^{j-1} c_{ik}^2} & i = j \neq 1, \\ \rho_{ij} \sqrt{1 - \sum_{k=1}^{j-1} c_{ik}^2} & i > j, \\ 0 & i < j. \end{cases}$$

Define $R_A = C_A C_A'$, and finally P_A are the eigenvectors of R_A so that $R_A = P_A \Lambda_A P_A'$. P_B can be similarly parameterized. Finally, $AA' = diag\left(\alpha_{ii}^{1/2}\right) P_A P_A' diag\left(\alpha_{ii}^{1/2}\right) = diag\left(\alpha_{ii}\right)$ and $BB' = diag\left(\beta_{ii}\right)$, so that the constraint needed for stationarity is that $\alpha_{ii} + \beta_{ii} < 1$ for all i, which also ensures that $I_d - AA' - BB'$ is positive definite.

2.3 Implied BEKK Model for r_t

For the conditional covariance matrix of the unrotated returns, H_t, the RARCH model in (1) and (2) implies that

$$H_t = \overline{H}^{1/2} G_t \overline{H}^{1/2} = \overline{CC}' + \overline{A} r_{t-1} r_{t-1}' \overline{A}' + \overline{B} H_{t-1} \overline{B}',$$

where

$$\overline{A} = \overline{H}^{1/2} A \overline{H}^{-1/2}, \quad \overline{B} = \overline{H}^{1/2} B \overline{H}^{-1/2}, \quad \overline{CC}' = \overline{H}^{1/2} \left(I_d - AA' - BB'\right) \overline{H}^{1/2}. \tag{6}$$

It is clear that the RARCH model implies a BEKK model for r_t but it is not an entirely general BEKK model; rather it is a constrained version since $\overline{A}$ and $\overline{B}$ are rotations of A and B which depend on the spectral decomposition of $\overline{H}$.

Bauwens et al. (2006) discuss the invariance of multivariate GARCH models to linear transformations of r_t. In their definition, invariance implies that the model for the transformed series is within the same model class (e.g. BEKK class) and retains the same dynamic specification (e.g. scalar, diagonal or full parameterization). This definition differs from the 'stability by aggregation' property discussed in Francq and Zakoian (2010, Ch. 11), which refers only to model class invariance. Bauwens et al. (2006) note that BEKK models are invariant to linear transformations except in the case of diagonal dynamics. For scalar dynamics, we further note that the scalar RARCH model will have the same dynamic parameters as the scalar BEKK model. This holds since $A = \alpha^{1/2} I_d$ implies $\overline{A} = \alpha^{1/2} P \Lambda^{1/2} P' P \Lambda^{-1/2} P' = \alpha^{1/2} I_d$, and the same applies to B.

In the case of diagonal specifications, which are not invariant to linear transformations, the diagonal RARCH model implies a full BEKK model for r_t. Suppose A is diagonal, then $\overline{A} = P\Lambda^{1/2}P'AP\Lambda^{-1/2}P'$ which is a full matrix that is asymmetric in general. The same applies to $\overline{B}$ when B is diagonal. This means that fitting a diagonal RARCH model implies rather rich dynamics for the unrotated returns. When the asymmetric square root is used, a diagonal RARCH model implies a full BEKK for r_t with $\overline{A} = P\Lambda^{1/2}A\Lambda^{-1/2}P' = PAP'$ which is symmetric, and $\overline{B}$ will also be symmetric. Thus, we prefer the symmetric square root given the generality it implies for the model of the unrotated returns since $\overline{A}$ and $\overline{B}$ will be asymmetric in this case.

2.4 The Rotated Conditional Correlation (RCC) Model

One shortcoming of the BEKK-type parameterization in the RARCH model is that the dynamics of $g_{ij,t}$ is linked to the dynamics of $g_{ii,t}$ and $g_{jj,t}$ for all i and j through cross-equation parameter restrictions. This is overcome if we apply the rotation to the returns after having them standardized in a first step by their fitted conditional variances. This is in the spirit of the DCC model of Engle (2002), thus we call the resulting model the rotated conditional correlation (RCC) model. As in DCC, RCC allows for the speed of change in the conditional correlations to be different than that of the individual volatilities, and also allows models to be fit in quite large dimensions.

Let $H_t = D_t C_t D_t$ where $D_t = diag(\sqrt{h_{ii,t}})$. Thus C_t is the conditional correlation matrix of r_t. In the DCC model the conditional variances, $\mathrm{Var}[r_{i,t}|\mathcal{F}^{r_i}_{t-1}] \equiv h_{ii,t}$, $i = 1, 2, ..., d$, are set as univariate GARCH processes, where $\mathcal{F}^{r_i}_{t-1}$ is asset i's own univariate information set, or 'natural filtration.' Then let

$$\varepsilon_t = D_t^{-1} r_t$$

denote the standardized, potentially correlated returns and let their unconditional covariance be written as

$$\Pi = \mathrm{Var}[\varepsilon_t].$$

The correlation dynamics are modelled as

$$C_t = (Q_t \circ \mathrm{I}_d)^{-\frac{1}{2}} Q_t (Q_t \circ \mathrm{I}_d)^{-\frac{1}{2}}, \tag{7}$$

where $\circ$ denotes the Hadamard elementwise product, with specifying a dynamic equation for Q_t. For instance, a correlation-targeting scalar DCC model is

$$Q_t = (1 - \alpha - \beta)\,\Pi + \alpha \varepsilon_{t-1}\varepsilon'_{t-1} + \beta Q_{t-1}, \tag{8}$$

where α and β satisfy restrictions similar to the scalar BEKK model. This ensures that C_t is a genuine correlation matrix.

From (8), it is clear that a rotation of the standardized returns, ε_t, enables us to fit specifications with flexible dynamics as demonstrated with the RARCH model for the non-standardized returns. Decompose $\Pi = P_C \Lambda_C P_C'$ where P_C contains the eigenvectors and Λ_C has the eigenvalues on the main diagonal. Then we construct the rotated standardized returns

$$\widetilde{e}_t = P_C \Lambda_C^{-1/2} P_C' \varepsilon_t.$$

where $\mathrm{Var}[\widetilde{e}_t] = I_d$. Then we model the dynamics of $\mathrm{Var}[\widetilde{e}_t | \mathcal{F}_{t-1}] \equiv Q_t^*$ as

$$Q_t^* = \left(I_d - AA' - BB'\right) + A\widetilde{e}_{t-1}\widetilde{e}_{t-1}'A + BQ_{t-1}^*B, \quad Q_0^* = I_d. \tag{9}$$

As shown for the RARCH model, Q_t is given by

$$Q_t = P_C \Lambda_C^{1/2} P_C' Q_t^* P_C \Lambda_C^{1/2} P_C',$$

which is used to construct the correlation matrix according to (7).

Aielli (2006) points to a bias problem in the DCC model when estimating Π in (8) using a moment estimator based on ε_t since $\mathrm{E}[\varepsilon_{t-1}\varepsilon_{t-1}'] \neq \mathrm{E}[Q_t]$. The RCC model is not subject to this bias since in (9) we have $\mathrm{E}[\widetilde{e}_{t-1}\widetilde{e}_{t-1}'] = \mathrm{E}[Q_t^*] = I_d$. This is an important advantage of the RCC model.

Setting $A = \alpha^{1/2}I_d$ and $B = \beta^{1/2}I_d$ gives the scalar RCC model. We can also introduce more flexible dynamics by allowing A and B to be diagonal or follow CP dynamics in the recursion for Q_t^*. The RCC model contributes to a recent research push aiming to increase the flexibility of the DCC model; see, for example, Cappiello et al. (2006), Billio et al. (2006), Billio and Caporin (2009) and Hafner and Franses (2009).

The main model in Hafner and Franses (2009) bears close resemblance to a diagonal BEKK-type model applied to (8). The primary difference is that we apply this specification to the rotated standardized returns, which has the important implication that unlike Hafner and Franses (2009), our model preserves the correlation targeting property and is also not subject to the bias problem discussed in Aielli (2006). The CP specification is also somewhat related to one of the proposed models in Hafner and Franses (2009), the distinction being that they impose a common smoothing parameter β on the system for the unrotated standardized returns, while we impose common persistence through λ on the system for the rotated standardized returns.

In the context of DCC models, Bauwens et al. (2011) also use the idea of orthogonalizing the daily returns in a first step. There are, however, two main differences between RCC and their model. First, they fit a DCC model to the rotated returns whereas we apply the rotation to the standardized returns, i.e. after the first step of fitting the conditional variances. Second, we use a constant rotation matrix for

13

ε_t, while they use a slowly-varying rotation matrix computed using a Nadaraya-Watson estimator applied to r_t. For future work, it would be interesting to compare the two models empirically.

2.5 Relation to Orthogonal GARCH Models

In this subsection, we discuss how the RARCH model differs from the class of OGARCH models introduced in Alexander and Chibumba (1997), and further extended and refined in van der Weide (2002), Lanne and Saikkonen (2007), Fan et al. (2008) and Boswijk and van der Weide (2011) among others. Consider general linear transformations of the form: $r_t = Ze_t$, where Z is some invertible matrix. Lanne and Saikkonen (2007) propose the polar decomposition for Z such that

$$Z = SU, \tag{10}$$

where S is a symmetric positive definite matrix, and U is an orthogonal matrix. Since $\text{Var}[e_t] = \text{I}_d$, we have $\text{Var}[r_t] \equiv \overline{H} = ZZ' = S^2$, thus S is the symmetric square root of $\overline{H}$ given by $P\Lambda^{1/2}P'$. Therefore part of the Z matrix can be estimated using only unconditional information.

