UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Scienze Economiche "Marco Fanno"

RANKING MULTIVARIATE GARCH MODELS
BY PROBLEM DIMENSION

MASSIMILIANO CAPORIN
University of Padova

MICHAEL MCALEER
Erasmus University Rotterdam

December 2010

Ranking Multivariate GARCH Models
by Problem Dimension*

Massimiliano Caporin

Department of Economics and Management "Marco Fanno"
University of Padova
Italy

Michael McAleer

Econometric Institute
Erasmus School of Economics
Erasmus University Rotterdam
The Netherlands
and
Tinbergen Institute
The Netherlands

First draft: January 2009
This version: July 2010

Abstract

In the last 15 years, several Multivariate GARCH (MGARCH) models have appeared in the literature. Some recent research has begun to examine MGARCH specifications in terms of their out-of-sample forecasting performance. In this paper, we provide an empirical comparison of a set of models, namely BEKK, DCC, Corrected DCC (cDCC) of Aeilli (2008), CCC, Exponentially Weighted Moving Average, and covariance shrinking, using the historical data of 89 US equities. Our methods follow some of the approach described in Patton and Sheppard (2009), and contribute to the literature in several directions. First, we consider a wide range of models, including the recent cDCC model and covariance shrinking. Second, we use a range of tests and approaches for direct and indirect model comparison, including the Weighted Likelihood Ratio test of Amisano and Giacomini (2007). Third, we examine how the model rankings are influenced by the cross-sectional dimension of the problem.

Keywords: Covariance forecasting, model confidence set, model ranking, MGARCH, model comparison.

JEL codes: C32, C53, C52.

1. Introduction

Multivariate Volatility Models (MVM) have attracted a considerable interest over the last decade. This may be associated with the increased availability of financial data, the increased computational powers of computers, and the fact that the financial industry has begun to realize the possible advantages of these models.

The recent literature on the topic has moved from the introduction of new models to the efficient estimation of existing models. Among the most highly cited topics are the "curse of dimensionality" and "feasible model estimation". In fact, the feasibility of model estimation is now of central interest, with many studies proposing appropriate parameterizations of known models (Billio et al, 2006, Billio and Caporin, 2009, Franses and Hafner, 2009, Caporin and Paruolo, 2009, Bonato et al., 2009, Asai et al., 2009), or focusing on special estimation methods (Engle and Kelly, 2008, Engle et al., 2008, Fan et al., 2007).

A second strand of the literature has focused on the statistical or asymptotic properties of the models and of the proposed estimators (Comte and Liebermann, 2003, Ling and McAleer, 2003, McAleer et al. 2008, Engle et al. 2008, Aielli, 2008, Caporin and McAleer, 2009, Hafner and Preminger, 2009). These studies point out that only in special cases the asymptotic properties are known under untestable moment restriction, or under claimed regularity conditions (see Caporin and McAleer (2009) for a detailed discussion).

Despite the theoretical properties typically being assumed under unstated and untestable regularity conditions, many proposed models have been used widely in empirical financial studies. Within this framework, a different problem arises: How can we compare and rank models characterized by a different structure? Some research has recently appeared in the literature to tackle the problem of evaluating alternative covariance models (see Engle and Colacito (2006), Engle and Sheppard (2008), Clements et al. (2009), and Patton and Sheppard (2009)). These papers present limited comparisons across a small range of models. Engle and Colacito (2006) compare only

the DCC model of Engle (2002) against a constant correlation model, and in a datasets with a cross-sectional dimension equal to 2 (two stock market or bond indices) or 34 (the same series used in Cappiello et al., 2006). Engle and Sheppard (2008) is a quite extensive study for the model considered, but uses a single cross-sectional dimension (50 sector indices defined within the perimeter of the S&P 500 index). Patton and Sheppard (2009) is a theoretical contribution on the approaches to be used for the evaluation of covariance forecasts, and does not include an empirical application (even with low cross-sectional dimensions) showing the arguments for and against the various methods. Clements et al. (2009) focus on dynamic correlation models, and present results for a cross-sectional dimension equal to 5 (five US based future contracts). Laurent et al. (2010) consider a moderately large set of models, but focus on a 10 assets example, and place emphasis on the model accuracy against a DCC benchmark. Furthermore, all of the previous papers include the DCC model of Engle (2002), and are thereby exposed to the possible estimation (in)consistency problems (see, for example, Aielli (2008)).

The methods of comparison used in the previous contributions could be viewed as two large classes (see Patton and Sheppard (2009)), namely the direct and indirect evaluation of volatility forecasts. The first group includes the Mincer-Zarnowitz regression (Mincer and Zarnowitz, 1969), Diebold-Mariano test (Diebold and Mariano, 1996, and West, 1996, 2006), Reality Check of White (2000), Superior Predictive Ability (SPA) test of Hansen (2005), and the Model Confidence Set (MCS) approach of Hansen et al. (2005). The second group includes approaches based on the comparison of loss functions adapted to the needs of covariance forecasts. This is the case, for instance, of asset allocation and risk management, where loss functions could be defined using global minimum variance portfolios returns, such as in Engle and Colacito (2006) and Patton and Sheppard (2009), or within a Value-at-Risk framework, as in Ferreira and Lopez (2005).

The tests directly comparing the covariance forecasts fit the general framework of loss-function comparison, as discussed in Clements et al. (2009) and Patton and Sheppard (2009). The Diebold-Mariano and West approaches are valid for pairwise comparisons

of the models, while Reality check and SPA require the identification of a benchmark model, whereas MCS does not require a benchmark specification. Overall, the MCS approach seems to be the preferred one and the most appropriate as it provides a statistical test and a method for determining which models are statistically equivalent with respect to a given loss function. Despite the use of a bootstrap method for the evaluation of test statistic, MCS is computationally feasible, efficient and statistically robust.

With respect to the indirect comparison of volatility forecasts, an interesting result has been shown in Clements et al. (2009), that illustrates how utility-based loss functions (in particular, quadratic utilities) make the impact of the covariance model very modest. The approach of Engle and Colacito (2006) should provide interesting results, even for large cross-sectional dimensions.

In this paper we contribute to the literature on covariance forecast evaluation in several ways. First, our selection of models to be compared differs from those of previous studies. Similarly to the literature we consider the CCC model of Bollerslev (1990), DCC model of Engle (2002), Scalar BEKK model with targeting of Ding and Engle (2002), and the naïve Exponentially Weighted Moving Average approach. We complement this set by including the cDCC model of Aielli (2008), and the covariance shrinking approach of Ledoit and Wolf (2003, 2004). The introduction of the cDCC model allows evaluation of the impact of both the lack of consistency and the existence of bias in the estimated parameters of the DCC model of Engle (2002). Aielli (2008) shows that the bias depends on the persistence of the DCC dynamic parameters. We are interested in evaluating if DCC could be used, regardless of its inconsistency. This fact is of interest as DCC has been proposed as a model with correlation targeting, whereas cDCC cannot be targeted, as discussed in Caporin and McAleer (2009).

By including the covariance shrinking method, we evaluate its advantages in large cross-sectional dimensions. Covariance shrinking is computationally feasible and may also reduce the problems associated with the inversion of large covariance matrices,

wherein inversion could be unstable due to the presence of small eigenvalues in the empirical covariances.

Furthermore, the presence in the model set of the Scalar BEKK allows determining if the separated estimation of variances and correlations (typical of CCC- and DCC-type models) has to be preferred to the joint estimation of the entire covariance (as in the BEKK-type models). This study could provide a confirmation of the result of Zumbach (2009) that shows evidence of a preference for covariance models with respect to variance and correlation specifications.

Second, we use the weighted likelihood ratio test of Amisano and Giacomini (2007), which is close to a loss-function based comparison of equal predictive ability based on a likelihood loss function. The test will be applied both in the direct evaluation of covariance forecasts and as an alternative to the Diebold-Mariano test.

Third, we will evaluate and rank the alternative models over different cross-sectional dimensions, starting from two assets, and up to 89 assets, which we select from the S&P100 constituents (a similar dataset has been used in Engle et al., 2008). By this additional study, we will determine if the cross-sectional dimension has a role in determining the preference ordering across models. In other words, by comparing models over an increasing number of variables, we will examine if estimation error and model error play a role in the forecasts of conditional covariance models. The financial literature has discussed extensively the impact of estimation error for the mean returns, leading to results suggesting its strong impact, and making naïve allocations preferable to optimal allocations because of the reduced impact of estimation error (see De Miguel et al., 2009). We draw a parallel within the MGARCH model set, and attempt to answer the question: If the cross-sectional dimension is large, does estimation error affect model performance? If simple or naïve models are preferred, we could interpret this as preliminary evidence in this direction.

We stress that we are comparing alternative feasible models for the evaluation of conditional covariance and/or correlation matrices. The models we consider all belong

to the GARCH and Dynamic Conditional Correlation families, thereby excluding Multivariate Stochastic Volatility models. From our perspective, these models, despite being theoretically appealing, suffer for the curse of dimensionality in a stronger way than MGARCH specifications, and their estimation in large cross-sectional dimensions is likely to be even more complicated than the models considered in this paper. For a survey of Multivariate Stochastic Volatility models, see Asai et al. (2006), and Chib et al. (2009).

Furthermore, we focus on extremely simple models (all are scalar representations), and follow the quasi-maximum likelihood estimation approach. We do not consider more complex parameterizations because we focus on simplicity. We are not interested in the determination of an optimal model or estimation method, but rather on baseline specifications, namely those that are the most common among practitioners, and try to verify if they are equivalent. Clearly, 89 assets is far from the traditional problem dimension for portfolio managers, but this sheds some light on a comparison of model performance across an increasing number of assets. Finally, we stress that our focus in on empirical application, and is not intended to provide a methodological contribution on the most appropriate methods for model comparison, which will be left for future research based on a simulation approach.

Our empirical results show that many models provide similar forecasting performance, both with direct and indirect evaluation methods. However, the set of preferred models is not stable over alternative approaches and sample periods. Furthermore, estimation error of model parameters plays a relevant role, and could be amplified when covariances are used to determine portfolio weights (thus referring to indirect model comparison). Finally, we note a marginal preference for models separately capturing the variance and correlation dynamics with respect to models directly estimating the covariance dynamics.

The paper proceeds as follow. Section 2 presents the model to motivate our choice, discuss the issue of covariance and correlation targeting and show the specifications we will estimate. Section 3 discusses the methods and approaches used to compare the

models. Section 4 presents the dataset used and reports the empirical results. Section 5 gives some concluding comments.

2. Feasible covariance and correlation models for large cross-sectional dimensions

This section briefly introduces the models that will be compared in the empirical application. Let x_t denote a k-dimensional vector of financial variables (returns), μ_t represent the expected mean of x_t from a conditional mean model, and ε_t the mean innovation vector. The following relations hold:

$$
\begin{aligned}
x_t \mid I^{t-1} &\sim D\left(\mu_t, \Sigma_t\right) \\
x_t - \mu_t = \varepsilon_t \mid I^{t-1} &\sim D\left(0, \Sigma_t\right)
\end{aligned}
\tag{1}
$$

where I^{t-1} is the information set at time t-1, $D(.)$ denotes a multivariate density and Σ_t represents the covariance matrix determined conditionally on the information set at time t-1.

In the following, we do not consider the effects of different mean specifications. The mean is fixed at the sample mean determined over the same sample used for the estimation of the parameters, such that $\hat{\mu}_t = \dfrac{1}{t-1} \sum_{i=1}^{t-1} x_i$. The mean could be based on a variety of time series or financial models, which are not the main concern in this paper. What is relevant is that, for each pair of covariance models compared, the mean models are identical. As a result, all forecast discrepancies are due to differences in the expected covariances, while all in-sample differences are due to differences in the estimated covariance models.

