Centre Interuniversitaire sur le Risque, les Politiques Économiques et l'Emploi

Cahier de recherche/Working Paper **10-21**

On the Forecasting Accuracy of Multivariate GARCH Models

Sébastien Laurent
Jeroen V.K. Rombouts
Francesco Violante

Mai/May 2010

Laurent: Maastricht University, The Netherlands; Université Catholique de Louvain, CORE, B-1348, Louvain-la-Neuve, Belgium
Rombouts: HEC Montréal, CIRANO, CIRPÉE and Université Catholique de Louvain (CORE).
Corresponding author. HEC Montréal, 3000 Chemin de la Côte-Ste-Catherine, Québec, Canada H3T 2A7
Tel. : +1 514 340-6466; Fax: +1 514 340-6469
jeroen.rombouts@hec.ca
Violante: Université de Namur, CeReFim. B-5000, Namur, Belgique; Université Catholique de Louvain, CORE, B-1348, Louvain-la-Neuve, Belgium

Financial support from CIRPÉE and the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State Prime Minister's Office, science policy programming, is gratefully acknowledged. We thank Lars Stentoft, the participants of the IFM2 Mathematical finance days and CIREQ 4th times series conference for their comments.

Abstract:
This paper addresses the question of the selection of multivariate GARCH models in terms of variance matrix forecasting accuracy with a particular focus on relatively large scale problems. We consider 10 assets from NYSE and NASDAQ and compare 125 model based one-step-ahead conditional variance forecasts over a period of 10 years using the model confidence set (MCS) and the Superior Predictive Ability (SPA) tests. Model performances are evaluated using four statistical loss functions which account for different types and degrees of asymmetry with respect to over/under predictions. When considering the full sample, MCS results are strongly driven by short periods of high market instability during which multivariate GARCH models appear to be inaccurate. Over relatively unstable periods, i.e. dot-com bubble, the set of superior models is composed of more sophisticated specifications such as orthogonal and dynamic conditional correlation (DCC), both with leverage effect in the conditional variances. However, unlike the DCC models, our results show that the orthogonal specifications tend to underestimate the conditional variance. Over calm periods, a simple assumption like constant conditional correlation and symmetry in the conditional variances cannot be rejected. Finally, during the 2007-2008 financial crisis, accounting for non-stationarity in the conditional variance process generates superior forecasts. The SPA test suggests that, independently from the period, the best models do not provide significantly better forecasts than the DCC model of Engle (2002) with leverage in the conditional variances of the returns.

Keywords: Variance matrix, forecasting, multivariate GARCH, loss function, model confidence set, superior predictive ability

JEL Classification: C10, C32, C51, C52, C53, G10

1 Introduction

Most financial applications are multivariate problems with volatility forecasts as one of the inputs. Forecasting sequences of variance matrices is relatively easily done using a multivariate GARCH model, i.e. the conditional variance matrix is modelled as a function of past returns. A large number of multivariate GARCH models have been proposed in the literature, see Bauwens, Laurent, and Rombouts (2006) and Silvennoinen and Terasvirta (2009) for extensive surveys. The first generation of models, for example the VEC model of Bollerslev, Engle, and Wooldridge (1988) and the BEKK model of Engle and Kroner (1995), are direct extensions of the univariate GARCH model of Bollerslev (1986). These models are very general and allow for rich and flexible dynamics for the conditional variance matrix. They have been extensively used to model volatility spillovers and in applications such as conditional CAPM and futures hedging. Examples are respectively Karolyi (1995) and Bali (2008). However, being heavily parameterized, they are tractable only for a small number of series, typically lower than four.

More recently, the focus has turned to larger scale problems such as dynamics of correlations between equity and bond returns, portfolio selection and Value at Risk, see Engle (2009) for examples. In these applications, the numerical evaluation of first generation models becomes unfeasible. Both the number of parameters and the number of operations required to evaluate the likelihood function tend to explode rapidly with the number of series. Alternative approaches for achieving more manageable and parsimonious specifications have been proposed. Feasible specifications can be obtained by imposing strong parameter restrictions on the BEKK model, e.g. the scalar BEKK model and the exponentially weighted moving average model proposed by J.P.Morgan (1996). Similarly, factor structures like in Engle and Gonzalez-Rivera (1991), the orthogonal models of Alexander and Chibumba (1997), van der Weide (2002) and Fan, Wang, and Yao (2008) have been proposed. Recently, increasing attention has been devoted to conditional correlation models because they can be estimated using a multi-step procedure. The first models have been introduced by Engle (2002) and Tse and Tsui (2002). Extensions of Engle (2002) are the asymmetric conditional correlation model of Cappiello, Engle, and Sheppard (2006), the consistent DCC of Aielli (2006) and the sequential conditional correlation model of Palandri (2009).

A priori it is difficult, if not impossible, to identify which model has the best out-of-sample forecasting performance. The evaluation of univariate volatility forecasts is well understood, see Hansen and Lunde (2005), Hansen, Lunde, and Nason (2003), Becker and Clements (2008) among others. However, although many multivariate GARCH models are available, from an applied viewpoint, there are no clear guidelines available on model evaluation and selection. Recent somewhat related studies include Clements, Doolan, Hurn, and Becker (2009) and Chiriac and Voev (2010). Though, their analysis usually involves a small number of alternative parametrizations and/or small cross sectional dimensions.

This paper addresses the selection of multivariate GARCH models in terms of conditional variance matrix out-of-sample forecasting accuracy with a focus on large scale problems. Another major innovation is that our comparison is based on large sets of competing model specifications. We first estimate a large variety of models and produce a set of out-of-sample model based forecasts. This can be easily done using standard econometric software packages which are today readily available to the forecaster. Second, we identify the set of models that show superior forecasting performance. These models can then be used either to produce combined forecasts or to select a particular preferred model.

Several approaches have been proposed with respect to the inference on the set of superior models. The testing procedure based on equal predictive ability (EPA) introduced by Diebold and Mariano (1995) to account for parameter uncertainty, allows for pair wise comparison of forecast performances across models. Important generalizations can be found West (1996), Clark and McCracken (2001), Clark and West (2006) and Clark and West (2007). See West (2009) for a survey. Giacomini and White (2006) develop a framework that allows the comparison between model based forecasts taking also into account the estimation method, estimation uncertainty, model misspecification and the choice of the sample size. Other alternatives are the reality check test for data snooping of White (2000) and the improved version proposed by Hansen (2005). These tests are based on superior predictive ability (SPA) and allow for multiple comparison but they require a benchmark model. An alternative approach is the model confidence set (MCS) test proposed by Hansen, Lunde, and Nason (2009). The MCS allows to identify, from a universe of model based forecasts, a subset of models, equivalent in terms of predictive ability, which outperform all the other competing models. In this paper, we use both the SPA and MCS tests to assess the forecast performance of multivariate GARCH models.

To measure out-of-sample forecasting performance, model based forecasts are usually compared to ex-post realizations as they become available. To do this, the forecaster needs to select a loss function and a proxy for the true conditional variance matrix which is unobservable even ex post. The question arises on which proxy to use and to what extent this substitution affects the forecast evaluation. Building on Hansen and Lunde (2006a) and Patton (2009), Laurent, Rombouts, and Violante (2009) address these questions in the case of the comparison of multivariate volatility models using statistical loss functions. They show that the substitution of the underlying volatility by a proxy may induce a distortion in the ranking i.e., the evaluation based on the proxy differs from the ranking that would be obtained if the true target was observable. However, such distortion can be avoided if the loss function has a particular functional form. In this paper, we use four robust loss functions which allow for various types of asymmetry in the way variances and variance matrix predictions are evaluated. With respect to the choice of the loss function, and within the MCS framework, we find that the Euclidean and Frobenius loss functions (both symmetric) appear to deliver a relatively large MCS, while the asymmetric loss functions, and in particular the Stein loss function, allow to identify sets of superior models which are systematically smaller. These results are consistent with the findings of Clements, Doolan, Hurn, and Becker (2009) in the multivariate setting and Hansen, Lunde, and Nason (2003) in the univariate settings.

We consider 10 series from the NYSE and the NASDAQ indices. The sample period is 21 years, from January 2, 1988 to December 31, 2008. The last 2486 trading days (from April 1, 1999 to December 27, 2008) constitute the sample for which we compute one-day ahead forecasts. We consider 125 multivariate GARCH model based forecasts. Laurent, Rombouts, and Violante (2009) underline the value of high precision proxies. In fact, when the set of competing models is characterized by a high degree of similarity, the availability of an accurate proxy makes it easier to discriminate between models. In this paper, model performances are evaluated using realized covariance based on intraday returns sampled at the 5 minute frequency. A robustness check with respect to the choice and the accuracy of the proxy is performed using intraday returns sampled at 1 minute and a realized kernel estimator based on intraday returns sampled at 1 and 5 minutes. Our results appear to be robust to the choice and the accuracy of the volatility proxy.

As pointed out by Hansen, Lunde, and Nason (2003), the MCS is specific to the set of candidate models and the sample period. Furthermore, the model selection can be misleading when the forecast sample consists of periods characterized by different types of dynamics. We illustrate how sensitive the MCS is with respect to the forecast sample under investigation by considering not only the full sample but also three sub-samples which are homogenous in their volatility dynamics. We find that over the dot-com bubble, the set of superior models is composed of more sophisticated models such as Orthogonal and dynamic conditional correlations, both with leverage effect in the conditional variances. Over calm periods, a simple assumption like constant conditional correlation and symmetry in the conditional variances cannot be rejected. Over the 2007-2008 financial crisis, accounting for non-stationarity in the conditional variance process generates superior forecasts.

In the last part of our application, we assess using SPA tests the predictive ability of six popular and parsimonious specifications selected with respect to two dimensions, the multivariate structure and symmetry in the dynamics of the variance processes. We find that the most valid alternative is represented by the Dynamic Conditional Correlation model of Engle (2002) when coupled with leverage effect in the conditional variances of the marginal processes. This model seems to capture well the dynamics of the conditional variance matrix consistently across the different sample periods. However, in line with the MCS results, simple hypotheses like constant correlation and/or symmetric variance process cannot be rejected only over periods of calm markets.

An alternative approach to evaluate variance matrix forecasts is to use an economic loss function such as asset allocation in Engle and Colacito (2006). Other examples are Value-at-Risk forecasting and derivative pricing. See also Voev (2009) for a related setting. However, as pointed out by Patton and Sheppard (2009) the main drawback of an evaluation of volatility forecasts based on economic criteria is that it generally relies on additional and application-specific assumptions, the ordering may not depend exclusively on the accuracy of the conditional variance matrix forecast and the criteria are generally non-robust, in the sense that imperfect forecasts can outperform the true conditional variance matrix.

The rest of the paper is organized as follows. Section 2 discusses the multivariate GARCH specifications, the proxies for the conditional variance, the loss functions and the MCS approach. Section 3 provides a description of the data and outlines some stylized facts. Section 4 presents the results for the multiple comparison based on the MCS and Section 5 for the comparison based on the SPA test. Section 6 concludes.

2 Methodology

In this section, we first introduce the multivariate GARCH models used for the forecasting exercise. Second, we define estimators of the underlying variance matrix used to compare the volatility forecasts. We conclude with a discussion on the properties of the loss functions used to evaluate the forecast errors and with a brief summary of the MCS approach.

2.1 Forecasting models set

Consider a N-dimensional vector stochastic process $r_t = \mu_t + \varepsilon_t$ and denote $\Im_{t-1}$ as the information set available at $t-1$. We are interested in modelling its conditional variance matrix $H_t = E(\varepsilon_t \varepsilon_t' | \Im_{t-1})$. Since the conditional mean μ_t is typically of minor importance for GARCH-type models, we assume a constant conditional mean for all assets, see also Hansen and Lunde (2005) and Becker and Clements (2008).

We consider parametric specifications for the conditional variance of the multivariate GARCH (MGARCH) type, i.e., H_t is a parametric function of past returns. To control for the number of parameters, we impose covariance or correlation targeting when possible, see Engle and Mezrich (1995). This means that H_t can be expressed in terms of the unconditional variance/correlation and other parameters, provided that the process is covariance stationary. Hence, it is possible to reparameterize the model and replace the unconditional covariance and/or correlation by a consistent estimator before maximizing the likelihood. The targeting ensures a reasonable value of the model-implied unconditional variance and, although it is not a maximum likelihood estimator (therefore asymptotically inefficient), the long run variance will be consistent even if the MGARCH model is misspecified. This solution also facilitates the numerical optimization of the remaining parameters by reducing the dimensionality of the parameter space. For the properties of the variance targeting estimator and a comparison with the standard quasi-maximum likelihood estimator in the univariate case, see Francq, Horvath, and Zakoian (2009).

