



Kent Academic Repository

Meligkotsidou, Loukia, Panopoulou, Ekaterini, Vrontos, Ioannis D. and Vrontos, Spyridon D. (2019) *Quantile Forecast Combinations in Realised Volatility Prediction*. *Journal of the Operational Research Society*, 70 (10). pp. 1720-1733. ISSN 0160-5682.

Downloaded from

<https://kar.kent.ac.uk/67216/> The University of Kent's Academic Repository KAR

The version of record is available from

<https://doi.org/10.1080/01605682.2018.1489354>

This document version

Author's Accepted Manuscript

DOI for this version

Licence for this version

UNSPECIFIED

Additional information

Versions of research works

Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in *Title of Journal*, Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).

Quantile Forecast Combinations in Realised Volatility Prediction

Loukia Meligkotsidou^a, Ekaterini Panopoulou^{b*}, Ioannis D.Vrontos^c, Spyridon D. Vrontos^d

^aDepartment of Mathematics, National and Kapodistrian University of Athens, Greece

^bKent Business School, University of Kent, Canterbury, United Kingdom

^cDepartment of Statistics, Athens University of Economics and Business, Greece

^dDepartment of Mathematical Sciences, University of Essex, United Kingdom

Abstract

This paper tests whether it is possible to improve point, quantile and density forecasts of realised volatility by conditioning on a set of predictive variables. We employ quantile autoregressive models augmented with macroeconomic and financial variables. Complete subset combinations of both linear and quantile forecasts enable us to efficiently summarise the information content in the candidate predictors. Our findings suggest that no single variable is able to provide more information for the evolution of the volatility distribution beyond that contained in its own past. The best performing variable is the return on the stock market followed by the inflation rate. Our complete subset approach achieves superior point, quantile and density predictive performance relative to the univariate models and the autoregressive benchmark.

JEL classification: G12; G22; C22; C53

Keywords: Forecasting; Realised volatility; Forecast combination; Predictive quantile regression; Subset quantile regressions.

*Corresponding author: Ekaterini Panopoulou, Kent Business School, University of Kent, Canterbury CT2 7FS, United Kingdom, Tel.: 0044 1227 824469, Email: A.Panopoulou@kent.ac.uk.

1 Introduction

Forecasting future volatility is essential for asset allocation, pricing, portfolio and risk management. Both the future level of volatility and its distribution play a central role in derivatives pricing, developing trading strategies, pricing and trading volatility derivatives, hedging portfolios and quantifying the uncertainty around the point forecast of volatility. A vast literature has investigated time variation in the level of volatility and the linkages between volatility and macroeconomic and financial variables. Early contributions include, among others, Schwert (1989) and Glosten et al. (1993) that find evidence of a significant link between mainly interest rate variables and future market volatility. More recently, Paye (2012) tests whether conditioning on macroeconomic variables can improve realised volatility forecasts and finds a link between several variables and stock market volatility. Christiansen et al. (2012) employ a comprehensive set of macro-finance variables to predict the realised volatility of four different asset classes: equities, commodities, foreign exchange rates, and bonds. Using Bayesian estimation techniques the authors identify the variables that are best in predicting realised volatility. By selecting the most important predictor variables, they find that their forecast models beat autoregressive benchmarks although this performance varies across asset classes and over time. Conrad and Loch (2015) confirm the counter-cyclical behaviour of stock market volatility for a broad set of macroeconomic variables. Their findings suggest that long-term volatility is mainly driven by information related to the current state of the economy, as well as to expectations regarding future macroeconomic conditions. Finally, Mitnik et al. (2015) show that boosting techniques substantially improve volatility forecasts for short- and long-run horizons and risk drivers affect future volatility in a nonlinear fashion.

While the aforementioned studies focus on the link between several macroeconomic and financial variables and the future level of volatility, little or no attention has been paid to the impact of these variables on the future volatility distribution. This paper complements these recent studies by focusing on the ability of macroeconomic and financial variables to provide superior quantile and density volatility forecasts. Our framework is a purely out-of-sample one based on quantile forecast combinations. As we show, aggregating these quantile forecasts can also provide us with accurate point volatility forecasts. In general, the forecasting framework we adopt is rooted in quantile predictive regressions, which have attracted a vast amount of attention since the seminal paper of Koenker and Bassett (1978).¹

We assume that future volatility quantiles depend on their own lags and macroeconomic and financial variables. By considering several conditional quantiles of future volatility, we allow for heterogeneous degrees of persistence of realised volatility and asymmetric dynamic responses of macroeconomic and financial variables at different parts of the distribution. For example, variables that are useful in predicting volatility surges (right tail of the volatility distribution) might be useless in calm periods (low volatility) and vice

¹Quantile regressions have been employed very recently in a forecasting context by Manzan and Zerom (2013), Meligkotsidou et al. (2014a, 2014b), Manzan (2015), and Pedersen (2015).

versa. Also the degree of volatility persistence can vary with the quantile to be predicted probably pointing to higher persistence in turbulent periods. We focus on US stock market volatility and update the dataset of Christiansen et al. (2012). Given the plethora of candidate predictors, a question that naturally arises is how to select the variables to include in the forecasting model without losing any useful information. The approach we follow belongs to the combination of forecasts literature. To allow for varying degrees of persistence in the distribution of volatility, we extend the Complete Subset Regression (CSR) framework of Elliott et al. (2013) to a Complete Subset (Quantile) AutoRegressive one, namely the CS(Q)AR approach. In the complete subset regression framework, forecasts are formed by combining forecasts from all possible linear or quantile regression models that keep the number of predictors fixed. This forecasting approach exploits the benefits emerging from three strands of the literature on out-of-sample forecasting. First, the quantile regression setting succeeds in producing accurate distribution, quantile and point forecasts (via efficiently aggregating quantile forecasts). Second, model uncertainty and parameter instability is reduced by employing quantile forecast combinations.² Finally, employing the complete subset approach induces shrinkage to the respective estimates and further helps reduce the effect of parameter estimation error.

To anticipate our key results, we find that no single variable is able to provide more information for the evolution of the volatility distribution beyond that contained in its own past. The best performing variable is the return on the stock market followed by the inflation rate. These variables along with the T-bill are the ones that improve point volatility forecasts with an out-of-sample R^2 greater than 2.33%. Our CSQAR approach achieves superior quantile, density and point predictive performance relative to univariate models and the CSAR approach. With respect to point forecasts, our quantile aggregation method can lead to substantial improvements judging from the out-of-sample R^2 values of about 9% relative to the autoregressive benchmark.

The outline of the paper is as follows. Section 2 describes our CSAR and CSQAR forecasting approaches and shows how point forecasts are generated via quantile forecast aggregation. Section 3 outlines the tests of forecasting accuracy used to evaluate the quantile, distribution and point forecasting approaches. Section 4 presents our empirical findings and finally, Section 5 concludes.

2 Forecasting approaches

Our aim is to evaluate the information content of financial variables for different parts of the future distribution of realised volatility. In this section, we describe our methodology for efficiently aggregating this information that leads to superior point, quantile and density forecasts.

²Timmermann (2006) provides a detailed review on forecast combination methodologies. See also Bates and Granger (1969) and Rapach et al. (2010).

2.1 Complete subset autoregressive approach (CSAR)

Suppose that we are interested in forecasting realised volatility, denoted by RV_t , using a set of K predictive variables. Since volatility is fairly persistent, any predictive model should include autoregressive terms. We investigate whether there is additional predictive content of the macroeconomic and financial variables that goes beyond the information contained in lagged volatility. To this end, we employ a set of predictive autoregressive models of time-varying order q_t , each including a subset of k predictors, $k \leq K$.

First we consider all possible (K) predictive autoregressive models ($AR(q_t)$) augmented with a single predictor, i.e. $k = 1$, of the form

$$RV_{t+1} = \alpha_i + \sum_{j=1}^{q_t} c_{ij} RV_{t-j+1} + \beta_i x_{it} + \varepsilon_{t+1}, \quad i = 1, \dots, K, \quad (1)$$

where RV_{t+1} is the observed realised volatility at time $t + 1$, RV_{t-j+1} is the observed realised volatility at time $t, t - 1, \dots, t - q_t + 1$ and x_{it} is the i th observed predictor at time $t, i = 1, \dots, K$. The error terms ε_{t+1} are assumed to be independent with mean zero and variance σ_i^2 . The predictive autoregressive models considered are estimated using the Maximum Likelihood (ML) approach under the assumption of a Gaussian error distribution with mean zero and constant conditional variance, i.e. $\varepsilon_{t+1} | \mathfrak{F}_t \sim N(0, \sigma_i^2)$, where \mathfrak{F}_t denotes the information set at time t . To recursively select the optimal lag length (q_t) over the out-of-sample period, we employ the Bayesian Information Criterion (BIC) and model (1) without any candidate predictors ($\beta_i = 0$). Model (1) nests the benchmark AR(1) and AR(2) specifications employed in Christiansen et al. (2012) and Paye (2012), respectively. Given the parametric distribution for the error terms, point, quantile and distribution forecasts of RV_{t+1} are straightforward to construct.

When the number of candidate predictors is large, finding a parsimonious way to include all relevant information in the forecasting model is challenging. Two broad approaches have been put forward in the literature; namely combination of forecasts and combination of information. Combination of forecasts combines forecasts generated from simple models each incorporating a part of the whole information set, while combination of information brings the entire information set into one elaborate model to generate an ultimate forecast (Huang and Lee, 2010; Panopoulou and Vrontos, 2015). The approach we follow belongs to the combination of forecasts strand and was proposed by Elliott et al. (2013) and extended to accommodate a large-dimensional set of predictors by Elliott et al. (2015).