2.1 Scalar BEKK

The first model we estimate is the Scalar BEKK with targeting constraint (see Engle and Kroner, 1995; Ding and Engle, 2001; Caporin and McAleer, 2008, 2009, 2010). The fully parameterized BEKK model with all orders set to 1 is given as

$$\Sigma_t = CC' + A\varepsilon_{t-1}\varepsilon_{t-1}'A' + B\Sigma_{t-1}B' \tag{2}$$

where A and B are square parameter matrices, and C is lower triangular. In the scalar BEKK model without targeting, the parameter matrices A and B in (2) are replaced by $A = \alpha^{1/2}I$ and $B = \beta^{1/2}I$, where I is an identity matrix, and α and β are scalar coefficients. When we introduce targeting, the intercept becomes an explicit function of the unconditional covariance matrix $\overline{\Sigma} = E[\varepsilon_t\varepsilon_t']$, and the model becomes

$$\Sigma_t = \overline{\Sigma} + \alpha\left(\varepsilon_{t-1}\varepsilon_{t-1}' - \overline{\Sigma}\right) + \beta\left(\Sigma_{t-1} - \overline{\Sigma}\right). \tag{3}$$

Scalar BEKK in (3) is feasible even for very large cross-sectional dimensions as it contains only two parameters that must be estimated by maximum likelihood, namely the parameters driving the model dynamics. The parameters in the unconditional covariance matrix could be estimated by the sample estimator. Notably, the Scalar BEKK model has standard asymptotic properties, as shown in Jeantheau (1998), Comte and Lieberman (2003), and McAleer et al. (2009), under the existence of 8th order moments.

Diagonal BEKK and VECH parameterizations are not considered as they are not feasible for large cross-sectional dimensions. Note that the VECH model adopted by Engle and Sheppard (2008) is equivalent to a scalar BEKK model.

2.2 Variance and correlation models

We will estimate three models based on a decomposition of the covariance matrices into variances and correlations. The first is the CCC model of Bollerslev (1990) which, starting from (1), assumes that the covariance matrix satisfies

$$\Sigma_t = D_t R D_t \tag{4}$$

where $D_t = diag\left(\sigma_{1,t}, \sigma_{2,t}, ... \sigma_{k,t}\right)$ is a diagonal matrix of conditional standard deviations, and R is an unconditional correlation matrix (the operator *diag(a)* generates a diagonal matrix, with the vector *a* along the main diagonal). We assume that all the conditional variances follow a simple GARCH(1,1) process without asymmetry in order to make the model directly comparable with Scalar BEKK. The addition of an asymmetric term would require an asymmetric BEKK model for a fair comparison.

In the CCC model, the correlation matrix R is determined using a traditional sample estimator. The model is estimated using a two-step approach, namely the conditional variances on each specific series, and then estimate the unconditional correlation matrix using the standardized residuals. This approach makes the model feasible, even with a large number of assets. The CCC model is a special case of the VARMA-GARCH model of Ling and McAleer (2003).

The DCC model of Engle (2002) was proposed as a generalization of the CCC model, and replaces the correlation matrix in (4) with a time-varying conditional correlation. The DCC is given as:

$$\Sigma_t = D_t R_t D_t \tag{5}$$

$$R_t = \bar{Q}_t^{-1} Q_t \bar{Q}_t^{-1}, \quad \bar{Q}_t = diag\left(dg\left(Q_t\right)\right)^{\frac{1}{2}} \tag{6}$$

$$Q_t = S + \alpha\left(D_{t-1}^{-1}\varepsilon_{t-1}\varepsilon_{t-1}'D_{t-1}^{-1} - S\right) + \beta\left(Q_{t-1} - S\right) \tag{7}$$

where D_t is the same as for the CCC model, the operator *dg(A)* extracts the main diagonal from matrix *A*, *S* is the unconditional correlation matrix, and α and β are the scalar parameters driving the model dynamics. Following Engle (2002), the model is estimated with a three-stage approach, namely estimate the conditional variance parameters and filter them, estimate *S* using the standardized residuals $D_t^{-1}\varepsilon_t$ to obtain a

sample estimator, and then conditionally on the previous estimates, maximize the conditional correlation log-likelihood with respect to the parameters driving the dynamics in (7). The introduction of a multi-step estimation method clearly reduces the efficiency, as shown in Engle and Sheppard (2001), but makes the model feasible with large cross-sectional dimensions. Note that the model in (5)-(7) includes targeting, as defined by Caporin and McAleer (2010) under assumptions which are deeply analysed and criticized in Aielli (2008). Without targeting, the model is inextricably exposed to the curse of dimensionality as the matrix *S* contains *0.5k(k-1)* parameters to be jointly estimated with *α* and *ß*.

Aielli (2008) shows that the sample estimator of *S* used in the second step of the DCC estimation method is inconsistent, thereby also affecting the consistency of the third step. In order to resolve this serious issue, Aielli (2008) introduces the cDCC model, which replaces (7) with

$$
Q_t = S + \alpha \left(\tilde{Q}_{t-1}^{\frac{1}{2}} D_{t-1}^{-1} \varepsilon_{t-1} \varepsilon_{t-1}' D_{t-1}^{-1} \tilde{Q}_{t-1}^{\frac{1}{2}} - S \right) + \beta \left(Q_{t-1} - S \right)
$$
$$
\tilde{Q}_{t-1} = diag \left(dg \left(Q_{t-1} \right) \right)
$$

(8)

where the parameter matrix *S* is symmetric, has unit elements over the main diagonal, and is now the covariance matrix of the innovations $\tilde{Q}_t^{\frac{1}{2}} D_t^{-1} \varepsilon_t$, which are not observable. The modification restores consistency, under unstated assumptions, but again exposes the model to the curse of dimensionality as the matrix *S* in (8) has to be estimated (see Aielli (2008) for further details). As noted in Caporin and McAleer (2010), correlation targeting is excluded for the cDCC model as *S* is not a correlation matrix and is not estimated using the available sample information. Aielli (2008) suggests a feasible estimation method.

For purposes of comparison with the other multivariate GARCH models examined, we will discuss briefly the estimation method of cDCC. Two steps are now required. In the first stage, we filter the conditional variances wherein the matrices D_t are estimated. The second step jointly determines the parameters driving the dynamics, α and β, and

the parameters in *S*, by maximizing the likelihood of (8). The core of the method is in the steps for the evaluation of the log-likelihood. Assuming normality, the log-likelihood of (8) is given as

$$LogL(\alpha,\beta,S) = \sum_{t=1}^{T}\left(-\frac{1}{2}\ln|R_t| - \frac{1}{2}\left(D_t^{-1}\varepsilon_t\right)' R_t^{-1} D_t^{-1}\varepsilon_t\right) \qquad (9)$$

where the conditional correlation matrix, R_t, follows from (6) and (8), and the innovations are given by the standardized residuals $D_t^{-1}\varepsilon_t$. Generally, to evaluate (9) conditionally on a set of parameters (α,β,S), we run the recursion associated with (8), determine the sequences of conditional correlation matrices R_t, and then compute the log-likelihood. A numerical optimization algorithm then iteratively evaluates (9) to determine the optimal estimates for (α,β,S).

In order to resolve the curse of dimensionality caused by *S*, Aielli (2008) suggests the following alternative steps for the evaluation of the likelihood in (9):

i) Conditionally on a set of parameters (α,β), run the recursions

$$q_{ii,t} = (1-\alpha-\beta) + \alpha q_{ii,t-1}\sigma_{i,t-1}^{-2}\varepsilon_{i,t-1}^2 + \beta q_{ii,t-1}, \qquad i = 1,2,...k$$

and compute the residuals $\tilde{Q}_t^{\frac{1}{2}}D_t^{-1}\varepsilon_t$;

ii) Evaluate S conditionally on the choice of (α,β) using the sample estimator

$$\hat{S}(\alpha,\beta) = \frac{1}{T}\sum_{t=1}^{T}\tilde{Q}_t^{\frac{1}{2}}D_t^{-1}\varepsilon_t\varepsilon_t' D_t^{-1}\tilde{Q}_t^{\frac{1}{2}};$$

iii) Conditionally on (α,β) and the estimate $\hat{S}(\alpha,\beta)$, evaluate the recursions

$$q_{ij,t} = \left(1 - \alpha - \beta\right)\hat{s}_{ij} + \alpha q_{ii,t-1}^{1/2}\sigma_{i,t-1}^{-1}\varepsilon_{i,t-1}\varepsilon_{j,t-1}\sigma_{j,t-1}^{-1}q_{jj,t-1}^{1/2} + \beta q_{ij,t-1}$$
$$i,j = 1,2,...k,\ \ j > i;$$

iv)   Evaluate the likelihood in (9).

Note that step i) does not depend on the elements in S as, by construction, S has unit elements along the main diagonal. Furthermore, the innovations $\tilde{Q}_t^{1/2}D_t^{-1}\varepsilon_t$ do not depend, by construction, on the off-diagonal elements of Q_t. The evaluation of the likelihood in (9) under the cDCC model requires the evaluation of a sample covariance, a limited additional effort with respect to the standard DCC model of Engle (2002). Note that the estimator of S provided by maximizing (9) following steps i)-iv) is conditional on the optimal choice of $\left(\alpha,\beta\right)$. In order to determine the optimal parameter values, a numerical algorithm should maximize (9) with respect to the parameters $\left(\alpha,\beta\right)$ and the recursions in steps i)-v). The estimation output provides the optimal values for $\left(\alpha,\beta\right)$ and the estimate of S associated with the optimal parameters driving the model dynamics. The cDCC model uses a different targeting approach, which we may define as "implicit targeting", as it uses a sample estimator conditional on a parameter choice within the evaluation of the log-likelihood function.

Finally, Aielli (2008) shows that the lack of consistency of the three-step DCC estimator depends strictly on the persistence of the parameters driving the correlation dynamics and on the relevance of the innovations. The bias is an increasing function of both α and $\alpha + \beta$. Not surprisingly, standard estimates obtained when fitting DCC models are small, and are close to 0 for α to 1 for $\alpha + \beta$, thereby leading to an opposite effect on the size of the bias. Therefore, in this paper we will determine if the bias is relevant in practical applications as a commentary on the inconsistent estimates of the standard scalar DCC model.

2.3 Naïve specifications

The last two models considered are the Exponentially Weighted Moving Average model and the Covariance Shrinking approach of Ledoit and Wold (2003, 2004). The EWMA model provides a recursion for the evaluation of the conditional covariance matrix, which is based on a single parameter λ:

$$\Sigma_t = (1-\lambda)\varepsilon_{t-1}\varepsilon_{t-1}' + \lambda\Sigma_{t-1} \tag{10}$$

The covariance matrix in (10) could be represented in a companion form highlighting its relationship with exponentially weighted sums of past innovation cross-products. In the empirical application, contrary to standard practice, we estimate the parameter λ, called the smoothing coefficient, as it requires a limited computational effort. By construction, the EWMA is feasible even for very large cross-sectional dimensions.

Finally, we consider the covariance shrinking approach of Ledoit and Wolf (2003, 2004). The authors proposed a method that is designed to find a compromise between the large estimation errors in the sample covariance and the misspecification error in the estimators of the covariance. Ledoit and Wolf (2003) suggest determining a covariance by combining a sample estimator of the covariance and a single index covariance (Ledoit and Wolf, 2003), or a constant correlation covariance (Ledoit and Wolf, 2004). Following the covariance shrinking approach, we define the expected covariance for time t as follows:

$$\Sigma_t = (1-\lambda)S_{t-1} + \lambda F_{t-1} \tag{11}$$

where S_{t-1} is the sample covariance matrix determined up to time t-1, and F_{t-1} is a structured estimator determined using the information set to time t-1, and is called shrinkage target. The coefficient λ, the shrinkage constant, has to be estimated, and depends on the form of the shrinkage target (for further details, see Ledoit and Wolf (2003, 2004)). In the following, we will consider as the shrinkage target the covariance with constant correlation, as described in Ledoit and Wolf (2004).

