Out-of-Sample Return Predictability: a Quantile Combination Approach

Luiz Renato Lima^{*a} and Fanning Meng^a

August 8, 2016

Abstract

This paper develops a novel forecasting method that minimizes the effects of weak predictors and estimation errors on the accuracy of equity premium forecasts. The proposed method is based on an averaging scheme applied to quantiles conditional on predictors selected by *LASSO*. The resulting forecasts outperform the historical average, and other existing models, by statistically and economically meaningful margins.

Keywords: LASSO, quantile regression, equity premium prediction, forecast combination. *JEL Codes:* C13, C14, C51, C53

We thank the editor Jonathan Wright and two anonymous referees for their valuable and insightful comments. We also appreciate feedbacks from seminar participants at the Midwest Economic Association, the International Symposium on Forecasting, the University of Tennessee and the University of Illinois at Urbana-Champaign. We thank Asa Lambert and John McMahan for their careful reading of the paper.

^{*}Corresponding author (e-mail: llima@utk.edu).

^aDepartment of Economics, University of Tennessee at Knoxville, TN 37996.

1 Introduction

Stock return is a key variable to firms' capital structure decisions, portfolio management, asset pricing and other financial problems. As such, forecasting return has been an active research area since Dow (1920). Rapach and Zhou (2013), Campbell (2000) and Goyal and Welch (2008) illustrate a number of macroe-conomic predictors and valuation ratios which are often employed in equity premium forecasting models. Valuation ratios include the dividend-price, earnings-price and book-to-market ratios an macroeconomic variables include nominal interest rates, the inflation rate, term and default spreads, corporate issuing activity, the consumption-wealth ratio, and stock market volatility. However, if such predictors are weak in the sense that their effects on the conditional mean of the equity premium are very small, then including them in the forecasting equation will result in low-accuracy forecasts which may be outperformed by the simplistic historical average (HA) model.

This paper develops a forecasting method that minimizes the negative effects of weak predictors and estimation errors on equity premium forecasts. Our approach relies on the fact that the conditional mean of a random variable can be approximated through the combination of its quantiles. This method has a long tradition in statistics and has been applied in the forecasting literature by Judge et al. (1988), Taylor (2007), Ma and Pohlman (2008) and Meligkotsidou et al. (2014). Our novel contribution to the literature is that we explore the existence of weak predictors in the quantile functions, which are identified through the ℓ_1 -penalized (*LASSO*) quantile regression method (Belloni and Chernozhukov (2011)). In applying such a method, we select predictors significant at the 5% level for the quantile functions. Next, we estimate quantile regressions with only the selected predictors, resulting in the post-penalized quantiles. These quantiles are then combined to obtain a point forecast of the equity premium, named the post-*LASSO* quantile combination (*PLQC*) forecast.

Our approach essentially selects a specification for the prediction equation of the equity premium. If a given predictor is useful to forecast some, but not all, quantiles of the equity premium, it is classified as partially weak. If the predictor helps forecast all quantiles, it is considered to be strong, whereas predictors that help predict no quantile are called fully weak predictors. The ℓ_1 -penalized method sorts the predictors according to this classification. The quantile averaging results in a prediction equation in which the coefficients of fully weak predictors are set to zero, while the coefficients of partially weak predictors are adjusted to reflect the magnitude of their contribution to the equity premium forecasts. Our empirical results show that of the 15 commonly used predictors that we examine, 9 are fully weak and 6 are partially weak. We show that failing to account for partially weak predictors results in misspecified prediction equations and, therefore, inaccurate equity premium forecasts.

We demonstrate that the proposed PLQC method offers significant improvements in forecast accuracy over not only the historical average, but also over many other forecasting models. This holds for both statistic and economic evaluations across several out-of-sample intervals. Furthermore, we develop a decomposition of the mean-square-prediction-error (*MSPE*) in order to summarize the contribution of each step of the proposed *PLQC* approach. In other words, we measure the additional loss that would arise from weak predictors and/or the estimation errors caused by extreme observations of equity premium. In particular, our results point out that in the 1967.1-1990.12 period, weak predictors explain about 15% of additional loss resultant from the non-robust forecast relative to the *PLQC* forecast. However, when we look at the 1991.1-2013.12 out-of-sample period, two-thirds of the loss of accuracy comes from the existence of weak predictors. Not surprisingly, the forecasts that fail to account for weak predictors are exactly the ones largely outperformed by the historical average during the 1991.1-2013.12 period.

Additionally, we conduct a robustness analysis by considering quantile combination models based on known predictors.¹ These models are not designed to deal with partial and fully weak predictors across quantiles and over time. Our empirical results show that equity premium forecasts obtained by combining quantile forecasts from such models are unable to provide a satisfactory solution to the original puzzle reported by Goyal and Welch (2008).

The remainder of this paper is organized as follows. Section 2 presents the econometric methodology and introduces the quantile combination approach. It also offers a comparison of the new and existing forecasting methods. Section 3 presents the main results about using a quantile combination approach to forecast the equity premium. Section 4 concludes.

¹We also conducted a robustness analysis using the "kitchen-sink" and common-factor models. Results are shown in the online appendix at http://econ.bus.utk.edu/department/faculty/lima.asp.

2 Econometric Methodology

Suppose that an econometrician is interested in forecasting the equity premium² of the S&P 500 index $\{r_{t+1}\}$, given the information available at time t, I_t . The data generating process (DGP) is defined as

$$r_{t+1} = X'_{t+1,t} \alpha + (X'_{t+1,t} \gamma) \eta_{t+1}, \qquad (1)$$

$$\eta_{t+1} | I_t \sim i.i.d. F_{\eta} (0, 1),$$

where $F_{\eta}(0,1)$ is some distribution with mean zero and unit variance that does not depend on I_t ; $X_{t+1,t} \in I_t$ is a $k \times 1$ vector of covariates available at time t; $\alpha = (\alpha_0, \alpha_1, ..., \alpha_{k-1})'$ and $\gamma = (\gamma_0, \gamma_1, ..., \gamma_{k-1})'$ are $k \times 1$ vectors of parameters, α_0 and γ_0 being intercepts. This is the conditional location-scale model that satisfies assumption D.2 of Patton and Timmermann (2007) and includes most common volatility processes, e.g. *ARCH* and stochastic volatility. Several special cases of model (1) have been considered in the forecasting literature³. In this paper, we consider another special case of model (1) by imposing $X'_{t+1,t} = X'_t$, a vector of predictors observable at time t. In this case, the conditional mean of r_{t+1} is given by $E(r_{t+1}|X_t) = X'_t \alpha$, whereas the conditional quantile of r_{t+1} at level $\tau \in (0, 1)$, $Q_{\tau}(r_{t+1}|X_t)$, equals:

$$Q_{\tau}\left(r_{t+1}|X_{t}\right) = X_{t}^{\prime}\alpha + X_{t}^{\prime}\gamma F_{\eta}^{-1}\left(\tau\right) = X_{t}^{\prime}\beta\left(\tau\right)$$

$$\tag{2}$$

where $\beta(\tau) = \alpha + \gamma F_{\eta}^{-1}(\tau)$, and $F_{\eta}^{-1}(\tau)$ is the unconditional quantile of η_{t+1} . Thus, this model generates a linear quantile regression for r_{t+1} , where the conditional mean parameters, α , enter in the definition of the quantile parameter, $\beta(\tau)$.⁴

Following the literature (Granger (1969), Granger and Newbold (1986), Christoffersen and Diebold (1997), and Patton and Timmermann (2007)), we assume that the loss function is defined as:

Assumption 1 (Loss Function) The loss function L is a homogeneous function solely of the forecast error $e_{t+1} \equiv r_{t+1} - \hat{r}_{t+1}$, that is, $L = L(e_{t+1})$, and L(ae) = g(a)L(e) for some positive function g.⁵

⁴Model (1) can be replaced with the assumption that the quantile function of r_{t+1} is linear. Another model that generates linear quantile regression is the random coefficient model studied by Gaglianone, Lima, Linton and Smith (2011).

²The equity premium is calculated by subtracting the risk-free return from the return of the S&P 500 index.

³Gaglianone and Lima (2012) and (2014) assume $X_{t+1,t} = (1, C_{t+1,t})'$ and $X_{t+1,t} = (1, f_{t+1,t}^1, ..., f_{t+1,t}^n)'$ respectively, where $C_{t+1,t}$ is the consensus forecast made at time t from the Survey of Professional Forecasts and $f_{t+1,t}^j$, j = 1, ..., n, are point forecasts made at time t by different economic agents.