We consider several families of MGARCH models which are revealed to be feasible in terms of numerical evaluation when the dimension of r_t is relatively large. According to the classification in Bauwens, Laurent, and Rombouts (2006), among the generalizations of the univariate standard GARCH model, we consider three specifications, namely the diagonal and scalar BEKK of Engle and Kroner (1995) and the multivariate RiskMetrics model of

J.P.Morgan (1996). In the BEKK model, the conditional variance is specified as

$$H_t = C + A\varepsilon_{t-1}\varepsilon_{t-1}' A' + BH_{t-1}B', \tag{1}$$

where C is a positive definite matrix and A and B are diagonal matrices of parameters in the diagonal BEKK and $A = aI$, $B = bI$, where a and b are scalars, in the scalar BEKK. In this model, variance targeting is imposed by setting $H = E(\varepsilon_t\varepsilon_t')$ and $C = I - AHA' - BHB'$ which implies $E(H_t) = H$. Note that the scalar BEKK model imposes the same dynamics to all the elements of H_t (and thus is equivalent to the scalar VEC model of Bollerslev, Engle, and Wooldridge, 1988). The RiskMetrics model has the same parametric form as defined in (1) but assumes that the conditional variance matrix is an integrated process, i.e., $a + b = 1$ and $C = 0$, governed by a fixed smoothing parameter, b equal to 0.96. This model, widely used by practitioners, does not require parameter estimation.

Among the MGARCH models that can be represented as linear combinations of univariate GARCH models, we consider the orthogonal GARCH (Ogarch) model of Kariya (1988) and Alexander and Chibumba (1997). In this model, the data are generated by an orthogonal transformation of $m \leq N$ (or a smaller number of) uncorrelated factors, f_t, which can be separately defined as any stationary univariate GARCH process. The model can be expressed as

$$V^{-1/2}\varepsilon_t = PL^{1/2}f_t \tag{2}$$

$$S_t = E_{t-1}(f_t f_t') = diag(\sigma^2_{f_{1,t}}, \ldots, \sigma^2_{f_{m,t}}) \tag{3}$$

$$H_t = V^{1/2}PL^{1/2}S_t PL^{1/2}V^{1/2}, \tag{4}$$

where $V = diag(v_1, ..., v_N)$, with $v_i = E(\varepsilon^2_{i,t})$, L and P are $m \times m$ and $N \times m$ matrices of the m largest eigenvalues of the unconditional correlation matrix and associated orthogonal eigenvectors, respectively. In the application, we set $m = N$. Other specifications belonging to this group are the generalized orthogonal GARCH model by van der Weide (2002), the full factor GARCH model by Vrontos, Dellaportas, and Politis (2003) and the conditionally uncorrelated components GARCH by Fan, Wang, and Yao (2008). However, these models are computationally challenging when the dimension is large.

The last category of models can be viewed as nonlinear combinations of univariate GARCH models. They allow to specify separately N individual, possibly different, univariate models for the conditional variances and a model for the conditional correlation matrix. The dynamic conditional correlation (DCC) model, in the formulation of Engle (2002) (DCCE), is defined as

$$
\begin{aligned}
H_t &= D_t^{1/2} R_t D_t^{1/2} && (5)\\
D_t &= diag(\sigma_{1,t}^2, \ldots, \sigma_{N,t}^2) && (6)\\
R_t &= diag(q_{11,t}^{-1/2} \ldots q_{NN,t}^{-1/2}) Q_t diag(q_{11,t}^{-1/2} \ldots q_{NN,t}^{-1/2}) && (7)\\
Q_t &= (1-\alpha-\beta)\bar{Q} + \alpha u_{t-1}u'_{t-1} + \beta Q_{t-1}, && (8)
\end{aligned}
$$

where $u_{i,t} = \varepsilon_{i,t}/\sigma_{i,t}$ define the devolatilized innovations. The constant conditional correlation (CCC) model of Bollerslev (1990), the asymmetric DCC (ADCC) model of Cappiello, Engle, and Sheppard (2006), the Dynamic Conditional Equi-Correlation (DECO) model of Engle and Kelly (2008) also belong to this family. To ensure positive definiteness, the correlation matrix is modeled as a transformation of a latent matrix Q_t which is a function of past devolatilized innovations.

While the CCC model of Bollerslev (1990) assumes time invariant, but pairise specific, correlations, which can be estimated by a consistent estimator for the unconditional correlation, the DECO model of Engle and Kelly (2008) assumes that correlations are time varying but equal across the N assets ($R_{ij,t} = \rho_t \ \forall i \neq j$). Interestingly, under some suitable conditions, the DECO model gives consistent estimators of the correlation dynamics (α, β) in (8) even when the equicorrelation assumption is not supported by the data. Since the hypothesis of equicorrelation is likely to be rejected, in this paper we use the DECO approach as a technique to estimate the correlation parameters α and β. We then use the DECO estimates to predict and forecast time varying and pairwise specific correlations. The ADCC extends the DCCE by accounting for asymmetries in the correlation dynamics through the additional term $\gamma(u_{t-1}u'_{t-1} \odot 1_{u_{t-1}<0}1'_{u_{t-1}<0})$ in (8) where $1_{u_{t-1}<0}$ is a vector of dimension N such that $[1_{u_{t-1}<0}]_i = 1$ if $u_{i,t-1} < 0$ and 0 otherwise. The main drawback of the DCCE, the DECO and the ADCC, is that, under variance/correlation targeting, the choice of the estimator for the long run target $\bar{Q}$ is not obvious as Q_t is not a conditional variance nor a correlation. Although inconsistent for the target, since the recursion in Q_t does not have a martingale difference representation, Engle and Sheppard (2001) suggest the use of the unconditional expectation of the outer product of devolatilized innovations, arguing that the impact of this choice is very small in practice.

An alternative formulation of the DCC model has been suggested by Tse and Tsui (2002) (DCCT). The conditional correlation R_t defined as:

$$R_t = (1 - \theta_1 - \theta_2)\bar{R} + \theta_1 \Psi_{t-1} + \theta_2 R_{t-1}, \tag{9}$$

with Ψ_{t-1} the $N \times N$ correlation matrix of ε_τ for $\tau = t-K, t-K+1, \ldots, t-1$ and $K \geq N$. Its i, j-th element is given by

$$\psi_{ij,t-1} = \frac{\sum_{m=1}^{K} u_{i,t-m} u_{j,t-m}}{\sqrt{(\sum_{m=1}^{K} u_{i,t-m}^2)(\sum_{m=1}^{K} u_{j,t-m}^2)}}, \tag{10}$$

where u_{it} is defined as above. In the application, we use $K = N$. In the DCCT the correlation matrix is modeled directly and depends on past local correlations of devolatilized innovations. Also in this case, under variance/correlation targeting, the choice of $\bar{R}$ is not obvious. We set $\bar{R}$ equal to the unconditional correlation of the devolatilized innovations.

One of the advantages of the conditional correlation models relies on the fact that the estimation problem can be carried out sequentially. This requires first the estimation of the N conditional variances of the assets, potentially preceded by the estimation of the variance target, second the estimation of the correlation target and third the parameters governing the dynamics of the conditional correlation. Although inefficient, this procedure is consistent and it dramatically reduces the computational burden of the likelihood. The univariate specification for the conditional variance that we include in the conditional correlation models are ARCH (Engle, 1982), GARCH (Bollerslev, 1986), GJR (Glosten, Jagannathan, and Runkle, 1992), Exponential GARCH (Nelson, 1991), Asymmetric Power ARCH (Ding, Granger, and Engle, 1993), Integrated GARCH (Engle and Bollerslev, 1986), RiskMetrics (J.P.Morgan, 1996), Hyperbolic GARCH (Davidson, 2004) and Fractionally Integrated GARCH (Baillie, Bollerslev, and Mikkelsen, 1996). With respect to the number of lags in the models, we fix both the ARCH (p) and the GARCH (q) orders to 1 for the scalar BEKK, multivariate RiskMetrics and the correlation specification in the DCC models. The univariate GARCH models for the conditional variances in the Orthogonal GARCH and DCC specifications include various combinations of the orders p, q. Table 1 summarizes the 125 multivariate GARCH configurations considered in the forecasting exercise.

Table 1: Forecasting models set

Conditional correlation type				Orthogonal GARCH				BEKK type			
Corr.	Variance	p	q		Variance	p	q			p	q
	Arch	1,2	-		Arch	1,2	-	BEKK	scalar	1	1
	Aparch	1	1		Aparch	1	1		diagonal	1	1
CCC,	Egarch	0,1,2	1,2	Orth.	Egarch	0,1,2	1,2	RM	-	1	1
DCCA,	Garch	1,2	1,2		Garch	1,2	1,2				
DCCE,	Gjr	1,2	1,2		Gjr	1,2	1,2				
DCCT,	Hgarch	1	1								
DECO	Igarch	1	1								
	Figarch	1	1								
	Rm	1	1								

2.2 Proxies for the conditional variance matrix

In our application, the daily realized covariance serves as a proxy for the true conditional variance matrix, Σ_t, when evaluating the forecasting performance of the different MGARCH models. Recent literature suggests several estimators. Examples are the well known realized variance, and its jump robust version bi-power covariation, see Barndorff-Nielsen and Shephard (2004a) and Barndorff-Nielsen and Shephard (2004b), the realized kernel estimators proposed by Zhou (1996), Hansen and Lunde (2006b), Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008a) and Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008b) which account for serial correlation in the high frequency returns. Parametric models, like vector moving average $RCov$ can be found in Hansen, Large, and Lunde (2008). Intraday returns are defined as $r_t = p_t - p_{t-\Delta}$ for $t = \Delta, 2\Delta, ..., T$, with $1/\Delta$ intervals per day. The daily realized variance ($RCov$) matrix (Andersen, Bollerslev, Diebold, and Labys, 2003 and Barndorff-Nielsen and Shephard, 2004a) is defined as

$$RCov^{(\Delta)} = \sum_{i=1}^{1/\Delta} r_{i\Delta} r'_{i\Delta}. \tag{11}$$

As the sampling frequency of the intraday returns increases ($\Delta \to 0$), $RCov^{(\Delta)}$ converges almost surely to Σ_t. See Barndorff-Nielsen and Shephard (2004b), Mykland and Zhang (2006), Andersen, Bollerslev, and Diebold (2002) and related references for details.

The definition of $RCov^{(\Delta)}$ requires the assumption that intraday returns are uncorrelated. However, failing this assumption, $RCov^{(\Delta)}$ would result in a biased estimator of Σ_t. Hence, we also consider a simple kernel estimator, defined as

$$RCovAC_q^{(\Delta)} = \lambda_0 + \sum_{i=1}^{q} (\lambda_{-i} + \lambda_i) + \sum_{i=q+1}^{2q} \left(1 - \frac{i-q}{q+1}\right)(\lambda_{-i} + \lambda_i) \quad (12)$$

$$\lambda_q = \begin{cases} \frac{1}{(1-q\Delta)} \sum_{i=q+1}^{1/\Delta} r_i r'_{i-q} & q \geq 0 \\ \frac{1}{(1-|q|\Delta)} \sum_{i=|q|+1}^{1/\Delta} r_{i-|q|} r'_i & q < 0 \end{cases} . \quad (13)$$

This estimator (see Zhou, 1996, Zhang, Mykland, and Ait-Sahalia, 2005, Hansen and Lunde, 2006b and Hansen, Large, and Lunde, 2008), based on the Newey and West (1987) variance estimator, is equal to the $RCov^{(\Delta)}$ plus a term that is a Bartlett-type weighted sum of higher-order autocovariances. More refined realized kernel estimators are recently proposed by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008a) and Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008b). Throughout the paper, unless explicitly mentioned, we will use the $RCov^{(5min)}$ estimator. $RCov^{(1min)}$, $RCovAC_q^{(1min)}$ and $RCovAC_q^{(5min)}$ will serve to check the robustness of the results to different proxies.

2.3 Loss functions

At the core of the forecasting comparison is the choice of the loss function. In this paper, we use the following loss functions

$$L_E = (\sigma_t - h_t)'(\sigma_t - h_t) \quad (14)$$

$$L_F = Tr[(\Sigma_t - H_t)'(\Sigma_t - H_t)] \quad (15)$$

$$L_S = Tr[H_t^{-1}\Sigma_t] - \log\left|H_t^{-1}\Sigma_t\right| - N \quad (16)$$

$$L_d = \frac{1}{d(d-1)} Tr(\Sigma_t^d - H_t^d) - \frac{1}{(d-1)} Tr(H_t^{d-1}(\Sigma_t - H_t)) \quad d \geq 3. \quad (17)$$

The first two loss functions belong to a family of quadratic loss functions based on the forecast error. L_E is the Euclidean distance in the vector space of $\sigma_t - h_t = vech(\Sigma_t - H_t)$, where $vech()$ is the operator that stacks the lower triangular portion of a matrix into a vector. Hence, L_E only considers the unique elements of the variance matrix and these elements are equally weighted. The Frobenius distance, L_F, is defined as the sum of the element-wise square differences of $\Sigma_t - H_t$ and is the natural extension to matrix spaces of the mean squared error. The relevant variable in the comparison is in this case the variance matrix itself and it corresponds to the loss function implied by the matrix Normal likelihood. Although closely related, it differs from L_E for double counting the loss associated to the conditional covariances. The Stein loss function L_S of James and Stein (1961) is a scale invariant loss function based on the standardized (in matrix sense) forecast error. It is the loss function implied by the Wishart density.