3. Comparing competing covariance and correlation models

We will present briefly the approaches to be used in comparing the models described in the previous section. Before moving to the methods, we introduce some notation.

It is assumed that the models are to be compared using out-of-sample forecasts, where forecasts are made one period ahead and for an evaluation period from *T+1* to *T+h*. Information to time *T* is used to estimate the various models and to produce the conditional forecasts for time *T+1*. The estimation sample is rolled forward, and information from time *2* to *T+1* is used to forecast the covariance matrix for time *T+2*, and so on, to time *T+h*. In order to avoid any dependence on the mean dynamics, we fit the mean using its sample estimator across all models (the sample mean is estimated with the same rolling approach). The one-step-ahead covariance forecasts for time *T+i* are denoted by $\hat{\Sigma}_{T+i}^{m}$, where *m* is the model index (*m=1,2,...M*). Note that, by construction, the forecasts are conditional on the information set at time *T+i-1*. The mean forecasts are denoted by $\hat{\mu}_{T+i}$, and do not depend on the model. For simplicity, we suppress the conditioning information set from the forecast notation.

We follow Patton and Sheppard (2009) and consider separately the direct and indirect evaluation methods. Within the first group, we include two approaches based on common loss functions, namely the Diebold-Mariano test, and the MCS approach of Hansen et al. (2005), and the test proposed by Amisano and Giacomini (2007).

For the Diebold-Mariano test, we consider the MSE loss function (see Patton and Sheppard, 2009):

$$lf_{m,T+i} = \frac{1}{k^2} \mathbf{1}_{k^2}' \left(vec\left(\hat{\Sigma}_{T+i}^{m} - e_{T+i}e_{T+i}' \right)' vec\left(\hat{\Sigma}_{T+i}^{m} - e_{T+i}e_{T+i}' \right) \right) \tag{12}$$

where $e_{T+i} = x_{T+i} - \hat{\mu}_{T+i}$ (note that the observed time *T+i* return is used), and the time *T+i* true volatility is approximated by $e_{T+i}e_{T+i}'$. The MSE loss function belongs to the

class of robust loss functions defined in Patton and Sheppard (2009) (see also Clements et al. (2009) and Laurent et al. (2009)). They also consider the QLIKE loss function of Patton (2010), but in the multivariate framework, as in Patton and Sheppard (2009), the QLIKE loss function is infeasible when the volatility proxy is the cross-product of realized returns (see also Laurent et al., 2009). Alternative realized covariance proxies could be used instead of $e_{T+i}e_{T+i}'$. A common choice is the realized covariance, as in Laurent et al. (2010). However, high frequency data on a large dimension dataset were not available to us, and we prefer to focus on model comparison for model dimensions larger than in Laurent et al. (2010).

The test of equal predictive ability corresponds to checking the null hypothesis of zero loss function differentials, $H_0 : E\left[\overline{lf}_j - \overline{lf}_l\right] = E\left[\overline{LF}_{jl}\right] = 0$, where i and j are two different model indices, $\overline{lf}_j = \frac{1}{h}\sum_{i=1}^{h} lf_{j,T+i}$, and $\overline{LF}_{jl} = \overline{lf}_j - \overline{lf}_l$ The test statistic is given as

$$t_{jl} = \frac{\sqrt{h}\,\overline{LF}_{jl}}{\sigma\left(\sqrt{h}\,\overline{LF}_{jl}\right)} \sim N(0,1) \tag{13}$$

where $\sigma\left(\sqrt{h}\,\overline{LF}_{jl}\right)$ is the heteroskedasticity and autocorrelation (HAC) consistent estimate of the asymptotic variance of $\sqrt{h}\,\overline{LF}_{jl}$.

The Amisano-Giacomini (2007) is based on the logarithmic scores of two competing models over the forecast evaluation period. We first denote the model m log-scores as:

$$\log f_{m,T+i} = -\log\left|\hat{\Sigma}_{T+i}^{m}\right| - e_{T+i}'\left(\hat{\Sigma}_{T+i}^{m}\right)^{-1}e_{T+i}, \tag{14}$$

where all symbols have the same meaning as in (13). Amisano and Giacomini (2007) then consider the following quantity:

$$L_{jl,T+i} = w(e_{T+i})\left[\log f_{j,T+i} - \log f_{l,T+i}\right] \tag{15}$$

where j and l represent two different models, and $w(e_{T+i})$ is a weighting function. The null hypothesis of equal predictive ability of the two models is $H_0 : E\left[\bar{L}_{jl,T}\right] = 0$, where $\bar{L}_{jl,T} = \dfrac{1}{h}\sum_{i=1}^{h} L_{jl,T+i}$, while the alternative hypothesis refers to different predictive ability.

The test statistic is given as:

$$t_{jl} = \frac{\bar{L}_{jl,T}}{\sigma\left(\bar{L}_{jl,T}\right)/\sqrt{h}} \sim N(0,1) \tag{16}$$

where $\sigma\left(\bar{L}_{jl,T}\right)$ is the heteroskedasticity and autocorrelation consistent estimate of the asymptotic variance $Var\left[\sqrt{h}\bar{L}_{jl,T}\right]$. If the null hypothesis is rejected, the test statistic sign could be used to determine the model preference (positive values suggests a preference for model j (see equation (22)). Amisano and Giacomini (2007) introduce in (15) a weighting function which may be used to focus the test on a specific region of the forecast density, such as the tail area. Within the direct model evaluation framework, we consider equal weights for all points over the forecast horizon ($w(e_{T+i}) = h^{-1}$). Note that the quantity in (14) can be considered as a gain, so that positive values of the test statistic in (16) are associated with a preference for the first model. In considering the Diebold-Mariano test, positive values of the test statistic show evidence for a preference for the second model (that is, with smaller losses).

The Diebold-Mariano and Amisano-Giacomini tests permit pairwise comparisons of models. However, the test outcomes do not ensure that an optimal test is clearly identified. For this reason, we consider the Model Confidence Set approach, which performs a joint forecast comparison across all models. The MCS performs an iterative selection procedure, testing at step j the null hypothesis of equal predicting ability of all

models included in a set $\mathcal{M}_\ell$ (the starting set $\mathcal{M}_0$ contains all models) under a given loss function. The null hypothesis has the form

$$H_0 : E\left[\overline{lf}_{jl} - \overline{lf}_{jl}\right] = E\left[\overline{LF}_{jl}\right] = 0, \quad j > l, \forall j, l \in \mathcal{M}_\ell \tag{17}$$

where the notation is the same as for the Diebold-Mariano test in (12) and (13).

Note that the same procedure, as well as all the subsequent statistics and tests, could also be used with the Amisano-Giacomini log-scores, with the caution of changing their sign (thereby translating them from gains to losses). Hansen et al. (2005) propose two statistics to test the null hypothesis:

$$t_R = \max_{j,l \in \mathcal{M}_\ell} \left| \frac{\overline{LF}_{jl}}{s\left(\overline{LF}_{jl}\right)} \right| \tag{18}$$

$$t_{SQ} = \sum_{j,l \in \mathcal{M}_\ell, j>l} \left(\frac{\overline{LF}_{jl}}{s\left(\overline{LF}_{jl}\right)} \right)^2 \tag{19}$$

where $s\left(\overline{LF}_{jl}\right)$ is a bootstrap estimate of the variance of $\overline{LF}_{jl}$, and the p-values of the test statistics are determined using a bootstrap approach. If the null hypothesis is rejected at a given confidence level, the worst performing model is excluded from the set (rejection is determined on the basis of bootstrap p-values under the null). Such a model is identified as follows:

$$j = \arg\max_{j \in \mathcal{M}_\ell} \left(\sum_{l \in \mathcal{M}_\ell} \overline{LF}_{jl} \right) \left(Var\left(\sum_{l \in \mathcal{M}_\ell} \overline{LF}_{jl} \right) \right)^{-1/2} \tag{20}$$

18

where the variance is computed again using a bootstrap method. In the empirical analysis given below, we will use both the Diebold-Mariano loss function in (12), as well as of the Amisano-Giacomini log-scores in (14) (minus the log-scores to transform them into losses) (see Hansen et al. (2005) for further details on MCS).

For the indirect evaluation of the multivariate models, we consider an asset allocation framework and compare the impact of model choice by contrasting the performances of specific portfolios: (i) equally weighted portfolio, denoted as EW, which is not exposed to the asset return mean estimation error, and is superior to many other portfolios (see De Miguel et al. (2009)); and (ii) global minimum variance portfolio with and without short selling constraints, denoted as GMV and GMVB, respectively. The weights of the equally weighted portfolio are $\mathbf{w} = k^{-1}\mathbf{1}$, where $\mathbf{1}$ is a k-dimensional vector of unit elements. The GMV weights are time- and model-dependent, and are based on the covariance forecasts:

$$\mathbf{w}_{T+i}^{m} = \frac{\left(\hat{\Sigma}_{T+i}^{m}\right)^{-1}\mathbf{1}}{\mathbf{1}'\left(\hat{\Sigma}_{T+i}^{m}\right)^{-1}\mathbf{1}} \tag{21}$$

Finally, GMVB weights $\hat{\mathbf{w}}_{T+i}^{m}$ are determined by solving the optimum problem:

$$\arg\min_{\mathbf{w}} \ \mathbf{w}'\hat{\Sigma}_{T+i}^{m}\mathbf{w}$$
$$\text{s.t. } w_l \geq 0, \ \ l = 1, 2, ...k \tag{22}$$
$$\text{and } \mathbf{w}'\mathbf{1} = 1$$

for each forecast evaluation period and for each model.

We then define the following quantities for the three portfolio strategies (based on the weights EW, GMV and GMVB, respectively):
(a) realized portfolio returns:

$$R_{T+i,EW} = \mathbf{w}'x_{T+i}, \ R_{T+i,GMV}^{m} = \mathbf{w}_{T+i}^{m}{}'x_{T+i}, \ R_{T+i,GMVB}^{m} = \hat{\mathbf{w}}_{T+i}^{m}{}'x_{T+i}, \ i = 1, 2, 3...h, \ m = 1, 2, 3,...M \ ;$$

(b) expected portfolio returns:

$$\hat{R}_{T+i,EW} = \mathbf{w}' \hat{\mu}_{T+i}, \ \hat{R}^m_{T+i,GMV} = \mathbf{w}^m_{T+i}{}' \hat{\mu}_{T+i}, \ \hat{R}^m_{T+i,GMVB} = \hat{\mathbf{w}}^m_{T+i}{}' \hat{\mu}_{T+i}, \ i = 1,2,3...h, \ m = 1,2,3,...M \ ;$$

(c) realized portfolio variances:

$$s_{T+i,EW} = \left(\mathbf{w}' \left(x_{T+i} - \hat{\mu}_{T+i} \right) \right)^2, \ s^m_{T+i,GMV} = \left(\mathbf{w}^m_{T+i}{}' \left(x_{T+i} - \hat{\mu}_{T+i} \right) \right)^2,$$

$$s^m_{T+i,GMVB} = \left(\hat{\mathbf{w}}^m_{T+i}{}' \left(x_{T+i} - \hat{\mu}_{T+i} \right) \right)^2, \ i = 1,2,3...h, \ m = 1,2,3,...M \ ;$$

(d) expected portfolio variances:

$$h^m_{T+i,EW} = \mathbf{w}' \hat{\Sigma}^m_{T+i} \mathbf{w}, \ h^m_{T+i,GMV} = \mathbf{w}^m_{T+i}{}' \hat{\Sigma}^m_{T+i} \mathbf{w}^m_{T+i}, \ h^m_{T+i,GMVB} = \hat{\mathbf{w}}^m_{T+i}{}' \hat{\Sigma}^m_{T+i} \hat{\mathbf{w}}^m_{T+i}, \ i = 1,2,3...h,$$

$$m = 1,2,3,...M \ .$$

Given the portfolio mean and variance forecasts and the realized portfolio returns, we define the following logarithmic scores (recall that the vector x_{T+i} contains the time $T+i$ observed returns for the analyses assets):

$$\log f_{m,EW,T+i} \left(R_{T+i,EW} \right) = -\frac{1}{2} \ln \left(h^m_{T+i,EW} \right) - \frac{1}{2} \left(R_{T+i,EW} - \hat{R}_{T+i,EW} \right)^2 \left(h^m_{T+i,EW} \right)^{-1} \qquad (23)$$

where *m* is the model index, and the portfolio strategy is EW. Equation (23) represents the logarithmic score when mean and variances forecasts are made conditionally on time *T+i-1*. Note that the logarithmic score is evaluated at the true observed values at time *T+i*. Such a quantity can be evaluated only for the EW portfolio strategy, which is the only strategy that provides a 'true' value $R_{T+i,EW}$. In fact, the GMV and GMVB strategies allow a determination of the realized returns, but these are exposed to the estimation error implicit in the determination of portfolio weights. As a result, in order to avoid introducing distortions in the test statistics, we consider only the EW strategy.

