A quantile regression approach to equity premium prediction

Loukia Meligkotsidou[1]
Ekaterini Panopoulou[2]
Ioannis D. Vrontos[3]
Spyridon D. Vrontos[4]

[1] Department of Mathematics, University of Athens, Panepistimiopolis
[2] Kent Business School, University of Kent
[3] Department of Statistics, Athens University of Economics and Business
[4] Westminster Business School, University of Westminster

This is the accepted version of the following article: Meligkotsidou L., Panopoulou E., Vrontos I. D., and Vrontos S. D. 2014. A Quantile Regression Approach to Equity Premium Prediction. Journal of Forecasting. DOI: 10.1002/for.2312, which has been published in final form at http://dx.doi.org/10.1002/for.2312

A Quantile Regression Approach to Equity Premium Prediction

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

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

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

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

[d]Westminster Business School, University of Westminster, United Kingdom

Abstract

We propose a quantile regression approach to equity premium forecasting. Robust point forecasts are generated from a set of quantile forecasts using both fixed and time-varying weighting schemes, thereby exploiting the entire distributional information associated with each predictor. Further gains are achieved by incorporating the forecast combination methodology into our quantile regression setting. Our approach using a time-varying weighting scheme delivers statistically and economically significant out-of-sample forecasts relative to both the historical average benchmark and the combined predictive mean regression modeling approach.

JEL classification: G11; G12; C22; C53

Keywords: Equity premium; Forecast combination; Predictive quantile regression; Robust point forecasts; Time-varying weights.

Acknowledgments: We benefited from discussions with seminar/ conference participants at the Department of Business Administration, Athens University of Economics and Business, the Department of Economics, University of Macedonia, the 6th Annual Methods in International Finance Network Workshop, the 4th International Conference of the Financial Engineering and Banking Society (F.E.B.S), the 3rd Humboldt–Copenhagen Conference on Financial Econometrics and the 4th International conference on Computational and Financial Econometrics (CFE). Ioannis Vrontos has received funding from the European Union's Seventh Framework Programme (FP7-SSH/2007-2013) for research, technological development and demonstration under grant agreement no 320270 – SYRTO. The usual disclaimer applies.

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

1 Introduction

Since the seminal contribution of Goyal and Welch (2008), equity premium predictability has attracted the attention of both academics and practitioners in finance.[1] The early contributions to equity premium predictability primarily focused on the in-sample predictive ability of the potential predictors and the development of the proper econometric techniques for valid inference.[2] Lately, interest has turned to the out-of-sample performance of the candidate variables. Goyal and Welch (2008) showed that their long list of predictors cannot deliver consistently superior out-of-sample performance. The authors employed a variety of predictive regression models ranging from those with a single variable to their 'kitchen sink' model that contains all of the predictors simultaneously. Campbell and Thompson (2008) showed that imposing simple restrictions suggested by economic theory on several coefficients improves the out-of-sample performance. Based on this result, these authors argue that market timing strategies can deliver profits to investors (see also Ferreira and Santa-Clara, 2011). More recently, Rapach et al. (2010) considered another approach for improving equity premium forecasts based on forecast combinations. The authors find that combinations of individual single-variable predictive regression forecasts, which help reduce the model uncertainty/parameter instability, significantly beat the historical average forecast. Finally, Ludvigson and Ng (2007) and Neely et al. (2013) adopted a diffusion index approach, which can conveniently track the key movements in a large set of predictors, and found evidence of improved equity premium predictability.

The empirical findings on equity premium predictability are mixed. The majority of studies on this topic have been conducted within linear regression frameworks. However, recent contributions to the literature have noted that the relationship between returns and predictors is not linear, and several approaches have been proposed to capture this nonlinearity. Markov-switching models are among the most popular models for forecasting stock returns (Guidolin and Timmermann, 2009; Henkel et al., 2011). Other well-known nonlinear specifications include threshold models (Franses and van Dijk, 2000; Terasvirta, 2006; Guidolin et al., 2009) and neural nets (Franses and van Dijk, 2000; Terasvirta, 2006; White, 2006). Non- or semi-parametric modeling represents another approach for approximating general functional forms for the relationship between the expected returns and the predictors (Chen and Hong, 2010; Ait-Sahalia and Brandt, 2001).

In this paper, we contribute to the equity premium predictability literature by considering predictive quantile regression models. We argue that due to nonlinearity and non-normality patterns, a linear regression approach might not be adequate for exploring the ability of various predictors to forecast the entire distribution of returns. Empirically, the evidence against normally distributed stock returns is overwhelming. The

[1] Following the related literature, equity premium is proxied by excess returns.

[2] Rapach and Zhou (2013) offer a detailed review on the issue of equity return predictability.

equity premium distribution exhibits time varying volatility, excess kurtosis (fat tails) and negative skewness, possibly induced by extreme market movements, business cycle fluctuations, institutional change, policy shocks, advances in information technology, and investor learning (Rapach et al., 2010). In this respect, we consider predictive quantile regression models, which enable us to have a more complete characterization of the conditional distribution of returns through a set of conditional quantiles. This approach is non-parametric, more flexible than other parametric approaches, such as linear regression, Markov-switching and threshold regression models, and is robust to deviations from normality, including the presence of outliers. Moreover, modeling just the conditional mean of the return series, through a standard or complex linear regression specification, may obscure interesting characteristics and lead us to conclude that a predictor has poor predictive performance, while it is actually valuable for predicting the lower or/and the upper quantiles of the returns. Our framework allows us to capture the asymmetric effect of candidate predictors (non-linear relationship) on the return distribution and as a result track different types of predictability. For example, Cenesizoglu and Timmermann (2008) find that high T-bill rates reduce the central and upper quantiles of the return distribution, while they don't have a similar effect on lower quantiles. In this respect, low T-bill rates are associated with strong market performance, while the converse is not true. To the extent that candidate predictors contain significant information for certain parts of the return distribution, but not for the entire distribution, a methodology that properly integrates this information would lead to additional benefits.

Since the seminal paper of Koenker and Bassett (1978), quantile regression models have attracted a vast amount of attention. Both theoretical and empirical research has been conducted in the area of quantile regression, including model extensions, new inferential procedures and numerous empirical applications; see, for example, Buchinsky (1994, 1995) and Yu et al. (2003) among others.[3] The paper most closely related to the present paper is that of Cenesizoglu and Timmermann (2008), who employ a quantile regression approach to capture the predictive ability of a list of state variables for the distribution of stock returns. The authors find quantile-varying predictability both in-sample and out-of-sample, which can be exploited in an asset allocation framework. In a follow-up paper, Cenesizoglu and Timmermann (2012) note that return prediction models that allow for a time-varying return distribution lead to better estimates of the tails of the distribution of the returns and suffer less from unanticipated outliers. Similar conclusions are reached by Pedersen (2010), who employs both univariate and multivariate quantile regressions to jointly model the distribution of stocks and bonds.

In this paper, we construct equity premium point forecasts by combining quantile forecasts obtained from a set of simple regressions (i.e., regressions with only one regres-

[3] Applications in the field of finance include Bassett and Chen (2001), Engle and Manganelli (2004), Meligkotsidou et al. (2009) and Chuang et al. (2009).

sor). To begin, each regressor $x_{i,t}$ ($i = 1, ..., N$) is used to predict the quantile of order τ_j ($j = 1, ..., J$) of the distribution of the excess return of the next period (r_{t+1}). Next, two approaches are explored to combine these quantile forecasts into a point forecast that is robust to non-normality and nonlinearity. One approach, which we name the Robust Forecast Combination (RFC), proceeds by first combining the quantile forecasts across all values of τ_j into point forecasts for each predictor $x_{i,t.}$ This step yields N robust point forecasts, which are combined into a final point forecast using either a fixed or a time-varying weighting scheme. An alternative approach, which we name the Quantile Forecast Combination (QFC), consists of first combining the predicted quantiles of the same order τ_j across all regressors. This step yields J quantile forecasts (one for each τ_j), which are then combined into a final robust point forecast using either a fixed or a time-varying weighting scheme. Note that both approaches (RFC and QFC) produce point forecasts of the expected value of r_{t+1}, conditional on the information available at time t.

For comparison purposes, we employ the updated Goyal and Welch (2008) dataset along with the standard linear regression predictive framework, as well as existing methods for combining individual forecasts from single predictor linear models. All different forecasts are evaluated against the benchmark of a constant equity premium, using both statistical and economic evaluation criteria. To anticipate our key results, we find considerable heterogeneity among the candidate variables, as far as their ability to predict the return distribution is concerned. More importantly, no single predictor proves successful in forecasting the entire return distribution. Overall, a superior predictive performance, both in terms of statistical and economic significance, is achieved under the QFC approach with time-varying weighting schemes. One might expect the latter approach to outperform the competing approaches considered, since it produces accurate quantile forecasts first, by synthesizing information from different predictors, thus producing an accurate estimate of the entire predictive distribution of the equity premium, and then it combines the quantile forecasts using an optimal scheme to produce a final point forecast.

The remainder of the paper is organized as follows. Section 2 describes the econometric models considered in this study, including the predictive mean and quantile regression models. Section 3 presents our forecasting approaches, and Section 4 discusses how these approaches can be combined. The framework for forecast evaluation, both in statistical and economic evaluation terms, is presented in Section 5, while our empirical results are reported in Section 6. Section 7 summarizes and presents the paper's conclusions.

2 Predictive Regressions

In this section we present the predictive regression models we employ to forecast the equity premium, denoted by r_t, using a set of N predictive variables.

2.1 Predictive mean regressions

First we consider all possible predictive mean regression models with a single predictor of the form:

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

where r_{t+1} is the observed excess return on a stock market index in excess of the risk-free interest rate at time $t+1$, $x_{i,t}$ are the N observed predictors at time t and the error terms ε_{t+1} are assumed to be independent with mean zero and variance σ^2. Equation (1) is the standard equity premium prediction model (see, for example, Rapach et al. 2010) and is estimated using the Ordinary Least Squares (OLS) method. Based on the least squares estimation, the expectation of a random variable r with distribution function F arises as the point estimate of r corresponding to the quadratic loss function $\rho(u) = u^2$, i.e., it arises as the value of $\bar{r}$, which minimizes the expected loss:

$$E\rho(r - \bar{r}) = \int \rho(r - \bar{r})dF(r).$$

Therefore, the OLS estimators $\hat{\alpha}_i, \hat{\beta}_i$ of the parameters in the predictive mean regression models in (1) can be estimated by minimizing the sample estimate of the quadratic expected loss, $\sum_{t=0}^{T-1} (r_{t+1} - \alpha_i - \beta_i x_{i,t})^2$, with respect to α_i, β_i.[4] Then, the point forecast of the equity premium at time $t+1$, based on the ith model specification, is obtained as: $\hat{r}_{i,t+1} = \hat{\alpha}_i + \hat{\beta}_i x_{i,t}$.

