Out-of-sample equity premium prediction: A complete subset quantile regression approach

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 extends the complete subset linear regression framework to a quantile regression setting. We employ complete subset combinations of quantile forecasts in order to construct robust and accurate equity premium predictions. We show that our approach delivers statistically and economically significant out-of-sample forecasts relative to both the historical average benchmark, the complete subset mean regression approach and the single-variable quantile forecast combination approach. Our 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: G11; G12; C22; C53

Keywords: Equity premium; Forecast combination; Predictive quantile regression; Robust point forecasts; Subset quantile regressions.

Acknowledgments: Ioannis Vrontos greatly acknowledges financial support from the Research Center of Athens University of Economics and Business under the grant EP-3043-01.

*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

The issue of forecasting equity returns is one of the most widely discussed topics in the finance literature mainly due to its central role in asset pricing, portfolio allocation and evaluation of investment managers. The in-sample predictive ability of a quite exhaustive list of potential predictors that typically contains valuation ratios, various interest rates and spreads, distress indicators, inflation rates along with other macroeconomic variables, indicators of corporate activity, etc. was the focus of the earlier studies.[1] However, since the seminal contribution of Goyal and Welch (2008) who show that their long list of predictors can not deliver consistently superior out-of-sample performance, attention has turned to the development of improved forecasting methods in order to establish the empirical validity of equity premium (proxied by excess returns) predictability. To mention a few, Campbell and Thompson (2008) show that when imposing simple restrictions, suggested by economic theory, on predictive regressions' coefficients, the out-of-sample performance improves (see also Ferreira and Santa-Clara, 2011). Ludvigson and Ng (2007) and Neely, Rapach, Tu and Zhou (2014) adopt a diffusion index approach, which can conveniently track the key movements in a large set of predictors, and find evidence of improved equity premium forecastability.[2]

In an attempt to reduce both model uncertainty and parameter instability, Rapach, Strauss and Zhou (2010, RSZ henceforth) employ forecast combinations of univariate equity premium models and find that combinations of individual single variable predictive regression models significantly beat the historical average forecast. Building on RSZ, Meligkotsidou, Panopoulou, Vrontos and Vrontos (2014, MPVV henceforth) incorporate the forecast combination methodology in a quantile regression setting. Their quantile regression approach allows them to cope with the non-linearity and non-normality patterns

[1] Commonly used valuation ratios are the dividend price/dividend yield ratio (see for example, Fama and French, 1988, 1989), the earnings price ratio (Campbell and Shiller, 1988, 1998), and the book-to-market ratio (Kothari and Shanken, 1997). Another strand of the literature includes macroeconomic/ financial variables such as inflation rates, short-term and long-term interest rates along with term and corporate bond spreads in the set of predictors (see e.g. Fama and Schwert, 1977; Campbell and Vuolteenaho, 2004; Campbell, 1987; Fama and French, 1989; Ang and Bekaert, 2007). A comprehensive list of variables that serve as predictors can be found in Goyal and Welch (2008).

[2] Rapach and Zhou (2012) offer a detailed review on the issue of equity return predictability.

that are evident in the relationship between stock returns and potential predictors. 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, Gargano and Timmermann (2013, EGT henceforth). 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.

This paper proposes a new forecasting approach based on complete subset quantile regressions. Specifically, we extend the framework of EGT to a quantile regression setting and adopt the methodology of MPVV in order to produce robust and accurate equity premium forecasts. Our proposed methodology merges three strands of the literature on out-of-sample forecasting and, as shown, exploits the benefits emerging from each one. First, we exploit the ability of the quantile regression setting to produce robust and accurate point forecasts. Second, we reduce model uncertainty and parameter instability by employing quantile forecast combinations. Finally, we employ complete subset quantile regressions which induces shrinkage to the respective estimates and further helps reduce the effect of parameter estimation error.

To be more specific, our forecasting framework is rooted in quantile predictive regressions, which have attracted a vast amount of attention since the seminal paper of Koenker and Bassett (1978). Incorporating the forecast combination approach (see RSZ) into our quantile regression setting helps reduce model uncertainty and deals with parameter instability.[3] MPVV propose two alternative ways to generate forecasts within the quantile regression setup. The first approach proceeds by first constructing robust point forecasts from a set of quantile predictions all of which are based on the same predictive variable. Next, it combines the robust forecasts obtained from different predictors using several existing combination methods in order to produce a final point forecast. The second approach consists of first combining all the predictions of the same quantile ob-

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

tained from different single predictor model specifications, in order to produce combined quantile forecasts. Then, robust point forecasts are obtained by operating either a fixed or a time-varying weighting scheme on the combined quantile forecasts.

The methodologies discussed so far employ single variable models in either a linear or a quantile regression framework. EGT abstract from the single predictor models and propose combining forecasts from all possible linear regression models that keep the number of predictors fixed. Their approach introduces a complex version of shrinkage to the respective estimates which helps reduce the effect of parameter estimation error. EGT show that the amount of shrinkage induced on least squares estimates from subset regressions is a function of the number of variables included in the model (k) and the total number of available predictors (K). Given that the amount of shrinkage depends on all the least squares estimates, it varies with each coefficient. Moreover, this methodology can cure the omitted variable bias especially in cases with strongly positively correlated regressors. The authors propose constructing forecasts based on a simple averaging scheme of all the possible models employed keeping the numbers of regressors fixed. In this paper, we extend the framework of EGT and MPVV to the quantile predictive regression framework discussed above. Similarly to EGT, we utilize information from all the predictors simultaneously in order to produce combined quantile forecasts from all quantile regressions that keep the number of predictors fixed. We also abstract from the simple averaging schemes and introduce several existing combination schemes into our setting. Then, the obtained quantile forecasts are synthesized to produce robust point forecasts of the variable of interest.

The empirical findings of both EGT and the present paper suggest that the predictive performance of subset regressions highly depend on the value of k. A further contribution of this paper is the development of a recursive algorithm for selecting k in real time, based on the past history of excess returns and predictive variables. The proposed algorithm is a likelihood-based method that chooses the best complete subset for a given quantile and is flexible enough to allow for variability of the selected value of k across quantiles. In this way, our approach incorporates information on the best subset for each quantile of the return distribution in real time and these 'optimal' quantile forecasts are appropriately combined to deliver robust equity premium forecasts.

To anticipate our key results, we find that our complete subset quantile regression framework achieves superior predictive performance, both in statistical and economic evaluation terms. More in detail, our proposed approach can lead to an out-of-sample R^2 of 5.64% (relative to the historical average benchmark) as opposed to 4.10% of the subset linear regression approach of EGT, 4.06% of the quantile combination approach of MPVV and 3.58% of the linear combination approach of RSZ. Our Model Confidence Set (MCS; Hansen, Lunde and Nason, 2011) and associated statistical significance results further complement and reinforce our findings. Specifically, the superiority of our methodology over the methodologies of EGT, MPVV and RSZ is evident as neither the linear single-variable or subset models, nor the quantile single-variable subset models are included in MCS. Tests of equal predictive ability indicate that while in a linear regression framework, subsets of two variables ($k = 2$) perform better than the remaining specifications, in our quantile regression framework subsets of either two or three variables ($k = 2, 3$) emerge as superior. Our real time recursive algorithm for selecting k across quantiles of returns succeeds in identifying the 'correct' value of k which is both time-varying and quantile-varying. When evaluating our forecasts from an economic perspective and specifically for a mean-variance investor, we also need return volatility forecasts, which we construct using the interval approximation approach of Pearson and Tukey (1965) and a set of predictive quantiles. Our economic evaluation results suggest that an investor that adopts our framework can gain sizeable benefits which range from 3.58% to an impressive 5.69% per year relative to a naive strategy based on the historical benchmark performance.

The outline of the paper is as follows. Section 2 describes the complete subset regression framework of EGT and introduces its extension to the quantile regression framework. The proposed methodology for robust estimation of the central location of the distribution of returns is outlined in Section 3. Section 4 presents our empirical findings, while section 5 describes the proposed methodology for the recursive selection of the number of predictors. Section 6 outlines the economic evaluation framework and presents the associated findings. Section 7 summarizes and concludes. Supplementary material is included in the Appendix (available from the authors upon request).

2 Complete Subset Quantile Regressions

2.1 Complete subset regressions

EGT 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.

More in detail, suppose that we are interested in forecasting the equity premium, denoted by r_t, using a set of K predictive variables. First we consider all possible predictive mean regression models with a single predictor, i.e. $k = 1$, of the form

$$r_{t+1} = \alpha_i + \beta_i x_{it} + \varepsilon_{t+1}, \ i = 1, \ldots, K, \tag{1}$$

where r_{t+1} is the observed excess return on a stock market index in excess of the risk-free interest rate at time $t + 1$, 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. Similarly, a regression of r_{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 $k \leq K$. EGT show that while subset regression combinations bear similarities to a complex version of shrinkage, they do not reduce to shrinking OLS estimates. 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).[4]

2.2 Complete subset quantile regressions

The above linear subset regression specification can only predict the mean and not the entire distribution of returns in the event that the joint distribution of r_{t+1} and x_{it} is not bivariate Gaussian and, therefore, their relationship is not linear. We adopt a more sophisticated approach to equity premium forecasting by employing predictive quantile regression models (Koenker and Bassett, 1978) and incorporating the complete subset combination framework of EGT in our quantile regression setting. The proposed approach is designed as follows.

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

$$r_{t+1} = \alpha_i^{(\tau)} + \beta_i^{(\tau)} x_{it} + \varepsilon_{t+1}, \ i = 1, \ldots, K, \tag{2}$$

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 r_{t+1} given x_{it} is $Q_\tau(r_{t+1}|x_{it}) = \alpha_i^{(\tau)} + \beta_i^{(\tau)} x_{it}$, where the intercept and the regression coefficients depend on τ. The $\beta_i^{(\tau)}$'s are likely to vary across τ's, revealing a larger amount of information about returns than the predictive mean regression model (Equation 1). Estimators of the parameters of the linear quantile regression models in (2), $\hat{\alpha}_i{}^{(\tau)}, \hat{\beta}_i^{(\tau)}$, can be obtained by minimizing the sum $\sum_{t=0}^{T-1} \rho_\tau \left(r_{t+1} - \alpha_i{}^{(\tau)} - \beta_i^{(\tau)} x_{it} \right)$,

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

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) with probability density function

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

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.

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

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

[5] 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.

2.3 Forecasting Approaches based on Complete Subset Quantile Regression

We construct equity premium 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 excess return (r_{t+1}). Next, two approaches are explored in order to combine these quantile forecasts into a point forecast that is robust to non-normality and non-linearity.

The first approach, which we name Robust Forecast Combination approach (RFC) proceeds by first combining the quantile forecasts across all values of τ into point forecasts for each complete subset of predictors. We employ Tukey's (1977) and Gastwirth's (1966) three-quantile estimators and the five-quantile estimator of Judge, Hill, Griffiths, Lutkepohl and Lee (1988) along with their time-varying counterparts developed in MPVV. This step yields $n_{k,K}$ point forecasts which are further combined in order to reduce uncertainty risk associated with each subset of the predictive variables. Except for the simple averaging scheme, suggested by EGT, we also employ the Trimmed Mean, the Median, the Discount Mean Squared Forecast Error (DMSFE) of Stock and Watson (2004) along with the Cluster combining method, introduced by Aiolfi and Timmermann (2006). These combining schemes utilize the Mean Squared Forecast Error (MSFE) as a loss function.

The second approach, which we name Quantile Forecast Combination (QFC) consists of first combining the predicted τ^{th} quantiles across all different subsets (k) of predictors ($n_{k,K}$ model specifications). 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 MPVV, we employ the Discount Asymmetric Loss Forecast Error (DALFE) and the Asymmetric Loss Cluster (AL Cluster) in order to construct subset quantile forecasts. 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).[6]

3 Point Forecasts based on Regression Quantiles

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. Lima and Meng (2017) provide a theoretical explanation for the use of combination of conditional quantiles to approximate the conditional mean forecast. The authors also argue that since the low-end (high-end) quantiles produce a downwardly (upwardly) biased forecast of the conditional mean, approximating the mean via conditional quantiles combines oppositely biased predictions, and these biases cancel each other out.

3.1 Point Forecasts based on a Fixed Weighting Scheme

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 form

$$\hat{r}_{t+1} = \sum_{\tau \in S} p_\tau \hat{r}_{t+1}(\tau), \quad \sum_{\tau \in S} p_\tau = 1,$$

where S denotes the set of quantiles that are combined, $\hat{r}_{t+1}(\tau)$ denotes the quantile forecasts associated with the τth quantile and $\hat{r}_{t+1}$ 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, the Gastwirth (1966) three-quantile estimator and the five-quantile estimator, suggested by Judge, Hill, Griffiths, Lutkepohl and Lee

[6] Details of the combining schemes are given in Appendix A.2.

