KU LEUVEN

DEPARTMENT OF ECONOMICS

DISCUSSION PAPER SERIES
DPS16.11

JUNE 2016

Sparse multivariate GARCH

Jianbin WU and Geert DHAENE

Econometrics

Faculty of Economics
and Business

SPARSE MULTIVARIATE GARCH

Jianbin Wu[†]

University of Leuven, Department of Economics, Naamsestraat 69, B-3000 Leuven, Belgium
jianbin.wu@kuleuven.be

Geert Dhaene

University of Leuven, Department of Economics, Naamsestraat 69, B-3000 Leuven, Belgium
geert.dhaene@kuleuven.be

June 15, 2016

We propose sparse versions of multivariate GARCH models that allow for volatility and correlation spillover effects across assets. The proposed models are generalizations of existing diagonal DCC and BEKK models, yet they remain estimable for high-dimensional systems of asset returns. To cope with the high dimensionality of the model parameter spaces, we employ the L_1 regularization technique to penalize the off-diagonal elements of the coefficient matrices. A simulation experiment for the sparse DCC model shows that the true underlying sparse parameter structure can be uncovered reasonably well. In an application to weekly and daily market returns for 24 countries using data from 1994 to 2014, we find that the sparse DCC model outperforms the standard DCC and the diagonal DCC models in and out of sample. Likewise, the sparse BEKK model outperforms the diagonal BEKK model.

Keywords: *multivariate GARCH, regularization, penalized estimation, volatility spillovers, correlation spillovers.*

1. INTRODUCTION

The estimation of conditional covariances between asset returns is central to many areas of empirical finance, including portfolio selection, asset pricing, and hedging. A large literature has developed exploring models of the multivariate GARCH family. Two widely used models are the BEKK model and the scalar DCC model, proposed by Engle and Kroner (1995) and Engle (2002), respectively. A shortcoming of these models is that they do not allow for volatility or correlation spillover effects across assets. In the scalar DCC model, the asset return correlations are assumed to evolve identically for all assets. This entails no restriction for bivariate systems, but when the number of assets is large this assumption is hard to defend. To address this problem, "diagonal" and "full" versions of the DCC model have been proposed; see, e.g., Engle (2002), Cappiello et al. (2006), Hafner and Franses (2009), and Billio and Caporin (2009). But these do not fully solve the problem. While the diagonal DCC model allows for idiosyncratic correlation dynamics, it still ignores correlation spillover effects. The full DCC model allows for correlation spillovers, but here the number of parameters is of order n^2, with n the number of assets considered, so estimation of the full DCC model is feasible only when n is small. Essentially the same holds for BEKK models, where the diagonal model version ignores volatility spillovers and the full model version allows them but runs into estimation problems unless n is small. In short, in multivariate GARCH modeling there is a conflict between flexibility and feasibility of estimation.

[†]Corresponding author.

In this paper we develop sparse versions of the DCC and BEKK models that seek to mitigate this conflict. The sparse models that we propose are more flexible than the diagonal models, yet more parsimonious than the full models. They are intended to capture correlation and volatility spillover effects while still being estimable when the dimension, n, is large ($n = 24$ in our application). Prior to estimation, the sparse models can be viewed as full models, with parameter restrictions imposed along the estimation in a data-compatible way. The sparse parameter structure is obtained by regularization. Specifically, we add a lasso penalty (Tibshirani 1996) to the log-likelihood function to penalize the off-diagonal elements of the coefficient matrices. This drives many of the off-diagonal elements to zero, so that a sparse structure of correlation or volatility spillover effects obtains.

Section 2 introduces the sparse DCC and BEKK models. Section 3 presents the results of a simulation experiment for the sparse DCC model. In Section 4 we estimate and evaluate sparse DCC and BEKK models for weekly and daily market returns for 24 countries using data from 1994 to 2014. We also compare the empirical performance of the sparse models with the diagonal BEKK model and with the scalar and diagonal DCC models using Diebold-Mariano tests. Section 5 concludes.

2. SPARSE MULTIVARIATE GARCH MODELS

2.1. Specification

Let r_t be the vector of returns on n assets in period t. We assume that $E_{t-1}r_t = 0$, where E_{t-1} is the conditional expectation given past information. Define the conditional and unconditional variance and correlation matrices

$$H_t = E_{t-1}(r_t r_t'), \qquad \overline{H} = E(r_t r_t'),$$
$$R_t = E_{t-1}(\varepsilon_t \varepsilon_t') = D_t^{-1} H_t D_t^{-1}, \qquad \overline{R} = E(\varepsilon_t \varepsilon_t'),$$

where ε_t is the vector of standardized returns and D_t is the diagonal matrix with the conditional standard deviations of the returns on the diagonal, i.e.,

$$\varepsilon_t = D_t^{-1} r_t, \qquad D_t = (I_n \odot H_t)^{1/2},$$

where $\odot$ is the Hadamard product. Multivariate GARCH models specify how H_t and R_t evolve over time, often through a first-order ARMA-type structure.

One challenge in multivariate GARCH modeling is to keep the model sufficiently flexible while preventing the number of parameters from growing too rapidly with n. See, for example, the discussion in Bauwens, Laurent, and Rombouts (2006). Leaving other differences aside, multivariate GARCH models typically come, in increasing order of generality, in "scalar", "diagonal", and "general" versions, with $O(1)$, $O(n)$, and $O(n^2)$ parameters, respectively. It is generally acknowledged that the richly parameterized models, with $O(n^2)$ parameters, can only be estimated sensibly when n is small enough (say, up to $n = 4$). For greater n, researchers tend to resort to scalar or diagonal model versions, with $O(1)$ or $O(n)$ free parameters. These more tightly parameterized models result from imposing prior restrictions on the coefficient matrices. Our strategy is to avoid imposing such restrictions. Starting from a rich model specification with $O(n^2)$ paramaters, we impose parameter sparsity through L_1 regularization. In this way, the sparsity structure is the result of a data-driven procedure instead of being imposed ex ante.

