Online appendix for "Out-of-Sample Return Predictability: a Quantile Combination Approach"

Luiz Renato Lima*[a]
and
Fanning Meng[a]

August 12, 2016

1 Monte Carlo Simulation

To better understand how the *PLQC* forecast works, we consider a Monte Carlo simulation experiment that allows us to study both the absolute forecast performance of the proposed method as well as its performance relative to some alternative methods. The main innovation is that we allow for the existence of both fully and partially weak predictors.

The Monte Carlo simulation is based on the following location-scale model:

$$r_{t+1} = \beta_0 + \sum_i \beta_i x_{i,t} + (\gamma_0 + \sum_i \gamma_i x_{i,t})\eta_{t+1} \tag{1}$$
$$i = 1, 2, \ldots, 6; \quad t = 1, 2, \ldots, 1000$$

where $\beta_0 = 1$, $\forall t$, $\eta_{t+1} \sim N(0, \sigma_\eta^2)$, and $\sigma_\eta = 0.75$. The total number of observations is $T = 1000$, from which the first 999 observations are used to estimate the coefficients and the last one is used to evaluate the one-step ahead forecasts. The number of potential predictors is 6. To illustrate the idea of fully weak predictors, we set $\beta_i = \gamma_i = 0$ for $i = 3, 4, \ldots, 6$ and all t. Hence, predictors $x_3, x_4, \ldots, x_6$ are fully weak because they never enter into the data generating process and therefore cannot predict any quantile of the distribution of r_{t+1} . In order to generate partially weak predictors, we set $\beta_1 = -1.5$ and $\gamma_1 = 5$, if $\eta_{t+1} \leq \phi^{-1}(0.5)$, the 50th percentile of a normal distribution with mean 0 and standard deviation σ_η; otherwise both parameters are set to zero. Similarly, $\beta_2 = 1.5$ and $\gamma_2 = 5$, if $\eta_{t+1} > \phi^{-1}(0.5)$. Hence, predictors x_1 and x_2 are partially weak since they affect some, but not all, quantiles of r_{t+1}. Finally, in order to create some outliers in the observations, we set $\gamma_0 = \sigma_\eta$ if $\eta_{t+1} < 1.96$, otherwise $\gamma_0 = 5\sigma_\eta$.

Notice that the weak predictability problem generated by this Monte Carlo simulation is quite severe since the predictors are either partially or fully weak, but never strong. In doing so, we want to simulate a situation that mimics the empirical problem reported in the next section where predictors of monthly equity premium became extremely weak after 1990s.

*Corresponding author (e-mail: llima@utk.edu).
[a]Department of Economics, University of Tennessee at Knoxville, TN 37996.

Table 1: Monte Carlo simulations

	$\rho = 0$		$\rho = 0.1$		$\rho = 0.25$		$\rho = 0.5$		$\rho = 0.95$	
Model	$R^2(\%)$	p-value	$R^2(\%)$	p-value	$R^2(\%)$	p-value	$R^2(\%)$	p-value	$R^2(\%)$	p-value
x_1	2.513	0.000	1.877	0.000	1.133	0.000	0.353	0.000	-0.120	0.552
x_2	2.945	0.000	2.276	0.000	1.464	0.000	0.566	0.000	-0.115	0.477
x_3	-0.104	0.554	-0.107	0.583	-0.113	0.647	-0.126	0.741	-0.130	0.681
x_4	-0.146	0.887	-0.138	0.845	-0.126	0.770	-0.116	0.667	-0.112	0.533
x_5	-0.077	0.282	-0.077	0.281	-0.082	0.317	-0.092	0.395	-0.114	0.534
x_6	-0.145	0.880	-0.151	0.902	-0.153	0.908	-0.162	0.930	-0.136	0.735
CSR k=1	1.693	0.000	1.307	0.000	0.837	0.000	0.307	0.000	-0.111	0.592
CSR k=2	3.043	0.000	2.427	0.000	1.698	0.000	0.866	0.000	-0.050	0.040
CSR k=3	4.052	0.000	3.310	0.000	2.418	0.000	1.342	0.000	-0.043	0.002
FOLS	5.272	0.000	4.418	0.000	3.333	0.000	1.911	0.000	-0.242	0.000
FQR	5.371	0.000	4.529	0.000	3.444	0.000	2.032	0.000	-0.069	0.000
PLQC	5.532	0.000	4.709	0.000	3.622	0.000	2.195	0.000	0.053	0.000

This table reports R^2_{OS} statistics (in%) and its significance through the p-value of the Clark and West (2007) test associated with each forecasting model over 5 different Monte-Carlo simulation experiments. $\rho_{ij} = (0, 0.1, 0.25, 0.50, 0.95)$ represent the Spearman correlation coefficient of 6 predictors (x_1, ..., x_6), from which the first 2 are partially weak and the others are fully weak.

We generate a 6×1 vector of predictors $X_t = (x_{1t}, ..., x_{6t})'$ where each predictor x_{it} is distributed uniformly over the interval $(0, 1)$ with Spearman correlation coefficient given by $\rho_{ij} = (0, 0.1, 0.25, 0.50, 0.95)$. Recall that the sign of the Spearman correlation indicates the direction of association between x_{it} and x_{jt} for $i \neq j$. If x_{it} tends to increase when x_{jt} increases, then the Spearman correlation coefficient is positive. A Spearman correlation of zero indicates that there is no tendency for x_{it} to either increase or decrease when x_{jt} increases. The Spearman correlation increases in magnitude as x_{it} and x_{jt} get closer to perfect monotone functions of each other. When x_{it} and x_{jt} are perfectly monotonically related, the Spearman correlation coefficient is 1.

Forecasts are computed using the single-predictor equation[1], the complete subset regressions (*CSR*) with $k = 1, 2$ and 3, *PLQC*, *FQR* and *FOLS* methods. The *PLQC* and *FQR* forecasts are obtained by combining 9 different quantile forecasts at levels $\tau = (0.1, 0.2, \ldots, 0.9)$ based on the time-invariant weighting scheme.[2] We report evaluation results, the out-of-sample R^2 and the p-value of the Clark-West test based on 25,000 out-of-sample forecasts (simulations).[3]

Table 1 displays the simulation results. When the Spearman correlation among x-variables is not very high, especially $\rho \leq 0.5$, the separation between partially and fully weak predictors is clear. Consequently, forecasts based on the single-predictor models using the partially weak predictors x_1 and x_2 have statistically significant positive R^2 values. As expected, forecasts based on fully weak predictors, $x_3, x_4, \ldots, x_6$ have a negative but insignificant R^2 value. None of the single-predictor forecasts outperform the combination methods represented by *CSR*, *PLQC*, *FQR* and *FOLS*. Our simulations also show that *FQR* outperforms the non-robust *FOLS* forecast. This confirms that robust estimation of the prediction equation can improve forecast accuracy.

Finally, if the Spearman correlation coefficient is very high, say $\rho = 0.95$, then it will be much more difficult to distinguish between partial and fully weak predictors. Consequently, our simulations show

[1] As shown in the paper, the single-predictor equation is represented by $r_{t+1} = \alpha_i + \beta_i x_{i,t} + \varepsilon_{i,t+1}$.

