A Quantile Regression Approach to Equity Premium Prediction

Loukia Meligkotsidou[a], Ekaterini Panopoulou[b],

Ioannis D.Vrontos[c]*, Spyridon D. Vrontos[b]

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

[b]Department of Statistics and Insurance Science, University of Piraeus, Piraeus, Greece

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

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, thus exploiting the entire distributional information associated with each predictor. Further gains are achieved by incorporating the forecast combination methodology in 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 mean predictive regression modeling approach.

JEL classification: G11; G12; C22; C53

Keywords: Conditional Quantiles; Equity Premium; Forecast Combination; Prediction; Time varying weights.

*Corresponding author. Department of Statistics, Athens University of Economics and Business, Patission 76, GR-10434 Athens, Greece. Tel.: +30-210-8203927. Email: vrontos@aueb.gr

1 Introduction

Equity premium predictability has attracted the attention of both academics and practitioners in finance. Results are mixed, since different techniques, variables and time periods are employed in the related research. The list of predictors is quite exhaustive and typically contains valuation ratios, various interest rates and spreads, distress indicators, inflation rates along with other macroeconomic variables, indicators of corporate activity, etc.[1] The early contributions to equity premium predictability mainly focused on the in-sample predictive ability of the potential predictors and the development of 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) show that their long list of predictors can not deliver consistently superior out-of-sample performance. The authors employ a variety of predictive regression models ranging from single variable ones to their 'kitchen sink' model that contains all their predictors simultaneously. Campbell and Thompson (2008) show that when imposing simple restrictions, suggested by economic theory, on predictive regressions' coefficients, the out-of-sample performance improves and market timing strategies can deliver profits to investors (see also Ferreira and Santa-Clara (2011)). More recently, Rapach, Strauss and Zhou (2010) consider 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 reducing model uncertainty/parameter instability, significantly beat the historical average forecast. Finally, Ludvigson and Ng (2007) and Neely,

[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)). Lettau and Ludvigson (2001) finds that the consumption to wealth ratio helps equity premium predictability, while corporate financing activity is exploited in Baker and Wurgler (2000). 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.

Rapach, Tu and Zhou (2011) adopt a diffusion index approach, which can conveniently track the key movements in a large set of predictors, and they find evidence of improved equity premium forecasting ability.

It still remains an open question whether there is clear evidence for predictability of the equity premium. Note that all the above regression specifications for the equity premium prediction can only model the conditional expectation of returns. That is, standard regression models describe only the average relationship of returns with the set of predictors. However, this approach might not be adequate for exploring equity premium predictability, since it can not reveal the predictive ability of various predictors at forecasting the entire distribution of returns. Looking at just the conditional mean of the return series may 'hide' interesting characteristics. For example, it can lead us to conclude that a predictor has poor predictive performance, while it is actually valuable for predicting the lower or/and the upper conditional quantiles of returns. To explore this possibility, we consider predictive quantile regression models for equity premium forecasting.

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, 1998) and Yu, Lu and Stander (2003) among others. Applications in the field of finance include work on risk measures (Taylor (1999), Chernozhukov and Umantsev (2001), Engle and Manganelli (2004)), asset management (Feng, Chen and Basset (2008), Ma and Pohlman (2008)) and the analysis of the cross section of stock market returns (Barnes and Hughes (2002)). Bassett and Chen (2001) propose the quantile regression method as an appropriate way for the classification of mutual fund investment styles and Meligkotsidou, Vrontos and Vrontos (2009) consider quantile regression models for hedge fund pricing allowing for model uncertainty. Chuang, Kuan and Lin (2009) examine the dynamic relationship of the quantiles of stock returns and trading volume and find an heterogeneous causal effect of volume across quantiles. Recently, Baur, Dimpfl and Jung (2012) provide a comprehensive description of the dependence pattern of stock returns by studying a range of quantiles of the conditional return distribution and find that lower quantiles exhibit positive dependence on past returns while upper quantiles are marked by negative dependence.

The paper more closely related to the present paper is 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) point out that return prediction models that allow for a time-varying return distribution lead to better estimates of the tails of the returns' distribution and suffer less from unanticipated outliers. Their empirical findings suggest that time-varying distribution forecasts can lead to an annual risk-adjusted return performance of 2% higher than a constant mean and volatility model. Similar conclusions are reached by Pedersen (2010) who employs both univariate and multivariate quantile regressions to jointly model the distribution of stocks and bonds.

The aim of our paper is to produce robust and accurate point forecasts of the equity premium, using both a fixed and a time-varying weighting scheme, based on the quantile forecasts obtained from a set of predictive quantile regressions. To this end, we utilize two different sources of information; distribution information, regarding how the relationship between the equity premium and a given predictive variable varies across the conditional quantiles of returns, as well as predictor information, regarding the different models that can be used for predictive inference. To take both sources of information into account we propose a two-stage approach. One stage is designed to construct single point forecasts of the equity premium from a set of quantile forecasts. At the other stage the forecasts obtained from different model specifications are combined in order to reduce uncertainty risk associated with a single predictive variable. Finally, we examine whether the order of applying the above two stages affects the performance of the proposed forecasting approach. 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 of combining individual forecasts from single predictor linear models. All different forecasts are evaluated against the benchmark 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 considered. More importantly, no single predictor proves successful in capturing the entire return distribution with the right tail being easier to predict in contrast to the left tail. Overall, superior predictive performance, both in statistical and economic evaluation terms, is achieved under the quantile regression approach as follows. First, various quantiles of the conditional distribution of returns are optimally predicted by combining information from different predictors using one of the existing forecast combination methods. Next, robust point forecasts of the equity premium are produced using time-varying weighting schemes.

The remainder of the paper is organized as follows. Section 2 describes the econometric models considered in this study, including standard conditional mean and quantile regression models. Section 3 outlines our proposed methodology for robust estimation of the central location of the distribution of returns. Section 4 discusses various methods of combining forecasts from different regression specifications, as well as from different conditional quantiles. Our dataset and the framework for forecast evaluation is presented in Section 5, while our empirical results are reported in Section 6. Section 7 outlines the economic evaluation framework and presents the associated findings. Section 8 summarizes and concludes.

2 Predictive Regressions

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

2.1 Quantile Regression Models

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

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

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_{it} 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, Strauss and Zhou, 2010), which suggests that the conditional mean of the random variable r_{t+1} given x_{it} is $E(r_{t+1} \mid x_{it}) = \alpha_i + \beta_i x_{it}$.

The above regression specifications can only model the conditional expectation and not the entire conditional distribution of returns. However, it is reasonable to believe that the effects of the predictors differ across quantiles of returns, especially if the return distribution deviates from normality. Equity premium returns exhibit a high degree of non-normality, fat-tails and skewness and, therefore, a more sophisticated approach to predictive inference on the return distribution should be useful. For this reason, we propose using quantile regression models (Koenker and Bassett (1978), Buchinsky

(1998), Yu, Lu and Stander (2003)) to predict the entire distribution of equity premium via modeling a set of conditional quantiles. It is well known that quantile regression estimators are more efficient and more robust than conditional mean regression estimators in cases that there exist deviations from normality. Information from different predictive quantile regression can be utilized with the aim to construct a robust and more accurate point forecast.

We consider quantile regression models with a single predictor of the form

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

where $\tau \in (0, 1)$ denotes the τth quantile of r_{t+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 conditional 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 coefficient $\beta_i^{(\tau)}$ shows how the ith variable predicts the equity premium at the level of the τth quantile. The $\beta_i^{(\tau)}$'s are likely to be different for different τ's, revealing a larger amount of information about returns than conditional mean regression, especially if the error distribution is not symmetric.

2.2 Inference on Predictive Regression Models

In this subsection we briefly present the problem of estimating the above predictive regression models. The conditional mean regression model can be estimated using the ordinary least squares method without making any particular assumptions for the error distribution. Least squares estimation is based on the fact that the expectation of a random variable r with distribution function F arises as the point estimate of r corresponding to the quadratic loss function $\rho(u) = u^2$, that is 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 ordinary least squares (OLS) estimators $\hat{\alpha}_i, \hat{\beta}_i$ of the parameters in the conditional mean regression models in (1) can be estimated by minimizing the sample estimate of the quadratic expected loss $T^{-1}\sum_{t=0}^{T-1}\rho\left(r_{t+1}-a_i-\beta_i x_{it}\right)$, with respect to α_i, β_i, or equivalently by minimizing the sum of squares[3]

$$\sum_{t=0}^{T-1}\left(r_{t+1}-\alpha_i-\beta_i x_{it}\right)^2.$$

A parametric approach to inference on the regression parameters can be followed if the functional form of the error distribution is specified. Most commonly the errors are assumed to follow the normal distribution with zero mean and variance σ^2, in which case the maximum likelihood estimates (MLEs) of the regression parameters are identical to the OLS estimates, $\hat{\alpha}_i, \hat{\beta}_i$. Then, the point forecast of the equity premium at time $t+1$, based on the ith model specification, is obtained as

$$\hat{r}_{i,t+1} = \hat{\alpha}_i + \hat{\beta}_i x_{it}.$$

Similarly to the expectation of the random variable r, its 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,

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

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

More in detail, 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. If a sample, $r_1, ..., r_T$, is drawn from F, an estimate of the τth quantile of r is obtained by minimizing the sample estimate of the expected loss, i.e. the function $T^{-1}\sum_{t=1}^{T}\rho_\tau\left(r_t - \bar{r}^{(\tau)}\right)$. The above idea can be used to estimate the parameters, ${\alpha_i}^{(\tau)}, \beta_i^{(\tau)}$, of the linear quantile regression models in (2). This can be done by minimizing the sum

$$\sum_{t=0}^{T-1}\rho_\tau\left(r_{t+1} - {\alpha_i}^{(\tau)} - \beta_i^{(\tau)}x_{it}\right), \tag{4}$$

where the check function $\rho_\tau(u)$ has been given in (3).

