Value at Risk and Expected Shortfall: A Forecast Combination Approach

Elena Andreou
Department of Economics
University of Cyprus*

Constantinos Kourouyiannis†
Department of Economics
University of Cyprus‡

Andros Kourtellos
Department of Economics
University of Cyprus§

January 2, 2010

Abstract

The recent financial crisis that started in 2007 provides a challenge for improving or proposing new methods in risk management. We study two measures of market risk, Value at Risk and Expected Shortfall and show that various forecast combination methods provide accurate forecasts. These methods address the model uncertainty due to alternative volatility model specifications and estimators, distributions of standardized returns and risk management methods. Our model space includes models from three alternative methods, parametric, filtered historical simulation and extreme value theory. We compare the out of sample performance of individual models and forecast combination methods using international stock market indices. Our empirical results show that individual models suffer from the problem of model risk and that forecast combinations can provide more accurate predictions of risk.

Keywords: Expected shortfall, forecast combinations, value at risk, volatility
JEL Classifications: C53, C52, C22

---

*P.O. Box 537, CY 1678, Nicosia, Cyprus, e-mail: elena.andreou@ucy.ac.cy
†Acknowledgements: We would like to thank the participants of the Computational and Financial Econometrics (CFE'09) conference for helpful comments. We would like to thank Valentina Corradi for important remarks. We are also greatful to the Cyprus Research Promotion Foundation and the School of Economics and Management of the University of Cyprus for financial support. The first author also acknowledges support of the European Research Council under the European Community FP7/2007-2013 ERC grant 209116.
‡Corresponding author: P.O. Box 537, CY 1678, Nicosia, Cyprus, e-mail:c.kourouyiannis@ucy.ac.cy
§P.O. Box 537, CY 1678, Nicosia, Cyprus, e-mail: andros@ucy.ac.cy.

1 Introduction

Financial disasters such as the UK market crash in 1987, the recent financial crisis that started in 2007, as well as the fallout of large financial institutions such as the Long Term Capital Management (LTCM) and Lehman Brothers, indicate the need to improve the existing risk management methods. Value at Risk (VaR) has become a benchmark in measuring financial risk and has been widely used by academics, practitioners and regulators (see for instance Jorion (2000) and Duffie and Pan (1997)). Given the importance of VaR as a risk measure, many approaches have been proposed for its estimation including parametric methods, semi-parametric methods (e.g. Filtered Historical Simulation and Extreme Value Theory) and non-parametric (Historical simulation). Kuester et al. (2006) provide a comparison of the out of sample performance of a large number of VaR models. VaR has nevertheless some important limitations, since it is not a coherent risk measure (due to the violation of the subadditivity property) and does not give any information about the potential losses beyond VaR. Artzner et al. (1999) propose Expected Shortfall (ES) as an alternative measure of risk, which overcomes these problems. In this paper we consider a forecast combination approach to both VaR and ES.

Despite the increasing popularity of VaR, the recent financial crisis indicated that even sophisticated VaR models cannot provide accurate forecasts, especially in periods where extreme events occur frequently. One reason for this is that models often suffer from various sources of misspecification, since they sometimes impose wrong assumptions related to the volatility or distribution and therefore their forecasting performance may vary across assets and periods. The objective of this paper is to use forecast combinations to provide VaR and ES forecasts that are robust to these sources of misspecification. Timmermann (2006) underlines the benefits of forecast combinations, which they take into account information from each model's forecast, they are robust to misspecification bias and measurement error of individual forecasts and provide diversification gains.

Forecast combinations have been used successfully in other areas of research, such as forecasting Real GDP (Stock and Watson, 2004), inflation (Stock and Watson, 2008), exchange rates (Wright, 2008) and stock returns (Avramov, 2002). Forecast Combinations have also been used in forecasting variables that are unobserved, such as volatility. For instance, Pesaran et al. (2008) use "thick" and Bayesian model averaging in the context of multi-asset volatility models and evaluate their forecasting accuracy based on a VaR diagnostic test. In addition, Liu and Maheu (2009) use Bayesian model averaging to forecast realized volatility and Patton (2009) combines realized volatility estimators and uses MSE and QLIKE distance measures for their evaluation. Forecast Combinations have also been used for testing conditional quantiles (Giacomini and Komunjer, 2005). However, there is no work that uses Forecast Combinations to directly predict VaR and ES.

Our methodology takes into account other sources of model uncertainty beyond volatility, such as the distribution of standardized returns and the risk management methods.

In this paper we use Forecast Combinations that have been proposed in the literature, such as Weighted BIC, Smoothed AIC, Bates-Granger, Granger-Ramanathan and Mallows Model Averaging to predict VaR. Using major stock market indices, such as S&P 500, NASDAQ Composite, DAX 30, FTSE 100, CAC 40 and Nikkei 225, we find that forecast combinations can give more reliable forecasts of VaR than individual models. We consider models from three broad categories, parametric, filtered historical simulation and extreme value theory and compare their forecasting performance using the Conditional Coverage test (Christoffersen, 2003) and Dynamic Quantile test (Engle and Manganelli, 2004). We find that parametric models based on the t-distribution perform best, but still they fail to give accurate forecasts when we use the FTSE 100 index. Forecast combinations perform well across all stock market indices and even at high confidence levels, where predictability of VaR becomes extremely difficult. We extend our analysis in the context of ES and find that forecast combinations outperform individual models. For the evaluation of ES forecasts we use alternative loss functions, such as MSE, QLIKE and LINEX and a more general criterion, maximum regret.

The paper is organized as follows. In Section 2 we discuss the methodology. In Section 3 we present the methods that we use for the evaluation of the performance of individual models and forecast combinations. In Section 4 we discuss our empirical findings using international stock market indices and in Section 5 we give our conclusions and future research.

2 Methodology

2.1 *Ex-post* vs *ex-ante* VaR

VaR at a confidence level α, is defined as the smallest number l, such as the probability that the loss L exceeds l is no larger than $1-\alpha$, $q^{\alpha}=\inf\{l \in \mathbb{R} : P(L>l) \leq 1-\alpha\}$(McNeil et al., 2005). We classify VaR in two categories: (a) *Ex-post* VaR, which uses information until time $t+1$ and realized volatility measures to give a proxy of VaR at time $t+1$ and (b) *Ex-ante* VaR, which uses information until time t and GARCH volatility models to give a forecast of VaR at time $t+1$.

Let p_t denote the logarithmic price of an asset. Then daily returns are given by $r_t = p_t - p_{t-1}$.We assume zero conditional mean of the return process : $\mu_{t/t-1} = E\left[r_t/\Im_{t-1}\right] = 0$, where $\Im_{t-1}$ reflects all relevant information through time $t-1$. This assumption is accepted empirically, and it is also consistent with the martingale difference assumption of weak Efficient Market Hypothesis. We use high frequency volatility measures to construct VaR proxies

$$q_t^a = \sigma_t F_v^{-1}(a) \tag{2.1}$$

where σ_t is an *ex-post* high frequency volatility measure and F_v is the empirical distribution of normalized returns $v_t = \frac{r_t}{\sigma_t}$.

We consider three realized volatility measures, Realized Volatility (RV), Realized Power Variation (RPV) and Realized Bipower Variation (RBP) . To define these measures, we normalize the daily time interval to unity and divide it into m periods. The length of each period is equal to $\Delta = \frac{1}{m}$ and the high-frequency log-returns or Δ period returns are given by $r_{t,j} = p_{t+j\Delta} - p_{t+(j-1)\Delta}$, $j = 1, ..., m$. RV is given by the sum of squared high-frequency log-returns and it is a consistent estimator of Quadratic Variation (QV). Under the assumption of no jumps in the price process it is also a consistent estimator of Integrated Volatility.

$$RV_{t+1} = \sum_{j=1}^{m} r_{t,j}^2 \xrightarrow{p} QV_{t+1} = \int_t^{t+1} \sigma_s^2 ds \ + \sum_{t<s\leq t+1,\ dq_s=1} \kappa_s^2 \tag{2.2}$$

RV is discussed extensively in Andersen et al. (2001a,b) and Barndoff-Nielsen and Shephard (2002a,b). Bardoff-Nielsen and Shephard (2004) proposed two other measures of volatility, RPV and RBP, which are robust to jumps and are consistent estimators of Integrated Power Variation (IPV).

$$RPV_{t+1} = \mu_1^{-1} \Delta^{1-p/2} \sum_{j=1}^{m} |r_{t,j}| \xrightarrow{p} IPV_{t+1}(1) = \int_t^{t+1} \sigma_s ds \tag{2.3}$$

$$RBP_{t+1} = \mu_1^{-2} \sum_{j=2}^{m} |r_{t,j-1}| \, |r_{t,j}| \xrightarrow{p} IPV_{t+1}(2) = \int_t^{t+1} \sigma_s^2 ds \tag{2.4}$$

In the class of *ex-ante* VaR forecasts we consider 24 different model specifications using the parametric, filtered historical simulation and extreme value theory methods. We consider 4 volatility models, GARCH, GJR-GARCH or Threshold - GARCH (TARCH), EGARCH and APARCH and 2 distributions the normal and *t*. The *ex-ante* VaR forecasts used in this paper are described in table 2.

2.2 Forecast Combinations

One of the most important problems in using Forecast Combinations, is that estimating all possible models is time consuming and sometimes not feasible. Some methods that have been proposed in the literature, deal with this problem by choosing the best models using model selection and then combine forecasts (e.g. using the "leaps and bounds" algorithm (Furnival and Wilson, 1974)). However, this method is subject to the pretesting criticism since it separates the model selection procedure from the forecasting procedure and treat the conditional estimates as unconditional. In order to overcome this problem and at the same time save a substantial amount of time we employ the orthogonalization as discussed in Magnus et al. (2009).

We consider the linear regression model:

$$q_t^\alpha = \beta_0 + \beta_1 q_{t-1}^\alpha + \sum_{j=2}^{k+1} \beta_j x_{t,j-1} + \varepsilon_t \tag{2.5}$$

where q_t is *ex-post* VaR, q_{t-1} the first lag of *ex-post* VaR and $x_{t,j-1},\ j = 2, ..., k+1$ the orthogonalized regressors, which span the same space as the *ex-ante* VaR forecasts. Following the terminology of Magnus et al. (2009) q_t and q_{t-1} are called "focus" regressors, since we always want to include them in the model and $x_{t,j-1},\ j = 2, ..., k+1$ are the "auxiliary" regressors, since we are less certain whether they can be useful in predicting the variable of interest.

Following Hansen (2008) we consider forecasting models that are strictly nested. Each model can be written as:

$$q_t^\alpha = \beta_{0,m} + \beta_{1,m} q_{t-1}^\alpha + \sum_{j\in\Omega} \beta_{j,m} x_{t,j-1} + \varepsilon_{t,m} \tag{2.6}$$

where $\mathbf{\Omega} = \{2, ..., k+1\}$. The least-squares forecast of q_t^α given by each model can be written as:

$$\widehat{f}_{T+1,m} = \widehat{\beta}_{0,m} + \widehat{\beta}_{1,m} q_{t-1}^\alpha + \sum_{j\in\Omega} \widehat{\beta}_{j,m} x_{t,j-1} \tag{2.7}$$

where $\widehat{\boldsymbol{\beta}}_{j,m}$ are the least-squares estimates of $\boldsymbol{\beta}_{j,m}$ of model m.The residuals of each model are given by $\widehat{e}_{t,m} = q_t^\alpha - \widehat{\beta}_{0,m} - \widehat{\beta}_{1,m} q_{t-1}^\alpha - \sum_{j\in\Omega} \widehat{\beta}_{j,m} x_{t,j-1}$ and the corresponding variance by:

$$\widehat{\sigma}_m^2 = \frac{1}{T} \sum_{t=1}^{T} \widehat{e}_{t,m}^2 \tag{2.8}$$

We can combine these forecasts given by each model by assigning a weight w_m to each forecast

$\widehat{f}_{T+1,m}$. So the forecast of q^{α}_{T+1} by combining individual forecasts is given by:

$$\widehat{\mathbf{f}}^{C}_{T+1}(\mathbf{w}) = \sum_{m=1}^{M} w_m \widehat{f}_{T+1,m} \tag{2.9}$$

For the estimation of the forecast error variance we use recursive estimates of $\boldsymbol{\beta}_m = [\beta_{1,m}, ...\beta_{k+1,m}]$ denoted by $\widehat{\boldsymbol{\beta}}_{t-1,m}$. The recursive least-squares forecasts of q^{α}_t given by each model is $\widehat{f}_{t,m} = \mathbf{x}_{t,m}{}'\widehat{\boldsymbol{\beta}}_{t-1,m}$, where $\mathbf{x}_{t,m} = [1, q^{\alpha}_{t-1}, x_{t,1}, ..., x_{t,k}]$ and the variance of forecast errors:

$$\widetilde{\sigma}^2_m = \frac{1}{k+1} \sum_{t=T-k}^{T} \left(q^{\alpha}_t - \widehat{f}_{t,m}\right)^2 \tag{2.10}$$

where k is chosen so that the variance is well defined[1].