⁵This is exactly the same Assumption L2 of Patton and Timmermann (2007). Although it rules out certain loss functions (e.g., those which also depend on the level of the predicted variable), many common loss functions are of this form, such as MSE, MAE, lin-lin, and asymmetric quadratic loss.

Proposition 1 presents our result on forecast optimality. It is a special case of the Proposition 3 of Patton and Timmermann (2007) in the sense that we assume a DGP with specific dynamic for the mean and variance. Under this case, we are able to show that the optimal forecast of the equity premium can be decomposed as the sum of its conditional mean and a bias measure

Proposition 1 Under DGP(1) with $X'_{t+1,t} = X'_t$ and a homogeneous loss function (Assumption 1), the optimal forecast will be

$$\widehat{r}_{t+1} = Q_{\tau} (r_{t+1} | X_t)$$
$$= E (r_{t+1} | X_t) + \kappa_{\tau}$$

where $\kappa_{\tau} = X'_t \gamma F_{\eta}^{-1}(\tau)$ is a bias measure relative to the conditional mean (MSPE) forecast. This bias depends on X_t , the distribution F_{η} and loss function L.⁶

The above result suggests that, when estimation of the conditional mean is affected by the presence of extreme observations as is the case with financial data, an approach to obtain robust *MSPE* forecasts of equity premium is through the combination of quantile forecasts. That is:

$$\sum_{\tau=\tau_{\min}}^{\tau_{\max}} \omega_{\tau} Q_{\tau} \left(r_{t+1} | X_t \right) = E \left(r_{t+1} | X_t \right) + \sum_{\tau=\tau_{\min}}^{\tau_{\max}} \omega_{\tau} \kappa_{\tau}$$
$$= E \left(r_{t+1} | X_t \right) + X_t' \gamma \sum_{\tau=\tau_{\min}}^{\tau_{\max}} \omega_{\tau} F_{\eta}^{-1} \left(\tau \right)$$

where ω_{τ} is the weight assigned to the conditional quantile $Q_{\tau}(r_{t+1}|X_t)$. Notice that the weights are quantile-specific since they are aimed at approximating the mean of η_{t+1} , which is zero. In the one-sample setting, integrating the quantile function over the entire domain [0, 1] yields the mean of the sample distribution (Koenker, 2005, pg 302). Thus, given that η_{t+1} is i.i.d., we have $E(\eta_{t+1}) = \int_0^1 F_{\eta}^{-1}(t) dt = 0.^7$ However, with finite sample, we need to consider a grid of quantiles ($\tau_{\min}, ..., \tau_{\max}$) and approximate $\int_0^1 F_{\eta}^{-1}(t) dt$ by $\sum_{\tau=\tau_{\min}}^{\tau_{\max}} \omega_{\tau} F_{\eta}^{-1}(\tau)$. The choice of the weight ω_{τ} reflects the potential asymmetry and excess kurtosis of the conditional distribution of η_{t+1}, F_{η} . In the simplest case when F_{η} is symmetric, assigning equal weight to quantiles located in the neighborhood of the median ($\tau = 0.5$)

⁶The proof is in the online appendix available at http://econ.bus.utk.edu/department/faculty/lima. asp.

⁷Recall that $F_{\eta}^{-1}(\tau) = Q_{\tau}(\eta_{t+1}).$

will suffice to guarantee that $\sum_{\tau=\tau_{\min}}^{\tau_{\max}} \omega_{\tau} Q_{\tau} (r_{t+1}|X_t) = E(r_{t+1}|X_t)$. However, when F_{η} is asymmetric, other weighting schemes should be used. In this paper, we consider two weighting schemes.

The robustness of this approach relies on the fact that $Q_{\tau}(r_{t+1}|X_t)$ are estimated using the quantile regression (*QR*) estimator, which is robust to estimation errors caused by occasional but extreme observations of equity premium.⁸ Since the low-end (high-end) quantiles produce downwardly (upwardly) biased forecast of the conditional mean, another insight from the combination approach is that the point forecast $\sum_{\tau=\tau_{\min}}^{\tau_{\max}} \omega_{\tau} Q_{\tau}(r_{t+1}|X_t)$ combines oppositely-biased predictions, and these biases cancel out each other. This cancelling out mitigates the problem of aggregate bias identified by Issler and Lima (2009).⁹

To our knowledge, the previous discussion is the first to provide a theoretical explanation of several empirical results which use the combination of conditional quantiles to approximate the conditional mean forecast (Judge et al. (1988), Taylor (2007), Ma and Pohlman (2008) and Meligkotsidou et al. (2014)). A common assumption in those papers is that the specification of the conditional quantile $Q_{\tau} (r_{t+1}|X_t)$ is fully known by the econometrician. However, *DGP* (1) is unknown. Therefore, the forecasting model based on the combination of conditional quantiles with fixed predictors is still potentially misspecified, especially when predictors are weak. In what follows, we explain how we address the problem of weak predictability in the conditional quantile function.

2.1 The ℓ_1 -penalized quantile regression estimator

Rewriting Equation 2, we have the conditional quantiles of r_{t+1} :

$$Q_{\tau}(r_{t+1}|X_t) = \beta_0(\tau) + x'_t \beta_1(\tau) \qquad \tau \in (0,1)$$

where $\beta_0(\tau) = \alpha_0 + \gamma_0 F_{\eta}^{-1}(\tau)$, $\beta_1(\tau) = \alpha_1 + \gamma_1 F_{\eta}^{-1}(\tau)$, and x_t is a $(k-1) \times 1$ vector of predictors (excluding the intercept).

In this paper, we identify weak predictors by employing a convex penalty to the quantile regression coefficients, leading to the ℓ_1 -penalized (*LASSO*) quantile regression estimator (Belloni and Chernozhukov (2011)). The *LASSO* quantile regression estimator solves the following problem :

$$\min_{\beta_0,\beta_1} \sum_{t} \rho_\tau(r_{t+1} - \beta_0(\tau) - x'_t \boldsymbol{\beta}_1(\tau)) + \lambda \frac{\sqrt{\tau(1-\tau)}}{m} \| \boldsymbol{\beta}_1(\tau) \|_{\ell_1}$$
(3)

⁸The robustness of an estimator can be obtained through what is known as an influence function. Following Koenker (2005, section 2.3), the influence function of the quantile regression estimator is bounded whereas that of the *OLS* estimator is not.

⁹Aggregate bias arises when we combine predictions that are mostly upwardly (downwardly) biased. In a case like that, the averaging scheme will not minimize the forecast bias.

where ρ_{τ} denotes the "tick" or "check" function defined for any scalar e as $\rho_{\tau}(e) \equiv [\tau - 1 (e \leq 0)] e$; 1 (·) is the usual indicator function; m is the size of the estimation sample; $\| \cdot \|_{\ell_1}$ is the ℓ_1 -norm, $\| \beta_1 \|_{\ell_1} = \sum_{i=1}^{k-1} |\beta_{1i}|; x_t = (x_{1,t}, x_{2,t}, ..., x_{(k-1),t})'.$

LASSO first selects predictor(s) from the information set $\{x_{i,t} : i = 1, 2, ..., (k-1)\}$ for each quantile τ at each time period t (Van de Geer (2008) and Manzan (2015)). As for the choice of the penalty level, λ , we follow Belloni and Chernozhukov(2011) and Manzan (2015). Next, we estimate a quantile regression with the selected predictors to generate a post-*LASSO* quantile forecast of the equity premium in t + 1, denoted as $f_{t+1,t}^{\tau} = \beta_0(\tau) + x_t^{*'}\beta(\tau)$, where x_t^* is(are) the predictor(s) selected at time t by the *LASSO* procedure with 5% significance level. This procedure is repeated to obtain a *PLQF* at various $\tau \in (0, 1)$. Finally, these *PLQFs* are combined to obtain the post-*LASSO* quantile combination (*PLQC*) forecast, $\hat{r}_{t+1} = \sum_{j=1}^{k} \omega_{\tau_j} f_{t+1,t}^{\tau_j}$. The (*PLQC*) is a point (MSPE) forecast of the equity premium in t + 1.

2.2 An alternative interpretation to the *PLQC* forecast

In this section, we show that the *PLQC* forecast can be represented by a prediction equation, which is robust to the presence of weak predictors and estimation errors.