A parametric approach to inference on the quantile regression parameters arises if the error distribution $g_\tau(\varepsilon)$ is specified. The error distribution that has been widely used for parametric inference in the quantile regression literature is the asymmetric Laplace distribution (for details, see Yu and Moyeed (2001) and Yu and Zhang (2005)). The advantage of this assumption is that the respective MLEs are identical to the estimates obtained by minimizing (4) and, therefore, inherit their asymptotic properties (for details on these properties see Koenker (2005)). Once the model parameters have been estimated, the forecast of the τth conditional quantile of the distribution of the equity premium at time $t+1$, based on the ith model specification, is obtained as

$$\hat{r}_{i,t+1}(\tau) = \hat{\alpha_i}^{(\tau)} + \hat{\beta}_i^{(\tau)}x_{it}.$$

In the subsequent sections we outline our methodology for producing robust and accurate point forecasts of the equity premium based on the above quantile forecasts. In this respect, we utilize two different sources of information; distribution information, regarding how the relationship between the equity premium and a given predictive variable varies across the conditional quantiles of returns, as well as predictor information, regarding the different models that can be used for predictive inference. To take account of both sources of information we propose a two-stage approach. One stage is designed to construct single point forecasts of the equity premium from a set of quantile forecasts. This is done by developing a fixed and a time-varying weighting scheme presented in Section 3. At the other stage the forecasts obtained from different model specifications are combined in order to reduce uncertainty risk associated with a single predictive variable (see Section 4). Finally, we examine possible impacts of the order of applying the above two stages to the performance of the proposed forecasting approach.

3 Robust Point Forecasts based on Regression Quantiles

In this section we consider the problem of finding robust point forecasts of the equity premium as an alternative to the standard approach which produces forecasts based on the conditional expectation model. As already mentioned, the special characteristics of the data, i.e. high volatility, fat tails, skewness and deviation from normality, render the standard conditional expectation approach inadequate for prediction of financial series. These special characteristics of financial series motivate the development of alternative forecasting methods that are robust to the distributional assumptions of asset returns. In such cases it seems reasonable to choose estimators/predictors which modify the conditional expectation prediction approach by putting reduced weight on extreme observations.

The need for robust alternatives to the conditional expectation approach has been apparent in the literature. The median, the trimmed mean and linear combinations of functions of a few quantiles (L-estimators) often provide efficient and robust, but simple, alternative ways of estimating or forecasting the location of a random variable. Gastwirth (1966) introduced a class of robust estimators of the location using a few sample quantiles ('quick estimators'), which do not weight the extreme quantiles too heavily. This type of estimators turns out to have good efficiency properties for a wide variety of distributions. Judge, Hill, Griffiths, Lutkepohl and Lee (1988) suggested an alternative five-quantile estimator attaching weight on extreme positive and negative events. Koenker and Basset (1978) suggest that robust point estimates of the central location of a distribution can be constructed as weighted averages of quantile estimators. In this spirit, several quantile forecasts $\hat{r}_{i,t+1}(\tau)$ can be combined to produce a single robust point forecast. All these approaches use a constant/fixed weighting scheme of different quantile forecasts to construct robust point forecasts.

Relaxing the assumption of a constant weighting scheme in the generation of point forecasts 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 (see also Dangl and Halling (2012)) can motivate the employment of time-varying schemes in the generation of robust point forecasts. In general, models with time-varying coefficients generate return predictions that are consistent with business cycle related patterns implied by asset pricing theory (e.g., Campbell and Cochrane (1999), Menzly, Santos, and Veronesi (2004)). The empirical evidence suggests that on average, predicted equity risk premia increase during a recession and peak around the trough. During expansions, predicted risk premia decrease and reach their lowest levels around the peak of the business cycle. Finally, an investor who relies on these predictions times the market very well, reducing her exposure around the peak of the business cycle and moving back into the market before the trough. Intuitively, in our setup, allowing for time-varying weights coincides with changing investors' expectations on the relative future outcome.

3.1 Point Forecasts based on a fixed weighting scheme

Point forecasts of the equity premium can be constructed as weighted averages of a set of quantile forecasts. First we employ standard estimators with fixed, prespecified, weights of the form

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

where S denotes the set of quantiles that will be used. 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. The above point forecasts incorporate information about how the effects of predictors vary across the distribution of returns. Therefore, they are robust and can be more accurate than the conditional mean forecasts, especially in cases that the return distribution clearly deviates from normality.

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

$$\text{FW1:} \qquad \widehat{r}_{i,t+1} = 0.25\hat{r}_{i,t+1}(0.25) + 0.50\hat{r}_{i,t+1}(0.5) + 0.25\hat{r}_{i,t+1}(0.75) \tag{5}$$

$$\text{FW2:} \qquad \widehat{r}_{i,t+1} = 0.3\hat{r}_{i,t+1}(1/3) + 0.4\hat{r}_{i,t+1}(0.5) + 0.3\hat{r}_{i,t+1}(2/3) \tag{6}$$

Furthermore, we use the alternative five-quantile estimator, suggested by Judge, Hill, Griffiths, Lutkepohl and Lee (1988), which attaches more weight on extreme positive and negative events as follows

$$\begin{aligned} \text{FW3: } \widehat{r}_{i,t+1} &= 0.05\hat{r}_{i,t+1}(0.10) + 0.25\hat{r}_{i,t+1}(0.25) + 0.40\hat{r}_{i,t+1}(0.5) \\ &+ 0.25\hat{r}_{i,t+1}(0.75) + 0.05\hat{r}_{i,t+1}(0.90) \end{aligned} \tag{7}$$

Additionally to the above three estimators, we consider a fourth one which combines information from a larger set of conditional quantiles. Specifically, we consider the following formula

$$\text{FW4: } \widehat{r}_{i,t+1} = 0.05\hat{r}_{i,t+1}(0.50) + 0.05\sum_{\tau\in S}\hat{r}_{i,t+1}(\tau), \tag{8}$$

where $S = \{0.05, 0.10, ..., 0.95\}$. All the above fixed-weights point forecasts (FW1-FW4) are estimators of the expected value of the return at time $t+1$, constructed using information from different parts of the return distribution. A subset of the above specifications has been employed by Taylor (2007) and Ma and Pohlman (2008) among others.

3.2 Point Forecasts based on a time-varying weighting scheme

Apart from generating forecasts based on fixed weights, we consider point forecasts based on time-varying weighting schemes. These weights are derived by some optimization procedure aiming at producing an empirical model that allows for economic changes over time and that is also capable of determining the 'right' parameter values in time to help investors (Spiegel (2008)). Our approach produces time-varying weighting schemes which combine different conditional quantiles of returns based on minimizing the mean square forecast error under reasonable constraints.

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

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

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 square forecast errors, $E(r_{t+1} - \hat{r}_{i,t+1})^2$, under an appropriate set of constraints.[4] Our optimization procedure is the analogue of the constrained Granger - Ramanathan method for quantile regression forecasts (Timmermann (2006), Granger and Ramanathan (1984), Hansen (2008) and Hsiao and Wan (2012)). These time-varying weights of the quantile estimates bear an interesting relationship to the portfolio weight constraints in finance (see Timmermann (2006)). In this sense the weights of the quantile estimates are constrained 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. 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. More specifically, FW1 with time-varying coefficients becomes

$$\text{TVW1: } \widehat{r}_{i,t+1} = p_{0.25,t}\hat{r}_{i,t+1}(0.25) + p_{0.50,t}\hat{r}_{i,t+1}(0.5) + p_{0.75,t}\hat{r}_{i,t+1}(0.75) \tag{9}$$

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

[4] Alternative loss functions can be also considered within our optimization procedure.

