DISCUSSION PAPER SERIES

No. 5279
CEPR/EABCN No. 22/2005

MODEL AVERAGING AND VALUE-AT-RISK BASED EVALUATION OF LARGE MULTI-ASSET VOLATILITY MODELS FOR RISK MANAGEMENT

M Hashem Pesaran and Paolo Zaffaroni

INDUSTRIAL ORGANIZATION and INTERNATIONAL MACROECONOMICS

€ABCN

Euro Area Business Cycle Network

www.eabcn.org

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP5279.asp**

ISSN 0265-8003

MODEL AVERAGING AND VALUE-AT-RISK BASED EVALUATION OF LARGE MULTI-ASSET VOLATILITY MODELS FOR RISK MANAGEMENT

M Hashem Pesaran, University of Cambridge
Paolo Zaffaroni, Banca d'Italia

Discussion Paper No. 5279
CEPR/EABCN No. 22/2005
October 2005

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 5279
CEPR/EABCN No. 22/2005

October 2005

ABSTRACT

Model Averaging and Value-at-Risk Based Evaluation of Large Multi-Asset Volatility Models for Risk Management*

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 based on 22 of Standard & Poor's 500 industry group indices over the period 1995-2003. We find strong evidence in support of 'thick' modelling proposed in the forecasting literature by Granger and Jeon (2004).

JEL Classification: C32, C52, C53 and G11
Keywords: decision-based evaluations, model averaging and value-at-risk

M Hashem Pesaran
Faculty of Economics and Politics
University of Cambridge
Austin Robinson Building
Sidgwick Avenue
CAMBRIDGE
CB3 9DE
Tel: (44 1223) 335216
Fax: (44 1223) 335471
Email: hashem.pesaran@econ.cam.ac.uk

Paolo Zaffaroni
Banca d'Italia
Via Nazionale 91
00184 Roma
ITALY
Tel: (39 06) 47921
Email: paolo.zaffaroni@econ.cam.ac.uk

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=100830

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=135577

* The publication of this Paper is funded by the Euro Area Business Cycle Network (www.eabcn.org). This Network provides a forum for the better understanding of the euro area business cycle, linking academic researchers and researchers in central banks and other policy institutions involved in the empirical analysis of the euro area business cycle. We would like to thank Gabriele Fiorentini and Frank Shorfeide for useful conversations, the Editor and an anonymous referee for useful suggestions that greatly improved the paper, and Vanessa Smith for her assistance with the data used in this study.

Submitted 16 September 2005

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 behaviour 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 monitoring of market risk in the case of large portfolios that are often encountered in practice. 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).[1] 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).[2] 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 either. Application of model selection procedures also involves additional risks when the number of assets is moderately large, and might very well be that no single model choice would

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

[2] 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.

be satisfactory in practice and could carry risks that are difficult to assess *a priori*. 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 averaging techniques to multi-asset volatility models. Frequently used model selection criteria are 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.[3] 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 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 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 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 Ling (1999). 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

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

in the paper, with a more detailed description available from the authors on request. 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 7 provides a detailed empirical analysis using daily returns on twenty two of Standard and Poor's 500 industry indices over the period January 2 1995 to October 13 2003. Section 8 concludes with a summary of the main results and suggestions for future research. The mathematical proofs are provided in an Appendix.

2 The Decision Problem

This paper is concerned with the decision of an individual fund manager who is interested in controlling the risk of a given portfolio over a given trading day. Denote the fund manager's asset positions at the close of business on day $t-1$ by the $N \times 1$ vector, $\mathbf{a}_{t-1} = (a_{1,t-1}, a_{2,t-1}, ..., a_{N,t-1})'$. The change in the value of this portfolio is given by $\Delta V_t = \sum_{j=1}^{N} (P_{jt} - P_{j,t-1}) a_{j,t-1} = \sum_{j=1}^{N} a_{j,t-1} P_{j,t-1} r_{jt}$, where P_{jt} is the price of the j^{th} asset at time t and $r_{jt} = (P_{jt} - P_{j,t-1}) / P_{j,t-1}$ is the associated daily rate of return, assuming negligible dividend payments for sake of simplicity. The rate of return of the portfolio can now be written as $\rho_t = \Delta V_t / V_{t-1} = \boldsymbol{\omega}'_{t-1}\mathbf{r}_t$, where $\mathbf{r}_t = (r_{1t}, r_{2t}, ..., r_{Nt})'$, and $\boldsymbol{\omega}_t = (\omega_{1t}, \omega_{2t}, ..., \omega_{Nt})'$, with $\omega_{it} = a_{it}P_{it} / \sum_{j=1}^{N} a_{jt}P_{jt}$. By construction $\boldsymbol{\tau}'\boldsymbol{\omega}_{t-1} = 1$ where $\boldsymbol{\tau}$ is an $N \times 1$ vector of unity. In the case of a fund manager who has been given the task of allocating a given sum, V_{t-1} on the N assets without the possibility of shorting, we have the additional non-negativity restrictions, $\omega_{it} \geq 0$, for all i.

The fund manager faces two different but closely related tasks, which we refer to as 'passive' and 'active' risk management problems. Under the latter the portfolio weights are treated as unknown and are determined by maximizing the expected utility of the portfolio, derived with respect to the conditional multivariate distribution of $\mathbf{r}_t$, subject to the non-negativity constraints (if applicable) and to the VaR constraint $\Pr\left(\rho_t < -\bar{\rho}_{t-1} \left| \mathcal{F}_{t-1}\right.\right) \leq \alpha$, where $\mathcal{F}_{t-1}$ is the available information, $\bar{\rho}_{t-1} > 0$ is a pre-specified rate of return and α is a probability value (typically taken to be 1%) which captures the trader's attitude to risk in the case of large losses. Under passive risk management $\boldsymbol{\omega}_{t-1}$ and α are assumed as given and the aim would to solve for $\bar{\rho}_{t-1}(\boldsymbol{\omega}_{t-1},\alpha)$ using

$$\Pr\left(\rho_t < -\bar{\rho}_{t-1}(\boldsymbol{\omega}_{t-1},\alpha) \left| \mathcal{F}_{t-1}\right.\right) \leq \alpha. \tag{1}$$

The capital at risk of the portfolio is then $L_{t-1}(\boldsymbol{\omega}_{t-1},\alpha) = V_{t-1}\bar{\rho}_{t-1}(\boldsymbol{\omega}_{t-1},\alpha)$, namely the maximum loss tolerated over day $t-1$ to t with probability α.

3 Multivariate Models of Asset Returns

For active risk management, a complete knowledge of the joint probability distribution of the vector of returns $\mathbf{r}_t$, conditional on available information, $\mathcal{F}_{t-1}$, would be needed. But for passive risk management it is clearly possible to work *directly* with the conditional distribution of $\rho_t = \boldsymbol{\omega}'_{t-1}\mathbf{r}_t$, with no apparent need for multivariate volatility modelling. Such a strategy is relatively simple to implement, but will be portfolio specific and could lead to contradictory outcomes if different portfolios are considered. Moreover, in comparing the risk of different portfolios it is important that the distribution of all portfolio returns are based on the same underlying multivariate model of $\mathbf{r}_t$.