We assume a vector of potential predictors $x_t = (1 \ x_{1,t} \ x_{2,t} \ x_{3,t})'$ available at time t and quantiles $\tau \in (\tau_1, ..., \tau_5)$. Based on x_t and τ , we obtain *PLQFs* of the equity premium in t+1, $f_{t+1,t}^{\tau_j}$, j = 1, 2, ..., 5:

$$\begin{pmatrix} f_{t+1,t}^{\tau_1} \\ f_{t+1,t}^{\tau_2} \\ f_{t+1,t}^{\tau_3} \\ f_{t+1,t}^{\tau_4} \\ f_{t+1,t}^{\tau_5} \\ f_{t+1,t}^{\tau_5} \end{pmatrix} = \begin{pmatrix} \beta_0(\tau_1) & \beta_1(\tau_1) & 0 & 0 \\ \beta_0(\tau_2) & \beta_1(\tau_2) & 0 & 0 \\ \beta_0(\tau_3) & \beta_1(\tau_3) & 0 & 0 \\ \beta_0(\tau_4) & 0 & \beta_2(\tau_4) & 0 \\ \beta_0(\tau_5) & 0 & \beta_2(\tau_5) & 0 \end{pmatrix} \times \begin{pmatrix} 1 \\ x_{1,t} \\ x_{2,t} \\ x_{3,t} \end{pmatrix}$$
(4)

In this example, $x_{3,t}$ is fully weak in the population because it does not help predict any quantile. In contrast, we define $x_{1,t}$ and $x_{2,t}$ as partially weak predictors because they help predict some, but not all quantiles. The *PLQC* forecast is generated based on Equation (5):

$$\widehat{r}_{t+1} = \sum_{j=1}^{5} \omega_{\tau_j} \beta_0(\tau_j) + \sum_{j=1}^{3} \omega_{\tau_j} \beta_1(\tau_j) x_{1,t} + \sum_{j=4}^{5} \omega_{\tau_j} \beta_2(\tau_j) x_{2,t}$$

$$= \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t}$$
(5)

where $\beta_0 = \sum_{j=1}^5 \omega_{\tau_j} \beta_0(\tau_j)$, $\beta_1 = \sum_{j=1}^3 \omega_{\tau_j} \beta_1(\tau_j)$ and $\beta_2 = \sum_{j=4}^5 \omega_{\tau_j} \beta_2(\tau_j)$.

Standard model selection procedures such as the one proposed by Koenker and Machado (1999) are not useful to select weak predictors for out-of-sample forecasting. Indeed, with only 3 predictors, 5 different quantile levels, $\tau \in (\tau_1, ..., \tau_5)$, and 300 time periods, there would potentially exist 12,000 models to be considered for estimation, which is computationally prohibitive. This is why we use the ℓ_1 -penalized quantile regression method to determine the most powerful predictors among all candidates. The ℓ_1 -penalized quantile regression method rules out the fully weak predictor $x_{3,t}$ from the prediction equation, whereas $x_{1,t}$ and $x_{2,t}$ are included but their contribution to the point forecast \hat{r}_{t+1} will reflect their partial weakness. Indeed, since the contribution of $x_{1,t}$ to predict $f_{t+1,t}^{\tau_4}$ and $f_{t+1,t}^{\tau_5}$ is weak, our forecasting device eliminates $\beta_1(\tau_4)$ and $\beta_1(\tau_5)$ from β_1 in Equation (5). The same rationale explains the absence of $\beta_2(\tau_1)$, $\beta_2(\tau_2)$ and $\beta_2(\tau_3)$ in β_2 .

Moreover, if we assume that τ_1 and τ_2 are low-end quantiles whereas τ_4 and τ_5 are high-end quantiles, the coefficient matrix (4) suggests that predictor $x_{1,t}$ is prone to make downwardly biased forecasts, whereas predictor $x_{2,t}$ is prone to make upwardly biased forecasts. These oppositely-biased forecasts are then combined by Equation (5) to generate a low-bias and low *MSPE* point forecast. Thus, we avoid the problem of aggregate bias that affects traditional forecast combination methods (Issler and Lima (2009)).

Two inefficient special cases may arise when one ignores the presence of partially weak predictors. In the first case, we estimate quantile regressions with the same predictors across $\tau \in (\tau_1, ..., \tau_5)$ to obtain the Fixed-predictor Quantile Regression (*FQR*) forecast:

$$\widehat{r}_{t+1} = b_0 + b_1 x_{1,t} + b_2 x_{2,t} \tag{6}$$

where $b_0 = \sum_{j=1}^5 \omega_{\tau_j} \beta_0(\tau_j)$, $b_1 = \sum_{j=1}^5 \omega_{\tau_j} \beta_1(\tau_j)$ and $b_2 = \sum_{j=1}^5 \omega_{\tau_j} \beta_2(\tau_j)$.

The second special case corresponds to the estimation of the prediction Equation (6) by *OLS* regression of r_{t+1} on the selected predictors, $x_{1,t}$, $x_{2,t}$, and an intercept, resulting in the Fixed *OLS* (*FOLS*) forecast. Although both *FQR* and *FOLS* forecasts rule out the fully weak predictors, they do not account for the presence of partially weak predictors in the population. Moreover, the *FOLS* forecasts will not be robust against extreme observations, since the influence function of the *OLS* estimator is unbounded.

To show the relative importance of accounting for partially weak predictors and estimation errors, we consider the following decomposition:

$$MSPE_{FOLS} - MSPE_{PLQC} = [MSPE_{FOLS} - MSPE_{FQR}] + [MSPE_{FQR} - MSPE_{PLQC}]$$
(7)

Hence, we decompose the MSPE difference between FOLS and PLQC forecasts into two elements. The first element on the righthand side of Equation (7) measures the additional loss of the *FOLS* forecast resulted from *OLS* estimator's lack of robustness to the estimation errors, while the second element represents the extra loss caused by the presence of partially weak predictors in the population. We will apply this decomposition later in the empirical section.

2.3 Weight selection

In this paper, we consider both time-invariant and time-variant weighting schemes. The former are simple averages of $f_{t+1,t}^{\tau}$. More specifically, we consider a discrete grid of quantiles $\tau \in (\tau_1, \tau_2, ..., \tau_J)$ and set equal weights $\omega_{\tau} = \omega$. Two leading examples are

$$PLQC_{1} : \frac{1}{3}f_{t+1,t}^{0.3} + \frac{1}{3}f_{t+1,t}^{0.5} + \frac{1}{3}f_{t+1,t}^{0.7}$$

$$PLQC_{2} : \frac{1}{5}f_{t+1,t}^{0.3} + \frac{1}{5}f_{t+1,t}^{0.4} + \frac{1}{5}f_{t+1,t}^{0.5} + \frac{1}{5}f_{t+1,t}^{0.6} + \frac{1}{5}f_{t+1,t}^{0.7}$$

Thus, the $PLQC_1$ and $PLQC_2$ attempt to approximate the point (*MSPE*) forecast by assigning equal weights to a discrete set of conditional quantiles. However, the importance of quantiles in the determination of optimal forecasts may not be equal and constant over time. To address this problem, we estimate the weights from a constrained *OLS* regression of r_{t+1} on $f_{t+1,t}^{\tau}$, $\tau \in (\tau_1, \tau_2, ..., \tau_J)$, with the following two leading examples:

$$PLQC_{3} : r_{t+1} = \sum_{\tau=\tau_{1}}^{\tau_{3}} \omega_{\tau} f_{t+1,t}^{\tau} + \varepsilon_{t+1} \quad \tau \in (0.3; 0.5; 0.7)$$

$$s.t. \quad \omega_{\tau_{1}} + \omega_{\tau_{2}} + \omega_{\tau_{3}} = 1$$

$$PLQC_{4} : r_{t+1} = \sum_{\tau=\tau_{1}}^{\tau_{5}} \omega_{\tau} f_{t+1,t}^{\tau} + \varepsilon_{t+1} \quad \tau \in (0.3; 0.4; 0.5; 0.6; 0.7)$$

$$s.t. \quad \omega_{\tau_{1}} + \omega_{\tau_{2}} + \omega_{\tau_{3}} + \omega_{\tau_{4}} + \omega_{\tau_{5}} = 1$$

$$(8)$$

Similar weighting schemes have been used in the forecasting literature by Judge et al. (1988), Taylor (2007), Ma and Pohlman (2008) and Meligkotsidou et al. (2014), among others.

2.4 The forecasting data, procedure and evaluation

Before explaining the forecasting data, we introduce the standard univariate predictive regressions estimated by *OLS* (Goyal and Welch (2008) and Rapach et al. (2010)). They are expressed as:

$$r_{t+1} = \alpha_i + \beta_i x_{i,t} + \varepsilon_{i,t+1} \tag{9}$$

where $x_{i,t}$ is a variable whose predictive ability is of interest; $\varepsilon_{i,t+1}$ is an i.i.d. error term; α_i and β_i are respectively the intercept and slope coefficients specific to model i = 1, ..., N. Each univariate model i yields its own forecast of r_{t+1} labeled as $f_{t+1,t}^i = \widehat{E}(r_{t+1}|X_t) = \widehat{\alpha}_i + \widehat{\beta}_i x_{i,t}$, where $\widehat{\alpha}_i$ and $\widehat{\beta}_i$ are *OLS* estimates of α_i and β_i .

