Economic Forecasting with Many Predictors

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Fanning Meng

May 2017

Dedication

This dissertation is dedicated to my beloved husband, Zhihao Hao, and my dearest son, Junqi Hao, for their support over these years. And I am also grateful to my parents and parents-in-law for taking care of my little boy during the days I'm away.

Acknowledgements

I would like to sincerely thank my advisor Dr. Luiz R. Lima for leading me onto the road of being a professional researcher. I also want to express my deep gratitude to Dr. Georg Schaur, Dr. Celeste Carruthers and Dr. Christian Vossler for their guidance and support on my research. Special thanks to Dr. Christian Vossler and Dr. Williams Neilson for their help with my sick leave requirements in Spring 2014 and Fall 2015. I would also like to thank Dr. Eric Kelley for serving on my committee and offering valuable comments on my research. Many thanks to Sherri Pinkston and Juliana Notaro for their wonderful assistance on many administrative issues. I want to express my appreciation to the whole Economics Department at Haslam College of Business for providing me the precious opportunity to achieve my academic goal.

Abstract

The dissertation is focused on the analysis of economic forecasting with a large number of predictors.

The first chapter 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.

In the second chapter, we find that incorporating distributional and high-frequency information into a forecasting model can produce substantial accuracy gains. Distributional information is included through a quantile combination approach, but estimation of quantile regressions with mixed-frequency data leads to a parameter proliferation problem. We consider extensions of the *MIDAS* and soft (hard) thresholding methods towards quantile regression. Our empirical study on *GDP* growth rate reveals a strong predictability gain when high-frequency and distributional information are adequately incorporated into the same forecasting model.

The third chapter analyzes the wage effects of college enrollment for returning adults based on the *NLSY79* data. To improve the estimation efficiency, we apply the double-selection model among time-varying features and individual fixed effects. The empirical results on hourly wage predictions show evidences towards the superiority

of double-selection model over a fixed-effect model. Based on the double-selection model, we find significant and positive returns on years of college enrollment for the returning adults. On average, one more year's college enrollment can increase hourly wage of returning adults by $1.12, an estimate that is about 7.7% higher than that from the fixed-effect model.

Table of Contents

List of Tables

List of Figures

Nomenclature

AL-LASSO	Asymmetric-loss *LASSO*
AR	Autoregressive
CPR	Correct Prediction Rates
CSR	Complete Subset Regressions
CW	Clark and West (2007) test
DGP	Data Generating Process
DM	Diebold-Mariano (2012) test
DS	Double-selection
EN	Elastic Net
FE	Fixed-effect
FOLS	Fixed OLS
FQR	Fixed Quantile Regression
GDP	Gross Domestic Production
HA	Historical Average
JAE	Journal of Applied Econometrics
LASSO	Least Absolute Shrinkage Selection Operator
MIDAS	Mixed Data Sampling
MSE	Mean Squared Error
MSPE	Mean-square-prediction-error
NLSY	National Longitudinal Survey of Youth
OLS	Ordinary Least Square

OOS	Out-of-sample
PC	Principal Components
PLQC	Post-*LASSO* Quantile Combination
PLQF	Post-*LASSO* Quantile Forecast
QCA	Quantile Combination Approach
QR	Quantile Regression
QS	Quantile Score
RFC	Robust Forecast Combination
RMSE	Root Mean Squared Error
RMSFE	Root Mean Squared Forecast Error

Chapter 1

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

1.1 Disclosure

This chapter is forthcoming in Journal of Applied Econometrics (*JAE*). This paper is coauthored with Luiz R. Lima. I mainly worked on the empirical analysis, including the construction, generation and evaluation of forecasting models. At the same time, Luiz developed a theoretical framework to explain the underlying working mechanism. We first submitted the article to *JAE* on September $8th$, 2015 and got the revise and resubmit decision in December, 2015. First revision was resubmitted in April $20th$, 2016 and got feedbacks in June, 2016. The last version was submitted on August $9th$, 2016. The paper was accepted on August $11th$, 2016.

1.2 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 macroeconomic 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 et al., 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. 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.

1.3 Econometric Methodology

Suppose that an econometrician is interested in forecasting the equity premium[1] 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

$$\begin{aligned} r_{t+1} &= X'_{t+1,t}\alpha + \left(X'_{t+1,t}\gamma\right)\eta_{t+1}, \\ \eta_{t+1}|I_t &\sim i.i.d.\ F_\eta\left(0,1\right), \end{aligned} \tag{1.1}$$

where $F_\eta\left(0,1\right)$ 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 (1.1) have been considered in the forecasting literature[2]. In this paper, we consider another special case of model (1.1) 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 $E\left(r_{t+1}|X_t\right) = X'_t\alpha$, whereas the conditional quantile of r_{t+1} at level $\tau \in (0,1)$, $Q_\tau\left(r_{t+1}|X_t\right)$, equals:

$$Q_\tau\left(r_{t+1}|X_t\right) = X'_t\alpha + X'_t\gamma F_\eta^{-1}\left(\tau\right) = X'_t\beta\left(\tau\right) \tag{1.2}$$

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

[2] Gaglianone and Lima (2012) and Gaglianone and Lima (2014) assume $X_{t+1,t} = (1, C_{t+1,t})'$ and $X_{t+1,t} = \left(1, f^1_{t+1,t}, ..., f^n_{t+1,t}\right)'$ respectively, where $C_{t+1,t}$ is the consensus forecast made at time t from the Survey of Professional Forecasts and $f^j_{t+1,t}$, $j = 1, ..., n$, are point forecasts made at time t by different economic agents.

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, $\beta(\tau)$.[3]

Following the literature Granger (1969); Granger and Newbold (1986); Christoffersen and Diebold (1997); 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 $e_{t+1} \equiv r_{t+1} - \widehat{r}_{t+1}$, that is, $L = L(e_{t+1})$, and $L(ae) = g(a)L(e)$ for some positive function g.[4]

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 1.1. *Under DGP(1.1) with $X'_{t+1,t} = X'_t$ and a homogeneous loss function (Assumption 1), the optimal forecast will be*

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

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.

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

[3] Model (1.1) 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 et al. (2011).

[4] This 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.

to obtain robust *MSPE* forecasts of equity premium is through the combination of quantile forecasts. That is:

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

where ω_τ is the weight assigned to the conditional quantile $Q_\tau\left(r_{t+1}|X_t\right)$. 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 $E\left(\eta_{t+1}\right) = \int_0^1 F_\eta^{-1}\left(t\right) dt = 0$.[5] However, with finite sample, we need to consider a grid of quantiles $(\tau_{\min}, ..., \tau_{\max})$ and approximate $\int_0^1 F_\eta^{-1}\left(t\right) dt$ by $\sum_{\tau=\tau_{\min}}^{\tau_{\max}} \omega_\tau F_\eta^{-1}\left(\tau\right)$. 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 ($\tau = 0.5$) will suffice to guarantee that $\sum_{\tau=\tau_{\min}}^{\tau_{\max}} \omega_\tau Q_\tau\left(r_{t+1}|X_t\right) = E\left(r_{t+1}|X_t\right)$. 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 $Q_\tau\left(r_{t+1}|X_t\right)$ are estimated using the quantile regression (QR) estimator, which is robust to estimation errors caused by occasional but extreme observations of equity premium.[6] 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 $\sum_{\tau=\tau_{\min}}^{\tau_{\max}} \omega_\tau Q_\tau\left(r_{t+1}|X_t\right)$ combines oppositely-biased predictions, and these

[5] Recall that $F_\eta^{-1}\left(\tau\right) = Q_\tau(\eta_{t+1})$.

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

biases cancel out each other. This cancelling out mitigates the problem of aggregate bias identified by Issler and Lima (2009).[7]

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 $Q_\tau\left(r_{t+1}|X_t\right)$ is fully known by the econometrician. However, *DGP* (1.1) 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.

1.3.1 The ℓ_1-penalized quantile regression estimator

Rewriting Equation 1.2, we have the conditional quantiles of r_{t+1}:

$$Q_\tau\left(r_{t+1}|X_t\right) = \beta_0\left(\tau\right) + x_t'\boldsymbol{\beta_1}(\tau) \qquad \tau \in (0,1)$$

where $\beta_0\left(\tau\right) = \alpha_0 + \gamma_0 F_\eta^{-1}\left(\tau\right)$, $\beta_1\left(\tau\right) = \alpha_1 + \gamma_1 F_\eta^{-1}\left(\tau\right)$, 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 et al., 2011). The *LASSO* quantile regression estimator solves the following problem :

$$\min_{\beta_0,\boldsymbol{\beta_1}} \sum_t \rho_\tau(r_{t+1} - \beta_0\left(\tau\right) - x_t'\boldsymbol{\beta_1}(\tau)) + \lambda\frac{\sqrt{\tau\left(1-\tau\right)}}{m} \quad \| \quad \boldsymbol{\beta_1}(\tau) \|_{\ell_1} \qquad (1.3)$$

[7] Aggregate 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 $\rho_\tau(e) \equiv [\tau - 1(e \leq 0)]\, e$; $1(\cdot)$ is the usual indicator function; m is the size of the estimation sample; $\| \cdot \|_{\ell_1}$ is the ℓ_1-norm, $\| \boldsymbol{\beta_1} \|_{\ell_1} = \sum_{i=1}^{k-1} |\boldsymbol{\beta_1}_i|$; $x_t = \left(x_{1,t}, x_{2,t}, ..., x_{(k-1),t}\right)'$.

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 et al. (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 $f^{\tau}_{t+1,t} = \beta_0(\tau) + x_t^{*\prime}\boldsymbol{\beta}(\tau)$, 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, $\widehat{r}_{t+1} = \sum_{j=1}^{k} \omega_{\tau_j} f^{\tau_j}_{t+1,t}$. The (*PLQC*) is a point (MSPE) forecast of the equity premium in $t+1$.

1.3.2 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 $t+1$, $f^{\tau_j}_{t+1,t}$, $j = 1, 2, ..., 5$:

$$\begin{pmatrix} f^{\tau_1}_{t+1,t} \\ f^{\tau_2}_{t+1,t} \\ f^{\tau_3}_{t+1,t} \\ f^{\tau_4}_{t+1,t} \\ f^{\tau_5}_{t+1,t} \end{pmatrix} = \begin{pmatrix} \beta_0(\tau_1) & \beta_1(\tau_1) & 0 & 0 \\ \beta_0(\tau_2) & \beta_1(\tau_2) & 0 & 0 \\ \beta_0(\tau_3) & \beta_1(\tau_3) & 0 & 0 \\ \beta_0(\tau_4) & 0 & \beta_2(\tau_4) & 0 \\ \beta_0(\tau_5) & 0 & \beta_2(\tau_5) & 0 \end{pmatrix} \times \begin{pmatrix} 1 \\ x_{1,t} \\ x_{2,t} \\ x_{3,t} \end{pmatrix} \tag{1.4}$$

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 (1.5):

$$\begin{aligned} \widehat{r}_{t+1} &= \sum_{j=1}^{5} \omega_{\tau_j} \beta_0 (\tau_j) + \sum_{j=1}^{3} \omega_{\tau_j} \beta_1 (\tau_j) x_{1,t} + \sum_{j=4}^{5} \omega_{\tau_j} \beta_2 (\tau_j) x_{2,t} \\ &= \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} \end{aligned} \tag{1.5}$$

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 $\widehat{r}_{t+1}$ will reflect their partial weakness. Indeed, since the contribution of $x_{1,t}$ to predict $f^{\tau_4}_{t+1,t}$ and $f^{\tau_5}_{t+1,t}$ is weak, our forecasting device eliminates $\beta_1 (\tau_4)$ and $\beta_1 (\tau_5)$ from β_1 in Equation (1.5). 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 (1.4) 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 (1.5) 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:

$$\widehat{r}_{t+1} = b_0 + b_1 x_{1,t} + b_2 x_{2,t} \tag{1.6}$$

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 (1.6) 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:

$$MSPE_{FOLS} - MSPE_{PLQC} = [MSPE_{FOLS} - MSPE_{FQR}] + [MSPE_{FQR} - MSPE_{PLQC}] \tag{1.7}$$

Hence, we decompose the MSPE difference between $FOLS$ and $PLQC$ forecasts into two elements. The first element on the righthand side of Equation (1.7) 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.