Note that since L_E only considers the unique elements of the forecast error matrix, it is symmetric in the sense that variances and covariances over/under predictions are equally penalized. L_F equally weights all elements of the forecast error matrix, thus double counting the covariances forecast errors. This means that L_F is symmetric with respect to the sign of the forecast error for a given element of the forecast error matrix, but it is asymmetric in the way that diagonal and off-diagonal elements of the forecast error matrix are weighted. The loss function L_S also considers the whole variance matrix as the variable of interest. This loss function is homogeneous of degree 0 (errors are measured in relative terms) and asymmetric with respect to over/under predictions (in matrix sense) and, in particular, under predictions are heavily penalized. Finally, in the same spirit, L_3 also accounts for asymmetry with respect to over/under predictions, but in the opposite direction, i.e. over predictions are penalized instead. L_d also allows to tune the degree of asymmetry, i.e. the weights given to over/under prediction, through the choice of the parameter d, which also represents its degree of homogeneity. We set $d = 3$ which implies a mild degree of asymmetry comparable to the one of L_S. See Laurent, Rombouts, and Violante (2009) for further details and examples.

2.4 The model confidence set

The MCS approach, introduced by Hansen, Lunde, and Nason (2009), is a testing procedure for superior predictive ability based on the reality check for data snooping of White (2000) and the superior predictive ability (SPA) test of Hansen (2005). The test allows to identify a subset of models equivalent in terms of predictive ability, that are superior to the other models. The advantage of the MCS procedure is that it does not require a benchmark model to be specified which is useful for applications without an objective benchmark.

Let us denote M^0 the initial set of models for which we compute one-step ahead conditional variance forecasts, denoted by $\hat{H}_{i,T+1}, ..., \hat{H}_{i,T+T^*+1}$ $i = 1, \ldots, M$ where T^* defines the forecasting sample length. The MCS procedure allows to identify a subset of models, M^*, which are superior, in terms of predictive ability, with respect to all the other models in M^0. To do this, we need an equivalence test, an elimination rule and an updating algorithm. The starting hypothesis is that all models in M^0 have equal forecasting performances as measured by a loss function $L_{i,t} = L(\Sigma_t, H_{i,t})$. If the null of equal predictive ability is rejected, then the elimination rule removes the model with the worst performing model. This process is repeated until the non-rejection of the null occurs (at a given confidence level). The set of surviving models is the MCS. More formally, we start by defining the relative performance at time t as $d_{ij,t} = L_{i,t} - L_{j,t}$ for all $i \neq j \in M^0$. Under the assumption that $d_{ij,t}$ is stationary, the null hypothesis takes the form $H_{0,M^0} : E(d_{ij,t}) = 0, \forall\, i \neq j \in M^0$. The deviation statistic is defined as $T_D = \frac{1}{M}\sum_{i\in M^0} t_i^2$, where $t_i = \frac{\sqrt{T^*}\bar{d}_i}{\omega_i}$ and $\bar{d}_i = M^{-1}\Sigma_{j\in M^0}\bar{d_{ij}}$ is the contrast of model i's sample loss with respect to the average across all models and $\bar{d_{ij}} = T^{*-1}\Sigma_{t=1}^{T^*} d_{ij,t}$ is the sample loss difference between model i and j. The variances $\omega_i^2 = lim_{t\to\infty} Var(\sqrt{T^*}\bar{d}_i)$ and the distribution of T_D can be obtained by a bootstrap scheme. If the null hypothesis is rejected, then we use as elimination rule $argmax_i t_i$ to exclude the weakest model from the set. The elimination rule excludes the model with the largest standardized excess loss relative to the average across models, that is $\bar{d}_i = \bar{L}_i - \bar{L} = \bar{L}_i - M^{-1}\Sigma_{j\in M^0}\bar{L}_j = M^{-1}\Sigma_{j\in M^0}(\bar{L}_i - \bar{L}_j)$. The MCS p-value is equal to $p_i = \max_{k\leq i} p(k)$ where $p(k)$ is the p-value of the test under the null H_{0,M^k} where k is the number of surviving models at step i of the iteration process. After the necessary iterations, the set of superior models is given by $\{i \in M_0 : E(d_{ij,t}) \leq 0$ $\forall\, i \neq j \in M^0\}$.

As argued by Hansen, Lunde, and Nason (2009), even an inferior model (a model with bad sample performance) may be included in the MCS. This is the case if the variance of its relative performance is large enough, i.e. the resulting standardized relative deviation, t_i, gets small enough to avoid being discarded by the elimination rule. Consider the following decomposition for $Var(\bar{d}_i)$

$$\begin{aligned} Var(\bar{d}_i) &= Var(\bar{L}) + Var(\bar{L}_i) - 2Cov(\bar{L}_i, \bar{L}) \\ &= Var(\bar{L}) + \left(1 + \frac{Var(\bar{L}_i)}{Var(\bar{L})} - 2\sqrt{\frac{Var(\bar{L}_i)}{Var(\bar{L})}} Corr(\bar{L}_i, \bar{L})\right). \end{aligned} \tag{18}$$

If we define an inferior model as a model with a sample performance worse than the average, that is $\bar{d}_i > 0$ or alternatively $\bar{L}_i > \bar{L}$ - such model enters the MCS at some given confidence level if and only if $Var(\bar{L}_i)$ is large enough and/or $Corr(\bar{L}_i, \bar{L})$ is small. However, in some specific cases this problem does not arise or it just marginally affects the elimination process. For example, if the set contains only two models, then $|\bar{d}_1| = |\bar{d}_2|$ and it follows that $Var(\bar{d}_1) =$

$Var(\bar{d}_2)$ and consequently the variance plays no role in the elimination. In such a case, for some level of confidence and given the elimination rule defined above, the model with the best sample performance is always preferred. In the case where the set contains more than two models, an inferior model might only be preferred to another inferior model with better sample performance but it will not outperform models for which $\bar{d}_i < 0$. By the same reasoning, if there is only one model in the set with $\bar{d}_i > 0$, it will always be excluded no matter how large its variance is. The decomposition of the variance of the relative performances plays a central role for understanding and disentangling the informativeness of the MCS, i.e., to assess whether weak models have been included in the set of superior models and the overall informativeness of the resulting MCS.

3 Data and forecasting scheme

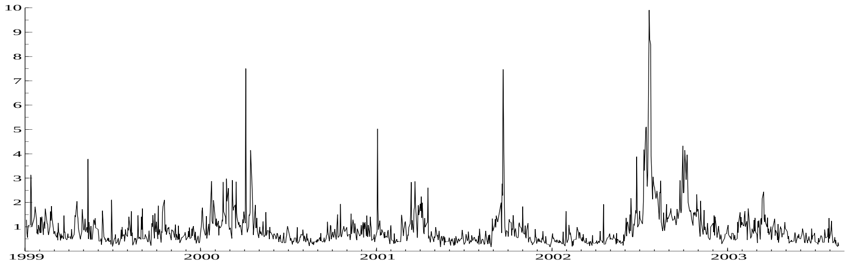
We consider stock returns from 10 assets traded in the NYSE and NASDAQ as detailed in Table 2. The sample period spans from March 02, 1988 to December 26, 2008, which amounts to 5226 trading days. The dataset has been cleaned from weekends, holidays and early closing days. Days with missing values and/or constant prices have also been removed. Following the approach of Andersen, Bollerslev, Frederiksen, and Nielsen (2010), the MGARCH models are estimated using daily open-to-close returns. As explained above, to reduce the computational burden, unconditional means are subtracted from each series of returns before proceeding to the estimation of the 125 multivariate GARCH models by quasi maximum likelihood. All programs are available from the authors on request. The initial estimation sample consists of the first 2740 daily observations, i.e. March 02, 1988 to December 31, 1999. The last 2486 trading days constitute the sample for which we compute one-day ahead forecasts. For computational convenience, we only re-estimate the model parameters every month (22 days) using a rolling window of the last 2740 observations. This rolling window of fixed size satisfies the assumptions required by the MCS test (Hansen, Lunde, and Nason, 2009), allows the comparison of nested models (Giacomini and White, 2006), as well as to compare results over sub-samples (since forecasts over different period are conditioned on the most recent information). The proxies for the conditional variance are based on intraday returns computed from five-minutes intervals last mid-quotes. Since the daily trading period of the NYSE and NASDAQ is 6.5 hours, this amounts to 78 intraday observations per day.

Table 2: Stock names and descriptive statistics

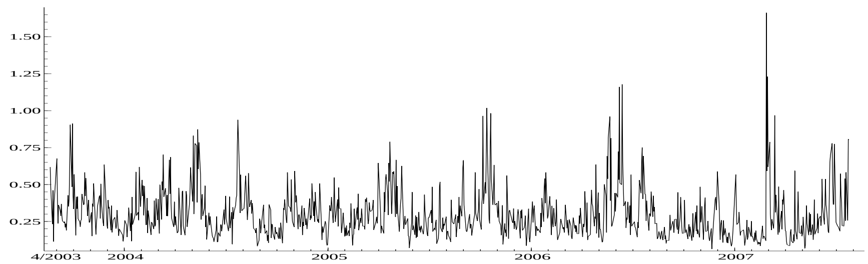
Name	Sector	Mean	Std. Dev.	Max	Min	Skewness	Kurtosis
Abbott Labs	Health Care	0.085	1.53	10.26	-9.47	-0.05	2.43
BP plc	Energy	0.013	1.17	10.27	-13.96	-0.22	11.83
Colgate-Palmolive	Consumer Stap.	0.073	1.40	16.51	-8.59	0.35	6.48
Eastman Kodak	Consumer Disc.	-0.043	1.74	12.76	-14.13	-0.14	6.42
FedEx Corp.	Industrials	0.068	1.79	12.58	-9.67	0.39	2.93
Coca Cola Co.	Consumer Stap.	0.067	1.38	8.92	-11.08	0.06	3.79
PepsiCo Inc.	Consumer Stap.	0.127	1.44	12.14	-13.78	-0.11	5.97
Procter & Gamble	Consumer Stap.	0.100	1.33	10.50	-9.05	0.00	5.01
Wal-Mart	Consumer Stap.	0.008	1.64	14.75	-8.71	0.27	4.35
Wyeth	Health Care	0.027	1.65	12.32	-15.42	-0.31	6.67

Note. Statistics based on the full sample (estimation plus forecast) of 5229 observations

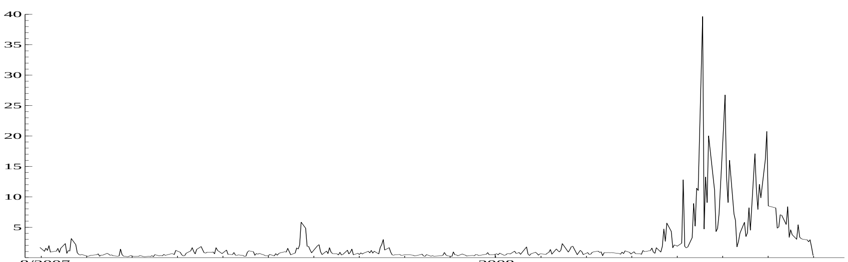
The sample period we consider is characterized by dramatic changes in volatility dynamics. To investigate the impact of this on the MCS results, the forecasting sample has been divided into three sub-samples. The first sub-sample (1050 obs.) identifies a period of widespread turbulence on the markets. Starting in January, 1999, and ending in March 2003, it includes the peak of the Dot-com boom (until March 2000), the burst and the aftermath of the bubble burst. Peaks in the volatility over this period correspond to the burst of the speculative bubble (March, 2000) and the attack to the twin towers (September, 2001). Towards the end of the period, the turmoil started with the bankruptcy of WorldCom (July, 2002) and ended in October, 2002, with a record low of the Dow Jones Industrial and Nasdaq (5- and 6-years low respectively). The second sub-sample (1080 obs.), from April 2003 to July, 2007, corresponds to a period of market stability. The third sub-sample (356 obs.) corresponds to the recent financial crisis. The beginning of the sample, August, 2007, coincides with the fall of Northern Rock when it became apparent that the financial turmoil, started with the subprime crisis in the US, had spread beyond US's borders. It is also the period when the crisis hits its peak in September and October 2008. To visualize the difference among the three sub periods, Figure 1 shows the realized variance of an equally weighted portfolio made of the 10 assets used in the application. It is clear from this figure that the volatility dynamics as well as its scale varies widely between periods.



(a) Dot-com bubble (99/01/04 - 03/03/31) - 1050 Obs.



(b) Calm period (03/04/01 - 07/07/31) - 1080 Obs.



(c) 2007-2008 financial crisis (07/08/01 - 08/12/26) - 356 Obs.

