Model Averaging in Risk Management with an Application to Futures Markets*

M. Hashem Pesaran
University of Cambridge, CIMF, GSA Capital and USC

Christoph Schleicher GSA Capital
Paolo Zaffaroni
Imperial College London and CIMF

January 2008

Abstract

This paper considers the problem of model uncertainty in the case of multi-asset volatility models and discusses the use of model averaging techniques as a way of dealing with the risk of inadvertently using false models in portfolio management. Evaluation of volatility models is then considered and a simple Value-at-Risk (VaR) diagnostic test is proposed for individual as well as 'average' models. The asymptotic as well as the exact finite-sample distribution of the test statistic, dealing with the possibility of parameter uncertainty, are established. The model averaging idea and the VaR diagnostic tests are illustrated by an application to portfolios of daily returns on six currencies, four equity indices, four ten year government bonds and four commodities over the period 1991-2007. The empirical evidence supports the use of 'thick' model averaging strategies over single models or Bayesian type model averaging procedures.

JEL Classifications: C32, C52, C53, G11
Key Words: Model Averaging, Value-at-Risk, Decision Based Evaluations.

*This is a substantially revised version of a paper previously circulated under the title of "Model Averaging and Value-at-Risk based Evaluation of Large Multi Asset Volatility Models for Risk Management". The present version deals with active as well as passive risk management problems under model uncertainty. We are to grateful to the Editor, three anonymous referees, and the participants of the 2007 London-Oxford Financial Econometrics Study Group, for constructive comments and suggestions. We would also like to thank GSA Capital for allowing us to use data from their futures data base for this paper.

1 Introduction

Multivariate models of conditional volatility are of crucial importance for optimal asset allocation, risk management, derivative pricing and dynamic hedging. However, their use in practice has been rather limited, particularly in the case of portfolios with a large number of assets. There are only a few published empirical studies that consider the performance of multivariate volatility models involving a large number of assets, and for operational reasons most of these studies focus on highly restricted versions of the multivariate generalized autoregressive conditional heteroscedastic (GARCH) model of Bollerslev (1986). The risk associated with possible model misspecification could then be sizeable. Also for risk-management purposes, the main focus is often on the tail behavior of the predictive density of the asset returns, and not simply to obtain the 'best' approximating volatility model. This in turn implies that a unified treatment of empirical portfolio analysis requires shifting the focus from a statistical to a decision-theoretic framework for model evaluation. This paper provides an integrated econometric approach to the portfolio optimization subject to the Value at Risk (VaR) constraint in the presence of model uncertainty, and the associated risk monitoring problem. In this paper we focus on uncertainty of multivariate volatility models and abstract from return prediction uncertainty already addressed extensively in the literature.[1] The various issues involved are discussed and evaluated in the context of an empirical application.

Many variants of the multivariate GARCH have been proposed in the literature. These include the conditionally constant correlation (CCC) model of Bollerslev (1990), the Riskmetrics specifications popularized by J.P.Morgan (1996) and used predominantly by practitioners, the orthogonal GARCH model of Alexander (2001), and the dynamic conditional correlation (DCC) model advanced by Engle (2002).[2] Recent surveys are provided in Bauwens, Laurent, and Rombouts (2003) and McAleer (2005). Multivariate stochastic volatility (SV) models have also been considered in the literature, with reviews by Ghysels, Harvey, and Renault (1995) and Shephard (2004).[3] We consider models frequently used by practitioners together with many models recently proposed in academic papers, and consider their empirical performance within a decision-theoretic framework.

The highly restricted nature of the multivariate volatility models advanced in the literature could present a high degree of model uncertainty which ought to be recognized at the outset. This is particularly important since due to data limitations and operational considerations it is not possible to subject these models to rigorous statistical testing. Application of model selection procedures also involves additional risks that are difficult to assess *a priori*. This is especially true when the number of assets is moderately large, and it might well be that no single model choice would be satisfactory in practice.

This paper considers model averaging as a risk diversification strategy in dealing with model uncertainty, and provides a detailed application of recent developments in model

[1]See, for example, Pesaran and Timmermann (1995).

[2]The DCC model is also related to the VCC model of Tse and Tsui (2002).

[3]So far the focus of the SV literature has been on univariate and multivariate models with a small number of assets, with the notable exceptions of Diebold and Nerlove (1989), Engle, Ng, and Rothschild (1990), King, Sentana, and Wadhwani (1994) and Harvey, Ruiz, and Shephard (1994), that are similar in structure to the class of factor GARCH models that we do consider below.

averaging techniques to multi-asset volatility models. Frequently used model selection criteria are the Akaike Information Criterion (AIC) and the Schwartz Bayesian Information Criterion (SBC). However, such a two-step procedure is subject to the pre-test (selection) bias problem and tends to under-estimate the uncertainty that surrounds the forecasts. Of course, the use of model averaging techniques in econometrics is not new and dates back to the work of Granger and Newbold (1977) on forecast combination.[4] However, this literature focusses on combining point forecasts and does not address the problem of combining forecast probability distribution functions which is relevant in risk management.

Concerning model evaluation, the standard forecast evaluation techniques that focus on metrics such as root mean square forecast errors (RMSFE), also run into difficulties when considering volatility models. Since volatility is not directly observable, it is often proxied by the square of daily returns or more recently by the standard error of intra-daily returns, known as realized volatility (see, for example, Andersen, Bollerslev, Diebold, and Labys (2003)). In multi-asset contexts the use of standard metrics such as RMSFE is further complicated by the need to select weights to be attached to errors in forecasts of individual asset volatilities and their cross-volatility correlations and the choice of such weights is not innocuous in a multivariate framework (see Pesaran and Skouras (2002)). Here we develop a simple criterion for evaluation of alternative volatility forecasts by examining the Value-at-Risk (VaR) performance of their associated portfolios. Our test, which can be applied to individual as well as to average models, belongs to a class of so-called unconditional coverage tests, the most important case of which is the Kupiec (1995) binomial test. In contrast to the existing literature, though, we formally establish both the asymptotic as well as the exact finite-sample distribution of our test statistics. Further, we provide formal conditions that permit to ignore the potential effect of the sampling variability associated with estimation. Conditional coverage tests (see Christoffersen (1998)) and density forecast tests (Crnkovic and Drachman (1997) and Berkowitz (2001)) could also be adapted to our model averaging framework, although the related distribution theory will need to be established. For a review of existing approaches to the evaluation of the VaR estimates see Andersen, Bollerslev, Christoffersen, and Diebold (2006). The VaR based diagnostic tests developed in this paper can be used both for risk monitoring of a given portfolio as well as for construction of optimal (in the VaR sense) portfolios.

The remainder of the paper is organized as follows: the decision problem that underlies the VaR analysis is set out in Section 2. Section 3 provides a brief outline of the different types of multivariate volatility models considered in the paper. Several approaches to model averaging are reviewed and discussed in Section 4. Section 5 introduces the Value-at-Risk (VaR) diagnostic test and establishes its finite-sample as well as its asymptotic distribution. Section 6 provides a detailed empirical analysis using daily returns for eighteen futures contracts covering equity indices, government bonds, exchange rates and commodities over the period 2 January 1991 to 11 July 2007. Section 7 concludes with a summary of the main results and suggestions for future research. The mathematical proofs and a description of the multivariate volatility models are provided in three appendices.

[4]For reviews of the forecast combination literature see Clemen (1989), Granger (1989), Diebold and Lopez (1996) and Hendry and Clements (2002).

2 The Decision Problem: Active Risk Management

Here we are concerned with the decision of a portfolio manager who is interested in controlling the risk of a given portfolio composed of N futures contracts over a given trading day. We refer to this portfolio decision problem as 'active risk management', and distinguish it from what might be called 'passive risk management' where the outcome of the portfolio decision is evaluated or monitored by a risk manager or by an outside supervisory financial institution. This distinction is important since the solution to the portfolio decision problem requires a complete knowledge of the conditional *multivariate* probability distribution of the $N \times 1$ vector of returns, $\mathbf{r}_t$. In contrast, for passive risk management it is clearly possible to work *directly* with the conditional *univariate* distribution of portfolio returns, ρ_t, with no apparent need for multivariate volatility modelling.

We suppose that the portfolio manager is allowed to hold long and short positions and the contracts could be in home currency (taken to be US dollar) or in foreign currencies. Denote the price of each contract (in local currency) on close of business day t by P_{jt}, and the US dollar exchange rate relevant to the j^{th} contract by E_{jt} (measured as the units of foreign currency in one US dollar), and the number of contracts held in the portfolio at yesterday's close by $n_{j,t-1}$. Abstracting from transaction costs, the change in the value of this portfolio in US dollar is given by

$$\Delta V_t = \sum_{j=1}^{N} n_{j,t-1} \left(\frac{P_{jt} - P_{j,t-1}}{E_{jt}} \right) = \sum_{j=1}^{N} \left(\frac{n_{j,t-1} P_{j,t-1}}{E_{j,t-1}} \right) \left(\frac{r_{jt}}{1 + r_{jt}^e} \right), \tag{1}$$

where n_{jt} and $r_{jt} = \left(P_{jt} - P_{j,t-1}\right)/P_{j,t-1}$ are the position size (number of contracts) and the one-day holding return of asset j, and $r_{jt}^e = (E_{jt} - E_{j,t-1})/E_{j,t-1}$ is the daily change in spot currency rate. Note that for US dollar denominated assets $r_{jt}^e = 0$. Since the second order terms $r_{jt}^e r_{jt}$, $\left(r_{jt}^e\right)^2 r_{jt}$, *etc.* are negligible the daily change in the value function can be simplified as

$$\Delta V_t \approx \sum_{j=1}^{N} \omega_{j,t-1} r_{jt} C_{t-1}, \tag{2}$$

where $\omega_{j,t-1} = n_{j,t-1} P_{j,t-1}/(E_{j,t-1} C_{t-1})$ is the value of the contracts in US dollar relative to notional capital, C_{t-1}, on close of day $t-1$. In what follows we suppose that a portfolio manager chooses these position sizes by solving a standard mean-variance problem subject to a daily value at risk (VaR) constraint. Let $\rho_t(\omega_{t-1}) = \omega_{t-1}' \mathbf{r}_t$ be the portfolio return, where $\omega_{t-1} = (\omega_{1,t-1}, \omega_{2,t-2}, ..., \omega_{N,t-1})'$ and $\mathbf{r}_t = (r_{1t}, r_{2t}, ..., r_{Nt})'$. Then the objective function of the mean-variance problem is given by

$$Q(\omega_{t-1} | \mathcal{M}, \mathcal{F}_{t-1}) = \omega_{t-1}' E(\mathbf{r}_t | \mathcal{M}, \mathcal{F}_{t-1}) - \frac{\delta_{t-1}}{2} \omega_{t-1}' V(\mathbf{r}_t | \mathcal{M}, \mathcal{F}_{t-1}) \omega_{t-1}, \tag{3}$$

where $\mathcal{F}_{t-1}$ is the available information, $\delta_{t-1} > 0$ is a (possibly time-varying) risk-aversion coefficient, and $\mathcal{M}$ denotes the assumed multivariate model of returns, characterized by the joint probability distribution of $\mathbf{r}_t$ conditional on $\mathcal{F}_{t-1}$, and denoted by $f_{\mathcal{M}}(\mathbf{r}_t \,| \mathcal{F}_{t-1})$. The VaR constraint is given by

$$\Pr\left(\omega_{t-1}' \mathbf{r}_t < -L_{t-1} \,| \mathcal{M}, \mathcal{F}_{t-1}\right) \leq \alpha, \tag{4}$$

where $L_{t-1} > 0$ is a pre-specified maximum daily loss (as a fraction of notional capital) and α is a probability value (typically taken to be 1%) which captures the fund manager's attitude towards risk in the case of large losses.

To obtain a feasible solution to the mean-variance problem we shall assume that conditional on $\mathcal{F}_{t-1}$ returns $\mathbf{r}_t$ have means $\mu_{\mathcal{M},t}$ and finite variance-covariances $\mathbf{\Sigma}_{\mathcal{M},t}$. The conditional probability distribution of the change in the portfolio value, $\rho_t = \omega'_{t-1}\mathbf{r}_t$, takes relatively simple forms when the distribution of returns are closed under linear transformations.[5] For example, in the case where the conditional distribution of $\mathbf{r}_t$ follows a multivariate t distribution with $v_{t-1} > 2$ degrees of freedom, ρ_t will also be t distributed with the same degree of freedom, and hence

$$\frac{\omega'_{t-1}\mathbf{r}_t - \omega'_{t-1}\mu_{\mathcal{M},t}}{\sqrt{\frac{v_{t-1}-2}{v_{t-1}}\omega'_{t-1}\mathbf{\Sigma}_{\mathcal{M},t}\omega_{t-1}}} \sim t_{v_{t-1}},$$

and the VaR constraint (4) simplifies to

$$\frac{-L_{t-1} - \omega'_{t-1}\mu_{\mathcal{M},t}}{\sqrt{\omega'_{t-1}\mathbf{\Sigma}_{\mathcal{M},t}\omega_{t-1}}} \leq \sqrt{\frac{v_{t-1}-2}{v_{t-1}}}T_v^{-1}(\alpha) = -\sqrt{\frac{v_{t-1}-2}{v_{t-1}}}c_{v_{t-1},\alpha} \equiv -\tilde{c}_{v_{t-1},\alpha}, \tag{5}$$

where $c_{v_{t-1},\alpha}$ ($c_{v_{t-1},\alpha} > 0$ for $\alpha < 0.5$) is the α% left tail of the Student t distribution with v_{t-1} degrees of freedom.

The optimal portfolio weights, $\omega^*_{t-1,\mathcal{M}}$, that maximize $Q(\omega_{t-1}|\mathcal{M}, \mathcal{F}_{t-1})$ subject to the VaR constraint in (5) are then given by[6]

$$\omega^*_{t-1,\mathcal{M}} = \begin{cases} \frac{1}{\delta_{t-1}}\mathbf{\Sigma}^{-1}_{\mathcal{M},t}\mu_{\mathcal{M},t}, & \text{if } \delta_{t-1} \geq \delta^*_{t-1} \\ \frac{1}{\delta^*_{t-1}}\mathbf{\Sigma}^{-1}_{\mathcal{M},t}\mu_{\mathcal{M},t}, & \text{otherwise}, \end{cases} \tag{6}$$

with

$$\delta^*_{t-1} \equiv \frac{s_{\mathcal{M},t}(\tilde{c}_{v_{t-1},\alpha} - s_{\mathcal{M},t})}{L_{t-1}},$$

where $s_{\mathcal{M},t} = \sqrt{\mu'_{\mathcal{M},t}\mathbf{\Sigma}^{-1}_{\mathcal{M},t}\mu_{\mathcal{M},t}}$ can be viewed as the *ex ante* daily Sharpe ratio of the portfolio. This solution shows that the VaR constraint will be binding only if the risk aversion coefficient is relatively small. In the case where the VaR constraint binds, the level of $\omega^*_{t-1,\mathcal{M}}$ is determined by the level of the risk capital, L_{t-1}, and the tail property of the underlying return distribution. In practice, to avoid negative values of δ^*_{t-1}, it would be advisable to cap $s_{\mathcal{M},t}$ so that it does not exceed $\tilde{c}_{v_{t-1},\alpha}$.

The solution to the constrained MV optimization problem is more complicated when the return distribution is constructed as an average of a number of Gaussian or t-distributed return distributions with different means and variances. We shall return to this problem below in section 4.1.

[5]The probability distribution of $\mathbf{r}_t$ is said to be closed under linear transformations if all linear combinations of $\mathbf{r}_t$ have the same distribution as the marginal distributions of the returns.

[6]For the details of the derivations see Appendix A.

3 Multivariate Models of Asset Returns

Our primary concern in this paper is on modelling and evaluation of alternative multivariate volatility models in a wider context that nests both passive and active risk management problems. Typically one would also need to address the uncertainty that surrounds the conditional mean returns, $E\left(\mathbf{r}_t \left|\mathcal{F}_{t-1}\right.\right) = \mu_t$. But given the focus of the present paper we shall abstract from this problem and throughout assume that mean returns can be characterized by first order autoregressive processes

$$r_{it} = a_{i0} + \alpha_{i1} r_{i,t-1} + \varepsilon_{it}, \tag{7}$$

such that $\mu_{it} = a_{i0} + \alpha_{i1} r_{i,t-1}$. Therefore, in what follows we shall focus on alternative specifications of the joint probability distribution of $\varepsilon_t = \mathbf{r}_t - \mu_t$, namely $f_{\mathcal{M}}(\mathbf{r}_t \left|\mathcal{F}_{t-1}\right.)$ for the model class $\mathcal{M}$. For this purpose it is convenient to work with the standardized unexpected returns, $\mathbf{z}_t$, defined by $\mathbf{z}_t = \mathbf{\Sigma}_t^{-\frac{1}{2}} \varepsilon_t$, where $\mathbf{\Sigma}_t = Var\left(\varepsilon_t \left|\mathcal{F}_{t-1}\right.\right)$.

A complete specification of $f_{\mathcal{M}}(\mathbf{r}_t \left|\mathcal{F}_{t-1}\right.)$ can be achieved by: (i) a non-singular choice of $\mathbf{\Sigma}_t$; (ii) specification of the distribution of standardized values, $\mathbf{z}_t$. For the latter, we focus on distributions that are closed under linear transformations. This includes the case of standard multivariate Gaussian, and the multivariate Student t with v degrees of freedom. These are the two specifications that are most commonly encountered in practice. In specifying $\mathbf{\Sigma}_t$, we focus on parametric volatility models, the classical example of which is the multivariate generalized autoregressive heteroskedasticity model of order $1,1$ (MGARCH$(1,1)$). In its most general form it is given by[7]

$$vech(\mathbf{\Sigma}_{MGARCH,t}) = \mathbf{a}_0 + \mathbf{A}_0 vech(\mathbf{\Sigma}_{MGARCH,t-1}) + \mathbf{B}_0 vech\left(\mathbf{r}_{t-1}\mathbf{r}'_{t-1}\right), \tag{8}$$

where $vech(\cdot)$ denotes the column stacking operator of the lower portion of a symmetric matrix, $\mathbf{a}_0$ is an $N(N+1)/2 \times 1$ vector, and $\mathbf{A}_0$, $\mathbf{B}_0$ are $N(N+1)/2 \times N(N+1)/2$ matrices of unknown coefficients. It is evident that even such a low-order model already contains a large number of parameters even for moderate values of N which renders model (8) effectively unfeasible for practical applications.

The different multivariate volatility models considered in this paper are special cases of the MGARCH$(1,1)$. These volatility models are denoted by M_i and the associated conditional covariance matrix by $\mathbf{\Sigma}_{it}$. Altogether we consider 53 different specifications of $\mathbf{\Sigma}_{it}$ that can be grouped into 9 different model types. We consider both econometric specifications advanced in the academic literature as well as *ad hoc* data filters more commonly used by practitioners.

Within the first group, we considered the constant conditional correlation (CCC(p,q)) model of Bollerslev (1990) and its more recent generalizations, namely the dynamic conditional correlation (DCC$(p,q,1,1)$) of Engle (2002) and the asymmetric dynamic conditional correlation (ADCC$(p,q,1,1)$) of Cappiello, Engle, and Sheppard (2006). We also consider the orthogonal GARCH (O-GARCH(p,q)) of Alexander (2001), the factor GARCH model of Harvey, Ruiz, and Sentana (1992) (factor GARCH$(p,q,1,1)$) and the Student t dynamic conditional correlation model of Pesaran and Pesaran (2007) (TDCC$(p,q,1,1)$).

[7] See Bollerslev, Engle, and Wooldridge (1988, equation (4)).

Within the second group we consider the equal-weighted moving average model (EQMA(n_0)), which is a rolling filter that puts equal weights on the n_0 most recent squared observations. We further consider the exponentially-weighted moving average (EWMA(n_0, λ_0)), well known as the Riskmetrics filter (see J.P.Morgan (1996)) and a number of its variants such as the two-parameter exponential-weighted moving average (EWMA (n_0, λ_0, ν_0)) (see De Santis, Litterman, Vesval, and Winkelmann (2003, p.14)). We also consider two hybrid filters: a mixed moving average (MMA(n_0, ν_0)) specification whereby the conditional variances are computed as in the EQMA(n_0) model but with the conditional covariances obtained using the Riskmetrics approach; and a generalized exponential-weighted moving average (EWMA(n_0, p, q, ν_0)) whereby conditional variances are modelled as univariate GARCH(p, q) with the conditional covariances specified using the Riskmetrics approach. More detailed accounts are given in Appendix B.