[2] Results based on other weighting schemes are similar. In our empirical analysis, we did not include lower quantile levels, 0.1 and 0.2, or higher quantile levels, 0.8 and 0.9, as we did here, because of our concern about relative small sample size in the equity premium data.

[3] This is similar to Elliot, Gargano and Timmermann (2013, page 32).

Table 2: Monte Carlo simulations

	$\rho = 0$		$\rho = 0.1$		$\rho = 0.25$		$\rho = 0.5$		$\rho = 0.95$	
Model	$R^2(\%)$	p-value	$R^2(\%)$	p-value	$R^2(\%)$	p-value	$R^2(\%)$	p-value	$R^2(\%)$	p-value
x_1	-0.108	0.651	-0.108	0.651	-0.108	0.651	-0.108	0.651	-0.108	0.651
x_2	-0.098	0.550	-0.098	0.548	-0.109	0.680	-0.120	0.779	-0.121	0.773
x_3	-0.067	0.250	-0.065	0.230	-0.094	0.513	-0.122	0.791	-0.134	0.869
x_4	-0.059	0.193	-0.055	0.162	-0.051	0.133	-0.072	0.290	-0.121	0.780
x_5	-0.076	0.298	-0.074	0.278	-0.091	0.441	-0.111	0.649	-0.128	0.808
x_6	-0.095	0.509	-0.097	0.528	-0.111	0.657	-0.130	0.822	-0.133	0.853
CSR k=1	-0.006	0.254	-0.004	0.230	-0.021	0.504	-0.052	0.772	-0.117	0.805
CSR k=2	-0.043	0.261	-0.041	0.243	-0.064	0.420	-0.094	0.603	-0.123	0.634
CSR k=3	-0.113	0.268	-0.111	0.255	-0.132	0.364	-0.152	0.465	-0.164	0.474
FOLS	-0.429	0.484	-0.426	0.472	-0.442	0.556	-0.460	0.635	-0.415	0.429
FQR	-0.140	0.065	-0.141	0.068	-0.149	0.090	-0.157	0.120	-0.122	0.038
PLQC	-0.045	0.044	-0.049	0.054	-0.048	0.062	-0.039	0.057	-0.028	0.058

This table reports R^2_{OS} statistics (in%) and its significance through the p-value of the Clark and West (2007) test associated with each forecasting model over 5 different Monte-Carlo simulation experiments. $\rho_{ij} = (0, 0.1, 0.25, 0.50, 0.95)$ represent the Spearman correlation coefficient of 6 fully weak predictors $(x_1, ..., x_6)$.

that all forecasting methods deteriorate substantially. The main message here is that weak predictors can be very harmful to a forecasting model but this problem can be minimized if an adequate forecasting method is employed. The proposed *PLQC* forecast performs very well because it is robust to estimation errors, avoids fully weak predictors, and accounts for the relative contribution of partially weak predictors to forecasting. Our Monte Carlo simulations suggest that this is the best approach to deal with this form of weak predictability.

In addition to the above design, we also consider two extreme cases in which all predictors are fully weak or partially weak. First, when all predictors are fully weak, the data generating process is simplified to:

$$r_{t+1} = \beta_0 + \gamma_0 \eta_{t+1} \quad t = 1, 2, \ldots, 1000 \tag{2}$$

where $\beta_0 = 1$, $\forall t$, $\eta_{t+1} \sim N(0, \sigma^2_\eta)$, and $\sigma_\eta = 0.75$. As before, we set $\gamma_0 = \sigma_\eta$ if $\eta_{t+1} < 1.96$, otherwise $\gamma_0 = 5\sigma_\eta$.

The results are shown in Table 2. All conditional models have negative $R^2(\%)$ across 5 different correlation parameters ρ. Overall, it is hard to find a model that can outperform the historical average model when the real data generating process does not have any strong or partially weak predictor.

The other extreme case considers a situation when all predictors are partially weak. Specifically, based on the DGP (1), we set $\beta_1 = -1.5$ and $\gamma_1 = 5$, if $\eta_{t+1} \leq \phi^{-1}(0.3)$. Similarly, $\beta_2 = -1.5$ and $\gamma_2 = 5$, if $\eta_{t+1} \leq \phi^{-1}(0.4)$; $\beta_3 = -1.5$ and $\gamma_3 = 5$, if $\eta_{t+1} \leq \phi^{-1}(0.5)$; $\beta_4 = -1.5$ and $\gamma_4 = 5$, if $\eta_{t+1} > \phi^{-1}(0.5)$; $\beta_5 = 1.5$ and $\gamma_5 = 5$, if $\eta_{t+1} \geq \phi^{-1}(0.6)$; $\beta_6 = 1.5$ and $\gamma_6 = 5$, if $\eta_{t+1} \geq \phi^{-1}(0.7)$. Hence, predictors $x_1, x_2, \ldots, x_6$ are all partially weak since they affect some, but not all, quantiles of r_{t+1}. All the rest parameters are the same as before.

Table 3 displays the results. When the Spearman correlation coefficient is relative small, especially lower than or equal to 0.25, the *PLQC* does best among all. This result suggests that if the correlation coefficient is high, and predictors are perfectly monotonically related, then *LASSO* does not choose the right predictor at each quantile, leading to a misspecified prediction equation. In general, none of the forecasting models perform quite well for the case with highly correlated partially weak predictors.

Table 3: Monte Carlo simulations

	$\rho = 0$		$\rho = 0.1$		$\rho = 0.25$		$\rho = 0.5$		$\rho = 0.95$	
Model	$R^2(\%)$	p-value	$R^2(\%)$	p-value	$R^2(\%)$	p-value	$R^2(\%)$	p-value	$R^2(\%)$	p-value
x_1	0.182	0.000	0.049	0.000	-0.035	0.002	-0.086	0.153	-0.105	0.332
x_2	0.252	0.000	0.064	0.000	-0.067	0.003	-0.155	0.532	-0.194	0.861
x_3	0.423	0.000	0.174	0.000	0.003	0.000	-0.113	0.190	-0.133	0.427
x_4	0.684	0.000	0.417	0.000	0.220	0.000	0.057	0.000	-0.003	0.003
x_5	0.521	0.000	0.331	0.000	0.187	0.000	0.060	0.000	-0.009	0.005
x_6	0.290	0.000	0.145	0.000	0.032	0.000	-0.062	0.065	-0.101	0.288
CSR k=1	0.849	0.000	0.500	0.000	0.253	0.000	0.056	0.000	-0.020	0.089
CSR k=2	1.514	0.000	0.913	0.000	0.495	0.000	0.178	0.000	0.057	0.000
CSR k=3	1.996	0.000	1.219	0.000	0.679	0.000	0.264	0.000	0.101	0.000
FOLS	2.342	0.000	1.328	0.000	0.572	0.000	-0.011	0.000	-0.195	0.000
FQR	2.344	0.000	1.373	0.000	0.625	0.000	0.043	0.000	-0.156	0.000
PLQC	2.525	0.000	1.496	0.000	0.678	0.000	0.132	0.000	-0.012	0.000

This table reports R^2_{OS} statistics (in%) and its significance through the p-value of the Clark and West (2007) test associated with each forecasting model over 5 different Monte-Carlo simulation experiments. $\rho_{ij} = (0, 0.1, 0.25, 0.50, 0.95)$ represent the Spearman correlation coefficient of 6 partially weak predictors ($x_1, ..., x_6$).