Consider the general DCC model with the correlation part specified as

$$R_t = (I_n \odot Q_t)^{-1/2} Q_t (I_n \odot Q_t)^{-1/2}, \tag{2.1}$$

$$Q_t = \overline{R} - A\overline{R}A' - B\overline{R}B' + A\varepsilon_{t-1}\varepsilon_{t-1}'A' + BQ_{t-1}B', \tag{2.2}$$

where A and B are coefficient matrices. This is the model of Cappiello, Engle, and Sheppard (2006) without the asymmetry term. For given $\overline{R}$ (which can be pre-estimated by correlation targeting), this model has $2n^2$ correlation parameters. The scalar version of the model is the standard DCC model of Engle (2002), with $A = aI_n$, $B = bI_n$, and scalar parameters a and b. The diagonal version restricts A and B to be diagonal matrices and has $2n$ correlation parameters (Cappiello, Engle, and Sheppard 2006; Hafner and Franses 2009). Other variants include models with regime switching correlations (Pelletier 2006) or a block structure on A and B, possibly obtained via clustering (Billio, Caporin, and Gobbo 2006; Billio and Caporin 2009; Otranto 2010). A common motivation in these papers is to specify the asset return correlation dynamics flexibly, yet tractably for estimation. Our approach is to obtain sparsely parameterized correlation dynamics via an L_1 penalized log-likelihood with penalty function

$$\text{pen}_{\lambda_A,\lambda_B}(A, B) = \lambda_A \sum_{i \neq j} |A_{ij}| + \lambda_B \sum_{i \neq j} |B_{ij}|$$

for chosen tuning parameters $\lambda_A > 0$ and $\lambda_B > 0$. Note that only the off-diagonal elements of A and B enter the penalty function. The effect of L_1 penalization is that estimates of the off-diagonal elements of A and B are being shrunk towards zero, typically resulting in many estimates being identically zero. Therefore, the estimated model lies between the diagonal and the general model versions.

The penalization approach can be applied in the same way to multivariate GARCH models that specify H_t directly. For example, the first-order BEKK model of Engle and Kroner (1995), subject to the variance targeting constraint, specifies

$$H_t = \overline{H} - A\overline{H}A' - B\overline{H}B' + Ar_{t-1}r_{t-1}'A' + BH_{t-1}B', \tag{2.3}$$

which is analogous to (2.2) and has analogous scalar and diagonal versions. Hence, penalization of A and B in the BEKK model can proceed in exactly the same way as in the DCC model.

The distinction between different types of tuning parameters (here, λ_A and λ_B) in the penalty function allows additional modeling and penalization flexibility. For example, setting $0 < \lambda_A < \infty$ and $\lambda_B = \infty$ imposes diagonality on B and sparsity on the off-diagonal elements of A. Furthermore, the model can easily be extended to incorporate slowly changing unconditional variances or correlations (Engle and Rangel 2008; Hafner and Linton 2010; Bauwens, Hafner, and Pierret 2013) or additional effects such as asymmetries (Cappiello, Engle, and Sheppard 2006). Additional effects typically entail additional parameter matrices, which may be penalized as above to the desired degree. Note, furthermore, that penalization of the diagonal model version also fits into our framework, by writing $A = aI_n + \text{diag}(\alpha)$ and $B = bI_n + \text{diag}(\beta)$, where a and b are scalars and α and β are vectors, and using $\text{pen}_{\lambda_\alpha,\lambda_\beta}(\alpha, \beta) = \lambda_\alpha \sum_i |\alpha_i| + \lambda_\beta \sum_i |\beta_i|$ as penalty function.

2.2. Estimation

To estimate the sparse DCC model we use the two-step procedure of Engle (2002), augmented with penalization in the second step. The volatility part of the DCC model consists of n univariate GARCH(1,1) models, one for each asset, with parameters denoted as θ. The correlation part consists of (2.1)–(2.2), with parameters $\phi = (A, B)$. The penalized Gaussian quasi log-likelihood, for given $\overline{R}$ and tuning parameters $\lambda = (\lambda_A, \lambda_B)$, is

$$\begin{aligned} l_{\text{pen}}(\theta, \phi) &= -\frac{1}{2} \sum_t (n \log(2\pi) + \log |H_t| + r_t' H_t^{-1} r_t) - \text{pen}_\lambda(\phi) \\ &= l_{\text{v}}(\theta) + l_{\text{c}}(\theta, \phi) - \text{pen}_\lambda(\phi), \end{aligned}$$

where l_{v} and l_{c} correspond to the volatility and correlation parts,

$$\begin{aligned} l_{\text{v}}(\theta) &= -\frac{1}{2} \sum_t (n \log(2\pi) + 2 \log |D_t| + r_t' D_t^{-2} r_t), \\ l_{\text{c}}(\theta, \phi) &= -\frac{1}{2} \sum_t (\log |R_t| + \varepsilon_t' R_t^{-1} \varepsilon_t - \varepsilon_t' \varepsilon_t), \end{aligned}$$

as in Engle (2002). In step one, $l_{\text{v}}(\theta)$ is maximized by fitting a GARCH(1,1) model for each asset separately. This gives $\widehat{\theta}$, $\widehat{D}_t$, $\widehat{\varepsilon}_t = \widehat{D}_t^{-1} r_t$, and $l_{\text{c}}(\widehat{\theta}, \phi)$, with $T^{-1} \sum_t \widehat{\varepsilon}_t \widehat{\varepsilon}_t'$ as the correlation-targeting estimate of $\overline{R}$. The second step is to solve