Let θ_{i0} be the $k_i \times 1$ vector of coefficients characterizing the true unknown parameters of the volatility model M_i, denoted by $\mathbf{\Sigma}_{it}= \mathbf{\Sigma}_{it}(\theta_{i0})$. For estimation of θ_{i0} we shall be using the Gaussian pseudo maximum likelihood estimator (PMLE), defined by

$$\hat{\theta}_{iT_0} = \arg\max_{\theta_i \in \Theta_i} \left\{ -\frac{1}{2} \sum_{t=\tau-T_0+1}^{\tau} \left[\log |\mathbf{\Sigma}_{it}(\theta_i)| + \varepsilon_t' \mathbf{\Sigma}_{it}^{-1}(\theta_i)\varepsilon_t \right] \right\}, \tag{9}$$

where Θ_i represents a suitable parameter space, τ is the end of the estimation period, T_0 is the size of the estimation period.[8] Correspondingly, let $\hat{\mathbf{\Sigma}}_{it} = \mathbf{\Sigma}_{it}(\hat{\theta}_{iT_0})$. We view Gaussian PMLE as a robust method, delivering consistent and asymptotically normal estimates of θ_i under the volatility model M_i even for non-Gaussian $\mathbf{z}_t$. In particular we shall assume that as $T_0 \to \infty$,

$$\hat{\theta}_{iT_0} \xrightarrow{p} \theta_{i0} \tag{10}$$

and

$$\sqrt{T_0}\left(\hat{\theta}_{iT_0} - \theta_{i0}\right) \mid M_i \xrightarrow{d} N\left[\mathbf{0}, \mathbf{\Omega}_i\left(\theta_{i0}\right)\right], \tag{11}$$

where $\mathbf{\Omega}_i\left(\theta_{i0}\right)$ is a positive definite matrix, $\xrightarrow{p}$ denotes convergence in probability and $\xrightarrow{d}$ convergence in distribution. The asymptotic properties of the Gaussian PMLE have been established for certain classes of multivariate GARCH-type volatility models (see Ling and McAleer (2001)) and it is reasonable to expect that results such as (10) and (11) would hold for the more general class of models considered in this paper, under suitable regularity conditions.[9]

In what follows we shall assume that under model M_i,

$$M_i: \quad \mathbf{r}_t = \mu_t + \mathbf{\Sigma}_{it}^{\frac{1}{2}}\mathbf{z}_{it}, \;\; \mathbf{z}_{it} \mid \mathcal{F}_{t-1} \sim (F_{it}, \; \mathbf{0}, \; \mathbf{I}_N), \tag{12}$$

meaning that $E\left(\mathbf{z}_{it} \left| \mathcal{F}_{t-1}, M_i\right.\right) = \mathbf{0}$, $E\left(\mathbf{z}_{it}\mathbf{z}_{it}' \left| \mathcal{F}_{t-1}, M_i\right.\right) = \mathbf{I}_N$, where $\mathbf{I}_N$ is the $N \times N$ identity matrix, and $F_{it}(.)$ is the the conditional joint probability distribution function of $\mathbf{z}_{it} = \mathbf{\Sigma}_{it}^{-\frac{1}{2}}\varepsilon_t$.

[8] An exception is the TDCC model, which is estimated under the assumption of a Student t distribution with ν degrees of freedom, where ν forms part of the parameter vector θ.

[9] Some of the models we consider do not require estimation. For instance Zaffaroni (2007) shows that the PMLE estimator of the Riskmetrics model fails even the consistency property.

Note that the above formulation allows the higher order moments of $\mathbf{z}_{it}$ to be time varying. This would be the case, for example, when $\mathbf{z}_{it}$ is distributed as the multivariate Student t with time varying degrees of freedom, v_{t-1}, conditional on $\mathcal{F}_{t-1}$.

4 Average Volatility Models

Considering the restrictive nature of the multivariate volatility models in the literature, model averaging techniques that explicitly allow for parameter and model uncertainty could be particularly important in risk management. Let $f(\mathbf{r}_t|\mathcal{F}_{t-1}, M_i)$ be the predictive density of $\mathbf{r}_t$ conditional on model M_i, $\mathcal{F}_{t-1}$ the in-sample available information, and $\mathcal{M} = \bigcup_{i=1}^m M_i$ the space of the models under consideration. Each *model* M_i is fully specified by the choice of the volatility model, $\mathbf{\Sigma}_{it}$, and of the conditional probability distribution, F_{it}, of devolatilized residuals, $\mathbf{z}_{it}$.

Model averaging implies a predictive density of $\mathbf{r}_t$ conditional on $\mathcal{F}_{t-1}$ given by

$$f(\mathbf{r}_t\,|\mathcal{F}_{t-1},\,\mathcal{M}) = \sum_{i=1}^{m} \lambda_{i,t-1} f(\mathbf{r}_t|\mathcal{F}_{t-1}, M_i), \tag{13}$$

where the set of weights $\lambda_{i,t-1}$ are pre-determined at the time the decision over the positions, $\omega_{j,t-1}$ $(j = 1, 2, ..., N)$, is taken. This is possible since it is assumed that there is no feedback from trade decisions to the probability models under consideration. One could consider attaching equal weights to all the models belonging to $\mathcal{M}$, yielding $\lambda_{i,t-1} = 1/m$. A further refinement would be to apply model averaging not to all of the models but only to a given number of top performing models. Therefore, one could pool different models by taking simple averages, but after 'trimming' models with poor past performances. Formally, this implies $\lambda_{i,t-1} = 1/n_{t-1}$ for $i \in \mathcal{N}_{t-1} \subset \mathcal{M}$, where n_{t-1} indicates the cardinality of the sequence of subsets of models $\mathcal{N}_{t-1}$. For $i \notin \mathcal{N}_{t-1}$, $\lambda_{i,t-1} = 0$. Such a procedure, often referred to as 'thick' modelling, has been proposed, among others, by Granger and Jeon (2004) who note that, standard two-stage procedures, such as selection methods based on the AIC or SBC, might exhibit poor performance simply because the 'true' model does not belong to the set of models under consideration.[10] Another example is the Bayesian Model Averaging (BMA) that combines the models under consideration using their respective posterior probabilities.[11] BMA requires $\lambda_{i,t-1} = \Pr(M_i\,|\mathcal{F}_{t-1})$, where the latter denotes the posterior probability of model M_i. The BMA approach requires specifications of the prior probability of model M_i and of the prior probability of θ_i conditional on M_i, for $i = 1, 2, ..., m$. BMA can be quite demanding computationally, particularly in the case of multi-variate volatility models with many unknown parameters. As a result, the model weights $\lambda_{i,t-1}$ are often approximated by the Akaike weights or the Schwartz weights. The

[10] See Stock and Watson (1999) for an application to macroeconomic time series and Aiolfi, Favero, and Primiceri (2001) for an application of 'thick' modelling to point forecasts of excess returns across different models.

[11] A formal Bayesian solution to the problem of model uncertainty is reviewed, for example, in Draper (1995) and Hoeting, Madigan, Raftery, and Volinsky (1999). Recent applications to time series econometrics are provided in Fernandez et al. (2001a,b), Garratt, Lee, Pesaran, and Shin (2003) and Godsill, Stone, and Weeks (2004).

latter gives a Bayesian approximation when the estimation sample, T_0, is sufficiently large.[12] In particular, setting $\lambda_{i,t-1} = \exp(\Delta_{i,t-1})/\sum_{j=1}^{m}\exp(\Delta_{j,t-1})$, in the case of AIC and SBC we have $\Delta_{i,t-1}= AIC_{i,t-1} - Max_j(AIC_{j,t-1}), \Delta_{i,t-1}= SBC_{i,t-1} - Max_j(SBC_{j,t-1})$, where in turn $AIC_{i,t-1} = LL_{i,t-1} - k_i$, $SBC_{i,t-1} = LL_{i,t-1} - \left(\frac{k_i}{2}\right)\ln(t-1)$, and $LL_{i,t-1}$ indicates the maximized logarithm of the joint probability distribution, with k_i parameters, of the observations $\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_{t-1}$ conditional on the given initial values $\mathbf{r}_0, ..., \mathbf{r}_{-s_i+1}$.[13]

In this paper, we implement both the 'thick' modelling and the (approximate) BMA procedures. The former is carried out by first ranking the individual models according the AIC or SBC criteria, and then constructing an 'average' model based on a given number of top-percentile (say the top 25%) of all the models under consideration. Therefore, we still make use of the information contained in AIC and SBC criteria, but only to trim-out the poorly performing models. Under this approach the models that survive will be given equal weights.

In contrast to applications that focus on point forecasts, in the case of density forecasting the choice of the number of models to be used in the model averaging process and the differences in their forecast error variances can have important implications for the shape of the resulting average model in general and the degree of its fat-tailness, in particular. Therefore, it seems likely that averaging across a very large number of models could be counter productive for density forecasting, whereas this might not be a problem for point forecasting. Further analysis of average models and their tail properties will be provided below in Section 5.3.

4.1 MV Optimization Subject to VaR Constraint in the Case of Average Models

Suppose the 'average' model is constructed using the probability weights, $\lambda_{i,t-1}$, applied, for example, to the following m Gaussian return distributions:

$$M_i : \mathbf{r}_t|\mathcal{F}_{t-1} \sim N(\mu_{it}, \mathbf{\Sigma}_{it}) \quad \text{for } i = 1, 2, ..., m. \tag{14}$$

The MV objective function in this case is given by

$$Q(\omega_{t-1}) = \omega'_{t-1}\bar{\mu}_t - \frac{\delta_{t-1}}{2}\omega'_{t-1}\bar{\mathbf{\Sigma}}_t\omega_{t-1},$$

where (see, for example, Draper (1995))

$$\begin{aligned} \bar{\mu}_t &= \sum_{i=1}^{m}\lambda_{i,t-1}\mu_{it}, \\ \bar{\mathbf{\Sigma}}_t &= \sum_{i=1}^{m}\lambda_{i,t-1}\mathbf{\Sigma}_{it} + \sum_{i=1}^{m}\lambda_{i,t-1}\left(\mu_{it} - \bar{\mu}_t\right)\left(\mu_{it} - \bar{\mu}_t\right)' \end{aligned}$$

[12] In the empirical applications to be discussed below T_0 is sufficiently large and parameter uncertainty is likely to be of second order importance. Also see Burnham and Anderson (1998, Chapter 4).

[13] We do, however, recognize that for small to moderate sample sizes used in macro-economic applications the choice of priors could be important, particularly if the object of exercise is the estimation of the marginal probability densities.

with the VaR constraint given by

$$\Pr(\omega_{t-1}'\mathbf{r}_t < -L_{t-1}|\mathcal{M},\mathcal{F}_{t-1}) = \sum_{i=1}^{m}\lambda_{i,t-1}\Phi\left(\frac{-\omega_{t-1}'\mu_{it}-L_{t-1}}{\sqrt{\omega_{t-1}'\boldsymbol{\Sigma}_{it}\omega_{t-1}}}\right) \leq \alpha, \tag{15}$$

where $\Phi(\cdot)$ is the distribution function of the standard normal variate.[14]

The Lagrangian for the above constrained optimization problem is given by

$$\begin{aligned}\mathcal{L}(\omega_{t-1},\psi_{t-1}) &= \omega_{t-1}'\bar{\mu}_t - (\delta_{t-1}/2)\omega_{t-1}'\bar{\boldsymbol{\Sigma}}_t\omega_{t-1} - \\ &\quad \psi_{t-1}\left\{\sum_{i=1}^{m}\lambda_{i,t-1}\Phi\left(\frac{-\omega_{t-1}'\mu_{it}-L_{t-1}}{\sqrt{\omega_{t-1}'\boldsymbol{\Sigma}_{it}\omega_{t-1}}}\right)-\alpha\right\},\end{aligned}$$

where ψ_{t-1} is the Lagrange multiplier which will be non-zero when the VaR constraint binds. The first-order necessary conditions for this optimization problem are given by

$$\frac{\partial\mathcal{L}(\omega_{t-1},\psi_{t-1})}{\partial\omega_{t-1}} = \bar{\mu}_t - \delta_{t-1}\bar{\boldsymbol{\Sigma}}_t\omega_{t-1} + \psi_{t-1}\left[\mathbf{g}_\mu(\omega_{t-1}) - \mathbf{G}_\sigma(\omega_{t-1})\omega_{t-1}\right] = 0, \tag{17}$$

$$\frac{\partial\mathcal{L}(\omega_{t-1},\psi_{t-1})}{\partial\psi_{t-1}} = \sum_{i=1}^{m}\lambda_{i,t-1}\Phi_i(\omega_{t-1}) - \alpha \leq 0, \tag{18}$$

and

$$\psi_{t-1}\frac{\partial\mathcal{L}(\omega_{t-1},\psi_{t-1})}{\partial\psi_{t-1}} = 0, \tag{19}$$

where

$$\begin{aligned}\mathbf{g}_\mu(\omega_{t-1}) &= \sum_{i=1}^{m}\frac{\lambda_{i,t-1}\phi_i(\omega_{t-1})\mu_{it}}{\left(\omega_{t-1}'\boldsymbol{\Sigma}_{it}\omega_{t-1}\right)^{1/2}},\\ \mathbf{G}_\sigma(\omega_{t-1}) &= \sum_{i=1}^{m}\frac{\lambda_{i,t-1}\phi_i(\omega_{t-1})\left(\omega_{t-1}'\mu_{it}+L_{t-1}\right)\boldsymbol{\Sigma}_{it}}{\left(\omega_{t-1}'\boldsymbol{\Sigma}_{it}\omega_{t-1}\right)^{3/2}},\\ \phi_i(\omega_{t-1}) &= \phi\left(\frac{-\omega_{t-1}'\mu_{it}-L_{t-1}}{\sqrt{\omega_{t-1}'\boldsymbol{\Sigma}_{it}\omega_{t-1}}}\right),\ \Phi_i(\omega_{t-1}) = \Phi\left(\frac{-\omega_{t-1}'\mu_{it}-L_{t-1}}{\sqrt{\omega_{t-1}'\boldsymbol{\Sigma}_{it}\omega_{t-1}}}\right),\end{aligned}$$

and $\phi(\cdot)$ is the density of the standard normal variate. The $m+1$ equations (17) and (18) in ω_{t-1} and ψ_{t-1} can be solved iteratively. Pre-multiplying (17) by ω_{t-1}' and solving for ψ_{t-1} in terms of ω_{t-1} we have

$$\psi_{t-1} = \frac{\omega_{t-1}'\bar{\mu}_t - \delta_{t-1}\omega_{t-1}'\bar{\boldsymbol{\Sigma}}_t\omega_{t-1}}{L_{t-1}\sum_{i=1}^{m}\lambda_{i,t-1}\phi_i(\omega_{t-1})\left(\omega_{t-1}'\boldsymbol{\Sigma}_{it}\omega_{t-1}\right)^{-1/2}} \geq 0. \tag{20}$$

[14] Alternatively, one could use any other set of return distributions, for example a set of t distribution with $\nu_{i,t-1}$ degrees of freedom ($\{T_{\nu_{i,t-1}}\}_{i=1}^{m}$). In this case the VaR constraint would be

$$\Pr(\omega_{t-1}'\mathbf{r}_t < -L_{t-1}|\mathcal{M},\mathcal{F}_{t-1}) = \sum_{i=1}^{m}\lambda_{i,t-1}T_{\nu_{i,t-1}}\left(\frac{-\omega_{t-1}'\mu_{i,t-1}-L_{t-1}}{\sqrt{\frac{\nu_{i,t-1}-2}{\nu_{i,t-1}}}\sqrt{\omega_{t-1}'\boldsymbol{\Sigma}_{i,t-1}\omega_{t-1}}}\right) \leq \alpha. \tag{16}$$

Solving for ω_{t-1} in terms of ψ_{t-1}

$$\omega_{t-1} = \left[\delta_{t-1}\bar{\mathbf{\Sigma}}_t + \psi_{t-1}\mathbf{G}_{\sigma}(\omega_{t-1})\right]^{-1}\left[\bar{\mu}_t + \psi_{t-1}\mathbf{g}_{\mu}(\omega_{t-1})\right]. \tag{21}$$

One could then check to see if the solution to the unconstrained problem, namely $\tilde{\omega}^*_{t-1} = (\delta_{t-1}\bar{\mathbf{\Sigma}}_t)^{-1}\bar{\mu}_t$, satisfies the VaR constraint (16). If affirmative, set $\omega^*_{t-1} = \tilde{\omega}^*_{t-1}$. Otherwise, use a standard root-finding algorithm such as the secant method[15] to search over different values of ψ_{t-1} in order to find a pair $(\psi^*_{t-1}, \omega^*_{t-1})$, such that ω^*_{t-1} is a function of ψ^*_{t-1} (via equation (21)) and satisfies the VaR constraint with equality.

5 Value-at-Risk Based Diagnostic Tests

This section examines the evaluation of multivariate volatility models from the perspective of risk management. First we consider the problem for a given model, M_i. Next, we describe how the analysis can be extended to models obtained by application of model averaging techniques.

5.1 VaR Diagnostics for Individual Models

In the econometric literature models are often evaluated by their out-of-sample forecast performance using standard metrics such as the RMSFE but, as noted earlier, the application of this approach to volatility models is subject to a number of difficulties. An alternative approach would be to employ decision-based evaluation techniques and compare different volatility models in terms of their performance in trading and risk management.[16] In this sub-section we propose simple examples of such a procedure based on the VaR problem set out in Section 2.

Consider first the VaR constraint (4) associated with the passive version of the risk management problem where the portfolio exposures, ω_{t-1}, are given, and suppose that the analysis is carried out conditional on model M_i. In this setting the VaR constraint becomes

$$\Pr\left(\rho_t < -\bar{\rho}_{i,t-1} \left| \mathcal{F}_{t-1}, M_i \right.\right) \le \alpha, \tag{22}$$

where $\bar{\rho}_{i,t-1}$ will be a function of α and the assumed volatility model, M_i. To fully specify the model, assume that the standardized returns, $\mathbf{z}_{it}$, have a joint cumulative distribution function $F_{it}(\cdot)$ which is closed under linear combinations so that $\mathbf{c}'\mathbf{z}_{it}$ also has (univariate) distribution $F_{it,\rho}(\cdot)$ of the same type for any fixed N-dimensional vector $\mathbf{c}$. A special case of our results is obtained if $\mathbf{z}_{it}$ is assumed to follow the multivariate normal or the Student t distribution. Conditional on $\mathcal{F}_{t-1}$ and model M_i being true, ρ_t will have mean $\mu_{\rho t} = \omega'_t\mu_t$ and variance $\sigma^2_{\rho t}(M_i) = \omega'_{t-1}\mathbf{\Sigma}_{it}\,\omega_{t-1}$. Therefore, under (12) we have

$$z_{\rho t}(M_i) = \frac{\omega'_{t-1}(\mathbf{r}_t - \mu_t)}{\sigma_{\rho t}(M_i)} \left| \mathcal{F}_{t-1}, M_i \right. \sim (F_{it}, 0, 1). \tag{23}$$

[15] See e.g. Burden and Faires (1997) for a description of the secant method.

[16] For a general discussion of decision-based evaluation techniques see Pesaran and Skouras (2002).

This implies that under M_i, $z_{\rho t}(M_i)$ is a martingale difference sequence with unit variance. Note, however, that $z_{\rho t}(M_i)$ need not be independent across time. Temporal dependence in $z_{\rho t}(M_i)$ could arise not only due to possible higher-order moment dependence of the underlying innovations $\mathbf{z}_{it}$, but also because of possible serial dependence of portfolio weights and the temporal dependence of $\mathbf{\Sigma}_{it}$.

