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

Luiz Lima and Fanning Meng

Journal of Applied Econometrics

Presented by
Harshvardhan
Yu Jiang

Business Analytics and Statistics,
Haslam College of Business,
University of Tennessee

Table of Content

Introduction

- Forecasting stock returns with macroeconomic predictors and valuation ratios is inaccurate due to presence of weak predictors
- Conditional mean forecasts use all predictors while quantile forecasts (FQC) use only strong predictors
- Lima and Meng (2016) propose Lasso based regularization forecasting model for identifying weak and partially weak predictors called PLQC
- PLQC outperforms most benchmark models over the time period 1967 and 2013 as found in the empirical study
- Effect is more pronounced in the presence of partially weak predictors which FQC cannot handle

Three Steps of PLQC Forecasting

Step 1

- Select statistically significant predictors using quantile regression with L1 regularization, Lasso

Step 2

- Rebuild quantile forecasting model with selected predictors to estimate post-penalized regression models at different quantiles, τ

Step 3

- Combine quantile regression point forecasts to obtain PLQC point estimate

Notions and Definitions

- **HA model**: Historical Average Model
- **LASSO**: Least Absolute Shrinkage and Selection Operator
- **PLQC**: Post-LASSO Quantile Combination
- **MSPE**: Mean Square Prediction Error
- **DGP**: Data-generating Process

- **Fully weak predictors**: A given predictor is useful to forecast no quantiles.
- **Partially weak predictors**: A given predictor is useful to forecast some, but not all, quantiles of the equity premium
- **Strong predictors**: A given predictor is useful to forecast all quantiles

Notions and Definitions

+ r_{t+1}: equity premium of the S&P 500 index
+ I_t: information available at time t
+ $F_\eta(0,1)$: some distribution with mean zero and unit variance that does not depend on I_t
+ $X_{t+1,t}$: $X_{t+1,t} \in I_t$, a k×1 vector of covariates available at time t
+ α: $\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_{k+1})'$ is k×1 vectors of parameters, and α_0 is the intercept
+ γ: $\gamma = (\gamma_0, \gamma_1, \ldots, \gamma_{k+1})'$ is k×1 vectors of parameters, and γ_0 is the intercept
+ X_t: $X'_{t+1,t} = X'_t$, a vector of predictors observable at time t
+ $Q_\tau(r_{t+1}|X_t)$: the conditional quantile of r_{t+1} at level $\tau \in (0,1)$
+ $\beta(\tau)$: $\beta(\tau) = \alpha + \gamma F_\eta^{-1}(\tau)$, $F_\eta^{-1}(\tau)$ is the unconditional quantile of η_{t+1}

Methodology

Estimator

Data Generating Process

- $r_{t+1} = X'_{t+1}\alpha + (X'_{t+1}\,\gamma)\eta_{t+1}$
- $\eta_{t+1}|\,I_t \sim iid\ F_\eta(0,1)$

Conditional Quantile and Optimal Forecast

- $\hat{r}_{t+1} = Q_t(r_{t+1}|X_t) = X'_t + X'_t\gamma F_\eta^{-1}(\tau) = X'_t\beta(\tau) = E(r_{t+1}|X_t) + \kappa\_\tau$

Combination of Forecasts

$$\sum_{\tau=\tau_{min}}^{\tau_{max}} \omega_\tau Q_\tau(r_{t+1}|X_t) = E(r_{t+1}|X_t) + \sum_{\tau=\tau_{min}}^{\tau_{max}} \omega_\tau \kappa_\tau$$

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

L1 Penalized Quantile Regression

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

where $\beta_0(\tau) = \alpha_0 + \gamma_0 F_\eta^{-1}(\tau)$ and $\beta_1 = \alpha_1 + \gamma_1 F_\eta^{-1}(\tau)$

Lasso Quantile Regression Estimator (Belloni and Chernozhukov, 2011)

$$\min_{\beta_0,\beta_1} \sum_t \rho_\tau\left(r_{t+1} - \beta_0(\tau) - x'_t\beta_1(\tau)\right) + \lambda \frac{\sqrt{\tau(1-\tau)}}{m} ||\beta_1(\tau)||_{l_1}$$

where ρ_τ denotes a "check" function such that $\rho_\tau(e) = [\tau - 1(e \leq 0)]e$, and $||.||_{l_1}$is L1 norm.