Figure 1: Daily realized volatility (computed from 5-min returns) of the 10 asset equally weighted portfolio

4 Multiple comparison based on conditional covariance forecasts

We describe the MCS results based on the conditional variance forecasts for four different forecasting samples described in the previous section, i.e. the full sample, the dot-com speculative bubble burst and aftermath, calm markets and the 2007-2008 financial crisis.

4.1 Full sample

The MCS results for the full forecast sample (2486 observations) are reported in Table 3 for the Euclidean (L_E), Stein (L_S) and L_3 loss functions. To save space, results for the Frobenius loss function (L_F) are not reported. Because of its similarity with L_E, results based on L_F are very similar in terms of ordering and MCS. However, in general we remark that the L_E MCS always includes the MCS obtained under the L_F loss function. Following Hansen, Lunde, and Nason (2003), we set the confidence level for the MCS to $\alpha = 0.25$. The number of bootstrap samples used to obtain the distribution under the null is set to $10,000$. The values reported for L_E are the average loss per element of $vech(\Sigma_t - H_t)$, i.e. the total loss is divided by $N(N+1)/2$ and N^2 respectively. For L_S, where the distance is measured in relative terms, the total loss is reported.

The MCS includes 39 models for L_E and is largely dominated by orthogonal and DECO models. We remark the following points with respect to the composition of the MCS. First, the family of orthogonal models exhibit the best sample performances. The flexibility of the orthogonal GARCH model seems therefore able to adapt to a sample that alternates periods of calm with periods of extremely high instability. The MCS also includes most specifications from the DECO family. Furthermore, the results suggest the rejection of the hypothesis of constant conditional correlation. Second, although the difference is not statistically significant, models allowing for asymmetry/leverage in the conditional variance systematically perform better than symmetric models with Gjr specifications showing the best sample performances. The same consideration holds with respect to longer versus shorter lags, with longer lag models showing in general better sample performances. Third, the MCS includes some specifications that allow for long memory and integrated conditional variances. This is the case for the DECO, DCCA and DCCE with hyperbolic GARCH conditional variances, DECO, DCCA and DCCT with fractionally integrated GARCH conditional variances, DECO with RiskMetrics conditional variances and the multivariate RiskMetrics model. Furthermore, if we focus only on the ranking based on sample performances, the specifications allowing for fractional integration or hyperbolic decay of shocks in the conditional variances exhibit the best sample performances within each family of models.

We next turn to the MCS under the two asymmetric loss functions for which we find substantially different results compared to L_E. Under L_S, the MCS includes 10 models, all belonging to the DCC family. Interestingly, the selected models focus on the long memory properties of the conditional variances rather than leverage, asymmetry or even time varying correlation. In fact, the MCS includes models from the CCC, DCCE, DCCA and DCCT

Table 3: MCS on full sample (99/01/04 - 08/12/26)

Euclidean distance (39 models)

MCS($\alpha = 25\%$)		Rnk	$\bar{L}_i$	T_D	p-val	VR	Corr
DCCA	Egarch (1,2)	48	3.880	1.165	0.27	1.302	0.999
	Figarch (1,1)	20	3.673	0.521	0.67	1.076	0.996
	Hgarch (1,1)	25	3.720	0.803	0.45	1.052	0.996
DCCT	Figarch (1,1)	38	3.823	1.089	0.30	1.159	0.994
DCCE	Egarch (1,2)	53	3.901	1.207	0.25	1.325	0.998
	Figarch (1,1)	18	3.661	0.406	0.71	1.075	0.996
	Hgarch (1,1)	24	3.719	0.766	0.47	1.057	0.996
DECO	Aparch (1,1)	27	3.735	0.848	0.42	1.111	0.998
	Egarch (0,1)	29	3.742	0.825	0.43	1.172	0.999
	(0,2)	30	3.747	0.877	0.40	1.163	0.999
	(1,2)	33	3.762	0.936	0.37	1.176	0.999
	Figarch (1,1)	2	3.478	0.004	0.94	0.906	0.997
	Garch (1,1)	34	3.768	0.906	0.38	1.171	0.998
	(1,2)	31	3.750	0.965	0.35	1.137	0.999
	(2,1)	28	3.737	0.993	0.34	1.125	0.999
	(2,2)	32	3.759	1.061	0.31	1.159	0.999
	Gjr (1,1)	22	3.692	0.603	0.60	1.090	0.998
	(1,2)	21	3.676	0.706	0.50	1.046	0.999
	(2,1)	14	3.635	0.521	0.67	0.991	0.999
	(2,2)	19	3.667	0.668	0.54	1.036	0.999
	Hgarch (1,1)	5	3.535	0.103	0.89	0.886	0.997
	Igarch (1,1)	35	3.783	1.018	0.33	1.061	0.993
	Rm (1,1)	23	3.699	0.545	0.64	1.117	0.998
Orth.	Aparch (1,1)	7	3.575	0.197	0.89	0.921	0.996
	Egarch (0,1)	17	3.660	0.628	0.58	1.019	0.998
	(0,2)	13	3.623	0.567	0.64	0.945	0.999
	(1,1)	15	3.647	0.735	0.50	0.933	0.998
	(1,2)	12	3.593	0.517	0.67	0.872	0.997
	(2,1)	26	3.726	1.037	0.32	1.066	0.999
	(2,2)	6	3.539	0.175	0.89	0.793	0.996
	Garch (1,1)	16	3.656	0.724	0.50	0.964	0.998
	(1,2)	11	3.589	0.594	0.67	0.870	0.998
	(2,1)	9	3.586	0.549	0.67	0.885	0.999
	(2,2)	8	3.580	0.466	0.69	0.865	0.998
	Gjr (1,1)	10	3.587	0.412	0.73	0.817	0.997
	(1,2)	3	3.507	0.169	0.89	0.713	0.996
	(2,1)	1	3.468	-	1.00	0.672	0.995
	(2,2)	4	3.509	0.116	0.89	0.730	0.996
RM	(1,1)	36	3.810	1.127	0.28	0.967	0.993

Stein distance (10 models)

MCS($\alpha = 25\%$)		Rnk	$\bar{L}_i$	T_D	p-val	VR	Corr
CCC	Figarch (1,1)	7	3.528	0.346	0.57	0.730	0.932
	Garch (2,1)	10	3.548	1.302	0.25	1.211	0.988
	Igarch (1,1)	3	3.501	0.546	0.69	1.119	0.985
DCCA	Igarch (1,1)	4	3.516	0.680	0.57	1.254	0.986
DCCT	Figarch (1,1)	5	3.518	0.232	0.69	0.743	0.931
	Garch (2,1)	9	3.541	0.880	0.36	1.223	0.989
	Igarch (1,1)	1	3.496	-	1.00	1.130	0.987
DCCE	Figarch (1,1)	6	3.525	0.381	0.57	0.789	0.929
	Garch (2,1)	8	3.535	0.561	0.49	1.255	0.989
	Igarch (1,1)	2	3.500	0.235	0.69	1.228	0.986

L_3 loss function (20 models)

MCS($\alpha = 25\%$)		Rnk	$\bar{L}_i$	T_D	p-val	VR	Corr
DECO	Figarch (1,1)	15	102.3	0.844	0.41	1.092	0.999
	Gjr (1,1)	23	105.0	1.063	0.29	1.156	1.000
	Hgarch (1,1)	17	102.5	0.883	0.39	1.082	0.999
	Igarch (1,1)	24	105.1	0.969	0.33	1.142	0.999
Orth.	Aparch (1,1)	10	98.83	0.689	0.50	1.023	0.999
	Egarch (0,1)	14	101.3	0.864	0.43	1.083	1.000
	(0,2)	11	99.98	1.001	0.43	1.033	1.000
	(1,1)	12	100.4	0.848	0.43	1.017	1.000
	(1,2)	6	98.45	0.988	0.43	0.977	0.999
	(2,1)	16	102.4	0.834	0.43	1.104	1.000
	(2,2)	4	95.68	0.609	0.53	0.898	0.999
	Arch (1)	18	103.8	0.923	0.43	0.914	0.990
	Garch (1,1)	13	101.1	0.851	0.43	1.053	1.000
	(1,2)	9	98.60	1.055	0.43	0.984	1.000
	(2,1)	7	98.59	0.928	0.43	0.997	1.000
	(2,2)	8	98.60	1.099	0.43	0.980	1.000
	Gjr (1,1)	5	97.69	0.887	0.43	0.954	0.999
	(1,2)	2	94.06	1.032	0.53	0.852	1.000
	(2,1)	1	91.98	-	1.00	0.801	1.000
	(2,2)	3	94.54	0.782	0.53	0.872	0.999

Note. Rnk: model i's ranking position based on average sample performances (out of 125 models); $\bar{L}_i$: model i's average sample performance; T_D: deviation statistic; p-val: MCS p-value; $VR : V(\bar{L}_i)/V(\bar{L})$ ratio between the variance of model i's loss and the average loss (across models); Corr: $Corr(\bar{L}_i, \bar{L})$ correlation between model i's loss and the average loss (across models).

families all with fractionally integrated and integrated GARCH or high order GARCH models for the conditional variances, with integrated models showing the best sample performances. When the evaluation is based on the L_3 loss function, the MCS contains 20 models. The MCS is in fact dominated by the orthogonal family of MGARCH, which scores the best sample performances. In line with the previous results, it includes also other specifications, all of which in the DECO family, which allow for long memory and integrated conditional variances.

It is worth noting that the results in terms of MCS are specific to the sample period (and the set of candidate models). As described in Section 3, the sample considered is characterized by dramatic changes in volatility dynamics, favoring long memory type models. Furthermore, relatively large average sample performances though close across models indicate that either all models under comparison fail in predicting accurately the conditional variance, i.e. the MCS is overall uninformative, or that this feature refers only to particular periods of time. In the next sections, MCS results are presented for three sub-samples. The aim is to verify to what extent different levels of market instability affect the forecasting performance of the models and the ability of the MCS procedure to separate between superior and inferior models.

4.2 Dot-com speculative bubble burst and aftermath

The MCS results are reported in Table 4 for the Euclidean (L_E), Stein (L_S) and L_3 loss functions. The MCS under L_E contains 38 models. As expected, there are differences with the MCS obtained for the full sample. First, modelling directly the conditional correlation and accounting for the leverage effect in the conditional variances becomes more important. To be precise, DCC type models with Egarch conditional variances dominate the MCS and show the smallest losses. This result is also confirmed by the fact that the MCS also contains two CCC specifications, both with Egarch dynamics for the conditional variances, which suggests that adequately modelling asymmetry in the conditional variances can in some cases compensate the restrictive assumption of no dynamics in the conditional correlation. Furthermore, the exclusion of other specifications that also specifically account for asymmetry/leverage in the variance, i.e. DCC type models with Aparch and Gjr dynamics for the conditional variances, suggests that the choice of the specific parametrization becomes important. Finally, as expected the relative importance of accounting for a (fractionally) integrated variance process, although still present, becomes less noticeable. In this case, we find only 4 specifications (out of the 38 models in the MCS) which allow for long memory and integrated conditional variances (against 10 out of 39 for the full sample).

The Stein loss function delivers a small MCS. The MCS consists of 2 models, namely the DCCE and the DCCT with integrated GARCH conditional variances. Although the MCS does not overlap with the one found under the symmetric loss function it is clear that when overweighting underpredictions the focus centers on the long memory properties of the conditional variance process. Table 4 also reports the best 10 models ordered in terms of

Table 4: MCS on first sub-sample. Dot-com bubble burst (99/01/04 - 03/03/31)

Euclidean distance (38 models)

MCS($\alpha = 25\%$)		Rnk	$\bar{L}_i$	T_D	p-val	VR	Corr
CCC	Egarch (0,1)	27	2.821	0.985	0.37	1.031	0.999
	(1,1)	41	2.844	1.170	0.29	1.150	0.996
DCCA	Egarch (0,1)	6	2.776	0.335	0.83	0.988	0.999
	(0,2)	18	2.801	0.588	0.65	1.030	0.999
	(1,1)	20	2.806	0.510	0.68	1.117	0.997
	(1,2)	17	2.799	0.545	0.66	1.012	0.999
	Figarch (1,1)	22	2.810	0.372	0.79	0.820	0.989
DCCT	Egarch (0,1)	23	2.811	0.658	0.57	1.026	0.999
	(1,1)	31	2.834	0.779	0.49	1.146	0.996
	Figarch (1,1)	44	2.849	0.839	0.45	0.855	0.989
DCCE	Egarch (0,1)	4	2.769	0.226	0.84	1.011	0.999
	(0,2)	13	2.794	0.404	0.77	1.052	0.999
	(1,1)	19	2.804	0.430	0.75	1.127	0.997
	(1,2)	10	2.783	0.331	0.83	1.019	0.999
	(2,2)	33	2.837	1.028	0.35	1.118	0.997
	Figarch (1,1)	14	2.796	0.343	0.83	0.832	0.990
	Gjr (2,1)	39	2.841	1.242	0.26	0.967	0.994
DECO	Egarch (0,1)	1	2.751	-	1.00	0.948	0.999
	(0,2)	7	2.776	0.290	0.83	0.991	0.999
	(1,1)	5	2.775	0.281	0.84	1.066	0.998
	(1,2)	2	2.760	0.322	0.88	0.961	0.999
	(2,1)	30	2.832	0.721	0.53	1.136	0.996
	(2,2)	21	2.807	0.605	0.62	1.055	0.998
	Figarch (1,1)	26	2.818	0.470	0.71	0.779	0.985
	Gjr (1,1)	43	2.848	1.125	0.30	0.875	0.993
	(2,1)	37	2.838	0.934	0.40	0.900	0.994
Orth.	Aparch (1,1)	3	2.764	0.089	0.88	0.976	0.992
	Egarch (0,1)	12	2.789	0.303	0.83	1.047	0.994
	(0,2)	16	2.797	0.364	0.79	1.083	0.996
	(1,1)	29	2.831	0.847	0.45	1.133	0.997
	(1,2)	25	2.817	0.604	0.62	1.099	0.996
	(2,2)	34	2.837	0.983	0.37	1.135	0.996
	Garch (2,1)	35	2.837	0.723	0.53	1.052	0.991
	(2,2)	24	2.815	0.567	0.65	1.044	0.993
	Gjr (1,1)	8	2.779	0.242	0.83	0.926	0.991
	(1,2)	9	2.780	0.256	0.83	0.933	0.992
	(2,1)	15	2.797	0.392	0.77	0.995	0.995
	(2,2)	11	2.785	0.280	0.83	0.991	0.994