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. Therefore, in what follows we shall focus on alternative specifications of the joint probability distribution of asset returns that we denote by $\Pr(\mathbf{r}_t |\mathcal{F}_{t-1})$. For this purpose it is convenient to work with the de-volatized returns, $\mathbf{z}_t$, defined by $\mathbf{z}_t = \boldsymbol{\Sigma}_t^{-\frac{1}{2}}\mathbf{r}_t$, where $\boldsymbol{\Sigma}_t = Var(\mathbf{r}_t |\mathcal{F}_{t-1})$ is the conditional covariance matrix of the returns assumed to be positive definite. Typically one would also need to model the conditional mean, $E(\mathbf{r}_t |\mathcal{F}_{t-1}) = \boldsymbol{\mu}_t$, although given the focus of the present paper on multivariate volatility models and the daily nature of the returns data that we shall be using to illustrate our approach we shall maintain that $\boldsymbol{\mu}_t = \mathbf{0}$, throughout.

A complete specification of $\Pr(\mathbf{r}_t |\mathcal{F}_{t-1})$ can be achieved by: (i) a non-singular choice of $\boldsymbol{\Sigma}_t$; (ii) specification of the distribution of de-volatized 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 $\boldsymbol{\Sigma}_t$, 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[4]

$$vech(\boldsymbol{\Sigma}_{MGARCH,t}) = \omega_0 + \mathbf{A}_0 vech(\boldsymbol{\Sigma}_{MGARCH,t-1}) + \mathbf{B}_0 vech\left(\mathbf{r}_{t-1}\mathbf{r}'_{t-1}\right), \qquad (2)$$

where $vech(.)$ denotes the column stacking operator of the lower portion of a symmetric matrix, ω_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 (2) 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

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

M_i and the associated conditional covariance matrix by $\mathbf{\Sigma}_{it}$. Altogether we consider 63 different specifications of $\mathbf{\Sigma}_{it}$ that can be grouped into ten 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 (2002). We also consider the orthogonal GARCH (O-GARCH(p,q)) of Alexander (2001), the factor GARCH model of Harvey, Ruiz, and Sentana (1992) (factor HRS $(p,q,1,1)$) and the factor GARCH of Diebold and Pesaran (1999) (factor DP $(p,q,1,1)$). Within the second group we consider equal-weighted moving average (EQMA(n_0)), which is a rolling filter that equally weights the most recent n_0 squared observations. We further consider the exponential-weighted moving average (EWMA$(n_0,\ \lambda_0)$), 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 $(n_0,\ \lambda_0,\nu_0)$) (see De Santis, Litterman, Vesval, and Winkelmann (2003, p.14)). We also consider two hybrid filters: a mixed moving average (MMA$(n_0,\ \nu_0)$) 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 a Supplement that is available from the authors on request.

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

$$\hat{\boldsymbol{\theta}}_{iT_0} = \arg\max_{\boldsymbol{\theta}_i \in \Theta_i} \left\{ -\frac{1}{2} \sum_{t=\tau-T_0+1}^{\tau} \left(\log \mid \mathbf{\Sigma}_{it}(\boldsymbol{\theta}_i) \mid + \mathbf{r}_t' \mathbf{\Sigma}_{it}^{-1}(\boldsymbol{\theta}_i) \mathbf{r}_t \right) \right\}, \qquad (3)$$

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. Correspondingly, let $\hat{\mathbf{\Sigma}}_{it} = \mathbf{\Sigma}_{it}(\hat{\boldsymbol{\theta}}_{iT_0})$. We view Gaussian PMLE as a robust method, delivering consistent and asymptotically normal estimates of $\boldsymbol{\theta}_i$ under the volatility model M_i even for non-Gaussian $\mathbf{z}_{it}$. In particular we shall assume that as $T_0 \to \infty$,

$$\hat{\boldsymbol{\theta}}_{iT_0} \xrightarrow{p} \boldsymbol{\theta}_{i0} \qquad (4)$$

and

$$\sqrt{T_0} \left(\hat{\boldsymbol{\theta}}_{iT_0} - \boldsymbol{\theta}_{i0} \right) \mid M_i \xrightarrow{d} N \left[\mathbf{0}, \mathbf{\Omega}_i \left(\boldsymbol{\theta}_{i0} \right) \right], \qquad (5)$$

where $\mathbf{\Omega}_i(\boldsymbol{\theta}_{i0})$ 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 (4) and (5) would hold for the more general class of models considered in this paper, under suitable regularity conditions.

In what follows we shall assume that under model M_i,

$$M_i: \quad \mathbf{r}_t = \mathbf{\Sigma}_{it}^{\frac{1}{2}}\mathbf{z}_{it}, \;\; \mathbf{z}_{it} \mid \; \mathcal{F}_{t-1} \sim (F_{it}, \; \mathbf{0}, \; \mathbf{I}_N), \tag{6}$$

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

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 $\Pr(\mathbf{r}_t \mid \mathcal{F}_{t-1}, M_i)$, be the predictive density of $\mathbf{r}_t$ conditional on model M_i and the in-sample available information, $\mathcal{F}_{t-1}$, and let the space of the models under consideration be $\mathcal{M} = \bigcup_{i=1}^{m} \{M_i\}$. 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

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

where the set of weights $\lambda_{i,t-1}$ are pre-determined at the time the decision over the the portfolio weights, $\omega_{j,t-1}$, $j = 1, 2, ..., N$, is taken. This is possible since it is assumed that there are no feedbacks from trade decisions to the probability models being considered. One could consider equally weighting 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 number of best performing models under consideration. Therefore, one could pool different models by taking simple averages, but after 'trimming' models with poor 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 subset of models $\mathcal{N}_{t-1}$. Such procedure, denominated as 'thick' modelling, has been proposed by Granger and Jeon (2004) who note that, among others, standard two-stage procedures, based on the AIC or SBC method, might exhibit poor performance simply because the 'true' model does not belong to the set of models under consideration.[5] Another example is the Bayesian Model Averaging (BMA) that combines the models under consideration using their respective posterior probabilities.[6] 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 $\boldsymbol{\theta}_i$ 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 latter gives a Bayesian approximation when the estimation sample, T_0, is sufficiently large.[7] 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 $\mathbf{r}_0, ..., \mathbf{r}_{-s_i+1}$.[8]

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 have important implications for the shape of the resulting average model in general and the degree of its fat-tailness, in particular. It seems likely that averaging across a very large number of models could be counter productive for density forecasting, although it might not be a problem in point forecasting.

[5] 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.

[6] 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).

[7] 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).

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

Further analysis of average models and their tail properties will be provided below in Section 6.