2 Measure of consistence in forecasting performance

In this section, we focus on the evaluation of *PLQC* forecasts' performance consistence compared to other conditional forecasts, as a supplement to the analysis of cumulative squared prediction errors (Figure 2 in the paper). First, we calculate the frequency of months (within the whole out-of-sample period 1967.1-2013.12), in which the squared prediction errors of the benchmark model are larger than those of a conditional forecast. Table 4 column 2 displays the percentage of months that a conditional model outperforms the *HA* in terms of squared prediction errors. Overall, *PLQC* forecasts have higher frequency than most single-predictor models, *CSR*, *FOLS* and *FQR* models.

Based on this frequency measure, we form a very basic impression about the consistence of the model's performance. However, this information is quite limited about the size of the squared prediction errors' gap between the benchmark and the conditional forecast. A positive gap indicates that the conditional forecast does better producing smaller squared prediction errors than the benchmark, while the opposite is true for negative gaps. To have a more comprehensive understanding regarding the consistence of the model's performance, we draw histograms of the squared prediction errors' gap. As shown in Figures 1 and 2, most of the single predictor models, the *CSR* models[4] and the *FOLS* models are skewed to the left, with more and/or larger negative gaps than positive ones. The distributions of *FQR3* and *FQR4* are roughly symmetric. In contrast, the distributions of *FQR1*, *FQR2* and all *PLQC* forecasts are skewed to the right, especially for those of *PLQC1* and *PLQC2*. Not only the *PLQC* forecasts have more positive gaps relative to negative ones, but also the upper tails of their distributions extend much further than those of the single-predictor and *CSR* forecasts. All of these help explain the better cumulative performance of the *PLQC* forecasts compared to other forecasting models as seen in Figures 1 and 2 of the paper.

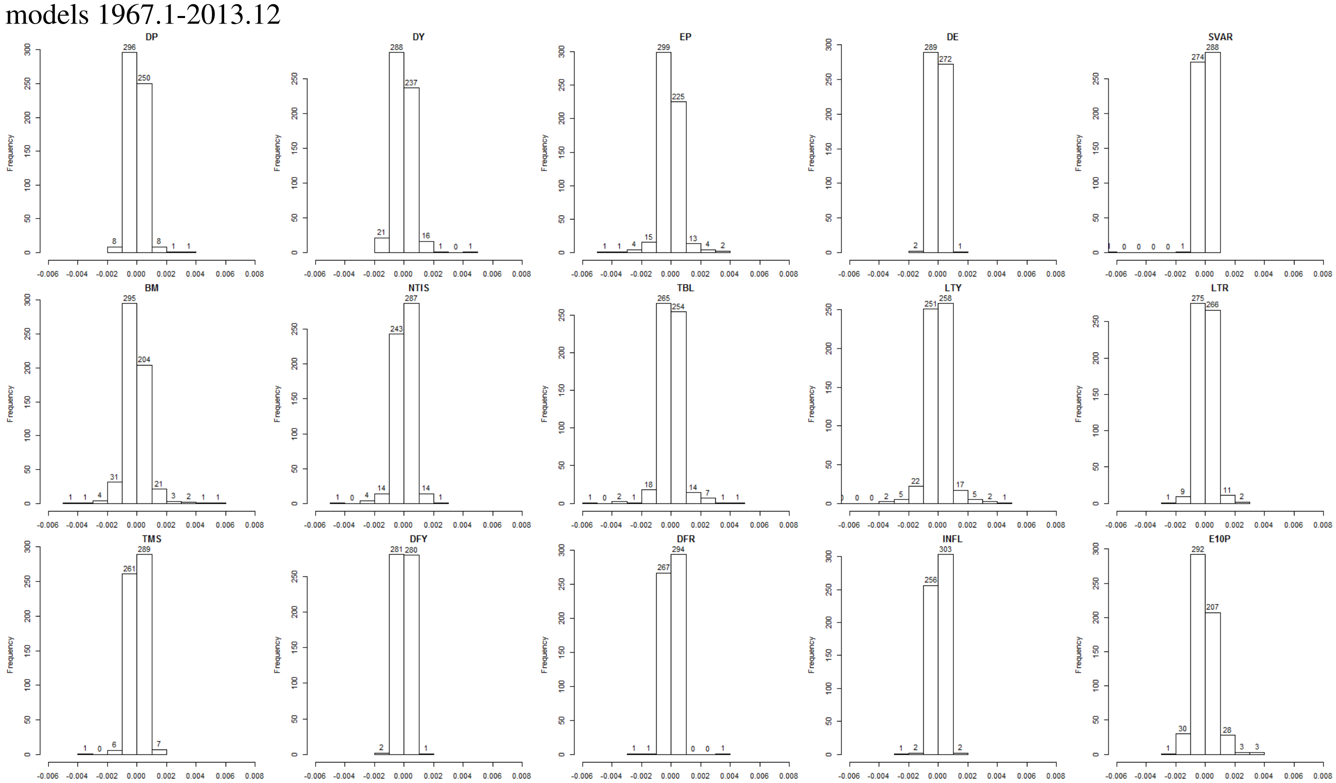
[4] To save space, we only include CSR k=1 and CSR k=2 in this figure. But CSR k=3 is similar to CSR k=2.

Table 4: Frequency of positive squared prediction errors' gaps

Model	Freq.
Single Predictor Model Forecasts	
DP	46.10
DY	45.21
EP	43.26
DE	48.40
SVAR	51.06
BM	41.13
NTIS	53.55
TBL	49.11
LTY	50.18
LTR	49.47
TMS	52.48
DFY	49.82
DFR	52.30
INFL	54.08
E10P	42.73
Complete Subset Regression Forecasts	
CSR k=1	47.52
CSR k=2	46.99
CSR k=3	45.57
Forecasts based on *LASSO*-Quantile Selection	
*FOLS*1	48.94
*FOLS*2	48.23
*FQR*1	54.61
*FQR*2	53.55
*FQR*3	50.35
*FQR*4	49.29
*PLQC*1	54.61
*PLQC*2	54.43
*PLQC*3	52.48
*PLQC*4	51.60