Stein distance (2 models)

MCS($\alpha = 25\%$)		Rnk	$\bar{L}_i$	T_D	p-val	VR	Corr
DCCE	Igarch (1,1)	1	3.268	-	1.00	0.999	0.999
DCCT	Igarch (1,1)	2	3.274	1.212	0.27	1.003	1.000
CCC	*Igarch (1,1)*	*3*	*3.283*	-	-	-	
DCCA	*Igarch (1,1)*	*4*	*3.293*	-	-	-	
DCCE	*Figarch (1,1)*	*5*	*3.439*	-	-	-	
DCCT	*Figarch (1,1)*	*6*	*3.444*	-	-	-	
DCCE	*Hgarch (1,1)*	*7*	*3.446*	-	-	-	
DCCT	*Hgarch (1,1)*	*8*	*3.454*	-	-	-	
DCCE	*Rm (1,1)*	*9*	*3.455*	-	-	-	
DCCE	*Egarch (1,2)*	*10*	*3.456*	-	-	-	

L_3 loss function (11 models)

MCS($\alpha = 25\%$)		Rnk	$\bar{L}_i$	T_D	p-val	VR	Corr
Orth.	Aparch (1,1)	1	16.394	-	1.00	0.918	0.999
	Egarch (0,1)	2	16.568	0.887	0.47	0.983	0.999
	(0,2)	3	16.664	0.688	0.47	1.031	1.000
	(1,1)	9	17.035	1.192	0.27	1.117	0.999
	(1,2)	7	16.918	0.996	0.33	1.082	0.999
	(2,2)	11	17.086	1.353	0.27	1.121	0.998
	Garch (2,2)	13	17.235	1.235	0.27	1.007	0.991
	Gjr (1,1)	4	16.733	1.255	0.33	0.876	0.998
	(1,2)	5	16.737	2.285	0.33	0.891	0.999
	(2,1)	8	17.012	1.394	0.27	0.998	0.998
	(2,2)	6	16.797	1.288	0.33	0.985	0.998

Notes. See Table 3.

sample performances. Even though statistically inferior, it is worth noting that the top of the classification is indoubitably dominated by models that account for this feature. On the other hand, the MCS under the L_3 loss function includes 8 models, all from the orthogonal GARCH family. Most models account for asymmetry in the variance processes of the components.

4.3 Calm markets

Results for the MCS for the second sub-sample are reported in Tables 5 and 6 for the Euclidean (L_E) and Stein (L_S), and the L_3 loss functions respectively. With the exception of the Stein loss function, the MCS obtained for this sample is the largest. This is not surprising because this period is characterized by relatively small and slow moving volatility. It is therefore reasonable to expect most of the MGARCH models under comparison are adequate to fit the dynamics of the conditional variance. In fact, if we look at the average sample performances over this period, they get close to zero showing a dramatic improvement over the evaluation based on full sample.

The MCS under L_E contains 74 models, about half of the models considered, and includes specifications from all the families of MGARCH models. As a general result, we can say that the over this period the data does not show evidence of dynamics in the correlation process or asymmetry/leverage or long memory in the conditional variance. However, when we look at the composition of the MCS, we can draw the following three conclusions. First, DECO type models are excluded from the set of superior models with the exception of DECO-Aparch and DECO-Rm. However, looking at the decomposition of the variance (columns 7 and 8) together with their ranking position, in both cases the information content of these models is doubtful. Both models show a relatively small correlation with the average performance of the other models. The same remark holds for the DCC type specifications with Risk-Metrics conditional variances. Second, a similar conclusion can be drawn for the orthogonal specifications. Although only Orth.-Gjr(p, q) models are statistically inferior, the remaining orthogonal specifications show the highest relative variance and smallest correlation with the average loss. Hence, it is possible that the orthogonal models end up in the MCS because the data does not contain sufficient information to infer that these models are inferior within the MCS. Third, the same remark holds for CCC/DCC type models with Riskmetrics and Gjr(p, q) ($p = 1$ and $q = 1, 2$) conditional variances. In particular, CCC/DCC-Gjr models show the poorest sample performances within the MCS, the largest relative variance (in average 25% larger than $Var(\bar{L})$) and the smallest correlation with $\bar{L}$.

We consider now the two asymmetric loss functions. Under L_S, the MCS contains 12 models. In line with previous results, the MCS shows no evidence of particular features in the variance process as dynamics in the correlation process or asymmetry/leverage or long memory in the conditional variance. The set of superior models is dominated by specifications within the conditional correlations family, namely CCC, DCCT and DCCE, with GARCH conditional variances, therefore the hypothesis of constant conditional correlation is difficult to reject. The MCS also includes two asymmetric specifications, i.e. DCCE-Gjr(1,1) and DCCT-Gjr(1,1), although both characterized by weaker sample performances within the MCS. Finally, under L_3, we obtain results similar to L_E both for the size and composition of the

Table 5: MCS on second sub-sample. Calm period (03/04/01 - 07/07/31)

Euclidean distance (74 models)

MCS(α = 25%)		Rnk	$\bar{L}_i$	T_D	p-val	VR	Corr
CCC	Aparch (1,1)	2	0.328	6.224	0.73	0.884	0.969
	Egarch (0,1)	9	0.345	0.695	0.73	0.975	0.997
	(0,2)	33	0.348	0.456	0.73	1.042	0.992
	(1,1)	18	0.346	0.895	0.73	1.098	0.982
	(1,2)	28	0.347	0.771	0.73	1.061	0.980
	(2,1)	7	0.344	0.967	0.73	1.082	0.987
	Figarch (1,1)	25	0.347	0.447	0.73	0.992	0.996
	Garch (1,1)	50	0.350	0.467	0.67	1.009	0.997
	(1,2)	46	0.350	0.446	0.68	1.021	0.997
	(2,1)	26	0.347	0.439	0.73	1.014	0.997
	(2,2)	11	0.345	0.612	0.73	0.982	0.998
	Gjr (1,1)	91	0.374	0.923	0.37	1.237	0.957
	(1,2)	85	0.372	0.619	0.54	1.260	0.961
	Hgarch (1,1)	55	0.351	0.454	0.68	0.940	0.995
	Rm (1,1)	65	0.356	0.507	0.63	0.990	0.967
DCCA	Aparch (1,1)	4	0.329	3.590	0.73	0.884	0.970
	Egarch (0,1)	20	0.346	0.497	0.73	0.977	0.997
	(0,2)	40	0.349	0.422	0.71	1.044	0.991
	(1,1)	32	0.348	0.626	0.73	1.101	0.981
	(1,2)	38	0.349	0.517	0.73	1.064	0.980
	(2,1)	16	0.346	0.869	0.73	1.084	0.986
	Figarch (1,1)	30	0.347	0.445	0.73	0.992	0.996
	Garch (1,1)	56	0.351	0.497	0.64	1.010	0.997
	(1,2)	53	0.351	0.488	0.65	1.021	0.997
	(2,1)	35	0.348	0.437	0.68	1.015	0.997
	(2,2)	19	0.346	0.504	0.73	0.983	0.998
	Gjr (1,1)	93	0.374	1.134	0.27	1.238	0.957
	(1,2)	89	0.373	0.830	0.41	1.262	0.960
	Hgarch (1,1)	49	0.350	0.439	0.69	0.942	0.995
	Rm (1,1)	64	0.356	0.482	0.65	0.989	0.967
DCCT	Aparch (1,1)	1	0.328	-	1.00	0.884	0.970
	Egarch (0,1)	8	0.345	0.710	0.73	0.975	0.997
	(0,2)	31	0.348	0.471	0.73	1.042	0.991
	(1,1)	17	0.346	1.031	0.73	1.098	0.982
	(1,2)	29	0.347	0.723	0.73	1.061	0.980
	(2,1)	6	0.344	0.959	0.73	1.082	0.987
	Figarch (1,1)	22	0.347	0.490	0.73	0.991	0.997
	Garch (1,1)	48	0.350	0.442	0.67	1.009	0.997
	(1,2)	39	0.349	0.439	0.67	1.021	0.997
	(2,1)	23	0.347	0.452	0.73	1.013	0.997
	(2,2)	10	0.345	0.678	0.73	0.982	0.998
	Gjr (1,1)	88	0.373	0.749	0.46	1.237	0.957
	(1,2)	82	0.372	0.570	0.57	1.261	0.960
	Hgarch (1,1)	43	0.350	0.440	0.67	0.940	0.995
	Rm (1,1)	5	0.340	1.288	0.73	0.957	0.971
DCCE	Aparch (1,1)	3	0.329	3.631	0.73	0.884	0.970
	Egarch (0,1)	15	0.346	0.598	0.73	0.977	0.997
	(0,2)	36	0.349	0.427	0.73	1.045	0.991
	(1,1)	24	0.347	0.813	0.73	1.101	0.981
	(1,2)	34	0.348	0.546	0.73	1.064	0.980
	(2,1)	12	0.345	0.943	0.73	1.084	0.986
	Figarch (1,1)	21	0.347	0.509	0.73	0.992	0.996
	Garch (1,1)	51	0.350	0.472	0.67	1.010	0.997
	(1,2)	47	0.350	0.450	0.68	1.022	0.997
	(2,1)	27	0.347	0.423	0.73	1.013	0.997
	(2,2)	13	0.345	0.587	0.73	0.984	0.998
	Gjr (1,1)	92	0.374	1.023	0.32	1.239	0.957
	(1,2)	86	0.373	0.678	0.50	1.262	0.960
	Hgarch (1,1)	42	0.349	0.434	0.67	0.942	0.995
	Rm (1,1)	63	0.355	0.461	0.67	0.989	0.967
DECO	Aparch (1,1)	14	0.346	0.956	0.73	0.902	0.970
	Rm (1,1)	45	0.350	0.459	0.73	0.974	0.973
Orth.	Aparch (1,1)	37	0.349	0.839	0.73	1.088	0.960
	Egarch (0,1)	44	0.350	0.611	0.73	1.095	0.960
	(0,2)	54	0.351	0.499	0.73	1.091	0.960
	(1,1)	57	0.351	0.450	0.73	1.097	0.960
	(1,2)	41	0.349	0.712	0.73	1.096	0.960
	(2,1)	60	0.352	0.432	0.71	1.087	0.961
	(2,2)	59	0.352	0.430	0.69	1.092	0.963
	Garch (1,1)	58	0.352	0.425	0.72	1.087	0.961
	(1,2)	61	0.352	0.441	0.67	1.090	0.960
	(2,1)	52	0.351	0.550	0.73	1.088	0.961
	(2,2)	62	0.353	0.443	0.67	1.086	0.962
SBEKK	(1,1)	67	0.363	0.534	0.60	0.955	0.952

L_3 loss function (74 models)