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.[9] 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 (1) associated with the passive version of the risk management problem where the portfolio weights, $\boldsymbol{\omega}_{t-1}$, are given, and suppose that the analysis is carried out conditional on model M_i. In this setting the VaR constraint is given by

$$\Pr\left(\rho_t < -\bar{\rho}_{i,t-1}\left|\mathcal{F}_{t-1}, M_i\right.\right) \leq \alpha, \tag{7}$$

and $\bar{\rho}_{i,t-1}$ will be a function of α and the assumed volatility model, M_i. To fully specify the model, assume that the de-volatized 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 $F_{it}(\cdot)$ 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 zero and variance $\sigma^2_{\rho t}(M_i) = \boldsymbol{\omega}'_{t-1}\boldsymbol{\Sigma}_{it}\,\boldsymbol{\omega}_{t-1}$. Therefore, under (6) we have

$$z_{\rho t}(M_i) = \frac{\boldsymbol{\omega}'_{t-1}\mathbf{r}_t}{\sigma_{\rho t}(M_i)}\left|\mathcal{F}_{t-1}, M_i\right. \sim (F_{it}, 0, 1). \tag{8}$$

This implies that under M_i, $z_{\rho t}(M_i)$ is a martingale difference sequence with a 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

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

because of possible serial dependence of portfolio weights and the temporal dependence of $\mathbf{\Sigma}_{it}$.

Denoting the value of $\bar{\rho}_{i,t-1}$ that satisfies (7) by $\bar{\rho}_{i,t-1}(\boldsymbol{\omega}_{t-1},\alpha)$ and assuming that (8) holds, then $F_{it}\left(-\bar{\rho}_{i,t-1}(\boldsymbol{\omega}_{t-1},\alpha)\sigma_{\rho t}^{-1}(M_i)\right) \leq \alpha$. But since $F_{it}(\cdot)$ is a continuous and monotonically non-decreasing function we have $-\bar{\rho}_{i,t-1}(\boldsymbol{\omega}_{t-1},\alpha)\sigma_{\rho t}^{-1}(M_i) = F_{it}^{-1}(\alpha) = -c_{it}(\alpha)$, or

$$\bar{\rho}_{i,t-1}(\boldsymbol{\omega}_{t-1},\alpha) = c_{it}(\alpha)\ \sigma_{\rho t}(M_i), \tag{9}$$

where $-c_{it}(\alpha)$ is the $\alpha\%$ 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 $\bar{\rho}_{i,t-1}(\boldsymbol{\omega}_{t-1},\alpha)$.

The above derivations hold even if the portfolio weights, $\boldsymbol{\omega}_{t-1}$, are derived conditional on model M_i. In that case the portfolio weights should be denoted by $\boldsymbol{\omega}_{i,t-1}$ 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.

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: $\hat{\pi}_i = \sum_{t\in\mathcal{T}_1} d_{it}(\hat{\boldsymbol{\theta}}_{iT_0})/T_1$, where $d_{it}(\hat{\boldsymbol{\theta}}_{iT_0}) = I[-\rho_t - c_{it}(\alpha)\ \hat{\sigma}_{\rho t}(M_i)]$ and $\hat{\sigma}_{\rho t}(M_i) = (\boldsymbol{\omega}_{t-1}'\hat{\mathbf{\Sigma}}_{it}\,\boldsymbol{\omega}_{t-1})^{\frac{1}{2}}$, $\hat{\mathbf{\Sigma}}_{it} = \mathbf{\Sigma}_{it}(\hat{\boldsymbol{\theta}}_{iT_0})$. Recall that $\hat{\boldsymbol{\theta}}_{iT_0}$ is the PMLE of the unknown parameters (if any) of $\mathbf{\Sigma}_{it}$ under model M_i (see (3)), 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

$$H_{i0}:\ \mathbf{\Sigma}_t = \ \mathbf{\Sigma}_{it} \text{ and } \ \mathbf{z}_{it} \mid \mathcal{F}_{t-1}, M_i \sim (F_{it},\ \mathbf{0}, \mathbf{I}_N). \tag{10}$$

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 $\hat{\pi}_i$

$$z_{\hat{\pi}_i} = \frac{\sqrt{T_1}(\hat{\pi}_i - \alpha)}{\sqrt{\alpha(1-\alpha)}} \tag{11}$$

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

Theorem 1 *(finite-T_1 distribution) Assume that* $\boldsymbol{\Sigma}_{it}(\theta_i)$ *is continuous in* θ_i *and that (5) holds. Let* $Bi(T_1, \alpha)$ *define a Binomial distribution with parameters* T_1 *and* α*. Then under* H_{i0},

$$T_1 \hat{\pi}_i \xrightarrow{d} Bi(T_1, \alpha), \qquad \text{as } T_0 \to \infty, \tag{12}$$

for any finite T_1, $0 < \alpha < 1$, *and any sequence of portfolio weights,* $\boldsymbol{\omega}_{t-1}$, $t = 0, \pm 1...$, *satisfying* $\| \boldsymbol{\omega}_{t-1} \| > 0$, *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 (5) holds and* $\boldsymbol{\theta}_{i0}$ *belongs to the interior of the compact set* Θ_i*; (iii)* $\boldsymbol{\Sigma}_{it}(\theta_i)$ *is twice continuously differentiable in* $\boldsymbol{\theta}_i$ *such that, for some* $\delta > 1$, $\inf_{\boldsymbol{\theta}_i \in \Theta_i} \underline{\lambda}_{it}(\boldsymbol{\theta}_i) > 0$, *a.s.*

$$E\{\sup_{\boldsymbol{\theta} \in \Theta_i} \frac{\| \partial \bar{\lambda}_{it}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} \|}{\underline{\lambda}_{it}^{\frac{1}{2}}(\boldsymbol{\theta}) \underline{\lambda}_{it}^{\frac{1}{2}}(\boldsymbol{\theta}_{i0})}\}^{\delta} = \mu_{it}, \quad \frac{1}{T_1} \sum_{t \in \mathcal{T}_1}^{T} \bar{f}_{it} \mu_{it}^{1/\delta} = O(1), \tag{13}$$

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

$$E \| \hat{\boldsymbol{\theta}}_{iT_0} - \boldsymbol{\theta}_{i0} \|^{\frac{\delta}{\delta - 1}} = O(T_0^{-\delta/(2(\delta-1))}). \tag{14}$$

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* $\boldsymbol{\omega}_{t-1}$, $t = 0, \pm 1...$, *satisfying* $\| \boldsymbol{\omega}_{t-1} \| > 0$.