Denoting the maximum value of $\bar{\rho}_{i,t-1}$ that satisfies (22) by $\bar{\rho}_{i,t-1}(\omega_{t-1},\alpha)$ and assuming that (23) holds, then $F_{it}\left((-\bar{\rho}_{i,t-1}(\omega_{t-1},\alpha) - \omega'_{t-1}\mu_t)\sigma^{-1}_{\rho t}(M_i)\right) = \alpha$. But since $F_{it}(\cdot)$ is a continuous and monotonically non-decreasing function we have $(-\bar{\rho}_{i,t-1}(\omega_{t-1},\alpha) - \omega'_{t-1}\mu_t)\sigma^{-1}_{\rho t}(M_i) = F^{-1}_{it}(\alpha) = -c_{it}(\alpha)$, or

$$\bar{\rho}_{i,t-1}(\omega_{t-1},\alpha) = -\omega'_{t-1}\mu_t + c_{it}(\alpha)\sigma_{\rho t}(M_i), \tag{24}$$

where $-c_{it}(\alpha)$ is the α per cent critical value of the distribution of $z_{\rho t}(M_i)$ conditional on model M_i and $\mathcal{F}_{t-1}$. Note that $c_{it}(\alpha)$ and $\sigma_{\rho t}(M_i)$ are based on observations available at time $t-1$, and this is highlighted in the notation used for $\bar{\rho}_{i,t-1}(\omega_{t-1},\alpha)$.[17]

The evaluation of model M_i can now proceed in the following manner. Suppose that the evaluation exercise starts on day $t = \tau + 1$ with the available sample of T observations split at this date into $T = T_0 + (T - T_0)$ for some $0 < T_0 < T$. Further suppose that the first T_0 observations before day $\tau + 1$ are used for estimation whereas the last $T_1 = T - T_0$ observations are used for evaluation purposes. Accordingly, we define the sets of estimation and evaluation dates by $\mathcal{T}_0 = \{\tau - T_0 + 1, \tau - T_0 + 2, ..., \tau\}$, and $\mathcal{T}_1 = \{\tau + 1, \tau + 2, ..., \tau + T_1\}$, respectively.

A simple test of the validity of model M_i from the perspective of the VaR can then be based on the proportion of days in the evaluation sample where the VaR constraint is violated: $\hat{\pi}_i = \sum_{t\in\mathcal{T}_1} d_{it}(\hat{\theta}_{iT_0})/T_1$, where $d_{it}(\hat{\theta}_{iT_0}) = I[-\rho_t + \omega'_{t-1}\mu_t - c_{it}(\alpha)\,\hat{\sigma}_{\rho t}(M_i)]$ and $\hat{\sigma}_{\rho t}(M_i) = (\omega'_{t-1}\hat{\mathbf{\Sigma}}_{it}\,\omega_{t-1})^{\frac{1}{2}}$, $\hat{\mathbf{\Sigma}}_{it} = \mathbf{\Sigma}_{it}(\hat{\theta}_{iT_0})$. Recall that $\hat{\theta}_{iT_0}$ is the PMLE of the unknown parameters (if any) of $\mathbf{\Sigma}_{it}$ under model M_i (see (9)), and $I(\cdot)$ as an indicator function.

We now present two Theorems. The first establishes the distribution of $T_1\hat{\pi}_i$ under the null hypothesis defined by

$$H_{i0}:\ \mathbf{\Sigma}_t = \ \mathbf{\Sigma}_{it} \text{ and } \ \mathbf{z}_{it} \mid \mathcal{F}_{t-1}, M_i \sim (F_{it},\ \mathbf{0}, \mathbf{I}_N). \tag{25}$$

for $T_1 < \infty$ and as $T_0 \to \infty$. The second Theorem establishes the asymptotic distribution of the following standardized test statistic based on $\hat{\pi}_i$

$$z_{\hat{\pi}_i} = \frac{\sqrt{T_1}(\hat{\pi}_i - \alpha)}{\sqrt{\alpha(1-\alpha)}} \tag{26}$$

under H_{i0}, and as $T_1/T_0 + 1/T_1 \to 0$. The proofs of both Theorems are provided in Appendix C.

[17] The above derivations hold even if the portfolio exposures, ω_{t-1}, are derived conditional on model M_i. In that case the portfolio weights could be denoted by $\omega_{i,t-1}$ to highlight their dependence on the choice of the volatility model. But to simplify the notations we continue to represent the portfolio weights without the subscript i.

Theorem 1 *(finite-T_1 distribution) Assume that* $\mathbf{\Sigma}_{it}(\theta_i)$ *is continuous in* θ_i *and that (11) holds. Let* $Bi(T_1, \alpha)$ *define a Binomial distribution with parameters* T_1 *and* α*. Then under* H_{i0}*,*

$$T_1 \hat{\pi}_i \xrightarrow{d} Bi(T_1, \alpha), \qquad as\ T_0 \to \infty, \tag{27}$$

for any finite T_1*,* $0 < \alpha < 1$*, and any sequence of portfolio exposures,* ω_{t-1}*,* $t = 0, \pm 1...$*, satisfying* $\| \omega_{t-1} \| > 0$*, with* $\| \cdot \|$ *being the Euclidean norm.*

Remark. This result is important for cases when T_1 is small or, alternatively, when one is interested in testing VaR performance of a given set of portfolios for small values of α. In such cases the asymptotic normal distribution presented below might not provide a sufficiently accurate approximation.

Theorem 2 *(asymptotic distribution) Assume that (i)* $f_{it}(\cdot) = F'_{it}(\cdot)$ *exists and* $\bar{f}_{it} = \sup_x f_{it}(x) < \infty$ *for any* t*; (ii) condition (11) holds and* θ_{i0} *belongs to the interior of the compact set* Θ_i*; (iii)* $\mathbf{\Sigma}_{it}(\theta_i)$ *is twice continuously differentiable in* θ_i *such that, for some* $\delta > 1$*,* $\inf_{\theta_i \in \Theta_i} \underline{\lambda}_{it}(\theta_i) > 0$*, a.s.*

$$E\{\sup_{\theta \in \Theta_i} \frac{\| \partial \bar{\lambda}_{it}(\theta)/\partial \theta \|}{\underline{\lambda}_{it}^{\frac{1}{2}}(\theta) \underline{\lambda}_{it}^{\frac{1}{2}}(\theta_{i0})}\}^{\delta} = \mu_{it}, \ \ \frac{1}{T_1} \sum_{t \in \mathcal{T}_1}^{T} \bar{f}_{it} \mu_{it}^{1/\delta} = O(1), \tag{28}$$

where $\bar{\lambda}_{it}(\theta_i)$ *and* $\underline{\lambda}_{it}(\theta_i)$ *define, respectively, the maximum and the minimum eigenvalues of* $\mathbf{\Sigma}_{it}(\theta_i)$*, (iv) for* T_0 *sufficiently large*

$$E \| \hat{\theta}_{iT_0} - \theta_{i0} \|^{\frac{\delta}{\delta-1}} = O(T_0^{-\delta/(2(\delta-1))}). \tag{29}$$

Under H_{i0}*,* $z_{\hat{\pi}_i} \xrightarrow{d} N(0,1)$ *as* $T_1/T_0 + 1/T_1 \to 0$*, any* $0 < \alpha < 1$*, for any sequence of portfolios* ω_{t-1}*,* $t = 0, \pm 1...$*, satisfying* $\| \omega_{t-1} \| > 0$.

Remarks:

(i) It is important to note that the null distribution of $z_{\hat{\pi}_i}$ does not depend on the portfolio exposures, ω_{t-1}, although the power of the test typically does depend on ω_{t-1}.

(ii) The mild condition for consistency of the test is that $\hat{\pi}_i$ does not converge in probability to α as $T_1/T_0 + 1/T_1 \to 0$. This can happen if either we use the wrong conditional covariance matrix or the wrong innovation distribution, or both. For example, in the first case, under $M_j : \mathbf{\Sigma}_{jt} \neq \mathbf{\Sigma}_{it}$ we have $E(\hat{\pi}_i | M_j) = \frac{1}{T_1} \sum_{t \in \mathcal{T}_1} E[F_{it}(-c_{it}(\alpha) q_{ij,t})]$, where $q_{ij,t} = (\omega'_{t-1} \hat{\mathbf{\Sigma}}_{it} \omega_{t-1} / \omega'_{t-1} \mathbf{\Sigma}_{jt} \omega_{t-1})^{1/2}$, for $t \in \mathcal{T}_1$. It is clear that under M_j, $q_{ij,t}$ does not tend to unity and in general $E(\hat{\pi}_i | M_j)$ will diverge from its hypothesized value of α, and the power of the test tends to unity with T_1.

(iii) Most likely, the assumptions required for (10) and (11) will imply (28) but we felt it is necessary to make the additional explicit assumptions since the former have been formally established only for a sub-class of multivariate volatility models considered in this paper.

(iv) When model M_i is not subjected to estimation, as is the case for some for some of the models we consider, such as the Riskmetrics model, then the Theorem applies by setting $\hat{\theta}_i = \theta_{i0}$ and the conditions (28) and (29) are no longer needed. In particular, the non-singularity condition of the model conditional covariance matrix is not required.

(v) Under the null hypothesis $H_{i0} : E(z_{\rho t}(M_i) \mid \mathcal{F}_{t-1}) = 0$. This is a key property since it implies that $I\left[-z_{\rho\, t}(M_i) - c_{it}(\alpha)\right] - \alpha$ is also a martingale difference process. Strict stationarity of the asset returns is not required.
(vi) The importance of the condition $T_1/T_0 \to 0$ in cross validation of forecasts was put forward by West (1996). McCracken (2000) extends West's framework to allow for non-differentiable loss functions in a regression set-up.

5.2 VaR-Based Diagnostics for Average Models

Suppose that set of m models is described by $\mathbf{r}_t|\mathcal{F}_{t-1}, M_i \sim (F_{it}, \mu_t, \mathbf{\Sigma}_{it})$ for $(i = 1, 2, ..., m)$. Therefore, $F_{it}(\cdot)$ defines the conditional distribution of the observed return $\mathbf{r}_t$, given $\mathcal{F}_{t-1}$ and the volatility model M_i.

The probability distribution function of the portfolio return, ρ_t, based on the average model obtained with respect to these models using the weights, $\lambda_{i,t-1}$, is then given by $\Pr(\rho_t < a\,|\mathcal{F}_{t-1}, M) = \sum_{i=1}^{m} \lambda_{i,t-1} F_{it}\left(\frac{a}{\sigma_{\rho t}(M_i)}\right)$. In cases where $\Pr(\rho_t < a\,|\mathcal{F}_{t-1}\,, M_i)$ does not have a closed form it needs to be computed by stochastic simulations, noting that conditional on model M_i we have, $J^{-1}\sum_{j=1}^{J} I(-\omega_{t-1}'\mathbf{r}_{jt}^{(i)} + a) \to \Pr(\rho_t < a\,|\mathcal{F}_{t-1}\,, M_i)$ almost surely as $J \to \infty$, where J is the number of replications and $\mathbf{r}_{jt}^{(i)}$ is the j^{th} draw from the assumed distribution of $\mathbf{r}_t$ under M_i. On the other hand, when the probability distribution of $\mathbf{r}_t$ under M_i is closed under linear transformations, as with Gaussian or multivariate t distribution, the computations can be simplified considerably by drawing from the distribution of $\rho_t = \omega_{t-1}'\mathbf{r}_t$ under M_i directly or using the closed-form expression when the latter exists.

It is now easy to generalize the diagnostic test statistics given by (26) for an individual model M_i, to the case of an average model. For a given α we need to find the maximum value $\bar{\rho}_{b,t-1}(\omega_{t-1}, \alpha)$, the VaR associated with the BMA forecast probabilities, for which $\sum_{i=1}^{m} \lambda_{i,t-1} F_{it}\left[\left(-\bar{\rho}_{b,t-1}(\omega_{t-1}, \alpha) - \omega'\mu_{it}\right)/\sigma_{\rho t}(M_i)\right] \leq \alpha$. To solve for $\bar{\rho}_{b,t-1}(\omega_{t-1},\alpha)$, let

$$g(\kappa) = \sum_{i=1}^{m} \lambda_{i,t-1} F_{it}\left(-\frac{\kappa + \omega'\mu_{it}}{\sigma_{\rho t}(M_i)}\right) - \alpha = 0, \tag{30}$$

and note that $g(\kappa) = 0$ has a unique positive solution under the additional assumptions that α is sufficiently small such that $g(0) > 0$, and the model densities $f_{it}(\cdot) = F_{it}'(\cdot)$ exist and $f_{it}(\cdot)$ is continuous and strictly positive for some i.[18] In this case $\bar{\rho}_{b,t-1}(\omega_{t-1},\alpha)$ can be easily computed using numerical techniques such as the Newton-Raphson iterative procedure. The VaR diagnostic statistic, given by (26), can then be computed for the average model using $\hat{d}_{bt} = I\left[-\rho_t - \bar{\rho}_{b,t-1}(\omega_{t-1},\alpha)\right]$, in place of $d_{it}(\hat{\theta}_{iT_0})$.

5.3 Tail Behavior of Average Volatility Models

It is well known that linear combinations (mixtures) of normal distributions are not normal, although the moments of the mixture distribution are effectively linear combinations of the corresponding moments of the individual normal distributions, with the same weights. For

[18] This result follows by noting that under the additional assumptions $g'(\kappa) < 0$, and $\lim_{\kappa\to\infty} g(\kappa) = -\alpha < 0$.

instance, the pooled volatility forecast of portfolio returns with zero conditional means is given by $V(\rho_t|F_{t-1},\mathcal{M}) = \sum_{i=1}^{m} \lambda_{it-1}\sigma^2_{\rho t}(M_i)$. However, tail probabilities using the mixture model and a Gaussian model with the same average volatility are not the same, namely

$$\sum_{i=1}^{m} \lambda_{it-1}\Phi\left[\frac{a}{\sigma_{\rho t}(M_i)}\right] \neq \Phi\left[\frac{a}{\sqrt{\sum_{i=1}^{m}\lambda_{it-1}\sigma^2_{\rho t}(M_i)}}\right], \tag{31}$$

unless $\Sigma_{it} = \Sigma_t$ for all i, where $\Phi(\cdot)$ defines the normal cumulative distribution function. The following Theorem, whose proof is reported in Appendix C, characterizes the direction of the bias. In risk management applications where $a < 0$ and one is interested in tail probabilities, it is easily seen that the correctly combined model, on the left hand side of (31), will be more fat-tailed than the associated Gaussian model with the same average volatility measure, on the right hand side of (31), so long as $a < -\sqrt{3}\sigma_{\rho t}(M_i)$, $i = 1, ..., m$. As we shall see this result has direct bearing on some of the empirical results that we shall be reporting below.

Theorem 3 *Let $f(x)$ be a differentiable real function, with f' denoting its first-derivative, with $\int_{-\infty}^{\infty} \mid f(u) \mid du < \infty$. Let $F(z) = \int_{-\infty}^{z} f(u)du$. Then, for any constant a and any finite sequence $b_1, b_2, ..., b_m$ of strictly positive constants satisfying*

$$a\left[(a/b_i^{\frac{1}{2}})f'(a/b_i^{\frac{1}{2}}) + 3f(a/b_i^{\frac{1}{2}})\right] > 0, \quad i = 1, 2, ..., m, \tag{32}$$

it follows that

$$\sum_{i=1}^{m} \lambda_i F\left[a/(b_i)^{\frac{1}{2}}\right] > F\left[a/(\sum_{i=1}^{m}\lambda_i b_i)^{\frac{1}{2}}\right], \tag{33}$$

for any finite sequence $\lambda_1, \lambda_2, ..., \lambda_m$ of non-negative constants such that $\lambda_1 + \lambda_2 + ... + \lambda_m = 1$, $\lambda_i < 1$, $i = 1, 2, ..., m$.

Remarks:
(i) When $f(u)$ is the standard normal density, for $a < 0$ condition (32) is

$$a/b_i^{\frac{1}{2}} < -\sqrt{3}, \quad i = 1, ..., m. \tag{34}$$

When $a > 0$ condition (32) is instead $0 < a/b_i^{\frac{1}{2}} < \sqrt{3}$ $(i = 1, 2, ..., m)$, although note that when $a > 0$, (33) expresses the case where the tail probability of the average model is smaller than for the model with the average parameter $\sum_{i=1}^{m}\lambda_i b_i$.
(ii) When $f(u)$ is the standardized Student t distribution with $\nu > 2$ degrees of freedom, for $a < 0$ the same condition (34) applies, independently from ν.

6 An Empirical Application

6.1 Data and Some Preliminary Analysis

The active and passive risk management procedures in the presence of model uncertainty developed in this paper can be applied to a variety of problems in finance. Here we shall consider a global macro portfolio of 18 futures contracts grouped into four equity futures indices (S&P, FTSE, DAX, NIKKEI), six currencies (GBP, EUR, JPY, CAD, AUD, CHF), four 10 year government bonds (US, EUR, Gilt, JGB), and four commodities (Gold, Silver, Wheat, Crude), yielding a reasonably diverse global macro portfolio. The overall portfolio return is measured in US dollar, with currencies defined as the number (fraction) of US dollars per unit of the foreign currency. The returns are daily and cover the period 2 January 1991 to 11 July 2007 (a total of $T = 4311$ daily observations). The source of the data is Datastream with returns on the futures contracts appropriately adjusted for roll overs. Since we are considering markets with different time zones and holidays, the return data are aligned by filling forward the missing asset prices due to differences in holidays in the US, euro area and Japan.

[Insert Table 1 around here.]

Daily returns are computed as $r_{jt} = 100\left(P_{jt} - P_{jt-1}\right)/P_{jt-1}$, $j = 1, ..., 18$, where P_{jt} is the j^{th} asset price. Table 1 gives the mean, standard deviation, skewness and kurtosis of asset returns together with estimates of a $t-$GARCH(1,1) model fitted to the individual returns over the full sample. The returns $\mathbf{r}_t = (r_{1t}, r_{2t}, ..., r_{18,t})'$ display the familiar stylized features – namely little evidence of skewness, possibly with the exception of JPY, Silver and Crude, but a substantial degree of fat-tailedness as measured by excess kurtosis. There are also important differences in the unconditional volatilities across asset classes, with bonds being least volatile followed by currencies, equities and commodities. The estimates of univariate t-GARCH models show a high degree of volatility persistence with the sum of the coefficients of $r^2_{i,t-1}$ and $\sigma^2_{j,t-1}$ being very close to unity. The estimates are also very similar across assets. The degrees of freedom of the Student t distribution, v, assumed for the innovations were closely clustered across assets, and ranged from 4.5 for Japanese Yen to 11.5 for FTSE with an average of 6.5, suggesting a significant degree of departure from normality, partly reflecting the relatively large estimates obtained for the kurtosis coefficients.

[Insert Table 2 around here.]

The unconditional return correlations across assets and asset classes are summarized in Table 2. The results show a relatively high degree of average pairwise correlations for assets within a given asset class and a relatively low average correlation across the asset classes with a few notable exceptions. Not surprisingly, gold and silver futures have a relatively high correlation with currencies, and amongst bonds, JGB is only weakly correlated with the returns on other bond futures.

These results further highlight the non-Gaussian nature of asset returns. But estimation of multivariate volatility models with non-Gaussian distributions present considerable technical difficulties and are unlikely to significantly affect the QMLE estimates that are computed assuming Gaussian errors. For risk management purposes, it seems justified to combine the QMLE estimates with multivariate Student t distributions with a low degree of freedom. Therefore, based on the univariate t-GARCH estimates we also consider multivariate volatility models where the innovations are t distributed with 7 degrees of freedom. This approach is followed for all the empirical results to be reported below, except for the TDCC model of Pesaran and Pesaran (2007) where the degrees of freedom of the underlying multivariate t distribution are estimated recursively.

6.2 Recursive Estimation of Multivariate Volatility Models

For each of the 9 types of multivariate volatility models listed in Appendix B, a number of variations were considered, depending on the choice of the window size (n_0) when applicable, the pre-specified parameters of the Riskmetrics specifications (λ_0, ν_0) and the orders of the multivariate GARCH models (p, q, r, s). In particular, we considered the following parameter values $n_0 = 50, 75, 125, 250$, $\lambda_0 = 0.94, 0.95, 0.96$, $\nu_0 = 0.6, 0.8, 0.94$, $p, q \in \{1, 2\}$ and $r = s = 1$.

To estimate the volatility models, we first obtained recursive forecasts of the individual mean returns using the AR(1) specification defined by (7), which we denote by $\hat{\mu}_{jt}$, for $j = 1, 2, ..., 18$. These AR specifications were estimated each day using a rolling window of size 800. The AR(1) autocorrelation coefficients of the individual returns were quantitatively small (ranging from -0.20 to 0.15 across all the assets and over the whole sample period), and were on average negative, suggesting some degree of market over-reaction.

The multivariate volatility models were estimated (when applicable) using the one-day ahead forecast errors, $\hat{\varepsilon}_{jt} = r_{jt} - \hat{\mu}_{jt}$, $j = 1, 2, ..., 18$, based on rolling samples of size 800 days. The re-estimations were carried out every 25 days. The first rolling sample covered the period 2 January 1991 to 25 January 1994, and the last estimation sample covered the period 17 June 2004 to 11 July 07; namely a total of 3512 rolling samples of size 800. Clearly, the parameters of the volatility models could have been also updated daily. The monthly updates of the parameters can be viewed as a plausible and practical solution to a highly computer intensive problem. Therefore, the models were estimated 144 times over the evaluation sample. Interestingly enough, the estimation procedure converged in the case of all volatility models with the exception of the ADCC models where they failed to converge in one sample period. For this period the parameters of the ADCC models were set equal to the ones obtained in the previous sample period.