dure

$$\mathbf{p}_t = \arg\min_{\mathbf{p}_t} E[r_{t+1} - (p_{0.25,t}\hat{r}_{i,t+1}(0.25) + p_{0.50,t}\hat{r}_{i,t+1}(0.5) + p_{0.75,t}\hat{r}_{i,t+1}(0.75))]^2$$

$$s.t.\ p_{0.25,t} + p_{0.50,t} + p_{0.75,t} = 1$$

$$0.20 \leq p_{0.25,t} \leq 0.40$$

$$0.40 \leq p_{0.50,t} \leq 0.60$$

$$0.20 \leq p_{0.75,t} \leq 0.40$$

Similarly, the FW2 model with time-varying coefficients becomes

$$\text{TVW2: } \widehat{r}_{i,t+1} = p_{1/3,t}\hat{r}_{i,t+1}(1/3) + p_{0.5,t}\hat{r}_{i,t+1}(0.5) + p_{2/3,t}\hat{r}_{i,t+1}(2/3) \tag{10}$$

where $p_{\tau,t}, \tau \in S = \{1/3, 0.5, 2/3\}$ are estimated by the following optimization procedure

$$\mathbf{p}_t = \arg\min_{\mathbf{p}_t} E[r_{t+1} - (p_{1/3,t}\hat{r}_{i,t+1}(1/3) + p_{0.5,t}\hat{r}_{i,t+1}(0.5) + p_{2/3,t}\hat{r}_{i,t+1}(2/3))]^2$$

$$s.t.\ p_{1/3,t} + p_{0.50,t} + p_{2/3,t} = 1$$

$$0.15 \leq p_{1/3,t} \leq 0.45$$

$$0.30 \leq p_{0.5,t} \leq 0.50$$

$$0.15 \leq p_{2/3,t} \leq 0.45$$

Finally, the FW3 model with time-varying coefficients becomes

$$\begin{aligned} \text{TVW3: } \widehat{r}_{i,t+1} &= p_{0.10,t}\hat{r}_{i,t+1}(0.10) + p_{0.25,t}\hat{r}_{i,t+1}(0.25) + p_{0.5,t}\hat{r}_{i,t+1}(0.5) \\ &+ p_{0.75,t}\hat{r}_{i,t+1}(0.75) + p_{0.90,t}\hat{r}_{i,t+1}(0.90) \end{aligned} \tag{11}$$

where $p_{\tau,t}, \tau \in S = \{0.10, 0.25, 0.5, 0.75, 0.90\}$ are estimated by the following optimization procedure

$$\mathbf{p}_t = \arg\min_{\mathbf{p}_t} E[r_{t+1} - (p_{0.10,t}\hat{r}_{i,t+1}(0.10) + p_{0.25,t}\hat{r}_{i,t+1}(0.25) +$$

$$+p_{0.5,t}\hat{r}_{i,t+1}(0.5) + p_{0.75,t}\hat{r}_{i,t+1}(0.75) + p_{0.90,t}\hat{r}_{i,t+1}(0.90))]^2$$

$$s.t.\ p_{0.10,t} + p_{0.25,t} + p_{0.50,t} + p_{0.75,t} + p_{0.90,t} = 1$$

$$0.00 \leq p_{0.10,t} \leq 0.10$$

$$0.15 \leq p_{0.25,t} \leq 0.35$$

$$0.40 \leq p_{0.50,t} \leq 0.60$$

$$0.15 \leq p_{0.75,t} \leq 0.35$$

$$0.00 \leq p_{0.90,t} \leq 0.10$$

4 Forecast Combination

Since Bates and Granger's (1969) seminal contribution, it has been known that combining individual models' forecasts can reduce uncertainty risk associated with a single predictive model and display superior predictive ability (see Rapach, Strauss and Zhou (2010), Hendry and Clements (2004)). Timmermann (2006) suggests that forecast combination can be thought of as a diversification strategy that improves forecasting performance in the same way as asset diversification improves portfolio performance. In the context of equity premium predictability, Rapach, Strauss and Zhou (2010) show that combination foreacsts of individual predictive models can consistently beat the benchmark. Following Rapach, Strauss and Zhou (2010), we also consider various combining methods, ranging from simple averaging schemes to more advanced ones, based on both the single predictor model specifications (conditional mean and quantile forecasts) of Section 2 and the robust point forecasts of Section 3. Below we discuss the design of our forecast experiment, which is identical to the one employed by Goyal and Welch (2008) and Rapach, Strauss and

Zhou (2010), and outline the combining methods employed.

We generate out-of-sample forecasts of the equity premium using a recursive (expanding) window. More specifically, 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 $\{\hat{r}_{i,t+1}\}_{t=K}^{T-1}$. This experiment simulates the situation of a forecaster in real time, since she employs data as soon as they become available.

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* combining weights at time t. Some of the combining methods described below require a holdout out-of-sample period during which the combining weights are estimated. The first P_0 out-of-sample observations are employed as the initial holdout period. In this respect, we compute combination forecasts over the post-holdout out-of-sample period, leaving us with a total of $T - (K + P_0) = P - P_0$ forecasts available for evaluation.

The simplest combining 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 combining scheme. The next schemes we employ are the trimmed mean and median ones. The trimmed mean combination forecast sets $w_{i,t}^{(C)} = 1/(N - 2)$ and $w_{i,t}^{(C)} = 0$ for the smallest and largest forecasts, while the median combination scheme is the median of $\{\hat{r}_{i,t+1}\}_{i=1}^{N}$ forecasts.

The second class of combining methods we consider, proposed by Stock and Watson (2004), suggests forming weights based on the historical performance of the individual models over the holdout out-of-sample period. Specifically, their discount MSFE (DMSFE) combining method suggests forming weights as follows

$$w_{i,t}^{(C)} = m_{i,t}^{-1} / \sum_{j=1}^{N} m_{j,t}^{-1}$$

where

$$m_{i,t} = \sum_{s=K}^{t-1} \psi^{t-1-s}(r_{s+1} - \widehat{r}_{i,s+1})^2,$$

where ψ is a discount factor which attaches more weight on the recent forecasting accuracy of the individual models in the cases where $\psi < 1$. 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) in the case of uncorrelated forecasts.

The third class of combining methods, namely the cluster combining method, was introduced by Aiolfi and Timmermann (2006). In order to create the cluster combining forecasts, we form L clusters of forecasts of equal size based on the MSFE performance. Each combination forecast is the average of the 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. In our analysis, we consider $L = 2, 3$.

Next, the principal component combining methods of Chan, Stock and Watson (1999) and Stock and Watson (2004) are 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 $s = K, ..., t-1$. 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 combining weights of the principal components. In order 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.

Finally, we employ the Bayesian model averaging (BMA) approach in order to produce combined forecasts. According to this approach, the weights $w_{i,t}^{(C)}$ are the posterior model probabilities obtained from a Bayesian model comparison exercise which compares the single predictor model specifications. The BMA method requires a parametric, likelihood-based, approach to inference. The posterior probabilities of the competing models are calculated as follows

$$w_{i,t}^{(C)} = P(m_i \mid r_1, \ldots, r_t) = \frac{P(m_i) f(r_1, \ldots, r_t \mid m_i)}{\sum_{j=1}^{N} P(m_j) f(r_1, \ldots, r_t \mid m_j)},$$

where m_i denotes the ith single predictor model, $P(m_i)$, $i = 1, \ldots, N$, are prior model probabilities and $f(r_1, \ldots, r_t \mid m_i)$ is the marginal likelihood of the data given by a specific model m_i, obtained by integrating the model parameters out of the likelihood function. Calculating the marginal likelihood of standard conditional mean regression models is straightforward (see, for example, O'Hagan and Forster, 2004), while for quantile regression models we adopt the Laplace approximation method of Meligkotsidou, Vrontos and Vrontos (2009) for estimating the marginal likelihood of the competing model specifications. The BMA approach is designed to assign large weights to the individual forecasts obtained by those predictors that provide better in-sample model fit.

5 Data and forecast evaluation

The data we employ are from Goyal and Welch (2008) who provide a detailed description of transformations and datasources.[5] The equity premium is calculated as the difference of the continuously compounded S&P500 returns, including dividends, and the Treasury Bill rate. As already mentioned, following the line of work of Goyal and Welch (2008), Rapach, Strauss and Zhou (2010) 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. Our forecasting experiment is conducted on a quarterly basis and 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, Strauss and Zhou (2010) covering 1965:1-2010:4.[6] 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 following variables.

- Dividend–price ratio (log), 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.
- Dividend yield (log), D/Y : Difference between the log of dividends and the log of lagged stock prices.
- Earnings–price ratio (log), 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.

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

[6] Please note that the out-of-sample period refer to the period used to evaluate the out-of-sample foreacts. We use the ten years (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.

- Dividend–payout ratio (log), D/E: Difference between the log of dividends and the log of earnings.

- Stock variance, SVAR: Sum of squared daily returns on the S&P 500 index.

- Book-to-market ratio, B/M: Ratio of book value to market value for the Dow Jones Industrial Average.

- 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.

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 as follows.

- Treasury bill rate, TBL: Interest rate on a three-month Treasury bill (secondary market).

- Long-term yield, LTY: Long-term government bond yield.

- Long-term return, LTR: Return on long-term government bonds.

- Term spread, TMS: Difference between the long-term yield and the Treasury bill rate.

- Default yield spread, DFY: Difference between BAA- and AAA-rated corporate bond yields.

- Default return spread, DFR: Difference between long-term corporate bond and long-term government bond returns.

To capture the overall macroeconomic environment, we employ the inflation rate and the investment-to-capital ratio defined as follows.

- Inflation, INFL: Calculated from the CPI (all urban consumers).
- Investment-to-capital ratio, I/K: Ratio of aggregate (private nonresidential fixed) investment to aggregate capital for the entire economy (Cochrane, 1991).

In our application, the natural benchmark forecasting model is the constant expected equity premium or prevailing mean (PM) model, which coincides with the estimate $\widehat{\alpha}_i$ in the linear regression model (1) when none of the predictive variables are included in the model. As a measure of forecast accuracy we employ the out-of-sample R-square (R^2_{OS}), which is defined as

$$R^2_{OS} = 1 - \frac{MSFE_i}{MSFE_{PM}} \tag{12}$$

where $MSFE_i$ is the Mean Square Forecast Error (MSFE) defined as the average squared forecast error over the out-of-sample period of any of our competing models and specifications and $MSFE_{PM}$ is the respective value for the prevailing mean (PM) or historical mean model. Positive values of the out-of-sample R-square are associated with superior forecasting ability of our proposed model/specification and vice-versa.