Remarks:
(i) It is important to note that the null distribution of $z_{\hat{\pi}_i}$ does not depend on the portfolio weights, $\boldsymbol{\omega}_{t-1}$, although the power of the test typically does depend on $\boldsymbol{\omega}_{t-1}$.
(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 : $\boldsymbol{\Sigma}_{jt} \neq \boldsymbol{\Sigma}_{it}$ we have $E(\hat{\pi}_i | M_j) = \frac{1}{T_1} \sum_{t \in T_1} E[F_{it}(-c_{it}(\alpha) q_{ij,t})]$, where $q_{ij,t} = (\boldsymbol{\omega}'_{t-1} \hat{\boldsymbol{\Sigma}}_{it} \boldsymbol{\omega}_{t-1} / \boldsymbol{\omega}'_{t-1} \boldsymbol{\Sigma}_{jt} \boldsymbol{\omega}_{t-1})^{1/2}$, 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 (4) and (5) will imply (13) 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, such as for some of the models we consider, then the theorem applies by setting $\hat{\boldsymbol{\theta}}_i = \boldsymbol{\theta}_{i0}$ and the conditions (13) and (14) 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 $\mathbf{r}_t|\mathcal{F}_{t-1}, M_i \sim (F_{it}, \mathbf{0}, \boldsymbol{\Sigma}_{it}),\ i = 1, 2, ..., m$. 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 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, $J^{-1}\sum_{j=1}^{J} I(-\boldsymbol{\omega}'_{t-1}\mathbf{r}_{jt}^{(i)}+a) \to \Pr(\rho_t < a\,|\mathcal{F}_{t-1}\ , M_i)$ 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 are closed under linear transformations, as with Gaussian or multivariate t distribution, the computations can be simplified considerably by drawing from the distribution of $\rho_t = \boldsymbol{\omega}'_{t-1}\mathbf{r}_t$ 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 (11) for an individual model M_i, to the case of an average model. For a given α we need to find the value for $\bar{\rho}_{b,t-1}(\boldsymbol{\omega}_{t-1}, \alpha)$, the VaR associated with the BMA forecast probabilities, for which $\sum_{i=1}^{m} \lambda_{i,t-1} F_{it}(-\bar{\rho}_{b,t-1}(\boldsymbol{\omega}_{t-1},\alpha)/\sigma_{\rho t}(M_i)) \leq \alpha$. To solve for $\bar{\rho}_{b,t-1}(\boldsymbol{\omega}_{t-1},\alpha)$, let

$$g(\kappa) = \sum_{i=1}^{m} \lambda_{i,t-1} F_{it}\left(-\frac{\kappa}{\sigma_{\rho t}(M_i)}\right) - \alpha = 0, \tag{15}$$

and note that $g(\kappa) = 0$ has a unique positive solution under the additional assumption that all the model densities $f_{it}(\cdot) = F'_{it}(\cdot)$ are differentiable and have a unique maximum at zero. In the case of such distributions $\bar{\rho}_{b,t-1}(\boldsymbol{\omega}_{t-1},\alpha)$ can be easily computed using numerical techniques such as the Newton-Raphson iterative procedure. The VaR diagnostic statistic, given by (11), can then be computed for the average model using $\hat{d}_{bt} = I\left[-\rho_t - \bar{\rho}_{b,t-1}(\boldsymbol{\omega}_{t-1},\alpha)\right]$, in place of $d_{it}(\hat{\boldsymbol{\theta}}_{iT_0})$.

6 Tail Behavior of Average Volatility Models

It is well known that linear combinations (mixtures) of normal distributions is 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 instance, the pooled volatility forecast of portfolio return, with zero conditional means, is given by $V(\rho_t|F_{t-1}, M) = \sum_{i=1}^{m} \lambda_{it-1}\sigma_{\rho t}^2(M_i)$. However, tail probabilities using the mixture model and a Guassian model with the same average volatility are not the same, namely

$$\sum_{i=1}^{m} \lambda_{it-1}\Phi\left[\frac{a}{\sigma_{\rho t}(M_i)}\right] \neq \Phi\left[\frac{a}{\sqrt{\sum_{i=1}^{m} \lambda_{it-1}\sigma_{\rho t}^2(M_i)}}\right], \tag{16}$$

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 the Appendix, 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 (16), will be more fat-tailed than the associated Gaussian model with the same average volatility measure, on the right hand side of (16), 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 $b_1, ..., b_N$ of strictly positive constants satisfying*

$$a\left[(a/b_i^{\frac{1}{2}})f'(a/b_i^{\frac{1}{2}}) + 3f(a/b_i^{\frac{1}{2}})\right] > 0, \quad i = 1, ..., N, \tag{17}$$

it follows that

$$\sum_{i=1}^{N} \lambda_i F\left[a/(b_i)^{\frac{1}{2}}\right] > F\left[a/(\sum_{i=1}^{N} \lambda_i b_i)^{\frac{1}{2}}\right], \tag{18}$$

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

Remarks:
(i) When $f(u)$ is the standard normal density, for $a < 0$ condition (17) is

$$a/b_i^{\frac{1}{2}} < -\sqrt{3}, \quad i = 1, ..., n. \tag{19}$$

When $a > 0$ condition (17) is instead $0 < a/b_i^{\frac{1}{2}} < \sqrt{3}$, $i = 1, ..., n$. although note that when $a > 0$ (18) expresses the case where the tail probability of

the average model is smaller than for the model with the average parameter $\sum_{i=1}^{n} \lambda_i b_i$.
(ii) When $f(u)$ is the standardized Student t with $\nu > 2$ degrees of freedom, for $a < 0$ the same condition (19) applies, independently from ν.

7 An Empirical Application

7.1 Data and Some Preliminary Analysis

The model averaging and the associated VaR evaluation tests developed in this paper can be applied to a variety of problems in finance. Here we shall consider the daily VaR of portfolios constructed from 22 main industry indices of the Standard & Poor's 500. The source of our data is Datastream, which provides twenty four S&P 500 industry price indices according to the Global Industry Classification Standard. To ensure a sufficiently long span of daily prices we have excluded the 'Semiconductors & Semiconductor Equipment' and 'Real Estates' from our analysis. The list of the $N = 22$ industries included in our analysis is given in the Note to Table 1. Our data set covers the industry indices from 2nd January 1995 to 13th October 2003 ($T = 2,291$ observation). Daily returns are computed as $r_{jt} = 100 \ln(P_{jt}/P_{jt-1})$, $j = 1, ..., 22$, where P_{jt} is the j^{th} price index. The realized returns $\mathbf{r}_t = (r_{1t}, r_{2t}, ..., r_{22,t})'$ exhibit all the familiar stylized features over our sample period. See Table 1. They are highly cross-correlated, with an average pair-wise cross-correlation coefficient of 0.5. A standard factor analysis yields that the two largest estimated eigenvalues are equal to 11.5 and 1.7, with the remaining being all smaller than unity. The unconditional daily volatility differs significantly across industries and lie in range of 1.13% (Food, Beverage & Tobacco) to 2.39% (Technology Hardware & Equipment). The first-order autocorrelation coefficients of the individual returns are quantitatively very small (ranging from -0.049 to 0.054) and are statistically significant only in the case of four out of the twenty two industries (Automobiles & Components, Health Care Equipment & Services, Diversified Financial, and Utilities). We decided not to filter out any serial correlation in the data since this would have probably induced a sizeable amount of noise, which could be more harmful than the small amount of serial correlation present in the case of four of the assets. We derived nonparametric estimates of the density functions for the standardized returns, confirming that the marginal distributions tend to be symmetric and slightly fat-tailed.