Note that portfolio weights do not have a proper source of estimation error. In fact, GMV weights are an analytical non-linear function of covariance forecasts, and are influenced by the estimation error of the model parameters. Moreover, GMVB weights are a non-linear function of covariance forecasts, but their evaluation is purely numerical, and this may include optimization and approximation errors. The dependence of weights on covariance forecasts could produce an amplification of parameter estimation error, given that portfolio variances are functions of covariance forecasts and portfolio weights.

The second approach we apply for the indirect model evaluation uses the MSE and QLIKE univariate loss functions:

$$lf_{m,p,T+i}^{1} = \left(h_{T+i,p}^{m} - s_{T+i,p}^{m} \right)^{2} \tag{24}$$

$$lf_{m,p,T+i}^{2} = \log \left(h_{T+i,p}^{m} \right) + s_{T+i,p}^{m} \left(h_{T+i,p}^{m} \right)^{-1} \tag{25}$$

The indirect evaluation of the models then proceeds using the Diebold-Mariano-type tests desicussed previously. Similarly, the loss functions in (24) and (25) are also used for the indirect evaluation by the MCS approach. Note that we do not compare multivariate models indirectly by mean of utility-based loss functions because Clements et al. (2009) show that these functions make the impact of the models very limited, thereby reducing the possibility of detecting discrepancies across models.

Finally, in the indirect comparison we also consider some of the model comparison approaches suggested in Engle and Colacito (2006). In particular, we report the out-of-sample averages of the expected variances $\bar{h}_{p}^{m}$ (with a preference for a lower average variance), and we also test the significance of the intercept of the regressions in:

$$\frac{s_{T+i,p}^{m}}{h_{T+i,p}^{m}} - 1 = \beta + \xi_{T+i} \tag{26}$$

where ξ_{T+i} is an innovation term, and robust HAC standard errors are required. If we consider (26), accurate models should have a zero intercept. Engle and Colacito (2006) also propose pairwise comparisons based on Diebold-Mariano type tests but, as they are closely related to the methods already described, we do not consider them in the empirical analysis below.

4. Data description and selected models

In order to compare the models presented in the previous sections, we have selected a dataset similar to that of Engle et al. (2009). We downloaded from Datastream the S&P100 constituents at the end of March 2009. Then we selected only those assets with total return indices available from the beginning of 1997 to the end of March 2009. The selected period contains 3194 daily returns. The list of the 89 selected stocks is reported in Appendix A.

We fit the following models (acronyms are given in parentheses): Scalar BEKK with covariance targeting (BEKK); Scalar DCC with correlation targeting (DCC); Scalar cDCC with implicit correlation targeting (cDCC); exponentially weighted moving average, with estimated smoothing coefficient (EWMA); constant conditional correlation model (CCC); and covariance shrinking (SHR) with constant correlation shrinkage target, as in Ledoit and Wolf (2004). The BEKK, DCC and cDCC models are estimated under both Multivariate Normal and Multivariate Student densities, using the Multivariate Student density adopted by Kawakatsu (2005), where the covariance forecast is not a function of the degrees of freedom. The covariance shrinking case and the EWMA with calibrated smoothing coefficient provide distribution-free covariance forecasts. Finally, in the CCC specification, the correlation matrix is, by construction, identical under both Multivariate Normality and Multivariate Student of Kawakatsu (2005), thereby providing identical covariance forecasts.

For the Multivariate Student cases, the degrees of freedom are estimated jointly with the other model parameters. When Normal or Student densities are used, model acronyms will be matched with N or T. Overall, we consider 9 different specifications, namely

EWMA, SHR, CCC, DCC(N), DCC(T), cDCC(N), cDCC(T), BEKK(N), and BEKK(T). In order to avoid dependence of the model comparison procedures with the mean return forecasts, these are always fixed at the sample mean.

The previous 9 specifications are evaluated for different problem dimensions, such that each model is estimated for 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80 and 89 assets. In these empirical applications, assets are ordered alphabetically, and we progressively enlarge the number of variables used in the model estimation and evaluation steps.

We estimate the models with a rolling approach, and then compare the out-of-sample forecasts for one year of daily observations. All models for all problem dimensions are estimated daily and are used to produce one-step-ahead forecasts. We consider two different out-of-sample evaluation periods. In the first, we focus on extreme market conditions and compare models for the period April 2008 – March 2009. This could be considered as a model stress-test to determine if more highly parameterized models are preferred to simpler or naïve specifications as they are not exposed to parameter uncertainty and instability. For this forecast range, we also perform a robustness check by inverting the asset order to verify the absence of distortions due to asset order. The second forecast evaluation period is for 2006, when the market was in a low volatility state and was trending upward. This second comparison allows verification of whether the model ranking might be affected by overall market conditions.

In the following section, we summarize the results in Tables 3 and 7, and report selected tables for model comparison. All the empirical results are collected in an Appendix, which is available from the authors upon request.

4.1 Results for direct model comparisons

The first result provided by the Diebold-Mariano and Amisano-Giacomini tests is the negative performance of the covariance shrinking approach over both evaluation periods

(see Table 1 and 2 for Diebold-Mariano results in the two forecast evaluation ranges - these results are not affected by assets order, as confirmed by the robustness check).

Table 1: Diebold-Mariano model comparison based on MSE loss function in the crisis range

		2	3	4	5	10	15	20	25	30	35	40	45	50	60	70	80	89
EWMA	SHR	-5.1	-5.8	-8.1	-10	-12.1	-12.9	-13	-14.5	-15	-15.9	-15.4	-15.6	-15.8	-15.6	-16	-16.7	-16.4
EWMA	CCC	-1.8	-1.9	-1.5	-2.1	-3.2	-2.5	-2.6	-2.8	-2.6	-2.6	-2.5	-2	-2.1	-2.2	-1.8	-1.8	-1.5
EWMA	DCC(N)	-1.8	-1.8	-1.5	-2.1	-3.3	-2.5	-2.5	-2.5	-2.3	-2.3	-2.1	-1.6	-1.7	-1.8	-1.5	-1.4	-1.2
EWMA	cDCC(N)	-1.8	-1.8	-1.5	-2.1	-3.2	-2.5	-2.5	-2.5	-2.4	-2.3	-2.1	-1.6	-1.7	-1.8	-1.5	-1.4	-1.3
EWMA	BEKK(N)	-1.1	-0.6	-0.1	-0.6	-0.5	0	-0.6	-0.2	-0.5	-1.3	-1.4	-2	-2.6	-3.7	-4.5	-5.8	-6.5
EWMA	DCC(T)	-1.6	-1.6	-1.6	-2.2	-3.2	-2.5	-2.5	-2.5	-2.4	-2.3	-2.1	-1.6	-1.7	-1.8	-1.4	-1.4	-1.2
EWMA	cDCC(T)	-1.7	-1.8	-1.9	-2.2	-3	-2.5	-2.5	-2.5	-2.4	-2.3	-2.2	-1.6	-1.7	-1.8	-1.5	-1.4	-1.2
EWMA	BEKK(T)	0.7	0.2	0.1	-0.4	-1.7	-2.8	-4.9	-6.4	-7.9	-9.3	-9.9	-10.8	-11.7	-12.5	-13.4	-14.6	-14.6
SHR	CCC	3.7	5.3	8.2	10	10.1	9.6	10.2	12.7	13.7	15.2	14.8	15.3	15.8	15.6	16.2	17	16.7
SHR	DCC(N)	3.7	5.3	8.1	9.9	10.1	9.6	10.2	12.7	13.8	15.2	14.9	15.4	15.9	15.8	16.3	17.2	16.8
SHR	cDCC(N)	3.7	5.3	8.1	9.9	10.1	9.6	10.2	12.8	13.8	15.2	14.9	15.4	15.9	15.8	16.3	17.2	16.9
SHR	BEKK(N)	5.9	6.9	9.3	11.5	13.9	15.1	15.3	16.7	17.3	18.2	18	18.3	18.9	19.2	20.3	21.8	21.8
SHR	DCC(T)	3.4	5.2	7.7	9.4	10.2	9.6	10.2	12.7	13.8	15.2	14.9	15.4	15.9	15.8	16.4	17.2	16.8
SHR	cDCC(T)	3.6	4.9	7.7	10	10.3	9.6	10.2	12.8	13.8	15.2	14.9	15.4	15.9	15.8	16.4	17.2	16.8
SHR	BEKK(T)	5.8	6.7	9.3	11.6	14.9	17.1	18.4	20.9	22.3	23.8	23.8	24.4	25.3	25.1	26	26.8	26.3
CCC	DCC(N)	-0.2	1.4	0.8	0.9	0.1	1.7	3.2	5	4.4	5.8	7.1	5.7	5.9	6.3	6.1	5.8	5.5
CCC	cDCC(N)	-0.1	1.2	0.6	0.9	0.4	2	3.1	4.9	4.3	5.5	6.9	5.4	5.5	5.9	5.8	5.4	5.2
CCC	BEKK(N)	1.2	1.2	1.3	1.5	2.7	2.2	1.8	2.1	1.8	1.2	1	-0.4	-1.4	-2.9	-4.2	-5.8	-6.6
CCC	DCC(T)	-1	0	-1	-1.4	1.6	0.9	2.1	4.4	3.9	5.4	6.5	5.3	5.5	6.1	5.9	5.9	5.5
CCC	cDCC(T)	-1	-1.1	-1.6	-0.1	1.9	1	2.2	4.4	3.7	5.1	6.2	4.9	5.1	5.7	5.5	5.5	5.2
CCC	BEKK(T)	1.7	1.7	1.4	1.5	1.9	0.7	-1.5	-3.6	-5.9	-8	-8.9	-10.3	-11.7	-12.5	-13.6	-15	-14.9
DCC(N)	cDCC(N)	0.5	-2.5	-2	-1.2	2	1.1	-1.3	-2.5	-2.9	-4.7	-5.9	-5.6	-6.4	-7	-7.1	-5.8	-5.1
DCC(N)	BEKK(N)	1.2	1.1	1.2	1.4	2.7	2.1	1.7	1.8	1.5	0.8	0.6	-0.8	-1.7	-3.2	-4.5	-6	-6.8
DCC(N)	DCC(T)	-1	-0.5	-1.3	-1.7	1.9	-0.4	-0.6	-0.5	-0.5	-0.8	-0.5	-0.1	-0.1	0.4	1	2.1	0.7
DCC(N)	cDCC(T)	-1	-1.5	-2	-0.6	1.9	0.1	-0.3	0.4	-1	-1.6	-1.6	-1.4	-1.6	-1.6	-1	0.4	-0.4
DCC(N)	BEKK(T)	1.7	1.6	1.4	1.5	1.9	0.7	-1.6	-3.8	-6	-8.1	-9	-10.5	-11.8	-12.7	-13.8	-15.2	-15
cDCC(N)	BEKK(N)	1.2	1.1	1.3	1.5	2.7	2.1	1.7	1.8	1.5	0.9	0.6	-0.8	-1.7	-3.2	-4.4	-6	-6.8
cDCC(N)	DCC(T)	-1	-0.4	-1.3	-1.7	1.7	-0.5	-0.4	-0.1	-0.1	0.1	0.6	1.1	1.3	2.4	3	3.4	1.5
cDCC(N)	cDCC(T)	-1	-1.4	-2	-0.6	1.9	0	-0.2	0.6	-0.6	-0.8	-0.5	-0.2	-0.2	0.2	0.8	2	0.6
cDCC(N)	BEKK(T)	1.7	1.6	1.4	1.5	1.9	0.7	-1.6	-3.8	-6.1	-8.1	-9	-10.5	-11.8	-12.7	-13.8	-15.2	-15
BEKK(N)	DCC(T)	-1.2	-1	-1.3	-1.6	-2.6	-2.1	-1.7	-1.9	-1.5	-0.8	-0.6	0.8	1.7	3.2	4.5	6	6.8
BEKK(N)	cDCC(T)	-1.2	-1.3	-1.5	-1.6	-2.5	-2.1	-1.7	-1.8	-1.5	-0.9	-0.6	0.8	1.7	3.2	4.5	6	6.8
BEKK(N)	BEKK(T)	2.1	2.3	1.9	1.3	-3.9	-5.7	-7.7	-9.2	-10.7	-11.8	-12.8	-13.8	-15.1	-16.2	-17.9	-19.8	-19.8
DCC(T)	cDCC(T)	0.8	-1.8	-0.6	1.6	1.3	0.5	0.7	1	-1.9	-5	-6.2	-5.9	-6.7	-7.5	-7.1	-5.6	-5.2
DCC(T)	BEKK(T)	1.5	1.4	1.4	1.6	1.8	0.7	-1.6	-3.8	-6	-8.1	-9	-10.5	-11.8	-12.7	-13.8	-15.2	-15
cDCC(T)	BEKK(T)	1.6	1.6	1.6	1.6	1.6	0.7	-1.6	-3.8	-6	-8.1	-9	-10.5	-11.8	-12.7	-13.8	-15.2	-15

