DEPARTMENT OF ECONOMICS
UNIVERSITY OF CYPRUS

ROBUST VOLATILITY FORECASTS IN THE PRESENCE OF STRUCTURAL BREAKS

Elena Andreou, Eric Ghysels and Constantinos Kourouyiannis

Discussion Paper 08-2012

P.O. Box 20537, 1678 Nicosia, CYPRUS Tel.: +357-22893700, Fax: +357-22895028
Web site: http://www.econ.ucy.ac.cy

Robust volatility forecasts in the presence of structural breaks

Elena Andreou* **Eric Ghysels†** **Constantinos Kourouyiannis‡§**

12 May 2012

Abstract

Financial time series often undergo periods of structural change that yield biased estimates or forecasts of volatility and thereby risk management measures. We show that in the context of GARCH diffussion models ignoring structural breaks in the leverage coefficient and the constant can lead to biased and inefficient AR-RV and GARCH-type volatility estimates. Similarly, we find that volatility forecasts based on AR-RV and GARCH-type models that take into account structural breaks by estimating the parameters only in the post-break period, significantly outperform those that ignore them. Hence, we propose a Flexible Forecast Combination method that takes into account not only information from different volatility models, but from different subsamples as well. This methods consists of two main steps: First, it splits the estimation period in subsamples based on estimated structural breaks detected by a change-point test. Second, it forecasts volatility weighting information from all subsamples by minimizing a particular loss function, such as the Square Error and QLIKE. An empirical application using the S&P 500 Index shows that our approach performs better, especially in periods of high volatility, compared to a large set of individual volatility models and simple averaging methods as well as Forecast Combinations under Regime Switching.

Keywords: forecast combinations, volatility, structural breaks

JEL Classifications: C53, C52, C58

---

*Department of Economics, University of Cyprus, P.O. Box 20537, CY 1678 Nicosia, Cyprus, e-mail: elena.andreou@ucy.ac.cy

†University of North Carolina, Department of Economics, Gardner Hall CB 3305, Chapel Hill, NC 27599-3305, e-mail: eghysels@unc.edu.

‡Department of Economics and Economics Research Centre, University of Cyprus, P.O. Box 20537, CY 1678 Nicosia, Cyprus, e-mail: c.kourouyiannis@ucy.ac.cy.

§Acknowledgements: We would like to thank Francesco Audrino, David Banks and Alex Karagregoriou as well as the participants of the Computational and Financial Econometrics (CFE'10) conference in London and the 4th International Conference on Risk Analysis (ICRA4) in Limassol for insightful comments. This work falls under the Cyprus Research Promotion Foundation's Framework Programme for Research, Technological Development and Innovation 2008 (DESMI 2008), co-funded by the Republic of Cyprus and the European Regional Development Fund, and specifically under Grant PENEK/ENISX/0308/60. The first author also acknowledges support of the European Research Council under the European Community FP7/2007-2013 ERC grant 209116.

1 Indroduction

Empirical evidence of the existence of structural breaks in financial time series made this area of research very active in the recent years. A lot of attention in the literature has been given to structural breaks in volatility, which imply change in the risk behavior of investors due to important financial events, such as the 1987 stock market crash, the dot-com bubble in 1995-2000 and the subprime mortgage crisis. The implications of ignoring structural breaks in the accuracy of volatility estimators and forecasts make the use of flexible methods that take into account structural change very appealing.

Structural breaks in simple Autoregressive (AR) models are investigated by Pesaran and Timmermann (2004) who deal with the choice of the estimation window in forecasting in the presence of breaks using AR models. Andreou and Ghysels (2009) provide a review of the literature of structural breaks, examine the implications of structural breaks and discuss change-point tests of a single or multiple breaks. One of the most popular change-point tests is the CUSUM test, which has been used in the context of ARCH models by Kokoszka and Leipus (2000). Andreou and Ghysels (2004) have also used the CUSUM test to estimate structural breaks in volatility and they found that using high frequency volatility estimators, such as Realized Volatility, can improve the power properties of the test. In this paper, we split our estimation period in subsamples based on structural breaks detected using the CUSUM type test, and we weight predictions based on these subsamples to forecast volatility.

According to Timmermann (2006), given the difficulty of detecting structural breaks in "real" time and the different response of individual models to breaks, forecast combinations that take into account information from different models, can provide more accurate forecasts than individual models. When structural change is present, Aiolfi and Timmermann (2004) found that simple weighting schemes can outperform the best performing individual models. Hansen (2009) proposed a method of averaging estimators of a linear regression with a possible structural break by minimizing the Mallows criterion. In the forecasting literature, methods that consider structural breaks have been proposed by Guidolin and Timmermann (2009) and Elliot and Timmermann (2005). Both papers consider structural breaks as being generated by a Markov switching model. The main differences of our approach are that (1) we do not specify the process that generates the regime changes but instead test for them using a change-point test and (2) we do not restrict our framework to Square Error but we also use asymmetric loss functions, such as QLIKE.

The main contributions of this paper are two. First, we investigate in a simulation study the effect of structural breaks in the constant and the leverage coefficients of a GARCH diffusion model to the performance of alternative volatility estimators. In this paper we consider Realized Volatility (RV), AR-RV, HAR-RV, LHAR-RV, a number of GARCH-type models and Rolling volatility. We use two approaches, the full sample, which uses all information in the sample to estimate the parameters of the volatility model and ignores the break, and the split sample, which estimates the parameters in the pre- and post-break samples. Breaks affect all volatility models except those that use information of a particular day to estimate volatility (Realized Volatility) or do not involve parameter estimation (rolling window). The HAR-RV type models are also less sensitive to structural breaks compared to other models. We also investigate the effect of these breaks in forecasting volatility, and we find that we have significant gains in the accuracy of the predictions when we use the split sample method, which takes into account structural breaks. Second, we propose a Flexible Forecast Combination method, which involves two main steps: In the first step, we use a CUSUM-type test to detect structural breaks and we split the estimation period in smaller subsamples based on these breaks. In the second step, we forecast volatility taking into account information from different individual models and different subsamples by minimizing a particular loss function, such as the Square Error and QLIKE. Using a simulation design with a GARCH diffusion process with or without breaks in the constant parameter as well as an empirical application based on the S&P 500 Index, we find that this Flexible Forecast Combination approach outperforms a large number of individual models and simple averaging methods as well as Forecast Combinations under regime switching. This is especially evident in the subsample that includes the subprime mortgage crisis, where our method significantly outperforms all other methods based on the QLIKE loss function.

The paper is organized as follows: In Section 2 we estimate structural breaks in the volatility of the S&P 500 using a CUSUM type test. In Section 3 we describe the Flexible Forecast Combination proposed in this thesis. In Section 4 we investigate the effect of structural breaks in volatility estimates and forecasts in a simulation study. Additionally, we compare the predictive performance of individual models and forecast combination methods based on a GARCH diffusion DGP with and without breaks in the constant parameter. In Section 5 we illustrate our approach in an empirical application based on the S&P 500 index. Finally, in Section 6 we summarize our results and we conclude.

2 Structural Breaks in Realized Volatility

In this section we test for structural breaks in Realized Volatility of the S&P 500 Index daily returns using the CUSUM type test of Kokoszka and Leipus (1999, 2000). We split the sample in smaller subsamples based on these breaks and we estimate the parameters of the TARCH and EGARCH models in each subsample. The most significant parameter estimates are those of the EGARCH model and in particular, the constant, the GARCH and leverage effect coefficient estimates. However, the most important changes in the parameter estimates of consequent subsamples are observed in the constant and the leverage coefficient parameters.

For the estimation of structural breaks in the volatility of the S&P 500 returns, we use the CUSUM type test (Kokoszka and Leipus, 1999, 2000). The use of Realized Volatility instead of square or absolute returns is based on the findings of Andreou and Ghysels (2004), who show that the use of high frequency volatility estimators improves the power of the CUSUM-type statistics. Using the following process and the corresponding distribution under the null hypothesis of no breaks, we can test for structural breaks in Realized Volatility.

$$U_T(k) = \left(\frac{1}{\sqrt{T}} \sum_{j=1}^{k} RV_j - \frac{k}{T\sqrt{T}} \sum_{j=1}^{T} RV_j \right) \stackrel{H_0}{\rightarrow} \sigma B(k) \tag{1}$$

where T is the sample size, RV_j is Realized Volatility based on 5 minute returns, $B(k)$ a Brownian bridge and $\sigma^2 = \sum_{j=-\infty}^{\infty} Cov(RV_j, RV_0)$. We estimate σ^2 using the Heteroskedasticity and Autocorrelation Consistent (HAC) Covariance estimator of Andrews (1991). The change-point of the break is given by the following CUSUM-type estimator:

$$\widehat{k} = \min \left\{ k : |U_T(k)| = \max_{1 \leq j \leq T} |U_T(k)| \right\} \tag{2}$$

First, we use the full sample of the S&P 500 Index that covers the period from February 3, 1986 to June 30, 2010 and we estimate a structural break in volatility on June 21, 1998 with a test statistic equal to 3.87, which indicates that the break is significant for 1% confidence level. This break is associated with the rapid increase of stock prices due to the substantial growth in the Internet sector, the well-known "dot-com bubble", which covers the period roughly from 1995 to 2000. Then we proceed by splitting the sample is smaller subsamples based on new estimated structural breaks[1].

Figure 1.1 shows the structural breaks of Realized Volatility based on the aforementioned procedure. Based on these breaks the initial sample is divided in smaller subsamples, which are characterized by low, high and extremely high volatility (the latter corresponds to a crisis period). The most interesting subsamples are the crisis subsamples that are characterized by very high volatility and extreme events. The first crisis subsample covers the period from February 3, 1986 to April 25, 1988, when the 1987 stock market crash took place and the second from January 4, 2008 to June 30, 2010, a subsample that is associated with the subprime mortgage crisis, which started with the drop in the housing prices in the US and peaked with the bankruptcy of a number of financial institutions (e.g. Lehman Brothers). Another interesting period which is characterized by high volatility is the subsample from July 21, 1998 to April 10, 2003, which covers a number of

[1]We stop when either the CUSUM test does not detect any other breaks or when the subsample becomes small (with less than 500 observations).

financial events, such as the rapid growth of the Internet sector, the accounting scandal of Enron and the terrorist attacks on September 11, 2001, which caused the destruction of the World Trade Center in New York.

Table 1 shows the procedure with the successive tests and the corresponding test statistics and break dates. All the breaks are significant at 1% confidence level. Tables 2 and 3 show the parameter estimates of the Normal TARCH(1,1) and Normal EGARCH(1,1) models, respectively, for each subsample and the full sample with the corresponding Bollerslev - Wooldridge standard errors. The parameters of the EGARCH model are more significant compared to the TARCH model given the logarithmic structure and the ability of the first to provide positive volatility predictions without any constraints in the parameters. For both models the most significant is the GARCH parameter that controls the persistence, followed by the leverage parameter and the constant. The ARCH parameters of the TARCH model are insignificant for all subsamples and the full sample. Even though the GARCH parameters are the most significant, there are not large changes in these parameters in the various subsamples. On the other hand, the changes in the leverage effect parameter and the constant are more noticeable. For example the constant parameter of the TARCH model becomes more than 5 times smaller from the high volatility subsample that includes the "dot-com bubble" and the LTCM crisis (July 21, 1998 - April 10, 2003) to the low volatility subsample before the subprime mortgage crisis (April 11, 2003 to January 3, 2008). Another example is the break in the leverage effect parameter, which becomes around 3 times smaller from the Stock market crash of 1987 subsample (February 3, 1986 - April 25, 1988) to the high volatility subsample which covers the period from April 26, 1988 to February 6, 1992. Motivated by the breaks in the volatility of the S&P 500 Index returns, we consider breaks in the constant and the leverage effect parameters of sizes 2 and 3 for the simulation design of this paper.

3 Methodology

In this section we describe a novel method of predicting volatility, the Flexible Forecast Combination (FFC). This method uses information from different models and subsamples and provides volatility predictions that are robust to the model uncertainty of the volatility model. The model space of the FFC approach includes *ex-ante* forecasts given by Autoregressive models of Realized Volatility, GARCH-type models as well as Rolling Volatility models.

3.1 Realized Volatility

We assume each daily interval is divided into m periods of length $\Delta = \frac{1}{m}$. Therefore, Δ period returns are given by $r_{t,j} = \log S_{t+j\Delta} - \log S_{t-1+(j-1)\Delta}$, $j = 1, ..., m$ and the daily returns by

$r_t = \sum_{j=1}^{m} r_{t,j}$. Quadratic Variation (QV) is given by the sum of Integrated Volatility and a jump factor that is equal to the sum of square jumps.

$$QV_t = \int_t^{t+1} \sigma_s^2 ds + \sum_{t<s<t+1, dq(s)=1} \kappa^2(s) \tag{3}$$

where $dq(t)$ is a counting process which takes the value of 1 when there is a break at time t and 0 otherwise and $\kappa(t)$ is the size of the realized jump. Given that there are no jumps in the price process of this simulation design, Quadratic Variation coincides with Integrated Volatility.

Given that Quadratic Variation is a latent variable, we use an *ex-post* estimator, namely Realized Volatility (RV) as a proxy, which is discussed extensively in Andersen, Bollerslev, Diebold and Ebens (2001), Andersen, Bollerslev, Diebold and Labys (2001), Bardorff-Nielsen and Shephard (2002) and Meddahi (2002). RV is given by the sum of square intra-daily returns and uses information only from a particular day:

$$RV_t = \sum_{j=1}^{m} r_{t,j}^2 \tag{4}$$

As we increase the number of daily intervals (m) by using finer intervals we get more accurate volatility estimates, since volatility occurs in continuous time. However, we use 5 minute returns to avoid microstructure noise which exists when we use data at higher frequencies (see Andersen, Bollerslev, Diebold and Labys, 2001).

3.2 Model space

First, we create the model space which includes forecasts from 17 different volatility models that are classified in four broad categories: (1) Autoregressive models of Realized Volatility (AR-RV), (2) Heterogeneous Autoregressive models of Realized Volatility (HAR-RV), (3) Parametric GARCH-type models and (4) Rolling Volatility models. The volatility forecasts h_{t+1} are obtained using a rolling window of intra-daily returns included in n trading days, i.e. $r_{t-n+1,1}, r_{t-n+1,2}, ..., r_{t-n+1,m}, ..., r_{t,1}, ..., r_{t,m}$, where the first index of intra-daily returns corresponds to the day and the second to the time within the day and takes values in the interval $[1, m]$. Based on these intra-daily returns we obtain the daily returns $r_{t-n+1}, ..., r_t$ and Realized Volatilities $RV_{t-n+1}, ..., RV_t$ that are used for the GARCH-type and AR-RV forecasts, respectively. Then we move the rolling window one day and use intra-daily returns $r_{t-n+2,1}, r_{t-n+2,2}, ..., r_{t-n+2,m}, ..., r_{t+1,1}, ..., r_{t+1,m}$ to obtain the volatility prediction h_{t+2}. The next section describes the method used by the FFC approach to combine these volatility forecasts.

The first category of models includes Autoregressive models of Realized volatility (AR(p)-RV) with 1, 5, 10 and 15 lags.

$$RV_t = \omega + \beta_1 RV_{t-1} + \beta_2 RV_{t-2} + ... + \beta_p RV_{t-p} + \varepsilon_t$$

These models take into account intra-daily information for the estimation of RV. They use a simple autoregressive model to capture the dependence in RV and provide predictions that quickly adapt to changes in volatility.

The second category of models consists of Heterogeneous Autoregressive models of Realized Volatility. The HAR-RV model proposed by Corsi (2009), instead of estimating the coefficient estimates of each lag of the AR process, it uses lags of Realized Volatility at daily, weekly and monthly aggregated periods and captures some well known features of financial returns such as long memory and fat tails.

$$RV_t = \omega + \beta_d RV_{t-1}^{(d)} + \beta_w RV_{t-1}^{(w)} + \beta_m RV_{t-1}^{(m)} + \varepsilon_t \tag{5}$$

where $RV_{t-1}^{(d)} = RV_{t-1}$, $RV_{t-1}^{(w)} = \frac{1}{5}\left(RV_{t-1} + RV_{t-2} + ... + RV_{t-5}\right)$ and $RV_{t-1}^{(m)} = \frac{1}{22}\left(RV_{t-1} + RV_{t-2} + ... + RV_{t-22}\right)$. This model can be extended to the Leverage Heterogeneous Autoregressive (LHAR-RV) model (Corsi and Reno, 2009) that takes into account the leverage effect of daily, weekly and monthly returns.

$$RV_t = \omega + \beta_d RV_{t-1}^{(d)} + \beta_w RV_{t-1}^{(w)} + \beta_m RV_{t-1}^{(m)} + \gamma_d r_{t-1}^{(d)-} + \gamma_w r_{t-1}^{(w)-} + \gamma_m r_{t-1}^{(m)-} + \varepsilon_t \tag{6}$$

where $r_{t-1}^{(d)-} = r_{t-1} I\left\{r_{t-1} < 0\right\}$, $r_{t-1}^{(w)-} = \frac{1}{5}\left(r_{t-1} + r_{t-2} + ... + r_{t-5}\right) I\left\{r_{t-1} + r_{t-2} + ... + r_{t-5} < 0\right\}$ and $r_{t-1}^{(m)-} = \frac{1}{22}\left(r_{t-1} + r_{t-2} + ... + r_{t-22}\right) I\left\{r_{t-1} + r_{t-2} + ... + r_{t-22} < 0\right\}$.

The third class of volatility models includes the GARCH (1,1)(Bollerslev,1986) and other extensions of the GARCH model that can capture the leverage effect, namely the TARCH (1,1) or GJR-GARCH (1,1) (Glosten, Jagannathan and Runkle, 1993), EGARCH (1,1,1) (Nelson, 1991) and APARCH(1,1,1) (Ding, Granger and Engle, 1993). The APARCH model does not restrict the power of returns and volatility to be equal to 2 and nests the GARCH and TARCH models. The EGARCH model has a logarithmic structure and therefore, gives positive values of volatility without imposing any restrictions to its parameters. We use two distributions for the innovations of the GARCH-type models, the normal and the t. In this class we also include the RiskMetrics (J.P. Morgan, 1996), which is a special case of the IGARCH(1,1) model with the constant restricted to be equal to zero. A more detailed description of these models as well as a comparison of their forecasting performance can be found in Hansen and Lunde (2005)[2].

The last class includes non-parametric volatility models using rolling windows of 30 and 60 daily observations. The advantage of these models is their non-parametric structure since they avoid imposing restrictive assumptions to daily returns. However, they fail to capture the rapid changes in volatility, especially in periods with numerous events that increase the volatility in the stock markets (e.g. during the subprime mortgage crisis in 2007-2010). The choice of the number

[2]We obtain the volatility forecasts of GARCH-type models using the Matlab codes developed by Kevin Sheppard and are described in the MFE MATLAB Function Reference (October 2009)

of daily observations in the rolling window is based on Patton (2011), Andreou and Ghysels (2002) and on the findings of Foster and Nelson (1996) who estimated the optimal window size.[3]

3.3 The Flexible Forecast Combination approach

The Flexible Forecast Combination (FFC) approach consists of two main steps: In the first step, structural breaks in Realized Volatility are estimated and the sample is divided in smaller subsamples using the estimated breaks detected by the change-point test as described in the previous section. In the second step, information from different models and subsamples is weighted to provide robust volatility forecasts.

First, the FFC approach splits the sample in smaller subsamples based on the estimated breaks in Realized Volatility using the CUSUM-type test and following the procedure discussed in Section 2. In each subsample the FFC approach makes inference form different models and thus, overcomes the misspecification of model uncertainty in the volatility model. The FFC approach also weights the information from different subsamples and therefore, can provide accurate volatility forecasts irrespective of the characteristics of the out-of-sample period, i.e. whether it is a low volatility or high volatility or crisis period.

In each subsample the combination weights are estimated by minimizing the distance of RV and the combined forecast. The model space includes volatility forecasts given by the 17 individual models using a rolling window approach as described in Section 3.2.

$$\mathbf{w}_i = \underset{\mathbf{w}_i \in H}{\arg\min} \frac{1}{T_i} \sum_{t \in A_i} L\left(RV_t, \mathbf{h}_t' \mathbf{w}_i\right) \tag{7}$$

where $\mathbf{w}_i' = [w_{i1}, ..., w_{im}]$ is the vector of combination weights of subsample i, $\mathbf{h}_t' = [h_{t1}, ..., h_{tm}]$ is the vector of individual volatility forecasts, m the number of individual forecasts in the model space, T_i is the size of the subsample i, A_i is the set with the indices of daily observations that belong to subsample i, RV_t is Realized Volatility based on 5 minutes returns, $L(.)$ is the loss function and the set $H = \left\{w_{ij} : 0 \leq w_{ij} \leq 1, \sum_{j=1}^{m} w_{ij} = 1\right\}$ corresponds to the weights that are restricted in the interior of the unit interval and to add up to 1. Thus, the FFC approach instead of relying to an individual model, it makes inference from various models with different characteristics and properties. Except from the diversification gains, the FFC approach has also the advantage of choosing the combination weights by minimizing the distance of RV and the combined forecast, which improves its performance compared to other simple averaging methods.

We consider the Square Error (SE) and the QLIKE loss functions, which according to Patton (2011) are robust to the noise of the volatility proxy in the sense that the rankings of competing

[3] Using data of the S&P 500 from January, 1928 to December, 1990, they found that the optimal rolling window size is 53 observations.

volatility forecasts are not distorted from the use of a conditionally unbiased estimator instead of the true conditional volatility. The first is symmetric whereas the latter gives more penalty to positive forecast errors which are essentially more important in risk management, since they are associated with under-estimation of risk.

$$SE: \; L\left(RV, \mathbf{h}\right) = \left(RV - \mathbf{h}\right)^2 \tag{8}$$

$$QLIKE: L\left(RV, \mathbf{h}\right) = \frac{RV}{\mathbf{h}} - \log\left(\frac{RV}{\mathbf{h}}\right) - 1 \tag{9}$$

Patton (2011) proposed a class of loss functions that are robust to the noise of the volatility proxy and they are homogeneous of degree $b+2$:

$$L\left(RV, h; b\right) = \begin{cases} \frac{1}{(b+1)(b+2)}\left(RV^{b+2} - h^{b+2}\right) - \frac{1}{b+1}h^{b+1}\left(RV - h^{b+1}\right)\left(RV - h\right) & for\ b \notin \{-1, -2\} \\ h - RV + RV\log\left(\frac{RV}{h}\right) & for\ b = -1 \\ \frac{RV}{h} - \log\left(\frac{RV}{h}\right) - 1 & for\ b = -2 \end{cases}$$

SE and QLIKE loss functions belong to this class for $b = 0$ and $b = -2$, respectively. In this paper we also consider the Homogeneous Robust loss function for $b = -1$, which also penalizes under-prediction of volatility more heavily but the degree of asymmetry is smaller compared to the QLIKE.

Once we obtain the weights of each subsample we construct n series of combined volatility forecasts.

$$h^c_{t,i} = \mathbf{h}_t'\mathbf{w}_i, \;\; t = 1, ...T \tag{10}$$

where $h^c_{t,i}$ is the volatility forecast at time t using the weights estimated in the subsample i and T the sample size of the estimation period, n is the number of subsamples included in the estimation period. Each series of combined volatility forecasts $h^c_{t,i}$ uses the weights from a particular subsample, which can be either a low volatility or high volatility or crisis subsample to predict volatility. So instead of using the weights of one subsample with some specific characteristics to predict volatility, we create n series of combined volatility predictions with weights that correspond to n different subsamples.

Next, we estimate some new combination weights which we denote by $\widetilde{\mathbf{w}}_i$ and are used to weight the volatility forecasts obtained by using information from different subsamples. Therefore, in this stage we combine forecasts that use information from different subsamples, whereas in the previous stage we combined forecasts based on different models in a particular subsample:

$$\widetilde{\mathbf{w}} = \arg\min_{\widetilde{\mathbf{w}}} \frac{1}{T}\sum_{t=1}^{T} L\left(RV_t, \mathbf{h}_t^{c\prime}\widetilde{\mathbf{w}}\right) \tag{11}$$

where $\widetilde{\mathbf{w}}' = [\widetilde{\mathbf{w}}_1, ..., \widetilde{\mathbf{w}}_n]$ is the vector of weights of each subsample and $\mathbf{h}_t^{c\prime} = \left[\mathbf{h}^c_{t,1}, ..., \mathbf{h}^c_{t,n}\right]$ the vector of volatility predictions based on the estimated weights of each subsample.

To forecast volatility at time $t+1$ we first obtain n predictions based on the different subsamples:

$$h^c_{t+1,i} = \mathbf{h}'_{t+1,i}\mathbf{w}_i \tag{12}$$

where $h^c_{t+1,i}$ is the volatility forecast based on the weights of subsample i and $\mathbf{w}_i$ the weights defined in equation 7. Therefore, at this stage we use information from alternative volatility models and we obtain n volatility forecasts (one for each subsample). Next, we combine these predictions using the weights $\widetilde{\mathbf{w}}$ in equation 11 to obtain a volatility forecast that uses information from different subsamples as well:

$$h^{FFC}_{t+1} = {\mathbf{h}^c_{t+1}}'\widetilde{\mathbf{w}}$$

In order to capture new structural breaks that may occur in the out-of-sample period, we revise the above steps for every 252 daily observations, which correspond to one trading year. Table 19 shows the dates of the breaks detected by the FFC method for every time that we update the data including an additional trading year in the sample.

3.4 Other Forecast Combination Methods

In this paper we also consider Forecast Combinations under Regime Switching[4] (FC-RS) with two states proposed by Elliot and Timmermann (2005) and simple averaging methods, namely the Mean, Median and Geometric Mean [5]. The FC-RS method assumes that the combination weights are time varying and driven by regime switching and the states are generated by a first order Markov chain. This approach is very appealing when there are structural breaks in the data. Elliot and Timmermann (2005) show that FC-RS with two states outperform other forecast combination methods in predicting a number of macroeconomic variables, namely the unemployment rate, inflation and GDP growth. They also find that the FC-RS approach performs well for a number of DGPs, such as those with persistence regimes, a single structural break and a time varying parameter process. Simple averaging methods have also been found to perform well in forecasting macroeconomic variables (e.g. Stock and Watson, 2004) since these forecasts are not subject to estimation error.

4 Simulation study

The purpose of the simulation study is two-fold. First, we investigate the effect of structural breaks in the constant and the leverage coefficient of the GARCH diffusion model in the estimation

[4] We implement Forecast Combinations under Regime Switching using the Matlab code that was compiled by Perlin (2010).

[5] We use the same model space of 17 individual models described in section 3.2 for all forecast combination methods, the FFC, FC-RS, Mean, Median and Geometric Mean.

and forecasting of volatility based on alternative parametric and non-parametric models. For the estimation of volatility we use two different approaches, the full and the split sample. The full sample approach uses all information in the sample to estimate the parameters of the models and ignores structural breaks. The split sample approach estimates the parameters in the pre- and post- break samples. We also evaluate the forecasting performance of alternative volatility models based on the full and the split sample approaches and test whether there are significant gains in the accuracy of the volatility forecasts when we take into account structural breaks. Second, we investigate the predictive performance of alternative volatility forecasts given by individual models, simple averaging methods, the FFC and FC-RS methods. The DGP is a GARCH diffusion model with or without breaks in the constant. For both cases we find that the FFC approach outperforms other volatility forecasts irrespective of whether the out-of-sample evaluation period is a low volatility or a crisis period.

4.1 Effect of structural breaks in volatility estimates

The simulated DGP for this simulation exercise is a GARCH (1,1) diffusion model (Andersen and Bollerslev, 1998, Andersen, Bollerslev and Meddahi, 2005). We use 1000 replications and a sample size of 3000 daily observations. We assume that the market is open for 6 hours and 30 minutes, which is equal to the time that the New York Stock Exchange (NYSE) and NASDAQ operate every day[6]. The pre-sample period is 1000 daily observations.

The price process of the GARCH diffusion model is given by:

$$d\log S_t = \sigma_t \left[\rho_1 dW_{1t} + \sqrt{1-\rho_1^2} dW_{2t} \right] \tag{13}$$

where dW_{1t} and dW_{2t} are independent Brownian motions.

The dynamics of the volatility process of the GARCH(1,1) diffusion are described by

$$d\sigma_t^2 = a_1 \left(a_2 - \sigma_t^2 \right) dt + a_3 \sigma_t^2 dW_{1t} \tag{14}$$

Under the null there is no break in the parameters of the GARCH(1,1) diffusion model. We use the same parameter values as in Andersen and Bollerslev (1998), Andersen, Bollerslev and Meddahi (2005):

$$H_0 \ : No\ break,\ \rho_1 = -.576,\ a_1 = 0.035,\ a_2 = 0.636\ and\ a_3 = 0.35 \tag{15}$$

Under the alternative there is a change-point in the middle of the sample in the constant or the leverage coefficient of the GARCH diffusion model. We consider two different break sizes equal to

[6] The NYSE and the NASDAQ Stock Exchanges are open from 9.30a.m. to 4.00p.m.

an increase twice and three times the aforementioned parameters under the null (with the other parameters being constant) as listed below:

$$H_{1a} : Break\ in\ the\ constant\ of\ size\ 2,\ a_2 = \begin{cases} 0.4 & if\ t \leq 0.5T \\ 0.8 & otherwise \end{cases} \tag{16}$$

$$H_{1b} : Break\ in\ the\ constant\ of\ size\ 3,\ a_2 = \begin{cases} 0.3 & if\ t \leq 0.5T \\ 0.9 & otherwise \end{cases} \tag{17}$$

$$H_{1c} : \ Break\ in\ the\ leverage\ coefficient\ of\ size\ 2,\ \rho_1 = \begin{cases} -0.4 & if\ t \leq 0.5T \\ -0.8 & otherwise \end{cases} \tag{18}$$

$$H_{1d} : Break\ in\ the\ leverage\ coefficient\ of\ size\ 3,\ \rho_1 = \begin{cases} -0.3 & if\ t \leq 0.5T \\ -0.9 & otherwise \end{cases} \tag{19}$$

First, we evaluate the performance of alternative volatility models under the null hypothesis of no break and the four alternatives of breaks of sizes 2 and 3 in the leverage coefficient and the constant of a GARCH(1,1) diffusion volatility model. We compare the performance of these volatility estimates based on the full sample approach that ignores structural breaks and the split sample approach that takes into account the break.