Our data¹⁰ contain monthly observations of the equity premium to the *S&P* 500 index and 15 predictors, which include Dividend-price ratio (*DP*), Dividend yield (*DY*), Earnings-price ratio (*EP*), Dividend-payout ratio (*DE*), Stock variance (*SVAR*), Book-to-market ratio (*BM*), Net equity expansion (*NTIS*), Treasury bill rate (*TBL*), Long-term yield (*LTY*), Long-term return (*LTR*), Term spread (*TMS*), Default yield spread (*DFY*), Default return spread (*DFR*), Inflation (*INFL*) and a moving average of Earning-price ratio (*E10P*), from December 1926 to December 2013. Contrary to Goyal and Welch (2008), we do not lag the predictor *INFL*, which implies that we are assuming adaptive expectations for future price changes.¹¹

In our empirical application, we generate out-of-sample forecasts of the equity premium, r_{t+1} , using (i) 15 single-predictor regression models based on Equation (9); (ii) the *PLQC* and *FQR* methods with the four weighting schemes presented above; (iii) the *FOLS* method; (iv) the complete subset regressions (*CSR*) with k = 1, 2 and 3. The *CSR* method (Elliott, Gargano and Timmermann (2013)) combine forecasts based on predictive regressions with k number of predictors. Hence, forecasts based on *CSR* with k = 1 correspond to an equal-weighted average of all possible forecasts from univariate prediction models (Rapach et al. (2010)). *CSR* models with k = 2 and 3 correspond to equal-weighted averages of all possible forecasts from bivariate and tri-variate prediction equations, respectively.

Following Rapach et al. (2010), Campbell and Thompson (2008) and Goyal and Welch (2008) among others, we use the historical average of equity premium, $\overline{r}_{t+1} = \frac{1}{t} \sum_{m=1}^{t} r_m$, as our benchmark model. If the information available at $X_t = (1, x_{1,t}, x_{2,t}, ..., x_{15,t})'$ is useful to predict equity premium, the forecasting models based on X_t should outperform the benchmark.

The forecasting procedure is based on recursive estimation window (Rapach et al. (2010)). Our estimation window starts with 361 observations from December 1926 to December 1956 and expands periodically as we move forward. The out-of-sample forecasts range from January 1957 to December 2013, corresponding to 684 observations. In addition, forecasts that rely on time-varying weighting

¹⁰The raw data come from Amit Goyal's webpage (http://www.hec.unil.ch/agoyal/).

¹¹A more complete definition for each variable can be found in the online appendix http://econ.bus.utk.edu/ department/faculty/lima.asp.

schemes ($PLQC_3$, $PLQC_4$, FQR_3 , FQR_4) require a holdout period to estimate the weights. Thus, we use the first 120 observations from the out-of-sample period as an initial holdout period, which also expands periodically. In the end, we are left with a total of 564 post-holdout out-of-sample forecasts available for evaluation.¹² In addition to the whole (long) out-of-sample period (January 1967 to December 2013), we test the robustness of our findings by considering the following out-of sample subperiods: January 1967 to December 1990, January 1991 to December 2013, and the most recent interval January 2008 to December 2013.

The first evaluation measure is the out-of-sample R^2 , R_{OS}^2 , which compares the forecast from a conditional model, \hat{r}_{t+1} , to that from the benchmark (unconditional) model \bar{r}_{t+1} (Campbell and Thompson (2008)). We report the value of R_{OS}^2 in percentage terms, R_{OS}^2 (%) = $100 \times R_{OS}^2$. Second, to test the null hypothesis $R_{OS}^2 \leq 0$, we apply both the Diebold-Mariano (1995) and the Clark and West (2007) tests¹³. Lastly, to evaluate the economic value of equity premium forecasts, we calculate the *certainty equivalent return (or utility gain)*, which can be interpreted as the management fee an investor would be willing to pay to have access to the additional information provided by the conditional forecast models relative to the information available in the benchmark model¹⁴.

3 Empirical Results

3.1 Out-of-sample forecasting results

In Figures 1 and 2, we present time series plots of the differences between the cumulative squared prediction error for the benchmark forecast and that of each conditional forecast. This graphical analysis informs the cumulative performance of a given forecasting model compared to the benchmark model over time. When the curve in each panel increases, the conditional model outperforms the benchmark, while the opposite holds when the curve decreases. Moreover, if the curve is higher at the end of the period, the conditional model has a lower *MSPE* than the benchmark over the whole out-of-sample period.

¹²This forecasting procedure follows exactly the same one adopted by Rapach et al. (2010).

¹³The Diebold and Mariano (1995) and West (1996) statistics are often used to test the null hypothesis, $R_{OS}^2 \leq 0$, among non-nested models. For nested models, as the ones in this paper, Clark and McCracken (2001) and McCraken (2007) show that these statistics have nonstandard distribution. Thus, the Diebold-Mariano (*DM*) and West tests can be severely undersized under the null hypothesis and have low power under the alternative hypothesis.

¹⁴For more detailed information on the calculation of utility gains, please refer to the online appendix at http://econ. bus.utk.edu/department/faculty/lima.asp.

Figure 1: Cumulative squared prediction error for the benchmark model minus the cumulative squared prediction errors for the single-predictor regression forecasting models, 1967.1-2013.12



A positively sloped curve in each panel indicates that the conditional model outperforms the *HA*, while the opposite holds for a downward sloping curve. Moreover, if the curve is higher at the end of the period, the conditional model has a lower *MSPE* than the benchmark over this period. Figure 1 shows that in terms of cumulative performance, few single-predictor models consistently outperform the benchmark.

In general, Figure 1 shows that in terms of cumulative performance, few single-predictor models consistently outperform the historical average.¹⁵ A number of the panels (such as the one based on TMS) exhibit increasing predictability in the first half of the sample period, but lose predictive strength thereafter. Also, the majority of the single-predictor forecasting models have a higher *MSPE* than the benchmark. Figure 1 looks very similar to that in Rapach et al. (2010, page 833) which uses quarterly data. Our results, which are based on monthly observations, show a significant deterioration of the single-predictor models after 1990.¹⁶ In sum, Figure 1 strengthens the arguments already reported throughout the literature (Goyal and Welch (2008), Rapach et al. (2010)), that it is difficult to identify individual predictors that help improve equity premium forecasts over time.

Figure 2 shows the same graphical analysis for $PLQC_j$, FQR_j , j = 1, 2, 3, 4, $FOLS_1$, $FOLS_2^{17}$ and

¹⁵One exception is the single predictor model based on *INFL*. Its curve is sloped upward for most of the time.

¹⁶Goyal and Welch (2008) as well as Rapach et al. (2010) considered quarterly forecasts of the equity premium.

¹⁷Recall that FOLS forecasts are based on the OLS estimation of an equation whose predictors are selected by the



Figure 2: Cumulative squared prediction error for the benchmark model minus the cumulative squared prediction errors for the *FQR*, *FOLS*, *CSR* and *PLQC* models, 1967.1-2013.12

A positively sloped curve in each panel indicates that the conditional model outperforms the *HA*, while the opposite holds for a downward sloping curve. Moreover, if the curve is higher at the end of the period, the conditional model has a lower *MSPE* than the benchmark over this period. Figure 2 shows that the *PLQC* forecasts are among top performers, especially after 1990.

CSR with k = 1, 2, 3. The curves for *PLQC_j* and *FQR_j* do not exhibit substantial falloffs as those observed in the single-predictor forecasting models (9). This indicates that the *PLQC_j* and *FQR_j* forecasts deliver out-of-sample gains on a considerably more consistent basis over time. The *PLQC* and *FQR* forecasts perform similarly, and *FQR* forecasts are only slightly better before 1990. Since the *PLQC* method accounts for partially weak predictors whereas *FQR* does not (this being the only difference between the two), the results shown in Figure 2 suggest that most of the predictors are not weak until 1990. The results in Figure 2 also provide the first empirical evidence about the ability of the *PLQC* model to efficiently predict monthly equity premium of the *S&P* 500 index¹⁸.

The comparison between FQR and FOLS shows the importance of using quantile regression to obtain a robust estimation of the prediction equation. Recall that FQR and FOLS rely on the same specification for the prediction equation, but they differ in how the coefficients are estimated. Comparing the panels corresponding to FQR and FOLS forecasts, we see how estimation errors in the prediction equation can result in a severe loss of forecasting accuracy. The curves of the FOLS forecasts are not only lower in magnitude but also much more erratic than the ones corresponding to the FQR forecasts. Finally, the CSRforecasts do not outperform the PLQC forecast. Besides being robust to the presence of weak predictors and estimation errors, the PLQC forecast results from the combination of different quantile forecasts, whose biases cancel out each other. This avoids the aggregate bias problem that affects most existing forecast combination methods including the CSR model (Issler and Lima (2009)).