The OGARCH model of Alexander and Chibumba (1997) and Alexander (2001) assumes $U = P$, hence $Z = P\Lambda^{1/2}$, which is the asymmetric square root of $\overline{H}$. In this case e_t is a vector of the standardized principal components of r_t which are unconditionally uncorrelated by construction. Alexander (2001) assumes that these standardized principal components are also *conditionally* uncorrelated with a diagonal conditional covariance matrix, which gives the following dynamic equation:

$$G_t = \left(\text{I}_d - \widetilde{A}\widetilde{A}' - \widetilde{B}\widetilde{B}' \right) + \widetilde{A}\widetilde{A}' \circ \left(e_{t-1}e'_{t-1} \right) + \widetilde{B}\widetilde{B}' \circ G_{t-1}, \tag{11}$$

where $\widetilde{A}$ and $\widetilde{B}$ are diagonal. Its popularity in applications relates to the convenient form of (11) which allows for estimating the diagonal elements of G_t individually as univariate GARCH models, and this is obviously helpful in large dimensions. However, assuming that G_t is diagonal is an ad hoc restriction and possibly a misspecification since we expect the standardized principal components to inherit the heteroskedastic properties of the original returns; Fan et al. (2008) argue similarly. The OGARCH model completely ignores the information in the cross products of the rotated returns, and we will show that this leads to statistically significant loss of information when modelling and forecasting the conditional correlations.

The GOGARCH model of van der Weide (2002), when modified according to Boswijk and van der Weide (2011), implies $Z = SU(\delta) = P\Lambda^{1/2}P'U(\delta)$, where the orthogonal matrix $U(\delta)$ is parameterized by a $d(d-1)/2 \times 1$ vector δ, with j-th element $-180 \leq \delta_j \leq 180$ which is a rotation angle. The dynamics of the resulting $e_t = Z^{-1}r_t$ are modelled as in (11).

14

Our model takes a different stand by assuming that G_t is a full matrix. Thus we include information from the cross products of the rotated returns which are ignored in the OGARCH and GOGARCH models. This means that our model and the OGARCH model are non-nested. With reference to (10), we simply assume $U = I_d$, which means that e_t is not a unique set of components that satisfies $\text{Var}[e_t] = I_d$. For instance, we can post-multiply $\overline{H}^{1/2} = P\Lambda^{1/2}P'$ by an arbitrary orthogonal matrix U^*, and still have $\text{Var}[e_t] = \text{E}[e_t e_t'] = \text{E}[U^{*\prime}\overline{H}^{-1/2} r_t r_t' \overline{H}^{-1/2} U^*] = I_d$. However, uniqueness (or identifiability) of e_t is not crucial since our objective is to simplify estimation, not get unique estimates of e_t. What is important is that for any model for e_t, it is straightforward to derive the implied model for r_t as shown in Section 2.3.

It is worth noting that Fan et al. (2008) make an important departure from the earlier OGARCH models by estimating U under the condition that the resulting components, e_t, are also conditionally uncorrelated. While the model of Fan et al. (2008) is conceptually appealing, a set of conditionally uncorrelated components may not exist. Thus, in practice, their model may only give components that are the *least* conditionally correlated in-sample. Boswijk and van der Weide (2011) adopt a closely related approach to estimate conditionally uncorrelated factors, which departs from earlier work on the GOGARCH model.

2.6   A Time-Varying-Weight Strict Factor Model Representation

The RARCH model can also be interpreted as a time-varying-weight strict factor model since it implies

$$H_t = \overline{H}^{1/2} G_t \overline{H}^{1/2} = P\Lambda^{1/2}P' G_t P\Lambda^{1/2}P'.$$

Suppose we take the spectral decomposition of G_t at each point in time such that $G_t = P_t^G \Lambda_t^G (P_t^G)'$, where P_t^G contains the eigenvectors of G_t and the diagonal matrix Λ_t^G has the eigenvalues of G_t along its main diagonal. Then we can write

$$H_t = P\Lambda^{1/2}P' P_t^G \Lambda_t^G (P_t^G)' P\Lambda^{1/2}P' = z_t \Lambda_t^G z_t',$$

where $z_t = P\Lambda^{1/2}P' P_t^G$ is a time-varying weight matrix.

This representation is reminiscent of strict factor models where the factors are not correlated, their conditional variances are given by the diagonal elements of the time-varying Λ_t^G, and there is no approximation error covariance since the number of factors is equal to the number of assets. The term strict factor model is typically used to characterize a model where the idiosyncratic components of asset returns are uncorrelated as in Ross (1976), for example. Here we adapt it to describe a model where the factors are uncorrelated both conditionally and unconditionally, and the factor loadings, z_t, are time-varying.

Note that OGARCH models assume that G_t is diagonal, and in this case $\Lambda_t^G = G_t$ while $P_t^G = \mathrm{I}_d$, hence they impose a fixed weight matrix $z = P\Lambda^{1/2}P'$. This representation provides an additional intuition behind our model, and explains why capturing the covariance dynamics of $e_{i,t}$, $i = 1, ..., d$, is important.

3 Inference and Model Evaluation

3.1 Maximum Likelihood Estimation

3.1.1 Optimization via Multisteps

For RARCH, OGARCH and GOGARCH models, let $\theta_S = vech(\overline{H})$ denote the model's static parameters, and let θ_D denote the vector of dynamic parameters indexing A and B in (2) for the RARCH model, or indexing $\widetilde{A}$ and $\widetilde{B}$ in (11) for OGARCH and GOGARCH. The full parameter vector is $\theta = \left(\theta_S', \theta_D'\right)'$, and the true parameter vector is denoted by θ_0. Typically the dimension of θ_S is large and potentially massive if d is large since it has $O(d^2)$ elements. The dimension of θ_D is often small with only $O(d)$ parameters in the specifications we consider.

The structure of these models allows for a 2-step estimation strategy to estimate θ. In the first step, we use a method of moments estimator:

$$\widehat{\overline{H}} = T^{-1}\sum_{t=1}^{T} r_t r_t',$$

implying $\widehat{\theta}_S$. This estimate is then decomposed into $\widehat{P}$ and $\widehat{\Lambda}$. Then we construct the time series of rotated returns, $e_t = \widehat{P}\widehat{\Lambda}^{-1/2}\widehat{P}'r_t$, $t = 1, 2, ..., T$. The second-step estimation is based on the quasi-likelihood

$$\log L(\theta_D; \widehat{\theta}_S) = \sum_{t=1}^{T} \log L_t(\theta_D; \widehat{\theta}_S) = -\frac{1}{2}\sum_{t=1}^{T}\left(d\log(2\pi) + \log|G_t| + e_t'G_t^{-1}e_t\right), \tag{12}$$

where G_t is given by (2).

This is optimized over θ_D keeping $\widehat{\theta}_S$ fixed, which delivers $\widehat{\theta}_D$. When estimating the OGARCH model, we use $e_t = \widehat{\Lambda}^{-1/2}\widehat{P}'r_t$ in (12) while the dynamic equation for G_t will be given by (11). For GOGARCH we use $e_t = \widehat{P}\widehat{\Lambda}^{-1/2}\widehat{P}'r_t$ in the following quasi-likelihood

$$\log L(\theta_D; \widehat{\theta}_S) = \sum_{t=1}^{T} \log L_t(\theta_D; \widehat{\theta}_S) = -\frac{1}{2}\sum_{t=1}^{T}\left(d\log(2\pi) + \log|G_t| + e_t'U(\delta)G_t^{-1}U(\delta)'e_t\right),$$

where the dynamic equation for G_t is also given by (11). In the GOGARCH case, the additional $d(d-1)/2$ parameters in δ are contained in θ_D.[10]

---

[10]When d is large, estimating δ is generally challenging thus we only include the GOGARCH model for comparison in our empirical analysis in the bivariate case.

