Quantile Forecast Combinations in Realised Volatility Prediction

Loukia Meligkotsidou[a], Ekaterini Panopoulou[b]*, Ioannis D.Vrontos[c], Spyridon D. Vrontos[d]

[a]Department of Mathematics, University of Athens, Greece

[b]Kent Business School, University of Kent, Canterbury, United Kingdom

[c]Department of Statistics, Athens University of Economics and Business, Greece

[d]Department of Mathematical Sciences, University of Essex, United Kingdom

Abstract

This paper tests whether it is possible to improve realised volatility forecasts by conditioning on macroeconomic and financial variables. We employ complete subset combinations of both linear and quantile forecasts in order to construct robust and accurate stock market volatility predictions. Our findings suggest that the complete subset approach delivers statistically significant out-of-sample forecasts relative to the autoregressive benchmark and traditional combination schemes. A recursive algorithm that selects, in real time, the best complete subset for each predictive regression quantile succeeds in identifying the best subset in a time- and quantile-varying manner.

JEL classification: G12; G22; C22; C53

Keywords: Realised volatility; Forecast combination; Predictive quantile regression; Robust point forecasts; Subset quantile regressions.

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

1 Introduction

Forecasting future volatility is essential for asset allocation, portfolio and risk management. A vast literature has investigated time variation in volatility and the linkages between volatility and macroeconomic and financial variables. Schwert (1989) relates the changes in the volatility of returns to the macroeconomic variables and addresses how bond returns, the short-term interest rate, producer prices or industrial production growth rate provide incremental information on monthly market volatility. Glosten et al. (1993) find evidence that short-term interest rates play an important role for future market variance. Whitelaw (1994) finds statistical significance for a commercial paper spread and the one-year treasury rate, while Brandt and Kang (2002) use the short-term interest rate, term premium and default premium, and find a significant effect. Other studies, including Hamilton and Lin (1996) and Perez-Quiros and Timmermann (2000), find evidence that the state of the economy is an important determinant in the volatility of the returns.

More recent contributions include Paye (2012) and Christiansen et al. (2012), who analyse the predictive content of financial and macroeconomic variables for monthly realized volatility. Specifically, Paye (2012) tests whether conditioning on macroeconomic variables can improve volatility forecasts and finds a link between several variables and stock market volatility. However, the author finds that improvements in out-of-sample forecasting accuracy mainly come from simple combinations of individual forecasts. Christiansen et al. (2012) employ a comprehensive set of macro-finance variables to predict the monthly realized 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. Specifically, they find the strongest predictive ability in variables associated with time-varying risk premia, leverage or financial distress. By selecting the most important predictor variables by means of Bayesian Model Averaging, they find that their forecast models beat autoregressive benchmarks although this performance varies across asset classes and over time.

This paper complements these recent contributions by focusing exclusively on an out-of-sample experiment conducted in a quantile forecast combination framework. Specifically, we apply the recently proposed methodology of Meligkotsidouet al. (2014b) to provide robust point forecasts of US stock market realised volatility. This forecasting approach is based on complete subset quantile regressions and exploits the benefits emerging from three strands of the literature on out-of-sample forecasting. First, the authors employ a quantile regression setting, which succeeds in producing robust and accurate point forecasts. Second, model uncertainty and parameter instability is reduced by employing quantile forecast combinations. Finally, employing complete subset quantile regressions induces shrinkage to the respective estimates and further helps reduce the effect of parameter estimation error.

In the context of equity premium predictability, Rapach et al. (2010) find that combinations of individual single variable predictive regression models significantly beat the historical average forecast as they reduce both model uncertainty and parameter instability.[1] Building on Rapach

[1] Timmermann (2006) provides a detailed review on forecast combination methodologies.

et al. (2010), Meligkotsidou et al. (2014a) incorporate the forecast combination methodology in a quantile regression setting. Their quantile regression approach to equity premium prediction allows them to cope with the non-linearity and non-normality patterns that are evident in the relationship between stock returns and potential predictors. In this way, robust and accurate equity premium forecasts are produced by combining a set of predictive quantile regressions in either a fixed or time-varying manner. A novel forecast combination method based on complete subset regressions is put forward by Elliott et al. (2013). The authors propose combining forecasts from all possible linear regression models that keep the number of predictors fixed. Their empirical application on equity premium predictability shows that subset combinations of up to four predictors generates superior forecast accuracy. Their approach introduces a complex version of shrinkage to the respective estimates which helps reduce the effect of parameter estimation error.[2] The authors propose constructing forecasts based on a simple averaging scheme of all the possible models employed keeping the numbers of regressors fixed. Finally, Meligkotsidou et al. (2014b) extend the framework of Elliott et al. (2013) to a quantile regression setting. They also utilise information from all the predictors simultaneously in order to produce combined quantile forecasts from all quantile regressions that keep the number of predictors fixed. Abstracting from the simple averaging schemes, the authors introduce several existing combination schemes into the quantile setting.

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).[3] Empirical contributions in the field of finance include Bassett and Chen (2001), Engle and Manganelli (2004), Meligkotsidou et al. (2009), Cenesizoglou and Timmermann (2012), Chuang et al. (2009) and Baur et al. (2012). The main advantage of the quantile regression framework lies in its ability to cope with non-linearity and non-normality patterns in the joint relationship between realised volatility and candidate predictors (see, inter alia, Guidolin and Timmermann, 2009; Guidolin et al., 2009; Henkel et al., 2011).

To anticipate our key results, we find that the complete subset quantile regression framework achieves superior predictive performance. In particular, our proposed approach can lead to an out-of-sample R^2 of 9.05% (relative to the autoregressive benchmark) as opposed to 6.90% of the BMA approach of Christiansen et al. (2012). The subset linear regression framework can also produce improved forecasts. While in the equity premium predictability literature, subsets of two or three variables perform better than the remaining specifications, in forecasting realised volatility subsets of six to ten variables (depending on the specification and combination scheme) emerge as superior. More importantly, the real time recursive algorithm for selecting k both in the linear and quantile framework, developed by Meligkotsidou et al. (2014b), succeeds in identifying the 'correct' value of k which is both time-varying and quantile-varying.

The outline of the paper is as follows. Section 2 describes the complete subset regression

[2] Shrinkage typically is employed in order to limit the number of parameters that have to be estimated when many potential predictors are available. Contributions to this field include the ridge regression (Hoerl and Kennard, 1970), model averaging (Bates and Granger, 1969; Raftery, Madigan and Hoeting, 1997), bagging (Breiman, 1996) and the Lasso (Tibshirani, 1996).

[3] See also Buchinsky (1994, 1995) and Yu, Lu and Stander (2003).

framework of Elliott et al. (2013) and introduces its extension to the quantile regression framework along with the proposed methodology for robust forecasting of realised volatility. Section 3 presents our empirical findings, while section 4 describes the proposed methodology for the recursive selection of the number of predictors and presents the associated findings. Section 5 summarizes and concludes.

2 Complete Subset Quantile Regressions

In this section we present the setup for our analysis. Section 2.1 outlines the Elliott et al. (2013) complete subset regressions framework and Section 2.2 extends this framework to subset quantile regressions.

2.1 Complete subset regressions

Elliott et al. (2013) propose a new method for combining forecasts based on complete subset regressions. For a given set of potential predictors, the authors propose combining forecasts from all possible linear regressions that keep the number of predictors fixed. For K possible predictors, there are K univariate models and $n_{k,K} = K!/((K-k)!k!)$ different $k-$variate models for $k \leq K$. The set of models for a fixed value of k is referred to as a complete subset and the authors propose using equal-weighted combinations of the forecasts from all models within these subsets indexed by k.

Suppose that we are interested in forecasting realized volatility, denoted by RV_t, using a set of K predictive variables. Since volatility is fairly persistent, we include an autoregressive term (lag=1) in the predictive regression to investigate whether there is additional predictive content of the macroeconomic and financial variables that goes beyond the information contained in lagged volatility. First we consider all possible predictive autoregressive models ($AR(1)$) with a single predictor, i.e. $k = 1$, of the form

$$RV_{t+1} = \alpha_i + c_i RV_t + \beta_i x_{it} + \varepsilon_{t+1}, \; i = 1, \ldots, K, \tag{1}$$

where RV_{t+1} is the observed realised volatility at time $t+1$, RV_t is the observed realised volatility at time t, x_{it} are the K observed predictors at time t, and the error terms ε_{t+1} are assumed to be independent with mean zero and variance σ^2. The predictive autoregressive models can be estimated using the Ordinary Least Squares (OLS) method by minimizing the sample estimate of the quadratic expected loss, $\sum_{t=0}^{T-1} (RV_{t+1} - \alpha_i - c_i RV_t - \beta_i x_{it})^2$, or the Maximum Likelihood (ML) approach after specifying the parametric form of the error distribution[4]. Similarly, a regression of RV_{t+1} can be run on a particular subset of the regressors and then average the forecasts across all k dimensional subsets to provide the forecast for the variable of interest, where $k \leq K$. Elliott et al. (2013) show that while subset regression combinations bear similarities to a complex version of shrinkage, they do not reduce to shrinking OLS estimates.

[4] The sample size T denotes any estimation sample employed in our recursive forecasting experiment. Details on the forecasting design are given in Section 3.

Rather the coefficient that controls shrinkage depends on all OLS 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. While the authors use equal-weighted combinations of forecasts within each subset along with approximate Bayesian Model Averaging, alternative weighting schemes can be employed. To this end, we also employ the Median, the Trimmed Mean, the Discount Mean Squared Forecast Error (DMSFE) of Stock and Watson (2004) along with the Cluster combining method, introduced by Aiolfi and Timmermann (2006).[5]

2.2 Complete subset quantile regressions

Following Meligkotsidou et al. (2014b), we incorporate the complete subset combination framework of Elliott et al. (2013) in a quantile regression setting. The proposed approach is designed as follows.