Elliott et al. (2013) employ complete subset regressions (CSR) as a means of combining forecasts. Specifically, for a given set of potential predictors, the authors propose combining forecasts from all possible linear regressions that keep the number of predictors fixed. The authors show that while subset regression combinations bear similarities to a complex version of shrinkage, they do not reduce to shrinking Ordinary Least Squares (OLS) estimates. Rather the coefficient that controls shrinkage depends on all estimates, the dimension of the subset and the number of included predictors. Only in the case of orthonormal

regressors does subset regression reduce to ridge regression. Moreover, the amount of shrinkage imposed on each coefficient differs with the coefficient at hand. More importantly, the authors show that in the case of strongly correlated predictors, subset regression can remedy the omitted variable bias and improve forecasts.

We adapt the CSR framework to an autoregressive one and employ complete subset autoregressive (CSAR) models. For the case of $k = 1$, all K available models given by Equation (1) are employed. For $k = 2$, Equation (1) is augmented with a second variable and all bivariate models are employed. We proceed in this way until $k = K$ where only one K -variate model is employed, referred to in the literature as the ‘kitchen sink’ model (Goyal and Welch, 2008). The set of models for a fixed value of k is referred to as a complete subset. Given that each subset produces $n_{k,K} = K!/((K - k)!k!)$ forecasts corresponding to the alternative models employed, these need to be aggregated (combined) into a single forecast per subset. In order to do so, we employ two combination schemes; an equally-weighted (EW) and a stochastic Bayesian combination scheme³.

Since the forecasting performance of the CSAR approach clearly varies with each subset employed, selecting the optimal value of k (k^*) is of utmost practical importance. We employ the likelihood-based (Bayesian) methodology of Meligkotsidou et al. (2014b) and select the most probable value of k at each point of time in the out of sample period based on the posterior probabilities of all values of k ($k \in \{1, 2, \dots, K\}$). Details of the algorithm are given in the Appendix.

2.2 Complete subset quantile autoregressive approach (CSQAR)

Similar to the CSAR approach, the baseline forecasting model we consider is the Quantile Autoregressive (QAR) specification developed by Koenker and Xiao (2006) augmented with one of K candidate predictor variables:

$$RV_{t+1} = \alpha_i^{(\tau)} + \sum_{j=1}^{q_t^{(\tau)}} c_{ij}^{(\tau)} RV_{t-j+1} + \beta_i^{(\tau)} x_{it} + \varepsilon_{t+1}, \quad i = 1, \dots, K, \quad (2)$$

where $\tau \in (0, 1)$, $\alpha_i^{(\tau)}$, $c_{ij}^{(\tau)}$, $\beta_i^{(\tau)}$ are quantile-varying parameters and $q_t^{(\tau)}$ is the lag order used to model the τ th quantile of realised volatility at time t , given the variable x_{it} . The errors ε_{t+1} are assumed independent from an error distribution $g_\tau(\varepsilon)$ with the τ th quantile equal to 0, i.e. $\int_{-\infty}^0 g_\tau(\varepsilon) d\varepsilon = \tau$. Model (2) suggests that the τ th quantile of RV_{t+1} given the past of realised volatility (RV_{t-j+1} , $j = 1, \dots, q_t^{(\tau)}$) and the candidate predictor variable, x_{it} , is $Q_\tau(RV_{t+1}|\mathfrak{S}_t) = \alpha_i^{(\tau)} + \sum_{j=1}^{q_t^{(\tau)}} c_{ij}^{(\tau)} RV_{t-j+1} + \beta_i^{(\tau)} x_{it}$, where \mathfrak{S}_t denotes the information set at time t with all the parameters depending on τ .

The augmented QAR model is the direct analogue of the AR model (Equation (1)) employed for the conditional mean. Since the response of past volatility realisations ($c_{ij}^{(\tau)}$) varies across quantiles, the model

³We thank an anonymous referee for suggesting this combination scheme. Details are given in the Appendix. For alternative quantile combination schemes see Shan and Yang (2009) and Tsiotas (2015).

allows for heterogeneous degrees of persistence of realised volatility at different parts of the distribution and asymmetric dynamic responses. Moreover, the inclusion of the conditioning variables (x_{it}) may not only shift the location of the distribution of RV_{t+1} , but also may alter the scale and shape of the conditional distribution, capturing systematic influences of conditioning variables on the future realised volatility distribution.

Estimators of the parameters of the quantile regression models in (2), $\hat{\alpha}_i^{(\tau)}$, $\hat{c}_{ij}^{(\tau)}$, $\hat{\beta}_i^{(\tau)}$, can be obtained by minimizing the sum $\sum_{t=0}^{T-1} \rho_\tau \left(RV_{t+1} - \alpha_i^{(\tau)} - \sum_{j=1}^{q_t^{(\tau)}} c_{ij}^{(\tau)} RV_{t-j+1} - \beta_i^{(\tau)} x_{it} \right)$, where $\rho_\tau(u)$ is the asymmetric linear loss function, usually referred to as the check function,

$$\rho_\tau(u) = u(\tau - I(u < 0)) = \frac{1}{2} [|u| + (2\tau - 1)u]. \quad (3)$$

In the symmetric case of the absolute loss function ($\tau = 1/2$) we obtain estimators of the median predictive regression models. For likelihood based inference, one may assume the asymmetric Laplace error distribution (for details, see Yu and Moyeed, 2001, and Yu and Zhang, 2005) for model (2).⁴ The advantage of likelihood-based inference is that it enables us to compare different quantile regression models, corresponding to different subsets of predictors, using criteria based on the likelihood function, for example BIC. This further enables us to establish an approach of selecting the best (in a likelihood based sense) complete subset on the basis of which forecasts are formed. Similar to the CSAR approach, we employ the BIC criterion to select the optimal lag length $q_t^{(\tau)}$ which is quantile- and time-varying.

Next, we focus on the complete subset framework, which we adapt for the QAR case and develop the complete subset quantile autoregressive approach (CSQAR). Specifically, similar to the CSAR approach, for each k and for each quantile of interest (τ), $n_{k,K}$ QAR models are run in order to predict the τ th quantile of the distribution of the next period's realised volatility (RV_{t+1}). Then we aggregate the information contained in each of the subsets by combining the $n_{k,K}$ τ th quantile forecasts. As with the CSAR approach, we employ both the EW and the Bayesian combining schemes. Finally, for each quantile, τ , we select the most probable value of the subset k and produce a quantile forecast at time $t + 1$, $\widehat{RV}_{t+1}(\tau)$, based on the selected complete subset obtained by our the recursive algorithm outlined in the Appendix. This algorithm selects the optimal subset k_τ^* in a quantile-varying manner, based on the respective posterior probabilities of all values of k ($k \in \{1, 2, \dots, K\}$) for a set of quantiles.

2.3 Point forecasts via quantile forecast aggregation

While the QAR methodology described so far provides a direct way for forecasting the quantiles of interest along with an approximate forecast distribution, forecasting the mean of future realised volatility is not

⁴A potential problem that may arise in small samples is the so-called quantile crossing, that is the estimated quantiles are not guaranteed to be monotonic in τ . If this occurs, the recently developed approach of Chernozhukov et al. (2010) can be employed to establish monotonicity of the estimated quantiles.

trivial. For a given model specification or a given complete subset that has been used for producing quantile forecasts, point forecasts can be constructed as weighted averages of a set of quantile forecasts (Meligkotsidou et al. 2014a). The weights represent probabilities attached to different quantile forecasts, suggesting how likely to predict the return at the next period each regression quantile is. More in detail, we employ standard estimators with fixed, pre-specified weights of the form

$$\widehat{RV}_{t+1} = \sum_{\tau \in S} p_{\tau} \widehat{RV}_{t+1}^{(\tau)}, \quad \sum_{\tau \in S} p_{\tau} = 1,$$

where S denotes the set of quantiles that are aggregated, $\widehat{RV}_{t+1}^{(\tau)}$ denotes the quantile forecasts associated with the τ th quantile and \widehat{RV}_{t+1} is the produced point forecast.

We consider Tukey's (1977) trimean and the Gastwirth (1966) three-quantile estimator given, respectively, by the following formulae

$$\text{FW1:} \quad \widehat{RV}_{t+1} = 0.25\widehat{RV}_{t+1}^{(0.25)} + 0.50\widehat{RV}_{t+1}^{(0.50)} + 0.25\widehat{RV}_{t+1}^{(0.75)}, \quad (4)$$

$$\text{FW2:} \quad \widehat{RV}_{t+1} = 0.30\widehat{RV}_{t+1}^{(1/3)} + 0.40\widehat{RV}_{t+1}^{(0.50)} + 0.30\widehat{RV}_{t+1}^{(2/3)}. \quad (5)$$

In order to attach more weight on extreme positive and negative events, we also use the five-quantile estimator, suggested by Judge et al. (1988)

$$\text{FW3:} \quad \widehat{RV}_{t+1} = 0.05\widehat{RV}_{t+1}^{(0.10)} + 0.25\widehat{RV}_{t+1}^{(0.25)} + 0.40\widehat{RV}_{t+1}^{(0.50)} + 0.25\widehat{RV}_{t+1}^{(0.75)} + 0.05\widehat{RV}_{t+1}^{(0.90)}. \quad (6)$$

3 Measuring Forecast Accuracy

In this section, we describe the criteria we employ to evaluate realised volatility forecasts. First, we focus on point forecasts constructed either directly (CSAR approach) or indirectly via aggregating quantile forecasts (CSQAR approach); then we focus on quantile and density forecasts. We divide the total sample of T observations into an in-sample portion of the first T_0 observations and an out-of-sample portion of $P = T - T_0$ observations used for forecasting. Out-of-sample forecasts are generated by continuously updating the estimation window, i.e. following a recursive (expanding) scheme, by adding one observation to the estimation sample at each step. As such, the coefficients in any predictive model employed are re-estimated after each step of the recursion. Proceeding in this way through the end of the out-of-sample period, we generate a series of P out-of-sample forecasts for the realised volatility $\left\{ \widehat{RV}_{i,t+1} \right\}_{t=T_0}^{T-1}$.