The table reports in the first and second columns the compared models. The first row reports the cross-sectional dimension. Bold dark-grey shaded numbers identify a preference for the first model, while bold numbers identify a preference for the second model.

Table 2: Diebold-Mariano model comparison based on MSE loss function over 2006

		2	3	4	5	10	15	20	25	30	35	40	45	50	60	70	80	89
EWMA	SHR	-2.1	-2.5	-2.3	-2.1	-2.4	-3.3	-3.6	-3.5	-3.4	-3.3	-3.2	-3.1	-3.1	-2.8	-3	-2.8	-2.9
EWMA	CCC	-0.7	-0.4	-1	-0.9	-1.1	-1.2	-1.4	-1.4	-1.5	-1.4	-1.4	-1.4	-1.4	-1.3	-1.3	-1.2	-1.3
EWMA	DCC(N)	0.2	0.7	0.1	0.3	0	-0.5	-0.8	-0.7	-0.9	-0.9	-0.9	-1	-1	-1	-1	-1	-1.1
EWMA	cDCC(N)	0.3	0.9	0.3	0.5	0.1	-0.4	-0.7	-0.7	-0.9	-0.9	-0.9	-1	-1	-1	-1	-1	-1.1
EWMA	BEKK(N)	0.3	-1	-0.9	-0.9	-1.1	-1.7	-2	-2.1	-2.1	-2.1	-2.1	-2.1	-2.1	-2	-2.1	-2.1	-2.2
EWMA	DCC(T)	0.3	0.8	0.2	0.3	0	-0.5	-0.8	-0.8	-0.9	-1	-0.9	-1	-1.1	-1	-1.1	-1	-1.1
EWMA	cDCC(T)	0.8	1	0.4	0.7	0.1	-0.4	-0.7	-0.7	-0.9	-0.9	-0.9	-1	-1	-1	-1	-1	-1.1
EWMA	BEKK(T)	-1	-1.1	-1	-1	-1.3	-1.9	-2.3	-2.4	-2.4	-2.5	-2.4	-2.4	-2.4	-2.3	-2.4	-2.4	-2.5
SHR	CCC	2.6	2.7	2.4	2.4	2.9	4	4.3	4	4	4	3.9	3.9	3.9	3.5	3.8	3.8	3.9
SHR	DCC(N)	2.4	2.6	2.3	2.3	2.8	3.9	4.2	3.9	3.9	3.9	3.9	3.8	3.8	3.5	3.7	3.7	3.8
SHR	cDCC(N)	2.4	2.6	2.3	2.2	2.8	3.8	4.2	3.9	3.9	3.9	3.9	3.8	3.8	3.5	3.7	3.7	3.8
SHR	BEKK(N)	2.4	2.8	2.5	2.4	2.9	4	4.3	4.3	4.3	4.3	4.4	4.4	4.4	4.3	4.5	4.4	4.5
SHR	DCC(T)	2.5	2.6	2.3	2.3	2.8	3.9	4.2	3.9	3.9	4	3.9	3.8	3.8	3.5	3.7	3.7	3.8
SHR	cDCC(T)	2.3	2.6	2.3	2.2	2.8	3.8	4.2	3.9	3.9	3.9	3.9	3.8	3.8	3.5	3.7	3.7	3.8
SHR	BEKK(T)	2.4	2.8	2.7	2.6	3.2	4.2	4.6	4.5	4.5	4.6	4.6	4.6	4.6	4.5	4.7	4.6	4.7
CCC	DCC(N)	1.7	1.6	1.7	1.6	2.2	2.6	3	2.9	3	2.9	2.9	2.9	2.9	2.8	2.9	2.8	2.8
CCC	cDCC(N)	1.7	1.5	1.6	1.5	2.1	2.6	3	2.9	2.9	2.9	2.9	2.9	2.9	2.8	2.9	2.8	2.8
CCC	BEKK(N)	1.4	-1.1	0.3	0.3	0.1	-1	-1.4	-2.1	-2.3	-2.5	-2.5	-2.6	-2.6	-2.3	-2.6	-2.8	-2.9
CCC	DCC(T)	1.8	1.6	1.7	1.6	2.2	2.6	3.1	2.9	2.9	2.9	2.9	2.9	2.9	2.8	2.8	2.7	2.7
CCC	cDCC(T)	1.3	1.5	1.6	1.6	2.1	2.6	3.1	2.9	3	3	2.9	2.9	2.9	2.8	2.8	2.7	2.7
CCC	BEKK(T)	-0.1	-1.2	-0.4	-0.5	-0.9	-2	-2.6	-2.8	-2.9	-3.1	-3.1	-3.1	-3.2	-2.8	-3.1	-3.2	-3.4
DCC(N)	cDCC(N)	0.9	1	1.1	1	1.3	2	2.6	2.4	2.5	2.5	2.6	2.6	2.7	2.6	2.7	2.8	2.9
DCC(N)	BEKK(N)	-0.1	-1.5	-0.9	-1	-1.4	-1.8	-2.1	-2.4	-2.5	-2.6	-2.6	-2.7	-2.7	-2.4	-2.7	-2.8	-2.9
DCC(N)	DCC(T)	0.9	1.3	1.6	1.3	-1.7	-2.3	-2.1	-2.9	-2.9	-2.9	-2.9	-3	-3	-2.6	-2.4	-0.5	-1.1
DCC(N)	cDCC(T)	0.9	1	1.2	1.3	1.2	2.2	3.1	2.5	2.1	0.4	0.8	1.4	0.3	1.7	0.3	1.5	0.1
DCC(N)	BEKK(T)	-1.1	-1.5	-1	-1.1	-1.6	-2.3	-2.8	-2.8	-2.9	-3.1	-3.1	-3.1	-3.2	-2.9	-3.1	-3.2	-3.4
cDCC(N)	BEKK(N)	-0.3	-1.5	-0.9	-1	-1.5	-1.8	-2.1	-2.4	-2.5	-2.6	-2.6	-2.7	-2.7	-2.4	-2.7	-2.8	-2.9
cDCC(N)	DCC(T)	0	-0.6	-0.9	-0.8	-1.4	-2.1	-2.6	-2.6	-2.7	-2.7	-2.8	-2.8	-2.9	-2.8	-2.7	-1.9	-1.7
cDCC(N)	cDCC(T)	0.8	1.1	1.3	1.1	-1.3	0.4	-0.4	-1.5	-2.1	-2.3	-2.5	-2.7	-2.8	-2	-1.9	-0.1	-0.9
cDCC(N)	BEKK(T)	-1.2	-1.5	-1.1	-1.2	-1.6	-2.3	-2.8	-2.8	-2.9	-3.1	-3.1	-3.1	-3.2	-2.9	-3.1	-3.2	-3.4
BEKK(N)	DCC(T)	0.2	1.5	0.9	1	1.4	1.8	2.1	2.3	2.5	2.6	2.6	2.7	2.7	2.4	2.6	2.8	2.9
BEKK(N)	cDCC(T)	0.7	1.4	0.9	1.1	1.5	1.8	2.1	2.4	2.5	2.6	2.6	2.7	2.7	2.4	2.6	2.8	2.9
BEKK(N)	BEKK(T)	-1.5	-1.3	-1.2	-1.2	-1.5	-2.5	-3.2	-3.4	-3.5	-3.6	-3.7	-3.8	-3.8	-3.8	-4.1	-4	-4.1
DCC(T)	cDCC(T)	0.8	0.9	1.1	1.3	1.3	2.3	2.9	2.9	3	3	2.8	2.8	2.8	2.7	2.9	2.8	2.9
DCC(T)	BEKK(T)	-1.1	-1.5	-1	-1.2	-1.6	-2.3	-2.8	-2.8	-2.9	-3.1	-3.1	-3.1	-3.2	-2.8	-3.1	-3.2	-3.4
cDCC(T)	BEKK(T)	-1	-1.5	-1.1	-1.2	-1.6	-2.3	-2.8	-2.8	-3	-3.1	-3.1	-3.1	-3.2	-2.9	-3.1	-3.2	-3.4

The table reports in the first and second columns the compared models. The first row reports the cross-sectional dimension. Bold dark-grey shaded numbers identify a preference for the first model, while bold numbers identify a preference for the second model.

For both tests and all problem dimensions, SHR always provides higher losses (lower log-scores) compared with the other models, apart from EWMA. In this last case, the tests are discordant: the Diebold-Mariano test is always in favour of EWMA, while the Amisano-Giacomini test provides evidence of model equivalence for medium problem dimensions, and a preference for SHR for large problem dimensions.

The previous finding is confirmed by MCS where, at the 5% level, the SHR model is included in the confidence set only for problem dimensions up to 5 during the crisis period, and are generally excluded in 2006 (unless differently specified results for MCS are equivalent under both test statistics).

With respect to EWMA, we note a striking difference by comparing the results of the Diebold-Mariano test with those of the Amisano-Giacomini procedure. With the latter, EWMA provides larger losses than all the other models when the number of assets exceeds 10 (it is also worse than SHR when the number of assets exceeds 35), and this result holds for both evaluation periods. If we consider the Diebold-Mariano test, EWMA is always equivalent to all the other models, apart from SHR, which is outperformed for more than 5 assets, but only during the crisis period. During the 2006 evaluation sample, there is additional evidence of a preference for EWMA for medium problem dimensions. Such behavior is confirmed by the MCS results, where EWMA is excluded from the confidence set at the 1% level under the Amisano-Giacomini loss function for more than 5 assets (for simplicity, when we refer to the Amisano-Giacomini loss function, this means minus the log-scores used in the Amisano-Giacomini test). For MSE loss, the confidence set always includes EWMA.