(1988) given, respectively, by the following formulae

$$\text{FW1: } \widehat{r}_{t+1} = 0.25\hat{r}_{t+1}(0.25) + 0.50\hat{r}_{t+1}(0.50) + 0.25\hat{r}_{t+1}(0.75)$$

$$\text{FW2: } \widehat{r}_{t+1} = 0.30\hat{r}_{t+1}(1/3) + 0.40\hat{r}_{t+1}(0.50) + 0.30\hat{r}_{t+1}(2/3).$$

$$\text{FW3: } \widehat{r}_{t+1} = 0.05\hat{r}_{t+1}(0.10) + 0.25\hat{r}_{t+1}(0.25) + 0.40\hat{r}_{t+1}(0.50) + 0.25\hat{r}_{t+1}(0.75) + 0.05\hat{r}_{t+1}(0.90).$$

3.2 Point Forecasts based on a Time-varying Weighting Scheme

Relaxing the assumption of a constant weighting scheme seems to be a natural extension. A number of factors, such as changes in regulatory conditions, market sentiment, monetary policies, institutional framework or even changes in macroeconomic interrelations (Campbell and Cochrane, 1999; Menzly, Santos and Veronesi, 2004; Spiegel, 2008; Dangl and Halling, 2012) can motivate the employment of time-varying schemes in the generation of robust point forecasts.

The variable of interest, r_{t+1}, is predicted using an optimal linear combination $\mathbf{p}_t=[p_{\tau,t}]_{\tau\in S}$ of the quantile forecasts $\hat{r}_{t+1}(\tau)$ given by

$$\hat{r}_{t+1} = \sum_{\tau\in S} p_{\tau,t}\hat{r}_{t+1}(\tau), \;\; \sum_{\tau\in S} p_{\tau,t} = 1.$$

The weights, $\mathbf{p}_t$, are estimated recursively using a holdout out-of-sample period continuously updated by one observation at each step. Optimal estimates of the weights are obtained by minimizing the mean squared forecast errors, $E_t(r_{t+1} - \hat{r}_{t+1})^2$, under an appropriate set of constraints. Our optimization procedure is the analogue of the constrained Granger and Ramanathan (1984) method for quantile regression forecasts (see also Timmermann, 2006; Hansen, 2008; Hsiao and Wan, 2014). Specifically, we employ constrained least squares using the quantile forecasts as regressors in lieu of a standard set of predictors. The time-varying weights on the quantile forecasts bear an interesting relationship to the portfolio weight constraints in finance. In this sense we constrain the weights to be non-negative, sum to one and not to exceed certain lower and upper bounds in order to reduce the weights' volatility and stabilize forecasts. In our empirical application, we employ three time-varying specifications which may be viewed as the time-varying counterparts of our FW1-FW3 schemes. More specifically, FW1 with time-varying coefficients becomes

$$\text{TVW1: } \widehat{r}_{t+1} = p_{0.25,t}\hat{r}_{t+1}(0.25) + p_{0.50,t}\hat{r}_{t+1}(0.50) + p_{0.75,t}\hat{r}_{t+1}(0.75),$$

$$\mathbf{p}_t = \arg\min_{\mathbf{p}_t} E[r_{t+1} - (p_{0.25,t}\hat{r}_{t+1}(0.25) + p_{0.50,t}\hat{r}_{t+1}(0.50) + p_{0.75,t}\hat{r}_{t+1}(0.75))]^2$$

$$s.t.\ p_{0.25,t} + p_{0.50,t} + p_{0.75,t} = 1, 0.20 \leq p_{0.25,t} \leq 0.40,\ 0.40 \leq p_{0.50,t} \leq 0.60, 0.20 \leq p_{0.75,t} \leq 0.40.$$

Similarly, the FW2 and FW3 schemes with time-varying coefficients become

$$\begin{aligned} \text{TVW2} &: \ \widehat{r}_{t+1} = p_{1/3,t}\hat{r}_{t+1}(1/3) + p_{0.5,t}\hat{r}_{t+1}(0.50) + p_{2/3,t}\hat{r}_{t+1}(2/3), \\ \text{TVW3} &: \ \widehat{r}_{t+1} = p_{0.10,t}\hat{r}_{t+1}(0.10) + p_{0.25,t}\hat{r}_{t+1}(0.25) + p_{0.5,t}\hat{r}_{t+1}(0.50) + \\ &\quad + p_{0.75,t}\hat{r}_{t+1}(0.75) + p_{0.90,t}\hat{r}_{t+1}(0.90), \end{aligned}$$

where the weights are estimated in a similar fashion to FW1.

4 Empirical findings

4.1 Data, forecast construction and forecast evaluation

The data we employ are from Goyal and Welch (2008) who provide a detailed description of transformations and datasources.[7] The equity premium is calculated as the difference of the continuously compounded S&P500 returns, including dividends, and the Treasury Bill rate over the sample period 1947:Q1 to 2017:Q4 ($T = 284$ observations). Following the line of work of Goyal and Welch (2008), RSZ and Ferreira and Santa-Clara (2011), out-of-sample forecasts of the equity premium are generated by continuously updating the estimation window, i.e. following a recursive (expanding) window. More specifically, our estimation window starts with T_0 =32 observations from 1947:Q1 to 1954:Q4 and expands by one quarter at a time as we move forward. The out-of-sample forecasts start from

[7] The data are available at http://www.hec.unil.ch/agoyal/. We thank Prof. Goyal for making them available to us.

1955:Q1 to 2017:Q4, corresponding to P =252 observations. In addition, constructing our time-varying robust forecasts and several forecast combination schemes require a holdout period to estimate the weights. To this end, we use the first P_0 =40 observations from the out-of-sample period as an initial holdout period (1955:Q1 to 1964:Q4), which also expands periodically. In the end, we are left with a total of 212 post-holdout out-of-sample forecasts available for evaluation. Our out-of-sample forecast evaluation period corresponds to the 'long' one analyzed by Goyal and Welch (2008) and RSZ.

The 12 economic variables employed in our analysis are related to stock-market characteristics, interest rates and broad macroeconomic indicators. With respect to stock market characteristics, we employ the Dividend–price ratio (log), D/P, the difference between the log of dividends paid on the S&P 500 index and the log of stock prices (S&P 500 index), where dividends are measured using a one-year moving sum; Dividend yield (log), D/Y, the difference between the log of dividends and the log of lagged stock prices; Earnings–price ratio (log), E/P, the difference between the log of earnings on the S&P 500 index and the log of stock prices, where earnings are measured using a one-year moving sum; Book-to-market ratio, B/M, the ratio of book value to market value for the Dow Jones Industrial Average and Net equity expansion, NTIS, the ratio of twelve-month moving sums of net issues by NYSE-listed stocks to total end-of-year market capitalization of NYSE stocks. Turning to interest-rate related variables, we employ five variables ranging from short-term government rates to long-term government and corporate bond yields and returns along with their spreads. These are the Treasury bill rate, TBL, the interest rate on a three-month Treasury bill (secondary market); 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; Default yield spread, DFY, the difference between BAA- and AAA-rated corporate bond yields; Default return spread, DFR, the difference between long-term corporate bond and long-term government bond returns; To capture the overall macroeconomic environment, we employ the inflation rate, INFL, calculated from the CPI (all urban consumers) and the investment-to-capital ratio, I/K, the ratio of aggregate (private nonresidential fixed) investment to aggregate capital for the entire economy.[8] Table 1 presents the descriptive statistics of the equity premium and the candidate predictors. On average, the quarterly equity premium reaches 1.7%, with a standard deviation of 7.7% and maximum and minimum values of 19.2% and -31.1%, respectively. In addition, the related distribution shows departures from normality as suggested by the negative skewness and excess kurtosis. Similar characteristics pertain with respect to the candidate predictor variables, the majority of which are leptokurtic and positively skewed.

[TABLE 1 AROUND HERE]

The natural benchmark forecasting model is the historical mean or prevailing mean (PM) model, according to which the forecast of the equity premium coincides with the constant in the linear regression 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 $R^2_{OS} = 1 - \frac{MSFE_i}{MSFE_{PM}}$, where $MSFE_i$ is the Mean Square Forecast Error associated with each of our competing models and specifications and $MSFE_{PM}$ is the respective value for the PM 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 both the Clark and West (2007) (CW) and the Diebold and Mariano (1995) (DM) test to compare our models/ specifications.[9]

The following subsections present an illustration of our proposed complete subset quantile regression approach to equity premium 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. Specifically, we examine the potential benefits of the subset quantile regression forecasts based on k-variate model forecasts ($k \geq 2$) under various combination methods (e.g. Mean, Median, Trimmed Mean, DMSFE, Cluster) relative to using subset linear regression forecasts

[8] Following EGT, we exclude the log dividend earnings ratio and the long term yield in order to avoid multicollinearity.

[9] A brief description of the Clark and West (2007) test and the Diebold and Mariano (1995) test is given in Appendices B.1 and B.2. Note that the CW test is intended to compare nested linear models, while the DM is more appropriate when comparing forecasts generated by non-nested models.

based on k-variate models as proposed in EGT or relative to several combination methods of univariate linear and/or quantile models as proposed in MPVV. To this end, we also test the forecasting accuracy of our proposed methodology relative to that of MPVV by adding as a benchmark the linear and subset quantile regression forecasts based on k-variate model forecasts for $k = 1$. Finally, we also employ the Model Confidence Set (MCS) approach of Hansen, Lunde and Nason (2011) to reveal the models with superior predictive ability without specifying a benchmark model. A MCS is a subset of models that contains the best model with a given level of confidence. The excluded models are viewed as significantly inferior models. Models with a larger MCS $p-$value show stronger predictive ability. We consider the significance (confidence) level of 10% (90%) and calculate the MCS $p-$values based on the range statistic using the circular block bootstrap and $MSFE$ as loss function. Hansen, Lunde and Nason (2011) argue that with informative data, the model confidence set consists only of the best model, whereas less informative data may result in a MCS with more than one model.[10]

4.2 Complete Subset Linear Regression Models

Before discussing the out-of-sample performance of the forecasts obtained by subset linear regressions under various combination schemes, we first present the forecasting ability of candidate predictor variables in a linear setting. Table 2 (second column) presents the related R^2_{OS} statistics relative to the historical average benchmark model for the out-of-sample period 1965:1-2017:4. The third and fourth column of the Table present the p-values from the CW and DM tests, respectively. Our findings confirm the consensus in the literature with respect to the scarce equity premium predictability. Specifically, only four variables, D/Y, DFR, INFL and I/K generate positive R^2_{OS} statistics, of which only D/Y and I/K are statistically significant on the basis of the CW test.[11]

[TABLE 2 AROUND HERE]

Table 3 presents the R^2_{OS} statistics of all subset regressions relative to the historical

[10] We thank the Associate Editor and an anonymous reviewer for these suggestions.

[11] TBL is also statistically significant at the 5% level, although the related R^2_{OS} is negative. This case can occasionally arise since the CW statistic is testing the one-sided null hypothesis of equal predictive accuracy in population, while the reported R^2_{OS} values reflect finite-sample performance.

average benchmark model (Panel A) and to the univariate subset ($k = 1$, Panel B). The second column of the Table reports the R^2_{OS} generated by simply averaging the forecasts (Mean combination method) produced by subset linear regressions for various values of k. This experiment coincides with the framework of EGT and suggests that the subset linear regression with $k = 2$ generates the largest R^2_{OS} value (3.84%). Similarly to EGT, subset regression forecasts with $k \leq 6$ produce positive R^2_{OS} values, while the out-of-sample forecasting ability of subsets deteriorates markedly for $k \geq 7$.

[TABLE 3 AROUND HERE]

Next, 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, the largest R^2_{OS} values occur for the $k = 2$ subset, with the exception of the Cluster schemes where the largest R^2_{OS} occur for $k = 1$. For these subsets ($k = 2$ or $k = 1$), most of the combining methods produce statistically significant positive R^2_{OS} values, while four of them, namely the Median, Trimmed Mean, DMSFE(0.9) and DMSFE(0.5) provide higher values of R^2_{OS} than that of the best ($k = 2$) subset regression based on the Mean combination scheme. A comparison of the different combination techniques suggests that the DMSFE(0.5) scheme, which penalizes more recent forecasting accuracy, ranks first followed by the Median combination scheme. These methods provide the highest R^2_{OS} values of 4.29% and 4.24%, respectively. To ascertain whether the forecasting models have a statistically significant difference in their out-of-sample performance, we also report the MCS findings. Specifically, bold indicates the best models that belong to the MCS. Out of 90 models, only 8 appear to be superior, excluding both the benchmark historical average model and the linear univariate subset models ($k = 1$), which coincides with the RSZ setup. The best performing models are the Mean, Median, Trimmed Mean, DMSFE (0.9) and DMSFE (0.5) for $k = 2$ and Median, Trimmed Mean, DMSFE (0.5) for $k = 3$. Turning to Panel B that reports the R^2_{OS} statistics of all subset regressions for ($k \geqslant 2$) relative to the univariate subset regression forecasts ($k = 1$), we have to note that positive R^2_{OS} are associated with $k = 2$ and 3 for all but the cluster methods and additionally for $k = 4$ for the Median and Trimmed

Mean combining methods. Statistical significance via the DM test, though, prevails only for the Median method and for subsets of 2 predictors. Relying on the CW test, we find that the majority of the models with positive R^2_{OS} statistics are statistically significant at the 10% level.

4.3 Complete Subset Quantile Regression Models

4.3.1 Robust Forecast Combination approach

Table 4 reports the R^2_{OS} statistics and the respective p-values of the CW and DM tests for the subset quantile regression models based on the RFC approach for the three fixed weighting schemes, i.e. the FW1 scheme (Panel A), the FW2 scheme (Panel B) and the FW3 scheme (Panel C). Results are reported for various combination methods, namely the Mean, Median, Trimmed Mean, DMSFE and Cluster, based on $k = 1$ to $k = 12$ subset quantile regression forecasts. As previously, bold denotes the models that belong to the MCS and Panels D-F report the R^2_{OS} statistics of all subset regressions for ($k \geqslant 2$) relative to the univariate subset regression forecasts ($k = 1$). To save space, we report the related figures for k up to 7. The full set of results is available in Appendix G.