3.1 Point forecast accuracy

Consistent with the existing literature and since volatility is persistent, the natural benchmark forecasting model is the $AR(1)$ model. To disentangle the relevant importance of the autoregressive part of the model vs the inclusion of candidate predictors, we also employ an $AR(q_t)$ model as a benchmark. In this case, the volatility forecast coincides with the forecast in the autoregressive model (1) when no predictor is included, i.e. $k = 0$. As a measure of forecast accuracy, we employ the out-of-sample R^2 computed as $R_{OS}^2 = 1 - \frac{MSFE_i}{MSFE_0}$, where $MSFE_i$ is the Mean Square Forecast Error associated with each of our competing models and specifications and $MSFE_0$ is the respective value for the benchmark model, both computed over the out-of-sample period. Positive values are associated with superior forecasting ability of our proposed model/specification. To evaluate the statistical significance of our conditional mean forecasts relative to the benchmark, we employ the Clark and West (2007) (CW) approximate normal test, which is an adjusted version of the Diebold and Mariano (1995) and West (1996) statistic. The test, in conjunction with the standard normal distribution, generates asymptotically valid inferences when comparing forecasts from nested linear models.⁵

3.2 Quantile forecast accuracy

While MSFE is an appropriate loss function for the evaluation of point (conditional mean) forecasts, it is not valid and informative for the evaluation of quantile forecasts. As proposed by Gneiting and Raftery (2007) and Gneiting and Ranjan (2011), the same loss function, i.e. the asymmetric linear loss function given by Equation (3) should be employed in both model estimation and forecast evaluation. Following Manzan (2015), we evaluate quantile forecasts using the quantile score (QS). The QS function focuses on a specific quantile τ and provides a local evaluation of the forecasts. The QS for the τ th quantile forecast of model/specification i is given by

$$QS_{t+1|t}^i(\tau) = \left[RV_{t+1} - \widehat{RV}_{i,t+1}^{(\tau)} \right] \left[\tau - I(RV_{t+1} - \widehat{RV}_{i,t+1}^{(\tau)} < 0) \right], \quad (7)$$

where I denotes the indicator function that takes the value of 1 if the argument is true and zero otherwise, RV_{t+1} is the actual realisation of volatility at time $t + 1$, and $\widehat{RV}_{i,t+1}^{(\tau)}$ is the related forecast for quantile τ . This scoring rule is defined in negative orientation coinciding with the notion of a loss function, so that when comparing two models, we prefer the one with the lowest score.

To evaluate the hypothesis of equal forecast accuracy of the quantile forecasts, we consider the approach proposed by Amisano and Giacomini (2007) and Giacomini and White (2006). Specifically, we employ the

⁵When comparing the aggregated quantile forecasts, we no longer have a nested environment. To this end, we also evaluate our forecasts on the basis of the Diebold and Mariano (1995) test.

test statistic

$$t = \frac{\overline{QS}_i(\tau) - \overline{QS}_0(\tau)}{\hat{\sigma}}, \quad (8)$$

where $\overline{QS}_i(\tau), \overline{QS}_0(\tau)$ denote the averages over the out-of-sample period of the quantile scores for a given quantile τ for the proposed model/ specification i relative to the benchmark model score and $\hat{\sigma}$ denotes the heteroscedasticity and autocorrelation consistent (HAC) standard error estimator of the quantile score difference.⁶ Assuming suitable regularity conditions, the statistic t is asymptotically standard normal under the null hypothesis of vanishing expected score differentials. In the case of rejection, model i is preferred to the benchmark if t is negative, and the benchmark model is preferred if t is positive.

3.3 Density forecast accuracy

The entire volatility density is not readily available from a quantile regression approach. However, it can be approximated by interpolating a set of volatility quantile forecasts. Such a method does not require a distributional assumption for the innovations of the conditional quantiles. To evaluate the performance of a model to forecast an area of the distribution, we employ the Weighted Quantile Score (WQS). The WQS is constructed by integrating the QS (given by Equation (7)) across τ with the score multiplied by a weight function that focuses on the specific part of the distribution as follows:

$$WQS_{t+1|t}^i = \int_0^1 QS_{t+1|t}^i(\tau)\omega(\tau)d\tau,$$

where $\omega(\tau)$ denotes a weight function in the unit interval.⁷ This accommodates the forecaster's interest to evaluate specific areas of the distribution. For example, a trader or a risk manager might be interested in the left/right tail of the realised volatility distribution, rather than a specific quantile. We employ four weight functions which are of particular interest: (1) $WQS_1 : \omega(\tau) = 1$ which provides an overall evaluation of the forecast distribution since it is a uniform weight, (2) $WQS_2 : \omega(\tau) = \tau(1 - \tau)$, which puts relative more weight in the middle of the distribution, (3) $WQS_3 : \omega(\tau) = (1 - \tau)^2$, which assigns more weight to the left tail of the distribution; and (4) $WQS_4 : \omega(\tau) = \tau^2$, which puts more weight on the right tail. In a similar vein, we perform tests of equal forecasting ability on the basis of Weighted Quantile Scores by replacing QS in Equation (8) with WQS .

Furthermore, to evaluate the performance of a model to forecast the entire distribution, we calculate the mean log predictive scores (MLPS) of candidate models and test for equal predictive accuracy by means of the Diebold and Mariano statistic. The test statistic is defined to be negatively oriented, i.e. a candidate model is preferred to the benchmark if the statistic is significantly negative.

⁶Since we focus on one step ahead forecasts, HAC adjustment is not necessary.

⁷In our forecasting experiment, we replace the continuous version of WQS with a discrete version summing over the quantiles of interest.

4 Empirical findings

4.1 Data and forecast construction

The data we employ comprises the ‘long’ (updated to December 2015) sample of Christiansen et al. (2012) who provide a detailed description of transformations and datasources. The variable of interest is the realised volatility (RV) of the S&P 500 index, defined as the square root of the realised variance, computed as the sum of squared intra-period (daily) returns, i.e.

$$RV_t = \sqrt{\sum_{j=1}^{M_t} r_{t;j}^2},$$

where $r_{t;j}$ is the j th daily continuously compounded stock market return in month t and M_t denotes the number of trading days during month t . As the number of intra-period observations becomes large, realised volatility is an accurate proxy for the true, but latent, integrated volatility (Andersen et al., 2003, 2006). Figure 1 plots the analysed realised volatility series. It can be seen that the series is highly variable with periods of high volatility corresponding to a number of financial crises.

[FIGURE 1 AROUND HERE]

Our forecasting experiment is conducted on a monthly basis and data span 1926:12 to 2015:12. The out-of-sample forecast evaluation period starts in January 1937, so the first estimation window, which is continuously updated consists of 120 observations. Following Christiansen et al. (2012), we rely on a set of 13 macroeconomic and financial predictive variables. Specifically, we consider stock valuation ratios such as the dividend price ratio (DP) and the earnings-price ratio (EP), commonly considered in predictive regressions for stock returns. To capture the leverage effect suggesting that negative returns are associated with higher subsequent volatility, we include lagged equity market returns (MKT). We also include the Fama and French (1993) risk factors, i.e. the size factor (SMB), the value factor (HML) and a short-term reversal factor (STR) which is related to market volatility and distress. Turning to interest-rate/bond related variables, we employ five variables ranging from short-term government rates to long-term government bond yields and returns along with their spreads. These are the Treasury bill rate (TB), the interest rate on a three-month Treasury bill; Long-term return (LTR), the return on long-term government bonds; Term spread (TMS), the difference between the long-term yield and the Treasury bill rate; Relative T-bill rate (RTB), the difference between the T-bill rate and its 12-month moving average and Relative Bond rate (RBR), the difference between LTR and its 12 month moving average. To proxy for credit risk, which tends to be higher in situations where leverage rises and should influence volatility, we rely on the yield spread between BAA and AAA rated bonds, i.e. the default spread (DEF). Finally, to capture the overall macroeconomic environment, we employ the inflation rate, INF, monthly growth rate of CPI (all urban consumers).

4.2 Point forecasts

First, we discuss the out-of-sample performance of point forecasts obtained by univariate linear and quantile autoregressive models employing Equations (1) and (2) and one predictor at a time.⁸ Table 1, Panel A (Column 1) presents the R_{OS}^2 statistics of all linear univariate models relative to the $AR(1)$ benchmark model. Positive values of R_{OS}^2 indicate superior forecasting performance of the predictive models with respect to the $AR(1)$ forecasts. The statistical significance of the corresponding forecasts is assessed by using the Clark and West (2007) MSFE-adjusted statistic (reported in Panel B of Table 1). Our findings suggest that all variables significantly outperform the predictions of the $AR(1)$ model with varying degrees of outperformance though. Equity market returns, inflation and the Treasury bill rate emerge as the most powerful predictors as improvements in R_{OS}^2 range from 5.98% to 7.76%. Our point volatility forecasts constructed via quantile forecast aggregation (Equations (4) -(6)), reported in columns 2-4, point to significant forecast improvements. The respective R_{OS}^2 are greater than 6.2% and reach 12.5% for MKT and the FW1 weighting method. All aggregation methods perform equally well, with FW1 being the most accurate, by a narrow margin though.

[TABLE 1 AROUND HERE]

Panels C and D of Table 1 report our findings with respect to the $AR(q)$ benchmark. In this way, we reveal the forecasting ability of predictors after we have taken into account the past history of realised volatility. Consequently and contrary to the $AR(1)$ case, only a few variables significantly outperform the predictions of the $AR(q)$ model (Panel C, Column 1) when the simple linear model is employed. Specifically, positive and significant R_{OS}^2 are associated with the three best performing predictors of the $AR(1)$ benchmark; namely MKT, INF and TB. More importantly, our quantile aggregated forecasts (FW1-FW3, Columns 2-4) outperform the benchmark irrespective of the candidate predictor employed. Forecast improvements range from 2.58% for DEF and FW2 to over 9% for MKT and FW1.