Estimates of univariate GARCH$(1, 1)$ models for the returns, not reported but available on request, also provide some support in favor of a Student t distribution with a low degree of freedom for the conditional distribution of the individual asset returns. The degrees of freedom estimated for the different assets lie in the narrow range of 5.2 to 11.7, with an average estimate of 7.3 and a mid-point value of 8.5. These results provide some support for our working assumption of zero conditional mean returns, and highlight the non-Gaussian and the highly cross correlated nature of the asset returns by industries. Estimation of multivariate volatility models with non-Gaussian distributions present considerable technical difficulties and are unlikely to significantly affect the QMLE estimates, computed assuming Gaussian errors. For risk management purposes, it seems justified to combine the QMLE estimates with multivariate Student t distributions with low degrees of freedom. Based on the univariate estimates, 6 and 8 degrees of freedom seem sensible choices and will be considered below.

7.2 Recursive Estimation of Multivariate Volatility Models

For each of the ten types of multivariate volatility models listed in Section 3, 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 $r = s = 1$. In the case of the factor models we considered only one factor.

All models were estimated recursively using an expanding window starting with 1784 observations as the first estimation sample, with the parameter values (when applicable) updated at monthly intervals. Clearly, the parameters of the volatility models could also have been updated daily. The monthly updates of the parameters can be viewed as a plausible and practical solution to a highly computer intensive problem.[10] Therefore, the models were estimated twenty-four times over the evaluation sample.

Since for certain values of p, q the estimation algorithm did not converge for all models and all data periods, we ended up with $m = 63$ different (nested and non-nested) models with convergent estimates. However, for some models the algorithm converged except for a few isolated time periods. In such cases the estimation results for the model in question was ignored by assigning a zero weight to it in the model averaging procedure for the non-convergent periods.[11]

The different volatility models were then evaluated over the last two years of data (from November 2, 2001 to October 13, 2003, inclusive), with $T_1 = 507$, using one-day ahead forecasts of $\mathbf{\Sigma}_t$ under M_i, denoted by $\hat{\mathbf{\Sigma}}_{it}$, $i = 1, ..., 63$,

[10] We also carried out a straightforward cross-validation test where all models (when relevant) were estimated once using the first $T_0 = 1784$ observations and then evaluated using the last $T_1 = 507$ observations. Perhaps not surprisingly, the results were generally less satisfactory than those based on the recursively computed parameter updates. These pure cross-validation results are available from the authors on request.

[11] All the computations have been carried out in MatLab and the codes are available upon request. For estimation of CCC, DCC and O-GARCH we used the UCSD_GARCH Toolbox developed by Sheppard (2002). All other codes are our own.

and for a given choice of the distribution of the $\mathbf{z}_{it}$ (which we took as Gaussian or Student t with 6 and 8 degrees of freedom). The evaluations were carried out with respect to an equal-weighted portfolio yielding the portfolio return $\rho_t = \boldsymbol{\omega}'_{t-1}\mathbf{r}_t = \bar{r}_t$. We also considered other portfolio weights, including portfolios with time varying weights, and obtained very similar results. To save space, however, we shall only report the results for the equal-weighted portfolio[12].

7.3 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 25% and 50% 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 63 specifications. See also Section 4.

When considering normal innovations, AIC selects the ADCC$(1,2,1,1)$ throughout the evaluation period whereas SBC first selects the ADCC$(1,1,1,1)$, then switches to ADCC$(1,2,1,1)$ from the middle of the sample onwards. In the case of models under multivariate Student t with 6 and 8 degrees of freedom, AIC selected the O-GARCH$(2,2)$ in the first three weeks of the evaluation sample, switching to DCC$(1,2,1,1)$ up to the middle of the sample, with ADCC$(1,2,1,1)$ being selected thereafter. Similar results were also obtained with SBC. With few exceptions, the DCC type models tended to dominate the remaining specifications. This outcome is particularly interesting since the evaluation sample includes the recent periods of large stock market falls and contrast the outcome of recursive modelling applied to S&P mean returns reported in Pesaran and Timmermann (1995) where the best model selected for the monthly excess returns tend to change quite frequently over time. This could be due to the relative stability of volatility models as compared to models of mean returns that are known to be subject to structural breaks.

To provide some idea of the extent to which the DCC type models dominate other specifications, in Table 2 we summarize selected values of the AIC-penalized log-likelihood values, $AIC_{i,t-1}$, for all the 63 models computed using a multivariate Student t distribution with 8 degrees of freedom. As can be seen the DCC (and CCC) type models systematically fit the data better than the other models, and the differences in the AIC-penalized log-likelihood values for

[12]It is also possible to use model-specific portfolio weights, $\boldsymbol{\omega}_{i,t-1}$, where the weights are determined recursively by a suitable expected utility maximization subject to the VaR constraint. Such an exercise would also involve modelling of the conditional mean returns, which has not been addressed in this paper. See Pesaran and Timmermann (2005) for a discussion of such an estimation strategy in real time.

the DCC and other models are sufficiently large for the model weights, $\lambda_{i,t-1}$, of the DCC models, to take the extreme value of unity for most periods.[13] It is also interesting to note that on average 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 or Factor GARCH models. Similar conclusions are also reached if one uses the Gaussian innovations or the SBC criteria.

7.4 VaR Diagnostic Test Results

For the individual and average models we recursively computed the VaR thresholds, $\bar{\rho}_{t-1}(\boldsymbol{\omega}_{t-1},\alpha)$ for the equal weight $\boldsymbol{\omega}_{t-1} = (1/22, 1/22, ..., 1/22)'$, assuming Gaussian and Student t distributed devolatized returns with 6 and 8 degrees of freedom and two different values of α, namely $\alpha = 1\%$ and $\alpha = 5\%$. Using these estimates we then computed, $\hat{\pi}$, the percentage of times that the VaR constraint were violated, and hence the VaR diagnostic statistic, $z_{\hat{\pi}}$, defined by (11). The results are summarized in Tables 3 for $\alpha = 1\%$ (Panel A) and $\alpha = 5\%$ (Panel B), respectively. In view of the model selection results discussed above the test outcomes are very similar, and in many instances are identical for the AIC and SBC selection criteria. In contrast, choice of the distribution of the devolatized returns appears important. For example, the 'best' modelling strategy is rejected by the VaR test when the underlying distribution is assumed to be Gaussian but not if the Student t is used.

Also in the present application 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 most 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.

