

# On the Forecasting Accuracy of Multivariate GARCH Models

Sébastien Laurent<sup>1</sup>, Jeroen V.K. Rombouts<sup>2</sup> and Francesco Violante<sup>3</sup>

January 28, 2011

## 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 the NYSE and compare 125 model based one, five and twenty-day 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. the dot-com bubble, the set of superior models is composed of 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

---

<sup>1</sup>Maastricht University, The Netherlands and Université catholique de Louvain, CORE, B-1348, Louvain-la-Neuve, Belgium.

<sup>2</sup>HEC Montréal, CIRANO, CIRPEE and Université catholique de Louvain, CORE, B-1348, Louvain-la-Neuve, Belgium.

<sup>3</sup>Université de Namur, CeReFim, B-5000, Namur, Belgium and Université catholique de Louvain, CORE, B-1348, Louvain-la-Neuve, Belgium.

Financial support from CIRPEE 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 Luc Bauwens, Christian Hafner, Jean-Michel Zakoian, Franz Palm, Lars Stentoft, the participants of the IFM2 Mathematical finance days, the CIREQ 4th times series conference, the 31st Annual International Symposium on Forecasting and CREATES seminar participants for their comments.

Correspondence to Jeroen Rombouts, HEC Montreal, 3000, chemin de la Cote-Sainte-Catherine. Tel.: +1 514 340-646. Fax.: +1 514 340-6469. E-mail: jeroen.rombouts@hec.ca

# 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 et al. (2006) and Silvennoinen and Terasvirta (2009b) for extensive surveys. The first generation of models, for example the VEC model of Bollerslev et al. (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 complexity of 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, which includes the diagonal and scalar BEKK models and the exponentially weighted moving average model proposed by J.P.Morgan (1996). Alternatively, factor structures like in Engle et al. (1990), the orthogonal models of Alexander and Chibumba (1997), Alexander (2000), van der Weide (2002), Lanne and Saikkonen (2007), and Fan et al. (2008) have been proposed. Furthermore, increasing attention has been devoted to conditional correlation models because they can be easily estimated using a multi-step procedure. These models have been first introduced by Engle (2002) and Tse and Tsui (2002). Extensions are the asymmetric conditional correlation model of Cappiello et al. (2006), the consistent DCC of Aielli (2006) and the sequential DCC 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 et al. (2003), Becker and Clements (2008) among others. In the multivariate setting, although many models are available, from an applied viewpoint, there are no clear guidelines available on model evaluation and selection.

This paper addresses the selection of multivariate GARCH models in terms of conditional variance matrix out-of-sample forecasting accuracy by providing a large scale analysis. We consider 10 assets from the NYSE, 125 multivariate GARCH specifications, 3 forecast horizons (1, 5 and 20-day ahead), 6 ex-post estimators (proxies) of the conditional covariance matrix, 4 statistical loss functions to measure model performances and 2 statistical tests for identification of the models with superior predictive performances. We also condition the analysis to the forecast sample period. We consider 3 different periods homogeneous in their volatility dynamics (calm, volatile and extremely volatile markets).<sup>1</sup>

Several approaches have been proposed with respect to the inference on the set of superior models. Testing procedures of equal predictive ability (EPA) based on pairwise comparison of forecast performances have been introduced by Diebold and Mariano (1995) and generalized by West (1996), Clark and McCracken (2001), Clark and West (2006), Giacomini and White (2006) and Clark and West (2007). See West (2006) for a survey. Since our aim is to compare a large number of model based forecasts in order to obtain a joint confidence interval for all possible pairwise comparisons, other alternatives based on multiple comparisons seem to be better suited to our analysis. The reality check test for data snooping of White (2000) and the improved version proposed by Hansen (2005) are based on superior predictive ability (SPA) and allow for multiple comparison against a prespecified benchmark model. Apart from the SPA test, we mainly follow the model confidence set (MCS) approach proposed by Hansen et al. (2010b). The MCS allows to identify, from a universe of model based forecasts, a subset of models, equivalent in terms of superior ability, which outperform all the other competing models. Note that, being tests of conditional predictive ability, SPA and MCS allow for a unified treatment of nested and non-nested models and to take into account estimation technique, parameter uncertainty, choice of the estimation and evaluation sample, and data heterogeneity.

To measure out-of-sample forecasting performance, model based forecasts are usually com-

---

<sup>1</sup>Recent somewhat related studies include Clements et al. (2009), Caporin and McAleer (2010) and Chiriac and Voev (2010), though their analysis usually involves a small number of alternative parameterizations and/or small cross sectional dimensions.

pared 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 et al. (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 et al. (2009) in the multivariate setting and Hansen et al. (2003) in the univariate settings.

Model performances are evaluated using the realized covariance estimator based on intraday returns sampled at the 5 minute frequency which serves as a proxy for the latent covariance matrix. Apart from the popular 5-minute frequency, which, given the characteristics of the assets selected should strike a good compromise between accuracy and microstructure bias (Andersen et al., 1999, Russell and Bandi, 2004), a robustness check with respect to the choice and the accuracy of the proxy is performed using the realized covariance estimator based on intraday returns sampled at 1 and 30 minutes and a realized kernel estimator based on intraday returns sampled at 1, 5 and 30 minutes, see de Pooter et al. (2008). Our results are robust to the choice and the accuracy of the volatility proxy.

As pointed out by Hansen et al. (2003), the MCS is specific to the set of candidate models and the sample period. We investigate the sensitivity of the models forecasting performances to the forecast evaluation sample by considering not only the full sample (from April 1, 1999 to December 27, 2008, totalling 2486 trading days) 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 sophisticated specifications 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.

With respect to the longer forecast horizons (5 and 20 day ahead), we find that while the composition of the MCS is in line with the one-step ahead case, the MCS reduces in size. The performances of models with similar properties and structure tend to cluster but differences between clusters increase. This, together with a substantial reduction of the variability of sample performances, due to the smoothness of longer horizon forecasts, makes it easier to separate between superior and inferior models.

In the last part of our study, 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 over periods of calm markets.

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. Detailed results for the MCS for the three subsamples are available in a supplementary appendix on the journal's website.

## 2 Methodology

### 2.1 Forecasting models set

Consider a  $N$ -dimensional vector stochastic process  $r_t = \mu_t + \epsilon_t$  and denote  $\mathfrak{S}_{t-1}$  the information set available at  $t - 1$ . Since the conditional mean  $\mu_t$  is typically of minor impor-

tance, following Hansen and Lunde (2005) and Becker and Clements (2008), we assume a constant conditional mean for all assets. The primary objective is the conditional variance matrix  $H_t = E(\epsilon_t \epsilon_t' | \mathfrak{S}_{t-1})$ , which is modeled using parametric specifications of the multivariate GARCH (MGARCH) type. To control for the number of parameters, we impose covariance or correlation targeting when possible, see Engle and Mezrich (1995).<sup>2</sup>

We consider several families of MGARCH models which are feasible in terms of numerical evaluation when the dimension of  $r_t$  is relatively large. According to the classification in Bauwens et al. (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 (RM) model of J.P.Morgan (1996). In the fully parameterized BEKK model with all orders set to 1, the conditional variance is given by

$$H_t = C + A\epsilon_{t-1}\epsilon_{t-1}'A' + BH_{t-1}B', \quad (1)$$

where  $C$  is a positive definite matrix and  $A$  and  $B$  are square parameter matrices. The full BEKK specification is not considered as it is not feasible for large cross-sectional dimensions. In the diagonal BEKK (DBEKK), the matrices of parameters  $A$  and  $B$  are diagonal, while in the scalar BEKK (SBEKK),  $A = aI_N$ ,  $B = bI_N$ , where  $a$  and  $b$  are scalars and  $I_N$  is the identity matrix. In these models, variance targeting is imposed by setting  $H = E(\epsilon_t \epsilon_t')$  and  $C = H - AHA' - BHB'$  which implies  $E(H_t) = H$ . The RM model assumes that the conditional variance matrix is an integrated process, i.e.,  $aI_N + bI_N = I_N$  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 model of Kariya (1988) and Alexander and Chibumba (1997). In this model, the data are generated by an orthogonal transformation of  $N$  uncorrelated factors,  $f_t$ , which can be separately defined as any stationary univariate GARCH process. The model can be expressed as

$$H_t = V^{1/2} P L^{1/2} S_t L^{1/2} P' V^{1/2'} \quad (2)$$

$$S_t = E_{t-1}(f_t f_t') = \text{diag}(\sigma_{f_{1,t}}^2, \dots, \sigma_{f_{N,t}}^2) \quad (3)$$