Next, we evaluate the ability of the equally weighted CSAR approach to provide more accurate point forecasts relative to both the $AR(1)$ and $AR(q)$ benchmark. Column 1 of Panels E and F report the related R_{OS}^2 values with respect to the $AR(1)$ and $AR(q)$ benchmarks (all of which are significant). The first cell of each panel corresponds to the combinations of single-variable models ($k = 1$) associated with the approach followed by Rapach et al. (2010). Interestingly, our results corroborate the existing literature on the increased benefits of forecast combinations. A sizable improvement of 4.78% and 1.08% over the benchmark $AR(1)$ and $AR(q)$, respectively, is evident. Focusing on forecasts produced by subset linear regressions for various values of k , we note that as we increase the number of subsets (k) the subset linear regression with $k \geq 2$ generates larger R_{OS}^2 values than the case of $k = 1$. In more detail, R_{OS}^2 is maximised

⁸Note that, since volatility can not be negative, all negative mean and quantile forecasts obtained by our predictive models are set to zero.

at $k = 7$ reaching the value of 9.71% (over the AR(1) benchmark) suggesting that averaging over all seven-variate models is optimal. Increasing k further leads to a deterioration in forecasting performance with a value of 5.25% for the Kitchen Sink model ($k = 13$). This predictive performance depends on the choice of the value of k . Therefore, we employ our algorithm in order to produce ‘optimal forecasts’, i.e. identify the best subset for each point of the out of sample period. The last two lines of Panels E and F report the related R_{OS}^2 values along with the CW t-statistic. Our results suggest that our likelihood-based approach to selecting k in real time is extremely successful, since the value of R_{OS}^2 obtained is equal to 7.95% and significant. With respect to the alternative $AR(q)$ benchmark, our findings, reported in Panel F, remain qualitatively similar.

We now turn our attention to evaluating the forecasting performance of the equally weighted CSQAR approach. Columns 2-4 (Table 1, Panel E) report our findings for the three weighting schemes (FW1-FW3) relative to the AR(1) benchmark respectively, while Columns 2-4 (Panel F) correspond to the $AR(q)$ benchmark. Our results indicate that the point forecasts generated via the CSQAR approach are superior to the linear ones. For example, averaging across k -variate quantile models generates R_{OS}^2 values that are greater than 8.59% as opposed to 4.78% for the k -variate linear models relative to the AR(1) benchmark. The FW3 weighting scheme is the one that emerges as superior judging from the higher R_{OS}^2 values in the majority of subsets. This superior performance stems from the fact that this method utilises a finer grid of quantiles and puts weights on the extreme 10% and 90% quantiles and as a result approximates the future realised volatility density more efficiently. Finally, it is evident that our algorithm is successful in identifying the best subset for each quantile and time period resulting in R_{OS}^2 values that range between 9.98% to 10.52%. Our findings are qualitatively similar when we employ the AR(q) benchmark.

Turning to the Bayesian combination scheme, Panels G and H (Table 1) report the related R_{OS}^2 values. Overall our findings suggest that this combination scheme further improves point forecast accuracy. For example, our likelihood-based approach to selecting k results in R_{OS}^2 that range between 10.03% to 12.78% relative to the AR(1) benchmark. Comparing the two combination schemes we observe that the Bayesian combination scheme improves equally weighted forecasts by more than 2%.

To shed light on the success of our algorithm, we plot in Figure 2 the selected values of k for both the CSAR (Panel A) and CSQAR (Panels B-H) approaches, under the prior distribution specified in the Appendix (with $\pi=1/2$). The proposed algorithm is flexible enough to allow for variability of the selected k across quantiles and, therefore, information on the best complete subset for each quantile of the volatility distribution can be incorporated within our approach. It is evident that for the CSAR approach the most frequently selected subset is the one corresponding to four-variate models. More importantly, the selected value of k varies across quantiles. In general, larger values of k are selected for the upper quantiles ($\tau=0.75$ and $\tau=0.90$). For this part of the distribution, the most frequently selected subsets are those of $k = 5$ to $k = 7$. For the central and lower quantiles, our algorithm tends to choose values of k between 3 and 5.

[FIGURE 2 AROUND HERE]

To further assess the performance of our CSQAR forecasts over time, Figure 3 plots the cumulative squared forecast error differences between the optimal CSQAR approaches (EW and Bayesian) and the AR(q) benchmark (Panels (a) and (c)), and the optimal CSAR (EW and Bayesian) benchmarks (Panels (b) and (d)). An increase in the line indicates a better performance of the proposed methodology over the respective benchmark. The figure confirms graphically the improved forecasting performance of our approach which is particularly positive around the period of 1987 crash and relatively stable in the more recent time period. It further illustrates the superiority of the Bayesian combination scheme for constructing point volatility forecasts.

[FIGURE 3 AROUND HERE]

Overall, our findings are in line with Paye (2012) and Christiansen et al. (2012) who find improved forecasting benefits for realised volatility when employing macro-finance variables. Specifically, Paye’s (2012) findings are similar to ours as the author also finds that individual forecasts rarely outperform an AR(2) benchmark while simple combination schemes are consistently successful. In a similar vein, Christiansen et al. (2012) account for model uncertainty by employing Bayesian techniques and find that their forecasting models beat the AR benchmarks. In our study, we also take into account predictor uncertainty in both a linear and quantile regression framework but focus more on uncertainty stemming from subset selection. Moreover, the flexibility of our approach combined with information from the whole future realised distribution leads to superior forecasts compared with the aforementioned studies.

4.3 Quantile forecasts

In this subsection, we evaluate the forecasting performance of our CSQAR approach relative to the $AR(q)$ and the $QAR(q)$ benchmarks. Our aim is to reveal whether our flexible approach of aggregating predictor information across quantiles is successful and to provide a guidance for selecting the best subset for each quantile of interest via our algorithm.

[TABLE 2 AROUND HERE]

Table 2 (Panel A) reports the Quantile Score test statistics for the null hypothesis of equal quantile forecast accuracy of the augmented QAR models relative to the benchmark $AR(q)$. Quantiles of the $AR(q)$ model are calculated assuming a Gaussian distribution with zero mean and constant conditional variance via Equation (1), where $\beta_i = 0$. The quantiles of interest are the ones employed in the point forecast aggregation, i.e. $\tau \in \{0.10, 0.25, 1/3, 0.50, 2/3, 0.75, 0.90\}$. The most important finding that emerges is that the augmented QAR models are more accurate than the $AR(q)$ benchmark. In more detail, forecast accuracy, judged by the statistical significance of the QS tests, is more pronounced for the left tail ($\tau = 0.10, \tau = 0.25$)

and the central to right tail part ($\tau = 2/3, \tau = 0.75$) of the realised volatility distribution. The central to left tail part ($\tau = 1/3, \tau = 0.50$) and the 0.90 quantile of the realised volatility distribution is the one exhibiting the weakest statistical significance. Overall, our findings suggest that quantile forecasting models can produce more accurate forecasts of the realised volatility future quantiles.

However, the results reported so far do not reveal any information on the ability of specific predictors to beat the simple QAR model. Panel B of Table 2 reports the related findings. Quite interestingly, no predictor succeeds in improving forecasts for all the quantiles of the distribution considered. Similarly to the linear case, the most useful predictor is MKT and to a lesser extent INF. Specifically, MKT can improve forecasts in all parts of the distribution with the exception of the lower left tail. With respect to INF, significant forecasting improvements are apparent for the right part of the distribution. The poor forecasting performance of individual predictors seeks for a way to aggregate predictor information and produce superior point and quantile forecasts. As we show our CSQAR approach can offer a leeway to this issue.

Table 2 (Panel C) reports the QS test statistics for the null hypothesis of equal quantile forecast accuracy of the EW CSQAR models relative to the $AR(q)$ benchmark. Several findings emerge from this comparison. First, irrespective of the subset employed, our EW CSQAR significantly outperforms the benchmark suggesting that the normality assumption fails to capture the characteristics of the future realised volatility distribution. Second, judging from the values of the test statistics, significance is more pronounced in the left tail ($\tau = 0.10, \tau = 0.25$) and the central to right tail part ($\tau = 2/3, \tau = 0.75$). The least significant values are associated with the right tail ($\tau = 0.90$) and the central quantiles of $\tau = 1/3$ and $\tau = 0.50$. This finding suggests that, although surges in volatility are difficult to forecast, the EW CSQAR adequately captures high future volatility realisations. Third, the best subset, as shown by the lowest t-statistics, varies across quantiles. For example, for the left tail ($\tau = 0.10$), the best subsets are the two- and three-variate ones, while this increases to five-variate models for the right tail quantiles. Finally, our algorithm succeeds in identifying the best subset in a time-varying and quantile-varying manner particularly well, since its performance is always close to the best subset. Our findings with respect to the Bayesian CSQAR approach, reported in Panel E, are qualitatively similar.

Next, we focus on the forecasting benefits of the EW CSQAR approach relative to the QAR one. Our findings (Table 2, Panel D) suggest that, depending on the quantile of interest, superiority of the EW CSQAR approach is evident for subsets lower than $k = 9$. Especially for the left tail ($\tau = 0.10$), this approach leads to statistically superior forecasts for subsets of less than 6 variables. On the other hand, for the right part of the distribution, improved forecasting performance of the EW CSQAR approach is evident for all subsets containing fewer than 10 variables. More importantly, our algorithm succeeds in selecting the optimal value of k since it produces more accurate forecasts than the simple QAR approach in all cases. However, focusing on the Bayesian approach, Panel F, our findings suggest that this approach produces

superior forecasts for more central quantiles ranging from $\tau = 0.25$ to $\tau = 0.75$.