Several findings emerge from this analysis. First, we observe that combining the forecasts of a subset of $k = 3$ quantile regression models produces higher R^2_{OS} values for almost all the combination methods with the exception of the Cluster(3) method (for which $k = 4$ emerges as superior) for all fixed weighting schemes. Second, for this subset, i.e. $k = 3$, all the combination methods based on the robust quantile regression models generate higher R^2_{OS} values than the corresponding combining methods based on the best $k = 2$ subset linear regression models, indicating the superior forecasting ability of the proposed RFC quantile approach. Third, a comparison of the different combination methods suggests that the Median combination technique outperforms the alternative combination methods for all FW schemes, generating R^2_{OS} values of 5.34%, 5.20% and 5.24%, respectively. Depending on the combination method and FW scheme, a small portion of models belong in the MCS (69 out of 298). In the majority of cases, this involves models generated for subsets of $k = 2$ to $k = 5$. As previously, cluster combinations prove inferior to even simpler methods such as the Mean or Median ones. Finally, even univariate subset forecasts combined via the DMSFE (0.5) method belong to the superior set of models. Testing for the improvements of our approach over MPVV, Panels D-F of Table 4 report the respective R^2_{OS} values along with the CW and DM $p-$values. For the majority of cases, subsets of k greater than or equal to two and less than six generate positive R^2_{OS} values. However, these are significant (on the basis of the DM test) only for the cases of $k = 2$ for all FW schemes and combining methods with the exception of the Cluster combination methods and DMSFE (0.5).

[TABLE 4 AROUND HERE]

Next, we present the out-of-sample performance of the subset quantile regression forecasts relative to the benchmark Prevailing Mean model based on the time-varying weighting schemes TVW1-TVW3 (Table 5, Panels A-C). Three combination methods can be used in the time-varying weighting framework; the Mean, Median and the Trimmed Mean. Based on the results of Table 5, we observe that the largest R^2_{OS} values occur for $k = 2$ or $k = 3$ subsets. For these subsets ($k = 2$ and $k = 3$), all the combining methods, i.e. the Mean, Median and Trimmed Mean, generate statistically significant positive R^2_{OS} values, which are higher than the corresponding R^2_{OS} values of the combining methods based on the best ($k = 2$) subset linear regression model (see Table 3). For these best subsets, the Median and the Trimmed Mean combination methods seem to outperform the Mean combination scheme since they produce higher R^2_{OS} values. The most striking result is the R^2_{OS} statistic of 5.64% obtained by the Median combination of forecasts of the $k = 2$ subset quantile regression models under the TVW1 scheme. Subsets of 2 and 3 variables belong to the MCS for TVW1 and TVW2 scheme along with $k = 4$ for TVW2 and the mean and Trimmed mean combination methods. On the other hand, TVW3 that involves a finer grid of quantiles, contributes to MCS with only two specifications; namely the Median and Trimmed Mean for $k = 2$ and $k = 3$, respectively. Turning to the $k = 1$ benchmark (Panels D - F), we note that subsets of up to 4 or 5 variables can improve forecasting performance as they are associated with positive R^2_{OS} values. However, judging from the related DM p-values, only subsets of $k = 2$ appear statistically superior for all combining methods on the basis of the TVW1 scheme and for the Median and Trimmed Mean for the remaining time varying schemes.

[TABLE 5 AROUND HERE]

4.3.2 Quantile Forecast Combination approach

We turn our attention to the results of the subset quantile regression models based on the QFC approach. Table 6 reports the out-of-sample performance of the subset quantile regression forecasts obtained by the QFC approach using fixed weighting schemes (FW1-FW3). The results of Table 6 (Panel A - Panel C) indicate that high positive R^2_{OS} values are obtained by using $k = 2$, $k = 3$ and $k = 4$ subsets for all weighting schemes FW1-FW3. In particular, for $k = 3$ subsets all of the combining methods produce the highest positive R^2_{OS} values, which are larger than those of the best ($k = 2$) subset linear regression model (see Table 3) and similar or even higher than the corresponding R^2_{OS} values of the best ($k = 3$) subset quantile regression forecasts based on the RFC approach (see Table 4). Among the various combination methods, the Median combination scheme ranks first, since, for the best $k = 3$ subset, generates the highest R^2_{OS} values ranging from 5.43% for FW3 to 5.56% for FW2 scheme. Second ranks the DALFE(0.5) method which produces R^2_{OS} values ranging from 5.12% for FW2 to 5.25% for the FW1 scheme. The MCS findings suggest that forecasts generated for $k = 3$ are optimal for all combining methods, but AL Cluster(3) and AL Cluster(2) for FW2 and FW3 forecasts. In the majority of cases, subsets of 2 and 4 variables also belong to MCS. Panels D-E report our findings for the $k = 1$ benchmark offering support to the proposed subset quantile approach over the univariate MPVV one. Similarly to our RFC approach, QFC-FW forecasts improve forecasts for subsets up to 5 or 6 depending on the combining method/scheme employed, as suggested by the positive R^2_{OS} values. Judging from the DM $p-$values, using $k = 2$ offers statistically significant improvements for all methods but the Cluster ones. In the case of FW3, the Median combination scheme for $k = 3$ is also marginally significant.

[TABLE 6 AROUND HERE]

Finally, Table 7 (Panels A-C) presents the results obtained by the subset QFC approach using time-varying weighting schemes (TVW1-TVW3). Three combination methods, namely the Mean, Median and Trimmed Mean, are used in this approach. Based on the results of Table 7, we observe that the subset quantile regression forecasts with $k = 2$ for QFC-TVW1 and QFC-TVW3 and with $k = 3$ for QFC-TVW2 generate the highest positive R^2_{OS} values. For these subsets ($k = 2$ or $k = 3$), the Median combination method outperforms the Mean and the Trimmed Mean combination schemes since it generates higher R^2_{OS} values. More importantly, the QFC-TVW1 approach based on the Median combination of $k = 2$ subsets of predictors produce the highest R^2_{OS} of 5.58% among the different forecasting approaches considered in our analysis (see Table 7, Panel A). Less than 15% (14/103) of the QFC-TVW models generate forecasts that belong to the MCS. These optimal models are mainly associated to two-variable subsets with the exception of TVW3 and the Mean and Trimmed Mean combining methods. Subsets of $k = 3$ join the best set of models for TVW2 and TVW1 for the Median and Trimmed mean combinations. For these combination schemes and TVW2 even $k = 4$ joins the set of best models. Turning to our comparisons with the MPVV approach (Panels D-F), we have to note that improvements to R^2_{OS} values prevail for subsets of up to $k = 4$ relative to $k = 1$, which are statistically significant for the majority of combination/weighting schemes for $k = 2$.[12],[13]

[TABLE 7 AROUND HERE]

4.4 Global MCS and summary

To further consolidate our findings, we also report MCS results employing all the forecasting models discussed above, i.e. the benchmark historical mean, the linear univariate models, the subset linear regression models for all k and combination methods along with the FW/ TVW RFC and QFC subset quantile models. Table 8 outlines the mod-

[12] We also evaluated the performance of our forecasting approaches by plotting the difference of the cumulative sum of squared errors (DSSE) of the null (historical mean) minus the cumulative sum of squared errors of the alternative proposed models. Overall, this set of results (reported in Appendix C) points to superiority of our approaches, which is consistent over time as shown by the upward sloping curve for the most part of the out-of-sample period.

[13] We also considered an alternative forecasting approach based on forecast encompassing (Appendix B.3). Our findings, reported in Appendix D, are qualitatively similar to our benchmark case involving all the candidate predictors and point to superiority of our quantile forecasting methods.

els, along with their MCS $p-$values and the associated R^2_{OS} values. Overall, 21 models appear superior and belong to the MCS with associated $p-$values greater than 0.958. As such, we are more that 95% sure that this set of models are superior to the ones excluded. Several interesting findings emerge from this set of results. First, neither the benchmark historical mean, nor the linear single-variable or subset models are included in MCS. This is evidence of the superiority of our methodology over the methodologies of EGT and RSZ. Second, no single-variable subsets of models ($k = 1$) appear in MCS, which provides evidence in favor of our framework over the linear RSZ and the MPVV one. Third, only quantile subset models with $k = 2$ and 3 are included and the associated R^2_{OS} values are greater than 5%. Fourth, our QFC approach ranks first as two thirds of the optimal models belong to this approach as opposed to one third from the RFC method. Similarly, the fixed weighting schemes appear superior to the time-varying ones. However, the best performing model is the RFC-TVW1 with the median combining scheme for $k = 2$. Finally, in more than half of the cases, the associated superior forecasts are generated with the Median combining scheme followed by the Trimmed mean and DALFE/DMSFE (0.5).

[TABLE 8 AROUND HERE]

4.5 Explaining the Benefits of the RFC and QFC Forecasts

In order to show the benefits of our subset quantile forecasts (RFC and QFC specifications), we decompose the MSFE of all forecasting models into two parts: the forecast variance and the squared forecast bias, along the lines of EGT, Lima and Meng (2017) and RSZ. The MSFE is calculated as $\frac{1}{P-P_0}\sum_{t=T_0+P_0}^{T-1}(r_{t+1}-\widehat{r}_{t+1})^2$ and the unconditional forecast variance as $\frac{1}{P-P_0}\sum_{t=T_0+P_0}^{T-1}\left(\widehat{r}_{t+1}-\frac{1}{P-P_0}\sum_{t=T_0+P_0}^{T-1}\widehat{r}_{t+1}\right)^2$ where $P-P_0$ is the total number of out-of-sample forecasts. The squared forecast bias is computed as the difference between MSFE and forecast variance. Figure 3 shows the relative forecast variance and squared forecast bias for the single-predictor models (green points in the scatter plot), the Historical Average (purple point in the scatter plot with label HA), the mean combination linear model ($k = 1$) of RSZ and subset linear model ($k = 2$) of EGT (orange points in the scatter plot with labels $L1 : L2$), the 21 MCS models (red points in the scatter plot with labels $M1 : M21$) and their analogues for $k = 1$ (blue points in the scatter plot with labels $K1 : K21$). The labels for the MCS models follow the order they are presented in Table 8; for example, $M21$ corresponds to the best model (highest R^2_{OS}), i.e. RFC TVW1 with median combining scheme and $k = 2$, while $K21$ denotes the respective $k = 1$ model. The relative forecast variance (squared bias) is calculated as the difference between the forecast variance (squared bias) of the ith model and the forecast variance (squared bias) of the benchmark historical average model. Points along the dotted straight line show model forecasts with the same MSFE as the HA benchmark, while points to the left (right) of the line show models that are superior (inferior) to the HA benchmark. Overall, a model exhibits strong predictability if it can produce forecasts in which the reduction in bias is greater than the increase in variance, relative to the HA forecast (a relation well-known in the literature as the bias-variance trade-off). It is clear that our proposed models in the cluster $M11$, $M13$, $M16$, $M20$, $M21$ significantly outperform the $k = 1$ mean combination of RSZ, the $k = 2$ subset combination of EGT and the respective analogues for $k = 1$ proposed by MPVV, as they substantially reduce the squared forecast bias at the expense of a very small increase in forecast variance. These models are QFC FW3 Median $k = 2$, QFC FW1 Median $k = 2$, QFC TVW2 Median $k = 2$, QFC TVW1 Median $k = 2$ and RFC TVW1 Median $k = 2$. It is worth noting that a median combining scheme is employed in all these specifications with $k = 2$, a QFC approach is used in four out of five cases and a TV weighting scheme in three out of five cases. The rest of the models identified by MCS form a second cluster which reduces substantially the forecast bias at the expense of a moderate increase in forecast variance. The main message from this plot is that our subset quantile forecasting models identified by MCS yield a sizable reduction in the forecast bias while keeping variance under control. In this way, they show improved forecast accuracy over HA, EGT, MPVV and RSZ.

[FIGURE 1 AROUND HERE]

4.6 Multiple hypothesis testing

Since we use multiple tests and multiple models to judge equity premium forecastability, it is not surprising that we have numerous cases with significant tests at conventional significance levels. The issue of multiple hypothesis testing is quite common in the statistics literature and in particular in applications related to genetics where thousands of genes are tested for specific features; however, it is still relatively new in the finance literature. Recent contributions include Harvey, Liu and Zhu (2016), Harvey and Liu (2018), Rapach, Strauss, Tu and Zhou (2019) who focus mainly on the in sample significance of potential factors/ variables. Multiple testing in an out of sample framework has received far less attention (see McCracken and Sapp (2005)). We follow Rapach, Strauss, Tu and Zhou (2019) and control for multiple testing using the Benjamini and Hochberg (2000) adaptive procedure. More in detail, Benjamini and Hochberg (1995) address multiple testing by controlling for the False Discovery Rate (FDR), i.e. the ratio of false rejections over total rejections. The authors develop a popular step up procedure to control FDR based on adjusted p-values. This procedure works well when the respective $p-$values are either independent or positively (regression) dependent (Benjamini and Yekutieli, 2001). However, when the number of true null hypotheses is less than the rejected null hypothesis ($\pi_0 = \frac{m_0}{m}$), the Benjamini and Hochberg (1995) procedure controls FDR at low levels. To this end, Benjamini and Hochberg (2000) adapt the procedure to the data by adjusting p-values with a conservative estimate of π_0. Simulations in Benjamini, Krieger and Yekutieli (2006) show that this adaptive procedure performs well in terms of FDR control and power, even in the case of dependent $p-$values.