---

<sup>2</sup>For the properties of the variance targeting estimator and a comparison with the standard quasi-maximum likelihood estimator in the univariate case, see Francq et al. (2009).

$$f_t = L^{-1/2} P' V^{-1/2} \epsilon_t, \quad (4)$$

where  $V = \text{diag}(v_1, \dots, v_N)$ , with  $v_i = E(\epsilon_{i,t}^2)$ ,  $i = 1, \dots, N$ ,  $L$  is a  $N \times N$  diagonal matrix holding the eigenvalues of the unconditional correlation matrix on the main diagonal and  $P$  is the matrix of the associated orthogonal eigenvectors. Other specifications belonging to this group but unfeasible for large dimensional systems are the generalized orthogonal GARCH model by van der Weide (2002) and Lanne and Saikkonen (2007), the full factor GARCH model by Vrontos et al. (2003) and the conditionally uncorrelated components GARCH by Fan et al. (2008).

The last category of models consists of nonlinear combinations of univariate GARCH models. They allow to specify separately  $N$ , possibly different, univariate models for the conditional variances,  $\sigma_{i,t}^2$ ,  $i = 1, \dots, N$ , and a model for the conditional correlation matrix,  $R_t$ . The dynamic conditional correlation (DCC) model proposed by Engle (2002) (DCCE), is defined as

$$H_t = D_t^{1/2} R_t D_t^{1/2} \quad (5)$$

$$R_t = (Q_t \odot I_N)^{-1/2} Q_t (Q_t \odot I_N)^{-1/2} \quad (6)$$

$$Q_t = (1 - \alpha - \beta) \bar{Q} + \alpha u_{t-1} u'_{t-1} + \beta Q_{t-1}, \quad (7)$$

where  $D_t = \text{diag}(\sigma_{1,t}^2, \dots, \sigma_{N,t}^2)$  and  $u_{i,t} = \epsilon_{i,t}/\sigma_{i,t}$ ,  $i = 1, \dots, N$  defines the devolatilized innovations. The Constant Conditional Correlation (CCC) model of Bollerslev (1990), the Asymmetric DCC (DCCA) model of Cappiello et al. (2006) and 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 pairwise 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 ( $\alpha$ ,  $\beta$ ) in (7) even when the equi-correlation assumption is not supported by the data. Since the hypothesis of equi-correlation is likely to be rejected, in this paper we use the DECO approach to estimate the correlation parameters  $\alpha$  and  $\beta$  which are then used to predict and forecast time

varying and pairwise specific correlations. The DCCA extends the DCCE by accounting for asymmetries in (7) through the additional term  $\gamma(u_{t-1}u'_{t-1} \odot 1_{u_{t-1}<0}1'_{u_{t-1}<0})$ , 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. Following Engle and Sheppard (2001), correlation targeting is imposed by setting the long run target  $\bar{Q}$  equal to  $E(u_{t-1}u'_{t-1})$ .<sup>3</sup>

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

$$R_t = (1 - \theta_1 - \theta_2)\bar{R} + \theta_1\Psi_{t-1} + \theta_2R_{t-1}, \quad (8)$$

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

$$\psi_{ij,t-1} = \frac{\sum_{m=1}^K u_{i,t-m}u_{j,t-m}}{\sqrt{(\sum_{m=1}^K u_{i,t-m}^2)(\sum_{m=1}^K u_{j,t-m}^2)}}, \quad (9)$$

where  $u_{it}$  is defined as above. In the DCCT, the correlation matrix is modeled directly and depends on past local correlations of devolatilized innovations. Under correlation targeting, we set  $\bar{R}$  equal to the unconditional correlation of devolatilized innovations.

An advantage of the conditional correlation models relies on the fact that the estimation can be carried out sequentially. This requires first the estimation of the  $N$  conditional variances of the assets, second the estimation of the correlation target and the conditional correlation process. Although inefficient, this procedure is consistent and it dramatically reduces the computational burden of the likelihood. The univariate specification for the conditional variance for the marginal processes in DCC and orthogonal models are: Arch (Engle, 1982), Garch (Bollerslev, 1986), Gjr (Glosten et al., 1992), Exponential (Egarch) (Nelson, 1991), Asymmetric Power (Aparch) (Ding et al., 1993), Integrated (Igarch) (Engle and Bollerslev, 1986), RiskMetrics (Rm) (J.P.Morgan, 1996), Hyperbolic (Hgarch) (Davidson, 2004) and Fractionally integrated (Figarch) (Baillie et al., 1996). With respect to the number of lags, we fix both the orders to 1 for the diagonal and scalar BEKK, RM and the correlation specification in all the DCC models. The univariate models for the conditional variances in the Orthogonal and DCC specifications include various combinations of the orders,  $p, q$ . Table 1 summarizes the 125 MGARCH configurations considered.

---

<sup>3</sup>The choice  $\bar{Q}$  is not obvious as  $Q_t$  is neither a conditional variance nor a correlation. Although  $E(u_{t-1}u'_{t-1})$  is inconsistent for the target since the recursion in  $Q_t$  does not have a martingale difference representation, Aielli (2006) shows that the bias is negligible in practice.

Table 1: Forecasting models set (125 models)

DCC (105 models)				Orthogonal (17 models)				BEKK-type (3 models)		
$R_t$	$\sigma_{i,t}^2$	$p$	$q$		$\sigma_{f_{i,t}}^2$	$p$	$q$	$H_t$	$p$	$q$
	Arch	1,2	-		Arch	1,2	-	SBEKK	1	1
	Aparch	1	1		Aparch	1	1	DBEKK	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

The true variance matrix,  $\Sigma_t$ , is unobservable, and therefore models' predictive accuracy has to be measured with respect to some ex-post estimator,  $\hat{\Sigma}_t$ . This estimator is typically based on intraday returns which are defined as  $r_i = p_i - p_{i-\Delta}$ , with  $i = \Delta, 2\Delta, \dots, T$  and where  $1/\Delta$  is the number of intervals per day. The realized variance estimator of Andersen et al. (2003), Barndorff-Nielsen and Shephard (2004b), denoted  $\hat{\Sigma}^{(\Delta)}$ , is defined as  $\hat{\Sigma}^{(\Delta)} = \sum_{i=1}^{\lfloor 1/\Delta \rfloor} r_{i\Delta} r'_{i\Delta}$ , where  $\lfloor 1/\Delta \rfloor$  represents the integer part of  $1/\Delta$ . Under suitable conditions,  $\hat{\Sigma}^{(\Delta)}$  converges almost surely to  $\Sigma_t$  as the sampling frequency of the intraday returns increases ( $\Delta \rightarrow 0$ ). See Barndorff-Nielsen and Shephard (2004a), Mykland and Zhang (2006), Andersen et al. (2009) and related references for details.