We simulate $T = 3000$ "daily" returns based on a GARCH diffusion volatility model with a break (alternative hypothesis) in the middle of the sample (at time $0.5T = 1500$) or without a break (null hypothesis). For the full sample approach we estimate the parameters of the volatility model using all the information available in the sample:

$$h_t^{full} = f\left(r_{1,1}, ..., r_{1,m}, r_{2,1}, ..., r_{2,m}, ..., r_{T,1}, ..., r_{T,m}, \theta\right) \tag{20}$$

where T is the total number of daily observations, m the number of intra-daily observations in one trading day and θ the parameters of each volatility model estimated using intra-daily returns from the full sample, i.e. $r_{t-n+1,1}, ..., r_{t,m}$. For the split sample approach we estimate the parameters in the pre- and post- break periods:

$$h_t^{split} = f\left(r_{1,1}, ..., r_{T,m}, \theta_{\mathbf{s_1}}, \theta_{\mathbf{s_2}}\right) \tag{21}$$

where $\theta_{\mathbf{s_1}}$ are the parameters of the volatility models estimated using the intra-daily returns of the pre-break subsample (s_1), i.e. $r_{1,1}, ..., r_{0.5T-1,m}$ and $\theta_{\mathbf{s_2}}$ the parameters estimated using the corresponding returns of the post-break subsample (s_2), i.e. $r_{0.5T,1}, ..., r_{T,m}$.

Tables 4 - 8 show the performance of alternative volatility models, based on the full and split sample approaches, in terms of Bias, Square Bias and MSE. Table 4 consists of the estimation results under the null hypothesis of no break and Tables 5 - 8 under the alternatives with breaks of sizes 2 and 3 in the leverage coefficient and the constant of the volatility process.

Under the null hypothesis, the best performing method in terms of MSE is the HAR-RV, followed by the LHAR-RV and the RV models. Next, we have the AR(p)-RV models with improvement in their performance as we increase the number of lags from 1 to 15. This indicates that there is useful information in the lags of the volatility, which is captured in a more parsimonious way by the HAR-RV and LHAR-RV models. The GARCH-type perform worse than the AR-RV type models, and finally the worst performing models are the RiskMetrics and the Rolling window models of 30 and 60 daily observations. Across GARCH-type models, those that take into account the leverage effect perform better, which is reasonable given the existence of the leverage coefficient in the DGP. However, the HAR-RV model performs better than the LHAR-RV despite the fact that only the latter captures the leverage effect. Given the GARCH-type structure of the DGP, the volatility models in the GARCH family gain more from the inclusion of parameters that control the leverage effect. Finally, the poor performance of the RiskMetrics and Rolling Volatility is expected given that they are inefficient estimates.

In terms of Bias, GARCH-type models perform better than the Autoregressive RV models (AR(p) and HAR). This is again due to the fact that the GARCH structure of the DGP yields an advantage to models in the GARCH family. RiskMetrics and rolling window models also perform well in terms of Bias, since they do not have any parameters to be estimated. Positive bias is observed in the Autoregressive-RV models as well as the EGARCH model, and negative bias is observed in the other GARCH-type models, RiskMetrics and rolling window. The opposite sign in the bias of the EGARCH and the other models in the GARCH family is due to the logarithmic structure of the EGARCH model, which is not nested (in contrast to the GARCH and TARCH) to the more general APARCH model.

Since there is no break in the DGP, there are small changes in the performance of the volatility models between the full and the split sample approaches in terms of MSE. In terms of Bias, the GARCH, TARCH and APARCH models perform better based on the full sample approach and the EGARCH performs better based on the split sample approach. This can be explained by the logarithmic structure of the EGARCH model, since it gives positive volatility estimates without any restriction in its parameters and hence, can give accurate estimates even in smaller samples. The other GARCH models require larger samples, since they impose restrictions in their parameters, which are necessary for positive variance.

Under the alternative hypotheses, there are no significant changes in the best performing methods in terms of MSE, since the HAR-RV outperforms all other volatility models. In terms of Bias, important changes in the rankings of volatility models based on the full and the split sample approaches are observed under the alternative hypothesis of a break in the constant. For example, when there is a break in the constant of the DGP of size 2, the rankings of the RiskMetrics, Rolling Volatility of 30 and 60 daily observations based on the full sample approach are 4, 5 and

3, respectively. Based on the split sample approach these rankings become 5, 9 and 8. When we increase the size of the break to 3, then these rankings for the full sample approach are 2, 3 and 1 and for the split sample approach 3, 9 and 6. The reason of these changes in the rankings of these models between the full and the split sample approaches is that they are not affected by the structural break in the constant, since there is no parameter estimation involved in these models and therefore, they outperform the GARCH-type models based on the full sample approach. For the split sample approach, the GARCH-type models take into account the break in the constant and consequently outperform the RiskMetrics and rolling window models.

Table 9 shows the ratios of the split and the full sample approaches for the null hypothesis of no break and the alternatives of break in the constant and the leverage coefficient. Realized Volatility is not affected by structural breaks since the volatility estimation is based on intra-daily data of a particular day. Similarly, the rolling window models are also robust to structural changes given the small sizes of their windows (30 and 60 daily observations). The RiskMetrics is also not affected by structural breaks in terms of MSE, since there are no parameters to be estimated. The only difference between the full and the split sample approaches for this model is the estimation of the initial value of volatility, which for the full sample approach is given by the sample variance of the first 100 observations used as estimation window (are not included in the 3000 observations of the full sample). On the other hand, the initial value of the variance of RiskMetrics for the split sample approach is updated for the post-break period and therefore, the split sample approach gives more accurate estimates when there is a break in the constant.

In terms of MSE, the benefits of taking into account structural breaks by using the split sample instead of the full sample approach are larger in the models of the GARCH-family compared to the AR-RV type models. When there is a break in the constant, there is improvement in the accuracy of the volatility estimates of all GARCH-type models of the split sample approach (compared to the full sample approach) and the gains are larger as we increase the size of the break. We have similar results for the AR-RV type models, except from the HAR-RV model, which is less sensitive to the break in the constant. The break in the leverage coefficient only affects the GARCH-type models with a leverage parameter, namely the TARCH, EGARCH and APARCH volatility models. As we increase the size of the break in the leverage coefficient there are more benefits in the accuracy of the volatility estimates of the split sample compared to the full sample approach.

4.2 Effect of structural breaks in volatility forecasts

In this section of the simulation study, we investigate the forecasting performance of alternative volatility models in the presence of structural breaks based on the full and the split sample approaches. We use the same DGP, i.e. GARCH diffusion with or without breaks in the constant and the leverage coefficient and the same model space (except from the Realized Volatility, which can be used only for estimation). We compare the predictive performance of the volatility models based on the full and the split sample approaches and test whether we have significant gains in the accuracy of our forecasts when we take into account structural breaks based on the Conditional Predictive Ability (CPA) test of Giacomini and White (2006).

As in the estimation exercise, we simulate a GARCH(1,1) diffusion process of sample $T = 3000$ with breaks in the constant of the volatility process and the leverage coefficient of the price process. We also use a period of 1000 observations for out-of-sample evaluation. The parameters of the GARCH(1,1) diffusion model in the out-of-sample period are the same with the post-break period. For the full sample approach we ignore the presence of a structural break in the sample and we estimate the parameters using all the observations of the full sample:

$$h_{t+1}^{full} = f\left(r_{1,1}, ..., r_{1,m}, ..., r_{t,1}, ..., r_{t,m}, \theta\right) \tag{22}$$

where $t > T$ and θ are the parameters of each volatility model estimated using intra-daily observations in the full sample, i.e. $r_{1,1}, ..., r_{T,m}$. For the split sample approach, we ignore the pre-break period and we use only the post-break period to estimate the parameters of the GARCH(1,1) diffusion model:

$$h_{t+1}^{split} = f\left(r_{0.5T,1}, ..., r_{0.5T,m}, ..., r_{t,1}, ..., r_{t,m}, \theta_{\mathbf{s_2}}\right) \tag{23}$$

where $t > T$ and $\theta_{\mathbf{s_2}}$ the parameters of each volatility model estimated using only returns form the post-break period, i.e. $r_{0.5T,1}, ..., r_{T,m}$.

For the comparison of the performance of volatility forecasts based on the full and the split sample approaches we use the Conditional Predictive Ability test proposed by Giacomini and White (2006). The null hypothesis of this test is given by:

$$H_0: \ E\left[L_{t+1}\left(IV_{t+1}, h_{t+1}^{full}\right) - L_{t+1}\left(IV_{t+1}, h_{t+1}^{split}\right)/\wp_t\right] = 0 \tag{24}$$

where IV_{t+1} is Integrated Volatility (which coincides with Quadratic Variation when there in no jump in the price process), h_{t+1}^{full} and h_{t+1}^{split} are the volatility forecasts given by the full and the split sample approaches, $L_{t+1}(.)$ is the loss function (in this paper we use Square Error) and $\wp_t$ the σ-algebra that consists of all the information from $\tau = 1, ...t$. When the null hypothesis is rejected we use the two step decision rule described in Giacomini and White (2006) to determine which approach gives the most accurate forecasts of volatility.

Tables 10 - 14 show the Bias and MSE of alternative volatility forecasts based on the full and the split sample approaches and the results of the CPA test under the null hypothesis of no break and the alternatives of breaks in the constant and the leverage coefficient. In terms of MSE, the AR-RV type models have the best rankings, followed by the GARCH-type models and the worst performing models are the rolling window models. In terms of Bias, we have similar results with the estimation exercise, since the GARCH-type models and rolling window have the best rankings. The first have an advantage compared to the other models because of the GARCH-type structure of the DGP.

Table 15 summarizes the results of the CPA test based on the Square Error loss function. The CPA test rejects the null hypothesis of equal predictive ability of the full and split sample approaches for the TARCH, EGARCH and APARCH when there is a break in the leverage coefficient of size 3. Based on the decision rule of Giacomini and White, volatility forecasts using the split sample approach perform better compared to those given by the full sample approach. When there is a break in the constant, the CPA test shows that the AR(p)-RV models based on the split sample approach significantly outperform the corresponding models based on the full sample approach. Furthermore, the two stage decision rule confirms the superior performance of the split sample compared to the full sample approach for all volatility models (except the HAR-RV type models) when there is a break in the constant of size 3. The way that the HAR-RV type models use information from different aggregation horizons is likely the reason that they give volatility forecasts that are less sensitive to structural breaks.

4.3 Comparison of the predictive performance of volatility forecasts

In this section we evaluate the performance of the FFC approach, as well as other forecast combination methods and individual models in predicting volatility in the presence of structural breaks. We have two simulation exercises: (1) In the first simulation exercise we consider the case that there is no break in the GARCH diffusion DGP and (2) in the second exercise the case with multiple breaks in the constant of the DGP based on the empirical findings of Section 2. Given that the effect of structural breaks in the performance of the individual volatility forecasts has already been investigated in the previous section, we give more emphasis to the performance of the FFC approach in various cases and we compare this approach with individual forecasts and other forecast combination methods. For the individual forecasts we take the most ideal case that the parameters are estimated in a period, where the GARCH diffusion DGP has the same parameters as in the out-of-sample evaluation period. Even under these circumstances, we find that the FFC approach has very good performance compared to individual forecasts.

The first simulation exercise is based on the GARCH diffusion DGP under the null hypothesis of no break (equations 13 and 14) of a total sample size of 4350 daily observations. This sample is divided as follows: (1)100 days used as pre-sample period, (2)750 days for the estimation of the parameters of the individual models, (3)3000 days as estimation period for the FFC and FC-RC and (4) 500 days for out-of-sample evaluation. Concerning the parameters of the GARCH diffusion DGP we consider two cases. In the first case we use the same parameters as in equation 15, which are the parameters also used by Andersen and Bollerslev (1998), Andersen, Bollerslev and Meddahi

(2005) and resemble a low volatility process. In the second case we give three times larger value to the constant (compared to the low volatility case), i.e. $a_2 = 1.908$ and simulate a process with increased volatility that resembles a crisis period. The intra-daily returns that correspond to the first 100 days, $r_{1,1}, ..., r_{T_0,m}$, $T_0 = 100$ are used as pre-sample period to avoid any possible bias in the initial observations of the GARCH diffusion DGP. The next 750 simulated daily returns and Realized Volatilities are used to estimate the parameters of the 17 individual models, i.e. $\theta = g(r_{T_0+1}, ..., r_{T_1}, RV_{T_0+1}, ..., RV_{T_1})$ or $\theta = g(r_{T_0+1,1}, ..., r_{T_1,m})$, $T_1 = 850$ and $t = T_0 + 1, ..., T_1$ given that both daily returns and Realized Volatilities are functions of intra-daily returns. Based on these parameter estimates we obtain forecasts of all individual models of size 3000 days, i.e. $h_t = f(r_{T_1+1,1}, ..., r_{T_2,m}, \theta)$, where $T_2 = 3850$ that will be used by the FFC and FC-RS approaches for estimating the combination weights. We also obtain volatility forecasts that correspond to another 500 observations, $h_t = f(r_{T_2+1,1}, ..., r_{T_3,m}, \theta)$, where $T_3 = 4350$ that will be used for out-of-sample evaluation of the individual models as well as to construct the simple averaging methods (Mean, Median and Geometric Mean). Given that these simulation exercises are computationally very intensive we use 250 simulations [7]. Table 16 shows the constant parameter of the GARCH diffusion process used in this simulation design.

In order to examine the performance of the FFC approach more thoroughly, we consider two different cases under the null hypothesis that there is no break in the DGP. In particular, we examine the cases where: (1) No breaks are detected by the FFC approach, (2) There is a misspecification in the first stage of the FFC approach and it detects a break in the middle of the sample that does not exist. As shown in Table 17, under the null hypothesis of no break in the GARCH diffusion DGP, the FFC approach outperforms all individual forecasts, simple averaging methods and Forecast Combinations under Regime Switching (FC-RS) for both cases, when the DGP resembles a low volatility and a crisis period. The success of the FFC approach is not only due to the diversification gains of taking into account information from different models, but also because of the way that estimates the combination weights by minimizing the distance between Realized Volatility and the combined forecast. Also the use of the Homogeneous Robust loss function is very appealing given that it is robust to the noise of the volatility proxy. Across individual forecasts, the LHAR-RV gives the most accurate predictions followed by the HAR-RV and the other AR-RV models. Additionally to all the advantages of the AR-RV models, the LHAR-RV is more parsimonious and also takes into account the leverage effect of returns at daily, weekly and monthly horizons. Regarding the other forecast combination methods the FC-RS outperforms the other simple averaging methods, but it is outperformed for all loss functions used in this paper (SE, HR b=-1 and QLIKE) and for both periods (low volatility and crisis) by the FFC approach and the best performing individual

[7] We examined the sensitivity of the simulation results to the number of simulations and we found that they are robust even for smaller number of simulations.

forecasts, LHAR-RV and HAR-RV.

The misspecification of detecting a break in the first stage of the FFC approach does not affect the performance of this approach since the difference in the loss is negligible. We can better understand this result by looking in Figures 1.2 and 1.3 which show the weights given to individual forecasts for the full sample when the FFC approach detects no break, and for the pre- and post-break samples when the FFC approach detects one break. The figures show that the weights are allocated with almost the same way across individual models in the full sample, pre- and post-break samples, which is expected since there is no break in the DGP. The LHAR-RV model that performs best across all individual forecasts has the largest weight for the three loss functions, followed by the HAR-RV and AR(1)-RV. Figure 1.8 also shows that the weights of the combination forecasts for the pre- and post-break samples are almost the same, which also can be explained by the fact that there is no break in the GARCH diffusion process.

In the second simulation exercise, we also simulate a GARCH diffusion DGP of a total number of daily observations 4350, but we split the estimation period in smaller subsamples based on breaks in the constant of the DGP. These subsamples resemble three different regimes, namely low volatility, high volatility and crisis periods. The low volatility and crisis subsmaples occur twice in the estimation period, whereas the high volatility subsample occurs only once. The simulation design is motivated by the structural breaks detected in the S&P 500 in Section 2. In particular, the high volatility subsample corresponds to the subsmaple of the S&P 500 that includes the dot-com bubble. The two simulated crisis subsamples correspond to the Stock Market Crash of 1987 and the subprime mortgage crisis. The constant parameter takes different values in the five subsamples so that the simulated process does not resemble a three state Markov Switching process. Table 16 shows the values of the constant parameter for each subsample.

We consider two different cases when the out-of-sample period resembles: (1) A low volatility period and (2) a crisis period. As in the previous simulation exercise with the GARCH diffusion without a break, the parameters of the DGP for the estimation period (the subsample used for the estimation of the parameters of the volatility models) is the same as the out-of-sample evaluation period. For the FFC approach we consider 5 different cases: (1) When the FFC approach ignores the presence of any level shifts in volatility and therefore, estimates the combination weights over the full sample of 3000 observations, denoted by FFC 1, (2) when the FFC approach estimates the breaks in the estimation period, denoted by FFC 2, (3) when the FFC approach ignores the low volatility subsamples and estimates the breaks and combination weights in the remaining estimation period, denoted by FFC 3, (4) when the FFC approach ignores the high volatility subsample and estimates the breaks and combination weights in the remaining estimation period, denoted by FFC 4 and (5) when the FFC approach ignores the crisis subsamples and estimates the breaks and combination weights in the remaining estimation period, denoted by FFC 5.

Table 18 shows the rankings of alternative volatility predictions given by individual models and forecast combinations for the GARCH diffusion DGP with multiple breaks. When the out-of-sample evaluation period is a low volatility regime, the best performing method is the FFC 5 that ignores the crisis subsamples for the three loss functions considered in this paper (Square Error, Homogeneous Robust for $b = -1$ and QLIKE). Similarly, when the out-of-sample evaluation period resembles a crisis period the best performing method based on the Homogeneous Robust for $b = -1$ and the QLIKE loss function is the FFC 3 approach that ignores the low volatility subsample (under the Square Error it ranks second). In both cases the FFC approach performs better when we ignore the subsamples that have different characteristics compared to the out-of-sample evaluation period. On the contrary, the FFC approach performs worse when subsamples with the same characteristics as the out-of-sample evaluation period are ignored. However, the differences in the performance of the FFC approach in the various cases is not significant and the FFC approach performs better than the other forecast combination methods and individual models almost for all cases. This shows an enormous robustness for the FFC approach.

Concerning the rankings of the other forecast combinations, the FC-RS outperforms all simple averaging methods, namely the Mean, Median and Geometric Mean for the case that the out-of-sample evaluation period is a low volatility period. On the contrary, when the out-of-sample period resembles a crisis period the FC-RS is outperformed by the three simple averaging methods. The FC-RS makes some restrictive assumptions that the states are generated by a first-order Markov chain and therefore, does not perform well when the out-of-sample period is characterized by high volatility, since it is outperformed even by simple averaging methods. On the contrary, the FFC approach is more flexible and thus performs well for both low volatility and crisis evaluation periods. Across individual models, LHAR-RV is the best performing model, followed by the HAR-RV and the other AR-RV models. The good performance of these models can be explained by the fact that they take into account the additional information in intra-daily returns and respond very fast to changes in volatility, with the LHAR-RV and HAR-RV being the most parsimonious models. The GARCH-type models make more restrictive assumptions and therefore, perform worse, and the rolling volatility models are the worst performing models given their slow respond to changes in volatility.

Figures 1.4 - 1.5 and 1.6 - 1.7 show the weights given to individual models by the FFC approach when the out-of-sample evaluation period resembles a low volatility and crisis period, respectively[8]. The LHAR-RV, which is the best performing model across the 17 individual models used in this paper has the highest weight. This result is more evident for the low volatility subsamples. When

[8]The FFC approach does not estimate the same breaks for all simulations. Thus, when we want to examine the weights given to the individual models and subsamples we split the estimation period in smaller subsamples based on the simulated breaks in the GARCH diffusion DGP without estimating them using a CUSUM-type test.

the subsamples are characterized by increased volatility and therefore predicting volatility is more challenging, the FFC approach gives more weight to other models such as the HAR-RV and the AR-RV models. The GARCH-type models do not provide any additional information to the FFC approach from the HAR-RV and AR-RV models and thus, receive very small weight. This result does not hold in the empirical application with the S&P 500 Index, where the DGP is more complicated and the FFC approach gives large weights to some GARCH-type models as well (e.g. EGARCH). The non-parametric models, namely the Rolling Volatility models of 30 and 60 daily observations receive some weight, although they are the worst performing models. This result indicates that the weights given to each model do not depend only on the performance of the models but on their structure as well which may enables them to provide some additional information to the FFC approach.

The characteristics of the out-of-sample evaluation period, namely whether it is a low volatility or crisis period, are not known to the FFC approach. Therefore, this approach does not necessarily gives the largest weight to the subsample with the same or similar characteristics with the out-of-sample evaluation period. For example, as shown in Table 1.9, FFC based on the Square Error loss function gives the most of the weight to the first subsample, which is a crisis subsample although the out-of-sample evaluation period is a low volatility subsample. Similarly, the FFC approach based on the QLIKE loss function gives the highest weight to the second subsample, which is a low volatility subsample, although the out-of-sample period corresponds to a crisis period. However, the combination weights of the FFC approach are allocated almost in all subsamples and therefore can provide accurate predictions of volatility irrespective of the characteristics of the out-of-sample period. This result is also evident by the rankings of the FFC approach as shown in Table 18. In Figures 1.4 - 1.7, we observe that the weights of individual models do not change significantly across the three loss functions used in this paper. This result does not hold for the combination weights of different subsamples since these weights behave differently according to the loss function which is used to minimize the distance between the volatility proxy and the combined forecast.

5 Empirical Application

The simulation results show that there are significant losses in the accuracy of volatility forecasts when we ignore structural breaks. These results also show that the FFC approach outperforms individual models, simple averaging methods and Forecast Combination under Regime Switching for various cases, e.g. based on a GARCH diffusion DGP with or without breaks in the constant parameter and for low volatility and crisis out-of-sample evaluation periods. In this section we use an empirical application based on the S&P 500 Index and we find that this method performs well, especially during the period of the subprime mortgage crisis, when other volatility models and forecast combination methods fail to predict volatility accurately.

5.1 Data

The data used in this paper is obtained from the Tick Data database. We use 5 minute returns of the S&P 500 Index for the estimation of Realized Volatility (RV) and the construction of volatility forecasts based on RV, namely AR-RV, HAR-RV and LHAR-RV. We also use daily returns to forecast volatility based on low frequency volatility estimators (GARCH-type and rolling window models). The sample covers the period from February 1, 1983 to June 30, 2010. Given that we use a rolling window of 750 observations for individual volatility models, we lose 3 years of daily returns and therefore, the initial date of the sample is February 3, 1986.

5.2 Combination weights of the FFC approach

Figures 1.10 - 1.12 show the weights ($\mathbf{w}$) given by the FFC approach to alternative volatility forecasts for different subsamples based on the SE, Robust (b=-1) and QLIKE loss functions. These figures correspond to the last time that the estimation period is updated with new data. In this case the estimation period spans February 3, 1986 to January 4, 2010 and the out-of-sample period spans January 5, 2010 to June 30, 2010. The weights behave differently across different subsamples and using alternative loss functions. In most of the cases the HAR-RV type models have the largest weight across all individual models. Normal EGARCH has the largest weight based on the SE loss function for the last subsample, something which is expected given the good performance of this model during the period of the subprime mortgage crisis.

Figures 1.13 - 1.15 show the weights ($\widetilde{\mathbf{w}}$) given to forecast combinations that make inference from different subsamples for the last 3 times that the estimation period of the FFC method is updated by one trading year. The out-of-sample period in Figures 1.13, 1.14 and 1.15 covers a big part of the subprime mortgage crisis period and spans January 4, 2008 - January 2, 2009, January 3, 2009 - January 4, 2010 and January 5, 2010 - June 30, 2010, respectively. The weights show the importance of the period that covers the 1987 stock market crisis (February 1986 to April 1991) in predicting volatility during the subprime mortgage crisis (January 2008 to June 2010). This is an important advantage for the FFC method, since it can make inference and predict volatility in high volatility subsamples using information from other subsamples with similar characteristics.

5.3 Performance of the FFC approach and comparison with other methods

For the evaluation of volatility predictions given by forecast combinations and individual models we use three alternative loss functions, which are three special cases of the Homogeneous Robust loss function, for $b = 0$ (Square Error), $b = -1$ and $b = -2$ (QLIKE). Based on these loss functions we estimate the losses by finding the distance between RV and the volatility forecast and report the corresponding rankings. We also compare the predictive performance of forecast combinations and individual models using the CPA test (Giacomini and White, 2006) based on the three aforementioned loss functions. The out-of-sample period of the S&P 500 is divided in two subsamples based on the structural break found in Section 2 on January 4, 2008. So the first subsample covers the period from January 2, 2004 to January 3, 2008 and is characterized by low volatility and the second subsample from January 4, 2008 to June 30, 2010 and it is a high volatility subsample, since it includes the events during the subprime mortgage crisis.

Tables 20 and 21 show the rankings of alternative volatility predictions given by individual models and forecast combinations for the two subsamples (low and high volatility) and the full sample, respectively. In the low volatility subsample the FFC SE method ranks second in terms of MSE. The group of the three best performing methods also includes the LHAR-RV model, which ranks first and the HAR-RV, which ranks third. Both HAR-RV and LHAR-RV forecast volatility by using information from the lags of RV in a parsimonious way, with the latter taking into account the leverage effect of daily, weekly and monthly returns as well. AR-RV type models perform better than GARCH-type models, since they use additional information from high frequency returns and make less restrictive assumptions. The RiskMetrics and rolling window models are the worst performing models. The simple averaging methods perform adequately well, and particularly the Geometric Mean ranks fourth followed by the FC-RS method, which ranks fifth. When we include asymmetries in the loss functions using the Robust loss function for b=-1 the group of the three best performing methods does not change. However, when we increase the asymmetry in the loss function by using the QLIKE loss function, the FFC QLIKE ranks first given the logarithmic structure and the good power properties of this loss function, which makes it very appealing in combining volatility forecasts. When there is asymmetry in the loss function, there is also a small improvement in the ranking of the FC-RS method since it outperforms the Geometric Mean and ranks fourth instead of fifth. However, there are no significance changes in the rankings of the other volatility forecasts as we increase the degree of asymmetry in the loss function.

This does not hold for the high volatility subsample, since the change in the value of the shape parameter (b) of the Homogeneous Robust loss function affects the rankings of the competing volatility forecasts. In particular, for $b = 0$ the three best performing methods are Normal EGARCH, FFC SE and Geometric Mean, for $b = -1$ the FFC Robust b=-1, LHAR-RV and HAR-RV and for $b = -2$ the FFC QLIKE, LHAR-RV and HAR-RV. For both subsamples, low and high volatility, the FFC approach is always in the two best performing methods for the three loss functions used in this paper. In particular, for the QLIKE loss function the FFC approach ranks first for both subsamples. This indicates that the asymmetries in the loss function improves the volatility prediction given by the FFC method compared to the other methods. This is consistent with the findings of Elliot and Timmermann(2003), who underline the importance of asymmetric loss functions in forecast combinations. The LHAR-RV and HAR-RV outperform all individual models for the high volatility subsample as well with only exception the Normal EGARCH under the SE loss function which ranks first. The Normal EGARCH has an advantage compared to the other GARCH-type models since given its logarithmic structure, it provides positive volatility forecasts without any restrictions in its parameters, something which is especially useful when volatility changes rapidly as in the high volatility subsample of our dataset. The FC-RS does not have the success in predicting volatility in the high volatility subsample as in forecasting macroeconomic variables (Elliot and Timmermann, 2005). In particular, the FC-RS is outperformed by a number of other methods such as the FFC, LHAR-RV, HAR-RV and simple averaging methods. Given the increased losses of all volatility forecasts in the high volatility subsample compared to the low volatility, the rankings of the full sample are similar to the high volatility subsample.

In Tables 22 - 30 there is a comparison of the forecast combination methods with the individual models that are included in their model space using the Conditional Predictive Ability (CPA) test (Giacomini and White, 2006) based on the Homogeneous Robust loss function for $b = 0$, $b = -1$ and $b = -2$. For the low volatility subsample of the S&P 500 (Tables 22 - 24) all forecast combination methods, namely the Mean, Median, Geometric Mean, FC-RS and FFC SE significantly outperform all GARCH-type, rolling window and the RiskMetrics models. Under the SE loss function, the FFC SE is the only approach that is not significantly outperformed by any individual model (it is outperformed only by LHAR-RV but not significantly). As we increase the degree of asymmetry of the Homogeneous Robust loss function for $b = -1$ and $b = -2$ the performance of the FFC improves even more since it outperforms significantly all individual models, including the AR-RV, HAR-RV and LHAR-RV models. The outstanding performance of the HAR-RV and LHAR-RV models is also shown in this table since both models significantly outperform all forecast combination methods except the FFC.

For the high volatility subsample (Tables 25 - 27) and the full sample (Tables 28 - 30) of the S&P 500, the FFC QLIKE also significantly outperforms all individual models based on the QLIKE loss function. The HAR-RV, LHAR-RV, AR(5)-RV, AR(10)-RV and AR(15)-RV significantly outperform all forecast combination methods except from the FFC QLIKE.

Tables 31 - 33 show how each forecast combination method is compared with the other forecast combination methods based on the CPA test. For the low volatility subsample the FFC method significantly outperforms all other methods under all three loss functions, SE, Homogeneous Robust for $b = -1$ and QLIKE. The FC-RS approach outperforms all simple averaging methods. Across the simple averaging methods, the Geometric Mean is the best performing method since it outperforms the Mean and the Median. For the high volatility subsample, the FFC SE significantly outperforms the Mean and FC-RS under the Square Error loss function. When we incorporate asymmetries in the loss function and use the Homogeneous Robust loss function for $b = -1$ and $b = -2$, the performance of the FFC improves since it significantly outperforms all other methods. The performance of the FC-RS for the high volatility subsample is worse, since it significantly outperforms only the Median, whereas in the low volatility subsample it outperforms all simple averaging methods. As in the low volatility subsample, the best performing method across the three simple averaging methods is the Geometric Mean.

In a nutshell, the FFC is the best performing method across all individual models and forecast combination methods when there are asymmetries in the loss function, i.e. based on the QLIKE loss function for both subsamples of the S&P 500. The FC-RS approach performs well only for the low volatility subsample and across simple averaging methods the best performing method is the Geometric Mean. The LHAR-RV and HAR-RV provide the most accurate volatility forecasts across a wide range of individual models. Furthermore, the aforementioned models outperform some forecast combination methods, such as the FC-RS, Mean, Median and Geometric Mean.

6 Conclusions

In this paper we investigate the effect of structural breaks in the constant and the leverage effect of a GARCH diffusion model in volatility estimates and forecasts given by a large set of individual models. Our model space includes AR-RV, HAR-RV, LHAR-RV, GARCH-type as well as rolling window models. We use two alternative approaches of estimating the parameters of the aforementioned models, the full sample that ignores structural breaks and the split sample which estimates the parameters in the pre- and post-break samples. We find significant gains in the accuracy of volatility estimates and forecasts of AR-RV and GARCH-type models when we take into account structural breaks in the parameters.

