Multivariate GARCH with dynamic beta

M. Raddant\textsuperscript{a} and F. Wagner\textsuperscript{b}

\textsuperscript{a}Department of Economics, Kiel University, Kiel, Germany; \textsuperscript{b}Institute for Theoretical Physics, Kiel University, Kiel, Germany

\textbf{ABSTRACT}

We investigate a solution for the problems related to the application of multivariate GARCH models to markets with a large number of stocks by restricting the form of the conditional covariance matrix and by introducing a system of recursion formulas. The model is based on a decomposition of the conditional covariance matrix into two components and requires only six parameters to be estimated. The first component can be interpreted as the market factor, all remaining components are assumed to be equal. This allow the analytical calculation of the inverse covariance matrix. The factors are dynamic and therefore enable to describe dynamic beta coefficients. We compare the estimated covariances for the S&P500 market with those of other GARCH models and find that they are competitive, despite the low number of parameters. As applications we use the daily values of beta coefficients to confirm a transition of the market in 2006. Furthermore we discuss the relationship of our model with the leverage effect.

\textbf{ARTICLE HISTORY}

Received 22 November 2019
Accepted 16 January 2021

\textbf{KEYWORDS}

Model construction and estimation < econometric modeling portfolio choice; investment decisions < general financial markets; asset pricing; trading volume; bond interest rates < general financial markets

\section{1. Introduction}

The estimation of co-movement between stocks is an essential problem in the analysis of asset returns, financial integration, and for portfolio management. On the one side, the statistical properties of asset returns necessitate to treat them in a setting that includes time-varying volatility. The most common setting thus are GARCH models. In many analysis of co-movement, one would however also like to analyze samples that cover entire markets and which can thus be very large. This is at odds with many multivariate GARCH models like DCC (Engle 2002) and BEKK (Engle and Kroner 1995).

In any multivariate GARCH model, the co-movement of stocks is described by the dynamics of the conditional covariance matrix $H_t$ and by its mean $\bar{H}$. When the number of stocks $N$ becomes large, the high number of so-called nuisance parameters in $\bar{H}$ typically leads to non-negligible biases in the maximum likelihood estimation, as it involves the estimation of $O(N^2)$ parameters. Another issue for the estimations are the parameters that govern the dynamics of $H_t$. In the VECCH model (Bollerslev, Engle, and M 1988) the elements of $H_t$ are treated as an $N(N+1)/2$ vector. In its scalar version (Ding and Engle 2001), only two parameters remain. The $K$-factor BEKK model (Engle and Kroner 1995) uses $N$-dimensional matrix multiplication. Still, in this model the number of parameters can become very large, unless one uses the scalar version with a manageable number of $2K$ parameters.

Additionally, a computational problem is that most methods require the repeated inversion of $H_t$ as part of the estimation, which can become computationally costly or even prohibitive for some settings. A partial solution is provided by methods that use covariance targeting (Engle and Mezrich 1996). Here $\bar{H}$ is replaced by the observed time-averaged covariance $C$. This method also allows for covariances and GARCH parameters to be estimated in two separate steps.

Several approaches have worked on the scalability of the estimation of multivariate GARCH models, most notably Aielli (2013) who proposed improvements in the modeling of the conditional covariance matrix and...
Engle et al. (2020) who developed an approach that uses a composite likelihood and builds on the estimation of pairs of assets. A different solution is the method proposed by Engle, Ledoit, and Wolf (2019) which renders improved estimates by using nonlinear shrinkage.

While these advances are very significant the underlying problem remains. The ratio of the number of observations to the number of variables is not favorable for the estimation of dynamic correlations in large asset markets (see also Raddant and Kenett 2021). One approach to deal with this problem and to describe the dynamics with fewer parameters is the use of factor models. An example is the OGARCH model by Alexander (2001) which allows to drastically reduce the number of parameters by describing the cross-section of asset returns by a number of factors $k$ that can typically be chosen much smaller than $N$. This reduces the number of parameters that have to be estimated substantially. A different approach to reduce the number of parameters is to restrict the form for $H$ as in the DECO model (Engle and Kelly 2012). Here we assume identical yet time-varying correlations between all pairs of stocks and calculate the covariance matrix based on univariate de-garched returns.

While it is appealing to model $H_t$ with a reduced number of parameters, the two above-mentioned approaches have to compromise on either the precision with respect to the dynamics or the precision with respect to the cross-section of asset returns. We believe that a significant improvement in this respect can therefore only be achieved by changing the backbone of the model, namely the recursion formula that is used to derive $H_t$. In order to overcome the problems related to large sample sizes, the volatility of the single stocks relative to the market has to become part of the recursion and part of covariance targeting. Hence, it must be treated analytically rather than by estimation.

We thus propose a model in which the volatility of individual stocks relative to the market is part of a system of GARCH recursions. In particular, we model two dynamic factors which describe the market volatility and the remaining volatility. The volatility of the individual stocks is comprised of varying contributions of these two factors. Since we deal with only two factors we can analytically derive a recursion that consistently describes the changes in these contributions (equivalent to changes in a stock’s $\beta$) together with the changes in $H_t$. What remains to be estimated are only six parameters. In particular, these are two GARCH parameters for each of the factors and two additional parameters that describe the change in the stocks’ betas.

The structure of the model is loosely based on the BEKK model, yet incorporates two dynamic factors. We define one common volatility factor $v_{0t}$ and $N−1$ degenerate equal factors $v_{1t}$. This setting ensures that $H_t$ can be inverted analytically. Additionally, a rotation of the eigenvector that belongs to the $v_{0t}$ factor is added to the recursion. In contrast to the GO-GARCH model (van der Weide 2002), this rotation is time dependent and is determined dynamically. This part of the model therefore delivers time-dependent beta coefficients relative to the market in the spirit of a CAPM. The model dynamics lead to a coupled system of GARCH recursions for the factors $v_{0,1t}$ and a vector recursion for $\beta_t$.

The paper is organized in the following way. Section 2 describes our model and sketches out the similarities with and differences to some other GARCH models. Section 3 then shows the estimation with data of the daily returns of 356 stocks from the S&P market in the years 1995–2013. We then compare the estimation results and especially the implied covariances with those of other models in Section 4. Section 5 contains two applications. By using the dynamic betas we verify a transitions in the market observed by Raddant and Wagner (2017). As a second application, we investigate the relationship of the leverage effect with the stock beta. The last section contains conclusion.

2. Multivariate GARCH with restricted covariance

2.1. Model for the covariance matrix

We consider a time series of returns for $N$ stocks with length $T$ denoted by $r_{it}$ with $i = 1 \ldots N$ and $t = 1 \ldots T$. In a multivariate GARCH model, the vector $r_i$ is related to a vector of noise $\varepsilon_t$ by

$$ r_t = H_t^{1/2} \varepsilon_t $$

where $E(\varepsilon_t) = 0$ and $\text{Var}(\varepsilon_t) = I_N$. The $N \times N$ matrix $H_t$ corresponds to the conditional covariance of $r_{it}$. $H_t^{1/2}$ is positive definite.
The dynamics of $H_t$ are expressed by a recursion formula that relates $H_t$ to the values of $H_{t-1}$ and the returns at time $t-1$. A well-known example for such a model is the $K$-component BEKK(1,1,K) model (Engle and Kroner 1995) in which the recursion is normally stated like this

$$ H_t = \Omega + \sum_{k} A_k (r_{t-1}' r_{t-1} - H_{t-1}) A_k' + B_k H_{t-1} B_k' $$

(2)

The model that we are going to introduce is different from the BEKK model, however, the structural form of the recursion for $H_t$ has some similarities. As a starting point, consider the following specification:

$$ H_t = H_{t-1} + \sum_{k=0}^{K-1} \left[ A_k (r_{t-1}' r_{t-1} - H_{t-1}) A_k + G_k (\tilde{H} - H_{t-1}) G_k \right] $$

(3)

where $A_k$ and $G_k$ are time-independent symmetric $N \times N$ matrices. $\tilde{H}$ denotes the expected value of $H_t$.