Remark 2 *The GOGARCH model uses a different rotation than the RARCH model, unless $\delta = 0$, and so the difference between the two is not only due to the non-zero conditional covariances in the RARCH model. Eq. (12) could be modified to allow for an estimated rotation matrix by modifying the final term to be $e_t' U(\delta) G_t^{-1} U(\delta)' e_t$ which would not pose an identification problem since the long-run covariance is I_d. We do not pursue this parameterization further since estimating the $d(d-1)/2$ parameters of $U(\delta)$ substantially increases the computational burden, except when d is small.*

For the RCC model, let $\widetilde{\theta}_M = (\widetilde{\theta}_1', ..., \widetilde{\theta}_d')'$ denote the parameter vector of the d marginal models, $\widetilde{\theta}_S = vech(\Pi)$ where Π is the unconditional covariance of the standardized returns, ε_t, and $\widetilde{\theta}_D$ is the vector of dynamic parameters in (9). The full parameter vector is $\widetilde{\theta} = (\widetilde{\theta}_M, \widetilde{\theta}_S, \widetilde{\theta}_D)$, and the true parameter vector is denoted by $\widetilde{\theta}_0$. The log-likelihood is

$$
\begin{aligned}
\log L(\widetilde{\theta}_M, \widetilde{\theta}_S, \widetilde{\theta}_D) &= \sum_{t=1}^{T} \log L_t(\widetilde{\theta}_M, \widetilde{\theta}_S, \widetilde{\theta}_D) \\
&= \sum_{t=1}^{T} \sum_{i=1}^{d} \log L_{i,t}(\widetilde{\theta}_i) + \sum_{t=1}^{T} \log L_{C,t}(\widetilde{\theta}_M, \widetilde{\theta}_S, \widetilde{\theta}_D).
\end{aligned}
\tag{13}
$$

where $\log L_{i,t}(\cdot)$ and $\log L_{C,t}(\cdot)$ denote the i-th margin and correlation log-likelihoods respectively. The two terms in (13) can be written as

$$
\begin{aligned}
\sum_{t=1}^{T} \sum_{i=1}^{d} \log L_{i,t}(\widetilde{\theta}_i) &= -\frac{1}{2} \sum_{t=1}^{T} \left(d \log(2\pi) + 2\log|D_t| + r_t' D_t^{-2} r_t \right) \\
\sum_{t=1}^{T} \log L_{C,t}(\widetilde{\theta}_M, \widetilde{\theta}_S, \widetilde{\theta}_D) &= -\frac{1}{2} \sum_{t=1}^{T} \left(-\varepsilon_t' \varepsilon_t + \log|C_t| + \varepsilon_t' C_t^{-1} \varepsilon_t \right).
\end{aligned}
$$

3.1.2 Asymptotic Distribution

To compute asymptotic standard errors, we cast the estimators into a method of moments framework. For RARCH, OGARCH and GOGARCH models, the moment conditions for the t-th observation are given by $m_t = (m_{S,t}', m_{D,t}')'$, where $m_{S,t} = \theta_S - vech(r_t r_t')$ and $m_{D,t} = \frac{\partial \log L_t(\theta_S, \theta_D)}{\partial \theta_D}$. For RCC, the moment conditions for the t-th observation are given by $\widetilde{m}_t = (\widetilde{m}_{M,t}', \widetilde{m}_{S,t}', \widetilde{m}_{D,t}')'$, where $\widetilde{m}_{M,t} = \left(\frac{\partial \log L_{1,t}(\widetilde{\theta}_1)}{\partial \widetilde{\theta}_1}, ..., \frac{\partial \log L_{d,t}(\widetilde{\theta}_d)}{\partial \widetilde{\theta}_d} \right)$ collects the scores of the marginal models, $\widetilde{m}_{S,t} = \widetilde{\theta}_S - vech(\varepsilon_t \varepsilon_t')$ and $\widetilde{m}_{D,t} = \frac{\partial \log L_{C,t}(\widetilde{\theta}_M, \widetilde{\theta}_S, \widetilde{\theta}_D)}{\partial \widetilde{\theta}_D}$.

It holds that $\mathrm{E}[m_t | \mathcal{F}_{t-1}] = 0$ and $\mathrm{E}[\widetilde{m}_t | \mathcal{F}_{t-1}] = 0$ at the true parameter values. In terms of asymptotic theory, for fixed d and $T \to \infty$, this is simply a 2-step moment estimator, e.g. Newey and McFadden (1994) and Pagan (1986), in the case of RARCH, OGARCH and GOGARCH. In the case of RCC, this is

a 3-step moment estimator. Under standard regularity conditions,[11] as $T \to \infty$ we have

$$\sqrt{T}\left(\widehat{\theta} - \theta_0\right) \xrightarrow{d} N(0, \mathcal{I}^{-1}\mathcal{J}(\mathcal{I}^{-1})') \tag{14}$$

where $\widehat{\theta} = \left(\widehat{\theta}'_S, \widehat{\theta}'_D\right)'$, and

$$\mathcal{J} = \text{Var}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} m_t\right], \quad \mathcal{I} = \text{E}\left[\frac{\partial m_t}{\partial \theta}\right], \tag{15}$$

and we use a HAC estimator, e.g. Newey and West (1987), to estimate $\mathcal{J}$. For RCC, the estimator of $\widetilde{\theta}$ follows (14) and (15) with $(\theta, \theta_0, \widehat{\theta}, m_t)$ replaced by $(\widetilde{\theta}, \widetilde{\theta}_0, \widehat{\widetilde{\theta}}, \widetilde{m}_t)$. It is worth noting that $\mathcal{I}$ will have a lower block diagonal structure in all cases. In the RCC case, the covariance matrix of the estimated dynamic parameters in (9) will account for the accumulated estimation error from estimating the marginal models and the unconditional covariance of ε_t.

For large d in diagonal and CP specifications for the RARCH model, the maximization by parts algorithm of Song et al. (2005) can be used. In the diagonal case, let $\theta_{D,i}$ denote the dynamic parameters which index the dynamics of the i-th series, $e_{i,t}$, such that $\theta_{D,i} = (\alpha_{ii}, \beta_{ii})$. The structure of the RARCH model is well-suited to this algorithm using a margins-copula likelihood decomposition while cycling and iterating over each $\theta_{D,i}$. The advantage of this is that each individual optimization is only 2-dimensional and, upon convergence, this iterative method yields the QML estimator; see Song et al. (2005) for details. For large-dimensional diagonal RCC models, the composite likelihood approach is suitable. This estimator, which is based on recent work by Engle et al. (2008) and Pakel et al. (2011), yields consistent parameter estimates but entails a loss in efficiency. See also the review in Engle (2009b) and its use by Hafner and Franses (2009).

3.2 Empirical Bayes and Limited Translation Estimators

In this subsection, we discuss the use of empirical Bayes (EB) methods to improve inference for the RARCH and RCC models.[12] The diagonal specification is undoubtedly the most flexible among the specifications we consider, yet there is a risk that some estimates of the individual α_{ii}'s and β_{ii}'s will be rather 'extreme' in the sense of being far away from their respective mean estimates. It is conceivable that these 'extreme' estimates may be well supported by the data, however they could also be a manifestation of imprecise point estimation since the different point estimates will have confidence intervals of varying lengths.

---

[11] Asymptotic theory for QML estimation of the BEKK and DCC models is discussed in Comte and Lieberman (2003) and Engle and Sheppard (2001), respectively. We note that RARCH and RCC models apply the same parameterizations but to rotated returns in case of RARCH, or to rotated standardized returns in case of RCC.

[12] The ideas discussed here are also applicable to the CP model; however, without loss of generality, we only focus on the diagonal model in the interest of brevity.

The structure of the RARCH and RCC models makes the use of empirical Bayes a natural extension. Suppose that in the diagonal specification the shock parameters (α_{ii}'s) and smoothing parameters (β_{ii}'s) are drawn from distributions that are centred around corresponding scalar parameters. Given this prior and the point estimates from QML estimation, denoted by $\alpha_{ii}^{\mathrm{ML}}$ and β_{ii}^{ML}, we can use EB methods to make inference about the posterior distribution of the parameters having observed $\alpha_{ii}^{\mathrm{ML}}$ and β_{ii}^{ML}. For a recent overview of empirical Bayes methods, see Efron (2010).

Since the shrinkage will be applied differently to the α_{ii}'s and β_{ii}'s, in what follows we use γ_{ii} to generically denote either parameter. Assume the following prior distribution: $\gamma_{ii} \overset{i.i.d.}{\sim} \mathrm{N}(\mu_\gamma, \nu_\gamma)$.[13] The ML estimator being consistent and asymptotically normal implies $\gamma_{ii}^{\mathrm{ML}}|\gamma_{ii} \overset{\cdot}{\sim} \mathrm{N}(\gamma_{ii}, \sigma^2_{\gamma_{ii}})$. Noting that estimates of $\sigma^2_{\gamma_{ii}}$ are available from QML estimation, then the posterior distribution of γ_{ii} is[14]

$$
\gamma_{ii}|\gamma_{ii}^{\mathrm{ML}} \overset{\cdot}{\sim} \mathrm{N}\left(\mu_\gamma + \frac{\nu_\gamma}{\left(\nu_\gamma + \sigma^2_{\gamma_{ii}}\right)}\left(\gamma_{ii}^{\mathrm{ML}} - \mu_\gamma\right), \frac{\nu_\gamma \sigma^2_{\gamma_{ii}}}{\left(\nu_\gamma + \sigma^2_{\gamma_{ii}}\right)}\right),
$$

which depends on the unknown parameters μ_γ and ν_γ. The EB estimator utilizes the information in $\gamma_{ii}^{\mathrm{ML}}$ to make inference about these parameters. Define

$$
\overline{\gamma} = d^{-1}\sum_{i=1}^{d}\gamma_{ii}^{\mathrm{ML}},
$$

then the EB estimator is

$$
\gamma_{ii}^{\mathrm{EB}} = \overline{\gamma} + \left(1 - \frac{(d-3)\sigma^2_{\gamma_{ii}}}{\sum_{i=1}^{d}(\gamma_{ii}^{\mathrm{ML}} - \overline{\gamma})^2}\right)\left(\gamma_{ii}^{\mathrm{ML}} - \overline{\gamma}\right). \tag{16}
$$

This EB estimator is itself the James and Stein (1961) shrinkage estimator. James and Stein (1961) showed that it everywhere dominates the ML estimator in terms of estimation risk under quadratic loss for dimensions larger than or equal to three. Empirical Bayes estimators utilize these shrinkage gains and they have proven useful in many applications; Morris (1983) and Efron (2010) provide overviews of the main results and discuss some applications. Here the Stein effect works for $d \geq 4$ because of the need to estimate μ_γ by the sample average, $\overline{\gamma}$.