We next turn to the analysis of all four out-of-sample periods. The results are displayed in Table 1. This table reports R_{OS}^2 statistics and its significance through the p-values of the Clark and West (2007) test (*CW*). It also displays the annual utility gain Δ (*annual*%) associated with each forecasting model and the p-value of the Diebold-Mariano (1995) test (*DM*). The results for the entire 1967.1:2013.12 out-of-sample period confirm that few single-predictor forecasting models have positive and significant R_{OS}^2 . The same thing happens to the *CSR* forecasts. The only exceptions in this long out-of-sample period are the *PLQC* and *FQR* forecasts. Their performance are similar in the sense that they both outperform the *FOLS* forecast in terms of R_{OS}^2 and utility gain Δ (*annual*%). Among the *PLQC* forecasts, we notice that

 $[\]ell_1$ -penalized quantile regression method. Since we have considered two sets of quantiles $\tau = (0.3, 0.5, 0.7)$ and $\tau = (0.3, 0.4, 0.5, 0.6, 07)$, there will be two such prediction equations and therefore two *FOLS* forecasts, denoted by *FOLS_j*, j = 1, 2.

¹⁸Based on a Monte-Carlo simulation experiment, we found that weak predictors can be harmful for forecasting. Our Monte-Carlo experiment can be found in the online appendix at http://econ.bus.utk.edu/department/faculty/ lima.asp.

the ones that rely on the combination of 5 quantiles perform better than those based on the combination of 3 quantiles during this period.

As for the subperiod 1967.1-1990.12, Table 1 shows that some single-predictor models perform well. In particular, forecasts from single-predictor models using either *DY* or *E10P* present positive and significant R_{OS}^2 and also sizable utility gains. The *CSR* forecasts are also reasonable and outperform the FOLS forecast. Recall that the difference between *PLQC* and *FQR* is that the latter ignores partially weak predictors, and therefore the result reported by Table 1 suggests that there is no advantage in using a forecasting device that is robust to (partially) weak predictors when predictors are actually strong. However, since *FQR* outperforms *OLS*-based *FOLS*, we conclude that forecasts which are robust against estimation errors provide a predictive advantage.

As for the 1991.1-2013.12 sub-period, we notice that the R_{OS}^2 of all single-predictor models fall substantially and become non-significant, suggesting that most of the predictors become weak after 1990. The same results for *CSR* forecasts indicate that this methodology is also affected by the presence of weak predictors. On the other hand, the results in Table 1 show that the R_{OS}^2 of the *PLQC* forecast does not fall much across the two sub-periods, confirming that this method is robust to weak predictors. Also, the R_{OS}^2 of the *FQR* forecasts falls on average by 0.22% whereas the R_{OS}^2 of the *PLQC* forecasts increases on average by 0.18%. This happens because the latter is robust to both fully and partially weak predictors whereas the former is only robust to fully weak predictors.

Finally, we look at the most recent out-of-sample subperiod, 2008.1-2013.12, characterized by the occurrence of the sub-prime crisis in the United States. A practitioner should expect that a good forecasting model would work reasonably well in periods of financial turmoil. However, the results in Table 1 suggest that none of the single-predictor models and the *CSR* forecasts perform well during this period of financial instability. In contrast, the statistic and economic measures of the *PLQC* forecasts are even better than those in other periods. More specifically, the R_{OS}^2 and utility gain statistics for $PLQC_j$ are at least twice as large as those for other out-of-sample periods. This suggests that the *PLQC* method works very well even during periods with multiple episodes of financial turmoil. These results provide strong evidence that we have identified an effective method for forecasting monthly equity premium on the *S&P* 500 index based on economic variables.

Table 2 shows the decomposition of the mean-square-prediction-error (*MSPE*) introduced in section 2.2. Recall that this decomposition measures the additional *MSPE* loss of *FOLS* forecasts relative to the *PLQC* forecasts. The first element on the right-hand side of equation measures the additional loss

	OOS: 1967.1 - 2013.12		OOS: 1967.1 - 1990.12			OOS:1991.1 - 2013.12				OOS: 2008.1 - 2013.12						
Model	$R_{OS}^{2}(\%)$	DM	CW	$\Delta(\text{annual}\%)$	$R_{OS}^{2}(\%)$	DM	CW	$\Delta(\text{annual}\%)$	$R_{OS}^2(\%)$	DM	CW	$\Delta(\text{annual}\%)$	$R_{OS}^{2}(\%)$	DM	CW	$\Delta(\text{annual}\%)$
Single Predictor Model Forecasts																
DP	-0.60	1.00	0.26	-0.10	1.31	0.30	0.03	1.72	-2.99	1.00	0.78	-1.99	-0.54	0.90	0.54	-0.73
DY	-1.01	1.00	0.22	0.22	1.55	0.38	0.02	2.16	-4.24	1.00	0.76	-1.79	-0.43	0.89	0.47	0.22
EP	-1.51	1.00	0.48	-0.29	-0.99	0.93	0.43	-0.67	-2.15	0.99	0.53	0.11	-3.17	0.92	0.61	2.36
DE	-0.71	0.98	0.98	-0.55	-0.90	1.00	1.00	-0.95	-0.48	0.57	0.73	-0.14	-1.15	0.84	0.75	-0.35
SVAR	-0.54	0.71	0.81	-0.26	-0.74	0.57	0.75	-0.10	-0.29	0.94	0.83	-0.43	-0.77	0.85	0.85	-1.53
BM	-3.54	1.00	0.62	-1.43	-2.75	0.95	0.44	-0.97	-4.53	1.00	0.79	-1.90	-0.85	0.95	0.52	0.11
NTIS	-0.96	0.58	0.36	-1.12	0.38	0.54	0.09	-0.34	-2.65	0.58	0.83	-1.92	-5.38	0.67	0.85	-6.82
TBL	0.09	0.24	0.07	2.09	0.37	0.27	0.06	4.11	-0.26	0.29	0.53	-0.01	0.56	0.04	0.24	1.25
LTY	-0.64	0.40	0.14	1.84	-1.08	0.38	0.13	3.60	-0.09	0.65	0.54	0.01	0.53	0.02	0.11	1.03
LTR	0.12	0.81	0.12	0.25	0.55	0.69	0.09	1.16	-0.43	0.79	0.42	-0.71	-0.08	0.56	0.40	-1.75
TMS	0.28	0.42	0.07	0.86	1.26	0.43	0.03	1.90	-0.96	0.46	0.53	-0.23	-0.24	0.13	0.45	-0.28
DFY	0.13	0.73	0.22	0.01	1.00	0.10	0.01	1.14	-0.97	0.99	0.87	-1.17	-1.16	0.53	0.75	-2.57
DFR	0.04	0.55	0.36	0.06	0.01	0.60	0.43	0.10	0.08	0.52	0.37	0.02	0.64	0.55	0.33	0.71
INFL	0.37	0.01	0.10	0.69	0.78	0.06	0.07	1.47	-0.14	0.03	0.51	-0.13	-1.07	0.34	0.83	-1.80
E10P	-1.42	1.00	0.17	0.06	1.32	0.58	0.04	1.84	-4.86	1.00	0.66	-1.78	0.03	0.86	0.33	0.70
				Com	plete Subse	et Regro	ession 1	Forecasts					-			
CSR k=1	0.39	0.86	0.06	0.49	1.29	0.07	0.00	1.58	-0.75	1.00	0.82	-0.65	-0.32	0.81	0.81	-0.54
CSR k=2	0.24	0.96	0.11	0.46	1.61	0.23	0.00	1.96	-1.48	1.00	0.83	-1.11	-0.49	0.70	0.71	-0.45
CSR k=3	-0.02	0.99	0.20	0.39	1.49	0.45	0.02	1.84	-1.93	1.00	0.82	-1.12	-0.56	0.57	0.60	0.56
Forecasts based on LASSO-Quantile Selection																
FOLS1	0.53	0.50	0.03	1.35	0.45	0.58	0.11	1.12	0.63	0.43	0.09	1.58	3.24	0.18	0.10	4.40
FOLS2	0.27	0.62	0.04	1.18	-0.19	0.75	0.17	0.71	0.84	0.40	0.07	1.67	3.69	0.20	0.09	4.16
FQR1	2.27	0.00	0.00	2.42	2.34	0.15	0.01	2.23	2.20	0.00	0.02	2.60	5.07	0.01	0.04	6.74
FQR2	2.10	0.02	0.00	2.25	1.96	0.38	0.02	2.15	2.29	0.01	0.02	2.34	5.33	0.01	0.04	5.79
FQR3	1.83	0.07	0.01	2.07	2.04	0.18	0.04	2.46	1.56	0.13	0.08	1.65	4.83	0.07	0.09	6.32
FQR4	1.56	0.13	0.02	1.83	1.82	0.33	0.05	2.27	1.24	0.12	0.11	1.37	3.29	0.08	0.14	4.57
PLQC1	2.12	0.01	0.01	1.59	1.81	0.16	0.05	1.03	2.50	0.02	0.04	2.19	6.50	0.07	0.06	5.19
PLQC2	2.27	0.00	0.01	1.86	2.23	0.09	0.04	1.76	2.31	0.01	0.04	1.96	5.84	0.04	0.06	4.71
PLQC3	1.62	0.09	0.04	1.69	1.61	0.13	0.10	2.46	1.62	0.21	0.11	1.79	5.63	0.17	0.10	3.55
PLQC4	2.16	0.06	0.03	2.16	2.20	0.16	0.08	2.97	2.11	0.11	0.08	1.32	6.08	0.10	0.08	4.51