The definition of  $\hat{\Sigma}^{(\Delta)}$  requires the intraday returns to be uncorrelated. When this assumption fails  $\hat{\Sigma}^{(\Delta)}$  would result in a biased estimator of  $\Sigma_t$ . For this reason we also consider as a second proxy a simple kernel estimator, denoted  $\hat{\Sigma}_{AC,q}^{(\Delta)}$ , based on the Newey and West (1987) variance estimator. The  $\hat{\Sigma}_{AC,q}^{(\Delta)}$  is equal to  $\hat{\Sigma}^{(\Delta)}$  plus a term that is a Bartlett-type weighted sum of  $q + 1$  higher-order autocovariances (see Zhou, 1996, Zhang et al., 2005, Hansen and Lunde, 2006b and Hansen et al., 2008).<sup>4</sup>

## 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

$$L_E = \text{vech}(\Sigma_t - H_t)' \text{vech}(\Sigma_t - H_t) \quad (10)$$

<sup>4</sup>For more refined realized kernels see Barndorff-Nielsen et al. (2008b) and Barndorff-Nielsen et al. (2008a).

$$L_F = \text{Tr}[(\Sigma_t - H_t)'(\Sigma_t - H_t)] \quad (11)$$

$$L_S = \text{Tr}[H_t^{-1}\Sigma_t] - \log |H_t^{-1}\Sigma_t| - N \quad (12)$$

$$L_b = \frac{1}{b(b-1)}\text{Tr}(\Sigma_t^b - H_t^b) - \frac{1}{(b-1)}\text{Tr}(H_t^{b-1}(\Sigma_t - H_t)) \quad b \geq 3. \quad (13)$$

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  $\text{vech}(\Sigma_t - H_t)$ , where  $\text{vech}(\cdot)$  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  $L_E$ . 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  $L_E$  by double counting the loss associated to the conditional covariances. The Stein loss function  $L_S$  (James and Stein, 1961) is a scale invariant loss function based on the standardized (in matrix sense) forecast error and it is the loss function implied by the Wishart density.  $L_S$  is homogeneous of degree 0 (errors are measured in relative terms) and asymmetric with respect to over/under predictions (in a matrix sense) with under predictions being heavily penalized. Finally, in the same spirit,  $L_b$  also accounts for asymmetry with respect to over/under predictions, but in the opposite direction, i.e. over predictions are penalized instead.  $L_b$  also allows to tune the degree of asymmetry through the choice of the parameter  $b$  (which also represents its degree of homogeneity). In this paper, we set  $b = 3$  which implies a mild degree of asymmetry comparable to  $L_S$ . See Laurent et al. (2009) for further details and examples. Note that all loss functions are normalized to zero when there is perfect fit.

## 2.4 The Model Confidence Set

The MCS approach of Hansen et al. (2010b) allows to identify, from an initial set, a subset of forecasts that contains the best forecast at a confidence level  $\alpha$ . Let us denote the initial set of  $k$ -step ahead conditional variance forecasts  $\mathcal{M}^0 : \{H_{i,t+k} \in \mathcal{M}^0 \forall i = 1, \dots, M\}$ , where  $t = 0, 1, \dots, T-1$  and  $T$  is the forecast sample size. The starting hypothesis is that all forecasts in  $\mathcal{M}^0$  have equal forecasting performance, measured by a loss function  $L_{i,t} = L(\Sigma_t, H_{i,t})$ . Let  $d_{ij,t} = L_{i,t} - L_{j,t} \forall i, j = 1, \dots, M$  define the relative performance of forecast  $i$  and  $j$ . The null hypothesis takes the form  $H_{0,\mathcal{M}^0} : E(d_{ij,t}) = 0, \forall i, j = 1, \dots, M$ . We use the ‘deviation’

statistic defined as  $T_D = M^{-1} \sum_{i \in \mathcal{M}^0} t_i^2$ , where  $t_i = \sqrt{T} \bar{d}_i / \sqrt{\text{Var}(\sqrt{T} \bar{d}_i)}$  represents the standardized relative performance of forecast  $i$  with respect to the average across forecasts,  $\bar{d}_i = M^{-1} \sum_{j \in \mathcal{M}^0} \bar{d}_{ij}$  and  $\bar{d}_{ij} = T^{-1} \sum_{t=1}^T d_{ij,t}$  is the sample loss difference between forecast  $i$  and  $j$ . A block bootstrap scheme is used to obtain the distribution under the null. If the null is rejected, an elimination rule removes the forecast with the largest  $t_i$ . This process is repeated until non-rejection of the null occurs, thus allowing to construct a  $(1 - \alpha)$ -confidence set for the best forecast in  $\mathcal{M}^0$ .

Being the statistic based on studentized quantities, the analysis of  $\text{Var}(\bar{d}_i)$  is crucial for evaluating the informativeness of the MCS. Hansen et al. (2010b) point out that an inferior forecast (i.e.,  $\bar{d}_i > 0$ ) may be included in the MCS if the variance of  $\bar{d}_i$  is large enough, i.e.  $t_i$  is sufficiently small to avoid being discarded. Given that  $\text{Var}(\bar{d}_i)$  can be decomposed as

$$\text{Var}(\bar{d}_i) = \text{Var}(\bar{L}) + \left( 1 + \frac{\text{Var}(\bar{L}_i)}{\text{Var}(\bar{L})} - 2 \sqrt{\frac{\text{Var}(\bar{L}_i)}{\text{Var}(\bar{L})}} \text{Corr}(\bar{L}_i, \bar{L}) \right), \quad (14)$$

where  $\bar{L}_i = T^{-1} \sum_{t=1}^T L_{i,t}$  and  $\bar{L} = M^{-1} \sum_{i \in \mathcal{M}} \bar{L}_i$ , this might be the case when  $\text{Var}(\bar{L}_i)$  is large enough and/or  $\text{Corr}(\bar{L}_i, \bar{L})$  is small. However, the risk of including uninformative models does not always arise. For example, when the residual set contains two forecasts,  $|\bar{d}_1| = |\bar{d}_2|$ , hence  $\text{Var}(\bar{d}_1) = \text{Var}(\bar{d}_2)$ , the variance plays no role in the elimination: the forecast with the largest loss is eliminated. Furthermore, when the set contains more than two forecasts, only one of which with  $\bar{d}_i > 0$ , the latter is always excluded. In other cases the risk of including uninformative models is only marginal: an inferior forecast can only be preferred to another inferior forecast with better sample performance but never to a forecast with  $\bar{d}_i < 0$ .

### 3 Data and forecasting scheme

Our data are extracted from the One-Minute Equity Data (OMED) database provided by Tick Data and consists of 1-minute intervals last trade prices for 10 stocks traded in the NYSE, see Table 2, for a period spanning from March 02, 1988 to December 27, 2008. To minimize microstructure noise components such as non-trading and non-synchronous trading which may induce bias in the volatility proxy, the stocks have been selected among the most liquid over the period analyzed.<sup>5</sup> We only use prices occurring within the official trading

---

<sup>5</sup>The non-synchronous trading may induce the empirical correlation between stock returns to converge to zero as the sampling frequency of the data increases (Epps, 1979). Preliminary analysis shows some evidence

day (9:30 a.m. - 4:00 p.m.), i.e., 390 observations per day. The data has been cleaned from weekends, holidays and early closing/late opening days. After removing days with missing values and/or constant prices, we retain 5226 trading days.

Proxies and model based variances are defined over the period of partially available intraday returns (open-to-close), i.e., excluding the portion of the daily price variation associated with the overnight return. Following Andersen et al. (2010), we treat the overnight returns as deterministically occurring jumps, so that the trading day returns equal the daily returns adjusted for the observed overnight jump. Thus,  $\hat{\Sigma}_t$  and  $H_t$  are measures of the volatility during the hours that the market is open. The MGARCH models are estimated by quasi maximum likelihood using trading day returns.<sup>6</sup> To reduce the computational burden, unconditional means are subtracted from each series of returns prior to the estimation. The initial estimation sample consists of the first 2740 daily observations, i.e. March 02, 1988 to March 31, 1999. The last 2486 trading days constitute the sample for which we compute 1, 5 and 20-day ahead forecasts. For computational convenience, we re-estimate the model parameters every month (22 days) using a rolling window of the last 2740 observations. Within the 22-day window, the parameters are kept fixed and only the data is updated. This mix of fixed and rolling forecast scheme satisfies the assumptions required by the MCS test (Hansen et al., 2010b), allows the comparison of nested models and to account for data heterogeneity (Giacomini and White, 2006, West, 2006), as well as to compare results over sub-samples (forecasts of different periods are conditioned on the most recent information). The two proxies for the conditional variance are computed using intraday returns sampled at 1, 5 and 30-minute frequency. Unless explicitly mentioned, the results reported in the following section refer to the realized variance estimator  $\hat{\Sigma}^{(5min)}_t$ , while  $\hat{\Sigma}_t^{(\Delta)}$ ,  $\Delta = 1min, 30min$  and  $\hat{\Sigma}_{AC,q,t}^{(\Delta)}$ ,  $\Delta = 1min, 5min, 30min$  serve to check the robustness of the results to different proxies.