2.2 Predictive quantile regressions

The model specification above is primarily devised to predict the mean of r_{t+1}, not its entire distribution. Hence, this model may fail to correctly predict the quantiles of the distribution of r_{t+1}, if the true relationship between r_{t+1} and $x_{i,t}$ is nonlinear or if r_{t+1} and $x_{i,t}$ are not jointly Gaussian. Following the literature on the nonlinear relationship between returns and predictors (Guidolin and Timmermann, 2009; Guidolin et al., 2009; Chen and Hong, 2010; Henkel et al., 2011), we adopt a more sophisticated approach to equity premium forecasting by employing predictive quantile regression models (Koenker and Bassett, 1978; Buchinsky, 1998; Yu et al., 2003). Quantile regression estimators are more efficient and more robust than mean regression estimators in cases that nonlinearities and deviations from normality, including the presence of outliers, exist. The fact that quantile regression estimators are not sensitive to outliers is particularly important in our forecasting context. It implies that the quantile forecasts are still accurate in the presence of extreme positive or negative returns in the sample and, therefore, the respective combined point forecasts are robust.

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

We consider single predictor quantile regression models of the form:

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

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 the τth quantile of r_{t+1}, given $x_{i,t}$ is $Q_\tau(r_{t+1}|x_{i,t}) = \alpha_i^{(\tau)} + \beta_i^{(\tau)} x_{i,t}$, where the intercept and the regression coefficients depend on τ. The $\beta_i^{(\tau)}$ values are likely to vary across the τ values, revealing a larger amount of information about returns than the predictive mean regression model.

Similar to the expectation of the random variable r, the τth quantile arises as the solution to a decision-theoretic problem; that of obtaining the point estimate of r corresponding to the asymmetric linear loss function, usually referred to as the check function:

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

More specifically, minimization of the expected loss:

$$E\rho_\tau(r - \bar{r}^{(\tau)}) = \int \rho_\tau(r - \bar{r}^{(\tau)}) dF(r),$$

with respect to $\bar{r}^{(\tau)}$ leads to the τth quantile. In the symmetric case of the absolute loss function ($\tau = 1/2$), we obtain the median. 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 $\sum_{t=0}^{T-1} \rho_\tau \left(r_{t+1} - \alpha_i{}^{(\tau)} - \beta_i^{(\tau)} x_{i,t} \right)$, where the check function $\rho_\tau(u)$ has been given in (3). Then, the forecast of the τth quantile of the distribution of the equity premium at time $t+1$, based on the ith model specification, is obtained as $\hat{r}_{i,t+1}(\tau) = \hat{\alpha_i}^{(\tau)} + \hat{\beta}_i^{(\tau)} x_{i,t}$.

3 Forecasting Approaches

In this section, we describe the forecasting approaches we follow. To facilitate the exposition of our approaches, we first describe the design of our forecast experiment, which is identical to the one employed by Goyal and Welch (2008) and Rapach et al. (2010), in order to ensure comparability of our results. Specifically, we generate out-of-sample forecasts of the equity premium using a recursive (expanding) window. In this way, all the data available at a point in time are used and the precision of the estimates increases as time evolves. We divide the total sample of T observations into an in-sample portion of the first K observations and an out-of-sample portion of $P = T - K$ observations used for forecasting. The estimation window is continuously updated following a recursive scheme, by adding one observation to the estimation sample at each step. As such, the coefficients in any predictive model employed are re-estimated after each step of the recursion. Proceeding in this way through the end of the out-of-sample period, we generate a series of P out-of-sample forecasts for the equity premium. The first P_0 out-of-sample observations serve as an initial holdout period for the methods that require one. In this respect, we evaluate $T-(K+P_0) = P-P_0$ forecasts of the equity premium $\{\hat{r}_{i,t+1}\}_{t=K+P_0}^{T-1}$ over the post-holdout out-of-sample period.

3.1 Forecasting approach based on mean regressions

Following Rapach et al. (2010), we exploit information across individual forecasts via forecast combinations.[5] Out-of-sample equity premium forecasts are generated in two steps. The first step generates forecasts by employing the N individual predictive regression models (1), i.e., each model is based on one of the candidate predictors. The next step expands into combinations of these forecasts by means of the schemes analyzed below. We refer to this forecasting approach as the Mean Forecast Combination (MFC) approach.

More specifically, the combination forecasts of r_{t+1}, denoted by $\hat{r}_{t+1}^{(C)}$, are weighted averages of the N single predictor individual forecasts, $\hat{r}_{i,t+1}$, $i = 1, \ldots, N$, of the form $\hat{r}_{t+1}^{(C)} = \sum_{i=1}^{N} w_{i,t}^{(C)} \hat{r}_{i,t+1}$, where $w_{i,t}^{(C)}, i = 1, ..., N$, are the *a priori* combination weights at time t. The simplest combination scheme is the one that attaches equal weights to all individual models, i.e., $w_{i,t}^{(C)} = 1/N$, for $i = 1, ..., N$, called the Mean combination scheme. This scheme is typically found to be a good forecast combination scheme as it reduces forecast variance and bias through averaging out individual model biases. Moreover, weights are known and don't suffer from estimation error. However, by attaching equal weights, little chance is given for a better model to work dominantly against bad models. One way to robustify the mean combination scheme and reduce its sensitivity to outlier forecasts is by employing either the Trimmed Mean or the Median combination schemes. The Trimmed Mean combination scheme sets $w_{i,t}^{(C)} = 0$ for the smallest and largest forecasts and $w_{i,t}^{(C)} = 1/(N-2)$ for the remaining ones, while the Median combination scheme employs the median of the $\{\hat{r}_{i,t+1}\}_{i=1}^{N}$ forecasts.

The second class of combination methods we consider, proposed by Stock and Watson (2004), suggests forming weights based on the historical performance of the individual models over a holdout out-of-sample period. Specifically, their Discount Mean Squared Forecast Error (DMSFE) combination method suggests forming weights as follows:

$$w_{i,t}^{(C)} = m_{i,t}^{-1} / \sum_{j=1}^{N} m_{j,t}^{-1}, \; m_{i,t} = \sum_{s=K}^{t-1} \psi^{t-1-s} (r_{s+1} - \widehat{r}_{i,s+1})^2, \; t = K + P_0, ..., T,$$

[5] Combining the forecasts of the individual models can reduce uncertainty risk associated with a single predictive model and display superior predictive ability (Bates and Granger, 1969, Hendry and Clements, 2004).

where ψ is a discount factor that attaches more weight to the recent forecasting accuracy of the individual models in the cases where $\psi \in (0, 1)$. The values of ψ we consider are 1.0 and 0.9. When ψ equals one, there is no discounting and the combination scheme coincides with the optimal combination forecast of Bates and Granger (1969) for the case of uncorrelated forecasts. Given that the performance of competing models changes over time, this method may improve on the equal weighting scheme by weighting improved forecasts progressively more heavily.

In a similar spirit, Aiolfi and Timmermann (2006) develop conditional combining methods exploiting persistence in forecasting performance. The authors argue that while it is difficult to identify the top model among forecasting models, it is possible to identify clusters of good and bad models. The Cluster combination scheme is the third class of combination schemes we employ. To create the Cluster combination forecasts, we form L clusters of forecasts of equal size based on past MSFE performance. To avoid estimation error of individual weights, each combination forecast is the average of the individual model forecasts in the best-performing cluster. This procedure begins over the initial holdout out-of-sample period and goes through the end of the available out-of-sample period using a rolling window of P_0 observations. In our analysis, we consider $L = 2, 3$. The rolling holdout window employed adds flexibility and ensures quick incorporation of good models in the forecast pool.

Finally, the Principal Components combination method of Chan et al. (1999) and of Stock and Watson (2004) is considered. In this case, a combination forecast is based on the fitted n principal components of the uncentered second moment matrix of the individual model forecasts, $\widehat{F}_{1,s+1}$, ..., $\widehat{F}_{n,s+1}$ for $s = K, ..., t-1$ and $t = K + P_0, ..., T$. The OLS estimates of $\varphi_1, ..., \varphi_n$ of the following regression: $r_{s+1} = \varphi_1 \widehat{F}_{1,s+1} + ... + \varphi_n \widehat{F}_{n,s+1} + \nu_{s+1}$ can be thought of as the individual combination weights of the principal components. The advantage of this method is that a large number of forecasts from individual models are reduced to a few principal components. As such it provides a convenient method for allowing some estimation of factor weights, yet reduces the number of weights that must be estimated. On the other hand, the performance of this method depends on the selection criterion for the number of principal components and the precision with which weights are estimated. To select the number n of principal components, we employ the IC_{p3} information criterion developed by Bai and Ng (2002) and set the maximum number of factors to 5.

3.2 Forecasting approaches based on Quantile Regressions

In this section, we describe two alternative quantile-based forecasting approaches. Both approaches generate a set of quantile forecasts of the distribution of the excess return of the next period (r_{t+1}), employing simple regressions (i.e., regressions with only one regressor, Equation 2). These approaches differ in the way that these quantile forecasts are combined into a point forecast.

3.2.1 The RFC approach (Robust Forecast Combination)

Our first approach, the RFC approach, proceeds by first combining the quantile forecasts, $\hat{r}_{i,t+1}(\tau)$, $\tau \in S$, where S denotes the set of quantiles considered, into point forecasts for each predictor $x_{i,t}$, $i = 1, ..., N$. These combinations are constructed via both fixed and time-varying schemes.