Given that point estimates of the R^2_{OS} are sample dependent, we need to evaluate the statistical significance of our forecasts. A disclaimer is in order here. Different evaluation techniques should be employed depending on whether the competing models are nested or not. Specifically, the PM model or constant equity premium model, which serves as the benchmark, is nested in the linear prediction models (Equation 1). The same holds for the alternative combining schemes developed under this setting. On the other hand, when it comes to the quantile predictive models and the estimators stemming from them, PM no longer belongs to this set of models. In such a setting, only the Prevailing Quantile (PQ) and the models specified by equation (2) could lead to a nested environment for forecast evaluation. To this end, we employ the Clark and West (2007) approximate normal test to compare nested models and the encompassing test of Harvey, Leybourne and Newbold (1998) for non-nested models. Next, we provide a brief description of these approaches.

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

$$f_{t+1} = (r_{t+1} - \hat{r}_{A,t+1})^2 - [(r_{t+1} - \hat{r}_{B,t+1})^2 - (\hat{r}_{A,t+1} - \hat{r}_{B,t+1})^2]$$

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

As already mentioned, we employ the encompassing test of Harvey, Leybourne and Newbold (1998) to compare non-nested models. The notion of forecast encompassing was developed in Granger and Newbold (1973) and Chong and Hendry (1986).[7] Consider forming a composite forecast $\hat{r}_{c,t+1}$ as a convex combination of the out-of-sample forecasts from a parsimonious model A, $\hat{r}_{A,t+1}$ and a larger model B, $\hat{r}_{B,t+1}$ in an optimal way so that $\hat{r}_{c,t+1} = \lambda \hat{r}_{A,t+1} + (1-\lambda)\hat{r}_{B,t+1}$, $0 \leq \lambda \leq 1$. If the optimal weight attached to model A forecast is zero, $\lambda = 0$, then model B forecasts encompass the competing forecasts from model A. Harvey, Leybourne and Newbold (1998) developed the encompassing test, denoted as $ENC-T$, based on the approach of Diebold and Mariano (1995) to test the null hypothesis that $\lambda = 0$, against the one-sided (upper tail) alternative hypothesis that $\lambda > 0$. 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

$$d_{t+1} = (u_{A,t+1} - u_{B,t+1})u_{A,t+1}.$$

The $ENC-T$ statistic is given by

$$ENC-T = \sqrt{(P-P_0)}\frac{\overline{d}}{\sqrt{\widehat{Var}(d)}} \quad (13)$$

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.[8] The $ENC-T$ statistic is asymptotically distributed as a standard normal variate under the null hypothesis. To improve finite sample performance Harvey, Leybourne and Newbold (1998) recommend employing the student's t distribution with $P-P_0-1$ degrees of freedom. As previously, this test is a one-sided test.

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

[8] For forecast horizons greater than one, an estimate of the long-run variance should be employed.

6 Empirical Results

6.1 A motivating illustration

Before presenting our empirical results, we provide an illustration on the sources of potential benefits of our proposed methodology. Our motivation stems from the empirical finding that a single mean curve can rarely provide an adequate summary of the evolution of the equity premium over time. For example, the most popular variables in the returns prediction literature, namely the dividend-price ratio and the term spread may capture different aspects of economic conditions. Furthermore, not only fluctuations of the business cycle induce a time-varying nature on mean predictive relationships, but also across quantiles, since there is no compeling theoretical reason slope coefficients are constant across quantiles. To the extent that candidate predictor variables contain significant information for some parts of the return distribution, but not for the whole, a methodology that properly integrates this information will lead to additional benefits. Our methodology bears a resemblance to combination forecasts that are employed to produce less volatile forecasts, stabilize individual forecasts, reduce forecast risk and improve forecast performance (Rapach, Strauss and Zhou (2010)).

As a first step, we generate forecasts employing Equation (2) with a single predictor at a time and calculate the MSFE associated with each specification. Then we calculate the MSFE associated with the prevailing quantile and estimated by the model that contains only a constant. This prevailing quantile serves as a benchmark in the same fashion as the historical average (prevailing mean) serves as a benchmark in typical mean predictive regressions. Table 1, Panel A illustrates our findings with highlighted (in grey) cells suggesting superior predictive ability. Overall, we observe 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 capturing the distribution of the returns. As is apparent, the right tail seems to be more easily predicted in contrast to the left tail. Such a finding is well documented in the literature as return prediction models perform better during expansions. More in detail, D/P, D/Y, E/P, D/E, TBL, LTY, DFR, INFL and I/K mainly help in predicting quantiles greater than the 40th quantile, while B/M, LTR, NTIS and TMS are more helpful when predicting quantiles lower than the 35th one. Notably, DFY contains no predictive ability for any part of the return distribution. The question we seek to answer is whether we can either 'horizontically' or 'vertically' combine the information content in the respective quantiles to provide robust point forecasts that improve on the mean predictive models. As already mentioned in Section 3, the 'horizontal' combination is done via the robust point forecasts. With respect to the 'vertical' combination of information, we employ a variety of combination methods. Their potential predictive ability is outlined in Table 1, Panel B. Our results suggest that combination methods contain significant information for a wider portion of the return distribution, albeit skewed towards the right tail. The two cluster combining methods are the only ones that cover the full range, while the principal components are successful in all parts of the distribution with the exception of the 35th to 45th quantile. The method that seems able to forecast the smallest fraction of the distribution is the BMA method that helps predicting the 45th to 85th quantile of the distribution.

[TABLE 1 AROUND HERE]

6.2 Out-of-sample performance of predictive regressions

In this subsection, we conduct an out-of-sample forecasting exercise with the aim to present and discuss the results of the proposed modeling approaches and the appropriateness of the forecasts obtained from the robust techniques described in Section 3. Table 2 reports the out-of-sample performance of the conditional expectation (mean) predictive regression model and combining methods. In particular, Table 2 presents the R^2_{OS} statistics 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; the statistical significance of the corresponding forecasts is assessed by using the Clark and West (2007) MSFE-adjusted statistic. Positive values of R^2_{OS} indicate 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 have a positive R^2_{OS} value, while three predictors, namely D/P (0.72%), D/Y (1.00%) and I/K (2.31%), have statistically significant positive values of R^2_{OS}. Among them the I/K predictor provides superior forecasting ability. Looking at the R^2_{OS} generated by the different combining techniques, we observe that almost all of them produce positive values of R^2_{OS}, with the exception of the principal components and the BMA method. Most of the combining methods have statistically significant positive R^2_{OS} values, while five of them, i.e. the mean, trimmed mean, DMSFE(1), DMSFE(0.9) and cluster 2, provide higher values of R^2_{OS} than that of the best I/K predictor. The results of Table 2 in general confirm the findings of Rapach, Strauss and Zhou (2010) who found that the D/P, D/Y and I/K predictors have significant forecasting ability, and that the combination methods outperform the individual predictive regression models.

[TABLE 2 AROUND HERE]

Next we present and discuss the out-of-sample performance of the robust point forecasts obtained by using fixed weights (FW) and time-varying weights (TVW) based on single predictor quantile regression models, as well as on their corresponding combining methods. Table 3 reports the R^2_{OS} statistics of the individual quantile regression models and of the combining methods, relative to the historical average benchmark model. Based on Panel A (left hand side) of Table 3, which reports the performance of the robust point forecasts formed by a fixed weighting scheme based on individual quantile predictive models, we observe that only three predictors (namely, D/P, D/Y and I/K) have positive and statistical significant values of R^2_{OS}, while among them the I/K predictor provides superior forecasting ability. These results are similar in spirit with those of the individual mean predictive regression model, and indicate a forecasting ability of these three predictors over the historical average using different weighting schemes for the corresponding quantiles. Note, however, that the values of R^2_{OS} of the robust point forecasts are larger than those of the individual mean regression of D/P and D/Y for all weighting schemes FW1-FW4, and of I/K for FW1, FW2 and FW4, indicating some improvement over the mean regression approach. Turning to the results of the combined forecasts of the individual predictive models with fixed weights in Panel B (left hand side) of Table 3, we observe that almost all the combining methods, except cluster 3 and principal components, provide positive and statistically significant R^2_{OS} values, indicating forecasting ability over the historical average. A comparison of the different combination techniques suggests that the DMSFE methods rank first followed by the mean combination method, since they generally provide higher positive R^2_{OS} values. Among the four fixed weighting schemes, the FW4 scheme produces in most of the cases higher R^2_{OS} values indicating an improved predictive performance, probably due to the fact that it utilizes distribution information obtained by a finer grid of conditional quantiles of returns.

[TABLE 3 AROUND HERE]

The results presented in Table 3 (right hand side) concern the out-of-sample performance of robust point forecasts with time-varying weights (TVW1-TVW3) based on quantile predictive regression models. The positive R^2_{OS} values of Panel A (right hand side) of Table 3 indicate that four predictors, namely the D/P, D/Y, DFR, and I/K, have significant forecasting ability with respect to the historical average. This approach, which uses time-varying weights, reveals that an additional predictor, the DFR, may have significant out-of-sample predictability, compared to the fixed weighting approach and the conditional mean predictive regression model which identified the DFR to have positive but not significant R^2_{OS}. The improved out-of-sample performance of the robust point forecasts using time-varying weights over the mean predictive regression model is also apparent since most of the negative R^2_{OS} values for the individual predictors (Panel A, right hand side) are smaller in terms of absolute value from the corresponding R^2_{OS} of the mean predictive models of Table 2.