Figure 2 reports our findings when controlling for multiple testing across all 1698 DM and CW outcomes. We show the first 1032 elements of both the unadjusted $p-$values and the related adjusted $p-$values based on the Benjamini and Hochberg (2000) adaptive procedure. Based on the unadjusted $p-$values, there are 22 (897) rejections of the null at the 1% (5%) significance level. When we control for multiple testing via the FDR adjusting the $p-$values, the related rejections of the null are equal to 0 and 909, respectively. Our results indicate that at a significance level of 3.66%, the lines of adjusted and unadjusted $p-$values cross and beyond this level, adjusted values are below the unadjusted ones favoring more of our proposed specifications. For example, at the 10% level, an additional 10 models provide rejections of the null of equal forecasting ability. Overall, our forecasting framework survives multiple-testing control and is not simply an artifact of data mining.

[FIGURE 2 AROUND HERE]

5 Real time Selection of k

Our empirical findings (Section 4) 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 excess returns and predictive variables, in order to produce 'optimal forecasts'. Our likelihood-based (Bayesian) 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 return distribution can be incorporated within our approach. 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 $t+1$, $\hat{r}_{t+1}(\tau)$, 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.[14]

Table 9 reports the out-of-sample performance of the 'optimal' QFC forecasts based both on fixed weighting schemes (FW1-FW3) and time-varying weights (TVW1-TVW3), under both prior specifications considered (i.e. π=1/2 (Panel A) and π=1/3 (Panel B)). The results of Table 9 reveal that our likelihood-based approach to selecting k in real time

[14]To save space, the details of the algorithm are presented in Appendix E. Appendix F plots the selected values of k over the out-of-sample period.

is extremely successful, since the values of R^2_{OS} obtained under all weighting schemes and for all combining methods are very high. Regarding the fixed weighting schemes, the largest R^2_{OS} values are obtained for the Median combining method, being in all cases close to or higher than 4%, with the highest value being equal to 4.46% (for the FW2 scheme, under π=1/2). In accordance with our statistical significance results (Table 6), the DALFE(0.5) method ranks second with R^2_{OS} values very close to those obtained by the Median combining method. It is interesting to note that the results of the recursive k-selection exercise are quite robust across the combining methods considered, apart from the AL Cluster(3) method. Moreover, it appears that the FW2 scheme constantly outperforms the other two schemes of producing robust point forecasts based on fixed weights. Similar findings pertain with respect to our TVW forecasts. More in detail, the largest R^2_{OS} value (4.67%) is obtained for the Median combining method and the TVW2 scheme under π=1/3. Overall, the TVW2 scheme constantly outperforms the other two time-varying weighting schemes and in this framework results are better in the case that the prior probability of inclusion is set to 1/3. This may be attributed to the fact that some very large values of k are selected throughout the holdout period, possibly due to weak likelihood information, especially in the case of π=1/2. Our MCS findings also point to superiority of π=1/3 as several models appear optimal. These are generally linked to the Median combining method and either FW2 or TVW2 weighting schemes.

[TABLE 9 AROUND HERE]

6 Economic Evaluation

Campbell and Thompson (2008) and RSZ suggest that even small predictability gains, in a statistical sense, can give an economically meaningful degree of return predictability providing increased portfolio returns for a mean-variance investor that maximizes expected utility. We follow this utility-based approach within this stylized asset allocation framework in order to rank the performance of competing models in a way that captures the risk return trade-off.

Consider a risk-averse investor who constructs a dynamically rebalanced portfolio consisting of the risk-free asset and one risky asset. Her portfolio choice problem is how to allocate wealth between the safe (risk-free Treasury Bill) and the risky asset (stock market), while risk stems from the uncertainty over the future path of the stock market (both in terms of future returns and the uncertainty surrounding them). This approach involves only one risky asset and as such it can be thought of as a standard exercise of market timing in the stock market. In a mean-variance framework, the solution to the maximization problem of the investor yields the following weight (w_t) on the risky asset

$$w_t = \frac{E_t(r_{t+1})}{\gamma Var_t(r_{t+1})} = \frac{\widehat{r}_{t+1}}{\gamma Var_t(r_{t+1})},$$

where E_t and Var_t denote the conditional expectation and variance operators, r_{t+1} is the equity premium and γ is the Relative Risk Aversion (RRA) coefficient that controls the investor's appetite for risk (Campbell and Viceira, 2002; Campbell and Thompson, 2008; RSZ). The conditional expectation $E_t(r_{t+1})$ of each model is given by the 'optimal' forecast from the specific model, $\widehat{r}_{t+1}$, and the variance, $Var_t(r_{t+1})$ is calculated using four alternative ways. The first method we employ is the ten-year rolling window of quarterly returns ($\widehat{\sigma}^2_{1,t+1}$). The remaining volatility forecasts are constructed using the interval approximation approach of Pearson and Tukey (1965). Specifically, we employ the following approximations to conditional standard deviation based on symmetrical quantiles as follows:

$$\widehat{\sigma}_{2,t+1} = \frac{\widehat{r}_{t+1}(0.99) - \widehat{r}_{t+1}(0.01)}{4.65}, \quad (6)$$

$$\widehat{\sigma}_{3,t+1} = \frac{\widehat{r}_{t+1}(0.975) - \widehat{r}_{t+1}(0.025)}{3.92}, \quad (7)$$

$$\widehat{\sigma}_{4,t+1} = \frac{\widehat{r}_{t+1}(0.95) - \widehat{r}_{t+1}(0.05)}{3.25}. \quad (8)$$

The denominators in the above formulae are based on the central distances between estimated quantiles under Pearson curves which are slightly different from a Gaussian curve. The forecasts for the quantiles of interest are based on the combination of quantile forecasts within the kth complete subset, with the values of k being optimally selected at each point of time employing our proposed selection algorithm. Optimal weights depend on both the conditional mean and variance and as a result on the respective forecasts each model/ specification gives. In this setting the optimally constructed portfolio gross return over the out-of-sample period, $R_{p,t+1}$, is equal to $R_{p,t+1} = w_t \cdot r_{t+1} + R_{f,t}$,where $R_{f,t} = 1 + r_{f,t}$ denotes the gross return on the risk-free asset from period t to $t+1$.[15] Over the forecast evaluation period the investor with initial wealth of W_o realizes an average utility of

$$\overline{U} = \frac{W_o}{(P - P_0)} \left[\sum_{t=0}^{P-P_0-1} (R_{p,t+1}) - \frac{\gamma}{2} \sum_{t=0}^{P-P_0-1} (R_{p,t+1} - \overline{R}_p)^2 \right], \quad (9)$$

where $R_{p,t+1}$ is the gross return on her portfolio at time $t+1$. At any point in time, the investor prefers the predictive model that yields the highest average realized utility.[16]

The economic value of our modeling approaches is assessed by comparing their average utility to the corresponding value obtained under the benchmark prevailing mean model. Our results are reported in the form of the annualized Certainly Equivalent Return (CER), i.e. the return that would leave an investor indifferent between using the prevailing mean forecasts versus the forecasts produced by one of our proposed approaches and is calculated as follows:

$$CER = \Delta\overline{U} = \overline{U^i} - \overline{U^{PM}}, \quad (10)$$

where $\overline{U^i}$ is the average realized utility over the out-of-sample period of any of our competing models/ specifications (i) and $\overline{U^{PM}}$ is the respective value for the prevailing mean (PM) model. If our proposed model does not contain any economic value, CER is negative, while positive values of the CER suggest superior predictive ability against the PM benchmark.

We assume that the investor dynamically rebalances her portfolio (updates the weights) quarterly over the out-of-sample period employing the forecasts given by the QFC ap-

[15] We constrain the portfolio weight on the risky asset to lie between 0% and 150% each month, i.e. $0 \leq w_t \leq 1.5$.

[16] We standardize the investor problem by assuming $W_o = 1$.

proach and our selection algorithm for $\pi = 1/2$ and $\pi = 1/3$. Similarly to Section 4 and 5, the out-of-sample period of evaluation is 1965:1-2017:4 and the benchmark strategy against which we evaluate our forecasts is the PM model. For every model/specification we calculate the CER associated with each strategy calculated from Equation (10) setting RRA (γ) equal to 3. Table 10 reports our findings for the aforementioned prior specifications. Panels A-C and Panels D-F report CER in annualized percentage points for the fixed weighting schemes and the time-varying weighting schemes, respectively under the alternative variance forecasts. The columns labeled σ_1 refer to the rolling variance forecast, while σ_2 to σ_4 refer to the robust subset variance forecasts given by equations (6)-(8). CER_1 and CER_2 refer to the prior specifications of $\pi = 1/2$ and $\pi = 1/3$, respectively.

The most striking feature of Table 10 is the robustness of benefits generated to an investor willing to adopt our modelling approaches which range from 3.58% to an impressive 5.69% per year. More in detail, the maximum CER is attained when the DALFE (0.5) FW2 scheme is employed in conjunction with a robust variance forecast given by (7) and a prior of $\pi = 1/3$, which penalizes large values of subsets. On the other hand, the minimum CER, albeit quite high, is attained under the FW scheme when the AL Cluster (3) FW3 scheme is employed combined with the rolling variance forecast and the same prior specification. Overall, the TVW schemes appear superior to their FW counterparts. The minimum benefits to an investor increase to 4.16% when TVW schemes are employed compared to 3.58% under FW specifications. When comparing the alternative prior specifications, the prior of $\pi = 1/2$ appears superior as it leads to greater gains in the majority of the approaches considered. In accordance with our findings from the statistical evaluation of the forecasts obtained under alternative combination methods, the DALFE(0.5) combining method emerges as the optimal one when FW schemes are considered, while the Median one generates the highest CERs among the TVW schemes. With respect to the alternative conditional variance specifications, we have to note that the proposed robust subset variance forecasts add significant economic value within our asset allocation framework. Further benefits are achieved when either σ_2 or σ_3 (given by equations (6)-(7)) are employed as opposed to σ_4 which employs closer to the central location quantile forecasts.

[TABLE 10 AROUND HERE]

7 Conclusions

In this study we propose a complete subset quantile regression approach to equity premium prediction. Our approach is based on the combination of the quantile forecasts, or the robust point 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 robust and accurate forecasts of the equity premium are constructed as weighted averages of a set of quantile forecasts by employing either fixed or time-varying weighting schemes. An important contribution of this study is the development of a likelihood-based method for selecting the value of k recursively. The proposed algorithm is able to identify the best subset for predicting each quantile of the return distribution in real time, based only on the past history of the data. Then, these 'optimal' quantile forecasts are combined to produce robust equity premium forecasts.

The results of our study are very promising. Our findings suggest that our complete subset quantile regression framework achieves superior predictive performance relative to the historical average benchmark, the combination approach, and the subset linear regression approach, both in statistical and economic evaluation terms. More importantly, our economic evaluation results suggest that a mean-variance investor that adopts our framework can gain sizable benefits which range from 3.58% to an impressive 5.69% per year relative to a naive strategy based on the historical benchmark performance.

An interesting avenue for future research is to extend our framework to a quantile on quantile one (see for example, Gupta, Pierdzioch, Selmi and Wohar, 2018). In this respect, we would be able to capture the entire dependence between the equity premium distribution and the distribution of the candidate predictors, while in the present setting we can only capture the response of equity premium to candidate predictors at various points of its distribution.[17]

References

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

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

[3] Benjamini, Y. and Y. Hochberg. (1995) "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." Journal of the Royal Statistical Society. Series B (Methodological), 57 (1), 289-300.

[4] Benjamini, Y. and Y. Hochberg. (2000) "On the Adaptive Control of the False Discovery Rate in Multiple Testing with Independent Statistics." Journal of Educational and Behavioral Statistics, 25 (1), 60-83.

[5] Benjamini, Y., A. M. Krieger, and D. Yekutieli. (2006) "Adaptive Linear Step-Up Procedures that Control the False Discovery Rate." Biometrika 93 (3), 491-507.

[6] Benjamini, Y. and D. Yekutieli. (2001) "The Control of the False Discovery Rate in Multiple Testing Under Dependency." Annals of Statistics 29 (4), 1165-1188.

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

[8] 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.

[9] Campbell, J. Y., and R. J. Shiller. (1988) "Stock Prices, Earnings, and Expected Dividends." Journal of Finance, 43, 661–76.

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

[11] 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.

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

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

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

[17] We thank an anonymous reviewer for this research suggestion.

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

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

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

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

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

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

[21] 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.

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

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

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

[25] Gupta, R., Pierdzioch, C., Selmi, R., and M. E. Wohar. (2018) "Does partisan conflict predict a reduction in US stock market (realized) volatility? Evidence from a quantile-on-quantile regression model."The North American Journal of Economics and Finance, 43, 87-96.

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

[27] Hansen, P. R., Lunde, A., and J. M. Nason, (2011). "The model confidence set." Econometrica, 79(2), 453-497.

[28] Harvey, C. R., Y. Liu, and H. Zhu, (2016). "... and the Cross-Section of Expected Returns." Review of Financial Studies 29 (1), 5-68.

[29] Harvey, C. R., and Y. Liu, (2018). "Lucky factors." Available at SSRN 2528780.

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

[31] 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.

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

[33] 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.

[34] Lima, L.R and F. Meng (2017) "Out-of-Sample Return Predictability: A Quantile Combination Approach." Journal of Applied Econometrics, 32, 877–895.