The sample period considered is characterized by dramatic changes in the 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 April, 1999, and ending in March 2003, it includes the peak of the Dot-com boom (until March 2000), the burst of the speculative

---

of this phenomenon when the covariance proxies are computed using 1-minute returns, while it is negligible when the data used to compute the covariance proxy is sampled at 5-minute frequency or lower.

<sup>6</sup>All programs have been written by the authors using OxMetrics 6 (Doornik, 2007) and G@RCH 6 (Laurent, 2009).

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 daily returns - March 02, 1988 to December 27, 2008 (5229 obs.)

bubble (March, 2000) and the aftermath of the bubble burst. Peaks in this period correspond to the 9/11 event (September, 2001). Towards the end of the period, the turmoil starts with

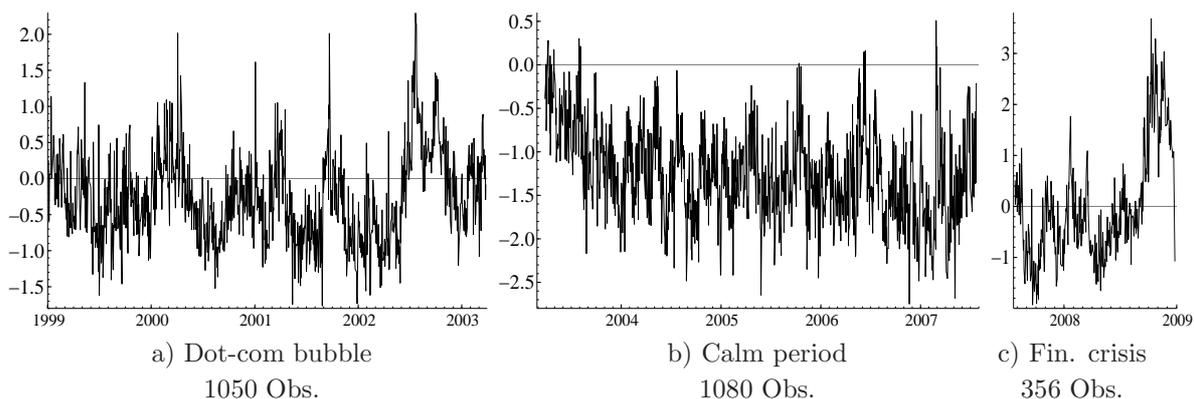


Figure 1: Daily realized volatility from 5-min returns (log-scale) - equally weighted portfolio the bankruptcy of WorldCom (July, 2002) and ends in October, 2002, with a 5- and 6-year record low of the Dow Jones Industrial and NASDAQ respectively. The second sub-sample (1080 obs.), from April 2003 to July, 2007, corresponds to a period of stable and upward trending markets. 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 the US's borders. The end of the period coincides with the peak of the crisis (September/October, 2008). Figure 1 shows the daily realized variance computed using 5-minute returns of an equally weighted portfolio of the 10 assets in Table 2. Clearly, the

volatility dynamics and its scale varies widely between periods.

## 4 Model Confidence Set results

### 4.1 Full sample

The MCS results for the full forecast sample (2486 obs.) are reported in Table 3.<sup>7</sup> Following Hansen et al. (2003), we set the confidence level for the MCS to  $\alpha = 0.25$ . The block length and the number of bootstrap samples are set respectively to 2 and 10,000.

The MCS includes 39 models for  $L_E$  and is largely dominated by orthogonal and DECO models. With respect to the composition of the MCS, we remark first, that the orthogonal models show the best sample performances. The flexibility of these models seems therefore able to better adapt to a sample that alternates periods of calm with periods of high instability. These results also support 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 outperform symmetric models with Gjr specifications showing the best sample performances. The same consideration holds for longer versus shorter lags, with higher order models showing in general better sample performances. Third, the MCS includes specifications that allow for long memory and integrated conditional variances, i.e. DECO, DCCA and DCCE with Hgarch conditional variances, DECO, DCCA and DCCT with Figarch conditional variances, DECO with Rm conditional variances and RM. Furthermore, if we focus on the sample performances, the specifications allowing for fractional integration or hyperbolic decay of shocks in the conditional variances exhibit the best performances within each family of models.

We next turn to the MCS under the two asymmetric loss functions. Under  $L_S$ , the MCS includes 10 models all belonging to the DCC family. The selected models focus on the long memory properties of the conditional variances rather than leverage, asymmetry or even time varying correlation. The MCS includes models from the CCC, DCCE, DCCA and DCCT families all with (fractionally) integrated variance processes or high order Garch, 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 dominated by orthogonal models,

---

<sup>7</sup>To save space, results for the Frobenius loss function ( $L_F$ ) are not reported. Because of its similarity with the Euclidean loss function ( $L_E$ ), results based on  $L_F$  are very similar in terms of ordering and, in general, we remark that the more conservative  $L_E$  MCS always includes the MCS obtained under the  $L_F$  loss function.

Table 3: MCS - full sample (1/04/99 - 27/12/08)

Euclidean distance (39 models)							Stein distance (10 models)								
MCS		Rnk	$\bar{L}_i$	$T_D$	p-val	VR	Corr	MCS		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	CCC	Figarch (1,1)	7	3.528	0.346	0.57	0.730	0.932
	Figarch (1,1)	20	3.673	0.521	0.67	1.076	0.996		Garch (2,1)	10	3.548	1.302	0.25	1.211	0.988
	Hgarch (1,1)	25	3.720	0.803	0.45	1.052	0.996		Igarch (1,1)	3	3.501	0.546	0.69	1.119	0.985
DCCT	Figarch (1,1)	38	3.823	1.089	0.30	1.159	0.994	DCCA	Igarch (1,1)	4	3.516	0.680	0.57	1.254	0.986
DCCCE	Egarch (1,2)	53	3.901	1.207	0.25	1.325	0.998	DCCCT	Figarch (1,1)	5	3.518	0.232	0.69	0.743	0.931
	Figarch (1,1)	18	3.661	0.406	0.71	1.075	0.996		Garch (2,1)	9	3.541	0.880	0.36	1.223	0.989
	Hgarch (1,1)	24	3.719	0.766	0.47	1.057	0.996		Igarch (1,1)	1	3.496	-	1.00	1.130	0.987
DECO	Aparch (1,1)	27	3.735	0.848	0.42	1.111	0.998	DCCCE	Figarch (1,1)	6	3.525	0.381	0.57	0.789	0.929
	Egarch (0,1)	29	3.742	0.825	0.43	1.172	0.999		Garch (2,1)	8	3.535	0.561	0.49	1.255	0.989
	(0,2)	30	3.747	0.877	0.40	1.163	0.999		Igarch (1,1)	2	3.500	0.235	0.69	1.228	0.986
	(1,2)	33	3.762	0.936	0.37	1.176	0.999	$L_3$ loss function (20 models)							
	Figarch (1,1)	2	3.478	0.004	0.94	0.906	0.997	MCS		Rnk	$\bar{L}_i$	$T_D$	p-val	VR	Corr
	Garch (1,1)	34	3.768	0.906	0.38	1.171	0.998	DECO	Figarch (1,1)	15	102.3	0.844	0.41	1.092	0.999
	(1,2)	31	3.750	0.965	0.35	1.137	0.999		Gjr (1,1)	23	105.0	1.063	0.29	1.156	1.000
	(2,1)	28	3.737	0.993	0.34	1.125	0.999		Hgarch (1,1)	17	102.5	0.883	0.39	1.082	0.999
	(2,2)	32	3.759	1.061	0.31	1.159	0.999	Igarch (1,1)	24	105.1	0.969	0.33	1.142	0.999	
	Gjr (1,1)	22	3.692	0.603	0.60	1.090	0.998	Aparch (1,1)	10	98.83	0.689	0.50	1.023	0.999	
	(1,2)	21	3.676	0.706	0.50	1.046	0.999	Egarch (0,1)	14	101.3	0.864	0.43	1.083	1.000	
	(2,1)	14	3.635	0.521	0.67	0.991	0.999	(0,2)	11	99.98	1.001	0.43	1.033	1.000	
	(2,2)	19	3.667	0.668	0.54	1.036	0.999	(1,1)	12	100.4	0.848	0.43	1.017	1.000	
	Hgarch (1,1)	5	3.535	0.103	0.89	0.886	0.997	(1,2)	6	98.45	0.988	0.43	0.977	0.999	
	Igarch (1,1)	35	3.783	1.018	0.33	1.061	0.993	(2,1)	16	102.4	0.834	0.43	1.104	1.000	
Rm (1,1)	23	3.699	0.545	0.64	1.117	0.998	(2,2)	4	95.68	0.609	0.53	0.898	0.999		
Orth.	Aparch (1,1)	7	3.575	0.197	0.89	0.921	0.996	Orth.	Arch (1)	18	103.8	0.923	0.43	0.914	0.990
	Egarch (0,1)	17	3.660	0.628	0.58	1.019	0.998	Garch (1,1)	13	101.1	0.851	0.43	1.053	1.000	
	(0,2)	13	3.623	0.567	0.64	0.945	0.999	(1,2)	9	98.60	1.055	0.43	0.984	1.000	
	(1,1)	15	3.647	0.735	0.50	0.933	0.998	(2,1)	7	98.59	0.928	0.43	0.997	1.000	
	(1,2)	12	3.593	0.517	0.67	0.872	0.997	(2,2)	8	98.60	1.099	0.43	0.980	1.000	
	(2,1)	26	3.726	1.037	0.32	1.066	0.999	Gjr (1,1)	5	97.69	0.887	0.43	0.954	0.999	
	(2,2)	6	3.539	0.175	0.89	0.793	0.996	(1,2)	2	94.06	1.032	0.53	0.852	1.000	
	Garch (1,1)	16	3.656	0.724	0.50	0.964	0.998	(2,1)	1	91.98	-	1.00	0.801	1.000	
	(1,2)	11	3.589	0.594	0.67	0.870	0.998	(2,2)	3	94.54	0.782	0.53	0.872	0.999	
	(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								