4.4 Density forecasts

We now compare the ability of our approach to produce superior distributional forecasts relative to the simple QAR model. Table 3 reports the related findings. Panel A focuses on the individual predictors and Panels B and C on our EW and Bayesian CSQAR approaches, respectively. The first four columns of each panel report the test statistics of the Weighted Quantile Scores, which test for equal forecasting ability for a particular part of the distribution. The last column reports the MLPS t -statistics.⁹ Similar to the findings reported so far and based on the WQS statistics, only MKT and INF produce superior density forecasts. However, the MLPS t -statistic also points to improved density forecasts for the TB as well. Turning to the EW and Bayesian CSQAR approaches, the WQS tests show significant improvements in all parts of the distribution for subsets lower than $k = 5$. With respect to MLPS, we may note that superiority of EW (Bayesian) approach is evident for subsets lower than $k = 5$ ($K = 1$). More importantly, the density forecasts obtained from our algorithm for selecting the optimal value of k provide highly significant improvement with respect to the QAR model according to the WQS statistics. However, the MPLS statistics are significant only for the EW combination scheme.

[TABLE 3 AROUND HERE]

Our findings have important implications for several empirical finance areas, such as risk management, asset allocation, pricing, and trading strategies. In all the aforementioned areas, constructing accurate point volatility forecasts is crucial. For example, pricing equity derivatives, developing equity derivative trading strategies and measuring portfolio risk require reliable volatility forecasts. The recent financial crisis has highlighted the importance of being able to anticipate changes in volatility. From a regulatory perspective, volatility predictions are important when projecting risk measures, such as portfolio Value at Risk or Expected Shortfall, stress-testing and the design of risk mitigation strategies. Furthermore, our approach that also allows for volatility quantile and density forecasting is important for trading/pricing volatility derivatives, designing volatility hedges for portfolios, and in general assessing investment decisions.

4.5 Robustness checks

In this section, we conduct a series of robustness tests. First, we consider two alternative out-of-sample periods corresponding to two different initialisation out-of-sample dates, namely January 1957 and January 1977. These periods coincide with the ones considered by Christiansen et al. (2012). Second, we address the issue of the logarithmic transformation of RV. Specifically, we examine the impact this transformation

⁹Please note that the WQS t -statistics are calculated from the seven quantile scores corresponding to the quantiles used in the point forecast construction. For calculating the MLPS t -statistics, a finer grid of quantiles ($\tau \in \{0.10, 0.15, \dots, 0.85, 0.90\}$) is employed to provide a better density approximation.

has on our findings. Finally, we consider a number of alternative combination schemes to produce point forecasts from quantile forecasts. The robustness results are qualitatively similar and are not reported for reasons of space.¹⁰

5 Conclusions

In this study we propose a quantile forecast combination approach to realised volatility prediction. The aim of our analysis is to construct point, quantile and density forecasts of realised volatility, which take into account the benefits emerging from the subset framework, the quantile regression framework and the information given by the potential predictors. We analyse the US stock market volatility series, assuming that quantiles of future volatility depend on their own lags and macroeconomic and financial variables. We consider several conditional quantiles of future volatility and allow for heterogeneous degrees of persistence of realised volatility and asymmetric dynamic responses of macroeconomic and financial variables at different parts of the distribution. To allow for varying degrees of persistence in the distribution of volatility, we extend the Complete Subset Regression (CSR) framework of Elliott et al. (2013) to a Complete Subset (Quantile) AutoRegressive one (CSQAR approach). The results of our study are very promising. Our findings suggest that no single variable is able to provide more information for the evolution of the volatility distribution beyond that contained in its own past. On the other hand, our CSQAR approach achieves superior point, quantile and density predictive performance relative to univariate models and the CSR approach.

Appendix. Bayesian weighting scheme and algorithm for selecting k

Here we describe the likelihood-based method used for obtaining the Bayesian weighting scheme and for selecting k in real time. Under the Bayesian approach to inference, uncertainty about any quantity of interest is represented by probability distributions. In regression variable selection problems there is uncertainty about the model specification. In our setting, it is also of particular interest to quantify the uncertainty about the complete subset that will be used for predicting either the mean or the quantiles of realised volatility at each time point in the out-of-sample period, indexed by $t + 1$. To this end, within our approach we need to compute the posterior probabilities of different model specifications as well as of all values of k ($k \in \{1, 2, \dots, K\}$), based on the data up to time t , select the most probable value of k and produce a point or quantile forecast at time $t + 1$, based on the selected complete subset.

In a Bayesian context, the random quantities of interest are the model specifications, representing the set of predictors included in the j th model and denoted by m_j , $j = 1, \dots, M$, $M = \sum_{i=1}^K n_{i,K}$, the value of k and the totality of the model parameters, denoted by θ_{m_j} . After specifying appropriate prior distributions

¹⁰This set of results are available from authors upon request.

for these quantities, $P(m_j)$, $P(k|m_j)$ and $f(\theta_{m_j}|m_j, k)$, their joint posterior distribution is given by

$$f(m_j, k, \theta_{m_j}|RV_{1:t}) \propto P(m_j)P(k|m_j)f(\theta_{m_j}|m_j, k)L(RV_{1:t}|m_j, k, \theta_{m_j}),$$

where $L(RV_{1:t}|m_j, k, \theta_{m_j})$ is the likelihood of the data up to time t under model specification m_j . This is the normal likelihood under the CSAR approach and the quantile specific Laplace likelihood under the CSQAR approach. Dependence on the set of predictors has been suppressed for simplicity. The posterior probability of a particular model specification conditional on the value of k is given by

$$P(m_j|k, RV_{1:t}) \propto P(m_j)P(k|m_j) \int f(\theta_{m_j}|m_j, k)L(RV_{1:t}|m_j, k, \theta_{m_j})d\theta_{m_j}.$$

These posterior probabilities are used as the respective Bayesian weights within our combination scheme. The weights are obtained by dividing the right-hand side of the above equation by the sum of all these quantities for the model specifications belonging to the same subset (k). Moreover, the marginal posterior distribution of k , is obtained as

$$P(k|RV_{1:t}) \propto \sum_{j=1}^M P(m_j)P(k|m_j) \int f(\theta_{m_j}|m_j, k)L(RV_{1:t}|m_j, k, \theta_{m_j})d\theta_{m_j}.$$

The integral $\int f(\theta_{m_j}|m_j, k)L(RV_{1:t}|m_j, k, \theta_{m_j})d\theta_{m_j}$ is the marginal likelihood of the data under the linear or quantile regression with k predictors and model specification m_j , i.e. $L(RV_{1:t}|m_j, k)$. In this paper, we estimate the marginal likelihood by the BIC approximation which is given by

$$\widehat{L}(RV_{1:t}|m_j, k) = \exp\{L(RV_{1:t}|m_j, k, \widehat{\theta}_{m_j}) - k \ln(t)/2\},$$

where $\widehat{\theta}_{m_j}$ denotes the ML estimate of θ_{m_j} . Alternatively, the marginal likelihood of quantile regression models can be estimated by Laplace approximation, while for linear regression models it can be computed analytically.

The prior specification we consider is the following. The prior probability of the j th model is taken to be $P(m_j) = \pi^{k_j}(1 - \pi)^{K-k_j}$, where π is the prior probability of including a predictor in the model, which is taken fixed and prespecified, and k_j is the number of predictors included in model m_j . We set π equal to $1/2$, thus reflecting prior ignorance about the number of predictors required to capture the volatility dynamics. The prior probability of k given the model specification m_j is then $P(k|m_j) = 1$, if $k_j = k$, and $P(k|m_j) = 0$, otherwise. This prior structure leads to the joint prior of k, m_j being $P(k, m_j) = \pi^{k_j}(1 - \pi)^{K-k_j}I(k_j = k)$ and to the natural Binomial (K, π) marginal prior on k . Then, the marginal posterior distribution of k , is given by

$$P(k|RV_{1:t}) \propto \pi^k(1 - \pi)^{K-k} \sum_{j=1}^M \widehat{L}(RV_{1:t}|m_j, k)I(k_j = k).$$

Acknowledgments

We would like to thank the editors and two anonymous referees for constructive comments and suggestions on an earlier version of the paper. Ioannis Vrontos greatly acknowledge financial support from the Research Center of the Athens University of Economics and Business under the grant no EP-2778-01.

References

- [1] Amisano, G., and R. Giacomini (2007). Comparing Density Forecasts via Weighted Likelihood Ratio Tests. *Journal of Business and Economic Statistics*, 25, 177–190.
- [2] Andersen, T.G., Bollerslev T. and F.X. Diebold (2003). Modeling and forecasting realized volatility. *Econometrica*, 71, 529–626.
- [3] Andersen, T.G., Bollerslev T., Christoffersen P. and F.X. Diebold (2006). Volatility and correlation forecasting. In *Handbook of Economic Forecasting*, Elliot G., Granger C.W., Timmermann A. (eds). Elsevier: Amsterdam; 777–878.
- [4] Bates, J.M., and C.W.J. Granger (1969). The combination of forecasts. *Operational Research Quarterly*, 20, 451–468.
- [5] Chernozhukov, C., Fernandez-Val, I., and A. Galichon (2010). Quantile and Probability Curves Without Crossing. *Econometrica*, 78, 1093–1125.
- [6] Christiansen, C., Schmeling, M. and A. Schrimpf (2012). A comprehensive look at financial volatility prediction by economic variables. *Journal of Applied Econometrics*, 27, 956–977.
- [7] Clark, T.E., and K.D. West (2007). Approximately Normal Tests for Equal Predictive Accuracy in Nested Models. *Journal of Econometrics*, 138, 291–311.
- [8] Conrad, C. and K. Loch (2015). Anticipating long-term stock market volatility. *Journal of Applied Econometrics*, 30, 7, 1090–1114.
- [9] Diebold, F.X., and R.S. Mariano (1995). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*, 13, 253–63.
- [10] Elliott, G., Gargano, A., and A. Timmermann (2013). Complete Subset Regressions. *Journal of Econometrics*, 177, 2, 357–373.
- [11] Elliott, G., A. Gargano, and A. Timmermann (2015). Complete subset regressions with large-dimensional sets of predictors. *Journal of Economic Dynamics and Control*, 54, 86–110.
- [12] Fama E.F. and K.R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33, 3–56.
- [13] Gastwirth, J.L. (1966). On robust procedures. *Journal of the American Statistical Association*, 61, 929–948.
- [14] Giacomini, R., and H. White (2006). Tests of Conditional Predictive Ability. *Econometrica*, 74, 1545–1578.
- [15] Glosten, L.R., Jagannathan, R., and D.E. Runkle (1993). On the relation between the expected value and the volatility of nominal excess return on stocks. *Journal of Finance* 48, 5, 1779–1801.
- [16] Gneiting, T., and A.E. Raftery (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, 102, 359–378.
- [17] Gneiting, T., and R. Ranjan (2011). Comparing Density Forecasts Using Threshold and Quantile Weighted Scoring Rules. *Journal of Business and Economic Statistics*, 29, 411–422.
- [18] Goyal, A., and I. Welch (2008). A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. *Review of Financial Studies*, 21, 1455–508.
- [19] Huang, H., and T.H. Lee (2010). To combine forecasts or to combine information. *Econometric Reviews*, 29, 534–571.