We propose a Flexible Forecast Combination (FFC) approach that predicts volatility using information from forecasts given by alternative models as well as different subsamples. The combination weights are estimated by minimizing the distance between RV and the combination forecasts, which is measured using the Homogeneous Robust loss function proposed by Patton (2011). Except from the property of giving rankings of competing volatility forecasts that are robust to the noise of the volatility proxy (e.g. RV), the Homogeneous Robust loss function is also appealing for combining volatility forecasts, especially when we incorporate asymmetries in the loss function. In particular, we find that the FFC approach based on the QLIKE loss function (a special case of the Homogeneous Robust loss function for $b = -2$) outperforms a large set of individual models as well as Forecast Combinations under Regime Switching and simple averaging methods. The structure of the FFC method captures structural breaks in volatility and therefore, it performs extremely well in the high volatility subsample of the S&P 500, which includes a number of events that are related to the subprime mortgage crisis, such as the Lehman Brothers bankruptcy. Another important advantage of this approach is that it can use information from previous subsamples to predict volatility in a period with similar characteristics. For example it can use information from the period of the 1987 stock market crash to predict volatility during the subprime mortgage crisis.

References

[1] Aiolfi, M., Timmermann, A., 2004, Structural Breaks and the Performance of Forecast Combinations, Working Paper.

[2] Andersen, T.G., T. Bollerslev, 1998, Answering the Sceptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts, *International Economic Review*, 39(4), 885-905.

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

[4] Andersen, T.G., T. Bollerslev, F.X. Diebold, P. Labys, 2003, Modeling and forecasting realized volatility, *Econometrica*, 71(2), 579-625.

[5] Andersen, T.G., T. Bollerslev, N. Meddahi, 2005, Correcting the errors: Volatility Forecast Evaluation Using High-Frequency Data and Realized Volatilities, *Econometrica*, 73, 279-296.

[6] Andreou, E. and E. Ghysels, 2002, Rollling-Sample Volatility Estimators: Some New Theoretical, Simulation, and Empirical Results, *Journal of Business and Economic Statistics*, 20(3), 363-377.

[7] Andreou, E. and E. Ghysels, 2004, The Impact of Sampling Frequency and Volatility Estimators on Change-Point Tests, *Journal of Financial Econometrics*, 2(2), 290-318.

[8] Andreou, E. and E. Ghysels, 2009, Structural Breaks in Financial Time Series, In: T.G.Anderson et.al. (Eds.), Handbook of Financial Time Series, Springer-Berlag Berlin Heidelberg, 839-870.

[9] Andrews, D. W. K., 1991, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation, *Econometrica*, 59(3), 817-858.

[10] Bardorff - Nielsen, O.E., Hansen, P.R., Lunde, A. and N. Shephard, 2008, Designing realized kernels to measure the ex-post variation of equity prices in the presence of noise, *Econometrica*, 76(6), 1481-1536.

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

[12] Bollerslev, T., 1986, Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, 31, 307-327.

[13] Corsi, F., 2009, A simple Approximate Long-Memory Model of Realized Volatility, *Journal of Financial Econometrics*, 1-23.

[14] Corsi, F., R. Reno, 2009, HAR volatility modelling with heterogeneous leverage and jumps, Working Paper.

[15] Ding, Z., W.J. Granger, R.F. Engle, 1993, A Long Memory Property of Stock Market Returns and a New Model, *Journal of Empirical Finance,* 1, 83-106.

[16] Elliot, G., A. Timmermann, 2004, Optimal forecast combination under general loss functions and forecast error distributions, *Journal of Econometrics*, 122, 47-79.

[17] Elliot, G., A. Timmermann, 2005, Optimal Forecast Combination under Regime Switching, *International Economic Review*, 46(4), 1081-1102.

[18] Foster, D.P. and D.B. Nelson, 1996, Continuous Record Asymptotics for Rolling Sample Variance Estimators, *Econometrica*, 64(1), 139-174.

[19] Giacomini, R. and H. White, 2006, Tests of Conditional Predictive Ability, *Econometrica*, 74(6), 1545-1578.

[20] Glosten, R., R. Jagannathan, D. Runkle, 1993, On the Relation Between Expected Value and the Volatility of the Nominal Excess Return on Stocks, *Journal of Finance*, 48, 1779-1801.

[21] Goncalves, S. and N. Meddahi, 2009, Bootstrapping Realized Volatility, *Econometric*a, 77(1), 283-306.

[22] Guidolin, M., A. Timmermann, 2009, Forecasts of US short-term interest rates: A flexible forecast combination approach, *Journal of Econometrics*, 150, 297-311.

[23] Hansen, B., 2009, Averaging Estimators for Regressions with a Possible Structural Break, *Econometric Theory*, 25, 1498 - 1514.

[24] J. P. Morgan,1996, *RiskMetrics Technical Documents* (Fourth Edition), New York, J. P. Morgan Co.

[25] Kokoszka, P., R. Leipus, 1999, Testing for Parameter Changes in ARCH Models, *Lithuanian Mathematical Journal*, 39, 231-247.

[26] Kokoszka, P., R. Leipus, 2000, Change-Point estimation in ARCH models, *Bernoulli*, 6(3), 513-539.

[27] Meddahi, N., 2002, A theoretical comparison between integrated and realized variances, *Econometrics Journal*, 6, 334-355.

[28] Nelson, D.B., 1991, Conditional Heteroskedasticity in Asset Returns: A New Approach, *Econometrica*, 59, 347-370.

[29] Patton, A. J., 2011, Volatility forecast comparison using imperfect volatility proxies, *Journal of Econometrics,* 160(1), 246-256.

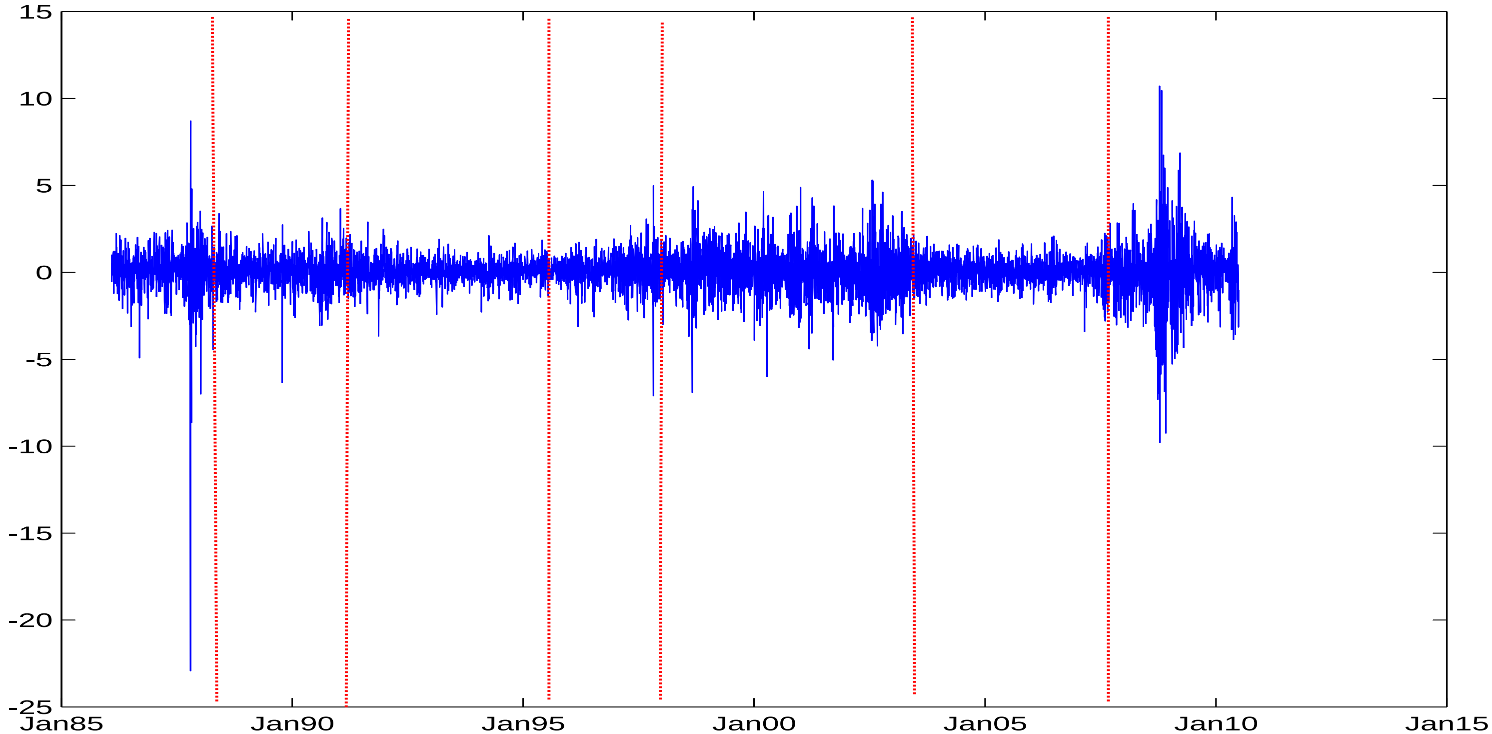
[30] Pesaran, M. H., A. Timmermann, 2005, Small sample properties of forecasts from autoregressive models under structural breaks, *Journal of Econometrics*, 129, 183-217.

[31] Stock, J.H., M.W. Watson, 2004, Combination Forecasts of Output Growth in a Seven-Country Data Set, *Journal of Forecasting*, 23, 405-430.

[32] Timmermann, A., 2006, Forecast Combinations, In: Elliot, G., Granger, C.W.J., Timmermann, A. (Eds.), Handbook of Economic Forecasting, Elsevier, Amsterdam in press.

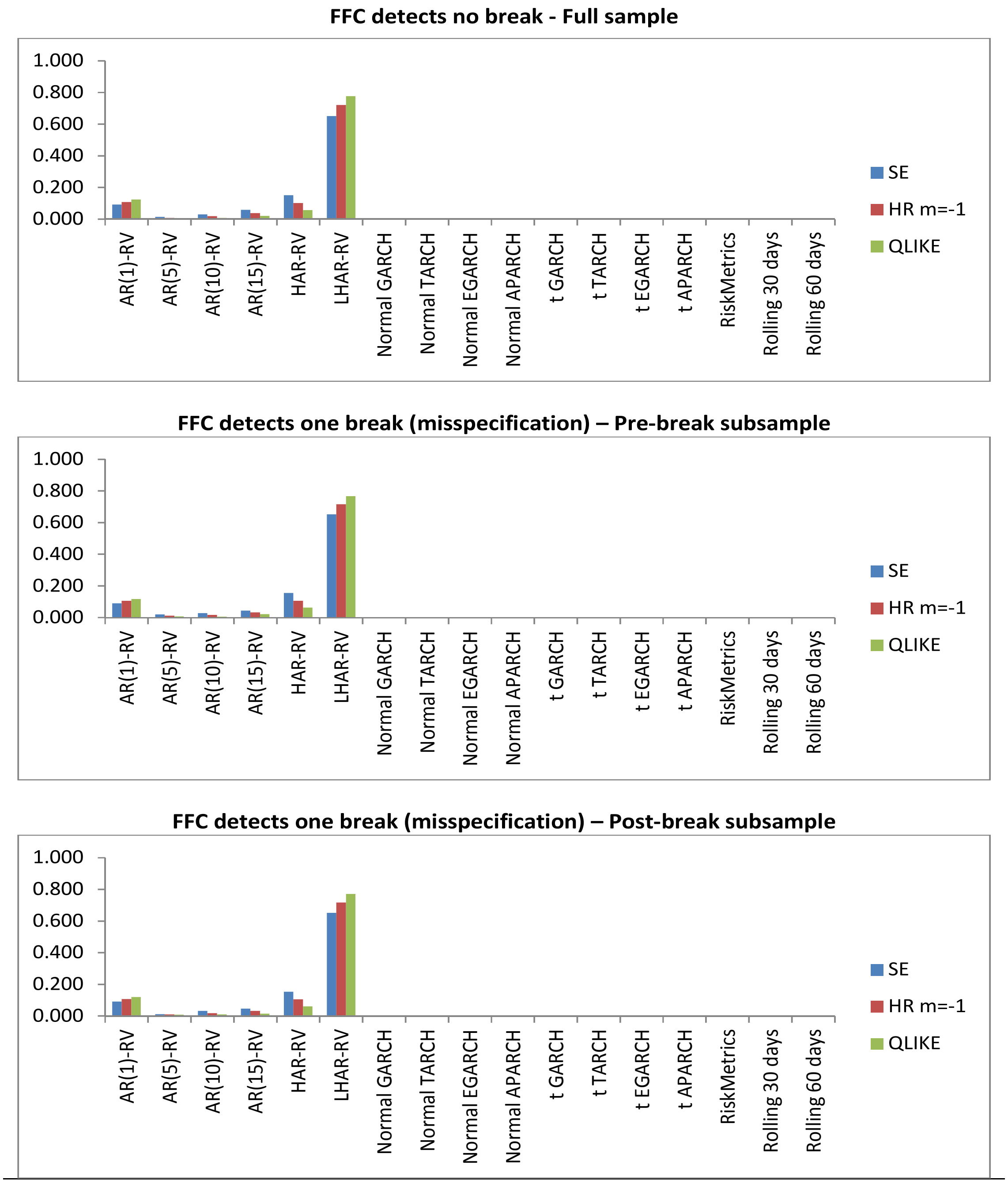
A Appendix

Figure 1.1: Structural breaks in Realized Volatility of the S&P 500 Index



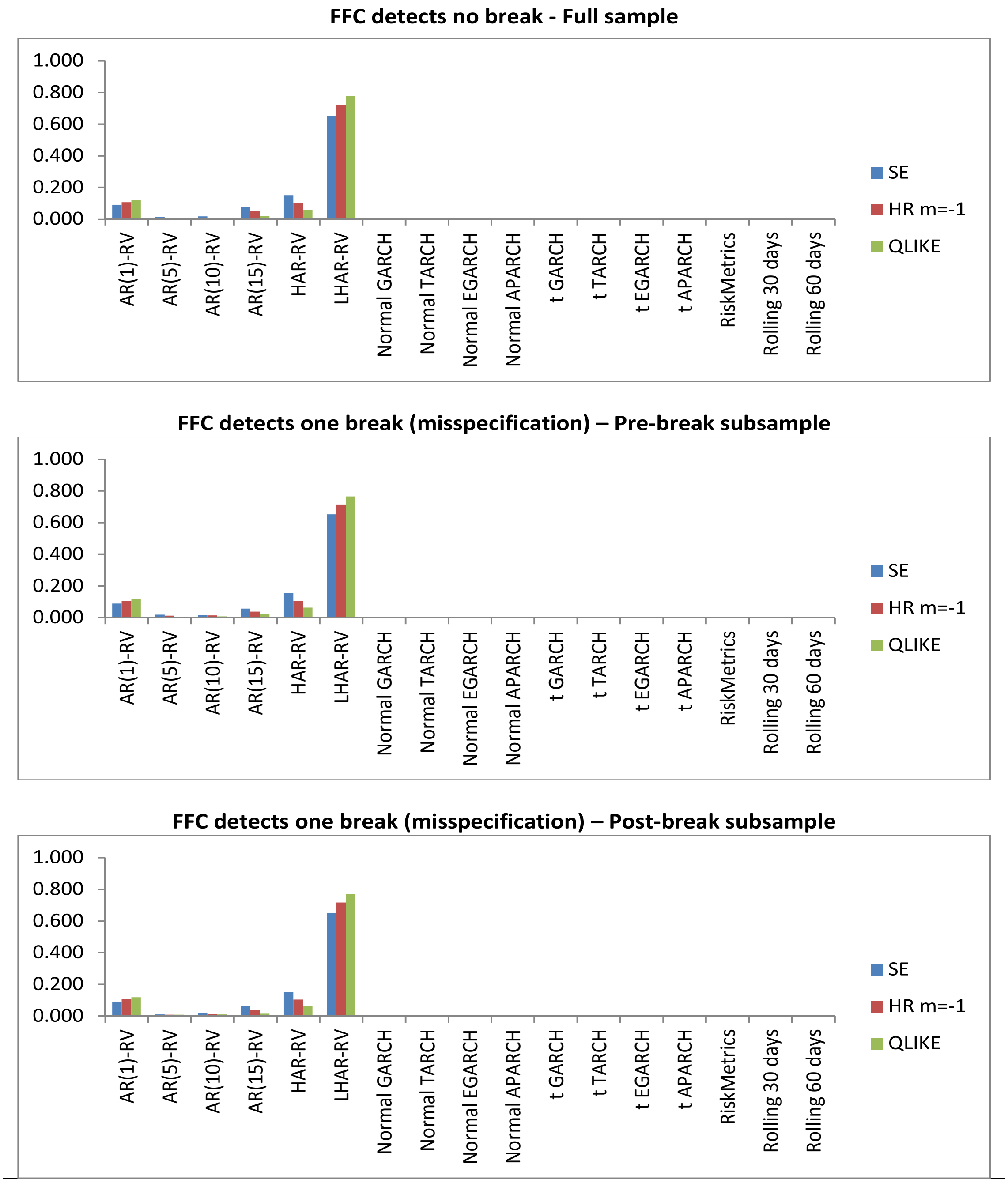
This figure shows the structural breaks in Realized Volatility of the S&P 500 Index from February 3, 1986 to June 30, 2010, estimated using the CUSUM type test of Kokoszka and Leipus (1999, 2000) test based on the HAC variance estimator (Andrews, 1991).

Figure 1.2: Weights of individual volatility forecasts of the FFC approach when the DGP is a GARCH diffusion process without breaks that resembles a low volatility period



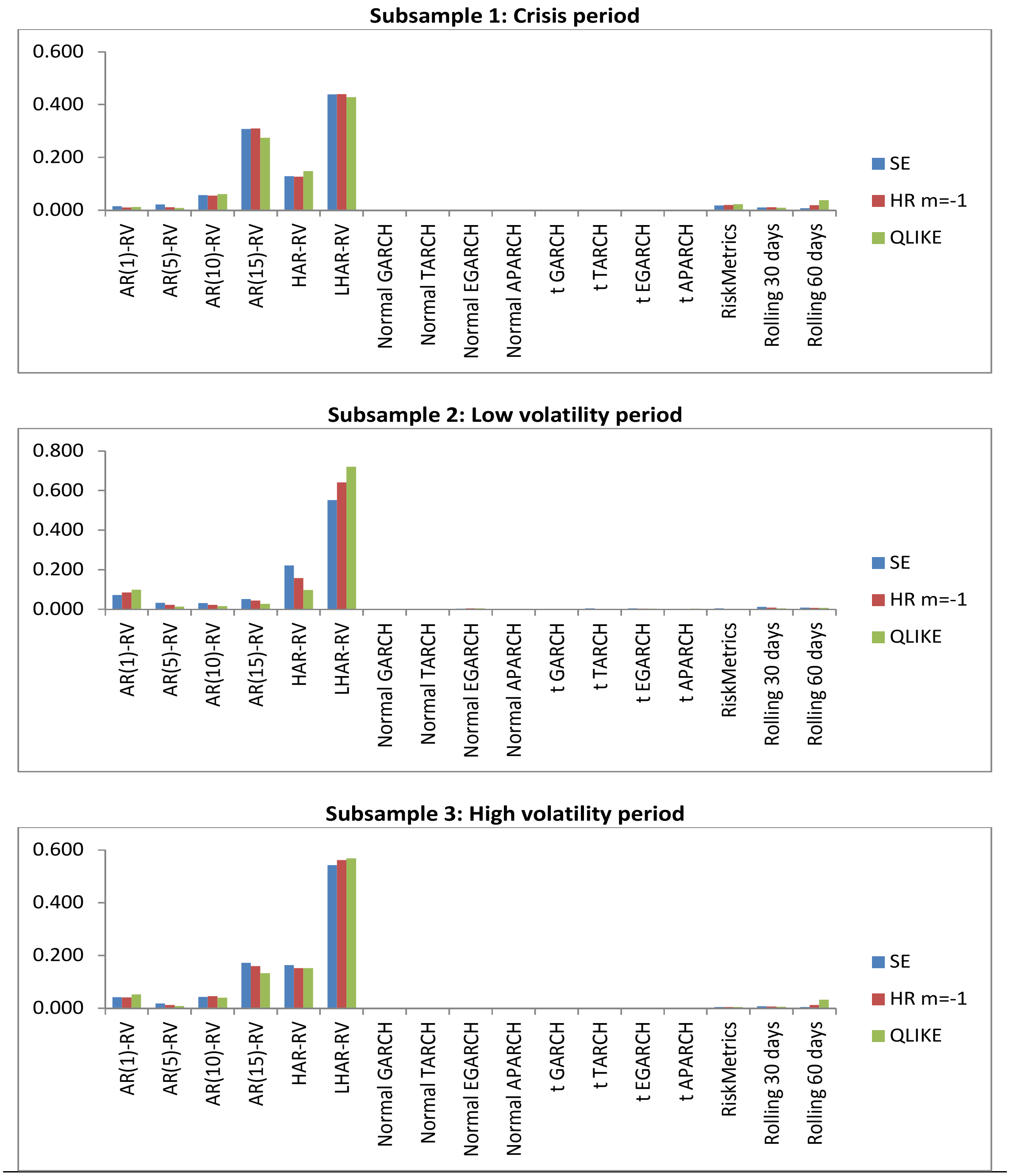
This figure shows the weights of individual forecasts for the full sample when the FFC detects no break in the DGP, for the pre- and post- breaks subsamples when there is a misspecification and the FFC detects a break in the middle of the sample. The DGP is a GARCH diffusion process without a break that resembles a low volatility period. The sample size is 3000 observations and the weights are based on 250 simulations.

Figure 1.3: Weights of individual volatility forecasts of the FFC approach when the DGP is a GARCH diffusion process without a break that resembles a crisis period



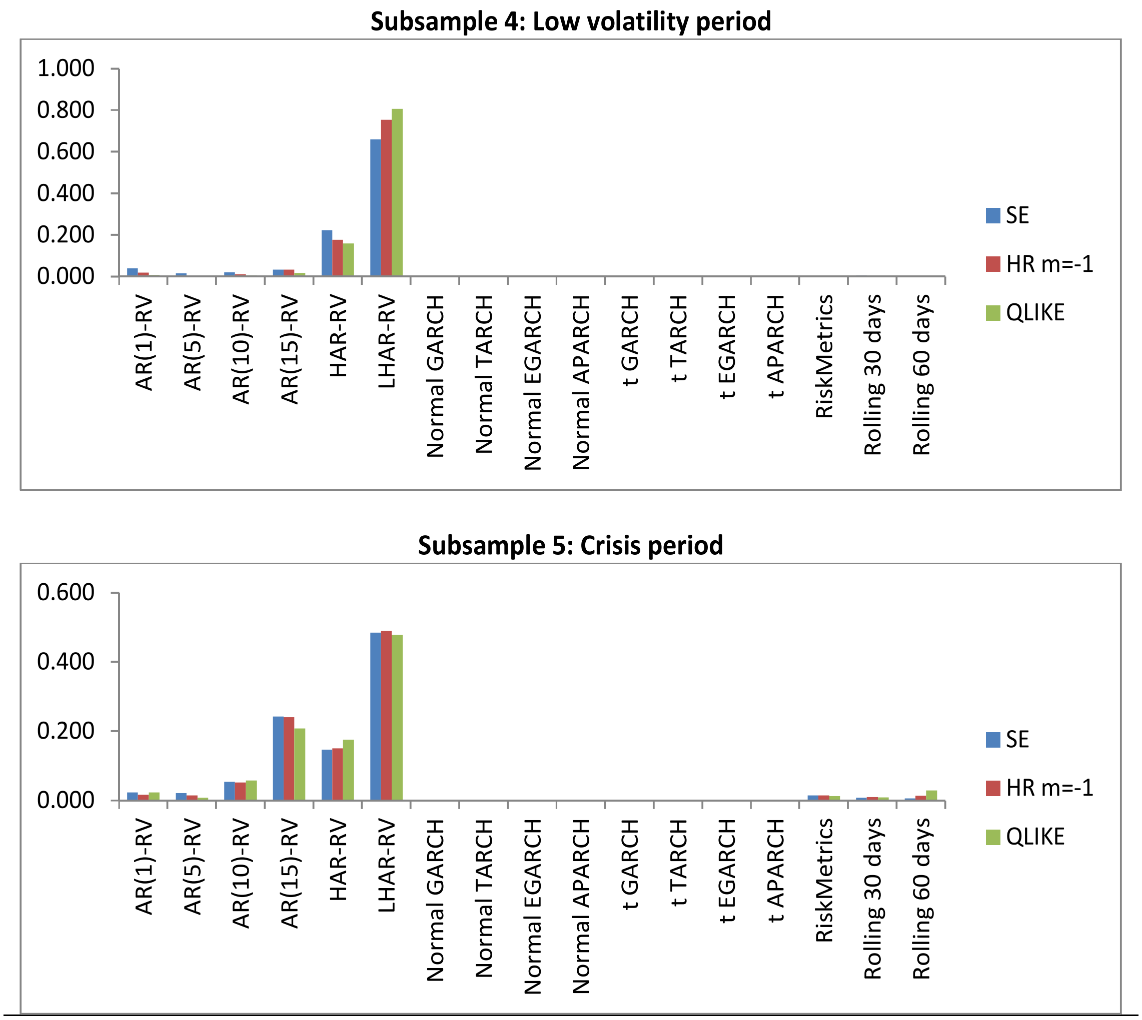
This figure shows the weights of individual forecasts for the full sample when the FFC detects no break in the DGP, for the pre- and post- breaks subsamples when there is a misspecification and the FFC detects a break in the middle of the sample. The DGP is a GARCH diffusion process without a break that resembles a crisis period. The sample size is 3000 observations and the weights are based on 250 simulations.

Figure 1.4: Weights of individual volatility forecasts of the FFC approach when the DGP is a GARCH diffusion process with multiple breaks and the out-of-sample evaluation period resembles a low volatility period (1)



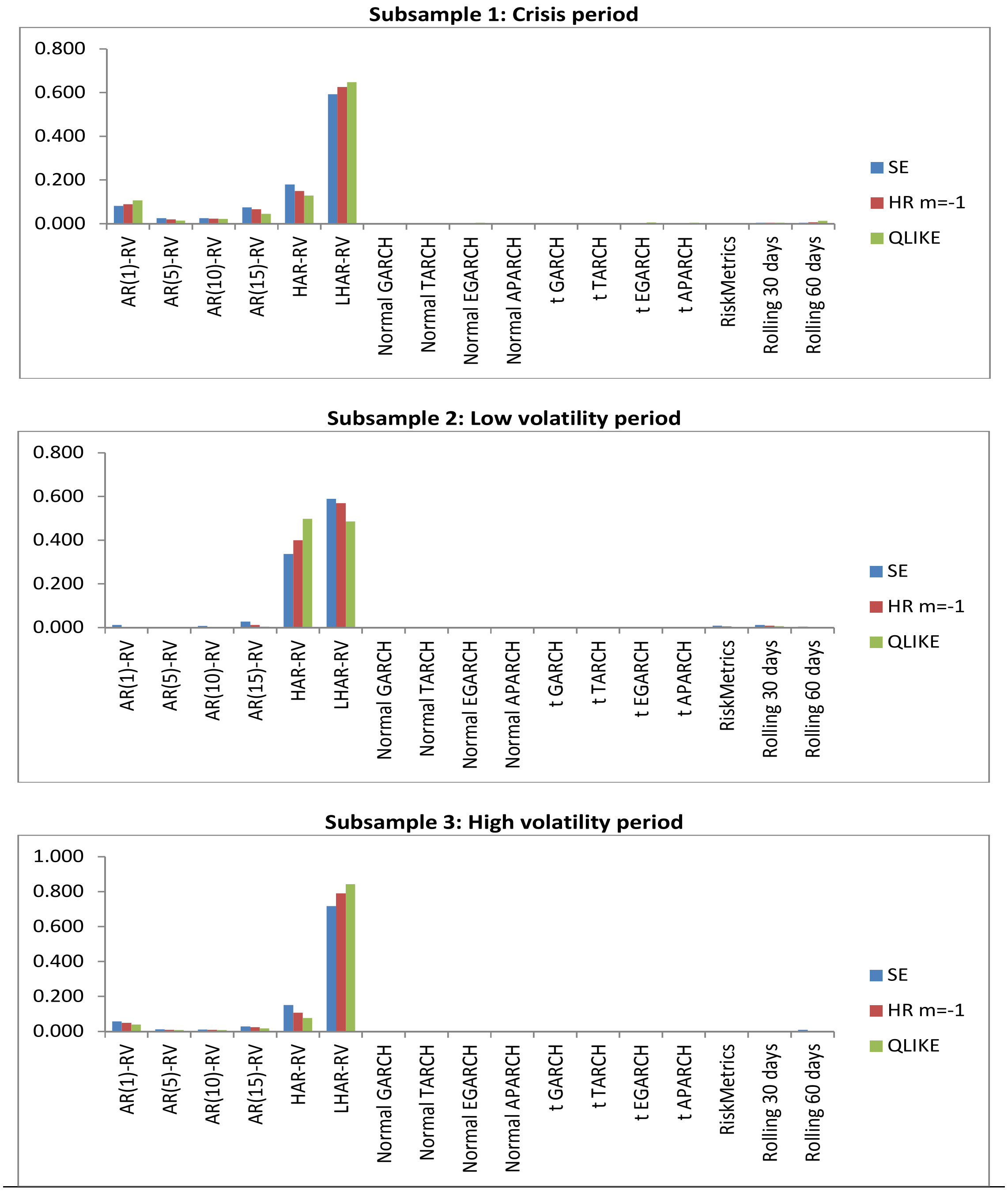
This figure shows the weights of individual forecasts of the FFC approach for the various subsamples of the GARCH diffusion DGP with multiple breaks. The sample size is 3000 observations and the weights are based on 250 simulations. The out-of-sample evaluation period resembles a low volatility period.