6.3 Modelling Strategies

A number of different modelling strategies may now be considered. One possibility would be to follow the classical approach and select the 'best' model from the set of models under consideration using model selection criteria such as AIC or SBC. Alternatively, the model uncertainty can be explicitly taken into account using 'thick' modelling or Bayesian type model averaging procedures. The former is implemented here using the top 10%, 25%, 50% and 75% of the models selected according to AIC or SBC. We refer to these as 'best', 'thick average', and 'Bayesian average' modelling strategies. As an extreme benchmark we also consider an equal-weighted average model using all the 53 specifications.

[Insert Tables 3 to 6 around here.]

In-sample penalized measures of fit for the various multivariate volatility models (as set out in Appendix B) for a selection of rolling windows are summarized in Tables 3 to 6. Tables 3 and 4 gives the AIC and SBC values assuming Gaussian innovations, whilst the results in Tables 5 and 6 report the in-sample measures assuming Student t innovations with 7 degrees of freedom. Each table gives the AIC (or SBC) values of the different models together with their rank in parentheses (1 for the best fitting model and 53 for the worst one) for the first sample (2-Jan-91 to 25-Jan-94, 800 observations), for the last sample, (17-June-04 to 11-July-07, 800 observations), and for the average AIC (or SBC) values over all 3,512 rolling samples.

Overall the DCC type models performed best, followed by CCC, and OGARCH specifications. Amongst the Riskmetrics type specifications, the simplest of the data filters namely the equal weighted moving average specification, EQMA, with $n_0 = 125$, or 250, do considerably better than the other filters and perform well even when compared to estimated models such as O-GARCH. Out of all the models considered the TDCC specification performed best irrespective of the penalization criteria, sample period or assumptions about the innovations. The better performance of the TDCC model could be partly due to the fact that the degrees of freedom of the underlying t distribution is updated every month rather than being set to 7 as in the case of the other specifications reported in Tables 5 and 6. However, it is unlikely that the large differences that exist between the AIC (or SBC) values of TDCC(1,1) and the second best model, ADCC (1,1), could be only due to the differences in the estimates of the degrees of freedom. Note that the average value of the AIC (across all the 144 rolling samples) for the TDCC model is -8676 as compared to the value of -8756 obtained for the ADCC with Student t (7) innovations (see Table 5). The apparent superiority of the TDCC over the other models is likely to be due to the way conditional correlations are defined in terms of de-volatized returns. As shown in Pesaran and Pesaran (2007) standardizing returns by realized volatility (estimated using daily returns) yields approximately Gaussian processes with respect to which correlations are likely to be more meaningful measures of dependence as compared to the standardization of returns by conditional volatilities as utilized in DCC type models.

The fact that the DCC models, and in particular the TDCC version, dominate the other models in terms of AIC or SBC also means that in Bayesian type model averaging the best model in the set of models under consideration will tend to get a weight that could be very close to unity. In the case of financial applications where the sample sizes are relatively large, the best model could totally dominate the other models. In our application where the average AIC (or SBC) value of the TDCC model exceeds the next best model by 80, and considering that the computation of posterior model probability weights involve exponentiating these differences, we typically end up giving a weight of unity to the best model and zero to the other models.[19] As a result, as we shall see below, portfolio outcomes and VaR diagnostics are almost identical for the best model and Bayesian type model averaging strategies.

6.4 Active Management: Performance of Optimal Portfolios and VaR Diagnostic Test Results

In this section we provide an out of sample, decision-based comparative analysis of the different multivariate volatility models, and different average models based on them. We use the same recursively computed one-step ahead mean forecasts, $\hat{\mu}_t = (\hat{\mu}_{1t}, \hat{\mu}_{2t}, ..., \hat{\mu}_{18,t})'$ in the case of all the 53 individual multivariate volatility models and their various averages.

[19]Notice that the model weights are obtained by exponentiation of the AIC-penalized log-likelihood values and even seemingly small differences in the average fit of the models can translate into major differences in model weights for sufficiently large sample sizes. See also Garratt, Lee, Pesaran, and Shin (2003).

In this way we are able to focus on the uncertainty of multivariate volatility models and abstract from the uncertainty associated with the mean returns. For each multivariate volatility model, M_i, we estimated the one-day ahead recursive forecasts of $\mathbf{\Sigma}_t$ denoted by $\hat{\mathbf{\Sigma}}_{it}$, $i = 1, ..., 53$. Using $\hat{\mu}_t$ and $\hat{\mathbf{\Sigma}}_{it}$, and for a given assumption regarding the distribution of innovations (Gaussian or a Student t with 7 degrees of freedom) we then computed the optimal portfolio weights, $\hat{\omega}^*_{t-1,M_i}$ using the closed form solution given by (6) and setting $\alpha = 1\%$, $\delta_{t-1} = 75$, and $L_{t-1} = 1\%$. Recall that α is the risk tolerance probability of the fund manager and L_{t-1} is the maximum permitted daily loss defined as the fraction of the notional capital, and δ_{t-1} is the coefficient of risk aversion. We calibrated δ in order to achieve a reasonable fraction of times where the VaR part of the optimization will be binding.

For each set of portfolio weights, we then computed the portfolio returns, $\rho_{t,M_i} = \mathbf{r}'_t \hat{\omega}^*_{t-1,M_i}$, and the associated performance statistics: the mean, standard deviation, the Information Ratio (IR), defined as the ratio of the mean to the standard deviation. All these statistics were computed recursively over the evaluation sample from 26-Jan-94 to 11-Jul-07, inclusive (3511 data points).

6.4.1 Individual Volatility Models

[Insert Table 7 around here.]

Table 7 summarizes the results for the individual volatility models and gives the annualized mean return, the IR and VaR diagnostic test statistics. These statistics are provided for Gaussian and Student $t(7)$ innovations. The percentage of times the VaR constraint (5) binds for the optimum solution is also given.

The results differ markedly across models, which highlights the important role the choice of multivariate volatility model can play in portfolio management. We also note that the VaR constraint tends to bind more often in the case of the Riskmetric filters relative to the OGARCH or DCC type models. The VaR constraint also binds more frequently when the innovations are t distributed.

The trading performance of the different volatility models, as measured by IR, vary considerably from a low of 0.61 for the MMA(250,0.95) specification with Gaussian innovations to a high of 1.51 for the CCC(1,2) specification with $t(7)$ innovations. Nevertheless, it is interesting that all volatility models generate a positive IR, despite the relatively simple model assumed for the return processes. Amongst the Riskmetric type specifications the simple EQMA filters performed best, which is in accordance with the in-sample results discussed above. The IR of the portfolios constructed assuming t-distributed innovations were also generally higher than those based on Gaussian innovations, although the magnitude of the difference is not that large. The TDCC model, which had performed best in-sample, continued to perform well in trading. However, the differences between models in trading turned out to be considerably less pronounced compared to their differences in terms of the statistical measures of penalized in-sample fit.

We next turn to the VaR diagnostics and use the portfolio returns for each volatility model, M_i, to compute (i) $\hat{\pi}_i$, the percentage of times the VaR constraint was violated

($\rho_{t,M_i} < L_{t-1} = 0.01$) and (ii) $z_{\hat{\pi}_i}$, the VaR diagnostic statistics defined by

$$z_{\hat{\pi}_i} = \frac{\sqrt{3511}(\hat{\pi}_i - .01)}{\sqrt{0.01 \times 0.99}}.$$

Under the null hypothesis that the underlying volatility model is correctly specified, $z_{\hat{\pi}_i}$ is approximately distributed as a standard normal variate. Since the parameters are estimated recursively every $T_1 = 25$ days which is small relative to the estimation sample of $T_0 = 800$, the conditions of the Theorem 1regarding T_1 and T_0 is likely to hold. The values of $\hat{\pi}_i$, and $z_{\hat{\pi}_i}$ for all the 53 individual volatility models are also summarized in Table 7.

The estimates of π_i are all biased upward, and are scattered over a relatively wide range, from a high value of 13.04% for the MMA(50,0.95) with Gaussian innovations to a low of 1.40 for DCC(1,1) with $t(7)$ innovations. Once again the DCC type models do considerably better than the Riskmetric filters in controlling the rate of VaR exceedences. The choice of the multivariate t distribution for the innovations helps reducing the bias of $\hat{\pi}_i$ for all i, but does not eliminate it. The null hypothesis that $\pi_i = 0.01$ is rejected at the 95% significance level for all individual volatility models under consideration.

We obtained similar results for other portfolios, although the extent of over-rejection were somewhat lower for the Riskmetric filters when an equal weighted portfolio was used. This is in line with the results in Theorems 1 and 2 where the distribution of the VaR diagnostic test is invariant to the choice of the portfolio weights, ω_{t-1}.

6.4.2 Modelling Strategies

As has been emphasized in this paper there are many ways in which the results from the 53 individual volatility models can be used/combined. We refer to these as *strategies* and distinguish between the standard classical strategy where the 'best' in-sample fitting model is selected, and alternative strategies that consist of combinations of the models. In particular we shall consider both Bayesian type and 'thick' model averaging procedures discussed in Section 4. In the case of thick modelling we focus on the top 10%, 20%, 50% and 75% of models ranked by AIC or SBC. The top echelon of selected models are then given equal weights in the averaging process. Finally, we included all the 53 models in an average 'All' group strategy.

[Insert Table 8 around here.]

For the model averaging strategies the optimal portfolio weights are computed iteratively as set out in Section 4.1, and the VaR diagnostic statistics are then computed with the help of (30) derived in Section 5.2. All computations are carried out recursively over the evaluation sample as in the case of the individual volatility models. The results are summarized in Table 8, using Gaussian and Student $t(7)$ distributions. Given the in-sample dominance of the DCC type models there are no differences in the test results for the average 'Bayesian' and the 'best' modelling strategies. As noted earlier, this is due to the fact that for almost all periods in the evaluation sample the 'best' model happens to totally dominate all other models, and as a result the average 'Bayesian' and the best models end up being the same for all practical purposes. This result suggests that the potential risk diversification benefits of Bayesian model averaging might be limited in financial applications where the available time series samples are typically rather large.

A comparison across strategies shows that all procedures yield very similar IRs, with thick strategies doing generally better than the 'best' models. The use of Student t innovations also seem to help improve the IRs irrespective of the strategy, although the 'best' models benefit more from switching to Student t innovations than to the thick modelling approaches. In terms of VaR exceedences again the thick modelling strategies tend to have lower rejection frequencies than the 'best' or the Bayesian average strategies. Amongst the various models and model strategies considered, the top 25% AIC thick modelling strategy with Student $t(7)$ innovations is the only strategy that yields VaR exceedences that are statistically insignificant ($\hat{\pi} = 1.20$ and $z_{\hat{\pi}} = 1.17$) whilst maintaining an IR that is comparable to that of other strategies ($IR = 1.37$). It seems that in the present application one needs both model averaging and Student t distributed innovations to deal with the fat tail nature of the underlying asset returns. A useful visual summary is provided in Figure 1 where the empirical VaR exceedances ($\hat{\pi}_i$) from Tables 7 and 8 are plotted against the IR. The results from the individual models are marked by empty circles, best and Bayesian average strategies are marked by empty triangles, and the results from the thick modelling strategies are marked by filled squares. The vertical line represents the tolerance probability, $\alpha = 1\%$. The results from thick modelling strategies stand out quite clearly, as they are clustered together close to the vertical line and display reasonably high IRs.

[Insert Figure 1 around here.]

The above findings are in line with the theoretical results discussed in Section 5.3, where it was shown that the average model will be more fat-tailed than the underlying Gaussian or Student t models with the same average volatility. When the underlying models are already fat tailed model averaging (without any single model dominating) can induce a further degree of fat-tailedness. This is evident in the case of the top 25% AIC thick modelling strategy with Student $t(7)$ innovations.

6.5 Statistical Diagnostic Test Results

The different volatility models and modelling strategies can also be evaluated using purely statistical techniques. A statistical procedure, which is close to ours, focuses on the probability density forecasts of a given portfolio return, $\rho_t = \omega_{t-1}'\mathbf{r}_t$, and considers the probability integral transforms $\hat{v}_{it} = \int_{-\infty}^{\rho_t} \hat{f}(x|\mathcal{F}_{t-1}, M_i)dx$, for $t = \tau+1, ..., \tau+T_1$, where $\hat{f}(x|\mathcal{F}_{t-1}, M_i)$ is the estimated probability density of ρ_t under model M_i and conditional on $\mathcal{F}_{t-1}$. Making use of a well-known result due to Rosenblatt (1952) it is easily seen that the sequence $\{\hat{v}_{it}, t \in \mathcal{T}_1\}$ will be *i.i.d.* uniformly distributed on the interval $[0,1]$ if $\hat{f}(x\,|\mathcal{F}_{t-1}, M_i)$ coincides with the 'true' but unknown conditional predictive density of ρ_t. For further discussions see Diebold, Gunther, and Tay (1998) and Diebold, Hahn, and Tay (1999).

To test the hypothesis that $\hat{v}_{it}$ are random draws from the uniform $[0,1]$ distribution, we consider the standard Kolmogorov-Smirnov test $KS = \max_{1\leq j\leq T_1}\left|\frac{j}{T_1} - \hat{v}_j^*\right|$ as well as the Kuiper test $Ku = \max_{1\leq j\leq T_1}(\frac{j}{T_1} - \hat{v}_j^*) + \max_{1\leq j\leq T_1}(\hat{v}_j^* - \frac{j}{T_1})$, where $\hat{v}_1^* \leq \hat{v}_2^* \leq ... \leq \hat{v}_{T_1}^*$

represent ordered values of $\hat{v}_{i\tau+1}, ..., \hat{v}_{i\tau+T_1}$. The Kuiper test has the added advantage of placing greater emphasis on the tail behavior of the distribution.

[Insert Figure 9 around here.]

Table 9 reports the p-values of these tests for the 53 individual multivariate models using equally weighted portfolios, defined by $\rho_t = (1/18)\sum_{j=1}^{18} r_{jt}$. The KS and Ku tests are rejected for all the individual models when the underlying innovations are assumed to be Gaussian, which is not surprising considering the known fat-tailed nature of the underlying returns. However, the results are mixed when the innovations are assumed to follow a Student t distribution. Although all volatility models continue to be rejected by the Ku test (possibly with the exception of the TDCC), none are rejected by the KS test at the 1% level.

The test results for the average modelling strategies are summarized in Table 10. The test outcomes are very similar to the ones obtained for the individual models in the sense that all average models are rejected by the Ku test, but none are rejected by the KS test at the 1% level of significance when the innovations are Student t distributed.

Overall, the statistical tests support the main conclusions reached using the VaR based diagnostics.

7 Summary and Conclusions

This paper considers the problem of model uncertainty in the context of multivariate volatility models and notes that it is particularly important given the highly restrictive nature of these models that are used in practice. To deal with model uncertainty we advocate the use of model averaging techniques where an 'average' model is constructed by combining the predictive densities of the models under consideration, using a set of weights that reflect the models' relative in-sample performance. We consider 'thick' modelling as well as (approximate) Bayesian modelling frameworks.

Second, the paper proposes a simple decision-based model evaluation technique that compares the volatility models in terms of their Value-at-Risk performance. The proposed test is applicable to individual as well as to average models, and can be used in a variety of contexts. Under mild regularity conditions, the test is shown to have a Binomial distribution when evaluation sample (T_1) is finite and T_0 (the estimation sample) is sufficiently large. The proposed test converges to a standard Normal variate provided $T_1/T_0 + 1/T_1 \rightarrow 0$, a condition also encountered in the forecast evaluation literature that uses the root mean square error as an evaluation criterion, as discussed in West (1996). The proposed VaR test is invariant to the portfolio weights and is shown to be consistent under departures from the null hypothesis. The Binomial version of the VaR test could have important applications in credit risk literature where the evaluation samples are typically short.

In the empirical application we experimented with AIC and SBC weights and found that, due to the large sample sizes available, they led to very similar results with the selected models often totally dominating the rest. The model most often selected by both criteria turned out to be the TDCC model. In out of sample evaluation, only the TDCC model managed to pass the VaR diagnostic tests. Interesting enough, the simplest of all data filters used in this paper, namely the Equal Weighted Moving Average filter also performed well; doing better than other data filters as well as the O-GARCH specifications. In general, the 'thick' modelling approach turned out to be the most reliable within the class of models and model average strategies that we considered. Thick model averaging strategies consistently had low VaR exceedance frequencies (relative to most single models), whilst retaining high information ratios. Overall, the only strategy that was not rejected by our VaR diagnostic tests was the equal-weighted average model based on the top 25 models (ranked by AIC) and assuming Student t innovations with 7 degrees of freedom.

Finally, while model averaging provides a useful alternative to the two-step model selection strategy, it is nevertheless subject to its own form of uncertainty, namely the choice of the space of models to be considered and their respective weights. It is therefore important that applications of model averaging techniques are investigated for their robustness to such choices. In the case of our application it is clearly desirable to consider also other forms of multivariate volatility models, which could be the subject of future research.

Appendix A Derivation of the Optimal Mean-Variance Portfolio Subject to the VaR Constraint

Under the assumption of a Student t distribution with ν_{t-1} degrees of freedom the Lagrangian of the mean-variance problem (3) subject to the value-at-risk constraint (5) is given by

$$\begin{aligned}\mathcal{L}(\omega_{t-1}|\mathcal{M},\mathcal{F}_{t-1}) &= \omega_{t-1}'\mu_{\mathcal{M},t} - \frac{\delta_{t-1}}{2}\omega_{t-1}'\mathbf{\Sigma}_{\mathcal{M},t}\omega_{t-1} \\ &\quad -\psi_{t-1}\left\{\tilde{c}_{\nu_{t-1},\alpha}\sqrt{\omega_{t-1}'\mathbf{\Sigma}_{\mathcal{M},t-1}\omega} - \omega_{t-1}'\mu_{\mathcal{M},t-1} - L_{t-1}\right\}\end{aligned} \tag{A.1}$$

The first-order conditions with respect to ω_{t-1} are

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial\omega_{t-1}} &= \mu_{\mathcal{M},t} - \delta_{t-1}\mathbf{\Sigma}_{\mathcal{M},t}\omega_{t-1} \\ &\quad -\psi_{t-1}\left\{-\tilde{c}_{\nu_{t-1},\alpha}\left(\omega_{t-1}'\mathbf{\Sigma}_{\mathcal{M},t-1}\omega\right)^{-0.5}\mathbf{\Sigma}_{\mathcal{M},t-1}\omega_{t-1} - \mu_{\mathcal{M},t-1}\right\} = 0,\end{aligned} \tag{A.2}$$

and the complementary slackness condition is

$$\psi_{t-1}\left\{\tilde{c}_{\nu_{t-1},\alpha}\sqrt{\omega_{t-1}'\mathbf{\Sigma}_{\mathcal{M},t-1}\omega} - \omega_{t-1}'\mu_{\mathcal{M},t-1} - L_{t-1}\right\} = 0. \tag{A.3}$$

If the VaR constraint does not bind $\psi_{t-1} = 0$ and the optimal solution is given by

$$\omega_{t-1}^* = \frac{1}{\delta_{t-1}}\mathbf{\Sigma}_{\mathcal{M},t-1}^{-1}\mu_{\mathcal{M},t-1}. \tag{A.4}$$

If, on the other hand, the VaR constraint binds $\psi_{t-1} < 0$ and the optimal solution is given by

$$\omega_{t-1}^* = \frac{1}{\delta_{t-1}^*}\mathbf{\Sigma}_{\mathcal{M},t-1}^{-1}\mu_{\mathcal{M},t-1}, \tag{A.5}$$

where $\delta_{t-1}^* \equiv \frac{\delta_{t-1}-\psi_{t-1}\left(\omega_{t-1}'\mathbf{\Sigma}_{\mathcal{M},t-1}\omega\right)^{-0.5}}{1+\psi_{t-1}} > \delta_{t-1}$. From the complementary slackness condition (A.3) we also have that

$$\tilde{c}_{\nu_{t-1},\alpha}\sqrt{\omega_{t-1}'\mathbf{\Sigma}_{\mathcal{M},t-1}\omega} - \omega_{t-1}'\mu_{\mathcal{M},t-1} - L_{t-1} = 0. \tag{A.6}$$

Substituting ω_{t-1}^* from (A.5) into (A.6) we get

$$\delta_{t-1}^* \equiv \frac{s_{\mathcal{M},t}(\tilde{c}_{v_{t-1},\alpha} - s_{\mathcal{M},t})}{L_{t-1}}, \tag{A.7}$$

where $s_{\mathcal{M},t} = \sqrt{\mu_{\mathcal{M},t}'\mathbf{\Sigma}_{\mathcal{M},t}^{-1}\mu_{\mathcal{M},t}}$.