With respect to fixed weighting schemes, robust point forecasts are formed as follows:

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

Here the weights, p_τ, represent the probabilities attached to different quantile forecasts, suggesting how likely it is for each regression quantile to predict the return in the next period. We consider Tukey's (1977) trimean and the Gastwirth (1966) three-quantile estimator given, respectively, by the following formulae:

$$\begin{aligned} &\text{FW1:} \qquad \widehat{r}_{i,t+1} = 0.25\hat{r}_{i,t+1}(0.25) + 0.50\hat{r}_{i,t+1}(0.50) + 0.25\hat{r}_{i,t+1}(0.75) \\ &\text{FW2:} \qquad \widehat{r}_{i,t+1} = 0.30\hat{r}_{i,t+1}(1/3) + 0.40\hat{r}_{i,t+1}(0.50) + 0.30\hat{r}_{i,t+1}(2/3). \end{aligned}$$

Furthermore, we use the alternative five-quantile estimator, suggested by Judge et al. (1988), which attaches more weight to extreme positive and negative events as follows:

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

The above three estimators have been proposed in the literature as methods to obtain robust point estimates of the central location of a distribution based on small sets of quantile estimates. To incorporate information from a larger set of quantiles, trying to obtain a more complete characterization of the distribution of interest, we also consider a fourth estimator of the form:

$$\text{FW4: } \widehat{r}_{i,t+1} = 0.05\hat{r}_{i,t+1}(0.50) + 0.05 \sum_{\tau \in S} \hat{r}_{i,t+1}(\tau), \text{ where } S = \{0.05, 0.10, ..., 0.95\}.$$

The above estimators belong to the class of L-estimators, consisting of estimators occurring as linear combinations of order statistics (here, linear combinations of quantiles). The weights reflect specific beliefs about how certain quantile estimates should affect the estimate of the central location. As Koenker and Bassett (1978) show, estimators of this type have high efficiency over a large class of distributions. A subset of the above specifications has been employed by Taylor (2007) and Ma and Pohlman (2008), among others.

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 et al., 2004; Dangl and Halling, 2012) can motivate the employment of time-varying schemes in the generation of robust point forecasts. The variable of interest, $r_{i,t+1}$, is predicted by minimizing the mean squared forecast errors, i.e., $E_t(r_{t+1} - \hat{r}_{i,t+1})^2 = \frac{1}{t-K} \sum_{s=K}^{t-1} (r_{s+1} - \widehat{r}_{i,s+1})^2$, $t = K + P_0, ..., T$ over a continuously updated (by one observation at each step) holdout out-of-sample period. In this way, an optimal linear combination $\mathbf{p}_t=[p_{\tau,t}]_{\tau\in S}$ of the quantile forecasts $\hat{r}_{i,t+1}(\tau)$ is obtained recursively under an appropriate set of constraints. This is given by:

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

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 to not exceed certain lower and upper bounds to reduce the volatility of the weights and stabilize the forecasts.

In our empirical application, we employ three time-varying specifications that may be viewed as the time-varying counterparts of our FW1-FW3 schemes.[6] More specifically, FW1 with time-varying coefficients becomes:

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

where $p_{\tau,t}, \tau \in S = \{0.25, 0.50, 0.75\}$ are estimated by the optimization procedure:

$$\mathbf{p}_t = \arg\min_{\mathbf{p}_t} E_t[r_{t+1} - (p_{0.25,t}\hat{r}_{i,t+1}(0.25) + p_{0.50,t}\hat{r}_{i,t+1}(0.50) + p_{0.75,t}\hat{r}_{i,t+1}(0.75))]^2$$
$$s.t.\ p_{0.25,t} + p_{0.50,t} + p_{0.75,t} = 1, 0.20 \le p_{0.25,t} \le 0.40,$$
$$0.40 \le p_{0.50,t} \le 0.60, 0.20 \le p_{0.75,t} \le 0.40.$$

[6]Since our methodology requires a holdout out-of-sample period during which the optimal linear combination $\mathbf{p}_t$ is estimated, a fourth specification based on FW4 is not employed due to the increased parameter space.

Similarly, the FW2 scheme with time-varying coefficients becomes:

$$\text{TVW2: } \widehat{r}_{i,t+1} = p_{1/3,t}\hat{r}_{i,t+1}(1/3) + p_{0.5,t}\hat{r}_{i,t+1}(0.50) + p_{2/3,t}\hat{r}_{i,t+1}(2/3),$$

where $p_{\tau,t}, \tau \in S = \{1/3, 0.50, 2/3\}$ are estimated by the following optimization procedure:

$$\begin{gathered}
\mathbf{p}_t = \arg\min_{\mathbf{p}_t} E_t[r_{t+1} - (p_{1/3,t}\hat{r}_{i,t+1}(1/3) + p_{0.5,t}\hat{r}_{i,t+1}(0.50) + p_{2/3,t}\hat{r}_{i,t+1}(2/3))]^2 \\
s.t.\ p_{1/3,t} + p_{0.50,t} + p_{2/3,t} = 1, 0.15 \leq p_{1/3,t} \leq 0.45, \\
0.30 \leq p_{0.5,t} \leq 0.50, 0.15 \leq p_{2/3,t} \leq 0.45.
\end{gathered}$$

Finally, the FW3 scheme with time-varying coefficients becomes:

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

where $p_{\tau,t}, \tau \in S = \{0.10, 0.25, 0.50, 0.75, 0.90\}$ are estimated by the following optimization procedure:

$$\begin{gathered}
\mathbf{p}_t = \arg\min_{\mathbf{p}_t} E_t[r_{t+1} - (p_{0.10,t}\hat{r}_{i,t+1}(0.10) + p_{0.25,t}\hat{r}_{i,t+1}(0.25) + \\
+ p_{0.5,t}\hat{r}_{i,t+1}(0.5) + p_{0.75,t}\hat{r}_{i,t+1}(0.75) + p_{0.90,t}\hat{r}_{i,t+1}(0.90))]^2 \\
s.t.\ p_{0.10,t} + p_{0.25,t} + p_{0.50,t} + p_{0.75,t} + p_{0.90,t} = 1 \\
0.00 \leq p_{0.10,t} \leq 0.10, 0.15 \leq p_{0.25,t} \leq 0.35, \\
0.40 \leq p_{0.50,t} \leq 0.60, 0.15 \leq p_{0.75,t} \leq 0.35, 0.00 \leq, p_{0.90,t} \leq 0.10.
\end{gathered}$$

Employing one of the weighting schemes outlined yields N robust point forecasts for each predictor $x_{i,t}, i = 1, ..., N$, which are then combined into a final point forecast using the combination schemes outlined in Section 3.1.

3.2.2 The QFC approach (Quantile Forecast Combination)

Our second approach, the QFC approach, proceeds by first combining the predicted quantiles $\hat{r}_{i,t+1}(\tau)$ of the same order τ across all candidate predictors N. To do so, we adjust the combination methods outlined in Section 3.1 to our quantile setting. The Mean, Trimmed Mean and Median combination schemes retain their validity in our framework because they do not rely on some measure of past performance. On the other hand, the DMSFE, Cluster and Principal Components combination methods, which are formed on the basis of past performance as measured by the MSFE, have to be modified. To do so, we replace the MSFE metric by a metric based on the asymmetric linear loss function

(Equation 3).

The combined quantile forecasts, $\hat{r}^{(C)}_{t+1}(\tau)$, are weighted averages of the form $\hat{r}^{(C)}_{t+1}(\tau) = \sum_{i=1}^{N} w^{(C)}_{i,t}\hat{r}_{i,t+1}(\tau)$, where $w^{(C)}_{i,t}$ denotes the combination weights. First, we introduce the Discount Asymmetric Loss Forecast Error (DALFE) combination method which suggests forming weights as follows:

$$w^{(C)}_{i,t} = m^{-1}_{i,t} / \sum_{j=1}^{N} m^{-1}_{j,t},\ m_{i,t} = \sum_{s=K}^{t-1} \psi^{t-1-s}\rho_\tau(r_{s+1} - \widehat{r}_{i,s+1}(\tau)), t = K + P_0, ..., T$$

where $\psi \in (0,1)$ is a discount factor. Similarly to the DMSFE combination method, the combination weights are computed based on the historical performance of the individual quantile regression models over the holdout out-of-sample period, and ψ is set equal to 0.9 and 1. In a similar manner, we modify the Cluster combination method by forming L clusters of forecasts based on their performance as measured by the asymmetric loss forecast error. The Asymmetric Loss Cluster (AL Cluster) combination forecast is the average of the individual quantile forecasts in the best performing cluster, which contains the forecasts with the lowest expected asymmetric loss values. We consider forming $L = 2, 3$ clusters.

Next, we introduce the Asymmetric Loss Principal Components method (AL Principal Components), under which the combination of forecasts is based on the fitted n principal components of the uncentered second moment matrix of the individual quantile forecasts, $\widehat{F}^{(\tau)}_{1,s+1}$, ..., $\widehat{F}^{(\tau)}_{n,s+1}$, for $s = K, ..., t-1$ and $t = K + P_0, ..., T$, where the combination weights are computed by minimizing the sum:

$$\sum_{s=K}^{t-1} \rho_\tau(r_{s+1} - \varphi_1\widehat{F}^{(\tau)}_{1,s+1} - ... - \varphi_n\widehat{F}^{(\tau)}_{n,s+1}).$$

The IC_{p3} information criterion is used to select the number n of the principal components.

Finally, we put forward two combination methods, under which optimal quantile forecasts, $\hat{r}^{(C)}_{t+1}(\tau)$, are obtained by minimizing an objective function based on the asymmetric linear loss. More specifically, we first consider the following optimization scheme, which is an analogue of the lasso quantile regression:

$$\mathbf{w}_t = \arg\min_{\mathbf{w}_t} \sum_t \rho_\tau\left(r_{t+1} - \sum_{i=1}^{N} w^{(C)}_{i,t}\hat{r}_{i,t+1}(\tau)\right) \quad s.t.\ \sum_{i=1}^{N} w^{(C)}_{i,t} = 1,\ \sum_{i=1}^{N} |w^{(C)}_{i,t}| \leq \delta_1,$$

where the parameter δ_1 is used as a control for the amount of shrinkage. We refer to this combination quantile forecast as the Asymmetric Loss Lasso (AL Lasso). We also consider the Asymmetric Loss Ridge (AL Ridge) optimization scheme, which is an analogue of the ridge quantile regression:

$$\mathbf{w}_t = \arg\min_{\mathbf{w}_t} \sum_t \rho_\tau \left(r_{t+1} - \sum_{i=1}^{N} w_{i,t}^{(C)} \hat{r}_{i,t+1}(\tau) \right) \quad s.t. \ \sum_{i=1}^{N} w_{i,t}^{(C)} = 1, \sum_{i=1}^{N} \left(w_{i,t}^{(C)} \right)^2 \leq \delta_2,$$

where the parameter δ_2 is used as a control for the amount of shrinkage. In our study, the parameters δ_1, δ_2 are set equal to 1.4 and 0.4, respectively.[7]

Once we obtain the set of combined quantile forecasts, we calculate a final robust point forecast using one of the fixed or time-varying weighting schemes outlined in Section 3.2.1.