Figure 1.5: Weights of individual volatility forecasts of the FFC approach when the DGP is a GARCH diffusion process with multiple breaks and the out-of-sample evaluation period resembles a low volatility period (2)



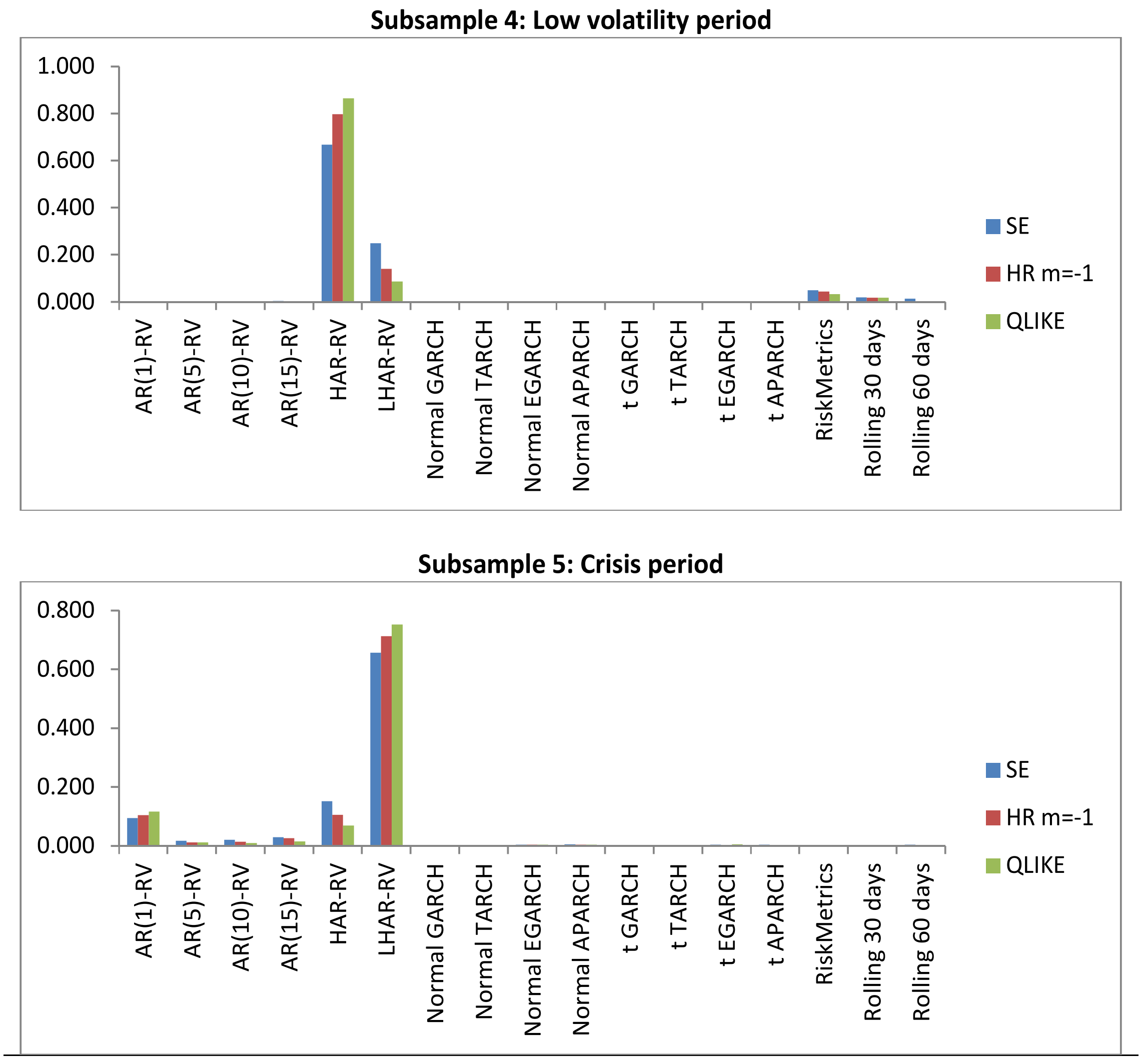
This figure shows the weights of individual forecasts of the FFC approach for the various subsamples of the GARCH diffusion DGP with multiple breaks. The sample size is 3000 observations and the weights are based on 250 simulations. The out-of-sample evaluation period resembles a low volatility period.

Figure 1.6: Weights of individual volatility forecasts of the FFC approach when the DGP is a GARCH diffusion process with multiple breaks and the out-of-sample evaluation period resembles a crisis period (1)



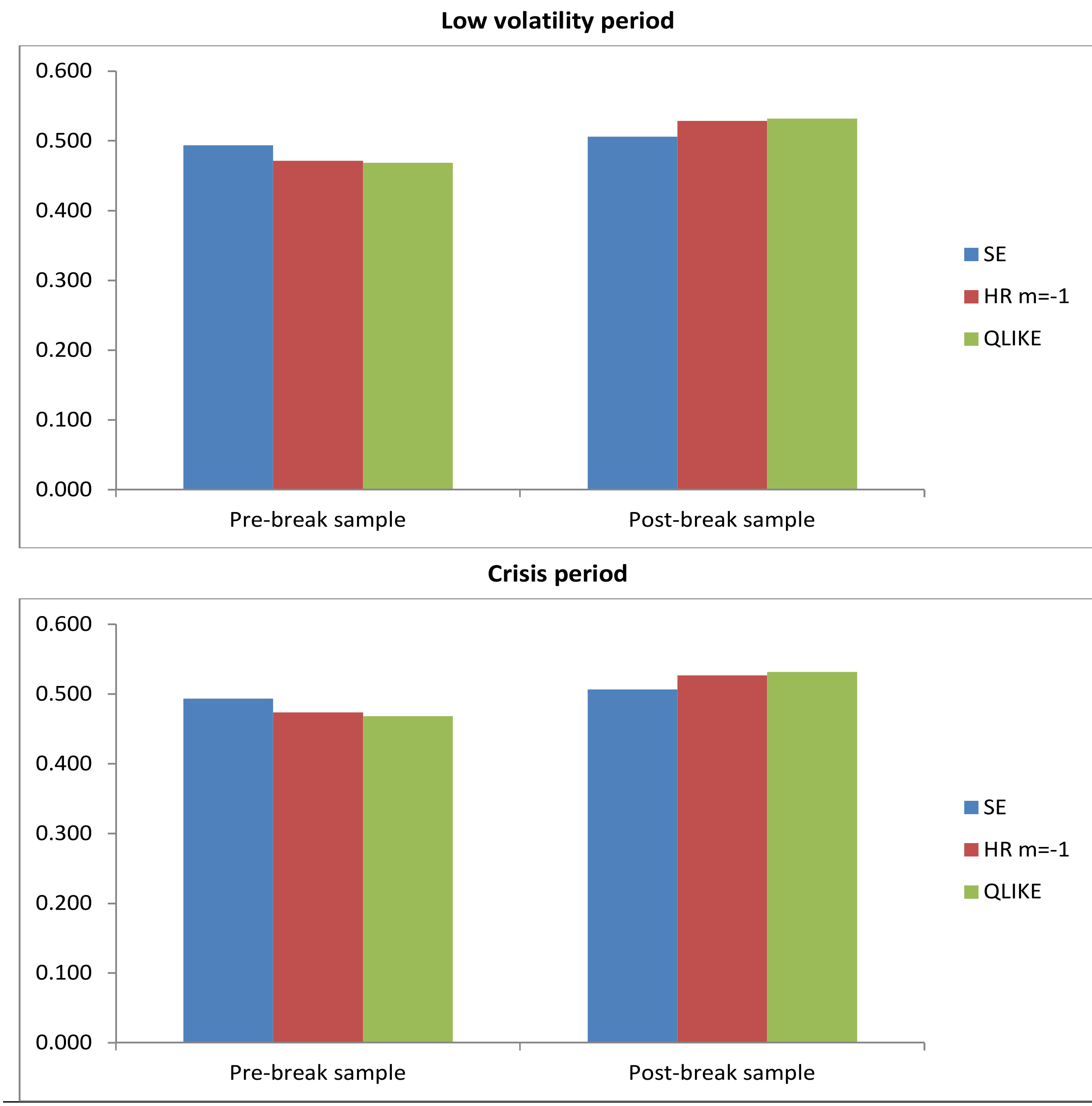
This figure shows the weights of individual forecasts of the FFC approach for the various subsamples of the GARCH diffusion DGP with multiple breaks. The sample size is 3000 observations and the weights are based on 250 simulations. The out-of-sample evaluation period resembles a crisis period.

Figure 1.7: Weights of individual volatility forecasts of the FFC approach when the DGP is a GARCH diffusion process with multiple breaks and the out-of-sample evaluation period resembles a crisis period (2)



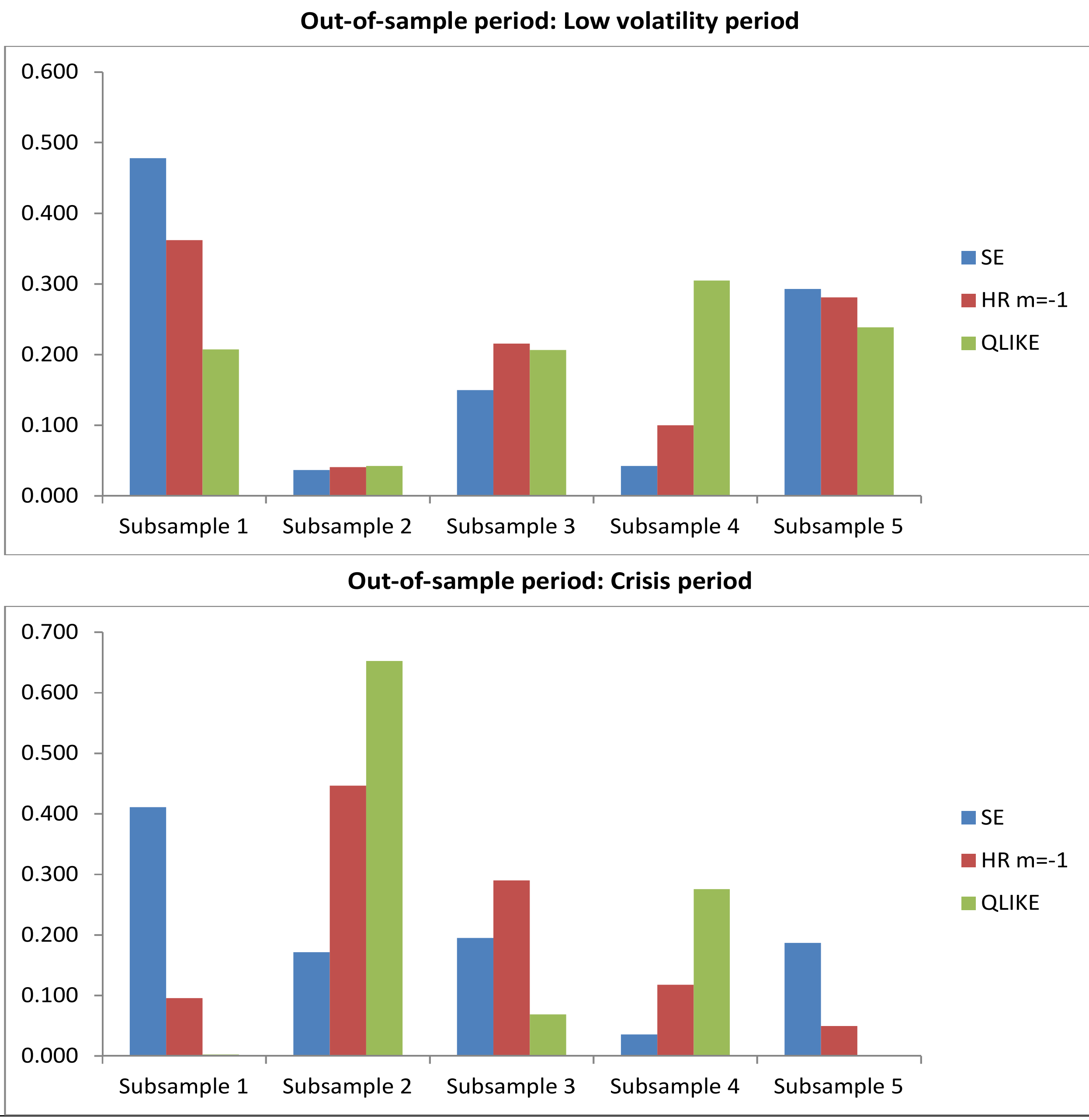
This figure shows the weights of individual forecasts of the FFC approach for the various subsamples of the GARCH diffusion DGP with multiple breaks. The sample size is 3000 observations and the weights are based on 250 simulations. The out-of-sample evaluation period resembles a crisis period.

Figure 1.8: Weights of combination volatility forecasts when the DGP is a GARCH diffusion process without a break but the FFC approach detects a break



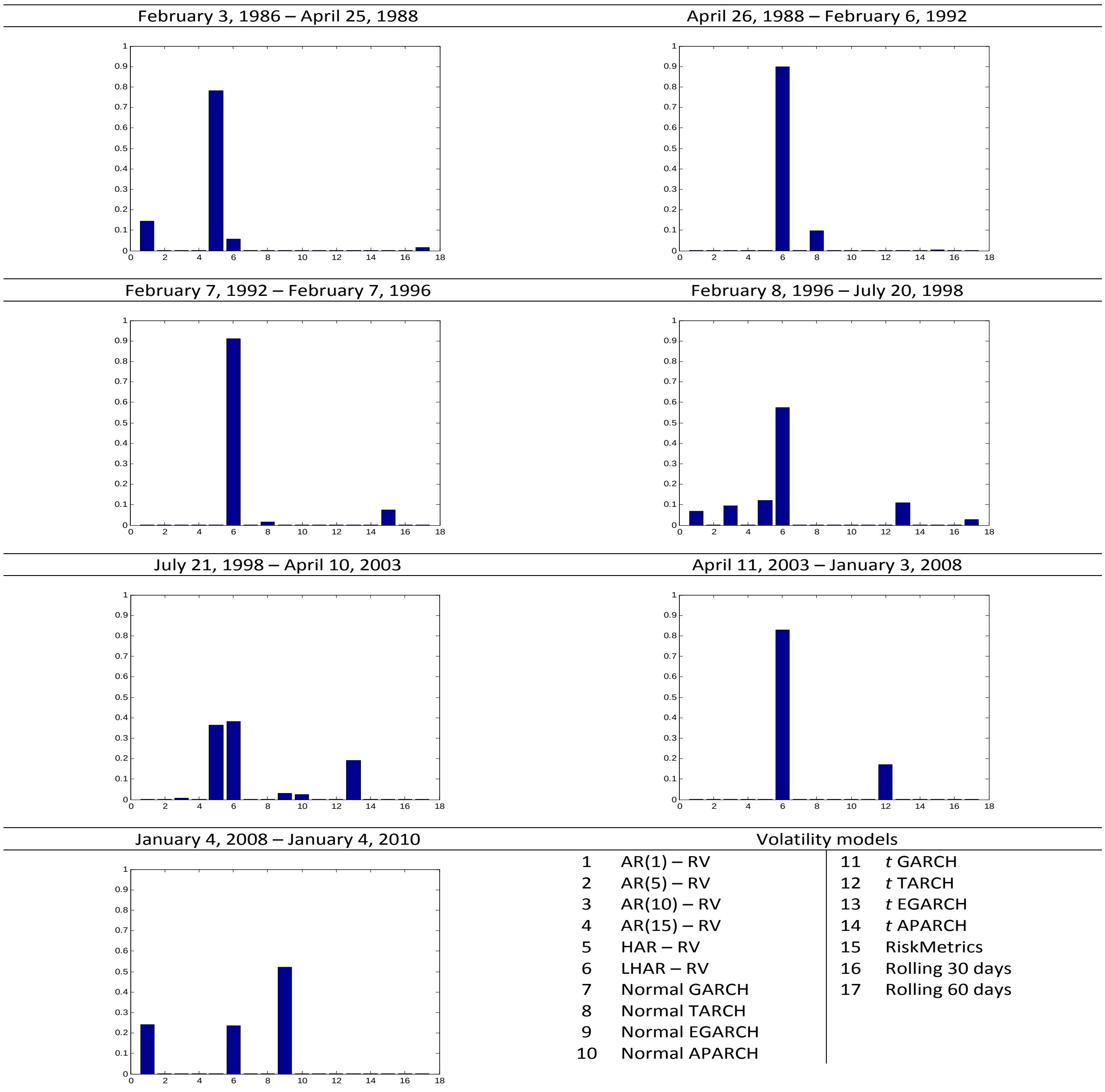
This figure shows the weights of combination weights of the FFC approach for the pre- and post-break samples. The DGP is a GARCH diffusion model without a break. We assume that there is a misspecification and the FFC approach detects a break in the middle of the sample. There are two cases, in the top panel the DGP resembles a low volatility period whereas in the bottom panel the DGP resembles a crisis period. The sample size is 3000 observations and the weights are based on 250 simulations.

Figure 1.9: Weights of combination volatility forecasts when the DGP is GARCH diffusion process with multiple breaks



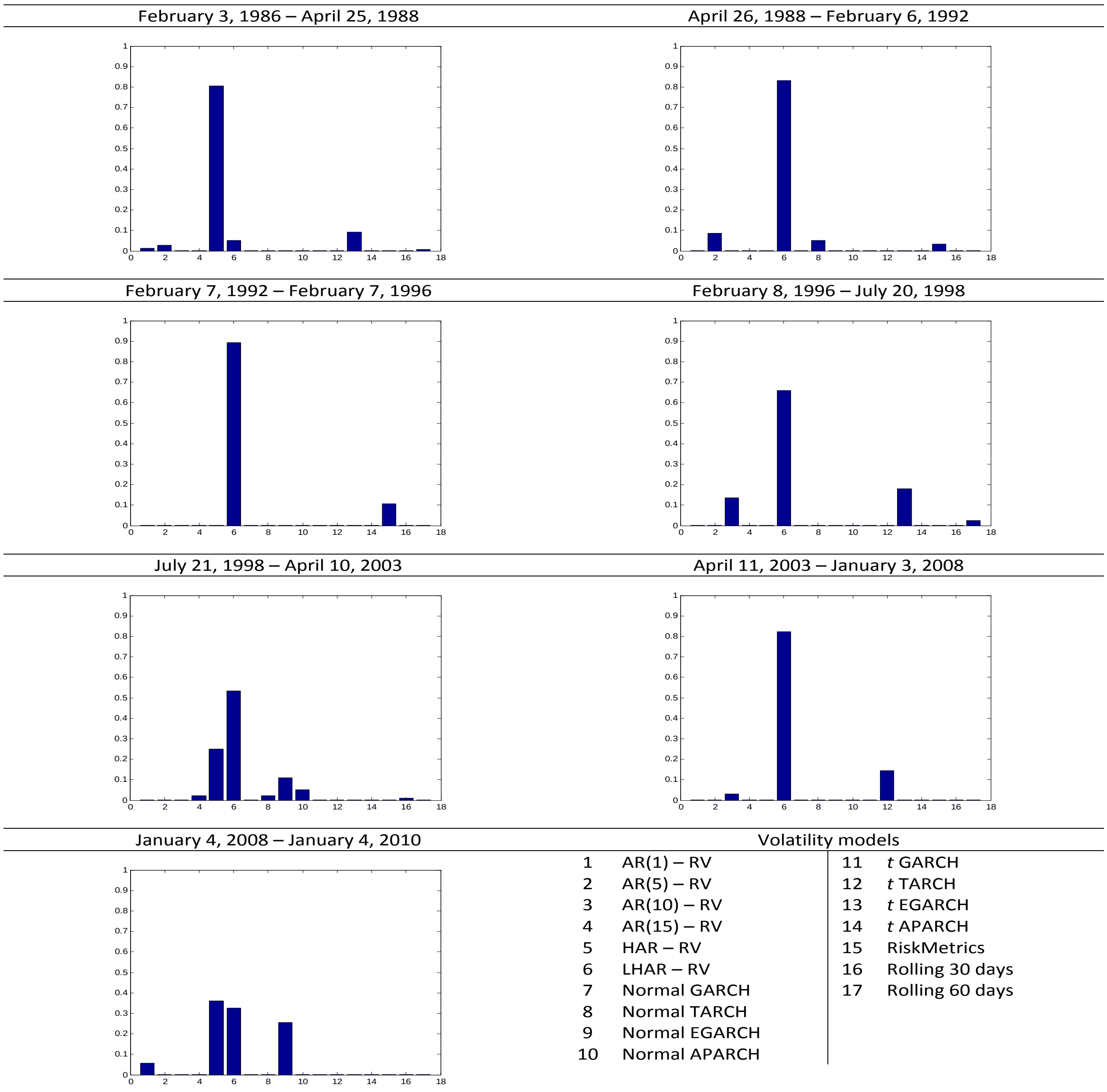
This figure shows the weights of the combination forecasts of the FFC approach for the five subsamples. The DGP is a GARCH diffusion model with multiple breaks. There are two cases, in the top panel the out-of-sample period resembles a low volatility period whereas in the bottom panel the out-of-sample period resembles a crisis period. Subsamples 1 and 5 resemble crisis periods, subsamples 2 and 4 low volatility periods and subsample 2 a high volatility period. The sample size is 3000 observations and the weights are based on 250 simulations.

Figure 1.10: Weights of individual volatility models for the S%P 500 Index based on the Square Error loss function



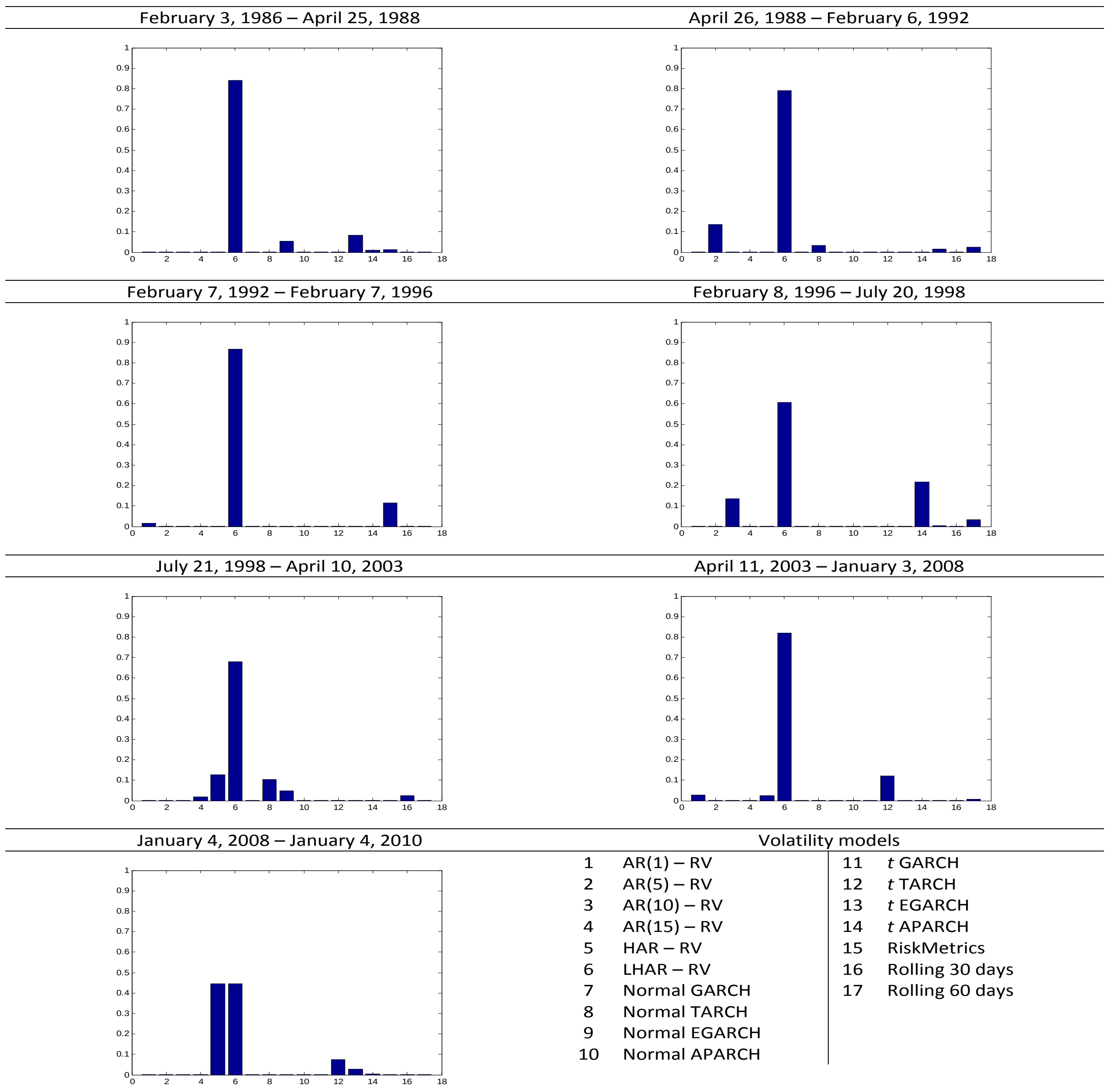
This figure shows the weights of each individual model for each subsample of the Flexible Forecast Combination method based on the Square Error loss function. The estimation period of the S&P 500 Index covers the period from February 3, 1986 to January 4, 2010.

Figure 1.11: Weights of individual volatility models for the S%P 500 Index based on the Homogeneous Robust loss function for b=-1



This figure shows the weights of each individual model for each subsample of the Flexible Forecast Combination method based on the Homogeneous Robust loss function for b=-1. The estimation period of the S&P 500 Index covers the period from February 3, 1986 to January 4, 2010.

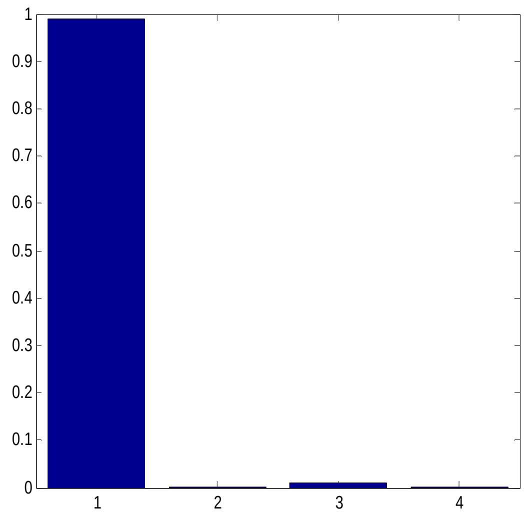
Figure 1.12: Weights of individual volatility models for the S%P 500 Index based on the QLIKE loss function



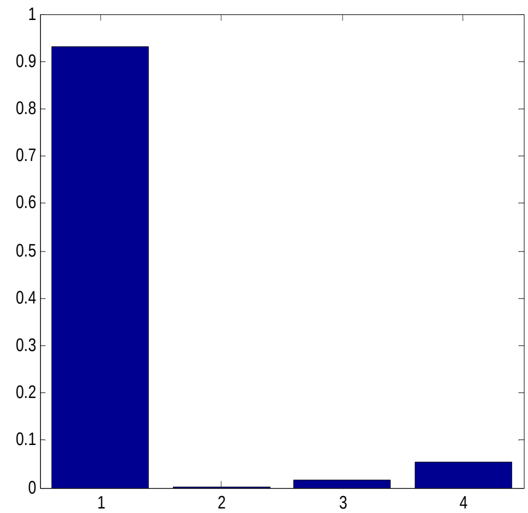
This figure shows the weights of each individual model for each subsample of the Flexible Forecast Combination method based on the QLIKE loss function. The estimation period of the S&P 500 Index covers the period from February 3, 1986 to January 4, 2010.

Figure 1.13: Weights of forecast combinations when the out-of-sample period of the S&P 500 Index spans January 4, 2008 to January 2, 2009

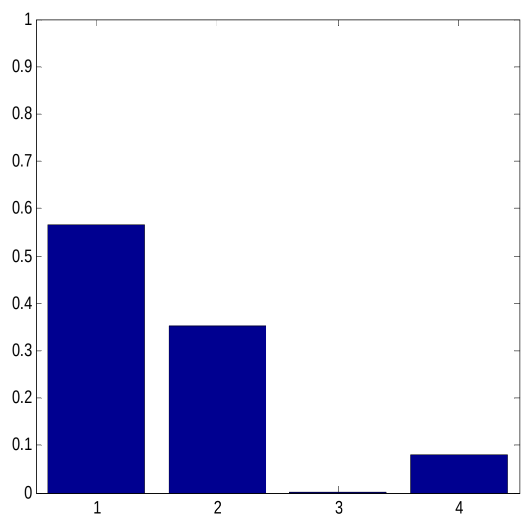
SE loss function



Homogeneous Robust for b=-1 loss function



QLIKE loss function



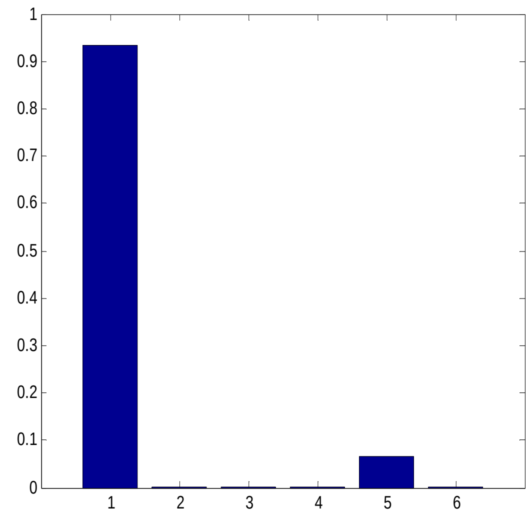
Subsamples

1	February 3, 1986 to April 12, 1991
2	April 15, 1991 to June 23, 1997
3	June 24, 1997 to April 10, 2003
4	April 11, 2003 to January 3, 2008

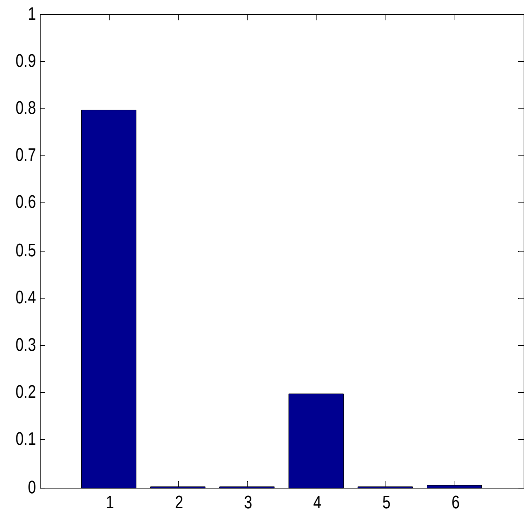
This figure shows the weights of forecast combinations for each subsample of the S&P 500 Index when the out-of-sample period spans January 4, 2008 to January 2, 2009. The estimation period spans February 3, 1986 to January 3, 2008. Each forecast combination uses information from a particular subsample.