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

[36] McCracken, M., and S. Sapp. (2005). "Evaluating the predictability of exchange rates using long horizon regression." Journal of Money, Credit and Banking, 37, 473-494.

[37] Meligkotsidou, L., E. Panopoulou, I. D. Vrontos, and S. D. Vrontos (2014). "A Quantile Regression Approach to Equity Premium Prediction", Journal of Forecasting, 33, 7 558-576.

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

[39] 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.

[40] 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.

[41] Rapach, D., J. Strauss, J. Tu and G. Zhou (2019) "Industry Return Predictability: A Machine Learning Approach." Journal of Financial Data Science, forthcoming.

[42] 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.

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

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

[45] 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.

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

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

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

Table 1. Descriptive Statistics

Variable	*Mean*	*Median*	*Standard Deviation*	*Kurtosis*	*Skewness*	*Min*	*Max*	Q_5	Q_{95}
Equity Premium	0.02	0.03	0.08	4.92	-0.94	-0.31	0.19	-0.12	0.12
D/P	-3.49	-3.46	0.44	2.36	-0.18	-4.49	-2.60	-4.26	-2.81
D/Y	-3.47	-3.45	0.44	2.43	-0.18	-4.50	-2.58	-4.24	-2.80
E/P	-2.76	-2.81	0.45	5.50	-0.65	-4.81	-1.77	-3.39	-2.01
B/M	0.54	0.52	0.25	2.48	0.47	0.13	1.20	0.18	1.02
NTIS	0.01	0.02	0.02	3.94	-1.02	-0.05	0.05	-0.02	0.04
TBL	0.04	0.04	0.03	4.17	0.94	0.00	0.15	0.00	0.09
LTR	0.02	0.01	0.05	6.12	0.99	-0.15	0.24	-0.06	0.10
TMS	0.02	0.02	0.01	3.19	-0.14	-0.04	0.05	0.00	0.04
FDY	0.01	0.01	0.00	8.38	1.91	0.00	0.03	0.00	0.02
DFR	0.00	0.00	0.02	15.70	0.32	-0.12	0.16	-0.03	0.03
INFL	0.01	0.01	0.01	6.49	0.53	-0.04	0.05	0.00	0.03
I/K	0.04	0.03	0.00	2.59	0.39	0.03	0.04	0.03	0.04

Notes: The Table reports the descriptive statistics for the Equity Premium and the predictors used.

Table 2. Out-of-sample performance of univariate regression models

Variable	R^2_{OS}	CW_{pv}	DM_{pv}
D/P	-0.11	0.06	0.52
D/Y	0.02	0.05	0.50
E/P	-1.16	0.31	0.66
B/M	-2.05	0.48	0.81
NTIS	-2.16	0.65	0.91
TBL	-2.00	0.04	0.63
LTR	-0.97	0.27	0.66
TMS	-2.55	0.06	0.71
DFY	-2.52	0.71	0.89
DFR	0.05	0.14	0.49
INFL	0.47	0.22	0.40
I/K	2.55	0.01	0.22

Notes: The Table reports the out-of-sample R^2 statistic with respect to the prevailing mean (PM) benchmark model for the out-of-sample period 1965:1-2017:4. Statistical significance for the R^2_{OS} statistic is based on the p-value of the Clark and West (2007) out-of-sample MSFE-adjusted statistic (CW_{pv}) and the Diebold and Mariano (1995) test (DM_{pv}).

Table 3. Out-of-sample performance of complete subset linear regression models

	Mean			*Median*			*Trimmed Mean*			$DMSFE(1)$			$DMSFE(0.9)$			$DMSFE(0.5)$			$Cluster(2)$			$Cluster(3)$		
	Panel A: Historical Average Benchmark																							
k	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}
1	2.82	0.00	0.01	2.43	0.00	0.00	2.69	0.00	0.01	2.83	0.00	0.01	2.86	0.00	0.02	**3.59**	0.01	0.03	2.40	0.01	0.09	1.19	0.08	0.30
2	**3.84**	0.00	0.05	**4.24**	0.00	0.02	**3.97**	0.00	0.04	3.78	0.00	0.05	**3.86**	0.01	0.05	**4.29**	0.01	0.07	2.27	0.03	0.21	0.20	0.11	0.47
3	3.66	0.01	0.13	**3.83**	0.00	0.12	**3.78**	0.01	0.12	3.56	0.01	0.14	3.70	0.01	0.13	**3.94**	0.01	0.15	1.60	0.03	0.33	0.32	0.06	0.47
4	2.78	0.01	0.25	2.49	0.01	0.28	2.81	0.01	0.25	2.67	0.01	0.26	2.80	0.01	0.25	3.00	0.01	0.25	1.22	0.03	0.39	0.08	0.04	0.49
5	1.52	0.01	0.37	1.01	0.02	0.42	1.47	0.01	0.38	1.43	0.02	0.38	1.48	0.02	0.38	1.68	0.02	0.37	0.16	0.03	0.49	-1.10	0.05	0.58
6	0.06	0.02	0.50	-0.53	0.02	0.54	-0.02	0.02	0.50	-0.02	0.02	0.50	-0.09	0.02	0.51	0.13	0.02	0.49	-1.43	0.04	0.59	-2.94	0.05	0.67
7	-1.58	0.03	0.60	-2.44	0.03	0.65	-1.68	0.03	0.61	-1.66	0.03	0.61	-1.84	0.03	0.61	-1.59	0.03	0.59	-3.41	0.05	0.68	-4.90	0.06	0.74
8	-3.45	0.04	0.69	-4.09	0.04	0.72	-3.57	0.04	0.70	-3.54	0.04	0.70	-3.80	0.04	0.70	-3.49	0.04	0.68	-5.76	0.06	0.76	-6.73	0.07	0.78
9	-5.67	0.05	0.77	-5.43	0.04	0.76	-5.85	0.05	0.78	-5.77	0.05	0.77	-6.06	0.05	0.77	-5.65	0.05	0.75	-8.03	0.07	0.81	-8.83	0.07	0.82
10	-8.38	0.06	0.83	-10.66	0.08	0.86	-8.65	0.06	0.84	-8.47	0.06	0.84	-8.73	0.06	0.83	-8.22	0.06	0.82	-11.36	0.08	0.86	-10.65	0.07	0.85
11	-11.76	0.07	0.89	-13.89	0.08	0.90	-12.24	0.07	0.89	-11.82	0.07	0.89	-11.97	0.08	0.88	-11.44	0.07	0.87	-13.82	0.09	0.90	-13.86	0.10	0.90
12	-15.99	0.09	0.93																					
	Panel B: $k = 1$ Benchmark																							
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2	1.05	0.08	0.18	1.86	0.03	0.09	1.31	0.06	0.15	0.97	0.09	0.20	1.03	0.09	0.18	0.73	0.15	0.27	-0.13	0.37	0.54	-1.00	0.55	0.75
3	0.86	0.11	0.34	1.44	0.05	0.30	1.11	0.09	0.32	0.75	0.12	0.36	0.86	0.11	0.34	0.36	0.18	0.43	-0.82	0.35	0.63	-0.88	0.34	0.65
4	-0.05	0.13	0.51	0.06	0.08	0.49	0.12	0.11	0.49	-0.17	0.14	0.52	-0.06	0.14	0.51	-0.61	0.20	0.58	-1.21	0.26	0.64	-1.13	0.23	0.63
5	-1.34	0.16	0.64	-1.45	0.10	0.63	-1.25	0.14	0.62	-1.44	0.17	0.65	-1.42	0.17	0.65	-1.98	0.23	0.70	-2.30	0.25	0.70	-2.33	0.24	0.70
6	-2.84	0.19	0.74	-3.03	0.13	0.72	-2.79	0.16	0.72	-2.94	0.20	0.74	-3.03	0.20	0.74	-3.59	0.26	0.78	-3.92	0.27	0.77	-4.18	0.27	0.78
7	-4.53	0.21	0.81	-4.99	0.15	0.80	-4.49	0.18	0.80	-4.62	0.22	0.81	-4.84	0.23	0.81	-5.37	0.29	0.84	-5.95	0.30	0.83	-6.17	0.29	0.83
8	-6.46	0.23	0.86	-6.68	0.16	0.85	-6.44	0.20	0.85	-6.56	0.24	0.86	-6.86	0.25	0.86	-7.34	0.31	0.88	-8.36	0.33	0.87	-8.02	0.29	0.86
9	-8.74	0.25	0.89	-8.06	0.14	0.86	-8.78	0.22	0.89	-8.85	0.26	0.90	-9.18	0.27	0.89	-9.59	0.33	0.91	-10.68	0.33	0.89	-10.14	0.30	0.88
10	-11.53	0.27	0.92	-13.41	0.22	0.92	-11.65	0.24	0.92	-11.63	0.28	0.92	-11.94	0.29	0.92	-12.25	0.35	0.93	-14.09	0.38	0.93	-11.98	0.29	0.90
11	-15.00	0.29	0.95	-16.72	0.21	0.94	-15.35	0.26	0.94	-15.08	0.30	0.95	-15.27	0.31	0.95	-15.59	0.37	0.95	-16.62	0.38	0.95	-15.24	0.35	0.94
12	-19.36	0.32	0.97																					

Notes: The Table reports the out-of-sample R^2 statistic with respect to the prevailing mean (PM) benchmark model and the $k = 1$ benchmark for the out-of-sample period 1965:1-2017:4. Statistical significance for the R^2_{OS} statistic is based on the p-value of the Clark and West (2007) out-of-sample MSFE-adjusted statistic (CW_{pv}) and the Diebold and Mariano (1995) test (DM_{pv}). Bold indicates that the model belongs to MCS.

Table 4. Out-of-sample performance of Robust Forecast Combination (RFC) approach-Fixed weighting (FW) schemes