4 Forecast Combinations

We now consider an amalgamation of the approaches considered so far, namely the MFC, RFC and QFC approaches.[8] To check whether potential benefits can arise from combining the three approaches, we employ the multiple forecast encompassing tests of Harvey and Newbold (2000). In the event that our three approaches contain distinct information about future excess returns, we suggest forming equally weighted composite forecasts.

The notion of forecast encompassing was developed by Granger and Newbold (1973) and Chong and Hendry (1986) through the formation of composite forecasts as weighted averages of the forecasts of two competing models.[9] Harvey and Newbold (2000) extend the pairwise encompassing tests (see Section 5.1) developed by Harvey et al. (1998) to compare three or more forecasts. We consider forming a composite forecast, $\hat{r}_{c,t+1}$, as an optimal combination of the forecasts of the predictive mean regressions, the robust forecast combinations and the quantile forecast combinations, i.e., $\hat{r}_{c,t+1} = \lambda_{MFC}\hat{r}_{MFC,t+1} + \lambda_{RFC}\hat{r}_{RFC,t+1} + \lambda_{QFC}\hat{r}_{QFC,t+1}$, where $\lambda_{MFC} + \lambda_{RFC} + \lambda_{QFC} = 1$. If $\lambda_{MFC} = 1$, and $\lambda_{RFC} = \lambda_{QFC} = 0$, the MFC forecasts encompass the RFC and QFC ones, as the RFC and QFC forecasts do not contain information useful for forecasting the equity premium other than that already employed in the linear model. In a similar manner, we can test whether the RFC model encompasses QFC and MFC and whether the QFC model encompasses the MFC and the RFC model.

Harvey and Newbold developed two test statistics, namely the $F-test$ statistic and the MS^* statistic, to test the null hypothesis of multiple forecast encompassing.[10] The authors show that the $F-test$ exhibits significant size distortions in small and moderate samples with non-normal errors, while the MS^* test exhibits good size and power

[7]The above two optimization schemes can be written equivalently using the L_1 norm for the lasso quantile regression and the L_2 norm for the ridge quantile regression in the objective function. More details on the lasso regression can be found in Tibsirani (1996), on the lasso quantile regression in Wu and Liu (2009) and on the ridge regression can be found in Hastie et al. (2009).

[8]The term 'amalgamation' is employed by Rapach and Strauss (2012) when considering combining three different econometric approaches to forecast US state employment growth.

[9]See also Clements and Hendry (1998).

[10]To save space, we do not report the explicit formulae of the tests.

properties in moderately large samples. To gain a more thorough understanding on the relationship between the rival models, we must employ each one of the models as the reference model and conduct the test as many times as the models considered. Failure to reject the null hypothesis does not necessarily imply that the reference model is strictly dominant to the competing forecasts. Rather, the forecasts may be highly correlated, in which case a combination of nearly identical or similar forecasts cannot improve upon any individual forecast. On the other hand, rejection of the null hypothesis in the encompassing test suggests that the forecasts of the reference model can be improved by combining them with the forecasts of the rival model.

5 Evaluation of forecasts

5.1 Statistical evaluation

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 estimate, $\widehat{\alpha}_i$, in the linear regression model (1) when no predictor is included. As a measure of the forecast accuracy, we employ the ratio $\frac{MSFE_i}{MSFE_{PM}}$, where $MSFE_i = \sum_{t=K+P_0}^{T-1}(r_{t+1} - \widehat{r}_{i,t+1})^2$ 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 of which are computed over the out-of-sample period. Values lower than 1 are associated with the superior forecasting ability of the respective models/specifications.

To compare the information content in our proposed models/specifications relevant to the benchmark PM model, we use encompassing tests. Specifically, consider forming a composite forecast, $\hat{r}_{c,t+1}$, as a convex combination of model A forecasts, $\hat{r}_{A,t+1}$, and the ones of model B, $\hat{r}_{B,t+1}$, in an optimal way so that $\hat{r}_{c,t+1} = \lambda_A \hat{r}_{A,t+1} + \lambda_B \hat{r}_{B,t+1}$, $\lambda_A + \lambda_B = 1$. If the optimal weight attached to model A forecasts is zero ($\lambda_A = 0$), then model B forecasts encompass model A forecasts in the sense that model B contains a significantly larger amount of information than that already contained in model A. Harvey et al. (1998) developed the encompassing test, denoted as $ENC-T$, based on the approach of Diebold and Mariano (1995) to test the null hypothesis that $\lambda_A = 0$, against the alternative hypothesis that $\lambda_A > 0$. Let $u_{A,t+1} = r_{t+1} - \hat{r}_{A,t+1}$, $u_{B,t+1} = r_{t+1} - \hat{r}_{B,t+1}$ denote the forecast errors of the competing models A and B, respectively and define $d_{t+1} = (u_{B,t+1} - u_{A,t+1})u_{B,t+1}$. The $ENC-T$ statistic is given by:

$$ENC - T = \sqrt{(P - P_0)} \frac{\overline{d}}{\sqrt{\widehat{Var}(d)}},$$

where $\overline{d}$ is the sample mean, $\widehat{Var}(d)$ is the sample-variance of $\{d_{s+1}\}_{s=K+P_0}^{T-1}$ and $P - P_0$ is the length of the out-of-sample evaluation window. The $ENC-T$ statistic is asymptotically distributed as a standard normal variate under the null hypothesis. To improve the finite sample performance, Harvey et al. (1998) recommend employing Student's t distribution with $P - P_0 - 1$ degrees of freedom. To render a model as superior in forecasting ability, one also needs to test whether model A forecasts encompass model B forecasts ($\lambda_B = 0$) by employing the $ENC - T$ statistic based on $d_{t+1} = (u_{A,t+1} - u_{B,t+1})u_{A,t+1}$. When both null hypotheses are rejected, then the competing models contain discrete information about the future and an optimal convex ($\lambda_A, \lambda_B \in (0,1)$) combination forecast can be formed. In the event that none of the hypotheses of interest is rejected, both models contain similar information and the competing models are equivalent in terms of forecasting ability. When one of the null hypotheses is rejected, then the respective model forecasts dominate the forecasts of the competing model.

5.2 Economic evaluation

While MSFE is the most popular measure of forecast accuracy, it is not necessarily the most relevant metric for assessing stock return forecasts, since it does not account for the risk borne by the investors over the out-of-sample period. To address this issue, we calculate realized utility gains for a mean-variance investor in real time. Following Campbell and Thompson (2008) and Rapach et al. (2010) we employ a mean-variance utility for an investor with relative risk aversion parameter γ who allocates her wealth between the safe (risk-free Treasury Bill) and the risky asset (stock market) quarterly employing equity premium forecasts based on the competing models/ specifications.[11] The investor decides at the end of each period t to allocate the following share (w_t) of her wealth to the risky asset:

$$w_t = \frac{E_t(r_{t+1})}{\gamma Var_t(r_{t+1})} = \frac{\widehat{r}_{i,t+1}}{\gamma Var_t(r_{t+1})}, \tag{4}$$

where E_t and Var_t denote the conditional expectation and variance of the equity premium (r_{t+1}) (Campbell and Viceira, 2002). The conditional expectation of each model/ specification is given by the forecast from the specific model and the variance is calculated using a ten-year rolling window of quarterly returns. Over the forecast evaluation period the investor with an initial wealth of W_0 realizes an average utility of:

$$\overline{U} = \frac{W_0}{(P - P_0)} \left[\sum_{t=0}^{P-P_0-1} (R_{p,t+1}) - \frac{\gamma}{2} \sum_{t=0}^{P-P_0-1} (R_{p,t+1} - \overline{R}_p)^2 \right], \tag{5}$$

[11] This utility-based approach, initiated by West et al. (1993), has been extensively employed in the literature as a measure for ranking the performance of competing models in a way that captures the trade-off between risk and return (Fleming et al., 2001; Marquering and Verbeek, 2004; Della Corte et al., 2009, 2010; Wachter and Warusawitharana, 2009). Alternative utility specifications may be employed such as power or log utility.

where $R_{p,t+1}$ is the gross return on her portfolio at time $t+1$.[12] In a similar manner, the investor can form her portfolio on the basis of the PM model, i.e. the historical average forecast. The utility gain is the difference between the average realized utility over the out-of-sample period of any of our i competing models/specifications ($\overline{U^i}$) and the respective value for the prevailing mean (PM) model ($\overline{U^{PM}}$). The utility gain can be interpreted as the portfolio management fee that an investor would be willing to pay to have access to the additional information available in our proposed specifications relative to the information in the historical equity premium. Following Campbell and Thompson (2008) and Rapach et al. (2010) we set γ equal to three and calculate this performance fee as follows:[13]

$$\Phi = \Delta\overline{U} = \overline{U^i} - \overline{U^{PM}}. \tag{6}$$

If our proposed model does not contain any economic value, the performance fee is negative ($\Phi \leq 0$), while positive values of the performance fee suggest superior predictive ability against the PM benchmark. Φ is reported in annualized basis points.

6 Empirical Application

6.1 The data

The data we employ are from Goyal and Welch (2008), who provide a detailed description of transformations and datasources.[14] The equity premium is calculated as the difference of the continuously compounded S&P 500 returns, including dividends, and the Treasury Bill rate. Our forecasting experiment is conducted on a quarterly basis and the data span 1947:1 to 2010:4. Our out-of-sample forecast evaluation period corresponds to the 'long' one analyzed by Goyal and Welch (2008) and Rapach et al. (2010), covering the period 1965:1-2010:4.[15]

The 15 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, D/P (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), the dividend yield, D/Y (difference between the log of dividends and the log of lagged stock prices), the earnings–

[12] We standardize the investor problem by assuming that $W_0 = 1$ and constrain the portfolio share on the risky asset to lie between 0% and 150% each month, i.e. $0 \leq w_t \leq 1.5$.

[13] A value of $\gamma = 3$ represents medium risk aversion. Alternatively, values of γ ranging from 1 (low risk aversion) to 10 (very high risk aversion) could be employed.