First, consider single predictor quantile autoregressive models ($k = 1$) of the form

$$RV_{t+1} = \alpha_i^{(\tau)} + c_i^{(\tau)} RV_t + \beta_i^{(\tau)} x_{it} + \varepsilon_{t+1}, \; i = 1, \dots, K, \tag{2}$$

where $\tau \in (0, 1)$ and the errors ε_{t+1} are assumed independent from an error distribution $g_\tau(\varepsilon)$ with the τth quantile equal to 0, i.e. $\int_{-\infty}^{0} g_\tau(\varepsilon) d\varepsilon = \tau$. Model (2) suggests that the τth quantile of RV_{t+1} given x_{it} and lagged realised volatility (RV_t) is $Q_\tau(RV_{t+1}|x_{it}, RV_t) = \alpha_i^{(\tau)} + c_i^{(\tau)} RV_t + \beta_i^{(\tau)} x_{it}$, where the intercept and the regression coefficients depend on τ. Both $c_i^{(\tau)}$ and $\beta_i^{(\tau)}$'s are likely to vary across τ's, revealing a larger amount of information about future realised volatility than the predictive autoregressive model (Equation 1). Estimators of the parameters of the quantile regression models in (2), $\hat{\alpha_i}^{(\tau)}, \hat{c}_i^{(\tau)}, \hat{\beta}_i^{(\tau)}$, can be obtained by minimizing the sum $\sum_{t=0}^{T-1} \rho_\tau \left(RV_{t+1} - \alpha_i{}^{(\tau)} - c_i^{(\tau)} RV_t - \beta_i^{(\tau)} x_{it} \right)$, where $\rho_\tau(u)$ is the asymmetric linear loss function, usually referred to as the check function,

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

In the symmetric case of the absolute loss function ($\tau = 1/2$) we obtain estimators of the median predictive regression models. A parametric approach to inference on the quantile regression parameters arises if the error distribution $g_p(\varepsilon)$ is specified. The error distribution that has been widely used for parametric inference in the quantile regression literature is the asymmetric Laplace distribution (for details, see Yu and Moyeed, 2001, and Yu and Zhang, 2005) with probability density function

$$g_\tau(\varepsilon) = \frac{\tau(1-\tau)}{\sigma^{(\tau)}} \exp \left[- \frac{|\varepsilon| + (2\tau - 1)\varepsilon}{2\sigma^{(\tau)}} \right], 0 < \tau < 1, \sigma^{(\tau)} > 0. \tag{4}$$

[5] To keep the analysis clear, Appendix A.1 provides a detailed description of the formation of these weighting schemes.

For $\tau = 1/2$, corresponding to the median regression, (4) becomes the symmetric Laplace density. A likelihood function can be formed by combining T independent asymmetric Laplace densities of the form (4), i.e.

$$L^{(\tau)}\left(RV_{1:T}|\alpha_i{}^{(\tau)}, \beta_i^{(\tau)}, \sigma^{(\tau)}\right) = \tag{5}$$
$$= \left(\frac{\tau(1-\tau)}{\sigma^{(\tau)}}\right)^T \exp\left\{-\frac{1}{\sigma^{(\tau)}}\sum_{t=0}^{T-1}\rho_\tau\left(RV_{t+1} - \alpha_i{}^{(\tau)} - c_i^{(\tau)}RV_t - \beta_i^{(\tau)}x_{it}\right)\right\}.$$

Then (5) can be used for likelihood based inference for the parameters $\alpha_i{}^{(\tau)}, c_i^{(\tau)}, \beta_i^{(\tau)}, \sigma^{(\tau)}$, for example for maximum likelihood estimation. The maximization of this likelihood function with respect to $\alpha_i{}^{(\tau)}, c_i^{(\tau)}, \beta_i^{(\tau)}$ is equivalent to minimizing the expected asymmetric linear loss, while the ML estimator of $\sigma^{(\tau)}$ is $\widehat{\sigma}^{(\tau)} = \frac{1}{T}\sum_{t=0}^{T-1}\rho_\tau\left(RV_{t+1} - \alpha^{(\tau)} - c_i^{(\tau)}RV_t - \beta_i^{(\tau)}x_{it}\right)$. Similarly to the predictive autoregressive case, the quantile regression (Equation 2) of RV_{t+1} can be run on a particular subset (k) of the regressors K, $k \leq K$, with the aim to produce realised volatility quantile forecasts.[6]

Next, we construct realised volatility point forecasts by combining quantile forecasts obtained from a set of complete subset regressions ($k - variate$ models with $k \leq K$). For each k, $n_{k,K}$ regressions 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 combine the predicted τ^{th} quantiles across all different subsets (k) of predictors ($n_{k,K}$ model specifications).This approach is the Quantile Forecast Combination (QFC) approach of Meligkotsiou et al. (2014b). With the exception of the Mean, Trimmed Mean and Median combining methods, the existing combination methods are not appropriate for combining predictor information in the quantile regression context. To this end, the MSFE loss function has to be replaced by a metric based on the asymmetric linear loss function (Equation 3). Following Meligkotsidou et al. (2014a), we employ the Discount Asymmetric Loss Forecast Error (DALFE) and the Asymmetric Loss Cluster (AL Cluster) in order to construct combined subset quantile forecasts (see Appendix A.2). This step yields a set of quantile forecasts (one for each τ_j), which are then combined into final robust point forecasts using either a fixed or a time-varying weighting scheme (see next section).

Finally, we consider the problem of constructing robust point forecasts of the equity premium based on a set of predictive quantile regressions as an alternative to the standard approach which produces forecasts based on the predictive mean regression model. Robust point estimates of the central location of a distribution can be constructed as weighted averages of a set of quantile estimators employing either fixed or time-varying weighting schemes as follows.

For a given model specification or a given complete subset that has been used for producing quantile forecasts, robust point forecasts can be constructed as weighted averages of a set of quantile forecasts. First, we employ standard estimators with fixed, prespecified weights of the

[6] The advantage of the parametric approach to 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 the Bayesian Information Criterion (BIC) or Bayesian model comparison. 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 (see Section 5).

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 combined, $\widehat{RV}_{t+1}^{(\tau)}$ denotes the quantile forecasts associated with the τth quantile and $\widehat{RV}_{t+1}$ is the produced robust point forecast. Here the weights represent probabilities attached to different quantile forecasts, suggesting how likely to predict the return at the next period each regression quantile is.

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

$$\text{FW1:} \qquad \widehat{RV}_{t+1} = 0.25\widehat{RV}_{t+1}^{(0.25)} + 0.50\widehat{RV}_{t+1}^{(0.50)} + 0.25\widehat{RV}_{t+1}^{(0.75)} \tag{6}$$

$$\text{FW2:} \qquad \widehat{RV}_{t+1} = 0.30\widehat{RV}_{t+1}^{(1/3)} + 0.40\widehat{RV}_{t+1}^{(0.50)} + 0.30\widehat{RV}_{t+1}^{(2/3)}. \tag{7}$$

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

$$\text{FW3:} \; \widehat{RV}_{t+1} = 0.05\widehat{RV}_{t+1}^{(0.10)} + 0.25\widehat{RV}_{t+1}^{(0.25)} + 0.40\widehat{RV}_{t+1}^{(0.50)} + 0.25\widehat{RV}_{t+1}^{(0.75)} + 0.05\widehat{RV}_{t+1}^{(0.90)}. \tag{8}$$

3 Empirical findings

3.1 Data, forecast construction and forecast evaluation

The data we employ comprises the 'long' sample of Christiansen et al. (2012) who provide a detailed description of transformations and datasources.[7] The variable of interest is the realised stock market volatility (RV) of the S&P 500 index. Realised volatility is defined as the log of the square root of the realised variance, computed as the sum of squared intra-period (daily) returns as follows

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

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).

Out-of-sample forecasts are generated by continuously updating the estimation window, i.e. following a recursive (expanding) window. More specifically, 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. The estimation window is continuously updated

[7] The data are available at the Journal of Applied Econometrics Data Archive (http://qed.econ.queensu.ca/jae/datasets/christiansen001/). We thank the authors for making them available to us.

following a recursive 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 $\left\{\widehat{RV}_{i,t+1}\right\}_{t=T_0}^{T-1}$. Our forecasting experiment is conducted on a monthly basis and data span 1926:12 to 2010:12. We consider three out-of-sample forecast evaluation periods corresponding to three initialization periods; namely January 1937, January 1957, January 1977. We use the five years (60 months) before the start of the out-of-sample evaluation period as the initial holdout out-of-sample period, required for the DMSFE/ DALFE and (AL)Cluster forecast combination schemes.

Following Christiansen et al. (2012), we rely on a comprehensive set of 13 macroeconomic and financial predictive variables. Some of these variables overlap with the predictive variables used in the comprehensive study on stock return predictability by Goyal and Welch (2008) as these are motivated via the risk premium channel (Mele, 2007). 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 (e.g. Campbell and Shiller, 1988; Goyal and Welch, 2008; Rapach et al. 2010). To capture the leverage effect (Black, 1976; Nelson, 1991; Glosten et al., 1993) suggesting that negative returns are associated with higher subsequent volatility, we also 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 as analyzed in Nagel (2012). 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 (TBL), the interest rate on a three-month Treasury bill (Ang and Bekaert, 2007); 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 (Campbell and Shiller, 1991); 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 (Merton, 1974) 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).

Since volatility is persistent, the natural benchmark forecasting model is the $AR(1)$ model. According to this model, 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 $R^2_{OS} = 1 - \frac{MSFE_i}{MSFE_{AR}}$, where $MSFE_i$ is the Mean Square Forecast Error associated with each of our competing models and specifications and $MSFE_{AR}$ is the respective value for the $AR(1)$ model, both computed over the out-of-sample period. Positive values are associated with superior forecasting ability of our proposed model/specification. Given that point estimates of the R^2_{OS} are sample dependent, we need to evaluate the statistical significance of our forecasts. To this end, we employ the Clark and West (2007) (CW) approximate normal test to compare our models/ specifications.[8]

The following subsections present an illustration of our proposed complete subset quantile regression approach to realised volatility forecasting. The aim of our analysis is to assess the predictive ability of the proposed forecasting approaches and to compare their performance against that of alternative approaches used in the literature.