Table 1: Out-of-sample Equity Premium Forecasting

This table reports R_{OS}^2 statistics (in%) and its significance through the p-values of the Clark and West (2007) test (*CW*). It also reports the p-value of the Diebold-Mariano (1995) test (*DM*) and the annual utility gain Δ (*annual*%) associated with each forecasting model over four out-of-sample periods. $R_{OS}^2 > 0$, if the conditional forecast outperforms the benchmark. The annual utility gain is interpreted as the annual management fee that an investor would be willing to pay in order to get access to the additional information from the conditional forecast model.

$MSPE_{FOLS} - MSPE_{PLQC} =$	$(MSPE_{FOLS} - MSPE_{FQR}) +$	$(MSPE_{FQR} - MSPE_{PLQC})$
OOS	% of total	% of total
1967.1 - 1990.12	84.3%	15.7%
1991.1 - 2013.12	31.3%	68.7%

 Table 2: Mean Squared Prediction Error (MSPE) Decomposition

The decomposition measures the additional *MSPE* loss of *FOLS* forecasts relative to the *PLQC* forecasts. The first element $(MSPE_{FOLS} - MSPE_{FQR})$ measures the additional loss from *OLS* estimator's lack of robustness to estimation errors, while the second element $(MSPE_{FQR} - MSPE_{PLQC})$ represents the extra loss caused by the presence of partially weak predictors in the population. Note: the *PLQC*, *FQR* and *FOLS* forecasts correspond to models noted as *PLQC*4, *FQR*4 and *FOLS*2 in the paper.

of the *FOLS* forecast resulted from *OLS* estimator's lack of robustness to the estimation errors, while the second element represents the extra loss caused by the presence of partially weak predictors in the population. For the 1967.1-1990.12 subperiod, the contribution of partially weak predictors is much smaller compared to that of estimation errors. This is consistent with the results shown in Figures 1 and 2 and also those in Table 1. In case of strong predictors, most of the loss will be explained by *OLS* estimator's lack of robustness to estimation errors. The situation changes dramatically when weak predictors become a more severe issue during the post-1990 out-of-sample period. As a result, the second element dominates, indicating that most of the forecast accuracy loss is ascribed to the presence of partially weak predictors.

In the next section, we provide more information that explains the benefits of the *PLQC* forecasts.

3.2 Explaining the benefits of the *PLQC* forecasts

In this section, we decompose the mean-square prediction error (*MSPE*) into two parts: the forecast variance and the squared forecast bias. We calculate the *MSPE* of any forecast \hat{r}_{t+1} as $\frac{1}{T^*} \sum_t (r_{t+1} - \hat{r}_{t+1})^2$ and the unconditional forecast variance as $\frac{1}{T^*} \sum_t (\hat{r}_{t+1} - \frac{1}{T^*} \sum_t \hat{r}_{t+1})^2$, where T^* is the total number of out-of-sample forecasts. The squared forecast bias is computed as the difference between *MSPE* and forecast variance (Elliott et al. (2013) and Rapach et al. (2010)).

Figures 3 and 4 depict the relative forecast variance and squared forecast bias of all single-predictor models, *CSR*, *FOLS*, *FQR* and *PLQC* models for two out-of-sample subperiods: 1967.1:1990.12 and 1991.1:2013.12. The relative forecast variance (squared bias) is calculated as the difference between the forecast variance (squared bias) of the *ith* model and the forecast variance (squared bias) of the historical



Figure 3: Scatterplot of forecast variance and squared forecast bias relative to historical average, 1967.1-1990.12

The y-axis and x-axis represent relative forecast variance and squared forecast bias of all single-predictor models, *CSR*, *FOLS*, *FQR* and *PLQC* models, calculated as the difference between the forecast variance (squared bias) of the conditional model and the forecast variance (squared bias) of the *HA*. Each point on the dotted line represents a forecast with the same *MSPE* as the *HA*; points to the right are forecasts outperformed by the *HA*, and points to the left represent forecasts that outperform the *HA*.

average (*HA*). Hence, the value of relative forecast variance (squared bias) for the *HA* is necessarily equal to zero. Each point on the dotted line represents a forecast with the same *MSPE* as the *HA*; points to the right of the line are forecasts outperformed by the *HA*, and points to the left represent forecasts that outperform the *HA*. Finally, both forecast variance and squared forecast bias are measured in the same scale so that it is possible to determine the trade-off between variance and bias of each forecasting model.

Since the *HA* forecast is a simple average of historical equity premium, it will have a very low variance but will be biased. Figure 3 shows that, in the 1967.1-1990.12 subperiod, most of the forecasts based on single-predictor models outperformed the *HA*. Combining this result with the empirical observation that the variances of forecasts based on single-predictor models are not lower than the variance of the *HA*, we conclude that such performance relies almost exclusively on a predictor's ability to lower forecast bias relative to that of *HA*. As a result, a predictor is classified as exhibiting strong predictability if it can



Figure 4: Scatterplot of forecast variance and squared forecast bias relative to historical average, 1991.1-2013.12

The y-axis and x-axis represent relative forecast variance and squared forecast bias of all single-predictor models, *CSR*, *FOLS*, *FQR* and *PLQC* models, calculated as the difference between the forecast variance (squared bias) of the conditional model and the forecast variance (squared bias) of the *HA*. Each point on the dotted line represents a forecast with the same *MSPE* as the *HA*; points to the right are forecasts outperformed by the *HA*, and points to the left represent forecasts that outperform the *HA*.

produce forecasts in which the reduction in bias is greater than the increase in variance, relative to the HA forecast.

The preceding discussion offers an explanation of the results presented in Figure 4. For the subperiod 1991.1-2013.12, almost all single-predictor models are outperformed by the *HA*, suggesting the presence of weak predictors. This weak performance is mainly driven by the substantial increase in the squared biases of such forecasts. We notice that when predictors are strong (in Figure 3), *PLQC* and *FQR* perform equally well. However, when predictors become weak (in Figure 4), the *PLQC* outperforms other forecasting methods.

Overall, the success of the *PLQC* forecast is explained by its ability to substantially reduce the squared forecast bias at the expense of a moderate increase in forecast variance. Additional reduction in the forecast variance of the *PLQC* forecasts can be obtained by increasing the number of quantiles used in the combination, as shown by points $PLQC_2$ and $PLQC_4$ in Figures 3 and 4. The main message is that

	1 2					
$\tau =$	SVAR	BM	LTR	DFY	INFL	E10P
30th	63.30%	0.18%	_	0.18%	79.61%	11.35%
40th	81.56%	_	_	_	100.00%	_
50th	40.78%	3.37%	_	-	88.65%	0.89%
60th	_	_	32.98%	-	89.18%	_
70th	_	20.74%	34.04%	_	_	55.32%

Table 3: Frequency of variables selected over OOS : Jan 1967 \sim Dec 2013

Table 3 presents the frequency with which each predictor is selected over the out-of-sample period (1967.1-2013.12) and across quantiles ($\tau = 0.3, 0.4, 0.5, 0.6, \text{ and } 0.7$)

the forecasting models that yield a sizeable reduction in the forecast bias while keeping variance under control are able to improve forecast accuracy over *HA*. This explains the superior performance of *PLQC* forecasts.