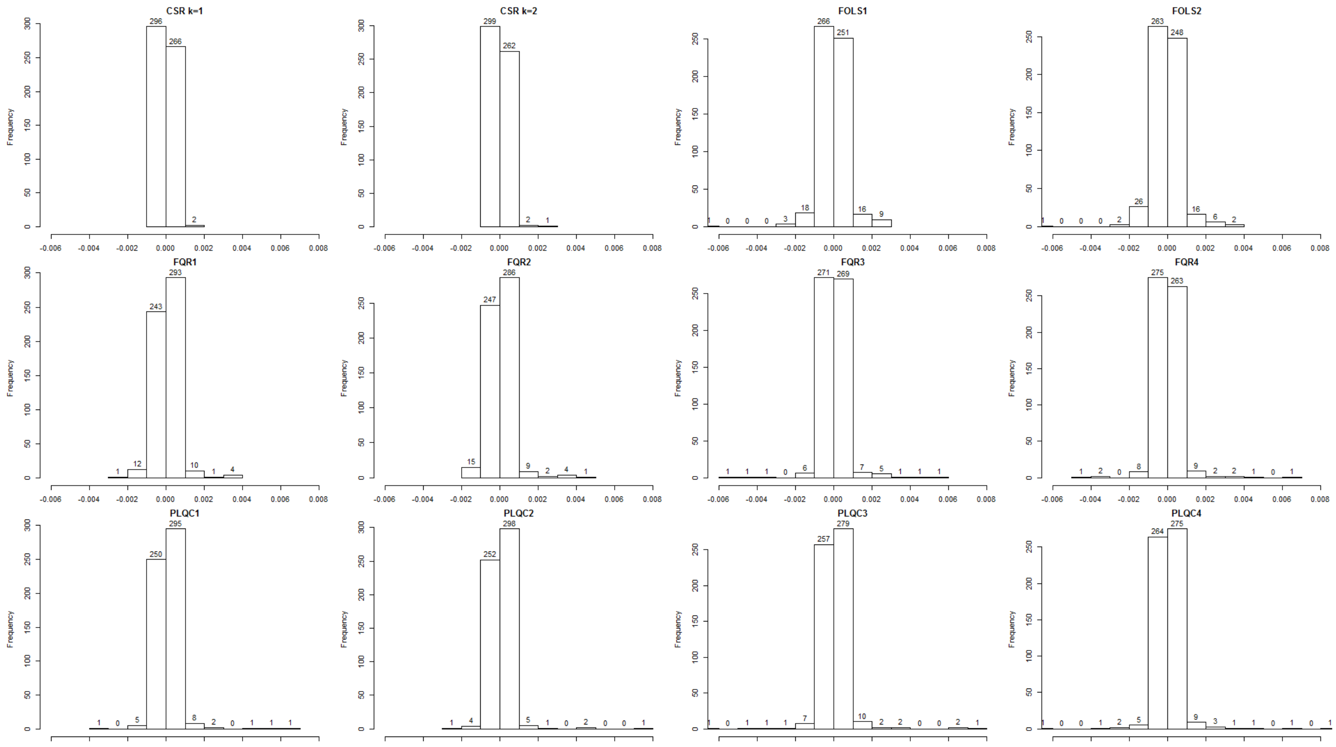
Column 2 illustrates the frequency of months (in%) in which the squared prediction errors of the benchmark are larger than those of a conditional forecast within the period (1967.1-2013.12).

Figure 1: Frequency of the squared prediction errors' gap between the benchmark model and the single-predictor models 1967.1-2013.12



The squared prediction errors' gap is calculated as the squared prediction error of the *HA* forecast minus that of the *PLQC* or *FQR* forecast. The figure shows the frequency of the errors' gap for each model.

Figure 2: Frequency of the squared prediction errors' gap between the benchmark model and the *CSR*, *FOLS*, *FQR* and *PLQC* models 1967.1-2013.12



The squared prediction errors' gap is calculated as the squared prediction error of the *HA* forecast minus that of the *CSR*, *FOLS*, *FQR* and *PLQC* forecasts. The figure shows the frequency of the errors' gap for each model.

3 Data description

In this section, we provide a complete description about our database. The data come from Amit Goyal's webpage (http://www.hec.unil.ch/agoyal/), which includes monthly observations of the returns on the *S&P* 500 index, the risk-free rate and of the following 15 predictors from December 1926 to December 2013:

· *Dividend-price ratio (log)*, *DP*: 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 one-year moving sum.

· *Dividend yield (log)*, *DY*: Difference between the log of dividends and the log of lagged stock prices.

· *Earnings-price ratio (log), EP*: Difference between the log of earnings on the *S&P* 500 index and the log of stock prices, where earnings are measured using one-year moving sum.

· *Dividend-payout ratio (log), DE*: 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, BM*: Ratio of book value to market value for the Dow Jones Industrial Average.

· *Net equity expansion, NTIS*: Ratio of the twelve-month moving sums of net issues by NYSE-listed stocks to total end-of-year market capitalization of NYSE stocks.

· *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.

· *Inflation*, *INFL*: Calculated from the CPI (all urban consumers). Since inflation at time t is only released at time $t+1$, we assumed adaptive expectations about future inflation. In other words, we predicted equity premium at time $t+1$ by using inflation rate from $t-1$ (released at time t) as the best prediction of the inflation rate at time t. Our results point out that adaptive expectations of inflation can play an important role in equity premium prediction.

· *Moving average of earning-price ratio*, *E10P*: ten-year moving average of earnings-price ratio.

4 Economic evaluation: utility gain

In the equity premium predictability literature, the percentage values of R^2_{OS} are typically small, but this does not mean that their economic values are insignificant. Indeed, as argued by Campbell and Thompson (2008), even a very small positive R^2_{OS}, such as 0.5% for monthly data or 1% for quarterly data, can still signal an economically meaningful degree of equity premium predictability in terms of increased annual portfolio returns for a mean-variance investor. To estimate the economic values of 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 historical average benchmark model.

In detail, assume a risk-averse investor who has a mean-variance utility function and considers how to optimally allocate the total wealth between a risky equity and a risk-free asset at time t based on current risk-free rate r^f_{t+1} and one-period ahead forecast of the equity premium, $\hat{r}_{t,t+1}$. Thus, the one-period ahead forecast of return on the risky equity is $\hat{R}_{t,t+1} = \hat{r}_{t,t+1} + r^f_{t+1}$. The weight allocated to the risky equity is calculated as $\omega_t = \frac{1}{\gamma}\frac{\hat{R}_{t,t+1}}{\hat{\sigma}^2_{t+1}}$, where γ is the risk-aversion parameter[5] and $\hat{\sigma}^2_{t+1}$ is the estimated variance of equity assets [6]. The more risk averse an investor is, the lower the weight on the risky asset. Also, the more volatile the equity return, the lower the weight on risky asset. In addition, we impose that $\omega_t \in (0, 1.5)$ to ensure no short sales or over leverage (Rapach and Zhou (2013)).

The realized portfolio return at time $t+1$ is $R^p_{t+1} = \omega_t R_{t+1} + (1-\omega_t) r^f_{t+1}$. Over T^* out-of-sample periods, an investor's utility (or certainty equivalent return) from this portfolio allocation can be calculated as:

$$U = \hat{\mu}_p - \frac{1}{2}\gamma\hat{\sigma}^2_p \tag{3}$$

where $\hat{\mu}_p = \frac{1}{T^*}\sum_t R^p_t$ and $\hat{\sigma}^2_p = Var(R^p) = \frac{1}{T^*}\sum_t (R^p_t - \hat{\mu}_p)^2$. The utility gain is the difference between utility derived based on a conditional forecast model and that from a historical average model. To facilitate the interpretation, we multiply the utility gains by 1200, which gives us the annual management fee that an investor would be willing to pay in order to get access to the additional information from that conditional forecast model. This same approach has also been used by Campbell and Thompson (2008), Rapach et al. (2010) and others.