Based on the choice of the weights assigned to each forecast, we consider 5 forecast combination methods, Weighted BIC or WBIC, Smoothed AIC or SAIC (Buckland et al. (1997) and Burnham and Anderson (2002)), Bates - Granger (1969), Granger-Ramanathan (1984) and Mallows Model Averaging or MMA (Hansen, 2008). For the WBIC and SAIC methods the weights are given by

$$WBIC\ and\ SAIC: \ w_m = \exp\left(-\frac{1}{2}C_m\right)\left[\sum_{j=1}^{M} \exp\left(-\frac{1}{2}C_m\right)\right]^{-1} \tag{2.11}$$

where C_m is the BIC and AIC, respectively[2]. The Bates - Granger method uses weights that are inversely proportional to the variance of forecast errors,

$$Bates - Granger: \ \widehat{w}_m = \left(\widetilde{\sigma}^2_m\right)^{-1}\left[\sum_{j=1}^{M} \left(\widetilde{\sigma}^2_j\right)^{-1}\right]^{-1} \tag{2.12}$$

In addition, we consider two versions of the Granger - Ramanathan approach, the unconstrained and constrained, where the weights are determined based on the following regressions:

[1] In this paper we use $k = T/2$.

[2] Note that the WBIC method has a Bayesian interpretation. Under diffused priors the Bayesian Model Averaging (BMA) weights can be approximated by the WBIC weights (2.11).

$$
\begin{aligned}
&Unconstrained\ Granger-Ramanathan: \quad q_t^\alpha = \sum_{j=1}^{M} \widehat{w}_j \widehat{f}_{t,m} + \varepsilon_t \\
&Constrained\ Granger-Ramanathan: \quad q_t^\alpha = \sum_{j=1}^{M} \widehat{w}_j \widehat{f}_{t,m} + \varepsilon_t, \quad s.t. \sum_{j=1}^{M} \widehat{w}_j = 1\ and\ 0 \leq w_j \leq 1
\end{aligned} \tag{2.13}
$$

Finally, we consider Mallows Model Averaging method which chooses optimal weights by minimizing the Mallows criterion:

$$
MMA: \widehat{\mathbf{w}} = \arg\min_{\widehat{\mathbf{w}}} \sum_{t=1}^{T} \left(q_t^\alpha - \widehat{\boldsymbol{\mu}}_t' \widehat{\mathbf{w}}\right)^2 + 2\sum_{m=1}^{M} \widehat{w}_m k_m s^2, \quad s.t. \sum_{j=1}^{M} \widehat{w}_j = 1\ and\ 0 \leq \widehat{w}_j \leq 1 \tag{2.14}
$$

where $\widehat{\boldsymbol{\mu}}_t = \left(\widehat{\mu}_{t,1}, ..., \widehat{\mu}_{t,M}\right)$, $\mu_{t,m} = \widehat{\beta}_{0,m} + \widehat{\beta}_{1,m} q_{t-1}^\alpha + \sum_{j\in\Omega} \widehat{\beta}_{j,m} x_{t,j-1}$ and $s^2 = \frac{1}{T-k_M}\sum_{t=1}^{T} \widehat{e}_{t,M}^2$ an estimate of the variance of the largest fitted model.

2.3 Expected Shortfall

Despite the popularity of the VaR as a risk measure to academics, practitioners and regulators, it is not a coherent risk measure, since it does not satisfy the subadditivity property. Artzner et al. (1999) proposed an alternative risk measure, the Expected Shortfall (ES), which satisfies all the properties of coherent risk measures and additionally incloses information for the potential size of the losses beyond VaR. ES is defined as

$$
m^\alpha = \frac{1}{1-\alpha} \int_\alpha^1 q^u du \tag{2.15}
$$

For the estimation of ES we use the lemma (McNeil et al., 2005, pp45), which says that ES is the mean of the losses given that the losses exceed VaR

$$
m^\alpha = E\left(L/L \geq q^\alpha\right) \tag{2.16}
$$

where L are the losses of a portfolio[3]. Therefore, once we obtain a forecast of VaR, based on an individual model or a forecast combination method, we can easily find the corresponding value of

[3] We use negative returns as the losses of the portfolio.

ES using

$$m_{t+1}^{\alpha} = \frac{\sum_{i=t-k+1}^{t} L_i 1\left\{L_i > q_{t+1}^{\alpha}\right\}}{\sum_{i=t-k+1}^{t} 1\left\{L_i > q_{t+1}^{\alpha}\right\}} \tag{2.17}$$

where k is the size of the rolling window.

3 Evaluation Methods

For the evaluation of the out of sample VaR forecasts we use traditional backtesting methods, such as the Conditional Coverage Test (Christoffersen, 2003) and Dynamic Quantile Test (Engle and Manganelli, 2004). For the evaluation of ES forecasts we consider symmetric (MSE) and asymmetric (QLIKE and LINEX) loss functions. We also use Maximum Regret to evaluate the performance of models and combinations across different confidence levels.

3.1 Backtesting VaR

While there are many tests that examine the accuracy of VaR, the majority of these tests consider the event that the loss on a portfolio exceeds its reported VaR. These tests use the hit sequence, an indicator function, which measures the number of violations of VaR. The hit sequence takes the value one when negative returns exceed VaR and zero otherwise.

$$I_{t+1} = 1\left\{-r_{t+1} > q_{t+1}^{\alpha}\right\} \tag{3.18}$$

Christoffersen (2003) notes that the problem of assessing the accuracy of VaR can be understood by studying two properties of the hit sequence (i) unconditional coverage: the average of the hit sequence (percentage of violations) should be equal to the confidence level and (ii) independence property : the hit sequence should be independent over time (VaR forecasts with clustered violations should be rejected). Although these properties are separate and distinct we employ the Conditional Coverage test, which combines the two properties in the following null:

$$H_0 : I_{t+1} \sim i.i.d.\ Bernoulli\,(\alpha) \tag{3.19}$$

We also use an alternative method of backtesting VaR, the Dynamic Quantile test, which is based on the following regression:

$$I_t - \alpha = \beta_0 + \sum_{i=1}^{p} \beta_i I_{t-i} + \beta_{p+1} \widehat{q}_t^{\alpha} + \varepsilon_t \tag{3.20}$$

The null hypothesis of this test is $H_0: \beta_0 = ... = \beta_{p+1} = 0$, which indicates that there is no explanatory power in the lags of the hit sequence and the VaR forecast. For testing the VaR models and forecast combinations we employ the specification of this test used by Kuester at al. (2006), in which the constant, four lags of the hit sequence and the contemporaneous VaR forecast are used as regressors.

3.2 Loss functions and Maximum Regret

For the evaluation of the ES forecasts given by individual models and forecast combinations we use alternative loss functions to evaluate the robustness of our results, as well as to address the fact that forecasters may dislike more positive than negative errors. For that reason we consider both symmetric (MSE) and asymmetric (QLIKE and LINEX) loss functions. Asymmetric loss functions that penalize under-prediction more heavily than over-prediction are especially useful in risk management due to the importance of downside risk.

The MSE and QLIKE loss functions are given by

$$MSE: \; L\left(m^{\alpha}, \widehat{m^{\alpha}}\right) = \left(m^{\alpha} - \widehat{m^{\alpha}}\right)^2 \tag{3.21}$$

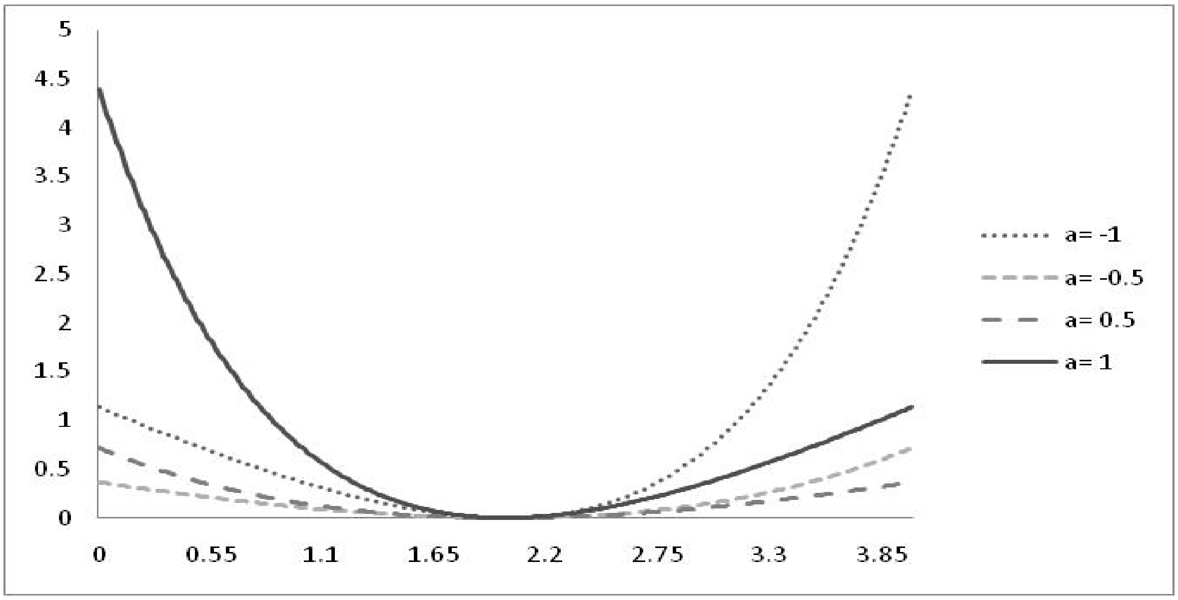
$$QLIKE: L\left(m^{\alpha}, \widehat{m^{\alpha}}\right) = \log\left(\widehat{m^{\alpha}}\right) + \frac{m^{\alpha}}{\widehat{m^{\alpha}}} \tag{3.22}$$

where m^{α} is a proxy of ES using high frequency *ex-post* volatility measures and $\widehat{m^{\alpha}}$ is a forecast of ES based on *ex-ante* ES forecasts. Patton (2008) showed that MSE and QLIKE loss functions are robust to noisy volatility proxies. Although MSE is the most popular symmetric loss function, the QLIKE distance is especially important in risk management since it penalizes under-prediction more heavily than over-prediction. Another loss function, which has the above property (for appropriate choice of its parameter) is the LINEX loss function (e.g. Varian (1974), Zellner (1986) and Christoffersen and Diebold (1997)):

$$LINEX: \; L\left(m^{\alpha}, \widehat{m^{\alpha}}, a\right) = \exp\left\{a\left(m^{\alpha} - \widehat{m^{\alpha}}\right)\right\} - a\left(m^{\alpha} - \widehat{m^{\alpha}}\right) - 1, \quad a \neq 0 \tag{3.23}$$

The choice of the parameter α determines whether under-prediction or over-prediction is more costly. We use $\alpha = 1$ so that the LINEX function assigns a higher penalty to errors due to under-prediction (as shown in Figure 1), which is essentially more important to policy makers, risk managers and regulators.

Figure 1: The LINEX loss function



This figure shows the LINEX loss function for alternative choices of the parameter a.