Forecasts

$$f_{t+1}^\tau = \beta_0(\tau) + x_t^{*\prime}\beta(\tau)$$

where x_t^* are predictors selected via Lasso at time t.

Alternative Interpretation

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

$$\begin{aligned} \hat{r}_{t+1} &= \sum_{j=1}^{5} \omega_\tau \beta_0(\tau_j) + \sum_{j=1}^{3} \omega_\tau \beta_1(\tau_j) x_{1,t} + \sum_{j=4}^{5} \omega_\tau \beta_2(\tau_j) x_{2,t} \\ &= \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} \end{aligned}$$

Partial weakness is apparent
Third predictor is fully ruled out
$x_{1,t}$ will produce downwardly biased forecasts
$x_{2,t}$ will produce upwardly biased forecasts
Problem of aggregation bias is avoided because of selective addition

Weights

Equal Weights

+ Assign equal weights for all quantiles

$$PLQC_1 : \frac{1}{3} f^{0.3}_{t+1,t} + \frac{1}{3} f^{0.5}_{t+1,t} + \frac{1}{3} f^{0.7}_{t+1,t}$$

$$PLQC_2 : \frac{1}{5} f^{0.3}_{t+1,t} + \frac{1}{5} f^{0.4}_{t+1,t} + \frac{1}{5} f^{0.5}_{t+1,t} + \frac{1}{5} f^{0.6}_{t+1,t} + \frac{1}{5} f^{0.7}_{t+1,t}$$

OLS-estimated Weights

+ Found from OLS regression of r_{t+1}on $f^{\tau}_{t+1}$for different τ

$$PLQC_3 : r_{t+1} = \sum_{\tau=\tau_1}^{\tau_3} \omega_\tau f^{\tau}_{t+1,t} + \varepsilon_{t+1} \quad \tau \in (0.3; 0.5; 0.7)$$

$$\text{s.t.} \quad \omega_{\tau_1} + \omega_{\tau_2} + \omega_{\tau_3} = 1$$

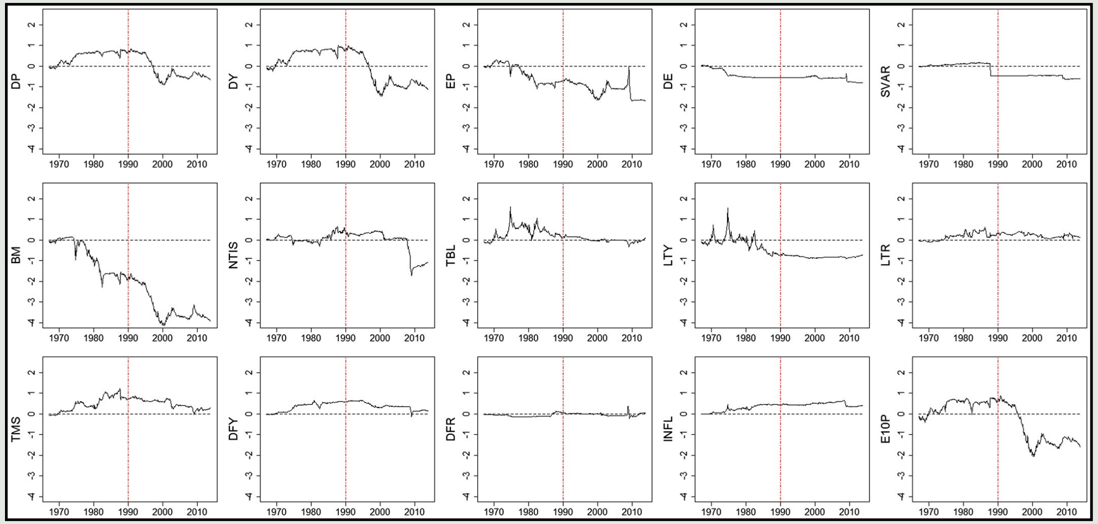
$$PLQC_4 : r_{t+1} = \sum_{\tau=\tau_1}^{\tau_5} \omega_\tau f^{\tau}_{t+1,t} + \varepsilon_{t+1} \quad \tau \in (0.3; 0.4; 0.5; 0.6; 0.7)$$

$$\text{s.t.} \quad \omega_{\tau_1} + \omega_{\tau_2} + \omega_{\tau_3} + \omega_{\tau_4} + \omega_{\tau_5} = 1$$