Appendix B Description of Volatility Models

Almost all the multivariate volatility models considered in the literature can be cast in terms of the following decomposition of the conditional volatility matrix, Σ_t, originally due to Bollerslev (1990):

$$\mathbf{\Sigma}_t = \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t, \tag{B.1}$$

where R_t is the one-step-ahead conditional correlation matrix with its (h, j)th element given by $\rho_{hj,t}$, and D_t is a diagonal matrix with $\sqrt{\sigma_{hh,t}}$ on its (h, h) th element. This is a convenient decomposition and allows separate specification of the conditional volatilities and conditional cross-asset returns correlations. The models used in our empirical applications also belong to the class of models spanned by different specifications of $\sqrt{\sigma_{hh,t}}$ and $\rho_{hj,t}$, which are computationally feasible for estimation and forecasting in the case of portfolios with a large number of assets ($N = 15$ in our application). In what follows ε_t denotes the $N \times 1$ vector of residuals from the OLS regressions of returns on a number of predictor variables. In our empirical application $N = 15$ and residuals are computed from first-order autoregressions of the individual return series. For the computation of the CCC, DCC, and ADCC models we have benefitted from Matlab code made available by Kevin Sheppard.

B.1 Equal-Weighted Moving Average (EQMA(n_0))

In the absence of reliable intra-daily observations on returns, a simple estimate of $\mathbf{\Sigma}_t$ can be obtained using the following rolling moment estimates based on the last n_0 observations:

$$\mathbf{\Sigma}_{1t} = \frac{1}{n_0} \sum_{s=1}^{n_0} \mathbf{r}_{t-s} \mathbf{r}'_{t-s}.$$

For $\mathbf{\Sigma}_{1t}$ to be positive definite we must have $n_0 > N$. In the empirical applications we consider four variants of $\mathbf{\Sigma}_{1t}$, using $n_0 = 50, 75, 125$, and 250. Subject to $n_0 > N$, care should be taken so that n_0 is not set too high; otherwise $\mathbf{\Sigma}_{1t}$ could behave like the unconditional variance matrix of the returns.

B.2 One and Two-Parameter Exponential-Weighted Moving Average (EWMA(n_0, λ_0, ν_0))

The one-parameter EWMA (setting $\lambda_0 = \nu_0$) is the popular *Riskmetrics* estimate of $\mathbf{\Sigma}_t$ (see J.P.Morgan (1996)) which is defined by the following recursion

$$\mathbf{\Sigma}_{2t} = \lambda_0 \mathbf{\Sigma}_{2,t-1} + \frac{(1-\lambda_0)}{(1-\lambda_0^{n_0})} \varepsilon_{t-1}\varepsilon'_{t-1} - \frac{(1-\lambda_0)}{(1-\lambda_0^{n_0})} \lambda_0^{n_0-1} \varepsilon_{t-n_0-1}\varepsilon'_{t-n_0-1}, \tag{B.2}$$

for a constant parameter $0 < \lambda_0 < 1$, and a window of size n_0. Typically, the initialization of the recursion in (B.2) is based on estimates of the unconditional variances using a pre-sample of data. For the (i, j)th entry of $\mathbf{\Sigma}_{2t}$ we have

$$\sigma_{2,ijt} = \frac{(1-\lambda_0)}{(1-\lambda_0^{n_0})} \sum_{s=1}^{n_0} \lambda_0^{s-1} \varepsilon_{i,t-s}\varepsilon_{j,t-s}.$$

The Riskmetrics model is characterized by the fact that n_0 and λ_0 are fixed *a priori.* Moreover, it has been recently pointed out that it is not possible to formally estimate the model statistically, due to its asymptotic degenerateness (see Zaffaroni (2007)). The value of $\lambda_0 = 0.94$ is suggested in J.P.Morgan (1996). In our analysis we shall consider the values $\lambda_0 = 0.94, 0.95$, and 0.96, and set $n_0 = 250$. We only consider one value for the window size since there is an obvious trade-off between λ_0 and n_0, with a small λ_0 yielding similar results to a small n_0. Note that for $\mathbf{\Sigma}_{2t}$ to be non singular requires $n_0 \geq N$. Nevertheless, the model does admit a well-defined forecasting function and indeed $\mathbf{\Sigma}_{2,t+1}$ represents the one-step ahead forecast of the conditional variance for period $t+1$, based on the information available up to time t.

Practitioners and academics have often pointed out that the effects of shocks on conditional variances and conditional correlations could decay at different rates, with correlations typically responding at a slower pace than volatilities (see De Santis and Gerard (1997)). This suggests using two different parameter values for the decay coefficients, one for volatilities and the other for correlations (see De Santis, Litterman, Vesval, and Winkelmann (2003, p.14)). This yields the two-parameter Exponential-Weighted Moving Average (EWMA (n_0, λ_0, ν_0)). Therefore, the diagonal elements of (B.2) define conditional variances $\sigma_{3,hht}$, $h = 1, ..., N$ the square-roots of which form the diagonal matrix $\mathbf{D}_{3t}$. The covariances are based on the same recursion as (B.2) but using a smoothing parameter ν_0, generally different from λ_0 ($\nu_0 \leq \lambda_0$) yielding

$$\sigma_{3,hjt} = \frac{(1-\nu_0)}{(1-\nu_0^{n_0})} \sum_{s=1}^{n_0} \nu_0^{s-1} \varepsilon_{h,t-s} \varepsilon_{j,t-s}, \text{ for } h \neq j.$$

We assume that the same window size, n_0, applies to variance and covariance recursions. The ratio

$$\sigma_{3,hjt} / \sqrt{\sigma_{3hh,t}\, \sigma_{3jj,t}} \tag{B.3}$$

represents the (h, j)th entry of the matrix $\mathbf{R}_{3t}$. $\mathbf{\Sigma}_{3t}$ is obtained by combining terms according to (B.1). The parameters ν_0 and λ_0 are not estimated but calibrated *a priori*, as for the one-parameter EWMA model.

B.3 Mixed Moving Average (MMA(n_0, ν_0))

This is a generalization of the equal-weighted MA model discussed above. Under this specification, the conditional variances are computed as in the equal-weighted MA model, the square root of which yields the diagonal matrix $\mathbf{D}_{4t}$. Then we estimate the conditional covariances using a Riskmetrics type filter: $\sigma_{4,hjt} = \frac{(1-\nu_0)}{(1-\nu_0^{n_0})} \sum_{s=1}^{n_0} \nu_0^{s-1} \varepsilon_{h,t-s} \varepsilon_{j,t-s}$, which after normalization according to (B.3) yields $\mathbf{R}_{4t}$. Re-combining the results according to (B.1) we then obtain $\mathbf{\Sigma}_{4t}$.

B.4 Generalized Exponential-Weighted Moving Average (EWMA(n_0, p, q, ν_0))

This is a generalization of the two-parameter EWMA. In the first stage N different univariate GARCH(p, q) volatility models are estimated for each r_{ht} by PMLE. The conditional covariances are then obtained using the Riskmetrics filter (B.2), with the parameters n_0 and ν_0 fixed *a priori*. The results are then normalized using (B.3), with the resultant variances and correlations re-combined according to (B.1), thus yielding $\mathbf{\Sigma}_{5t}$. The estimated number of parameters of this model is $k_5 = N(1 + p + q)$, which will be used in the computation of AIC and SBC.

B.5 Constant Conditional Correlation (CCC(p, q))

Bollerslev (1990) introduced a multivariate GARCH model with the simplifying assumption that the one-step ahead conditional correlations are constant. Under this model, (B.1) takes the form $\mathbf{\Sigma}_{6t} = \mathbf{D}_{6t}\mathbf{R}_6\mathbf{D}_{6t}$, where $\mathbf{D}_{6t}$ is a diagonal matrix containing the square-root of the $\sigma_{6,hht}$, each of which follow the GARCH(p, q)

model of Bollerslev (1986)

$$\sigma_{6,hht} = c_{0h} + \sum_{k=1}^{q} \alpha_{0hk}\varepsilon_{h,t-k}^2 + \sum_{j=1}^{p} \beta_{0hj}\sigma_{6,hht-j},$$

for constant positive parameters c_{0h}, $\alpha_{0h1}, ..., \alpha_{0hq}, \beta_{0h1}, ..., \beta_{0hp}$. Positivity of these parameters is sufficient but not necessary to ensure $\sigma_{6,hht} > 0$ *a.s.* (see Nelson and Cao (1992)). The positive definite matrix $\mathbf{R}_6$, made by $N(N-1)/2$ constant parameters, contains the (constant) conditional correlations of the $\varepsilon_{ht}, h = 1, 2, ..., N$.

Bollerslev (1990) proposed to estimate the model by the PMLE and noting that (9) simplifies due to the constant correlation assumption. The estimated number of parameters of this model is given by $k_6 = N(p + q + 1) + N(N-1)/2$.

B.6 Orthogonal GARCH (O-GARCH(p, q))

This model is proposed by Alexander (2001) and uses a static principle component decomposition of standardized residuals defined by

$$\tilde{\varepsilon}_{it} = \frac{\varepsilon_{it} - \bar{\varepsilon}_{iT}}{s_{iT}},\ t = 1, 2, ..., T,$$

where $\bar{\varepsilon}_{iT}$ and s_{iT} are the sample mean and standard deviations of the returns. Denote the sample covariance matrix of the standardized returns by

$$\tilde{\mathbf{S}}_T = \frac{\sum_{t=1}^{T} \tilde{\varepsilon}_t \tilde{\varepsilon}_t'}{T},\ \tilde{\varepsilon}_t = (\tilde{\varepsilon}_{1t}, , ..., \tilde{\varepsilon}_{Nt})'.$$

Then

$$\tilde{\mathbf{S}}_T \mathbf{W}_T = \mathbf{W}_T \mathbf{\Lambda}_T, \tag{B.4}$$

where and $\mathbf{W}_T$ and $\mathbf{\Lambda}_T$ are the corresponding $N \times N$ matrices of eigenvectors and eigenvalues, respectively. Then setting (see Alexander (2001))

$$\mathbf{\Sigma}_{7t}(u) = \mathbf{V}\,\mathbf{W}(u)\mathbf{\Gamma}_t(u)\mathbf{W}(u)'\mathbf{V},$$

where $\mathbf{W}(u) = (\mathbf{w}_1, ..., \mathbf{w}_u)$ denotes the $N \times u$ matrix of eigenvectors corresponding to the first largest u eigenvalues, $\mathbf{V}$ is a diagonal matrix with the sample standard deviation of r_{ht} on the (h, h)th entry and $\mathbf{\Gamma}_t(u)$ is a $u \times u$ diagonal matrix whose (j, j)th entry, γ_{jt}, $j = 1, ..., u$, is assumed to satisfy the following univariate GARCH(p, q) specification

$$\gamma_{jt} = c_{0j} + \alpha_{0j1}s_{jt-1}^2 + ... + \alpha_{0jp}s_{jt-p}^2 + \beta_{0j1}\,\gamma_{jt-1} + ... + \beta_{0jq}\,\gamma_{jt-q},\ \ j = 1, ..., u,$$

where $\mathbf{s}_j = (\varepsilon_1, ..., \varepsilon_T)'\,\mathbf{w}_j$, $j = 1, ..., N$. Note that this method makes use of the fact that the factors are unconditionally orthogonal, but there is no guarantee that they will also be conditionally orthogonal. Also to ensure that $\mathbf{\Sigma}_{7t}(u)$ is non-singular we must have $u = N$, which is the value considered here, yielding $\mathbf{\Sigma}_{7t} = \mathbf{\Sigma}_{7t}(N)$. Hence for the O-GARCH$(p, q)$ specification we have $k_7 = N(p + q + 1)$.

B.7 Dynamic Conditional Correlation (DCC$(p, q, 1, 1)$)

Engle (2002) relaxed the assumption of constant conditional correlation of the CCC model of Bollerslev (1990). The conditional variances of individual returns are estimated as univariate GARCH(p, q) specifications, and the diagonal matrix, $\mathbf{D}_{8t}$, is formed with their square roots. Unlike the CCC, the conditional correlations are now allowed to be time-varying and are obtained as follows. Starting with the standardized residuals, $\tilde{\varepsilon}_{8t} = (\mathbf{D}_{8t})^{-1}\,\varepsilon_t$, the DCC model assumes that the (h, j)th entry of the conditional covariance matrix of $\tilde{\varepsilon}_{8t}$, namely $\mathbf{R}_{8t}$, is given by $q_{hjt}/\sqrt{q_{hht}\, q_{jjt}}$, where q_{hjt} is the (h,j)th element of matrix $\mathbf{Q}_t$ defined by

$$\mathbf{Q}_t = \overline{\mathbf{Q}}\,(1-\gamma_{01}-\delta_{01}) + \gamma_{01}\tilde{\varepsilon}_{9,t-1}\tilde{\varepsilon}'_{9,t-1} + \delta_{01}\mathbf{Q}_{t-1}.$$

for a fixed positive definite matrix $\overline{\mathbf{Q}}$, and positive parameters satisfying $\gamma_{01} + \delta_{01} < 1$. Finally, $\mathbf{\Sigma}_{8t}$ is obtained re-combining $\mathbf{D}_{8t}$ and $\mathbf{R}_{8t}$ based on (B.1). The estimation of the parameters of the DCC model is carried out using a two-stage Gaussian PMLE procedure. The log-likelihood function is first optimized with respect to the parameters driving the individual conditional variances. Conditional on these parameter estimates, in the second step the log-likelihood function is maximized with respect to the parameters driving conditional correlations. See Engle (2002, Section 4) for details. For this model we have $k_8 = N(p+q+1) + N(N+1)/2 + 2$.

B.8 Asymmetric Dynamic Conditional Correlation (ADCC$(p,q,1,1)$)

Cappiello, Engle, and Sheppard (2006) generalized the DCC allowing for the possibility of asymmetric effects on conditional variances and correlations. The conditional variances of the individual returns are specified using the specification advanced by Glosten, Jagannathan, and Runkle (1993) given by:

$$\sigma_{9,hht} = c_{0h} + \sum_{k=1}^{q}\alpha_{0hk}\varepsilon^2_{h,t-k} + \sum_{k=1}^{q}\vartheta_{0hk}I(\varepsilon_{h,t-k}<0)\varepsilon^2_{h,t-k} + \sum_{j=1}^{p}\beta_{0hj}\sigma_{9,hh,t-j},$$

where $I(\mathcal{A})$ denotes the indicator function which takes the value of unity if $\mathcal{A} > 0$, and zero otherwise. Let $\tilde{\varepsilon}_{9t} = (\mathbf{D}_{9t})^{-1}\varepsilon_t$, where $\mathbf{D}_{9t}$ is the diagonal matrix formed with the square roots of $\sigma_{9,hht}$. The ADCC model assumes that the (h,j)th entry of the conditional covariance matrix of $\tilde{\mathbf{r}}_{9t}$, namely $\mathbf{R}_{9t}$, is given by $q_{hjt}/\sqrt{q_{hht}\, q_{jjt}}$ where q_{hjt} is the (h,j)th element of matrix $\mathbf{Q}_t$ defined by

$$\mathbf{Q}_t = \overline{\mathbf{Q}}\,(1-\gamma_{01}-\delta_{01}-\vartheta_{01}) + \gamma_{01}\tilde{\varepsilon}_{9,t-1}\tilde{\varepsilon}'_{9,t-1} + \delta_{01}\mathbf{Q}_{t-1} + \vartheta_{01}\underline{\varepsilon}_{9,t-1}\underline{\varepsilon}'_{9,t-1}$$

where $\underline{\varepsilon}_{9t} = \tilde{\varepsilon}_{9t} \odot I(\varepsilon_{9,t-1} < 0)$ (here $\odot$ denotes the Hadamard product), $\overline{\mathbf{Q}}$ is a fixed positive definite matrix, and γ_{01}, δ_{01}, and ϑ_{01} are positive parameters satisfying $\gamma_{01} + \delta_{01} + \vartheta_{01} < 1$. Finally, $\mathbf{\Sigma}_{9t}$ is constructed using $\mathbf{D}_{9t}$ and $\mathbf{R}_{9t}$ as in (B.1). The estimation of the parameters of the ADCC model is carried out as for the DCC, where now we have $k_9 = N(p+2q+1) + N(N+1)/2 + 3$.

B.9 t-Dynamic Conditional Correlation (TDCC(p,q))

Pesaran and Pesaran (2007) modify the DCC model of Engle (2002) by basing the stochastic process of the conditional correlation matrix on *devolatized* residuals $\check{\varepsilon}_{10t}$ rather than *standardized* residuals $\tilde{\varepsilon}_{8t}$. Whereas the standardized residuals are obtained by dividing $\tilde{\varepsilon}_t$ by the conditional standard deviations from a first-stage GARCH(p,q) model, devolatized residuals are computed by dividing $\tilde{\varepsilon}_t$ by the square root of a k-day moving average of squared residuals, including the contemporaneous observation,

$$\check{\varepsilon}_{10jt} = \frac{\varepsilon_{jt}}{\sigma^{realized}_{jt}}, \tag{B.5}$$

$$\sigma^{realized}_{jt} = \sqrt{\frac{\sum_{j=0}^{k-1}\varepsilon^2_{j,t-k}}{k}}, \tag{B.6}$$

which renders them approximately Gaussian. The conditional correlation matrix of $\check{\varepsilon}_t$, namely $\mathbf{R}_{10t}$, is given by $q_{hjt}/\sqrt{q_{hht}\, q_{jjt}}$, where q_{hjt} is the (h,j)th element of matrix $\mathbf{Q}_t$ defined by

$$\mathbf{Q}_t = \overline{\mathbf{Q}}\,(1-\gamma_{01}-\delta_{01}) + \gamma_{01}\check{\varepsilon}_{10,t-1}\check{\varepsilon}'_{10,t-1} + \delta_{01}\mathbf{Q}_{t-1}.$$

for a fixed positive definite matrix $\overline{\mathbf{Q}}$, and positive parameters satisfying $\gamma_{01} + \delta_{01} < 1$. Finally, $\mathbf{\Sigma}_{10t}$ is obtained re-combining $\mathbf{D}_{8t}$ and $\mathbf{R}_{10t}$ based on (B.1). As in the DCC model the conditional variances of individual returns are estimated as univariate GARCH (p, q) specifications, and the diagonal matrix, $\mathbf{D}_{10t}$, is formed with their square roots. The parameters of the TDCC model are estimated using maximum likelihood based on a Student t-distribution with ν_{t-1} degrees of freedom. The number of parameters to be estimated is $k_{10} = N(p + q + 1) + N(N + 1)/2 + 3$.