3.2 Performance of Complete Subset Regression Models

First, we discuss the out-of-sample performance of the forecasts obtained by single-variable autoregressive models. Table 1 presents the R^2_{OS} statistics of all single-variable models relative to the $AR(1)$ benchmark model for the three out-of-sample periods considered (columns 2 - 4). Positive values of R^2_{OS} indicate superior forecasting performance of the predictive models with respect to the $AR(1)$ forecast. The statistical significance of the corresponding forecasts is assessed by using the Clark and West (2007) MSFE-adjusted statistic. Our findings suggest that only three variables, namely MKT, DEF and STR can consistently outperform the predictions of the $AR(1)$ model. In the recent out-of-sample period, this set is enriched with the EP variable. Both DEF and MKT are associated with the leverage effect (Black, 1976; Nelson, 1991; Glosten et al. 1993). Specifically, as firm leverage and credit risk increases, default spreads also increase and signal an increase in future volatility. In a similar mode, bear stock markets (low past returns) precipitate higher subsequent stock market volatility. With respect to the short-term reversal factor, its forecasting ability is related to market illiquidity. Nagel (2012) found that STR is related to market stress and the supply of liquidity. The predictive ability of the earnings to price ratio appears only in the recent out-of-sample period and is associated with time-varying risk premia (Mele, 2007). More importantly, these economic variables coincide with the ones identified as useful predictors by Christiansen et al. (2012) via their BMA approach.

[TABLE 1 AROUND HERE]

Given that only three to four out of thirteen variables improve realised volatility forecasts, we turn our attention to the performance of subset linear regressions under various combination schemes. Table 2 (Panels A-C) reports the related R^2_{OS} values (all of which are significant) for the three evaluation periods considered. The first line in each panel reports our findings associated with the approach followed by Rapach et al (2010), i.e. a variety of combinations of single-variable models ($k = 1$). Interestingly, our results corroborate the existing literature on the increased benefits of forecast combinations. Irrespective of the method employed, R^2_{OS} values are positive and significant. Simply averaging forecasts (Mean combination method) can lead to R^2_{OS} ranging from 1.34% to 1.01% for the January 1937 initialisation period to the January 1977 one, respectively. Marginal benefits appear when employing the DMSFE combination technique that forms weights based on the historical performance of the individual models. On the other hand, superior forecasting performance is associated with the cluster

[8] A brief description of the Clark and West (2007) test is given in Appendix B.

combining scheme. Classifying predictors in 'good' and 'bad' can lead to an R^2_{OS} value of 2.47% for the longest evaluation period. Increasing the clusters to four can lead to an R^2_{OS} value of 5.02%, which is quite impressive given the relative ease this method can be applied.[9] Ex post this is expected as forming clusters of the three bext performing predictors (CL(4)) is optimal as our single-variable analysis showed (Table 1).

Next, we focus on forecasts generated by simply averaging the forecasts (Mean combination method) produced by subset linear regressions for various values of k (column 2, Table 2). This experiment coincides with the framework of Elliott et al. (2013) and suggests that as we increase the number of subsets (k) the subset linear regression with $k \geq 2$ generates larger R^2_{OS} value than the case of $k = 1$.In more detail, for the January 1937 initialisation out of sample period (Panel A), R^2_{OS} is maximised at $k = 10$ reaching the value of 7.02% suggesting that averaging over all 10-variable models is optimal. Increasing k leads to a deterioration in forecasting performance with a value of 4.86% for the Kitchen Sink model ($k = 13$). This behaviour is markedly different from Elliott et al. (2013) and Meligkotsidou et al. (2014b) who found that subset regression forecasts for the equity premium with $k \leq 6$ produce positive R^2_{OS} values, while the out-of-sample forecasting ability of subsets deteriorates significantly for $k \geq 7$. Similar findings pertain for the other evaluation periods. Specifically, for the January 1957 evaluation period (Panel B), averaging over 10-variable models yields the maximum R^2_{OS} of 4.53%, while for the most recent evaluation sample (Panel C), the Kitchen Sink model that includes all predictors in one regression appears optimal as it is assocated with an R^2_{OS} value of 6.13%. For this case, the values of R^2_{OS} are an increasing function of k.

[TABLE 2 AROUND HERE]

Finally, we focus on alternative (to the Mean) combination methods such as the Median, Trimmed Mean, DMSFE and the Cluster combining schemes within the subset linear regression approach. Overall and similar to the case of $k = 1$, the best performing combining schemes are the Cluster ones. The largest R^2_{OS} value of 8.82% occurs for $k = 7$ and for the longest evaluation period under the Cluster(5) combining scheme. The DMSFE methods perform satisfactorily and reach an R^2_{OS} of 7.57% at an increased value of k equal to 10. With respect to the Trimmed mean and Median combining schemes, we should note that they cannot outperform the simple mean combination scheme. Turning to the most recent evaluation samples (Panels B-C), our results are qualitatively similar to the ones for the longest evaluation period. While the ranking of our combining methods remains unchanged, the forecasting ability of all the methods considered is lower. For example, the best performing cluster method (Cluster(4)) is associated with R^2_{OS} values of 6.79% and 6.63% for the two evaluation periods, respectively. Similar to our previous findings, the shortest evaluation period requires more heavily parameterised models as the optimal k of 11 or 12 suggests. The only exception is CL(5) for which $k = 8$ emerges as optimal.

[9] Please note that for single-variable models (k =1), CL(4) and CL(5) coincide as the algorithm rounds up the number of models included in each cluster. In our case, for k =1, both CL(4) and CL(5) methods identify the 3 best performing variables over the holdout out-of-sample period. For $k \geq 2$,these methods are distinct as the model space increases.

Overall, our results of Table 2 in general indicate that employing alternative weighting schemes under the subset regression approach can lead to improved forecasting performance relative to simple averaging of single-variable model forecasts.

3.3 Performance of Complete Subset Quantile Regression Models

In this subsection, we evaluate the forecasting performance of the proposed subset quantile regression models. Employing the Quantile Forecast Combination (QFC) approach, the quantile forecasts, obtained from different k-variate predictive model specifications, are first combined employing several combination schemes. These schemes are either simple methods such as the Mean, Median and Trimmed Mean, or are based on the asymmetric linear loss function such as the DALFE and the AL Cluster methods. Then, robust point forecasts are obtained by synthesizing the quantile forecasts (corresponding to different parts of the realised volatility ditribution) employing the fixed weighting schemes (FW1-FW3) given by equations (6) - (8).

Table 3 (Panels A - C) reports the out-of-sample performance of the subset quantile regression forecasts for the long evaluation period for the three weighting schemes (FW1-FW3), respectively. Our results indicate that high positive R^2_{OS} values are obtained by using $k = 7, 8$, or 9 subsets for all weighting schemes. Overall, our quantile forecasts are superior to the linear ones (Table 2) irrespective of the combining or weighting scheme employed. For example, averaging across $k-$variate quantile models generates R^2_{OS} values that are greater than 7.5% as opposed to 7% for the $k-$variate linear models. The ranking of the combining methods remains roughly unchanged with the AL_Cluster(5) combining method achieving superior performance. The associated R^2_{OS} values exceed 9% for the FW1 and FW2 weighting methods. It is also worth noting that this perfromance is achieved for the lowest subset, $k = 7$, in the case of the FW2 weighting method. This weighting scheme is the one that emerges as superior judging by the higher R^2_{OS} values in the majority of the combining schemes. On the other hand, the FW3 method that utilises a finer grid of quantiles and puts weights on the extreme 10% and 90% quantiles is the one lagging in performance, by a small margin though. This may be due to the relative imprecision that extreme quantiles are estimated and/or the log transformation of the realised volatility that is closer to being normally distributed (compared to the raw untransformed series).[10]

[TABLE 3 AROUND HERE]

Table 4 reports our findings for the forecast evaluation period initialised at January 1957. Overall, while our findings are similar to the ones for the long evaaluation period, the predictive ability of the quantile forecast combinations is lower by approximately 1.5%. However, compared to the same evaluation period and the subset regression models, we observe an increase in R^2_{OS} values by about 1%. As previously, among the various combination methods, the cluster schemes rank first followed by the DALFE ones. In particular, the CL(5) combination scheme ranks first,

[10] The latter issue constitutes an issue for future research.

since, for the best $k = 8$ subset, generates the highest R^2_{OS} values ranging from 7.35% for FW3 to 7.82% for the FW2 scheme.

[TABLE 4 AROUND HERE]

Finally, Table 5 (Panels A-C) presents the results obtained by our quantile approach for the more recent evaluation period. Similar to the subset linear regression case, this out-of-sample period favours heavily parameterised models. This is especially true when the simple combination schemes or the kitcehn Sink model are employed. Superior performance is achieved by the FW1 method and the CL(2) combining scheme for which the R^2_{OS} value reached 5.94% ($k = 12$). Our findings suggest that increasing the number of clusters leads to a decrease in k. For example, the best performance for the FW1 method and the CL(3), CL(4) and CL(5) method is achieved for $k = 10, k = 9$ and $k = 9$, respectively.

[TABLE 5 AROUND HERE]

4 Real time Selection of k

Our empirical findings (Section 3) suggest that the predictive performance of our subset quantile regression approach depends on the choice of the value of k. Therefore, it is important to develop a real time algorithm of selecting k recursively, based on the past history of volatility and predictive variables, in order to produce 'optimal forecasts'. Since our proposed methodology involves forecasting an array of quantiles, it is quite interesting to examine whether the selected value of k varies across quantiles of volatility, thus revealing a further source of information that can be exploited within our proposed framework. Our 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.