The recursion exhibits clearly the fixed point $H_t = \tilde{H}$ if $r_t'r_t$ is replaced by its conditional expected value.

The parameters in $A_k$ describe how fast $H_t$ returns to its mean $\tilde{H}$ after a disturbance by $r_t'r_t$, $G_k$ describes the persistence of $H_t$ (similar yet not identical to $B$ in other GARCH models).

In the following, we will modify (3) by defining time-varying $A_k, G_k$ and by decomposing $H_t$ by using two projection matrices. The linearity of Equation (3) allows the application of some linear algebra. We can use this property by applying the following decomposition of symmetric matrices $M$ with dimension $N$ into products of the $K$ eigenvalues and $N \times N$ projection matrices $P_k$

$$ M = \sum_{k=0}^{K-1} \lambda_k P_k $$

(4)

which satisfy the orthogonality relations

$$ P_k \cdot P_l = P_k \delta_{lk} $$

(5)

and the completeness relation with the unit matrix $I$

$$ \sum_{k=0}^{K-1} P_k = I $$

(6)

The trace of $P_k$ gives the multiplicity of $\lambda_k$. Hence, in the non-degenerate case ($K = N$) this is just an alternative form to state the eigen-decomposition of a symmetric matrix and the $P_k$ are the dyadic products of the eigenvectors of $M$. For $K < N$, $P_k$ are independent of the particular choice of the eigenvectors. This formalism is particularly useful in cases where we have only one large eigenvalue $\lambda_0$ and $N-1$ degenerate (much smaller) eigenvalues $\lambda_1$. We can then apply Equation (6) and concentrate on only two projection matrices, namely $P_0$ and $P_1 = I - P_0$. This allows an analytic treatment of the system by using functions of $M$ (or $H_t$) where

$$ f(M) = \sum_{k=0}^{K-1} f(\lambda_k) P_k $$

(7)

The eigenvalue-spectrum of the covariance matrix is related to the projection matrices by

$$ C = \frac{1}{T} \sum_{t=1}^{T} r_t r_t' = \sum_{k=0}^{N-1} \lambda_k P_k $$

(8)

In Figure 1, we show the eigenvalue spectrum for the S&P data ($N = 356, T = 4782$). $C$ exhibits one large eigenvalue $\lambda_0$ and $N-1$ eigenvalues of order $\lambda_0/N$. Within the framework of a multivariate GARCH model, one can
interpret this behavior as if the true covariance matrix $E[H]$ had a single eigenvalue $\lambda_0$ and $N-1$ degenerate eigenvalues. The noise transforms them into an MP spectrum as described by Marčenko and Pastur (1967), which qualitatively agrees with the observed spectrum of $C$.

To arrive at a new form of $H_t$, we assume a two-component structure as in $C$ at each $t$ and write $H_t$ as

$$H_t = N v_0 t P_0 t + v_1 t (I - P_0 t)$$

with the projector $P_0 t$

$$P_0 t = \frac{1}{N} \beta t \beta t'$$

in terms of the first eigenvector $\beta$, which is normalized to $\beta t \beta t' = N$.

The factors are the eigenvalues of $H_t$. The first factor is $N v_0 t$. This component of the large eigenvalue of $C$ can be regarded as the market (see, e.g. Laloux et al. 1999) and therefore the first part in Equation (9) can be interpreted as a time-dependent market factor. When we project $r_t$ on the eigenvector $\beta t$ we obtain a market return $r_{Mt}$

$$r_{Mt} = \frac{1}{N} \beta t r_t$$

The $\beta$ of an asset $i$ is commonly defined as the volatility of the asset compared to the volatility of the market, in our case $r_M$, thus

$$\beta_i = \frac{\text{Cov}(r_i, r_M)}{\text{Var}(r_M)}$$

This is similar to the definition of a beta in a single-index model (Sharpe 1963) or the CAPM (Sharpe 1964; Lintner 1965) where beta is used to describe the risk of an investment relative to the market portfolio. The first (normalized) eigenvector of $H$ is identical to the statement in (12). An equivalent relationship also holds for the daily $\beta t$ and $H_t$ (see also Engle 2016, for a similar definition of conditional betas).

The second factor of our model is composed of the remaining eigenvalues of $H_t$. The $N-1$ degenerate factors $v_1 t$ together form the second component and show as the second term in Equation (9).

We can now apply this two component structure and modify the recursion into a factor model (see Lin 1992) with time-dependent $A_k t$ and $G_k t$ with

$$A_k t = \sqrt{\alpha_k} P_{k t} \quad \text{and} \quad G_k t = \sqrt{\gamma_k} P_{k t}$$

In order to make also the $\beta t$ in the recursion dynamic we have to add the off-diagonal terms that allow for a transition between $P_0 t$ and $P_1 t$. These terms contribute only to $r_t r_t'$ and $H$. After adding those terms our recursion
reads

\[ H_t = H_{t-1} + \sum_{k=0,1} \beta_t \left( r_{t-1} r_{t-1}' - H_{t-1} \right) P_{kt-1} \]
\[ + P_{0t-1} \left[ \alpha_0 r_{t-1} r_{t-1}' + \gamma_0 H \right] P_{1t-1} \]
\[ + P_{1t-1} \left[ \alpha_0 r_{t-1} r_{t-1}' + \gamma_0 H \right] P_{0t-1} \] (14)

This recursion for \( H_t \) obviously boils down to a system of recursions for \( v_{0,1t} \) and \( \beta_t \). By performing the operation \( tr \ P_{0,1t} \) on both sides of Equation (14) one obtains two equations for \( v_{0,1t} \). When we apply both sides of Equation (14) on \( \beta_{it} \) we find a vector \( D_{it} \) which determines the change of \( \beta_t \)

\[ D_{it} = \alpha_0 \ r_{Mt} (r_{it} - r_{Mt} \beta_{it}) + \gamma_0 \bar{m}_t \bar{v}_0 (\beta_t - \bar{m}_t \beta_{it}) \] (15)

The details of this derivation are described in Appendix 1. We state them assuming \( N \gg 1 \), which for all practical purposes is sufficient. The recursion for \( v_{0t} \) is then given by

\[ m_{t-1}^2 v_{0t} = R_{0t-1} \] (16)

with \( R_{0t} \) denoting the operation \( tr \ P_{0t} \) on the r.h.s. of Equation (14).

\[ R_{0t} = v_{0t} + \gamma_0 (\bar{m}_t^2 \bar{v}_0 - v_{0t}) + \alpha_0 (r_{Mt}^2 - v_{0t}) \] (17)

Equation (16) depends on the angle \( \phi_{t-1} \) between \( \beta_t \) and \( \beta_{t-1} \) given by the overlap \( \cos(\phi_{t-1}) = m_{t-1} = \beta_t \cdot \beta_{t-1} / N \). Similarly, \( \bar{m}_t = \bar{\beta} \cdot \beta_{t} / N \) corresponds to the angle between \( \beta_t \) and \( \bar{\beta} \). The overlap \( m_t \) is determined from the normalization condition \( \beta_t^2 \cdot \beta_t = N \) with

\[ m_t^2 = \left[ 1 + \frac{D_t \cdot D_t}{N R_{0t}^2} \right]^{-1} \] (18)

The recursion for \( \beta_t \) can then be expressed in terms of \( D_{t-1} \) as

\[ \beta_{it} = m_{t-1} \left[ \beta_{it-1} + \frac{D_{it-1}}{R_{0t-1}} \right] \] (19)

\( \beta_{it} \) can change only for deviations of \( r_{t-1} \) from the market return \( r_{Mt-1} \) and deviations of \( \beta_{t-1} \) from the mean value \( \bar{\beta} \). Since \( D_t \cdot D_t \propto N \) there is no \( N \)-dependence in Equation (18). When \( v_{0t} \) is known we can obtain the recursion for \( v_{1t} \):

\[ v_{1t} = v_{1t-1} - (1 - m_{t-1}^2) v_{0t} + \alpha_1 \left( \frac{r_{t-1}^2}{N} - r_{Mt-1}^2 - v_{1t-1} \right) \]
\[ + \gamma_1 (\bar{v}_1 + (1 - \bar{m}_{t-1}^2) \bar{v}_0 - v_{1t-1}) \] (20)

The returns \( r_t \) appear in the recursions (16) through the market component \( r_{Mt} \) and in (15) and (20) through the component perpendicular to \( \beta_t \).