$$\max_\phi \left\{ l_{\text{c}}(\widehat{\theta}, \phi) - \text{pen}_\lambda(\phi) \right\},$$

for which we use the block-coordinate update method. Dividing $\phi = (\phi_{\text{pen}}, \phi_{\text{unp}})$ into a block of penalized parameters ϕ_{pen} and a block of unpenalized parameters ϕ_{unp}, we update one block at the time, cycling over the two blocks until convergence. We update ϕ_{unp} with the Newton-Raphson method and ϕ_{pen} with the coordinate ascent optimization algorithm (given that $\text{pen}_\lambda(\phi)$ is not differentiable in ϕ_{pen} at the origin). This algorithm updates one parameter at the time, with all others held fixed, cycling over the penalized parameters until convergence. The parameter update of $\phi_j \in \phi_{\text{pen}}$ is as follows: if $|\nabla_{\phi_j} l_{\text{c}}(\widehat{\theta}, \phi)|_{\phi_j = 0}$ is less than the tuning parameter, ϕ_j is set to zero; else ϕ_j is set to $\arg\max_{\phi_j} l_{\text{c}}(\widehat{\theta}, \phi)$.

The sparse BEKK model with volatility specification (2.3) and parameters $\phi = (A, B)$ can be estimated along the same lines in one step by maximizing $l_{\text{pen}}(\phi) = l_{\text{v}}(\phi) - \text{pen}_\lambda(\phi)$, where $l_{\text{v}}(\phi) = -\frac{1}{2} \sum_t (n \log(2\pi) + \log |H_t| + r_t' H_t^{-1} r_t)$ and with $T^{-1} \sum_t r_t r_t'$ as the variance-targeting estimate of $\overline{H}$.

At each iteration along the optimization, we impose positive definiteness on Q_t or H_t (in the DCC or BEKK model, respectively) for all t in the estimation sample. This guarantees positive definiteness at the converged estimates in the estimation sample, but not necessarily outside the estimation sample, although we did not encounter this problem. Should it occur, one may impose a positive lower bound on the eigenvalues of Q_t or H_t. Without further restriction, the sparse GARCH models do not guarantee positive definiteness.

Multivariate GARCH models with high-dimensional parameters are numerically challenging to estimate. The penalization step is numerically slow, adding to the challenge. Furhtermore, the degree of regularization is controlled by the tuning parameters λ_A and λ_B, which have to be chosen. At the present stage, we set $\lambda_B = \infty$ and impose the further restriction $B = bI_n$, where b is a scalar. Hafner and Franses (2009) noted that, in the diagonal DCC model, the parameters associated with the autoregressive part Q_{t-1} are less varying than those associated with the innovations $\varepsilon_t\varepsilon_t'$. So, broadly speaking, B may be more tightly parameterized than A. With $B = bI_n$, we have $\phi = (A, b)$ and ϕ_{pen} consists of the off-diagonal elements of A only. We choose λ_A by cross-validation, using approximately the first 90% of the data as training sample and the remaining 10% as validation sample, involving the following steps:

(i) based on the training sample, we estimate ϕ_{unp} with ϕ_{pen} set to zero;

(ii) at these values of ϕ_{unp} and ϕ_{pen}, we compute the log-likelihood gradient vector for ϕ_{pen}, that is, $G_{\phi_{\text{pen}}} = \nabla_{\phi_{\text{pen}}} l_{\text{c}}(\widehat{\theta}, \phi)$ (in the DCC model) or $G_{\phi_{\text{pen}}} = \nabla_{\phi_{\text{pen}}} l_{\text{v}}(\phi)$ (in the BEKK model);

(iii) we compute the 68–96th percentiles, in steps of 4%, of the elements of $|G_{\phi_{\text{pen}}}|$;

(iv) for λ_A equal to each of these percentiles, we compute the penalized estimate of ϕ based on the training sample and evaluate the unpenalized log-likelihood on the validation sample at this value of ϕ;

(v) we choose the value of λ_A that maximizes this log-likelihood.

3. SIMULATIONS

This section reports on simulations for the sparse DCC model. The simulation setup broadly mimics the dimension and properties of the daily market index return data of 24 developed countries for 1994–2014 that we use in the empirical application discussed in the next section. Our aim here is to explore how well the estimator can detect the sparse parameter structure in a large-dimensional, highly parameterized DCC model. As in Hafner and Franses (2009), we focus on the model's correlation part only, ignoring the volatility part. So we set $D_t = I_n$ and only carried out step 2 of the estimation. We generated data $r_t = \varepsilon_t$ for $t = 1, \ldots, 5000$ (and a burn-in sample of 1000 periods) according to

$$\varepsilon_t \sim N(0, R_t), \qquad R_t = (I_n \odot Q_t)^{-1/2} Q_t (I_n \odot Q_t)^{-1/2},$$
$$Q_t = \overline{R} - A\overline{R}A' - b^2\overline{R} + A\varepsilon_{t-1}\varepsilon_{t-1}'A' + b^2 Q_{t-1},$$

with $n = 24$, $b^2 = 0.995$, $\overline{R}$ equal to the empirical daily return correlation matrix, and A chosen as follows. We drew the diagonal elements of A from the uniform distribution $U[.8c, 1.2c]$ with mean $c = .07$, set 20 randomly chosen off-diagonal elements of A equal to the values in the set $\pm c \cdot \{.01, .02, .1, .15, .2\}$ (each value being repeated twice), and set the other 532 off-diagonal elements of A equal to zero. We generated 20 simulated data sets in this way.