4.1 Algorithm for selecting k

In this subsection we propose a likelihood-based (Bayesian) method for selecting k in real time. The experiment we conduct is naturally designed in the context of our QFC forecasting approach. At each time point in the out-of-sample period, indexed by $t + 1$, we compute the posterior probabilities of all values of k ($k \in \{1, 2, ..., K\}$), based on the data up to time t, for a set of quantiles. Then, for each quantile, τ, we select the most probable value of k and produce a quantile forecast at time $t + 1$, $\widehat{RV}_{t+1}(\tau)$, based on the selected complete subset. These quantile forecasts are then combined according to the fixed weighting and time-varying weighting schemes of Section 3 in order to produce 'optimal' QFC forecasts 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 of particular interest to quantify the uncertainty about the complete subset that will be used for predicting each volatility quantile.

Therefore, in a Bayesian context, the random quantities of interest are the model specification, representing the set of predictors included in the jth model and denoted by m_j, $j = 1, ..., M$, $M = \sum_{i=1}^{K} n_{i,K}$, the value of k and the totality of the model parameters associated with the τth quantile regression, denoted by $\theta^{(\tau)}$. After specifying appropriate prior distributions for these quantities, $P(m_j)$, $P(k|m_j)$ and $f(\theta^{(\tau)}|m_j, k)$, their joint posterior distribution is given by

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

where $L^{(\tau)}(RV_{1:t}|m_j, k, \theta^{(\tau)})$ is the likelihood of the data up to time t under the τth quantile regression (Equation 5), based on the asymmetric Laplace density (4). Dependence on the set of predictors has been suppressed for simplicity. Then, the marginal posterior distribution of k, under the τth quantile regression, is obtained as

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

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

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

where $\widehat{\theta}^{(\tau)}$ denotes the ML estimate of $\theta^{(\tau)}$, obtained as discussed in Subsection 2.2. Alternatively, the marginal likelihood of quantile regression models can be estimated by Laplace approximation (see Meligkotsidou et al., 2009).

The prior specification we consider is the following. The prior probability of the jth 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 $2/3$, thus reflecting the need of models with many predictors in order 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, under the τth quantile regression is given by

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

Below we present and discuss the results of our likelihood-based approach to selecting k for the fixed and time-varying weighting schemes of Section 3 and the respective combining methods (see Appendix A.2).

4.2 Algorithm Performance

Table 6 reports the out-of-sample performance under the prior specification considered (i.e. π=2/3) of the 'optimal' CSR forecasts (Panel A) and the 'optimal' QFC forecasts based on fixed weighting schemes (FW1-FW3) (Panel B) for the three out-of-sample evaluation periods. The results of Table 6 reveal that our likelihood-based approach to selecting k in real time is extremely successful, since the values of R^2_{OS} obtained under all weighting schemes and for all combining methods are very high, especially for the evaluation periods starting January 1937 and January 1957. However for the most recent evaluation period, starting January 1977, the performance of QFC is inferior to the ones of CSR. Regarding the first evaluation period, the largest R^2_{OS} values are obtained for the Cluster combining methods (CL(5)), being in all cases higher than 8.5%, with the highest value equal to 9.02% for the FW2 scheme. Regarding the first dataset, the largest R^2_{OS} values are obtained again for the Cluster combining methods (CL(5)), being in all cases higher than 6.39%, with the highest value being equal to 7.51%, for the FW2 scheme. For the most recent evaluation period, we observe that the largest R^2_{OS} values are obtained for the Cluster combining method (CL(5)) and the CSR approach (6.07%) and second best is the Cluster combining method (CL(5)) under the FW3 scheme (5.65%).

[TABLE 6 AROUND HERE]

It is interesting to note that the results of the recursive k-selection exercise are quite robust across the combining methods considered, as Cluster combining methods are the best, and then we have DM, Mean, Tr. Mean and Median method. Moreover, it appears that all the quantile schemes outperform the linear ones in 2 of the 3 evaluation periods and the FW2 scheme outperforms the other two FW schemes in two of the three out-of-sample periods.

In conclusion, let us note that the findings of our recursive experiment are very encouraging, since they show that the proposed approach of selecting k in real time, based only on the past history of the data, produces particularly well-performing forecasts and that these results are very robust to the choice of weighting scheme and combining method.

5 Conclusions

In this study we propose a quantile forecast combination approach to realised volatility prediction. The aim of our analysis is to construct volatility forecasts, which take into account the benefits emerging from the subset framework, the quantile regression framework and the information given by the potential predictors.

The quantile predictive approach proposed in this paper is based on the combination of the quantile forecasts across complete subsets of model specifications that keep the number of predictors, k, fixed. Forecast combination is based on several well-established combining methods, while weighted averages of a set of combined quantile forecasts produce robust and accurate forecasts of US equity realised volatility. We employ the likelihood-based method developed by Meligkotsidou et al. (2014b) to select the value of k recursively. This algorithm is able to identify the best subset for predicting each quantile of the realised volatility distribution in real time, based only on the past history of the data. Then, these ‘optimal’ quantile forecasts are combined to produce robust volatility forecasts.

The results of our study are very promising. Our findings suggest that our quantile forecast combinations produced in the complete subset quantile regression framework achieves superior predictive performance relative to the autoregressive benchmark, the combination approach, and the subset linear regression approach. specifically, our approach can lead to sizable benefits that exceed 9% in terms of R^2_{OS}. Extending our framework to allow for higher autoregressive lag orders, rolling window forecasts and a large dimensional set of predictors is a promising route for future research.

Appendix A. Forecast Combination Schemes

Combining individual models' forecasts can reduce uncertainty risk associated with a single predictive model and display superior predictive ability (Bates and Granger, 1969; Hendry and Clements, 2004). In Appendix A.1, we briefly discuss existing combination schemes that are appropriate for combining subset mean regression forecasts, while in Appendix A.2 we present the respective combining methods that are appropriate for producing combined subset quantile forecasts (QFC approach).

A.1. Combination Methods for Mean forecasting

The combination forecasts of RV_{t+1}, denoted by $\widehat{RV}^{(C)}_{t+1}$, are weighted averages of the $k-$variate predictor individual forecasts within each subset, $\widehat{RV}_{i,t+1}$, $i = 1, \ldots, n_{k,K}$, of the form $\widehat{RV}^{(C)}_{t+1} = \sum_{i=1}^{n_{k,K}} w^{(C)}_{i,t} \widehat{RV}_{i,t+1}$, where $w^{(C)}_{i,t}, i = 1, ..., n_{k,K}$, are the *a priori* combining weights at time t for each specific subset, $k, k \leq K$.

The simplest combining scheme is the one that attaches equal weights to all k-variate models for a specific k, i.e. $w^{(C)}_{i,t} = 1/n_{k,K}$, for $i = 1, ..., n_{k,K}$, called the Mean combining scheme. The next schemes we employ are the Trimmed Mean and Median ones. The Trimmed Mean combination scheme sets $w^{(C)}_{i,t} = 0$ for the smallest and largest forecasts and $w^{(C)}_{i,t} = 1/(n_{k,K} - 2)$ for the remaining ones, while the Median combination scheme employs the median of the $\left\{\widehat{RV}_{i,t+1}\right\}_{i=1}^{n_{k,K}}$ forecasts.

The methods we describe below require a holdout out-of-sample period during which the combining weights are estimated. To this end, the first P_0 out-of-sample observations are employed as the initial holdout period over which we construct combination forecasts and the remaining $T - (T_0 + P_0) = P - P_0$ forecasts are available for evaluation. The second class of combining methods we consider, proposed by Stock and Watson (2004), suggests forming weights based on the historical performance of the individual models over the holdout out-of-sample period. Specifically, their Discount Mean Squared Forecast Error (DMSFE) combining method suggests forming weights as follows

$$w^{(C)}_{i,t} = m^{-1}_{i,t} / \sum_{j=1}^{n_{k,K}} m^{-1}_{j,t}, \; m_{i,t} = \sum_{s=T_0}^{t-1} \psi^{t-1-s} (RV_{s+1} - \widehat{RV}_{i,s+1})^2,$$

where ψ is a discount factor which attaches more weight on the recent forecasting accuracy of the individual models in the cases where $\psi \in (0, 1)$. The values of ψ we consider are 1.0, 0.9 and 0.5. When ψ equals one, there is no discounting and the combination scheme coincides with the optimal combination forecast of Bates and Granger (1969) in the case of uncorrelated forecasts.

Finally, the third class of combining methods, namely the Cluster combining method, was introduced by Aiolfi and Timmermann (2006). In order to create the Cluster combining forecasts, we form L clusters of forecasts of equal size based on the MSFE performance. Each combination forecast is the average of the k-variate model forecasts in the best performing cluster. This procedure begins over the initial holdout out-of-sample period and goes through the end of the available out-of-sample period using a rolling window. In our analysis, we consider

$L = 2, 3, 4, 5$.

A.2. Combination Methods for Quantile Forecasting

The DMSFE, Cluster and Principal Components combining methods have been designed in the framework of standard linear regression, in order to construct forecasts that exploit the entire set of predictive variables. The combining weights, $w_{i,t}^{(C)}$, are computed based on the MSFE, that is on a quadratic loss function that measures how close to the realized excess returns the individual forecasts are. These methods are appropriate within the linear regression (subset) framework. However, these combining schemes are not appropriate for combining predictor information within the QFC approach since variable information is now combined in the context of forecasting several quantiles of returns rather than producing point forecasts. In this case, the MSFE is no longer suitable for measuring the performance of the produced forecasts and has to be replaced by a metric based on the asymmetric linear loss function.