Note. Rnk: model  $i$ 's ranking position based on average sample performances,  $\bar{L}_i$  (out of 125 models);  $T_D$ : deviation statistic; p-val: MCS p-value;  $VR \equiv Var(L_i)/Var(L)$ ;  $Corr \equiv Corr(L_i, L)$ . The values reported for  $L_E$  and  $L_3$  are the average loss per element of the forecast error matrix considered, i.e. the average loss is divided by  $N(N+1)/2$  and  $N^2$  respectively. For  $L_S$ , where the distance is measured in relative terms, the average loss is reported.

which score the best sample performances. The MCS also includes other specifications, all of which in the DECO family, allowing 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 or, most likely, that this failure refers only to particular periods of time where the inadequacy of the MGARCH estimator is so striking to drive the result even when very long evaluation samples

are considered. In the next sections, we discuss the MCS results for three sub-samples. The results are reported in Tables A1, A2, A3 and A4 of the supplementary appendix.<sup>8</sup> 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

For the Dot-com period, the MCS under  $L_E$  contains 38 models, see Table A1 of the supplementary appendix. As expected, there are differences with the MCS obtained for the full sample. First, modeling directly the conditional correlation and accounting for the leverage effect in the conditional variances becomes more important. DCC type models with Egarch conditional variances dominate the MCS and show the smallest losses. Among these we find two CCC specifications, both with Egarch dynamics for the conditional variances, which suggests that adequately modeling 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, e.g. DCC with Aparch and Gjr dynamics for the conditional variances, underlines the importance of the specific choice of the parametrization. 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) allowing for long memory and integrated conditional variances (against 10 out of 39 for the full sample).

The Stein loss function delivers a small MCS which consists of 2 models, namely the DCCE and the DCCT with Igarch 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. In fact, although statistically inferior, the top of the classification is dominated by models accounting for this feature. The MCS under the  $L_3$  loss function includes 8 models, all from the orthogonal GARCH family, most of which accounting for asymmetry in the variance processes of the components. Hence, consistently with the results for the full

---

<sup>8</sup>The supplementary appendix is available on the journal's website.

sample, orthogonal models seem to systematically underpredict the conditional variance.

### 4.3 Calm markets

The MCSs for the second sub-sample are reported in Tables A2 and A3 of the supplementary appendix. The MCS under  $L_E$  contains 74 models, about 60% of the models considered and includes specifications from all families of models. As a general result, the data does not show evidence of dynamics in the correlation process and asymmetry/leverage or long memory in the conditional variance. This period being characterized by a relatively small and slow moving volatility, the result is not surprising and as expected most of the MGARCH model based forecasts show a better fit, e.g.,  $L_E$  is on average about 0.3 against 2.7 in the Dot-com period. Looking at the composition of the MCS, we can draw the following conclusions. First, DECO type models are excluded from the MCS except for DECO-Aparch and DECO-Rm. However, since both models show a relatively small correlation with the average across models, the information content of these models is doubtful. Second, although only Orth.-Gjr models are statistically inferior, the remaining orthogonal specifications show the highest relative variance and smallest correlation with the average loss. Hence, it is likely that these models end up in the MCS because the data does not contain sufficient information to infer that they are inferior within the MCS. Third, similar conclusions hold for DCC type models with Rm and Gjr conditional variances. In particular, the latter show by far the poorest performances within the MCS, the largest relative variance and the smallest correlation with  $\bar{L}$ .

Under  $L_S$ , the MCS contains 12 models, and consists of CCC, DCCT and DCCE specifications, with Garch conditional variances, confirming that the hypotheses of constant conditional correlation and symmetry cannot be rejected. It also includes two asymmetric specifications, i.e. DCCE-Gjr(1,1) and DCCT-Gjr(1,1), both characterized by weaker sample performances within the MCS. The MCS under  $L_3$  is similar in size and composition to the set obtained under  $L_E$ . Although the type of asymmetry accounted for by  $L_3$  is not statistically relevant, we observe changes in the ordering. The orthogonal models, ranking between 37th and 62nd under  $L_E$ , are at the 6th and the 19th position under  $L_3$ , suggesting that, also over calm periods, these models tend to underpredict the conditional variance.

#### 4.4 2007-08 financial crisis

The MCSs for the last sub-sample are reported in Table A4 of the supplementary appendix. The MCS under  $L_E$  contains 39 models. In line with the results obtained for the full sample, the MCS is dominated by specifications in the DECO and the Orthogonal families. Other DCC type specifications are included when accounting for long memory and integrated conditional variances. Furthermore, although the MCS includes models that account for asymmetry/leverage, contrary to the Dot-com bubble burst period, specifications with Egarch dynamics for the variance processes are systematically rejected.

Under  $L_S$ , the results are also consistent with the full sample, though the MCS is larger (26 models). The models in the MCS belong to the DCC family and account for long-memory in volatility or leverage effect. The non-rejection of some CCC specifications, which is surprising in this case, suggests that adequately modeling the conditional variances of the returns can compensate the loss of forecasting accuracy induced by the restrictive assumption of constant correlation.

Also in line with the results obtained for the full sample, the MCS under  $L_3$  contains 26 models and is dominated by orthogonal and DECO specifications, the former showing the best sample performances. Among the DECO specifications included in the MCS, we find both evidence of long memory/integrated conditional variances and leverage effect.

In this sub-sample, the average loss (irrespective of the choice of the loss function) is much larger than in the first two, e.g.,  $L_E$  is on average about 14.5 against 0.3 in the Calm period. Hence, MGARCH models have difficulties to accurately predict the conditional variance in turbulent periods. As a consequence, large losses accumulated over short periods of high instability tend to drive the MCS results even when long forecasting periods are considered. Hence, a careful evaluation of the trade off between forecast sample length (to reduce sampling variability) and the informativeness and accuracy of the selection appears to be crucial in this setting.

#### 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  $\hat{\Sigma}(\Delta)$ , computed using 1 and 30 minute returns and  $\hat{\Sigma}_{AC,q=1}^{(\Delta)}$ , computed using 1, 5 and 30 minutes returns. The MCS is robust in terms of size and composition to the

alternative volatility proxies. In particular, when the proxy is based on higher frequency returns we generally find smaller MCSs.