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

[15] Please note that the out-of-sample period refers to the period used to evaluate the out-of-sample forecasts. We use the ten years 1955:1 to 1964:4 (40 quarters) before the start of the out-of-sample evaluation period as the initial holdout out-of-sample period, required for both constructing our time-varying robust forecasts and for several forecast combination schemes.

price ratio, E/P (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), the dividend–payout ratio, D/E (difference between the log of dividends and the log of earnings), the stock variance, SVAR (sum of squared daily returns on the S&P 500 index), the book-to-market ratio, B/M (ratio of book value to market value for the Dow Jones Industrial Average) and the net equity expansion, NTIS (ratio of twelve-month moving sums of net issues by NYSE-listed stocks to total end-of-year market capitalization of NYSE stocks). This set of variables, consisting mainly of valuation ratios, aims to capture some measure of 'fundamental' value to market value since these ratios are widely used to relate stock valuation to actual cash flows, profits or firm values.

Turning to interest-rate related variables, we employ six variables ranging from short-term government rates to long-term government and corporate bond yields and returns along with their spreads; namely the Treasury bill rate, TBL (Interest rate on a three-month Treasury bill), the long-term yield, LTY (Long-term government bond yield), the long-term return, LTR (return on long-term government bonds), the term spread, TMS (difference between the long-term yield and the Treasury bill rate), the default yield spread, DFY (difference between BAA- and AAA-rated corporate bond yields), the default return spread, DFR (difference between long-term corporate bond and long-term government bond returns). These variables capture level or slope stock market effects from the term structure, since for example the short term rate is linked with firms' financing costs, while the long term interest rate is associated with long term growth prospects. A positive term spread is associated with future expansions, while a widening default spread is linked to increased equity default risk and recessions. In this respect, they act as business cycle variables along with the the inflation rate, INFL (calculated from the CPI- all urban consumers) and the investment-to-capital ratio, I/K (ratio of aggregate -private nonresidential- fixed investment to aggregate capital for the entire economy), which aim to capture the overall macroeconomic environment.

6.2 Empirical results

6.2.1 A motivating illustration

Before presenting our empirical results, we provide an illustration of the sources of the potential benefits of our proposed methodology. The aim of this exercise is to assess the predictive ability of the individual predictor variables, $x_{i,t}$, to forecast the τth quantile. To this end, we generate forecasts employing a single predictor at a time, $\widehat{r}^{(\tau)}_{i,t+1} = \widehat{\alpha}^{(\tau)}_i + \widehat{\beta}^{(\tau)}_i x_{i,t}$, $i = 1, ..., N$, and calculate the expected asymmetric loss, $\sum_t \rho_\tau \left(r_{t+1} - \widehat{r}^{(\tau)}_{i,t+1} \right)$, associated with each model specification. Next, we calculate the expected loss associated with the quantile forecasts, $\widehat{r}^{(\tau)}_{t+1} = \widehat{\alpha}^{(\tau)}$, obtained from the Prevailing Quantile (PQ) model, i.e., the model that contains only a constant. This prevailing quantile model serves as a benchmark in the same fashion as the historical average (prevailing mean) serves as a benchmark in typical predictive mean regressions. Table 1, Panel A illustrates our findings with highlighted (in grey) cells suggesting a superior predictive ability, i.e., lower out-of-sample values of the expected asymmetric loss. Overall, we observe considerable heterogeneity among the candidate variables as far as their ability to predict the return distribution is concerned. For example, the D/P and D/Y variables display predictive ability for the 10th and 15th quantile, but mainly for the central and some right-tail quantiles of the distribution of returns, i.e., from the 45th to the 80th quantiles. On the other hand, DFR, INFL and I/K are valuable predictors for the left-tail and central quantiles of the return distribution. Finally, D/E, SVAR and DFY help in predicting some upper quantiles and TBL the 30th to 45th quantiles. It is apparent that no single predictor proves successful in predicting the entire distribution of returns.

[TABLE 1 AROUND HERE]

We now examine whether combining the information from different predictors to predict each quantile enhances our ability to forecast the quantiles of the return distribution. For this purpose, we employ the appropriate combination methods for combining quantile forecasts, as described in Section 3.2.2. The potential predictive ability of the combination schemes considered is outlined in Table 1, Panel B. Our results suggest that combination methods outperform single variable models over the whole range of the future return distribution. The Mean, Trimmed Mean, DALFE and AL Ridge methods cover the full range of the distribution, while the Median and the AL Cluster methods are successful in all parts of the distribution, with the exception of the 90th and the 5th quantile, respectively. The AL Principal Components combination method does not outperform the PQ model in terms of predictive ability, except for the 30th and 40th quantile. Finally, the AL Lasso method is superior to the PQ model at forecasting the left part of the return distribution and some right-tail quantiles.

6.2.2 Statistical evaluation of alternative approaches

Table 2 reports the out-of-sample performance of both the single predictor mean regression forecasts and forecasts obtained using the MFC approach. In particular, Table 2 presents the MSFE ratios of each of the individual predictive regression models relative to the historical average benchmark model for the out-of-sample period 1965:1-2010:4. Values lower than 1 indicate a superior forecasting performance of the predictive models with respect to the historical average forecast. We observe that only four out of the 15 individual predictors, namely D/P, D/Y, DFR and I/K, have lower than one MSFE ratios, indicating superior predictive ability.

To assess the statistical significance of the out-of-sample forecasts of the various competing models with respect to the PM forecasts, we use the encompassing test. In Table 2, λ_A denotes the parameter associated with the test that examines whether the PM forecasts encompass the forecasts taken from the individual predictive models, whereas λ_B denotes the parameter associated with the test that examines whether the individual predictive model forecasts encompass the PM ones. Our findings suggest that the D/P, D/Y and I/K predictors contain useful forecasting information beyond what is already contained in the PM model. On the other hand, the PM forecasts dominate the D/E, B/M, NTIS, LTY, LTR and DFY forecasts. Our findings with respect to the MFC approach suggest that all of the combination schemes (except for the Principal Components method) produce lower than unity MSFE ratios. The encompassing test confirms the statistical significance of our forecasts obtained by this approach (with the exception of the Principal Components method). Overall, the results of Table 2 are in agreement with the findings of Rapach et al. (2010), who found that D/P, D/Y and I/K have significant forecasting ability and that the combination methods outperform the individual predictive regression models.

[TABLE 2 AROUND HERE]

Table 3 reports the MSFE ratios and the results of the encompassing test for the RFC approach forecasts (Panel A) and the QFC approach forecasts (Panel B), under both fixed and time-varying weighting schemes, relative to the historical average (PM) benchmark model.[16] Based on Panel A of Table 3, we may draw the following conclusions. First, regarding the results of the RFC approach under fixed weighting schemes (FW1-FW4), we observe that almost all of the combination methods, except for the Principal Components method, and in some cases the Cluster 3 method, provide MSFE ratios below unity and, hence, their forecasts dominate the PM forecast. The related encompassing tests confirm the statistical significance of these forecasts. A comparison of the different combination techniques suggests that the DMSFE methods rank first, followed by the mean combination method. Among the four fixed weighting schemes, the FW4 scheme produces, in most of the cases, lower MSFE ratios, indicating improved predictive performance, most likely because it utilizes distributional information obtained from a finer grid of return quantiles. Second, the results of the RFC approach under time-varying weighting schemes (TVW1-TVW3) are more striking. The MSFE ratios in this case are all below unity, ranging from 0.976 for the Median-TVW2 combination method to 0.963 for the Mean-TVW3 combination method.[17] Moreover, all of the MSFE ratios for the RFC approach that are based on time-varying weights are lower than the corresponding MSFE ratios of both the MFC (Table 2) and the fixed weighting RFC approach (Table 3, Panel B). The encompassing tests suggest that the RFC forecasts dominate the forecasts of the PM model.

[16]The respective results for single predictor robust point forecasts are available from the authors upon request. These results indicate superior forecasting ability of four predictors over the historical average, namely D/P, D/Y, DFR and I/K and they show improved out-of-sample performance over the mean regression approach, especially in the case of time-varying weighting schemes.

[17]Since the time-varying weighting schemes require a holdout out-of-sample period, they can only be used together with combination methods that do not require a holdout period.

Panel B of Table 3 presents the out-of-sample performance of the QFC robust point forecasts obtained under fixed (FW1-FW4) and time-varying weighting schemes (TVW1-TVW3). The results of Panel B suggest that the QFC forecasts that are based on fixed weighting schemes, with the exception of the AL Principal Components combination method, provide MSFE ratios below unity, indicating a superior performance relative to the historical average benchmark. A comparison of the different combination methods reveals that the AL Ridge method ranks first, followed by the DALFE, the Mean and the AL Cluster 2. It is interesting to observe that more promising results arise from the use of time-varying weighting schemes of the proposed QFC approach. Specifically, the QFC-TVW approach generates MSFE ratios below unity, and in many cases, the lowest ratios among the different forecasting approaches considered in our analysis. The results of Table 3 suggest that the best out-of-sample performance is obtained by applying the Mean QFC approach using time-varying weights.

[TABLE 3 AROUND HERE]

Pairwise Encompassing Tests

Our analysis so far has shown that the proposed forecasting methods based on quantile regression (i.e., the RFC and QFC approaches) using time-varying weighting provide superior forecasts compared to the standard MFC approach. Below, we present and discuss a more formal comparison of the MFC approach with the two alternative approaches proposed in this paper via a series of encompassing tests. Specifically, we compare all pairs of forecasts obtained by the MFC, the time-varying RFC and the time-varying QFC approaches using pairwise encompassing tests. The results of these tests are shown in Table 4 (Panel A). The comparison of MFC with RFC shows that the MFC forecasts are dominated by the RFC forecasts under the first weighting scheme, if either the Mean or the Trimmed Mean combination method is used, and under the third weighting scheme, if the Median combination method is used. Similarly, the MFC forecasts are dominated by the QFC forecasts under both the first and the second weighting schemes for all the combination methods considered. Quite importantly, the MFC forecasts do not prove more accurate than any of the proposed forecasting approaches based on quantile regression. Finally, the comparison of the two robust forecasting approaches with each other shows that the QFC forecasts are superior to the RFC forecasts for the Mean and Median combination methods under the third time-varying weighting scheme.