Comparing across strategies, the best outcome is found with respect to the thick modelling strategy when averaging across the best 15 models (top 25 percentile). The test results are quite robust with respect to the choice of the conditional distribution of the innovations, although they deteriorate as we move from the normal distribution towards Student t with 6 degrees of freedom. This is in line with the theoretical result discussed in Section 6, 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. In cases where the underlying models are already fat tailed, the model averaging (without any single model dominating) can induce an excessive degree of fat-tailness.

[13] 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.

As can be seen from the results in Tables 3, this tendency is accentuated as the coverage of 'thick' modelling is increased, and is most acute when all the 63 models are included.

7.5 Statistical Diagnostic Test Results

The different 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, $\rho_t = \boldsymbol{\omega}'_{t-1}\mathbf{r}_t$, 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 now 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 the ordered values of the $\hat{v}_{i\tau+1}, ..., \hat{v}_{i\tau+T_1}$. The Kuiper test has the added advantage of placing greater emphasis on the tail behavior of the distribution.

Table 4 reports the p-values of these tests, computed using the analytic approximations provided in Stephens (1970), for the three modelling strategies. The test results for the 'best' and the 'average' modelling strategies are identical, for the same reasons as noted above, and indicate a mild rejection (at 7 to 9 per cent levels) of the models with Gaussian de-volatilized returns if the Kuiper test is used. However, the test results strongly favor the Student t distribution with 8 degrees of freedom, in particular. None of the specifications are rejected by the KS test. The Student t distribution is favored when considering the 'thick' modelling approach which includes the best 15 models but tend to be rejected when the average include a larger number of models. The opposite is observed with respect to the normal distribution.

Overall, the statistical tests support the main conclusions reached using the VaR based diagnostics, although they appear to be less informative and less clear cut as far as the tail properties of the portfolio return distributions are concerned.

8 Summary and Conclusions

The 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 $T_1/T_0 + 1/T_1 \to 0$, a condition also encountered in forecast evaluation literature that uses root mean square error as evaluation criteria, as discussed in West (1996). The proposed VaR test is also 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 Asymmetric Dynamic Conditional Correlation (ADCC) model of Cappiello, Engle and Sheppard (2002). In the out of sample evaluation tests, only the multivariate Student t version of the ADCC model with 8 degrees of freedom 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 remaining estimated models, namely O-GARCH and Factor GARCH specifications.

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 that other forms of multivariate volatility models are also considered, which could be the subject of future research.

Appendix: Mathematical Proofs

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 $n \geq 1$,

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

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 $E\left(d_{it_1}d_{it_2}...d_{it_n}\right)$ in (A.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 $t_1,...,t_n$. 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 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 (10) holds. Set $q_{it} = q_{it}(\hat{\boldsymbol{\theta}}_{iT_0}, \boldsymbol{\theta}_{i0}) = (\hat{\sigma}_{\rho t}(M_i)/\sigma_{\rho t}(M_i)) = (\boldsymbol{\omega}_{t-1}'\hat{\boldsymbol{\Sigma}}_{it}\,\boldsymbol{\omega}_{t-1}/\boldsymbol{\omega}_{t-1}'\boldsymbol{\Sigma}_{it}\,\boldsymbol{\omega}_{t-1})^{1/2}$. Then $E[d_{it}(\hat{\boldsymbol{\theta}}_{iT_0})|\mathcal{F}_{t-1}, M_i] = F_{it}(-c_{it}(\alpha)q_{it})$ 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 $T_0 \to \infty$, $\hat{\boldsymbol{\theta}}_{iT_0} \xrightarrow{p} \boldsymbol{\theta}_{i0}$ and since $\boldsymbol{\Sigma}_{it}(\boldsymbol{\theta}_i)$ is a continuous function of $\boldsymbol{\theta}_i$ it also follows that $q_{it}(\hat{\boldsymbol{\theta}}_{iT_0}, \boldsymbol{\theta}_{i0}) \xrightarrow{p} 1$, 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

$$\sqrt{T_1}(\hat{\pi}_i - \alpha) = \sqrt{T_1}(\pi_i - \alpha) + \sqrt{T_1}(\hat{\pi}_i - \pi_i), \quad \text{(A.2)}$$

where $\pi_i = \frac{1}{T_1}\sum_{t\in\mathcal{T}_1} d_{it}(\boldsymbol{\theta}_{i0})$. 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 $X_{it,T_0} = \sqrt{T_0}[d_{it}(\hat{\boldsymbol{\theta}}_{iT_0}) - d_{it}(\boldsymbol{\theta}_{i0})]$. 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 $t \in \mathcal{T}_1$, $\Pr\left(|X_{it,T_0}| = \sqrt{T_0}\,|\mathcal{F}_{t-1}, M_i\right) \leq |F_{it}(-c_{it}(\alpha)q_{it}(\hat{\boldsymbol{\theta}}_{iT_0}, \boldsymbol{\theta}_{i0})) - F_{it}(-c_{it}(\alpha))|$, and consequently $E(|X_{it,T_0}| \ \ |\mathcal{F}_{t-1}, M_i) \leq \sqrt{T_0}|F_{it}(-c_{it}(\alpha)q_{it}(\hat{\boldsymbol{\theta}}_{iT_0}, \boldsymbol{\theta}_{i0})) - F_{it}(-c_{it}(\alpha))|$. Using the mean-value expansion of $F_{it}(-c_{it}(\alpha)q_{it}(\hat{\boldsymbol{\theta}}_{iT_0}, \boldsymbol{\theta}_{i0}))$ around $\hat{\boldsymbol{\theta}}_{iT_0}$ one gets $F_{it}(-c_{it}(\alpha)q_{it}(\hat{\boldsymbol{\theta}}_{iT_0}, \boldsymbol{\theta}_{i0})) = F_{it}(-c_{it}(\alpha)) - c_{it}(\alpha)f_{it}(-c_{it}(\alpha)q_{it}(\bar{\boldsymbol{\theta}}_i, \boldsymbol{\theta}_{i0}))\partial q_{it}(\bar{\boldsymbol{\theta}}_i, \boldsymbol{\theta}_{i0})/\partial\hat{\boldsymbol{\theta}}_{iT_0}'(\hat{\boldsymbol{\theta}}_{iT_0} - \boldsymbol{\theta}_{i0})$, where the elements of $\bar{\boldsymbol{\theta}}_i$ are convex combinations of the corresponding elements of $\hat{\boldsymbol{\theta}}_{iT_0}$ and $\boldsymbol{\theta}_{i0}$. By the Holder's inequality for norm of matrices, since $\parallel \boldsymbol{\omega}_t \parallel > 0$, we have $E(|X_{it,T_0}| \ \ |\mathcal{F}_{t-1}, M_i)$ $\leq c_{it}(\alpha)f_{it}(-c_{it}(\alpha)q_{it}(\bar{\boldsymbol{\theta}}_i, \boldsymbol{\theta}_{i0})) \parallel \partial q_{it}(\bar{\boldsymbol{\theta}}_i,\ \boldsymbol{\theta}_{i0})/\partial\hat{\boldsymbol{\theta}}_{iT_0} \parallel \sqrt{T_0} \parallel \hat{\boldsymbol{\theta}}_{iT_0} - \boldsymbol{\theta}_{i0} \parallel \leq c_{it}(\alpha)f_{it}(-c_{it}(\alpha)q_{it}(\bar{\boldsymbol{\theta}}_i, \boldsymbol{\theta}_{i0}))$ $\{\sup_{\theta\in\Theta_i} \parallel \partial\bar{\lambda}_{it}(\boldsymbol{\theta})/\partial\boldsymbol{\theta} \parallel /\underline{\lambda}_{it}^{\frac{1}{2}}(\theta)\underline{\lambda}_{it}^{\frac{1}{2}}(\theta_0)\}\sqrt{T_0} \parallel \hat{\boldsymbol{\theta}}_{iT_0} - \boldsymbol{\theta}_{i0} \parallel$. Taking the unconditional mean and using the Holder inequality again yields $E(|X_{it,T_0}||M_i)$ $\leq c_{it}(\alpha)\sup_x f_{it}(x)(E|\sup_{\theta\in\Theta_i} \parallel \partial\bar{\lambda}_{it}(\boldsymbol{\theta})/\partial\boldsymbol{\theta} \parallel /\underline{\lambda}_{it}^{\frac{1}{2}}(\theta)\underline{\lambda}_{it}^{\frac{1}{2}}(\theta_0)|^{\delta})^{\frac{1}{\delta}} \sqrt{T_0}(E\parallel \hat{\boldsymbol{\theta}}_{iT_0} - \boldsymbol{\theta}_{i0} \parallel^{\frac{\delta}{\delta-1}})^{1-1/\delta}$. Therefore, $T_1^{-1}\sum_{t\in\mathcal{T}_1} X_{it,T_0} = O_p(1)$ and the second term in (A.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 (18) 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 $g(x) \equiv F(a/\sqrt{x})$. Jensens'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 (17).■