For each simulated data set, we estimated $\phi = (A, b)$ as outlined above, with the correlation matrix of the simulated data as an estimate of $\overline{R}$. To reduce the computation time, we fixed the tuning parameter λ_A at the 88th percentile of $|G_{\phi_{\text{pen}}}|$ (computed from the full simulated data set) instead of determining λ_A by cross-validation. Table 1 is a contingency table of the true and estimated off-diagonal elements of A, averaged across the simulations. As the table shows, the underlying sparsity structure is uncovered reasonably well, with two thirds of the "large" nonzero parameter values (those in $\pm c \cdot \{.1, .15, .2\}$) being detected and 95% of the zeros being estimated at zero. As expected, "small" nonzero parameter values (those in $\pm c \cdot \{.01, .02\}$) are much harder to detect: only 8% are estimated to be nonzero.

Table 1. Estimated versus true sparsity

	estimated zero	estimated nonzero	total
true zero	503.85	28.15	532
true "small"	7.35	0.65	8
true "large"	4.00	8.00	12
total	515.20	36.80	552

"Small" refers to values in $\pm c \cdot \{.01, .02\}$. "Large" refers to values in $\pm c \cdot \{.1, .15, .2\}$.

Table 2 gives details for each simulation separately. The first three columns pertain to the off-diagonal elements of A, giving the number of true zeros estimated at zero, and the number of "small" and "large" values, respectively, estimated to be nonzero. The last four columns report the true and estimated values of the average of the diagonal elements of A and b^2. These estimates are very close to the true values, although $\widehat{b}^2$ tends to slightly underestimate b^2.

Table 2. Estimation results for each simulation

zero	"small"	"large"	$\sum_i A_{ii}/n$	$\sum_i \widehat{A}_{ii}/n$	b^2	$\widehat{b}^2$
503	1	9	0.0697	0.0688	0.995	0.9945
498	1	7	0.0689	0.0666	0.995	0.9946
506	2	10	0.0714	0.0709	0.995	0.9948
507	0	5	0.0700	0.0718	0.995	0.9943
500	0	7	0.0719	0.0726	0.995	0.9943
508	0	9	0.0687	0.0679	0.995	0.9946
502	0	9	0.0660	0.0650	0.995	0.9946
502	2	9	0.0666	0.0641	0.995	0.9946
511	1	9	0.0683	0.0683	0.995	0.9946
508	1	7	0.0670	0.0674	0.995	0.9947
505	1	9	0.0688	0.0697	0.995	0.9944
502	1	7	0.0710	0.0691	0.995	0.9947
505	0	10	0.0687	0.0690	0.995	0.9946
517	0	9	0.0690	0.0695	0.995	0.9944
503	0	7	0.0685	0.0690	0.995	0.9944
484	0	5	0.0666	0.0648	0.995	0.9946
513	0	9	0.0697	0.0703	0.995	0.9943
504	1	8	0.0668	0.0665	0.995	0.9948
502	0	8	0.0674	0.0664	0.995	0.9946
497	2	7	0.0705	0.0688	0.995	0.9949

The first three columns refer to the off-diagonal elements of A (532 zeros, 8 "small" values, and 12 "large" values) and report the numbers of zeros estimated as zero, "small" values estimated as nonzero, and "large" values estimated as nonzero, respectively.

4. APPLICATION TO MARKET RETURNS FOR 24 COUNTRIES, 1994-2014

We estimate and compare scalar, diagonal, and sparse multivariate GARCH models for weekly and daily market returns of 24 countries with developed stock markets over the period March 1, 1994, to July 7, 2014. The countries are Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Hong Kong, Ireland, Israel, Italy, Japan, the Netherlands, New Zealand, Norway, Portugal, Singapore, Spain, Sweden, Switzerland, the United Kingdom, and the United States. We use the MSCI market index data, expressed in U.S. dollars, as obtained from Datastream with code names 'MS****$(MSPI)', e.g., 'MSAUST$(MSPI)' for Australia. The returns are computed in logarithmic form and demeaned prior to the analysis.

4.1. Results for weekly returns

We divide the data into 3 parts, with the first 80% as the in-sample training data, the next 10% as the in-sample validation data, and the remaining 10% as the out-of-sample testing data. The procedure outlined above selects the tuning parameter λ_A as the 68th percentile of $|G_{\phi_{\text{pen}}}|$ for the sparse DCC model and as the 76th percentile for the sparse BEKK model.

Using the in-sample training and validation data, we estimate the sparse DCC and BEKK models (with the corresponding λ_A obtained), the scalar and diagonal DCC models, and the diagonal BEKK model. Table 3 reports a summary of the parameter estimates and the in-sample and out-of-sample average log-likelihood per observation for each model. The sparse DCC model has the greatest in-sample and out-of-sample average log-likelihood values and the diagonal BEKK model has the least.

Further, we compare each pair of GARCH models with the Diebold and Mariano (1995) test, with minus the out-of-sample log-likelihood as the loss function. Table 4 reports the t statistics of the Diebold-Mariano test. The sparse DCC model is significantly better than the scalar DCC model and the sparse BEKK model is significantly better than the diagonal BEKK model.