[TABLE 4 AROUND HERE]

Multiple Encompassing Tests and an Amalgam Forecast

To check whether potential benefits can arise from combining the three approaches, namely the MFC, the RFC and the QFC approach, we employ the multiple forecast encompassing tests of Harvey and Newbold (2000). Given the abundance of the models we have considered so far, we only report multiple forecast encompassing tests for the models employed in the pairwise encompassing tests. Table 4, Panel B (columns 2-4) reports the respective MS^* test statistics. Overall, non-rejections of the null dominate our findings, pointing to similarities in the forecasting ability of our competing models and possibly non-gains from considering forming composite forecasts. More specifically, the only case that the MS^* test rejects the null of multiple encompassing is when the Mean combination scheme is employed and the robust point forecasts are generated by the TVW3 scheme. Forming composite forecasts of the three approaches considered can help us gain more insight into the nature of our forecasts. Given that our experiment should be in real time, we do not estimate the weights in forming our composite forecasts, rather we attach a weight of 1/3 to each of our competing models. Table 4, Panel B (column 5) reports the MSFE ratio of our amalgam forecasts along with the related encompassing tests (columns 6-7). Overall, the MSFE ratio ranges from 0.964 for the amalgam forecast formed on the basis of Mean combination schemes and TVW1 robust forecasts to 0.983 for the forecasts formed based on the Median combination schemes and TVW3. More importantly, all amalgam forecasts dominate the benchmark forecasts of the historical average as indicated by the encompassing tests. However, no amalgam forecast proves more accurate than the forecasts of the QFC and/or RFC methods, lending support to the superiority of our proposed approaches. Even in the case that the MS^* test pointed to benefits to combining methods, namely the Mean combination scheme with the robust point forecasts generated by TVW3, the amalgam forecast is superior to the MFC forecasts but not superior to the RFC or QFC forecasts.

6.2.3 Economic evaluation of alternative approaches

We begin our analysis with the economic evaluation of the MFC approach (Table 5, Panel A, column 2). Our results suggest that, regardless of the method employed, an investor enjoys utility gains ranging from 145 (Median) to 321 (DMSFE(0.9)). Quite interestingly, while the Principal Components method is not statistically superior to the benchmark model, its employment can generate profits to an investor amounting to 236 bps. The combination methods with the highest ability to time the market are the DMSFE methods, followed by the Mean and the Trimmed Mean. Next, we turn our attention to the economic performance of the fixed weighting RFC approach (Table 5, Panel A, columns 3-6). Overall, our results suggest that an investor who employs the RFC approach will always generate positive abnormal returns, which are nearly as good as the MFC ones. The lowest utility gains are observed in the Median method ranging from 18 bps to 108 bps, whereas the highest utility gains are attained by the Principal Components, Cluster 2 and DMSFE(0.9). A comparison of the four weighting schemes reveals that FW4, which aggregates information of quantiles over a finer grid, provides the investor with more utility gains, and the highest performance fee of 275 bps is achieved when the investor employs FW4 with DMSFE(0.9). Turning to our time-varying RFC approach (Table 5, Panel A, columns 7-9), our findings indicate that the time-varying

RFC approach outperforms both the MFC and the fixed weighting RFC approaches. The utility gains range from 159 bps (TVW2 Median) to 395 bps (TVW3 Mean), which is the highest value of the utility gain attained so far.

[TABLE 5 AROUND HERE]

Panels B and C (Table 5) address the issue of the economic evaluation of the QFC approach and the amalgam forecasts, respectively. The overall picture that emerges confirms the robustness of our proposed methodology. More specifically, the performance fee that an investor would be willing to pay to utilize our proposed models (with the exception of the FW-Median combination method) ranges from 158 bps for the Trimmed Mean QFC-FW1 to 425 bps for QFC-TVW1 and the Mean combination method. When considering the fixed weighting schemes, the best performance is achieved by AL Lasso (QFC-FW1), AL Cluster 3 (QFC-FW2) and AL Ridge (QFC-FW3 and QFC-FW4). More importantly, when an investor employs any of the QFC-TVW models, she can enjoy benefits ranging from 243 bps to 425 bps. Superior performance is achieved by the QFC-TVW1 scheme, regardless of the combination method employed. Comparing our QFC to the RFC time-varying approaches, we find that when employing either TVW1 or TVW2, QFC is to be preferred, while the opposite holds for TVW3. Finally, it is interesting to note that the amalgam forecasts attain a satisfactory performance ranging from a fee of 236 bps to 375 bps, with the exception of TVW3 and the Median combination scheme.

The key findings and implications of both the statistical and economic evaluation of our approaches can be summarized as follows. First, equity premium forecasts generated by combining quantile forecasts outperform both the mean forecast combination approach and the historical average by statistically and economically meaningful margins. Second, combining predictor information first, in order to produce accurate quantile forecasts of the equity premium distribution which are then employed in the point forecast construction seems to be the optimal approach. Benefits are more pronounced when these point forecasts are generated in a time-varying weighting manner. Finally, the robustness of our findings suggests that even simple combination schemes, such as the mean, combined with the employment of as few as three quantile forecasts can generate significant benefits.

7 Conclusions

In this study, we propose a quantile regression approach to equity premium prediction. We develop two forecasting approaches that produce robust to nonlinearity, non-normality and outliers point forecasts of the equity premium. Both approaches combine quantile forecasts obtained from a set of single-variable regressions. The first approach (RFC approach) proceeds by first combining a set of quantile forecasts into robust point forecasts, one for each candidate predictor, by employing either a fixed or a time-varying scheme. Next, these forecasts are combined into a final point forecast using existing forecast combination schemes. The second approach that we propose (the QFC approach) first combines predictor information to produce a composite quantile forecast using suitably modified combination schemes. Next, this set of quantile forecasts is combined into a final robust point forecast via a fixed or a time-varying scheme.

Our approaches are able to capture the nonlinear relationship of returns with predictors and to identify potential differences in the ability of predictors to forecast various quantiles of returns. While no single predictor proves successful in forecasting the entire return distribution, our analysis suggests that predictors exist with superior predictive ability for lower or/and upper quantiles of returns. Overall, a superior predictive performance, in terms of both statistical and economic significance, is achieved under the QFC approach with time-varying weighting schemes. Our findings suggest that in order to approximate the equity premium process, which is a highly uncertain, complex, and constantly evolving one, quantile forecasts should be generated by the combination of information contained in a rich set of predictors. Then, time varying weighting schemes aimed at capturing the evolution of the equity premium should be employed in order to produce robust point forecasts. Our approach reduces uncertainty associated with a candidate predictor through combination of forecasts, addresses the complexity of the return process by forecasting various parts of the return distribution and finally weighs these constantly evolving parts by time-varying schemes.

Rapach et al. (2010) state that 'applied asset pricing models could benefit from the consideration of more complex data-generating processes with more variables that better mimic time varying fluctuations in expected returns related to the real economy'. Similarly to their combination strategy, the approaches used in the present paper provide a tractable way of doing this. Our asset allocation experiment showed that a mean-variance investor who adopts our framework can gain sizable benefits that range from 243 bps to 425 bps per year relative to a naive strategy based on the historical mean benchmark. What is more promising is the fact that our methodology can be easily extended to reveal predictable patterns in the higher order moments of the equity premium distribution like the variance, skewness and kurtosis. We expect that quantile forecasts of higher moments will be more precise than conventional ones and will enable the investor to gain a relatively complete picture of the expected return distribution which can be used for portfolio selection and asset pricing.

References

Ait-Sahalia Y, Brandt MW. 2001. Variable selection for portfolio choice. *Journal of Finance* **56:** 1297–1351.

Aiolfi M, Timmermann A. 2006. Persistence in Forecasting Performance and Conditional Combination Strategies. *Journal of Econometrics* **135**: 31–53.

Bai J, Ng S. 2002. Determining the Number of Factors in Approximate Factor Models.

Econometrica **70**: 191–221.

Bassett WG, Chen H-L. 2001. Portfolio style: Return-based attribution using quantile regression. *Empirical Economics* **26**: 293-305.

Bates JM, Granger CWJ. 1969. The combination of forecasts. *Operational Research Quarterly* **20**: 451–468.

Buchinsky M. 1994. Changes in U.S. Wage Structure 1963-1987: An application of Quantile Regression. *Econometrica* **62**: 405-458.

Buchinsky M. 1995. Quantile Regression Box-Cox Transformation model, and the U.S. wage structure, 1963-1987. *Journal of Econometrics* **65**: 109-154.

Buchinsky M. 1998. Recent Advances in Quantile Regression Models: A Practical Guidline for Empirical Research. *Journal of Human Resources* **33:** 88-126.

Campbell JY, Cochrane JH. 1999. By force of habit: A consumption-based explanation of aggregate stock market behavior. *The Journal of Political Economy* **107**: 205–251.

Campbell, JY, Thompson S B. 2008. Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? *Review of Financial Studies* **21**: 1509–31.

Campbell, JY, Viceira L. 2002. *Strategic Asset Allocation.* Oxford University Press, Oxford.

Cenesizoglu T., Timmermann A. 2008. Is the distribution of stock returns predictable?. Available online at SSRN http:// ssrn.com/abstract=1107185.

Cenesizoglu T., Timmermann A. 2012. Do Return Prediction Models Add Economic Value? *Journal of Banking and Finance* **36**(11): 2974-2987.

Chan YL, Stock JH, Watson MW 1999. A dynamic factor model framework for forecast combination. *Spanish Economic Review* **1**: 21–121.

Chen Q, Hong Y. 2010. Predictability of equity returns over different time horizons: a nonparametric approach. Manuscript, Cornell University.

Chong YY, Hendry DF. 1986. Econometric evaluation of linear macroeconomic models. *Review of Economic Studies* **53:** 671-690.

Chuang CC, Kuan CM, Lin H-Y. 2009. Causality in quantiles and dynamic stock return-volume relations. *Journal of Banking and Finance* **33**(7): 1351-1360.

Clements MP, Hendry DF. 1998. *Forecasting Economic Time Series.* Cambridge University Press, Cambridge.

Dangl T, Halling M. 2012. Predictive regressions with time-varying coefficients. *Journal of Financial Economics* **106:** 157-181.

Della Corte P, Sarno L, Tsiakas I. 2009. An Economic Evaluation of Empirical Exchange Rate Models. *Review of Financial Studies* **22:** 3491-3530.

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

Diebold FX, Mariano RS. 1995. Comparing Predictive Accuracy. *Journal of Business and Economic Statistics* **13:** 253–63.

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

Ferreira MI, Santa-Clara P. 2011. Forecasting stock market returns: the sum of the parts is more than the whole. *Journal of Financial Economics* **100**: 514–537.

Fleming J, Kirby C, Ostdiek B. 2001. The Economic Value of Volatility Timing. *Journal of Finance* **56:** 329-352.