Appendix C Proofs of the Theorems

Proof of Theorem 1. As $T_0 \to \infty$, $\hat{\pi}_i \to_p \pi_i = \frac{1}{T_1} \sum_{t \in \mathcal{T}_1} d_{it}$, $d_{it} = I\left(-\rho_t - c_{it}(\alpha)\sigma_{\rho t}(M_i)\right)$. Consider now the moments of $T_1\pi_i$ and note that for any integer $n \geq 1$,

$$E\left(T_1\pi_i\right)^n = \sum_{t_1, t_2, \ldots, t_n \in \mathcal{T}_1} \left\{E\left(d_{it_1} d_{it_2} \ldots d_{it_n}\right)\right\}. \tag{C.1}$$

However, for any $\delta > 0$ we have $E(d_{it}^{\delta} \mid \mathcal{F}_{t-1}, M_i) = \alpha$, or unconditionally $E(d_{it}^{\delta} \mid M_i) = \alpha$. Hence, all the terms $E\left(d_{it_1} d_{it_2} \ldots d_{it_n}\right)$ in (C.1) coincide with the case when the d_{it_j}, $j = 1, .., n$, are *i.i.d* Bernoulli distributed random variables with parameter α, for any choice of $t_1, \ldots, t_n$. Also, since $T_1 < \infty$, the support of the distribution of $T_1\pi_i$ is bounded and as a consequence its moment generating function exists and is the same as that of a Binomial distribution with parameters T_1 and α. Therefore, by the method of moments (see Billingsley (1986, Theorem 30.1)), $T_1\pi_i$ will also have a Binomial distribution. ■

Proof of Theorem 2. Assume H_{i0} defined by (25) holds. Set $q_{it} = q_{it}(\hat{\theta}_{iT_0}, \theta_{i0}) = (\hat{\sigma}_{\rho\, t}(M_i)/\sigma_{\rho t}(M_i)) = (\omega_{t-1}' \hat{\boldsymbol{\Sigma}}_{it}\, \omega_{t-1}/\omega_{t-1}' \boldsymbol{\Sigma}_{it}\, \omega_{t-1})^{1/2}$. Then

$$E[d_{it}(\hat{\theta}_{iT_0})|\mathcal{F}_{t-1}, M_i] = F_{it}(-c_{it}(\alpha)q_{it})$$

and

$$E[\hat{\pi}_i|M_i] = \frac{1}{T_1} \sum_{t \in \mathcal{T}_1} E\{F_{it}(-c_{it}(\alpha)q_{it})\}.$$

As $T_0 \to \infty$, $\hat{\theta}_{iT_0} \xrightarrow{p} \theta_{i0}$ and since $\boldsymbol{\Sigma}_{it}(\theta_i)$ is a continuous function of θ_i it also follows that $q_{it}(\hat{\theta}_{iT_0}, \theta_{i0}) \xrightarrow{p} 1$, for all values of $t \in \mathcal{T}_1$. Hence, for any given *finite* evaluation sample size, T_1, and as $T_0 \to \infty$, $E\left(\hat{\pi}_i|M_i\right) = \frac{1}{T_1} \sum_{t \in \mathcal{T}_1} E\{F_{it}(-c_{it}(\alpha)q_{it})\} \xrightarrow{p} F_{it}(-c_{it}(\alpha)) = \alpha$. Consider now the statistic $\sqrt{T_1}(\hat{\pi}_i - \alpha)$ and write it as

$$\sqrt{T_1}(\hat{\pi}_i - \alpha) = \sqrt{T_1}(\pi_i - \alpha) + \sqrt{T_1}(\hat{\pi}_i - \pi_i), \tag{C.2}$$

where $\pi_i = \frac{1}{T_1} \sum_{t \in \mathcal{T}_1} d_{it}(\theta_{i0})$. Also note that $\sqrt{T_1}(\hat{\pi}_i - \pi_i) = \sqrt{T_1/T_0}(\sum_{t \in \mathcal{T}_1} X_{it,T_0}/T_1)$, where $X_{it,T_0} = \sqrt{T_0}[d_{it}(\hat{\theta}_{iT_0}) - d_{it}(\theta_{i0})]$. But it is easily seen that,

$$|X_{it,T_0}| = \begin{cases} \sqrt{T_0} & \text{if } (\rho_t + c_{it}(\alpha)\hat{\sigma}_{\rho t}(M_i))(\rho_t + c_{it}(\alpha)\sigma_{\rho\, t}(M_i)) < 0, \\ 0 & \text{otherwise.} \end{cases}$$

Hence, for all $t \in \mathcal{T}_1$, $\Pr\left(|X_{it,T_0}| = \sqrt{T_0}\, |\mathcal{F}_{t-1}, M_i\right) \leq |F_{it}(-c_{it}(\alpha)q_{it}(\hat{\theta}_{iT_0}, \theta_{i0})) - F_{it}(-c_{it}(\alpha))|$, and consequently $E(|X_{it,T_0}| \ \ |\mathcal{F}_{t-1}, M_i) \leq \sqrt{T_0}|F_{it}(-c_{it}(\alpha)q_{it}(\hat{\theta}_{iT_0}, \theta_{i0})) - F_{it}(-c_{it}(\alpha))|$. Using the mean-value expansion of $F_{it}(-c_{it}(\alpha)q_{it}(\hat{\theta}_{iT_0}, \theta_{i0}))$ around $\hat{\theta}_{iT_0}$ one gets

$$F_{it}(-c_{it}(\alpha)q_{it}(\hat{\theta}_{iT_0}, \theta_{i0})) = F_{it}(-c_{it}(\alpha)) - c_{it}(\alpha)f_{it}(-c_{it}(\alpha)q_{it}(\bar{\theta}_i, \theta_{i0}))\partial q_{it}(\bar{\theta}_i, \theta_{i0})/\partial\hat{\theta}_{iT_0}'(\hat{\theta}_{iT_0} - \theta_{i0}),$$

where the elements of $\bar{\theta}_i$ are convex combinations of the corresponding elements of $\hat{\theta}_{iT_0}$ and θ_{i0}. By Hölder's inequality for the norm of matrices, since $\| \omega_t \| > 0$, we have $E(|X_{it,T_0}| \ \ |\mathcal{F}_{t-1}, M_i)$
$\leq c_{it}(\alpha)f_{it}(-c_{it}(\alpha)q_{it}(\bar{\theta}_i, \theta_{i0})) \, \| \partial q_{it}(\bar{\theta}_i, \ \theta_{i0})/\partial\hat{\theta}_{iT_0} \| \ \sqrt{T_0} \, \| \hat{\theta}_{iT_0} - \theta_{i0} \| \leq c_{it}(\alpha)f_{it}(-c_{it}(\alpha)q_{it}(\bar{\theta}_i, \theta_{i0}))$
$\{\sup_{\theta \in \Theta_i} \| \partial\bar{\lambda}_{it}(\theta)/\partial\theta \| /\underline{\lambda}_{it}^{\frac{1}{2}}(\theta)\underline{\lambda}_{it}^{\frac{1}{2}}(\theta_0)\}\sqrt{T_0} \ \| \hat{\theta}_{iT_0} - \theta_{i0} \|$. Taking the unconditional mean and using the Hölder inequality again yields $E(|X_{it,T_0}||M_i)$
$\leq c_{it}(\alpha) \sup_x f_{it}(x)(E| \sup_{\theta \in \Theta_i} \| \partial\bar{\lambda}_{it}(\theta)/\partial\theta \| /\underline{\lambda}_{it}^{\frac{1}{2}}(\theta)\underline{\lambda}_{it}^{\frac{1}{2}}(\theta_0)|^{\delta})^{\frac{1}{\delta}} \sqrt{T_0}(E\| \ \hat{\theta}_{iT_0} - \theta_{i0} \ \|^{\frac{\delta}{\delta-1}})^{1-1/\delta}$. Therefore, $T_1^{-1} \sum_{t \in \mathcal{T}_1} X_{it,T_0} = O_p(1)$ and the second term in (C.2) vanishes as $T_1/T_0 + 1/T_1 \to 0$. Hence $\sqrt{T_1}(\hat{\pi}_i - \alpha) - \sqrt{T_1}(\pi_i - \alpha) = o_p(1)$, where $\sqrt{T_1}(\pi_i - \alpha) = \frac{1}{\sqrt{T_1}} \sum_{t \in \mathcal{T}_1} g_{it}$, $g_{it} = I\left(-\rho_t - c_{it}(\alpha)\sigma_{\rho t}(M_i)\right) - \alpha$. Therefore, it remains to establish the asymptotic distribution of $\sqrt{T_1}(\pi_i - \alpha)$. This easily follows by the martingale central

limit theorem of Brown (1971, Theorem 2) since the g_{it} are a bounded, martingale difference sequence with the constant variance $\alpha(1-\alpha)$.■

Proof of Theorem 3: Inequality (33) can be expressed as $\sum_{i=1}^{N} \lambda_i g(b_i) > g(\sum_{i=1}^{N} \lambda_i b_i)$, for the function $g(x) \equiv F(a/\sqrt{x})$. Jensen's inequality ensures that the latter inequality is satisfied whenever $g(\cdot)$ is strictly convex. Since $g(\cdot)$ is twice differentiable by construction, we just need to check the conditions such that the second derivative of $g(x)$ satisfies $g''(x) > 0$. Straightforward calculations yield the required condition (32).■

Table 1: Summary Statistics and Univariate GARCH Models

(in per cent)

Asset	Mean ($\times$ 100)	St.Dev.	Skewness	Kurtosis	t-GARCH(1,1) $\hat{\alpha}_j$	$\hat{\beta}_j$	$\hat{\nu}_j$
Currencies							
GBP	0.938	0.577	-0.121	6.542	0.031	0.963	5.349
EUR	0.337	0.647	0.052	5.254	0.026	0.971	6.130
JPY	-0.715	0.697	0.860	12.504	0.039	0.945	4.535
CAD	0.425	0.376	0.048	4.882	0.038	0.960	7.600
AUD	1.040	0.621	-0.073	6.101	0.026	0.971	6.142
CHF	-0.212	0.706	0.151	4.967	0.022	0.974	5.958
mean	*0.302*	*0.604*	*0.153*	*6.708*	*0.030*	*0.964*	*5.952*
Equities							
SP	3.114	1.005	-0.072	7.843	0.051	0.944	5.846
FTSE	2.137	1.057	-0.023	5.902	0.069	0.923	11.491
DAX	3.526	1.398	-0.285	9.407	0.074	0.919	7.360
NIKKEI	-0.047	1.392	0.072	5.051	0.061	0.930	7.214
mean	*2.182*	*1.213*	*-0.077*	*7.051*	*0.064*	*0.929*	*7.978*
Bonds							
10Y US	1.373	0.364	-0.250	4.824	0.031	0.962	6.624
10Y EUR	1.309	0.313	-0.278	5.099	0.041	0.950	8.193
10Y Gilt	1.009	0.403	0.035	7.489	0.035	0.960	6.592
10Y JGB	1.665	0.279	-0.460	7.295	0.062	0.930	5.163
mean	*1.339*	*0.340*	*-0.238*	*6.177*	*0.042*	*0.951*	*6.643*
Commodities							
Gold	0.084	0.872	0.108	13.646	0.053	0.944	4.115
Silver	2.099	1.553	-0.523	9.834	0.034	0.961	4.033
Wheat	2.088	1.241	0.267	5.036	0.058	0.919	7.636
Crude	5.067	2.041	-0.870	19.479	0.049	0.942	6.396
mean	*2.334*	*1.427*	*-0.254*	*11.999*	*0.049*	*0.941*	*5.545*
mean(all assets)	*1.402*	*0.863*	*-0.076*	*7.842*	*0.044*	*0.948*	*6.465*

Notes: Columns 2 to 5 report the sample mean, standard deviation, skewness and kurtosis of returns computed as $r_{jt} = 100 \times (P_{jt} - P_{j,t-1})/P_{j,t-1}$, where P_{jt} is the close price of the j^{th} asset. Columns 7 to 9 report the estimated parameters of GARCH(1,1) models with Student t-innovations:

$$\mathrm{V}(r_{jt}|\Omega_{t-1}) = \sigma^2_{jt} = \bar{\sigma}^2_j(1-\alpha_j-\beta_j) + \alpha_j r^2_{j,t-1} + \beta_j \sigma^2_{j,t-1}, \text{ where } \frac{r_{jt}}{\sigma_{jt}}\sqrt{\frac{\nu_{j,t-1}-2}{\nu_{j,t-1}}} \sim t_{\nu_{j,t-1}},$$

and $\bar{\sigma}^2_j$ is the unconditional variance of r_{jt}. All estimated coefficients are significant at the 1% level.

The full sample period is January 2, 1991 to July 11, 2007 (4311 observations). The currencies in order are British pound, Euro, Japanese Yen, Canadian dollar, Australian dollar, and Swiss Franc.

Source: *Datastream*

Table 2: Average Pairwise Correlations of Returns Within and Across Asset Classes

Asset	Foreign Exchange	Equity Indices	10Y Bonds	commodities
GBP	0.427	-0.081	0.060	0.106
EUR	0.493	-0.120	0.078	0.119
JPY	0.310	-0.038	0.001	0.094
CAD	0.175	0.091	0.049	0.138
AUD	0.261	0.059	0.016	0.196
CHF	0.486	-0.158	0.077	0.125
SP	-0.038	0.322	0.007	-0.038
FTSE	-0.076	0.440	0.017	-0.012
DAX	-0.083	0.454	-0.026	-0.038
NIKKEI	0.033	0.180	-0.058	0.044
10Y US	0.097	-0.047	0.331	-0.023
10Y EUR	0.059	0.005	0.438	-0.003
10Y Gilt	0.021	0.035	0.388	-0.005
10Y JGB	0.009	-0.053	0.078	-0.019
Gold	0.242	-0.043	0.014	0.306
Silver	0.190	0.012	-0.014	0.299
Wheat	0.053	0.006	-0.014	0.075
Crude	0.034	-0.018	-0.036	0.125

Notes: This table reports the average pairwise correlation of returns of futures contract i with the returns of all contracts ($j \neq i$) in each asset class. The sample period is January 2, 1991 to July 11, 2007 (4311 observations).

Table 3: AIC Values for Multivariate Volatility Models under Standard Normal Innovations

Model type	Sample periods							Sample periods					
	25-Jan-94		11-Jul-07		Average			25-Jan-94		11-Jul-07		Average	
	(1)		(2)		(3)			(4)		(5)		(6)	
EQMA							(1,1,0.97)	-13653	(22)	-10007	(16)	-13532	(23)
(n_0)							(1,2,0.97)	-13585	(20)	-10117	(19)	-13510	(20)
(250)	-13283	(19)	-9533	(14)	-13169	(18)	(2,1,0.97)	-13663	(23)	-10025	(17)	-13523	(22)
(125)	-13131	(18)	-9623	(15)	-13449	(19)	(2,2,0.97)	-13614	(21)	-10138	(20)	-13516	(21)
(75)	-13875	(26)	-10373	(28)	-14243	(28)							
(50)	-15617	(39)	-11966	(42)	-15872	(42)	OGARCH						
							(p,q)						
EWMA							(1,1)	-12397	(14)	-10148	(21)	-13063	(16)
(n_0,λ_0,ν_0)							(1,2)	-12402	(15)	-10157	(22)	-13067	(17)
(250,0.95,0.95)	-15825	(41)	-11762	(39)	-15532	(38)	(2,1)	-12415	(16)	-10292	(26)	-13020	(15)
(250,0.97,0.95)	-15907	(42)	-11713	(38)	-15537	(39)	(2,2)	-12419	(17)	-10271	(25)	-12982	(14)
(250,0.95,0.97)	-13877	(27)	-10179	(23)	-13775	(25)							
(250,0.97,0.97)	-13908	(28)	-10100	(18)	-13747	(24)	CCC						
(125,0.95,0.95)	-15626	(40)	-11835	(41)	-15634	(40)	(p,q)						
(125,0.97,0.95)	-15592	(38)	-11782	(40)	-15664	(41)	(1,1)	-12235	(13)	-9150	(10)	-12456	(11)
(125,0.95,0.97)	-13837	(25)	-10338	(27)	-13992	(27)	(1,2)	-12207	(11)	-9250	(12)	-12453	(10)
(125,0.97,0.97)	-13766	(24)	-10253	(24)	-13977	(26)	(2,1)	-12226	(12)	-9167	(11)	-12460	(13)
(75,0.95,0.95)	-16011	(43)	-12306	(45)	-16147	(43)	(2,2)	-12198	(10)	-9273	(13)	-12457	(12)
(75,0.97,0.95)	-16052	(44)	-12287	(44)	-16212	(44)							
(75,0.95,0.97)	-14548	(31)	-11038	(32)	-14760	(31)	DCC						
(75,0.97,0.97)	-14521	(30)	-10976	(31)	-14765	(32)	(p,q)						
(50,0.95,0.95)	-17995	(51)	-13846	(51)	-17905	(51)	(1,1)	-12034	(9)	-8956	(6)	-12266	(7)
(50,0.97,0.95)	-18072	(52)	-13872	(52)	-18000	(52)	(1,2)	-12002	(7)	-9043	(8)	-12261	(6)
(50,0.95,0.97)	-16734	(48)	-12776	(49)	-16688	(47)	(2,1)	-12030	(8)	-8972	(7)	-12270	(9)
(50,0.97,0.97)	-16733	(47)	-12748	(48)	-16714	(48)	(2,2)	-11999	(6)	-9065	(9)	-12267	(8)
MMA							ADCC						
(n_0,ν_0)							(p,q)						
(250,0.95)	-16995	(50)	-12418	(46)	-16566	(46)	(1,1)	-11962	(2)	-8882	(2)	-12197	(2)
(250,0.97)	-14773	(32)	-10643	(30)	-14570	(30)	(1,2)	-11964	(3)	-8891	(3)	-12224	(3)
(125,0.95)	-16336	(45)	-12163	(43)	-16455	(45)	(2,1)	-11981	(4)	-8908	(4)	-12238	(5)
(125,0.97)	-14293	(29)	-10501	(29)	-14553	(29)	(2,2)	-11982	(5)	-8919	(5)	-12237	(4)
(75,0.95)	-16588	(46)	-12539	(47)	-16722	(49)							
(75,0.97)	-14875	(33)	-11122	(33)	-15120	(37)	TDCC						
(50,0.95)	-18331	(53)	-14027	(53)	-18310	(53)	(p,q)						
(50,0.97)	-16869	(49)	-12818	(50)	-16907	(50)	(1,1)	-11451	(1)	-8676	(1)	-11860	(1)
GEWMA													
(p,q,ν_0)													
(1,1,0.95)	-15365	(36)	-11474	(34)	-15104	(36)							
(1,2,0.95)	-15276	(34)	-11616	(36)	-15074	(33)							
(2,1,0.95)	-15385	(37)	-11495	(35)	-15088	(35)							
(2,2,0.95)	-15312	(35)	-11641	(37)	-15078	(34)							

Notes: This table reports the Akaike Information Criteria (AIC) of the volatility models described in Appendix B under the assumption of standard normal innovations (with the exception of the TDCC model). The AIC is computed as $AIC_{it} = LL_{it} - k_i$, where LL_{it} is the maximized log-likelihood value at time t for model i, and k_i is the number of parameters estimated under model i. All multivariate volatility models (when applicable) were estimated based on one-day ahead forecast errors, $\hat{\varepsilon}_{it} = r_{it} - \hat{\mu}_{it}$, using rolling windows of size 800 days every 25 days. The one-step ahead forecasts, $\hat{\mu}_{it}$, were obtained by estimating an AR(1) model also on a window of size 800 recursively and updated daily. For the TDCC model the distribution is Student t with ν_{t-1} degrees of freedom, where ν_{t-1} is re-estimated every 25 days. Columns 1 and 4 report the AIC values of the first sample (2-Jan-91 to 25-Jan-94, 800 observations). Columns 2 and 5 report the AIC values of the last sample (17-Jun-04 to 11-Jul-07, 800 observations). Columns 3 and 6 report the average AIC values over all 3512 rolling samples, with 800 observations each, extracted from the full sample of data (2-Jan-91 to 11-Jul-07). The models' rank is given in parentheses.