Considering the non-naïve models, the Diebold-Mariano and Amisano-Giacomini test results are substantially similar, which leads to the following observations:

i) All models are equivalent if we consider up to 10 assets when the models are compared over extreme market conditions, while up to 15 if we compare models over the year 2006; if we consider the 2006 evaluation period, many more models are statistically equivalent, even for medium problem dimensions; when the overall market volatility is lower, more parameterized specifications would seem to be equivalent to the naïve models;

ii) Moving from the Normal to the Student density does not improve the forecast accuracy of the BEKK, DCC and cDCC specifications (whether, in sample, the Student density outperforms the Normal by information criteria; these are not reported);

iii) Scalar BEKK is always inferior to the CCC, DCC, and cDCC specifications, irrespective of the density; DCC and cDCC outperform CCC, even for smaller dimensions (from the 4 assets case); cDCC outperforms DCC under both Normal and Student densities over the extreme market condition case, while the reverse holds for the 2006 forecast evaluation period.

These findings are confirmed by the MCS approach, but only for the Amisano-Giacomini log-scores (minus). In fact, in this case and the crisis forecast range, the following results hold: the confidence set includes (at the 1% level) both DCC(N), cDCC(N), and cDCC(T) for large cross sectional dimensions (from 60 to 89); cDCC(T) is included in the confidence set only up to models with 25 assets, while DCC(T) is included only up to 10 assets. For the MSE loss function under the crisis forecast evaluation range, and at the 1% level, all models are substantially equivalent for all problem sizes. Some models are excluded only at the 5% level, in particular SHR, and BEKK(T).

If we compare models during the year 2006, results for the MSE and Amisano-Giacomini loss functions are much closer: SHR is almost always excluded from the confidence set; EWMA and BEKK(T) are excluded for medium and large problem dimensions; for large problem dimensions, the set of equivalent models includes only DCC and cDCC for the Amisano-Giacomini loss function, while it contains DCC only for the MSE loss function. In comparing directly the model forecasts, the Diebold-Mariano, Amisano-Giacomini, and MCS approaches provide interesting elements for model rankings.

The main message is the equivalence across many models when the market is in a low volatility state, and the preference for dynamic conditional correlation models. Moreover, when the market is experiencing large and sudden changes in volatility, dynamic conditional correlation models may be preferred, but the results are not consistent across all model comparison methods. Furthermore, DCC and cDCC are substantially equivalent. Table 3 reports a summary of the previous results.

Table 3: Summary of results for direct model comparisons

Forecast sample and comparison	January to December 2006	April 2008 to March 2009
DM test (MSE loss)	Most models are equivalent up to 15 asset dimension; SHR underperforms all models (EWMA excluded); *EWMA outperforms in some cases;* Student density does not improve over Normal; CCC outperforms BEKK; DCC outperforms BEKK and CCC; cDCC outperforms BEKK, CCC and DCC;	Most models are equivalent up to 10 asset dimension; SHR underperforms all models (EWMA excluded); *EWMA is equivalent to most models;* Student density does not improve over Normal; CCC outperforms BEKK; DCC outperforms BEKK and CCC; cDCC outperforms BEKK, CCC and DCC;
AG	Most models are equivalent up to 15 asset dimension; SHR underperforms all models (EWMA excluded); *EWMA underperforms with more than 10 assets;* Student density does not improve over Normal; CCC outperforms BEKK; DCC outperforms BEKK and CCC; cDCC outperforms BEKK, CCC and DCC;	Most models are equivalent up to 10 asset dimension; SHR underperforms all models (EWMA excluded); *EWMA underperforms with more than 10 assets;* Student density does not improve over Normal; CCC outperforms BEKK; DCC outperforms BEKK and CCC; cDCC outperforms BEKK, CCC and DCC;
MCS with MSE loss	*SHR is not included (at the 1% level); EWMA is always included; DCC is the best model for larger problem dimensions; Many models equivalent for medium problem dimensions;*	*SHR and BEKK(T) are included only for small problem dimensions (at the 5% level); All models are equivalent at the 1% level;*
MCS with AG loss	*SHR is included only for small problem dimensions; EWMA is included only for small problem dimensions; DCC and cDCC are the best models for larger problem dimensions; Many models are equivalent for medium problem dimensions.*	*All models are equivalent up to 5 assets; DCC(N) and cDCC(N) are the best models for larger problem dimensios.;*

Notes: The first column reports the quantities used for the direct model comparisons. In the table, AG = Amisano-Giacomini, DM = Diebold-Mariano, MSC = Model Confidence Set, and MSE = Mean Squared Error. The second and third columns report a summary of results for the two out-of-sample periods. When the density is not reported for DCC, cDCC and BEKK, the comments refer to both densities. Comments in italics identify different behavior across samples or loss functions.

4.2 Results for indirect model comparisons

As discussed in Section 3, we compare indirectly the fitted models by focusing on three portfolio allocation strategies, namely EW, GMV and GMVB. In particular, we evaluate the discrepancies between the expected and realized variances of portfolios managed using the three allocation strategies.

We start by analyzing the results obtained using the quantities suggested by Engle and Colacito (2006). The average expected portfolio variance $\bar{h}_p^m$ is a measure of model appropriateness, such that the lower is the quantity, the better is the model. We use the values of $\bar{h}_p^m$ to rank models over several cross-sectional dimensions, and report in Table 4 the average rank of the models over the problem dimension.

We note a difference with respect to the direct comparisons: DCC and cDCC specifications have high ranks, while SHR and CCC are, respectively, the first and second most preferred models over the three portfolio strategies during the crisis period. On the contrary, SHR is the worst model if we compare the results over the year 2006, while EWMA performs the best. Several checks have been made on the estimates and implementation of the models, which confirm these results. A common pattern is observed for CCC and DCC, which are associated with low ranks.

A more complete picture would be obtained by testing model accuracy. The results are even more puzzling: most models are inaccurate in forecasting portfolio variances from moderate cross-sectional dimensions, 10 assets, onward, under all portfolio strategies; few exceptions are given by EWMA under the EW and GMVB strategies, and by BEKK specifications under GMV and GMVB strategies for the crisis period; all models underestimate portfolio variances during the crisis period (an expected result); during 2006 all models overestimate variances under the EW strategy (EWMA excluded), while the results are mixed under GMV and GMVB strategies; in the last case, the overestimation of portfolio variances increases with model dimension. We may link this result to the reduced model estimation error affecting the BEKK and EWMA

specifications compared with CCC, DCC, and cDCC, given these also include univariate GARCH components (note these are multi-stage estimation approaches). Differently, the results for SHR may reveal inappropriateness of the model, where the underlying covariance is dynamic (rolling estimation methods of shrinkage target and covariance are not sufficient to quickly capture the evolution of the covariance). Therefore, SHR could be better exploited by combining it to dynamic covariance models, for instance, following Hafner and Reznikova (2010).

If we consider the Diebold-Mariano tests, the results are somewhat similar to those obtained from the direct model comparisons. First, during both the crisis and the 2006 out-of-sample periods, the SHR model underperforms all other specifications, even for small cross-sectional dimensions when we consider the QLIKE loss function (the result is slightly weaker for the GMV strategy and large problem dimensions). Such a finding is confirmed by the MSE loss function for the 2006 out-of-sample period, while during the crisis period, the underperformance of SHR is confirmed but only under the EW portfolio strategy. This is likely to be an outcome of the large increase in the variances during the 2008 financial crisis.

Table 4: Summary of results for model rankings based on average out-of-sample portfolio variances

	EWMA	SHR	CCC	DCC(N)	cDCC(N)	BEKK(N)	DCC(T)	cDCC(T)	BEKK(T)
Out-of-sample period: January to December 2006									
EW	**1.00**	9.00	6.29	**3.35**	4.41	5.71	**3.35**	5.24	6.65
GMV	**1.00**	9.00	**2.94**	**4.12**	4.35	7.18	4.35	4.82	7.24
GMVB	**1.00**	9.00	**3.12**	**3.41**	4.71	7.18	3.82	5.53	7.24
Out-of-sample period: April 2008 to March 2009									
EW	9.00	**1.00**	**3.00**	6.06	7.53	4.59	4.59	6.35	**2.88**
GMV	**2.71**	**2.41**	**2.88**	7.29	7.18	4.35	5.59	5.71	6.88
GMVB	9.00	**1.00**	**2.00**	4.88	6.29	7.71	**3.29**	4.88	5.94
Out-of-sample period: April 2008 to March 2009 (assets in reverse order)									
EW	9.00	**1.00**	**2.76**	5.59	7.06	6.00	4.53	5.88	**3.18**
GMV	4.59	**1.82**	**2.41**	5.41	6.82	7.00	**4.12**	5.53	7.29
GMVB	9.00	**1.00**	**2.00**	4.71	6.00	8.00	**3.47**	4.76	6.06

Note: The first row reports the fitted models, while the first column reports the three portfolio strategies considered. Bold values denote lower rankings (first to third), with lower values associated with better models.

In turn, this makes the estimation of portfolio weights extremely noisy, particularly affecting the results based on the MSE loss function. In support of this interpretation, we observe that the Diebold-Mariano test results for both loss functions for the GMV

and GMVB strategies during the two different out-of-sample periods behave differently. With the exclusion of SHR and EWMA for the QLIKE loss functions, all models are equivalent when compared for April 2008-March 2009, while some statistically significant differences appear when the comparison is based on 2006 (with BEKK providing statistically greater losses than the other dynamic models).

These results raise some doubts about the usefulness of dynamic covariance and correlation models of a relatively complex nature when they are used to determine portfolio weights. The joint effect of the estimation error on the model coefficients and of their amplification made by estimated portfolio weights would seem to make complex models virtually equivalent to simple models, such that increasing the complexity of a model does not improve the efficiency of an allocation strategy. On the contrary, when one of the two sources of estimation error is sterilized, by means of an EW strategy, some discrepancies seem to appear: CCC and BEKK underperform DCC and cDCC; cDCC outperforms DCC for large problem dimensions during the crisis, while the opposite holds during 2006 for the MSE loss function; the Multivariate Student-based models underperform Multivariate Normal-based models. We presume this last finding is due, in addition to the morivation made for the direct comparison, to the aggregation of many variables (the assets) into a single element (the portfolio) by a weighted average, where a central limit theorem may play a role.

Following the direct comparison, we determine the model confidence sets over several problem dimensions (see Tables 5 and 6 for two examples). For the crisis period (April 2008 to March 2009) and the EW strategy, the model confidence set includes all models at the 1% level for the MSE and QLIKE loss functions. Minor differences appear at the 5% level: for MSE, with only SHR marginally excluded; for QLIKE, the set of equivalent models includes only EWMA, cDCC(N), and DCC(N) (the last is marginally included in the confidence set).

Considering the MCS results for both test statistics and for the GMV and GMVB strategies, all models are equivalent, apart from the exclusion of EWMA at the 5% level

under the GMV allocation rule. This result confirms our earlierresults that models are equivalent when portfolio weights are estimated.