Of course, the EB method is related to the heterogeneity of the dynamic and persistence parameters as discussed in Section 2.2.2 to motivate the CP specification. It is clear from (16) that the EB estimator can shrink some parameters drastically if they are far away from the mean estimate and have a relatively

---

[13]Independence among the γ_{ii}'s is a simplifying assumption which enables us to utilize the empirical Bayes estimator in its classical formulation. The alternative route of utilizing the covariance between the QML estimates would complicate the derivation of the posterior distribution.

[14]See Efron (2010) for details.

large variance, $\sigma^2_{\gamma_{ii}}$. To control for this effect, Efron and Morris (1972) introduced the limited translation (LT) estimator:

$$\gamma_{ii}^{\text{LT}} = \left\{ \begin{array}{lll} \max(\gamma_{ii}^{\text{EB}}, \gamma_{ii}^{\text{ML}} - D\sigma_{\gamma_{ii}}) & \text{for} & \gamma_{ii}^{\text{ML}} > \overline{\gamma}, \\ \min(\gamma_{ii}^{\text{EB}}, \gamma_{ii}^{\text{ML}} + D\sigma_{\gamma_{ii}}) & \text{for} & \gamma_{ii}^{\text{ML}} \leq \overline{\gamma}, \end{array} \right. \tag{17}$$

where D is a choice parameter. For instance, $D = 1$ means that γ_{ii}^{LT} will never deviate more than $\sigma_{\gamma_{ii}}$ from γ_{ii}^{ML}. We will use the EB and LT estimators to assess the poolability of the dynamic parameters as they are shrunk towards a vector of parameters that is possibly close to, but not as extreme as the scalar specification.

3.3   Model Evaluation

3.3.1   Non-Nested Comparison via Predictive Ability

We use a quasi-likelihood criterion for r_t to compare the forecasts of the RARCH, OGARCH, GOGARCH and RCC models, which means we will focus on the 1-step predictive ability of the models using a Kullback-Leibler distance. Given the likelihood for e_t in the RARCH, OGARCH and GOGARCH models, it is straightforward to compute the likelihood for r_t since the Jacobian of the transformation is $\frac{\partial r_t}{\partial e_t} = P\Lambda^{1/2}P'$, and its determinant is $\left|P\Lambda^{1/2}P'\right| = \left|\Lambda^{1/2}\right|$, where the second equality follows from P being orthogonal. Thus for a time series of length T, we have $\log L_r = \log L_e + \frac{T}{2}\log|\Lambda|$, where $\log L_r$ and $\log L_e$ denote the log-likelihoods for r_t and e_t, respectively. For the RCC model, we have $\log L_\varepsilon = \log L_{\widetilde{\varepsilon}} + \frac{T}{2}\log|\Lambda_C|$, and $\log L_r$ is computed by adding the sum of the log-likelihoods of the marginal models for the conditional variances.

Let $\log L^a_{r,t}$ denote the t-th observation log-likelihood for r_t based on model a. To compare two models, a and b, we look at the average log-likelihood difference

$$l_{a,b} = \frac{1}{T}\sum_{t=1}^{T} l_{a,b,t}, \quad l_{a,b,t} = \log L^a_{r,t} - \log L^b_{r,t}. \tag{18}$$

We then test if $l_{a,b}$ is statistically significantly different from zero by computing a HAC estimator of the variance of $l_{a,b}$. This predictive ability test was introduced by Diebold and Mariano (1995). Using a quasi-likelihood criterion is valid for non-nested and misspecified models; see Cox (1961) and Vuong (1989) for in-sample model comparison, and Amisano and Giacomini (2007) for out-of-sample model selection. For the set of models under comparison, we choose the diagonal specification within each class (RARCH, OGARCH, GOGARCH and RCC) since it is the most flexible specification, and then test for equal predictive ability. We will use either OGARCH or GOGARCH in a comparison, but not both since the GOGARCH nests OGARCH and thus this test would not be appropriate.[15]

---

[15] If interest is in testing nested models, the approach of Giacomini and White (2006) can be adopted by using rolling-window estimation. This allows for pairwise comparisons of the predictive ability of all the four classes of models as well as the different specifications under each class.

3.3.2 Margin-Copula Predictive Likelihood Decomposition

Noureldin et al. (2011) propose a margins-copula predictive likelihood decomposition to assess the fit of multivariate volatility models by separating the gains/losses due to the forecast variances and correlations. Let $\log L_i = \sum_{t=1}^{T} \log f(r_{i,t}|\mathcal{F}_{t-1})$ be the marginal log-likelihood for the i-th asset, where we have conditioned on the entire filtration, not just its own natural filtration. The copula log-likelihood is then given by $\log L_r - \sum_{i=1}^{d} \log L_i$, where $\log L_r$ is the model's joint log-likelihood. Under the assumption of conditional normality for r_t, the copula parameter is the conditional correlation matrix of the unrotated returns. This decomposition appears naturally in the RCC model due to the separate estimation of the margins in a first step, and here we apply it to the RARCH, OGARCH and GOGARCH models. Margin-based and copula-based predictive ability tests can be conducted as outlined in Noureldin et al. (2011). These tests are useful to show the improvement of RARCH and RCC over OGARCH/GOGARCH in terms of predictive power for the conditional correlations.

4 Empirical Analysis

4.1 Data

We use close-to-close daily returns data on ten stocks from the DJIA index. These are: Alcoa (AA), American Express (AXP), Bank of America (BAC), Coca Cola (KO), Du Pont (DD), General Electric (GE), International Business Machines (IBM), JP Morgan (JPM), Microsoft (MSFT), and Exxon Mobil (XOM). The sample period is 1/2/2001 to 31/12/2009 and the source of the data is Yahoo!Finance, which is accessible online. We use close prices adjusted for dividends and splits.

Our primary empirical example in Section 4.3 focuses on the pair XOM-AA, which we use to present the models' main features. In Section 4.4 we estimate the models using all 10 stocks from the DJIA index.

4.2 Competing Models

For RARCH models, where the dynamic equation is given by (2), we fit the following specifications:

- Scalar RARCH (S-RARCH). $A = \alpha^{1/2}\mathrm{I}_d$ and $B = \beta^{1/2}\mathrm{I}_d$.

- Diagonal RARCH (D-RARCH). $A = diag(\alpha_{ii}^{1/2})$ and $B = diag(\beta_{ii}^{1/2})$.

- Common persistence RARCH (CP-RARCH). $A = diag(\alpha_{ii}^{1/2})$ and λ is the common persistence parameter.

For comparison, we also report results for these three specifications when applied to OGARCH and GOGARCH models, where in the latter models it is assumed that G_t is diagonal. The diagonal OGARCH

	Returns				GARCH(1,1) innovations			
	Covariance		Eigenvectors		Covariance		Eigenvectors	
XOM	3.069	2.918	0.394	-0.919	0.981	0.470	0.706	-0.708
AA	2.918	8.633	0.919	0.394	0.470	0.983	0.708	0.706
Eigenvalues			9.882	1.820			1.452	0.512

Table 1: Dataset: XOM and AA daily returns 1/2/2001-31/12/2009. Left-hand side is the unconditional covariance of returns, together with their eigenvalues and (normalized) eigenvectors. On the right-hand side is the unconditional covariance of the returns when standardized by the conditional variance from a GARCH(1,1) model.

and GOGARCH models (with unconstrained $\alpha_{ii}^{1/2}$ and $\beta_{ii}^{1/2}$) correspond to the models of Alexander (2001) and Boswijk and van der Weide (2011), respectively, while the other specifications are novel in this context.

For the RCC model, we consider the following specifications for the dynamic equation (9):

- Scalar RCC (S-RCC). $A = \alpha^{1/2}\mathrm{I}_d$ and $B = \beta^{1/2}\mathrm{I}_d$.

- Diagonal RCC (D-RCC). $A = diag(\alpha_{ii}^{1/2})$ and $B = diag(\beta_{ii}^{1/2})$.

- Common persistence RCC (CP-RCC). $A = diag(\alpha_{ii}^{1/2})$ and λ is the common persistence parameter.

4.3 Bivariate Model: XOM-AA

We will start out with a detailed bivariate example: the daily returns of Exxon Mobil (XOM) and Alcoa (AA). The unconditional covariance matrix of the returns is given in Table 1. The first eigenvector looks like a market factor, while the second gives the weights from a long/short portfolio. Figure 3 shows the unrotated returns in the upper panel, and the rotated returns in the lower panel.