4.1 Robustness analysis: a common factor and "kitchen sink" model

In this section, we did the robustness analysis on a common-factor model for the mean and quantile functions and the "kitchen-sink" model.

In the absence of notion about $E(r_{t+1}|X_t)$ or the underlying DGP, Capistrán and Timmermann (2009) propose a common factor model that is sufficiently rich to cover a variety of empirically relevant scenarios. Accordingly, let r_{t+1} and the individual forecast, $f^i_{t+1,t}$, be driven by the following common factor model:

$$\begin{aligned} r_{t+1} &= \mu_r + \beta_{rH} H_{t+1,t} + \eta_{t+1}, \eta_{t+1} \sim N\left(0, \sigma^2_\eta\right) \\ f^i_{t+1,t} &= \mu_i + \beta_{iH} H_{t+1,t} + \varepsilon_{i,t+1}, \varepsilon_{i,t+1} \sim N\left(0, \sigma^2_i\right) \end{aligned} \tag{4}$$

where $H_{t+1,t}$ is a common factor predictable at time t. $E(r_{t+1}|X_t) = \mu_r + \beta_{rH} H_{t+1,t}$. Assuming $H_{t+1,t} = H_t$, we apply model (4) to estimate the conditional mean $E(r_{t+1}|X_t)$ directly (Neely et al. (2014)). More specifically, we assume that there is only one factor estimated by principal component and labeled *PC*. We also estimate the quantile function $Q_\tau(r_{t+1}|X_t) = \mu_\tau + \beta_\tau H_t$ using the standard quantile regression estimator (Zhao (2013)) and apply the same weighting schemes as described in section 2 of the paper to combine $Q_\tau(r_{t+1}|X_t)$. We label these new quantile forecast combinations as QPC_1 and QPC_2 with fixed-weighting schemes and as QPC_3 and QPC_4 with time-variant weighting schemes.

Following Goyal and Welch (2008) and Rapach et al. (2010), we additionally consider a "kitchen sink" (*KS*) model that includes all 15 economic variables into a multiple predictive regression model.

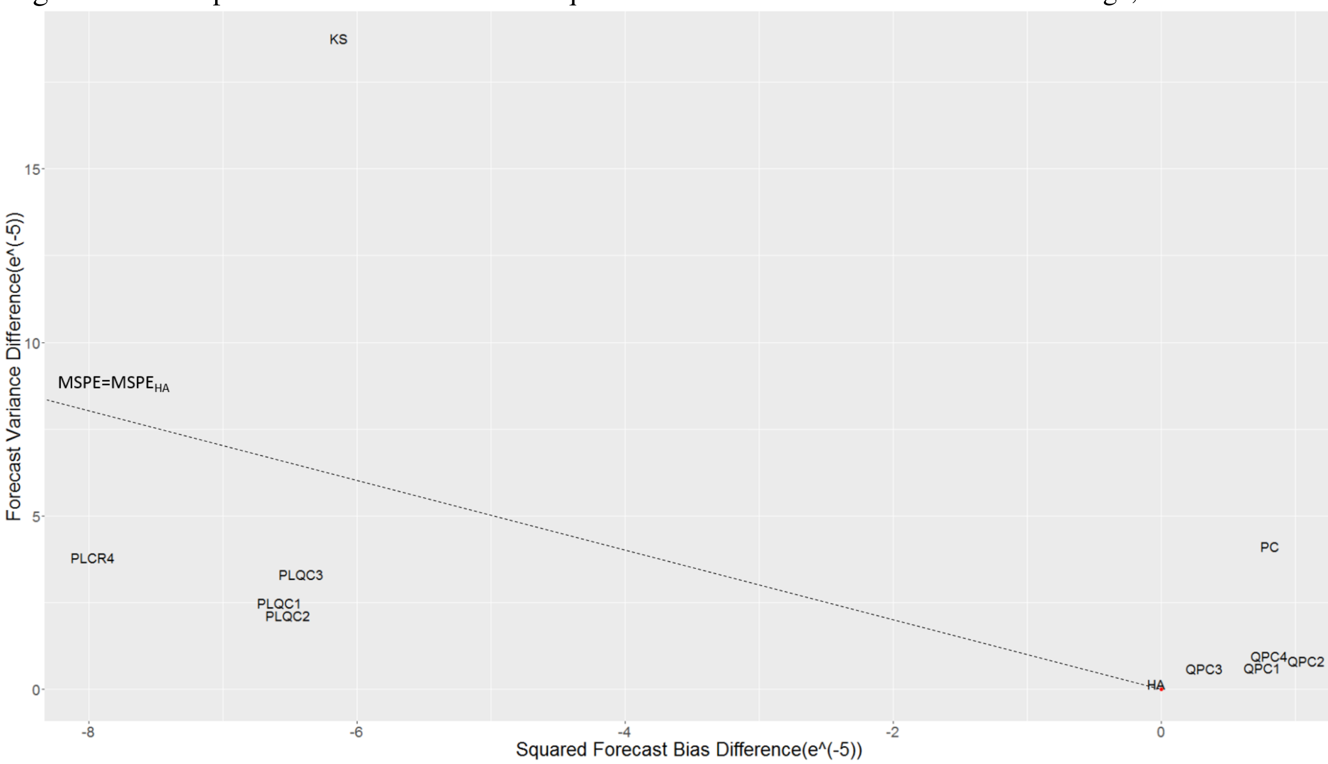
Our results are displayed in Figure 3. In addition to a higher variance, the *PC* forecast is even more biased than *HA*. Thus, it is easily outperformed by the benchmark. Similarly, the *QPC* forecasts are also outperformed by the *HA*. As for the *KS* model, although it does reduce bias substantially, it does so at the expense of a large increase in forecast variance.

In sum, the empirical results show that: (i) the low bias of the "kitchen-sink" model is obtained at the expense of a very high forecast variance in a way that does not reduce *MSPE* compared to the historical

[5] In order to ensure a moderate risk averse preference, we follow the literature by setting γ equal to 3.

[6] Following Campbell and Thompson (2008), we use a 10-year moving window to estimate the variance of equity returns.

Figure 3: Scatterplot of forecast variance and squared forecast bias relative to historical average, 1967.1-2013.12



The y-axis and x-axis represent relative forecast variance and squared forecast bias of *PLQC*, *QPC*, *PC* and *KS* 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 forecast; (ii) The quantile forecast combination based on a common factor quantile model cannot outperform the *HA* benchmark, supporting *LASSO* as an efficient model selection tool for forecasting equity premium.