We also examine the relative performance of the GARCH models with the asset-allocation methodology proposed by Engle and Colacito (2006). Consider an asset allocation problem for n assets with return vector r_t whose conditional variance matrix is H_t. The variance minimization problem is

$$\min_{w_t} w_t' H_t w_t \qquad \text{subject to } w_t' 1_n = 1,$$

where 1_n is an $n \times 1$ vector of ones. The solution is

$$w_t = \frac{H_t^{-1} 1_n}{1_n' H_t^{-1} 1_n}$$

and the minimum-variance portfolio has return $w_t' r_t$. With the out-of-sample squared return $(w_t' r_t)^2$ as the loss function, we compare each pair of GARCH models using the Diebold-Mariano test. In addition to the portfolios constructed from the GARCH models, we also consider the equally-weighted portfolio, with weights $w_t = n^{-1} 1_n$, and the constantly-weighted portfolio with weights $w_t = (1_n' \overline{H}^{-1} 1_n)^{-1} \overline{H}^{-1} 1_n$ based on the unconditional variance of r_t. Table 4 shows that in terms of asset allocation the diagonal DCC model performs best, followed by the sparse DCC model. The sparse BEKK outperforms the diagonal BEKK model, but is dominated by the DCC models. With a few exceptions, however, the differences between the GARCH models are not statistically significant. The equally-weighted portfolio performs worst: is dominated by the GARCH-based portfolios and the dominations are statistically significant.

Table 3. Weekly returns: model estimates and log-likelihood values

model	log-likelihood in sample	log-likelihood out-of-sample	number of parameters	a or $\sum_i A_{ii}/n$	b or $\sum_i B_{ii}/n$
sparse DCC	−49.7362	−43.8634	104 + 72	0.0946	0.9867
diagonal DCC	−49.9267	−43.8900	25 + 72	0.0986	0.9864
scalar DCC	−50.0578	−43.9560	2 + 72	0.0974	0.9896
sparse BEKK	−50.3628	−44.7004	118	0.1351	0.9752
diagonal BEKK	−50.5471	−44.8087	48	0.1429	0.9721

DCC specifications: $Q_t = \overline{R} - A\overline{R}A' - b^2\overline{R} + A\varepsilon_{t-1}\varepsilon_{t-1}'A' + b^2Q_{t-1}$ with $A = aI_n$ in the scalar DCC, A diagonal in the diagonal DCC, and A unrestricted in the sparse DCC. BEKK specifications: $H_t = \overline{H} - A\overline{H}A' - B\overline{H}B' + Ar_{t-1}r_{t-1}'A' + BH_{t-1}B'$ with A and B diagonal in the diagonal BEKK, and A unrestricted and B diagonal in the sparse BEKK. For the DCC models the number of parameters is split between those in the correlation part and the $3 \times 24 = 72$ parameters in the volatility part.

Table 4. Weekly returns: Diebold-Mariano tests based on the out-of-sample log-likelihood value

	sparse DCC	diag. DCC	scalar DCC	sparse BEKK	diag. BEKK
sparse DCC	–	−1.0110	−2.4585	−3.6973	−4.1673
diagonal DCC	1.0110	–	−2.0843	−3.4874	−3.9360
scalar DCC	2.4585	2.0843	–	−3.1148	−3.5421
sparse BEKK	3.6973	3.4874	3.1148	–	−4.5769
diagonal BEKK	4.1673	3.9360	3.5421	4.5769	–
average loss	43.8634	43.8900	43.9560	44.7004	44.8087

Entries: t statistics of the Diebold-Mariano test of the null that the corresponding row and column models have equal expected loss, with minus the out-of-sample log-likelihood as the loss function. A positive t statistic indicates that the column model is better than the row model in that it has the least loss of the two models.

Table 5. Weekly returns: Diebold-Mariano tests based on out-of-sample asset allocation

	sparse DCC	diagonal DCC	scalar DCC	sparse BEKK	diagonal BEKK	constant weight	equal weight
sparse DCC	–	0.8869	−0.8994	−0.2511	−0.3004	−0.8123	−3.3329
diag. DCC	−0.8869	–	−2.5607	−0.4149	−0.4651	−0.9792	−3.4462
scalar DCC	0.8994	2.5607	–	−0.0743	−0.1349	−0.6853	−3.3061
sparse BEKK	0.2511	0.4149	0.0743	–	−0.3983	−1.8908	−3.3305
diag. BEKK	0.3004	0.4651	0.1349	0.3983	–	−1.9458	−3.3412
const. weight	0.8123	0.9792	0.6853	1.8908	1.9458	–	−3.1579
equal weight	3.3329	3.4462	3.3061	3.3305	3.3412	3.1579	–
average loss	1.6088	1.5826	1.6428	1.6562	1.6679	1.7932	3.2492

Entries: t statistics of the Diebold-Mariano test of the null that the corresponding row and column models have equal expected loss, with the out-of-sample squared portfolio return (expressed in %) as the loss function. A positive t statistic indicates that the column model is better than the row model in that it has the least loss of the two models.

4.2. Results for daily returns

For the daily returns we use the same 80-10-10% division of the data into in-sample training, in-sample validation, and out-of-sample testing data. The tuning parameter, λ_A, is selected as the 68th percentile of $|G_{\phi_{\text{pen}}}|$ for the sparse DCC and sparse BEKK models. Table 6 gives a summary of the estimated parameters and the average log-likelihood values. Again, the sparse DCC model has the greatest average log-likelihood, both in and out of sample, and the diagonal BEKK model has the least.

Tables 7 and 8 report Diebold-Mariano model comparison tests based on out-of-sample log-likelihood values and asset allocation, parallelling the earlier Tables 4 and 5. When the loss function is minus the log-likelihood (Table 7), again the sparse DCC model performs best and the sparse BEKK model outperforms the diagonal BEKK model. Surprisingly, though, the statistical significance of the t statistics has gone down. With the squared portfolio return as loss function (Table 8), we now see that the sparse BEKK model almost significantly outperforms all other models, followed by the diagonal BEKK model. The equally-weighted and constantly-weighted portfolios are significantly dominated by GARCH-based portfolios.