1.3.3 Weight selection

In this paper, we consider both time-invariant and time-variant weighting schemes. The former are simple averages of $f^{\tau}_{t+1,t}$. 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

$$
\begin{aligned}
PLQC_1 \quad &: \quad \frac{1}{3}f^{0.3}_{t+1,t} + \frac{1}{3}f^{0.5}_{t+1,t} + \frac{1}{3}f^{0.7}_{t+1,t} \\
PLQC_2 \quad &: \quad \frac{1}{5}f^{0.3}_{t+1,t} + \frac{1}{5}f^{0.4}_{t+1,t} + \frac{1}{5}f^{0.5}_{t+1,t} + \frac{1}{5}f^{0.6}_{t+1,t} + \frac{1}{5}f^{0.7}_{t+1,t}
\end{aligned}
$$

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 $f^{\tau}_{t+1,t}$, $\tau \in (\tau_1, \tau_2, ..., \tau_J)$, with the following two leading examples:

$$
\begin{aligned}
PLQC_3 \quad &: \quad r_{t+1} = \sum_{\tau=\tau_1}^{\tau_3} \omega_\tau f^{\tau}_{t+1,t} + \varepsilon_{t+1} \quad \tau \in (0.3; 0.5; 0.7) \\
s.t. \quad &\quad \omega_{\tau_1} + \omega_{\tau_2} + \omega_{\tau_3} = 1 \\
PLQC_4 \quad &: \quad r_{t+1} = \sum_{\tau=\tau_1}^{\tau_5} \omega_\tau f^{\tau}_{t+1,t} + \varepsilon_{t+1} \quad \tau \in (0.3; 0.4; 0.5; 0.6; 0.7) \\
s.t. \quad &\quad \omega_{\tau_1} + \omega_{\tau_2} + \omega_{\tau_3} + \omega_{\tau_4} + \omega_{\tau_5} = 1
\end{aligned}
\tag{1.8}
$$

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.

1.3.4 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:

$$
r_{t+1} = \alpha_i + \beta_i x_{i,t} + \varepsilon_{i,t+1} \tag{1.9}
$$

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 $f^i_{t+1,t} = \widehat{E}(r_{t+1}|X_t) = \widehat{\alpha}_i + \widehat{\beta}_i x_{i,t}$, where $\widehat{\alpha}_i$ and $\widehat{\beta}_i$ are *OLS* estimates of α_i and β_i.

Our data [8] 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.

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 (1.9); (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 et al., 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); 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.

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

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 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.[9] 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^2_{OS}, which compares the forecast from a conditional model, $\widehat{r}_{t+1}$, to that from the benchmark (unconditional) model $\overline{r}_{t+1}$ (Campbell and Thompson (2008)). We report the value of R^2_{OS} in percentage terms, $R^2_{OS}(\%) = 100 \times R^2_{OS}$. Second, to test the null hypothesis $R^2_{OS} \leq 0$, we apply both the Diebold and Mariano (2012) and Clark and West (2007) tests[10]. 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 model.

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

[10]The Diebold and Mariano (2012) and West (1996) statistics are often used to test the null hypothesis, $R^2_{OS} \leq 0$, 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 and Mariano (2012) (DM) and Clark and West (2007) tests can be severely undersized under the null hypothesis and have low power under the alternative hypothesis.

1.4 Empirical Results

1.4.1 Out-of-sample forecasting results

In Figures A.1 and A.2[11], 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.

In general, Figure A.1 shows that in terms of cumulative performance, few single-predictor models consistently outperform the historical average.[12] 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 A.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.[13] In sum, Figure A.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 A.2 shows the same graphical analysis for $PLQC_j$, FQR_j, $j = 1, 2, 3, 4$, $FOLS_1$, $FOLS_2$[14] and CSR with $k = 1, 2, 3$. The curves for $PLQC_j$ and FQR_j do not

[11] Figures and Tables are all in the appendix, labeled as A., B., C. for chapter 1, 2 and 3 respectively.

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

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

[14] 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

exhibit substantial falloffs as those observed in the single-predictor forecasting models (1.9). 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 A.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 index[15].

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 CSR forecasts 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 A.1. This table reports R^2_{OS} 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

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 $FOLS_j$, $j = 1, 2$.

[15]Based on a Monte-Carlo simulation experiment, we found that weak predictors can be harmful for forecasting.

Diebold and Mariano (2012) 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^2_{OS}. 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^2_{OS} and utility gain $\Delta\,(annual\%)$. Among the $PLQC$ forecasts, we notice that 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 A.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^2_{OS} 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 A.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^2_{OS} 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 A.1 show that the R^2_{OS} 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^2_{OS} of the FQR forecasts falls on average by 0.22% whereas the R^2_{OS} 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 A.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^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. 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 A.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 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 A.1. In case of strong predictors, most of the loss will be explained by *OLS* estimator's lack of robustness to estimation errors, so using quantile regression presents an advantage in that it avoids 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. 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.

1.4.2 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 $\widehat{r}_{t+1}$ as $\frac{1}{T^*}\sum_t(r_{t+1}-\widehat{r}_{t+1})^2$ and the unconditional forecast variance as $\frac{1}{T^*}\sum_t(\widehat{r}_{t+1}-\frac{1}{T^*}\sum_t\widehat{r}_{t+1})^2$, 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 A.3 and A.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 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 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 A.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 A.3), *PLQC* and *FQR* perform equally well. However, when predictors become weak (in Figure A.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 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 A.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 A.3 reports selection frequency for only 6 predictors, meaning that 9 (out of 15) predictors are fully weak. Thus, the prediction Equation (1.5) 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 A.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 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.

1.4.3 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:

$$\theta_t = \arg\min_{\theta_t \in R^{15}} \sum_t \rho_\tau \left(r_{t+1} - \sum_{i=1}^{15} \theta_{i,t} \widehat{r}_{i,t+1}(\tau) \right) \text{ s.t. } \sum_{i=1}^{15} \theta_{i,t} = 1; \quad \sum_{i=1}^{15} |\theta_{i,t}| \leq \delta_1 \tag{1.10}$$

where $\rho_\tau(\cdot)$ is the asymmetric loss, $\widehat{r}_{i,t+1}(\tau)$ is the quantile function obtained from a single-predictor quantile model, i.e., $\widehat{r}_{i,t+1}(\tau) = \alpha_i(\tau) + \beta_i(\tau) x_{i,t}$ 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 (1.10) results in an estimation of the τth conditional quantile of r_{t+1}, $\widehat{r}_{t+1}(\tau) = \sum_{i=1}^{15} \widehat{\theta}_{i,t}\widehat{r}_{i,t+1}(\tau)$. 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*, $f^{\tau}_{t+1,t}$ outperform other single-predictor quantile forecasts $\widehat{r}_{i,t+1}(\tau)$, $i = 1, ..., 15$ 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:

$$QS^k(\tau) = \frac{1}{T^*}\sum_{t=1}^{T^*}(r_{t+1} - \hat{Q}^k_{t+1,t}(\tau))(1.(r_{t+1} \leq \hat{Q}^k_{t+1,t}(\tau)) - \tau) \tag{1.11}$$

where T^* is the number of out-of-sample forecasts, r_{t+1} is the realized value of equity premium, $\hat{Q}^k_{t+1,t}(\tau)$ represents the quantile forecast at level τ of model k, and indicator function 1.() equals 1 if $r_{t+1} \leq \hat{Q}^k_{t+1,t}(\tau)$; 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 A.4 shows the *QS* for each single-predictor quantile model ($\widehat{r}_{i,t+1}(\tau)$), *AL-LASSO* ($\widehat{r}_{t+1}(\tau)$) and *PLQF* ($f^{\tau}_{t+1,t}$), 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.

Figure A.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 models[16]. 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 A.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.[17] In other words, the *AL-LASSO* method assumes quantile additivity, $\widehat{r}_{t+1}(\tau) = \sum_{j=1}^{15} \widehat{\theta}_{i,t}\widehat{r}_{i,t+1}(\tau)$, 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.

1.5 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

[16]For 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.

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

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.

Chapter 2

Quantile Forecasting with Mixed-Frequency Data

2.1 Introduction

A central issue in out-of-sample point forecasting is how to minimize the effect of estimation error on predictive accuracy measured in terms of mean squared error (*MSE*). Indeed, Elliott and Timmermann (2016, pp. 9-10) consider a simple example where the sample mean is used as an optimal point forecast for an independently and identically distributed random variable. They show that the estimation error improves in-sample accuracy but increases out-of-sample *MSE*. Elliott and Timmermann (2016) also point out that although the effect of estimation error on forecasting accuracy is of small order and so it disappears asymptotically, it may still have an important statistical and economic effect in many out-of-sample forecasting problems.

A potential source of estimation error resides in the lack of robustness of the *OLS* estimator to outliers or extreme observations. It is well known that the influence function of the *OLS* estimator is unbounded, implying that estimation of the conditional mean by *OLS* may lead to large estimation errors when data are not Gaussian (Zhao and Xiao, 2014). A second form of estimation error occurs when one

wants to use high frequency information to predict low-frequency variables. It is also well known that the inclusion of high-frequency predictors into a forecasting model leads to the so called parameter proliferation problem (Andreou et al., 2011, 2013) which corresponds to a situation where the number of parameters requiring estimate is very large compared to the sample size. In this case, Elliott and Timmermann (2016) notice that although over-parameterized models improve the in-sample *MSE*, they also reduce out-of-sample predictive accuracy.

Existing solutions to these two sources of estimation errors have been considered by the literature under independent setups. For instance, there is a long tradition in statistics that the conditional mean of a random variable can be approximated through the combination of its quantiles. Thus, under the presence of non-Gaussian observations, optimal *MSE* forecasts can still be obtained by combining different quantile forecasts. This approach is known as the quantile combination approach (*QCA*) and has been used in the forecasting literature by Judge et al. (1988); Taylor (2007); Ma and Pohlman (2008); Meligkotsidou et al. (2014) and Lima and Meng (2016), among others. However, the *QCA* itself cannot address the second source of estimation error caused by the parameter proliferation problem.

On the other hand, the *MIDAS* approach has been employed successfully in the forecasting literature as an efficient solution to the estimation error caused by the parameter proliferation problem, especially in models with first and second moment dynamics. Indeed, as shown by Pettenuzzo et al. (2016), including high-frequency predictors in models with time-varying volatility leads to more accurate forecasts in terms of *MSE*. This result suggests that high frequency information can also be useful for quantile forecasting, but little is known about how to make quantile forecasting with mixed-frequency data.