One approach is to evaluate the performance of ES forecasts using alternative loss functions for a given confidence level. However, if we want find an ES method that performs well across all confidence levels we need to use a more general evaluation criterion, such as Maximum Regret. The calculation of maximum regret consists of two steps: First, we compute the regret of each procedure for a given confidence level, which is given by the difference of the risk (MSE, QLIKE or LINEX loss function) and the best achievable risk between all methods. Second, we find the maximum regret of each procedure across different confidence levels.

4 Empirical Application

4.1 Data

Our database involves 5 minute data of major stock market indices from the US, European and Asian stock markets. We use the S&P 500 index, NASDAQ Composite for the US stock markets, the German DAX 30, the UK FTSE 100, the French CAC 40 and the Japanese NIKKEI 225. The S&P 500 index covers the period January 2, 1991 to August 31, 2009 and it is the relatively longer historical time series. The rest of the international stock market indices are available from July 1, 2003 to August 31, 2009. The data source is the Tick Data database. Table 1 shows the descriptive statistics of these stock market indices at daily frequency. The most important characteristics of the data are : (a) mean very close to zero, (b) mild skewness and (c) substantial kurtosis. The latter characteristic indicates that the assumption of Normality may be violated since the data exhibit fat tails.

Our objective is to forecast daily VaR and ES using 5 minute data for the estimation of *ex-post* high frequency volatility measures[4]. For the empirical applications using the S&P 500 index we use forecasts of VaR and ES based on a rolling window of 1000 observations. For the other international stock market indices we use a rolling window of 500 observations, given the smaller sample size.

Table 1: Descriptive Statistics of daily log returns

Index	Mean*	Variance*	Skewness	Kurtosis
S&P 500	2.424	1.372	−0.1516	12.003
NASDAQ	1.310	2.131	−0.0494	10.090
DAX 30	3.539	2.080	0.2686	13.295
FTSE 100	1.382	1.570	−0.1023	13.237
CAC 40	1.223	1.919	0.0810	12.266
NIKKEI 225	0.823	2.796	−0.4832	11.703

Table 1 shows the descriptive statistics of daily log returns of international stock market indices. S&P 500 covers the period from January 1991 to August 2009 and the other stock market indices from July 2003 to August 2009. * indicates that the actual numbers are multiplied by 10^{-4}.

4.2 Empirical Analysis based on the VaR forecasts

First we use the S&P 500 index and test the performance of VaR forecasts given by individual models. Model risk is present almost in all methods, especially for parametric GARCH-type models that are based on the Normal distribution. As shown in table 3, parametric models based on the Normal distribution pass the Conditional Coverage test (at 1% confidence level) only for 90% and 95% confidence levels. Given the exponential tails of the Normal distribution, these models fail to give accurate forecasts as we move further to the tail from 90% to 99.5%, where extreme events take place. From the large number of VaR models used in this paper only few perform well across all confidence levels. In contrast, parametric GARCH-type models based on the t-distribution turn out to be the top ranking models, since they can capture the fat tails of the loss distribution. Generally, models based on the FHS method perform well at lower confidence levels (e.g. 95%),

[4]The choice of 5 minute data is based on the empirical findings of other studies (e.g. Andersen et al., 2001b) which report that at this frequency there is no evidence of microstructure noise.

and models based on the EVT approach perform better at higher confidence levels (eg. 99.5%). The above results are also summarized in figures 3a and 3b, where the relative deviation from the expected percentage of violations is lower at the 99.5% confidence level for EVT and at the 97.5% for FHS. As expected, as we move further to the tail the Normal parametric models deviate from the expected percentage of violations. An interesting finding is that this deviation is positive for all models, which indicates under-estimation of risk, except of the parametric models based on the t distribution.

On the other hand, forecast combinations provide more accurate VaR forecasts. As shown in table 4, VaR forecasts based on Weighted BIC, Smoothed AIC, Bates-Granger, Granger-Ramanathan, Constrained Granger-Ramanathan and Mallows Model Averaging, pass the Conditional Coverage test for all confidence levels(except in two cases)[5]. In addition, Figure 4 shows that the deviations from the expected percentage of violations is close to zero, which indicates that forecast combinations can provide accurate forecasts of risk. Interestingly, the error of forecast combinations has the opposite direction than that of most of the individual models, since there is over-estimation of risk, especially at higher confidence levels. Here, it is important to mention that under-estimation of risk is more costly than over-estimation, especially for policy makers, since it can drive a bank to bankruptcy when the capital holdings are not sufficient to satisfy its obligations. Over-estimation of risk can also be costly for financial institutions since large capital holdings would not allow them to use this capital for investment and this can decrease profitability. Nevertheless, we find that the prediction error of forecast combination methods is relatively small.

Our empirical analysis is also extended to international stock market indices, such as the NASDAQ Composite, DAX 30, FTSE 100, CAC 40 and NIKKEI 225 (tables 5-6). Using the Dynamic Quantile test we find that only 3 models, FHS Normal APARCH, FHS t-APARCH and EVT Normal APARCH, pass the test for all stock market indices at a 1% confidence level. The results are especially interesting for the FTSE 100, since only 7 out of 24 models are not rejected by the test and even the parametric models based on the t distribution, which generally perform well, fail to pass the test. On the other hand, the performance of forecast combinations is superior for all stock market indices, since there are only a few rejections. When we use Realized Bipower Variation, which is robust to jumps, all forecast combination methods pass the Dynamic Quantile test at 1% confidence level. As shown in Figures 2a and 2b, VaR forecasts given by the Mallows Model Averaging approach respond almost immediately to rapid changes in volatility and can provide robust forecasts even in the presence of extreme events. This holds for the other forecast combination methods as well.

[5] Weighted BIC at the 99% confidence level when Realized Power Variation is used as *ex-post* high frequency volatility measure and Bates-Granger method at the 99% confidence level when Realized Bipower Variation is used as *ex-post* high frequency volatility measure.

4.3 Empirical Analysis based on the ES forecasts

We now turn to the empirical analysis of one of the most popular coherent risk measures, the Expected Shortfall (ES). Tables 7 and 8 show the mean 99% ES of the S&P 500 index using individual models and forecast combinations, respectively. As expected, as we move closer to the tail of the loss distribution the ES increases. Analyzing the ES forecasts given by the various models we get similar conclusions to those of the VaR. For example parametric models based on the Normal distribution give the lowest forecasts of ES, whereas the t-GARCH model gives the highest ES forecasts for all confidence levels.

Table 9 presents the bias and the loss of the ES forecasts given by individual models and combinations. There is positive bias in all models, except the t-GARCH, which indicates under-estimation of risk. On the contrary, the bias of the ES forecasts given by combinations in negative (and smaller in absolute value from the t-GARCH). The rankings given by symmetric and asymmetric loss functions are similar, despite that they penalize positive and negative forecast errors differently. This can be explained by the fact that the positive bias of individual models is significantly higher in absolute value from the negative bias of forecast combinations, and therefore the rankings do not change much when we use asymmetric loss functions instead of symmetric. The best performing methods using the MSE, QLIKE and LINEX loss function are forecast combinations. Particularly, when we use the MSE and QLIKE loss functions the best performing method is the Constrained Granger-Ramanathan and when we use the QLIKE loss function the Weighted BIC. From the individual models the best performing models are those based on the t distribution (t-TARCH has the best performance). There is no clear superiority between the models of the FHS and EVT families. From both FHS and EVT families the model with the Normal GARCH volatility structure has the highest rank. Parametric Normal models are the worst performing models. Here it is important to mention that the problem of under-estimation of risk observed in Normal GARCH-type parametric models is not due to the volatility structure but to the use of the inverse normal distribution function for the standardized returns. For example it can be proved that the Normal GARCH volatility model can capture heavy tails. This is the reason that FHS and EVT methods that use the Normal GARCH volatility model perform well.

In Table 10 we use the maximum regret of the above loss functions across various confidence levels (90%, 95%, 97.5%, 99% and 99.5%), which is a more robust evaluation measure. As noted by Hansen (2008) a forecast with low maximum regret does not perform significantly worse than the best performing model for each confidence level. Based on the maximum regret measure (of the MSE, QLIKE and LINEX loss functions) forecast combinations have the best performance. The Constrained Granger-Ramanathan is the best performing method among the other forecast combinations. The parametric t distributed models follow. The models from the EVT family generally perform better than those from the FHS family. From the EVT family the model with t-TARCH volatility structure has superior performance than the other models and from the FHS class the t-APARCH ranks first when we use the MSE and LINEX loss functions and Normal GARCH when we use the QLIKE loss function. Again the parametric models that assume Normality have the worst performance.

We now consider two subsamples of the S&P 500 using various forecast combination methods. The first subsample covers the period from January 2, 1996 to September 30, 2002 and the second subsample from October 1, 2002 to August 31, 2009. The most important financial event in the first subsample is the dot-com bubble and in the second the recent financial crisis. Figures 5a and 5b show the percentiles of the 95% ES of the two subsamples. The percentiles 0.005, 0.025, 0.25, 0.5 and 0.75 of the conditional ES in the first subsample are lower than those in the second subsample. In contrast, the 0.975 and 0.995 percentiles of the conditional ES in the second subsample are higher. This implies that the losses during the recent financial crisis are more severe than any other recent financial event, reflected by the upper percentiles of conditional ES.

5 Conclusions and Future Research

In this paper we test the performance of VaR and ES models following three alternative methods, the parametric, filtered historical simulation and extreme value theory, using some of the major stock market indices from the US, European and Asian stock markets. We rank the performance of these methods using backtesting, alternative loss functions and maximum regret. The results indicate that parametric GARCH models based on the t distribution rank first, models from the EVT family rank second, followed by models from the FHS family. As expected parametric models that assume Normality have the worst performance. We find that only a handful of models perform well across all the major stock market indices and across different confidence levels. In particular, as we increase the confidence level, it becomes more difficult to obtain accurate VaR and ES forecasts relying only on one individual model.

Instead we find that forecast combinations that take into account information from various methods, provide robust forecasts of VaR and ES across all confidence levels. The forecast combination method with the lowest maximum regret of ES forecasts is the Constrained Granger-Ramanathan. The use of data that involves the recent financial crisis, indicates that forecast combinations perform well even in the presence of extreme events, such as bankruptcies of banks and financial institutions (e.g. Lehman Brothers), where forecasting risk is even more challenging. In a nutshell our empirical results provide strong evidence that forecast combinations yield more accurate forecasts of risk than individual models and can deal with the problem of model uncertainty, which is an essential issue in modern risk management.

Given that financial decisions are often based on multi-period-ahead forecasts of risk, finding methods to forecast VaR in longer horizons than one day becomes very important. For example the Basel Committee on Banking Supervision requires banks to meet market risk capital requirements based on 10-day VaR estimates. Ongoing work involves using forecast combinations to predict 10-day VaR.

References

[1] Andersen, T.G.,T. Bollerslev, F. X. Diebold, H. Ebens, 2001a, The distribution of realized stock return volatility, *Journal of Financial Economics*, 61, 43-76.

[2] Andersen, T.G., T. Bollerslev, F.X. Diebold, P. Labys, 2001b, The distribution of exchange rate volatility, *Journal of the American Statistical Association*, 96, 42-55.

[3] Artzner, P., F. Delbaen, J.M.Eber and D. Heath, 1999, Coherent Measures of Risk, *Math Finance*, 9, 203-228.

[4] Avramov, D., 2002, Stock return predictability and model uncertainty, *Journal of Financial Economics*, 64, 423-458.

[5] Bardorff-Nielsen O.E., N. Shephard, 2002a, Econometric analysis of realized volatility and its use in estimating volatility models, *Journal of the Royal Statistical Society*, Series B, 64, 253 - 280.

[6] Bardorff-Nielsen O.E., N. Shephard, 2002b, Estimating quadratic variation using realized variance, *Journal of Applied Econometrics*, 17, 457-477.

[7] Bardorff-Nielsen O.E., N. Shephard, 2004, Power and bipower variation with stochastic volatility and jumps, *Journal of Financial Econometrics*, 2(1), 1-37.