Figure 1.14: Weights of forecast combinations when the out-of-sample period of the S&P 500 spans January 3, 2009 to January 4, 2010

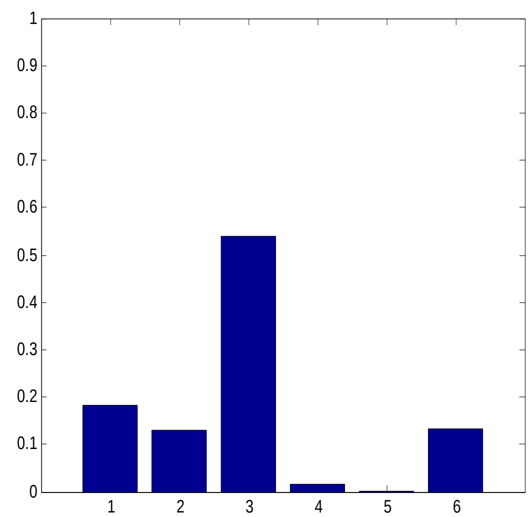
SE loss function



Homogeneous Robust for b=-1 loss function



QLIKE loss function



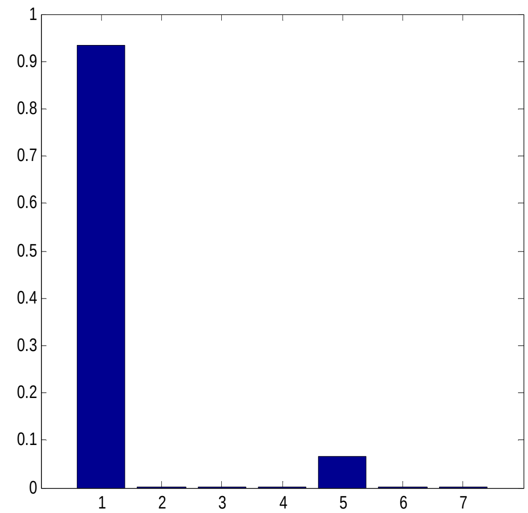
Subsamples

1	February 3, 1986 to April 25, 1988
2	April 26, 1988 to February 6, 1992
3	February 7, 1992 to February 7, 1996
4	February 8, 1996 to July 20, 1998
5	July 21, 1998 to April 10, 2003
6	April 11, 2003 to January 2, 2009

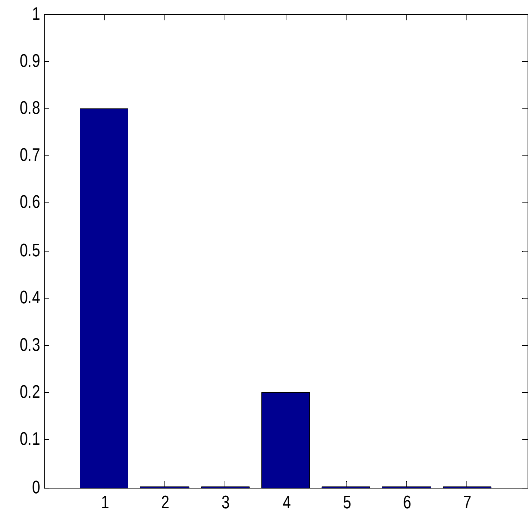
This figure shows the weights of forecast combinations for each subsample of the S&P 500 Index when the out-of-sample period spans January 3, 2009 to January 4, 2010. The estimation period spans February 3, 1986 to January 2, 2009. Each forecast combination uses information from a particular subsample.

Figure 1.15: Weights of forecast combinations when the out-of-sample period of the S&P 500 spans January 5, 2010 to June 30, 2010

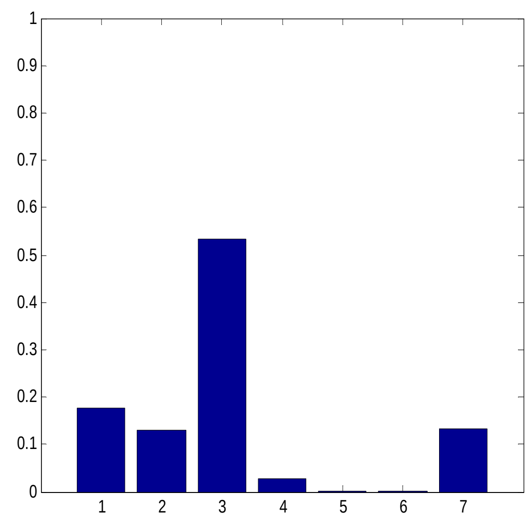
SE loss function



Homogeneous Robust for b=-1 loss function



QLIKE loss function



Subsamples

1	February 3, 1986 to April 25, 1988
2	April 26, 1988 to February 6, 1992
3	February 7, 1992 to February 7, 1996
4	February 8, 1996 to July 20, 1998
5	July 21, 1998 to April 10, 2003
6	April 11, 2003 to January 3, 2008
7	January 4, 2008 to January 4, 2010

This figure shows the weights of forecast combinations for each subsample of the S&P 500 Index when the out-of-sample period spans January 5, 2010 to June 30, 2010. The estimation period spans February 3, 1986 to January 4, 2010. Each forecast combination uses information from a particular subsample.

Table 1: CUSUM- type test for multiple breaks in the Realized Volatility of the S&P 500 Index

Breaks	Start date	End date	Sample size	Test statistic	Break date
1	03/02/1986	30/06/2010	6154	3.8665***	21/07/1998
2	03/02/1986	20/07/1998	3149	1.8193***	26/04/1988
3	21/07/1998	30/06/2010	3005	2.8663***	04/01/2008
4	26/04/1988	20/07/1998	2586	3.4589***	08/02/1996
5	21/07/1998	03/01/2008	2378	4.8512***	11/04/2003
6	26/04/1988	07/02/1996	1969	3.9743***	07/02/1992

This table shows the structural breaks in Realized Volatility of the S&P 500 Index from February 3, 1986 to June 30, 2010. These breaks are estimated using the CUSUM type test of Kokoszka and Leipus (1998, 2000) based on the HAC variance estimator (Andrews, 1991). *, ** and *** correspond to rejections of the null hypothesis of no break in the volatility for 10%, 5% and 1% confidence levels with corresponding critical values of 1.22, 1.36 and 1.63, respectively.

Table 2: Parameter estimates of the TARCH model fitted to the subsamples of the S&P 500 Index

Subsample	Description	Start date	End date	Sample size	ω	α	β	θ
1	Crisis period Stock market crash 1987	03/02/1986	25/04/1988	563	0.178 (0.144)	0.052 (0.090)	0.732*** (0.129)	0.272 (0.257)
2	High volatility subsample	26/04/1988	06/02/1992	958	0.146 (0.122)	0.000 (0.018)	0.773*** (0.165)	0.097 (0.064)
3	Low volatility subsample	07/02/1992	07/02/1996	1011	0.030** (0.014)	0.000 (0.017)	0.860*** (0.055)	0.111** (0.044)
4	High volatility subsample	08/02/1996	20/07/1998	617	0.074 (0.047)	0.000 (0.025)	0.815*** (0.097)	0.250* (0.148)
5	High volatility subsample Dotcom and LTCM	21/07/1998	10/04/2003	1188	0.075 (0.061)	0.000 (0.080)	0.880*** (0.103)	0.163*** (0.037)
6	Low volatility subsample	11/04/2003	03/01/2008	1190	0.013 (0.009)	0.000 (0.021)	0.934*** (0.037)	0.087*** (0.027)
7	Crisis period Recent financial crisis	04/01/2008	30/06/2010	627	0.033** (0.014)	0.000 (0.029)	0.903*** (0.021)	0.165*** (0.043)
8	Full sample	03/02/1986	30/06/2010	6152	0.020*** (0.007)	0.012 (0.008)	0.908*** (0.023)	0.131*** (0.035)

This table shows the subsamples of the S&P 500 from February 3, 1986 to June 30, 2010 based on structural breaks estimated using the CUSUM test and the corresponding parameter estimates of the TARCH (1,1) model: $h_t = \omega + \alpha r_{t-1}^2 + \gamma r_{t-1}^2 I\{r_{t-1} < 0\} + \beta_{t-1} h_{t-1}$ for each subsample. Entries in parenthesis correspond to the Bollerslev - Wooldridge standard errors. *,** and *** refer to significant levels of 10%, 5% and 1%, respectively.

Table 3: Parameter estimates of the EGARCH model fitted to the subsamples of the S&P 500 Index

Subsample	Description	Start date	End date	Sample size	ω	α	β	θ
1	Crisis period Stock market crash 1987	03/02/1986	25/04/1988	563	0.059* (0.032)	0.230** (0.096)	0.930*** (0.036)	-0.168* (0.089)
2	High volatility subsample	26/04/1988	06/02/1992	958	-0.043 (0.072)	0.078 (0.057)	0.775** (0.356)	-0.131 (0.129)
3	Low volatility subsample	07/02/1992	07/02/1996	1011	-0.123** (0.061)	0.079** (0.035)	0.881*** (0.056)	-0.134*** (0.039)
4	High volatility subsample	08/02/1996	20/07/1998	617	0.001 (0.012)	0.135** (0.054)	0.917*** (0.042)	-0.192*** (0.063)
5	High volatility subsample Dotcom and LTCM	21/07/1998	10/04/2003	1188	0.014* (0.008)	0.068*** (0.022)	0.968*** (0.012)	-0.129*** (0.022)
6	Low volatility subsample	11/04/2003	03/01/2008	1190	-0.012 (0.008)	0.072*** (0.016)	0.971*** (0.015)	-0.109*** (0.028)
7	Crisis period Recent financial crisis	04/01/2008	30/06/2010	627	0.021*** (0.008)	0.141*** (0.033)	0.976*** (0.008)	-0.147*** (0.029)
8	Full sample	03/02/1986	30/06/2010	6152	0.008*** (0.003)	0.134*** (0.027)	0.979*** (0.006)	-0.101*** (0.020)

This table shows the subsamples of the S&P 500 from February 3, 1986 to June 30, 2010 based on structural breaks estimated using the CUSUM type test and the corresponding parameter estimates of the EGARCH (1,1) model: $\log h_{t+1} = \omega + \alpha \left[\frac{|r_t|}{\sqrt{h_t}} - E\left(\frac{|r_t|}{\sqrt{h_t}}\right)\right] + \beta \log h_t + \theta \frac{r_t}{\sqrt{h_t}}$. Entries in parenthesis refer to the Bollerslev - Wooldridge standard errors. *,** and *** refer to significant levels of 10%, 5% and 1%, respectively.

Table 4: Evaluation of alternative volatility estimates when the DGP is a GARCH diffusion volatility model without breaks

	Full Sample			*Split Sample*			*Bias²(Split)/*	*Full Sample*		*Split Sample*		*MSE(Split)/*
	Bias	*Bias²*	*Rank*	*Bias*	*Bias²*	*Rank*	*Bias²(Full)*	*MSE*	*Rank*	*MSE*	*Rank*	*MSE(Full)*
RV	8.0688	65.1047	12	8.0688	65.1047	12	1.0000	**14.9825**	**3**	**14.9825**	**3**	1.0000
AR(1) - RV	8.1150	65.8524	16	8.1063	65.7117	16	**0.9979**	21.0646	7	20.9719	7	0.9956
AR(5) - RV	8.0875	65.4079	14	8.0762	65.2258	14	**0.9972**	16.9870	6	16.9490	6	0.9978
AR(10) - RV	8.0880	65.4158	15	8.0810	65.3021	15	**0.9983**	16.9659	5	16.9203	5	0.9973
AR(15) - RV	8.0822	65.3214	13	8.0745	65.1968	13	**0.9981**	16.9582	4	16.9099	4	0.9972
HAR - RV	26.8489	720.8646	18	26.6006	707.5945	18	**0.9816**	**11.2407**	**1**	**10.9564**	**1**	0.9747
LHAR - RV	24.8200	616.0338	17	24.6347	606.8701	17	**0.9851**	**12.5147**	**2**	**12.1258**	**2**	0.9689
Normal GARCH	**-0.8137**	**0.6622**	**2**	-1.3386	1.7920	4	2.7063	81.7425	15	81.7181	15	0.9997
Normal TARCH	-3.5833	12.8401	11	-3.9819	15.8555	11	1.2348	70.2619	13	70.6677	13	1.0058
Normal EGARCH	1.7551	3.0805	5	**1.1366**	**1.2919**	**1**	**0.4194**	70.0602	10	68.7240	10	0.9809
Normal APARCH	-0.8505	0.7234	4	-1.6912	2.8601	5	3.9537	67.5818	9	67.6830	9	1.0015
t GARCH	**-0.6190**	**0.3831**	**1**	**-1.1477**	**1.3173**	**2**	3.4384	81.6266	14	81.5450	14	0.9990
t TARCH	-3.4981	12.2364	10	-3.8954	15.1744	10	1.2401	70.1268	11	70.4629	12	1.0048
t EGARCH	1.8311	3.3528	6	**1.2010**	**1.4424**	**3**	**0.4302**	70.1356	12	68.8162	11	0.9812
t APARCH	**-0.8209**	**0.6739**	**3**	-1.7004	2.8914	6	4.2907	67.5420	8	67.6109	8	1.0010
RiskMetrics	-2.0261	4.1051	8	-1.9847	3.9389	7	**0.9595**	96.6442	16	96.6798	16	1.0004
Rolling 30 days	-1.9888	3.9555	7	-1.9888	3.9555	8	1.0000	115.1399	17	115.1399	17	1.0000
Rolling 60 days	-2.0497	4.2013	9	-2.0497	4.2013	9	1.0000	139.7926	18	139.7926	18	1.0000

This table shows the Bias, Square Bias and MSE of volatility estimates based on alternative individual models using the full and split sample approaches. The DGP is a GARCH diffusion volatility model without breaks. The sample size is 3000 observations. The parameters of the volatility models for the full sample approach are estimated using all available information in the full sample and for the split sample approach they are estimated for the pre- and post-break periods. Bias and MSE are multiplied by 1000 and square Bias by 10^6. Entries in bold are the three best performing methods in estimating volatility in terms of Square Bias and MSE and the ratios of the cases that the split sample outperforms the full sample.

Table 5: Evaluation of alternative volatility estimates when the DGP is a GARCH diffusion volatility model with break in the leverage coefficient of size 2

	Full Sample			Split Sample			$Bias^2(Split)/$	Full Sample		Split Sample		$MSE(Split)/$
	Bias	$Bias^2$	*Rank*	*Bias*	$Bias^2$	*Rank*	$Bias^2(Full)$	*MSE*	*Rank*	*MSE*	*Rank*	*MSE(Full)*
RV	8.1412	66.2794	12	8.1412	66.2794	12	1.0000	**14.9910**	**3**	**14.9910**	**3**	1.0000
AR(1) - RV	8.1924	67.1158	16	8.1839	66.9766	16	**0.9979**	21.1458	7	21.0511	7	**0.9955**
AR(5) - RV	8.1622	66.6210	15	8.1516	66.4487	15	**0.9974**	17.0570	6	17.0163	6	**0.9976**
AR(10) - RV	8.1617	66.6126	14	8.1515	66.4476	14	**0.9975**	17.0242	5	16.9721	5	**0.9969**
AR(15) - RV	8.1559	66.5190	13	8.1484	66.3957	13	**0.9981**	17.0184	4	16.9631	4	**0.9968**
HAR - RV	26.8994	723.5804	18	26.6535	710.4105	18	**0.9818**	**11.3517**	**1**	**11.0561**	**1**	**0.9740**
LHAR - RV	24.6199	606.1376	17	24.0812	579.9022	17	**0.9567**	**13.2320**	**2**	**12.8302**	**2**	**0.9696**
Normal GARCH	**-1.2503**	**1.5632**	**3**	-1.8439	3.4001	4	2.1751	79.7003	15	80.1865	15	1.0061
Normal TARCH	-4.2395	17.9734	11	-4.9806	24.8059	11	1.3801	68.6515	13	68.4492	13	**0.9971**
Normal EGARCH	1.3913	1.9358	5	**1.0440**	**1.0899**	**1**	**0.5630**	66.4330	10	62.8238	9	**0.9457**
Normal APARCH	-1.3904	1.9331	4	-2.3197	5.3811	6	2.7837	65.3551	9	64.1070	11	**0.9809**
t GARCH	**-0.8975**	**0.8054**	**1**	**-1.4941**	**2.2324**	**3**	2.7717	79.5282	14	79.7273	14	1.0025
t TARCH	-4.0293	16.2354	10	-4.7411	22.4781	10	1.3845	68.3678	12	67.8391	12	**0.9923**
t EGARCH	1.5656	2.4511	6	**1.2359**	**1.5276**	**2**	**0.6232**	66.7153	11	62.7837	8	**0.9411**
t APARCH	**-1.2425**	**1.5438**	**2**	-2.2108	4.8878	5	3.1662	65.1925	8	63.6970	10	**0.9771**
RiskMetrics	-2.5622	6.5650	7	-2.5114	6.3072	7	**0.9607**	96.8047	16	96.7970	16	**0.9999**
Rolling 30 days	-2.5960	6.7393	9	-2.5960	6.7393	9	1.0000	117.3219	17	117.3219	17	1.0000
Rolling 60 days	-2.5861	6.6881	8	-2.5861	6.6881	8	1.0000	142.2775	18	142.2775	18	1.0000

This table shows the Bias, Square Bias and MSE of volatility estimates based on alternative individual models using the full and split sample appraoches. The DGP is a GARCH diffusion volatility model with break in the leverage coefficient of size 2. The sample size is 3000 observations. The parameters of the volatility models for the full sample approach are estimated using all available information in the full sample and for the split sample approach they are estimated for the pre- and post-break periods. Bias and MSE are multiplied by 1000 and square Bias by 10^6. Entries in bold are the three best performing methods in estimating volatility in terms of Square Bias and MSE and the ratios of the cases that the split sample outperforms the full sample.

Table 6: Evaluation of alternative volatility estimates when the DGP is a GARCH diffusion volatility model with break in the leverage coefficient of size 3

	Full Sample			*Split Sample*			*Bias²(Split)/*	*Full Sample*		*Split Sample*		*MSE(Split)/*
	Bias	*Bias²*	*Rank*	*Bias*	*Bias²*	*Rank*	*Bias²(Full)*	*MSE*	*Rank*	*MSE*	*Rank*	*MSE(Full)*
RV	8.1099	65.7699	12	8.1099	65.7699	12	1.0000	**15.0120**	**3**	**15.0120**	**3**	1.0000
AR(1) - RV	8.1599	66.5835	16	8.1508	66.4350	16	**0.9978**	21.1398	7	21.0468	7	**0.9956**
AR(5) - RV	8.1300	66.0973	14	8.1195	65.9267	14	**0.9974**	17.0396	6	16.9999	6	**0.9977**
AR(10) - RV	8.1300	66.0975	15	8.1196	65.9272	15	**0.9974**	17.0067	5	16.9565	5	**0.9970**
AR(15) - RV	8.1242	66.0023	13	8.1172	65.8886	13	**0.9983**	17.0006	4	16.9472	4	**0.9969**
HAR - RV	26.8699	721.9929	18	26.6234	708.8056	18	**0.9817**	**11.3329**	**1**	**11.0405**	**1**	**0.9742**
LHAR - RV	24.5269	601.5688	17	23.4254	548.7510	17	**0.9122**	**13.1966**	**2**	**13.5484**	**2**	1.0267
Normal GARCH	-1.7823	3.1767	4	-2.3629	5.5833	4	1.7576	78.9602	15	79.4174	15	1.0058
Normal TARCH	-4.7027	22.1151	11	-5.7021	32.5140	11	1.4702	68.4605	13	66.7934	13	**0.9757**
Normal EGARCH	**0.9016**	**0.8128**	**1**	**1.2948**	**1.6765**	**1**	2.0626	65.9838	10	58.7182	9	**0.8899**
Normal APARCH	-1.9586	3.8362	6	-2.7296	7.4508	6	1.9423	65.1926	9	60.8508	11	**0.9334**
t GARCH	**-1.4375**	**2.0663**	**3**	**-2.0235**	**4.0946**	**3**	1.9816	78.8222	14	78.9912	14	1.0021
t TARCH	-4.5435	20.6433	10	-5.4766	29.9932	10	1.4529	68.1941	12	66.1755	12	**0.9704**
t EGARCH	**1.0206**	**1.0415**	**2**	**1.4888**	**2.2165**	**2**	2.1281	66.2369	11	58.6822	8	**0.8859**
t APARCH	-1.8598	3.4590	5	-2.6609	7.0806	5	2.0470	64.9646	8	60.5289	10	**0.9317**
RiskMetrics	-3.1143	9.6986	7	-3.0625	9.3789	7	**0.9670**	96.2682	16	96.2588	16	**0.9999**
Rolling 30 days	-3.1399	9.8591	8	-3.1399	9.8591	8	1.0000	116.8952	17	116.8952	17	1.0000
Rolling 60 days	-3.1582	9.9742	9	-3.1582	9.9742	9	1.0000	142.0929	18	142.0929	18	1.0000

This table shows the Bias, Square Bias and MSE of volatility estimates based on alternative individual models using the full and split sample appraoches. The DGP is a GARCH diffusion volatility model with break in the leverage coefficient of size 3. The sample size is 3000 observations. The parameters of the volatility models for the full sample approach are estimated using all available information in the full sample and for the split sample approach they are estimated for the pre- and post-break periods. Bias and MSE are multiplied by 1000 and square Bias by 10^6. Entries in bold are the three best performing methods in estimating volatility in terms of Square Bias and MSE and the ratios of the cases that the split sample outperforms the full sample.

Table 7: Evaluation of alternative volatility estimates when the DGP is a GARCH diffusion volatility model with break in the constant of size 2

	Full Sample			Split Sample			$Bias^2$(Split)/	Full Sample		Split Sample		MSE(Split)/
	Bias	$Bias^2$	Rank	Bias	$Bias^2$	Rank	$Bias^2$(Full)	MSE	Rank	MSE	Rank	MSE(Full)
RV	7.5790	57.4415	13	7.5790	57.4415	14	1.0000	**14.5300**	**3**	**14.5300**	**3**	1.0000
AR(1) - RV	7.7914	60.7060	14	7.8685	61.9126	16	1.0199	20.7808	7	20.3029	7	**0.9770**
AR(5) - RV	7.5486	56.9808	12	7.6019	57.7881	15	1.0142	16.5567	6	16.4005	6	**0.9906**
AR(10) - RV	7.5092	56.3880	11	7.5789	57.4404	13	1.0187	16.5199	5	16.3707	5	**0.9910**
AR(15) - RV	7.4767	55.9007	10	7.5635	57.2063	12	1.0234	16.5055	4	16.3608	4	**0.9912**
HAR - RV	24.6499	607.6197	18	25.0230	626.1485	18	1.0305	**10.4040**	**1**	**10.6542**	**1**	1.0241
LHAR - RV	22.4611	504.5005	17	23.1787	537.2521	17	1.0649	**11.9617**	**2**	**11.6264**	**2**	**0.9720**
Normal GARCH	-6.0447	36.5384	9	**-0.8962**	**0.8032**	**2**	**0.0220**	83.9091	15	79.3482	15	**0.9456**
Normal TARCH	-8.8145	77.6953	15	-3.3207	11.0271	11	**0.1419**	74.6280	13	68.4924	13	**0.9178**
Normal EGARCH	**-0.4602**	**0.2118**	**1**	1.3693	1.8750	6	8.8549	70.4288	9	66.6899	10	**0.9469**
Normal APARCH	-4.8426	23.4512	6	**-1.0728**	**1.1509**	**3**	**0.0491**	70.9218	11	65.7873	9	**0.9276**
t GARCH	-5.9444	35.3356	8	**-0.7604**	**0.5783**	**1**	**0.0164**	83.6840	14	79.1969	14	**0.9464**
t TARCH	-8.9103	79.3937	16	-3.2661	10.6671	10	**0.1344**	74.4402	12	68.2580	12	**0.9170**
t EGARCH	**-0.6271**	**0.3932**	**2**	1.4118	1.9932	7	5.0687	70.3364	8	66.7561	11	**0.9491**
t APARCH	-5.0411	25.4128	7	-1.1036	1.2180	4	**0.0479**	70.7746	10	65.7165	8	**0.9285**
RiskMetrics	-1.6466	2.7113	4	-1.1592	1.3437	5	**0.4956**	93.8575	16	93.8596	16	1.0000
Rolling 30 days	-1.6826	2.8311	5	-1.6826	2.8311	9	1.0000	112.0110	17	112.0110	17	1.0000
Rolling 60 days	**-1.5637**	**2.4453**	**3**	-1.5637	2.4453	8	1.0000	134.9806	18	134.9806	18	1.0000

This table shows the Bias, Square Bias and MSE of volatility estimates based on alternative individual models using the full and split sample approaches. The DGP is a GARCH diffusion volatility model with break in the constant of size 2. The sample size is 3000 observations. The parameters of the volatility models for the full sample approach are estimated using all available information in the full sample and for the split sample approach they are estimated for the pre- and post-break periods. Bias and MSE are multiplied by 1000 and square Bias by 10^6. Entries in bold are the three best performing methods in estimating volatility in terms of Square Bias and MSE and the ratios of the cases that the split sample outperforms the full sample.

Table 8: Evaluation of alternative volatility estimates when the DGP is a GARCH diffusion volatility model with break in the constant of size 3

	Full Sample			*Split Sample*			*Bias^2(Split)/*	*Full Sample*		*Split Sample*		*MSE(Split)/*
	Bias	*Bias^2*	*Rank*	*Bias*	*Bias^2*	*Rank*	*Bias^2(Full)*	*MSE*	*Rank*	*MSE*	*Rank*	*MSE(Full)*
RV	7.5590	57.1390	9	7.5590	57.1390	14	1.0000	**16.2329**	**3**	**16.2329**	**3**	1.0000
AR(1) - RV	7.7780	60.4968	10	7.9783	63.6540	16	1.0522	23.5271	7	22.6703	7	**0.9636**
AR(5) - RV	7.4571	55.6080	8	7.5913	57.6276	15	1.0363	18.5800	6	18.3105	6	**0.9855**
AR(10) - RV	7.3980	54.7310	7	7.5543	57.0671	13	1.0427	18.5226	5	18.2763	5	**0.9867**
AR(15) - RV	7.3512	54.0404	6	7.5337	56.7564	12	1.0503	18.4980	4	18.2652	4	**0.9874**
HAR - RV	23.9433	573.2804	18	24.9773	623.8670	18	1.0882	**11.2119**	**1**	**11.9157**	**1**	1.0628
LHAR - RV	21.4115	458.4506	17	23.1380	535.3648	17	1.1678	**13.4930**	**2**	**12.9415**	**2**	**0.9591**
Normal GARCH	-12.3669	152.9409	14	**-0.7297**	**0.5325**	**2**	**0.0035**	99.1976	15	88.6780	15	**0.8940**
Normal TARCH	-16.1068	259.4299	16	-3.1051	9.6419	11	**0.0372**	90.8923	13	76.4950	13	**0.8416**
Normal EGARCH	-4.5155	20.3897	4	1.4446	2.0869	7	**0.1024**	81.5064	9	74.5137	10	**0.9142**
Normal APARCH	-10.4459	109.1166	11	-0.8091	0.6547	4	**0.0060**	85.5064	11	73.5276	9	**0.8599**
t GARCH	-12.1560	147.7673	13	**-0.6304**	**0.3975**	**1**	**0.0027**	98.8614	14	88.5089	14	**0.8953**
t TARCH	-16.0934	258.9972	15	-3.0757	9.4602	10	**0.0365**	90.6073	12	76.2152	12	**0.8412**
t EGARCH	-4.7681	22.7347	5	1.4819	2.1960	8	**0.0966**	81.2660	8	74.5784	11	**0.9177**
t APARCH	-10.8055	116.7592	12	-0.8336	0.6948	5	**0.0060**	85.3444	10	73.4376	8	**0.8605**
RiskMetrics	**-1.5031**	**2.2593**	**2**	**-0.7917**	**0.6269**	**3**	**0.2775**	104.9355	16	104.9434	16	1.0001
Rolling 30 days	**-1.5748**	**2.4800**	**3**	-1.5748	2.4800	9	1.0000	125.3038	17	125.3038	17	1.0000
Rolling 60 days	**-1.3674**	**1.8698**	**1**	-1.3674	1.8698	6	1.0000	150.6084	18	150.6084	18	1.0000

This table shows the Bias, Square Bias and MSE of volatility estimates based on alternative individual models using the full and split sample approaches. The DGP is a GARCH diffusion volatility model with break in the constant of size 3. The sample size is 3000 observations. The parameters of the volatility models for the full sample approach are estimated using all available information in the full sample and for the split sample approach they are estimated for the pre- and post-break periods. Bias and MSE are multiplied by 1000 and square Bias by 10^6. Entries in bold are the three best performing methods in estimating volatility in terms of Square Bias and MSE and the ratios of the cases that the split sample outperforms the full sample.

Table 9: Comparison of alternative volatility estimates based on the split and the full samples when the DGP is a GARCH diffusion volatility model

	$Bias^2$(Split)/$Bias^2$(Full)					*MSE(Split)/MSE(Full)*				
	No break	*Break leverage Size=2*	*Break leverage Size = 3*	*Break constant Size = 2*	*Break constant Size = 3*	*No break*	*Break leverage Size=2*	*Break leverage Size = 3*	*Break constant Size = 2*	*Break constant Size = 3*
RV	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
AR(1) - RV	**0.9979**	**0.9979**	**0.9978**	1.0199	1.0522	**0.9956**	**0.9955**	**0.9956**	**0.9770**	**0.9636**
AR(5) - RV	**0.9972**	**0.9974**	**0.9974**	1.0142	1.0363	**0.9978**	**0.9976**	**0.9977**	**0.9906**	**0.9855**
AR(10) - RV	**0.9983**	**0.9975**	**0.9974**	1.0187	1.0427	**0.9973**	**0.9969**	**0.9970**	**0.9910**	**0.9867**
AR(15) - RV	**0.9981**	**0.9981**	**0.9983**	1.0234	1.0503	**0.9972**	**0.9968**	**0.9969**	**0.9912**	**0.9874**
HAR - RV	**0.9816**	**0.9818**	**0.9817**	1.0305	1.0882	**0.9747**	**0.9740**	**0.9742**	1.0241	1.0628
LHAR - RV	**0.9851**	**0.9567**	**0.9122**	1.0649	1.1678	**0.9689**	**0.9696**	1.0267	**0.9720**	**0.9591**
Normal GARCH	2.7063	2.1751	1.7576	**0.0220**	**0.0035**	**0.9997**	1.0061	1.0058	**0.9456**	**0.8940**
Normal TARCH	1.2348	1.3801	1.4702	**0.1419**	**0.0372**	1.0058	**0.9971**	**0.9757**	**0.9178**	**0.8416**
Normal EGARCH	**0.4194**	**0.5630**	2.0626	8.8549	**0.1024**	**0.9809**	**0.9457**	**0.8899**	**0.9469**	**0.9142**
Normal APARCH	3.9537	2.7837	1.9423	**0.0491**	**0.0060**	1.0015	**0.9809**	**0.9334**	**0.9276**	**0.8599**
t GARCH	3.4384	2.7717	1.9816	**0.0164**	**0.0027**	**0.9990**	1.0025	1.0021	**0.9464**	**0.8953**
t TARCH	1.2401	1.3845	1.4529	**0.1344**	**0.0365**	1.0048	**0.9923**	**0.9704**	**0.9170**	**0.8412**
t EGARCH	**0.4302**	**0.6232**	2.1281	5.0687	**0.0966**	**0.9812**	**0.9411**	**0.8859**	**0.9491**	**0.9177**
t APARCH	4.2907	3.1662	2.0470	**0.0479**	**0.0060**	1.0010	**0.9771**	**0.9317**	**0.9285**	**0.8605**
RiskMetrics	**0.9595**	**0.9607**	**0.9670**	**0.4956**	**0.2775**	1.0004	**0.9999**	**0.9999**	1.0000	1.0001
Rolling 30 days	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Rolling 60 days	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

This table shows the ratios of the Square Bias and MSE of alternative volatility estimates based on the split sample compared to the full sample approach. The DGP is a GARCH diffusion volatility model without breaks, with break in the leverage coefficient of sizes 2 and 3 and with break in the constant of sizes 2 and 3. The break occurs in the middle of the sample. The sample size is 3000 observations. The parameters of the volatility models for the full sample approach are estimated using all information in the full sample and for the split sample approach they are estimated in the pre- and post-break samples. Entries in bold correspond to the ratios of the cases that the split sample outperforms the full sample approach.