The most striking result can be drawn from panel B (right hand side) of Table 3; the R^2_{OS} generated by the combining methods of the individual quantile predictive models with time-varying weights (TVW1-TVW3) are all positive and statistically significant, ranging from 2.40% for the median combination method using TVW2 to 3.67% for the mean combination method using TVW3.[9] Moreover, all the R^2_{OS} values for the combination forecasts are greater than the corresponding R^2_{OS} values of the combination methods based on mean predictive models (see Table 2) as well as based on fixed weights. These results indicate superior predictive ability of forecasts obtained by first using robust techniques that take into account the distribution information regarding how different conditional quantiles of equity premium are affected by individual predictors, and then incorporating predictor information in order to produce combined forecasts from different individual models.

[TABLE 4 AROUND HERE]

Next, we present our empirical findings with respect to the investigation of the predictive ability of forecasts which are obtained by first utilizing the predictor information to produce combined quantile forecasts from different individual models and then synthesizing this distributional information through robust forecasting weighting schemes. This procedure aims at providing optimal forecasts of each part of the conditional distribution

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

of the equity premium, by appropriately, combining individual quantile forecasts, before using these quantiles to construct robust point forecasts. In Table 4, we present the out-of-sample performance of these robust point forecasts using a fixed (FW1-FW4) and a time-varying weighting scheme (TVW1-TVW3) based on the combined information. The results of Table 4 (left hand side) suggest that the forecasts of the fixed weighting schemes based on different combination methods provide positive and statistically significant R^2_{OS} in all cases except for the principal component combination method, indicating a superior performance relative to the historical average benchmark. It is interesting to observe that more promising results in favor of our robust quantile regression approach arise from the use of time-varying weighting schemes (TVW1-TVW3). Looking at the right hand side of Table 4 one may observe that all of the R^2_{OS} values are positive and statistically significant and in many cases we obtain the largest R^2_{OS} values among the different modeling approaches that have been used in our analysis. The results of Table 4 suggest that the best out-of-sample performance is obtained by applying the robust point forecasts which use time-varying weights based on the mean combination method.

To conclude, the results of our analysis indicate that superior predictive performance is achieved under the quantile regression approach as follows. First, various quantiles of the conditional distribution of returns are optimally predicted by combining information from different predictors using some forecast combination method. Next, robust point forecasts of the equity premium are produced using time-varying weighting schemes.

In what follows, we evaluate the economic significance of our proposed specifications against the benchmark historical average.

7 Economic evaluation

As Campbell and Thompson (2008) and Rapach, Strauss and Zhou (2010) suggest, even small values of R^2_{OS} can give an economically meaningful degree of return predictability that could result in increased portfolio returns for a mean-variance investor that maximizes expected utility. Within this stylized asset allocation framework, this utility-based approach, initiated by West, Edison and Cho (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, Kirby and Ostdiek (2001), Marquering and Verbeek (2004), Della Corte, Sarno, and Tsiakas (2009), Della Corte, Sarno and Valente (2010), Wachter and Warusawitharana (2009)).

7.1 The framework for measuring economic value

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 the only source of risk stems from the uncertainty over the future path of the stock market. Since only one risky asset is involved, this approach could 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

$$w_t = \frac{E_t(r_{t+1})}{\gamma Var_t(r_{t+1})} \tag{14}$$

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), Rapach, Strauss and Zhou (2010)). The conditional variance of the portfolio is approximated by the historical variance of the stock market return and is estimated using a ten-year rolling window of quarterly returns.[10] In this way, the optimal weights vary with the degree the conditional mean varies, i.e. the forecast each model/ specification

[10] See Campbell and Thomson (2008) and Rapach, Strauss and Zhou (2010).

gives.[11] Under 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 $t+1$.[12]

Assuming quadratic utility, over the forecast evaluation period the investor with initial wealth of W_o realizes an average utility of

$$\overline{U} = \frac{W_o}{(P - P_0)} \sum_{t=0}^{P-P_0-1} (R_{p,t+1} - \frac{\gamma}{2(1+\gamma)} R_{p,t+1}^2) \tag{15}$$

where $R_{p,t+1}$ is the gross return on her portfolio at time $t+1$.[13] At any point in time, the investor prefers the model for conditional returns that yields the highest average realized utility. The first measure of economic significance we employ is the utility gain calculated by the difference between the average realized utilities of competing models.[14] The utility gain is given by the following formula

$$\Delta\overline{U} = (\frac{\overline{U^i} - \overline{U^{PM}}}{\overline{U^{PM}}}) \cdot 400 \tag{16}$$

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) or historical mean model. We multiply this ratio by 400 to express it to average annualized percentage return (basis points, bps).

[11] Alternatively, one could make use of information about the entire distribution provided by the quantile regression predictive models.

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

[13] One could instead employ other utility functions that belong to the constant relative risk aversion (CRAA) family such as power or log utility. However, quadratic utility allows for nonormality in the return distribution while remaining in the mean-variance framework.

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

Furthermore, given that a better model requires less wealth to attain a given level of $\overline{U}$ than an alternative model, a risk-averse investor will be willing to pay to have access to this superior model which would be subject to management fees as opposed to the simple PM model. In the event that the superior model is one of our proposed i specifications, the investor would pay a performance fee to switch from the portfolio constructed based on the historical average to the i specification. This performance fee, denoted by Φ, is the fraction of the wealth which when subtracted from the i proposed portfolio returns equates the average utilities of the competing models. In our set-up the performance fee is calculated as the difference between the realized utilities as follows:

$$\frac{1}{(P-P_0)}\sum_{t=0}^{P-P_0-1}\left\{(R^i_{p,t+1}-\Phi)-\frac{\gamma}{2(1+\gamma)}(R^i_{p,t+1}-\Phi)^2\right\}= \tag{17}$$
$$\frac{1}{(P-P_0)}\sum_{t=0}^{P-P_0-1}\left\{R^{PM}_{p,t+1}-\frac{\gamma}{2(1+\gamma)}(R^{PM}_{p,t+1})^2\right\}.$$

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.

As a complement to the performance fee measure, we also employ the manipulation-proof performance measure proposed by Goetzmann, Ingersoll, Spiegel and Welch (2007). This measure can be interpreted as a portfolio's premium return after adjusting for risk and it remedies potential caveats associated with the popular Sharpe ratio such as the effect of non-normality (Jondeau and Rockinger (2006)), the underestimation of the performance of dynamic strategies (Marquering and Verbeek (2004), Han (2006)) and the choice of utility function (Della Corte, Sarno and Sestieri (2012)). This measure is defined as

$$M(R_p)=\frac{1}{1-\gamma}\ln\left\{\frac{1}{(P-P_0)}\sum_{t=0}^{P-P_0-1}\left(\frac{R_{p,t+1}}{r_{f,t}}\right)^{1-\gamma}\right\}.$$

The difference, Θ, between the $M(R_p)s$ of competing models calculated as follows is employed to assess the most valuable model

$$\Theta = M(R_p)^i - M(R_p)^{PM}. \tag{18}$$

Both Φ and Θ are reported in annualized basis points.

7.2 Empirical evidence on the economic value of predictive regressions

We assume that the investor dynamically rebalances her portfolio (updates the weights) quarterly over the out-of-sample period employing the respective recursive forecasts for all the models/specifications under consideration. Her precision of estimates/ forecasts normally increases as more information (data) become available. Similarly to Section 6.1, the out-of-sample period of evaluation is 1965:1-2010:4 and the benchmark strategy against which we evaluate our forecasts is the constant expected risk premium. For every model/specification we calculate the average annualized utility gain (Equation 16), the performance fee associated with each strategy calculated from Equation (17) and the manipulation-proof performance measure (Equation 18). Following Campbell and Thompson (2008) and Rapach, Strauss and Zhou (2010) we set RRA (γ) equal to 3.[15]

We begin our analysis with the economic evaluation of the mean predictive regression models. Table 5 reports the respective figures for both the single-variable models and the models employing combination forecasts. Contrary to the statistical evaluation results, utility losses are associated with only four single-variable models, namely the models formed on the basis of SVAR, B/M, NTIS and DFY. The remaining specifications yield utility gains ($\Delta\overline{U}$) that range from 0.168% (D/E) to 1.488% (TMS). Furthermore, our results suggest that an investor would be willing to pay sizable annual performance fees

[15] The utility function we employ differs from the one in Rapach, Strauss and Zhou (2010) and Campbell and Thompson (2008), so our figures are not directly comparable.

(Φ) that exceed 200 basis points (bp) to switch from the constant expected risk premium strategy to a strategy that times the market conditioning on one of the following variables: D/P, D/Y, TBL, LTY, TMS, INFL and I/K. Overall, the manipulation-free performance fee measure (Θ) paints a qualitatively similar picture suggesting that the best performing model is the one based on D/Y closely followed by I/K.

[TABLE 5 AROUND HERE]

Turning to the economic value of combination forecasts, our results suggest that, irrespective of the method employed, an investor enjoys utility gains ranging from 0.616% (BMA) to 1.508% (Principal Components). Quite interestingly, while both BMA and principal components are associated with a negative R^2_{OS} suggesting poor forecasting performance, their employment can generate profits to an investor. The superiority of the principal components combining method may seem quite puzzling. However, unreported results show that its performance comes from its ability to forecast the direction of change of the market in more than 80% of the cases.[16] This ability to time the market is depicted in a remarkable performance fee of 394 bps, followed by the DMSFE and mean combining methods. The ranking of our combining methods slightly changes when we employ the manipulation-free performance fee measure (Θ). Specifically, the DMSFE methods rank first followed by the mean and the principal components methods, while the median and the BMA methods achieve the worst performance.