### 2.2. Stability and estimation

The recursions in this model differ from those in factor models where the eigenvectors of \( H \) are constant. It is important to understand that \( \bar{v}_{0,1} \) and \( \bar{\beta} \) together with \( \bar{H} \) are determined by covariance targeting from the empirical covariance matrix \( C \). Also, due to the non-linearity of the recursions the results for \( \bar{v}_{0,1} \) and \( \bar{\beta} \) may differ from their expected values. The parameters \( \bar{v}_{0,1} \) and \( \bar{\beta} \) can be determined from the observed covariance
matrix \( C \) with a relative accuracy of \( \frac{1}{\sqrt{T}} \) (by comparing the eigenvalues of \( C \) and \( \tilde{C} \)), assuming slowly varying \( v_{0t} \) and \( \beta_t \). Details are provided in Appendix 3.

For the mean value \( \tilde{H} \), we assume the same restriction as for \( H_t \), in particular

\[
\tilde{H} = \tilde{v}_0 \tilde{\beta} \tilde{\beta}' + \tilde{v}_1 \left( I - \frac{\tilde{\beta} \tilde{\beta}'}{N} \right)
\]

(21)

For the stability of the recursions, we have to assume that \( \alpha_{01} \) is small, which implies autocorrelation of \( \varepsilon_t \), in line with the data. The recursions then have a fixed point solution \( v_{0,1t} = \tilde{v}_{0,1} \) and \( \beta_t = \tilde{\beta} \). A necessary condition for stability are the following inequalities for the GARCH parameters \( \alpha_k \) and \( \gamma_k \)

\[
0 < \gamma_k < \gamma_k + \alpha_k < 1
\]

(22)

The six parameters of our model can be estimated by maximum likelihood, the log likelihood \( L \) is given by

\[
L = \frac{1}{2} \sum t \left[ \sum_i \ln f(\varepsilon_{it}) - \ln(Nv_{0t}) - (N - 1) \ln v_{1t} \right]
\]

(23)

For details, see Appendix 2. We do not discuss consistency and efficiency of the estimator at this point. Valuable insights on the properties of estimators for multivariate GARCH models have been demonstrated by Pedersen and Rahbek (2014) and Francq, Horvath, and Zakoian (2016), their results are however not directly applicable to our model.

As a result of the estimation one obtains time series for \( v_{0,1} \) and \( \beta \). By using the relationships stated in Equations (9) and (10) one can then construct \( H \) element-wise (or time-series by time-series) for the stocks \( i \) and \( j \) as

\[
H_{ijt} = \frac{v_{0t} - v_{1t}}{N - 1} \beta_{it} \beta_{jt} \quad \text{for} \quad i \neq j
\]

\[
H_{ii} = \frac{v_{0t} - v_{1t}}{N - 1} \beta_{it} \beta_{jt} + \frac{Nv_{1t}}{N - 1} \quad \text{for} \quad i = j
\]

(24)

Since our model relies on the identification of the market factor as the first eigenvector of the covariance matrix, we will at last have a look at the properties of this eigenvector with respect to the sample size. In particular, we will use comparisons of beta values. This will give as a criterion to judge how well the market factor in our model represents the true market. First, we can compare the beta values based on the actual S&P500 index with those based on the first eigenvector. The left panel of Figure 2 shows a comparison. The plot confirms that the market factor and the S&P index contain very similar information since the beta values are a close fit. Second, we can compare beta values based on the first eigenvector from samples with a limited size with beta values from the full sample and beta values based on the S&P500 index. For this analysis, we draw 60 randomized samples for each sample size \( N \) and calculate the average absolute difference for each \( N \). The results are shown in the right panel of Figure 2. The difference \( |\beta_N - \beta_{S&P}| \) shows the difference in the values for beta that is due to the sample size and the difference of the market benchmark. Its minimum can be regarded as an approximation for the typical error in the determination of beta. The difference \( |\beta_N - \beta_{evo}| \) shows differences that only result from sample size variations, assuming that the first eigenvector represents the true market. As an indication for a reasonable minimum sample size for our model one can argue that \( N > 50 \) should result in an acceptable description of the market, since the changes in both measures after this point are getting smaller. Only little improvement in the absolute error can be expected for \( N > 100 \). Also, at this point the error caused by the sample size becomes relatively small compared to the error caused by using a different definition of the market index.

2.3. Relation to other models

A common class of multivariate GARCH models suitable for large samples are factor models. They are characterized by using constant \( P_k \) with rank 1 obtained from \( C \) by Equation (8) with the \( K \) largest eigenvalues. A model
structure similar to OGARCH (Alexander 2001) could be recovered from our model by defining $A_k = \sqrt{\alpha_k} P_k$ and $G_k = \sqrt{\gamma_k} P_k$ with time-dependent $\lambda_{kt}$. A conditional covariance matrix would then be calculated from the largest eigenvectors, which appear as the projection matrices in $H_t = \sum \lambda_{kt} P_k$. A slightly more complex approach are full-factor models like Vrontos, Dellaportas, and Politis (2003) where the conditional correlations obey a GARCH behavior, a feature which has some similarity with the behavior of our dynamic $\beta$.

The GO-GARCH model (van der Weide 2002) is applied to the de-correlated returns $\tilde{r}_t = C^{-1/2} r_t$. The $K$ principal components of $\tilde{r}$ are rotated by a $K$-dimensional rotation. The additional $K(K - 1)/2$ angles are additional parameters in MLE. Our model therefore may be called an effective two-factor model. It differs from these factor models by two features. The first is that our model always has two factors. Second, we use time-dependent $P_{kt}$. Equation (19) (or Equation (A11) in Appendix 1) corresponds to a time-dependent rotation with angle $\cos(\phi_t) = m_t$ as in a $K = 2$ GO-GARCH. However, $\phi_t$ is determined by the recursion and not by estimation.

For the large literature on other factor ARCH, GARCH and DCC models the reader is referred to Bollerslev, Chou, and Kroner (1992) and Zhang and Chan (2009).

Another class of models avoids numerical calculation of $H_t^{-1}$ by a restricted $H_t$ as in our model. An example is the DECO model of Engle and Kelly (2012). In the conditional correlation matrix $R_t$ pairwise correlations are equal

$$R_t = (1 - \rho_t) I + \rho_t$$

which allows analytical calculation of $R_t^{-1}$. In this model, first the diagonal matrix of conditional expectation values $D_t^2$ of $r_t^2$ is determined by $N$ univariate GARCH models. With this one can then calculate $H_t = D_t R_t D_t$ and obtain the average correlation $\rho_t$ from the DCC recursion. Our model could reproduce a DECO type behavior by setting $\beta_t = 1$ with

$$\rho_t = \left(1 + \frac{\nu_{1t}}{\nu_{0t}}\right)^{-1}$$

### 3. Estimation results for S&P stocks

#### 3.1. Preliminary fit and noise dependence

In this section, we describe the maximum likelihood estimation with data of daily returns of 356 stocks that were constituents of the S&P index for the years 1995–2013. For our analysis, we use data from Thompson Reuters on the closing price of stocks which were continuously traded with sufficient volume throughout the sample period and had a meaningful market capitalization.\(^3\)
Figure 3. The left panel shows the pdf of all GARCH filtered absolute returns $\eta_G(t_i)$ obtained from the raw returns by applying (27). The black line corresponds to a fitted t-distribution with $\nu = 3.32$, the broken gray line shows a Gaussian distribution. In the right panel, two typical pdfs for stocks of the S&P market are compared with the predicted pdf using t-distributed (black) or Gaussian (gray) noise (see Appendix 4 for calculation of errors). The values (log of the pdf) $|r| > 2.5$ are multiplied by a factor of 10. The theoretical curves are based on the two-parameter fits as shown in Table 1.