This paper proposes an unified approach that minimizes the effect of estimation error on predictive accuracy. The proposed approach relies on distributional information to overcome the lack of robustness of the *OLS* estimator and on ℓ_1-penalized quantile regressions to include high-frequency information into a quantile

forecasting model. We show that, contrary to the literature, the *MIDAS* approach is not useful to address the parameter proliferation problem in quantile regression because different quantiles are probably affected by different high-frequency predictors over time, making the data transformation proposed by *MIDAS* too restrictive for quantile forecasting.

We address the above problem by proposing a new data transformation based on the combination of soft and hard thresholding rules suggested by Bai and Ng (2008). In other words, we first select the most predictive high-frequency variables and use them to compute the most significant principal components. The novelty here is that this selection procedure is performed across different quantiles, giving rise to quantile forecasts that are based on different high-frequency predictors. In the final step, we combine the quantile forecasts to obtain a robust approximation of the *MSE* forecast. Thus, this approach incorporates both distributional and high-frequency information into the same forecasting model and hence contributes to simultaneously eliminate two important sources of estimation error that affect out-of-sample predictive accuracy.

GDP forecasting is essential to policy-makers as a tool to monitor the efficiency of monetary and fiscal policies. However, *GDP* is observable at a quarterly frequency while many potentially informative predictors are monthly or daily variables. Moreover, time series observations of *GDP* growth are sometimes characterized by the presence of extreme positive or negative values generated by periods of abrupt economic recession and expansion. The existence of high-frequency predictors and occasional but extreme observations suggests that we can apply the proposed approach to improve forecasts of *GDP* growth rates.

Our results show that including high-frequency and distributional information into the forecasting model produces a substantial gain in terms of forecasting accuracy of *GDP* growth rates. This is true across all methods used to address the parameter proliferation problem and along short and long forecast horizons. In fact, the average *MSE* of models that explore both distributional and high-frequency information is up to 18% lower than the average *MSE* from forecasts that do not use distributional

information. In general, our results support the conclusion that distributional and high-frequency information can improve forecasting accuracy, but *MIDAS* may not be the best approach to incorporate high-frequency predictors in quantile forecasting models.

We also show that the *QCA* does not come without a price, since it requires an absence of structural breaks in the quantile function. This condition is itself more stringent than requiring an absence of structural breaks in the mean function. When a break in the quantile function occurs (without a break in the mean), we interpret the *QCA* as being a misspecified model for the conditional mean. Indeed, we show in the empirical section that an equal-weight combination of *MSE* forecasts obtained using *QCA* works very well to forecast GDP growth during the subprime crisis period, suggesting that forecast combination can be used to attenuate the loss of accuracy caused by structural breaks in the quantile function.

The remainder of this paper is organized as follows. The econometric model and the quantile combination approach are introduced in Section 2. Section 3 discusses some existing solutions to the parameter proliferation problem. Section 4 presents an empirical analysis on *GDP* forecasting and Section 5 concludes.

2.2 The Econometric Model

Suppose that an agent is interested in forecasting a second-order stationary time series, y_{t+h}, using information available at time t, I_t. This information set can include predictors that are sampled at a higher frequency than y_{t+h}. For example, y_{t+h} could be a quarterly variable, while some of the predictors available at time t could be monthly or daily variables.

The success of high-frequency information to forecast low-frequency economic time series has been largely documented by the literature since the seminal paper by Ghysels et al. (2004). The paper by Pettenuzzo et al. (2016) suggests the possibility that high-frequency predictors can affect moments of the conditional distribution

other than the mean. In this paper, we consider the following data-generating process (DGP), where high-frequency predictors can affect not only the conditional mean but also the conditional quantiles of y_{t+h}.

$$\begin{aligned} y_{t+h} &= X_t'\boldsymbol{\alpha} + (X_t'\boldsymbol{\gamma})\eta_{t+h} \\ \eta_{t+h}|X_t &\sim i.i.d.\ F_\eta(0,1) \end{aligned} \tag{2.1}$$

where η_{t+h} is assumed to follow a unknown distribution F_η with mean 0 and unit variance. $X_t \in I_t$ is a vector of predictors observable at time t. The conditional mean of y_{t+h} is $E(y_{t+h}|X_t) = X_t'\boldsymbol{\alpha}$, and the conditional quantile of y_{t+h} is $Q_\tau\left(y_{t+h}|X_t\right) = X_t'\boldsymbol{\beta}(\tau)$, where $\boldsymbol{\beta}(\tau) = \boldsymbol{\alpha} + \boldsymbol{\gamma}F_\eta^{-1}(\tau)$, and $F_\eta^{-1}(\tau)$ represents the unconditional quantile of η_{t+h}. Similar DGPs has been considered in the forecasting literature by Gaglianone and Lima (2012, 2014) and Lima and Meng (2016). A novelty here is that we allow X_t to include mixed-frequency data. The assumption that y_{t+h} is second-order stationary implies that we do not allow for structural breaks in the quantile function of y_{t+h}.

In addition, we assume that the loss function $L(\cdot)$ is defined as in Patton and Timmermann (2007)[1], that is:

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

Based on DGP 2.1 and loss function $L(\cdot)$, Lima and Meng (2016) showed that the optimal forecast is given by:

$$\begin{aligned} \widehat{y}_{t+h,t} = Q_\tau\left(y_{t+h}|X_t\right) &= X_t'\boldsymbol{\alpha} + X_t'\boldsymbol{\gamma}F_\eta^{-1}(\tau) \\ &= E(y_{t+h}|X_t) + \kappa_\tau \end{aligned} \tag{2.2}$$

[1]The assumption on loss function is the same as the Assumption L2 of Patton and Timmermann (2007). It nests many loss functions including the most common MSE, MAE, lin-lin, and asymmetric quadratic loss functions.

where $\kappa_\tau = X_t'\gamma F_\eta^{-1}(\tau)$ is a bias measure with respect to the conditional mean and depends on X_t, the unknown error distribution F_η and loss function $L(\cdot)$.

If the loss function is mean squared error (MSE), then an optimal forecast corresponds to the conditional mean $E(y_{t+h}|X_t)$. In practice, the conditional mean is estimated by the usual ordinary least squares (OLS) estimator. However, as shown by Zhao and Xiao (2014), if the data are not normally distributed, OLS estimation is usually less efficient than methods that exploit distributional information. Nonetheless, given a set of quantile levels $\boldsymbol{\tau} = (\tau_1, \tau_2, \ldots, \tau_n)$, one can still approximate the conditional mean (MSE forecast) by combining n quantile forecasts,

$$\begin{aligned} \sum_{\tau=\tau_1}^{\tau_n} \omega_\tau Q_\tau(y_{t+h}|X_t) &= E(y_{t+h}|X_t) + \sum_{\tau=\tau_1}^{\tau_n} \omega_\tau \kappa_\tau \\ &= E(y_{t+h}|X_t) + X_t'\gamma \sum_{\tau=\tau_1}^{\tau_n} \omega_\tau F_\eta^{-1}(\tau) \end{aligned} \tag{2.3}$$

where $\omega_\tau \in (0,1)$ is the weight assigned to the quantile forecast at level τ. Notice that the weights are quantile-specific, since they are aimed at approximating the mean of η_{t+h}, 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), page 302). Thus, given that η_{t+h} is i.i.d., we have $E(\eta_{t+h}) = \int_0^1 F_\eta^{-1}(t)\,dt = 0$.[2] However, in a finite sample we need to consider a grid of quantiles $\boldsymbol{\tau} = (\tau_1, \tau_2, \ldots, \tau_n)$ and approximate $\int_0^1 F_\eta^{-1}(t)dt$ by $\sum_{\tau=\tau_1}^{\tau_n} \omega_\tau F_\eta^{-1}(\tau)$.

In the empirical analysis, we choose equal weights so that $\omega_\tau = \frac{1}{n}$, for $\forall\tau$.[3] A similar strategy was used by Pettenuzzo et al. (2016) that computes point forecasts by averaging multiple draws from the posterior distribution of the target variable y_{t+h}. Lima and Meng (2016) show that since the low-end (high-end) quantiles produce

[2] Recall that $F_{\eta,h}^{-1}(\tau) = Q_\tau(\eta_{t+h})$.

[3] As shown in Lima and Meng (2016), the difference between time-varying and time-fixed weight scheme does not affect the results in a significant way. We also tried time-varying weights with a constraint that the weights sum up to 1, but no better results are found.

downwardly (upwardly) biased forecast of the conditional mean, another insight from the quantile combination approach is that the point forecast $\sum_{\tau=\tau_1}^{\tau_n} \omega_\tau Q_\tau\left(y_{t+h}|X_t\right)$ 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)[4]. Unlike the rest of the literature, we allow for high-frequency predictors to affect the forecast of y_{t+h}. This creates a parameter proliferation problem in the quantile function (2.2) which we discuss next.

2.3 The Parameter Proliferation Problem

In the optimal forecast (2.2), if we restrict the predictors X_t to those sampled at the same frequency as y_{t+h}, we would miss other critical information resources, especially the ones updated more frequently. In order to capture more informative resources, we assume that $X_t = \left(W_t, X_t^M\right)'$ where W_t is a $k \times 1$ vector of pre-determined predictors, such as lags of y_{t+h} as well as predictors sampled at the same frequency as y_{t+h}, and $X_t^M = \left(X_t^1, X_t^2, ..., X_t^p\right)'$ is $p \times 1$ vector of high-frequency predictors with X_t^j, $j = 1, 2, \ldots, p$, being updated m_j times between $t-1$ and t, where $m_j > 1$ for $\forall j$. For example, if y_t is a quarterly variable and X_t^j is a monthly variable, we would have $m_j = 3$. Moreover, if y_{t+h} is affected by four ($q = 4$) of its own lags, then there would be $p_j = q * m_j = 12$ lags of the monthly predictor X_t^j, that is, $X_t^j = (x_t^j, x_{t-\frac{1}{3}}^j, x_{t-\frac{2}{3}}^j, x_{t-1}^j, \ldots, x_{t-\frac{11}{3}}^j)'$.

Given p high-frequency predictors with p_j lag predictors each, there would be up to $\sum_{j=1}^p p_j = P$ parameters. Without additional restrictions, the total number of coefficients to be estimated equals $K = 1 + k + \sum_{j=1}^p p_j = 1 + k + P$.[5] In the case when K is close to or even larger than the total number of observations T, we are faced with a parameter proliferation problem.

[4] Aggregate 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.

[5] Recall $Q_\tau\left(y_{t+h}|X_t\right) = X_t'\boldsymbol{\alpha} + X_t'\boldsymbol{\gamma}F_\eta^{-1}\left(\tau\right) = X_t'\boldsymbol{\beta}(\tau)$. The vector of quantile coefficients $\boldsymbol{\beta}(\tau)$ includes K coefficients, where $\beta_k(\tau) = \alpha_k + \gamma_k F_\eta^{-1}\left(\tau\right)$, $k = 0, 1, \ldots, K-1$.

2.3.1 Solutions to the Parameter Proliferation Problem

The bridge model was used as a first attempt to include high-frequency predictors into a forecasting model. It simply connects low-frequency target variables to high-frequency predictors through aggregation (Baffigi et al., 2004). However, valuable information could be lost during the aggregation process. Another approach is through the state-space model (Mariano and Murasawa, 2003; Bai et al., 2013), where a system of two equations, measurement equation and state equation, is estimated. The major challenge for the state-space model is that it requires a large number of parameters to be estimated.