Table 6. Daily returns: model estimates and log-likelihood values

model	log-likelihood in sample	log-likelihood out-of-sample	number of parameters	a or $\sum_i A_{ii}/n$	b or $\sum_i B_{ii}/n$
sparse DCC	−31.1275	−25.2177	97 + 72	0.0700	0.9960
diagonal DCC	−31.1797	−25.2299	25 + 72	0.0698	0.9963
scalar DCC	−31.2522	−25.3101	2 + 72	0.0665	0.9974
sparse BEKK	−31.8493	−25.4974	99	0.1088	0.9921
diagonal BEKK	−31.8993	−25.5155	48	0.1090	0.9922

DCC specifications: $Q_t = \overline{R} - A\overline{R}A' - b^2\overline{R} + A\varepsilon_{t-1}\varepsilon'_{t-1}A' + b^2Q_{t-1}$ with $A = aI_n$ in the scalar DCC, A diagonal in the diagonal DCC, and A unrestricted in the sparse DCC. BEKK specifications: $H_t = \overline{H} - A\overline{H}A' - B\overline{H}B' + Ar_{t-1}r'_{t-1}A' + BH_{t-1}B'$ with A and B diagonal in the diagonal BEKK, and A unrestricted and B diagonal in the sparse BEKK. For the DCC models the number of parameters is split between those in the correlation part and the $3 \times 24 = 72$ parameters in the volatility part.

Table 7. Daily returns: Diebold-Mariano tests based on the out-of-sample log-likelihood value

	sparse DCC	diag. DCC	scalar DCC	sparse BEKK	diag. BEKK
sparse DCC	–	−0.6149	−2.4440	−1.8176	−1.8890
diagonal DCC	0.6149	–	−2.4001	−1.6575	−1.7340
scalar DCC	2.4440	2.4001	–	−1.0458	−1.1277
sparse BEKK	1.8176	1.6575	1.0458	–	−1.4975
diagonal BEKK	1.8890	1.7340	1.1277	1.4975	–
average loss	25.2177	25.2299	25.3101	25.4974	25.5155

Entries: t statistics of the Diebold-Mariano test of the null that the corresponding row and column models have equal expected loss, with minus the out-of-sample log-likelihood as the loss function. A positive t statistic indicates that the column model is better than the row model in that it has the least loss of the two models.

Table 8. Daily returns: Diebold-Mariano tests based on out-of-sample asset allocation

	sparse DCC	diagonal DCC	scalar DCC	sparse BEKK	diagonal BEKK	constant weight	equal weight
sparse DCC	–	−0.3639	−1.1187	1.8451	1.5121	−2.1703	−6.5138
diag. DCC	0.3639	–	−0.8987	1.8507	1.5290	−2.1173	−6.5660
scalar DCC	1.1187	0.8987	–	2.1551	1.8319	−1.8202	−6.3710
sparse BEKK	−1.8451	−1.8507	−2.1551	–	−1.5919	−4.9064	−6.6528
diag. BEKK	−1.5121	−1.5290	−1.8319	1.5919	–	−4.7745	−6.5977
const. weight	2.1703	2.1173	1.8202	4.9064	4.7745	–	−6.1165
equal weight	6.5138	6.5660	6.3710	6.6528	6.5977	6.1165	–
average loss	0.2367	0.2371	0.2401	0.2219	0.2239	0.2646	0.6335

Entries: t statistics of the Diebold-Mariano test of the null that the corresponding row and column models have equal expected loss, with the out-of-sample squared portfolio return (expressed in %) as the loss function. A positive t statistic indicates that the column model is better than the row model in that it has the least loss of the two models.

4.3. Daily volatility and correlation spillover effects

The main advantage of the sparse BEKK and DCC models, relative to their scalar and diagonal versions, is that they allow volatility and correlation spillovers through the off-diagonal elements of A. Consider the sparse BEKK model. If A_{ij} is non-zero, then a shock to market j's return at time $t-1$ will affect market i's volatility at time t. Figure 1 depicts the estimated volatility spillover effects graphically. Each directed arrow corresponds to a non-zero estimated off-diagonal element of A in the sparse BEKK model for daily returns, with thicker lines representing stronger effects. The estimates suggest that there are no or few volatility spillover effects from and to the stock markets in Australia, Israel, Japan, and Singapore. In contrast, most European stock markets and the German market in particular appear to have strong spillover effects with each other. The volatility of the German stock market seems to be affected by spillover effects from the U.S., France, Ireland, the Netherlands, and Spain, and to exhibit spillover effects to Japan and Switzerland.

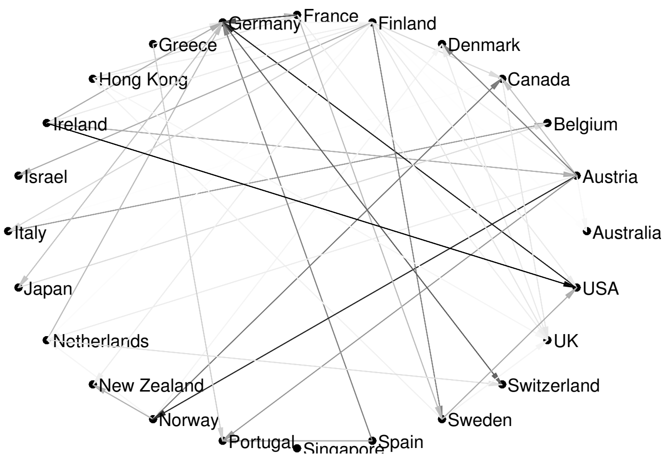

Figure 1. Volatility spillover estimates in the sparse BEKK model (daily returns)

In similar fashion, the sparse DCC model is able to capture correlation spillovers through the off-diagonal elements of A. Figure 2 shows the correlation spillover effects, analogous to Figure 1, based on the estimated sparse DCC model for daily returns. Again, many intra-European spillover effects are found. In particular, there seem to be strong correlation spillovers from Sweden to Finland, from Austria to Denmark and Ireland, and from Greece to Portugal. The German stock market also appears to have many correlation spillover effects with other countries.