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.

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.

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 (2002): "Asymmetric dynamics in the conditional correlations of global equity and bond returns," *NYU* , Preprint.

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., 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.

Diebold, F., and M. Pesaran (1999): "The latent - factor GARCH model of asset return volatility: estimation using a market index," *University of Cambridge and University of Pennsylvania* , Mimeo.

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* , forthcoming.

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.

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. (1999): "On the probabilistic properties of a double threshold ARMA conditional heteroskedasticity model," *Journal of Applied Probability*, 36, 688–705.

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.

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.

PESARAN, M., AND A. TIMMERMANN (1995): "Predictability of stock returns: robustness and economic significance," *Journal of Finance*, 50, 1201–1228.

PESARAN, M., AND A. TIMMERMANN (2005): "Real time econometrics," Econometric Theory, 21, 213–231.

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.

SHEPPARD, K. (2002): "UCSD_GARCH Toolbox, Version 2.0.4 19-JUN-2002," .

STEPHENS, M. (1970): "Use of Kolmogorov-Smirnov, Cramer-Von Mises and related statistics without extensive tables," *Journal of the Royal Statistical Society, Series B (Methodological)*, 32, 115–122.

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.

Table 1: Summary statistics

Sector	Mean	St.Dev.	Skewness	Kurtosis	Ljung-Box(20)
EN	0.031	1.386	0.049	5.435	40.3
MA	0.015	1.367	0.141	6.347	22.1
IC	0.040	1.395	-0.156	6.784	33.1
CS	0.022	1.318	-0.466	8.777	18.7
TRN	0.027	1.407	-0.501	10.644	28.4
AU	0.011	1.628	-0.172	7.017	36.3
LP	0.015	1.194	-0.099	6.750	25.1
HR	0.034	1.422	-0.393	9.241	16.4
ME	0.030	1.660	-0.056	8.168	37.2
MS	0.057	1.739	0.017	6.120	48.5
FD	0.028	1.328	-0.217	6.597	30.8
FBT	0.032	1.132	0.008	6.312	32.4
HHPE	0.042	1.445	-1.581	30.256	55.1
HC	0.039	1.274	-0.295	7.008	57.4
PHB	0.054	1.472	-0.172	5.821	53.1
BK	0.051	1.590	0.045	5.324	37.0
DF	0.075	1.840	0.036	5.013	47.4
INSC	0.044	1.549	0.415	11.045	38.8
IS	0.062	2.246	0.060	5.019	32.6
TEHW	0.043	2.393	0.165	5.719	30.6
TS	0.000	1.605	-0.072	5.969	22.7
UL	0.005	1.197	-0.363	9.881	25.8

Note: Columns 2 to 4 report the sample mean, standard deviation, skewness and kurtosis. Column 5 reports the Ljung-Box statistic of order 20 for testing autocorrelations in individual asset returns. The critical value of χ^2_{20} at the 1 % significance level is 37.56. The sample period is January 2, 1995 to October 13, 2003. Column 1 gives the Industry codes taken from REUTERS for the S&P 500 industry groups according to the Global Industry Classification Standard: EN for Energy, MA for Materials, IC for Capital Goods, CS for Commercial Services & Supplies,TRN for Transportation, AU for Automobiles & Components,LP for Consumer Durables & Apparel, HR for Hotels, Restaurants & Leisure, ME for Media, MS for Retailing, FD for Food & Staples Retailing, FBT for Food, Beverage & Tobacco, HHPE for Household& Personal Products, HC for Health Care , Equipment & Services, PHB for Pharmaceuticals & Biotechnology, BK for Banks, DF for Diversified Financials, INSC for Insurance, IS for Software & Services, TEHW for Technology Hardware & Equipment, TS for Telecommunication Services, UL for Utilities.
Source: *Datastream.*

Table 2: AIC-penalized Quasi log-likelihood values

Model Type		2 Nov'01	13 Oct'03	Average
EQMA	(n_0)			
	(50)	-52921	-69477	-61666
	(75)	-50167	-65931	-58500
	(125)	-48815	-64185	-56943
	(250)	-48096	-63437	-56182
EWMA	(λ_0, ν_0)			
	(0.96,0.94)	-53681	-70350	-62493
	(0.96,0.80)	-87212	-114517	-101336
	(0.96,0.60)	-167254	-220654	-194364
	(0.95,0.94)	-53696	-70367	-62505
	(0.95,0.80)	-87180	-114515	-101309
	(0.95,0.60)	-167207	-220659	-194334
	(0.94,0.94)	-53726	-70406	-62537
	(0.94,0.80)	-87142	-114495	-101274
	(0.94,0.60)	-167143	-220626	-194279
	(0.95,0.95)	-52279	-68510	-60868
	(0.96,0.96)	-50970	-66802	-59361
MMA	(n_0, ν_0)			
	(50,0.60)	-169511	-221486	-195659
	(75,0.60)	-169503	-221370	-195615
	(125,0.60)	-169533	-221210	-195607
	(250,0.60)	-169754	-221114	-195832
	(50,0.80)	-90156	-116706	-103702
	(75,0.80)	-90089	-116523	-103586
	(125,0.80)	-90130	-116411	-103591
	(250,0.60)	-90317	-116311	-103764
	(50,0.94)	-58962	-76663	-68266
	(75,0.94)	-57016	-74134	-66038
	(125,0.94)	-56667	-73624	-65604
	(250,0.94)	-58962	-76663	-68266
Gen. EWMA	(p, q, ν_0)			
	(2,2,0.94)	-53821	-63768	-59212
	(2,2,0.80)	-86819	-94612	-91299
	(2,2,0.60)	-167006	-202297	-185255
	(1,2,0.94)	-53767	-63550	-59072
	(1,2,0.80)	-86705	-93863	-90911
	(1,2,0.60)	-166798	-201250	-184666
	(2,1,0.94)	-53751	-63638	-59108
	(2,1,0.80)	-86708	-94096	-91028
	(2,1,0.60)	-166863	-201696	-184932
	(1,1,0.94)	-53773	-63607	-59102
	(1,1,0.80)	-86782	-94010	-91036
	(1,1,0.60)	-166953	-201546	-184936