Table 4: SBC Values for Multivariate Volatility Models under Standard Normal Innovations

Model type	Sample periods							Sample periods					
	25-Jan-94		11-Jul-07		Average			25-Jan-94		11-Jul-07		Average	
	(1)		(2)		(3)			(4)		(5)		(6)	
EQMA							(1,1,0.97)	-13779	(22)	-10133	(17)	-13659	(20)
(n_0)							(1,2,0.97)	-13754	(20)	-10285	(22)	-13679	(21)
(250)	-13283	(19)	-9533	(10)	-13169	(14)	(2,1,0.97)	-13832	(24)	-10194	(19)	-13691	(22)
(125)	-13131	(18)	-9623	(11)	-13449	(19)	(2,2,0.97)	-13825	(23)	-10349	(25)	-13727	(23)
(75)	-13875	(26)	-10373	(26)	-14243	(28)							
(50)	-15617	(39)	-11966	(42)	-15872	(42)	OGARCH						
							(p,q)						
EWMA							(1,1)	-12523	(10)	-10275	(21)	-13189	(16)
(n_0, λ_0, ν_0)							(1,2)	-12571	(11)	-10325	(23)	-13235	(18)
(250,0.95,0.95)	-15825	(41)	-11762	(37)	-15532	(38)	(2,1)	-12583	(12)	-10461	(27)	-13189	(15)
(250,0.97,0.95)	-15907	(42)	-11713	(36)	-15537	(39)	(2,2)	-12629	(13)	-10482	(28)	-13193	(17)
(250,0.95,0.97)	-13877	(27)	-10179	(18)	-13775	(25)							
(250,0.97,0.97)	-13908	(28)	-10100	(16)	-13747	(24)	CCC						
(125,0.95,0.95)	-15626	(40)	-11835	(40)	-15634	(40)	(p,q)						
(125,0.97,0.95)	-15592	(38)	-11782	(38)	-15664	(41)	(1,1)	-12720	(14)	-9635	(12)	-12941	(10)
(125,0.95,0.97)	-13837	(25)	-10338	(24)	-13992	(27)	(1,2)	-12734	(15)	-9777	(14)	-12980	(11)
(125,0.97,0.97)	-13766	(21)	-10253	(20)	-13977	(26)	(2,1)	-12753	(16)	-9694	(13)	-12987	(12)
(75,0.95,0.95)	-16011	(43)	-12306	(45)	-16147	(43)	(2,2)	-12767	(17)	-9842	(15)	-13026	(13)
(75,0.97,0.95)	-16052	(44)	-12287	(44)	-16212	(44)							
(75,0.95,0.97)	-14548	(31)	-11038	(32)	-14760	(31)	DCC						
(75,0.97,0.97)	-14521	(30)	-10976	(31)	-14765	(32)	(p,q)						
(50,0.95,0.95)	-17995	(51)	-13846	(51)	-17905	(51)	(1,1)	-12165	(3)	-9087	(3)	-12398	(3)
(50,0.97,0.95)	-18072	(52)	-13872	(52)	-18000	(52)	(1,2)	-12175	(4)	-9217	(7)	-12434	(4)
(50,0.95,0.97)	-16734	(48)	-12776	(49)	-16688	(47)	(2,1)	-12203	(6)	-9146	(5)	-12444	(6)
(50,0.97,0.97)	-16733	(47)	-12748	(48)	-16714	(48)	(2,2)	-12214	(7)	-9281	(9)	-12483	(7)
MMA							ADCC						
(n_0, ν_0)							(p,q)						
(250,0.95)	-16995	(50)	-12418	(46)	-16566	(46)	(1,1)	-12138	(2)	-9058	(2)	-12373	(2)
(250,0.97)	-14773	(32)	-10643	(30)	-14570	(30)	(1,2)	-12182	(5)	-9109	(4)	-12442	(5)
(125,0.95)	-16336	(45)	-12163	(43)	-16455	(45)	(2,1)	-12241	(8)	-9168	(6)	-12498	(8)
(125,0.97)	-14293	(29)	-10501	(29)	-14553	(29)	(2,2)	-12284	(9)	-9221	(8)	-12539	(9)
(75,0.95)	-16588	(46)	-12539	(47)	-16722	(49)							
(75,0.97)	-14875	(33)	-11122	(33)	-15120	(33)	TDCC						
(50,0.95)	-18331	(53)	-14027	(53)	-18310	(53)	(p,q)						
(50,0.97)	-16869	(49)	-12818	(50)	-16907	(50)	(1,1)	-11542	(1)	-8767	(1)	-11952	(1)
GEWMA													
(p,q,ν_0)													
(1,1,0.95)	-15491	(35)	-11600	(34)	-15230	(34)							
(1,2,0.95)	-15444	(34)	-11785	(39)	-15243	(35)							
(2,1,0.95)	-15554	(37)	-11663	(35)	-15256	(36)							
(2,2,0.95)	-15523	(36)	-11852	(41)	-15289	(37)							

Notes: This table reports the Schwartz Bayesian Criteria (SBC) of the volatility models described in Appendix B under the assumption of standard normal innovations (with the exception of the TDCC model). The SBC is computed as $SBC_{it} = LL_{it} - 0.5k_i \log(T)$, where LL_{it} is the maximized log-likelihood at time t for model i, k_i is the number of parameters estimated under model i, and $T = 800$ is the sample size. For the TDCC model the distribution is Student t with ν_{t-1} degrees of freedom, where ν_{t-1} is re-estimated every 25 days. Columns 1 and 4 report the SBC values of the first sample (2-Jan-91 to 25-Jan-94, 800 observations). Columns 2 and 5 report the SBC values of the last sample (17-Jun-04 to 11-Jul-07, 800 observations). Columns 3 and 6 report the average SBC values over all 3512 rolling samples, with 800 observations each, extracted from the full sample of data (2-Jan-91 to 11-Jul-07). The models' rank is given in parentheses.

Table 5: AIC Values for Multivariate Volatility Models under Student t Innovations

Model type	Sample periods						Model type	Sample periods					
	25-Jan-94		11-Jul-07		Average			25-Jan-94		11-Jul-07		Average	
	(1)		(2)		(3)			(4)		(5)		(6)	
EQMA							(1,1,0.97)	-12823	(30)	-9669	(20)	-12862	(22)
(n_0)							(1,2,0.97)	-12793	(28)	-9769	(23)	-12856	(20)
(250)	-12477	(21)	-9184	(14)	-12543	(14)	(2,1,0.97)	-12841	(31)	-9674	(21)	-12862	(23)
(125)	-12175	(18)	-9225	(15)	-12695	(19)	(2,2,0.97)	-12811	(29)	-9776	(25)	-12865	(24)
(75)	-12522	(22)	-9720	(22)	-13163	(28)							
(50)	-13345	(36)	-10627	(41)	-13960	(42)	OGARCH						
							(p, q)						
EWMA							(1,1)	-11907	(14)	-9911	(27)	-12637	(17)
(n_0, λ_0, ν_0)							(1,2)	-11918	(15)	-9923	(28)	-12642	(18)
(250,0.95,0.95)	-13643	(40)	-10471	(37)	-13735	(38)	(2,1)	-11933	(16)	-10030	(30)	-12628	(16)
(250,0.97,0.95)	-13670	(43)	-10465	(36)	-13752	(39)	(2,2)	-11939	(17)	-10015	(29)	-12608	(15)
(250,0.95,0.97)	-12765	(24)	-9632	(18)	-12870	(25)							
(250,0.97,0.97)	-12766	(25)	-9592	(16)	-12862	(21)	CCC						
(125,0.95,0.95)	-13250	(33)	-10430	(35)	-13793	(40)	(p, q)						
(125,0.97,0.95)	-13281	(34)	-10424	(34)	-13816	(41)	(1,1)	-11754	(11)	-8994	(10)	-12144	(10)
(125,0.95,0.97)	-12463	(20)	-9649	(19)	-13000	(27)	(1,2)	-11749	(10)	-9101	(12)	-12147	(11)
(125,0.97,0.97)	-12461	(19)	-9613	(17)	-12997	(26)	(2,1)	-11765	(13)	-9007	(11)	-12150	(12)
(75,0.95,0.95)	-13365	(37)	-10678	(44)	-14015	(43)	(2,2)	-11755	(12)	-9120	(13)	-12156	(13)
(75,0.97,0.95)	-13405	(38)	-10681	(45)	-14045	(44)							
(75,0.95,0.97)	-12762	(23)	-10064	(32)	-13402	(31)	DCC						
(75,0.97,0.97)	-12772	(27)	-10041	(31)	-13407	(32)	(p, q)						
(50,0.95,0.95)	-14123	(50)	-11353	(51)	-14696	(51)	(1,1)	-11568	(7)	-8806	(6)	-11961	(5)
(50,0.97,0.95)	-14162	(51)	-11365	(52)	-14726	(52)	(1,2)	-11560	(5)	-8905	(8)	-11964	(7)
(50,0.95,0.97)	-13693	(45)	-10931	(49)	-14258	(48)	(2,1)	-11580	(9)	-8818	(7)	-11969	(8)
(50,0.97,0.97)	-13706	(47)	-10923	(48)	-14267	(49)	(2,2)	-11569	(8)	-8924	(9)	-11973	(9)
MMA							ADCC						
(n_0, ν_0)							(p, q)						
(250,0.95)	-14243	(52)	-10817	(47)	-14257	(47)	(1,1)	-11526	(2)	-8756	(2)	-11912	(2)
(250,0.97)	-13299	(35)	-9897	(26)	-13336	(30)	(1,2)	-11535	(3)	-8771	(3)	-11941	(3)
(125,0.95)	-13641	(39)	-10626	(40)	-14127	(45)	(2,1)	-11553	(4)	-8783	(4)	-11956	(4)
(125,0.97)	-12767	(26)	-9769	(24)	-13270	(29)	(2,2)	-11560	(6)	-8801	(5)	-11962	(6)
(75,0.95)	-13646	(41)	-10813	(46)	-14241	(46)							
(75,0.97)	-12962	(32)	-10134	(33)	-13566	(33)	TDCC						
(50,0.95)	-14299	(53)	-11452	(53)	-14843	(53)	(p, q)						
(50,0.97)	-13806	(49)	-10978	(50)	-14354	(50)	(1,1)	-11451	(1)	-8676	(1)	-11860	(1)
GEWMA													
(p, q, ν_0)													
(1,1,0.95)	-13697	(46)	-10531	(38)	-13716	(36)							
(1,2,0.95)	-13660	(42)	-10632	(42)	-13709	(34)							
(2,1,0.95)	-13720	(48)	-10534	(39)	-13715	(35)							
(2,2,0.95)	-13680	(44)	-10637	(43)	-13717	(37)							

Notes: This table reports the AIC of the volatility models described in Appendix B and estimated (when applicable) assuming normal innovations as in Table 3 but evaluated here using a Student t distribution with 7 degrees of freedom (with the exception of the TDCC model). For the TDCC model the distribution is Student t with ν_{t-1} degrees of freedom, where ν_{t-1} is re-estimated every 25 days. Columns 1 and 4 report the AIC values of the first sample (2-Jan-91 to 25-Jan-94, 800 observations). Columns 2 and 5 report the AIC values of the last sample (17-Jun-04 to 11-Jul-07, 800 observations). Columns 3 and 6 report the average AIC values over all 3512 rolling samples, with 800 observations each, extracted from the full sample of data (2-Jan-91 to 11-Jul-07). The models' rank is given in parentheses.

Table 6: SBC Values for Multivariate Volatility Models under Student t Innovations

Model type	Sample periods							Sample periods					
	25-Jan-94		11-Jul-07		Average			25-Jan-94		11-Jul-07		Average	
	(1)		(2)		(3)			(4)		(5)		(6)	
EQMA							(1,1,0.97)	-12950	(28)	-9795	(22)	-12988	(22)
(n_0)							(1,2,0.97)	-12961	(29)	-9937	(25)	-13025	(25)
(250)	-12477	(21)	-9184	(10)	-12543	(10)	(2,1,0.97)	-13010	(31)	-9842	(23)	-13031	(26)
(125)	-12175	(14)	-9225	(11)	-12695	(14)	(2,2,0.97)	-13022	(32)	-9987	(26)	-13076	(27)
(75)	-12522	(22)	-9720	(20)	-13163	(28)							
(50)	-13345	(36)	-10627	(39)	-13960	(42)	OGARCH						
							(p, q)						
EWMA							(1,1)	-12034	(10)	-10038	(27)	-12764	(16)
(n_0, λ_0, ν_0)							(1,2)	-12086	(11)	-10092	(30)	-12811	(18)
(250,0.95,0.95)	-13643	(40)	-10471	(37)	-13735	(34)	(2,1)	-12102	(12)	-10199	(32)	-12796	(17)
(250,0.97,0.95)	-13670	(42)	-10465	(36)	-13752	(35)	(2,2)	-12150	(13)	-10226	(33)	-12819	(19)
(250,0.95,0.97)	-12765	(24)	-9632	(17)	-12870	(21)							
(250,0.97,0.97)	-12766	(25)	-9592	(14)	-12862	(20)	CCC						
(125,0.95,0.95)	-13250	(33)	-10430	(35)	-13793	(36)	(p, q)						
(125,0.97,0.95)	-13281	(34)	-10424	(34)	-13816	(37)	(1,1)	-12239	(15)	-9479	(12)	-12628	(11)
(125,0.95,0.97)	-12463	(20)	-9649	(18)	-13000	(24)	(1,2)	-12276	(16)	-9628	(16)	-12674	(12)
(125,0.97,0.97)	-12461	(19)	-9613	(15)	-12997	(23)	(2,1)	-12292	(17)	-9534	(13)	-12677	(13)
(75,0.95,0.95)	-13365	(37)	-10678	(41)	-14015	(43)	(2,2)	-12325	(18)	-9690	(19)	-12725	(15)
(75,0.97,0.95)	-13405	(38)	-10681	(42)	-14045	(44)							
(75,0.95,0.97)	-12762	(23)	-10064	(29)	-13402	(31)	DCC						
(75,0.97,0.97)	-12772	(27)	-10041	(28)	-13407	(32)	(p, q)						
(50,0.95,0.95)	-14123	(50)	-11353	(51)	-14696	(51)	(1,1)	-11699	(2)	-8937	(3)	-12093	(3)
(50,0.97,0.95)	-14162	(51)	-11365	(52)	-14726	(52)	(1,2)	-11734	(4)	-9078	(7)	-12137	(4)
(50,0.95,0.97)	-13693	(43)	-10931	(49)	-14258	(48)	(2,1)	-11753	(6)	-8992	(5)	-12142	(5)
(50,0.97,0.97)	-13706	(44)	-10923	(48)	-14267	(49)	(2,2)	-11784	(7)	-9139	(9)	-12189	(7)
MMA							ADCC						
(n_0, ν_0)							(p, q)						
(250,0.95)	-14243	(52)	-10817	(46)	-14257	(47)	(1,1)	-11702	(3)	-8931	(2)	-12088	(2)
(250,0.97)	-13299	(35)	-9897	(24)	-13336	(30)	(1,2)	-11753	(5)	-8989	(4)	-12159	(6)
(125,0.95)	-13641	(39)	-10626	(38)	-14127	(45)	(2,1)	-11813	(8)	-9043	(6)	-12216	(8)
(125,0.97)	-12767	(26)	-9769	(21)	-13270	(29)	(2,2)	-11863	(9)	-9103	(8)	-12264	(9)
(75,0.95)	-13646	(41)	-10813	(45)	-14241	(46)							
(75,0.97)	-12962	(30)	-10134	(31)	-13566	(33)	TDCC						
(50,0.95)	-14299	(53)	-11452	(53)	-14843	(53)	(p, q)						
(50,0.97)	-13806	(45)	-10978	(50)	-14354	(50)	(1,1)	-11542	(1)	-8767	(1)	-11952	(1)
GEWMA													
(p, q, ν_0)													
(1,1,0.95)	-13824	(46)	-10658	(40)	-13842	(38)							
(1,2,0.95)	-13829	(47)	-10801	(44)	-13877	(39)							
(2,1,0.95)	-13888	(48)	-10702	(43)	-13884	(40)							
(2,2,0.95)	-13891	(49)	-10847	(47)	-13927	(41)							

Notes: This table reports the SBC of the volatility models described in Appendix B and estimated (when applicable) assuming normal innovations as in Table 4 but evaluated here using a Student t distribution with 7 degrees of freedom (with the exception of the TDCC model). For the TDCC model the distribution is Student t with ν_{t-1} degrees of freedom, where ν_{t-1} is re-estimated every 25 days. Columns 1 and 4 report the SBC values of the first sample (2-Jan-91 to 25-Jan-94, 800 observations). Columns 2 and 5 report the SBC values of the last sample (17-Jun-04 to 11-Jul-07, 800 observations). Columns 3 and 6 report the average SBC values over all 3512 rolling samples, with 800 observations each, extracted from the full sample of data (2-Jan-91 to 11-Jul-07). The models' rank is given in parentheses.

Table 7: Information Ratios and VaR Diagnostic Tests for Optimal Portfolio (Individual Multivariate Models)

Model type		Normal					Student (7)				
		mean return	IR	$\hat{\pi}$	$z_{\hat{\pi}}$	% VaR binds	mean return	IR	$\hat{\pi}$	$z_{\hat{\pi}}$	% VaR binds
EQMA	(n_0)										
	(250)	11.66	1.38	2.99	11.85	22	11.43	1.43	2.45	8.63	34
	(125)	13.62	1.42	3.73	16.26	28	13.15	1.46	3.33	13.89	42
	(75)	13.86	1.24	5.61	27.46	36	13.34	1.29	5.04	24.07	51
	(50)	16.52	1.18	9.14	48.49	48	15.67	1.22	8.15	42.55	64
EWMA	(n_0, λ_0, ν_0)										
	(250,0.95,0.95)	14.16	1.05	8.74	46.12	48	13.69	1.11	7.72	40.01	63
	(250,0.97,0.95)	13.69	0.99	8.94	47.30	47	13.31	1.06	7.97	41.54	62
	(250,0.95,0.97)	13.28	1.28	5.33	25.76	35	12.84	1.32	4.67	21.86	51
	(250,0.97,0.97)	12.96	1.24	5.24	25.25	34	12.56	1.29	4.24	19.32	49
	(125,0.95,0.95)	14.21	1.05	8.86	46.80	48	13.74	1.11	7.80	40.52	63
	(125,0.97,0.95)	13.79	0.99	9.09	48.15	47	13.39	1.06	8.15	42.55	62
	(125,0.95,0.97)	13.54	1.27	5.72	28.14	36	13.08	1.31	5.04	24.07	52
	(125,0.97,0.97)	13.26	1.23	5.64	27.63	36	12.83	1.28	4.87	23.05	51
	(75,0.95,0.95)	14.37	1.01	9.97	53.41	51	13.87	1.06	8.54	44.93	66
	(75,0.97,0.95)	14.04	0.96	10.08	54.09	51	13.61	1.02	8.86	46.80	65
	(75,0.95,0.97)	13.86	1.16	7.35	37.81	41	13.31	1.20	6.44	32.38	58
	(75,0.97,0.97)	13.62	1.12	7.46	38.48	41	13.12	1.17	6.41	32.21	57
	(50,0.95,0.95)	16.28	0.95	12.70	69.69	60	15.55	1.01	11.51	62.57	73
	(50,0.97,0.95)	15.87	0.90	12.76	70.03	60	15.25	0.96	11.28	61.21	73
	(50,0.95,0.97)	16.43	1.08	10.97	59.35	53	15.64	1.13	9.46	50.36	69
	(50,0.97,0.97)	16.22	1.05	11.02	59.69	53	15.45	1.10	9.60	51.21	69
MMA	(n_0, ν_0)										
	(250,0.95)	10.39	0.61	9.77	52.22	46	10.45	0.69	8.66	45.61	59
	(250,0.97)	11.36	0.95	5.58	27.29	34	11.12	1.00	4.84	22.88	47
	(125,0.95)	12.61	0.78	9.85	52.73	48	12.38	0.85	8.54	44.93	61
	(125,0.97)	12.96	1.08	6.10	30.34	37	12.67	1.15	5.27	25.42	51
	(75,0.95)	12.86	0.79	10.45	56.29	51	12.84	0.88	9.51	50.70	65
	(75,0.97)	13.18	1.02	7.92	41.20	42	12.85	1.08	6.95	35.43	57
	(50,0.95)	14.36	0.75	13.04	71.73	59	14.29	0.85	11.65	63.42	72
	(50,0.97)	15.64	0.97	11.08	60.03	53	15.01	1.03	9.97	53.41	68
GEWMA	(p, q, ν_0)										
	(1,1,0.95)	12.54	0.95	8.57	45.10	43	12.17	1.00	7.41	38.14	58
	(1,2,0.95)	13.12	1.01	8.54	44.93	43	12.71	1.06	7.32	37.64	57
	(2,1,0.95)	12.72	0.94	8.37	43.91	43	12.43	1.00	7.23	37.13	57
	(2,2,0.95)	12.94	0.96	8.77	46.29	43	12.61	1.02	7.60	39.33	57
	(1,1,0.97)	11.65	1.16	4.90	23.22	31	11.34	1.20	4.10	18.47	45
	(1,2,0.97)	12.08	1.21	4.90	23.22	30	11.75	1.26	4.07	18.30	45
	(2,1,0.97)	12.08	1.18	5.07	24.24	31	11.77	1.23	4.24	19.32	45
	(2,2,0.97)	12.19	1.19	5.10	24.41	32	11.88	1.24	4.50	20.84	45

Nominal tolerance probability $\alpha = 1\%$

Notes: This table reports the mean return, information ratio (IR), VaR exceedance ratio ($\hat{\pi}$) and its $z_{\hat{\pi}}$-statistic, as well as the percentage of times the VaR constraint is binding, of portfolios that were constructed based on the estimated mean return (μ_t – generated by an AR(1) model) and the covariance matrix ($\mathbf{\Sigma}_{it}$ – generated by the individual multivariate volatility models described in Appendix B). The risk-aversion parameter δ is set to 75 and the nominal VaR frequency α is 1%. We assume that innovations are either normal or Student t with 7 degrees of freedom. For the TDCC model the Student t distribution has ν_{t-1} degrees of freedom (instead of 7), where ν_{t-1} is re-estimated every 25 days. The multivariate models were estimated using a rolling window of 800 observations over the period 2-Jan-91 to 11-Jul-07, 3511 rolling samples in total.