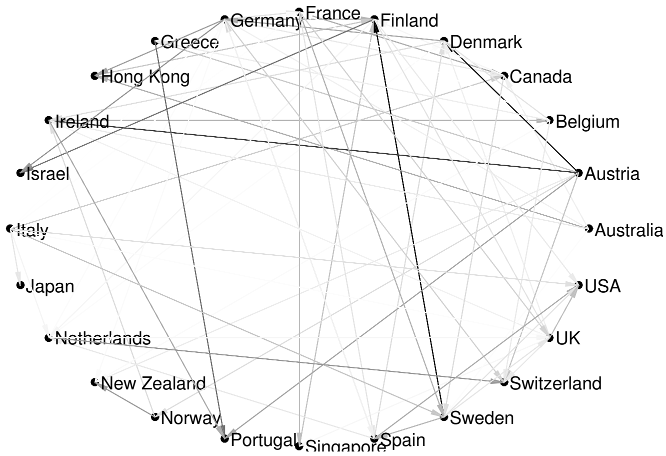

Figure 2. Correlation spillover estimates in the sparse DCC model (daily returns)

4.4. Daily conditional volatilities, co-volatilities, and correlations

Here we present and briefly discuss a selection of the time series of daily conditional volatilities, co-volatilities, and correlations as implied by the sparse BEKK and DCC models.

Figure 3 displays the conditional co-volatilities $(H_t)_{ij}$, based on the sparse BEKK model, for a selection of country pairs (i, j). There were extremely large covariances between the markets around November 2008. Only the Japanese stock market had less dramatic covariances with the other markets during this period of crisis. The covariances between European markets were still high in the period 2009–2013, while those between Japan, Hong Kong, Singapore, and New Zealand quickly came back to near pre-crisis levels.

Figure 4 presents a selection of conditional correlations $(R_t)_{ij}$ based on the sparse DCC model. Many correlations reached their peak in November 2008. This suggests that the high conditional covariances in 2008 were not only driven by the high volatilities per se, but also by increased correlations. In addition, many correlations show an upward trend over the twenty-year period considered here, although the phenomenon is not universal across countries. A particularly striking example is the German and French stock markets, where the sparse DCC model indicates a rapidly increasing correlation between 1997 and 1999, followed by a further gradual increase towards almost 1. The pattern of increased correlations over time for some countries and not for others is in line with the findings of Bekaert et al. (2009). They showed that there is a statistically significant upward trend in the correlations among European countries and a nearly significant upward trend between Europe and the U.S., while for the other correlations the upward trend is weaker and statistically not significant.

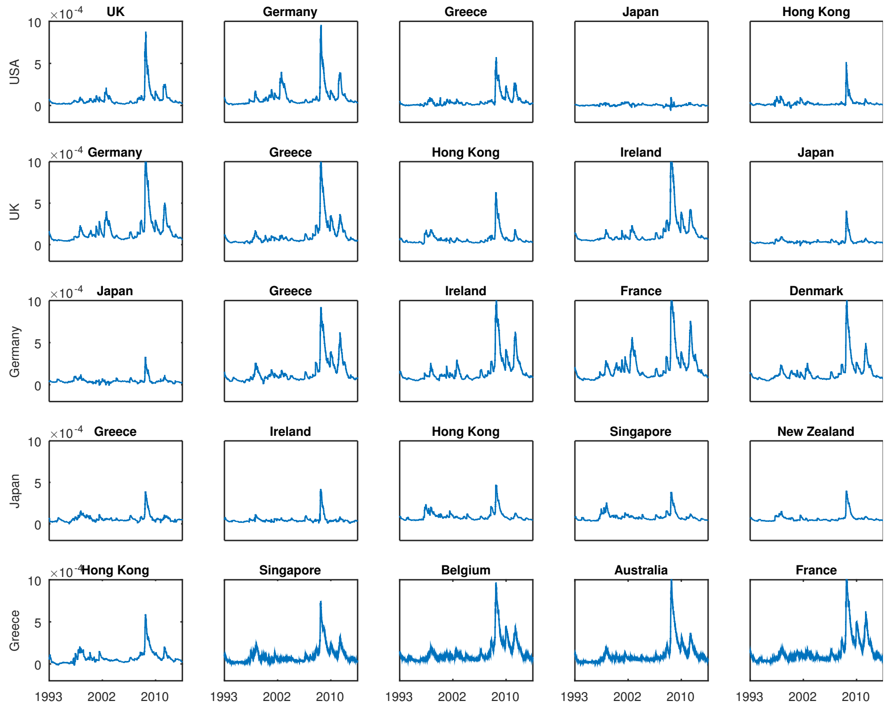

Figure 3. Daily conditional co-volatilities in the sparse BEKK model

Figure 4. Daily conditional correlations in the sparse DCC model

Figure 5 shows the conditional volatilities $(H_t)_{ii}$ of the stock markets in the U.S., UK, Germany, Japan, and Greece, based on the sparse DCC model. The volatilities were indeed dramatically high during the 2008 credit crunch, which is probably the main driving factor of the high conditional covariances in 2008. The volatility of the Greek stock market was higher than that of the other markets and remained high in 2009–2013.