[8] Bates, J.M., C.M.W. Granger, 1969, The combination of forecasts, *Operations Research Quarterly*, 20, 451-468.

[9] Christoffersen, P., 2003, Elements of Financial Risk Management, Academic Press.

[10] Christoffersen, P.F., F.X. Diebold, 1997, Optimal prediction under asymmetric loss, *Econometric Theory,* 13, 808–817.

[11] Clement, R.T., 1989, Combining forecasts: A review and annotated bibliography, International Journal of Forecasting, 5, 559-581.

[12] Duffie, D., J. Pan, 1997, An Overview of Value at Risk, *Journal of Derivatives*, 4, 7-49.

[13] Engle, R.F. and S. Manganelli, 2004, CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles, *Journal of Business and Economic Statistics*, 22, 367-381.

[14] Furnival, G.M. and R.W. Wilson, 1974, Regression by leaps and bounds, *Technometrics*, 16, 499-511.

[15] Giacomini R., I. Komunjer, 2005, Evaluation and Combination of Conditional Quantile Forecasts, *Journal of Business and Economic Statistics*, 23(4), 416-431.

[16] Granger, C.W.J., R. Ramanathan, 1984, Improved methods of combining forecast accuracy, *Journal of Forecasting*, 19, 197-204.

[17] Hansen, B.E., Least-squares forecast averaging, Journal of Econometrics, 146(2), 342-350.

[18] Jorion, P., 2000, Value at Risk: The New Benchmark for Managing Financial Risk (2nd edition), McGraw-Hill.

[19] Kuester K., S. Mittnik and M. S. Paollela, 2006, Value-at-Risk Prediction: A Comparison of Alternative Strategies, *Journal of Financial Econometrics*, 4(1), 53-89.

[20] Liu C., J.M. Maheu, 2009, Forecasting realized volatility: a Bayesian model-averaging approach, *Journal of Applied Econometrics*, 24(5), 709-733.

[21] Magnus, J.R., O. Powell, P. Prufer, 2009, A comparison of two model averaging techniques with an application to growth empirics, *forthcoming in Journal of Econometrics.*

[22] McNeil, A.J., R. Frey, P. Embrechts, 2005, Quantitative Risk Management, Princeton University Press.

[23] Patton, A., K. Sheppard, 2009, Optimal Combinations of realized volatility estimators, *International Journal of Forecasting*, 25, 218-238.

[24] Patton, A., 2008, Volatility Forecast Comparison using Imperfect Volatility Proxies, working paper, University of Oxford.

[25] Pesaran, M.H., C. Schleicher, P. Zaffaroni, 2008, *Journal of Empirical Finance*, 16, 280-305.

[26] Stock, J.H., M. Watson, 2004, Combination forecasts of output growth in a seven-country data set, *Journal of Forecasting*, 23, 405-430.

[27] Stock, J.H., M. Watson, 2008, Philips Curve Inflation Forecasts, Working Paper, National Bureau of Economic Research.

[28] Timmermann, A., 2006, Forecast Combinations, In: Elliott, G., Granger, C.W.J., Timmermann, A. (Eds.), *Handbook of Economic Forecasting*, vol.1, Elsevier, Amsterdam, 135-196.

[29] Varian, H.R., 1974, A Bayesian approach to real estate assessment, In: Fienberg, S.E., Zellner, A. (Eds.), Studies in Bayesian Econometrics and Statistics in Honor of Leonard J. Savage, North-Holland, Amsterdam, pp. 195–208.

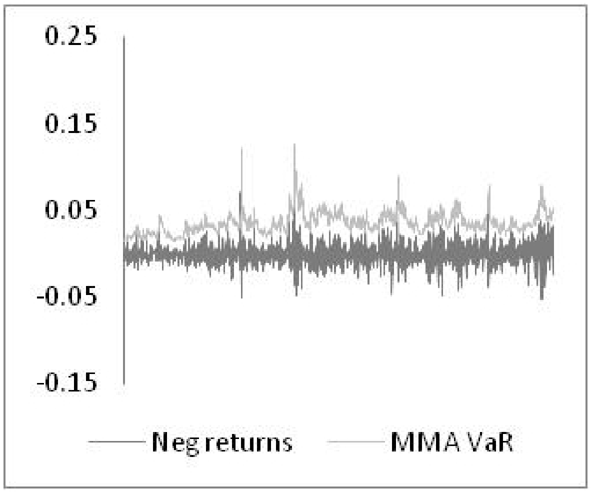
[30] Wright, J.H., 2008, Bayesian Model Averaging and exchange rate forecasts, *Journal of Econometrics*, 146, 329-341.

[31] Zellner, A., 1986, Bayesian estimation and prediction using asymmetric loss functions, *Journal of the American Statistical Association,* 81, 446–451.

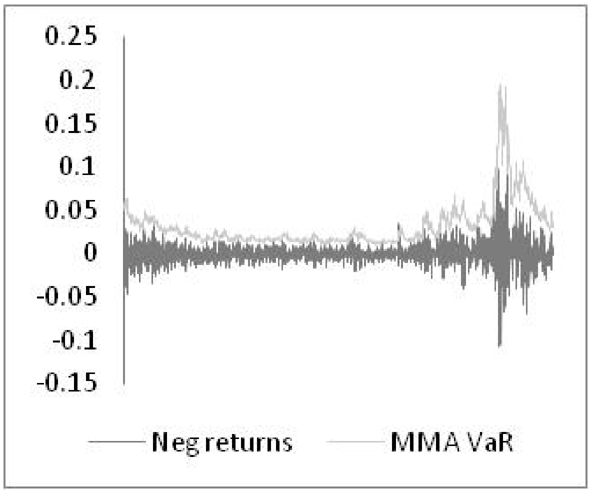
APPENDIX

Figure 2a: 99% VaR forecasts using the Mallows Model Averaging method.

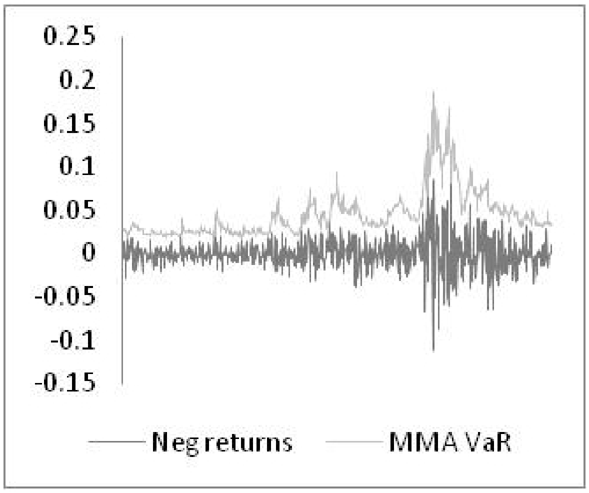
S&P 500: January 1996 - September 2002



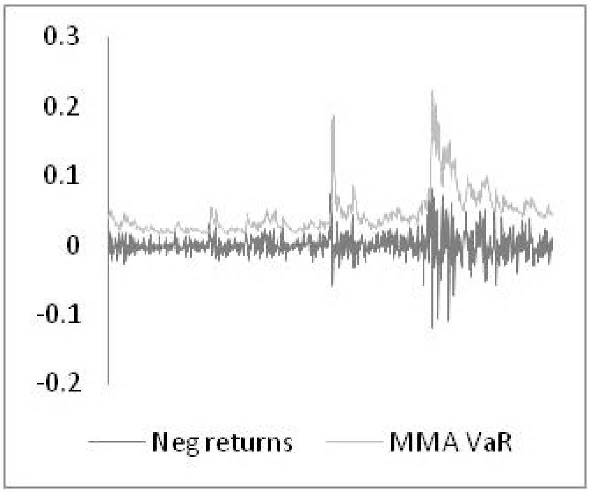
S&P 500: October 2002 - August 2009



NASDAQ: June 2006 - August 2009



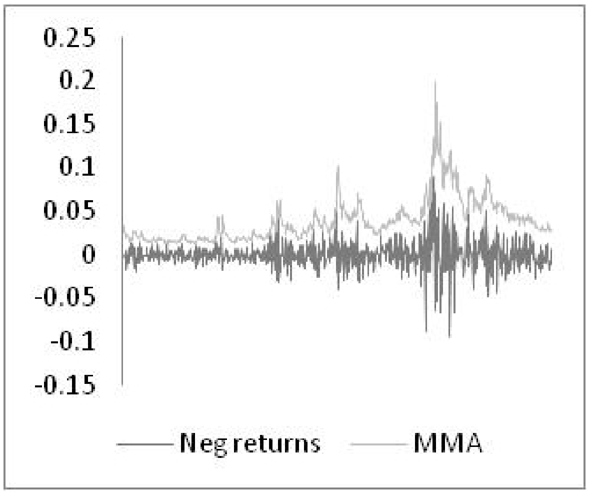
DAX 30: June 2006 - August 2009



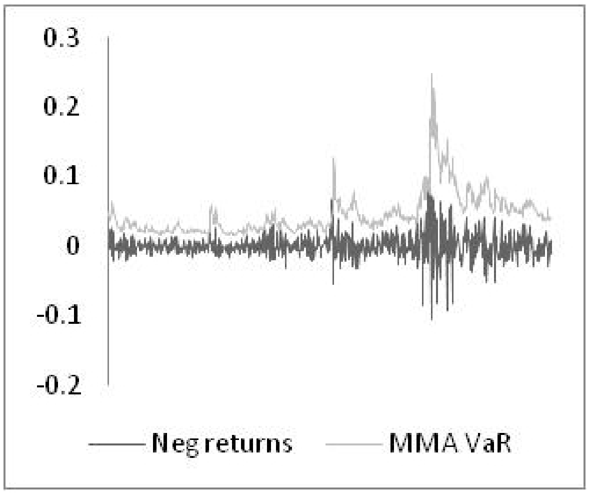
This figure shows the 99% VaR forecasts obtained by Mallows Model Averaging for the S&P 500, NASDAQ Composite and DAX 30 stock market indices. The *ex-post* volatility measure is Realized Bipower Variation. The VaR forecasts are based on a rolling window of 1000 observations for the S&P 500 and a rolling window of 500 observations for the NASDAQ and DAX 30 stock market indices.