	Mean			*Median*			*Trimmed Mean*			$DMSFE(1)$			$DMSFE(0.9)$			$DMSFE(0.5)$			$Cluster(2)$			$Cluster(3)$		
k	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}
Panel A: RFC-FW1 (Historical Average Benchmark)																								
1	2.40	0.01	0.04	1.84	0.02	0.07	2.19	0.01	0.05	2.50	0.01	0.04	2.56	0.01	0.04	**3.56**	0.01	0.04	2.76	0.01	0.05	1.71	0.06	0.22
2	**4.47**	0.00	0.03	**4.88**	0.00	0.01	**4.63**	0.00	0.02	**4.47**	0.00	0.03	**4.50**	0.00	0.03	**4.85**	0.00	0.06	3.33	0.01	0.13	2.27	0.03	0.25
3	**5.03**	0.00	0.06	**5.34**	0.00	0.06	**5.18**	0.00	0.06	**4.97**	0.00	0.07	**5.03**	0.00	0.07	**5.04**	0.00	0.11	3.56	0.01	0.17	2.58	0.02	0.25
4	**4.66**	0.00	0.13	**4.61**	0.00	0.15	**4.73**	0.00	0.13	**4.57**	0.00	0.13	**4.66**	0.00	0.13	**4.55**	0.00	0.17	3.29	0.01	0.22	2.77	0.01	0.26
5	**3.67**	0.00	0.22	3.49	0.00	0.24	**3.66**	0.00	0.22	**3.58**	0.00	0.23	3.67	0.00	0.22	3.52	0.00	0.25	2.84	0.01	0.28	1.85	0.01	0.36
6	2.41	0.00	0.33	2.12	0.00	0.35	2.36	0.00	0.33	2.33	0.00	0.33	2.39	0.00	0.33	2.24	0.01	0.35	1.77	0.01	0.38	1.15	0.01	0.42
7	1.06	0.01	0.43	0.38	0.01	0.47	1.01	0.01	0.43	0.98	0.01	0.43	1.01	0.01	0.43	0.90	0.01	0.44	0.46	0.01	0.47	-0.27	0.01	0.52
Panel B: RFC-FW2 (Historical Average Benchmark)																								
1	2.29	0.01	0.06	1.47	0.03	0.14	2.13	0.01	0.07	2.38	0.01	0.06	2.41	0.01	0.06	**3.30**	0.01	0.06	2.66	0.01	0.06	0.82	0.13	0.37
2	**4.30**	0.00	0.04	**4.65**	0.00	0.01	**4.48**	0.00	0.03	**4.32**	0.00	0.04	**4.34**	0.00	0.04	**4.70**	0.00	0.08	**3.44**	0.02	0.13	1.82	0.05	0.30
3	**4.95**	0.00	0.07	**5.20**	0.00	0.06	**5.09**	0.00	0.07	**4.90**	0.00	0.07	**4.95**	0.00	0.08	**4.94**	0.00	0.12	**4.06**	0.01	0.14	2.80	0.02	0.23
4	**4.63**	0.00	0.13	**4.69**	0.00	0.14	**4.69**	0.00	0.13	**4.56**	0.00	0.14	**4.65**	0.00	0.14	**4.52**	0.00	0.18	**3.87**	0.01	0.19	3.10	0.01	0.24
5	**3.71**	0.00	0.22	**3.65**	0.00	0.24	**3.68**	0.00	0.23	**3.64**	0.00	0.23	3.75	0.00	0.23	**3.61**	0.00	0.26	3.40	0.00	0.25	2.48	0.01	0.32
6	2.49	0.00	0.32	2.16	0.00	0.35	2.42	0.00	0.33	2.43	0.00	0.33	2.53	0.00	0.33	2.43	0.00	0.34	2.45	0.01	0.33	1.68	0.01	0.39
7	1.02	0.00	0.43	0.34	0.01	0.48	0.96	0.00	0.44	0.96	0.00	0.44	1.05	0.01	0.43	1.02	0.01	0.44	0.89	0.01	0.45	0.46	0.01	0.47
Panel C: RFC-FW3 (Historical Average Benchmark)																								
1	2.60	0.00	0.02	1.99	0.01	0.04	2.35	0.01	0.03	2.68	0.00	0.02	2.71	0.00	0.02	**3.64**	0.01	0.03	2.38	0.02	0.08	2.08	0.05	0.18
2	**4.58**	0.00	0.02	**4.98**	0.00	0.01	**4.73**	0.00	0.02	**4.56**	0.00	0.02	**4.57**	0.00	0.03	**4.92**	0.00	0.05	2.90	0.02	0.16	2.00	0.04	0.27
3	**5.07**	0.00	0.06	**5.24**	0.00	0.06	**5.22**	0.00	0.05	**4.98**	0.00	0.06	**5.02**	0.00	0.07	**5.08**	0.00	0.10	3.11	0.01	0.19	2.09	0.02	0.29
4	**4.62**	0.00	0.12	**4.50**	0.00	0.14	**4.69**	0.00	0.12	**4.52**	0.00	0.13	**4.59**	0.00	0.13	**4.55**	0.00	0.16	3.04	0.01	0.23	2.41	0.01	0.29
5	**3.60**	0.00	0.22	3.27	0.00	0.25	**3.58**	0.00	0.22	**3.50**	0.00	0.23	3.56	0.00	0.23	3.50	0.00	0.25	2.43	0.01	0.31	1.45	0.01	0.39
6	2.32	0.01	0.33	2.00	0.01	0.35	2.27	0.01	0.33	2.23	0.01	0.33	2.26	0.01	0.34	2.23	0.01	0.35	1.29	0.01	0.41	0.60	0.01	0.46
7	0.99	0.01	0.43	0.37	0.01	0.47	0.93	0.01	0.44	0.91	0.01	0.44	0.90	0.01	0.44	0.91	0.01	0.44	0.02	0.01	0.50	-0.79	0.02	0.55
Panel D: RFC-FW1 ($k = 1$ Benchmark)																								
k	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2	2.12	0.01	0.04	3.10	0.00	0.02	2.50	0.01	0.03	2.03	0.02	0.05	1.98	0.02	0.05	1.34	0.07	0.16	0.59	0.19	0.36	0.57	0.17	0.36
3	2.69	0.02	0.12	3.56	0.01	0.12	3.06	0.02	0.11	2.54	0.02	0.13	2.53	0.02	0.13	1.53	0.06	0.25	0.83	0.10	0.37	0.89	0.10	0.35
4	2.31	0.02	0.23	2.82	0.01	0.24	2.60	0.02	0.22	2.13	0.03	0.25	2.15	0.03	0.24	1.02	0.07	0.37	0.55	0.08	0.43	1.08	0.07	0.36
5	1.30	0.03	0.37	1.67	0.01	0.36	1.50	0.03	0.36	1.11	0.03	0.38	1.14	0.04	0.38	-0.05	0.08	0.50	0.09	0.07	0.49	0.14	0.08	0.49
6	0.01	0.04	0.50	0.28	0.02	0.48	0.18	0.03	0.48	-0.17	0.04	0.52	-0.18	0.04	0.52	-1.37	0.09	0.62	-1.01	0.07	0.58	-0.57	0.07	0.55
7	-1.37	0.05	0.61	-1.49	0.02	0.61	-1.20	0.04	0.59	-1.55	0.05	0.62	-1.59	0.05	0.62	-2.76	0.10	0.71	-2.36	0.08	0.67	-2.02	0.08	0.64
Panel E: RFC-FW2 ($k = 1$ Benchmark)																								
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2	2.06	0.02	0.05	3.22	0.00	0.02	2.40	0.01	0.05	1.99	0.02	0.06	1.97	0.02	0.06	1.45	0.07	0.15	0.81	0.17	0.32	1.01	0.11	0.26
3	2.72	0.02	0.12	3.78	0.01	0.11	3.02	0.02	0.12	2.58	0.02	0.13	2.60	0.02	0.13	1.70	0.06	0.24	1.44	0.08	0.29	2.00	0.04	0.18
4	2.40	0.02	0.23	3.26	0.01	0.21	2.61	0.02	0.23	2.23	0.03	0.24	2.29	0.03	0.24	1.26	0.06	0.35	1.25	0.06	0.35	2.30	0.02	0.21
5	1.45	0.03	0.36	2.22	0.01	0.32	1.58	0.03	0.35	1.29	0.03	0.37	1.37	0.03	0.36	0.32	0.07	0.47	0.77	0.05	0.42	1.68	0.03	0.33
6	0.20	0.04	0.48	0.70	0.01	0.45	0.29	0.03	0.48	0.05	0.04	0.50	0.12	0.04	0.49	-0.90	0.07	0.58	-0.21	0.05	0.52	0.87	0.03	0.43
7	-1.31	0.04	0.60	-1.15	0.02	0.58	-1.20	0.04	0.59	-1.45	0.05	0.61	-1.40	0.05	0.60	-2.36	0.08	0.67	-1.82	0.06	0.63	-0.36	0.04	0.53
Panel F: RFC-FW3 ($k = 1$ Benchmark)																								
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2	2.03	0.01	0.04	3.05	0.00	0.02	2.44	0.01	0.04	1.93	0.02	0.05	1.90	0.02	0.06	1.33	0.07	0.15	0.53	0.20	0.37	-0.08	0.30	0.52
3	2.53	0.02	0.13	3.31	0.01	0.14	2.94	0.02	0.12	2.37	0.03	0.14	2.37	0.03	0.14	1.49	0.07	0.25	0.75	0.12	0.38	0.01	0.20	0.50
4	2.07	0.03	0.25	2.56	0.01	0.26	2.40	0.02	0.24	1.89	0.03	0.27	1.92	0.03	0.26	0.95	0.07	0.38	0.68	0.08	0.41	0.34	0.11	0.46
5	1.02	0.04	0.39	1.30	0.02	0.39	1.26	0.03	0.38	0.85	0.04	0.41	0.87	0.04	0.41	-0.14	0.09	0.52	0.05	0.08	0.49	-0.65	0.11	0.57
6	-0.28	0.05	0.53	0.00	0.02	0.50	-0.08	0.04	0.51	-0.45	0.05	0.54	-0.46	0.06	0.54	-1.47	0.10	0.63	-1.11	0.09	0.59	-1.51	0.11	0.62
7	-1.65	0.06	0.63	-1.66	0.03	0.62	-1.45	0.05	0.61	-1.82	0.06	0.64	-1.86	0.07	0.64	-2.84	0.12	0.72	-2.42	0.09	0.67	-2.93	0.11	0.70

Notes: See Notes in Table 3.

Table 5. Out-of-sample performance of Robust Forecast Combination (RFC) approach-Time-varying weighting (TVW) schemes

	Historical Average Benchmark									$k=1$ Benchmark								
	Panel A: RFC-TVW1									Panel D: RFC-TVW1								
	Mean			*Median*			*Trimmed Mean*			*Mean*			*Median*			*Trimmed Mean*		
k	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}
1	3.42	0.00	0.01	2.52	0.00	0.01	3.26	0.00	0.00	-	-	-	-	-	-	-	-	-
2	**4.79**	0.00	0.02	**5.64**	0.00	0.00	**4.94**	0.00	0.02	1.41	0.04	0.09	3.20	0.00	0.01	1.74	0.03	0.08
3	**4.84**	0.00	0.07	**4.96**	0.00	0.07	**4.97**	0.00	0.06	1.46	0.06	0.24	2.50	0.01	0.18	1.77	0.04	0.22
4	4.14	0.00	0.15	3.81	0.00	0.18	4.22	0.00	0.15	0.75	0.07	0.40	1.32	0.02	0.36	0.99	0.05	0.37
5	2.94	0.01	0.26	2.65	0.01	0.29	2.95	0.01	0.26	-0.50	0.09	0.56	0.14	0.03	0.49	-0.32	0.07	0.53
6	1.52	0.01	0.38	1.37	0.01	0.40	1.49	0.01	0.39	-1.97	0.10	0.68	-1.18	0.03	0.60	-1.82	0.08	0.66
7	0.09	0.01	0.49	-0.25	0.01	0.52	0.07	0.01	0.50	-3.46	0.11	0.76	-2.85	0.04	0.70	-3.29	0.09	0.75
	Panel B: RFC-TVW2									Panel E: RFC-TVW2								
1	3.44	0.00	0.00	2.40	0.00	0.00	3.26	0.00	0.00	-	-	-	-	-	-	-	-	-
2	**4.72**	0.00	0.02	**4.97**	0.00	0.01	**4.89**	0.00	0.02	1.33	0.05	0.12	2.63	0.01	0.05	1.68	0.04	0.10
3	**4.92**	0.00	0.07	**4.94**	0.00	0.07	**5.06**	0.00	0.06	1.54	0.06	0.24	2.60	0.02	0.19	1.86	0.04	0.22
4	**4.33**	0.00	0.15	4.23	0.00	0.16	**4.39**	0.00	0.15	0.92	0.06	0.38	1.87	0.02	0.32	1.17	0.05	0.36
5	3.22	0.00	0.25	3.14	0.00	0.26	3.21	0.00	0.25	-0.23	0.07	0.52	0.75	0.02	0.44	-0.06	0.05	0.51
6	1.87	0.01	0.37	1.69	0.01	0.38	1.83	0.01	0.37	-1.63	0.08	0.64	-0.73	0.02	0.56	-1.48	0.06	0.62
7	0.29	0.01	0.48	-0.31	0.01	0.52	0.29	0.01	0.48	-3.26	0.08	0.74	-2.78	0.03	0.69	-3.07	0.07	0.72
	Panel C: RFC-TVW3									Panel F: RFC-TVW3								
1	3.06	0.01	0.04	2.13	0.02	0.10	2.72	0.01	0.05	-	-	-	-	-	-	-	-	-
2	4.35	0.00	0.03	**4.66**	0.00	0.02	4.41	0.00	0.03	1.32	0.04	0.11	2.59	0.00	0.02	1.74	0.02	0.07
3	4.46	0.00	0.08	4.38	0.00	0.09	**4.57**	0.00	0.08	1.44	0.05	0.24	2.30	0.01	0.18	1.90	0.03	0.20
4	3.75	0.01	0.17	3.26	0.01	0.21	3.82	0.01	0.16	0.71	0.06	0.40	1.16	0.02	0.37	1.13	0.04	0.36
5	2.61	0.01	0.28	2.18	0.01	0.32	2.60	0.01	0.28	-0.47	0.07	0.55	0.05	0.03	0.50	-0.12	0.05	0.51
6	1.24	0.01	0.40	0.99	0.01	0.42	1.20	0.01	0.41	-1.88	0.08	0.67	-1.16	0.03	0.60	-1.56	0.06	0.64
7	-0.18	0.01	0.51	-0.69	0.02	0.55	-0.21	0.01	0.51	-3.34	0.09	0.76	-2.88	0.03	0.71	-3.01	0.06	0.73

Notes: See Notes in Table 3.

Table 6. Out-of-sample performance of Quantile Forecast Combination (QFC) approach-Fixed weighting (FW) schemes