As an example (complete results are available upon request), consider  $L_E$ , under  $\hat{\Sigma}^{(1min)}$  ( $\hat{\Sigma}_{AC,q=1}^{(1min)}$ ). 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. The robustness of these results is implied by the consistency of the loss function. Thus, 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. When the evaluation is based on  $\hat{\Sigma}_{AC,q=1}^{(5min)}$ , consistently with the results obtained under  $\hat{\Sigma}^{(5min)}$ , 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 30 minutes returns we find 41 (40) models for the full sample and 41 (59), 73 (66) and 37 (35) for the three sub-samples respectively.

Our results show that the use of a high frequency proxy ensures the elimination of uninformative models while the consistency with the results obtained using relatively low frequency proxies shows that the potential microstructure bias is negligible. This result underlines the value of high precision proxies, in particular when the set of competing models is characterized by a high degree of similarity, see Laurent et al. (2009) and Patton and Sheppard (2009).

#### 4.6 Multiple comparison based on longer forecast horizons

The MCS under  $L_E$  for the multi-step (5 and 20 days) forecast evaluation over the full sample are reported in Table 4. Detailed results for the three sub-samples are given in Table A5 of the supplementary appendix. As expected, when the forecast horizon increases the average loss increases, and this irrespectively of the evaluation period or the choice of the loss function. Furthermore, the MCS reduces in size, which seems to be a specific feature here. This result is due to two reasons. First, the performances of models with similar properties and structure tend to cluster (convergence to the same long run variance matrix) but differences between clusters increase (different specifications can imply different levels for the long run variance). Second, longer horizon forecasts are generally smoother, which substantially reduces the variability of the relative performances,  $\bar{d}_i$ , making it easier to separate between models. The interaction between the two effects is particularly strong for the calm period. The results are in line with the conclusion drawn for the 1-step ahead

Table 4: MCS ( $L_E$ ) - Multistep ahead covariance forecasts - Full sample

5-day ahead forecast horizon (16 models)							20-day ahead forecast horizon (12 models)								
MCS		Rnk	$\bar{L}_i$	$T_D$	p-val	VR	Corr	MCS		Rnk	$\bar{L}_i$	$T_D$	p-val	VR	Corr
DCCA	Rm (1,1)	15	4.508	0.907	0.40	1.039	0.998	DECO	Rm (1,1)	9	5.117	1.821	0.25	0.960	0.999
DECO	Gjr (1,1)	11	4.468	1.082	0.31	1.043	0.999	Orth.	Aparch (1,1)	1	5.020	-	1.00	1.006	1.000
	(2,1)	14	4.495	1.211	0.25	1.047	0.999		Egarch (0,1)	7	5.102	1.131	0.29	1.026	0.999
Orth.	Rm (1,1)	4	4.366	0.224	0.85	0.962	0.998	Garch (1,1)	(0,2)	13	5.172	1.209	0.25	1.044	0.999
	Aparch (1,1)	2	4.312	0.073	0.94	0.963	0.998		(1,1)	10	5.121	1.482	0.25	1.031	0.999
Orth.	Egarch (0,1)	6	4.375	0.446	0.69	1.007	0.999	(2,1)	(2,1)	11	5.144	1.308	0.25	1.034	0.999
	(0,2)	23	4.569	1.160	0.27	1.141	0.996		Gjr (1,1)	(1,1)	3	5.039	0.351	0.69	0.989
Orth.	(1,1)	9	4.410	0.608	0.60	1.026	0.999	(1,2)		(1,2)	5	5.092	1.495	0.25	1.002
	(2,1)	10	4.465	0.662	0.57	1.061	0.998		(2,1)	(2,1)	8	5.105	1.630	0.25	0.990
Orth.	(2,2)	12	4.470	0.819	0.45	0.985	0.999	Gjr (1,1)		(1,1)	2	5.024	0.019	0.89	0.967
	Garch (1,1)	5	4.369	0.677	0.60	0.967	0.998		(1,2)	(1,2)	4	5.068	0.963	0.38	0.991
Orth.	(1,2)	13	4.472	1.001	0.35	1.035	0.999	(2,1)		(2,1)	6	5.096	1.432	0.25	0.978
	(2,1)	8	4.399	0.737	0.57	0.974	0.999		Gjr (1,1)	(1,1)	1	4.300	-	1.00	0.912
Orth.	Gjr (1,1)	1	4.300	-	1.00	0.912	0.996	(1,2)		(1,2)	7	4.390	0.616	0.60	0.990
	(1,2)	7	4.390	0.616	0.60	0.990	0.999		(2,1)	(2,1)	3	4.316	0.052	0.94	0.915
Orth.	(2,1)	3	4.316	0.052	0.94	0.915	0.997	Gjr (1,1)		(1,1)	1	4.300	-	1.00	0.912
	(2,1)	3	4.316	0.052	0.94	0.915	0.997		(1,2)	(1,2)	7	4.390	0.616	0.60	0.990
Orth.	(2,1)	3	4.316	0.052	0.94	0.915	0.997	(2,1)		(2,1)	3	4.316	0.052	0.94	0.915

Note. See Table 3.

forecast comparison, i.e., the constant correlation hypothesis cannot be rejected, but the size of the MCS reduces to only the CCC-Egarch(2,1) model for 5-day ahead horizon and the CCC-Egarch(2,1), the CCC-Garch(1,2) and the DCCT-Garch(1,2) models for the 20-day ahead horizon. This is because over this period the variability of  $\bar{d}_i$  reduces so much that even small differences in performances become highly significant.

In sharp contrast with the 1-step ahead case, we find that non stationary models are rejected most of the time. This is because, longer horizon forecasts for these types of models typically exhibit an explosive pattern. Exceptions are the RiskMetrics type models (RM and DCC type with Rm) and the conditional correlation models with Igarch conditional variances, but only over the 2007-2008 financial crisis period. The non exclusion of these specifications indicates the inadequacy of MGARCH models in periods of extreme market instability. In fact, the  $k$ -step ahead forecast for the RiskMetrics type models (except for the correlation component in the DCC type) is uninformative because it coincides with the 1-step ahead forecast independently from the forecast horizon. This also holds for the models allowing for integrated conditional variances, whose intercept over this period is insignificant in most cases and numerically close to zero (0.003 on average). Similar considerations and qualitatively the same results, not reported but available upon request, are also found for the  $L_S$  and  $L_3$  loss functions.

## 5 Setting the 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 alternatives. As benchmarks we choose simple and parsimonious specifications taking into account two dimensions: the assumption on the multivariate structure (CCC, DCCE and Orthogonal) and on the dynamics of the variance of the marginal processes/principal components (Garch(1,1) vs. 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 other DCC specifications 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 modeling the dynamics of the conditional covariances and correlations. 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). This test generates the probability distribution of the model which performs best relative to the benchmark. Using the notation introduced in Section 2.4, let us define  $d_{0j,t} = L_{0,t} - L_{j,t}$ ,  $j = 1, \dots, M$ , 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  $\omega_j^2$  denotes the asymptotic variance of  $\lambda_j$ . The test statistic is  $\sqrt{T} \max_{j \in M} \frac{\bar{d}_{0j}}{\bar{\omega}_j}$  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 Table 5. Consistently with the 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. With respect to the type of multivariate model, the Orthogonal approach