Table 5: Model confidence set – QLIKE loss function – 2006 evaluation period

	EWMA	SHR	CCC	DCC(N)	cDCC(N)	BEKK(N)	DCC(T)	cDCC(T)	BEKK(T)
Equally weighted									
2	**1**	0.003	**0.343**	**0.586**	**0.367**	**0.96**	**0.96**	**0.586**	**0.96**
3	**0.968**	0.006	**0.89**	**0.915**	**0.831**	**0.94**	**1**	**0.968**	**0.968**
4	**0.815**	0.003	**0.815**	**0.815**	**0.815**	**0.787**	**1**	**0.815**	**0.815**
5	**1**	0	**0.927**	**0.927**	**0.927**	**0.927**	**0.963**	**0.963**	**0.913**
10	**0.728**	0	**0.179**	**0.414**	**0.179**	**1**	**0.647**	**0.157**	**0.179**
15	**0.827**	0	**0.827**	**0.962**	**0.962**	**1**	**0.962**	**0.962**	**0.335**
20	**0.713**	0	**0.713**	**0.886**	**0.886**	**0.886**	**0.886**	**1**	**0.021**
25	**0.306**	0	**0.47**	**0.47**	**0.47**	**0.47**	**0.47**	**1**	0.002
30	**0.453**	0	**0.456**	**0.538**	**0.456**	**1**	**0.556**	**0.556**	0.001
35	**0.516**	0	**0.849**	**0.849**	**0.849**	**1**	**0.849**	**0.849**	0.001
40	**0.404**	0	**0.836**	**0.906**	**0.906**	**1**	**0.906**	**0.906**	0
45	**0.19**	0	**0.645**	**0.966**	**0.966**	**0.966**	**0.966**	**1**	0
50	**0.034**	0	**0.5**	**0.913**	**0.923**	**0.337**	**0.923**	**1**	0
60	**0.049**	0	**0.162**	**0.748**	**0.571**	**0.162**	**1**	**0.6**	0
70	0.002	0	**0.335**	**1**	**0.582**	**0.027**	**0.582**	**0.526**	0
80	0	0	**0.555**	**1**	**0.886**	0.001	**0.886**	**0.886**	0
89	0	0	**0.694**	**1**	**0.704**	0.001	**0.704**	**0.704**	0
Global minimum variance portfolio with short selling constraints									
2	**0.86**	**0.013**	**0.069**	**0.069**	**0.055**	**0.263**	**0.115**	**0.084**	**1**
3	**1**	0.001	**0.062**	**0.227**	**0.088**	**0.205**	**0.526**	**0.083**	**0.531**
4	**1**	0	**0.285**	**0.691**	**0.479**	**0.479**	**0.691**	**0.361**	**0.691**
5	**1**	0	**0.087**	**0.195**	**0.141**	**0.195**	**0.141**	**0.141**	**0.195**
10	**1**	0	**0.144**	**0.144**	**0.144**	**0.144**	**0.144**	**0.144**	**0.067**
15	**1**	0	**0.019**	**0.069**	**0.069**	**0.069**	**0.06**	**0.058**	0.007
20	**1**	0	**0.02**	**0.063**	**0.058**	**0.063**	**0.058**	**0.058**	0.003
25	**1**	0	**0.012**	**0.075**	**0.061**	**0.075**	**0.061**	**0.026**	0.002
30	**1**	0	0.009	**0.032**	**0.025**	**0.025**	**0.025**	**0.025**	0
35	**1**	0	0.006	**0.033**	**0.028**	0.006	**0.028**	**0.023**	0
40	**1**	0	0.007	**0.046**	**0.036**	0.006	**0.036**	**0.027**	0
45	**1**	0	0.009	**0.046**	**0.039**	0.008	**0.039**	**0.027**	0
50	**1**	0	**0.015**	**0.085**	**0.066**	0.007	**0.066**	**0.053**	0
60	**1**	0	**0.011**	**0.16**	**0.08**	0.005	**0.16**	**0.057**	0.001
70	**1**	0	0.008	**0.179**	**0.079**	0.001	**0.179**	**0.091**	0
80	**1**	0	0.003	**0.14**	**0.074**	0.001	**0.14**	**0.078**	0
89	**1**	0	0.004	**0.115**	**0.115**	0.001	**0.115**	**0.111**	0

The first row reports the fitted models while the first column the cross-sectional dimension. The two panels report the p-values for the construction of the model confidence set for the statistic R for the equally weighted strategy and the global minimum variance without short selling strategy. Bold dark grey shaded number over rows denote models included in the confidence set at the 5% confidence level. Additional models included in the confidence set at the 1% confidence level are identified by bold numbers.

Table 6: Model confidence set – QLIKE loss function – crisis period

	EWMA	SHR	CCC	DCC(N)	cDCC(N)	BEKK(N)	DCC(T)	cDCC(T)	BEKK(T)
Equally weighted									
2	0.757	0.132	0.35	0.696	0.757	0.757	0.757	1	0.558
3	1	0.082	0.194	0.287	0.452	0.249	0.376	0.613	0.258
4	0.232	0.052	0.121	0.232	0.439	0.211	0.232	1	0.121
5	0.273	0.071	0.153	0.273	0.44	0.226	0.273	1	0.147
10	1	0.054	0.118	0.94	0.995	0.29	0.368	0.736	0.227
15	1	0.027	0.063	0.317	0.55	0.141	0.144	0.55	0.063
20	1	0.016	0.025	0.041	0.283	0.034	0.034	0.283	0.025
25	1	0.016	0.023	0.062	0.19	0.03	0.03	0.187	0.023
30	1	0.016	0.024	0.056	0.126	0.024	0.024	0.065	0.024
35	1	0.022	0.032	0.06	0.151	0.033	0.033	0.055	0.032
40	1	0.02	0.027	0.064	0.138	0.027	0.027	0.027	0.027
45	1	0.021	0.032	0.059	0.139	0.032	0.032	0.035	0.032
50	1	0.022	0.03	0.051	0.112	0.03	0.03	0.03	0.03
60	1	0.021	0.028	0.039	0.1	0.028	0.028	0.039	0.028
70	1	0.016	0.021	0.035	0.094	0.021	0.021	0.021	0.021
80	1	0.025	0.036	0.052	0.097	0.036	0.036	0.052	0.036
89	1	0.028	0.038	0.044	0.082	0.038	0.038	0.044	0.037
Global minimum variance without short selling									
2	0.915	0.102	0.236	0.416	0.516	1	0.599	0.915	0.774
3	1	0.177	0.579	0.947	0.959	0.959	0.959	0.959	0.928
4	1	0.1	0.308	0.885	0.885	0.885	0.846	0.885	0.795
5	0.522	0.128	0.333	0.634	0.634	0.493	0.634	1	0.292
10	1	0.139	0.242	0.976	0.976	0.494	0.494	0.494	0.494
15	1	0.087	0.137	0.407	0.407	0.407	0.293	0.407	0.293
20	1	0.066	0.11	0.308	0.308	0.308	0.229	0.308	0.229
25	1	0.045	0.053	0.183	0.183	0.136	0.136	0.136	0.136
30	1	0.069	0.089	0.379	0.43	0.241	0.241	0.241	0.241
35	1	0.072	0.084	0.545	0.62	0.222	0.222	0.222	0.222
40	1	0.061	0.073	0.443	0.534	0.217	0.217	0.217	0.217
45	1	0.074	0.101	0.586	0.586	0.236	0.236	0.236	0.231
50	1	0.047	0.062	0.467	0.525	0.174	0.174	0.222	0.061
60	1	0.054	0.066	0.335	0.335	0.178	0.178	0.335	0.066
70	1	0.056	0.07	0.23	0.23	0.123	0.123	0.23	0.07
80	1	0.068	0.085	0.219	0.219	0.172	0.172	0.219	0.085
89	1	0.064	0.082	0.226	0.226	0.155	0.155	0.226	0.082

The first row reports the fitted models while the first column the cross-sectional dimension. The two panels report the p-values for the construction of the model confidence set for the statistic R for the equally weighted strategy and the global minimum variance without short selling strategy. Bold dark grey shaded number over rows denotes models included in the confidence set at the 5% confidence level. Additional models included in the confidence set at the 1% confidence level are identified by bold numbers.

We then analyze the results for the 2006 out-of-sample period. First, if we focus on both the GMV and GMVB strategies, we note the following: SHR is always excluded from the confidence set, apart for very small problem dimensions under MSE loss; EW and BEKK(T) are excluded from the confidence set for large problem dimensions; all other

models are equivalent. Overall, the results confirm the previous findings that many models are equivalent for estimated portfolio weights. However, we presume that the greater flexibility in separately estimating variances and correlations provides some benefits over naïve and general covariance specifications, in particular, when the problem dimension is large.

If we consider the EW portfolio strategy, the results are slightly different; first, we have a confirmation of the exclusion of the SHR and BEKK specifications; second, for medium problem dimensions, EWMA seems to be the best model (at the 5% level), while for large problem dimensions, the confidence set also includes DCC and cDCC; and CCC is generally excluded from the confidence set for medium and large problem dimensions.

We interpret the results as follows: for small and medium problem dimensions, the estimation error has a relevant role, making more complex covariance and correlation models almost equivalent, if not worse, than EWMA; when the number of variables is increased, the flexibility of DCC and cDCC models becomes even more relevant than the estimation error.

In summary (see also Table 7), the indirect comparisons suggest that model performances are affected by several sources of error. The estimation error of model parameters is always present. Estimation error of portfolio weights may play a relevant role during extreme market conditions, and be so relevant as to make many models statistically equivalent in terms of forecasts. Finally, the choice of the covariance proxy may also have a role, and may affect the results. However, since we compared models using the same proxy, namely the returns cross-product (which was the only one available), it has the same effect on all the results. However, additional studies evaluating the impact of the proxy could shed additional light on cases where many models have been labelled as equivalent (for instance, the model confidence set for the EW strategy).

Table 7: Summary of results of indirect model comparisons based on Diebold-Mariano and Model Confidence Set

Forecast sample and comparison	January to December 2006	April 2008 to March 2009
DM test (MSE loss)	SHR underperforms all models; *BEKK underperforms under GMV and GMVB;* DCC and cDCC outperform BEKK and CCC under EW; *DCC outperforms cDCC under EW;* Student density does not improve over Normal under EW;	*SHR underperforms all models under EW;* Under GMV and GMVB all models are equivalent; DCC and cDCC outperform BEKK and CCC under EW; cDCC outperforms DCC under EW; Student density does not improve over Normal under EW;
DM test (QLIKE loss)	SHR underperforms all models; EWMA underperforms under GMV; *BEKK models underperform under GMVB;* DCC and cDCC outperform BEKK and CCC under EW; *DCC outperforms cDCC under EW;* Student density does not improve over Normal under EW;	SHR underperforms all models; EWMA underperforms under GMV; *Most models are equivalent under GMV and GMVB;* DCC and cDCC outperform BEKK and CCC under EW; cDCC outperforms DCC under EW; Student density does not improve over Normal under EW;
MCS (MSE loss)	*SHR is always excluded; EWMA and BEKK(T) are excluded for large problem dimensions under GMV and GMVB; BEKK and CCC are excluded under EW; EWMA, DCC and cDCC are included for large problem dimensions;*	*All models are equivalent at the 1% level; At the 5% level under EW, SHR is excluded for large problem dimensions;*
MCS (QLIKE loss)	*SHR is always excluded; EWMA and BEKK(T) are excluded for large problem dimensions under GMV and GMVB; BEKK and CCC are excluded under EW; EWMA, DCC and cDCC are included for large problem dimensions.*	*All models are equivalent at the 1% level; At the 5% level under EW, the set includes only EWMA, DCC(N) and cDCC(N).*

Notes: The first column reports the quantities used for the indirect model comparisons. In the table, DM = Diebold-Mariano, MSC = Model Confidence Set, MSE and QLIKE denote the two loss functions, while EW, GMV and GMVB identify the portfolio strategies considered. The second and third columns report a summary of results for the two out-of-sample periods. When the density is not reported for DCC, cDCC and BEKK, the comments refer to both densities. If portfolio strategies are not reported, the comments apply to all strategies. Comments in italics identify different behavior across samples or loss functions.