	Mean			*Median*			*Trimmed Mean*			$DALFE(1)$			$DALFE(0.9)$			$DALFE(0.5)$			*AL Cluster*(2)			*AL Cluster*(3)		
k	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}
Panel A: QFC -FW1 (Historical Average Benchmark)																								
1	2.40	0.01	0.04	1.65	0.02	0.09	2.14	0.01	0.05	2.45	0.01	0.04	2.49	0.01	0.04	3.09	0.01	0.04	2.45	0.01	0.07	2.69	0.01	0.08
2	4.47	0.00	0.03	**5.10**	0.00	0.01	**4.62**	0.00	0.02	4.47	0.00	0.03	4.52	0.00	0.03	**4.89**	0.00	0.04	3.82	0.01	0.08	3.15	0.01	0.15
3	**5.03**	0.00	0.06	**5.44**	0.00	0.05	**5.20**	0.00	0.06	**5.00**	0.00	0.06	**5.08**	0.00	0.07	**5.25**	0.00	0.08	**4.64**	0.00	0.10	3.87	0.01	0.15
4	**4.66**	0.00	0.13	**4.68**	0.00	0.14	**4.76**	0.00	0.12	4.61	0.00	0.13	**4.70**	0.00	0.13	**4.76**	0.00	0.14	4.28	0.00	0.16	3.67	0.01	0.21
5	3.67	0.00	0.22	3.72	0.00	0.22	3.71	0.00	0.22	3.62	0.00	0.22	3.70	0.00	0.22	3.72	0.00	0.23	3.32	0.00	0.25	2.74	0.01	0.30
6	2.41	0.00	0.33	2.29	0.00	0.33	2.39	0.00	0.33	2.36	0.00	0.33	2.42	0.00	0.33	2.41	0.00	0.33	1.90	0.01	0.37	1.33	0.01	0.41
7	1.06	0.01	0.43	0.73	0.01	0.45	1.03	0.01	0.43	1.02	0.01	0.43	1.04	0.01	0.43	1.03	0.01	0.43	0.50	0.01	0.47	-0.07	0.01	0.50
Panel B: QFC -FW2 (Historical Average Benchmark)																								
1	2.29	0.01	0.06	1.28	0.05	0.17	2.02	0.01	0.08	2.32	0.01	0.06	2.37	0.01	0.06	2.91	0.01	0.05	2.33	0.01	0.10	3.21	0.01	0.09
2	4.30	0.00	0.04	**4.89**	0.00	0.01	4.43	0.00	0.03	4.30	0.00	0.04	4.35	0.00	0.04	**4.72**	0.00	0.05	3.72	0.01	0.10	3.23	0.01	0.15
3	**4.95**	0.00	0.07	**5.56**	0.00	0.05	**5.10**	0.00	0.07	**4.92**	0.00	0.07	**4.99**	0.00	0.07	**5.12**	0.00	0.09	4.54	0.00	0.11	4.01	0.01	0.15
4	**4.63**	0.00	0.13	**4.77**	0.00	0.14	**4.71**	0.00	0.13	**4.60**	0.00	0.14	**4.69**	0.00	0.14	**4.68**	0.00	0.16	4.50	0.00	0.15	3.93	0.00	0.20
5	3.71	0.00	0.22	3.76	0.00	0.23	3.72	0.00	0.22	3.68	0.00	0.23	3.78	0.00	0.22	3.73	0.00	0.24	3.77	0.00	0.23	3.29	0.00	0.27
6	2.49	0.00	0.32	2.36	0.00	0.34	2.46	0.00	0.33	2.46	0.00	0.33	2.56	0.00	0.32	2.51	0.00	0.33	2.68	0.00	0.32	2.20	0.00	0.36
7	1.02	0.00	0.43	0.62	0.01	0.46	1.00	0.00	0.43	0.99	0.00	0.43	1.08	0.00	0.43	1.08	0.00	0.43	1.20	0.00	0.43	0.74	0.01	0.46
Panel C: QFC -FW3 (Historical Average Benchmark)																								
1	2.60	0.00	0.02	1.81	0.01	0.05	2.32	0.01	0.02	2.64	0.00	0.02	2.65	0.00	0.02	3.24	0.00	0.02	2.50	0.01	0.06	2.67	0.01	0.07
2	**4.58**	0.00	0.02	**5.09**	0.00	0.01	**4.71**	0.00	0.02	**4.57**	0.00	0.02	**4.59**	0.00	0.02	**4.92**	0.00	0.04	3.82	0.01	0.07	3.10	0.01	0.13
3	**5.07**	0.00	0.06	**5.43**	0.00	0.05	**5.24**	0.00	0.05	**5.03**	0.00	0.06	**5.08**	0.00	0.06	**5.24**	0.00	0.07	4.52	0.00	0.09	3.71	0.01	0.15
4	**4.62**	0.00	0.12	**4.65**	0.00	0.13	**4.73**	0.00	0.12	4.57	0.00	0.13	**4.63**	0.00	0.13	**4.66**	0.00	0.14	4.10	0.00	0.16	3.47	0.01	0.21
5	3.60	0.00	0.22	3.62	0.00	0.22	3.63	0.00	0.22	3.54	0.00	0.22	3.58	0.00	0.22	3.54	0.00	0.24	3.08	0.01	0.26	2.48	0.01	0.31
6	2.32	0.01	0.33	2.16	0.01	0.34	2.30	0.01	0.33	2.27	0.01	0.33	2.27	0.01	0.33	2.20	0.01	0.35	1.62	0.01	0.38	1.05	0.01	0.43
7	0.99	0.01	0.43	0.62	0.01	0.46	0.95	0.01	0.43	0.95	0.01	0.43	0.91	0.01	0.44	0.82	0.01	0.45	0.25	0.01	0.48	-0.37	0.01	0.52
Panel D: QFC -FW1 ($k = 1$ Benchmark)																								
k	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2	2.12	0.01	0.04	3.51	0.00	0.01	2.53	0.01	0.03	2.07	0.01	0.04	2.08	0.02	0.05	1.85	0.03	0.08	1.40	0.08	0.17	0.47	0.19	0.38
3	2.69	0.02	0.12	3.85	0.01	0.10	3.12	0.01	0.11	2.62	0.02	0.12	2.66	0.02	0.12	2.23	0.03	0.17	2.24	0.04	0.18	1.22	0.08	0.31
4	2.31	0.02	0.23	3.07	0.01	0.22	2.68	0.02	0.21	2.22	0.02	0.24	2.27	0.03	0.24	1.73	0.04	0.29	1.88	0.04	0.28	1.01	0.07	0.38
5	1.30	0.03	0.37	2.10	0.01	0.32	1.60	0.02	0.35	1.20	0.03	0.38	1.24	0.03	0.37	0.65	0.05	0.43	0.89	0.04	0.41	0.05	0.07	0.50
6	0.01	0.04	0.50	0.65	0.02	0.45	0.25	0.03	0.48	-0.09	0.04	0.51	-0.08	0.04	0.51	-0.70	0.06	0.56	-0.57	0.05	0.55	-1.39	0.08	0.61
7	-1.37	0.05	0.61	-0.94	0.02	0.57	-1.14	0.04	0.59	-1.47	0.05	0.62	-1.49	0.05	0.62	-2.12	0.07	0.66	-2.00	0.06	0.64	-2.83	0.08	0.68
Panel E: QFC -FW2 ($k = 1$ Benchmark)																								
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2	2.06	0.02	0.05	3.66	0.00	0.02	2.46	0.01	0.04	2.03	0.02	0.05	2.03	0.02	0.06	1.87	0.03	0.09	1.42	0.09	0.18	0.02	0.28	0.49
3	2.72	0.02	0.12	4.33	0.00	0.08	3.14	0.01	0.11	2.66	0.02	0.13	2.69	0.02	0.13	2.28	0.04	0.18	2.26	0.04	0.18	0.83	0.09	0.36
4	2.40	0.02	0.23	3.54	0.01	0.19	2.75	0.02	0.22	2.33	0.02	0.23	2.38	0.02	0.23	1.83	0.04	0.29	2.22	0.03	0.25	0.74	0.06	0.41
5	1.45	0.03	0.36	2.51	0.01	0.30	1.74	0.02	0.34	1.39	0.03	0.36	1.44	0.03	0.36	0.85	0.04	0.42	1.47	0.03	0.36	0.08	0.06	0.49
6	0.20	0.04	0.48	1.09	0.01	0.42	0.46	0.03	0.46	0.14	0.04	0.49	0.19	0.04	0.48	-0.40	0.05	0.53	0.36	0.03	0.47	-1.04	0.06	0.58
7	-1.31	0.04	0.60	-0.67	0.02	0.55	-1.04	0.03	0.58	-1.37	0.04	0.60	-1.32	0.04	0.60	-1.88	0.06	0.64	-1.15	0.04	0.58	-2.56	0.06	0.66
Panel F: QFC -FW3 ($k = 1$ Benchmark)																								
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2	2.03	0.01	0.04	3.35	0.00	0.02	2.45	0.01	0.03	1.98	0.02	0.05	1.99	0.02	0.05	1.73	0.04	0.09	1.36	0.07	0.16	0.44	0.20	0.38
3	2.53	0.02	0.13	3.69	0.01	0.10	2.99	0.02	0.11	2.45	0.02	0.14	2.49	0.02	0.13	2.07	0.04	0.19	2.06	0.04	0.19	1.07	0.10	0.33
4	2.07	0.03	0.25	2.89	0.01	0.23	2.47	0.02	0.23	1.98	0.03	0.26	2.03	0.03	0.26	1.46	0.05	0.32	1.63	0.04	0.30	0.82	0.08	0.40
5	1.02	0.04	0.39	1.85	0.02	0.34	1.34	0.03	0.37	0.93	0.04	0.40	0.95	0.04	0.40	0.31	0.06	0.47	0.59	0.05	0.44	-0.20	0.09	0.52
6	-0.28	0.05	0.53	0.36	0.02	0.47	-0.02	0.04	0.50	-0.37	0.05	0.53	-0.39	0.05	0.54	-1.08	0.08	0.60	-0.90	0.07	0.58	-1.66	0.10	0.63
7	-1.65	0.06	0.63	-1.21	0.03	0.59	-1.40	0.05	0.61	-1.74	0.06	0.64	-1.80	0.06	0.64	-2.51	0.09	0.69	-2.31	0.08	0.67	-3.12	0.10	0.71

Notes: See Notes in Table 3.

Table 7. Out-of-sample performance of Quantile Forecast Combination (QFC) approach-Time-varying weighting (TVW) schemes

Panel A: QFC-TVW1 (Historical Average Benchmark)										Panel D: QFC-TVW1 ($k=1$ Benchmark)									
	Mean			*Median*			*Trimmed Mean*				*Mean*			*Median*			*Trimmed Mean*		
k	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}		R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}
1	3.68	0.00	0.01	3.14	0.00	0.00	3.42	0.00	0.01	1	-	-	-	-	-	-	-	-	-
2	**4.83**	0.00	0.03	**5.58**	0.00	0.01	**5.03**	0.00	0.02	2	1.20	0.06	0.14	2.52	0.01	0.04	1.67	0.04	0.09
3	4.51	0.00	0.09	**4.82**	0.00	0.08	**4.73**	0.00	0.08	3	0.87	0.10	0.34	1.74	0.03	0.27	1.36	0.07	0.28
4	3.84	0.01	0.17	3.89	0.01	0.18	3.95	0.00	0.17	4	0.17	0.10	0.48	0.78	0.04	0.42	0.55	0.07	0.43
5	2.61	0.01	0.29	2.69	0.01	0.29	2.64	0.01	0.29	5	-1.11	0.11	0.62	-0.46	0.05	0.54	-0.80	0.09	0.58
6	1.23	0.01	0.41	1.12	0.01	0.42	1.21	0.01	0.41	6	-2.54	0.12	0.72	-2.08	0.06	0.67	-2.29	0.09	0.70
7	-0.25	0.01	0.52	-0.28	0.01	0.52	-0.27	0.01	0.52	7	-4.08	0.12	0.80	-3.52	0.06	0.74	-3.81	0.10	0.77
Panel B:QFC-TVW2 (Historical Average Benchmark)										Panel E: QFC-TVW2 ($k=1$ Benchmark)									
1	3.76	0.00	0.00	3.04	0.00	0.00	3.46	0.00	0.00	1	-	-	-	-	-	-	-	-	-
2	**4.86**	0.00	0.02	**5.41**	0.00	0.01	**5.02**	0.00	0.02	2	1.14	0.07	0.17	2.45	0.02	0.07	1.63	0.04	0.12
3	**4.93**	0.00	0.07	**5.50**	0.00	0.06	**5.13**	0.00	0.07	3	1.21	0.08	0.29	2.54	0.02	0.20	1.74	0.05	0.24
4	4.21	0.00	0.16	**4.46**	0.00	0.15	**4.31**	0.00	0.15	4	0.47	0.09	0.44	1.46	0.03	0.36	0.88	0.06	0.40
5	3.00	0.01	0.27	3.18	0.01	0.26	3.06	0.01	0.27	5	-0.79	0.10	0.58	0.14	0.04	0.49	-0.41	0.07	0.54
6	1.69	0.01	0.38	1.70	0.01	0.38	1.70	0.01	0.38	6	-2.15	0.10	0.69	-1.38	0.04	0.61	-1.82	0.07	0.65
7	0.14	0.01	0.49	-0.35	0.01	0.52	0.16	0.01	0.49	7	-3.76	0.10	0.77	-3.50	0.05	0.72	-3.42	0.08	0.74
Panel C: QFC-TVW3 (Historical Average Benchmark)										Panel F:QFC-TVW3 ($k=1$ Benchmark)									
1	2.82	0.01	0.08	2.34	0.02	0.07	2.59	0.02	0.08	1	-	-	-	-	-	-	-	-	-
2	4.26	0.00	0.05	**4.96**	0.00	0.02	4.47	0.00	0.04	2	1.48	0.03	0.10	2.69	0.01	0.03	1.93	0.02	0.06
3	4.34	0.00	0.09	4.59	0.00	0.09	4.57	0.00	0.08	3	1.56	0.05	0.23	2.31	0.02	0.19	2.03	0.03	0.19
4	3.70	0.01	0.18	3.75	0.01	0.18	3.78	0.01	0.17	4	0.90	0.05	0.38	1.44	0.02	0.34	1.22	0.04	0.35
5	2.37	0.01	0.30	2.59	0.01	0.29	2.37	0.01	0.31	5	-0.47	0.07	0.55	0.26	0.03	0.48	-0.22	0.05	0.52
6	0.90	0.01	0.43	0.92	0.01	0.43	0.88	0.01	0.43	6	-1.98	0.08	0.67	-1.45	0.04	0.62	-1.76	0.06	0.65
7	-0.58	0.01	0.54	-0.68	0.01	0.55	-0.61	0.01	0.54	7	-3.51	0.08	0.76	-3.09	0.04	0.71	-3.29	0.06	0.74

Notes: See Notes in Table 3.