5 Proof for Proposition 1

In this section, we provide detailed proof corresponding to proposition 1 in the paper. Recall that the data generating process (*DGP*) is defined as

$$\begin{aligned} r_{t+1} &= X'_{t+1,t}\alpha + \left(X'_{t+1,t}\gamma\right)\eta_{t+1}, \\ \eta_{t+1}|I_t &\sim i.i.d.\ F_\eta\left(0,1\right), \end{aligned} \tag{5}$$

Proof. The proof is similar to the one shown by Granger (1969), Christoffersen and Diebold (1997) and Patton and Timmermann (2007) in the first part of their Proposition 2. Thus, by homogeneity of the loss function and *DGP* (5) we have that the optimal forecast is given by[7]

$$\begin{aligned} \widehat{r}_{t+1,t} &= \underset{\widehat{r}}{\operatorname{argmin}} \int L(r-\widehat{r})dF_{t+1,t}(r) \\ & \underset{\widehat{r}}{\operatorname{argmin}} \int \left[g\left(\frac{1}{X'_{t+1,t}\gamma}\right)\right]^{-1} L\left(\frac{1}{X'_{t+1,t}\gamma}\left(r-\widehat{r}\right)\right)dF_{t+1,t}(r) \\ &= \underset{\widehat{r}}{\operatorname{argmin}} \int L\left(\frac{1}{\left(X'_{t+1,t}\gamma\right)}\left(r-\widehat{r}\right)\right)dF_{t+1,t}(r) \\ &= \underset{\widehat{r}}{\operatorname{argmin}} \int L\left(\frac{1}{\left(X'_{t+1,t}\gamma\right)}\left(X'_t\alpha+\left(X'_t\gamma\right)\eta_{t+1}-\widehat{r}\right)\right)dF_{t+1,t}(r). \end{aligned}$$

Let us represent a forecast by $X'_t\alpha + \left(X'_t\gamma\right)\widehat{\delta}_{t+1,t}$. In this way, it follows that:

$$\begin{aligned} \widehat{r}_{t+1,t} &= X'_t\alpha + \left(X'_t\gamma\right)\cdot \underset{\widehat{\delta}}{\operatorname{argmin}} \int L\left(\frac{1}{\left(X'_t\gamma\right)}\left(X'_t\alpha\right.\right. \\ &\quad \left.\left. + \left(X'_t\gamma\right)\eta_{t+1} - X'_t\alpha - \left(X'_t\gamma\right)\widehat{\delta}\right)\right)dF_\eta(\eta) \\ &= X'_t\alpha + \left(X'_t\gamma\right)\cdot \underset{\widehat{\delta}}{\operatorname{argmin}} \int L\left(\eta_{t+1}-\widehat{\delta}\right)dF_\eta(\eta) \\ &= X'_t\alpha + \kappa_\tau \\ &= E\left(r_{t+1}|X_t\right) + \kappa_\tau, \end{aligned}$$

where $\kappa_\tau = \left(X'_t\gamma\right)\delta^*$ and $\delta^* = \operatorname{argmin}_{\widehat{\delta}} \int L\left(\eta_{t+1}-\widehat{\delta}\right)dF_\eta(\eta)$.

Given this optimality result, it is now important to show the relationship between δ^* and the quantiles of η_{t+1}. Let $F_{t+1,t}$ be the conditional distribution of r_{t+1}. Thus,

[7] We omit some time subscripts in the proof to save space.

$$
\begin{aligned}
F_{t+1,t}\left(\widehat{r}_{t+1}\right) &= \Pr\left(r_{t+1} < \widehat{r}_{t+1} | X_t\right) \\
&= \Pr\begin{pmatrix} r_{t+1} = X_t'\alpha + \left(X_t'\gamma\right)\eta_{t+1} < \\ < X_t'\alpha + \left(X_t'\gamma\right)\delta^* | X_t \end{pmatrix} \\
&= \Pr\left(\eta_{t+1} < \delta^* | X_t\right) = F_\eta\left(\delta^*\right) = \tau \in (0,1)
\end{aligned} \tag{6}
$$

So $\delta^* = F_\eta^{-1}(\tau)$, which corresponds to the τth quantile of η_{t+1}. ■

As a corollary of the above result, we have that the optimal forecast should be equal to

$$
\begin{aligned}
\widehat{r}_{t+1} &= Q_\tau\left(r_{t+1}|X_t\right), \text{for some } \tau \in (0;1) \\
&= E\left(r_{t+1}|X_t\right) + \kappa_\tau
\end{aligned} \tag{7}
$$

If L is the *MSPE* loss function then $\delta^* = 0$ and the optimal forecast $\widehat{r}_{t+1,t} = E\left(r_{t+1}|X_t\right)$. Therefore, κ_τ will capture deviations from the standard *MSPE* loss function. For instance, if L corresponds to the mean absolute error (MAE) loss, then $\delta^* = F_\eta^{-1}(0.5) = median\,(\eta)$ and $\widehat{r}_{t+1} = median\,(r_{t+1}|X_t)$. Finally, if L is the asymmetric lin-lin loss, then $\delta^* = F_\eta^{-1}(\tau)$ and $\widehat{r}_{t+1} = Q_\tau\left(r_{t+1}|X_t\right)$ for $\tau \neq 0.5$.

6 Replication with stock return

In this section, we demonstrate the same empirical analysis with stock return, instead of equity premium, as the forecasting variable. Generally speaking, the results are quite similar to what we see with equity premium in the paper.

In Figures 4 and 5, we present time series plots of the differences between the cumulative squared prediction error for the benchmark forecast and that of each conditional forecast as Figures 1 and 2 shown in the paper.