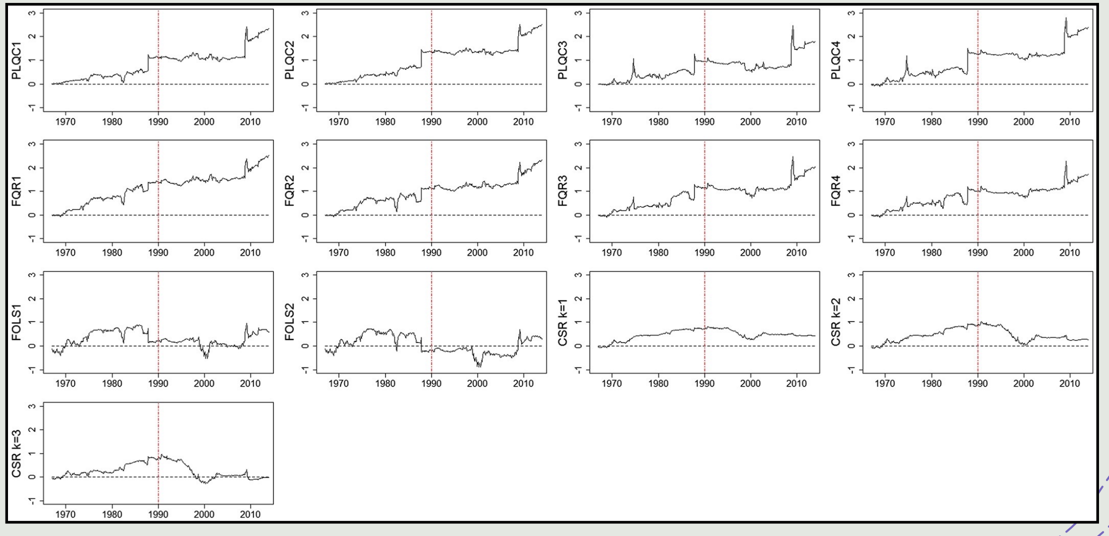
Empirical Results

Model performance compared with benchmark models

Benchmark models perform better than single-predictor models



PLQC Model performs great, especially after 1990



Model performance is judged as cumulative squared prediction error for the benchmark model minus cumulative squared prediction errors. Positive slope indicates that conditional model outperforms the historical average and vice-versa.

Quantile forecasting generates robust forecasts

Strong and Weak Predictors

- Between 1967 and 1990, the single predictor model performs well
- When predictors are strong, we don't need models that are robust to (partially) weak predictors, except for the predictive advantage
- Between 1990 and 2013, there were more weak predictors, and PLQC performs best

Error Decomposition

Table II. Mean squared prediction error (MSPE) decomposition

$MSPE_{FOLS} - MSPE_{PLQC} =$	$(MSPE_{FOLS} - MSPE_{FQR}) +$	$(MSPE_{FQR} - MSPE_{PLQC})$
OOS	% of total	% of total
1967:1–1990:12	84.3%	15.7%
1991:1–2013.12	31.3%	68.7%