The parameter estimates of the RARCH, OGARCH and GOGARCH models are given in Table 2 together with the associated log-likelihood values for the unrotated returns evaluated at $(\widehat{\theta}_S, \widehat{\theta}_D)$. The joint log-likelihood is decomposed to indicate the performance in terms of the margins and the copula. In the RARCH class, the D-RARCH model provides a moderate improvement in fit compared to S-RARCH. This is due to the diagonal parameters freely fitting each conditional variance. The D-RARCH model shows that α_{11} and α_{22} are an order of magnitude different than the S-RARCH's α, so that XOM's conditional variance dynamics are much more responsive to its own shock, while the estimates for the conditional variance of AA are smoother. Of course, these estimates also fit the conditional covariance dynamics given the cross-equation parameter restrictions of the D-RARCH model.

Using (6), the parameters of the BEKK model for r_t implied by the D-RARCH model are

$$\overline{A} = \overline{H}^{1/2} A \overline{H}^{-1/2} = \begin{pmatrix} 0.275 & -0.019 \\ 0.034 & 0.182 \end{pmatrix}, \quad \overline{B} = \overline{H}^{1/2} B \overline{H}^{-1/2} = \begin{pmatrix} 0.951 & 0.007 \\ -0.012 & 0.983 \end{pmatrix},$$
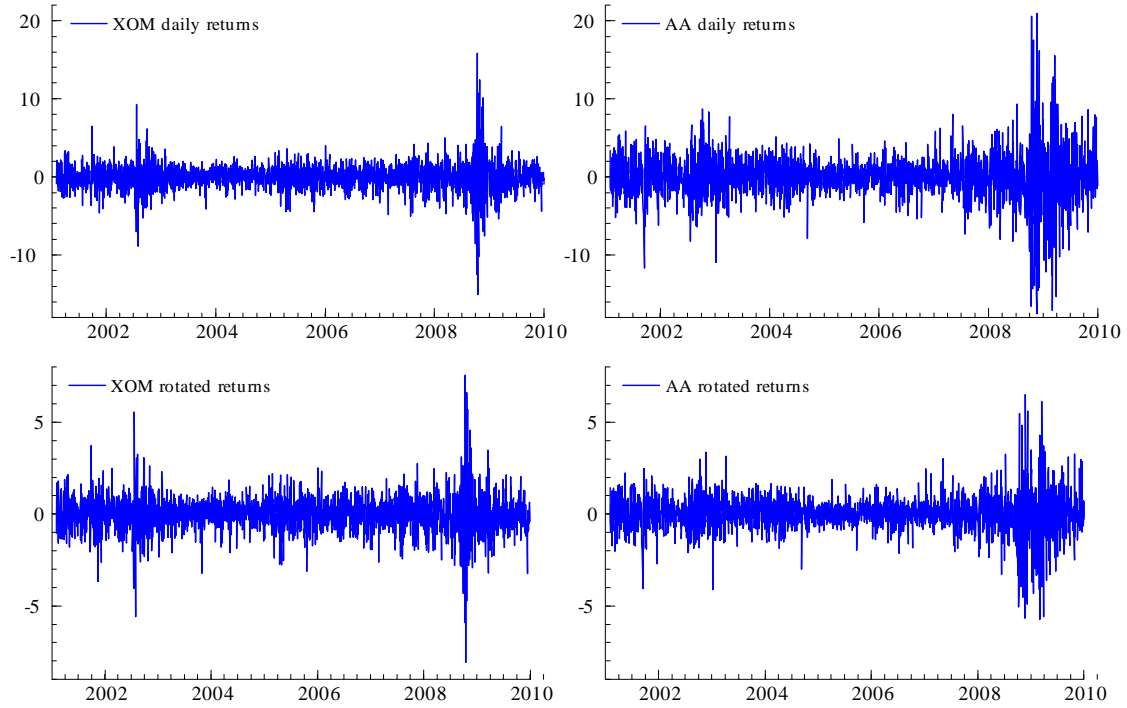
22

Figure 3: XOM and AA series: Top panel plots the daily returns (r_t). Bottom panel plots the rotated returns (e_t).

implying a full BEKK model for r_t. Asymmetry of $\overline{A}$ and $\overline{B}$ follows from specifying $\overline{H}^{1/2}$ using the symmetric square root. The CP-RARCH model estimates imply roughly the same level of persistence in the elements of G_t as the S-RARCH and D-RARCH models. The picture for OGARCH and GOGARCH is rather similar but indicating a slightly lower level of persistence.

	RARCH			OGARCH			GOGARCH			RCC		
	S	D	CP	S	D	CP	S	D	CP	S	D	CP
Marginal parameters												
α_{XOM}										0.084 (0.015)	0.084 (0.015)	0.084 (0.015)
β_{XOM}										0.901 (0.019)	0.901 (0.019)	0.901 (0.019)
α_{AA}										0.048 (0.014)	0.048 (0.014)	0.048 (0.014)
β_{AA}										0.948 (0.017)	0.948 (0.017)	0.948 (0.017)
Dynamic parameters												
α	0.050 (0.011)			0.066 (0.017)			0.063 (0.017)			0.015 (0.019)		
β	0.943 (0.013)			0.924 (0.020)			0.927 (0.022)			0.977 (0.006)		
α_{11}		0.072 (0.015)	0.055 (0.013)		0.060 (0.019)	0.076 (0.023)		0.088 (0.016)	0.067 (0.019)		0.006 (0.012)	0.005 (0.008)
α_{22}		0.036 (0.008)	0.044 (0.010)		0.082 (0.020)	0.059 (0.016)		0.041 (0.012)	0.055 (0.020)		0.037 (0.107)	0.054 (0.062)
β_{11}		0.909 (0.020)			0.934 (0.022)			0.884 (0.023)			0.993 (0.075)	
β_{22}		0.960 (0.009)			0.887 (0.029)			0.955 (0.014)			0.960 (0.068)	
λ			0.993 (0.003)			0.989 (0.005)			0.991 (0.005)			0.992 (0.017)
δ							0.015 (0.040)	−0.008 (0.073)	0.030 (0.379)			
LL decomposition												
Margin (XOM)	−4,034	−4,030	−4,032	−4,028	−4,031	−4,027	−4,029	−4,030	−4,028	−4,026	−4,026	−4,026
Margin (AA)	−5,098	−5,098	−5,099	−5,099	−5,126	−5,102	−5,100	−5,097	−5,099	−5,096	−5,096	−5,096
Copula	284	288	284	233	270	235	258	266	258	306	307	307
Total LL	−8,848	−8,840	−8,847	−8,894	−8,887	−8,893	−8,870	−8,861	−8,869	−8,816	−8,815	−8,815

Table 2: Dataset: XOM and AA daily returns 1/2/2001-31/12/2009. Parameter estimates of the scalar (S), diagonal (D) and common persistence (CP) specifications. Top panel: Marginal parameter estimates are of the variance targeting GARCH(1,1) models for the RCC margins. Dynamic parameters are estimates of the RARCH, OGARCH and GOGARCH models, and the correlation dynamics for RCC. λ is the common persistence parameter, while δ is the rotation angle in the GOGARCH model. Standard errors are reported in parentheses. Bottom panel: Log-likelihood decomposition at the estimated parameter values.

Interestingly, the GOGARCH model's estimated rotation angle is very close to zero and statistically insignificant. This implies that $U(\delta) \approx I_d$, making the e_t series from the GOGARCH model very close to those from the RARCH model. The primary difference, in this case, between the two models is that GOGARCH assumes that $g_{12,t}$ is zero, which is reflected in RARCH's superior copula fit.

The RARCH models provide an increase in the likelihood compared to OGARCH and GOGARCH. The increase in the log-likelihood in RARCH models is primarily due to an increase in the copula fit, implying that capturing the conditional correlations in the rotated returns (which is not the case in OGARCH and GOGARCH) does improve the modelling of the conditional correlations of the unrotated returns. There is a small loss in fit in the first margin (XOM) when using the RARCH model, however this is more than compensated through capturing the conditional correlation dynamics with RARCH models providing an overall gain in fit.

The last three columns of table 2 give estimates of the different RCC specifications. When estimating the variance targeting GARCH(1,1) models for the margins, we first standardize the returns of XOM and AA by their respective unconditional variances, fit variance targeting GARCH(1,1) models for these standardized returns and report the log-likelihood for the original returns as the marginal log-likelihood. The estimates suggest different dynamics for the two series, which can already be inferred from the improvement offered by the diagonal RARCH, OGARCH and GOGARCH specifications.

The estimates for the RCC dynamics suggest only a marginal improvement by the D-RCC and CP-RCC over S-RCC. With the margins fit freely, there seems to be no additional improvement from further enriching the RCC dynamics in this case. This is in contrast to the RARCH model results, but it is perhaps unsurprising since there is a single conditional correlation to fit in this case. Overall the estimates suggest that the conditional correlation matrix is quite persistent. The log-likelihood decomposition results indicate a rather significant improvement in the overall fit compared to the RARCH, OGARCH and GOGARCH models, especially in comparison to OGARCH.