Figure 4. Market factor $\sqrt{\sigma^2}$ (top) and market return $|r_M|$ (bottom). A value of 2 is added to $\sqrt{\sigma^2}$.

Table 1. Maximum likelihood estimates of the parameters for the S&P market. Column 1 states the number of parameters, column 2 the values of the log likelihood per time relative to the six parameter fit. Standard errors are given in parentheses. The results in the two top rows are obtained with Gaussian noise, in rows 3–8 with t-distributed noise.

| $N_{par}$ | $L/T$ | $\alpha_0 \cdot 10^2$ | $\gamma_0 \cdot 10^2$ | $\alpha_1 \cdot 10^1$ | $\gamma_1 \cdot 10^2$ | $\alpha_0 \cdot 10^2$ | $\gamma_0 \cdot 10^2$ |
|----------|-------|-----------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| 2        | 52.2  | 4.871                 | 0.383                  | $\alpha_0$             | $\gamma_0$             | $\alpha_0$             | $\gamma_0$             |
|          |       | (0.068)               | (0.014)                |                        |                        |                        |                        |
| 2        | -2.57 | 3.132                 | 0.284                  | $\alpha_0$             | $\gamma_0$             | $\alpha_0$             | $\gamma_0$             |
|          |       | (0.048)               | (0.011)                |                        |                        |                        |                        |
| 4        | -0.10 | 1.592                 | 0.328                  | 2.472                  | 0.766                  | $\alpha_0$             | $\gamma_0$             |
|          |       | (0.042)               | (0.016)                | (0.050)                | (0.078)                |                        |                        |
| 6        | 0.00  | 5.29                  | 3.25                   | 2.490                  | 0.784                  | 1.732                  | 0.328                  |
|          |       | (0.16)                | (0.31)                 | (0.041)                | (0.065)                | (0.044)                | (0.014)                |

The scale of returns is arbitrary. We choose to normalize $r_i$ such that the average of $r_{it}^2$ over all $i$ and $t$ is equal to 1. This affects the scale in Figures 3, 4, 7, 9 and A1.

The initial values for $\nu_{0,1}$ and $\beta$ have been determined from the observed covariance matrix $C$ in the first 4 years. In a preliminary fit, we use only the two GARCH parameters of the first component so that $\alpha_0 = \alpha_1 = \alpha_{01}$ and $\gamma_0 = \gamma_1 = \gamma_{01}$ together with a Gaussian distributed noise. The resulting log likelihood (divided by $T$) and the parameter values are given in Table 1.
As a check on the property of this fit, we consider the so-called de-garched returns $\eta_{Gt}$. These are obtained by inverting Equation (1)

$$\eta_{Gt} = H_t^{-1/2} r_t$$

(27)

If the GARCH model represents the data exactly $\eta_{Gt}$ should be again Gaussian distributed. This is not the case, as the pdf in the left panel of Figure 3 shows. The pdf for $\eta_{Gt}$ is, however, well described by a Student’s t-distribution with a tail index of $\nu = 3.32$.

In the right panel, we show two typical pdfs of absolute returns and compare them with the predicted density either using t-distributed (black line) or Gaussian (broken gray line) noise. The latter describes the tails but fails grossly for the Gaussian region near $r = 0$ and in the transition to the tail.

In the following, we therefore use t-distributed noise in the estimation of our model. The estimated value of $\nu = 3.35$ agrees within the errors with the value obtained before. This improves the likelihood (2nd column in Table 1) dramatically. Even a very conservative evaluation using the probability change per $t$ given by $\exp(-\Delta L/T)$ is highly significant.

Such a strong noise dependence is not present in applications of univariate GARCH(1,1) models. There, the tail index can be reproduced either by t-distributed or Gaussian noise. However, since our model (hereafter called RMG for restricted matrix GARCH) has only four GARCH parameters, $\alpha_{0,1}$ and $\gamma_{0,1}$, the tail cannot be matched for all $N$ stocks when Gaussian noise is used (see also Wagner, Milaković, and Alfarano 2010, for a non-parametric moment analysis of stocks and indices).

### 3.2. Fit of the full model and evaluation

To make sure that our model describes the distribution of returns adequately we chose to simulate the returns and to compare them to the empirical ones.

Equation (1) predicts $r_t$ for given $H_t$ from any GARCH model and noise at time $t$. This relationship can easily be utilized for a Monte Carlo simulation. By multiplying $H_t^{-1/2}$ with a vector of noise we obtain simulated predicted returns $\hat{r}_t$. When we repeat this process with different noise $n_{sim}$ times we obtain a range of possible values for the returns. By performing this procedure at all $t$ we obtain $n_{sim}T$ predictions for $\hat{r}$. Individual time series cannot be obtained, however, average values as pdf should agree with the observed quantities.

As we will see in Section 4, this procedure can also be used to compare distributions of covariances for different models. To obtain smooth distributions we found that $n_{sim} = 10$ is sufficient for the pdf of $\hat{r}$, covariances require $n_{sim} = 40$.

The drawback of the preliminary two parameter fit consists in the small value for $\gamma_0$. The estimate corresponds to autocorrelation that lasts for 339 days, which is roughly twice as long as we would expect from the observed individual $|r_{it}|$. The likelihood improves when we add $\alpha_{01} = \alpha_0$ and $\gamma_{01} = \gamma_0$ to the estimation (third row in Table 1). However, only after including all six $\alpha$ and $\gamma$ we obtain a reasonable value for $\gamma_0$ together with another increase in the likelihood. The estimates for the parameters that govern the dynamics of $\beta$ are smaller by comparison, indicating that $\beta$ varies much less than the factors $v_{0,1}$. As in any GARCH model, these exhibit much less fluctuations than the underlying returns. This is shown in Figure 4 where the market factor $\sqrt{v_{0t}}$ is compared with $r_{Mt}$.

More details on the autocorrelation are shown in Figure 5, where we compare the GARCH filtered returns $|\eta_{it}|$ with $|r_{it}|$ averaged over all stocks $i$. The large statistics allow to resolve the peaks at multiples of 3 months, which are caused by the dividend pay days. These correlations are outside of the realm of any GARCH and survive the filtering. One can, however, see that the autocorrelation in $\eta$ is not perfectly removed. The reason is that with just six parameters the model produces the correct volatility for the stocks on average, but mutes (inflates) episodes with very high (low) volatility since the betas move slower than $v_0$ and some idiosyncratic effects are missed. As we will see in Section 4.2, this particular weakness however does not carry over to the estimated covariances.
Figure 5. Autocorrelation for returns $|r_t|$ averaged over all stocks (top) and the averaged GARCH filtered returns $|h_t|$ (bottom).

Figure 6. Beta values from the RMG model. The top panels and the bottom left panel show beta values for different stocks, averaged over a 60-day window. The bottom right panel shows the dynamics of the distribution of beta values by the quantiles, using a 20-day window.

3.3. Beta values for financials and IT companies

Figure 6 finally gives an overview about the beta values that can be obtained from the RMG. Note that these betas are normalized to $\beta_t' \cdot \beta_t = N$.

The top left panel shows the development of beta values for some financial stocks, the top right shows the beta of IT-related companies. Stocks from both groups show strong similarity within their group. The bottom left panel shows the beta for stocks from other sectors. They develop much more diverse. The bottom right panel illustrates how the distribution of beta values has developed over time. At the times of the IT bubble and the Lehmann crisis, we observe peaks for the beta values and also a much wider distribution.