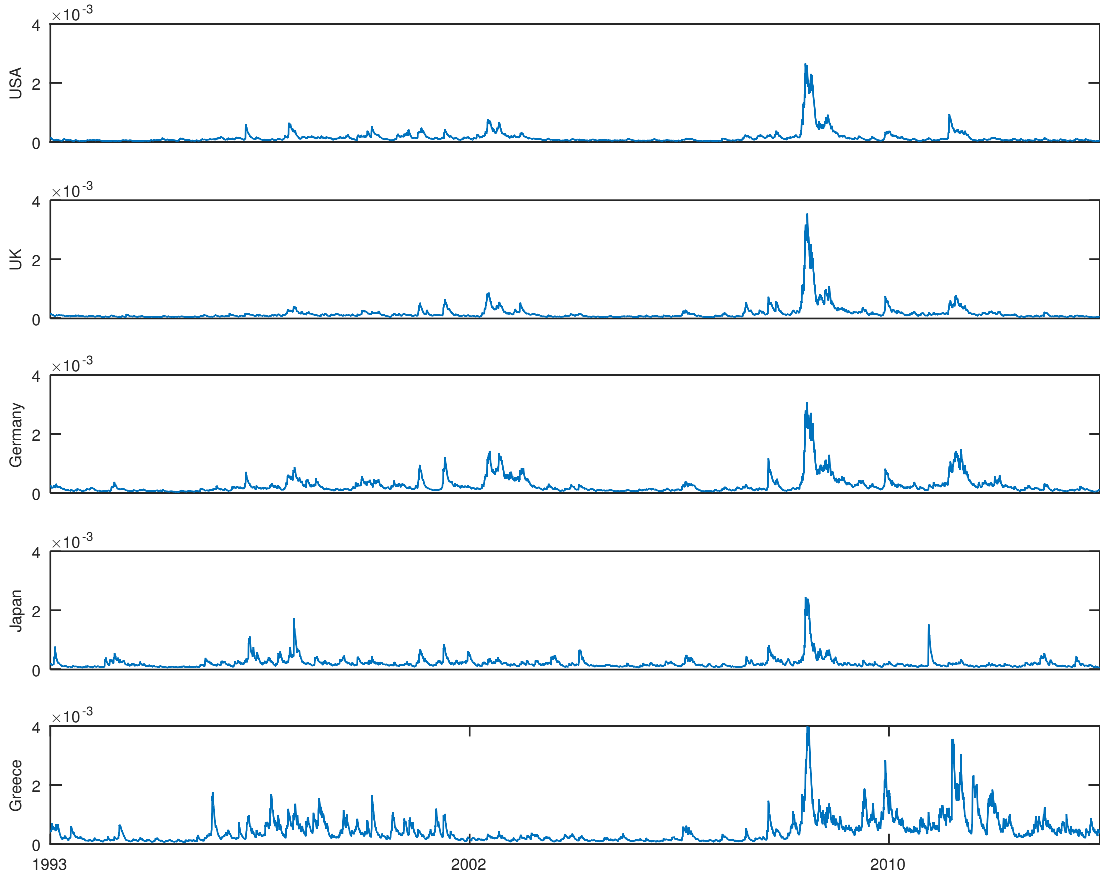

Figure 5. Conditional daily volatilities in the sparse DCC model

5. CONCLUDING REMARKS

In this paper, we propose a sparse modeling approach towards multivariate GARCH. The focus is on GARCH$(1,1)$ structures, the generalization to higher orders being obvious. Our approach allows to explore the dynamics of large-dimensional financial time series, with particular attention to uncovering volatility or correlation spillover effects. As the number of potential spillover effects increases quadratically with the dimension of the system, some form of regularization is needed, resulting in a sparse structure of identified spillover effects.

In our application to weekly and daily market returns for 24 countries over the last two decades, we find that the sparse DCC model systematically outperforms the DCC models that exclude correlation spillover effects. The sparse BEKK model, likewise, performs better than the diagonal BEKK model, which excludes volatility spillovers.

Our empirical study further indicates that European stock markets have pronounced volatility and correlation spillovers to each other. The model estimates suggest, in particular, strong volatility and correlation spillovers from and to the German stock market. On the other hand, the stock markets in Singapore, Australia, Japan, and Israel appear relatively more isolated, with few spillover effects. The sparse model estimates also indicate that, at the high of the 2008 credit crunch, the conditional covariances of the stock markets were dramatically high, partly caused by the conditional correlations being at their peak over the last twenty years.

One of the present limitations of our approach is that due to the lasso regularization technique, it is difficult to construct parameter confidence sets and to carry out statistical tests on the parameters.

Acknowledgment. Jianbin Wu gratefully acknowledges financial support from the China Scholarship Council.

REFERENCES

Bauwens, L., C. M. Hafner, and D. Pierret (2013). Multivariate volatility modeling of electricity futures. *Journal of Applied Econometrics 28*, 743–761.

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

Bekaert, G., R. J. Hodrick, and X. Zhang (2009). International stock return comovements. *The Journal of Finance 64*, 2591–2626.

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.

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.

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

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

Engle, R. and R. Colacito (2006). Testing and valuing dynamic correlations for asset allocation. *Journal of Business & Economic Statistics 24*, 238–253.

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

Engle, R. F. and J. G. Rangel (2008). The spline-GARCH model for low-frequency volatility and its global macroeconomic causes. *Review of Financial Studies 21*, 1187–1222.

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.

Hafner, C. M. and O. Linton (2010). Efficient estimation of a multivariate multiplicative volatility model. *Journal of Econometrics 159*, 55–73.

Otranto, E. (2010). Identifying financial time series with similar dynamic conditional correlation. *Computational Statistics & Data analysis 54*, 1–15.

Pelletier, D. (2006). Regime switching for dynamic correlations. *Journal of Econometrics 131*, 445–473.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B 58*, 267–288.