5. Concluding Remarks

From the empirical point of view, Multivariate GARCH models suffer seriously from the so-called curse of dimensionality. For this reason, several simple specifications are typically used, including the CCC, DCC and Scalar BEKK models. Alternatively, naïve methods could be used, such as EWMA or the Covariance Shrinking approach. However, few studies have considered a detailed out-of-sample comparison of these models. This paper has shed light on this topic, but the outcome is far from conclusive. By using alternative evaluation methods, including the direct and indirect approaches, pairwise and multivariate methodologies, and different out-of-sample evaluation periods, the results are mixed. The only common finding is that covariance shrinking methods underperform the dynamic models, even for small cross-sectional dimensions, at least for the dataset and periods considered in the paper. Less common outcomes suggest, for small problem dimensions, there is a higher probability that alternative approaches will provide substantially equivalent covariance forecasts.

This finding is less evident for large problem dimensions, where simple dynamic specifications, despite being highly restrictive, may be superior to naïve specifications based on calibrated coefficients. In this case, models separately capturing the variance and correlation dynamics are marginally preferred to pure covariance models. Furthermore, the impact of several sources of error, such as estimation error of the model parameters, possibly amplified by estimated portfolio weights for indirect comparison, and errors associated with the choice of proxy, come into play and can affect the outcomes, thereby suggesting the need for further analysis. Overall, we do not find confirmation of the result of Zumabch (2009), which suggested a preference for covariance models. Furthermore, we provide evidence that naïve allocation strategies are less affected by estimation error of covariance models; this is a result similar to that obtained by Uppal et al. (2009) in a different framework.

Finally, it should be emphasized that the main message from this analysis of the paper is that there is no optimal model. The best model must be chosen with respect to a sample period and by using selection criteria that match the purpose of the analysis. It is clear

that direct and indirect evaluations can provide markedly different results. This may be read as further confirmation of the widely held view that "all models are wrong, but some are useful", wherein usefulness may change over time and for different applications.

Additional research on the topic is needed, and should focus on the methodological approaches for model comparison, on the rubustness of model ranking over different forecast horizons (longer that the one-day horizon we uses), and for even larger problem dimensions.

References

Aielli, 2008, Consistent estimation of large scale dynamic conditional correlations, Working paper n. 47, Department of Economics, Statistics, Mathematics and Sociology, University of Messina.

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

Asai, M., Caporin, M., and McAleer, M., 2009, Block structure multivariate stochastic volatility, Available at SSRN: http://ssrn.com/abstract=1524667.

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

Bauwens, L., Laurent, S., and Rombouts, J.V.K., 2006, Multivariate GARCH models: A survey, Journal of Applied Econometrics, 21, 79-109.

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

Bollerslev T., 1990, Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH approach, Review of Economic and Statistics, 72, 498-505.

Bonato, M., Caporin, M., and Ranaldo, A., 2009, Forecasting realized covariances with a Block structure WAR model, Swiss National Bank Working Paper 2009-03.

Caporin, M. And McAleer, M., 2008, Scalar BEKK and indirect DCC, Journal of Forecasting, 27-6, 537-549.

Caporin, M. And McAleer, M., 2009, Do we really need both BEKK and DCC? A tale of two covariance models, Available at SSRN: http://ssrn.com/abstract=1338190.

Caporin, M. And McAleer, M., 2010, Do we really need both BEKK and DCC? A tale of two multivariate GARCH models, Available at SSRN: http://ssrn.com/abstract=1549167.

Caporin, M., and Paruolo, P., 2009, Structured multivariate volatility models, Available at SSRN: http://ssrn.com/abstract=1318639.

Cappiello, L., Engle, R.F., and Sheppard, K., 2006, Asymmetric Dynamics in the Correlations of Global Equity and Bond Returns, Journal of Financial Econometrics, 4, 537-572.

Chib, S., Omori, Y., and Asai, M., 2009, Multivariate Stochastic Volatility, in Andersen, T.G., Davis, R.A., Kreiß, J.P., and Mikosch, T. (eds.), Handbook of Financial Time Series, Springer.

Clements, A., Doolan, M., Hurn, S., and Becker, M., 2009, On the efficacy of techniques for evaluating multivariate volatility forecasts, NCER working paper series.

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

De Miguel, V., Garlappi, L., and Uppal, R., 2009, Optimal versus naïve diversification: how inefficient is the 1/N portfolio strategy?, Review of Financial Studies, 22, 1915-1953.

Diebold, F.X. and Mariano, R.S., 1995, Comparing predictive accuracy, Journal of Business and Economic Statistics, 13-3, 253-263.

Ding, Z. and Engle, R., 2001, Large scale conditional covariance modelling, estimation and testing, Academia Economic Papers, 29, 157-184.

Engle, R.F., 2002, Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models, Journal of Business and Economic Statistics, 20, 339-350.

Engle, R.F., and Colacito, R., 2006, Testing and valuing dynamic correlations for asset al location, Journal of Business and Economic Statistics, 24, 238-253.

Engle, R.F., and Kelly, B., 2008, Dynamic Equicorrelation, New York University Working Paper FIN-08-038.

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

Engle, R.F., and Sheppard, K., 2001, Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH, Working Paper 2001-15, University of California at San Diego.

Engle, R.F., and Sheppard, K., 2008, Evaluating the specification of covariance models for large portfolios, available at www.kevinsheppard.net.

Engle, R.F., Shephard, N., and Sheppard, K., 2008, Fitting vast dimensional time-varying covariance models, Oxford Financial Research Centre, Financial Economics Working Paper n. 30.

Fan, Y., Pastorello, S., and Renault, E., 2007, Maximization by parts in Extremum Estimation, Mimeo, University of North Carolina in Chapel Hill.

Ferreira, M.A. and Lopez, J.A., 2005, Evaluating interest rate covariance models within a value-at-risk framework, Journal of Financial Econometrics, 3, 126-168.

Franses, P.H., and Hafner, C.M., 2009, A Generalized Dynamic Conditional Correlation Model: Simulation and Application to Many Assets, Econometric Reviews, 28, 612-631.

Hafner, C.M., and Herwartz, H., 2008, Analytical quasi maximum likelihood inference in multivariate volatility models, Metrika, 67, 219-239.

Hafner, C.M., and Preminger, A., 2009, On asymptotic theory for multivariate GARCH models, Journal of Multivariate Analysis, 100, 2044-2054.

Hafner, C.M, and Reznikova, O., 2010, On the estimation of dynamic conditional correlation models, Université Catholique de Louvain, Institute of Statistics, Discussion Paper DS1006.

Hansen, P.R., 2005, A test for superior predictive ability, Journal of Business and Economic Statistics, 23-4, 365-380.

Hansen, P.R., Lunde, A. and Nason, J.M., 2005, Model confidence sets for forecasting models, Federal Reserve Bank of Atlanta Working Paper 2005-7.

Jeantheau, T., 1998, Strong consistency of estimators for multivariate ARCH models, Econometric Theory, 14, 70-86.

Laurent, S., Rombouts, J.V.K., and Violante, F., 2009, On loss functions and ranking forecasting performances of multivariate GARCH models, CIRANO working paper 2009s-45.

Laurent, S., Rombouts, J.V.K., and Violante, F., 2010, On the forecasting accuracy of multivariate GARCH models, unpublished manuscript.

Ledoit, O., and Wolf, M., 2003, Improved estimation of the covariance matrix of stock returns with an application to portfolio selection, Journal of Empirical Finance, 10, 603-621.

Ledoit, O., and Wolf, M., 2004, Honey, I shrunk the sample covariance matrix, Journal of Portfolio Management, Summer 2004, 110-119.

Ling, S. and McAleer, M., 2003, Asymptotic theory for a vector ARMA-GARCH model, Econometric Theory, 19, 278-308.

McAleer, M., 2005, Automated inference and learning in modeling finanical volatility, Econometric Theory, 21,232-261.

McAleer, M., Chan, F., Hoti, S., and Lieberman, O., 2008, Generalized autoregressive conditional correlation, Econometric Theory, 24-6, 1554-1583.

McAleer M. and M. Medeiros, 2008, Realized volatility: A review, Econometric Reviews, 27, 10-45.

Mincer, J., and Zarnowitz, V., 1969, The evaluation of economic forecasts. In: Mincer J (ed) Economic Forecasts and Expectations, Columbia University Press.

Patton, A.J., 2010, Volatility forecast comparison using imperfect volatility proxies, Journal of Econometrics, forthcoming.

Patton, A.J., and Sheppard, K., 2009, Evaluating volatility and correlation forecasts, in Andersen, T.G., Davis, R.A., Kreiß, J.P., and Mikosch, T., (eds.), Handbook of Financial Time Series, Springer.

Silvennoinen, A., and Terasvirta, T., 2009, Multivariate GARCH models, in Andersen, T.G., Davis, R.A., Kreiß, J.P., and Mikosch, T. (eds.), Handbook of Financial Time Series, Springer.

West, K.D., 1996, Asymptotic inference about predictive ability, Econometrica, 64, 1067-1084.

West, K.D., 2006, Forecast evaluation, In: Elliott G, Granger C, Timmermann A (eds) Handbook of Economic Forecasting, North Holland Press, Amsterdam.

White, H., 2000, A reality check for data snooping, Econometrica, 68-5, 1097-1126

Zumbach, G., 2009, The empirical properties of large covariance matrices, RiskMetrics Journal, 9-1 Winter 2009, 31-54.

Appendix A: List of equities included in the empirical analysis

The following list contains the names of the 89 companies whose stock total returns have been used in the empirical analysis of the paper.

3M, ABBOTT LABORATORIES, ALCOA, ALLSTATE, ALTRIA GROUP, AMER.ELEC.PWR., AMERICAN EXPRESS, AMGEN, APPLE, AT&T, AVON PRODUCTS, BAKER HUGHES, BANK OF AMERICA, BANK OF NEW YORK MELLON, BAXTER INTL., BOEING, BRISTOL MYERS SQUIBB, BURL.NTHN.SANTA FE C, CAMPBELL SOUP, CAPITAL ONE FINL., CATERPILLAR, CHEVRON, CISCO SYSTEMS, CITIGROUP, COCA COLA, COLGATE-PALM., COMCAST 'A', CONOCOPHILLIPS, COSTCO WHOLESALE, CVS CAREMARK, DELL, DOW CHEMICAL, E I DU PONT DE NEMOURS, EMC, ENTERGY, EXELON, EXXON MOBIL, FEDEX, FORD MOTOR, GENERAL DYNAMICS, GENERAL ELECTRIC, GILEAD SCIENCES, HALLIBURTON, HEWLETT-PACKARD, HJ HEINZ, HOME DEPOT, HONEYWELL INTL., INTEL, INTERNATIONAL BUS.MCHS., JOHNSON & JOHNSON, JP MORGAN CHASE & CO., LOCKHEED MARTIN, LOWE'S COMPANIES, MCDONALDS, MEDTRONIC, MERCK & CO., MICROSOFT, MORGAN STANLEY, NATIONAL OILWELL VARCO, NIKE 'B', NORFOLK SOUTHERN, OCCIDENTAL PTL., ORACLE, PEPSICO, PFIZER, PROCTER & GAMBLE, QUALCOMM, RAYTHEON 'B', REGIONS FINL.NEW, SARA LEE, SCHERING-PLOUGH, SCHLUMBERGER, SOUTHERN, SPRINT NEXTEL, TARGET, TEXAS INSTS., TIME WARNER, UNITED TECHNOLOGIES, UNITEDHEALTH GP., US BANCORP, VERIZON COMMUNICATIONS, WAL MART STORES, WALGREEN, WALT DISNEY, WELLS FARGO & CO, WEYERHAEUSER, WILLIAMS COS., WYETH, XEROX