Figure 2b: 99% VaR forecasts using the Mallows Model Averaging method

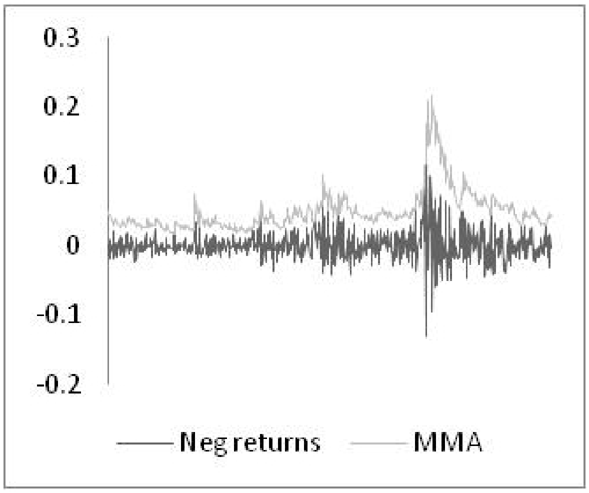
FTSE 100: June 2006 - August 2009



CAC 40: June 2006 - August 2009

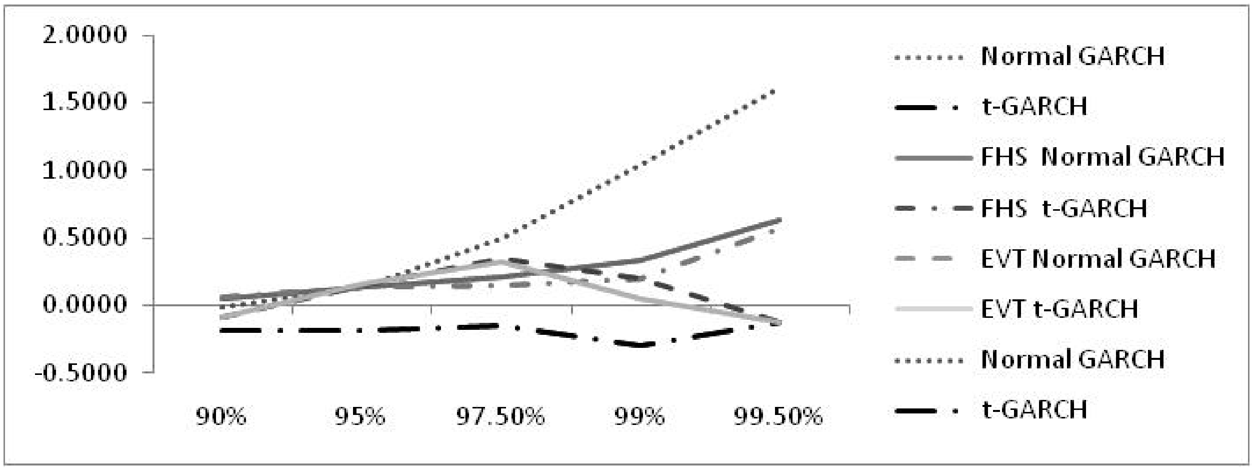


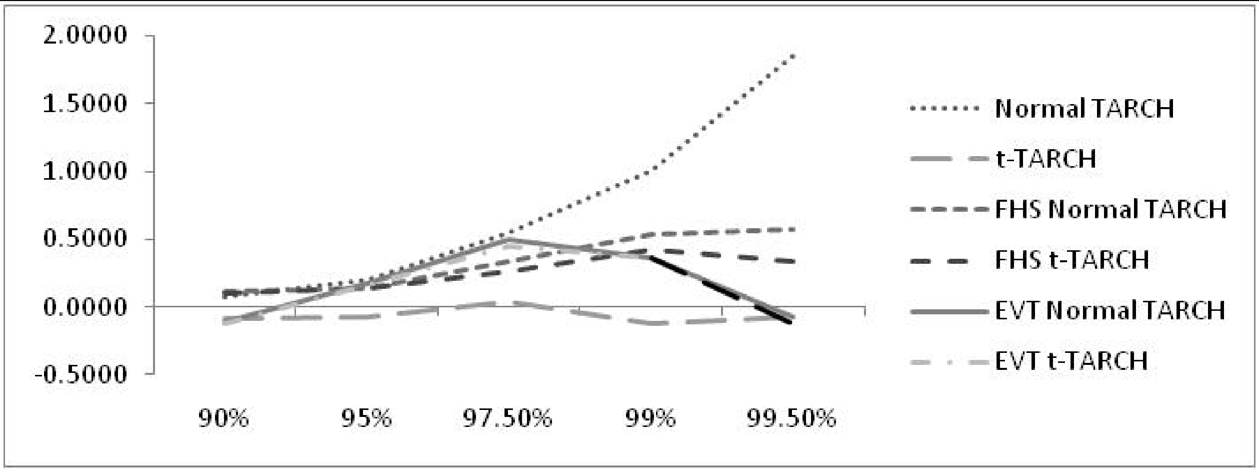
NIKKEI 225: June 2006 - August 2009



This figure shows the 99% VaR forecasts obtained by Mallows Model Averaging for the FTSE 100, CAC 40 and NIKKEI 225 stock market indices. The *ex-post* volatility measure is Realized Bipower Variation. The VaR forecasts are based on a rolling window of 500 observations.

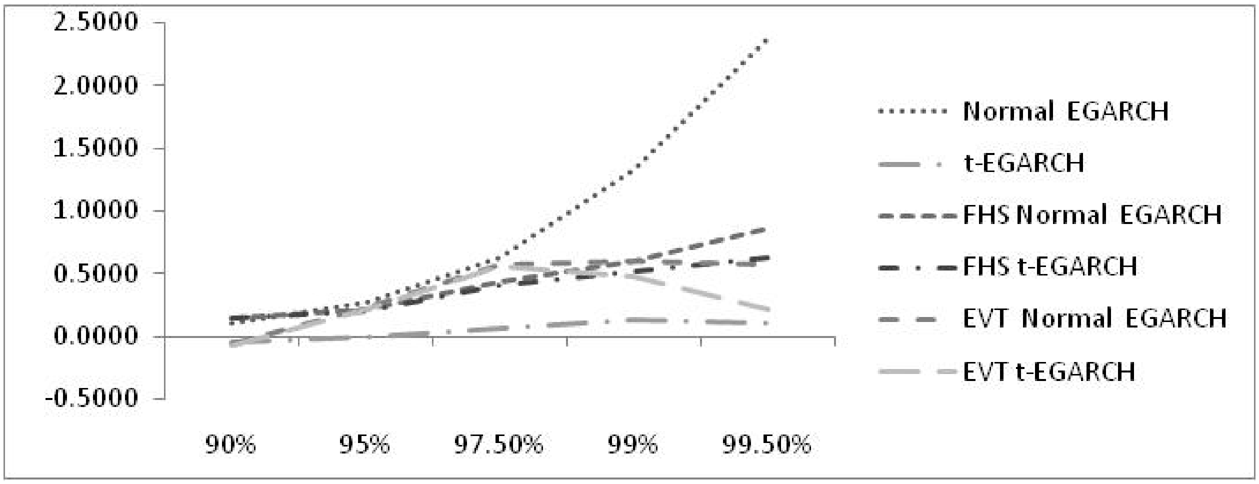
Figure 3a: Relative Deviations from the expected percentage of violations of VaR forecasts of individual models using the S&P 500 index

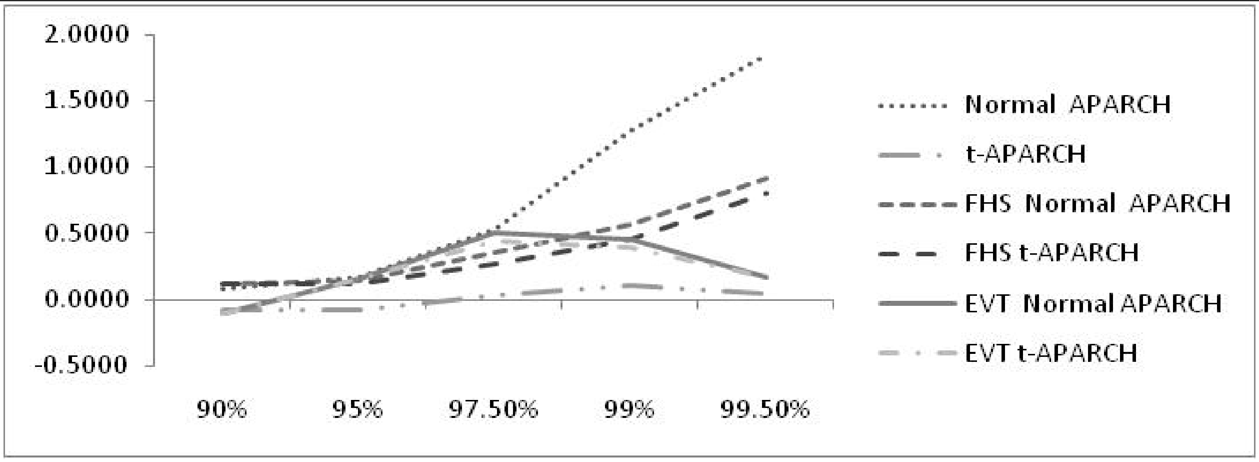




This figure shows the relative deviations from the expected percentage of VaR violations across different confidence levels for individual models using the S&P 500 index. These deviations are given by $\frac{E(I_t)-\alpha}{\alpha}$ and can be estimated by $\frac{\frac{1}{T}\sum_{t=1}^{T} I_t-\alpha}{\alpha}$, where I_t is the hit sequence and α the confidence level. The data sample covers the period from January 1996 to August 2009 (excluding the estimation window). The VaR forecasts are based on a rolling window of 1000 observations.

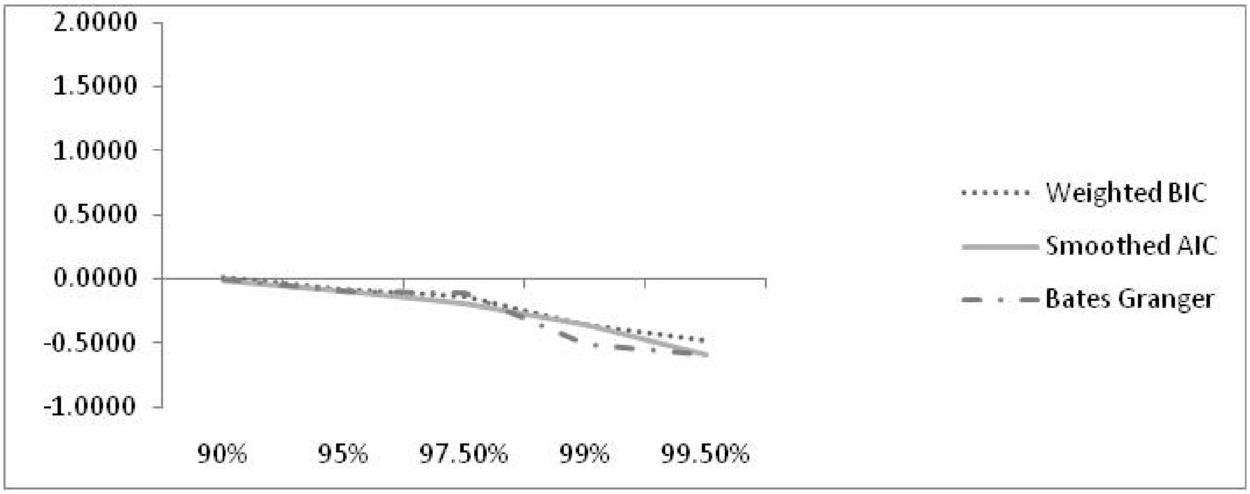
Figure 3b: Relative Deviations from the expected percentage of violations of VaR forecasts of individual models using the S&P 500 index.

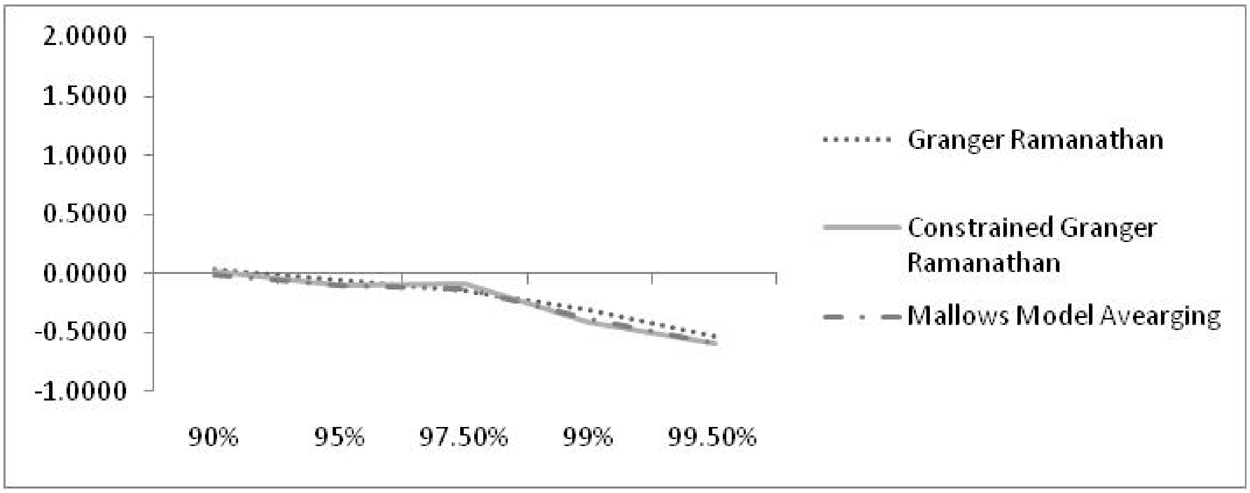




This figure shows the relative deviations from the expected percentage of VaR violations across different confidence levels for individual models using the S&P 500 index. These deviations are given by $\frac{E(I_t)-\alpha}{\alpha}$ and can be estimated by $\frac{\frac{1}{T}\sum_{t=1}^{T} I_t-\alpha}{\alpha}$, where I_t is the hit sequence and α the confidence level. The data sample covers the period from January 1996 to August 2009 (excluding the estimation window). The VaR forecasts are based on a rolling window of 1000 observations.