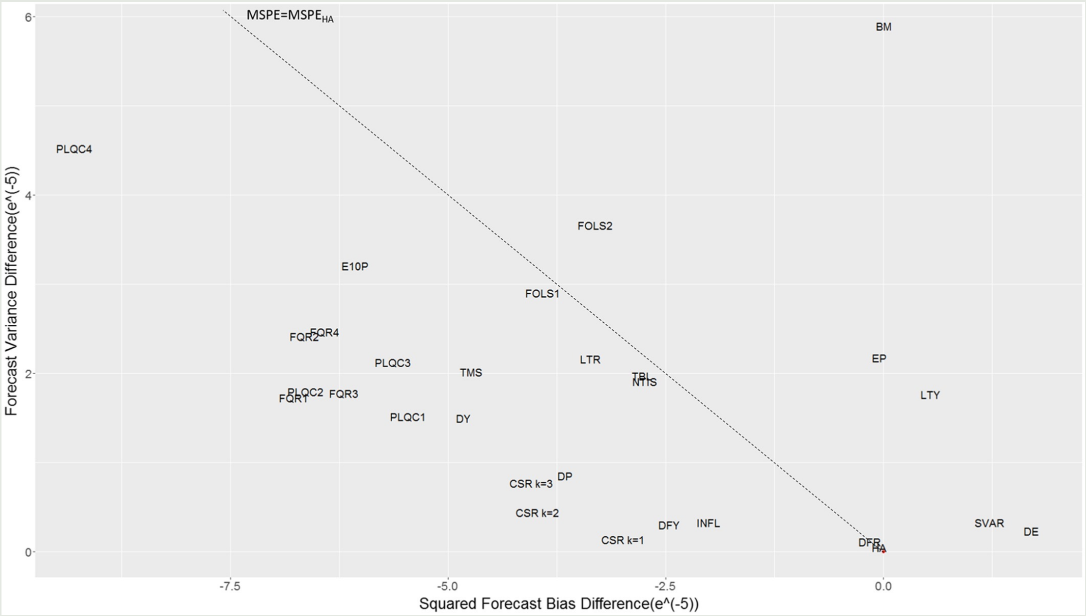
- Quantile regression avoids the effects of **estimation errors** and **aggregation bias**

Bias-variance Decomposition of Error

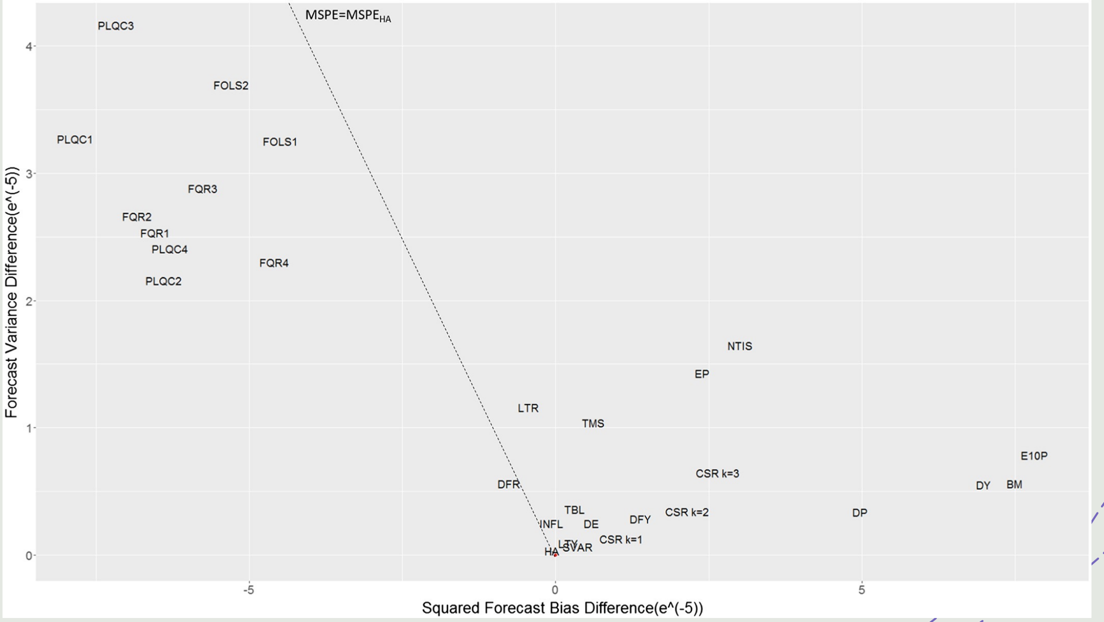
1. MSPE = $\frac{1}{T^*}\sum_t \left(r_{t+1} - \hat{r}_{t+1}\right)^2$
2. Unconditional Forecast Variance = $\frac{1}{T^*}\sum_t \left(\hat{r}_{t+1} - \frac{1}{T^*}\sum_t \hat{r}_{t+1}\right)^2$
3. $Bias^2 = (1) - (2)$