MCS(α = 25%)		Rnk	$\bar{L}_i$	T_D	p-val	VR	Corr
CCC	Aparch (1,1)	2	0.631	1.090	0.49	0.792	0.910
	Egarch (0,1)	22	0.718	3.320	0.49	1.005	0.997
	(0,2)	52	0.756	1.586	0.49	1.391	0.961
	(1,1)	63	0.784	0.882	0.49	1.777	0.918
	(1,2)	61	0.783	1.015	0.49	1.745	0.917
	(2,1)	57	0.772	1.327	0.49	1.615	0.936
	Figarch (1,1)	31	0.731	0.731	0.49	1.025	0.998
	Garch (1,1)	34	0.732	0.725	0.48	0.998	0.997
	(1,2)	40	0.739	0.752	0.44	1.050	0.998
	(2,1)	42	0.740	0.731	0.46	1.081	0.996
	(2,2)	23	0.718	0.873	0.49	0.980	0.999
	Gjr (1,2)	95	0.876	0.947	0.34	2.098	0.897
	Hgarch (1,1)	49	0.747	0.825	0.39	1.043	0.996
	Igarch (1,1)	93	0.874	0.852	0.38	1.352	0.791
	Rm (1,1)	15	0.674	4.335	0.49	0.823	0.912
DCCA	Aparch (1,1)	4	0.638	6.954	0.49	0.790	0.910
	Egarch (0,1)	29	0.727	0.840	0.49	1.009	0.997
	(0,2)	55	0.767	0.723	0.49	1.407	0.958
	(1,1)	71	0.794	0.748	0.49	1.801	0.915
	(1,2)	70	0.794	0.730	0.49	1.770	0.913
	(2,1)	60	0.781	0.787	0.49	1.635	0.934
	Figarch (1,1)	37	0.737	0.741	0.45	1.031	0.998
	Garch (1,1)	38	0.738	0.767	0.43	1.000	0.997
	(1,2)	47	0.745	0.859	0.38	1.054	0.998
	(2,1)	50	0.748	0.811	0.40	1.086	0.996
	(2,2)	28	0.727	0.736	0.49	0.983	0.999
	Hgarch (1,1)	51	0.749	0.840	0.38	1.051	0.995
	Igarch (1,2)	97	0.882	1.097	0.26	1.345	0.792
	Rm (1,1)	18	0.677	3.386	0.49	0.818	0.914
DCCT	Aparch (1,1)	1	0.631	-	1.00	0.791	0.910
	Egarch (0,1)	24	0.718	4.498	0.49	1.007	0.997
	(0,2)	53	0.757	1.150	0.49	1.398	0.960
	(1,1)	64	0.784	0.934	0.49	1.785	0.916
	(1,2)	62	0.784	0.880	0.49	1.754	0.915
	(2,1)	58	0.773	2.041	0.49	1.622	0.935
	Figarch (1,1)	30	0.731	0.749	0.49	1.027	0.998
	Garch (1,1)	32	0.732	0.733	0.49	0.999	0.997
	(1,2)	39	0.738	0.745	0.44	1.052	0.998
	(2,1)	41	0.740	0.732	0.47	1.083	0.996
	(2,2)	25	0.719	0.872	0.49	0.981	0.999
	Gjr (1,2)	94	0.876	0.894	0.36	2.107	0.897
	Hgarch (1,1)	45	0.744	0.775	0.42	1.047	0.996
	Igarch (1,1)	92	0.873	0.844	0.38	1.349	0.791
	Rm (1,1)	5	0.653	7.324	0.49	0.814	0.914
DCCE	Aparch (1,1)	3	0.636	6.797	0.49	0.790	0.910
	Egarch (0,1)	27	0.724	0.877	0.49	1.010	0.997
	(0,2)	54	0.765	0.769	0.49	1.408	0.958
	(1,1)	68	0.792	0.814	0.49	1.801	0.915
	(1,2)	67	0.791	0.810	0.49	1.771	0.913
	(2,1)	59	0.779	0.843	0.49	1.635	0.934
	Figarch (1,1)	35	0.733	0.728	0.47	1.029	0.998
	Garch (1,1)	36	0.736	0.736	0.46	1.000	0.997
	(1,2)	44	0.743	0.799	0.40	1.054	0.998
	(2,1)	43	0.743	0.759	0.43	1.074	0.996
	(2,2)	26	0.724	0.786	0.49	0.983	0.999
	Gjr (1,2)	96	0.881	1.003	0.31	2.119	0.895
	Hgarch (1,1)	48	0.746	0.786	0.42	1.051	0.996
	Rm (1,1)	12	0.671	4.628	0.49	0.819	0.913
DECO	Aparch (1,1)	21	0.694	1.725	0.49	0.815	0.910
	Rm (1,1)	20	0.686	2.173	0.49	0.827	0.918
Orth.	Aparch (1,1)	6	0.661	6.560	0.49	0.874	0.900
	Egarch (0,1)	8	0.666	5.965	0.49	0.880	0.899
	(0,2)	9	0.666	5.685	0.49	0.867	0.900
	(1,1)	16	0.676	4.113	0.49	0.887	0.899
	(1,2)	14	0.673	4.443	0.49	0.885	0.899
	(2,1)	17	0.676	3.965	0.49	0.879	0.899
	(2,2)	19	0.679	3.378	0.49	0.883	0.903
	Garch (1,1)	7	0.665	5.720	0.49	0.877	0.900
	(1,2)	11	0.670	4.466	0.49	0.880	0.898
	(2,1)	10	0.668	5.554	0.49	0.881	0.899
	(2,2)	13	0.672	4.258	0.49	0.882	0.900
DBEKK	(1,1)	46	0.745	0.841	0.49	0.820	0.898
SBEKK	(1,1)	33	0.732	0.93	0.486	0.837	0.891
RM	(1,1)	56	0.772	0.73	0.466	0.879	0.913

Notes. See Table 3.

MCS. However, although over this sample the type of asymmetry accounted for by L_3 is not statistically relevant, i.e., does not impact on the composition of the MCS, we observe changes in the ordering of the models. For example, the Orthogonal type models included in both MCSs, while ranking between 37th and 62nd under L_E, figure between the 6th and the 19th position of the overall ranking under L_3. Given the asymmetry of L_3, we can deduce that Orthogonal models tend to underestimate the conditional variance. The only differences in terms of MCS with the outcome obtained under the symmetric loss functions are: i) the inclusion of DCC type specifications with integrated conditional variances, which, however, appear to be quite uninformative since they show very poor sample performances (within the MCS) and show among the largest relative variances and the smallest correlations with the average loss; ii) the inclusion of all BEKK type models.

Table 6: MCS-second sub-sample. Calm period (03/04/01 - 07/07/31) (Cont.)

Stein distance (12 models)							
MCS(α = 25%)		Rnk	$\bar{L}_i$	T_D	p-val	VR	Corr
CCC	Garch (1,1)	5	3.180	0.285	0.72	0.948	0.999
	(1,2)	10	3.193	1.253	0.26	1.168	0.996
	(2,1)	3	3.175	0.476	0.74	1.033	0.998
DCCT	Garch (1,1)	6	3.183	0.413	0.61	0.935	0.999
	(1,2)	8	3.191	0.683	0.47	1.154	0.996
	(2,1)	2	3.174	0.265	0.74	1.022	0.998
	(2,2)	7	3.189	1.265	0.29	1.027	0.998
	Gjr (1,1)	16	3.203	1.171	0.26	0.806	0.982
DCCE	Garch (1,1)	4	3.179	0.307	0.74	0.967	0.998
	(1,2)	12	3.194	1.101	0.30	1.198	0.996
	(2,1)	1	3.171	-	1.00	1.065	0.998
	Gjr (1,1)	15	3.201	1.084	0.29	0.834	0.982

Notes. See Table 3.

4.4 2007-08 financial crisis

Results for the MCS for the last sub-sample are reported in Table 7 for the Euclidean (L_E), Stein (L_S) and L_3 loss functions. The MCS under L_E contains 39 models which is in line with the results obtained for full sample. The MCS is dominated by specifications in the DECO and the orthogonal GARCH families, while other DCC type specifications are selected only when they account for long memory and integrated conditional variances. Indeed, with respect to the full sample (and in sharp contrast with the Dot-com speculative bubble burst period) modelling long memory and integrated conditional variances becomes more important. On the other hand, although we find in the MCS models that account for asymmetry/leverage in the conditional variance of the returns, models with exponential GARCH dynamics for the

Table 7: MCS-third sub-sample: 2007-2008 financial crisis (07/08/01 - 08/12/26)

Euclidean distance (39 models)

MCS($\alpha = 25\%$)		Rnk	$\bar{L}_i$	T_D	p-val	VR	Corr
CCC	Hgarch (1,1)	40	17.172	1.034	0.32	1.171	0.995
DCCA	Figarch (1,1)	28	16.345	0.880	0.39	1.099	0.997
	Hgarch (1,1)	21	16.162	0.678	0.50	1.072	0.997
	Rm (1,1)	35	16.954	0.892	0.38	1.264	0.998
DCCT	Figarch (1,1)	43	17.283	1.207	0.25	1.184	0.995
	Hgarch (1,1)	38	17.086	0.992	0.33	1.154	0.995
DCCE	Figarch (1,1)	25	16.305	0.826	0.42	1.097	0.997
	Hgarch (1,1)	22	16.208	0.797	0.44	1.076	0.997
	Rm (1,1)	44	17.376	1.157	0.27	1.307	0.999
DECO	Aparch (1,1)	27	16.317	0.886	0.39	1.122	0.997
	Figarch (1,1)	5	14.919	0.063	0.90	0.922	0.998
	Garch (1,1)	32	16.661	0.884	0.39	1.187	0.997
	(1,2)	29	16.492	0.887	0.39	1.153	0.998
	(2,1)	31	16.583	0.938	0.36	1.141	0.999
	(2,2)	33	16.713	0.962	0.34	1.175	0.999
	Gjr (1,1)	23	16.237	0.828	0.42	1.104	0.998
	(1,2)	16	16.043	0.787	0.44	1.058	0.999
	(2,1)	14	15.879	0.780	0.44	1.001	0.999
	(2,2)	17	16.048	0.892	0.39	1.048	0.999
	Hgarch (1,1)	2	14.816	0.061	0.90	0.899	0.997
	Igarch (1,1)	24	16.275	0.808	0.44	1.071	0.992
	Rm (1,1)	19	16.076	0.444	0.68	1.132	0.998
Orth.	Aparch (1,1)	13	15.791	0.596	0.58	0.918	0.996
	Egarch (0,1)	26	16.308	0.914	0.38	1.020	0.998
	(0,2)	15	16.026	0.890	0.39	0.942	0.999
	(1,1)	20	16.088	0.881	0.39	0.928	0.998
	(1,2)	12	15.757	0.891	0.44	0.868	0.998
	(2,1)	30	16.562	1.068	0.30	1.067	0.999
	(2,2)	6	15.316	0.282	0.79	0.784	0.996
	Arch (2)	71	18.210	1.128	0.28	1.101	0.971
	Garch (1,1)	18	16.052	0.869	0.39	0.963	0.997
	(1,2)	9	15.618	0.827	0.44	0.867	0.998
	(2,1)	10	15.644	0.814	0.44	0.884	0.999
	(2,2)	11	15.666	0.874	0.44	0.861	0.998
	Gjr (1,1)	7	15.391	0.405	0.71	0.812	0.996
	(1,2)	3	14.853	0.120	0.90	0.705	0.997
	(2,1)	1	14.577	-	1.00	0.660	0.996
	(2,2)	4	14.895	0.070	0.90	0.720	0.997
RM	(1,1)	8	15.464	0.153	0.86	0.973	0.992

Stein distance (26 models)

MCS($\alpha = 25\%$)		Rnk	$\bar{L}_i$	T_D	p-val	VR	Corr
CCC	Aparch (1,1)	21	4.773	0.992	0.32	1.098	0.990
	Egarch (0,1)	14	4.712	0.579	0.46	0.991	0.986
	(0,2)	16	4.716	0.569	0.46	1.006	0.985
	(1,2)	10	4.665	0.587	0.48	0.954	0.990
	Figarch (1,1)	2	4.531	3.442	0.48	0.781	0.942
	Hgarch (1,1)	9	4.663	0.623	0.47	0.784	0.931
DCCA	Aparch (1,1)	30	4.843	1.099	0.29	1.417	0.991
	Egarch (0,1)	20	4.766	0.626	0.44	1.286	0.987
	(0,2)	23	4.787	0.678	0.42	1.313	0.984
	(1,2)	17	4.722	0.586	0.46	1.229	0.991
	Figarch (1,1)	6	4.585	1.143	0.48	0.959	0.939
	Hgarch (1,1)	8	4.631	0.684	0.48	0.861	0.930
DCCT	Aparch (1,1)	19	4.758	0.814	0.37	1.145	0.992
	Egarch (0,1)	11	4.669	0.550	0.48	1.031	0.989
	(0,2)	13	4.678	0.534	0.48	1.048	0.987
	(1,2)	7	4.623	0.636	0.48	0.995	0.993
	Figarch (1,1)	1	4.511	-	1.00	0.816	0.940
	Gjr (1,2)	24	4.802	1.192	0.26	1.214	0.990
	Hgarch (1,1)	4	4.566	4.693	0.48	0.737	0.931
DCCE	Aparch (1,1)	22	4.787	0.743	0.40	1.337	0.991
	Egarch (0,1)	15	4.714	0.578	0.47	1.203	0.987
	(0,2)	18	4.727	0.562	0.47	1.228	0.984
	(1,2)	12	4.671	0.635	0.48	1.151	0.991
	Figarch (1,1)	3	4.543	2.103	0.48	0.927	0.939
	Garch (2,1)	28	4.834	0.903	0.34	1.306	0.981
	Hgarch (1,1)	5	4.578	0.897	0.48	0.824	0.934