A more successful solution used by the forecasting literature is the mixed data sampling (*MIDAS*) regression model, where one assumes a restriction on the form of how the distributed lags are included in the regression equation (Ghysels et al., 2004, 2005; Clements and Galvão, 2008, 2009; Kuzin et al., 2011). One caveat of the *MIDAS* approach is that it is limited by the number of high-frequency predictors that can be included. In other words, *MIDAS* is able to reduce the number of parameters if just a few high-frequency predictors are included in the regression model. Hence, we follow Pettenuzzo et al. (2016) and focus only on the analysis of one high-frequency predictor at a time. In particular, we use a special case of the *MIDAS* approach by which lags of X_t^j are included in the regression model through the following polynomial:

$$B(L^{\frac{1}{m_j}}; \boldsymbol{\theta_j}) \quad = \quad \sum_{i=0}^{p_j-1} b(i; \boldsymbol{\theta_j}) L^{\frac{i}{m_j}}$$

where $L^{\frac{i}{m_j}}$ is the lag operator and $L^{\frac{i}{m_j}} X_t^j = x^j_{t-\frac{i}{m_j}}$. And $b(i; \boldsymbol{\theta_j})$ is the weight assigned to the i^{th} lagged term, $i = (1, 2, \ldots, p_j)$, which is assumed to take an Almon form, that is:

$$b(i, \boldsymbol{\theta_j}) = \sum_{k=0}^{q_j-1} \theta_k i^k \quad i = 1, \ldots, p_j$$

where $\boldsymbol{\theta_j} = (\theta_0, \theta_1, \ldots, \theta_{q_j-1})$, $q_j << p_j$ and q_j is the order of the polynomial. Thus, the $p_j \times 1$ vector of high frequency data X_t^j can be converted to a vector of q_j transformed variables $\tilde{X}_t^j$ based on the matrix $\boldsymbol{Q}$

$$\boldsymbol{Q} = \begin{bmatrix} 1 & 1 & 1 & \ldots & 1 \\ 1 & 2 & 3 & \ldots & p_j \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 2^{q_j-1} & 3^{q_j-1} & \ldots & p_j^{q_j-1} \end{bmatrix}.$$

After this transformation, the vector of predictors included in the location-scale model (2.1) becomes equal to $X_t^* = (W_t, \tilde{X}_t^j)'$, where $\tilde{X}_t^j = QX_t^j$. Notice that the total number of parameters considered for estimation is substantially reduced to $K^* = 1 + k + q_j$ and the optimal forecast (2.2) becomes

$$\begin{aligned} \widehat{y}_{t+h} &= Q_\tau\left(y_{t+h}|X_t^*\right) = X_t^{*\prime}\alpha^* + X_t^{*\prime}\gamma^* F_\eta^{-1}(\tau) \\ &= E\left(y_{t+h}|X_t^*\right) + \kappa_\tau^*, \end{aligned} \tag{2.4}$$

where $Q_\tau\left(y_{t+h}|X_t^*\right) = X_t^{*\prime}\beta^*(\tau)$ is the τ^{th} quantile of y_{t+h} conditional on X_t^* and $\beta^*(\tau) = \alpha^* + \gamma^* F_\eta^{-1}(\tau)$ can now be estimated by the standard quantile regression estimator since the parameter proliferation problem has been eliminated. Notice that $E\left(y_{t+h}|X_t^*\right)$ corresponds exactly to the optimal MSE forecast obtained from Equation (8) in Pettenuzzo et al. (2016). Thus, the above optimal forecast is equal to the MSE forecast plus a bias measure $\kappa_\tau^* = X_t^{*\prime}\gamma^* F_\eta^{-1}(\tau)$ that depends on the loss function, the error distribution F_η and X_t^*. If the loss function corresponds to the mean-square-error, then $\widehat{y}_{t+h} = E\left(y_{t+h}|X_t^*\right)$ which can be efficiently estimated by OLS if the data are Gaussian. In the absence of Gaussianity, efficient estimation of $E\left(y_{t+h}|X_t^*\right)$ can be obtained by the quantile combination approach explained in the previous section.

Unfortunately, the $MIDAS$ approach is likely to be undermined by the "ad-hoc" choice of the Q matrix. Moreover, the same vector of transformed predictors $\tilde{X}_t^j = QX_t^j$ is used across all quantiles as shown in Equation (2.4) no matter whether they

have predictive power or not. In other words, the *MIDAS* approach does not always select the predictors that are really useful to obtain accurate forecasts across different quantiles.

To overcome the above problems, we adopt a new solution based on the combination of soft and hard thresholding rules. Let the transformation $T_t = \omega_t X_t^M$ map the data X_t^M from an original space of P covariates to a new space of P variables, which are uncorrelated over the data set. These new transformed variables are called principal components but not all of them need to be kept. Indeed, keeping only the first L principal components gives rise to a truncated transformation $T_{Lt} = \omega_{Lt} X_t^M$, where the matrix ω_{Lt} performs a linear transformation in the original data X_t^M. This approach is commonly adopted in the literature of factor analysis.

Bai and Ng (2008) propose a transformation $T_{Lt}^* = \omega_{Lt}^* X_t^M$, where ω_{Lt}^* is a sparse matrix of ω_{Lt}. This is equivalent to saying that ω_{Lt}^* places zero weight on some predictors, so that the linear transformation resulting from T_{Lt}^* will exclude some predictors. This is known as soft thresholding (Bai and Ng, 2008).

In addition to this soft thresholding rule, we consider a truncated version of ω_{Lt}^*, denoted as $\omega_{\ell t}^*$, which is equivalent to select fewer principal components based on an additional hard thresholding rule. When these same data transformations are performed across quantiles, one obtains $\ell(\tau)$ principal components per quantile, that is:

$$T_{\ell t}^*(\tau) = \omega_{\ell t}^*(\tau) X_t^M \tag{2.5}$$

where the quantile-specific matrix $\omega_{\ell t}^*(\tau)$ effectively puts zero weights on those predictors not useful to forecast the τ^{th} conditional quantile of y_{t+h}. Put differently, the above approach is equivalent to selecting the most powerful predictors for each τ^{th} quantile before computing the first $\ell(\tau)$ principal components. These selections are based on soft and hard thresholding rules, which are described in the next section.

2.3.2 Soft Thresholding

We first apply a penalization method (either *LASSO* or *Elastic Net*) at each quantile level τ to select the most powerful subset of (lags of) high-frequency predictors, from which we construct orthogonal factors based on the principal components method. This first step was named soft thresholding rule by Bai and Ng (2008). In the second step, we use a hard thresholding rule to determine which and how many factors from the first step will be used to forecast the quantile of y_{t+h}.

Given the original vector of predictors that originates the parameter proliferation problem, i.e. $X_t = \left(W_t, X_t^M\right)'$, we rewrite the optimal forecast (2.2) as follows

$$Q_\tau\left(y_{t+h}|W_t, X_t^M\right) = W_t'\boldsymbol{\beta}(\boldsymbol{\tau}) + X_t^{M\prime}\boldsymbol{\phi}(\boldsymbol{\tau}) \tag{2.6}$$

In order to solve the parameter proliferation problem, we first apply the soft thresholding method, which selects (lags of) high-frequency predictors in X_t^M by either *LASSO* or *Elastic Net* (*EN*). Then common factors are computed using principal components based on the selected predictors. In what follows, we explain how we use *LASSO* and *Elastic Net* (*EN*) to estimate the quantile function (2.6).

The *LASSO* Quantile Estimator

The *LASSO* estimator of the quantile function at level $\tau \in (\tau_1, \ldots, \tau_n)$ was developed by Belloni et al. (2011) and can be summarized as follows:

$$\min_{\boldsymbol{\beta},\boldsymbol{\phi}} \sum_t \rho_\tau(y_{t+h} - W_t'\boldsymbol{\beta}(\boldsymbol{\tau}) - X_t^{M\prime}\boldsymbol{\phi}(\boldsymbol{\tau})) + \lambda_\tau \frac{\sqrt{\tau\left(1-\tau\right)}}{m} \quad \| \quad \boldsymbol{\phi}(\boldsymbol{\tau}) \|_{\ell_1} \tag{2.7}$$

$$\| \boldsymbol{\phi}(\boldsymbol{\tau}) \|_{\ell_1} = \sum_{i=1}^{\boldsymbol{P}} |\boldsymbol{\phi}(\boldsymbol{\tau})_i|$$

where the first part, $\rho_\tau(e) \equiv [\tau - 1(e \leq 0)]e$, represents the standard quantile estimation and the function $1(e \leq 0)$ equals 1 if $e \leq 0$, otherwise equals 0. The *LASSO* estimator imposes a penalty, $\lambda_\tau \frac{\sqrt{\tau(1-\tau)}}{m}$, on the coefficients of all (lags of)

high-frequency predictors, where m stands for the estimation sample size. The optimal value of parameter λ_τ for each quantile τ is obtained following Belloni et al. (2011).

One distinguishing feature of *LASSO* is that coefficients of insignificant predictors can be exactly zero, leaving forecasts unaffected by uninformative predictors. Although *LASSO* is successful at variable selection, it also has two potential limitations that can affect forecasting accuracy. In particular, in the case that the number of parameters K is larger than the sample size, T, as in our empirical exercise, *LASSO* selects at most T variables before it saturates. Second, if there is a group of variables among which the pairwise correlations are very high, then *LASSO* tends to select one variable from the group and does not care which one is selected (Zou and Hastie, 2005).

Hence, since lags of high-frequency predictors are highly correlated, *LASSO* will tend to select one and drop the rest, failing to take full advantage of the high-frequency information. In our empirical analysis, among lags of macroeconomic and financial predictors, the correlation between any two lag predictors in X_t^M can be very close to 1. Table B.1 reports the correlation coefficients among 6 randomly selected lags of macro predictors and among 6 lags of the financial predictor *IRS*[6]. As we see, some of them are very close to one.

The *LASSO* penalty is convex, but not strictly convex. Zou and Hastie (2005) showed that strict convexity enforces the grouping effect, assigning similar coefficients to highly correlated predictors. In what follows we show how to incorporate strict convex penalty in the *LASSO* quantile regression estimator.

[6]The 6 randomly selected macro lag predictors are the 1^{st} lag of variable "personal income" (lag1PI), 2^{nd} lag of variable "manufacturing and trade inventories" (lag2MTinvent), 4^{th} lag of variable "sales of retail stores" (lag4Retailsales), 5^{th} lag of variable "currency held by the public" (lag5currency), 7^{th} lag of variable "industrial production index-consumer goods" (lag7IPconsgds) and 8^{th} lag of variable "depository Inst Reserves: nonborrowed, adjusted for reserve requirement changes" (lag8Reservesnonbor) (Ludvigson and Ng, 2009). The 6 lags of *IRS* correspond respectively to $1^{st}, 41^{th}, 81^{th}, 121^{th}, 161^{th}$ and 201^{th} lag.