Weak predictors showcase benefits of using PLQC model

1967-90: Single-period forecasts outperform the historical average, where reduction in bias is greater than increase in variance



1991-2013: PLQC and FQR perform equally well until the noted presence of weak predictors



Each point on the dotted line represents a forecast model with the same MSPE as the historical average (HA). Models to the right are outperformed by HA, while models to left outperform HA.

Variables selected by Lasso

Table III. Frequency of variables selected over OOS: January 1967 to December 2013

$\tau =$	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%

Note: The table presents the frequency with which each predictor is selected over the out-of-sample period 1967:1–2013:12 and across quantiles ($\tau = \mathbf{0.3, 0.4, 0.5, 0.6}$, and 0.7).

- There are no strong predictors, though INFL is strongest in the study
- All predictors are partially weak, with DFY being useful only for predicting 0.3 quantile
- SVAR is strong only in
- BM and E10P are only strong in forecasting 0.3, 0.5 and 0.7 quantiles

Robustness Analysis

- Meligkotsidou et al. (2014): the asymmetric-loss LASSO (AL-LASSO) model

$$\theta_t = \arg\min_{\theta_t \in R^{15}} \sum_t \rho_\tau \left(r_{t+1} - \sum_{i=1}^{15} \theta_{i,t} \widehat{r}_{i,t+1}(\tau) \right) \text{ s.t. } \sum_{i=1}^{15} \theta_{i,t} = 1; \quad \sum_{i=1}^{15} |\theta_{i,t}| \leq \delta_1$$

- $\rho_\tau(\cdot)$: the asymmetric loss
- $\widehat{r}_{i,t+1}(\tau)$: the quantile function obtained from a single-predictor quantile model
- δ_1: it controls the level of shrinkage
- A solution to AL-LASSO model is an estimation of the τ^{th} conditional quantile of r_{t+1}, $\hat{r}_{t+1}(\tau) = \sum_{i=1}^{15} \hat{\theta}_{i,t} \hat{r}_{i,t+1}(\tau)$, which is repeated for every $\tau \in (0.3, 0.4, 0.5, 0.6, 0.7)$

