

# Comparing Multivariate GARCH Models by Problem Dimension

Massimiliano Caporin and Michael McAleer

**Abstract** In the last 15 years, several Multivariate GARCH (MGARCH) models have appeared in the literature. The two most widely known and used are the Scalar BEKK model and the DCC model. 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 MGARCH models, namely BEKK, DCC, Corrected DCC (cDCC), CCC of Bollerslev, Exponentially Weighted Moving Average, and covariance shrinking, using the historical data of 89 US equities. Our methods 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. 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, model comparison, MGARCH, time series analysis.

## 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

---

<sup>1</sup> Massimiliano Caporin, University of Padova, Department of Economics and Management “Marco Fanno”; email: massimiliano.caporin@unipd.it.

Michael McAleer, Erasmus University Rotterdam; email: michael.mcaleer@gmail.com

“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, 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).

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 Sheppard (2008), Clements et al. (2009), and Patton and Sheppard (2009)). These papers present limited comparisons across a small range of models.

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.

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.

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).

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 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.

## 2 Feasible covariance and correlation models for large problem dimensions

The conditional density of a  $k$ -dimensional vector of financial variables (returns) in deviation from their mean follows:

$$x_t - \mu_t = \varepsilon_t \mid I^{t-1} \sim D(0, \Sigma_t) \quad (1)$$

where  $I^{t-1}$  is the information set at time  $t-1$ ,  $D(\cdot)$  denotes a multivariate density with time-varying covariance matrix. We do not consider the effects of different mean specifications and assume the mean is fixed at the sample mean for a range of alternative covariance models. 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.

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):

$$\Sigma_t = \bar{\Sigma} + \alpha \left( \varepsilon_{t-1} \varepsilon_{t-1}' - \bar{\Sigma} \right) + \beta \left( \Sigma_{t-1} - \bar{\Sigma} \right). \quad (2)$$

where the intercept is equal to the unconditional covariance matrix  $\bar{\Sigma} = E[\varepsilon_t \varepsilon_t']$ .

Scalar BEKK in (2) 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.

We then consider 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  $\Sigma_t = D_t R D_t$  where  $D_t$  is a diagonal matrix of conditional standard deviations, and  $R$  is an unconditional correlation matrix. 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.

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 DCC model of Engle (2002) was proposed as a generalization of the CCC model, and assumes a time-varying conditional correlation in  $\Sigma_t = D_t R_t D_t$  following

$$R_t = \bar{Q}_t^{-1} Q_t \bar{Q}_t^{-1}, \quad \bar{Q}_t = (Q_t \odot I)^{1/2} \quad (3)$$

$$Q_t = S + \alpha \left( D_{t-1}^{-1} \varepsilon_{t-1} \varepsilon_{t-1}' D_{t-1}^{-1} - S \right) + \beta (Q_{t-1} - S) \quad (4)$$

where  $D_t$  is the same as for the CCC model,  $S$  is the unconditional correlation matrix, and  $\alpha$  and  $\beta$  are the scalar parameters driving the model dynamics. Following Engle (2002), the model is estimated with a multi-step approach that clearly reduces the

efficiency but makes the model feasible with large cross-sectional dimensions. Note that the model in (3)-(4) includes targeting, as defined by Caporin and McAleer (2010). Aielli (2008) shows that the estimator of  $S$  used in the multi-step estimation of the DCC is inconsistent. In order to resolve this serious issue, Aielli (2008) introduces the cDCC model, which replaces (4) with

$$Q_t = S + \alpha \left( \bar{Q}_{t-1} D_{t-1}^{-1} \varepsilon_{t-1} \varepsilon_{t-1}' D_{t-1}^{-1} \bar{Q}_{t-1} - S \right) + \beta (Q_{t-1} - S) \quad (5)$$

where the parameter matrix  $S$  is symmetric, has unit elements over the main diagonal, and is now the covariance matrix of the innovations  $\bar{Q}_t D_t^{-1} \varepsilon_t$ , which are not observable. The modification restores consistency but again exposes the model to the curse of dimensionality as the matrix  $S$  in (5) has to be estimated. Aielli (2008) also suggests a feasible estimation method for large problem dimensions.

The last two models considered are the Exponentially Weighted Moving Average model and the Covariance Shrinking approach of Ledoit and Wolf (2003, 2004). The EWMA model provides a recursion for the evaluation of the conditional covariance matrix, which is based on a single parameter  $\lambda$  which we fix at 0.94:

$$\Sigma_t = (1 - \lambda) \varepsilon_{t-1} \varepsilon_{t-1}' + \lambda \Sigma_{t-1} \quad (6)$$

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. Following the covariance shrinking approach, we define the expected covariance for time  $t$  as follows:

$$\Sigma_t = (1 - \lambda) S_{t-1} + \lambda F_{t-1} \quad (7)$$

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  $\lambda$ , 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

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  $T+h$ . Models are estimated with a rolling approach and, 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 ( $m=1,2,\dots,M$ ).