The *Elastic Net* Quantile Estimator

When the predictors are highly correlated, a possible solution is to select them through "Elastic Net" (*EN*). The idea behind the *EN* procedure is to stretch the fishing net to retain all the "big fish". Like *LASSO*, *EN* simultaneously shrinks the estimates and performs model selection. The *EN* objective function is

$$\min_{\boldsymbol{\beta},\boldsymbol{\phi}} \sum_{t} \rho_\tau(y_{t+h} - W_t'\boldsymbol{\beta}(\boldsymbol{\tau}) - X_t^{M\prime}\boldsymbol{\phi}(\boldsymbol{\tau})) + \lambda_{1\tau} \parallel \boldsymbol{\phi}(\boldsymbol{\tau}) \parallel_{\ell_1} + \lambda_{2\tau} \parallel \boldsymbol{\phi}(\boldsymbol{\tau}) \parallel_{\ell_2} \quad (2.8)$$

where $\parallel \boldsymbol{\phi}(\boldsymbol{\tau}) \parallel_{\ell_1} = \sum_{i=1}^{\boldsymbol{P}} |\boldsymbol{\phi}(\boldsymbol{\tau})_i|$ and $\parallel \boldsymbol{\phi}(\boldsymbol{\tau}) \parallel_{\ell_2} = \sum_{i=1}^{P} \boldsymbol{\phi}(\boldsymbol{\tau})_i^2$. Thus, "Elastic Net" nests both *LASSO* and *Ridge* quantile regressions. When the two parameters $\lambda_{1\tau}$ and $\lambda_{2\tau}$ satisfy the relation $\frac{\lambda_{2\tau}}{\lambda_{1\tau}+\lambda_{2\tau}} > 0$, the *EN* penalty is strictly convex, which enforces highly-correlated predictors to have similar coefficients. As a result, *EN* can capture all significant predictors even if they are highly correlated.

To increase computational efficiency, we follow Bai and Ng (2008)[7] and reformulate the *EN* as a *LASSO* problem. It has a computationally appealing property because we can solve the *EN* objective function by using algorithms of *LASSO* proposed by Belloni et al. (2011). In order to implement this representation, we define new variables (assuming for simplicity that the W variables are absent):

$$y_t^+ = \begin{pmatrix} y_t \\ \boldsymbol{O_P} \end{pmatrix} \quad X_t^+ = \frac{1}{\sqrt{1+\lambda_2}} \begin{pmatrix} X_t^M \\ \sqrt{\lambda_2}\boldsymbol{I_P} \end{pmatrix}$$

Where $\boldsymbol{O_P}$ represents $P \times 1$ vector of zeros and $\boldsymbol{I_p}$ is a $P \times P$ identity matrix. Note that the sample size is now equal to $T + P$ which means that the elastic net can potentially select all P high-frequency predictors in all situations.

Based on the new variables, y_t^+ and X_t^+, the *EN* objective function (2.8) can be rewritten in terms of the ℓ_1-penalized quantile regression method studied by Belloni

[7]Different from us, Bai and Ng (2008) estimate the conditional mean directly based on the *MSE* loss function.

et al. (2011), that is:

$$\min_{\boldsymbol{\beta},\boldsymbol{\phi}} \sum_{t=1}^{T^*+P} \rho_\tau(y_{t+h}^+ - X_t^{+\prime}\boldsymbol{\phi}(\boldsymbol{\tau})) + \gamma_\tau \parallel \boldsymbol{\phi}(\boldsymbol{\tau}) \parallel_{\ell_1} \tag{2.9}$$

where $\gamma_\tau = \frac{\lambda_{1\tau}}{\sqrt{1+\lambda_{2\tau}}}$. As noticed by Zou and Hastie (2005), the optimizer in (2.9) is probably different from the one in (2.8), but since our interest at this stage is variable selection, we follow Bai and Ng (2008) and consider only the ordering of variables provided by the *Elastic Net* in (2.9). The optimal value of $\lambda_{2\tau}$[8] is obtained by minimizing the mean cross-validated errors of the forecasting model with the *EN* mixing parameter restricted to $\alpha = 0.5$ (Friedman et al., 2010)[9]. The penalty level γ_τ is obtained as in the previous section 2.3.2.

For both *LASSO* and *EN*, the soft thresholding rule is that by which a predictor will be selected if the absolute value of its estimated coefficient is significantly different from 0. In the empirical analysis, we define a coefficient estimate as significantly different from 0 if its absolute value is larger than a certain threshold. Specifically, we choose one-twentieth of the largest absolute value of the estimated coefficients as our selection threshold.

Thus, given s_τ predictors $X_t^{S_\tau}$ selected by using either *LASSO* or *Elastic Net*, we construct a vector of common factors, $F_t^{S_\tau} = (F_{t,1}, F_{t,2}, \ldots, F_{t,s_\tau})$, from the principal components of $X_t^{S_\tau\prime} X_t^{S_\tau}$ and restrict the maximum number of common factors to 20. This corresponds to what we have named soft-thresholding at the quantile level τ.

2.3.3 Hard Thresholding

Among the s_τ orthogonal common factors (principal components), we apply a second round of selection based on a hard thresholding rule. In other words, a common

[8] Since a random number is involved in the estimation of $\lambda_{2\tau}$ during the cross-validation process, we repeat the process 100 times in order to reduce the randomness in the results.

[9] The penalty of *EN* models with parameter λ and mixing parameter α can be represented by $\frac{1-\alpha}{2}\lambda \parallel \boldsymbol{\phi} \parallel_{\ell_2} + \alpha\lambda \parallel \boldsymbol{\phi} \parallel_{\ell_1}$. Hence, we have a *LASSO* estimator if $\alpha = 1$; if $\alpha = 0$, we have *Ridge*. To combine *LASSO* and *Ridge*, we chose $\alpha = 0.5$.

factor $F_{t,s}$, $s \in (1, \ldots, s_\tau)$, will be classified as significant if its p-value is no larger than a certain threshold c, $|pval_s| \leq c$, otherwise it will be dropped.[10] This last step selects $\ell(\tau)$ common factors at the quantile level τ, leading to the data transformation represented by Equation (2.5). Thus, the optimal forecast (2.2) becomes

$$Q_\tau(y_{t+h}|W_t, PC_t^\tau) = W_t'\boldsymbol{\beta}(\boldsymbol{\tau}) + PC_t^{\tau\prime}\boldsymbol{\varphi}(\boldsymbol{\tau}) \tag{2.10}$$

where PC_t^τ is the vector of $\ell(\tau)$ selected common factors at the quantile level τ.

Finally, for n widely spread quantile levels, we generate a conditional-mean (MSE) forecast of y_{t+h} by equal weighting its corresponding quantile forecasts, that is:

$$f_{t+h,t} = \sum_{\tau=\tau_1}^{\tau_n} \frac{1}{n} Q_\tau(y_{t+h}|W_t, PC_t^\tau) \tag{2.11}$$

We name these two conditional-mean forecasts $f_{t+h,t} : LASSO$ and $f_{t+h,t} : EN$, respectively, depending on whether $LASSO$ or EN is used to select the predictors in the first round. Likewise, we name $f_{t+h,t}^{MIDAS}$ the following conditional mean forecast

$$f_{t+h,t}^{MIDAS} = \sum_{\tau=\tau_1}^{\tau_n} \frac{1}{n} Q_\tau\left(y_{t+h}|W_t, \widetilde{X}_t^j\right) \tag{2.12}$$

where $\tilde{X}_t^j$ is the transformed vector of high-frequency predictors obtained using the Almon lag polynomial function. In the next section, we are going to employ these two methods to forecast the real GDP growth rate.

2.4 Empirical Analysis

2.4.1 Data

We obtain quarterly real GDP data (Y_t) from the Federal Reserve Bank of St. Louis ranging from $1961Q4$ to $2012Q1$. Then we compute the annualized log change of

[10] In the empirical analysis, we set $c = 0.01$.

GDP as $y_t = ln(\frac{Y_t}{Y_{t-1}}) * 400$. This is our target variable and there are 201 quarterly observations in total.

Figure B.1 shows the annualized quarterly GDP growth rates from $1963Q1 - 2012Q1$. Extreme large negative growth rates appear in $1980Q2$ and $2008Q4$, and extreme large positive ones appear in $1971Q1$ and $1978Q2$. The existence of extreme observations signals that the data may violate the normal distribution assumptions required for the efficiency of OLS estimations.

To see what the distribution looks like, Figure B.2 displays the histogram plot of the annualized quarterly GDP growth rate, compared to a normal distribution curve (the blue line). Normality tests[11] significantly reject the null hypothesis of normal distribution for GDP growth rates. Recall that in the absence of Gaussianity, the OLS estimator is usually less efficient than methods that exploit distributional information. Without some regularity conditions, the OLS estimator may not even be consistent as in the case when the data are generated from a Cauchy distribution (Zhao and Xiao, 2014).

Robust forecasts of the conditional mean based on the combination of quantile forecasts do not come without a price. Requiring an absence of structural breaks in the quantile function is stronger than requiring an absence of structural breaks in the mean function, since the former could be violated while the latter still holds.

To check whether the series of GDP growth rate is stationary or not, we apply the Xiao and Lima (2007) test for the second-order stationarity and the Kwiatkowski et al. (1992) test for the first-order stationarity. The stationarity test by Kwiatkowski et al. (1992) provides an answer about whether or not the mean function has broken, whereas the test by Xiao and Lima (2007) provides an answer to whether or not a break has occurred at the scale of the distribution of y_t. We suspect that the underlying distributions may be distorted by the financial crisis in 2008, so we applied

[11]The normality tests we did include the Jarque-Bera Jarque and Bera (1987) normality test, adjusted Jarque-Bera (Urzúa, 1996) normality test, skewness normality test and kurtosis normality test (Shapiro et al., 1968).

the stationarity tests to the whole sample as well as to a sample excluding the post-2008 observations. The test statistics and critical values at 5% level are reported in Table B.2. We can see that there is an absence of structural breaks in the mean function during the two periods of our analysis, but there seems to be a break in the scale of the distribution, when the post-crisis observations are added to the sample. This stationarity tests suggest that point forecast based on the quantile combination approach may perform quite well before the subprime crisis, but not very well after that due to the lack of second-order stationarity.

We consider 132 monthly macroeconomic time series, from $1960m1$ to $2011m12$, as explanatory variables[12]. It contains 6 categories of macroeconomic quantities, "output and income", "labor market", "housing", "consumption, orders and inventories", "money and credit" and "bond and exchange rates" as recorded in Ludvigson and Ng (2010) and Jurado et al. (2015).

In addition, we also explore the value of daily financial predictors in *GDP* forecasting (Andreou et al., 2013; Pettenuzzo et al., 2014). Specifically, we analyzed 6 financial predictors: the excess return on the market (*RET*); three Fama/French factors (*SMB, HML, MOM*); the Interest Rate Spread (*IRS*); and the Effective Federal Funds Rate (*DFF*). Detailed descriptions on each of these financial predictors are displayed in Table B.3. The Interest Rate Spread (*IRS*) and the Effective Federal Funds Rate (*DFF*) are obtained from the Federal Reserve Bank of St. Louis whereas *RET, SMB, HML and MOM* are available at the Kenneth French's website [13].

We follow Pettenuzzo et al. (2014) and assume four autoregressive lags of y_t ($q = 4$) and twelve months of past macro and financial predictors. In sum, there are $12 \times 132 = 1584$ lagged macroeconomic predictors. For daily financial predictors, since the number of missing observations varies due to different months and years, we consider the most recent 222 lags of *RET*, 224 lags for predictors *SMB, HML, MOM*, 246 lags for *IRS* and 365 lags for *DFF*. For example, for one-quarter ahead *GDP*

[12]The raw data is obtained from Jurado et al. (2015).

[13]http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. We thank Kenneth French for making the data publicly available.

forecast in the first quarter of 1963, the forecasting equation includes 4 autoregressive lags $1962Q1-1962Q4$, 12 lags of each macro predictor from January 1962 to December 1962 and lags of financial predictors starting at 12/31/1962 and counting backwards until the 222^{th} lag of predictor *RET*, 224^{th} lag of *SMB, HML, MOM*, 246^{th} lag of *IRS* and 365^{th} lag of *DFF*.