4. Comparison with other GARCH models

In the following, we will compare the estimation results of our model with those of other multivariate GARCH models. We will look at the predicted returns and covariances. Here we are interested in the practical implications
of the dynamics of the model for application. In particular, we are interested in a rough approximation of the precision of the estimated covariances with respect to the empirical data.

4.1. Comparison of distributions of returns

We start by comparing the predicted pdf of stock returns from the RMG model discussed in Section 3 with those of the univariate GARCH(1,1) (abbreviated UVG) and OGARCH model (hereafter abbreviated OG). In both cases we use i.i.d. Gaussian \( \varepsilon \) for the noise.

The left part of Figure 7 shows that the pdf for the same two stocks as in Figure 3 is in reasonable agreement with the three models. RMG and OG fail in cases where the leading eigenvector does not dominate \( H_t \), as the third example for PG&E shows.

For a more comprehensive comparison, we use the \( \chi^2/n_d \) ratio, which is the sum of quadratic deviation in units of the squared error divided by the number \( n_d \) of bins. A histogram that shows how well the distributions of predicted returns fit the data measured by these ratios for all 356 stocks of the S&P market is shown for RMG, OG and UVG in the right panel of Figure 7. All three distributions exhibit a peak around a value of 4–5 which corresponds to the 5% confidence level. Values between 10–20 lead still to a qualitative description (see Appendix 4 for details on the calculation of the statistics). In contrast to UVG, both RMG and OG have a small fraction (5%) of outliers mainly from the energy sector as PG&E. On average RMG performs better than OG which may be due to the systematic error by covariance targeting. We stress that for UVG and OG 712 parameters have to be determined. The only six parameters used in RMG lead to a much more parsimonious description of the data.

4.2. Comparison of covariances

In the following, we compare the covariances obtained from the RMG model with those obtained from the DCC, OG and DECO models, and those implied by the empirical data. The form of \( H_t \) in RMG, OG and DECO is in each case in some way restricted, which influences the estimated covariances. We therefore consider the standard DCC model, applied to subsamples of the data, as a benchmark estimation.

Even if the DCC model allows relatively large \( N \), it is admitted not designed for a sample like ours. It is typically applied for (and works best) for \( N \leq 10 \) (see Aielli 2013; Zhang and Chan 2009). For approaches to overcome this problem by using modified estimators, composite likelihood and shrinkage methods we refer the reader to Engle et al. (2020) and Engle, Ledoit, and Wolf (2019). Including these models into the comparison would require an in-depth model comparison and selection process that is outside the scope of this paper. We can expect that these models would yield estimation results that are marginally better than those of the DCC presented here, which is however not decisive for our comparison. For computational reasons, the comparisons
in this section (with exception of Figure 8) are based on subsamples of covariances. We can therefore apply the DCC model to blocks of 8 stocks. The division of the stocks into these blocks is random. With a total number of 44 blocks, we cover 352 stocks and 1232 covariances. The RMG, OG and DECO models have been estimated for the entire sample of stocks, the comparison of covariances however is always based on the exact same 1232 covariances. By repeating the sampling, we have verified that the results presented in this section do not depend on the sampling.

We will first have a look at the dynamics of the covariances. Figure 8 shows the average correlation $\rho_t$ together with the 90% interval for DCC, RMG and OG. The average correlation from DECO (not shown) is almost identical to that of the DCC model. The overall dynamics of the DCC and the RMG model seem to be rather similar. RMG exhibits larger fluctuations since it describes all pairs with six parameters. The OG model leads to qualitatively very different dynamics of the correlations.

It is not possible to compare the dynamics of the covariances to the data directly since we do not observe the true $H_t$. We can, however, judge the models by comparing the distributions of simulated $\hat{r}_{it}\hat{r}_{jt}$, based on the estimation, with distributions of the observed $r_{it}r_{jt}$. Further we can compare the single elements of the simulated covariances matrix $\hat{C}_{ij}$ (the time-averaged covariance for the stocks $i$ and $j$) with the observed covariance $C_{ij}$. To achieve this comparison, we thus perform the Monte Carlo simulation described in Section 3 to obtain all $\hat{C}_{ij}$ and distributions of $\hat{r}_{it}\hat{r}_{jt}$.

In Figure 9, we show scatterplots comparing $\hat{C}_{ij}$ for the models with the observed time averaged covariances $C_{ij}$. Since $C$ is input in OG and is used for the mean of $H$ in DCC by covariance targeting we observe almost straight lines broadened by noise for these two cases. The somewhat larger deviation seen for RMG is partly due to the systematic error of the random matrix $C$. In RMG only the properties of the large eigenvalue of $C$ are used. The performance of DECO is similar and is governed by the assumption of stock-independent correlations. We can quantify this observation by calculating the differences between simulated and empirical covariances as $RMSE(C_{ij} - \hat{C}_{ij})$ given in the top part of Table 2.
Table 2. The upper part of the table shows the differences between the estimated versus the empirical covariances. We also report the average percentage deviation of this difference with respect to $C_{ij}$. The bottom part of the table reports the differences in the distributions of the covariances, applying Cliff’s delta test. The first line gives the average value, the second line the average absolute difference together with the standard deviation.

|          | RMG    | DECO   | DCC    | OG     |
|----------|--------|--------|--------|--------|
| RMSE$(\hat{C}_{ij} - C_{ij})$ | 2.1390 | 3.0633 | 1.4018 | 0.9513 |
| $|\hat{C}_{ij} - C_{ij}|$   | 6.88  % | 11.03 % | 5.97  % | 4.44  % |
| $\Delta_{ij}$          | 0.0091 | 0.0570 | 0.0558 | 0.1395 |
| std$(\cdot)$            | 0.079  | 0.031  | 0.031  | 0.054  |
| $|\Delta_{ij}|$         | 0.0650 | 0.0575 | 0.0562 | 0.1395 |
| std$(\cdot)$            | 0.046  | 0.030  | 0.030  | 0.054  |

A more interesting comparison is to look at the distributions of the covariances. These distributions have pronounced peaks around 0 which are very dominant but not very decisive for our comparison. We have therefore decided to look at absolute covariances, in particular $\log(|r_{it}r_{jt}|)$ for $|r_{it}r_{jt}| > 0$. Three examples for this distribution are shown in the figure in Appendix 5.

We measure the difference between the predicted distributions and those given by the data by calculating Cliff (1999)’s delta, denoted by $\Delta_{ij}$. This statistic relies on comparing the elements in both distributions and counting how often each element is larger (smaller) than any element from the other distribution. Delta is bound within $[-1,1]$, a value of 0 signal distributions that are indistinguishable while deviations from 0 signal some degree of imperfection in the overlap of the two distributions.

$$\Delta_{ij} = \frac{1}{n_t n_T} \sum_{i,t} [|\hat{r}_{it}\hat{r}_{jt}| > |r_{it}r_{jt}|] - [|\hat{r}_{it}\hat{r}_{jt}| < |r_{it}r_{jt}|]$$

In the bottom part of Table 2, we report the mean of $\Delta_{ij}$ and $|\Delta_{ij}|$ as well as the respective standard deviation. RMG has a $\Delta_{ij}$ very close to zero, while the values of DECO and DCC are slightly larger than zero. The dispersion of $\Delta_{ij}$ however is larger for RMG than for the other models. Judging by the absolute $\Delta_{ij}$ we can see that RMG achieves results that are only marginally behind DECO and DCC. The OG model for comparison performs noticeably worse. Hence, while RMG is not the winner in this comparison, the results indicate that it does achieve a level of precision that is competitive with other GARCH models and is thus a useful alternative when dealing with large sample sizes.