We apply the 1-step predictive ability test outlined in Section 3.3 to the D-RARCH, D-OGARCH and D-RCC models which are the most flexible in each class. Comparing D-RARCH to D-OGARCH gives a t-statistic of 2.81 which is statistically significant at 1 percent, indicating that D-RARCH provides superior 1-step forecasts. Comparing D-RCC to D-RARCH and D-OGARCH gives t-statistics equal to 2.24 and 3.74, respectively, indicating that D-RCC outperforms both models. For the correlation gains, the t-statistics of the copula-based predictive ability tests are 1.04 for D-RARCH versus D-OGARCH, and 1.8 and 2.4 for D-RCC versus D-OGARCH and D-RARCH, respectively.

Figure 4 plots the conditional correlations from the diagonal models which provided the best fit in each
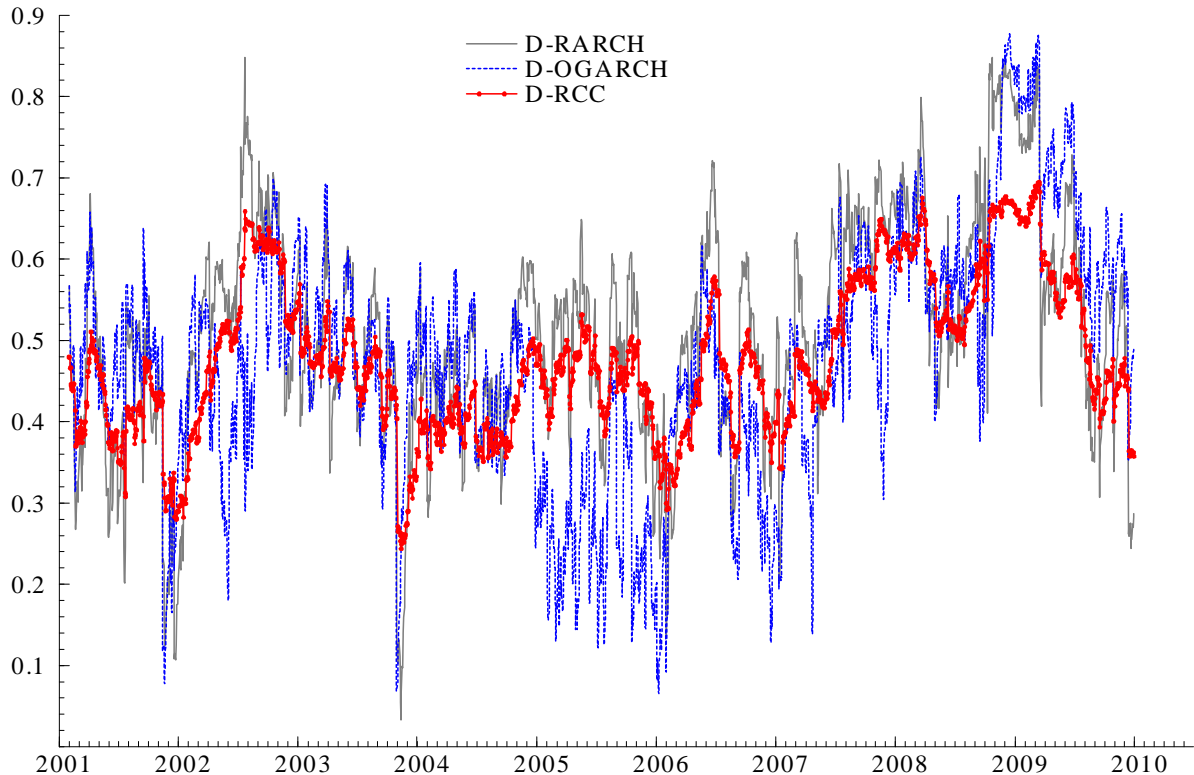
Figure 4: Conditional correlations from the diagonal RARCH, OGARCH and RCC models.

model class. The D-RCC conditional correlation is the most persistent and lies within a tighter range. It appears to be generally lower than the conditional correlation from the D-RARCH and D-OGARCH model, with the exception of the year 2005 where D-OGARCH conditional correlation was noticeably lower. This observation is perhaps most evident during the latter part of the financial crisis, roughly starting 2009, with a large difference in the level of the conditional correlations.

4.4 10-Dimensional Model

In this subsection we analyze all 10 stocks from the DJIA index. The first two eigenvectors, corresponding to the two largest eigenvalues of the unconditional covariance matrix of the returns, are reported in Table 3. The first eigenvector looks roughly like a market factor and the second is a market portfolio that is short (long) in financial stocks (BAC, JPM and AXP) and long (short) in the other stocks.[16] The two largest eigenvalues are 35.93 and 6.85, and they account for 73 percent of the total variation in the returns, where total variation is measured by the trace of $\overline{H}$.

Table 4 shows the estimated parameters for the scalar, diagonal and common persistence models.

---

[16]Note that when the eigenvalues are distinct, the eigenvectors are unique up to sign.

	BAC	JPM	IBM	MSFT	XOM	AA	AXP	DD	GE	KO
Eigenvector 1	0.505	0.439	0.182	0.218	0.180	0.360	0.392	0.242	0.292	0.106
Eigenvector 2	-0.584	-0.288	0.177	0.259	0.255	0.582	-0.033	0.223	0.077	0.129

Table 3: Dataset: 10 DJIA stocks daily returns 1/2/2001-31/12/2009. The first two (normalized) eigenvectors correspond to the two largest eigenvalues of the unconditional covariance matrix of the returns.

Moving from the scalar to the diagonal models seems to pay off with a considerable improvement in overall fit in D-RARCH, and less so for D-OGARCH. The RARCH models provide a significant overall gain in the log-likelihood over OGARCH, all due to improving the copula fit. Note that RARCH loses in the margins to OGARCH as the RARCH parameters provide a fit to both the variance and covariance elements of G_t.

Of course, the RCC models provide the best fit since the margins are freely estimated. The overall gain compared to RARCH and OGARCH is quite impressive, and the RCC gains are uniform across all margins and the copula. Unlike RARCH and OGARCH cases, moving from S-RCC to D-RCC does not improve the copula fit massively. In this moderately large dimension, the favourable performance of the CP specification is evident, particularly in the RCC model where moving from D-RCC to CP-RCC decreases the log-likelihood only marginally.

Given that both the scalar and CP specifications are nested in the diagonal model, we can use a likelihood ratio (LR) test. The scalar specification imposes $2d$ restrictions on the diagonal model, and according to the LR test, the reduction in fit is statistically significant at 5 percent in all three models. The CP specification imposes $d(d + 1)/2$ restrictions on the parameters of the diagonal model, and the loss in fit when moving from D-RARCH to CP-RARCH is statistically significant at 5 percent, while this is not the case in the OGARCH and RCC models. This suggests that common persistence dynamics is not a restrictive assumption in the OGARCH and RCC cases.

This is an interesting result since the number of dynamic parameters in the CP model is $d+1$ compared to $2d$ dynamic parameters in the diagonal model. This is potentially due to differences in the heterogeneity of the persistence parameters, λ_{ii}, among the RARCH, OGARCH and RCC diagonal specifications. We find that $\sigma_\lambda = 0.011$ for D-RARCH, $\sigma_\lambda = 0.008$ for D-OGARCH and $\sigma_\lambda = 0.019$ for D-RCC implying that common persistence may not be a reasonable assumption for RCC; however, we note that the D-RCC parameters are subject to confidence intervals of varying lengths and that is why the empirical Bayes analysis adds further insight into the poolability of the dynamic parameters, as we discuss later. It is also worth noting that $\sigma_\alpha = 0.014$ and $\sigma_\beta = 0.023$ for D-RARCH, $\sigma_\alpha = 0.020$ and $\sigma_\beta = 0.026$ for D-OGARCH, and $\sigma_\alpha = 0.004$ and $\sigma_\beta = 0.020$ for D-RCC.

Compared to the D-OGARCH specifications, D-RARCH produces superior 1-step forecasts with a

27

	RARCH			OGARCH			RCC		
	S	D	CP	S	D	CP	S	D	CP
Dynamic parameters									
α	0.020 (0.002)			0.045 (0.009)			0.007 (0.007)		
β	0.978 (0.003)			0.952 (0.010)			0.980 (0.007)		
$\min \alpha_{ii}$		0.009 (0.002)	0.010 (0.002)		0.027 (0.015)	0.025 (0.009)		0.003 (0.003)	0.002 (0.003)
$\max \alpha_{ii}$		0.054 (0.017)	0.054 (0.009)		0.097 (0.121)	0.095 (0.018)		0.018 (0.022)	0.015 (0.021)
$\min \beta_{ii}$		0.905 (0.035)			0.869 (0.177)			0.925 (0.079)	
$\max \beta_{ii}$		0.989 (0.003)			0.967 (0.020)			0.993 (0.028)	
λ			0.998 (0.001)			0.996 (0.002)			0.987 (0.018)
LL decomposition									
Margin (BAC)	-4,496	-4,355	-4,373	-4,416	-4,351	-4,361	-4,350	-4,350	-4,350
Margin (JPM)	-4,769	-4,719	-4,734	-4,706	-4,695	-4,700	-4,671	-4,671	-4,671
Margin (IBM)	-4,058	-4,085	-4,092	-4,025	-4,025	-4,023	-4,011	-4,011	-4,011
Margin (MSFT)	-4,449	-4,488	-4,482	-4,438	-4,431	-4,433	-4,424	-4,424	-4,424
Margin (XOM)	-4,090	-4,067	-4,084	-4,040	-4,032	-4,035	-4,026	-4,026	-4,026
Margin (AA)	-5,115	-5,130	-5,132	-5,097	-5,097	-5,096	-5,096	-5,096	-5,096
Margin (AXP)	-4,665	-4,648	-4,652	-4,620	-4,705	-4,654	-4,599	-4,599	-4,599
Margin (DD)	-4,249	-4,310	-4,299	-4,231	-4,247	-4,232	-4,228	-4,228	-4,228
Margin (GE)	-4,291	-4,299	-4,300	-4,263	-4,327	-4,314	-4,257	-4,257	-4,257
Margin (KO)	-3,556	-3,558	-3,562	-3,528	-3,542	-3,542	-3,520	-3,520	-3,520
Copula	4,640	4,860	4,807	3,888	4,040	3,963	4,919	4,946	4,939
Total LL	-39,098	-38,798	-38,904	-39,475	-39,413	-39,426	-38,263	-38,236	-38,244