Franses PH, van Dijk D. 2000. *Nonlinear Time Series Models in Empirical Finance.* Cambridge University Press, Cambridge, United Kingdom.

Gastwirth JL. 1966. On robust procedures. *Journal of the American Statistical Association* **61:** 929-948.

Goyal A, Welch I. 2008. A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. *Review of Financial Studies* **21:** 1455–508.

Granger CWJ, Newbold P. 1973. Some Comments on the Evaluation of Economic Forecasts. *Applied Economics* **5:** 35-47.

Granger CWJ, Ramanathan R. 1984. Improved methods of combining forecasts. *Journal of Forecasting* **3:** 197-204.

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

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

Hansen B. 2008. Least-squares forecast averaging. J*ournal of Econometrics* **146**: 342-350.

Harvey DI, Leybourne SJ, Newbold P. 1998. Tests for Forecast Encompassing. *Journal of Business and Economic Statistics* **16**: 254–59.

Harvey D, Newbold P. 2000. Tests for multiple forecast encompassing. *Journal of Applied Econometrics* **15**: 471–482.

Hastie T, Tibshirani R, Friedman J. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer.

Hendry DF, Clements MP. 2004. Pooling of Forecasts. *Econometrics Journal* **7:** 1–31.

Henkel SJ, Martin JS, Nadari F. 2011. Time-varying short-horizon predictability. *Journal of Financial Economics* **99**: 560–580.

Hsiao C, Wan SK. 2014. Is there an optimal forecast combination?. *Journal of Econometrics* **178**: 294-309.

Judge GG, Hill RC, Griffiths WE, Lutkepohl H, Lee T-C. 1988. *Introduction to the Theory and Practice of Econometrics.* Wiley, New York.

Koenker R, Bassett G. 1978. Regression Quantiles. *Econometrica* **46**: 33-50.

Ludvigson SC, Ng S. 2007. The empirical risk-return relation: a factor analysis approach. *Journal of Financial Economics* **83:** 171–222.

Ma L, Pohlman L. 2008. Return forecasts and optimal portfolio construction: a quantile regression approach. *European Journal of Finance* **14(**5): 409–425.

Marquering W, Verbeek M. 2004. The Economic Value of Predicting Stock Index Returns and Volatility. *Journal of Financial and Quantitative Analysis* **39**: 407–29.

Meligkotsidou L, Vrontos ID, Vrontos SD. 2009. Quantile Regression Analysis of Hedge Fund Strategies. *Journal of Empirical Finance* **16**: 264-279.

Menzly L, Santos T, Veronesi P. 2004. Understanding Predictability. *Journal of Political Economy* **112(**1): 1-47.

Neely CJ, Rapach DE, Tu J, Zhou G. 2013. Forecasting the equity risk premium: the role of technical indicators. *Management Science*, forthcoming.

Pedersen, TQ. 2010. Predictable return distributions. Available online at SSRN http://ssrn.com /abstract=1658394.

Rapach D, Strauss J. 2012. Forecasting U.S. State-Level Employment Growth: An Amalgamation Approach. *Journal of International Forecasting* **28:** 315–327.

Rapach D, Strauss J, Zhou G. 2010. Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy. *Review of Financial Studies* **23**(2): 821-862.

Rapach D, Zhou G. 2013. Forecasting stock returns. *Handbook of Economic Forecasting* Amsterdam: Elsevier; Vol. 2A, 328-383.

Stock JH, Watson MW. 2004. Combination Forecasts of Output Growth in a Seven-Country Data Set. *Journal of Forecasting* **23**: 405–30.

Taylor JW 2007. Forecasting daily supermarket sales using exponentially weighted quantile regression. *European Journal of Operational Research* **178**: 154-167.

Terasvirta T. 2006. Forecasting economic variables with nonlinear models. In *Handbook of Economic Forecasting* Vol. I, 413-457, Amsterdam, Elsevier.

Tibshirani R. 1996. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society* Series B, **58**: 267-288.

Timmermann A. 2006. Forecast combinations. In *Handbook of Economic Forecasting* Vol. I, 135-196, Amsterdam, Elsevier.

Tukey JW. 1977. *Explanatory Data Analysis*. Addison-Wesley, Reading, MA (1977).

Yu K, Lu Z, Stander J. 2003. Quantile regression: applications and current research areas. *The Statistician*, **52**: 331-350.

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

West K, Edison H, Cho D. 1993. A Utility-based Comparison of Some Models of Exchange Rate Volatility. *Journal of International Economics*, **35**: 23-46.

White H. 2006. Approximate nonlinear forecasting methods. In *Handbook of Economic Forecasting*, Amsterdam, Elsevier; Vol. I, 459-512.

Wu Y, Liu Y. 2009. Variable selection in quantile regression. *Statistica Sinica*, **19:** 801-817.

Table 1. Conditional Quantile Predictive Ability

Panel A: Individual predictive models

Predictor	**Q5**	**Q10**	**Q15**	**Q20**	**Q25**	**Q30**	**Q35**	**Q40**	**Q45**	**Q50**	**Q55**	**Q60**	**Q65**	**Q70**	**Q75**	**Q80**	**Q85**	**Q90**	**Q95**
D/P																			
D/Y																			
E/P																			
D/E																			
SVAR																			
B/M																			
NTIS																			
TBL																			
LTY																			
LTR																			
TMS																			
DFY																			
DFR																			
INFL																			
I/K																			

Panel B: Combination Methods

Mean																			
Median																			
Trimmed Mean																			
DALFE(1)																			
DALFE(0.9)																			
AL Cluster 2																			
AL Cluster 3																			
AL Principal Components																			
AL Lasso																			
AL Ridge																			

Notes: Q5- Q95 denote the 5% to 95% quantiles of the return distribution. Grey cells denote superior predictive ability, i.e. lower out-of-sample values of the expected asymmetric loss, $\sum_t \rho_\tau \left(r_{t+1} - \hat{r}_{t+1}^{(\tau)} \right)$, associated with the quantile forecasts of each model specification or combination method (shown in the first column of the table), than the value associated with the forecasts of the prevailing quantile (PQ) model.

Table 2. Out-of-sample performance of individual predictive mean regression models and Mean Forecast Combination (MFC) approach

Predictor	MSFE Ratio	λ_A	λ_B	MFC approach	MSFE Ratio	λ_A	λ_B
D/P	0.9928	0.580**	0.420	Mean	0.9703	2.597***	-1.597
D/Y	0.9900	0.590**	0.410	Median	0.9781	3.209***	-2.209
E/P	1.0109	0.287	0.713	Trimmed Mean	0.9715	2.943***	-1.943
D/E	1.0160	0.177	0.824*	DMSFE(1)	0.9704	2.463***	-1.463
SVAR	1.0665	0.110	0.890	DMSFE(0.9)	0.9702	2.444***	-1.444
B/M	1.0180	0.089	0.911**	Cluster 2	0.9766	1.244**	-0.244
NTIS	1.0210	-0.192	1.192**	Cluster 3	0.9878	0.766*	0.234
TBL	1.0243	0.406**	0.594**	Principal Components	1.0169	0.347	0.653*
LTY	1.0259	0.360	0.640**				
LTR	1.0115	0.261	0.739*				
TMS	1.0265	0.373*	0.627***				
DFY	1.0271	-0.398	1.398**				
DFR	0.9909	0.627	0.373				
INFL	1.0076	0.340	0.660				
I/K	0.9768	0.693***	0.307				

Notes: The table reports the MSFE ratios of the individual predictive mean regression models and of the Mean Forecast Combination (MFC) approach with respect to the prevailing mean (PM) benchmark model for the out-of-sample period 1965:1-2010:4. The MSFE of the PM model is equal to 0.0071. Values of the MSFE ratio below unity indicate superior forecasting performance of the predictive models with respect to the historical average forecast. Statistical significance of the out-of-sample forecasts is assessed by pairs of encompassing tests: (i) one for testing if the PM model forecasts encompass the forecasts of the individual predictive models or the MFC approach (associated with the parameter λ_A), and (ii) a second one for testing if the individual predictive models' or the MFC approach's forecasts encompass the PM model forecasts (associated with the parameter λ_B). *, **, *** indicate significance at the 10%, 5% and 1% confidence levels, respectively.

Table 3. Out-of-sample performance of the Robust Forecast Combination (RFC) approach and the Quantile Forecast Combination (QFC) approach