Another analysis that we find interesting is the identification of which predictors are chosen by the ℓ_1 -penalized method across quantiles and over time. This analysis was originally suggested by Pesaran and Timmermann (1995) for the mean function. Table 3 shows the frequency with which each predictor is selected over the out-of-sample period, 1967.1-2013.12, and across the quantiles used to compute the *PLQC* forecast, i.e. $\tau = 0.3, 0.4, 0.5, 0.6$, and 0.7. Recall from section 2 that a predictor is defined to be partially weak if it is useful to forecast some, but not all, quantiles of the equity premium. If it helps forecast all quantiles, it is considered to be strong, whereas if it helps predict no quantile, it is fully weak. Notice that Table 3 reports selection frequency for only 6 predictors, meaning that 9 (out of 15) predictors are fully weak. Thus, the prediction Equation (5) that results from this selection procedure will include at most 6 predictors but these predictors are not equally important due to their different levels of partial weakness. For instance, the selection frequency for *DFY* is no more than 1% at some quantiles. Whereas the predictor *INFL* seems to be strong at almost all quantiles, except $\tau = 0.7$. Failing to account for partially weak predictors results in misspecified prediction equations and, therefore, inaccurate forecasts of equity premium as shown before.

Figure 5 shows in detail how the proposed selection procedure works over time and across quantiles. There are 5 charts, one for each quantile used to compute the *PLQC* forecast. For each chart, we list 15 predictors on the vertical axis. The horizontal axis shows the out-of-sample period. Red dots inside the charts indicate that a predictor was selected to forecast a given quantile of the equity premium at time t. Figure 5 shows that predictor *INFL* is useful for forecasting almost all quantiles until 2010 (with noted



Figure 5: Variables selected by *PLQC* for quantile levels $\tau = 0.3, 0.4, 0.5, 0.6, 0.7$ over OOS 1967.1-2013.12

The 5 charts, one for each quantile used in the *PLQC* forecast, display the selected predictor(s) at each time point t over the out-of-sample period, 1967.1-2013.12.

exceptions at $\tau = 0.7$), but it loses predictability power after that. Other predictors, such as *LTR*, *BM* and *SVAR*, are not important at the beginning of the period but become useful for forecasting after 1985, whereas predictor *E10P* seems to be very useful only for forecasting the two most extreme quantiles $\tau = 0.3$ and $\tau = 0.7$. Thus, by carefully excluding fully weak predictors and identifying the relative importance of partially weak predictors, our forecasting approach can yield much better out-of-sample forecasts, which helps us understand why models that overlook weak predictors are outperformed by the proposed *PLQC* method.

3.3 Robustness analysis: other quantile forecasting models

Meligkotsidou et al. (2014) propose the asymmetric-loss *LASSO* (*AL-LASSO*) model, which estimates the conditional quantile function as a weighted sum of quantiles by using *LASSO* to select the weights, that is:

$$\theta_{t} = \arg\min_{\theta_{t} \in R^{15}} \sum_{t} \rho_{\tau} \left(r_{t+1} - \sum_{i=1}^{15} \theta_{i,t} \widehat{r}_{i,t+1} \left(\tau \right) \right) \text{ s.t. } \sum_{i=1}^{15} \theta_{i,t} = 1; \quad \sum_{i=1}^{15} |\theta_{i,t}| \le \delta_{1}$$
(10)

where $\rho_{\tau}(\cdot)$ is the asymmetric loss, $\hat{r}_{i,t+1}(\tau)$ is the quantile function obtained from a single-predictor quantile model, i.e., $\hat{r}_{i,t+1}(\tau) = \alpha_i(\tau) + \beta_i(\tau) x_{i,t}$ and $x_{i,t} \in X_t = (x_{1,t}, ..., x_{15,t})'$. The parameter δ_1 controls for the level of shrinkage. A solution to problem (10) results in an estimation of the τth conditional quantile of r_{t+1} , $\hat{r}_{t+1}(\tau) = \sum_{i=1}^{15} \hat{\theta}_{i,t} \hat{r}_{i,t+1}(\tau)$. This process is repeated for every $\tau \in (0.3, 0.4, 0.5, 0.6, 0.7)$.

As the first robustness test, we investigate whether our *PLQF*, $f_{t+1,t}^{\tau}$ outperform other single-predictor quantile forecasts $\hat{r}_{i,t+1}(\tau)$, i = 1, ..., 15 and the *AL-LASSO* based on the quantile score (*QS*) function (Manzan (2015)). The *QS* represents a local out-of-sample evaluation of the forecasts in the sense that rather than providing an overall assessment of the distribution, it concentrates on a specific quantile. The higher the *QS*, the better the model does in forecasting a given quantile. It is computed as:

$$QS^{k}(\tau) = \frac{1}{T^{*}} \sum_{t=1}^{T^{*}} (r_{t+1} - \hat{Q}_{t+1,t}^{k}(\tau)) (1 \cdot (r_{t+1} \le \hat{Q}_{t+1,t}^{k}(\tau)) - \tau)$$
(11)

where T^* is the number of out-of-sample forecasts, r_{t+1} is the realized value of equity premium, $\hat{Q}_{t+1,t}^k(\tau)$ represents the quantile forecast at level τ of model k, and indicator function 1.() equals 1 if $r_{t+1} \leq \hat{Q}_{t+1,t}^k(\tau)$; otherwise it equals 0. As a result, quantile scores are always negative. Thus, the larger the QS is, i.e., the closer it is to zero, the better.

Table 4 shows the QS for each single-predictor quantile model $(\hat{r}_{i,t+1}(\tau))$, AL-LASSO $(\hat{r}_{t+1}(\tau))$ and PLQF $(f_{t+1,t}^{\tau})$, over the full out-of-sample period 1967.1-2013.12. We see that AL-LASSO does not perform well because its quantile scores are among the lowest ones for most quantiles τ . On the other hand, PLQF possesses one of the highest quantile scores across the same quantiles τ . Moreover, none of the single-predictor quantile forecasts consistently outperform PLQF across τ . Since accurate quantile forecasts are essential to yield successful point forecasts in the second step, the success of the PLQC point forecast relative to other quantile combination based models is explained by the fact that it averages the most accurate quantile forecasts of equity premium.

No. Model	$QS~(imes 10^{-2})$ across quantile levels $ au$							
$\tau =$	0.3	0.4	0.5	0.6	0.7			
PLQF	-1.499	-1.641	-1.672	-1.615	-1.457			
AL-LASSO	-1.532	-1.652	-1.708	-1.637	-1.464			
DP	-1.532	-1.675	-1.692	-1.626	-1.459			
DY	-1.533	-1.676	-1.692	-1.624	-1.460			
EP	-1.539	-1.677	-1.695	-1.626	-1.461			
DE	-1.568	-1.697	-1.698	-1.625	-1.450			
SVAR	-1.512	-1.660	-1.688	-1.628	-1.449			
BM	-1.526	-1.674	-1.694	-1.632	-1.467			
NTIS	-1.527	-1.672	-1.689	-1.623	-1.449			
TBL	-1.527	-1.658	-1.679	-1.619	-1.462			
LTY	-1.529	-1.664	-1.689	-1.625	-1.467			
LTR	-1.536	-1.678	-1.696	-1.617	-1.448			
TMS	-1.532	-1.668	-1.689	-1.626	-1.462			
DFY	-1.537	-1.673	-1.690	-1.618	-1.446			
DFR	-1.523	-1.670	-1.692	-1.621	-1.454			
INFL	-1.517	-1.648	-1.674	-1.610	-1.445			
E10P	-1.529	-1.673	-1.696	-1.627	-1.466			

 Table 4: Quantile Scores

Table 4 shows the QS for each single-predictor quantile model, AL-LASSO and PLQF models. Quantile scores are always negative. Thus, the larger the QS is, i.e., the closer it is to zero, the better. The quantile scores of AL-LASSO are among lowest ones for most quantiles τ . On the other hand, PLQF possesses one of the highest quantile scores across the same quantiles.

Figure 6 shows the cumulative squared forecast error of the *HA* minus the cumulative squared forecast errors of point forecasts obtained by combining quantile forecasts from *PLQF*, *AL-LASSO* and single-predictor quantile models¹⁹. We additionally report what Meligkotsidou et al. (2014) called robust forecast combination (RFC_1) forecast, which is computed by averaging all the 15 point forecasts obtained from the single-predictor quantile forecasting models.

Figure 6 suggests that the point forecasts obtained from single-predictor quantile models and *AL-LASSO* are still unable to outperform the *HA* consistently over time in terms of their cumulative performance. The RFC_1 hardly outperforms the historical average in any consistent basis of time. This happens because, unlike the *PLQC* forecast, these models are not designed to deal with partially and fully weak predictors across quantiles and over time, and thus are severely affected by misspecification. The failure of the *AL-LASSO* can also be explained by that quantiles are not additive.²⁰ In other words, the *AL-LASSO*

¹⁹For the sake of brevity and without affecting our conclusions, we only use the first weighting scheme to compute these point forecasts. Each single-predictor quantile forecasting model generates one point forecast. Thus, there will be 15 such point forecasts.