Figure 4: Relative Deviations from the expected percentage of violations of VaR forecasts of forecast combinations using the S&P 500 index.

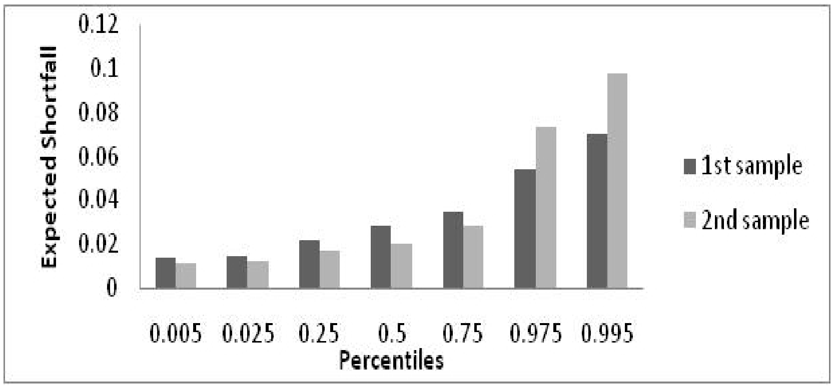




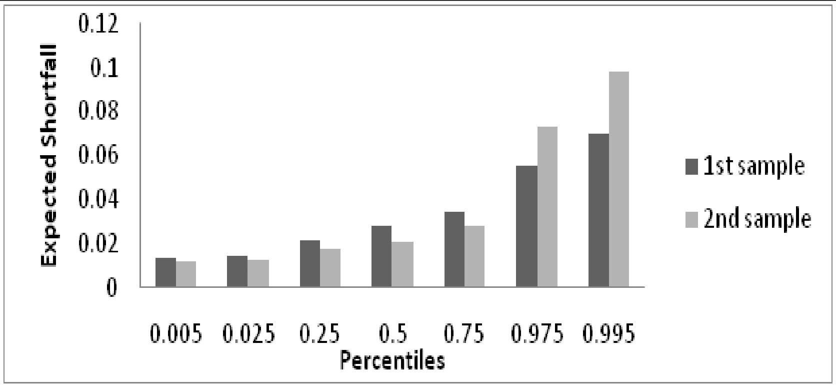
This figure shows the relative deviations from the expected percentage of VaR violations across different confidence levels for forecast combinations using the S&P 500 index. These deviations are given by $\frac{E(I_t)-\alpha}{\alpha}$ and can be estimated by $\frac{\frac{1}{T}\sum_{t=1}^{T} I_t-\alpha}{\alpha}$, where I_t is the hit sequence and α the confidence level. The data sample covers the period from January 1996 to August 2009 (excluding the estimation window). Realized Bipower Variation is used as *ex-post* volatility measure. The VaR forecasts are based on a rolling window of 1000 observations.

Figure 5a: Percentiles of the 95% Expected Shortfall of two subsamples of the S&P 500 index.

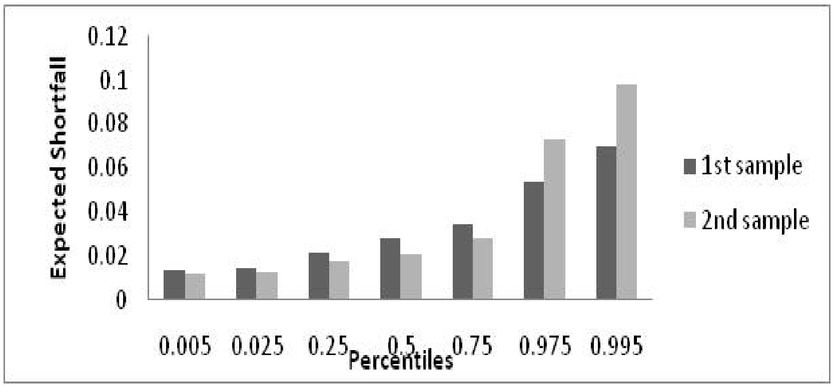
Weighted BIC



Smoothed AIC



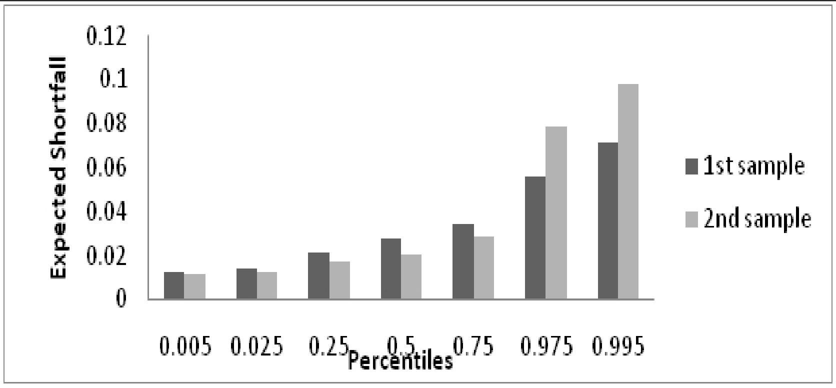
Bates - Granger



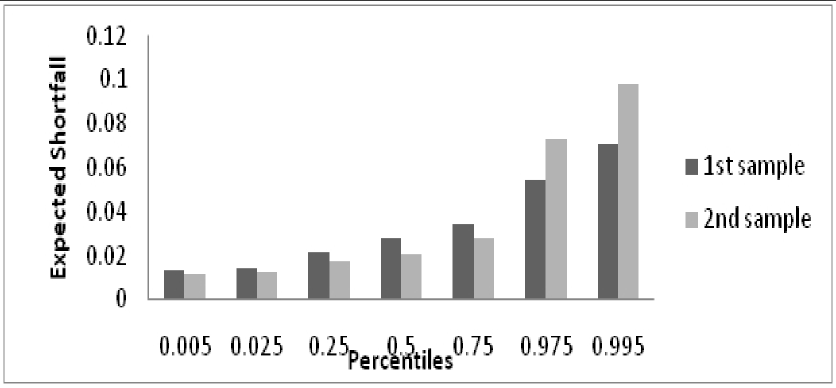
This figure shows the percentiles of the 95% Expected Shortfall of two subsamples of the S&P 500 index given by forecast combinations. The first subsample covers the period from January 1996 to September 2002 and the second subsample from October 2002 to August 2009. Realized Bipower Variation is used as *ex-post* high frequency volatility measure. The size of the rolling window is 1000 observations.

Figure 5b: Percentiles of the 95% Expected Shortfall of two subsamples of the S&P 500 Index.

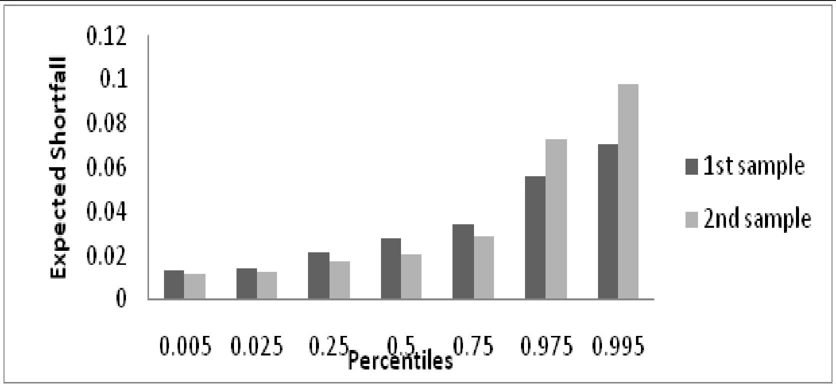
Granger - Ramanathan



Constrained Granger - Ramanathan



Mallows Model Averaging



This figure shows the percentiles of the 95% Expected Shortfall of two subsamples of the S&P 500 index given by forecast combinations. The first subsample covers the period from January 1996 to September 2002 and the second subsample from October 2002 to August 2009.Realized Bipower Variation is used as *ex-post* high frequency volatility measure. The size of the rolling window is 1000 observations.

Table 2: *Ex-ante* VaR forecasts.

1	Historical Simulation	$\widehat{F}_{r_t}^{-1}(\alpha)$	14	FHS t-GARCH	$\sigma_{5,t+1}\widehat{F}_{\varepsilon_t}^{-1}(\alpha)$
2	Normal GARCH	$\sigma_{1,t+1}\Phi^{-1}(\alpha)$	15	FHS t-TARCH	$\sigma_{6,t+1}\widehat{F}_{\varepsilon_t}^{-1}(\alpha)$
3	Normal TARCH	$\sigma_{2,t+1}\Phi^{-1}(\alpha)$	16	FHS t-EGARCH	$\sigma_{7,t+1}\widehat{F}_{\varepsilon_t}^{-1}(\alpha)$
4	Normal EGARCH	$\sigma_{3,t+1}\Phi^{-1}(\alpha)$	17	FHS t-APARCH	$\sigma_{8,t+1}\widehat{F}_{\varepsilon_t}^{-1}(\alpha)$
5	Normal APARCH	$\sigma_{4,t+1}\Phi^{-1}(\alpha)$	18	EVT Normal GARCH	$\sigma_{1,t+1}\widehat{F}_{EVT}^{-1}(\alpha)$
6	t-GARCH	$\sigma_{5,t+1}t_v^{-1}(\alpha)$	19	EVT Normal TARCH	$\sigma_{2,t+1}\widehat{F}_{EVT}^{-1}(\alpha)$
7	t-TARCH	$\sigma_{6,t+1}t_v^{-1}(\alpha)$	20	EVT Normal EGARCH	$\sigma_{3,t+1}\widehat{F}_{EVT}^{-1}(\alpha)$
8	t-EGARCH	$\sigma_{7,t+1}t_v^{-1}(\alpha)$	21	EVT Normal APARCH	$\sigma_{4,t+1}\widehat{F}_{EVT}^{-1}(\alpha)$
9	t-APARCH	$\sigma_{8,t+1}t_v^{-1}(\alpha)$	22	EVT t-GARCH	$\sigma_{5,t+1}\widehat{F}_{EVT}^{-1}(\alpha)$
10	FHS Normal GARCH	$\sigma_{1,t+1}\widehat{F}_{\varepsilon_t}^{-1}(\alpha)$	23	EVT t-TARCH	$\sigma_{6,t+1}\widehat{F}_{EVT}^{-1}(\alpha)$
11	FHS Normal TARCH	$\sigma_{2,t+1}\widehat{F}_{\varepsilon_t}^{-1}(\alpha)$	24	EVT t-EGARCH	$\sigma_{7,t+1}\widehat{F}_{EVT}^{-1}(\alpha)$
12	FHS Normal EGARCH	$\sigma_{3,t+1}\widehat{F}_{\varepsilon_t}^{-1}(\alpha)$	25	EVT t-APARCH	$\sigma_{8,t+1}\widehat{F}_{EVT}^{-1}(\alpha)$
13	FHS Normal APARCH	$\sigma_{4,t+1}\widehat{F}_{\varepsilon_t}^{-1}(\alpha)$			

This table consists of the model specifications of the ex ante VaR forecasts that are used as regressors in the linear model.