5. Applications

5.1. Market transition

One benefit of our model is that the large number of stocks in our sample allows to search for group-specific regularities. An obvious question is if the correlations that can be extracted from $H_t$ differ for stocks from specific sectors, and more interestingly, how they develop over time. Our sample period covers a time period during which we have seen pronounced changes in the market, including the IT bubble and the financial crisis. The visual inspection of some stocks’ betas in Section 4 has already hinted at changes in the IT and the financial sector, which we want to analyze in more detail. For this reason, we use the estimated $H_t$ and then derive the conditional correlation matrices from $D_t^{-1}H_tD_t^{-1}$, where $D$ is a diagonal matrix with the square root of $H_{ii}$ on the main diagonal. We can use these correlations to calculate the median correlation for pairs of stocks from specific sectors (GICS classification).

Figure 10 shows some of these median correlations. We observe very high average within-sector-correlations for stocks in the IT and financial sectors. But also some correlations between stocks of different sectors are rather high, for example when the consumer or materials sectors are involved.

In general, we observe two important changes over time. First, the overall level of correlations has shifted upwards from 2002 until 2007. The second observation is that the relative contribution of different sectors to
THE EUROPEAN JOURNAL OF FINANCE

5.1. Development of median correlation by sector, averaged over 50 days. We show the average of all stock correlations (bold gray line), as well as some of the sectors with the strongest correlations. The sector with the strongest inter-sector correlation has for a long time been the IT sector (blue), later the financial sector (red) has taken over this role. The overall correlation has changed. The most important change is that the financial sector surpasses the IT sector in terms of correlation around 2006.

These changes can be analyzed in more detail. In the analysis of Raddant and Wagner (2017) of the US, the UK, and the German stock markets the same change has been found in the behavior based on the stock’s beta values, derived under the assumption that the covariance matrix of the returns has one large eigenvalue already at a time window sizes of 3 years. In the years 1994–2006 stocks with high trading volume and high beta mainly came from the information technology sector, whereas in 2006–2012 such stocks came mainly from the financial sector. Since a \( \beta > 1 \) signals a risky investment, a market risk measure \( \hat{R} \) has been defined for the sectors \( s \) by multiplying \( \beta_i > 1 \) with the number \( V_{it} \) of traded shares in each time window.

\[
\hat{R}_s = A_S \sum_{i \in s} \theta (\beta_i - 1.0) \beta_{it} V_{it}
\]

(29)

The normalization constant \( A_S \) is chosen to have \( \sum_s \hat{R}_s = 1 \). When we apply this measure, we see that before 2006 only the information technology sector and after 2006 only the financial sector exhibit large values of the risk measure.

However, the time and the duration of this transition had a systematic error of 1.5 years due to the window size. Repeating the calculation of \( \hat{R} \) with the \( \beta \) obtained from the RMG model serves two purposes. First, it is a check whether in RMG the market property can be reproduced. Second, the transition time can be determined more accurately, since daily \( \beta \) are known from RMG. To reduce the noise on \( \beta \) we average \( \hat{R}_s \) over 1 month. In Figure 11, the risk parameters from Equation (29) for the S&P market are shown as a function of time. The agreement with the previous determination (dashed lines) is good. With a time resolution of 1 month we can now safely say that the transition happens during the year 2006.

5.2. Leverage effect

The leverage effect consists in a negative correlation between volatility and future returns (Black 1976; Bekaert and Wu 2000). It is a relatively small effect (Schwert 1989), but important for the estimation of risk. Since GARCH models provide a measurement of the daily volatility they are well suited for an analysis of this effect.

In a first step, we determine for each stock the time correlation \( L_{it} \) between the market volatility \( v_0 \) and the observed returns \( r_{it} \):

\[
L_{it} = \frac{1}{N_L} \sum_{t'} (v_{0(t'-t)} - \bar{v}_0) r_{it'}
\]

(30)
with the normalization factor \( N_L^2 = T \cdot \text{var}(v_0) \sum_t r_t^2 \). Empirically the \( L_t \cdot \text{sign}(t) \) are dominantly negative with large fluctuations. To improve the sensitivity, we use the asymmetry defined by

\[
A_i = \frac{1}{t_m} \sum_{t=1}^{t_m} \left( L_t - L_{i(-t)} \right)
\]

with a maximum of the time lag \( t_m \) of 2 months. \( A_i \) corresponds to the difference in the area under \( L_t \) for positive and negative \( t \). Due to the time symmetry of our model, using either \( r_t \) from Equation (1) or the filtered returns \( \eta \) should eliminate the leverage effect.

It has been suggested by Black that the leverage effect is related to risk. To test this suggestion, we show in the left panel of Figure 12 the asymmetry as a function of the mean value \( \bar{\beta}_i \). Since \( \bar{\beta}_i \) changes around 2006 we use two time series, one covering the years 1995 until 2005 and a second one covering 2007 until 2013. The asymmetries are clearly negative and increase slightly with \( \bar{\beta} \) as indicated by the line connecting the mean of \( A \). By replacing \( r_t \) by \( \eta \) in Equation (30) the effect should disappear. As shown in the right panel this is in fact the case.

The leverage effect can be included in the recursion of our model. Analogous to the GJR-GARCH model (Glosten, Jagannathan, and Runkle 1993) an additional matrix proportional to \( P_0, t \cdot \text{GJR} \cdot P_1, t \) with \( \text{GJR}_{ij} = \delta_{ij} |r_{it}|/|r_{it}| \) can be added on the r.h.s. of Equation (A8). For large \( N \) the recursions for \( v_0 \) and \( \beta \) are unchanged,
only $v_1$ is affected. We have repeated the fit including such a term. The likelihood improves, however, the values of $\alpha$ and $\gamma$ are inside the errors the same. Also the results contained in Figure 12 remain the same.

6. Conclusion

In our GARCH model, we split the return into a market component and the remainder, similar to a two-factor model. The parametrization of the covariance matrix $H$ avoids the numerical inversion of $H$. The small number of parameters and a computing time $T_{\text{comp}} \propto N$ allow the application to markets with a large number of stocks. The off-diagonal elements of our $H_t$ have a precision that is competitive with other established models. Further, our model has the advantage that daily $\beta$ values relative to the market are determined. We found that replacing Gaussian by t-distributed noise is essential for the quality of the estimation results.

A possible development of the model would be to follow a similar strategy like DECO and to estimate univariate $h_t$ in a first stage and to base the actual estimation of the model on residuals in a second stage.

Notes

1. Note that in order to ease notation for the rest of the paper we directly use $r$ in these recursion. Hence, we pretend that these $r$ are identical to the increments of some vector stochastic process $\{y_t\}$. Note also that (3) only becomes a properly defined process for the covariance matrix once we define the process for the market volatility, which indirectly imposes restrictions on $A$, $G$ and $H(0)$ and leads to the recursion in (14), see also Appendices 1 and 3.

2. Our qualitative findings hold for variations of the length of the sample size. Similar conclusion on the effect of the sample size can also be drawn from comparing the relative size of the leading eigenvalue.

3. We excluded stocks which price did not change for more than 8% of the trading days, or which were exempt from trading or for which no trading was recorded for more than 10 days in a row. We manually deleted 15 stocks which price movements at some point showed similarities to penny stocks and/or which market capitalization was very low.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Aielli, Gian Piero. 2013. “Dynamic Conditional Correlation: On Properties and Estimation.” Journal of Business & Economic Statistics 31 (3): 282–299. doi:10.1080/07350015.2013.771027

Alexander, C. 2001. Market Models. John Wiley: New York.

Bekaert, Geert, and Guojun Wu. 2000. “Asymmetric Volatility and Risk in Equity Markets.” The Review of Financial Studies 13 (1): 1–42.

Black, Fischer. 1976. “Studies of Stock Market Volatility Changes.” In Proceedings of the American Statistical Association, Business and Economic Statistics Section, 177–181.

Bollerslev, Tim, Robert F. Engle, and Jeffry M. 1988. “Wooldridge. A Capital Asset Pricing Model with Time-varying Covariances.” Journal of Political Economy 96 (1): 116–131.