Table 8. Global MCS

Method	*Scheme*	*Combination Scheme*	*Subset k*	R^2_{OS}	*MCSpv*
QFC	TVW2	Trimmed Mean	3	5.13%	0.958
RFC	FW3	DMSFE(0.5)	3	5.08%	0.958
RFC	FW1	Trimmed Mean	3	5.18%	0.958
RFC	FW1	DMSFE(0.5)	3	5.04%	0.958
QFC	FW1	Trimmed Mean	3	5.20%	0.958
QFC	FW2	DALFE(0.5)	3	5.12%	0.958
RFC	FW3	Trimmed Mean	3	5.22%	0.958
QFC	FW3	Trimmed Mean	3	5.24%	0.958
QFC	FW3	DALFE(0.5)	3	5.24%	0.958
RFC	FW3	Median	3	5.24%	0.958
QFC	FW3	Median	2	5.09%	0.958
QFC	FW1	DALFE(0.5)	3	5.25%	0.958
QFC	FW1	Median	2	5.10%	0.958
RFC	FW1	Median	3	5.34%	0.958
QFC	FW3	Median	3	5.43%	0.958
QFC	TVW2	Median	2	5.41%	0.958
QFC	FW1	Median	3	5.44%	0.958
QFC	TVW2	Median	3	5.50%	0.995
QFC	FW2	Median	3	5.56%	0.995
QFC	TVW1	Median	2	5.58%	0.995
RFC	TVW1	Median	2	5.64%	1.000

Notes: The Table reports the models that belong to the Model Confidence Set (MCS), *Method* denotes if the method used is Robust Forecast Combination (RFC) or Quantile Forecast Combination (QFC), *Scheme* denotes the Fixed Weighting (FW) or the Time-Varying Weighting (TVW) scheme employed, *Subset k* denotes the number of predictors used in the subset, R^2_{OS} denotes the out-of-sample R^2 statistic of the specific method and *MCSpv* denotes the associated p-value of the MCS test.

Table 9. Out-of-sample performance of the 'optimal' QFC forecasts

	Mean			*Median*			*Trimmed Mean*			*DALFE*(1)			*DALFE*(0.9)			*DALFE*(0.5)			*AL Cluster*(2)			*AL Cluster*(3)		
	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}	R^2_{OS}	CW_{pv}	DM_{pv}
Panel A: $\pi = 1/2$																								
FW1	3.86	0.00	0.15	**4.20**	0.00	0.14	3.98	0.00	0.15	3.82	0.00	0.16	3.90	0.00	0.16	**4.10**	0.00	0.17	3.42	0.01	0.21	2.67	0.01	0.27
FW2	**4.14**	0.00	0.15	**4.46**	0.00	0.14	**4.22**	0.00	0.15	4.11	0.00	0.15	**4.19**	0.00	0.15	**4.25**	0.00	0.17	3.92	0.00	0.18	3.26	0.01	0.24
FW3	3.73	0.00	0.16	**4.02**	0.00	0.14	3.84	0.00	0.15	3.68	0.01	0.16	3.73	0.01	0.16	3.88	0.01	0.18	3.11	0.01	0.22	2.31	0.02	0.30
TVW1	2.80	0.01	0.23	3.09	0.01	0.21	2.90	0.01	0.22															
TVW2	3.71	0.01	0.17	4.03	0.00	0.16	3.82	0.01	0.17															
TVW3	2.85	0.01	0.22	3.26	0.01	0.19	2.91	0.01	0.22															
Panel B: $\pi = 1/3$																								
FW1	3.12	0.01	0.14	**3.56**	0.01	0.10	**3.24**	0.01	0.13	3.11	0.01	0.15	3.17	0.01	0.15	**3.46**	0.01	0.16	2.87	0.02	0.19	1.77	0.03	0.31
FW2	**3.71**	0.01	0.11	**4.28**	0.00	0.07	**3.81**	0.00	0.11	**3.70**	0.01	0.11	**3.80**	0.01	0.11	**4.06**	0.01	0.13	**3.73**	0.01	0.14	**2.92**	0.01	0.22
FW3	3.17	0.01	0.14	**3.50**	0.01	0.10	**3.29**	0.01	0.12	3.15	0.01	0.14	3.17	0.01	0.14	**3.42**	0.01	0.15	2.75	0.02	0.20	1.69	0.04	0.32
TVW1	2.83	0.02	0.19	**3.37**	0.01	0.14	**3.02**	0.02	0.18															
TVW2	**4.06**	0.01	0.10	**4.67**	0.00	0.06	**4.21**	0.01	0.09															
TVW3	2.68	0.02	0.20	2.85	0.02	0.18	2.72	0.02	0.19															

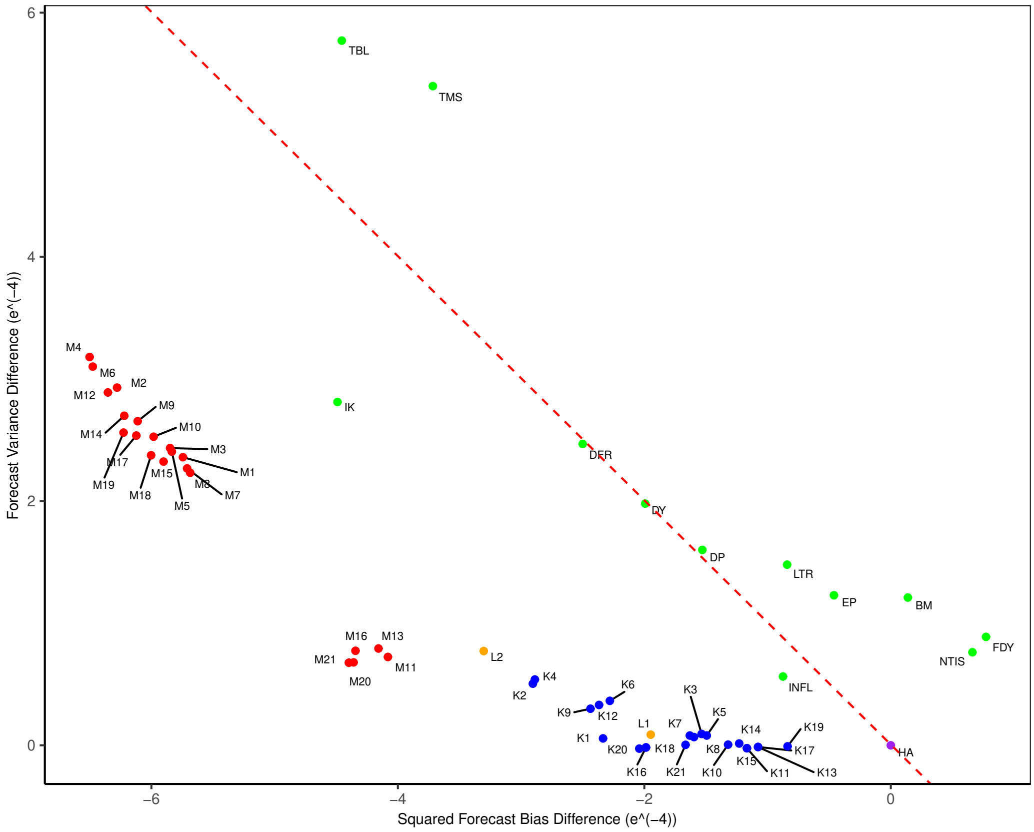
Notes: The Table reports the out-of-sample R^2 statistic of the Quantile Forecast Combination (QFC) approach under fixed (FW) and time-varying weighting (TVW) schemes with respect to the prevailing mean (PM) benchmark model for the out-of-sample period 1965:1-2017:4. Statistical significance for the R^2_{OS} statistic is based on the p-value of the Clark and West (2007) out-of-sample MSFE-adjusted statistic (CW_{pv}) and the Diebold and Mariano (1995) test (DM_{pv}).

Table 10. Economic evaluation of the 'optimal' QFC forecasts

	Panel A: FW1								Panel D: TVW1							
	σ_1		σ_2		σ_3		σ_4		σ_1		σ_2		σ_3		σ_4	
	CER_1	CER_2	CER_1	CER_2	CER_1	CER_2	CER_1	CER_2	CER_1	CER_2	CER_1	CER_2	CER_1	CER_2	CER_1	CER_2
Mean	4.81	4.22	5.22	4.62	5.12	4.82	4.81	4.46	4.57	4.34	4.80	5.06	5.10	5.28	4.91	4.74
Median	4.94	4.45	5.46	4.72	5.46	4.95	5.18	4.48	4.76	4.70	5.26	5.46	5.58	5.49	5.39	5.08
Trimmed Mean	4.84	4.28	5.26	4.59	5.18	4.84	4.79	4.47	4.59	4.42	4.86	5.12	5.11	5.33	4.90	4.81
DALFE(1)	4.80	4.24	5.22	4.73	5.15	4.91	4.81	4.49								
DALFE(0.9)	5.01	4.43	5.43	4.88	5.31	4.96	4.92	4.71								
DALFE(0.5)	5.25	4.86	5.53	5.39	5.44	5.47	5.26	5.36								
AL Cluster(2)	4.67	4.19	5.18	5.05	5.15	5.14	5.12	5.26								
AL Cluster(3)	4.28	3.70	4.81	4.75	5.02	4.98	5.10	5.26								
	Panel B: FW2								Panel E: TVW2							
Mean	4.81	4.56	5.43	4.87	5.35	5.07	4.96	4.93	4.70	4.67	5.24	5.15	5.21	5.43	5.06	5.03
Median	4.94	4.94	5.48	4.89	5.53	5.18	5.15	4.99	4.78	4.92	5.45	5.42	5.36	5.48	5.05	5.08
Trimmed Mean	4.84	4.63	5.47	4.86	5.41	5.09	5.01	4.98	4.72	4.72	5.33	5.22	5.25	5.47	5.05	5.08
DALFE(1)	4.80	4.57	5.41	4.97	5.37	5.12	4.95	4.95								
DALFE(0.9)	5.04	4.78	5.60	5.14	5.54	5.30	5.16	5.19								
DALFE(0.5)	5.16	5.09	5.58	5.54	5.55	5.69	5.33	5.65								
AL Cluster(2)	4.96	4.48	5.51	5.29	5.49	5.48	5.57	5.71								
AL Cluster(3)	4.57	4.09	5.20	5.08	5.44	5.40	5.59	5.67								
	Panel C: FW3								Panel F: TVW3							
Mean	4.70	4.19	5.03	4.55	4.98	4.84	4.78	4.35	4.44	4.16	4.75	5.03	5.10	5.19	4.86	4.55
Median	4.83	4.31	5.36	4.72	5.37	4.82	5.22	4.44	4.66	4.35	5.07	5.17	5.44	5.27	5.23	4.71
Trimmed Mean	4.73	4.23	5.12	4.53	5.03	4.84	4.79	4.38	4.44	4.19	4.77	5.04	5.09	5.20	4.82	4.58
DALFE(1)	4.69	4.20	5.07	4.64	5.00	4.91	4.79	4.39								
DALFE(0.9)	4.87	4.36	5.23	4.76	5.11	4.91	4.85	4.55								
DALFE(0.5)	5.13	4.80	5.41	5.29	5.28	5.28	5.08	5.20								
AL Cluster(2)	4.45	4.03	4.95	4.82	4.86	4.85	4.77	4.84								
AL Cluster(3)	4.04	3.58	4.53	4.55	4.63	4.66	4.70	4.87								

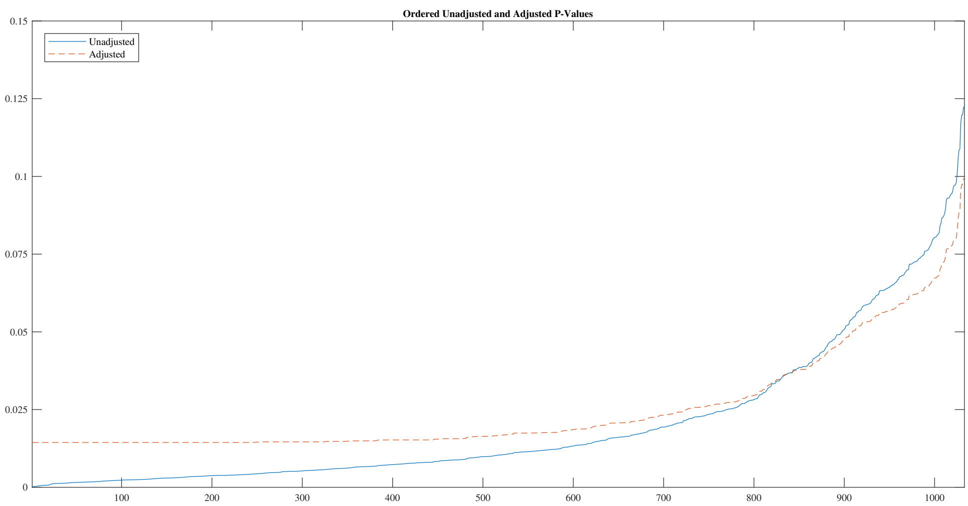
Notes: CER denotes the Certainty Equivalent Return (reported in annualized percentage points) that an investor with mean-variance preferences and risk aversion coefficient of three would gain when employing the alternative specifications. CER_1 and CER_2 correspond to the selection of the best subset k on the basis of a prior of $\pi = 1/2$ and $\pi = 1/3$, respectively. The weight on stocks in the investor's portfolio is restricted to lie between zero and 1.5.

Figure 1: Scatterplot of forecast variances and squared forecast biases



Notes: We depict the Historical Average (purple point), the k=1 model of Rapach, Strauss and Zhou (2010) and k=2 of EGT (orange points with labels L1:L2), the 21 MCS models (red points with labels M1:M21) and their analogues for k=1 (blue points with labels K1:K21). The other points correspond to the individual predictive regression model forecasts.

Figure 2: Multiple hypothesis testing



Notes: The figure depicts the first 1032 elements for the sequence of ordered p-values for testing the significance of R^2_{OS}. The solid line delineates unadjusted p-values. The dashed line delineates adjusted p-values based on the Benjamini and Hochberg (2000) adaptive procedure.