Below we describe how we modify the existing combining methods in order to produce quantile forecasts that exploit variable information. The combined quantile forecasts, $\widehat{RV}_{t+1}^{(C)}(\tau)$, are weighted averages of the form $\widehat{RV}_{t+1}^{(C)}(\tau) = \sum_{i=1}^{n_{k,K}} w_{i,t}^{(C)} \widehat{RV}_{i,t+1}(\tau)$, where the combining weights, $w_{i,t}^{(C)}$, have to be computed based on the check function (3).

First, we introduce the Discount Asymmetric Loss Forecast Error (DALFE) combining method which suggests forming weights as follows

$$w_{i,t}^{(C)} = m_{i,t}^{-1} / \sum_{j=1}^{n_{k,K}} m_{j,t}^{-1}, \; m_{i,t} = \sum_{s=T_0}^{t-1} \psi^{t-1-s} \rho_\tau(RV_{s+1} - \widehat{RV}_{i,s+1}(\tau)),$$

where $\psi \in (0, 1)$ is a discount factor. The combining weights are computed based on the historical performance of the individual quantile regression models over the holdout out-of-sample period and ψ is set equal to 0.5, 0.9 and 1.

We also modify the Cluster combining method by forming L clusters of forecasts based on their performance as measured by the asymmetric loss forecast error. The Asymmetric Loss Cluster (AL Cluster) combination forecast is the average of the individual quantile forecasts in the best performing cluster which contains the forecasts with the lower expected asymmetric loss values. We consider forming $L = 2, 3, 4, 5$ clusters.

Appendix B. The Clark and West (2007) test of equal forecasting ability.

Clark and West (2007) develop an adjusted version of the Diebold and Mariano (1995) and West (1996) statistic, namely the MSFE-adjusted statistic, which in conjunction with the standard normal distribution generates asymptotically valid inferences when comparing forecasts from nested linear models. Suppose that we want to evaluate the forecasts of a parsimonious model A relative to a larger model B. Under the null hypothesis of equal MSFE, model B should generate larger MSFE than model A, due to the estimation of additional parameters that introduces noise into the forecasts while these do not improve predictions. A smaller MSFE should not be considered as evidence of superiority of model A over B. In this respect, the testing procedure of Clark and West (2007) aims at correcting for the inflation in the MSFE of the larger model before evaluating the relative forecasting accuracy of the two models. Let $\widehat{RV}_{A,t+1}$ and $\widehat{RV}_{B,t+1}$ denote the one-step ahead forecasts for r_t obtained from models A and B respectively. We define

$$f_{t+1} = (RV_{t+1} - \widehat{RV}_{A,t+1})^2 - [(RV_{t+1} - \widehat{RV}_{B,t+1})^2 - (\widehat{RV}_{A,t+1} - \widehat{RV}_{B,t+1})^2]$$

The test statistic of Clark and West, denoted as $MSFE - adjusted$, is given by the standard $t - statistic$ of the regression of $\{f_{s+1}\}_{s=T_0+P_0}^{T-1}$on a constant. Given that under the alternative hypothesis of the test, model B has lower MSFE than model A, this is an one-sided test. Clark and West (2007) recommend using 1.282, 1.645 and 2.326 as critical values for a 0.10, 0.05 and 0.01 test, respectively. Extensive simulations performed by them, which consider a variety of different processes and settings show that the aforementioned critical values provide reliable results.

References

[1] Aiolfi, M., and A. Timmermann.(2006) "Persistence in Forecasting Performance and Conditional Combination Strategies." Journal of Econometrics, 135, 31–53.

[2] Andersen TG, Bollerslev T and FX Diebold (2003). "Modeling and forecasting realized volatility." Econometrica, 71, 529–626.

[3] Andersen TG, Bollerslev T, Christoffersen P and FX Diebold FX. (2006). Volatility and correlation forecasting. In Handbook of Economic Forecasting, Elliot G, Granger CW, Timmermann A (eds). Elsevier: Amsterdam; 777–878.

[4] Ang, A., and G. Bekaert. (2007) "Return Predictability: Is It There?." Review of Financial Studies, 20, 651–707.

[5] Baker, M., and J. Wurgler. (2000) "The Equity Share in New Issues and Aggregate Stock Returns." Journal of Finance, 55, 2219-2257.

[6] Bassett, W. G., and H-L. Chen. (2001) "Portfolio style: Return-based attribution using quantile regression." Empirical Economics, 26, 293-305.

[7] Bates, J. M., and C.W.J. Granger. (1969) "The combination of forecasts." Operational Research Quarterly, 20, 451–468.

[8] Baur, D. G.; T. Dimpfl; and R. Jung. (2012) "Stock return autocorrelations revisited: A quantile regression approach." Journal of Empirical Finance, 19(2), 254-265.

[9] Black F. (1976). "Studies in stock price volatility changes." In Proceedings of the American Statistical Association, Business and Economic Statistics Section; 177–181.

[10] Brandt MW and Q. Kang (2002). "On the relationship between the conditional mean and volatility of stock returns: a latent VAR approach." Wharton School, Philadelphia, PA.

[11] Breiman, L. (1996)."Bagging predictors" Machine Learning, 36, 105–139.

[12] Buchinsky, M. (1994) "Changes in U.S. Wage Structure 1963-1987: An application of Quantile Regression." Econometrica, 62, 405-458.

[13] Buchinsky, M. (1995) "Quantile Regression Box-Cox Transformation model, and the U.S. wage structure, 1963-1987." Journal of Econometrics, 65, 109-154.

[14] Buchinsky, M. (1998), "Recent Advances in Quantile Regression Models: A Practical Guidline for Empirical Research." Journal of Human Resources, 33, 88-126.

[15] Campbell, J. Y. (1987) "Stock Returns and the Term Structure." Journal of Financial Economics, 18, 373–99.

[16] Campbell, J. Y., and J. H. Cochrane. (1999) "By force of habit: A consumption-based explanation of aggregate stock market behavior." The Journal of Political Economy, 107, 205–251.

[17] Campbell JY, Shiller RJ. (1988). "The dividend–price ratio and expectations of future dividends and discount factors."Review of Financial Studies 1(3), 195–228.

[18] Campbell JY, Shiller RJ. (1991). "Yield spreads and interest rate movements: a bird's eye view." Review of Economic Studies 58, 495–514.

[19] Campbell, J. Y., and R. J. Shiller. (1998) "Valuation Ratios and the Long-Run Stock Market Outlook." Journal of Portfolio Management, 24, 11–26.

[20] Campbell, J. Y., and S. B. Thompson.(2008) "Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?" Review of Financial Studies, 21, 1509–31.

[21] Campbell, J.Y., and L. Viceira. (2002) "Strategic Asset Allocation." Oxford University Press, Oxford.

[22] Campbell, J. Y., and T. Vuolteenaho. (2004) "Inflation Illusion and Stock Prices." American Economic Review, 94, 19–23.

[23] Cenesizoglu, T., and A. Timmermann. (2012) "Do Return Prediction Models Add Economic Value?" Journal of Banking and Finance, 36, 2974-2987.

[24] Chen, Q. and Y. Hong. (2010) "Predictability of equity returns over different time horizons: a nonparametric approach." Manuscript, Cornell University.

[25] 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.

[26] Chuang, C.-C.; C.-M. Kuan; and H.-Y. Lin. (2009) "Causality in quantiles and dynamic stock return-volume relations." Journal of Banking and Finance, 33(7), 1351-60.

[27] Clark, T. E., and K. D. West.(2007) "Approximately Normal Tests for Equal Predictive Accuracy in Nested Models." Journal of Econometrics, 138, 291–311.

[28] Dangl, T., and M. Halling (2012), Predictive regressions with time-varying coefficients." Journal of Financial Economics, 106 (1), 157-181.

[29] Della Corte, P., L. Sarno, and I. Tsiakas (2009). "An Economic Evaluation of Empirical Exchange Rate Models." Review of Financial Studies, 22 (9) 3491-3530.

[30] Della Corte, P., L. Sarno, and G. Valente (2010) "A Century of Equity Premium Predictability and Consumption-Wealth Ratios: An International Perspective." Journal of Empirical Finance, 17, 313-331.

[31] Diebold, F. X., and R. S. Mariano.(1995), "Comparing Predictive Accuracy." Journal of Business and Economic Statistics, 13, 253–63.

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

[33] Elliott, G., A. Gargano, and A. Timmermann. (2013) "Complete Subset Regressions." Journal of Econometrics, 177(2), 357-373.

[34] Fama, E. F., and K. R. French. (1988) "Dividend Yields and Expected Stock Returns." Journal of Financial Economics, 22, 3–25.

[35] Fama, E. F., and K. R. French. (1989) "Business Conditions and Expected Returns on Stocks and Bonds." Journal of Financial Economics, 25, 23–49.

[36] Fama EF, French KR. (1993). "Common risk factors in the returns on stocks and bonds." Journal of Financial Economics, 33, 3–56.

[37] Fama, E. F., and G. W. Schwert. (1977) "Asset Returns and Inflation." Journal of Financial Economics, 5, 115–46.

[38] Ferreira, M.I., and P. Santa-Clara. (2011) "Forecasting stock market returns: the sum of the parts is more than the whole." Journal of Financial Economics, 100, 514–537.

[39] Fleming, J.; C. Kirby; and B. Ostdiek.(2001), "The Economic Value of Volatility Timing." Journal of Finance, 56, 329-352.

[40] Gastwirth, J.L. (1966), "On robust procedures." Journal of the American Statistical Association, 61, 929-948.

[41] Glosten, L. R., R. Jagannathan, 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.

[42] Goyal, A. and I. Welch.(2008) A Comprehensive Look at the Empirical Performance of Equity Premium Prediction." Review of Financial Studies, 21, 1455–508.

[43] Granger, C. W. J., and R. Ramanathan. (1984) "Improved methods of combining forecasts." Journal of Forecasting, 3, 197-204.