Let $r_{t+1} = \sigma_{t+1}\varepsilon_{t+1}$.We use 8 different volatility models,

(1) Normal GARCH(1,1): $\sigma_{1,t+1}^2 = \omega + \alpha r_t^2 + \beta\sigma_{1,t}^2$ $\quad \varepsilon_{t+1} \sim N(0,1)$

(2) Normal TARCH(1,1): $\sigma_{2,t+1}^2 = \omega + \alpha r_t^2 + \beta\sigma_{2,t}^2 + \theta r_t^2 1_{\{r_t<0\}}$ $\quad \varepsilon_{t+1} \sim N(0,1)$

(3) Normal EGARCH(1,1): $\log\left(\sigma_{3,t+1}^2\right) = \omega + \alpha\left[\frac{|r_t|}{\sigma_{3,t}} - E\left\{\frac{|r_t|}{\sigma_{3,t}}\right\}\right] + \beta\log\left(\sigma_{3,t}^2\right) + \theta\frac{r_t}{\sigma_{3,t}}$ $\quad \varepsilon_{t+1} \sim N(0,1)$

(4) Normal APARCH(1,1): $\sigma_{4,t+1}^\delta = \omega + \alpha\left(|r_t| - \theta r_t\right)^\delta + \beta\sigma_{4,t}^\delta$ $\quad \varepsilon_{t+1} \sim N(0,1)$

(5) t-GARCH(1,1): $\sigma_{5,t+1}^2 = \omega + \alpha r_t^2 + \beta\sigma_{5,t}^2$ $\quad \varepsilon_{t+1} \sim t_v$

(6) t-TARCH(1,1): $\sigma_{6,t+1}^2 = \omega + \alpha r_t^2 + \beta\sigma_{6,t}^2 + \theta r_t^2 1_{\{r_t<0\}}$ $\quad \varepsilon_{t+1} \sim t_v$

(7) t-EGARCH(1,1): $\log\left(\sigma_{7,t+1}^2\right) = \omega + \alpha\left[\frac{|r_t|}{\sigma_{7,t}} - E\left\{\frac{|r_t|}{\sigma_{7,t}}\right\}\right] + \beta\log\left(\sigma_{7,t}^2\right) + \theta\frac{r_t}{\sigma_{7,t}}$ $\quad \varepsilon_{t+1} \sim t_v$

(8) t-APARCH(1,1): $\sigma_{8,t+1}^\delta = \omega + \alpha\left(|r_t| - \theta r_t\right)^\delta + \beta\sigma_{8,t}^\delta$ $\quad \varepsilon_{t+1} \sim t_v$

We also use the inverse of the following distribution functions for each method

(1) Parmatric $(Normal)$ $\quad \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du$

(2) Parametric (t) $\quad t_v(x) = \int_{-\infty}^{x} \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)} \frac{1}{\sqrt{v\pi}} \frac{1}{\left(1+\left(\frac{u^2}{v}\right)\right)^{(v+1)/2}} du$

(3) Filtered Historical Simulation (FHS) $\quad \widehat{F}_{\varepsilon_t}(\alpha) = \frac{1}{k}\sum_{s=t-k+1}^{t} 1_{\{\varepsilon_s \leq a\}}$

(4) Extreme Value Theory (EVT) $\quad \widehat{F}_{EVT}(x) = 1 - \frac{T_u}{T}\left(\frac{x}{u}\right)^{-1/\xi}$

where T is the total number of observations, T_u is the number of observations beyond the threshold u and ξ is the shape parameter of the Generalized Pareto Distribution.

Table 3: P-values of the Conditional Coverage Test for individual VaR models using the S&P 500 index across different confidence levels

Model	$\alpha = 90\%$	$\alpha = 95\%$	$\alpha = 97.5\%$	$\alpha = 99\%$	$\alpha = 99.5\%$
Normal GARCH	**0.4050**	**0.1016**	0	0	0
Normal TARCH	**0.1599**	**0.0342**	0	0	0
Normal EGARCH	**0.1031**	0.0013	0	0	0
Normal APARCH	**0.1645**	**0.0466**	0	0	0
t-GARCH	0.0001	**0.0113**	**0.0265**	**0.1439**	**0.8087**
t-TARCH	**0.0924**	**0.3689**	**0.1040**	**0.4005**	**0.8894**
t-EGARCH	**0.5872**	**0.4625**	**0.2760**	**0.4733**	**0.8202**
t-APARCH	**0.1578**	**0.4905**	**0.2675**	**0.5429**	**0.8926**
FHS Normal GARCH	**0.2808**	**0.1736**	**0.0242**	**0.1499**	**0.0452**
FHS Normal TARCH	0.0077	**0.0928**	0.0060	**0.0124**	**0.0736**
FHS Normal EGARCH	**0.0130**	**0.0144**	0.0004	0.0050	0.0045
FHS Normal APARCH	**0.0202**	**0.0637**	0.0026	0.0080	0.0023
FHS t-GARCH	**0.3355**	**0.1249**	**0.0371**	**0.4416**	**0.0736**
FHS t-TARCH	0.0085	**0.1549**	**0.0234**	**0.0588**	**0.3511**
FHS t-EGARCH	**0.0118**	**0.0198**	0.0010	**0.0189**	**0.0452**
FHS t-APARCH	**0.0420**	**0.1279**	**0.0234**	**0.0411**	0.0084
EVT Normal GARCH	**0.0498**	**0.0533**	0.0026	**0.4416**	**0.8087**
EVT Normal TARCH	0.0006	**0.0302**	0	**0.0472**	**0.8894**
EVT Normal EGARCH	**0.2076**	0.0039	0	0.0050	**0.0736**
EVT Normal APARCH	**0.0191**	**0.0533**	0	**0.0411**	**0.7144**
EVT t-GARCH	**0.0924**	**0.0678**	0.0020	**0.6693**	**0.8087**
EVT t-TARCH	0.0022	**0.0917**	0.0003	**0.0472**	**0.8087**
EVT t-EGARCH	**0.1083**	**0.0122**	0	**0.0282**	**0.5914**
EVT t-APARCH	**0.0110**	**0.0533**	0.0003	**0.0822**	**0.7144**

This table consists of the p-values of the Conditional Coverage test of 99% VaR forecast of individual models using the S&P 500 Composite Index. The bold numbers are the p-values greater than 0.01, which indicate that the VaR forecasts pass the test at 1% confidence level. The data sample covers the period from January 1996 to August 2009 (excluding the estimation window). The out of sample VaR forecasts are based on a rolling window of 1000 observations.

Table 4: P-values of the Conditional Coverage test for forecast combinations using the S&P 500 index across different confidence levels.

	$\alpha = 90\%$	$\alpha = 95\%$	$\alpha = 97.5\%$	$\alpha = 99\%$	$a = 99.5\%$
Ex post VaR	Realized Volatility (RV)				
Weighted BIC	**0.4321**	**0.0796**	**0.0550**	**0.0142**	**0.0908**
Smoothed AIC	**0.1816**	**0.0389**	**0.0758**	**0.0142**	**0.0450**
Bates-Granger	**0.4154**	**0.0604**	**0.1053**	**0.0249**	**0.0196**
Granger-Ramanathan	**0.1426**	**0.0743**	**0.2215**	**0.0994**	**0.0196**
Constrained Granger-Ramanathan	**0.5809**	**0.0487**	**0.1591**	**0.0249**	**0.0196**
Mallows Model Averaging	**0.2301**	**0.0389**	**0.0502**	**0.0249**	**0.0450**
Ex post VaR	Realized Power Variation (RPV)				
Weighted BIC	**0.8257**	**0.1838**	**0.0265**	0.0077	**0.0908**
Smoothed AIC	**0.5359**	**0.1838**	**0.1053**	**0.0414**	**0.0196**
Bates-Granger	**0.8518**	**0.1560**	**0.0653**	**0.0249**	**0.0196**
Granger-Ramanathan	**0.7345**	**0.3262**	**0.0653**	**0.0994**	**0.0908**
Constrained Granger-Ramanathan	**0.9439**	**0.2148**	**0.1304**	**0.0249**	**0.0908**
Mallows Model Averaging	**0.5775**	**0.1838**	**0.0653**	**0.0249**	**0.0196**
Ex post VaR	Realized Bipower Variation (RBP)				
Weighted BIC	**0.8513**	**0.2159**	**0.0344**	**0.0657**	**0.0908**
Smoothed AIC	**0.9630**	**0.0777**	**0.0989**	**0.0657**	**0.0196**
Bates-Granger	**0.9647**	**0.1361**	**0.0680**	0.0039	**0.0196**
Granger-Ramanathan	**0.6864**	**0.3949**	**0.0836**	**0.1439**	**0.0450**
Constrained Granger-Ramanathan	**0.8837**	**0.1148**	**0.0827**	**0.0249**	**0.0196**
Mallows Model Averaging	**0.9069**	**0.1361**	**0.2929**	**0.0414**	**0.0196**

This table consists of the p-values of the Conditional Coverage test of 99% VaR forecast of forecast combinations using the S&P 500 Composite index. The bold numbers are the p-values greater than 0.01, which indicate that the VaR forecasts pass the test at 1% confidence level. The data sample covers the period from January 1996 to August 2009 (excluding the estimation window). The out of sample VaR forecasts are based on a rolling window of 1000 observations.

Table 5: P-values of the Dynamic Quantile Test for individual VaR models using international stock market indices for a 99% confidence level.

Model	NASDAQ	DAX 30	FTSE 100	CAC 40	NIKKEI 225
Normal GARCH	**0.0691**	**0.0817**	0	**0.1146**	**0.0155**
Normal TARCH	**0.0302**	**0.0204**	0	**0.0328**	0.0034
Normal EGARCH	0	0.0002	0	0	0
Normal APARCH	0.0003	0.0024	0	0	0.0001
t-GARCH	**0.9699**	**0.9527**	0.0076	**0.0298**	**0.0826**
t-TARCH	**0.8780**	**0.9630**	0.0028	**0.1597**	**0.7250**
t-EGARCH	**0.7131**	**0.8624**	0.0043	**0.3808**	**0.1173**
t-APARCH	**0.8480**	**0.7001**	0	0	**0.1111**
FHS Normal GARCH	**0.6807**	**0.9029**	0.0075	**0.1006**	**0.0661**
FHS Normal TARCH	**0.7452**	**0.1726**	0.0079	**0.1341**	**0.9257**
FHS Normal EGARCH	0.0028	**0.1469**	**0.1117**	0.0019	**0.1602**
FHS Normal APARCH	**0.0529**	**0.4628**	**0.0170**	**0.1373**	**0.5041**
FHS t-GARCH	**0.6952**	**0.9214**	0.0075	**0.1032**	**0.0817**
FHS t-TARCH	**0.6850**	**0.1757**	0.0078	**0.1366**	**0.9769**
FHS t-EGARCH	0.0056	**0.0446**	**0.0885**	**0.0234**	**0.0137**
FHS t-APARCH	**0.3673**	**0.5010**	**0.0171**	**0.1449**	**0.4642**
EVT Normal GARCH	**0.4909**	**0.9763**	0.0081	**0.1011**	**0.0881**
EVT Normal TARCH	**0.7656**	**0.2350**	0.0089	**0.1380**	**0.1093**
EVT Normal EGARCH	**0.0999**	**0.3127**	**0.1261**	0.0077	**0.0485**
EVT Normal APARCH	**0.6060**	**0.3263**	**0.0221**	**0.0321**	**0.0603**
EVT t-GARCH	**0.9407**	**0.9786**	0.0083	**0.1041**	**0.0835**
EVT t-TARCH	**0.9652**	**0.1704**	0.0089	**0.1411**	**0.3997**
EVT t-EGARCH	**0.1002**	**0.1287**	**0.1208**	**0.0800**	**0.0763**
EVT t-APARCH	**0.7816**	**0.8682**	0.0085	**0.0328**	**0.0832**

This table consists of the p-values of the Dynamic Quantile test of 99% VaR forecast of individual models using international stock market indices. The bold numbers are the p-values greater than 0.01, which indicate that the VaR forecasts pass the test at 1% confidence level. The data sample covers the period from June 2006 to August 2009 (excluding the estimation window). The out of sample VaR forecasts are based on a rolling window of 1000 observations.

Table 6: P-values of the Dynamic Quantile test for forecast combinations using international stock market indices at 99% confidence level.