Table 7 cont'd: Information Ratios and VaR Diagnostic Tests for Optimal Portfolio (Individual Multivariate Models)

Model type		Normal					Student (7)				
		mean return	IR	$\hat{\pi}$	$z_{\hat{\pi}}$	% VaR binds	mean return	IR	$\hat{\pi}$	$z_{\hat{\pi}}$	% VaR binds
OGARCH	(p,q)										
	(1,1)	9.70	1.19	2.73	10.33	16	9.59	1.24	2.34	7.95	27
	(1,2)	9.34	1.13	2.62	9.65	16	9.22	1.17	2.45	8.63	27
	(2,1)	9.29	1.15	2.68	9.99	16	9.16	1.19	2.31	7.78	27
	(2,2)	9.63	1.20	2.51	8.97	16	9.48	1.24	2.28	7.61	27
CCC	(p,q)										
	(1,1)	9.95	1.43	1.79	4.73	13	9.82	1.46	1.48	2.86	23
	(1,2)	10.23	1.47	1.82	4.90	12	10.10	1.51	1.62	3.71	23
	(2,1)	10.13	1.42	2.08	6.43	13	9.98	1.46	1.82	4.90	24
	(2,2)	10.21	1.42	1.94	5.58	14	10.05	1.46	1.77	4.56	24
DCC	(p,q)										
	(1,1)	9.41	1.37	1.57	3.37	13	9.31	1.41	1.40	2.36	24
	(1,2)	9.65	1.42	1.68	4.05	12	9.55	1.45	1.48	2.86	23
	(2,1)	9.65	1.38	1.85	5.07	14	9.54	1.42	1.62	3.71	25
	(2,2)	9.69	1.38	1.88	5.24	14	9.57	1.42	1.59	3.54	24
ADCC	(p,q)										
	(1,1)	8.76	1.27	1.74	4.39	14	8.70	1.31	1.57	3.37	25
	(1,2)	8.92	1.24	1.99	5.92	14	8.93	1.30	1.74	4.39	25
	(2,1)	9.36	1.30	1.91	5.41	15	9.24	1.34	1.57	3.37	27
	(2,2)	9.31	1.25	2.05	6.26	16	9.31	1.32	1.77	4.56	27
TDCC	(p,q)										
	(1,1)	-	-	-	-	-	9.21	1.39	1.57	3.37	21

Nominal tolerance probability $\alpha = 1\%$

Table 8: Information Ratios and VaR Diagnostic Tests for Optimal Portfolio (Multivariate Modelling Strategies)

Model type	Normal					Student (7)				
	mean return	IR	$\hat{\pi}$	$z_{\hat{\pi}}$	% VaR binds	mean return	IR	$\hat{\pi}$	$z_{\hat{\pi}}$	% VaR binds
Best (AIC)	9.30	1.29	1.91	5.41	15	9.30	1.34	1.59	3.54	26
Best (BIC)	9.24	1.27	2.05	6.26	15	9.39	1.36	1.62	3.71	26
Bayesian Average (AIC)	9.33	1.29	1.91	5.41	15	9.29	1.34	1.59	3.54	26
Bayesian Average (BIC)	9.27	1.28	2.08	6.43	15	9.40	1.36	1.62	3.71	26
Top 10% (AIC)	9.24	1.38	1.59	3.54	13	8.83	1.34	1.40	2.36	24
Top 10% (BIC)	9.05	1.34	1.57	3.37	13	8.74	1.33	1.48	2.86	25
Top 25% (AIC)	8.70	1.34	1.37	2.19	13	8.72	1.37	1.20	1.17	24
Top 25% (BIC)	9.04	1.37	1.51	3.03	13	8.90	1.38	1.31	1.85	24
Top 50% (AIC)	9.18	1.37	1.57	3.37	15	9.08	1.40	1.28	1.68	26
Top 50% (BIC)	9.19	1.38	1.54	3.20	15	9.10	1.41	1.28	1.68	26
Top 75% (AIC)	9.68	1.36	1.82	4.90	20	9.56	1.39	1.40	2.36	32
Top 75% (BIC)	9.68	1.35	1.79	4.73	20	9.56	1.39	1.42	2.53	31
All	10.11	1.32	2.36	8.12	25	9.95	1.36	1.79	4.73	36

Nominal tolerance probability $\alpha = 1\%$

Notes: Table 8 reports the mean return, information ratio (IR), VaR exceedance frequency ($\hat{\pi}$) and its $z_{\hat{\pi}}$-statistic, as well as the percentage of times the VaR constraint is binding, of portfolios that were constructed based on the estimated mean return (μ_t – generated by an AR(1) model) and a set of covariance matrices $\{\mathbf{\Sigma}_{it}\}$ with weights $\{\lambda_{it}\}$. The risk-aversion parameter δ is set to 75 and the nominal VaR frequency α is 1%. We assume that innovations are either normal or Student t with 7 degrees of freedom. The multivariate models were estimated using a rolling window of 800 observations over the period 2-Jan-91 to 11-Jul-07, 3511 rolling samples in total.

Table 9: Kuiper and Kolmogorov-Smirnov Tests of the Validity of the Individual Multivariate Models Using an Equally Weighted Portfolio

Model type	normal		t7		Model type	normal		t7	
	Ku	KS	Ku	KS		Ku	KS	Ku	KS
EQMA					(1,1,0.97)	0.000	0.005	0.003	0.052
(n_0)					(1,2,0.97)	0.000	0.006	0.003	0.052
(250)	0.000	0.004	0.012	0.104	(2,1,0.97)	0.000	0.006	0.002	0.048
(125)	0.000	0.006	0.004	0.052	(2,2,0.97)	0.000	0.008	0.002	0.048
(75)	0.000	0.007	0.001	0.030					
(50)	0.000	0.007	0.000	0.033	OGARCH				
					(p,q)				
EWMA					(1,1)	0.000	0.015	0.000	0.002
(n_0,λ_0,ν_0)					(1,2)	0.000	0.015	0.000	0.002
(250,0.95,0.95)	0.000	0.008	0.000	0.022	(2,1)	0.000	0.015	0.000	0.003
(250,0.97,0.95)	0.000	0.005	0.000	0.048	(2,2)	0.000	0.015	0.000	0.001
(250,0.95,0.97)	0.000	0.008	0.000	0.027					
(250,0.97,0.97)	0.000	0.004	0.001	0.048	CCC				
(125,0.95,0.95)	0.000	0.008	0.000	0.022	(p,q)				
(125,0.97,0.95)	0.000	0.006	0.000	0.040	(1,1)	0.000	0.008	0.002	0.036
(125,0.95,0.97)	0.000	0.008	0.000	0.024	(1,2)	0.000	0.008	0.001	0.036
(125,0.97,0.97)	0.000	0.005	0.001	0.043	(2,1)	0.000	0.010	0.001	0.030
(75,0.95,0.95)	0.000	0.009	0.000	0.018	(2,2)	0.000	0.010	0.001	0.022
(75,0.97,0.95)	0.000	0.006	0.000	0.030					
(75,0.95,0.97)	0.000	0.010	0.000	0.020	DCC				
(75,0.97,0.97)	0.000	0.007	0.000	0.024	(p,q)				
(50,0.95,0.95)	0.000	0.012	0.000	0.008	(1,1)	0.000	0.007	0.002	0.043
(50,0.97,0.95)	0.000	0.008	0.000	0.012	(1,2)	0.000	0.008	0.001	0.040
(50,0.95,0.97)	0.000	0.013	0.000	0.009	(2,1)	0.000	0.008	0.002	0.043
(50,0.97,0.97)	0.000	0.010	0.000	0.012	(2,2)	0.000	0.008	0.001	0.030
MMA					ADCC				
(n_0,ν_0)					(p,q)				
(250,0.95)	0.000	0.004	0.002	0.075	(1,1)	0.001	0.012	0.000	0.016
(250,0.97)	0.000	0.004	0.010	0.096	(1,2)	0.000	0.009	0.000	0.012
(125,0.95)	0.000	0.005	0.000	0.052	(2,1)	0.000	0.010	0.000	0.016
(125,0.97)	0.000	0.005	0.001	0.075	(2,2)	0.000	0.012	0.000	0.016
(75,0.95)	0.000	0.008	0.000	0.040					
(75,0.97)	0.000	0.007	0.000	0.043	TDCC				
(50,0.95)	0.000	0.005	0.000	0.040	(p,q)				
(50,0.97)	0.000	0.006	0.000	0.036	(1,1)	0.000	0.008	0.012	0.096
GEWMA									
(p,q,ν_0)									
(1,1,0.95)	0.000	0.006	0.001	0.043					
(1,2,0.95)	0.000	0.005	0.001	0.048					
(2,1,0.95)	0.000	0.008	0.001	0.036					
(2,2,0.95)	0.000	0.008	0.001	0.043					

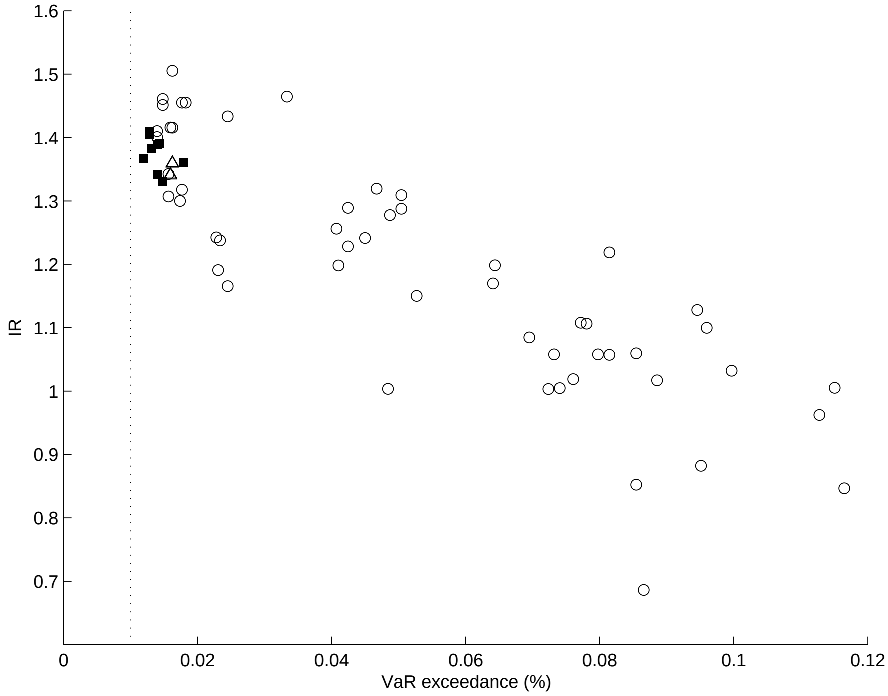
Notes: This table reports the probability values for the Kuiper and the Kolmogorov-Smirnov tests for the different multivariate volatility models considered in this paper. We assume that the innovations are either normal or follow a Student t distribution with 7 degrees of freedom. For the TDCC model the Student t distribution has ν_{t-1} degrees of freedom (instead of 7), where ν_{t-1} is re-estimated every 25 days. The multivariate models were estimated using a rolling window of 800 observations over the period 2-Jan-91 to 11-Jul-07, 3511 rolling samples in total.

Table 10: Kuiper and Kolmogorov-Smirnov Tests of the Validity of the Multivariate Model Averaging Strategies Using an Equally Weighted Portfolio

Strategy type	normal		t7	
	Ku	KS	Ku	KS
Best (AIC)	0.000	0.012	0.000	0.024
Best (BIC)	0.000	0.012	0.000	0.024
Bayesian Average (AIC)	0.000	0.012	0.000	0.024
Bayesian Average (BIC)	0.000	0.012	0.000	0.024
Top 10% (AIC)	0.000	0.009	0.000	0.022
Top 10% (BIC)	0.000	0.012	0.000	0.022
Top 25% (AIC)	0.001	0.013	0.000	0.015
Top 25% (BIC)	0.000	0.009	0.000	0.027
Top 50% (AIC)	0.000	0.010	0.000	0.022
Top 50% (BIC)	0.000	0.010	0.000	0.024
Top 75% (AIC)	0.000	0.010	0.000	0.022
Top 75% (BIC)	0.000	0.010	0.000	0.024
All	0.001	0.010	0.000	0.020

Notes: This table reports the probability values for the Kuiper and the Kolmogorov-Smirnov tests for the average multivariate volatility models considered in this paper. We assume that the innovations are either normal or follow a Student t distribution with 7 degrees of freedom. The multivariate models were estimated using a rolling window of 800 observations over the period 2-Jan-91 to 11-Jul-07, 3511 rolling samples in total.

Figure 1: Scatter Plot of Information Ratios and VaR Exceedance Frequencies Across Models and Modelling Strategies



References

Aiolfi, M., C. Favero, and G. Primiceri (2001): "Recursive 'thick' modeling of excess returns and portfolio allocation," IGIER Working Paper 197, IGIER.

Alexander, C. (2001): "Orthogonal GARCH," in *Mastering risk*, ed. by C. Alexander, vol. 2, pp. 21–38. London: Financial Times - Prentice Hall.

Andersen, A., T. Bollerslev, F. Diebold, and P. Labys (2003): "Modeling and forecasting realized volatility," *Econometrica*, 71, 579–626.

Andersen, T., T. Bollerslev, P. Christoffersen, and F. Diebold (2006): "Practical Volatility and Correlation Modeling for Financial Market Risk Management," in *Risks of Financial Institutions*, ed. by M. Carey,and R. Stulz, pp. 513–548. University of Chicago Press for NBER.

Bauwens, L., S. Laurent, and J. Rombouts (2003): "Multivariate GARCH models: a survey," *CORE*, Preprint.

Berkowitz, J. (2001): "Testing the accuracy of density forecasts, with applications to risk management," *Journal of Business and Economic Statistics*, 19, 465–474.

Billingsley, P. (1986): *Probability and measure*. New York: Wiley, second edn.

Bollerslev, T. (1986): "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 31, 302–327.

Bollerslev, T. (1990): "Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH," *Review of Economics and Statistics*, 72, 498–505.

Bollerslev, T., R. Engle, and J. Wooldridge (1988): "A capital asset pricing model with time varying covariances," *Journal of Political Economy*, 96, 116–131.

Brown, B. (1971): "Martingale central limit theorem," *The Annals of Mathematical Statistics*, 35, 1296–1303.

Burden, R. L., and J. D. Faires (1997): *Numerical Analysis*. Pacific Grove: Brooks-Cole Publishing, 6 edn.

Burnham, K., and D. Anderson (1998): *Model Selection and Inference: A Practical Information-Theoretic Approach*. New York: Springer Verlag.

Cappiello, L., R. Engle, and K. Sheppard (2006): "Asymmetric dynamics in the conditional correlations of global equity and bond returns," *Journal of Financial Econometrics*, 4(4), 537–572.

Christoffersen, P. (1998): "Evaluating interval forecasts," *International Economic Review*, 39, 841–862.

Clemen, R. (1989): "Combining forecasts: a review and annotated bibliography," *International Journal of Forecasting*, 5, 559–583.

Crnkovic, C., and J. Drachman (1997): "Quality control," *Risk*, 9, 139–143.

De Santis, G., and B. Gerard (1997): "International asset pricing and portfolio diversification with time-varying risk," *Journal of Finance*, 52, 1881–1912.

De Santis, G., R. Litterman, A. Vesval, and K. Winkelmann (2003): "Covariance matrix estimation," in *Modern investment management: an equilibrium approach*, ed. by R. Litterman. London: Wiley.

Diebold, F., T. Gunther, and A. Tay (1998): "Evaluating density forecasts with application to financial risk management," *International Economic Review*, 39, 863–883.

Diebold, F., J. Hahn, and A. Tay (1999): "Multivariate density forecast evaluation and calibration in financial risk management: high-frequency returns on foreign exchange," *The Review of Economics and Statistics*, 81, 661–673.

Diebold, F., and J. Lopez (1996): "Forecast evaluation and combination," in *Handbook of Statistics*, ed. by G. G.S. Maddala,and C. Rao, pp. 214–268. Amsterdam: North Holland.

Diebold, F., and M. Nerlove (1989): "The dynamics of exchange rate volatility: a multivariate latent factor ARCH model," *Journal of Applied Econometrics*, 4, 1–21.

Draper, D. (1995): "Assessment and propagation of model uncertainty (with discussion)," *Journal of the Royal Statistical Society, Series B*, 57, 45–97.

Engle, R. (2002): "Dynamic conditional correlation - a simple class of multivariate GARCH models," *Journal of Business and Economic Statistics* , (add issue and page numbers).

Engle, R., V. Ng, and M. Rothschild (1990): "Asset pricing with a FACTOR-ARCH covariance structure: empirical estimates for treasury bills," *Journal of Econometrics*, 45, 213–237.

Fernandez, C., E. Ley, and M. F. J. Steel (2001a): "Benchmark priors for bayesian model averaging," *Journal of Econometrics*, 100, 381–427.

Fernandez, C., E. Ley, and M. F. J. Steel (2001b): "Model uncertainty in cross-country growth regressions," *Journal of Applied Econometrics*, 16, 563–576.

Garratt, A., K. Lee, M. H. Pesaran, and Y. Shin (2003): "Forecast uncertainties in macroeconometric modelling: an application to the UK economy," *Journal of the American Statistical Association*, 98, 829–838.

Ghysels, E., A. Harvey, and E. Renault (1995): "Stochastic Volatility," in *Handbook of Statistics vol 14*. Amsterdam: North Holland.

Glosten, L., R. Jagannathan, and D. Runkle (1993): "On the relation between the expected value and the volatility of nominal excess return on stocks," *Journal of Finance*, 48, 1779–1801.

Godsill, S., M. Stone, and M. Weeks (2004): "Assessing the impact of private sector balance sheets on financial crises: a comparison of Bayesian and information-theoretic measures of model uncertainty," *University of Cambridge*, Preprint.

Granger, C. (1989): "Combining forecasts - Twenty years later," *Journal of Forecasting*, 8, 167–173.

Granger, C., and Y. Jeon (2004): "Thick modeling," *Economic Modeling*, 21, 323–343.

Granger, C., and P. Newbold (1977): *Forecasting economic time series.* New York; London: Academic Press.

Harvey, A., E. Ruiz, and E. Sentana (1992): "Unobservable components time series models with ARCH disturbances," *Journal of Econometrics*, 52, 129–158.

Harvey, A., E. Ruiz, and N. Shephard (1994): "Multivariate Stochastic Variance Models," *Review of Economic Studies*, 61, 247–264.

Hendry, D., and M. Clements (2002): "Pooling of forecasts," *Econometrics Journal*, 5, 1–26.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999): "Bayesian model averaging: a tutorial," *Statistical Science*, 14, 382–417.

J.P.Morgan (1996): "RiskMetrics^{TM} - Technical Document, Fourth Edition," .

King, M., E. Sentana, and S. Wadhwani (1994): "Volatility and links between national stock markets," *Econometrica*, 62, 901–933.

Kupiec, P. (1995): "Techniques for verifying the accuracy of risk management models," *The Journal of Derivatives*, 3, 73–84.

Ling, S., and M. McAleer (2001): "Necessary and sufficient moment conditions for the GARCH(r, s) and the asymmetric power GARCH(r, s) models," *Econometric Theory* , forthcoming.

McAleer, M. (2005): "Automated Inference and Learning in Modeling Financial Volatility," *Econometric Theory*, 21, 232–261.

McCracken, M. (2000): "Robust out-of-sample inference," *Journal of Econometrics*, 99, 195–223.

Nelson, D., and C. Cao (1992): "Inequality constraints in the univariate GARCH model," *Journal of Business & Economic Statistics*, 10, 229–235.

Pesaran, B., and M. H. Pesaran (2007): "Modelling Volatilities and Conditional Correlations in Futures Markets with a Multivariate t Distribution," *CESifo Working Paper No. 2056*.

Pesaran, M., and S. Skouras (2002): "Decision-based methods for forecast evaluation," in *A Companion to Economic Forecasting*, ed. by M. Clements,and D. Hendry, pp. 241–267. Oxford: Basil Blackwell.

Rosenblatt, M. (1952): "Remarks on a multivariate transformation," *Annals of Mathematical Statistics*, 23, 470–472.

Shephard, N. (2004): *Stochastic volatility: selected readings*. Oxford: Oxford University Press, forthcoming.

Stock, J., and M. Watson (1999): "A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series," in *Cointegration, causality and forecasting*, ed. by R. Engle,and H. White. Oxford: Oxford University Press, Chapter 1.

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

West, K. (1996): "Asymptotic inference about predictive ability," *Econometrica*, 64, 1067–1084.

Zaffaroni, P. (2007): "Estimating and forecasting volatility with large scale models: theoretical appraisal of professionals' practice," *forthcoming in the Journal of Time Series Analysis*.