Bollerslev, Tim, Ray Y. Chou, and Kenneth F. Kroner. 1992. “ARCH Modeling in Finance.” Journal of Econometrics 52: 5–59.

Cliff, Norman. 1999. “Dominance Statistics: Ordinal Analyses to Answer Ordinal Questions.” Psychological Bulletin 114 (3): 1–42.

Ding, Zhuanxin, and Robert F. Engle. 2001. “Large Scale Conditional Covariance Matrix Modeling, Estimation and Testing.” working paper FLN-01-029, NYU Stern School of Business.

Engle, Robert F. 2002. “Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models.” Journal of Business & Economic Statistics 20 (3): 339–350.

Engle, Robert F. 2016. “Dynamic Conditional Beta.” Journal of Financial Econometrics 14 (4): 643–667.

Engle, Robert F, and Bryan Kelly. 2012. “Dynamic Equicorrelation.” Journal of Business & Economic Statistics 30: 212–228.

Engle, Robert F., and Kenneth F. Kroner. 1995. “Multivariate Simultaneous Generalized ARCH.” Economic Theory 11: 122–150.

Engle, Robert F., and Joseph Mezrich. 1996. “GARCH for Groups.” Risk (Concord, NH) 9: 36–40.

Engle, Robert F., Olivier Leduc, and Michael Wolf. 2019. “Large Dynamic Covariance Matrices.” Journal of Business & Economic Statistics 37: 363–375.

Engle, Robert F, Cavit Pakel, Kevin K. Shephard, and Neil Shephard. 2020. “Fitting Vast Dimensional Time-varying Covariance Models.” Journal of Business and Economic Statistics forthcoming. doi:10.1080/07350015.2020.1713795

Franco, Christian, Lajos Horvath, and Jean-Michel Zakoian. 2016. “Variance Targeting Estimation of Multivariate GARCH Models.” Journal of Financial Econometrics 14: 353–382.
Appendices

Appendix 1. Derivation of the recursions

We calculate $v_{0,1t}$ and $\beta_t$ contained in the conditional covariance matrix

$$H_t = Nv_{0t}P_{0t} + v_{1t}(I - P_t)$$  \hspace{1cm} (A1)

from the recursion (14)

$$H_t = H_{t-1} + \sum_{k=0,1} P_{kt-1} \left[ \alpha_k \left( r_{t-1}r_{t-1}' - H_{t-1} \right) + \gamma_k \left( \tilde{H} - H_{t-1} \right) \right] P_{kt-1}$$

$$+ P_{0t-1} \left[ \alpha_{10} r_{t-1} r_{t-1}' + \gamma_{10} \tilde{H} \right] P_{1t-1}$$

$$+ P_{1t-1} \left[ \alpha_{10} r_{t-1} r_{t-1}' + \gamma_{10} \tilde{H} \right] P_{0t-1}$$  \hspace{1cm} (A2)

The mean $\tilde{H}$ is given by Equation (21). We apply the operation $\text{trace}P_{kt-1}$ on both sides of Equation (A2). The computation can be done using the algebraic properties of $P_k$ shown in Equations (5) and (6). This leads to

$$\text{trace}H_tP_{0t-1} = Nm_{t-1}^2 v_{0t} + v_{1t}(1 - m_{t-1}^2)$$

$$= N R_{0t-1}$$  \hspace{1cm} (A3)

$$\text{trace}H_tP_{1t-1} = N(1 - m_{t-1}^2) v_{0t} + v_{1t}(N - 2 + m_{t-1}^2)$$

$$= N R_{1t-1}$$  \hspace{1cm} (A4)

where $m_{t-1} = \beta'_{t-1} \cdot \beta_{t-1}/N$ denotes the overlap between $\beta_{t-1}$ and $\beta_t$. The $R_{0,1t}$ represents the result of the same operation on the r.h.s. of Equation (A2) involving only the diagonal terms given by

$$R_{0t} = (1 - \alpha_0 - \gamma_0) v_{0t} + \alpha_0 r_{3tt}^2 + \frac{\gamma_0}{N} (N\bar{m}^2_t \bar{v}_0 - (1 - \bar{m}^2_t) \bar{v}_1)$$  \hspace{1cm} (A5)

$$R_{1t} = \frac{N - 1}{N} (1 - \alpha_1 - \gamma_1) v_{1t} + \alpha_1 \left( \frac{r_{3tt}^2}{N} - r_{3tt}^2 \right)$$

$$+ \frac{\gamma_1}{N} (N(1 - \bar{m}^2_t) \bar{v}_0 + (N - 2 + \bar{m}^2_t) \bar{v}_1)$$  \hspace{1cm} (A6)
The determination of the mean values \( \bar{v}_k \) and \( \bar{\beta} \) from \( \bar{H} \) in Equation (21) is described in Appendix 3. \( \bar{m}_t \) denotes the overlap between \( \beta_t \) and \( \bar{\beta} \). Equations (A7) and (A8) determine \( \bar{v}_k \) from the quantities \( R_{kt-1} \) and the overlap \( m_t^{2-1} \).

\[
\bar{v}_0 = \frac{1}{N m_t^{2-1}} \left[ (N - 2 + m_t^{2-1}) R_0 - (1 - m_t^{2-1}) R_1 \right] \quad (A7)
\]

\[
\bar{v}_1 = \frac{1}{N m_t^{2-1}} \left[ m_t^{2-1} R_1 - (1 - m_t^{2-1}) R_0 \right] \quad (A8)
\]

To find \( \beta_t \) and thereby \( m_t^{2-1} \) we apply \( H_t \) to \( \beta_{t-1} \) and subtract \( \beta_{t-1} \) trace \( P_{0t-1} H_t \) to obtain

\[
\frac{1}{N} \left( H_t \beta_{t-1} - \beta_{t-1} \text{trace} \, H_t P_{0t-1} \right) = \bar{w}_i (\beta_{t-1} - m_t^{2-1} \beta_t) = D_{t-1} \quad (A9)
\]

where we used the abbreviation \( \bar{w}_i = \bar{v}_0 - \bar{v}_1 / N \). \( D_{t-1} \) represents the same operation on the r.h.s. of Equation (A2) involving only the non-diagonal terms. It is given by

\[
D_{t-1} = \alpha_0 r_M (r_{t-1} - r_M \beta_{t-1}) + r_{t-1} \beta_{t-1} \quad (A10)
\]

This leads to the following recursion for \( \beta_t \).

\[
\beta_{t} = m_{t-1} \beta_{t-1} + \frac{D_{t-1}}{w_i m_{t-1}} \quad (A11)
\]

The unknown \( m_t \) follows from the normalization of \( \beta_t \cdot \beta_t = N \). By some miracle the equation for \( m_t \) is only quadratic in \( m_t^2 \) despite the complicated structure of Equations (A7) and (A8). It reads

\[
\left( R_{0t} - \frac{R_{0t} + R_{t-1}}{N} \right)^2 m_t^2 (1 - m_t^2) = \frac{D_t^2}{D_{t-1}} \frac{D_t}{R_{0t}} \left( m_t^2 - \frac{1}{N} \right)^2 \quad (A12)
\]

One of the two solutions with \( N m_t^2 \leq 1 \) is to be excluded. Since \( H_t \) is insensitive to the sign of \( \beta_t \) we always have a positive overlap \( m_{t-1} \). With the correct solution of \( m_t \) the set of Equations (A7), (A8), (A11) and (A12) yields \( v_{0,1t} \) and \( \beta_t \) in terms of \( t_{t-1} \) and \( \bar{H} \). The explicit solution looks rather complicated due to the dependence on \( N \) in Equation (A12). For our purpose the limit \( N \gg 1 \) is sufficient. In this case we have

\[
m_t^{2-1} \bar{v}_0 = R_{t-1} \quad (A13)
\]

With \( D_{t-1} \) from Equation (A10) we obtain \( m_t \) by

\[
m_t^2 = \left( 1 + \frac{D_t^2}{N R_{0t}} \right)^{-1} \quad (A14)
\]

Finally \( \beta_t \) and \( v_{11} \) are found by

\[
\beta_{11} = m_{t-1} \left( \beta_{t-1} + \frac{D_{t-1}}{R_{0t-1}} \right) \quad (A15)
\]

Inserting \( R_{0t,1t} \) leads to the formula given in Section 2.