Table 10: Evaluation of alternative volatility forecasts when the DGP is a GARCH diffusion volatility model without breaks

	Full Sample			*Split Sample*			*Full Sample*		*Split Sample*		*CPA - MSE*	
	Bias	*Bias^2*	*Rank*	*Bias*	*Bias^2*	*Rank*	*MSE*	*Rank*	*MSE*	*Rank*	*% rej.*	*I_{sf}*
AR(1) - RV	9.1630	83.9600	15	10.1984	104.0080	15	32.0605	6	32.1897	6	**0.7170**	0.4677
AR(5) - RV	8.7321	76.2503	14	9.3593	87.5965	13	**28.1966**	**3**	**28.3064**	**3**	0.3990	0.4153
AR(10) - RV	8.7081	75.8314	13	9.3071	86.6227	11	28.1993	4	28.3243	4	0.3840	0.4112
AR(15) - RV	8.6936	75.5790	12	9.2778	86.0772	10	28.2081	5	28.3368	5	0.3940	0.4110
HAR - RV	28.1177	790.6063	17	29.1558	850.0604	17	**20.2216**	**1**	**20.5870**	**2**	0.4730	0.4561
LHAR - RV	25.4725	648.8470	16	26.3520	694.4279	16	**21.2176**	**2**	**20.3974**	**1**	0.3650	0.4172
Normal GARCH	2.3345	5.4499	8	5.3030	28.1221	8	83.1026	14	85.8594	13	0.4450	0.4616
Normal TARCH	**-0.9120**	**0.8318**	**2**	**1.6239**	**2.6372**	**1**	72.4755	10	76.0479	8	0.4300	0.4517
Normal EGARCH	5.7736	33.3347	10	9.3294	87.0380	12	75.7759	12	80.0010	11	0.4300	0.4488
Normal APARCH	2.2567	5.0929	6	4.6197	21.3420	6	71.8640	8	77.4187	10	0.4220	0.4391
t GARCH	2.5597	6.5523	9	5.6246	31.6357	9	83.0966	13	85.8942	14	0.4490	0.4608
t TARCH	**-0.8409**	**0.7072**	**1**	**1.8149**	**3.2938**	**3**	72.3689	9	75.9396	7	0.4350	0.4524
t EGARCH	5.8403	34.1092	11	9.5188	90.6073	14	75.7001	11	80.0102	12	0.4320	0.4497
t APARCH	2.2647	5.1288	7	4.6883	21.9799	7	71.6230	7	76.6012	9	0.4250	0.4393
RiskMetrics	-1.9838	3.9356	5	-1.9502	3.8031	5	95.9927	15	96.0054	15	0.0030	0.4751
Rolling 30 days	-1.8220	3.3195	4	-1.8220	3.3195	4	121.6196	16	121.6196	16	0.0000	0.0000
Rolling 60 days	**-1.7759**	**3.1539**	**3**	**-1.7759**	**3.1539**	**2**	146.5359	17	146.5359	17	0.0000	0.0000

This table shows the Bias, Square Bias, MSE and the corresponding rankings of volatility forecasts based on alternative individual models using the full and split sample approaches. The DGP is a GARCH diffusion volatility model without breaks. The in-sample size is 3000 daily observations and 1000 observations are used for out-of-sample evaluation. The parameters of the volatility models for the full sample approach are estimated using all available information in the full sample and for the split sample approach using only the post-break sample. Bias and MSE are multiplied by 1000 and Square Bias by 10^6. CPA-MSE column corresponds to the (i) % of rejections of the null hypothesis of equal predictive ability between the volatility forecasts based on the full and split sample approaches using the CPA test for a confidence level of 5% and (ii) to the percentage of cases (I_{sf}) that the split sample outperforms the full sample approach using the MSE loss function, respectively. Entries in bold are the three best performing methods in forecasting volatility based on MSE, the cases that the percentage of rejections is higher than 50% and the cases that the split sample outperforms on average the full sample approach.

Table 11: Evaluation of alternative volatility forecasts when the DGP is a GARCH diffusion volatility model with break in the leverage coefficient of size 2

	Full Sample			Split Sample			Full Sample		Split Sample		CPA - MSE	
	Bias	*Bias²*	*Rank*	*Bias*	*Bias²*	*Rank*	*MSE*	*Rank*	*MSE*	*Rank*	*% rej.*	I_{sf}
AR(1) - RV	9.3639	87.6828	13	10.3835	107.8165	15	31.9499	6	32.0685	5	**0.6890**	0.4658
AR(5) - RV	8.9681	80.4275	12	9.6062	92.2796	14	28.1930	4	**28.3107**	**2**	0.3770	0.4170
AR(10) - RV	8.9508	80.1161	11	9.5662	91.5116	13	**28.1873**	**3**	**28.3405**	**3**	0.3980	0.4094
AR(15) - RV	8.9362	79.8563	10	9.5233	90.6931	12	28.1961	5	28.3475	4	0.4060	0.4126
HAR - RV	28.3153	801.7573	17	29.3310	860.3077	17	**20.2757**	**1**	**20.6539**	**1**	0.4520	0.4589
LHAR - RV	23.0084	529.3855	16	23.1966	538.0837	16	**20.3473**	**2**	33.2175	6	0.2150	0.2947
Normal GARCH	**1.6397**	**2.6886**	**1**	4.7281	22.3551	8	79.0686	13	82.1232	14	0.4110	0.4634
Normal TARCH	-11.5793	134.0798	14	-2.9279	8.5728	5	56.4052	12	56.6802	10	0.4670	**0.5354**
Normal EGARCH	-3.2950	10.8570	6	9.1243	83.2533	10	56.3280	11	56.9543	11	0.4810	**0.5424**
Normal APARCH	-7.8824	62.1319	8	4.1597	17.3033	6	51.9914	8	53.1386	8	0.4550	**0.5485**
t GARCH	**1.7837**	**3.1815**	**2**	4.9286	24.2915	9	79.1033	14	81.9306	13	0.4140	0.4639
t TARCH	-11.5955	134.4553	15	**-2.8202**	**7.9533**	**3**	56.2828	10	56.5890	9	0.4730	**0.5347**
t EGARCH	-3.3188	11.0143	7	9.2004	84.6476	11	56.2029	9	57.3163	12	0.4750	**0.5406**
t APARCH	-8.0164	64.2628	9	4.2452	18.0216	7	51.7354	7	53.0744	7	0.4480	**0.5481**
RiskMetrics	-2.9129	8.4847	5	-2.8736	8.2575	4	93.4855	15	93.5200	15	0.0040	0.4820
Rolling 30 days	**-2.7262**	**7.4323**	**3**	**-2.7262**	**7.4323**	**1**	122.2867	16	122.2867	16	0.0000	0.0000
Rolling 60 days	-2.7337	7.4731	4	**-2.7337**	**7.4731**	**2**	147.5143	17	147.5143	17	0.0000	0.0000

This table shows the Bias, Square Bias, MSE and the corresponding rankings of volatility forecasts based on alternative individual models using the full and split sample approaches. The DGP is a GARCH diffusion volatility model with break in the leverage coefficient of size 2. The in-sample size is 3000 daily observations and 1000 observations are used for out-of-sample evaluation. The parameters of the volatility models for the full sample approach are estimated using all available information in the full sample and for the split sample approach using only the post-break sample. Bias and MSE are multiplied by 1000 and Square Bias by 10^6. CPA-MSE column corresponds to the (i) % of rejections of the null hypothesis of equal predictive ability between the volatility forecasts based on the full and split sample approaches using the CPA test for a confidence level of 5% and (ii) to the percentage of cases (I_{sf}) that the split sample outperforms the full sample approach using the MSE loss function, respectively. Entries in bold are the three best performing methods in forecasting volatility based on MSE, the cases that the percentage of rejections is higher than 50% and the cases that the split sample outperforms on average the full sample approach.

Table 12: Evaluation of alternative volatility forecasts when the DGP is a GARCH diffusion volatility model with break in the leverage coefficient of size 3

	Full Sample			*Split Sample*			*Full Sample*		*Split Sample*		*CPA - MSE*	
	Bias	*Bias^2*	*Rank*	*Bias*	*Bias^2*	*Rank*	*MSE*	*Rank*	*MSE*	*Rank*	*% rej.*	*I_{sf}*
AR(1) - RV	9.3368	87.1758	11	10.3425	0.1070	15	31.8178	6	31.9298	5	**0.6840**	0.4653
AR(5) - RV	8.9333	79.8040	10	9.5658	0.0915	14	28.0865	4	**28.1943**	**2**	0.3990	0.4237
AR(10) - RV	8.9169	79.5105	9	9.5232	0.0907	13	**28.0795**	**3**	**28.2236**	**3**	0.4010	0.4150
AR(15) - RV	8.9024	79.2521	8	9.4849	0.0900	12	28.0883	5	28.2352	4	0.4030	0.4182
HAR - RV	28.2915	800.4067	17	29.3008	0.8585	17	**20.1947**	**2**	**20.5626**	**1**	0.4590	0.4589
LHAR - RV	21.4487	460.0463	16	20.3841	0.4155	16	**18.8183**	**1**	50.5962	12	0.1930	0.3430
Normal GARCH	**0.6989**	**0.4885**	**1**	**3.6168**	**0.0131**	**3**	75.3786	14	78.1909	14	0.4230	0.4624
Normal TARCH	-16.4999	272.2456	14	-5.3516	0.0286	9	48.9879	12	47.4628	11	**0.5150**	**0.5848**
Normal EGARCH	-7.9806	63.6894	6	9.2966	0.0864	11	47.1124	10	43.1426	8	**0.6050**	**0.6240**
Normal APARCH	-13.2336	175.1280	12	**2.7795**	**0.0077**	**1**	43.0335	8	39.4439	6	**0.5660**	**0.6263**
t GARCH	**0.8580**	**0.7362**	**2**	3.8329	0.0147	4	75.3729	13	77.9272	13	0.4300	0.4628
t TARCH	-16.6625	277.6381	15	-5.2548	0.0276	8	48.7166	11	47.2892	10	**0.5060**	**0.5820**
t EGARCH	-8.1523	66.4594	7	9.1658	0.0840	10	46.8424	9	43.6552	9	**0.6040**	**0.6202**
t APARCH	-13.5192	182.7699	13	**2.8502**	**0.0081**	**2**	42.6859	7	39.5393	7	**0.5580**	**0.6245**
RiskMetrics	-4.2530	18.0882	5	-4.2184	0.0178	7	89.8372	15	89.8734	15	0.0050	0.4777
Rolling 30 days	**-4.0363**	**16.2917**	**3**	-4.0363	0.0163	5	119.1892	16	119.1892	16	0.0000	0.0000
Rolling 60 days	-4.0398	16.3200	4	-4.0398	0.0163	6	145.5981	17	145.5981	17	0.0000	0.0000

This table shows the Bias, Square Bias, MSE and the corresponding rankings of volatility forecasts based on alternative individual models using the full and split sample approaches. The DGP is a GARCH diffusion volatility model with break in the leverage coefficient of size 3. The in-sample size is 3000 daily observations and 1000 observations are used for out-of-sample evaluation. The parameters of the volatility models for the full sample approach are estimated using all available information in the full sample and for the split sample approach using only the post-break sample. Bias and MSE are multiplied by 1000 and Square Bias by 10^6. CPA-MSE column corresponds to the (i) % of rejections of the null hypothesis of equal predictive ability between the volatility forecasts based on the full and split sample approaches using the CPA test for a confidence level of 5% and (ii) to the percentage of cases (I_{sf}) that the split sample outperforms the full sample approach using the MSE loss function, respectively. Entries in bold are the three best performing methods in forecasting volatility based on MSE, the cases that the percentage of rejections is higher than 50% and the cases that the split sample outperforms on average the full sample approach.

Table 13: Evaluation of alternative volatility forecasts when the DGP is a GARCH diffusion volatility model with break in the constant of size 2

	Full Sample			*Split Sample*			*Full Sample*		*Split Sample*		*CPA - MSE*	
	Bias	*Bias²*	*Rank*	*Bias*	*Bias²*	*Rank*	*MSE*	*Rank*	*MSE*	*Rank*	*% rej.*	I_{sf}
AR(1) - RV	30.2645	0.9159	7	13.5889	0.1847	15	51.6160	6	50.9544	6	**0.8530**	**0.6151**
AR(5) - RV	21.0127	0.4415	6	11.9070	0.1418	12	45.0834	5	**44.7847**	**3**	**0.7280**	**0.5666**
AR(10) - RV	20.2939	0.4118	5	11.7806	0.1388	11	**45.0721**	**3**	44.8133	4	**0.7120**	**0.5658**
AR(15) - RV	19.8595	0.3944	4	11.7041	0.1370	10	45.0784	4	44.8322	5	**0.7060**	**0.5645**
HAR - RV	42.3020	1.7895	13	36.9055	1.3620	17	**31.4010**	**1**	**32.7179**	**2**	**0.6390**	**0.5434**
LHAR - RV	37.7939	1.4284	10	33.3249	1.1105	16	**36.7180**	**2**	**32.4042**	**1**	**0.5320**	**0.6529**
Normal GARCH	36.3824	1.3237	8	7.4083	0.0549	7	135.5474	14	135.7227	13	0.4690	0.4966
Normal TARCH	37.9245	1.4383	11	2.8065	0.0079	4	123.1391	12	120.0322	8	0.4840	**0.5151**
Normal EGARCH	61.1828	3.7433	16	12.4964	0.1562	13	118.5572	9	126.6211	12	**0.5810**	0.4737
Normal APARCH	48.3791	2.3405	15	7.2510	0.0526	6	118.7264	10	122.8142	10	0.4790	0.4931
t GARCH	36.9212	1.3632	9	7.8258	0.0612	9	135.2478	13	135.7859	14	0.4630	0.4962
t TARCH	38.0147	1.4451	12	3.0528	0.0093	5	122.6838	11	119.8587	7	0.4810	**0.5149**
t EGARCH	61.2405	3.7504	17	12.7402	0.1623	14	118.1117	8	126.5858	11	**0.5680**	0.4729
t APARCH	48.3083	2.3337	14	7.4099	0.0549	8	118.0074	7	121.5160	9	0.4820	0.4930
RiskMetrics	**0.7403**	**0.0005**	**1**	**-2.2899**	**0.0052**	**3**	152.0045	15	151.8237	15	0.0030	**0.5168**
Rolling 30 days	**-2.2477**	**0.0051**	**3**	**-2.2477**	**0.0051**	**2**	192.3244	16	192.3244	16	0.0000	0.0000
Rolling 60 days	**-2.1830**	**0.0048**	**2**	**-2.1830**	**0.0048**	**1**	231.7480	17	231.7480	17	0.0000	0.0000

This table shows the Bias, Square Bias, MSE and the corresponding rankings of volatility forecasts based on alternative individual models using the full and split sample approaches. The DGP is a GARCH diffusion volatility model with break in the constant of size 2. The in-sample size is 3000 daily observations and 1000 observations are used for out-of-sample evaluation. The parameters of the volatility models for the full sample approach are estimated using all available information in the full sample and for the split sample approach using only the post-break sample. Bias and MSE are multiplied by 1000 and Square Bias by 10^6. CPA-MSE column corresponds to the (i) % of rejections of the null hypothesis of equal predictive ability between the volatility forecasts based on the full and split sample approaches using the CPA test for a confidence level of 5% and (ii) to the percentage of cases (I_{sf}) that the split sample outperforms the full sample approach using the MSE loss function, respectively. Entries in bold are the three best performing methods in forecasting volatility based on MSE, the cases that the percentage of rejections is higher than 50% and the cases that the split sample outperforms on average the full sample approach.

Table 14: Evaluation of alternative volatility forecasts when the DGP is a GARCH diffusion volatility model with break in the constant of size 3

	Full Sample			Split Sample			Full Sample		Split Sample		CPA - MSE	
	Bias	*Bias^2*	*Rank*	*Bias*	*Bias^2*	*Rank*	*MSE*	*Rank*	*MSE*	*Rank*	*% rej.*	*I_{sf}*
AR(1) - RV	35.9492	1.2923	13	15.5581	0.2421	15	65.8371	6	64.5155	6	**0.8250**	**0.6599**
AR(5) - RV	24.2645	0.5888	10	13.4426	0.1807	12	57.3387	5	**56.7013**	**3**	**0.7280**	**0.5616**
AR(10) - RV	23.2143	0.5389	9	13.2782	0.1763	11	**57.3240**	**3**	56.7378	4	**0.7220**	**0.5617**
AR(15) - RV	22.5343	0.5078	8	13.1784	0.1737	10	57.3273	4	56.7611	5	**0.7220**	**0.5619**
HAR - RV	43.9274	1.9296	15	41.6362	1.7336	17	**38.5437**	**1**	**41.5388**	**2**	**0.5220**	0.4969
LHAR - RV	37.6308	1.4161	14	37.5715	1.4116	16	**47.5727**	**2**	**41.0813**	**1**	0.4390	**0.6500**
Normal GARCH	14.5455	0.2116	6	8.5210	0.0726	6	180.9746	14	171.6365	13	0.3230	**0.5330**
Normal TARCH	12.2466	0.1500	4	3.3565	0.0113	4	170.5950	12	151.6938	8	0.3590	**0.5590**
Normal EGARCH	44.7225	2.0001	17	14.1315	0.1997	13	150.2948	8	160.0988	11	0.4330	**0.5099**
Normal APARCH	26.4481	0.6995	12	8.7437	0.0765	7	160.8053	10	155.8588	10	0.3680	**0.5349**
t GARCH	15.3604	0.2359	7	8.9656	0.0804	9	180.4785	13	171.7056	14	0.3180	**0.5320**
t TARCH	12.6879	0.1610	5	3.6067	0.0130	5	169.9801	11	151.4646	7	0.3520	**0.5587**
t EGARCH	44.6246	1.9914	16	14.4098	0.2076	14	149.4956	7	160.2127	12	0.4350	**0.5092**
t APARCH	26.1122	0.6818	11	8.8957	0.0791	8	160.0696	9	153.9970	9	0.3690	**0.5353**
RiskMetrics	**2.0857**	**0.0044**	**1**	**-2.4760**	**0.0061**	**3**	192.6955	15	192.2006	15	0.0030	**0.5195**
Rolling 30 days	**-2.4690**	**0.0061**	**3**	**-2.4690**	**0.0061**	**2**	243.4843	16	243.4843	16	0.0000	0.0000
Rolling 60 days	**-2.3852**	**0.0057**	**2**	**-2.3852**	**0.0057**	**1**	293.3970	17	293.3970	17	0.0000	0.0000

This table shows the Bias, Square Bias, MSE and the corresponding rankings of volatility forecasts based on alternative individual models using the full and split sample approaches. The DGP is a GARCH diffusion volatility model with break in the constant of size 3. The in-sample size is 3000 daily observations and 1000 observations are used for out-of-sample evaluation. The parameters of the volatility models for the full sample approach are estimated using all available information in the full sample and for the split sample approach using only the post-break sample. Bias and MSE are multiplied by 1000 and Square Bias by 10^6. CPA-MSE column corresponds to the (i) % of rejections of the null hypothesis of equal predictive ability between the volatility forecasts based on the full and split sample approaches using the CPA test for a confidence level of 5% and (ii) to the percentage of cases (I_{sf}) that the split sample outperforms the full sample approach using the MSE loss function, respectively. Entries in bold are the three best performing methods in forecasting volatility based on MSE, the cases that the percentage of rejections is higher than 50% and the cases that the split sample outperforms on average the full sample approach.

Table 15: Comparison of alternative volatility forecasts based on the full and the split sample approaches using the CPA test

	No break	*Break leverage Size=2*	*Break leverage Size = 3*	*Break constant Size = 2*	*Break constant Size = 3*
AR(1) - RV	**0.7170**	**0.6890**	**0.6840**	**0.8530**	**0.8250**
	(0.4677)	(0.4658)	(0.4653)	**(0.6151)**	**(0.6599)**
AR(5) - RV	0.3990	0.3770	0.3990	**0.7280**	**0.7280**
	(0.4153)	(0.4170)	(0.4237)	**(0.5666)**	**(0.5616)**
AR(10) - RV	0.3840	0.3980	0.4010	**0.7120**	**0.7220**
	(0.4112)	(0.4094)	(0.4150)	**(0.5658)**	**(0.5617)**
AR(15) - RV	0.3940	0.4060	0.4030	**0.7060**	**0.7220**
	(0.4110)	(0.4126)	(0.4182)	**(0.5645)**	**(0.5619)**
HAR - RV	0.4730	0.4520	0.4590	**0.6390**	**0.5220**
	(0.4561)	(0.4589)	(0.4589)	**(0.5434)**	(0.4969)
LHAR - RV	0.3650	0.2150	0.1930	**0.5320**	0.4390
	(0.4172)	(0.2947)	(0.3430)	**(0.6529)**	**(0.6500)**
Normal GARCH	0.4450	0.4110	0.4230	0.4690	0.3230
	(0.4616)	(0.4634)	(0.4624)	(0.4966)	**(0.5330)**
Normal TARCH	0.4300	0.4670	**0.5150**	0.4840	0.3590
	(0.4517)	**(0.5354)**	**(0.5848)**	**(0.5151)**	**(0.5590)**
Normal EGARCH	0.4300	0.4810	**0.6050**	**0.5810**	0.4330
	(0.4488)	**(0.5424)**	**(0.6240)**	(0.4737)	**(0.5099)**
Normal APARCH	0.4220	0.4550	**0.5660**	0.4790	0.3680
	(0.4391)	**(0.5485)**	**(0.6263)**	(0.4931)	**(0.5349)**
t GARCH	0.4490	0.4140	0.4300	0.4630	0.3180
	(0.4608)	(0.4639)	(0.4628)	(0.4962)	**(0.5320)**
t TARCH	0.4350	0.4730	**0.5060**	0.4810	0.3520
	(0.4524)	**(0.5347)**	**(0.5820)**	**(0.5149)**	**(0.5587)**
t EGARCH	0.4320	0.4750	**0.6040**	**0.5680**	0.4350
	(0.4497)	**(0.5406)**	**(0.6202)**	(0.4729)	**(0.5092)**
t APARCH	0.4250	0.4480	**0.5580**	0.4820	0.3690
	(0.4393)	**(0.5481)**	**(0.6245)**	(0.4930)	**(0.5353)**
RiskMetrics	0.0030	0.0040	0.0050	0.0030	0.0030
	(0.4751)	(0.4820)	(0.4777)	**(0.5168)**	**(0.5195)**

This table shows the percentage of rejections of the CPA test based on the Square Error (SE) loss function of the null hypothesis of equal predictive ability between the volatility forecasts based on the full and the split sample approaches for 5% confidence level. The entries in parenthesis are the percentage of cases that the split sample outperforms the full sample approach. Entries in bold are the cases that the percentage of rejections is higher than 50% and the cases that the split sample outperforms on average the full sample.

Table 16: Parameters of the GARCH diffusion DGP

Simulation exercise 1: GARCH diffusion without breaks

Out-of-sample evaluation period		Low volatility		Crisis	
Subsample	Sample size	Characteristics	α_2	Characteristics	α_2
Pre-sample	100	Low volatility	0.636	Crisis	1.908
Estimation	750	Low volatility	0.636	Crisis	1.908
1 -5	3000	Low volatility	0.636	Crisis	1.908
Out-of-sample	500	Low volatility	0.636	Crisis	1.908

Simulation exercise 2: GARCH diffusion with multiple breaks

Out-of-sample evaluation period		Low volatility		Crisis	
Subsample	Sample size	Characteristics	α_2	Characteristics	α_2
Pre-sample	100	Low volatility	0.636	Crisis	1.908
Estimation	750	Low volatility	0.636	Crisis	1.908
1	500	Crisis	2.544	Crisis	2.544
2	500	Low volatility	0.636	Low volatility	0.636
3	750	High volatility	1.272	High volatility	1.272
4	750	Low volatility	0.318	Low volatility	0.318
5	500	Crisis	1.908	Crisis	1.908
Out-of-sample	500	Low volatility	0.636	Crisis	1.908

This table shows the different values of the constant of the GARCH diffusion DGP in the various subsamples that resemble low volatility, high volatility and crisis subsamples. The GARCH diffusion DGP is given by the following formula: $d\log S_t = \sigma_t \left[\rho_1 dW_{1t} + \sqrt{1-\rho_1^2} dW_{2t}\right]$, where $d\sigma_t^2 = a_1\left(\alpha_2 - \sigma_t^2\right) dt + \alpha_3 \sigma_t^2 dW_{1t}$, $\rho_1 = -.576,\ \alpha_1 = 0.035\ and\ \alpha_3 = 0.35$. The constant parameter α_2 changes for the various subsamples of the simulation process as shown in the table above. The pre-sample period is used to avoid any possible bias in the initial value of the GARCH diffusion process. The estimation period is used for the estimation of the parameters of the volatility models. Subsamples 1-5 are used for the estimation of the combination weights of the FFC and FC-RS approaches. The out-of-sample period is used for the evaluation of the alternative volatility forecasts.

Table 17: Evaluation of the performance of alternative volatility forecasts based on the GARCH diffusion DGP without breaks

Out-of-sample period	**Low volatility**						**Crisis**					
Method	*MSE*	*Rank*	*HR m=-1*	*Rank*	*QLIKE*	*Rank*	*MSE*	*Rank*	*HR m=-1*	*Rank*	*QLIKE*	*Rank*
AR(1)-RV	0.0245	8	0.0163	8	0.0264	8	0.2207	9	0.0489	9	0.0265	9
AR(5)-RV	0.0213	5	0.0143	5	0.0232	5	0.1942	5	0.0429	6	0.0234	7
AR(10)-RV	0.0216	7	0.0143	6	0.0233	6	0.1944	6	0.0431	7	0.0234	6
AR(15)-RV	0.0216	6	0.0143	7	0.0234	7	0.1948	7	0.0431	8	0.0234	8
HAR-RV	0.0145	4	0.0097	4	0.0158	4	0.1314	4	0.0291	4	0.0158	4
LHAR-RV	**0.0134**	**3**	**0.0088**	**3**	**0.0145**	**3**	**0.1212**	**3**	**0.0265**	**3**	**0.0145**	**3**
Normal GARCH	0.0637	19	0.0440	20	0.0702	20	0.5925	20	0.1322	20	0.0703	19
Normal TARCH	0.0571	16	0.0376	15	0.0604	14	0.5175	16	0.1129	14	0.0604	14
Normal EGARCH	0.0560	13	0.0376	16	0.0623	17	0.5102	13	0.1136	16	0.0625	17
Normal APARCH	0.0569	15	0.0379	17	0.0614	16	0.5147	15	0.1141	17	0.0614	15
t GARCH	0.0642	20	0.0440	19	0.0702	19	0.5848	19	0.1320	19	0.0707	20
t TARCH	0.0574	17	0.0375	14	0.0600	13	0.5176	17	0.1135	15	0.0601	13
t EGARCH	0.0565	14	0.0369	13	0.0611	15	0.5124	14	0.1123	13	0.0616	16
t APARCH	0.0576	18	0.0383	18	0.0625	18	0.5187	18	0.1151	18	0.0627	18
RiskMetrics	0.0741	21	0.0503	21	0.0832	21	0.6683	21	0.1522	21	0.0835	21
Rolling 30 days	0.0902	22	0.0636	22	0.1041	22	0.8129	22	0.1914	22	0.1041	22
Rolling 60 days	0.1032	23	0.0707	23	0.1131	23	0.9541	23	0.2153	23	0.1131	23
Mean	0.0325	10	0.0219	10	0.0356	10	0.2929	10	0.0661	10	0.0359	10
Median	0.0394	12	0.0267	12	0.0433	12	0.3548	12	0.0811	12	0.0434	12
Geometric Mean	0.0325	11	0.0219	11	0.0356	11	0.2929	11	0.0661	11	0.0359	11
FC RS	0.0292	9	0.0186	9	0.0282	9	0.1982	8	0.0408	5	0.0220	5
FFC 1 (detects no break)	**0.0132**	**2**	**0.0086**	**2**	**0.0138**	**1**	**0.1187**	**2**	**0.0257**	**1**	**0.0138**	**1**
FFC 2 (detects one break)	**0.0131**	**1**	**0.0085**	**1**	**0.0138**	**2**	**0.1183**	**1**	**0.0257**	**2**	**0.0138**	**2**

This table shows the performance of volatility forecasts given by individual models, simple averaging methods and other forecast combination methods, such as Forecast Combinations under Regime Switching (FC-RS) and the Flexible Forecast Combination Approach (FFC). The DGP is a GARCH diffusion process without a break that resembles a low volatility or crisis period depending on the choice of the constant parameter. The losses are measured based on three different loss functions, namely MSE, Homogeneous Robust (HR) for b=-1 and QLIKE loss functions. The estimation period consists of 3000 simulated days and the out-of-sample period 500 days. We use 250 simulations. FFC1 corresponds to the case that the FFC approach detects no break and the FFC2 to the case that there is misspecification and the FFC approach detects a break in the middle of the estimation period.