- [20] Judge, G.G., Hill, R.C., Griffiths, W.E., Lutkepohl, H., and T.-C. Lee (1988). *Introduction to the Theory and Practice of Econometrics*. New York, Wiley.
- [21] Koenker, R., and G. Bassett (1978). Regression Quantiles. *Econometrica*, 46, 33-50.
- [22] Koenker, R., and Z. Xiao (2006). Quantile Autoregressions. *Journal of the American Statistical Association*, 101, 980–990.
- [23] Manzan, S. (2015). Forecasting the distribution of economic variables in a data-rich environment. *Journal of Business and Economic Statistics*, 33, 144–164.
- [24] Manzan, S., and D. Zerom (2013). Are Macroeconomic Variables Useful for Forecasting U.S. Inflation?. *International Journal of Forecasting*, 29, 469–478.
- [25] Meligkotsidou, L., E. Panopoulou, I.D. Vrontos, and S.D. Vrontos (2014a). A Quantile Regression Approach to Equity Premium Prediction, *Journal of Forecasting*, 33, 7, 558-576.
- [26] Meligkotsidou, L., E. Panopoulou, I.D. Vrontos, and S.D. Vrontos (2014b). Out-of-sample equity premium prediction: A complete subset quantile regression approach. Kent Business School working paper, available at <https://kar.kent.ac.uk/45149/>.
- [27] Mittnik, S., N. Robinsonov, and M. Spindler (2015). Stock market volatility: Identifying major drivers and the nature of their impact. *Journal of Banking and Finance*, 58, 1-14.
- [28] Panopoulou, E. and S. Vrontos (2015). Hedge fund return predictability; To combine forecasts or combine information?. *Journal of Banking and Finance*, 56, 103-122.
- [29] Paye, B.S. (2012). 'Deja Vol': Predictive regressions for aggregate stock market volatility using macroeconomic variables. *Journal of Financial Economics*, 106, 527-546.
- [30] Pedersen, T.Q. (2015). Predictable return distributions. *Journal of Forecasting*, 34, 114–132.
- [31] Rapach, D., Strauss, J., and G. Zhou (2010). Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy. *Review of Financial Studies*, 23, 2, 821-862.
- [32] Schwert, G. (1989). Why does stock market volatility change over time?. *Journal of Finance*, 44, 1115-1153.
- [33] Shan, K., and Y. Yang (2009). Combining Regression Quantile Estimators. *Statistica Sinica*, 19, 1171-1191.
- [34] Timmermann, A. (2006). Forecast combinations. In *Handbook of Economic Forecasting*, Vol. I, G. Elliott, C.W.J. Granger and A. Timmermann, eds. Amsterdam, Elsevier.
- [35] Tsiotas, G. (2015). A quasi-Bayesian model averaging approach for conditional quantile models. *Journal of Statistical Computation and Simulation*, 85, 10, 1963-1986.
- [36] Tukey, J.W. (1977). *Explanatory Data Analysis*. Addison-Wesley, Reading, MA.
- [37] West, K.D. (1996). Asymptotic Inference About Predictive Ability. *Econometrica*, 64, 1067–84.
- [38] Yu, K., and J. Zhang (2005). A Three-Parameter Asymmetric Laplace Distribution and Its Extension. *Communications in Statistics - Theory and Methods*, 34, 1867-1879.
- [39] Yu, K., and R.A. Moyeed (2001). Bayesian quantile regression. *Statistics and Probability Letters*, 54, 437-447.

Table 1. Performance of univariate AR, QAR, CSAR and CSQAR point forecasts

Panel A. R_{OS}^2 of AR and QAR vs. AR(1) benchmark					Panel B. CW $t - stats$ of AR and QAR			
Predictor	Linear	FW1	FW2	FW3	Linear	FW1	FW2	FW3
<i>DP</i>	1.58	7.78	7.72	7.53	2.28	2.96	3.07	2.92
<i>EP</i>	3.93	7.77	7.33	7.83	2.67	3.03	3.07	3.01
<i>MKT</i>	7.76	12.52	11.82	12.31	3.59	3.75	3.80	3.74
<i>DEF</i>	1.93	6.50	6.22	6.60	2.14	2.93	2.89	2.90
<i>HML</i>	3.43	7.83	7.31	7.78	2.45	3.06	3.07	3.02
<i>INF</i>	5.98	10.61	10.50	10.52	3.15	3.63	3.64	3.58
<i>LTR</i>	3.03	8.09	7.71	7.92	2.33	3.09	3.16	3.05
<i>RBR</i>	3.74	8.07	7.92	7.94	2.49	3.09	3.13	3.06
<i>RTB</i>	3.15	7.59	7.39	7.55	2.36	3.03	3.09	3.01
<i>SMB</i>	2.98	7.66	7.54	7.60	2.31	3.01	3.04	2.98
<i>STR</i>	3.22	7.90	7.87	7.80	2.41	3.08	3.13	3.01
<i>TB</i>	6.50	8.84	8.33	8.92	3.16	3.31	3.33	3.29
<i>TMS</i>	3.77	8.32	7.78	8.21	2.60	3.11	3.10	3.08

Panel C. R_{OS}^2 of AR and QAR vs. AR(q) benchmark					Panel D. CW $t - stats$ of AR and QAR			
Predictor	Linear	FW1	FW2	FW3	Linear	FW1	FW2	FW3
<i>DP</i>	-2.24	4.20	4.14	3.94	-0.63	3.57	3.66	3.54
<i>EP</i>	0.20	4.18	3.73	4.25	1.53	3.67	3.59	3.76
<i>MKT</i>	4.18	9.12	8.39	8.90	5.67	5.90	5.68	6.25
<i>DEF</i>	-1.88	2.87	2.58	2.97	-1.71	3.22	3.07	3.24
<i>HML</i>	-0.32	4.25	3.71	4.20	-1.17	3.78	3.64	3.82
<i>INF</i>	2.33	7.13	7.02	7.04	4.25	5.27	5.07	5.43
<i>LTR</i>	-0.74	4.52	4.12	4.35	-2.08	3.87	3.83	3.89
<i>RBR</i>	0.01	4.50	4.35	4.37	0.82	3.94	3.85	4.00
<i>RTB</i>	-0.61	4.00	3.80	3.96	-0.26	3.65	3.57	3.71
<i>SMB</i>	-0.79	4.08	3.95	4.02	-0.93	3.60	3.48	3.64
<i>STR</i>	-0.53	4.33	4.30	4.22	-1.29	3.76	3.78	3.73
<i>TB</i>	2.87	5.30	4.77	5.39	4.15	4.30	4.20	4.39
<i>TMS</i>	0.03	4.76	4.19	4.64	1.30	3.89	3.68	3.95

Panel E. R_{OS}^2 of EW CSAR and EW CSQAR vs. AR(1)					Panel F. R_{OS}^2 of EW CSAR and EW CSQAR vs. AR(q)			
k	CSAR	FW1	FW2	FW3	CSAR	FW1	FW2	FW3
1	4.78	8.91	8.59	8.83	1.08	5.37	5.04	5.29
2	5.90	9.47	9.17	9.46	2.24	5.96	5.65	5.95
3	7.02	9.94	9.67	9.99	3.41	6.44	6.16	6.50
4	8.04	10.36	10.11	10.48	4.47	6.88	6.62	7.00
5	8.87	10.72	10.48	10.89	5.34	7.25	7.00	7.43
6	9.44	10.98	10.76	11.20	5.93	7.53	7.30	7.75
7	9.71	11.17	10.93	11.42	6.21	7.72	7.47	7.98
8	9.69	11.26	11.00	11.54	6.18	7.82	7.54	8.10
9	9.39	11.29	10.96	11.59	5.87	7.85	7.50	8.15
10	8.86	11.30	10.86	11.61	5.32	7.86	7.40	8.18
11	8.07	11.32	10.76	11.63	4.50	7.88	7.30	8.20
12	6.93	11.38	10.58	11.68	3.31	7.94	7.11	8.25
13	5.25	11.25	10.09	11.51	1.57	7.81	6.59	8.08
k^*	7.95	10.31	9.98	10.52	4.38	6.83	6.49	7.05
<i>CW</i>	3.36	3.40	3.42	3.41	5.50	4.67	4.54	4.90