[44] Guidolin, M.; S. Hyde; D. McMillan and S. Ono. (2009) "Non-linear predictability in stock and bond returns: When and where is it exploitable?" International Journal of Forecasting, 25, 373–399.

[45] Guidolin, M. and A. Timmermann. (2009) "Forecasts of US short-term interest rates: A flexible forecast combination approach." Journal of Econometrics, 150, 297–311.

[46] Hamilton JD and G. Lin (1996). "Stock market volatility and the business cycle." Journal of Applied Econometrics, 11, 573–593.

[47] Hansen, B. (2008) "Least-squares forecast averaging." Journal of Econometrics, 146, 342-350.

[48] Hendry, D. F., and M. P. Clements. (2004) "Pooling of Forecasts." Econometrics Journal, 7, 1–31.

[49] Henkel, S.J.; J.S. Martin; and F. Nadari. (2011), "Time-varying short-horizon predictability." Journal of Financial Economics, 99, 560–580.

[50] Hoerl, A.E., Kennard, R.W. (1970) "Ridge regression: biased estimation for nonorthogonal problems" Technometrics 12, 55–67.

[51] Hsiao, C., and S. K. Wan. (2014) "Is there an optimal forecast combination?." Journal of Econometrics,178(2), 294-309.

[52] Judge, G. G, R. C. Hill, W. E. Griffiths, H. Lutkepohl; and T.-C. Lee.(1988) "Introduction to the Theory and Practice of Econometrics." New York, Wiley.

[53] Koenker, R., and G. Bassett (1978). "Regression Quantiles." Econometrica, 46, 33-50.

[54] Kothari, S., and J. Shanken. (1997) "Book-to-Market, Dividend Yield, and Expected Market Returns: A Time-Series Analysis." Journal of Financial Economics, 44, 169–203.

[55] Lettau, M., and S. C. Ludvigson. (2001) "Consumption, Aggregate Wealth, and Expected Stock Returns." Journal of Finance, 56, 815–49.

[56] Ludvigson, S. C., and S. Ng. (2007) "The empirical risk-return relation: a factor analysis approach." Journal of Financial Economics, 83, 171–222.

[57] Marquering, W., and M. Verbeek. (2004) "The Economic Value of Predicting Stock Index Returns and Volatility." Journal of Financial and Quantitative Analysis, 39, 407–29.

[58] Mele A. (2007). "Asymmetric stock market volatility and the cyclical behavior of expected returns." Journal of Financial Economics, 86, 446–478.

[59] Meligkotsidou, L., I. D. Vrontos, and S. D. Vrontos. (2009) "Quantile Regression Analysis of Hedge Fund Strategies." Journal of Empirical Finance, 16, 264-279.

[60] 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.

[61] 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/.

[62] Menzly, L.; T. Santos; and P. Veronesi. (2004), "Understanding Predictability." Journal of Political Economy, 112, 1-47.

[63] Merton RC. (1974.) On the pricing of corporate debt: the risk structure of interest rates. Journal of Finance, 29, 449–470.

[64] Nagel S. (2012). "Evaporating liquidity." Review of Financial Studies, 25(7), 2005–2038.

[65] Nelson DB. (1991). "Conditional heteroskedasticity in asset returns." Econometrica, 59, 347–370.

[66] Neely, C. J., D. E. Rapach, J. Tu, and G. Zhou. (2014) "Forecasting the equity risk premium: the role of technical indicators." Management Science 60(7), 1772-91.

[67] Paye, B. S. (2012). "'Deja Vol': Predictive regressions for aggregate stock market volatility using macroeconomic variables. Journal of Financial Economics, 106, 527-546.

[68] Pearson, E. S. and J. W. Tukey (1965) "Approximate Means and Standard Deviations Based on Distances between Percentage Points of Frequency Curves." Biometrika 52, 533-546.

[69] Perez-Quiros G and A. Timmermann (2000). "Firm size and cyclical variations in stock returns." Journal of Finance, 55, 1229–1262.

[70] Rapach, D., J. Strauss, 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.

[71] Rapach, D., and G. Zhou. (2012) "Forecasting stock returns." In preparation for G. Elliott and A. Timmermann, Eds., Handbook of Economic Forecasting, Vol. 2.

[72] Raftery, A., Madigan, D., Hoeting, J. (1997) "Bayesian Model Averaging for linear regression models". Journal of the American Statistical Association 97, 179–191.

[73] Schwert, G. (1989) Why does stock market volatility change over time? Journal of Finance, 44, 1115-1153.

[74] Spiegel, M.(2008) Forecasting the equity premium: Where we stand today. The Review of Financial Studies, 21, 1453–1454.

[75] Stock, J. H., and M. W. Watson (2004) Combination Forecasts of Output Growth in a Seven-Country Data Set. Journal of Forecasting, 23, 405–30.

[76] Tibshirani, R. (1996) "Regression shrinkage and selection via the Lasso." Journal of the Royal Statistical Society. Series B 58, 267–288.

[77] Timmermann, A. (2006) "Forecast combinations." In Handbook of Economic Forecasting, Vol. I, G. Elliott, C. W. J. Granger and A. Timmermann, eds. Amsterdam, Elsevier.

[78] Tukey, J. W (1977) "Explanatory Data Analysis." Addison-Wesley, Reading, MA.

[79] Wachter, J., and M. Warusawitharana. (2009) "Predictable returns and asset allocation: Should a skeptical investor time the market?" Journal of Econometrics, 148 (2), 162-178.

[80] West, K. D. (1996) "Asymptotic Inference About Predictive Ability. Econometrica", 64, 1067–84.

[81] West, K., H. Edison and D. Cho (1993) "A Utility-based Comparison of Some Models of Exchange Rate Volatility." Journal of International Economics, 35, 23-46.

[82] Whitelaw R. (1994). "Time variations and covariations in the expectation and volatility of stock returns." Journal of Finance, 49, 515–541.

[83] Yu, K., and J. Zhang (2005). "A Three-Parameter Asymmetric Laplace Distribution and Its Extension." Communications in Statistics - Theory and Methods, 34, 1867-1879.

[84] Yu, K., Z. Lu, and J. Stander (2003), "Quantile regression: applications and current research areas." The Statistician, 52, 331-350.

[85] Yu, K., and R. A. Moyeed (2001), "Bayesian quantile regression." Statistics and Probability Letters, 54, 437-447.

Table 1. Performance of single-variable autoregressive models

	January 1937	January 1957	January 1977
AR MSFE	0.1079	0.0998	0.0961
Predictor			
DP	-0.77	-0.36	-2.99
EP	-0.15	-2.27	**0.69**
MKT	**1.37**	**1.79**	**1.54**
DEF	**5.20**	**1.88**	**1.67**
HML	-0.19	-0.16	-0.15
INF	-0.17	-0.29	-0.04
LTR	-0.46	-0.16	-0.90
RBR	-0.43	-0.58	-0.43
RTB	-0.12	-0.52	0.16
SMB	-0.64	-0.17	0.08
STR	**1.12**	**1.37**	**1.86**
TB	-0.28	-0.25	-0.89
TMS	-0.09	-0.64	-0.99

Notes: The Table reports the out-of-sample R^2 statistic with respect to the $AR(1)$ benchmark model for the out-of-sample period 1977:1-2010:12. Bold indicates significance (5% level) based on the p-value of the Clark and West (2007) out-of-sample MSFE-adjusted statistic (CW_{pv})..

Table 2. Out-of-sample performance of complete subset regression models

Panel A: Start: January 1937										
k	Mean	Median	Tr.Mean	DM(1)	DM(0.9)	DM(0.5)	CL(2)	CL(3)	CL(4)	CL(5)
1	1.34	0.45	0.93	1.41	1.44	1.37	2.47	4.16	5.02	5.02
2	2.60	1.52	2.38	2.71	2.78	2.61	4.62	5.80	6.46	6.93
3	3.74	2.83	3.59	3.88	3.99	3.72	5.99	7.15	7.66	7.88
4	4.72	4.00	4.66	4.86	5.02	4.69	7.01	7.88	8.17	8.30
5	5.51	4.88	5.50	5.63	5.85	5.51	7.73	8.31	8.51	8.56
6	6.10	6.12	6.13	6.20	6.48	6.19	8.19	8.57	8.70	8.75
7	6.53	6.70	6.55	6.60	6.95	6.74	8.47	**8.72**	**8.77**	**8.82**
8	6.81	6.84	6.82	6.87	7.28	7.17	**8.62**	8.68	8.76	8.81
9	6.98	**6.96**	6.96	7.02	7.50	7.47	8.61	8.54	8.63	8.60
10	**7.02**	6.63	**6.98**	**7.04**	**7.56**	**7.57**	8.20	8.34	8.37	8.36
11	6.83	6.27	6.77	6.82	7.34	7.38	7.70	7.77	8.05	8.00
12	6.21	5.18	6.01	6.19	6.58	6.69	6.69	7.00	6.85	6.85
13	4.86									
Panel B: Start: January 1957										
k	Mean	Med.	Tr.Mean	DM(1)	DM(0.9)	DM(0.5)	CL(2)	CL(3)	CL(4)	CL(5)
1	1.05	0.42	0.81	1.09	1.15	1.11	1.87	3.27	3.88	3.88
2	1.87	1.47	1.79	1.94	2.07	2.00	3.43	4.26	4.67	4.96
3	2.52	2.34	2.48	2.61	2.80	2.73	4.22	5.02	5.41	5.57
4	3.02	2.56	3.01	3.11	3.37	3.34	4.83	5.47	5.68	5.70
5	3.39	2.95	3.40	3.49	3.82	3.83	5.31	5.71	5.86	5.92
6	3.69	3.57	3.69	3.78	4.18	4.25	5.71	5.94	6.11	6.21
7	3.94	3.90	3.92	4.04	4.49	4.62	6.04	6.29	6.38	6.46
8	4.18	4.04	4.15	4.27	4.78	4.95	6.37	6.51	6.65	6.74
9	4.40	**4.18**	4.36	4.48	5.05	5.22	**6.59**	6.63	**6.79**	**6.76**
10	**4.53**	3.95	**4.47**	**4.58**	**5.21**	**5.33**	6.45	**6.64**	6.67	6.72
11	4.40	3.79	4.32	4.41	5.05	5.14	6.00	6.40	6.64	6.58
12	3.75	2.53	3.47	3.72	4.24	4.37	4.81	5.42	5.52	5.52
13	2.23									
Panel C: Start: January 1977										
k	Mean	Med.	Tr.Mean	DM(1)	DM(0.9)	DM(0.5)	CL(2)	CL(3)	CL(4)	CL(5)
1	1.01	0.40	0.71	1.01	1.09	1.05	1.29	2.28	2.75	2.75
2	1.95	1.34	1.79	1.95	2.10	1.96	2.66	3.37	3.79	4.22
3	2.79	2.34	2.66	2.79	3.01	2.78	3.71	4.61	5.14	5.30
4	3.50	2.76	3.42	3.52	3.79	3.51	4.73	5.46	5.64	5.63
5	4.07	3.51	4.05	4.11	4.42	4.15	5.56	5.93	5.99	5.99
6	4.51	4.36	4.52	4.56	4.90	4.67	6.09	6.22	6.23	6.35
7	4.84	4.96	4.86	4.91	5.25	5.07	6.44	6.49	6.43	6.47
8	5.10	5.36	5.11	5.17	5.50	5.35	6.62	6.51	6.59	**6.58**
9	5.33	5.72	5.33	5.39	5.70	5.54	6.63	6.53	6.55	6.44
10	5.57	5.96	5.58	5.63	5.89	5.67	6.52	6.52	6.41	6.33
11	5.83	6.06	5.91	5.88	6.09	5.80	6.45	**6.53**	**6.63**	6.48
12	6.08	**6.10**	**6.24**	**6.10**	**6.24**	**5.99**	**6.66**	6.23	6.19	6.19
13	**6.13**									