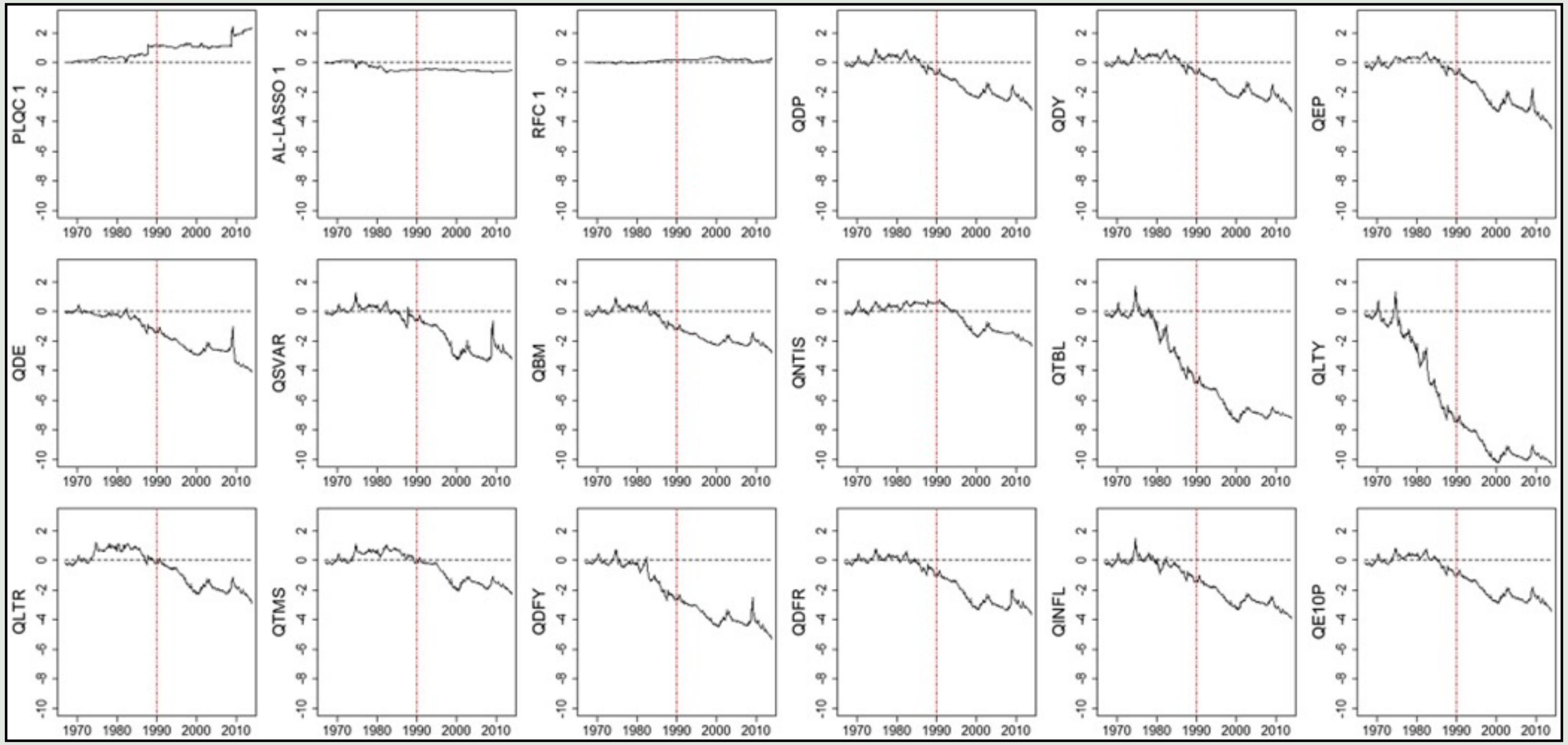
- $QS^k(\tau) = \frac{1}{T^*}\sum_{t=1}^{T^*}\left(r_{t+1} - \hat{Q}_{t+1,t}^k(\tau)\right)\left(1.\left(r_{t+1} \leq \hat{Q}_{t+1,t}^k(\tau)\right) - \tau\right)$
- QS: it represents a local out-of-sample evaluation of the forecasts. The higher the QS, the better the model does in forecasting a given quantile.
- T^*: the number of out-of-sample forecasts.
- r_{t+1}: the realized value of equity premium.
- $\hat{Q}_{t+1,t}^k(\tau)$: it represents the quantile forecast at level τ of model k.
- $1.()$: the indicator function, it equals to 1 if $r_{t+1} \leq \hat{Q}_{t+1,t}^k(\tau)$; otherwise it equals to 0.

Table IV. Quantile scores

No. model	QS ($\times 10^{-2}$) across quantile levels τ				
τ=	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

Note: The table 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.

PLQC models are built to handle partially and fully weak predictors across quantiles and over time



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.