2.4.2 Estimation Scheme and forecasting models

The accuracy of quantile estimates requires a large degree of freedom. Thus, constrained by our sample size, we leave a major part of data for estimation and a smaller portion for out-of-sample analysis. In other words, given 197 observations[14] of *GDP* growth, we use the first 152 observations from $1963Q1$ to $2000Q4$ as an initial estimation sample. For longer forecast horizons $h = 3$ and 6, the initial estimation windows are shortened, ranging from $1962Q4 + h$ to $2001Q1 - h$. This estimation sample is periodically expanded as the out-of-sample forecast moves forwards. To facilitate the comparison with the literature and also isolate the effects of the 2008 financial crisis on forecasting accuracy, we first focus on the $2001Q1-2008Q4$ out-of-sample (OOS) period and then, as a robustness check, we conduct an OOS analysis for the $2008Q1 - 2012Q1$ post-crisis period.

In order to explore the value of high-frequency and distributional information in improving out-of-sample forecasts, we consider 5 different forecasts. The first two corresponds to $f_{t+h,t}$ and $f^{MIDAS}_{t+h,t}$ described by Equations (2.11) and (2.12), respectively. The next two models, labeled $y_{t+h,t}$ and $y^{MIDAS}_{t+h,t}$, are obtained by using *OLS* to estimate location models, where the soft (hard) thresholding and *MIDAS* approaches are used to solve the parameter proliferation problem. Hence, $y_{t+h,t}$ and $y^{MIDAS}_{t+h,t}$ are especial cases of $f_{t+h,t}$ and $f^{MIDAS}_{t+h,t}$ respectively in the sense that they do not take advantage of distributional information to enhance forecasting accuracy.

[14]Recall that we include 4 autoregressive lags of y_t. So the dependent variable starts at $1963Q1$.

Finally, the last model is an autoregressive model with 4 lags, *AR(4)*[15] which does not use both high-frequency and distributional information, and therefore will be used as a benchmark in our forecasting analysis. The idea is that if neither high-frequency nor distributional information is important to enhance forecasting accuracy of *GDP* growth rates, then forecasts based on *AR(4)* would outperform forecasts that rely on these two additional pieces of information. In sum, there are two models that take advantage of both high-frequency and distributional information ($f_{t+h,t}$ and $f_{t+h,t}^{MIDAS}$); two models that only take advantage of high-frequency information ($y_{t+h,t}$ and $y_{t+h,t}^{MIDAS}$); and one model that rely on neither high-frequency nor distributional information, *AR(4)*.

Since there are 6 high-frequency financial predictors, we will index the forecasting models using the acronym of each of these predictors. For example, $f_{t+h,t}$:*EN-RET* refers to a forecasting model that uses *EN* to select lags of the financial predictor *RET* as well as lags of macro predictors. Likewise, $f_{t+h,t}$:*LASSO-RET* refers to a forecasting model that uses *LASSO* to select lags of *RET* as well as lags of macro predictors. In case X_t^M contains only lags of macro predictors, the forecasts based on the quantile combination approach will be labelled as $f_{t+h,t}$:*LASSO-MAC* and $f_{t+h,t}$:*EN-MAC*, respectively. The same indexing will be used for the forecasting models "$y_{t+h,t}$". For *MIDAS* models, we apply the data transformation $\tilde{X}_t^j = QX_t^j$ only to financial predictors (Andreou et al., 2013; Pettenuzzo et al., 2014). The monthly predictors are aggregated to a quarterly frequency, from which two common factors are computed using principal components and two lags of these factors are included into the forecasting equation. This is a standard procedure for *MIDAS* models in the forecasting literature. Finally, in order to include distributional information, we combine the quantile forecasts at levels $\tau = (0.05, 0.06, \ldots, 0.95)$ with equal weights.

[15]It is a common approach to use standard autoregressive model (AR) as benchmark in this literature (Andreou et al., 2013; Pettenuzzo et al., 2016). Specifically, it is estimated from $y_{t+h} = \mu^h + \sum_{i=0}^{q-1} \mu_i^h y_{t-i} + \epsilon_{t+h}$, where q is the number of autoregressive lags and $q = 4$ for our $AR(4)$ model.

2.4.3 Forecast Evaluation

To evaluate the performance of different models in forecasting GDP growth rate, we compute the root mean squared forecast error ($RMSFE$) of each model $\tilde{y}^j_{t+h,t} \in (f_{t+h,t},\ y_{t+h,t},\ f^{MIDAS}_{t+h,t},\ y^{MIDAS}_{t+h,t})$, relative to the benchmark $AR(4)$ model, $\hat{y}_{t+h,t}$. The $RMSFE$ is calculated as follows:

$$RMSFE^h_j = \frac{\sqrt{\sum_{t=1}^{T^*}(y_{t+h} - \tilde{y}^j_{t+h,t})^2}}{\sqrt{\sum_{t=1}^{T^*}(y_{t+h} - \hat{y}_{t+h,t})^2}} \tag{2.13}$$

where T^* is the number of h-step ahead out-of-sample (OOS) forecasts. If the value of $RMSFE^j$ is lower than 1, then model j outperforms the benchmark model in terms of RMSE, producing better GDP growth forecasts.

To test whether a conditional model j produces significantly better forecasts, we test the null hypothesis of equal predictability proposed by Clark and West (2007). We choose this test because the benchmark $AR(4)$ is nested by all other forecasting models[16].

2.4.4 Empirical Results

In this section, we report our results for the $2001Q1 - 2008Q4$ out-of-sample period. Table B.4 summarizes the results for 3 forecast horizons $h = 1, 3, 6$. The odd columns $1, 3, 5$ show the root mean squared forecast error ($RMSFE$) of each conditional model relative to the benchmark *AR(4)*. If the value of $RMSFE$ is less than 1, it indicates that the underlying model outperforms the benchmark during the out-of-sample period. To test whether a conditional model can produce significantly better forecasts than the benchmark model, we rely on the Clark and West (2007) test. P-values of one-sided test are reported in even columns, $2, 4, 6$. A p-value lower than 0.1 suggests

[16]The test by Diebold and Mariano (2012) is designed to compare non-nested models. If the forecasting models are nested, then the DM test may be undersized under the null and may have low power under the alternative hypothesis.

that the underlying model produces a significantly better forecast of GDP growth rates over the $AR(4)$ at a 10% significance level.

Panel B of Table B.4 shows the results for forecasting models that do not explore distributional information. One can see that, among those forecasts, $y_{t+h,t}^{MIDAS}$ using DFF, RET or MOM as financial predictors outperform the $AR(4)$ benchmark at almost all forecasting horizons h. It is also clear that the forecasts $y_{t+h,t} : LASSO$ and $y_{t+h,t} : EN$ rarely outperform $y_{t+h,t}^{MIDAS}$. This result suggests that when distributional information is not considered for forecasting, the $MIDAS$ approach can be used to produce the best forecast of GDP growth using only high-frequency information.

Panel A of Table B.4 shows the results for forecasting models that consider both distributional and high-frequency information. When we compare $f_{t+h,t}^{MIDAS}$ in Panel A to $y_{t+h,t}^{MIDAS}$ in Panel B, one sees that incorporating distributional information to the $MIDAS$ model produces a small accuracy gain. Indeed, the average $RMSFE$ of $f_{t+h,t}^{MIDAS}$ across all financial predictors and forecasting horizons is only 2% less than that of $y_{t+h,t}^{MIDAS}$. As we explained in the previous section, the $MIDAS$ model uses the same set of transformed data to forecast different quantiles of y_{t+h}. Since different quantiles may be affected by different high-frequency predictors over time, the data transformation imposed by the $MIDAS$ approach may be too restrictive for quantile forecasting.

On the other hand, the approach proposed in this paper produces a strong forecasting accuracy gain as shown in Table B.4. In fact, the average $RMSFE$ of $f_{t+h,t}$:$LASSO$ ($f_{t+h,t}$:EN) across all financial predictors and forecasting horizons is 16% (18%) less than that of $y_{t+h,t}$:$LASSO$ ($y_{t+h,t}$:EN). These results suggest that distributional and high-frequency information can improve forecasting accuracy, but $MIDAS$ may not be the best approach to incorporate these two pieces of relevant information into a forecasting model.

Figure B.3 provides a detailed view of the results presented above. Each row corresponds to a forecasting horizon. In each row, there are three charts, each representing a different forecasting method. The horizontal axis in each chart displays

the high-frequency predictors included in the forecasting models. The dotted (solid) line depicts the average *RMSFE* for the group of $f_{t+h,t}$ ($y_{t+h,t}$) forecasts. For instance, at $h = 1$, the average *RMSFE* for $f_{t+h,t} : EN$ models across all high-frequency predictors is 0.96, so the dotted line starts at 0.96 on the vertical axis of that chart. Figure B.3 points out that, regardless the forecasting horizon, the average performance of $f_{t+h,t}$:*LASSO* and $f_{t+h,t}$:*EN* is better than the average performance of all other forecasts, including $f_{t+h,t}^{MIDAS}$ and the ones that rely only on high-frequency information. Once more, these results provide support to the conclusion that incorporating distributional information and high-frequency predictors into a forecasting model can produce substantial accuracy gains if high-frequency predictors are correctly selected across quantiles.

The above conclusion is further corroborated by the results of the Clark-West (*CW*) test reported in Table B.5. The null hypothesis corresponds to equal predictability of $f_{t+h,t}$ and $y_{t+h,t}$ models across the three major methods. One can see that the p-values of one-sided *CW* test is close to zero when *EN* and *LASSO* are used to select the predictors of the quantile function, but the same does not happen when *MIDAS* is used. As said before, in using the same set of transformed data to forecast different quantiles, the *MIDAS* model impedes the possibility that different quantiles can be affected by different predictors over time. This may be too restrictive for quantile forecasting.

2.4.5 Forecast Combination

This paper covers two different groups of forecasts $f_{t+h,t}$ and $y_{t+h,t}$, which differ in terms of the inclusion of high-frequency predictors and whether or not distributional information is considered for forecasting. Thus, a forecaster will potentially face model uncertainty which could be minimized if he/she adopts a strategy of forecast combination as in Andreou et al. (2013) and Pettenuzzo et al. (2016). Due to sample size restriction, we combine forecasts using equal weights but other weighting schemes

could be considered as, for example, the squared discounted *MSFE* weight used by Andreou et al. (2013). The equal-weight forecast is defined as follows:

$$\tilde{y}^{EW}_{t+h,t} = \sum_{j=1}^{m} \frac{1}{m} \tilde{y}^{j}_{t+h,t}.$$

where $\tilde{y}^{j}_{t+h,t}$ represents a $f_{t+h,t}$ and/or $y_{t+h,t}$ forecast. Thus, our combination includes forecasts from each individual group $f_{t+h,t}$ or $y_{t+h,t}$ as well as from both groups.

Table B.6 displays the *RMSFE* of the forecast combinations relative to the *AR(4)* model. A value less than 1 suggests that the combined forecasts outperform *AR(4)*. The p-value of the Diebold-Mariano (*DM*) test is presented next to each *RMSFE*. If the p-value is lower than 0.1, then we conclude that the forecast combination model produces significantly better forecasts than the $AR(4)$ benchmark over the $2001Q1 - 2008Q4$ out-of-sample period.