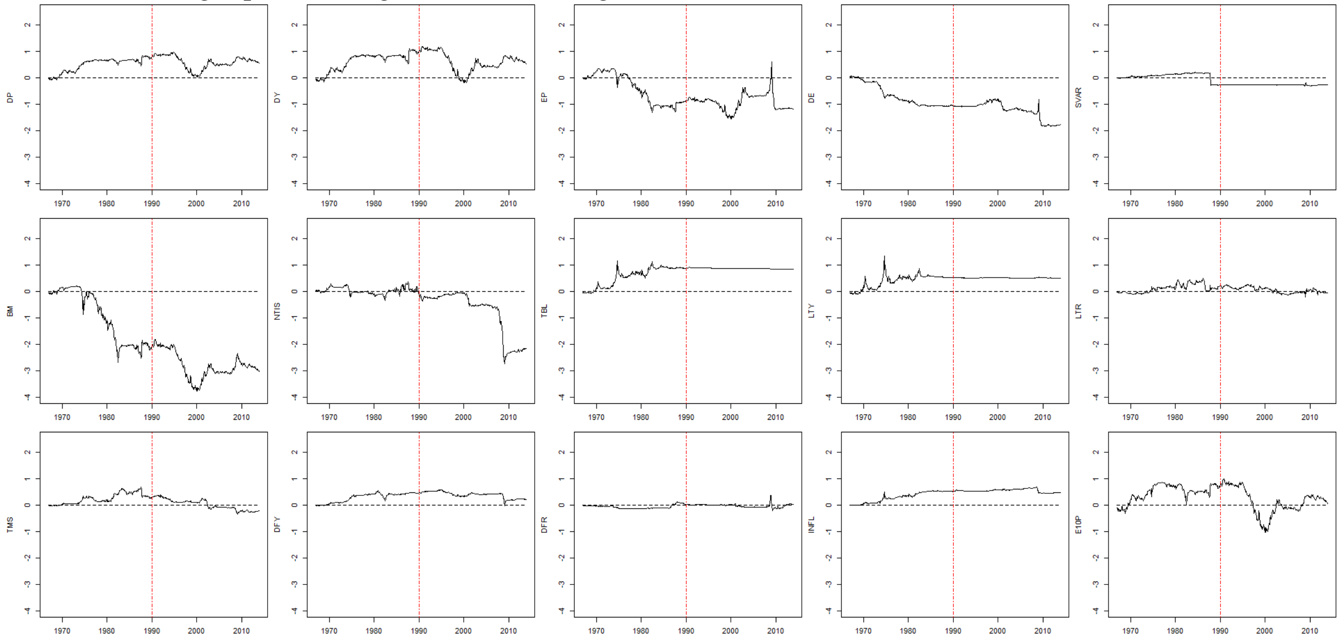
In general, Figure 4 shows that in terms of cumulative performance, few of the single-predictor models consistently outperforms the historical average. Some of the panels have positively sloped curves during some periods of time, but eventually display negatively sloped curves in the end, such as the one based on *E10P*. Also, the majority of the single-predictor forecasting models have a higher *MSPE* than the benchmark.

Figure 5 shows the same graphical analysis for $PLQC_j$, FQR_j, $j = 1, 2, 3, 4$, $FOLS_1$, $FOLS_2$[8] and *CSR* with $k = 1, 2, 3$. The curves for $PLQC_j$ and FQR_j do not exhibit substantial falloffs as those observed Figure 4. It 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 until 1990.12, but *FQR* is outperformed by *PLQC* during the 1991.1-2013.12 period. The results shown in Figure 2 suggest that most of the predictors become weak after 1990, as we found in the paper with equity premium.

Table 5 reports R^2_{OS} statistics and its significance through the p-values of the Clark and West (2007) test (*CW*). Ialso shows the p-values of the Diebold-Mariano (1995) test (*DM*) and the annual utility gain $\Delta\,(annual\%)$ associated with each forecasting model. The results for the entire 1967.1:2013.12 out-of-sample period confirm that few single-predictor forecasting models have a positive R^2_{OS}, and their

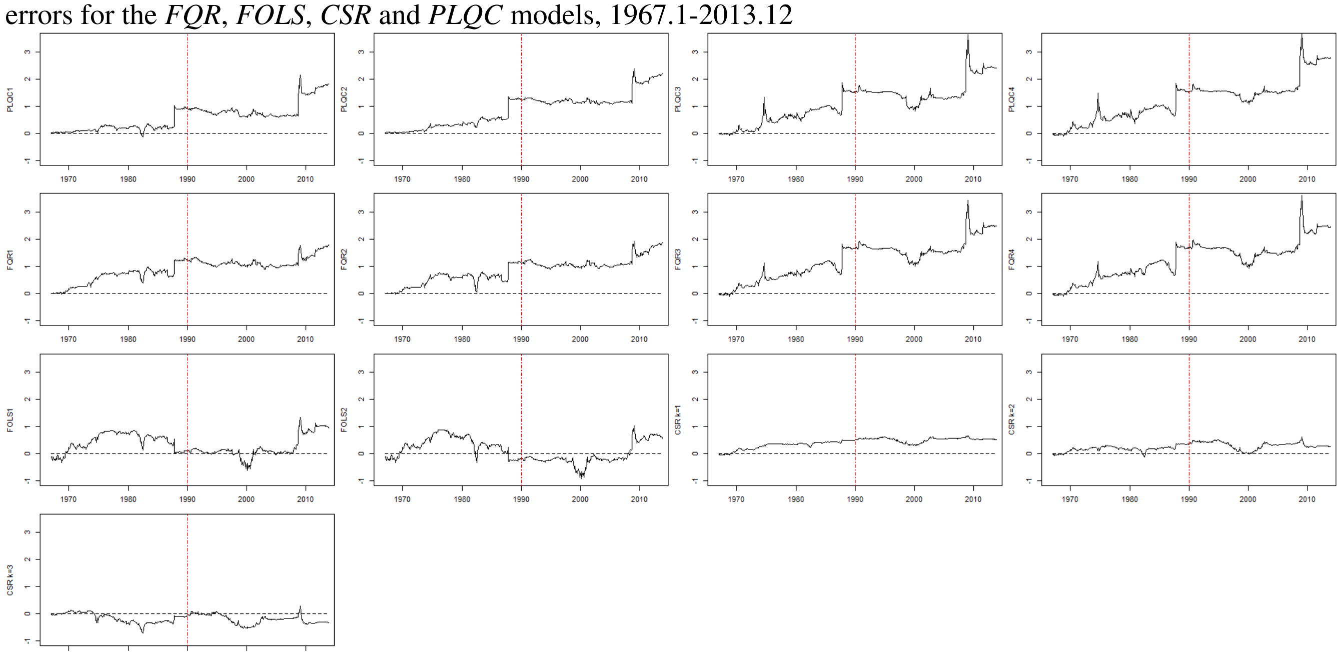
[8]Recall that *FOLS* forecasts are based on the *OLS* estimation of an equation whose predictors are selected by the ℓ_1-penalized quantile regression method. Since we have considered two sets of quantiles $\tau = (0.3, 0.5, 0.7)$ and $\tau = (0.3, 0.4, 0.5, 0.6, 07)$, there will be two such prediction equations and therefore two *FOLS* forecasts, denoted by $FOLS_j$, $j = 1, 2$.

Figure 4: 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 positive 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, none of the single-predictor models consistently outperforms the benchmark.

Figure 5: 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 positive 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* forecast is a top performer, especially after 1990.

p-values are mostly not statistically significant. The same happens to the *CSR* forecasts. The only exceptions in this long out-of-sample period are the *PLQC*, *FQR* and *FOLS* forecasts, but the *PLQC* forecasts outperform the rest generally.

As for the subperiod 1967.1-1990.12, Table 5 shows that some single-predictor models performed well, as well as *CSR* forecasts. The performance of *PLQC* and *FQR* forecasts are quite similar. *FQR* slightly outperforms the *PLQC*, suggesting if predictors are strong, there will be no advantage to use a forecasting device that is robust against (partially) weak predictors. However, since the quantile-regression-based *FQR* forecasts outperform *OLS*-based *FOLS*, we conclude that there is still an advantage to consider forecasts that are robust against estimation errors.

As noted earlier, most predictors become weaker after 1990. Indeed, all single-predictor models have non-significant R^2_{OS} values and negative utility gains in the 1991.1-2013.12 subperiod. The same results for *CSR* forecasts confirm that this methodology is also affected by the presence of weak predictors. The *PLQC* forecasts dominate any other forecasting method in terms of R^2_{OS} and utility gains. The performance of *FQR* deteriorates significantly during the 1991.1-2013.12 subperiod.