So far, our results confirm the findings of Rapach, Strauss and Zhou (2010) on the economic benefits of combining. Next, we turn our attention to the economic performance of robust point forecasts formed by a fixed weighting scheme given by equations (5) to (8). Our results, reported in Table 6 (Panels A and B for the single predictor models and their combinations, respectively), may be summarized as follows:

[16] These results are not reported for brevity but are available from the authors upon request.

- The economic performance of robust point forecasts is nearly as good as the performance of the mean predictive models. Overall, similarly to the mean forecasts, utility losses are associated with SVAR, B/M, NTIS and DFY irrespective of the weighting scheme, while TMS, IK, TBL, LTY and D/Y emerge as the most powerful predictors with utility gains exceeding 0.8% in most cases.

- A comparison of the four weighting schemes in a univariate environment suggests that FW4 that aggregates information of quantiles over a finer grid provides the investor with more utility gains. However, the higher performance fee of 383 bps is achieved when the investor conditions on TMS and employs FW2 to aggregate information from quantile predictions.

- Turning to Panel B and the combination methods, our results suggest that an investor that approximates the distribution of returns by combining quantile information and then creates the respective point forecast will always generate positive abnormal returns. Utility gains, with the exception of median combinations, range from 1.104% for the Cluster 2 method and FW4 to 1.664% for the Principal components method and FW2. In order to switch from the constant expected risk premium strategy to the first case, the investor would be willing to pay an annual performance fee of 289 bps while for the latter one a fee of 436 bps.

- While our methods for producing robust point forecasts yield quite similar results, FW2 and FW4 seem to generate superior forecasts relative to FW1 and FW3.

- Overall, the risk-adjusted abnormal return Θ is fully-consistent (size and sign) with the results obtained from the performance fees for both the single predictor variable model (Panel A) and the combining methods (Panel B).

[TABLE 6 AROUND HERE]

Our statistical evaluation tests (Section 6.2) have shown that allowing for a time-varying weighting scheme in the generation process of robust point forecasts can lead to out-of-sample benefits. Next, we repeat this analysis in an economic evaluation framework in order to show whether these benefits are economically significant. Table 7 reports the respective results for both the univariate models (Panel A) and the combining methods (Panel B). As already mentioned, in the current setting our combination methods are limited to the three simpler ones since they do not require an extra holdout out-of-sample period.

[TABLE 7 AROUND HERE]

The robust point forecasts based on a time-varying weighting scheme offer positive utility gains on the basis of 11 out of 15 variables for TVW1 and 12 out of 15 variables for TVW2 and TVW3 specifications. Quite interestingly, while the predictive ability of DP is well established in a statistical context and in an economic context when fixed weights are employed, this time-varying environment leads to utility losses in all three specifications. On the other hand, variables such as TMS and TBL which are countercyclical by nature are associated with sizable utility gains of around 2%. Our estimates suggest that an investor would pay up to 553 bps to switch from the constant expected risk premium to robust point forecasts conditioned on TMS in a time varying nature. Sizable benefits are also depicted when LTY, LTR, INFL and I/K are employed with utility gains exceeding 1%. Attempting to evaluate the alternative time-varying procedures, our results suggest that they are broadly equivalent. Furthermore, our results in panel B verify the effectiveness of combining methods at this time-varying framework as well. Specifically, employing the mean combining method works quite well and beats the historical average by significant margins. Similar performance is obtained by the trimmed mean combining method while the median one ranks third. However, even with this combining method, an investor can enjoy benefits of up to 325 bps.

Finally, Table 8 addresses the issue of forming robust point forecasts, either in a fixed or time-varying manner, based on combined information. To this end, as already mentioned, each quantile forecast is constructed based on the information contained in the single predictor variables and then these composite quantile forecasts form the robust point forecasts.[17] The overall picture that emerges confirms the robustness of our proposed methodology. The performance fee that an investor would be willing to pay to utilize our proposed models ranges from 141 bps for FW1 and the median combining method to 478 bps for TVW1 and the mean combining method. When considering the fixed weighting schemes, the best performance is achieved by FW2. Within this setting the BMA combining method yields the superior performance closely followed by the Principal Components method and the DMSFE (0.9). With respect to time-varying schemes, TVW1 achieves the best performance when mean and trimmed mean combination forecasts are employed. Finally, forecasts associated with the time-varying scheme are in general superior to the fixed weighting scheme.

[TABLE 8 AROUND HERE]

8 Conclusions

In this study we investigate whether there is evidence of out-of-sample predictive ability of various economic variables for the equity premium. We develop an alternative modeling approach which is based on quantile predictive regression models and produces robust and accurate point forecasts of the equity premium, based on the quantile forecasts, by using fixed and time-varying weighting schemes. To take into account the findings of recent academic studies which suggest that forecast combinations improve the out-of-sample equity premium prediction, we propose to utilize different combination methods

[17] This issue does not apply to forecasts formed on the basis of mean combination forecasts and fixed weighting schemes.

based on the robust quantile forecasts. Thus, in our analysis, the crucial issue under consideration is to examine whether the framework that adopts two different sources of information, i.e. distribution information and predictor information, is able to deliver more accurate out-of-sample forecasts of the equity premium.

The results of our analysis suggest that there is evidence of equity premium predictability. Specifically, we find that the alternative predictive approach based on quantile regressions using a time-varying weighting scheme outperforms the historical average benchmark and the combined mean predictive regression modeling approach. Moreover, our study contributes to the growing empirical literature on equity premium predictability by proposing to use a two stage forecasting approach. First, we recommend combining individual forecasts from different single predictor quantile regressions, thus incorporating information from various economic variables in order to produce accurate quantile predictions. Then, at a second stage, we propose to construct robust point forecasts of the equity premium by adopting a time-varying weighting scheme which combines a set of quantile forecasts, thus incorporating information from the conditional distribution of returns. We also show that the predictive ability of the proposed approach has substantial statistical and economic value over the standard predictive modeling approaches.

The usefulness of the proposed modeling approach stems from the highly complex and dynamic nature of equity returns. Our model is able to capture different characteristics of the return series (such as deviation from normality, fat tails and skewness) and to identify potential differences in predictors across quantiles of returns. For example, our analysis suggests that there exist predictors that have superior predictive ability for lower or/and upper conditional quantiles of returns. Thus, the quantile regression approach is able to uncover interesting distributional information, and also from an economic perspective to incorporate meaningful business conditions information.

Our study extends previous work on equity premium predictability by proposing to use a quantile predictive regression modeling approach and a two stage weighting scheme of constructing robust point forecasts. We believe that there exist several potential applications (for example, to other asset classes such as mutual fund and hedge fund investments) and extensions of our modeling approach in the area of empirical finance. Interesting extensions of the proposed model can be the development of break-point quantile regression models that take into account the empirical evidence of structural instability in macroeconomic relations using individual predictive models. Another possible extension could be to restrict the sign and/or the magnitude of the quantile regression coefficients motivated by economic theory. Clearly, many interesting questions remain open and various topics for future research arise in this context by using more complex and flexible modeling approaches.

References

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

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

Bai, J., and S. Ng . "Determining the Number of Factors in Approximate Factor Models." Econometrica, 70 (2002), 191–221.

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

Barnes, M. L., and A. W. Hughes. "A Quantile Regression Analysis of the Cross Section of Stock Market Returns." Working Paper 02-2, Federal Reserve Bank of Boston, (2002).

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

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

Baur, D. G.; T. Dimpfl; and R. Jung. "Stock return autocorrelations revisited: A quantile regression approach." Journal of Empirical Finance, (2012), forthcoming.

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

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

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

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

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

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

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

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

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

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

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

Cenesizoglu, T., and A. Timmermann. "Do Return Prediction Models Add Economic Value?." Journal of Banking and Finance, forthcoming (2012).

Chan, Y. L.; J. H. Stock; and M. W. Watson. "A dynamic factor model framework for forecast combination." Spanish Economic Review, 1 (1999), 21–121.

Chernozhukov, V., and L. Umantsev. "Conditional Value-at-Risk: Aspects of Modeling and Estimation." Empirical Economics, 26 (2001), 271-292.

Chong, Y.Y., and D.F. Hendry. "Econometric evaluation of linear macroeconomic models." Review of Economic Studies, 53 (1986), 671-690.

Chuang, C.-C.; C.-M. Kuan; and H.-Y. Lin. "Causality in quantiles and dynamic stock return-volume relations." Journal of Banking and Finance, 33, 7 (2009), 1351-1360.

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

Clements, M. P., and D. F. Hendry. "Forecasting Economic Time Series." Cambridge University Press, Cambridge (1998).

Cochrane, J. H. "Production-Based Asset Pricing and the Link between Stock Returns and Economic Fluctuations." Journal of Finance, 46 (1991), 209–37.

Dangl, T., and M. Halling. "Predictive regressions with time-varying coefficients." Journal of Financial Economics, (2012), forthcoming.

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

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

Della Corte, P.; L. Sarno; and G. Sestieri. "The Predictive Information Content of External Imbalances for Exchange Rate Returns: How Much is it Worth?." Review of Economics and Statistics, 94, 1 (2012), 100-115.

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

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

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

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

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

Feng, Y.; R. Chen; and G. W. Basset. "Quantile momentum." Statistics and its Interface, 1 (2008), 243–254.