Table 18: Evaluation of the performance of alternative volatility forecasts based on the GARCH diffusion DGP with multiple breaks

Out-of-sample period	**Low volatility**						**Crisis**					
Method	*MSE*	*Rank*	*HR m=-1*	*Rank*	*QLIKE*	*Rank*	*MSE*	*Rank*	*HR m=-1*	*Rank*	*QLIKE*	*Rank*
AR(1)-RV	0.0319	11	0.0179	11	0.0265	11	0.2205	11	0.0488	11	0.0264	11
AR(5)-RV	0.0282	10	0.0158	8	0.0232	8	0.1930	8	0.0429	8	0.0232	8
AR(10)-RV	0.0281	9	0.0159	9	0.0233	9	0.1943	9	0.0430	9	0.0233	9
AR(15)-RV	0.0281	8	0.0159	10	0.0233	10	0.1945	10	0.0430	10	0.0234	10
HAR-RV	0.0189	7	0.0106	7	0.0159	7	0.1308	7	0.0291	7	0.0158	7
LHAR-RV	0.0182	4	0.0101	6	0.0146	5	0.1209	6	**0.0263**	**3**	**0.0145**	**2**
Normal GARCH	0.0814	23	0.0498	22	0.0746	23	0.5829	22	0.1321	23	0.0706	22
Normal TARCH	0.0710	16	0.0432	17	0.0625	17	0.5156	18	0.1129	16	0.0605	17
Normal EGARCH	0.0747	21	0.0457	21	0.0658	20	0.5064	15	0.1131	18	0.0623	20
Normal APARCH	0.0747	20	0.0450	19	0.0655	18	0.5135	17	0.1140	19	0.0619	19
t GARCH	0.0810	22	0.0501	23	0.0746	22	0.5814	21	0.1320	22	0.0712	23
t TARCH	0.0711	17	0.0432	16	0.0623	16	0.5169	19	0.1130	17	0.0602	16
t EGARCH	0.0742	19	0.0455	20	0.0660	21	0.5105	16	0.1114	15	0.0612	18
t APARCH	0.0741	18	0.0446	18	0.0657	19	0.5184	20	0.1149	20	0.0628	21
RiskMetrics	0.0988	24	0.0582	24	0.0841	24	0.6675	23	0.1515	24	0.0836	24
Rolling 30 days	0.1188	25	0.0719	25	0.1074	25	0.8122	25	0.1911	25	0.1042	25
Rolling 60 days	0.1551	26	0.0885	26	0.1181	26	0.9413	26	0.2138	26	0.1126	26
Mean	0.0410	12	0.0247	12	0.0362	12	0.2925	12	0.0659	12	0.0359	12
Median	0.0477	15	0.0289	15	0.0427	15	0.3547	14	0.0805	14	0.0435	14
Geometric Mean	0.0410	13	0.0247	13	0.0362	13	0.2925	13	0.0659	13	0.0359	13
FC RS	0.0450	14	0.0271	14	0.0409	14	0.7754	24	0.1246	21	0.0546	15
FFC 1 (detects no break)	**0.0182**	**2**	**0.0098**	**3**	0.0144	4	**0.1195**	**3**	**0.0263**	**2**	**0.0149**	**3**
FFC 2 (estimates breaks)	0.0183	6	**0.0098**	**2**	**0.0144**	**3**	**0.1193**	**1**	0.0264	4	0.0149	4
FFC 3 (- low volatility)	**0.0182**	**3**	0.0100	5	0.0148	6	**0.1195**	**2**	**0.0258**	**1**	**0.0139**	**1**
FFC 4 (- high volatility)	0.0183	5	0.0099	4	**0.0143**	**2**	0.1205	5	0.0265	5	0.0152	6
FFC 5 (- crisis)	**0.0175**	**1**	**0.0097**	**1**	**0.0141**	**1**	0.1202	4	0.0269	6	0.0152	5

This table shows the performance of volatility forecasts given by individual models, simple averaging methods and other forecast combination methods, such as Forecast Combinations under Regime Switching (FC-RS) and the Flexible Forecast Combination Approach (FFC). The DGP is a GARCH diffusion process without a break that resembles a low volatility or crisis period depending on the choice of the constant parameter. The losses are measured based on three different loss functions, namely MSE, Homogeneous Robust (HR) for b=-1 and QLIKE loss functions. The estimation period consists of 3000 simulated days and the out-of-sample period 500 days. We use 250 simulations. FFC1 corresponds to the case that the FFC approach detects no break and the FFC2 to the case that there is misspecification and the FFC approach detects a break in the middle of the estimation period.

Table 19: Dates of structural breaks in Realized Volatility detected by the Flexible Forecast Combination approach

Estimation period	03/02/1986 – 31/12/2003	03/02/1986 – 31/12/2004	03/02/1986 – 30/12/2005	03/02/1986 – 03/01/2007	03/02/1986 – 03/01/2008	03/02/1986 – 02/01/2009	03/02/1986 – 04/01/2010
1	15/04/1991	15/04/1991	15/04/1991	15/04/1991	15/04/1991	26/04/1988	26/04/1988
2	24/06/1997	24/06/1997	24/06/1997	24/06/1997	24/06/1997	07/02/1992	07/02/1992
3	.	.	09/04/2003	11/04/2003	11/04/2003	08/02/1996	08/02/1996
4	.	.	.	.	.	21/07/1998	21/07/1998
5	.	.	.	.	.	11/04/2003	11/04/2003
6	.	.	.	.	.	.	04/01/2008

This table shows the dates of the structural breaks detected by the flexible forecast combination method while the estimation period is updated for every 252 daily observations. The estimation of these breaks is based on the CUSUM type test (Kokoszka and Leipus, 1998, 2000).

Table 20: Rankings of alternative volatility forecasts for the two subsamples of the S&P 500 Index

	1st subsample: Jan 2, 2004 – Jan 3, 2008						2nd subsample: Jan 4, 2008 – Jun 30, 2010					
Method	*Square Error*	*Rank*	*Robust b=-1*	*Rank*	*QLIKE*	*Rank*	*Square Error*	*Rank*	*Robust b=-1*	*Rank*	*QLIKE*	*Rank*
AR(1) - RV	0.131	15	0.088	11	0.196	11	11.165	8	0.708	12	0.276	15
AR(5) - RV	0.105	9	0.072	8	0.168	7	11.137	7	0.574	6	0.196	4
AR(10) - RV	0.103	8	0.069	6	0.162	5	11.943	11	0.599	7	0.196	5
AR(15) - RV	0.102	6	0.069	5	0.162	6	11.917	10	0.600	8	0.197	6
HAR - RV	**0.095**	**3**	**0.064**	**3**	**0.150**	**3**	9.593	4	0.483	3	**0.154**	**3**
LHAR - RV	**0.076**	**1**	**0.056**	**1**	**0.139**	**2**	19.554	18	0.473	2	**0.144**	**2**
Normal GARCH	0.144	18	0.111	19	0.257	20	19.048	15	1.052	18	0.329	18
Normal TARCH	0.119	11	0.093	12	0.221	14	15.152	14	0.824	15	0.267	14
Normal EGARCH	0.126	14	0.095	15	0.216	13	**7.821**	**1**	0.533	4	0.226	8
Normal APARCH	0.122	12	0.094	14	0.221	15	14.559	13	0.780	14	0.250	12
t GARCH	0.152	19	0.113	20	0.258	21	19.095	16	1.059	19	0.333	20
t TARCH	0.137	17	0.100	17	0.227	17	19.876	19	0.991	16	0.289	17
t EGARCH	0.126	13	0.094	13	0.215	12	9.605	5	0.631	9	0.243	11
t APARCH	0.134	16	0.099	16	0.226	16	21.461	20	0.999	17	0.277	16
RiskMetrics	0.166	20	0.110	18	0.234	18	19.310	17	1.100	20	0.332	19
Rolling 30 days	0.208	22	0.128	22	0.249	19	22.736	21	1.235	21	0.360	21
Rolling 60 days	0.177	21	0.125	21	0.266	22	26.379	22	1.549	22	0.444	22
Mean	0.103	7	0.077	9	0.186	9	10.562	6	0.634	10	0.229	9
Median	0.107	10	0.079	10	0.191	10	12.389	12	0.699	11	0.240	10
Geometric Mean	0.098	4	0.072	7	0.175	8	**9.461**	**3**	0.553	5	0.205	7
FC-RS	0.101	5	0.069	4	0.160	4	11.388	9	0.738	13	0.254	13
FFC SE	**0.093**	**2**	.	.	.	.	**8.519**	**2**	.	.	.	.
FFC Robust b=-1	.	.	**0.060**	**2**	.	.	.	.	**0.393**	**1**	.	.
FFC QLIKE	.	.	.	.	**0.132**	**1**	.	.	.	.	**0.130**	**1**

This table shows the rankings of alternative volatility forecasts of individual models (AR-RV, HAR-RV, LHAR-RV, GARCH type, RiskMetrics and rolling window, simple averaging methods (Mean, Median and Geometric Mean), Forecast Combinations under Regime Switching (FC-RS), Flexible Forecast Combinations based on the Square Error, Homogeneous Robust for b=-1 and QLIKE loss functions (FFC SE, FFC Robust b=-1 and FFC QLIKE). We evaluate the performance of the aforementioned methods using two subsamples of the S&P 500 Index, the low volatility subsample from January 2, 2004 to January 3, 2008 and the high volatility subsample from January 4, 2008 to June 30, 2010.

Table 21: Rankings of alternative volatility forecasts for the full sample of the S&P 500 Index

	Full sample: Jan 2, 2004 – Jun 30, 2010					
Method	*Square Error*	*Rank*	*Robust b=-1*	*Rank*	*QLIKE*	*Rank*
AR(1) - RV	4.362	8	0.326	13	0.227	13
AR(5) - RV	4.335	7	0.264	6	0.179	6
AR(10) - RV	4.643	11	0.272	7	0.175	4
AR(15) - RV	4.633	10	0.273	8	0.175	5
HAR - RV	3.738	4	**0.225**	**3**	**0.151**	**3**
LHAR - RV	7.545	18	**0.216**	**2**	**0.141**	**2**
Normal GARCH	7.393	15	0.472	18	0.284	19
Normal TARCH	5.884	14	0.373	15	0.239	15
Normal EGARCH	**3.077**	**1**	0.263	5	0.220	11
Normal APARCH	5.658	13	0.357	14	0.233	14
t GARCH	7.416	16	0.476	19	0.287	20
t TARCH	7.707	19	0.442	16	0.251	17
t EGARCH	3.761	5	0.300	10	0.226	12
t APARCH	8.312	20	0.444	17	0.246	16
RiskMetrics	7.507	17	0.490	20	0.271	18
Rolling 30 days	8.847	21	0.552	21	0.291	21
Rolling 60 days	10.225	22	0.671	22	0.334	22
Mean	4.114	6	0.290	9	0.203	9
Median	4.817	12	0.317	11	0.210	10
Geometric Mean	**3.688**	**3**	0.256	4	0.186	7
FC-RS	4.430	9	0.326	12	0.196	8
FFC SE	**3.324**	**2**	.	.	.	.
FFC Robust b=-1	.	.	**0.188**	**1**	.	.
FFC QLIKE	.	.	.	.	**0.131**	**1**

This table shows the rankings of alternative volatility forecasts of individual models (AR-RV, HAR-RV, LHAR-RV, GARCH type, RiskMetrics and rolling window)), simple averaging methods (Mean, Median and Geometric Mean), Forecast Combinations under Regime Switching (FC-RS), Flexible Forecast Combinations based on the Square Error, Homogeneous Robust for b=-1 and QLIKE loss functions (FFC SE, FFC Robust b=-1 and FFC QLIKE). We evaluate the performance of the aforementioned methods using the full sample of the S&P 500 Index from January 2, 2004 to June 30, 2010.

Table 22: Comparison of forecast combination methods with individual forecasts using the CPA test based on the Square Error loss function for the low volatility subsample of the S&P 500 Index

A. Square Error loss function

Method	*Mean*	*Median*	*Geometric Mean*	*FC-RS*	*FFC SE*
AR(1) - RV	0.183	0.225	**0.087**$^{+}$	**0.003**$^{+}$	**0.000**$^{+}$
	(0.896)	(0.880)	**(0.948)**	**(0.989)**	**(0.984)**
AR(5) - RV	0.199	0.166	0.239	0.742	0.134
	(0.407)	(0.235)	(0.627)	(0.867)	(0.942)
AR(10) - RV	**0.075**$^{-}$	**0.068**$^{-}$	0.240	0.670	0.312
	(0.313)	**(0.213)**	(0.508)	(0.741)	(0.929)
AR(15) - RV	**0.062**$^{-}$	**0.055**$^{-}$	0.205	0.663	0.339
	(0.297)	**(0.204)**	(0.473)	(0.697)	(0.928)
HAR - RV	**0.016**$^{-}$	**0.031**$^{-}$	**0.037**$^{-}$	**0.042**$^{-}$	**0.057**$^{+}$
	(0.181)	**(0.146)**	**(0.234)**	**(0.076)**	**(0.894)**
LHAR - RV	**0.012**$^{-}$	**0.005**$^{-}$	**0.034**$^{-}$	**0.070**$^{-}$	0.184
	(0.000)	**(0.013)**	**(0.020)**	**(0.006)**	(0.014)
Normal GARCH	**0.000**$^{+}$	**0.000**$^{+}$	**0.000**$^{+}$	**0.000**$^{+}$	**0.000**$^{+}$
	(0.966)	**(0.957)**	**(0.959)**	**(0.871)**	**(0.894)**
Normal TARCH	**0.001**$^{+}$	**0.003**$^{+}$	**0.001**$^{+}$	**0.003**$^{+}$	**0.002**$^{+}$
	(0.925)	**(0.930)**	**(0.937)**	**(0.857)**	**(0.907)**
Normal EGARCH	**0.004**$^{+}$	**0.005**$^{+}$	**0.001**$^{+}$	**0.006**$^{+}$	**0.002**$^{+}$
	(0.849)	**(0.835)**	**(0.890)**	**(0.894)**	**(0.938)**
Normal APARCH	**0.000**$^{+}$	**0.001**$^{+}$	**0.000**$^{+}$	**0.005**$^{+}$	**0.005**$^{+}$
	(0.970)	**(0.966)**	**(0.967)**	**(0.879)**	**(0.939)**
t GARCH	**0.000**$^{+}$	**0.000**$^{+}$	**0.000**$^{+}$	**0.003**$^{+}$	**0.000**$^{+}$
	(0.921)	**(0.938)**	**(0.924)**	**(0.867)**	**(0.889)**
t TARCH	**0.004**$^{+}$	**0.005**$^{+}$	**0.003**$^{+}$	**0.009**$^{+}$	**0.007**$^{+}$
	(0.956)	**(0.950)**	**(0.956)**	**(0.905)**	**(0.933)**
t EGARCH	**0.002**$^{+}$	**0.002**$^{+}$	**0.001**$^{+}$	**0.007**$^{+}$	**0.004**$^{+}$
	(0.868)	**(0.862)**	**(0.911)**	**(0.901)**	**(0.938)**
t APARCH	**0.005**$^{+}$	**0.008**$^{+}$	**0.004**$^{+}$	**0.009**$^{+}$	**0.008**$^{+}$
	(0.952)	**(0.958)**	**(0.958)**	**(0.904)**	**(0.932)**
RiskMetrics	**0.000**$^{+}$	**0.000**$^{+}$	**0.000**$^{+}$	**0.008**$^{+}$	**0.005**$^{+}$
	(0.867)	**(0.850)**	**(0.893)**	**(0.880)**	**(0.901)**
Rolling 30 days	**0.000**$^{+}$	**0.000**$^{+}$	**0.000**$^{+}$	**0.001**$^{+}$	**0.000**$^{+}$
	(0.899)	**(0.889)**	**(0.924)**	**(0.891)**	**(0.905)**
Rolling 60 days	**0.000**$^{+}$	**0.000**$^{+}$	**0.000**$^{+}$	**0.000**$^{+}$	**0.000**$^{+}$
	(0.932)	**(0.924)**	**(0.940)**	**(0.904)**	**(0.917)**

This table shows the p-values of the null hypothesis of the CPA test of equal predictive ability of alternative volatility predictions given by forecast combinations with individual models using the low volatility subsample of the S&P 500 Index (January 2, 2004 to January 3, 2008). We consider three different loss functions for the CPA test, namely the Square Error, Homogeneous Robust for b=-1 and QLIKE. The entries in parenthesis are the percentage of cases that the column method outperforms the row method in forecasting volatility. Entries in bold correspond to the cases that the null hypothesis of the CPA test is rejected. The sign + (-) indicates that column method significantly outperforms (is outperformed) by the row method in forecasting volatility.

Table 23: Comparison of forecast combination methods with individual forecasts using the CPA test based on the Homogeneous Robust loss function for $b = -1$ for the low volatility subsample of the S&P 500 Index

B. Homogeneous Robust loss function for b=-1

Method	*Mean*	*Median*	*Geometric Mean*	*FC-RS*	*FFC Robust b=-1*
AR(1) - RV	0.192	0.224	**0.075^{+}**	**0.000^{+}**	**0.000^{+}**
	(0.672)	(0.628)	**(0.842)**	**(0.909)**	**(1.000)**
AR(5) - RV	**0.026^{-}**	**0.011^{-}**	0.148	0.743	**0.001^{+}**
	(0.249)	**(0.208)**	(0.357)	(0.935)	**(0.985)**
AR(10) - RV	**0.001^{-}**	**0.001^{-}**	**0.009^{-}**	0.849	**0.007^{+}**
	(0.225)	**(0.189)**	**(0.282)**	(0.673)	**(0.970)**
AR(15) - RV	**0.001^{-}**	**0.001^{-}**	**0.008^{-}**	0.816	**0.010^{+}**
	(0.236)	**(0.196)**	**(0.287)**	(0.621)	**(0.964)**
HAR - RV	**0.000^{-}**	**0.000^{-}**	**0.001^{-}**	**0.030^{-}**	**0.000^{+}**
	(0.182)	**(0.173)**	**(0.200)**	**(0.022)**	**(0.992)**
LHAR - RV	**0.000^{-}**	**0.000^{-}**	**0.000^{-}**	**0.002^{-}**	**0.096^{+}**
	(0.024)	**(0.046)**	**(0.000)**	**(0.006)**	**(0.100)**
Normal GARCH	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**
	(0.955)	**(0.947)**	**(0.954)**	**(0.861)**	**(0.905)**
Normal TARCH	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**
	(0.880)	**(0.909)**	**(0.905)**	**(0.831)**	**(0.865)**
Normal EGARCH	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**
	(0.759)	**(0.764)**	**(0.804)**	**(0.827)**	**(0.850)**
Normal APARCH	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**
	(0.899)	**(0.927)**	**(0.932)**	**(0.852)**	**(0.904)**
t GARCH	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**
	(0.922)	**(0.937)**	**(0.923)**	**(0.853)**	**(0.889)**
t TARCH	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**
	(0.908)	**(0.920)**	**(0.919)**	**(0.852)**	**(0.879)**
t EGARCH	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**
	(0.728)	**(0.737)**	**(0.785)**	**(0.828)**	**(0.861)**
t APARCH	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**
	(0.903)	**(0.929)**	**(0.929)**	**(0.854)**	**(0.894)**
RiskMetrics	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**
	(0.787)	**(0.758)**	**(0.836)**	**(0.842)**	**(0.875)**
Rolling 30 days	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**
	(0.771)	**(0.785)**	**(0.835)**	**(0.859)**	**(0.881)**
Rolling 60 days	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**
	(0.877)	**(0.874)**	**(0.906)**	**(0.873)**	**(0.908)**

This table shows the p-values of the null hypothesis of the CPA test of equal predictive ability of alternative volatility predictions given by forecast combinations with individual models using the low volatility subsample of the S&P 500 Index (January 2, 2004 to January 3, 2008). We consider three different loss functions for the CPA test, namely the Square Error, Homogeneous Robust for b=-1 and QLIKE. The entries in parenthesis are the percentage of cases that the column method outperforms the row method in forecasting volatility. Entries in bold correspond to the cases that the null hypothesis of the CPA test is rejected. The sign + (-) indicates that column method significantly outperforms (is outperformed) by the row method in forecasting volatility.

Table 24: Comparison of forecast combination methods with individual forecasts using the CPA test based on the QLIKE loss function for the low volatility subsample of the S&P 500 Index

C. QLIKE loss function					
Method	*Mean*	*Median*	*Geometric Mean*	*FC-RS*	*FFC QLIKE*
AR(1) - RV	0.304	0.258	0.184	**0.000$^+$**	**0.000$^+$**
	(0.507)	(0.487)	(0.653)	**(0.840)**	**(0.990)**
AR(5) - RV	**0.000$^-$**	**0.000$^-$**	**0.002$^-$**	0.344	**0.000$^+$**
	(0.233)	**(0.206)**	**(0.296)**	(0.638)	**(0.980)**
AR(10) - RV	**0.000$^-$**	**0.000$^-$**	**0.000$^-$**	0.695	**0.000$^+$**
	(0.211)	**(0.190)**	**(0.254)**	(0.507)	**(0.871)**
AR(15) - RV	**0.000$^-$**	**0.000$^-$**	**0.000$^-$**	0.730	**0.000$^+$**
	(0.223)	**(0.197)**	**(0.259)**	(0.530)	**(0.934)**
HAR - RV	**0.000$^-$**	**0.000$^-$**	**0.000$^-$**	**0.028$^-$**	**0.000$^+$**
	(0.167)	**(0.160)**	**(0.186)**	**(0.064)**	**(0.991)**
LHAR - RV	**0.000$^-$**	**0.000$^-$**	**0.000$^-$**	**0.000$^-$**	**0.004$^+$**
	(0.076)	**(0.085)**	**(0.058)**	**(0.000)**	**(0.996)**
Normal GARCH	**0.000$^+$**	**0.000$^+$**	**0.000$^+$**	**0.000$^+$**	**0.000$^+$**
	(0.952)	**(0.946)**	**(0.954)**	**(0.880)**	**(0.952)**
Normal TARCH	**0.000$^+$**	**0.000$^+$**	**0.000$^+$**	**0.000$^+$**	**0.000$^+$**
	(0.866)	**(0.909)**	**(0.899)**	**(0.827)**	**(0.904)**
Normal EGARCH	**0.000$^+$**	**0.000$^+$**	**0.000$^+$**	**0.000$^+$**	**0.000$^+$**
	(0.733)	**(0.744)**	**(0.797)**	**(0.838)**	**(0.899)**
Normal APARCH	**0.000$^+$**	**0.000$^+$**	**0.000$^+$**	**0.000$^+$**	**0.000$^+$**
	(0.857)	**(0.921)**	**(0.895)**	**(0.849)**	**(0.930)**
t GARCH	**0.000$^+$**	**0.000$^+$**	**0.000$^+$**	**0.000$^+$**	**0.000$^+$**
	(0.950)	**(0.943)**	**(0.942)**	**(0.874)**	**(0.939)**
t TARCH	**0.000$^+$**	**0.000$^+$**	**0.000$^+$**	**0.000$^+$**	**0.000$^+$**
	(0.875)	**(0.914)**	**(0.904)**	**(0.837)**	**(0.906)**
t EGARCH	**0.000$^+$**	**0.000$^+$**	**0.000$^+$**	**0.000$^+$**	**0.000$^+$**
	(0.707)	**(0.724)**	**(0.758)**	**(0.832)**	**(0.899)**
t APARCH	**0.000$^+$**	**0.000$^+$**	**0.000$^+$**	**0.000$^+$**	**0.000$^+$**
	(0.861)	**(0.920)**	**(0.898)**	**(0.841)**	**(0.923)**
RiskMetrics	**0.000$^+$**	**0.000$^+$**	**0.000$^+$**	**0.000$^+$**	**0.000$^+$**
	(0.841)	**(0.789)**	**(0.866)**	**(0.816)**	**(0.897)**
Rolling 30 days	**0.000$^+$**	**0.000$^+$**	**0.000$^+$**	**0.000$^+$**	**0.000$^+$**
	(0.727)	**(0.725)**	**(0.768)**	**(0.834)**	**(0.897)**
Rolling 60 days	**0.000$^+$**	**0.000$^+$**	**0.000$^+$**	**0.000$^+$**	**0.000$^+$**
	(0.811)	**(0.802)**	**(0.843)**	**(0.860)**	**(0.930)**

This table shows the p-values of the null hypothesis of the CPA test of equal predictive ability of alternative volatility predictions given by forecast combinations with individual models using the low volatility subsample of the S&P 500 Index (January 2, 2004 to January 3, 2008). We consider three different loss functions for the CPA test, namely the Square Error, Homogeneous Robust for b=-1 and QLIKE. The entries in parenthesis are the percentage of cases that the column method outperforms the row method in forecasting volatility. Entries in bold correspond to the cases that the null hypothesis of the CPA test is rejected. The sign + (-) indicates that column method significantly outperforms (is outperformed) by the row method in forecasting volatility.

Table 25: Comparison of forecast combination methods with individual forecasts using the CPA test based on the Square Error loss function for the high volatility subsample of the S&P 500 Index

A. Square loss function

Method	*Mean*	*Median*	*Geometric Mean*	*FC-RS*	*FFC SE*
AR(1) - RV	0.111	**0.067$^-$**	0.153	0.916	0.123
	(0.733)	**(0.088)**	(0.845)	(0.062)	(0.976)
AR(5) - RV	0.117	0.126	0.359	0.957	**0.098$^+$**
	(0.682)	(0.094)	(0.898)	(0.050)	**(0.987)**
AR(10) - RV	0.118	0.092	0.282	0.827	**0.029$^+$**
	(0.762)	(0.155)	(0.848)	(0.944)	**(0.984)**
AR(15) - RV	0.149	0.102	0.335	0.887	**0.032$^+$**
	(0.768)	(0.147)	(0.856)	(0.957)	**(0.986)**
HAR - RV	0.114	0.114	0.253	0.230	0.441
	(0.061)	(0.051)	(0.524)	(0.026)	(0.978)
LHAR - RV	0.620	0.536	0.544	0.519	0.315
	(0.976)	(0.976)	(0.978)	(0.981)	(0.979)
Normal GARCH	**0.003$^+$**	**0.001$^+$**	**0.000$^+$**	**0.072$^+$**	**0.001$^+$**
	(0.984)	**(0.982)**	**(0.974)**	**(0.949)**	**(0.966)**
Normal TARCH	**0.007$^+$**	**0.088$^+$**	**0.042$^+$**	**0.053$^+$**	0.112
	(0.957)	**(0.952)**	**(0.957)**	**(0.931)**	(0.952)
Normal EGARCH	**0.063$^-$**	**0.057$^-$**	**0.074$^-$**	0.186	0.575
	(0.133)	**(0.081)**	**(0.091)**	(0.099)	(0.208)
Normal APARCH	**0.068$^+$**	0.204	**0.074$^+$**	0.100	0.120
	(0.942)	(0.970)	**(0.968)**	(0.936)	(0.960)
t GARCH	**0.002$^+$**	**0.000$^+$**	**0.000$^+$**	**0.075$^+$**	**0.001$^+$**
	(0.986)	**(0.979)**	**(0.970)**	**(0.947)**	**(0.965)**
t TARCH	**0.008$^+$**	**0.025$^+$**	**0.013$^+$**	0.108	**0.068$^+$**
	(0.965)	**(0.968)**	**(0.970)**	(0.946)	**(0.962)**
t EGARCH	**0.060$^-$**	**0.097$^-$**	0.579	0.120	0.175
	(0.236)	**(0.075)**	(0.867)	(0.155)	(0.930)
t APARCH	**0.013$^+$**	**0.027$^+$**	**0.014$^+$**	0.149	**0.061$^+$**
	(0.962)	**(0.976)**	**(0.971)**	(0.954)	**(0.965)**
RiskMetrics	**0.002$^+$**	**0.001$^+$**	**0.000$^+$**	**0.045$^+$**	**0.001$^+$**
	(0.990)	**(0.986)**	**(0.978)**	**(0.946)**	**(0.971)**
Rolling 30 days	**0.003$^+$**	**0.002$^+$**	**0.001$^+$**	**0.042$^+$**	**0.002$^+$**
	(0.987)	**(0.968)**	**(0.973)**	**(0.954)**	**(0.978)**
Rolling 60 days	**0.006$^+$**	**0.007$^+$**	**0.001$^+$**	**0.025$^+$**	**0.001$^+$**
	(0.978)	**(0.973)**	**(0.973)**	**(0.955)**	**(0.986)**

This table shows the p-values of the null hypothesis of the CPA test of equal predictive ability of alternative volatility predictions given by forecast combinations with individual models using the high volatility subsample of the S&P 500 Index (January 4, 2008 to June 30, 2010). We consider three different loss functions for the CPA test, namely the Square Error, Homogeneous Robust for b=-1 and QLIKE. The entries in parenthesis are the percentage of cases that the column method outperforms the row method in forecasting volatility. Entries in bold correspond to the cases that the null hypothesis of the CPA test is rejected. The sign + (-) indicates that column method significantly outperforms (is outperformed) by the row method in forecasting volatility.

Table 26: Comparison of forecast combination methods with individual forecasts using the CPA test based on the Homogeneous Robust loss function for $b = -1$ for the high volatility subsample of the S&P 500 Index

B. Homogeneous Robust loss function for b=-1

Method	*Mean*	*Median*	*Geometric Mean*	*FC-RS*	*FFC Robust b=-1*
AR(1) - RV	**0.061**$^{+}$	**0.034**$^{-}$	0.140	0.687	**0.000**$^{+}$
	(0.581)	**(0.487)**	(0.744)	(0.104)	**(0.965)**
AR(5) - RV	**0.021**$^{-}$	**0.035**$^{-}$	**0.065**$^{-}$	0.134	**0.000**$^{+}$
	(0.187)	**(0.152)**	**(0.471)**	(0.062)	**(0.971)**
AR(10) - RV	**0.006**$^{-}$	**0.015**$^{-}$	**0.016**$^{-}$	0.257	**0.001**$^{+}$
	(0.225)	**(0.190)**	**(0.449)**	(0.062)	**(0.978)**
AR(15) - RV	**0.007**$^{-}$	**0.015**$^{-}$	**0.025**$^{-}$	0.267	**0.001**$^{+}$
	(0.220)	**(0.184)**	**(0.457)**	(0.061)	**(0.979)**
HAR - RV	**0.075**$^{-}$	**0.057**$^{-}$	0.169	**0.008**$^{-}$	0.115
	(0.093)	**(0.096)**	(0.131)	**(0.022)**	(0.971)
LHAR - RV	**0.021**$^{-}$	**0.008**$^{-}$	**0.076**$^{-}$	**0.003**$^{-}$	0.460
	(0.157)	**(0.139)**	**(0.366)**	**(0.129)**	(0.898)
Normal GARCH	**0.000**$^{+}$	**0.000**$^{+}$	**0.000**$^{+}$	**0.006**$^{+}$	**0.000**$^{+}$
	(0.986)	**(0.981)**	**(0.962)**	**(0.877)**	**(0.955)**
Normal TARCH	**0.000**$^{+}$	**0.006**$^{+}$	**0.000**$^{+}$	**0.017**$^{+}$	**0.000**$^{+}$
	(0.890)	**(0.930)**	**(0.915)**	**(0.786)**	**(0.920)**
Normal EGARCH	**0.013**$^{-}$	**0.045**$^{-}$	**0.004**$^{-}$	0.130	**0.027**$^{+}$
	(0.390)	**(0.265)**	**(0.497)**	(0.286)	**(0.939)**
Normal APARCH	**0.014**$^{+}$	**0.046**$^{+}$	**0.002**$^{+}$	**0.031**$^{+}$	**0.001**$^{+}$
	(0.819)	**(0.917)**	**(0.922)**	**(0.770)**	**(0.925)**
t GARCH	**0.000**$^{+}$	**0.000**$^{+}$	**0.000**$^{+}$	**0.005**$^{+}$	**0.000**$^{+}$
	(0.984)	**(0.973)**	**(0.954)**	**(0.879)**	**(0.952)**
t TARCH	**0.000**$^{+}$	**0.001**$^{+}$	**0.000**$^{+}$	**0.020**$^{+}$	**0.000**$^{+}$
	(0.922)	**(0.950)**	**(0.942)**	**(0.872)**	**(0.927)**
t EGARCH	**0.008**$^{+}$	**0.025**$^{-}$	**0.025**$^{+}$	**0.087**$^{-}$	**0.003**$^{+}$
	(0.538)	**(0.337)**	**(0.821)**	**(0.404)**	**(0.931)**
t APARCH	**0.000**$^{+}$	**0.001**$^{+}$	**0.000**$^{+}$	**0.039**$^{+}$	**0.000**$^{+}$
	(0.907)	**(0.946)**	**(0.941)**	**(0.885)**	**(0.927)**
RiskMetrics	**0.000**$^{+}$	**0.000**$^{+}$	**0.000**$^{+}$	**0.005**$^{+}$	**0.000**$^{+}$
	(0.987)	**(0.974)**	**(0.966)**	**(0.871)**	**(0.958)**
Rolling 30 days	**0.000**$^{+}$	**0.000**$^{+}$	**0.000**$^{+}$	**0.013**$^{+}$	**0.000**$^{+}$
	(0.971)	**(0.954)**	**(0.958)**	**(0.890)**	**(0.971)**
Rolling 60 days	**0.000**$^{+}$	**0.000**$^{+}$	**0.000**$^{+}$	**0.006**$^{+}$	**0.000**$^{+}$
	(0.973)	**(0.947)**	**(0.966)**	**(0.904)**	**(0.982)**

This table shows the p-values of the null hypothesis of the CPA test of equal predictive ability of alternative volatility predictions given by forecast combinations with individual models using the high volatility subsample of the S&P 500 Index (January 4, 2008 to June 30, 2010). We consider three different loss functions for the CPA test, namely the Square Error, Homogeneous Robust for b=-1 and QLIKE. The entries in parenthesis are the percentage of cases that the column method outperforms the row method in forecasting volatility. Entries in bold correspond to the cases that the null hypothesis of the CPA test is rejected. The sign + (-) indicates that column method significantly outperforms (is outperformed) by the row method in forecasting volatility.