Table 4: Dataset: 10 DJIA stocks daily returns 1/2/2001-31/12/2009. Parameter estimates of the scalar (S), diagonal (D), and common persistence (CP) specifications. Top panel: estimates of the dynamic parameters; α and β are the parameters of the scalar specifications, and α_{ii} and β_{ii} are those of the diagonal specifications. For CP, λ (the common persistence parameter) and α_{ii} are reported. Standard errors are reported in parentheses. Lower panel: Log-likelihood decomposition at the estimated parameter values.

t-statistic equal to 3.49. The D-RCC model outperforms both D-OGARCH and D-RARCH with t-statistics equal to 7.09 and 2.66, respectively. For the correlations, D-RARCH significantly improves over D-OGARCH with a t-statistic of 6.58, while D-RCC's predictive ability tests against D-OGARCH and D-RARCH give t-statistics equal to 5.44 and 0.54, respectively. Hence the overall gains of D-RCC over D-RARCH seem to be due to better forecasts of the conditional variances. These results reflect the importance of the information in the cross products of the returns, which is ignored in the OGARCH model.

Table 5 (top panel):

	RARCH α_{ii}^{ML}	RARCH β_{ii}^{ML}	RARCH α_{ii}^{LT}	RARCH β_{ii}^{LT}	OGARCH α_{ii}^{ML}	OGARCH β_{ii}^{ML}	OGARCH α_{ii}^{LT}	OGARCH β_{ii}^{LT}	RCC α_{ii}^{ML}	RCC β_{ii}^{ML}	RCC α_{ii}^{LT}	RCC β_{ii}^{LT}
BAC	0.053 (0.017)	0.946 (0.017)	0.036	0.954	0.080 (0.014)	0.918 (0.015)	0.073	0.922	0.005 (0.011)	0.992 (0.047)	0.010	0.969
JPM	0.030 (0.004)	0.969 (0.004)	0.029	0.969	0.054 (0.012)	0.942 (0.013)	0.055	0.941	0.007 (0.012)	0.993 (0.028)	0.013	0.979
IBM	0.016 (0.008)	0.979 (0.012)	0.017	0.977	0.042 (0.014)	0.955 (0.015)	0.046	0.950	0.006 (0.006)	0.978 (0.026)	0.009	0.972
MSFT	0.011 (0.004)	0.985 (0.005)	0.012	0.985	0.075 (0.109)	0.917 (0.130)	0.020	0.982	0.009 (0.011)	0.970 (0.018)	0.003	0.971
XOM	0.054 (0.017)	0.905 (0.035)	0.037	0.940	0.058 (0.071)	0.937 (0.078)	0.044	0.927	0.018 (0.022)	0.960 (0.035)	0.007	0.977
AA	0.016 (0.003)	0.981 (0.004)	0.017	0.981	0.042 (0.029)	0.955 (0.033)	0.056	0.939	0.008 (0.011)	0.990 (0.007)	0.005	0.988
AXP	0.025 (0.003)	0.974 (0.007)	0.025	0.973	0.050 (0.049)	0.938 (0.072)	0.073	0.926	0.007 (0.011)	0.925 (0.079)	0.010	0.964
DD	0.009 (0.002)	0.989 (0.003)	0.009	0.989	0.097 (0.121)	0.869 (0.177)	0.037	0.958	0.006 (0.010)	0.977 (0.022)	0.011	0.973
GE	0.019 (0.005)	0.979 (0.005)	0.020	0.978	0.027 (0.015)	0.967 (0.020)	0.035	0.957	0.007 (0.011)	0.949 (0.034)	0.008	0.966
KO	0.020 (0.005)	0.976 (0.007)	0.020	0.976	0.036 (0.057)	0.955 (0.084)	0.064	0.913	0.003 (0.003)	0.989 (0.023)	0.004	0.977

LL decomposition

	RARCH ML	RARCH LT	OGARCH ML	OGARCH LT	RCC ML	RCC LT
Margin (BAC)	-4,355	-4,703	-4,351	-4,417	-4,350	-4,350
Margin (JPM)	-4,719	-4,724	-4,695	-4,705	-4,671	-4,671
Margin (IBM)	-4,085	-4,076	-4,025	-4,024	-4,011	-4,011
Margin (MSFT)	-4,488	-4,483	-4,431	-4,713	-4,424	-4,424
Margin (XOM)	-4,067	-4,070	-4,032	-4,065	-4,026	-4,026
Margin (AA)	-5,130	-5,129	-5,097	-5,097	-5,096	-5,096
Margin (AXP)	-4,648	-4,652	-4,705	-4,606	-4,599	-4,599
Margin (DD)	-4,310	-4,308	-4,247	-4,234	-4,228	-4,228
Margin (GE)	-4,299	-4,297	-4,327	-4,330	-4,257	-4,257
Margin (KO)	-3,558	-3,556	-3,542	-3,541	-3,520	-3,520
Copula	4,860	4,855	4,040	4,000	4,946	4,909
Total LL	-38,798	-39,144	-39,413	-39,732	-38,236	-38,272

Table 5: Dataset: 10 DJIA stocks daily returns 1/2/2001-31/12/2009. Maximum likelihood (ML) and limited translation (LT) parameter estimates for the diagonal RARCH, OGARCH and RCC models. Top panel: ML and LT parameter estimates with standard errors reported in parentheses underneath the ML estimates. Bottom panel: Log-likelihood decomposition at the estimated parameter values using ML and LT estimators.

In Table 5 we analyze the results of the EB and LT estimators. As discussed in Section 3.2, the EB estimator may result in strong shrinkage effects, depending on the parameters' standard errors and distance from the mean estimate, which may result in violations of the parameter restrictions required for the diagonal model. In all three cases, we used the LT estimator to cushion against such effects. In the RARCH model, we set $D = 1$, while for the OGARCH and RCC models, we set $D = 0.5$; see (17). The table reports the QML estimates along with their standard errors, and the LT estimates. The parameters in the RARCH and OGARCH models are directly linked to the respective assets. In the case of RCC, the parameters of BAC, for instance, are those which parameterize the correlation dynamics of BAC with the remaining assets, and so on.

In the RARCH model, the effect of shrinkage is strongest in the parameters of BAC and XOM. The $\alpha_{ii}^{\mathrm{ML}}$ is quite high for these assets, implying less smoothness in the conditional variances and stronger-than-average responsiveness to the shocks. Shrinkage leads to a notable deterioration in the fit of BAC's conditional variance, while XOM's fit is almost unaffected. Interestingly the copula fit is not drastically affected, and so the overall deterioration in the joint log-likelihood is mostly due to BAC's marginal log-likelihood. Testing the predictive ability of D-RARCH according to the QML estimates against the LT estimates gives a t-statistic of 5.81 suggesting that shrinkage is not suitable in this case. This mirrors the results in Table 4 where it is shown that moving from a diagonal to a scalar RARCH (implying shrinkage to scalar parameters) leads to a statistically significant reduction in predictive ability.

In the OGARCH case where there is higher heterogeneity among the dynamic parameters, noticeable shrinkage is applied to almost half of the assets. Given the OGARCH model's structure, this results in a worse fit mostly in the margins while the copula log-likelihood only decreases marginally. Shrinkage leads to a statistically significant reduction in predictive ability with a t-statistic of 3.66. Again, this mirrors the result when imposing common dynamic parameters as in the S-OGARCH model. In the case of RCC, the $\alpha_{ii}^{\mathrm{LT}}$ estimates show considerable difference compared to $\alpha_{ii}^{\mathrm{ML}}$ for many of the assets as they are shrunk towards the mean estimate. These parameters control only the evolution of the conditional correlations and so the fit of the margins is unaffected. There is a slight deterioration in the copula fit but it is not statistically significant when using the predictive ability test. For the joint log-likelihood, the predictive ability test for the unshrunk versus the shrunk model gives a t-statistic of -0.10 implying that RCC dynamic parameters are not adversely affected by shrinkage.

5   Conclusion

This paper advocates a rotation technique for raw returns which leads to easy-to-fit multivariate volatility models via covariance targeting. We introduce the RARCH and RCC models to study the dynamics of the conditional covariance matrix of daily asset returns, highlighting the similarities and differences between our approach and the class of OGARCH models.