L_3 loss function (26 models)

MCS($\alpha = 25\%$)		Rnk	$\bar{L}_i$	T_D	p-val	VR	Corr
DECO	Aparch (1,1)	26	682.5	1.116	0.28	1.128	1.000
	Figarch (1,1)	17	660.0	0.769	0.48	1.062	1.000
	Garch (1,1)	29	687.4	1.189	0.26	1.147	1.000
	Gjr (1,1)	24	680.2	1.076	0.30	1.121	1.000
	(1,2)	22	677.6	1.152	0.27	1.105	1.000
	Hgarch (1,1)	15	656.9	0.820	0.48	1.052	1.000
	Igarch (1,1)	21	675.9	0.896	0.38	1.108	0.999
	Rm (1,1)	25	681.9	1.025	0.32	1.136	1.000
Orth.	Aparch (1,1)	10	641.6	0.795	0.48	0.991	0.999
	Egarch (0,1)	16	658.6	0.782	0.47	1.050	1.000
	(0,2)	11	648.8	0.787	0.48	1.001	1.000
	(1,1)	12	650.5	0.808	0.44	0.986	1.000
	(1,2)	8	637.2	1.057	0.48	0.949	0.999
	(2,1)	18	664.1	0.841	0.42	1.072	1.000
	(2,2)	4	617.4	0.667	0.49	0.871	0.999
	Arch (1)	19	665.5	0.962	0.48	0.890	0.989
	(2)	27	684.5	0.952	0.35	1.013	0.994
	Garch (1,1)	13	652.6	0.797	0.48	1.021	0.999
	(1,2)	6	635.9	1.087	0.48	0.955	1.000
	(2,1)	7	637.1	0.993	0.48	0.968	1.000
	(2,2)	9	637.4	1.146	0.48	0.952	1.000
	Gjr (1,1)	5	631.2	0.980	0.48	0.924	0.999
	(1,2)	2	605.8	1.224	0.49	0.825	0.999
	(2,1)	1	590.6	-	1.00	0.776	0.999
	(2,2)	3	609.2	0.858	0.49	0.846	0.999
RM	(1,1)	14	654.4	0.886	0.48	1.044	0.998

Notes. See Table 3

conditional variances are detected as inferior and excluded from the MCS. Note that for the dot-com bubble period, we find the opposite result.

Under L_S the results are also (qualitatively) consistent with the ones obtained for the full sample, though the MCS gets larger (26 models). The models in the MCS all belong to the DCC family and account for long-memory in volatility (i.e., CCC, DCCE, DCCA and DCCT with hyperbolic and fractionally integrated GARCH dynamics for the variances) and/or leverage effect (note that Egarch models perform better than Aparch models, while Gjr models are mostly excluded from the MCS). The non rejection of some CCC specifications, which is surprising in this case, illustrates that adequately modelling the conditional variances of the returns can compensate the loss in forecasting accuracy induced by the restrictive assumption of constant conditional correlation.

For the second asymmetric loss function L_3 the results are also in line with the full sample. The MCS contains 26 models and is dominated by orthogonal and DECO specifications with the former showing the best sample performances. Among the DECO specifications included in the MCS we find both evidence of long memory and integrated conditional variances and of leverage effect when modelled with Aparch and Gjr dynamics. As in the MCS under L_E, we also find two orthogonal Arch specifications, but unlike in the previous case, there is no clear evidence that either of the two models is inferior within the MCS.

Finally, the average loss over the last sub-sample is much larger than in the first two periods (irrespectively of the choice of the loss function). We conclude first that in turbulent periods GARCH models do not seem to be well suited to adequately estimate the conditional variance. Second, the large losses accumulated over short periods of high instability tend to drive the MCS results even when long forecasting periods are considered. In fact, there is a trade off between the forecast sample length (to reduce sampling variability) and the informativeness of the selection.

4.5 Robustness check to the use of alternative proxies

To verify the robustness of our results to the choice of the volatility proxy, we repeat the analysis using $RCov$, see (11), computed using 1 and 15 minute returns and $RCovAC_{q=1}$, see (13), computed using 1, 5 and 15 minutes returns. The results in terms of MCS are robust in terms of size and composition to the alternative volatility proxies. In particular, when the proxy is based on higher frequency returns, i.e., $RCov$ and $RCovAC_{q=1}$ based on one minute returns, we generally find smaller MCS. The use of a higher frequency proxy ensures the elimination of uninformative models. As an example (complete results are available upon request), if we consider the Euclidean distance (L_E), under $RCov^{(1min)}$ ($RCovAC_{q=1}^{(1min)}$) we find 25 (35) models for the full sample, 26 (33) for the dot-com bubble burst period, 60 (71) for the calm period and 47 (38) for the 2007-2008 financial crisis sub-sample. In accordance with the literature, the robustness of these results is implied by the consistency of the loss function. The higher accuracy of the proxy only translates into a lower variability of the sample evaluation of the models which makes easier to effectively discriminate between models. Along the same line, and consistently with the results obtained under $RCov^{(5min)}$, when the evaluation is based on $RCovAC_{q=1}^{(5min)}$ and L_E we find 40 models for the full sample, and 30, 71 and 38 for the three sub-samples respectively. Finally, when we use proxies based on 15 minutes returns we find 40 (41) models for the full sample and 39 (50), 73 (68) and 37 (37) for the three sub-samples respectively.

5 Setting a benchmark: the predictive ability of the DCCE

In this section, we focus on the predictive ability of a predefined benchmark model with respect to all other models. As benchmarks we choose simple and parsimonious specifications and take into account two dimensions: the assumption on the multivariate structure (CCC, DCCE and Orthogonal) and on the dynamics of variance of the marginal processes/principal components (Garch(1,1) and Egarch(0,1)). The CCC-Garch(1,1) model represents the simplest alternative and allows to test simple hypotheses such as constant correlation and symmetric variances for the marginal processes. The choice of the DCCE among the DCC specification introduced in Section 2.1 is not coincidental: this model has been increasingly popular because of its flexibility and straightforward interpretation. The DCCE-Garch(1,1) therefore serves as a benchmark to assess whether relaxing the assumption of constant correlation is sufficient to improve predictive ability. Finally, the Orthogonal-Garch(1,1) model represents a simple and parsimonious alternative to direct modelling of the dynamics of the conditional covariance and correlation. In a univariate setting, Hansen and Lunde (2005) suggest that the absence of leverage effect is likely to be rejected on stock market returns. To validate this result in the multivariate framework, we also couple the three multivariate models with the Egarch(0,1) specifications for the conditional variance processes.

The predictive ability of our benchmarks is evaluated using the test for superior predictive ability (SPA) proposed by Hansen (2005). Using the notation introduced in Section 2.4, let us define $d_{0j,t} = L_{0,t} - L_{j,t}$, $j = 1, ..., M$, the relative performance of model j with respect to the benchmark model (indexed by 0). Under reasonable assumptions $\lambda_j = E[d_{0j,t}]$ is well defined. The null hypothesis is expressed with respect to the best alternative model, i.e. $H_{0,M} : \max_{j \in M} \frac{\lambda_j}{\omega_j} \leq 0$, where ω_j^2 denotes the asymptotic variance of λ_j. The corresponding test statistic is $\sqrt{T^*}\left[\max_{j \in M} \frac{\bar{d}_{0j}}{\hat{\omega}_j}\right]$ where $\bar{d}_{0j} = T^{*-1} \sum_{t=1}^{T^*} d_{0j,t}$ is the sample loss differential between the benchmark and model j. P-values for the test are obtained by bootstrap.

The results for the six different benchmarks are reported in Tables 8 and 9. Consistently

Table 8: SPA test: symmetric variance

Benchmark 1: CCC-Garch(1,1)	L_E			L_S			L_3		
	p_L	p_C	p_U	p_L	p_C	p_U	p_L	p_C	p_U
Full sample	0.003	0.003	0.003	0.000	0.000	0.000	0.026	0.027	0.027
Dot-com bubble	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Calm period	0.018	0.020	0.023	0.434	**0.817**	0.963	0.170	**0.211**	0.259
07-08 financial crisis	0.015	0.016	0.016	0.000	0.000	0.000	0.019	0.019	0.019
Benchmark 2: DCCE-Garch(1,1)	L_E			L_S			L_3		
	p_L	p_C	p_U	p_L	p_C	p_U	p_L	p_C	p_U
Full sample	0.061	0.064	0.067	0.000	0.000	0.000	0.095	0.098	0.101
Dot-com bubble	0.001	0.002	0.002	0.000	0.000	0.000	0.003	0.003	0.003
Calm period	0.108	**0.115**	0.170	0.384	**0.825**	0.982	0.092	**0.102**	0.141
07-08 financial crisis	0.023	0.024	0.024	0.008	0.009	0.009	0.037	0.038	0.038
Benchmark 3: Orth.-Garch(1,1)	L_E			L_S			L_3		
	p_L	p_C	p_U	p_L	p_C	p_U	p_L	p_C	p_U
Full sample	0.087	**0.118**	0.120	0.000	0.000	0.000	0.191	**0.276**	0.280
Dot-com bubble	0.070	0.081	0.090	0.000	0.000	0.000	0.031	0.034	0.037
Calm period	0.001	0.002	0.003	0.000	0.000	0.000	0.010	0.013	0.021
07-08 financial crisis	0.257	**0.321**	0.332	0.003	0.003	0.003	0.357	**0.488**	0.494

Note. p_C: consistent p-value, p_L and p_U upper and lower bound for the consistent p-value respectively. See Hansen (2005) for further details. Consistent p-values in bold indicate the non rejection of the null at confidence level $\alpha = 0.10$.

with the MCS results in Section 4, the hypothesis of constant correlation (Benchmark 1 and 4), as well as of symmetric dynamics for the variance matrix (Benchmark 2 and 5) is always rejected except when forecasts are compared over calm periods. However, the hypothesis of symmetric dynamics for the variances of the assets returns considered is rather weak. Evidence of the leverage effect is much stronger (e.g., Benchmark 5) when the comparison is taken over periods of market instability. Also, allowing for dynamic correlation significantly improves models' forecasting ability.

Table 9: SPA test: asymmetric variance

Benchmark 4: CCC-Egarch(0,1)									
	L_E			L_S			L_3		
	p_L	p_C	p_U	p_L	p_C	p_U	p_L	p_C	p_U
Full sample	0.014	0.016	0.016	0.000	0.000	0.000	0.043	0.043	0.044
Dot-com bubble	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Calm period	0.100	**0.164**	0.237	0.000	0.000	0.000	0.242	**0.423**	0.547
07-08 financial crisis	0.016	0.016	0.016	0.046	0.056	0.085	0.018	0.019	0.019
Benchmark 5: DCCE-Egarch(0,1)									
	L_E			L_S			L_3		
	p_L	p_C	p_U	p_L	p_C	p_U	p_L	p_C	p_U
Full sample	0.100	**0.115**	0.136	0.001	0.002	0.002	0.082	0.084	0.086
Dot-com bubble	0.403	**0.746**	0.909	0.000	0.000	0.000	0.080	0.091	0.131
Calm period	0.154	**0.227**	0.386	0.000	0.000	0.000	0.023	0.029	0.035
07-08 financial crisis	0.035	0.037	0.037	0.165	**0.235**	0.459	0.035	0.036	0.036
Benchmark 6: Orth.-Egarch(0,1)									
	L_E			L_S			L_3		
	p_L	p_C	p_U	p_L	p_C	p_U	p_L	p_C	p_U
Full sample	0.243	**0.372**	0.400	0.000	0.000	0.000	0.297	**0.524**	0.546
Dot-com bubble	0.341	**0.522**	0.597	0.000	0.000	0.000	0.217	**0.723**	0.838
Calm period	0.004	0.006	0.009	0.000	0.000	0.000	0.000	0.000	0.000
07-08 financial crisis	0.189	**0.220**	0.229	0.003	0.003	0.004	0.336	**0.489**	0.503

Notes. See Table 8.

With respect to the type of multivariate model, the Orthogonal approach (in particular with leverage) exhibits superior performance exclusively over turbulent periods while it is systematically outperformed over calm periods. As underlined in Section 4 the fact that this model is preferred under the L_3 criterion suggests that it is likely to underestimate the covariance matrix (Benchmark 3 and 6). In this application, the most valid specification is the DCCE-Egarch(0,1). It captures well the dynamics of the covariance matrix across the different samples. Its performances are not statistically worse than any of the 124 competing models, both when considering the full sample or any of the sub-samples. Note that for the 2007-08 financial crisis period the null is rejected under L_E but not under L_S, i.e. the DCCE-Egarch(0,1) possibly tends to overestimate the variance matrix during periods of extreme market instability.

6 Conclusion

Several multivariate GARCH models exist in the literature. However, from an applied viewpoint no guidelines are available on forecasting performances evaluation and model selection. We apply the model confidence set approach (MCS), which allows to isolate superior models in terms of predictice ability, to 125 multivariate GARCH model based forecasts. We consider 10 assets from NYSE and NASDAQ for which we forecast the conditional variance matrix from January 4, 1999 to December 26, 2008. The evaluation is based on two symmetric and two asymmetric loss functions and the ex-post underlying volatility is approximated by the realized covariance estimator based on intraday returns sampled at 5 minute frequency.