Table 5: SPA test

Symmetric dynamics for the marginal variances									Asymmetric dynamics for the marginal variances										
Benchmark 1: CCC-Garch(1,1)									Benchmark 4: CCC-Egarch(0,1)										
$L_E$			$L_S$			$L_3$			$L_E$			$L_S$			$L_3$				
$p_L$	$p_C$	$p_U$	$p_L$	$p_C$	$p_U$	$p_L$	$p_C$	$p_U$	$p_L$	$p_C$	$p_U$	$p_L$	$p_C$	$p_U$	$p_L$	$p_C$	$p_U$		
Full sample	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.02	0.02	Full sample	0.01	0.02	0.02	0.00	0.00	0.00	0.04	0.04	0.04
Dot-com	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	Dot-com	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Calm period	0.02	0.02	0.02	0.43	<b>0.82</b>	0.96	0.17	<b>0.21</b>	0.26	Calm period	0.10	<b>0.16</b>	0.24	0.00	0.00	0.00	0.24	<b>0.42</b>	0.55
Fin. crisis	0.01	0.02	0.02	0.00	0.00	0.00	0.02	0.02	0.02	Fin. crisis	0.02	0.02	0.02	0.05	0.06	0.08	0.09	0.02	0.02
Benchmark 2: DCCE-Garch(1,1)									Benchmark 5: DCCE-Egarch(0,1)										
$L_E$			$L_S$			$L_3$			$L_E$			$L_S$			$L_3$				
$p_L$	$p_C$	$p_U$	$p_L$	$p_C$	$p_U$	$p_L$	$p_C$	$p_U$	$p_L$	$p_C$	$p_U$	$p_L$	$p_C$	$p_U$	$p_L$	$p_C$	$p_U$		
Full sample	0.06	0.06	0.07	0.00	0.00	0.00	0.09	<b>0.10</b>	0.10	Full sample	0.10	<b>0.11</b>	0.14	0.00	0.00	0.00	0.08	0.08	0.09
Dot-com	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	Dot-com	0.40	<b>0.75</b>	0.91	0.00	0.00	0.00	0.08	0.09	0.13
Calm period	0.11	<b>0.11</b>	0.17	0.38	<b>0.82</b>	0.98	0.09	<b>0.10</b>	0.14	Calm period	0.15	<b>0.23</b>	0.39	0.00	0.00	0.00	0.02	0.03	0.03
Fin. crisis	0.02	0.02	0.02	0.01	0.01	0.01	0.04	0.04	0.04	Fin. crisis	0.03	0.04	0.04	0.16	<b>0.23</b>	0.46	0.03	0.04	0.04
Benchmark 3: Orth.-Garch(1,1)									Benchmark 6: Orth.-Egarch(0,1)										
$L_E$			$L_S$			$L_3$			$L_E$			$L_S$			$L_3$				
$p_L$	$p_C$	$p_U$	$p_L$	$p_C$	$p_U$	$p_L$	$p_C$	$p_U$	$p_L$	$p_C$	$p_U$	$p_L$	$p_C$	$p_U$	$p_L$	$p_C$	$p_U$		
Full sample	0.09	<b>0.12</b>	0.12	0.00	0.00	0.00	0.19	<b>0.28</b>	0.28	Full sample	0.24	<b>0.37</b>	0.40	0.00	0.00	0.00	0.30	<b>0.52</b>	0.55
Dot-com	0.07	0.08	0.09	0.00	0.00	0.00	0.03	0.03	0.04	Dot-com	0.34	<b>0.52</b>	0.60	0.00	0.00	0.00	0.22	<b>0.72</b>	0.84
Calm period	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.02	Calm period	0.00	0.006	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Fin. crisis	0.26	<b>0.32</b>	0.33	0.00	0.00	0.00	0.36	<b>0.49</b>	0.49	Fin. crisis	0.19	<b>0.22</b>	0.23	0.00	0.00	0.00	0.34	<b>0.49</b>	0.50

Note.  $p_C$  consistent p-value,  $p_L$  and  $p_U$  lower and upper 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  $\alpha = 0.10$ . The number of bootstrap samples used to obtain the distribution under the null is 10,000 and the block length is 2.

(Benchmark 3 and 6) exhibits superior performance only 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.

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. 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 (MGARCH) 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 predictive ability, to 125 MGARCH model based forecasts. We consider 10 assets from NYSE for which we forecast 1, 5 and 20-day ahead conditional variance matrices from April 1, 1999 to December 27, 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 to the properties of the loss function, we conclude that 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.

We illustrate how sensitive the MCS is with respect to the forecast sample under investigation by considering not only the full forecast sample but also by investigating sub-samples which are homogenous in their volatility dynamics. We find that 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.

With respect to the longer forecast horizons (5 and 20 day ahead), we find that while the composition of the MCS is in line with the 1-step ahead case, the MCS reduces in size. The performances of models with similar properties and structure tend to cluster but differences between clusters increase. This, together with a substantial reduction of the variability of sample performances, due to the smoothness of longer horizon forecasts, makes it easier to separate between superior and inferior models.

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 equi-correlation, de-

livers 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 leverage effect in the variance of the marginal processes is not outperformed by the other 124 specifications cannot be rejected at standard levels irrespectively of the evaluation period.

This paper considers only forecasts based on MGARCH models. It would be interesting to compare the performances of this class of volatility models with other approaches such as heterogeneous autoregression based on historical values of ex-post measures of the conditional variance as in the model proposed by Corsi (2009), models that combine ARMA structures for both the conditional variance and realized measures of volatility as in Hansen et al. (2010a) or yet multivariate stochastic volatility (Gourieroux et al., 2009) and regime switching models as in Silvennoinen and Terasvirta (2009a). Other problems like the evaluation of forecast performances of correlation matrices and high dimensional applications (hundreds of series) also merit more attention.

## References

- G. P. Aielli. Consistent estimation of large scale dynamic conditional correlations. Department of Statistics, University of Florence, 2006.
- C. Alexander. Orthogonal methods for generating large positive semi-definite covariance matrices. Henley Business School, Reading University, 2000.
- C.O. Alexander and A.M. Chibumba. Multivariate orthogonal factor GARCH. University of

- Sussex, Mimeo, 1997.
- T.G. Andersen, T. Bollerslev, F.X. Diebold, and P. Labys. (understanding, optimizing, using and forecasting) realized volatility and correlation. Technical report, New York University, Leonard N. Stern School of Business, 1999.
- T.G. Andersen, T. Bollerslev, F.X. Diebold, and P. Labys. Modeling and forecasting realized volatility. *Econometrica*, 71:579–625, 2003.
- T.G. Andersen, T. Bollerslev, and F.X. Diebold. Parametric and non parametric volatility measurement. In L.P. Hansen and Y. Ait-Sahalia, editors, *Handbook of Financial Econometrics*. Elsevier, 2009.
- T.G. Andersen, T. Bollerslev, P. Frederiksen, and M.O. Nielsen. Continuous-time models, realized volatilities and testable distributional implications for daily stock returns. *Journal of Applied Econometrics*, 25:233–261, 2010.
- R.T. Baillie, T. Bollerslev, and H.O. Mikkelsen. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74:3–30, 1996.
- T.G. Bali. The intertemporal relation between expected returns and risk. *Journal of Financial Economics*, 87:101–131, 2008.
- O. Barndorff-Nielsen and N. Shephard. Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, 2:1–37, 2004a.
- O. Barndorff-Nielsen and N. Shephard. Econometric analysis of realized covariation: High frequency covariance, regression and correlation in financial economics. *Econometrica*, 72:885–925, 2004b.
- O. Barndorff-Nielsen, P.R. Hansen, A. Lunde, and N. Shephard. Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *DP, Oxford University*, 2008a.
- O. Barndorff-Nielsen, P.R. Hansen, A. Lunde, and N. Shephard. Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica*, 76(6):1481–1536, 2008b.

- L. Bauwens, S. Laurent, and J.V.K. Rombouts. Multivariate GARCH models: A survey. *Journal of Applied Econometrics*, 21:79–109, 2006.
- R. Becker and A.E. Clements. Are combination forecasts of S&P 500 volatility statistically superior? *International Journal of Forecasting*, 24:122–133, 2008.
- T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327, 1986.
- T. Bollerslev. Modeling the coherence in short-run nominal exchange rates: A multivariate generalized arch model. *Review of Economics and Statistics*, 72:498–505, 1990.
- T. Bollerslev, R.F. Engle, and J.M. Wooldridge. A capital asset pricing model with time varying covariances. *Journal of Political Economy*, 96:116–131, 1988.
- M. Caporin and M. McAleer. Ranking multivariate GARCH models by problem dimension. *SSRN eLibrary*, 2010.
- L. Cappiello, R.F. Engle, and K. Sheppard. Asymmetric dynamics in the correlations of global equity and bond returns. *Journal of Financial Econometrics*, 4(4):537–572, 2006.
- R. Chiriac and V. Voev. Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics*, 2010. forthcoming.
- T.E. Clark and M.W. McCracken. Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics*, 105:85–110, 2001.
- T.E. Clark and K.D. West. Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis. *Journal of Econometrics*, 135:155–186, 2006.
- T.E. Clark and K.D. West. Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, 138:291–311, 2007.
- A. Clements, M. Doolan, S. Hurn, and R. Becker. Evaluating multivariate volatility forecasts. *NCER Working Paper Series 41*, 2009.
- F. Corsi. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7:174–196, 2009.