Notes: The Table reports the out-of-sample R^2 statistic with respect to the $AR(1)$ benchmark model for the out-of-sample period 1977:1-2010:12. The reported R^2_{OS} statistics are statistically significant based on the p-value of the Clark and West (2007) out-of-sample MSFE-adjusted statistic (CW_{pv}).

Table 3. Out-of-sample performance of subset quantile regression models (start date: January 1937)

Panel A: FW1 method										
k	Mean	Median	Tr.Mean	DM(1)	DM(0.9)	DM(0.5)	CL(2)	CL(3)	CL(4)	CL(5)
1	2.45	1.65	2.11	2.49	2.50	2.38	3.42	4.76	5.38	5.38
2	3.53	2.70	3.36	3.59	3.62	3.43	5.19	6.16	6.79	7.14
3	4.57	3.79	4.45	4.63	4.68	4.45	6.41	7.29	7.69	7.91
4	5.50	4.93	5.43	5.56	5.63	5.37	7.19	7.88	8.24	8.38
5	6.28	5.90	6.25	6.33	6.42	6.15	7.80	8.30	8.54	8.68
6	6.89	6.99	6.89	6.93	7.05	6.79	8.24	8.63	8.79	8.91
7	7.32	7.48	7.34	7.35	7.50	7.27	8.55	8.85	8.95	9.00
8	7.60	7.63	7.62	7.62	7.79	7.59	8.74	**8.95**	**8.99**	**9.02**
9	**7.70**	**7.73**	**7.71**	**7.71**	**7.90**	**7.74**	**8.74**	8.84	8.87	8.91
10	7.62	7.51	7.62	7.62	7.82	7.70	8.54	8.60	8.57	8.58
11	7.26	7.05	7.23	7.26	7.44	7.36	8.00	8.19	8.20	8.11
12	6.38	5.72	6.22	6.37	6.50	6.48	6.88	7.27	7.51	7.51
13	4.87									
Panel B: FW2 method										
k	Mean	Median	Tr.Mean	DM(1)	DM(0.9)	DM(0.5)	CL(2)	CL(3)	CL(4)	CL(5)
1	2.50	1.65	2.19	2.53	2.55	2.47	3.48	4.64	5.29	5.29
2	3.64	2.86	3.48	3.69	3.73	3.60	5.30	6.11	6.70	6.93
3	4.70	4.03	4.58	4.76	4.82	4.63	6.50	7.28	7.62	7.81
4	5.64	5.16	5.58	5.69	5.77	5.54	7.31	7.91	8.21	8.35
5	6.41	6.16	6.40	6.46	6.56	6.31	7.90	8.34	8.61	8.76
6	7.00	7.10	7.01	7.03	7.16	6.91	8.30	8.68	8.86	8.96
7	7.41	7.49	7.43	7.44	7.58	7.35	8.56	8.89	**8.99**	**9.05**
8	7.67	7.60	7.68	7.68	7.85	7.64	**8.69**	**8.92**	8.98	9.02
9	**7.75**	**7.63**	**7.76**	**7.76**	**7.94**	**7.76**	8.65	8.73	8.81	8.87
10	7.63	7.42	7.62	7.63	7.82	7.66	8.39	8.44	8.49	8.48
11	7.19	6.91	7.14	7.18	7.36	7.25	7.84	8.02	7.98	8.02
12	6.16	5.44	5.95	6.15	6.28	6.22	6.69	7.19	7.51	7.51
13	4.59									
Panel C: FW3 method										
k	Mean	Median	Tr.Mean	DM(1)	DM(0.9)	DM(0.5)	CL(2)	CL(3)	CL(4)	CL(5)
1	2.24	1.39	1.86	2.28	2.28	2.15	3.25	4.61	5.27	5.27
2	3.36	2.47	3.17	3.43	3.45	3.24	5.10	6.09	6.73	7.08
3	4.43	3.60	4.30	4.50	4.55	4.29	6.37	7.26	7.68	7.90
4	5.36	4.75	5.30	5.42	5.50	5.22	7.16	7.88	8.23	8.38
5	6.13	5.73	6.11	6.18	6.29	5.99	7.75	8.29	8.52	8.66
6	6.72	6.82	6.73	6.76	6.90	6.62	8.16	8.57	8.75	8.87
7	7.13	7.31	7.15	7.16	7.33	7.08	8.44	8.77	8.89	8.95
8	7.39	7.44	7.41	7.41	7.59	7.39	8.61	**8.85**	**8.92**	**8.97**
9	**7.49**	**7.50**	**7.49**	**7.50**	**7.70**	**7.55**	**8.61**	8.77	8.81	8.86
10	7.42	7.29	7.41	7.43	7.63	7.53	8.45	8.54	8.54	8.55
11	7.11	6.92	7.09	7.11	7.31	7.25	7.97	8.16	8.15	8.08
12	6.33	5.71	6.19	6.33	6.47	6.46	6.89	7.33	7.58	7.58
13	4.92									

Notes: The Table reports the out-of-sample R^2 statistic with respect to the $AR(1)$ benchmark model for the out-of-sample period 1977:1-2010:12. The reported R^2_{OS} statistics are statistically significant based on the p-value of the Clark and West (2007) out-of-sample MSFE-adjusted statistic (CW_{pv}).

Table 4. Out-of-sample performance of subset quantile regression models (start date: January 1957)

Panel A: FW1 method										
k	Mean	Median	Tr.Mean	DM(1)	DM(0.9)	DM(0.5)	CL(2)	CL(3)	CL(4)	CL(5)
1	1.66	1.01	1.46	1.69	1.70	1.59	2.49	3.61	4.14	4.14
2	2.40	2.00	2.30	2.44	2.48	2.33	3.86	4.61	5.10	5.37
3	3.10	2.81	3.03	3.15	3.21	3.05	4.73	5.44	5.83	6.04
4	3.72	3.31	3.69	3.78	3.88	3.72	5.34	6.04	6.36	6.46
5	4.26	3.90	4.24	4.32	4.45	4.31	5.89	6.44	6.64	6.79
6	4.72	4.68	4.71	4.78	4.94	4.81	6.38	6.82	6.98	7.15
7	5.10	5.10	5.09	5.15	5.35	5.22	6.78	7.18	7.33	7.44
8	5.39	5.31	5.39	5.44	5.66	5.54	7.08	7.43	7.54	**7.63**
9	5.58	**5.55**	5.58	5.62	5.86	5.73	**7.22**	**7.47**	**7.56**	7.62
10	**5.59**	5.47	**5.59**	**5.62**	**5.87**	**5.75**	7.15	7.34	7.40	7.45
11	5.32	5.13	5.30	5.33	5.58	5.48	6.66	6.99	7.07	7.05
12	4.55	3.99	4.43	4.54	4.74	4.69	5.49	5.98	6.39	6.39
13	3.13									
Panel B: FW2 method										
k	Mean	Median	Tr.Mean	DM(1)	DM(0.9)	DM(0.5)	CL(2)	CL(3)	CL(4)	CL(5)
1	1.69	0.96	1.49	1.71	1.74	1.68	2.59	3.62	4.26	4.26
2	2.51	2.11	2.43	2.55	2.61	2.53	4.11	4.78	5.24	5.30
3	3.26	3.05	3.21	3.31	3.39	3.30	5.00	5.59	5.83	5.97
4	3.91	3.67	3.89	3.96	4.08	3.98	5.61	6.12	6.38	6.48
5	4.46	4.35	4.46	4.51	4.66	4.56	6.12	6.54	6.77	6.96
6	4.91	4.94	4.91	4.97	5.14	5.03	6.55	6.97	7.15	7.33
7	5.29	5.27	5.28	5.34	5.53	5.42	6.90	7.32	7.51	7.64
8	5.59	5.45	5.59	5.64	5.85	5.71	7.17	**7.58**	7.72	**7.82**
9	**5.79**	**5.68**	5.78	**5.82**	6.04	**5.88**	**7.29**	7.57	**7.73**	7.82
10	5.79	5.59	**5.78**	5.81	**6.05**	5.88	7.18	7.40	7.55	7.56
11	5.44	5.17	5.40	5.44	5.68	5.54	6.63	7.02	7.13	7.17
12	4.48	3.79	4.28	4.47	4.66	4.59	5.33	6.01	6.55	6.55
13	2.88									
Panel C: FW3 method										
k	Mean	Median	Tr.Mean	DM(1)	DM(0.9)	DM(0.5)	CL(2)	CL(3)	CL(4)	CL(5)
1	1.57	0.90	1.35	1.60	1.60	1.48	2.42	3.53	4.05	4.05
2	2.33	1.89	2.22	2.37	2.41	2.23	3.80	4.55	5.02	5.30
3	3.01	2.69	2.94	3.06	3.13	2.93	4.68	5.38	5.77	5.98
4	3.59	3.15	3.56	3.65	3.75	3.56	5.25	5.96	6.28	6.37
5	4.07	3.69	4.05	4.13	4.27	4.10	5.74	6.32	6.51	6.66
6	4.45	4.40	4.44	4.51	4.69	4.54	6.16	6.62	6.80	6.97
7	4.76	4.78	4.75	4.82	5.03	4.89	6.49	6.91	7.08	7.20
8	5.00	4.93	4.99	5.05	5.29	5.16	6.73	7.11	7.24	**7.35**
9	5.15	**5.10**	5.14	5.20	5.45	5.34	**6.85**	**7.15**	**7.26**	7.32
10	**5.17**	5.01	**5.16**	**5.20**	**5.47**	**5.37**	6.82	7.04	7.11	7.16
11	4.95	4.76	4.92	4.95	5.22	5.15	6.40	6.72	6.77	6.77
12	4.26	3.71	4.14	4.25	4.46	4.44	5.27	5.82	6.20	6.20
13	2.88									