Panel G. R_{OS}^2 of Bayesian CSAR and CSQAR vs. AR(1)					Panel H. R_{OS}^2 of Bayesian CSAR and CSQAR vs. AR(q)			
k	CSAR	FW1	FW2	FW3	CSAR	FW1	FW2	FW3
1	9.18	11.52	11.05	11.41	5.66	8.08	7.60	7.97
2	10.61	12.25	12.10	12.43	7.14	8.84	8.68	9.03
3	11.73	12.73	12.47	12.87	8.30	9.34	9.07	9.49
4	11.28	12.83	12.60	12.93	7.83	9.45	9.21	9.55
5	10.45	12.68	12.36	12.69	6.98	9.29	8.96	9.30
6	9.30	12.44	12.08	12.51	5.78	9.04	8.67	9.11
7	8.49	12.25	11.71	12.36	4.94	8.85	8.29	8.96
8	7.81	11.94	11.40	12.09	4.23	8.52	7.96	8.67
9	7.23	11.69	11.11	11.89	3.63	8.26	7.66	8.47
10	6.73	11.58	10.88	11.81	3.11	8.15	7.42	8.39
11	6.27	11.57	10.79	11.82	2.63	8.14	7.33	8.40
12	5.79	11.51	10.55	11.76	2.13	8.08	7.08	8.33
13	5.25	11.25	10.09	11.51	1.57	7.81	6.59	8.08
k^*	10.03	12.74	12.34	12.78	6.54	9.35	8.94	9.39
<i>CW</i>	5.01	4.04	4.11	4.08	7.54	5.99	6.02	6.23

The Table reports the out-of-sample R^2 statistic of the univariate linear autoregressive model (AR), the quantile autoregressive models (QAR), the complete subset autoregressive models (CSAR) and the complete subset quantile autoregressive models (CSQAR, Equally Weighted, EW, and Bayesian weighting schemes) with respect to the AR(1) and AR(q) benchmark models for the out-of-sample period 1937:1-2015:12. Statistical significance for the R_{OS}^2 statistic is based on the Clark and West (2007) out-of-sample t-statistic (*CW*).

Table 2. Performance of univariate quantile models and CSQAR approach (quantile forecasts)

Panel A. Univariate models vs. AR(q) benchmark														Panel B. Univariate models vs. QAR(q) benchmark													
Predictor	0.10	0.25	1/3	0.50	2/3	0.75	0.90	0.10	0.25	1/3	0.5	2/3	0.75	0.90													
<i>DP</i>	-15.46	-9.45	-4.57	-6.56	-9.27	-9.93	-5.22	1.55	-0.36	1.29	1.34	0.84	0.91	0.69													
<i>EP</i>	-15.51	-9.86	-4.88	-6.74	-9.26	-10.03	-5.48	0.36	-0.63	1.74	0.64	0.76	0.18	-1.10													
<i>MKT</i>	-15.75	-10.99	-6.71	-8.37	-10.88	-11.40	-6.02	-0.21	-4.43	-2.67	-2.95	-3.07	-3.48	-1.68													
<i>DEF</i>	-14.91	-9.68	-5.05	-6.41	-8.91	-9.59	-4.98	0.77	0.19	0.94	0.64	1.22	0.08	-0.39													
<i>HML</i>	-14.97	-9.92	-5.34	-7.08	-9.03	-9.71	-5.08	-0.56	-0.55	1.71	-0.19	0.88	0.01	-1.69													
<i>INF</i>	-15.20	-10.49	-5.90	-7.72	-9.89	-10.50	-5.92	-0.87	-1.86	-1.13	-1.82	-1.86	-2.17	-2.48													
<i>LTR</i>	-15.17	-10.08	-5.32	-6.90	-9.04	-9.43	-4.36	0.85	-0.56	1.33	-0.27	0.99	0.90	1.99													
<i>RBR</i>	-16.03	-9.95	-5.35	-6.98	-9.09	-9.91	-4.68	-0.56	-1.20	1.32	-0.15	0.83	-0.39	0.53													
<i>RTB</i>	-15.47	-9.81	-5.27	-6.51	-8.54	-9.01	-4.72	1.79	-0.54	0.92	0.42	0.74	0.96	0.44													
<i>SMB</i>	-15.36	-9.79	-5.18	-6.67	-8.54	-8.97	-4.57	1.41	-0.76	0.85	0.43	1.06	1.19	0.88													
<i>STR</i>	-15.25	-9.88	-5.31	-7.04	-9.10	-9.39	-4.72	0.51	-0.52	0.49	-0.34	0.39	0.41	0.33													
<i>TB</i>	-15.96	-10.08	-5.21	-6.64	-9.03	-9.60	-5.25	-1.82	-1.43	-0.74	-0.56	-0.16	-0.35	-0.49													
<i>TMS</i>	-14.03	-10.09	-5.45	-6.92	-9.08	-9.32	-4.66	-0.21	-1.49	-0.04	0.62	0.00	0.23														

Panel C. EW CSQAR vs. AR(q) benchmark								Panel D. EW CSQAR vs. QAR(q) benchmark							
<i>k</i>	0.10	0.25	1/3	0.50	2/3	0.75	0.90	0.10	0.25	1/3	0.5	2/3	0.75	0.90	
1	-15.97	-10.49	-5.96	-7.66	-9.64	-10.34	-5.38	-3.99	-4.13	-2.83	-2.96	-1.66	-2.95	-2.34	
2	-16.10	-10.78	-6.34	-7.90	-9.94	-10.53	-5.62	-3.75	-4.58	-3.67	-3.60	-2.34	-3.07	-2.59	
3	-16.10	-10.95	-6.59	-7.93	-10.18	-10.68	-5.79	-3.12	-4.42	-3.80	-3.58	-2.73	-3.15	-2.63	
4	-16.08	-11.03	-6.69	-7.85	-10.34	-10.79	-5.89	-2.63	-4.12	-3.64	-3.43	-2.95	-3.25	-2.64	
5	-16.01	-11.03	-6.66	-7.66	-10.40	-10.82	-5.90	-2.15	-3.80	-3.34	-3.15	-3.00	-3.29	-2.57	
6	-15.91	-10.94	-6.53	-7.43	-10.35	-10.77	-5.83	-1.76	-3.46	-3.02	-2.88	-2.96	-3.22	-2.43	
7	-15.73	-10.73	-6.33	-7.14	-10.20	-10.62	-5.68	-1.37	-3.03	-2.69	-2.55	-2.81	-3.06	-2.24	
8	-15.45	-10.43	-6.06	-6.82	-9.95	-10.40	-5.53	-0.93	-2.59	-2.33	-2.13	-2.58	-2.83	-2.08	
9	-15.09	-10.06	-5.68	-6.49	-9.65	-10.12	-5.41	-0.46	-2.14	-1.86	-1.74	-2.30	-2.54	-1.98	
10	-14.61	-9.66	-5.24	-6.14	-9.31	-9.84	-5.31	0.05	-1.65	-1.33	-1.33	-1.99	-2.27	-1.88	
11	-14.07	-9.23	-4.76	-5.80	-8.97	-9.53	-5.11	0.56	-1.14	-0.77	-0.91	-1.63	-1.92	-1.62	
12	-13.43	-8.75	-4.16	-5.46	-8.48	-9.16	-4.82	1.10	-0.64	-0.02	-0.52	-0.99	-1.49	-1.23	
13	-12.31	-8.04	-3.24	-4.99	-7.79	-8.37	-4.17	1.66	0.21	1.14	-0.05	-0.10	-0.54	-0.40	
<i>k*</i>	-16.05	-10.81	-6.49	-7.81	-10.43	-10.84	-5.49	-2.66	-3.83	-3.46	-3.16	-2.88	-3.15	-1.94	

Panel E. Bayesian CSQAR vs. AR(q) benchmark								Panel F. Bayesian CSQAR vs. QAR(q) benchmark							
<i>k</i>	0.10	0.25	1/3	0.50	2/3	0.75	0.90	0.10	0.25	1/3	0.5	2/3	0.75	0.90	
1	-15.00	-10.86	-6.73	-8.35	-9.99	-9.98	-5.78	-1.90	-2.99	-2.50	-3.37	-2.48	-1.61	-1.27	
2	-15.71	-10.66	-6.01	-7.68	-10.67	-10.48	-5.80	-1.49	-2.71	-2.46	-2.94	-2.93	-2.33	-1.63	
3	-15.40	-10.26	-5.98	-7.54	-10.45	-10.43	-5.50	-1.12	-2.44	-2.26	-2.95	-2.81	-2.23	-1.37	
4	-14.89	-10.06	-5.72	-7.17	-10.45	-10.22	-4.80	-0.49	-2.30	-1.98	-2.50	-2.64	-2.01	-0.78	
5	-14.58	-9.86	-5.48	-6.79	-10.17	-9.96	-4.20	-0.00	-2.02	-1.66	-2.07	-2.33	-1.97	-0.37	
6	-14.55	-9.58	-5.19	-6.44	-9.87	-9.82	-4.35	0.22	-1.69	-1.33	-1.66	-2.14	-1.94	-0.70	
7	-14.35	-9.29	-4.89	-6.14	-9.57	-9.70	-4.58	0.46	-1.34	-1.00	-1.35	-1.87	-1.98	-1.01	
8	-14.09	-9.09	-4.64	-5.90	-9.26	-9.45	-4.46	0.68	-1.13	-0.72	-1.10	-1.60	-1.81	-0.85	
9	-13.83	-8.91	-4.43	-5.70	-8.91	-9.18	-4.26	0.84	-0.96	-0.47	-0.89	-1.34	-1.58	-0.59	
10	-13.56	-8.71	-4.26	-5.55	-8.60	-9.02	-4.11	1.01	-0.75	-0.23	-0.70	-1.09	-1.43	-0.39	
11	-13.28	-8.51	-4.07	-5.45	-8.33	-8.84	-4.14	1.19	-0.50	0.03	-0.57	-0.80	-1.22	-0.40	
12	-12.92	-8.32	-3.74	-5.27	-8.05	-8.65	-4.16	1.45	-0.24	0.45	-0.37	-0.45	-0.93	-0.39	
13	-12.31	-8.04	-3.24	-4.99	-7.79	-8.37	-4.17	1.66	0.21	1.14	-0.05	-0.10	-0.54	-0.40	
<i>k*</i>	-14.88	-9.94	-5.78	-7.27	-10.11	-10.19	-4.43	-0.32	-2.11	-2.03	-2.58	-2.21	-2.01	-0.70	

The Table reports the Quantile Score t-statistics for the null hypothesis of equal predictive ability of the univariate QAR models, Equally Weighted (EW) and Bayesian Combination schemes relative to the AR(q) and QAR(q) benchmark models for the out-of-sample period 1937:1-2015:12. Values less than -1.645 indicate that the proposed model outperforms the benchmark one, at the 5% level.