Model Type		2 Nov'01	13 Oct'03	Average
DCC	(p, q, r, s)			
	(1,1,1,1)	-47851	-61964	-55310
	(2,1,1,1)	-47849	-61998	-55326
	(1,2,1,1)	-47843	-61895	-55272
	(2,2,1,1)	-47876	-62044	-55371
ADCC	(p, q, r, s)			
	(1,1,1,1)	-48126	-61856	-55356
	(2,1,1,1)	-48199	-62141	-55521
	(1,2,1,1)	-48099	-61737	-55282
	(2,2,1,1)	-48194	-62018	-55462
CCC	(p, q)			
	(1,1)	-48299	-62847	-55980
	(2,1)	-48298	-62881	-55997
	(1,2)	-48293	-62780	-55944
	(2,2)	-48069	-62698	-55785
O-GARCH	(p, q, r, s)			
	(1,1,1,1)	-47899	-66423	-57518
	(2,1,1,1)	-47926	-66529	-57670
	(1,2,1,1)	-47909	-66818	-57588
	(2,2,1,1)	-47703	-66180	-57271
Factor HRS	(p, q, r, s)			
	(1,1,1,1)	-50625	-66506	-59004
	(2,1,1,1)	-50119	-65747	-58411
	(1,2,1,1)	-50102	-65698	-58382
	(2,2,1,1)	-50116	-65733	-58403
Factor DP	(p, q, r, s)			
	(1,1,1,1)	-50180	-65722	-58445
	(2,1,1,1)	-50207	-65759	-58479
	(1,2,1,1)	-50183	-65726	-58449
	(2,2,1,1)	-50190	-65735	-58457

Note: The figures report the maximized values of the Student t (8) log likelihoods, penalized by the AIC criterion: $AIC_{i,t-1} = LL_{i,t-1} - k_i$, where $LL_{t-1,i}$ is the maximized log likelihood at time t for model i and k_i is the number of estimated parameters of model i.
Columns 2 and 6 report the AIC-penalized log likelihood at the initial date of the evaluation period (2 Nov '01), columns 3 and 7 report the AIC-penalized log likelihood at the final date of the evaluation period (13 Oct '03), and columns 4 and 8 report the average AIC-penalized log likelihood values over the days between these two dates.For a brief description of the models see Section 3.

Table 3: VaR Diagnostic Tests

Recursive estimation with expanding window for an equal-weighted portfolio

Panel A: $\alpha = 1\%$

	Normal		Student (8)		Student (6)	
	$\hat{\pi}$	$z_{\hat{\pi}}$	$\hat{\pi}$	$z_{\hat{\pi}}$	$\hat{\pi}$	$z_{\hat{\pi}}$
Modelling Strategies						
Best						
AIC	1.972	2.200	0.592	-0.924	0.394	-1.370
SBC	1.972	2.200	0.592	-0.924	0.197	-1.817
'Bayesian' Average						
AIC	1.972	2.200	0.592	-0.924	0.394	-1.370
SBC	1.972	2.200	0.592	-0.924	0.197	-1.817
Thick Average						
AIC best 15 (25%)	0.394	-1.370	0.197	-1.817	0	-2.263
SBC best 15 (25%)	0.394	-1.370	0.197	-1.817	0	-2.263
AIC best 32 (50%)	0.592	-0.924	0.197	-1.817	0	-2.263
SBC best 32 (50%)	0.592	-0.924	0.197	-1.817	0	-2.263
All (100%)	0.197	-1.817	0	-2.263	0	-2.263

Panel B: $\alpha = 5\%$

	Normal		Student (8)		Student (6)	
	$\hat{\pi}$	$z_{\hat{\pi}}$	$\hat{\pi}$	$z_{\hat{\pi}}$	$\hat{\pi}$	$z_{\hat{\pi}}$
Modelling Strategies						
Best Models						
AIC	7.100	2.170	4.733	-0.275	3.550	-1.497
SBC	6.903	1.966	4.733	-0.275	3.353	-1.701
'Bayesian' Average Models						
AIC	7.100	2.170	4.733	-0.275	3.550	-1.497
SBC	6.903	1.966	4.733	-0.275	3.353	-1.701
Thick Average Models						
AIC best 15 (25%)	4.536	-0.478	4.142	-0.886	3.155	-1.905
SBC best 15 (25%)	4.536	-0.478	4.142	-0.886	2.958	-2.109
AIC best 32 (50%)	4.733	-0.275	2.761	-2.312	2.169	-2.924
SBC best 32 (50%)	4.733	-0.275	2.761	-2.312	2.169	-2.924
All (100%)	3.155	-1.905	1.577	-3.535	1.577	-3.535

Table 4: Probability Values for Kupier and Kolmogorov-Smirnov Tests
(Recursive wstimation with expanding window applied to an equal-weighted portfolio)

	Normal		Student (8)		Student (6)	
Modelling Strategy	Ku	KS	Ku	KS	Ku	KS
Best						
AIC	0.091	0.158	0.485	0.303	0.179	0.174
SBC	0.091	0.181	0.362	0.247	0.292	0.279
'Bayesian' Average						
AIC	0.091	0.158	0.485	0.303	0.179	0.174
SBC	0.091	0.181	0.362	0.247	0.292	0.279
Thick Average						
AIC (25%)	0.003	0.004	0.154	0.187	0.052	0.107
SBC (25%)	0.003	0.004	0.148	0.163	0.025	0.080
AIC (50%)	0.143	0.066	0.005	0.003	0.000	0.016
SBC (50%)	0.143	0.066	0.006	0.003	0.000	0.016
All (100%)	0.016	0.049	0.000	0.002	0.000	0.014

Note: The columns indicated by KS report the p-values of the Kolmogorov-Smirnov test and the columns indicated by Ku reports the p-values of the Kupier test (cf. Section 7.5).