Conclusion

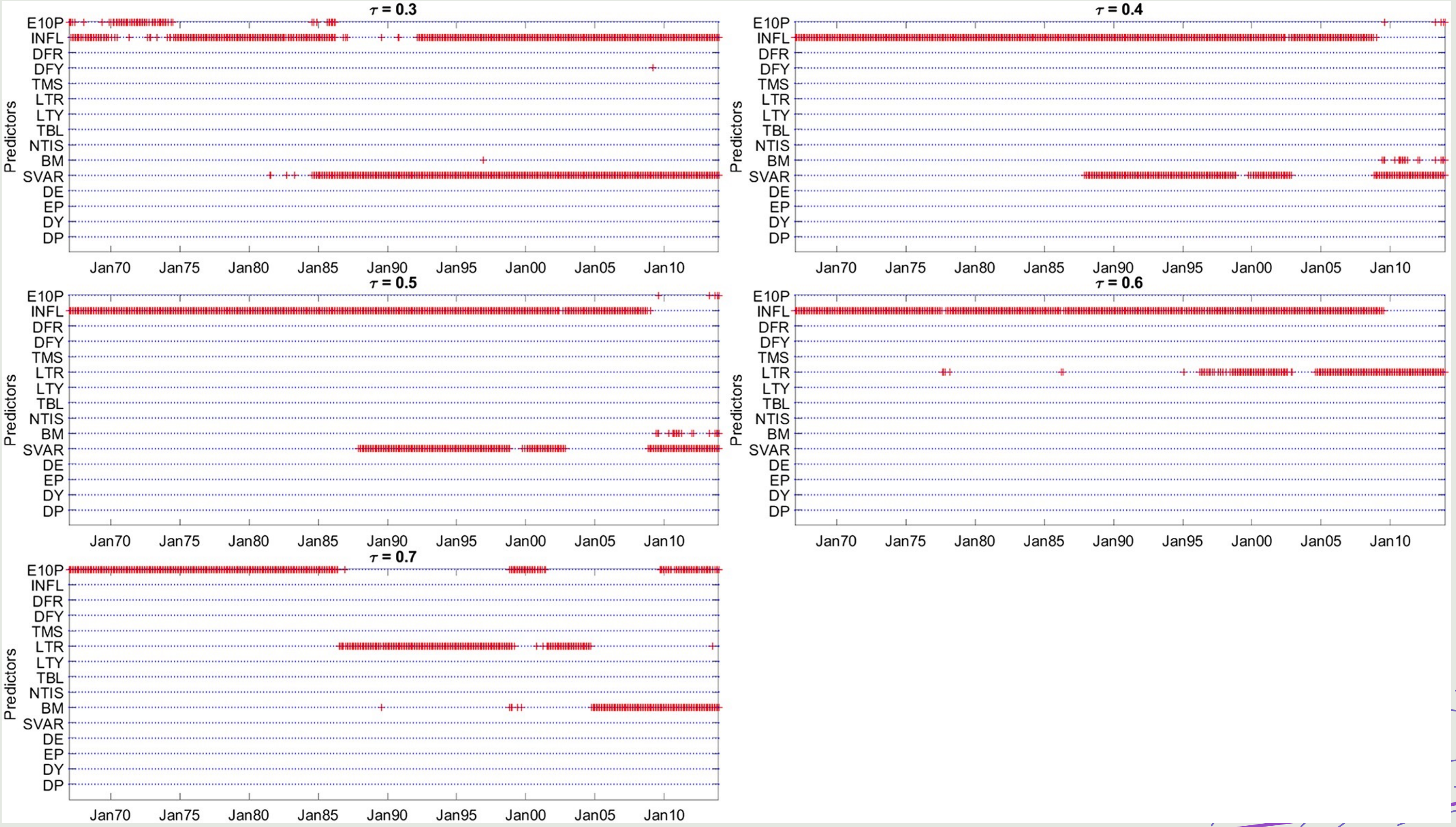
Goal: improve equity premium forecasts.

Method: propose a new model post-LASSO quantile forecast(PLQF) to produce more accurate forecasts by minimizing the negative effects of weak predictors and estimation errors.

Result: PLQF forecast outperforms the historical average(HA) and other methods since it can reduce variance and lower bias by integrating LASSO estimation and quantile combination.

Appendix

Variables selected for different quantiles over 1963-2013 period



Out of sample premium forecasting

Table I. Out-of-sample equity premium forecasting.

	OOS: 1967:1–2013:12				OOS: 1967:1–1990:12				OOS: 1991:1–2013:12				OOS: 2008:1–2013:12			
Model	R^2_{OS}(%)	DM	CW	Δ(annual%)	R^2_{OS}(%)	DM	CW	Δ(annual%)	R^2_{OS}(%)	DM	CW	Δ(annual%)	R^2_{OS}(%)	DM	CW	Δ(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
Complete subset regression 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

Note: This table reports R^2_{OS} 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 and Mariano (1995) test (DM) and the annual utility gain Δ (annual%) associated with each forecasting model over four out-of-sample periods. $R^2_{OS} > 0$, 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.