Table 3. Performance of univariate quantile models and CSQAR approach (density forecasts)

Predictor	Panel A. Univariate models vs. QAR(q) benchmark					Panel B. EW CSQAR vs. QAR(q) benchmark					Panel C. Bayesian CSQAR vs. QAR(q) benchmark						
	WQS_1	WQS_2	WQS_3	WQS_4	MLPS	k	WQS_1	WQS_2	WQS_3	WQS_4	MLPS	k	WQS_1	WQS_2	WQS_3	WQS_4	MLPS
<i>DP</i>	1.33	1.28	1.13	1.26	-1.06	1	-6.82	-6.46	-6.39	-5.19	-2.08	1	-3.86	-3.90	-3.55	-2.93	-2.29
<i>EP</i>	0.45	0.81	0.54	-0.21	-0.54	2	-5.99	-5.75	-5.77	-4.76	-2.34	2	-3.62	-3.69	-3.02	-3.07	-0.52
<i>MKT</i>	-4.12	-4.13	-3.69	-3.72	-2.39	3	-5.45	-5.29	-5.14	-4.51	-2.44	3	-3.15	-3.29	-2.69	-2.66	-1.55
<i>DEF</i>	0.90	0.98	0.90	0.55	1.99	4	-5.05	-4.93	-4.61	-4.33	-2.18	4	-2.62	-2.82	-2.23	-2.16	-0.89
<i>HML</i>	-0.04	0.33	-0.08	-0.31	1.04	5	-4.62	-4.53	-4.07	-4.10	-1.90	5	-2.15	-2.38	-1.77	-1.75	0.10
<i>INF</i>	-2.76	-2.59	-2.25	-2.77	-1.68	6	-4.20	-4.13	-3.58	-3.82	-0.94	6	-1.95	-2.10	-1.41	-1.80	0.18
<i>LTR</i>	1.81	1.48	0.63	2.14	0.62	7	-3.71	-3.67	-3.08	-3.44	-0.82	7	-1.76	-1.84	-1.06	-1.82	-0.21
<i>RBR</i>	0.21	0.20	-0.37	0.37	-0.97	8	-3.21	-3.18	-2.54	-3.07	-0.61	8	-1.45	-1.53	-0.75	-1.57	0.57
<i>RTB</i>	1.02	0.94	1.11	0.91	1.28	9	-2.71	-2.67	-1.98	-2.72	-0.20	9	-1.14	-1.23	-0.50	-1.27	1.44
<i>SMB</i>	1.20	1.10	0.87	1.30	2.43	10	-2.20	-2.17	-1.39	-2.39	0.50	10	-0.86	-0.96	-0.25	-1.02	1.55
<i>STR</i>	0.41	0.31	0.14	0.48	0.76	11	-1.63	-1.61	-0.77	-1.95	0.75	11	-0.61	-0.70	0.02	-0.85	1.47
<i>TB</i>	-1.12	-0.98	-1.50	-0.60	-1.95	12	-0.94	-0.92	-0.09	-1.37	1.31	12	-0.27	-0.34	0.39	-0.62	1.59
<i>TMS</i>	0.01	0.04	-0.54	0.31	1.28	13	0.20	0.16	0.92	-0.33	1.86	13	0.20	0.16	0.92	-0.33	1.86
						k^*	-4.50	-4.55	-4.52	-3.52	-1.86	k^*	-2.45	-2.67	-2.10	-1.97	-0.35

The Table reports the Weighted Quantile Score and MLPS t-statistics for the null hypothesis of equal predictive ability of the univariate QAR models, the EW and Bayesian CSQAR approaches relative to the QAR(q) benchmark for the out-of-sample period 1937:1-2015:12. Values less than -1.645 indicate that the proposed model outperforms the benchmark one at 5% level.

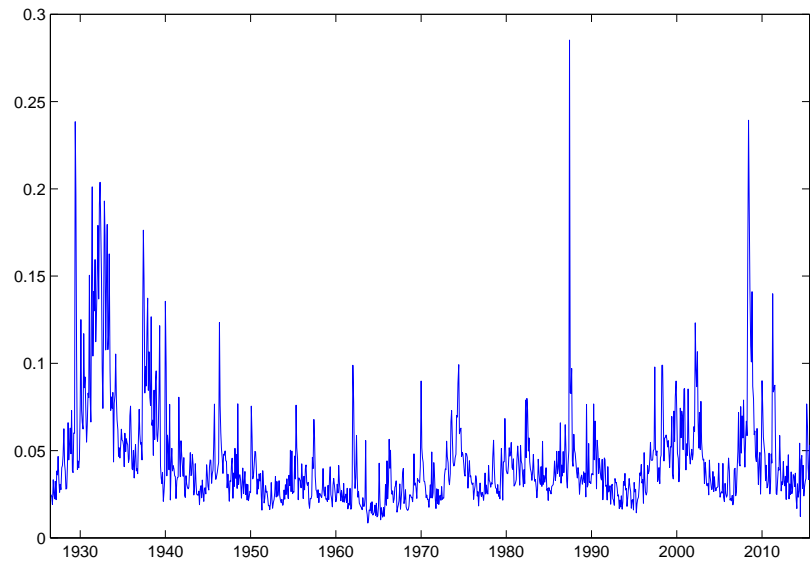


Figure 1: Plot of the realised volatility series.

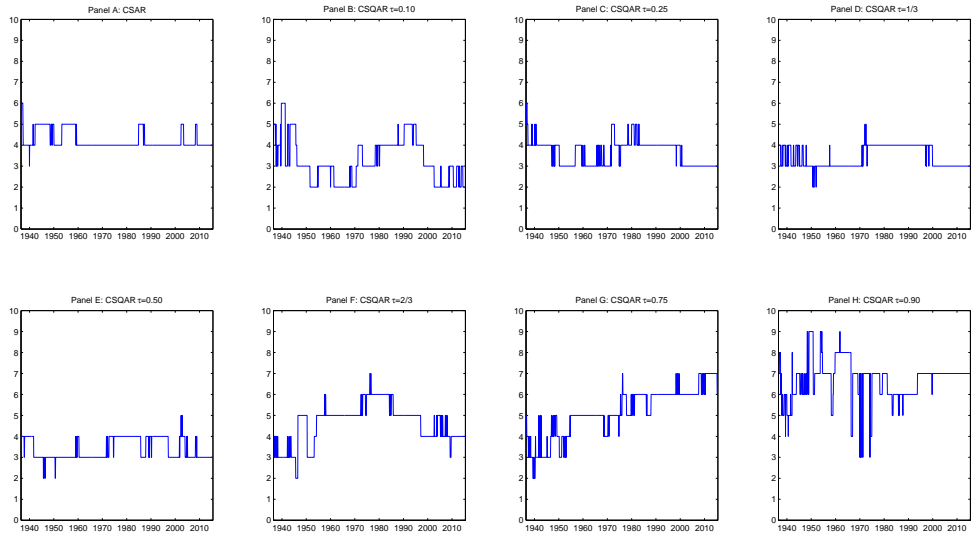


Figure 2: Selection of optimal subset (k^*) over the out-of-sample period.

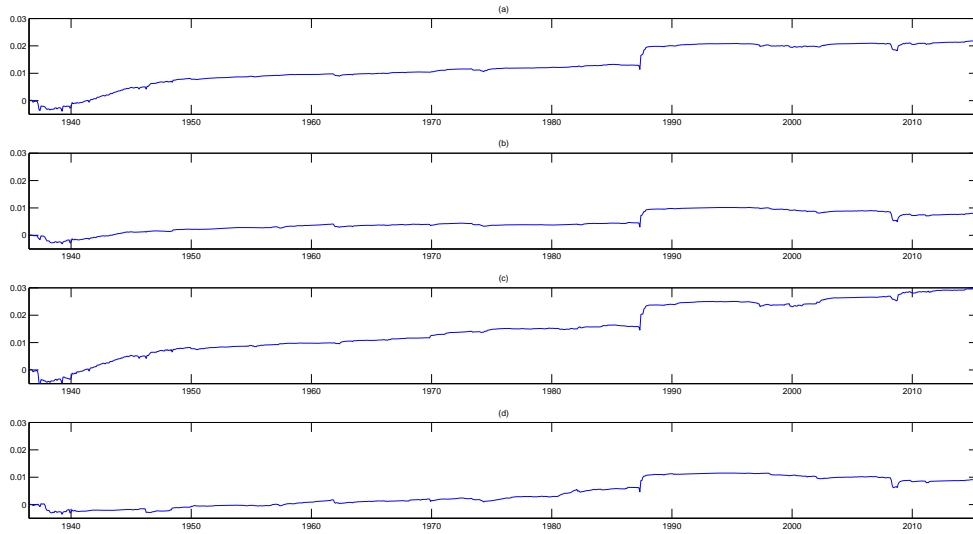


Figure 3: Cumulative sum of forecasts squared errors of (a) the optimal EW CSQAR forecasts vs the AR(q) benchmark, (b) the optimal EW CSQAR forecasts vs the optimal EW CSAR forecasts, (c) the optimal Bayesian CSQAR forecasts vs the AR(q) benchmark and (d) the optimal Bayesian CSQAR forecasts vs the optimal Bayesian CSAR forecasts over the out-of-sample period.