Table B.6 suggests that combining forecasts from models that incorporate high-frequency and distributional information, FC-$f_{t+h,t}$, produces sizable accuracy gains relative to a benchmark $AR(4)$ model. A smaller but significant accuracy gain is also obtained if we only combine forecasts that use high-frequency information, FC-$y_{t+h,t}$. Finally, combinations that includes forecasts from both groups, FC-both, also produces sizable accuracy gains, confirming the idea that forecast combination can be seen as a hedge against misspecified models. The analysis of the *DM* test also suggests that the simple equal-weighted forecast combination cannot be statistically worse than the $AR(4)$ benchmark.

2.4.6 Robustness Analysis

In this section, we present results for the $2008Q1 - 2012Q1$ out-of-sample interval, which corresponds to the post-crisis period. Recall that in this period the hypothesis of mean stationarity is not rejected, but the hypothesis of second-order stationarity

is rejected. These results suggest the presence of structural breaks in the quantile function, but an absence of structural breaks in the mean function. Breaks in the quantile function undermine the performance of the quantile combination approach used in this paper as an robust method for M forecasts.

The new results are shown in Figure B.4. Contrary to what we see in Figure B.3, forecasts that explore both distributional and high-frequency information do not always outperform forecast that rely only on high-frequency information. If a break in the quantile function occurs, the combination of different quantile forecasts based on past information could fail to approximate correctly the conditional mean of y_{t+h}. For $h = 1$ and $h = 3$, the $f_{t+h,t} : EN$ forecasts not only outperform the benchmark model but also works better than $y_{t+h,t}$:EN. However, as we move to $h = 6$, the $f_{t+h,t}$:EN performs quite poorly. In sum, during the volatile $2008Q1 - 2012Q1$ period, when economic forecasting becomes much more challenging, the approach that explores high-frequency and distributional information using EN to select the predictors still contribute to improve short-horizon forecasts, but dramatically loses power for longer horizons.

The above result suggests that when the null hypothesis of second-order stationarity is rejected, the quantile combination approach will be a misspecified model of the conditional mean. The forecasting literature suggests that forecast combination can be a solution to model misspecification (Rossi, 2013). For this reason, we compute equal-weighted forecast combinations using $f_{t+h,t}$ and/or $y_{t+h,t}$ for the post-crisis period. The results presented in Table B.7 show a sizable accuracy gain relative to the $AR(4)$ benchmark resulting from the forecast combination. Indeed, when all $f_{t+h,t}$ and $y_{t+h,t}$ forecasts are combined (FC-both), the resulting forecast outperforms the benchmark by a large margin across different forecast horizons. We also see that, for $h = 1$ and $h = 3$, the combination based on forecasts that include distributional and high-frequency information, FC-both, cannot be outperformed by the combination that includes only high-frequency information, FC-$y_{t+h,t}$. These results suggest that a simple equal-weighted average of point forecasts can be used to attenuate the loss

of accuracy caused by structural breaks in the quantile function while highlighting the importance of considering both distributional and high-frequency information for out-of-sample predictability.

2.5 Conclusion

This paper develops an unified solution to address two forms of estimation errors that affect out-of-sample predictive accuracy. Specifically, we minimize the effect of non-Gaussian observations on predictive accuracy by using a combination of quantile forecasts. However, the inclusion of high frequency predictors in a quantile forecasting model is more challenging than in location models. Indeed, we show that the parameter proliferation problem in the context of quantile regression cannot be fully addressed by the *MIDAS* approach. This problem arises because different quantiles are probably affected by different high-frequency predictors over time, making the data transformation proposed by *MIDAS* too restrictive for quantile forecasting. We address this problem by proposing a new data transformation based on the combination of soft and hard thresholding rules suggested by Bai and Ng (2008). The novelty here is that we go one step further by performing such a selection procedure across different quantiles. Thus, this approach incorporates both distributional and high-frequency information into the same forecasting model and hence contributes to simultaneously eliminate two important sources of estimation error that affect out-of-sample predictive accuracy.

Our empirical analysis with *GDP* growth rates suggests that the new approach can improve forecasting accuracy by a substantial margin over short, intermediate and long horizons. Specifically, the average *MSE* of models that explore both distributional and high-frequency information is up to 18% lower than the average *MSE* from forecasts that do not use distributional information. However, the proposed method comes with a price since it does require an absence of structural breaks in the quantile function. This condition is itself more stringent than requiring

an absence of structural breaks in the mean function. When breaks in the quantile function occur (without a break in the mean), we showed that one can interpret the new approach as a "misspecified model" of the conditional mean. In this case, we can combine *MSE* forecasts from these misspecified models to obtain better forecasts. Indeed, by combining *MSE* forecasts from models that explore high-frequency and distributional information, we obtain *GDP* growth forecasts that outperform the benchmark by a large margin. These results support the conclusion that forecast combination can be used to attenuate the loss of forecasting accuracy caused by structural breaks in the quantile function.

In addition to *GDP*, the proposed forecasting model could also be used to predict other economic variables, such as monthly inflation and industrial production(Pettenuzzo et al., 2016). In this case, with relative larger sample sizes, other forms of weighting schemes could be consider to obtain robust point forecasts.

Chapter 3

Wage Returns to Adult Education: New Evidence from Feature-Selection

3.1 Introduction

People recognize higher education's importance (Pascarella et al., 2005). However, most related analyses focus on college returns for traditional students (Kane and Rouse, 1995; Heckman et al., 2006; Belfield et al., 2014; Turner, 2016), or on factors affecting traditional students' enrollment choices (Dynarski, 2003; Lovenheim and Owens, 2014). In contrast, this paper focuses on estimating college enrollment's returns for returning adults who didn't attend college within two years of their high school graduation (Seftor and Turner, 2002; Stenberg and Westerlund, 2008). Therefore, to isolate college enrollment's returns for returning adults from continuing students, we focus exclusively on high school graduates who either never enrolled in college or first enrolled at least two years after high school graduation. The outcome variable is the hourly wage rate, and the treatment variable is years of college enrollment.

As documented in the literature (Leigh and Gill, 1997; Berker et al., 2003), adult education has positive wage effects. For example, based on cross-sectional data,

Leigh and Gill (1997) discover community college's positive returns for returning adults. After 20 years, we have access to longitudinal data on both employment and college-enrollment information for the 1979 cohorts of the National Longitudinal Survey of Youth (*NLSY79*). One important advantage of the panel data is that we can control for unobserved time-invariant factors, such as ability, which could be correlated with the treatment variable or the outcome variable and could cause omitted-variable problems if left in the error term. This control is also impossible for cross-sectional analysis.

A conventional approach to control for these time-invariant variables is the fixed-effect model (Jepsen et al., 2014; Belfield et al., 2014). However, given thousands of individual fixed effects, the model's efficiency is a concern. Literature has proposed several shrinkage-based solutions to solve the efficiency problem in the panel data analysis. Lamarche (2010) discuss the optimal degree of quantile shrinkage among specific effects to minimize the estimated panel data's asymptotic variance. Lu and Su (2016) and Qian and Su (2016) apply the group *LASSO* or adaptive group fused *LASSO* in the panel data's estimations.

Although traditional *LASSO* shrinkage models improve estimation efficiency by reducing covariates' dimensionality, they cannot provide valid causal inference for the shrunken coefficients. Moreover, for the post-selection models, the estimated coefficients potentially suffer from misspecification problems caused by the *LASSO* selection of outcome variable. Inferences based on the feature-selection model are potentially invalid if model-selection mistakes are made. In practice, distinguishing coefficients close to zero from those exactly at zero is very difficult. Excluding the nonzero coefficient variable results in model-selection mistakes, which could contaminate estimations and inferences. In summary, although previous shrinkage-based selection methods address the efficiency problem, they encounter new problems.

This paper introduces another selection mechanism among time-varying control variables and individual fixed effects, the double-selection model proposed by Belloni et al. (2014), to draw valid reference on the treatment variable without sacrificing

estimation efficiency. To account for the correlations among the outcome variable, the treatment effect, and other control variables in the post-double-selection estimation, we include all influential variables: those having significant effects not only on the outcome variable (i.e., hourly wage) but also on the treatment variable (i.e., years of college enrollment).

To compare the double-selection model's performance to that of the fixed-effect model, we first estimate both models based on a training data set and then generate the hourly wage predictions for a reserved out-of-sample set of observations corresponding to each selected individual. In our empirical analysis, the wage predictions obtained from the double-selection model are more accurate than those from the fixed-effect model. Furthermore, based on the double-selection model's estimations, we find college enrollment's significant and positive wage effects for returning adults. On average, one more year's college enrollment can increase returning adults' future hourly wages by about \$1.12, which is about 7.7% higher than the fixed-effect model's estimate.

3.2 Data

The data source is the 79 cohorts of the National Longitudinal Survey of Youth (*NLSY79*), including 12,686 individuals from 1979 to 2011 discontinuously. To facilitate the analysis of college enrollment's wage effects for returning adults, several restrictions are imposed on the sample selection

First, we limit the sample to high school graduates eligible for college enrollment. About 12% individuals are dropped because of this restriction. We also exclude students who completed more than four years of college, including graduate students, law students, and medical students. Furthermore, we distinguish continuing students from non-continuing based on the gap between high school graduation and first-time college enrollment. Students first enrolled in college within two years of high school graduation are categorized as continuing students. Otherwise, they are categorized as

non-continuing students. About 60% of the individuals fall into the non-continuing category. The rest are dropped.

Based on this restricted sample, we generate the treatment variable: the number of years a returning adult is enrolled in college[1]. For time-varying control variables, we generate the enrollment indicator, which is 1 if an individual enrolled in college that year otherwise, the indicator is 0. We also generate a trend index to control for the U-shaped wage curve caused by first-time returning to college, noted as $\frac{1}{K}$, where K indicates the number of years until or since the individual first enrolled in college. For years before college enrollment, the indicator is negative; it is positive for years after enrollment. It equals 1 for the year of first-time college enrollment. If an individual never enrolled in college, it equals 0. In addition, the age of individuals AGE_{it} is included as a control variable.

Since the outcome variable is the hourly wage, we are restricted to observations with non-missing values for the outcome variable. Furthermore, following the literature (Leigh and Gill, 1997), we consider an hourly wage between 1.67 and 100. Unusually high or low wages are dropped. For the purposes of estimation and prediction, we randomly separate observations of an individual: the training data set being two-thirds of the observations and a reserved out-of-sample data set being the remaining one-third. Therefore, to ensure that we have enough hourly wage observations to separate into the two parts, we limit the sample to individuals with at least 6 non-missing wage observations after high school graduation. Also, between high school graduation and first-time college enrollment, we restrict the sample to at least two non-missing wage observations. In the end, the data include 3742 individuals with over $55,000$ observations.

Two figures need to be addressed. Figure C.1 displays the percentage of individuals falling into each category of highest grades completed for the whole sample with $12,686$ individuals. About 5% individuals never finished high school. For over 40%

[1]This variable is calculated based on the reported enrollment status, which could be treated as a lower boundary of the true college-enrollment years because of missing observations.

individuals, the 12^{th} grade was their highest grade completed. Approximately 11% individuals completed two years of college, while about 15% completed four years. Figure C.2 conveys a similar message for the restricted sample with 3,742 individuals. Over 80% of the non-continuing high school graduates never completed one year of college. In other words, fewer than 20% non-continuing students returned to college. Compared to the continuing students, the non-continuing students are much less likely to complete their college education, perhaps because that they are faced with a higher-opportunity cost to return to school. Even returning adults are most likely to complete just one or two years of college. This evidence suggests that substantial differences exist between continuing and non-continuing students in terms of the college-enrollment decision. This evidence also explains why we focus exclusively on the wage effects of non-continuing students' college enrollment.