- J. Davidson. Moment and memory properties of linear conditional heteroscedasticity models, and a new model. *Journal of Business and Economic Statistics*, 22:16–29, 2004.
- M. de Pooter, M. Martens, and D. van Dijk. Predicting the daily covariance matrix for S&P 100 stocks using intraday data - but which frequency to use? *Econometric Reviews*, 27(1-3):199–229, 2008.
- F.X. Diebold and R.S. Mariano. Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13:253–263, 1995.
- Z. Ding, C. W. J. Granger, and R. F. Engle. A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1:83–106, 1993.
- J.A. Doornik. *Object-Oriented Matrix Programming Using Ox*. Timberlake Consultants Press, 2007.
- R.F. Engle. Autoregressive conditional heteroscedasticity with estimates of variance of united kingdom inflation. *Econometrica*, 50:987–1008, 1982.
- R.F. Engle. Dynamic conditional correlation - a simple class of multivariate GARCH models. *Journal of Business and Economic Statistics*, 20:339–350, 2002.
- R.F. Engle. *Anticipating correlations*. Princeton Univ. Press, 2009.
- R.F. Engle and T. Bollerslev. Modelling the persistence of conditional variances. *Econometric Reviews*, 5:1–50, 1986.
- R.F. Engle and B.T. Kelly. Dynamic equicorrelation. Stern School of Business, 2008.
- R.F. Engle and F.K. Kroner. Multivariate simultaneous generalized ARCH. *Econometric Theory*, 11:122–150, 1995.
- R.F. Engle and J. Mezrich. Grappling with GARCH. *Risk*, pages 112–117, 1995.
- R.F. Engle and K. Sheppard. Fitting vast dimensional time varying covariance models. Mimeo, UCSD, 2001.
- R.F. Engle, V.K. Ng, and M. Rothschild. Asset pricing with a factor-ARCH covariance structure: Empirical estimates for treasury bills. *Journal of Econometrics*, 45:213–238, 1990.

- T.W. Epps. Comovements in stock prices in the very short run. *Journal of the American Statistical Association*, 74:291–298, 1979.
- J. Fan, M. Wang, and Q. Yao. Modelling multivariate volatilities via conditionally uncorrelated components. *Journal of Royal Statistical Society B*, 70:679–702, 2008.
- C. Francq, L. Horvath, and J.M. Zakoian. Merits and drawbacks of variance targeting in GARCH models. MPRA Paper n.15143, 2009.
- G. Giacomini and H. White. Tests of conditional predictive ability. *Econometrica*, 74:1545–1578, 2006.
- L.R. Glosten, R. Jagannathan, and D.E. Runkle. On the relation between the expected value and volatility of the nominal excess return on stocks. *Journal of Finance*, 46:1779–1801, 1992.
- C. Gourieroux, J. Jasiak, and R. Sufana. The wishart autoregressive process of multivariate stochastic volatility. *Journal of Econometrics*, 150(2):167–181, 2009.
- P.R. Hansen. A test for superior predictive ability. *Journal of Business and Economic Statistics*, 23:365–380, 2005.
- P.R. Hansen and A. Lunde. A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? *Journal of Applied Econometrics*, 20:873–889, 2005.
- P.R. Hansen and A. Lunde. Consistent ranking of volatility models. *Journal of Econometrics*, 131:97–121, 2006a.
- P.R. Hansen and A. Lunde. Realized variance and market microstructure noise. *Journal of Business and Economic Statistics*, 24:127–218, 2006b.
- P.R. Hansen, A. Lunde, and J.M. Nason. Choosing the best volatility models: The model confidence set approach. Federal Reserve of Atlanta Working Paper, 2003.
- P.R. Hansen, J. Large, and A. Lunde. Moving average-based estimators of integrated variance. *Econometric Reviews*, 27:79–111, 2008.
- P.R. Hansen, Z. Huang, and H. Shek. Realized GARCH: A Complete Model of Returns and Realized Measures of Volatility. *SSRN*, 2010a.

- P.R. Hansen, A. Lunde, and J.M. Nason. Model confidence sets. *Econometrica*, 2010b. forthcoming.
- W. James and C. Stein. Estimation with quadratic loss. *Proc. Fourth Berkley Symp. on Math. Statist. and Prob.*, 1:361–379, 1961.
- J.P.Morgan. *Riskmetrics Technical Document, 4th ed.* J.P.Morgan, New York, 1996.
- T. Kariya. MTV model and its application to the prediction of stock prices. In T. Pullila and S. Puntanen, editors, *Proceedings of the Second International Tampere Conference in Statistics*, Finland, 1988. University of Tampere.
- G.A. Karolyi. 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, 1995.
- M. Lanne and P. Saikkonen. A multivariate generalized orthogonal factor GARCH model. *Journal of Business and Economic Statistics*, 25:61–75, 2007.
- S. Laurent. *G@RCH 6. Estimating and Forecasting GARCH Models*. Timberlake Consultants Ltd., 2009.
- S. Laurent, J.V.K. Rombouts, and F. Violante. On loss functions and ranking forecasting performances of multivariate volatility models. Cirano discussion paper 2009-45, 2009.
- P.A. Mykland and L. Zhang. Anova for diffusions and Ito processes. *The Annals of Statistics*, 34:1931–1963, 2006.
- D.B. Nelson. Conditional heteroskedasticity in asset returns: a new approach. *Econometrica*, 59:349–370, 1991.
- W.K. Newey and K.D. West. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–08, 1987.
- A. Palandri. Sequential conditional correlations: Inference and evaluation. *Journal of Econometrics*, 153(2):122–132, 2009.
- A.J. Patton. Volatility forecast comparison using imperfect volatility proxies. *Forthcoming in Journal Econometrics*, 2009.

- A.J. Patton and K. Sheppard. Evaluating volatility and correlation forecasts. In T.G. Andersen, R.A. Davis, J.P. Kreiss, and T. Mikosch, editors, *Handbook of Financial Time Series*. Springer, 2009.
- J.R. Russell and F.M. Bandi. Microstructure noise, realized volatility, and optimal sampling. Econometric Society 2004 Latin American Meetings 220, Econometric Society, 2004.
- A. Silvennoinen and T. Terasvirta. Modeling multivariate autoregressive conditional heteroskedasticity with the double smooth transition conditional correlation GARCH model. *Journal of Financial Econometrics*, 7:373–411, 2009a.
- A. Silvennoinen and T. Terasvirta. Multivariate GARCH models. In T.G. Andersen, R.A. Davis, J.P. Kreiss, and T. Mikosch, editors, *Handbook of Financial Time Series*. Springer, 2009b.
- Y.K. Tse and A.K.C Tsui. A multivariate GARCH model with time-varying correlations. *Journal of Business and Economic Statistics*, 20:351–362, 2002.
- R. van der Weide. GO-GARCH: A multivariate generalized orthogonal GARCH model. *Journal of Applied Econometrics*, 17:549–564, 2002.
- I.D. Vrontos, P. Dellaportas, and D.N. Politis. A full-factor multivariate GARCH model. *Econometrics Journal*, 6:311–333, 2003.
- K.D. West. Asymptotic inference about predictive ability. *Econometrica*, 64:1067–1084, 1996.
- K.D. West. Forecast evaluation. In G. Elliott, C.W.J. Granger, and A. Timmermann, editors, *Handbook of Economic Forecasting*. Elsevier, 2006.
- H. White. Reality check for data snooping. *Econometrica*, 68:1097–1126, 2000.
- L. Zhang, P.A. Mykland, and Y. Ait-Sahalia. A tale of two time scales: Determining integrated volatility with noisy high frequency data. *Journal of The American Statistical Association*, 100 (472):1394–1411, 2005.
- B. Zhou. High-frequency data and volatility in foreign exchange rates. *Journal of Business & Economic Statistics*, 14:45–52, 1996.