	NASDAQ	DAX 30	FTSE 100	CAC 40	NIKKEI 225
Ex post VaR	Realized Volatility (RV)				
Weighted BIC	**0.7636**	**0.8892**	0.0047	**0.8455**	**0.9853**
Smoothed AIC	**0.7730**	**0.8900**	**0.0283**	**0.8207**	0.0018
Bates-Granger	**0.7693**	**0.9760**	**0.0237**	**0.8294**	**0.9739**
Granger-Ramanathan	**0.7709**	**0.0256**	**0.9883**	0.0013	**0.8843**
Constrained Granger-Ramanathan	**0.7682**	**0.9742**	**0.0236**	**0.9623**	**0.9378**
Mallows Model Averaging	**0.7711**	**0.9772**	**0.0233**	**0.8281**	**0.9921**
Ex post VaR	Realized Power Variation (RPV)				
Weighted BIC	**0.7616**	**0.9823**	**0.1048**	**0.9651**	**0.9838**
Smoothed AIC	**0.7731**	**0.9670**	**0.0651**	**0.9623**	**0.9576**
Bates-Granger	**0.7684**	**0.8888**	**0.1026**	**0.9633**	**0.9537**
Granger-Ramanathan	0.0001	**0.0295**	**0.0170**	**0.8919**	**0.9872**
Constrained Granger-Ramanathan	**0.7667**	**0.9391**	**0.0537**	**0.9642**	**0.9209**
Mallows Model Averaging	**0.7708**	**0.9671**	**0.0261**	**0.9632**	**0.9626**
Ex post VaR	Realized Bipower Variation (RBP)				
Weighted BIC	**0.7624**	**0.9811**	**0.1356**	**0.9639**	**0.9956**
Smoothed AIC	**0.7735**	**0.9831**	**0.0283**	**0.9609**	**0.9738**
Bates-Granger	**0.7690**	**0.9813**	**0.1307**	**0.9624**	**0.9821**
Granger-Ramanathan	**0.9130**	**0.0297**	**0.0470**	**0.7056**	**0.8331**
Constrained Granger-Ramanathan	**0.7659**	**0.9819**	**0.1061**	**0.8924**	**0.9747**
Mallows Model Averaging	**0.7710**	**0.9832**	**0.0560**	**0.9620**	**0.9769**

This table consists of the p-values of the Dynamic Quantile test of 99% VaR forecasts of forecast combinations using international stock market indices. The bold numbers are the p-values greater than 0.01, which indicate that the VaR forecasts pass the test at 1% confidence level. The data sample covers the period from June 2006 to August 2009 (excluding the estimation window). The out of sample VaR forecasts are based on a rolling window of 1000 observations.

Table 7: Mean 99% ES forecasts of individual models using the S&P 500 index.

Model	$\alpha = 90\%$	$\alpha = 95\%$	$\alpha = 97.5\%$	$\alpha = 99\%$	$\alpha = 99.5\%$
Normal GARCH	22059	26186	29641	33911	36624
Normal TARCH	21812	25636	28825	32579	34895
Normal EGARCH	21138	**24888**	**28218**	**32197**	**34619**
Normal APARCH	21508	25399	28691	32709	35015
t-GARCH	**23644**	**29119**	**34399**	**41649**	**45700**
t-TARCH	23163	27977	32440	38534	41754
t-EGARCH	22429	27259	31932	38068	41454
t-APARCH	22688	27565	32100	38099	41388
FHS Normal GARCH	21209	25986	31193	37380	42836
FHS Normal TARCH	21199	25916	30003	35405	40865
FHS Normal EGARCH	**20510**	25159	29188	34830	41385
FHS Normal APARCH	21003	25756	29882	35272	41637
FHS t-GARCH	21326	25942	31102	37390	43096
FHS t-TARCH	21298	26011	30258	35967	41440
FHS t-EGARCH	20668	25343	29344	35161	41820
FHS t-APARCH	21135	25855	30111	35679	42042
EVT Normal GARCH	22497	25813	30058	38107	44533
EVT Normal TARCH	22897	25766	29301	35696	40776
EVT Normal EGARCH	22249	25048	28563	34794	39858
EVT Normal APARCH	22775	25626	29115	35462	40562
EVT t-GARCH	22545	25767	30048	38178	44818
EVT t-TARCH	22948	25858	29456	36104	41487
EVT t-EGARCH	22348	25194	28781	35175	40431
EVT t-APARCH	22861	25726	29347	35885	41186

This table consists of the mean of the conditional 99% Expected Shortfall of individual models using the S&P 500 index. The data sample covers the period from January 1996 to August 2009 (excluding the estimation window). The out of sample forecasts are based on a rolling window of 1000 observations. The bold numbers show the lowest and highest mean ES for each confidence level.

Table 8: Mean 99% ES forecasts of forecast combinations using the S&P 500.

	$\alpha = 90\%$	$\alpha = 95\%$	$\alpha = 97.5\%$	$\alpha = 99\%$	$a = 99.5\%$
Ex post VaR	Realized Volatility (RV)				
Weighted BIC	22315	27811	32979	41651	46072
Smoothed AIC	**22668**	**27979**	32918	41609	45712
Bates-Granger	22444	27812	32883	41787	**46104**
Granger-Ramanathan	22540	27968	32719	41426	45212
Constrained Granger-Ramanathan	22377	27809	32958	**41817**	45983
Mallows Model Averaging	22574	27921	32932	41622	45848
Ex post VaR	Realized Power Variation (RPV)				
Weighted BIC	22103	27616	**33241**	41400	45717
Smoothed AIC	22279	27746	32922	41251	45287
Bates-Granger	22125	27636	32889	41403	45646
Granger-Ramanathan	22239	27687	32966	41253	45234
Constrained Granger-Ramanathan	22065	27522	32963	41385	45576
Mallows Model Averaging	22226	27691	32955	41313	45403
Ex post VaR	Realized Bipower Variation (RBP)				
Weighted BIC	21851	27429	33098	41294	45360
Smoothed AIC	22142	27580	32750	41177	44949
Bates-Granger	21940	27440	**32662**	41441	45373
Granger-Ramanathan	22015	27454	32765	**41163**	**44891**
Constrained Granger-Ramanathan	**21850**	**27307**	32795	41400	45317
Mallows Model Averaging	22060	27528	32757	41241	45043

This table consists of the mean of the conditional 99% Expected Shortfall of forecast combinations using the S&P 500 index. The data sample covers the period from January 1996 to August 2009 (excluding the estimation window). The out of sample forecasts are based on a rolling window of 1000 observations. The bold numbers show the lowest and highest mean ES for each confidence level.

Table 9: Evaluation of the 99% ES forecasts using the S&P 500 index.

Model	Bias*	MSE*	Rank	QLIKE	Rank	LINEX*	Rank
Normal GARCH	5.9347	0.1583	27	-2.2751	27	0.0798	27
Normal TARCH	7.2364	0.1746	29	-2.2642	29	0.0880	29
Normal EGARCH	7.6422	0.1838	30	-2.2608	30	0.0927	30
Normal APARCH	7.1126	0.1722	28	-2.2653	28	0.0868	28
t-GARCH	-1.8182	0.1034	11	-2.2953	8	0.0516	11
t-TARCH	1.2937	0.0940	7	-2.2957	7	0.0471	7
t-EGARCH	1.7732	0.0997	9	-2.2931	12	0.0500	9
t-APARCH	1.7267	0.0962	8	-2.2948	9	0.0483	8
FHS Normal GARCH	2.4861	0.1074	13	-2.2929	13	0.0539	13
FHS Normal TARCH	4.4207	0.1246	21	-2.2836	21	0.0627	21
FHS Normal EGARCH	5.0029	0.1309	26	-2.2801	26	0.0659	26
FHS Normal APARCH	4.5425	0.1252	22	-2.2828	22	0.0630	22
FHS t-GARCH	2.4679	0.1083	14	-2.2927	14	0.0544	14
FHS t-TARCH	3.8444	0.1161	17	-2.2867	17	0.0584	17
FHS t-EGARCH	4.6634	0.1260	24	-2.2817	24	0.0634	24
FHS t-APARCH	4.1244	0.1188	18	-2.2854	19	0.0597	18
EVT Normal GARCH	1.7477	0.1028	10	-2.2940	10	0.0516	10
EVT Normal TARCH	4.1274	0.1195	19	-2.2855	18	0.0601	19
EVT Normal EGARCH	5.0432	0.1297	25	-2.2811	25	0.0653	25
EVT Normal APARCH	4.3557	0.1226	20	-2.2842	20	0.0617	20
EVT t-GARCH	1.6619	0.1045	12	-2.2938	11	0.0524	12
EVT t-TARCH	3.7086	0.1125	15	-2.2885	15	0.0566	15
EVT t-EGARCH	4.6556	0.1255	23	-2.2823	23	0.0632	23
EVT t-APARCH	3.9242	0.1156	16	-2.2869	16	0.0581	16
Weighted BIC	-1.1587	0.0799	4	-2.3019	1	0.0398	4
Smoothed AIC	-1.0413	0.0810	5	-2.3009	6	0.0405	5
Bates-Granger	-1.3059	0.0787	2	-2.3018	3	0.0392	2
Granger-Ramanathan	-1.0274	0.0820	6	-2.3010	5	0.0409	6
Constrained Granger-Ramanathan	-1.2646	0.0783	1	-2.3018	2	0.0391	1
Mallows Model Averaging	-1.1052	0.0794	3	-2.3014	4	0.0396	3

This figure shows the Bias, MSE, QLIKE and LINEX (a=1) loss functions of the 99% ES forecasts using the S&P 500 Index from January 1996 to August 2009. The size of the rolling window is 1000 observations. Realized Bipower Variation is used as ex post high frequency volatility measure. * means that the actual numbers are multiplied by 1000.

Table 10: Maximum Regret of the 99% ES forecasts using the S&P 500 index.

Model	MSE*	Rank	QLIKE	Rank	LINEX*	Rank
Normal GARCH	1.1825	27	0.0343	27	0.6003	27
Normal TARCH	1.6722	29	0.0555	29	0.8488	29
Normal EGARCH	1.7423	30	0.0576	30	0.8840	30
Normal APARCH	1.6165	28	0.0530	28	0.8204	28
t-GARCH	0.3499	9	0.0097	7	0.1741	8
t-TARCH	0.3201	7	0.0098	8	0.1653	7
t-EGARCH	0.4147	15	0.0136	12	0.2109	15
t-APARCH	0.3664	10	0.0116	9	0.1868	10
FHS Normal GARCH	0.4239	16	0.0132	10	0.2147	16
FHS Normal TARCH	0.5537	25	0.0189	21	0.2817	25
FHS Normal EGARCH	0.5253	24	0.0218	26	0.2681	24
FHS Normal APARCH	0.4685	19	0.0191	22	0.2392	19
FHS t-GARCH	0.4630	17	0.0141	13	0.2343	17
FHS t-TARCH	0.5152	23	0.0175	19	0.2621	23
FHS t-EGARCH	0.4768	21	0.0201	24	0.2434	21
FHS t-APARCH	0.4041	14	0.0164	18	0.2065	14
EVT Normal GARCH	0.3923	13	0.0150	15	0.1986	13
EVT Normal TARCH	0.4700	20	0.0163	17	0.2398	20
EVT Normal EGARCH	0.5660	26	0.0208	25	0.2880	26
EVT Normal APARCH	0.4631	18	0.0177	20	0.2361	18
EVT t-GARCH	0.3915	12	0.0153	16	0.1984	12
EVT t-TARCH	0.3437	8	0.0134	11	0.1757	9
EVT t-EGARCH	0.5120	22	0.0195	23	0.2605	22
EVT t-APARCH	0.3792	11	0.0150	14	0.1935	11
Weight-ed BIC	0.0197	4	0.0005	3	0.0096	4
Smoothed AIC	0.0368	5	0.0013	5	0.0185	5
Bates-Granger	0.0151	2	0.0002	2	0.0075	2
Granger-Ramanathan	0.0579	6	0.0021	6	0.0289	6
Constrained Granger-Ramanathan	0.0023	1	0.0002	1	0.0012	1
Mallows Model Averaging	0.0189	3	0.0006	4	0.0095	3

This figure shows the Maximum Regret based on the MSE, QLIKE and LINEX (a=1) loss functions of the 99% ES forecasts using the S&P 500 index from January 1996 to August 2009. The size of the rolling window is 1000 observations. Realized Bipower Variation is used as *ex-post* high frequency volatility measure. * means that the actual numbers are multiplied by 10000.