Notes: The Table reports the out-of-sample R^2 statistic with respect to the $AR(1)$ benchmark model for the out-of-sample period 1977:1-2010:12. The reported R^2_{OS} statistics, which are greater than 0.50, are statistically significant based on the p-value of the Clark and West (2007) out-of-sample MSFE-adjusted statistic (CW_{pv}).

Table 5. Out-of-sample performance of subset quantile regression models (start date: January 1977)

Panel A: FW1 method										
k	Mean	Median	Tr.Mean	DM(1)	DM(0.9)	DM(0.5)	CL(2)	CL(3)	CL(4)	CL(5)
1	0.34	-0.15	0.14	0.33	0.36	0.28	0.67	1.37	1.76	1.76
2	1.05	0.49	0.91	1.05	1.11	0.96	1.87	2.49	2.99	3.26
3	1.77	1.31	1.63	1.77	1.85	1.64	2.88	3.52	4.00	4.24
4	2.44	1.74	2.35	2.45	2.55	2.30	3.66	4.37	4.71	4.77
5	3.04	2.42	2.99	3.07	3.18	2.90	4.40	4.90	5.02	5.10
6	3.55	3.33	3.54	3.59	3.71	3.41	4.99	5.31	5.34	5.45
7	3.97	4.01	3.97	4.01	4.14	3.83	5.37	5.63	5.66	5.68
8	4.31	4.42	4.33	4.36	4.49	4.19	5.63	5.84	5.86	5.85
9	4.59	4.88	4.62	4.64	4.77	4.48	5.82	5.90	**5.93**	**5.91**
10	4.87	5.21	4.91	4.91	5.03	4.75	5.94	**5.93**	5.87	5.91
11	5.16	5.35	5.25	5.20	5.30	5.06	5.89	5.90	5.84	5.82
12	5.43	**5.65**	**5.56**	**5.44**	**5.52**	**5.36**	**5.94**	5.51	5.55	5.55
13	**5.75**									

Panel B: FW2 method										
k	Mean	Median	Tr.Mean	DM(1)	DM(0.9)	DM(0.5)	CL(2)	CL(3)	CL(4)	CL(5)
1	0.33	-0.28	0.13	0.33	0.37	0.34	0.73	1.21	1.77	1.77
2	1.13	0.52	1.00	1.14	1.21	1.14	2.04	2.52	2.97	3.05
3	1.88	1.51	1.77	1.90	2.00	1.88	3.08	3.53	3.89	4.05
4	2.57	2.08	2.49	2.59	2.71	2.54	3.87	4.35	4.63	4.69
5	3.16	2.78	3.12	3.19	3.32	3.11	4.55	4.88	5.06	5.18
6	3.64	3.51	3.64	3.68	3.81	3.57	5.03	5.34	5.38	5.47
7	4.01	4.08	4.03	4.06	4.19	3.93	5.32	5.58	5.65	5.67
8	4.32	4.43	4.35	4.36	4.49	4.22	5.48	5.74	5.79	5.81
9	4.55	4.83	4.58	4.60	4.71	4.43	5.57	**5.73**	**5.81**	**5.81**
10	4.77	5.06	4.81	4.80	4.90	4.63	**5.59**	5.67	5.71	5.67
11	4.96	5.13	5.03	4.99	5.07	4.83	5.47	5.61	5.59	5.54
12	5.06	**5.26**	**5.17**	**5.07**	**5.12**	**4.96**	5.55	5.21	5.18	5.18
13	**5.26**									

Panel C: FW3 method										
k	Mean	Median	Tr.Mean	DM(1)	DM(0.9)	DM(0.5)	CL(2)	CL(3)	CL(4)	CL(5)
1	0.51	-0.01	0.28	0.50	0.53	0.42	0.85	1.54	1.95	1.95
2	1.27	0.69	1.12	1.26	1.32	1.13	2.11	2.72	3.20	3.51
3	1.99	1.53	1.86	1.99	2.07	1.81	3.12	3.79	4.25	4.47
4	2.63	1.93	2.55	2.65	2.74	2.44	3.87	4.60	4.91	4.93
5	3.18	2.59	3.14	3.21	3.32	2.98	4.54	5.07	5.16	5.22
6	3.62	3.41	3.61	3.66	3.78	3.43	5.06	5.38	5.42	5.52
7	3.95	4.01	3.96	3.99	4.13	3.78	5.36	5.60	5.64	5.66
8	4.21	4.32	4.23	4.26	4.40	4.07	5.53	5.73	5.76	5.76
9	4.43	4.68	4.44	4.48	4.62	4.31	5.66	5.75	**5.78**	**5.77**
10	4.66	4.96	4.69	4.71	4.84	4.56	5.76	**5.76**	5.71	5.75
11	4.95	5.17	5.03	4.99	5.11	4.87	5.76	5.75	5.68	5.64
12	5.25	**5.49**	**5.40**	**5.27**	**5.36**	**5.19**	**5.83**	5.49	5.51	5.51
13	**5.54**									

Notes: The Table reports the out-of-sample R^2 statistic with respect to the $AR(1)$ benchmark model for the out-of-sample period 1977:1-2010:12. The reported R^2_{OS} statistics are statistically significant based on the p-value of the Clark and West (2007) out-of-sample MSFE-adjusted statistic (CW_{pv}).

Table 6. Out-of-sample performance of the 'optimal' CSR and QFC forecasts

	Mean	Median	Tr.Mean	DM(1)	DM(0.9)	DM(0.5)	CL(2)	CL(3)	CL(4)	CL(5)
Panel A: Complete Subset Regression										
January 1937	5.00	4.38	4.93	5.12	5.34	5.01	7.37	8.14	8.50	8.56
January 1957	3.75	3.46	3.73	3.84	4.15	4.12	5.74	6.15	6.39	6.39
January 1977	4.03	3.47	3.99	4.07	4.36	4.04	5.47	6.00	6.13	6.07
Panel B: Quantile Forecast Combinations- FW1										
	Mean	Median	Tr.Mean	DM(1)	DM(0.9)	DM(0.5)	CL(2)	CL(3)	CL(4)	CL(5)
January 1937	6.40	6.09	6.36	6.45	6.56	6.30	7.99	8.53	8.80	8.96
January 1957	4.67	4.44	4.63	4.72	4.87	4.73	6.37	6.93	7.21	7.39
January 1977	3.21	2.65	3.15	3.23	3.35	3.07	4.55	5.07	5.32	5.45
Panel C: Quantile Forecast Combinations- FW2										
	Mean	Median	Tr.Mean	DM(1)	DM(0.9)	DM(0.5)	CL(2)	CL(3)	CL(4)	CL(5)
January 1937	6.46	6.23	6.43	6.50	6.61	6.36	7.97	8.50	8.83	9.02
January 1957	4.85	4.80	4.83	4.89	5.05	4.93	6.45	6.96	7.29	7.51
January 1977	3.35	3.18	3.33	3.38	3.50	3.28	4.60	5.07	5.38	5.55
Panel D: Quantile Forecast Combinations- FW3										
	Mean	Median	Tr.Mean	DM(1)	DM(0.9)	DM(0.5)	CL(2)	CL(3)	CL(4)	CL(5)
January 1937	6.38	6.13	6.35	6.43	6.55	6.27	8.04	8.59	8.84	8.99
January 1957	4.56	4.36	4.53	4.61	4.77	4.61	6.30	6.87	7.14	7.31
January 1977	3.49	3.04	3.45	3.52	3.64	3.31	4.84	5.34	5.54	5.65

Notes: The Table reports the out-of-sample R^2 statistic of the optimal Complete Subset Regression (CSR) and Quantile Forecast Combination (QFC) approaches with respect to the $AR(1)$ benchmark model for the three out-of-sample periods considered. The reported R^2_{OS} statistics are statistically significant based on the p-value of the Clark and West (2007) out-of-sample MSFE-adjusted statistic (CW_{pv}).