²⁰It means that for two random variables X and Y, $Q_{\tau}(X+Y)$ is not necessarily equal to $Q_{\tau}(X) + Q_{\tau}(Y)$.



Figure 6: Cumulative squared prediction error for the benchmark model minus the cumulative squared prediction errors for the *PLQC* 1, AL-LASSO 1, *RFC* 1, and singe-predictor quantile forecasting models, 1967.1-2013.12

A positively sloped curve in each panel indicates that the conditional model outperforms the HA, while the opposite holds for a downward sloping curve. Moreover, if the curve is higher at the end of the period, the conditional model has a lower MSPE than the benchmark over this period. In this figure, the cumulative performance of single-predictor quantile models, AL-LASSO and RFC_1 hardly beat that of the HA consistently over time, as the PLQC forecast does.

method assumes quantile additivity, $\hat{r}_{t+1}(\tau) = \sum_{j=1}^{15} \hat{\theta}_{i,t} \hat{r}_{i,t+1}(\tau)$, which may not hold in practice. The cumulative performance of *PLQC* forecast beats the *HA* over time and shows a clear superiority over other point (*MSPE*) forecasts obtained from a combination of quantile forecasts.

4 Conclusion

This paper studies equity premium forecasting using monthly observations of returns to the S&P 500 from 1926.12 to 2013.12. A common feature of existing models is that they produce inaccurate forecasts due to the presence of weak predictors and estimation errors. We propose a model selection procedure to identify partially and fully weak predictors, and use this information to make optimal *MSPE* forecasts based on

an averaging scheme applied to quantiles. The quantiles combination, as a robust approximation to the conditional mean, avoids accuracy loss caused by estimation errors. The resulting *PLQC* forecasts achieve a middle ground in terms of variance versus bias whereas existing methods reduce forecast variance significantly but are unable to lower bias by a large scale. For this reason, the *PLQC* forecast outperforms the historical average, and other existing forecasting models by statistically and economically meaningful margins.

In the robustness analysis, we consider other quantile forecasting models based on fixed predictors. These models are not designed to deal with partial and fully weak predictors across quantiles and over time. The empirical results show that the quantile forecasts from such models are outperformed by the proposed post-*LASSO* quantile forecast (*PLQF*). Moreover, the point forecasts obtained from the combination of such quantile forecasts are still unable to provide a solution to the original puzzle reported by Goyal and Welch (2008).

In conclusion, equity premium forecasts can be improved if a method minimizes the effect of misspecification caused by weak predictors and estimation errors. Our results support the conclusion that an optimal *MSPE* out-of-sample forecast of the equity premium can be achieved when we integrate *LASSO* estimation and quantile combination into the same framework.

References

- Belloni, A., and Chernozhukov, V. (2011), "*l*1-penalized quantile regression in high-dimensional sparse models," *The Annals of Statistics*, 39, 82-130.
- [2] Campbell, J. Y. (2000), "Asset pricing at the millennium," The Journal of Finance, 55, 1515-1567.
- [3] Campbell, J. Y., and Thompson, S. B. (2008), "Predicting excess stock returns out of sample: Can anything beat the historical average?" *Review of Financial Studies*, 21, 1509-1531.
- [4] Capistrán, C. and Timmermann, A. (2009), "Forecast combination with entry and exit of experts,"*Journal of business and economic statistics*, 27(4), 428-440.
- [5] Christoffersen, P. F., and Diebold, F. X. (1997), "Optimal prediction under asymmetric loss," *Econometric theory*, 13(06), 808-817.

- [6] Clark, T. E., and McCracken, M. W. (2001), "Tests of equal forecast accuracy and encompassing for nested models," *Journal of econometrics*, 105(1), 85-110.
- [7] Clark, T. E., and West, K. D. (2007), "Approximately normal tests for equal predictive accuracy in nested models," *Journal of econometrics*, 138, 291-311.
- [8] Diebold, F. X. and Mariano, R. S. (1995), "Comparing predictive accuracy," *Journal of Business and Economic Statistics*, 13, 253-263.
- [9] Dow, C. H. (1920), "Scientific stock speculation," Magazine of Wall Street.
- [10] Elliott, G., Gargano, A., and Timmermann, A. (2013), "Complete subset regressions," *Journal of Econometrics*, 177(2), 357-373.
- [11] Gaglianone, W., and Lima, L. R. (2012), "Constructing density forecasts from quantile regression," *Journal of Money, Credit and Banking*, 44, 1589-1607.
- [12] Gaglianone, W. P., and Lima, L. R. (2014), "Constructing Optimal Density Forecasts From Point Forecast Combinations," *Journal of Applied Econometrics*, 29(5), 736-757.
- [13] Granger, C. W. (1969), "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica: Journal of the Econometric Society*, 424-438.
- [14] Granger, C. W. and Newbold, P. (1986), "Forecasting time series, 2nd edition," Academic Press.
- [15] Issler, J. V., and Lima, L. R. (2009), "A panel data approach to economic forecasting: The biascorrected average forecast," *Journal of Econometrics*, 152(2), 153-164.
- [16] Judge, G. G., R. C. Hill, W. E. Griffiths, H. Lutkepohl and T.-C. Lee. (1988), "Introduction to the Theory and Practice of Econometrics," New York, Wiley.
- [17] Koenker, R., and Machado, J. A. (1999). "Goodness of fit and related inference processes for quantile regression," *Journal of the american statistical association*, 94(448), 1296-1310.
- [18] Koenker, R. (2005), "Quantile regression," No. 38, Cambridge university press.
- [19] Koenker, R. and Basset, G. (1978), "Regression Quantiles", Econometrica, 46, 33-50.

- [20] Ma, L., and L. Pohlman. (2008), "Return forecasts and optimal portfolio construction: aquantile regression approach." *European Journal of Finance*, 14, 5, 409–425
- [21] Manzan, S. (2015), "Forecasting the distribution of economic variables in a data-rich environment," *Journal of Business & Economic Statistics*, 33(1), 144-164.
- [22] McCracken, M. W. (2007), "Asymptotics for out of sample tests of Granger causality," *Journal of Econometrics*, 140(2), 719-752.
- [23] Meligkotsidoua, L., Panopouloub, E., Vrontosc, I., and Vrontosd, S. (2014), "A Quantile Regression Approach to Equity Premium Prediction." *Journal of Forecasting*, 33, 7, 558-576.
- [24] Neely, C. J., Rapach, D. E., Tu, J., and Zhou, G. (2014), "Forecasting the equity risk premium: the role of technical indicators," *Management Science*, 60, 1772-1791.
- [25] Patton, A. J., and Timmermann, A. (2007), "Properties of optimal forecasts under asymmetric loss and nonlinearity," *Journal of Econometrics*, 140(2), 884-918.
- [26] Patton, A. J., and Timmermann, A. (2007), "Testing forecast optimality under unknown loss," *Journal of the American Statistical Association*, 102(480), 1172-1184.
- [27] Pesaran, M. Hashem, and Allan Timmermann (1995), "Predictability of stock returns: Robustness and economic significance," *The Journal of Finance*, 50.4, 1201-1228.
- [28] Rapach, D., and Zhou, G. (2013), "Forecasting stock returns," *Handbook of Economic Forecasting*, Part A, 328-383.
- [29] Rapach, D. E., Strauss, J. K., & Zhou, G. (2010), "Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy," *The Review of Financial Studies*, 821-862.
- [30] Taylor, J.W. (2007), "Forecasting daily supermarket sales using exponentially weighted quantile regression." European Journal of Operational Research, 178, 154-167.
- [31] Tibshirani, R. (1996), "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society*. Series B, 267-288.
- [32] Timmermann, A. (2006), "Forecast combinations," Handbook of economic forecasting, 1, 135-196.

- [33] Van de Geer, S. A. (2008), "High-dimensional generalized linear models and the lasso," *The Annals of Statistics*, 614-645.
- [34] Weiss, A. A. (1996), "Estimating time series models using the relevant cost function," *Journal of Applied Econometrics*, 11, 539-560.
- [35] Welch, I., and Goyal, A. (2008), "A comprehensive look at the empirical performance of equity premium prediction," *Review of Financial Studies*, 21, 1455-1508.
- [36] West, K. D. (1996), "Asymptotic inference about predictive ability," *Econometrica: Journal of the Econometric Society*, 1067-1084.
- [37] Zhao, Y. (2013), "Forecasting the Stock Return Distribution Using Macro Finance Variables," Job Market Paper.