	MSFE Ratio	λ_A	λ_B	MSFE Ratio	λ_A	λ_B	MSFE Ratio	λ_A	λ_B	MSFE Ratio	λ_A	λ_B
Panel A: Robust Forecast Combination (RFC) approach												
	RFC-FW1			**RFC-FW2**			**RFC-FW3**			**RFC-FW4**		
Mean	0.9761	2.050**	-1.050	0.9768	1.861**	-0.861	0.9741	2.456***	-1.456	0.9720	3.144***	-2.144
Median	0.9865	1.458*	-0.458	0.9893	1.257*	-0.257	0.9848	1.850**	-0.850	0.9794	3.600***	-2.600
Trimmed Mean	0.9778	2.255**	-1.255	0.9786	2.057**	-1.057	0.9761	2.751***	-1.751	0.9737	3.757***	-2.757
DMSFE(1)	0.9755	2.081**	-1.081	0.9763	1.878**	-0.878	0.9737	2.441***	-1.441	0.9719	2.986***	-1.986
DMSFE(0.9)	0.9747	2.022**	-1.022	0.9760	1.814**	-0.814	0.9731	2.343**	-1.343	0.9716	2.842***	-1.842
Cluster 2	0.9726	1.446**	-0.446	0.9778	1.280**	-0.280	0.9744	1.394**	-0.394	0.9769	1.317**	-0.317
Cluster 3	1.0059	0.393	0.608	0.9992	0.517	0.484	1.0017	0.466	0.534	0.9861	0.791*	0.209
Prin. Components	1.0289	0.317	0.683**	1.0256	0.332	0.668**	1.0284	0.318	0.682**	1.0287	0.295	0.705**
	RFC-TVW1			**RFC-TVW2**			**RFC-TVW3**					
Mean	0.9635	2.829***	-1.829	0.9654	2.907***	-1.907	0.9633	1.817***	-0.817			
Median	0.9718	3.756***	-2.756	0.9760	5.199***	-4.199	0.9669	1.660***	-0.660			
Trimmed Mean	0.9650	3.037***	-2.037	0.9677	3.314***	-2.314	0.9667	1.730***	-0.730			
Panel B: Quantile Forecast Combination (QFC) Approach												
	QFC -FW1			**QFC -FW2**			**QFC -FW3**			**QFC -FW4**		
Mean	0.9761	2.050**	-1.050	0.9768	1.861**	-0.861	0.9741	2.456***	-1.456	0.9720	3.144***	-2.144
Median	0.9886	1.354*	-0.354	0.9903	1.208	-0.208	0.9866	1.768**	-0.768	0.9830	3.498***	-2.498
Trimmed Mean	0.9785	2.214**	-1.214	0.9791	2.044**	-1.044	0.9768	2.727***	-1.727	0.9746	3.808***	-2.808
DALFE(1)	0.9758	2.047**	-1.047	0.9766	1.846**	-0.846	0.9738	2.430***	-1.430	0.9719	3.043***	-2.043
DALFE(0.9)	0.9752	2.011**	-1.011	0.9760	1.825**	-0.825	0.9731	2.419***	-1.419	0.9711	3.092***	-2.092
AL Cluster 2	0.9768	1.331**	-0.331	0.9809	1.115**	-0.115	0.9754	1.489**	-0.489	0.9733	1.864**	-0.864
AL Cluster 3	0.9798	0.965*	0.035	0.9753	1.084**	-0.084	0.9787	1.051**	-0.051	0.9785	1.237**	-0.237
AL Prin. Comp.	1.0079	0.448*	0.552*	1.0160	0.417*	0.583**	1.0062	0.456*	0.544*	1.0181	0.383**	0.617**
AL Lasso	0.9777	0.747**	0.253	0.9899	0.592**	0.408	0.9782	0.755**	0.245	0.9866	0.657**	0.343
AL Ridge	0.9696	1.157**	-0.157	0.9719	1.022**	-0.022	0.9680	1.234**	-0.234	0.9705	1.215**	-0.215
	QFC -TVW1			**QFC -TVW2**			**QFC -TVW3**					
Mean	0.9594	2.138***	-1.138	0.9619	2.553***	-1.553	0.9677	1.387**	-0.387			
Median	0.9669	2.748***	-1.748	0.9717	3.736***	-2.736	0.9746	1.495**	-0.495			
Trimmed Mean	0.9619	2.216***	-1.216	0.9648	2.799***	-1.799	0.9702	1.382**	-0.382			

Notes: The table reports the MSFE ratios of the Robust Forecast Combination (RFC) and the Quantile Forecast Combination (QFC) approach, under fixed weighting (FW) and time-varying weighting (TVW) schemes, with respect to the prevailing mean (PM) benchmark model for the out-of-sample period 1965:1-2010:4. Values of the MSFE ratio below unity indicate superior forecasting performance of the predictive models with respect to the historical average forecast. Statistical significance of the out-of-sample forecasts is assessed by pairs of encompassing tests: (i) one for testing if the PM model forecasts encompass the RFC or QFC forecasts (associated with the parameter λ_A), and (ii) a second one for testing if the RFC or QFC forecasts encompass the PM model forecasts (associated with the parameter λ_B). *, **, *** indicate significance at the 10%, 5% and 1% confidence levels, respectively.

Table 4. Encompassing tests and amalgam forecasts

Panel A: Encompassing tests for pairs of forecasts from the MFC, RFC and QFC approaches

	Mean Combination		**Median Combination Method**		**Trimmed Mean Combination Method**	
	λ_A	λ_B	λ_A	λ_B	λ_A	λ_B
MFC, RFC-TVW1	6.262**	-5.262	3.069	-2.069	5.492*	-4.492
MFC, RFC-TVW2	2.373	-1.373	1.291	-0.291	1.964	-0.964
MFC, RFC-TVW3	1.704	-0.704	1.194*	-0.194	1.174	-0.174
MFC, QFC-TVW1	3.625**	-2.625	3.647**	-2.647	3.179*	-2.179
MFC, QFC-TVW2	4.552**	-3.552	3.280*	-2.280	3.520*	-2.520
MFC, QFC-TVW3	0.723	0.277	0.771	0.229	0.606	0.394
RFC-TVW1, QFC-TVW1	1.827	-0.827	2.509	-1.509	1.672	-0.672
RFC-TVW2, QFC-TVW2	2.250	-1.250	2.611	-1.611	2.102	-1.102
RFC-TVW3, QFC-TVW3	-3.394	4.394*	-1.573	2.573*	-3.023	4.023

Panel B: Multiple encompassing tests and amalgam forecasts

	MS* **MFC**	***MS**** **RFC-TVW**	***MS**** **QFC-TVW**	**MSFE Ratio**	λ_A	λ_B
Mean Combination Methods						
MFC, RFC-TVW1, QFC-TVW1	1.981	1.644	1.310	0.9640	2.577**	-1.577
MFC, RFC-TVW2, QFC-TVW2	1.418	1.174	0.875	0.9655	2.807**	-1.807
MFC, RFC-TVW3, QFC-TVW3	3.549**	3.381**	3.669**	0.9661	1.907**	-0.907
Median Combination Methods						
MFC, RFC-TVW1, QFC-TVW1	1.642	1.293	0.880	0.9676	3.176**	-2.176
MFC, RFC-TVW2, QFC-TVW2	0.999	0.744	0.489	0.9683	1.940**	-0.940
MFC, RFC-TVW3, QFC-TVW3	1.486	0.522	1.329	0.9827	2.529**	-1.529
Trimmed Mean Combination Methods						
MFC, RFC-TVW1, QFC-TVW1	1.553	1.194	0.944	0.9746	3.172**	-2.172
MFC, RFC-TVW2, QFC-TVW2	0.955	0.729	0.495	0.9737	3.508**	-2.508
MFC, RFC-TVW3, QFC-TVW3	1.209	0.920	1.229	0.9728	3.760**	-2.760

Notes: The table reports results on the encompassing tests for all pairs of forecasts obtained by the Mean Forecast Combination (MFC) approach, the Robust Forecast Combination (RFC) approach and the Quantile Forecast Combination (QFC) approach, as well as results on multiple encompassing tests employed to compare the forecasts obtained by the MFC, the RFC and the QFC approach, under the three time-varying weighting schemes (TVW1-TVW3). In Panel A, for each pair of approaches, shown in the first column, statistical significance of the out-of-sample forecasts is assessed by pairs of encompassing tests: (i) one for testing if the forecasts produced by the first approach encompass the forecasts produced by the second (associated with parameter λ_A), and (ii) a second one for testing if the forecasts produced by the second approach encompass the forecasts produced by the first (associated with parameter λ_B). In Panel B, columns (2) - (4) report the *MS** statistics to test the null of multiple forecast encompassing. The test is conducted three times for every triad by employing the model in the first row as the reference model. Columns (5) - (7) report the MSFE ratios of an amalgam forecast constructed by averaging the forecasts of the three approaches, shown in the first column. Statistical significance of the out-of-sample forecasts is assessed by pairs of encompassing tests: (i) one for testing if the amalgam forecasts encompass the PM forecasts (associated with the parameter λ_A), and (ii) a second one for testing if the PM forecasts encompass the amalgam forecasts (associated with the parameter λ_B). *, **, *** indicate significance at the 10%, 5% and 1% confidence levels, respectively.

Table 5. Economic evaluation

Panel A: Mean Forecast Combination & Robust Forecast Combination

	MFC	RFC-FW1	RFC-FW2	RFC-FW3	RFC-FW4	RFC-TVW1	RFC-TVW2	RFC-TVW3
Mean	297.41	186.79	190.6	207.32	236.59	371.49	338.90	394.85
Median	145.30	59.01	18.42	76.89	108.40	231.73	158.60	327.55
Trimmed Mean	270.86	159.17	162.38	178.16	216.80	352.71	306.40	373.76
DMSFE(1)	304.75	204.59	206.26	223.17	249.46			
DMSFE(0.9)	320.79	245.70	242.50	258.98	275.16			
Cluster 2	248.87	252.01	229.06	236.64	217.11			
Cluster 3	242.84	175.07	224.49	216.39	263.24			
Principal Components	235.71	250.88	255.38	252.54	238.68			

Panel B: Quantile Forecast Combination

	QFC-FW1	QFC-FW2	QFC-FW3	QFC-FW4	QFC - TVW1	QFC - TVW2	QFC - TVW3
Mean	186.79	190.60	207.32	236.59	424.97	383.53	358.33
Median	28.75	14.76	45.79	83.49	314.28	242.92	294.50
Trimmed Mean	157.37	161.32	176.63	211.95	405.73	356.23	342.67
DALFE(1)	195.92	197.97	215.21	242.23			
DALFE(0.9)	220.51	223.47	240.17	263.13			
AL Cluster 2	250.28	243.45	248.28	261.00			
AL Cluster 3	276.53	338.32	270.11	257.20			
AL Principal	262.00	289.17	217.64	173.78			
AL Lasso	328.87	317.67	292.53	226.94			
AL Ridge	316.86	310.97	314.81	295.4			

Panel C: Amalgam Forecasts

	Mean Combination Method	Median Combination Method	Trimmed Mean Combination Method
MFC, RFC-TVW1, QFC-TVW1	372.38	316.12	198.58
MFC, RFC-TVW2, QFC-TVW2	346.79	352.95	214.42
MFC, RFC-TVW3, QFC-TVW3	374.93	79.06	235.60

Notes: The table reports the performance fee, Φ, which is the difference between the realized utilities of competing models, $\Phi = \Delta U = \overline{U}^{i} - \overline{U}^{PM}$, where $\overline{U}^{i}, \overline{U}^{PM}$ denote the average mean-variance utility of an investor with a risk aversion coefficient of three over the forecast evaluation period from using the *i*th model/specification and the historical average benchmark model (PM), respectively. The weight on stocks in the investor's portfolio is restricted to lie between zero and 1.5. The mean-variance utility for the *i*th model/specification is given by:

$$\overline{U}^{i} = \frac{W_0}{P-P_0}\left[\sum_{t=0}^{P-P_0-1}(R^{i}_{p,t+1}) - \frac{\gamma}{2}\sum_{t=0}^{P-P_0-1}(R^{i}_{p,t+1} - \overline{R}^{i}_{p})^2\right],$$

where $P-P_0$ is the number of out-of-sample forecasts, W_0 is the initial wealth of the investor and γ denotes the coefficient of relative risk aversion. Φ is reported in annualized basis points.