Finally, we look at the most recent out-of-sample subperiod, 2008.1-2013.12. 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^2_{OS} 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.

Table 6 shows the decomposition of the mean-square-prediction-error (*MSPE*). Recall that this decomposition measures the additional *MSPE* loss of *FOLS* forecasts relative to the *PLQC* forecasts. The first element on the righthand side of equation 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. For the out-of-sample period, 1967.1-1990.12, 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 Table 5. In case of strong predictors, most of the loss will be explained by *OLS* estimator's lack of robustness to estimation errors, so there is still an advantage to use quantile regression methods to avoid the effect of estimation errors. The situation changes dramatically when weak predictors become a more severe issue during the post-1990 out-of-sample period[9]. As a result, the second element dominates, indicating that most of the forecast accuracy loss is ascribed to the presence of partially weak predictors.

[9]During this period, the *MSPE* of *FQR* model is larger than that of the *FOLS*, leading to negative number in Table 6. But again, this emphasize the major role played by the presence of partially weak predictors.

Table 5: Out-of-sample Returns 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	0.48	0.95	0.04	0.43	1.45	0.14	0.01	1.86	-0.74	1.00	0.37	-1.05	0.12	0.87	0.36	-0.12
DY	0.46	0.97	0.03	0.67	1.85	0.23	0.01	2.67	-1.30	1.00	0.35	-1.39	0.36	0.84	0.30	0.34
EP	-1.07	0.99	0.29	0.63	-1.22	0.92	0.50	0.14	-0.88	0.96	0.25	1.13	-2.90	0.91	0.54	3.66
DE	-1.59	0.98	0.96	-0.63	-1.74	1.00	1.00	-1.57	-1.41	0.66	0.70	0.35	-2.33	0.84	0.69	0.02
SVAR	-0.24	0.51	0.67	-0.02	-0.45	0.44	0.70	-0.04	0.03	0.66	0.43	0.01	0.05	0.55	0.45	0.14
BM	-2.72	1.00	0.41	-0.98	-3.03	0.95	0.49	0.01	-2.33	1.00	0.34	-2.00	0.06	0.91	0.35	0.16
NTIS	-1.93	0.93	0.68	-0.82	-0.48	0.80	0.23	-0.72	-3.77	0.91	0.92	-0.92	-7.07	0.85	0.91	-4.25
TBL	0.76	0.09	0.05	1.36	1.45	0.08	0.05	2.75	-0.11	0.88	0.96	-0.09	-0.07	0.18	0.77	-0.04
LTY	0.44	0.13	0.12	1.06	0.82	0.12	0.11	2.13	-0.04	0.82	0.74	-0.05	-0.09	0.95	0.73	-0.17
LTR	-0.06	0.85	0.17	0.25	0.33	0.65	0.13	1.10	-0.56	0.89	0.46	-0.65	-0.11	0.65	0.41	-1.48
TMS	-0.19	0.74	0.34	0.26	0.54	0.53	0.12	0.91	-1.11	0.81	0.83	-0.42	-0.59	0.16	0.68	-0.87
DFY	0.18	0.70	0.18	0.09	0.79	0.14	0.02	1.20	-0.59	0.97	0.75	-1.07	-1.24	0.64	0.79	-2.92
DFR	0.03	0.64	0.37	0.27	-0.01	0.71	0.47	0.06	0.07	0.58	0.38	0.49	0.50	0.57	0.35	2.21
INFL	0.43	0.01	0.06	0.60	0.90	0.02	0.04	1.33	-0.16	0.12	0.60	-0.17	-0.88	0.59	0.86	-1.57
E10P	0.04	0.99	0.03	0.49	1.50	0.52	0.03	2.34	-1.81	1.00	0.22	-1.41	0.94	0.79	0.18	1.14
Complete Subset Regression Forecasts																
CSR k=1	0.46	0.73	0.03	0.39	0.91	0.10	0.01	1.04	-0.11	0.97	0.44	-0.29	-0.33	0.90	0.71	-0.54
CSR k=2	0.23	0.95	0.13	0.25	0.76	0.49	0.07	1.02	-0.45	0.99	0.48	-0.56	-0.61	0.86	0.65	-0.68
CSR k=3	-0.29	0.99	0.32	0.02	0.12	0.81	0.26	0.53	-0.80	0.99	0.51	-0.50	-0.88	0.80	0.60	0.25
Forecasts based on *LASSO*-Quantile Selection																
*FOLS*1	0.86	0.61	0.02	1.77	0.26	0.65	0.14	1.32	1.61	0.52	0.04	2.25	4.57	0.20	0.07	5.96
*FOLS*2	0.51	0.61	0.03	1.66	-0.22	0.74	0.18	1.17	1.43	0.41	0.05	2.16	3.68	0.22	0.09	4.69
*FQR*1	1.60	0.04	0.01	2.58	2.04	0.20	0.02	2.67	1.05	0.05	0.08	2.48	3.89	0.02	0.07	7.39
*FQR*2	1.67	0.05	0.01	2.45	1.91	0.27	0.03	2.66	1.36	0.04	0.07	2.23	4.13	0.03	0.08	5.61
*FQR*3	2.22	0.16	0.01	2.65	2.94	0.07	0.03	3.89	1.31	0.50	0.11	1.36	4.66	0.20	0.13	5.84
*FQR*4	2.18	0.23	0.02	2.44	2.99	0.12	0.03	3.88	1.16	0.56	0.12	0.95	4.52	0.26	0.14	4.64
*PLQC*1	1.63	0.12	0.03	1.57	1.45	0.29	0.10	0.95	1.87	0.14	0.08	2.22	6.17	0.11	0.08	6.11
*PLQC*2	1.97	0.04	0.02	1.96	2.05	0.13	0.07	1.70	1.85	0.08	0.07	2.23	5.50	0.08	0.08	5.80
*PLQC*3	2.15	0.17	0.02	2.10	2.65	0.07	0.06	3.51	1.52	0.53	0.11	0.63	5.50	0.30	0.11	3.08
*PLQC*4	2.49	0.14	0.02	2.54	2.73	0.13	0.05	4.00	2.18	0.37	0.09	1.03	6.31	0.23	0.10	3.74

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 shows the p-value of the Diebold-Mariano (1995) test (*DM*) and the annual utility gain $\Delta\,(annual\%)$ associated with each forecasting model over four out-of-sample periods. $R^2_{OS} > 0$, if the conditional forecast outperforms the benchmark. To test the null hypothesis $R^2_{OS} \leq 0$, the p-values for one-sided (upper-tail) Diebold-Mariano (1995) test (*DM*) and Clark and West (2007) test (*CW*) are obtained. 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.

Table 6: 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	74.42%	25.58%
1991.1 - 2013.12	-91.39%	191.39%

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*3, *FQR*3, and *FOLS*1 respectively in the paper. Results are similar based on *PLQC*4, *FQR*4 and *FOLS*2.