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

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

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

Goetzmann, W.; J. Ingersoll; M. Spiegel; and I. Welch. "Portfolio Performance Manipulation and Manipulation-proof Performance Measures." Review of Financial Studies, 20 (2007), 1503-1546.

Goyal, A. and I. Welch. "A Comprehensive Look at the Empirical Performance of

Equity Premium Prediction." Review of Financial Studies, 21 (2008), 1455–508.

Granger, C. W. J., and P. Newbold . "Some Comments on the Evaluation of Economic Forecasts." Applied Economics, 5 (1973), 35-47.

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

Han, Y. "Asset Allocation with a High Dimensional Latent Factor Stochastic Volatility Model." Review of Financial Studies, 19 (2006), 237-271.

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

Harvey, D. I.; S. J. Leybourne; and P. Newbold. "Tests for Forecast Encompassing." Journal of Business and Economic Statistics, 16 (1998), 254–59.

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

Hsiao, C., and S. K. Wan. "Is there an optimal forecast combination?." Journal of Econometrics, (2012), forthcoming.

Jondeau, E., and M. Rockinger. "The Economic Value of Distributional Timing". Swiss Finance Institute Research Paper No. 06-35 (2006).

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

Koenker, R. "Quantile regressions." Cambridge University Press, Cambridge (2005).

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

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

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

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

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

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

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

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

Neely, C. J.; D. E. Rapach; J. Tu; and G. Zhou. "Forecasting the equity risk premium: the role of technical indicators." Federal Reserve Bank of St. Louis Working Paper 2010-008C, (2011).

O'Hagan, A., and J. Forster. "Bayesian Inference." In Kendall's Advanced Theory of Statistics, II edn, Vol. 2B, Arnold, London (2004).

Pedersen, T. Q. "Predictable return distributions." Available online at SSRN http://ssrn.com /abstract=1658394, (2010).

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

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

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

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

Taylor, J. "A Quantile Regression approach to estimating the distribution of multi-period returns." Journal of Derivatives, 7 (1999), 64-78.

Taylor, J. W. "Forecasting daily supermarket sales using exponentially weighted quantile regression." European Journal of Operational Research, 178 (2007), 154-167.

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

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

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

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

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

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

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

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

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: Combining Methods

Mean						■	■	■	■	■	■	■	■	■	■	■	■	■	
Median					■	■	■	■	■	■	■	■	■	■	■	■	■	■	
Trimmed Mean						■	■	■	■	■	■	■	■	■	■	■	■	■	
DMSFE(1)					■	■	■	■	■	■	■	■	■	■	■	■	■	■	
DMSFE(0.9)				■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	
Cluster 2	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Cluster 3	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Principal Components	■	■	■	■	■	■				■	■	■	■	■	■	■	■	■	■
BMA									■	■	■	■	■	■	■	■	■		

Notes: Q5- Q95 denote the 5% to 95% quantiles of the return distribution. Grey cells denote superior predictive ability (lower MSFE) of the respective i specification (1st column) over the prevailing quantile (PQ) model.

Table 2. Out-of-sample performance of mean predictive regression models and combining methods

Individual predictive models		**Combination forecasts**	
D/P	0.72**	Mean	2.97***
D/Y	1.00**	Median	2.19***
E/P	-1.09	Trimmed Mean	2.64***
D/E	-1.60	DMSFE(1)	2.96***
SVAR	-6.65	DMSFE(0.9)	2.98***
B/M	-1.80	Cluster 2	2.34**
NTIS	-2.11	Cluster 3	1.22*
TBL	-2.43	Principal Components	-1.69
LTY	-2.59	BMA	-2.02
LTR	-1.15		
TMS	-2.65		
DFY	-2.71		
DFR	0.90		
INFL	-0.76		
I/K	2.31***		

Notes: The table reports the out-of-sample R^2 statistic of Campbell and Thompson (2008). Positive values indicate that the model given in Column (1) and (3) outperforms the historical average benchmark model. Statistical significance of the out-of-sample R^2 statistic is based on the p-value of the Clark and West (2007) out-of-sample *MSFE-adjusted* statistic. *, **, *** indicate significance at the 10%, 5% and 1% confidence levels, respectively.

Table 3. Out-of-sample performance of robust point forecasts and combining methods

Panel A: Individual Predictive Models

	FW1	FW2	FW3	FW4	TVW1	TVW2	TVW3
D/P	1.59*	1.61*	1.58**	1.01**	1.21*	1.48*	0.95*
D/Y	2.14**	2.47**	2.11**	1.63**	1.24*	2.14**	1.18*
E/P	-1.65	-1.90	-1.43	-1.56	-0.62	-0.75	-1.11
D/E	-0.15	-0.48	-0.62	-1.31	-0.35	-0.27	-0.46
SVAR	-9.96	-10.43	-9.27	-9.04	-3.99	-6.43	-1.46
B/M	-3.70	-4.11	-3.10	-2.31	-0.64	-1.45	-0.20
NTIS	-4.76	-5.43	-4.05	-3.17	-0.44	-1.68	0.73*
TBL	-1.31	-2.04	-1.31	-2.39	-1.00	-1.58	-1.98
LTY	-1.38	-2.08	-1.56	-2.43	-1.40	-2.22	-1.68
LTR	-5.23	-5.58	-4.26	-2.51	-2.30	-2.56	-1.40
TMS	-6.42	-6.38	-5.76	-4.33	-1.98	-3.14	-0.85
DFY	-7.18	-6.55	-5.81	-3.44	-2.61	-2.52	-1.48
DFR	-0.57	-0.26	-0.55	0.42	1.80*	2.39*	1.35*
INFL	-0.67	-1.03	-0.71	-0.83	0.40	-0.83	0.03
I/K	2.34***	2.37***	2.26***	2.58***	1.79**	1.73**	1.35**

Panel B: Combination forecasts

	FW1	FW2	FW3	FW4	TVW1	TVW2	TVW3
Mean	2.39**	2.32**	2.59***	2.80***	3.65***	3.46***	3.67***
Median	1.35*	1.07*	1.52**	2.06***	2.82***	2.40***	3.31***
Trimmed Mean	1.90**	1.81**	2.09**	2.37***	3.27***	2.98***	3.17**
DMSFE(1)	2.45**	2.37**	2.63***	2.81***	---	---	---
DMSFE(0.9)	2.53**	2.40**	2.69***	2.84***	---	---	---
Cluster 2	2.74**	2.22**	2.56**	2.31**	---	---	---
Cluster 3	-0.59	0.09	-0.17	1.39**	---	---	---
Principal Components	-2.89	-2.56	-2.84	-2.87	---	---	---

Notes: The table reports the out-of-sample R^2 statistic of Campbell and Thompson (2008). Positive values indicate that the model given in Columns (2) to (8) outperforms the historical average benchmark model. Statistical significance of the out-of-sample R^2 statistic is based on the p-value of the Harvey, Leybourne and Newbold (1998) out-of-sample *ENC-T* statistic. *, **, *** indicate significance at the 10%, 5% and 1% confidence levels, respectively. FW1-FW4 correspond to the fixed weighting schemes given by equations (5)-(8) of Section 3.1, while TVW1-TVW2 denote the time-varying weighting schemes given by equations (9)-(11) of Section 3.2.

Table 4. Out-of-sample performance of robust point forecasts based on combined information

	FW1	FW2	FW3	FW4	TVW1	TVW2	TVW3
Mean	2.39**	2.32**	2.59***	2.80***	4.06***	3.81***	3.23**
Median	1.14*	0.97*	1.34**	1.70***	3.31***	2.83***	2.55**
Trimmed Mean	1.82**	1.76**	2.02**	2.27***	3.61***	3.30***	2.81**
DMSFE(1)	2.46**	2.37**	2.66***	2.85***	---	---	---
DMSFE(0.9)	2.57**	2.44**	2.76***	2.86***	---	---	---
Cluster 2	2.78**	2.42**	2.78**	2.57**	---	---	---
Cluster 3	1.73*	1.75*	1.87*	2.24**	---	---	---
Principal Components	-2.24	-1.42	-1.60	-1.33	---	---	---
BMA	2.46**	1.58**	2.87**	2.12**	1.78**	2.06**	1.84**

Notes: See Table 3.