Table 27: Comparison of forecast combination methods with individual forecasts using the CPA test based on the QLIKE loss function for the high volatility subsample of the S&P 500 Index

C. QLIKE loss function

Method	*Mean*	*Median*	*Geometric Mean*	*FC-RS*	*FFC QLIKE*
AR(1) - RV	**0.001^{+}**	**0.001^{+}**	**0.001^{+}**	0.618	**0.000^{+}**
	(0.649)	**(0.617)**	**(0.776)**	(0.989)	**(0.989)**
AR(5) - RV	**0.003^{-}**	**0.000^{-}**	**0.052^{-}**	**0.006^{-}**	**0.000^{+}**
	(0.228)	**(0.212)**	**(0.316)**	**(0.118)**	**(0.989)**
AR(10) - RV	**0.002^{-}**	**0.000^{-}**	**0.041^{-}**	**0.003^{-}**	**0.000^{+}**
	(0.230)	**(0.208)**	**(0.304)**	**(0.128)**	**(0.946)**
AR(15) - RV	**0.003^{-}**	**0.000^{-}**	**0.047^{-}**	**0.003^{-}**	**0.000^{+}**
	(0.230)	**(0.203)**	**(0.297)**	**(0.139)**	**(0.944)**
HAR - RV	**0.001^{-}**	**0.000^{-}**	**0.011^{-}**	**0.000^{-}**	**0.028^{+}**
	(0.123)	**(0.139)**	**(0.137)**	**(0.010)**	**(0.883)**
LHAR - RV	**0.000^{-}**	**0.000^{-}**	**0.002^{-}**	**0.000^{-}**	**0.000^{+}**
	(0.097)	**(0.112)**	**(0.085)**	**(0.005)**	**(0.944)**
Normal GARCH	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**
	(0.968)	**(0.971)**	**(0.958)**	**(0.794)**	**(0.960)**
Normal TARCH	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**
	(0.920)	**(0.947)**	**(0.944)**	**(0.661)**	**(0.946)**
Normal EGARCH	**0.000^{-}**	**0.000^{-}**	**0.000^{+}**	**0.000^{-}**	**0.000^{+}**
	(0.447)	**(0.318)**	**(0.695)**	**(0.466)**	**(0.922)**
Normal APARCH	**0.000^{+}**	**0.002^{+}**	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**
	(0.751)	**(0.808)**	**(0.879)**	**(0.578)**	**(0.931)**
t GARCH	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**
	(0.966)	**(0.971)**	**(0.952)**	**(0.802)**	**(0.960)**
t TARCH	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.001^{+}**	**0.000^{+}**
	(0.935)	**(0.954)**	**(0.942)**	**(0.746)**	**(0.949)**
t EGARCH	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**
	(0.633)	**(0.559)**	**(0.808)**	**(0.538)**	**(0.936)**
t APARCH	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**
	(0.864)	**(0.899)**	**(0.906)**	**(0.704)**	**(0.942)**
RiskMetrics	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**
	(0.930)	**(0.903)**	**(0.915)**	**(0.784)**	**(0.946)**
Rolling 30 days	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**
	(0.907)	**(0.863)**	**(0.909)**	**(0.770)**	**(0.938)**
Rolling 60 days	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**
	(0.935)	**(0.909)**	**(0.936)**	**(0.831)**	**(0.963)**

This table shows the p-values of the null hypothesis of the CPA test of equal predictive ability of alternative volatility predictions given by forecast combinations with individual models using the high volatility subsample of the S&P 500 Index (January 4, 2008 to June 30, 2010). We consider three different loss functions for the CPA test, namely the Square Error, Homogeneous Robust for b=-1 and QLIKE. The entries in parenthesis are the percentage of cases that the column method outperforms the row method in forecasting volatility. Entries in bold correspond to the cases that the null hypothesis of the CPA test is rejected. The sign + (-) indicates that column method significantly outperforms (is outperformed) by the row method in forecasting volatility.

Table 28: Comparison of forecast combination methods with individual forecasts using the CPA test based on the Square Error loss function for the full sample of the S&P 500 Index

A. Square Error loss function

Method	*Mean*	*Median*	*Geometric Mean*	*FC-RS*	*FFC SE*
AR(1) - RV	0.135	**0.091^{-}**	0.178	0.917	0.136
	(0.843)	**(0.068)**	(0.916)	(0.052)	(0.986)
AR(5) - RV	0.135	0.152	0.361	0.954	0.107
	(0.802)	(0.062)	(0.932)	(0.029)	(0.989)
AR(10) - RV	0.135	0.116	0.307	0.820	**0.042^{+}**
	(0.846)	(0.099)	(0.918)	(0.968)	**(0.987)**
AR(15) - RV	0.164	0.128	0.360	0.884	**0.046^{+}**
	(0.854)	(0.096)	(0.919)	(0.974)	**(0.988)**
HAR - RV	0.133	0.143	0.263	0.235	0.452
	(0.043)	(0.040)	(0.702)	(0.015)	(0.986)
LHAR - RV	0.621	0.537	0.554	0.523	0.332
	(0.990)	(0.987)	(0.990)	(0.991)	(0.991)
Normal GARCH	**0.007^{+}**	**0.002^{+}**	**0.001^{+}**	**0.092^{+}**	**0.003^{+}**
	(0.992)	**(0.991)**	**(0.985)**	**(0.970)**	**(0.984)**
Normal TARCH	**0.011^{+}**	0.109	**0.053^{+}**	**0.075^{+}**	0.141
	(0.958)	(0.972)	**(0.966)**	**(0.951)**	(0.967)
Normal EGARCH	**0.087^{-}**	**0.087^{-}**	**0.086^{-}**	0.207	0.571
	(0.100)	**(0.062)**	**(0.073)**	(0.069)	(0.147)
Normal APARCH	**0.086^{+}**	0.236	**0.090^{+}**	0.122	0.148
	(0.954)	(0.979)	**(0.977)**	(0.958)	(0.969)
t GARCH	**0.005^{+}**	**0.001^{+}**	**0.000^{+}**	**0.093^{+}**	**0.001^{+}**
	(0.990)	**(0.990)**	**(0.985)**	**(0.966)**	**(0.982)**
t TARCH	**0.013^{+}**	**0.034^{+}**	**0.019^{+}**	0.137	**0.087^{+}**
	(0.972)	**(0.982)**	**(0.979)**	(0.966)	**(0.969)**
t EGARCH	**0.072^{-}**	0.126	0.591	0.133	0.187
	(0.160)	(0.061)	(0.938)	(0.108)	(0.961)
t APARCH	**0.021^{+}**	**0.039^{+}**	**0.022^{+}**	0.175	**0.080^{+}**
	(0.974)	**(0.982)**	**(0.982)**	(0.972)	**(0.972)**
RiskMetrics	**0.004^{+}**	**0.002^{+}**	**0.001^{+}**	**0.063^{+}**	**0.003^{+}**
	(0.991)	**(0.991)**	**(0.985)**	**(0.965)**	**(0.984)**
Rolling 30 days	**0.006^{+}**	**0.004^{+}**	**0.002^{+}**	**0.057^{+}**	**0.004^{+}**
	(0.987)	**(0.978)**	**(0.982)**	**(0.971)**	**(0.986)**
Rolling 60 days	**0.010^{+}**	**0.012^{+}**	**0.003^{+}**	**0.038^{+}**	**0.004^{+}**
	(0.986)	**(0.976)**	**(0.981)**	**(0.974)**	**(0.991)**

This table shows the p-values of the null hypothesis of the CPA test of equal predictive ability of alternative volatility predictions given by forecast combinations with individual models using the full sample of the S&P 500 Index (January 2, 2004 to June 30, 2010). We consider three different loss functions for the CPA test, namely the Square Error, Homogeneous Robust for b=-1 and QLIKE. The entries in parenthesis are the percentage of cases that the column method outperforms the row method in forecasting volatility. Entries in bold correspond to the cases that the null hypothesis of the CPA test is rejected. The sign + (-) indicates that column method significantly outperforms (is outperformed) by the row method in forecasting volatility.

Table 29: Comparison of forecast combination methods with individual forecasts using the CPA test based on the QLIKE loss function for the full sample of the S&P 500 Index

B. Homogeneous Robust loss function for b=-1

Method	*Mean*	*Median*	*Geometric Mean*	*FC-RS*	*FFC Robust b=-1*
AR(1) - RV	**0.082^{+}**	**0.050^{+}**	0.146	0.917	**0.000^{+}**
	(0.673)	**(0.524)**	(0.821)	(0.515)	**(0.961)**
AR(5) - RV	**0.029^{-}**	**0.048^{-}**	**0.069**	0.159	**0.000^{+}**
	(0.185)	**(0.142)**	**(0.480)**	(0.075)	**(0.960)**
AR(10) - RV	**0.010^{-}**	**0.023^{-}**	**0.020**	0.257	**0.001^{+}**
	(0.222)	**(0.171)**	**(0.468)**	(0.051)	**(0.973)**
AR(15) - RV	**0.011^{-}**	**0.022^{-}**	**0.029**	0.267	**0.001^{+}**
	(0.222)	**(0.166)**	**(0.470)**	(0.050)	**(0.976)**
HAR - RV	**0.075^{-}**	**0.058^{-}**	0.151	**0.009^{-}**	**0.087^{+}**
	(0.080)	**(0.091)**	(0.119)	**(0.027)**	**(0.973)**
LHAR - RV	**0.010^{-}**	**0.004^{-}**	**0.040**	**0.004^{-}**	0.501
	(0.125)	**(0.125)**	**(0.256)**	**(0.090)**	(0.894)
Normal GARCH	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.006^{+}**	**0.000^{+}**
	(0.981)	**(0.975)**	**(0.960)**	**(0.905)**	**(0.958)**
Normal TARCH	**0.000^{+}**	**0.004^{+}**	**0.000^{+}**	**0.021^{+}**	**0.000^{+}**
	(0.878)	**(0.921)**	**(0.913)**	**(0.830)**	**(0.914)**
Normal EGARCH	**0.007^{-}**	**0.056^{-}**	**0.001^{+}**	0.115	**0.003^{+}**
	(0.440)	**(0.315)**	**(0.645)**	(0.312)	**(0.953)**
Normal APARCH	**0.003^{+}**	**0.033^{+}**	**0.000^{+}**	**0.029^{+}**	**0.000^{+}**
	(0.845)	**(0.925)**	**(0.930)**	**(0.830)**	**(0.930)**
t GARCH	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.005^{+}**	**0.000^{+}**
	(0.981)	**(0.969)**	**(0.954)**	**(0.907)**	**(0.951)**
t TARCH	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.025^{+}**	**0.000^{+}**
	(0.918)	**(0.950)**	**(0.944)**	**(0.897)**	**(0.920)**
t EGARCH	**0.009^{+}**	**0.030^{-}**	**0.016^{+}**	**0.059**	**0.001^{+}**
	(0.618)	**(0.379)**	**(0.868)**	**(0.433)**	**(0.943**
t APARCH	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.044^{+}**	**0.000^{+}**
	(0.903)	**(0.948)**	**(0.944)**	**(0.911)**	**(0.924)**
RiskMetrics	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.006^{+}**	**0.000^{+}**
	(0.979)	**(0.966)**	**(0.950)**	**(0.902)**	**(0.950)**
Rolling 30 days	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.011^{+}**	**0.000^{+}**
	(0.968)	**(0.952)**	**(0.959)**	**(0.913)**	**(0.962)**
Rolling 60 days	**0.000^{+}**	**0.000^{+}**	**0.000^{+}**	**0.004^{+}**	**0.000^{+}**
	(0.972)	**(0.963)**	**(0.966)**	**(0.919)**	**(0.974)**

This table shows the p-values of the null hypothesis of the CPA test of equal predictive ability of alternative volatility predictions given by forecast combinations with individual models using the full sample of the S&P 500 Index (January 2, 2004 to June 30, 2010). We consider three different loss functions for the CPA test, namely the Square Error, Homogeneous Robust for b=-1 and QLIKE. The entries in parenthesis are the percentage of cases that the column method outperforms the row method in forecasting volatility. Entries in bold correspond to the cases that the null hypothesis of the CPA test is rejected. The sign + (-) indicates that column method significantly outperforms (is outperformed) by the row method in forecasting volatility.

Table 30: Comparison of forecast combination methods with individual forecasts using the CPA test based on the QLIKE loss function for the full sample of the S&P 500 Index

C. QLIKE loss function

Method	*Mean*	*Median*	*Geometric Mean*	*FC-RS*	*FFC QLIKE*
AR(1) - RV	**0.004⁺**	**0.004⁺**	**0.001⁺**	**0.006⁺**	**0.000⁺**
	(0.582)	**(0.561)**	**(0.725)**	**(0.950)**	**(0.997)**
AR(5) - RV	**0.000⁻**	**0.000⁻**	**0.001⁻**	**0.098⁻**	**0.000⁺**
	(0.234)	**(0.207)**	**(0.305)**	**(0.212)**	**(0.993)**
AR(10) - RV	**0.000⁻**	**0.000⁻**	**0.002⁻**	**0.022⁻**	**0.000⁺**
	(0.233)	**(0.207)**	**(0.283)**	**(0.220)**	**(0.958)**
AR(15) - RV	**0.000⁻**	**0.000⁻**	**0.002⁻**	**0.021⁻**	**0.000⁺**
	(0.234)	**(0.208)**	**(0.287)**	**(0.233)**	**(0.949)**
HAR - RV	**0.000⁻**	**0.000⁻**	**0.000⁻**	**0.000⁻**	**0.000⁺**
	(0.152)	**(0.157)**	**(0.166)**	**(0.045)**	**(0.900)**
LHAR - RV	**0.000⁻**	**0.000⁻**	**0.000⁻**	**0.000⁻**	**0.000⁺**
	(0.102)	**(0.114)**	**(0.084)**	**(0.011)**	**(0.993)**
Normal GARCH	**0.000⁺**	**0.000⁺**	**0.000⁺**	**0.000⁺**	**0.000⁺**
	(0.957)	**(0.957)**	**(0.953)**	**(0.849)**	**(0.946)**
Normal TARCH	**0.000⁺**	**0.000⁺**	**0.000⁺**	**0.000⁺**	**0.000⁺**
	(0.891)	**(0.927)**	**(0.916)**	**(0.782)**	**(0.919)**
Normal EGARCH	**0.000⁺**	**0.000⁺**	**0.000⁺**	**0.000⁺**	**0.000⁺**
	(0.647)	**(0.620)**	**(0.770)**	**(0.706)**	**(0.910)**
Normal APARCH	**0.000⁺**	**0.000⁺**	**0.000⁺**	**0.000⁺**	**0.000⁺**
	(0.825)	**(0.890)**	**(0.889)**	**(0.771)**	**(0.924)**
t GARCH	**0.000⁺**	**0.000⁺**	**0.000⁺**	**0.000⁺**	**0.000⁺**
	(0.955)	**(0.954)**	**(0.946)**	**(0.851)**	**(0.945)**
t TARCH	**0.000⁺**	**0.000⁺**	**0.000⁺**	**0.000⁺**	**0.000⁺**
	(0.903)	**(0.933)**	**(0.921)**	**(0.813)**	**(0.920)**
t EGARCH	**0.000⁺**	**0.000⁺**	**0.000⁺**	**0.000⁺**	**0.000⁺**
	(0.693)	**(0.681)**	**(0.794)**	**(0.729)**	**(0.912)**
t APARCH	**0.000⁺**	**0.000⁺**	**0.000⁺**	**0.000⁺**	**0.000⁺**
	(0.868)	**(0.910)**	**(0.905)**	**(0.804)**	**(0.923)**
RiskMetrics	**0.000⁺**	**0.000⁺**	**0.000⁺**	**0.000⁺**	**0.000⁺**
	(0.870)	**(0.834)**	**(0.885)**	**(0.812)**	**(0.914)**
Rolling 30 days	**0.000⁺**	**0.000⁺**	**0.000⁺**	**0.000⁺**	**0.000⁺**
	(0.823)	**(0.790)**	**(0.838)**	**(0.813)**	**(0.910)**
Rolling 60 days	**0.000⁺**	**0.000⁺**	**0.000⁺**	**0.000⁺**	**0.000⁺**
	(0.897)	**(0.884)**	**(0.910)**	**(0.849)**	**(0.938)**

This table shows the p-values of the null hypothesis of the CPA test of equal predictive ability of alternative volatility predictions given by forecast combinations with individual models using the full sample of the S&P 500 Index (January 2, 2004 to June 30, 2010). We consider three different loss functions for the CPA test, namely the Square Error, Homogeneous Robust for b=-1 and QLIKE. The entries in parenthesis are the percentage of cases that the column method outperforms the row method in forecasting volatility. Entries in bold correspond to the cases that the null hypothesis of the CPA test is rejected. The sign + (-) indicates that column method significantly outperforms (is outperformed) by the row method in forecasting volatility.

Table 31: Comparison of volatility predictions given by forecast combination methods using the CPA test for the low volatility subsample of the S&P 500 Index

A. Square Error loss function

Method	*Mean*	*Median*	*Geometric Mean*	*FC-RS*	*FFC SE*
Mean	.	**0.026**$^{-}$	**0.004**$^{+}$	**0.014**$^{+}$	**0.012**$^{+}$
	.	**(0.182)**	**(0.918)**	**(0.711)**	**(0.843)**
Median	**0.026**$^{+}$	.	**0.004**$^{+}$	**0.020**$^{+}$	**0.026**$^{+}$
	(0.818)	.	**(0.893)**	**(0.768)**	**(0.872)**
Geometric Mean	**0.004**$^{-}$	**0.004**$^{-}$	.	**0.023**$^{+}$	**0.020**$^{+}$
	(0.082)	**(0.107)**	.	**(0.587)**	**(0.802)**
FC-RS	**0.014**$^{-}$	**0.020**$^{-}$	**0.023**$^{-}$	.	**0.014**$^{+}$
	(0.289)	**(0.232)**	**(0.413)**	.	**(0.934)**
FFC SE	**0.012**$^{-}$	**0.026**$^{-}$	**0.020**$^{-}$	**0.014**$^{-}$	.
	(0.157)	**(0.128)**	**(0.198)**	**(0.066)**	.

B. Homogeneous Robust loss function for b=-1

Method	*Mean*	*Median*	*Geometric Mean*	*FC-RS*	*FFC Robust b=-1*
Mean	.	**0.000**$^{-}$	**0.000**$^{+}$	**0.000**$^{+}$	**0.000**$^{+}$
	.	**(0.321)**	**(0.900)**	**(0.741)**	**(0.844)**
Median	**0.000**$^{+}$	.	**0.000**$^{+}$	**0.001**$^{+}$	**0.000**$^{+}$
	(0.679)	.	**(0.823)**	**(0.758)**	**(0.846)**
Geometric Mean	**0.000**$^{-}$	**0.000**$^{-}$	.	**0.001**$^{+}$	**0.000**$^{+}$
	(0.100)	**(0.177)**	.	**(0.694)**	**(0.840)**
FC-RS	**0.000**$^{-}$	**0.001**$^{-}$	**0.001**$^{-}$	.	**0.001**$^{+}$
	(0.259)	**(0.242)**	**(0.306)**	.	**(0.910)**
FFC Robust b=-1	**0.000**$^{-}$	**0.000**$^{-}$	**0.000**$^{-}$	**0.001**$^{-}$	.
	(0.156)	**(0.154)**	**(0.160)**	**(0.090)**	.

C. QLIKE loss function

Method	*Mean*	*Median*	*Geometric Mean*	*FC-RS*	*FFC QLIKE*
Mean	.	**0.000**$^{-}$	**0.000**$^{+}$	**0.000**$^{+}$	**0.000**$^{+}$
	.	**(0.356)**	**(0.896)**	**(0.764)**	**(0.925)**
Median	**0.000**$^{+}$	.	**0.000**$^{+}$	**0.000**$^{+}$	**0.000**$^{+}$
	(0.644)	.	**(0.807)**	**(0.781)**	**(0.913)**
Geometric Mean	**0.000**$^{-}$	**0.000**$^{-}$	.	**0.000**$^{+}$	**0.000**$^{+}$
	(0.104)	**(0.193)**	.	**(0.738)**	**(0.940)**
FC-RS	**0.000**$^{-}$	**0.000**$^{-}$	**0.000**$^{-}$	.	**0.000**$^{+}$
	(0.236)	**(0.219)**	**(0.262)**	.	**(0.927)**
FFC QLIKE	**0.000**$^{-}$	**0.000**$^{-}$	**0.000**$^{-}$	**0.000**$^{-}$	.
	(0.075)	**(0.087)**	**(0.060)**	**(0.073)**	.

This table shows the p-values of the null hypothesis of the CPA test of equal predictive ability of alternative volatility predictions given by forecast combinations using the low volatility subsample of the S&P 500 Index (January 2, 2004 to January 3, 2008). We consider three different loss functions for the CPA test, namely the Square Error, Homogeneous Robust for b=-1 and QLIKE. The entries in parenthesis are the percentage of cases that the column method outperforms the row method in forecasting volatility. Entries in bold correspond to the cases that the null hypothesis of the CPA test is rejected. The sign + (-) indicates that column method significantly outperforms (is outperformed) by the row method in forecasting volatility.

Table 32: Comparison of volatility predictions given by forecast combination methods using the CPA test for the high volatility subsample of the S&P 500 Index

A. Square Error loss function

Method	*Mean*	*Median*	*Geometric Mean*	*FC-RS*	*FFC SE*
Mean	.	0.104	**0.061**$^{+}$	0.111	**0.087**$^{+}$
	.	(0.027)	**(0.939)**	(0.227)	**(0.949)**
Median	0.104	.	**0.007**$^{+}$	**0.061**$^{+}$	0.164
	(0.973)	.	**(0.954)**	**(0.885)**	(0.955)
Geometric Mean	**0.061**$^{-}$	**0.007**$^{-}$	.	0.157	0.186
	(0.061)	**(0.046)**	.	(0.120)	(0.950)
FC-RS	0.111	**0.061**$^{-}$	0.157	.	**0.026**$^{+}$
	(0.773)	**(0.115)**	(0.880)	.	**(0.986)**
FFC SE	**0.087**$^{-}$	0.164	0.186	**0.026**$^{-}$	.
	(0.051)	(0.045)	(0.050)	**(0.014)**	.

B. Homogeneous Robust loss function for b=-1

Method	*Mean*	*Median*	*Geometric Mean*	*FC-RS*	*FFC Robust b=-1*
Mean	.	**0.081**$^{-}$	**0.001**$^{+}$	**0.084**$^{-}$	**0.004**$^{+}$
	.	**(0.105)**	**(0.968)**	**(0.403)**	**(0.915)**
Median	**0.081**$^{+}$	.	**0.000**$^{+}$	**0.037**$^{+}$	**0.000**$^{+}$
	(0.895)	.	**(0.887)**	**(0.545)**	**(0.933)**
Geometric Mean	**0.001**$^{-}$	**0.000**$^{-}$	.	0.137	**0.000**$^{+}$
	(0.032)	**(0.113)**	.	(0.204)	**(0.955)**
FC-RS	**0.084**$^{+}$	**0.037**$^{-}$	0.137	.	**0.000**$^{+}$
	(0.597)	**(0.455)**	(0.796)	.	**(0.979)**
FFC Robust b=-1	**0.004**$^{-}$	**0.000**$^{-}$	**0.000**$^{-}$	**0.000**$^{-}$	.
	(0.085)	**(0.067)**	**(0.045)**	**(0.021)**	.

C. QLIKE loss function

Method	*Mean*	*Median*	*Geometric Mean*	*FC-RS*	*FFC QLIKE*
Mean	.	**0.000**$^{-}$	**0.000**$^{+}$	**0.000**	**0.000**$^{+}$
	.	**(0.284)**	**(0.903)**	**(0.500)**	**(0.936)**
Median	**0.000**$^{+}$	.	**0.000**$^{+}$	**0.000**$^{+}$	**0.000**$^{+}$
	(0.716)	.	**(0.853)**	**(0.526)**	**(0.925)**
Geometric Mean	**0.000**$^{-}$	**0.000**$^{-}$	.	**0.000**$^{-}$	**0.000**$^{+}$
	(0.097)	**(0.147)**	.	**(0.345)**	**(0.965)**
FC-RS	**0.000**	**0.000**$^{-}$	**0.000**$^{+}$	.	**0.000**$^{+}$
	(0.500)	**(0.474)**	**(0.655)**	.	**(0.995)**
FFC QLIKE	**0.000**$^{-}$	**0.000**$^{-}$	**0.000**$^{-}$	**0.000**$^{-}$	.
	(0.064)	**(0.075)**	**(0.035)**	**(0.005)**	.

This table shows the p-values of the null hypothesis of the CPA test of equal predictive ability of alternative volatility predictions given by forecast combinations using the high volatility subsample of the S&P 500 Index (January 4, 2008 to June 30, 2010). We consider three different loss functions for the CPA test, namely the Square Error, Homogeneous Robust for b=-1 and QLIKE. The entries in parenthesis are the percentage of cases that the column method outperforms the row method in forecasting volatility. Entries in bold correspond to the cases that the null hypothesis of the CPA test is rejected. The sign + (-) indicates that column method significantly outperforms (is outperformed) by the row method in forecasting volatility.

Table 33: Comparison of volatility predictions given by forecast combination methods using the CPA test for the full sample of the S&P 500 Index

A. Square Error loss function

Method	*Mean*	*Median*	*Geometric Mean*	*FC-RS*	*FFC SE*
Mean	.	0.132	**0.068**$^{+}$	0.129	0.106
	.	(0.018)	**(0.971)**	(0.145)	(0.964)
Median	0.132	.	**0.011**$^{+}$	**0.081**$^{+}$	0.198
	(0.982)	.	**(0.971)**	**(0.922)**	(0.971)
Geometric Mean	**0.068**$^{-}$	**0.011**$^{-}$	.	0.175	0.201
	(0.029)	**(0.029)**	.	(0.080)	(0.964)
FC-RS	0.129	**0.081**$^{-}$	0.175	.	**0.042**$^{+}$
	(0.855)	**(0.078)**	(0.920)	.	**(0.990)**
FFC SE	0.106	0.198	0.201	**0.042**$^{-}$	.
	(0.036)	(0.029)	(0.036)	**(0.010)**	.

B. Homogeneous Robust loss function for b=-1

Method	*Mean*	*Median*	*Geometric Mean*	*FC-RS*	*FFC Robust b=-1*
Mean	.	**0.078**$^{-}$	**0.000**$^{+}$	**0.091**$^{-}$	**0.002**$^{+}$
	.	**(0.094)**	**(0.968)**	**(0.383)**	**(0.912)**
Median	**0.078**$^{+}$	.	**0.000**$^{+}$	**0.041**$^{+}$	**0.000**$^{+}$
	(0.906)	.	**(0.890)**	**(0.588)**	**(0.931)**
Geometric Mean	**0.000**$^{-}$	**0.000**$^{-}$	.	0.155	**0.000**$^{+}$
	(0.032)	**(0.110)**	.	(0.204)	**(0.949)**
FC-RS	**0.091**$^{+}$	**0.041**$^{-}$	0.155	.	**0.000**$^{+}$
	(0.617)	**(0.412)**	(0.796)	.	**(0.980)**
FFC Robust b=-1	**0.002**$^{-}$	**0.000**$^{-}$	**0.000**$^{-}$	**0.000**$^{-}$	.
	(0.088)	**(0.069)**	**(0.051)**	**(0.020)**	.

C. QLIKE loss function

Method	*Mean*	*Median*	*Geometric Mean*	*FC-RS*	*FFC QLIKE*
Mean	.	**0.000**$^{-}$	**0.000**$^{+}$	**0.000**$^{+}$	**0.000**$^{+}$
	.	**(0.321)**	**(0.895)**	**(0.655)**	**(0.921)**
Median	**0.000**$^{+}$	.	**0.000**$^{+}$	**0.000**$^{+}$	**0.000**$^{+}$
	(0.679)	.	**(0.819)**	**(0.678)**	**(0.910)**
Geometric Mean	**0.000**$^{-}$	**0.000**$^{-}$	.	**0.000**$^{+}$	**0.000**$^{+}$
	(0.105)	**(0.181)**	.	**(0.557)**	**(0.945)**
FC-RS	**0.000**$^{-}$	**0.000**$^{-}$	**0.000**$^{-}$	.	**0.000**$^{+}$
	(0.345)	**(0.322)**	**(0.443)**	.	**(0.994)**
FFC QLIKE	**0.000**$^{-}$	**0.000**$^{-}$	**0.000**$^{-}$	**0.000**$^{-}$	.
	(0.079)	**(0.090)**	**(0.055)**	**(0.006)**	.

This table shows the p-values of the null hypothesis of the CPA test of equal predictive ability of alternative volatility predictions given by forecast combinations using the full sample of the S&P 500 Index (January 2, 2004 to June 30, 2010). We consider three different loss functions for the CPA test, namely the Square Error, Homogeneous Robust for b=-1 and QLIKE. The entries in parenthesis are the percentage of cases that the column method outperforms the row method in forecasting volatility. Entries in bold correspond to the cases that the null hypothesis of the CPA test is rejected. The sign + (-) indicates that column method significantly outperforms (is outperformed) by the row method in forecasting volatility.