Necessary conditions on the GARCH parameters \( \alpha_0 \) and \( \gamma_k \) are obtained by setting \( r_{M1}^2 = \bar{v}_0 \) and \( r_{1}^2 = N(\bar{v}_0 + \bar{v}_1) \) to its mean values. Then the recursions have a fixed point \( v_{11t} = \bar{v}_k \) and \( \beta_t = \bar{\beta} \) which is stable for

\[
0 < \gamma_k < \gamma_k + \alpha_k < 1 \quad (A16)
\]

**Appendix 2. Calculation of the likelihood**

The log likelihood with noise distribution \( f(\varepsilon) \) can be computed using

\[
\varepsilon_t = H_t^{-1/2} r_t \quad (A17)
\]

With a spectral decomposition for \( H_t \) functions of \( H_t \) can be calculated analytically. For \( H_t^{-1/2} \) we obtain

\[
H_t^{-1/2} = \frac{1}{\sqrt{N v_0}} P_{0t} + \frac{1}{\sqrt{v_1}} P_{1t} \quad (A18)
\]

It is easy to verify \( (H_t^{-1/2})^2 \cdot H_t = 1 \) with the orthogonality relations for \( P_{0t,1t} \). For \( \varepsilon \) we obtain

\[
\varepsilon(t) = \frac{r_M}{\sqrt{N v_0}} \beta_t + \frac{1}{\sqrt{v_1}} (r_t - r_M \beta_t) \quad (A19)
\]
The log likelihood $L$ is given by

$$L = \frac{1}{2} \sum_t \left[ \sum_i \ln f(\varepsilon_{it}) - \ln(N\nu_{it}) - (N - 1) \ln \nu_{1t} \right]$$  \hspace{1cm} (A20)$$

For Gaussian noise calculation of the $\ln(f)$ can be avoided. The complication of using $t$-distributed noise leads to an negligible increase of computing time compared to the calculation of $v_{0,1t}$ and $\beta_t$. In any case the computing time increases only with $T \cdot N$.

**Appendix 3. Covariance targeting for the restricted covariance matrix**

For the time averaged covariance matrix $\tilde{C}$ of our model computed from

$$r_t = \tilde{H}^{1/2} \tilde{\varepsilon}_i$$  \hspace{1cm} (A21)$$

we decompose the noise $\varepsilon$ into a component $\varepsilon_{Mt}$ parallel to $\tilde{\beta}$ and a component $\varepsilon_{it}$ perpendicular to $\tilde{\beta}$. Both are i.i.d. with mean zero and variance one for large $N$. For $\tilde{C}$ we obtain

$$\tilde{C} = \frac{1}{T} \sum_t \left[ N \nu_{0t} P_{0t} \varepsilon_{Mt}^2 + \nu_{1t} \varepsilon_{0t} \varepsilon_{it} \right]$$
$$+ \sqrt{\nu_{0t} \nu_{1t}} \varepsilon_{Mt} (\beta_t \tilde{\varepsilon}_i + \tilde{\varepsilon}_i \beta'_t)$$  \hspace{1cm} (A22)$$

The statistical uncertainty of $\tilde{C}$ is mainly due to the noise and much less due to the fluctuations of the slowly varying $v_{0,1t}$ and $\beta_t$. For a rough estimate we neglect in the time average the $\varepsilon$-dependence in the latter. For the $tr(\tilde{C})$ we obtain for large $N$

$$tr(\tilde{C}) = \frac{1}{T} \sum_t \left[ N \nu_{0t} \varepsilon_{Mt}^2 + \nu_{1t} \sum_i \varepsilon_{it}^2 \right]$$  \hspace{1cm} (A23)$$

The law of large numbers leads for the average over the noise to

$$tr(\tilde{C}) = \frac{N}{T} \sum_t (\nu_{0t} + \nu_{1t}) = N(\tilde{\nu}_0 + \tilde{\nu}_1)$$  \hspace{1cm} (A24)$$

where we replaced the time average of $v_{0,1t}$ by its mean $\tilde{\nu}_{0,1}$. Comparing Equation (A24) with the observed covariance matrix gives one relation for $\tilde{\nu}_{0,1}$ with a relative error of $1/\sqrt{T}$. The same procedure leads for the third term in Equation (A22) to a contribution of order $1/\sqrt{T}$ since the expectation value of $\varepsilon_{it} \varepsilon_{Mt}$ vanishes. Replacing the second term $\nu_{1t}$ by $\tilde{\nu}_1$ leads to a random matrix $R$ with a Marčenko–Pastur-spectrum. The average of the first term in Equation (A22) is equal to $N\tilde{\nu}_0 \tilde{P}_0$. The resulting $\tilde{C}$ reads as

$$\tilde{C} = N \tilde{\nu}_0 \tilde{P}_0 + R$$  \hspace{1cm} (A25)$$

Since $|R| < < N$ we can determine $\tilde{\nu}_0$ and $\tilde{\beta}$ by comparing the leading eigenvalue and eigenvector of $\tilde{C}$ and the observed $C$. The statistical error is in the order of $1/\sqrt{T}$.

**Appendix 4. Experimental distributions**

For comparing empirical data with simulations, we apply a qualitative criterion which allows to locate eventual disagreement and is less sensitive to systematic errors as outliers. We assume $T$ independent measurements of an observable $x(t)$ ($t = 1, T$). The values $x(t)$ are binned into $M$ bins of width $\Delta x_m$ centered around $x_m$ containing $n_m$ events. The probability density $f_m$ at $x_m$ is given by the relative frequency

$$f_m = \frac{1}{\Delta x_m} \frac{n_m}{T}$$  \hspace{1cm} (A26)$$

Errors of $f_m$ can be estimated assuming a Gaussian distribution by

$$\Delta f_m = \frac{1}{\sqrt{n_m - 1}} f_m$$  \hspace{1cm} (A27)$$

Good plots of the distribution can be achieved by varying $\Delta x_m$ and by using a minimum for $n_m \geq 6$. In this way, the empirical distributions in Figures 1, 3, and 7 were created.

To estimate the quality of agreement of $f_m$ with a simulated density $F(x)$ we adopt the following qualitative measure: the average mean squared deviation also called $\chi^2$ ratio

$$\chi^2 = \frac{1}{M} \sum_{m=1}^{M} \left( \frac{F(x_m) - f_m}{\Delta f_m} \right)^2$$  \hspace{1cm} (A28)$$

A value of $\chi^2 < 4$ can be interpreted as agreement with 5% confidence level and values of $\chi^2 < 10$ still signal qualitative agreement.
Figure A1. Simulated distributions of covariances. We show the distributions of the predicted covariances versus the original data. The dotted line connects the empirical pdf and a two standard deviations range. The broken and dotted black line corresponds to the simulated predicted correlations of RMG. The results for the DCC model are represented by the broken gray line. The DECO model is represented by the gray line and the OGARCH model by the broken black line. Note that the horizontal axis shows the logarithm of the absolute values of $r_{ij}$, therefore cases where the correlation is 0 are omitted.

Appendix 5. Distributions of covariances

For a visual comparison of the distributions, we use the variable $\log(|r_{ij}|)$ for $|r_{ij}| > 0$ in order to be sensitive to large values of $r_{ij}$ and not to be overwhelmed by the many small $r_{ij}$. Three typical examples of these distributions for pairs of stocks are shown in Figure A1. The results for OG are often outside the empirical errors, whereas RMG and DECO agree with DCC and the data.