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:

$$l_{m,T+i} = \frac{1}{k^2} \mathbf{1}_{k^2}' \left( \text{vec} \left( \hat{\Sigma}_{T+i}^m - e_{T+i} e_{T+i}' \right) \text{vec} \left( \hat{\Sigma}_{T+i}^m - e_{T+i} e_{T+i}' \right) \right) \quad (8)$$

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 test of equal predictive ability corresponds to checking the null hypothesis of zero loss function differentials,  $H_0 : E[\bar{l}_j - \bar{l}_i] = E[\bar{LF}_{jl}] = 0$ , where  $i$  and  $j$  are two

different model indices,  $\bar{l}_j = \frac{1}{h} \sum_{i=1}^h l_{m,T+i}$ , and  $\bar{LF}_{jl} = \bar{l}_j - \bar{l}_i$ . The test statistic is given as  $t_{jl,h} = \sqrt{h} \bar{LF}_{jl} \sigma^{-1}(\sqrt{h} \bar{LF}_{jl}) \sim N(0,1)$  where  $\sigma(\sqrt{h} \bar{LF}_{jl})$  is the heteroskedasticity and autocorrelation (HAC) consistent estimate of the asymptotic variance of  $\sqrt{h} \bar{LF}_{jl}$ .

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:

$$\log f_{m,T+i} = \log \left| \hat{\Sigma}_{T+i}^m \right| + e_{T+i}' \left( \hat{\Sigma}_{T+i}^m \right)^{-1} e_{T+i}. \quad (9)$$

Amisano and Giacomini (2007) then consider the quantity  $L_{jl,T+i} = w(e_{T+i}) [\log f_{j,T+i} - \log f_{l,T+i}]$  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  $H_0 : E[\bar{L}_{jl,T,h}] = 0$ , where  $\bar{L}_{jl,T,h} = \frac{1}{h} \sum_{i=1}^h L_{jl,T+i}$ , while the alternative hypothesis refers to different predictive ability. The test statistic is given as:

$t_{jl,h} = \bar{L}_{jl,T,h} \sqrt{h} \sigma^{-1}(\bar{L}_{jl,T,h}) \sim N(0,1)$  where  $\sigma(\bar{L}_{jl,T,h})$  is the heteroskedasticity and autocorrelation consistent estimate of the asymptotic variance  $\text{Var}[\sqrt{h} \bar{L}_{jl,T,h}]$ . 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  $j$ ). Within the direct model evaluation framework, we consider equal weights for all points over the forecast horizon. Positive values of the test statistic 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 model 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

$$H_0 : E[\overline{lf}_{jl} - \overline{lf}_{il}] = E[\overline{LF}_{jl}] = 0, \quad j > i, \forall j, i \in \mathcal{M}_\ell.$$

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:  $t_{R,h} = \max_{j,l \in \mathcal{M}_\ell} \left| \overline{LF}_{jl} s^{-1}(\overline{LF}_{jl}) \right|$  and  $t_{SQ,h} = \sum_{j,l \in \mathcal{M}_\ell, j > l} \left( \overline{LF}_{jl} s^{-1}(\overline{LF}_{jl}) \right)^2$

where  $s(\overline{LF}_{jl})$  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. Such a model is identified as  $j = \arg \max_{j \in \mathcal{M}_\ell} \left( \sum_{l \in \mathcal{M}_\ell} \overline{LF}_{jl} \right) \left( \text{Var} \left( \sum_{l \in \mathcal{M}_\ell} \overline{LF}_{jl} \right) \right)^{-1/2}$

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, as well as of the Amisano-Giacomini log-scores (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.

Given the portfolio mean and variance forecasts (the last are model dependent), and the realized portfolio returns, we compare models using the Amisano-Giacomini test for the EW strategy (GMV and GMVB have been excluded to avoid distortions in the test statistic due to the estimation error of portfolio weights). Furthermore, we compare models using both the MSE loss function as well as the QLIKE loss function defined in Patton and Sheppard (2009) within a Diebold-Mariano framework as well as for the Model Confidence Set.

## 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.

On these data we fit the models defined in Section 2, using both Multivariate Normal and Multivariate Student densities for BEKK, DCC and cDCC. The 9 specifications considered 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.

All the empirical results are reported in a set of tables and graphs in the following sections. Tables are summarized and reported in the paper, while the entire set of empirical results is available from the authors upon request.

## 5 Results for direct model comparison

A summary of the results is reported in Table 1. The Diebold-Mariano and Amisano-Giacomini tests point out the negative performance of the covariance shrinking approach over both evaluation periods. 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. These findings are confirmed by MCS. 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, and this result holds for both evaluation periods. If we consider the MSE loss function, 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. 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;
- ii) Moving from the Normal to the Student density does not improve the forecast accuracy of the BEKK, DCC and cDCC specifications;
- 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 results are confirmed by the MCS approach, but only for the Amisano-Giacomini loss function. 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.

If we compare models during the year 2006, results for the MSE and Amisano-Giacomini loss functions are much closer.

The main message we extrapolate from our results 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.

## 6 Results for indirect model comparison

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. The crisis makes the estimation of portfolio weights extremely noisy, particularly affecting the results based on the MSE loss function. 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.

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 the estimation error of 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 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.

Following the direct comparison, we determine the model confidence sets over several problem dimensions. 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). 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 earlier results that models are equivalent when portfolio weights are estimated.

We then analyze the results for the 2006 out-of-sample period. 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. Furthermore, 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, 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.

## 6 Conclusions

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, estimation error of 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.

## 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>.

- 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 Rinaldo, 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>.
- 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 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., 2008, Evaluating the specification of covariance models for large portfolios, available at [www.kevinsheppard.net](http://www.kevinsheppard.net).
- Fan, Y., Pastorello, S., and Renault, E., 2007, Maximization by parts in Extremum Estimation, Mimeo, University of North Carolina in Chapel Hill.
- 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.
- 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.
- 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.
- 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., 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.
- 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

Table 1: Summary of results for direct model comparisons

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

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.

Table 2: Summary of results of indirect model comparisons based on Diebold-Mariano and Model Confidence Set

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

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.