In line with recent literature, we find the Euclidean and Frobenius loss functions (both symmetric) to deliver relatively large MCS, from about one half to one fourth of the total number of models, while the two asymmetric loss functions identify sets of superior models systematically smaller. The MCS is composed of sophisticated specifications such as orthogonal and dynamic conditional correlation (DCC), both with long memory in the conditional variances With respect loss function choice we find that while Orthogonal and DECO models tend to underestimate the conditional covariance, the DCC of Engle (2002) (as well as its asymmetric version) and the DCC of Tse and Tsui (2002) tend to overestimate.

The model selection can be misleading when the forecast sample consists of periods characterized by different types of dynamics. We illustrate how sensitive the MCS is with respect to the forecast sample under investigation by considering not only the full sample but also by investigating sub-samples which are homogenous in their volatility dynamics. Over the dot-com bubble burst and aftermath period, the set of superior models is composed by rather sophisticated models such as DCC and Orthogonal, both with leverage effect in the conditional variances of returns and principal components, respectively. Over calm periods, a simple assumption like constant conditional correlation and symmetry in the conditional variances cannot be rejected. Finally, over the 2007-2008 financial crisis, accounting for non-stationarity in the conditional variance process significantly improves models' forecasting performances.

Focussing on the DCC class of models we can draw the following conclusions. First, the DECO model, which is estimated under the assumption of cross sectional equicorrelation, delivers superior forecasts over periods of market instability, but performs rather poorly during calm periods. Second, modeling the asymmetric response of shocks in the conditional correlation with a single parameter does not seem to significantly improve models' forecasting performances with respect to the standard DCC of Engle (2002). Third, when comparing the DCC of Engle (2002) with the DCC of Tse and Tsui (2002), we can conclude that, although statistically equivalent in terms of forecasting ability, while the first shows better sample performances over turbulent periods, the second performs better over calm periods. Fourth, we find that the most valid specification is represented by the DCC model of Engle (2002) when coupled with leverage effect in the conditional variances of the marginal processes. This model captures well the dynamics of the variance matrix consistently across the different sample periods. The latter result is confirmed by the Superior Predictive Ability (SPA) test. The null hypothesis that the DCC of Engle (2002) with exponential GARCH dynamics is not outperformed by the other 124 specifications cannot be rejected at standard critical levels.

This paper considers only one-step ahead forecasts of conditional variance matrices. It would be interesting to construct sets of superior models based on multiple step-ahead forecasts. Other issues like forecasting correlation matrices and high dimensional applications (hundreds of series) merit more attention.

References

Aielli, G. P. (2006): "Consistent Estimation of Large Scale Dynamic Conditional Correlations," Department of Statistics, University of Florence.

Alexander, C., and A. Chibumba (1997): "Multivariate Orthogonal Factor GARCH," University of Sussex, Mimeo.

Andersen, T., T. Bollerslev, and F. Diebold (2002): *Parametric and non Parametric Volatility Measurement*. Handbook of Financial Econometrics.

Andersen, T., T. Bollerslev, F. Diebold, and P. Labys (2003): "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579–625.

Andersen, T., T. Bollerslev, P. Frederiksen, and M. Nielsen (2010): "Continuous-time Models, Realized Volatilities and Testable Distributional Implications for Daily Stock Returns," *Journal of Applied Econometrics*, 25, 233261.

Baillie, R., T. Bollerslev, and H. Mikkelsen (1996): "Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 74, 3–30.

Bali, T. (2008): "The Intertemporal Relation between Expected Returns and Risk," *Journal of Financial Economics*, 87, 101–131.

Barndorff-Nielsen, O., P. Hansen, A. Lunde, and N. Shephard (2008a): "Designing Realized Kernels to Measure the ex post Variation of Equity Prices in the Presence of Noise," *Econometrica*, 76(6), 1481–1536.

——— (2008b): "Multivariate Realised Kernels: Consistent Positive Semi-Definite Estimators of the Covariation of Equity Prices with Noise and Non-Synchronous Trading," *DP, Oxford University*.

Barndorff-Nielsen, O., and N. Shephard (2004a): "Econometric Analysis of Realized Covariation: High Frequency Covariance, Regression and Correlation in Financial Economics," *Econometrica*, 72, 885–925.

——— (2004b): "Power and Bipower Variation with Stochastic Volatility and Jumps," *Journal of Financial Econometrics*, 2, 1–37.

Bauwens, L., S. Laurent, and J. Rombouts (2006): "Multivariate GARCH Models: A Survey," *Journal of Applied Econometrics*, 21, 79–109.

Becker, R., and A. Clements (2008): "Are Combination Forecasts of S&P 500 Volatility Statistically Superior?," *International Journal of Forecasting*, 24, 122–133.

Bollerslev, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

——— (1990): "Modeling the Coherence in Short-run Nominal Exchange Rates: A Multivariate Generalized ARCH model," *Review of Economics and Statistics*, 72, 498–505.

Bollerslev, T., R. Engle, and J. Wooldridge (1988): "A Capital Asset Pricing Mode1 with Time Varying Covariances," *Journal of Political Economy*, 96, 116–131.

Cappiello, L., R. Engle, and K. Sheppard (2006): "Asymmetric Dynamics in the Correlations of Global Equity and Bond Returns," *Journal of Financial Econometrics*, 4(4), 537–572.

Chiriac, R., and V. Voev (2010): "Modelling and Forecsting Multivariate Realized Volatility," *forthcoming in Journal of Applied Econometrics*.

Clark, T., and M. McCracken (2001): "Tests of Equal Forecast Accuracy and Encompassing for Nested Models," *Journal of Econometrics*, 105, 85–110.

Clark, T., and K. West (2006): "Using Out-of-sample Mean Squared Prediction Errors to test the Martingale Difference Hypothesis," *Journal of Econometrics*, 135, 155–186.

——— (2007): "Approximately Normal Tests for Equal Predictive Accuracy in Nested Models," *Journal of Econometrics*, 138, 291–311.

Clements, A., M. Doolan, S. Hurn, and R. Becker (2009): "Evaluating Multivariate Volatility Forecasts," *NCER Working Paper Series 41*.

Davidson, J. (2004): "Moment and Memory Properties of Linear Conditional Heteroscedasticity Models, and a New Model," *Journal of Business and Economic Statistics*, 22, 16–29.

Diebold, F., and R. Mariano (1995): "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13, 253–263.

Ding, Z., C. W. J. Granger, and R. F. Engle (1993): "A Long Memory Property of Stock Market Returns and a New Model," *Journal of Empirical Finance*, 1, 83–106.

Engle, R. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1008.

——— (2002): "Dynamic Conditional Correlation - a Simple Class of Multivariate GARCH Models," *Journal of Business and Economic Statistics*, 20, 339–350.

——— (2009): *Anticipating correlations*. Princeton Univ. Press.

Engle, R., and T. Bollerslev (1986): "Modelling the Persistence of Conditional Variances," *Econometric Reviews*, 5, 1–50.

ENGLE, R., AND R. COLACITO (2006): "Testing and Valuing Dynamic Correlations for Asset Allocation," *Journal of Business and Economic Statistics*, 24, 238–253.

ENGLE, R., AND G. GONZALEZ-RIVERA (1991): "Semiparametric ARCH Models," *Journal of Business and Economic Statistics*, 9(4), 345–59.

ENGLE, R., AND B. KELLY (2008): "Dynamic Equicorrelation," Stern School of Business.

ENGLE, R., AND F. KRONER (1995): "Multivariate Simultaneous Generalized ARCH," *Econometric Theory*, 11, 122–150.

ENGLE, R., AND J. MEZRICH (1995): "Grappling with GARCH," *Risk*, pp. 112–117.

ENGLE, R., AND K. SHEPPARD (2001): "Fitting Vast Dimensional Time Varying Covariance Models," Mimeo, UCSD.

FAN, J., M. WANG, AND Q. YAO (2008): "Modelling Multivariate Volatilities via Conditionally Uncorrelated Components," *Journal of Royal Statistical Society B*, 70, 679–702.

FRANCQ, C., L. HORVATH, AND J.-M. ZAKOIAN (2009): "Merits and Drawbacks of Variance Targeting in GARCH Models," .

GIACOMINI, G., AND H. WHITE (2006): "Tests of Conditional Predictive Ability," *Econometrica*, 74, 1545–1578.

GLOSTEN, L., R. JAGANNATHAN, AND D. RUNKLE (1992): "On the Relation Between the Expected Value and Volatility of the Nominal Excess Return on Stocks," *Journal of Finance*, 46, 1779–1801.

HANSEN, P. (2005): "A Test for Superior Predictive Ability," *Journal of Business and Economic Statistics*, 23, 365–380.

HANSEN, P., J. LARGE, AND A. LUNDE (2008): "Moving Average-Based Estimators of Integrated Variance," *Econometric Reviews*, 27, 79–111.

HANSEN, P., AND A. LUNDE (2005): "A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)," *Journal of Applied Econometrics*, 20, 873–889.

——— (2006a): "Consistent Ranking of Volatility Models," *Journal of Econometrics*, 131, 97–121.

——— (2006b): "Realized Variance and Market Microstructure Noise," *Journal of Business and Economic Statistics*, 24, 127–218.

Hansen, P., A. Lunde, and J. Nason (2003): "Choosing the Best Volatility Models: The Model Confidence Set Approach," Federal Reserve of Atlanta Working Paper.

——— (2009): "Model Confidence Sets," Federal Reserve of Atlanta Working Paper.

James, W., and C. Stein (1961): "Estimation with Quadratic Loss," *Proc. Fourth Berkley Symp. on Math. Statist. and Prob.*, 1, 361–379.

J.P.Morgan (1996): *Riskmetrics Technical Document, 4th ed.* J.P.Morgan, New York.

Kariya, T. (1988): "MTV Model and its Application to the Prediction of Stock Prices," in *Proceedings of the Second International Tampere Conference in Statistics*, ed. by T. Pullila, and S. Puntanen, Finland. University of Tampere.

Karolyi, G. (1995): "A Multivariate GARCH Model of International Transmissions of Stock Returns and Volatility: The Case of the United States and Canada," *Journal of Business and Economic Statistics*, vol. 13(1), 11–25.

Laurent, S., J. Rombouts, and F. Violante (2009): "On Loss Functions and Ranking Forecasting Performances of Multivariate Volatility Models," Cirano discussion paper 2009-45.

Mykland, P., and L. Zhang (2006): "Anova for Diffusions and Ito Processes," *The Annals of Statistics*, 34, 1931–1963.

Nelson, D. (1991): "Conditional Heteroskedasticity in Asset Returns: a New Approach," *Econometrica*, 59, 349–370.

Newey, W., and K. West (1987): "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55(3), 703–08.

Palandri, A. (2009): "Sequential Conditional Correlations: Inference and Evaluation," *Journal of Econometrics*, 153(2), 122–132.

Patton, A. (2009): "Volatility Forecast Comparison Using Imperfect Volatility Proxies," *Forthcoming in Journal Econometrics*.

Patton, A., and K. Sheppard (2009): "Evaluating Volatility and Correlation Forecasts," in *Handbook of Financial Time Series*, ed. by T. Andersen, R. Davis, J. Kreiss, and T. Mikosch. Springer.

Silvennoinen, A., and T. Terasvirta (2009): "Multivariate GARCH Models," in *Handbook of Financial Time Series*, ed. by T. Andersen, R. Davis, J. Kreiss, and T. Mikosch. Springer.

Tse, Y., and A. Tsui (2002): "A Multivariate GARCH Model with Time-Varying Correlations," *Journal of Business and Economic Statistics*, 20, 351–362.

van der Weide, R. (2002): "GO-GARCH: A Multivariate Generalized Orthogonal GARCH Model," *Journal of Applied Econometrics*, 17, 549–564.

Voev, V. (2009): "On the Economic Evaluation of Volatility Forecasts," *CREATES Research Paper 2009-56*.

Vrontos, I., P. Dellaportas, and D. Politis (2003): "A Full-Factor Multivariate GARCH Model," *Econometrics Journal*, 6, 311–333.

West, K. (1996): "Asymptotic Inference About Predictive Ability," *Econometrica*, 64, 1067–1084.

——— (2009): "Forecast Evaluation," in *Handbook of Economic Forecasting*, ed. by G. Elliott, C. Granger, and A. Timmermann. Elsevier.

White, H. (2000): "Reality Check for Data Snooping," *Econometrica*, 68, 1097–1126.

Zhang, L., P. Mykland, and Y. Ait-Sahalia (2005): "A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High Frequency Data," *Journal of The American Statistical Association*, 100 (472), 1394–1411.

Zhou, B. (1996): "High-frequency Data and Volatility in Foreign Exchange Rates," *Journal of Business & Economic Statistics*, 14, 45–52.