Table 5. Economic evaluation of mean predictive regression models and combining methods

Model	ΔU	Φ	Θ	Combining Method	ΔU	Φ	Θ
D/P	0.796	202.57	212.35	Mean	1.356	352.43	432.83
D/Y	1.316	336.10	417.10	Median	0.840	219.20	281.46
E/P	0.604	154.22	142.64	Trimmed Mean	1.272	330.33	411.38
D/E	0.168	43.47	54.77	DMSFE(1)	1.388	360.59	441.45
SVAR	-0.412	-106.56	-180.89	DMSFE(0.9)	1.460	380.18	461.25
B/M	-0.040	-10.65	-31.90	Cluster 2	1.220	317.20	363.15
NTIS	-0.344	-88.84	-161.97	Cluster 3	1.132	295.83	326.34
TBL	0.828	211.59	252.36	Principal Components	1.508	394.21	422.12
LTY	0.984	250.60	319.30	BMA	0.616	155.29	210.62
LTR	0.368	95.38	133.76				
TMS	1.488	389.33	388.03				
DFY	-0.432	-111.46	-136.51				
DFR	0.432	111.48	127.63				
INFL	0.860	221.64	276.54				
I/K	1.400	359.20	401.13				

Notes: The utility gain ΔU is calculated by $\Delta U = \frac{\overline{U}^{i} - \overline{U}^{PM}}{\overline{U}^{PM}} * 400$ where $\overline{U}^{i}, \overline{U}^{PM}$ denote the average quadratic utility of an investor with risk aversion coefficient of three over the forecast evaluation period from using the *i* 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 performance fee, Φ , is the fraction of the wealth which when subtracted from the *i* proposed portfolio returns equates the average utilities of the competing model. The performance fee, Φ , is calculated as the difference between the realized utilities of competing models as follows:

$$\frac{1}{P-P_0}\sum_{t=0}^{P-P_0-1}\{(R_{p,t+1}^{i}-\Phi)-\frac{\gamma}{2(1+\gamma)}(R_{p,t+1}^{i}-\Phi)^2)\}=\frac{1}{P-P_0}\sum_{t=0}^{P-P_0-1}\{R_{p,t+1}^{PM}-\frac{\gamma}{2(1+\gamma)}(R_{p,t+1}^{PM})^2)\}$$

where *P* is the number of out-of-sample forecasts, γ denotes the coefficient of relative risk aversion. Θ is the difference between the manipulation-proof performance measure of competing models $\Theta = M(R_p)^{i} - M(R_p)^{PM}$ where $M(R_p) = \frac{1}{1-\gamma}\ln\left\{\frac{1}{P-P_0}\sum_{t=0}^{P-P_0-1}\left(\frac{R_{p,t+1}^{PM}}{i_t}\right)^{1-\gamma}\right\}$. ΔU , Φ and Θ are reported in annualized basis points.

Table 6. Economic evaluation of robust point forecasts and combining methods (Fixed weights)

	FW1			FW2			FW3			FW4		
Panel A: Univariate Models	**ΔU**	**Φ**	**Θ**	**ΔU**	**Φ**	**Θ**	**ΔU**	**Φ**	**Θ**	**ΔU**	**Φ**	**Θ**
D/P	0.460	118.31	98.44	0.428	109.99	96.44	0.596	153.04	134.23	0.528	134.78	127.51
D/Y	0.728	186.77	196.79	0.824	210.85	240.39	0.812	209.69	208.00	1.004	256.55	311.73
E/P	0.472	122.92	96.68	0.472	121.87	96.09	0.568	147.47	125.65	0.440	113.04	90.44
D/E	0.332	86.00	69.03	0.272	70.00	53.82	0.136	34.75	16.58	0.148	38.26	30.87
SVAR	-0.428	-112.45	-223.69	-0.396	-103.24	-207.98	-0.476	-124.32	-234.53	-0.580	-150.67	-244.33
B/M	-0.108	-27.68	-54.53	-0.004	-1.06	-23.31	-0.236	-61.88	-98.86	0.008	2.17	-18.37
NTIS	-0.208	-54.83	-153.66	-0.276	-71.76	-174.16	-0.284	-74.78	-175.48	-0.384	-100.19	-196.10
TBL	0.884	227.86	251.33	0.876	225.54	249.17	0.880	226.22	252.14	0.860	220.63	252.82
LTY	0.972	248.54	304.26	0.996	254.84	313.79	0.876	224.22	276.48	0.980	250.56	311.88
LTR	0.220	57.08	82.19	0.292	76.29	105.55	0.044	11.70	15.98	0.376	97.76	132.57
TMS	1.416	373.67	359.60	1.448	382.53	368.90	1.328	349.67	334.63	1.436	378.23	364.23
DFY	-0.884	-231.11	-346.23	-0.876	-228.08	-330.02	-0.748	-195.84	-307.72	-0.552	-143.22	-208.32
DFR	0.484	125.76	128.86	0.412	106.57	104.37	0.452	117.30	122.86	0.384	98.93	102.70
INFL	0.648	168.31	194.66	0.672	174.35	202.80	0.616	159.81	180.28	0.692	179.23	211.90
I/K	1.296	334.47	351.27	1.296	334.23	352.63	1.276	330.04	336.30	1.344	346.10	363.83
Panel B: Combining Method												
Mean	1.252	331.29	393.73	1.292	342.18	404.80	1.268	334.31	401.05	1.280	335.55	409.60
Median	0.756	199.73	250.54	0.616	163.80	198.93	0.844	222.77	276.58	0.844	221.87	275.99
Trimmed Mean	1.160	307.22	375.05	1.176	311.39	381.85	1.188	312.96	383.61	1.208	317.40	394.36
DMSFE(1)	1.328	351.32	416.12	1.360	360.40	424.64	1.336	352.53	421.31	1.332	349.27	424.73
DMSFE(0.9)	1.500	396.47	461.06	1.512	400.49	464.24	1.484	391.65	460.93	1.432	376.34	452.68
Cluster 2	1.464	385.07	411.83	1.296	340.49	368.95	1.372	359.50	388.12	1.104	288.53	319.81
Cluster 3	1.116	291.79	322.80	1.200	314.99	345.08	1.184	309.08	339.97	1.300	340.14	376.57
Principal Components	1.644	431.32	461.80	1.664	436.44	464.99	1.660	436.00	467.54	1.648	432.60	463.38

Notes: See Table 5.

Table 7. Economic evaluation of robust point forecasts and combining methods (Time varying weights)

	TVW1			TVW2			TVW3		
Panel A: Univariate Models	**ΔU**	**Φ**	**Θ**	**ΔU**	**Φ**	**Θ**	**ΔU**	**Φ**	**Θ**
D/P	-0.228	-58.35	-36.51	-0.080	-21.01	-51.44	-0.340	-86.61	-69.97
D/Y	0.096	24.87	86.33	0.252	64.85	115.50	0.088	23.07	84.20
E/P	-0.120	-31.11	-61.59	-0.108	-27.91	-74.53	-0.368	-93.94	-108.50
D/E	0.788	203.79	249.59	0.712	185.23	229.66	0.972	250.31	308.63
SVAR	0.240	63.01	27.74	0.212	55.73	12.20	0.416	107.81	149.64
B/M	-0.392	-102.43	-137.07	-0.556	-145.09	-180.72	-0.664	-171.06	-211.52
NTIS	0.340	89.42	3.92	0.204	53.89	-28.31	0.680	175.60	225.20
TBL	1.916	502.73	539.42	1.872	491.11	526.69	2.064	544.76	573.92
LTY	1.604	419.28	477.80	1.448	378.52	423.89	1.492	389.70	430.37
LTR	1.472	388.09	465.84	1.296	342.03	419.04	1.320	345.22	419.84
TMS	2.052	542.88	573.92	2.084	552.75	585.27	1.756	460.75	489.98
DFY	-0.028	-7.59	23.45	0.088	22.87	44.30	0.400	102.72	182.27
DFR	0.684	177.66	237.69	0.716	186.26	244.82	0.696	178.09	258.51
INFL	1.608	417.64	516.24	1.592	415.45	515.46	1.644	426.44	523.45
I/K	0.992	255.28	308.74	1.044	269.24	309.43	1.052	270.15	320.04
Panel B: Combining Method									
Mean	1.780	465.01	551.74	1.736	454.71	538.34	1.652	427.58	514.98
Median	1.136	295.49	368.83	0.960	251.36	325.03	1.264	325.50	410.78
Trimmed Mean	1.712	446.27	533.89	1.620	423.88	510.59	1.584	409.18	496.63

Notes: See Table 5.

Table 8. Economic evaluation of robust point forecasts based on combined information

	FW1			FW2			FW3			FW4		
Panel A	**ΔU**	**Φ**	**Θ**	**ΔU**	**Φ**	**Θ**	**ΔU**	**Φ**	**Θ**	**ΔU**	**Φ**	**Θ**
Mean	1.252	331.29	393.74	1.292	342.16	404.80	1.268	334.31	401.05	1.280	335.56	409.63
Median	0.532	141.49	167.08	0.464	122.88	135.56	0.568	149.91	179.39	0.644	169.12	212.66
Trimmed Mean	1.148	303.61	371.31	1.168	309.80	379.06	1.160	306.43	376.27	1.164	305.22	380.93
DMSFE(1)	1.328	351.65	415.66	1.356	359.39	422.13	1.344	354.16	422.11	1.328	349.05	423.39
DMSFE(0.9)	1.504	397.63	462.71	1.524	403.48	466.21	1.480	390.30	460.48	1.404	368.18	445.80
Cluster 2	1.456	383.37	414.24	1.464	385.70	414.24	1.408	370.49	405.24	1.344	352.18	414.01
Cluster 3	0.980	256.29	287.44	1.072	280.52	309.96	0.992	259.34	295.86	1.096	285.38	338.62
Principal Components	1.380	359.82	384.69	1.524	398.04	423.43	1.428	372.48	398.05	1.504	393.63	420.88
BMA	1.600	410.57	471.59	1.640	422.45	483.24	1.320	337.78	394.93	1.148	293.66	347.76
Panel B	**TVW1**			**TVW2**			**TVW3**					
Mean	1.840	477.53	566.67	1.796	468.37	557.36	1.540	396.80	485.88			
Median	1.324	343.17	429.34	1.156	300.64	388.00	1.232	316.44	404.86			
Trimmed Mean	1.780	460.66	549.05	1.716	446.62	537.82	1.496	384.96	473.33			
BMA	1.320	340.72	407.90	1.176	302.83	375.89	1.376	355.33	427.07			

Notes: See Table 5.