We focus on diagonal structures and introduce the common persistence specification as a more-tightly-parameterized alternative to the diagonal specification. Quasi-maximum likelihood estimation for our model is computationally attractive, thanks to the convenient form of covariance targeting with a long-run identity matrix. We also discuss the use of empirical Bayes methods for estimation, which provides additional insight into the poolability of the dynamic parameters.

Empirical analysis shows that our model leads to statistically significant gains in the 1-step predictive joint likelihood compared to OGARCH and GOGARCH, and that capturing the dynamics of the covariances of the rotated returns does improve the prediction of the conditional correlations. Given its flexibility, the RCC model performs best in the 10-dimensional case we study, and our proposed common persistence specification performs quite favourably in comparison to the diagonal specification.

Acknowledgements

References

Aielli, G. P. (2006). Consistent estimation of large scale dynamic conditional correlations. Unpublished paper: Department of Statistics, University of Florence.

Alexander, C. (2001). Orthogonal GARCH. In *Mastering Risk*, Volume 2, pp. 21–38. Prentice Hall.

Alexander, C. O. and A. M. Chibumba (1997). Multivariate orthogonal factor GARCH. Mimeo: University of Sussex, UK.

Amisano, G. and R. Giacomini (2007). Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business & Economic Statistics 25*, 177–190.

Bauwens, L., C. M. Hafner, and D. Pierret (2011). Multivariate volatility modelling of electricity futures. *Journal of Applied Econometrics*. Forthcoming.

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

Billio, M. and M. Caporin (2009). A generalized dynamic conditional correlation model for portfolio risk evaluation. *Mathematics and Computers in Simulation 79*, 2566–2578.

Billio, M., M. Caporin, and M. Gobbo (2006). Flexible dynamic conditional correlation multivariate GARCH models for asset allocation. *Applied Financial Economics Letters 2*, 123–130.

Boswijk, H. P. and R. van der Weide (2011). Method of moments estimation of GO-GARCH models. *Journal of Econometrics*. Forthcoming.

Brownlees, C. T. (2010). On the relation between firm characteristics and volatility dynamics with an application to the 2007-2009 financial crisis. Working paper: Stern School of Business, New York University.

Cappiello, L., R. F. Engle, and K. Sheppard (2006). Asymmetric dynamics in the correlations of global equity and bond returns. *Journal of Financial Econometrics 4*, 537–572.

Colacito, R., R. F. Engle, and E. Ghysels (2011). A component model for dynamic correlations. *Journal of Econometrics 164*, 45–59.

Comte, F. and O. Lieberman (2003). Asymptotic theory for multivariate GARCH processes. *Journal of Multivariate Analysis 84*, 61–84.

Cox, D. R. (1961). Tests of separate families of hypotheses. *Proceedings of the Berkeley Symposium 4*, 105–123.

Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics 13*, 253–263.

Efron, B. (2010). *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing and Prediction*. Cambridge: Cambridge University Press.

Efron, B. and C. N. Morris (1972). Limiting the risk of Bayes and empirical Bayes estimators - part II. *Journal of the American Statistical Association 67*, 130–139.

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. (2009a). *Anticipating Correlations*. Princeton University Press.

Engle, R. F. (2009b). High dimensional dynamic correlations. In J. L. Castle and N. Shephard (Eds.), *The Methodology and Practice of Econometrics: Papers in Honour of David F Hendry*, pp. 122–148. Oxford University Press.

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

Engle, R. F. and G. G. J. Lee (1999). A permanent and transitory component model of stock return volatility. In R. F. Engle and H. White (Eds.), *Cointegration, Causality, and Forecasting. A Festschrift in Honour of Clive W.J. Granger*, pp. 475–497. Oxford: Oxford University Press.

Engle, R. F. and J. Mezrich (1996). GARCH for groups. *Risk 9*, 36–40.

Engle, R. F., N. Shephard, and K. K. Sheppard (2008). Fitting vast dimensional time-varying covariance models. Unpublished paper: Oxford-Man Institute, University of Oxford.

Engle, R. F. and K. K. Sheppard (2001). Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. Unpublished paper: UCSD.

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

Francq, C. and J. M. Zakoian (2010). *GARCH Models*. Chichester: Wiley.

Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica 74*, 1545–1578.

Hafner, C. M. and P. H. Franses (2009). A generalized dynamic conditional correlation model: simulation and application to many assets. *Econometric Reviews 28*, 612–631.

James, W. and C. Stein (1961). Estimation with quadratic loss. In *Proceedings of the Fourth Berkeley Symposium in Math Statist. Prob*, Volume 1, pp. 361–380. University of California.

Lanne, M. and P. Saikkonen (2007). A multivariate generalized orthogonal factor GARCH model. *Journal of Business and Economic Statistics 25*, 61–75.

Morris, C. N. (1983). Parametric empirical Bayes inference: theory and applications. *Journal of the American Statistical Association 78*, 47–55.

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. In R. F. Engle and D. McFadden (Eds.), *The Handbook of Econometrics, Volume 4*, pp. 2111–2245. North-Holland.

Newey, W. K. and K. D. West (1987). A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*, 703–708.

Noureldin, D., N. Shephard, and K. Sheppard (2011). Multivariate high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics*. Forthcoming.

Pagan, A. (1986). Two stage and related estimators and their applications. *Review of Economic Studies 53*, 517–538.

Pakel, C., N. Shephard, and K. Sheppard (2011). Nuisance parameters, composite likelihoods and a panel of GARCH models. *Statistica Sinica 21*, 307–329.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory 13*, 341–360.

Silvennoinen, A. and T. Teräsvirta (2009). Multivariate GARCH models. In T. G. Andersen, R. A. Davis, J. P. Kreiss, and T. Mikosch (Eds.), *Handbook of Financial Time Series*, pp. 201–229. Springer-Verlag.

Song, P. X. K., Y. Fan, and J. D. Kalbfleisch (2005). Maximization by parts in likelihood inference. *Journal of the American Statistical Association 100*, 1145–1169.

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

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica 57*, 307–333.

A   List of Stocks Used in Figure 1

The following is the list of stocks, with the ticker symbol in parentheses, from the S&P100 index used in Figure 1: 3M Co. (MMM), AT&T Inc. (T), Abbott Laboratories (ABT), Alcoa (AA), Allstate Corp. (ALL), Altria Group (MO), Amazon.com (AMZN), American Electric Power Co. (AEP), American Express Inc. (AXP), Amgen Inc. (AMGN), Apple Inc. (AAPL), Avon Products Inc. (AVP), Baker Hughes (BHI), Bank of America Corp. (BAC), Bank of New York (BK), Baxter International Inc. (BAX), Boeing Co. (BA), Bristol-Myers Squibb (BMY), Burlington Northern Santa Fe Corp. (BNI), CVS Caremark (CVS), Campbell Soup Co. (CPB), Capital One Financial Corp (COF), Chevron (CVX), Caterpillar (CAT), Cisco Systems (CSCO), Citigroup Inc. (C), The Coca Cola Co. (KO), Colgate-Palmolive Co. (CL), Comcast Corp. (CMCSA), ConocoPhillips (COP), Costco (COST), Dell (DELL),

Devon Energy (DVN), DuPont (DD), Dow Chemical (DOW), EMC Corp. (EMC), Entergy (ETR), Exelon (EXC), Exxon Mobil Corp. (XOM), FedEx (FDX), Ford Motor (F), General Dynamics (GD), General Electric (GE), Gilead Sciences (GILD), Goldman Sachs (GS), Halliburton (HAL), H. J. Heinz Co. (HNZ), Hewlett Packard Co. (HPQ), Home Depot (HD), Honeywell International (HON), Intel Corp. (INTC), International Business Machines (IBM), JP Morgan Chase & Co. (JPM), Johnson & Johnson Inc. (JNJ), Lockheed Martin (LMT), Lowe's Companies Inc. (LOW), McDonald's Corp. (MCD), Medtronic Inc. (MDT), Merck & Co. (MRK), Microsoft (MSFT), Morgan Stanley (MS), Nike (NKE), National Oilwell Varco Inc. (NOV), Norfolk Southern Corp (NSC), Occidental Petroleum Corp. (OXY), Oracle Corp. (ORCL), Pepsico Inc. (PEP), Pfizer Inc. (PFE), Procter & Gamble Co. (PG), Qualcomm Inc. (QCOM), Raytheon Co. (RTN), Regions Financial (RF), Sara Lee Corp. (SLE), Schering-Plough Corporation (SGP), Schlumberger (SLB), Southern Co. (SO), Sprint Nextel Corp. (S), Target Corp. (TGT), Texas Instruments (TXN), Time Warner Inc. (TWX), Tyco International (TYC), US Bancorp (USB), United Parcel Service Inc. (UPS), United Technologies Corp. (UTX), UnitedHealth Group Inc. (UNH), Verizon Communications Inc. (VZ), Wal-Mart Stores Inc. (WMT), Walgreen co. (WAG), The Walt Disney Co. (DIS), Wells Fargo & Co. (WFC), Weyerhaeuser Co. (WY), Williams Companies (WMB), Wyeth (WYE) and Xerox (XRX).