3.3 Econometric Models

The conventional fixed-effect regression model takes the regular form:

$$w_{it} = \eta_0 ENROLLyears_{it} + \beta_1 ENROLL_{it} + \beta_2 \frac{1}{K}_{it} + \beta_3 AGE_{it} + \alpha_i + \epsilon_{it} \qquad (3.1)$$

where w_{it} stands for the hourly wage received by individual i at year t. $ENROLLyears_{it}$, is the treatment variable of interest, indicating the years of college enrollment recorded by the survey. For example, $ENROLLyears_{it} = 1$ indicates that an individual i has one year of college enrollment records at time t. The dummy variable $ENROLL_{it} = 1$ if individual i was enrolled in college at time t; otherwise, it equals zero, controlling for college enrollment's temporarily negative impact on income and reflecting the opportunity costs of returning to college. Moreover, to capture any U-shaped trend of income following the returning behavior, we add the indicator $\frac{1}{K}_{it}$, where K_{it} represents years until or since the individual first enrolled in college. For those never enrolled, we have $\frac{1}{K}_{it} = 0$. For the first-time enrollment year, we have

$\frac{1}{K}_{it} = 1$. Therefore, this index's value is largest and equal to 1 around the first year of enrollment. For years before first-time enrollment, the index value negatively approximates zero with the years moving backward. In contrast for years after first-time enrollment, the index value positively approximates zero in the years moving forward. AGE_{it} is the individual i's age at time t, an additional time-varying control variable. α_i represents the individual fixed effects. ϵ_{it} is the white noise.

Given 3742 individuals, the selection among time-varying features and individual fixed effects can be quite time-consuming in the selection-based approach. To improve computation efficiency without loss of generality, for each time, we randomly select 1000 individuals from the sample and randomly separate their observations into the training and reserved out-of-sample data sets. Then we use both double-selection and fixed-effect regressions(Jepsen et al., 2014; Belfield et al., 2014) to fit the training data and estimate the returns to years of college enrollment for returning adults. To evaluate the two methods' performance, we compare the out-of-sample predictive accuracy for wages. Specifically, we calculate the mean-squared-forecast-errors ($MSFE$) based on the predicted wages generated from the testing sample. For generalization purposes, we repeat the procedure 100 times.

3.3.1 Double-selection (DS) Regression

Although we could directly estimate Equation 3.1, a selection among time-varying control variables and individual fixed effects could be helpful to improve efficiency, given the individual fixed effects' high dimensions. Furthermore, To draw a valid causal inference on the treatment variable $ENROLLyears_{it}$, following Belloni et al. (2014), we adopt the double-selection procedure, in which we select regarding both the outcome variable w_{it} and the treatment variable $ENROLLyears_{it}$ as shown in

Equation 3.2 and 3.3.

$$ENROLLyears_{it} = X_i'\boldsymbol{\beta_1} + \epsilon_{it} \quad s.t.\alpha_1||\beta_{\mathbf{1}}||_1 + (1-\alpha_1)/2||\beta_{\mathbf{1}}||_2^2 \leq \lambda_1 \quad (3.2)$$

$$w_{it} = X_i'\boldsymbol{\beta_2} + \varepsilon_{it} \quad s.t.\alpha_2||\beta_{\mathbf{2}}||_1 + (1-\alpha_2)/2||\beta_{\mathbf{2}}||_2^2 \leq \lambda_2 \quad (3.3)$$

where $X_i = (ENROLL_{it}, \frac{1}{K}_{it}, AGE_{it}, \alpha_i)'$, including both time-varying control variables and individual fixed effects. $\lambda_1, \lambda_2, \alpha_1, \alpha_2$ are the tuning parameters used in the double-selection procedure [2].

Let I_1 and I_2 denote the sets of selected variables from Equations 3.2 and 3.3, respectively. The union set of selection variables, $I = I_1 \cup I_2$, will be included in the post-double-selection regression:

$$w_{it} = \eta_0 ENROLLyears_{it} + I_i'\boldsymbol{\beta} + \nu_{it} \quad (3.4)$$

Adult education's wage effect is reflected by the estimation and inference of η_0. I_i stands for the selected explanatory variables: either time-varying or time-invariant, or both.

Compared to single-selection with only the outcome regression Equation 3.3, the double-selection is more robust to model-selection mistakes caused by excluding variables with small but nonzero coefficients. Incorporating the union set of selected variables I into the final regression 3.4 helps reduce the omitted-variable problem and also produces the correct inference for the treatment variable $ENROLLyears_{it}$.

3.4 Empirical Results

To explore double-selection's effectiveness among time-varying control variables and individual fixed effects, we construct and compare the wage predictions based on

[2]The choices of the tuning parameters $\lambda_1, \lambda_2, \alpha_1, \alpha_2$ rely on a cross-validation approach minimizing the cross-validated errors. For the cross-validation process, we use 10-fold with repetition 5 times to save computation time. We find that the choices of folds and repetition times do not have a significant impact on the optimal parameters.

models with and without double-selection, respectively. With randomly generated training and reserved out-of-sample data sets, we calculate the mean-squared forecast errors ($MSFE$) as:

$$MSFE = \frac{1}{\sum_i T_i} \sum_{i=1}^{N} \sum_{t=1}^{T_i} (w_{it} - \hat{w}_{it})^2 \tag{3.5}$$

where $N = 1000$ is the number of individuals selected for the training and reserved out-of-sample data sets. T_i is the number of yearly observations for individual i in the reserved out-of-sample data. $\sum_i T_i$ represents the number of observations reserved in the out-of-sample data for N individuals. $\hat{w}_{it}$ and w_{it} are predicted wages and true wages reported for individual i at year t. The larger the value is, the worse the predictions are. For each set of training and reserved out-of-sample data, we generate a pair of $MSFE$ corresponding to the double-selection model and the fixed-effect model. Based on 100 rounds of sample selections, we have 100 discrete values of $MSFE$ as shown in Figure C.3.

In Figure C.3, the red curve represents the fitted distribution of $MSFE$ based on the double-selection (DS) model, while the green curve stands for the fitted distribution of the $MSFE$ from the fixed-effects model (FE) without feature selection. The $MSFE$ of double-selection model is consistently smaller than those without selection under both definitions.

Table C.1 shows a detailed summary of statistics based on the $MSFE$ from the double-selection model and the fixed-effect model without selection. For example, the average value of $MSFE$ based on DS model is about 4.03, while it is 4.84 for FE a model without selection. In other words, on average, the DS model reduces $MSFE$ by 16.8% compared to the FE model. We see similar results for other quartiles. In sum, we conclude that wage predictions based on the double-selection model are more accurate.

For the analysis of returns related to years of college education, we focus on the treatment variable's coefficient estimates $ENROLLyears_{it}$. Figure C.4,shows the coefficient estimates for the treatment variable based on two estimation models:

double-selection and fixed-effect models. A red star corresponds to a significant coefficient estimate, indicating a significant wage effect, whereas a green circle represents insignificant wage effects. Within 100 rounds of sampling, the wage effects on years of college enrollment for returning adults are significant and positive according to both models.

Table C.2 displays the summary statistics for the double-selection model's and the fixed-effect model's estimated wage effects. On average, one more year of adult education can increase future hourly wages from \$1.04 to \$1.12. College enrollment's estimated average return based on the double-selection model is about 7.7% higher than that based on the fixed-effect model.

3.5 Conclusion

In this paper, we study college enrollment's impact on returning adults' wages. Specifically, we are interested in those students who didn't enroll in college within at least two years of high school graduation. After imposing several restrictions on the sample selection for *NLSY79* data, we select 3742 high school graduates as non-continuing high school graduates, among whom about 15% returned to college as returning adults with a degree earned or not. Some of them returned multiple times.

To estimate college enrollment's wage effect with increased efficiency, we apply the most recently proposed double-selection model to select among time-varying control variables and individual fixed effects. To explore the double-selection estimation's effectiveness compared to that of the conventional fixed-effect model, we compare the predictive accuracy for wages based on a reserved out-of-sample set of observations. We find that the double-selection model's wage predictions are much more accurate than those of the fixed-effect model. Based on the double-selection model's coefficient estimates, we see that the years of college enrollmentŚs effects on wages for returning adults are most likely to be positive and significant. On average, one more year's college enrollment can increase future hourly wages for returning adults by about

$1.12. Moreover, although the two models' wage-effect estimates are close to each other, the double-selection model's average return estimate is about 7.7% higher than the fixed-effect model's.

Bibliography

Andreou, E., Ghysels, E., and Kourtellos, A. (2011). Forecasting with mixed-frequency data. *Oxford handbook of economic forecasting*, pages 225–245. 25

Andreou, E., Ghysels, E., and Kourtellos, A. (2013). Should macroeconomic forecasters use daily financial data and how? *Journal of Business & Economic Statistics*, 31(2):240–251. 25, 40, 42, 45, 46

Baffigi, A., Golinelli, R., and Parigi, G. (2004). Bridge models to forecast the euro area gdp. *International Journal of forecasting*, 20(3):447–460. 31

Bai, J., Ghysels, E., and Wright, J. H. (2013). State space models and midas regressions. *Econometric Reviews*, 32(7):779–813. 31

Bai, J. and Ng, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, 146(2):304–317. 26, 33, 34, 36, 37, 48

Belfield, C., Liu, Y. T., and Trimble, M. J. (2014). The medium-term labor market returns to community college awards: Evidence from north carolina. a capsee working paper. *Center for Analysis of Postsecondary Education and Employment*. 50, 51, 55

Belloni, A., Chernozhukov, V., et al. (2011). ?1-penalized quantile regression in high-dimensional sparse models. *The Annals of Statistics*, 39(1):82–130. 2, 7, 8, 34, 35, 36

Belloni, A., Chernozhukov, V., and Hansen, C. (2014). High-dimensional methods and inference on structural and treatment effects. *The Journal of Economic Perspectives*, 28(2):29–50. 51, 55

Berker, A., Horn, L., and Carroll, C. D. (2003). Work first, study second: Adult undergraduates who combine employment and postsecondary enrollment. postsecondary educational descriptive analysis reports. 50

Campbell, J. Y. (2000). Asset pricing at the millennium. *The Journal of Finance*, 55(4):1515–1567. 2

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(4):1509–1531. 12

Christoffersen, P. F. and Diebold, F. X. (1997). Optimal prediction under asymmetric loss. *Econometric theory*, 13(06):808–817. 5

Clark, T. E. and West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of econometrics*, 138(1):291–311. 13, 43, 81, 83

Clements, M. P. and Galvão, A. B. (2008). Macroeconomic forecasting with mixed-frequency data: Forecasting output growth in the united states. *Journal of Business & Economic Statistics*, 26(4):546–554. 31

Clements, M. P. and Galvão, A. B. (2009). Forecasting us output growth using leading indicators: An appraisal using midas models. *Journal of Applied Econometrics*, 24(7):1187–1206. 31

Diebold, F. X. and Mariano, R. (2012). Comparing predictive accuracy. *Journal of Business & economic statistics*. 13, 16, 43, 83

Dow, C. H. (1920). *Scientific stock speculation*. Magazine of Wall Street. 1

Dynarski, S. M. (2003). Does aid matter? measuring the effect of student aid on college attendance and completion. *American Economic Review*, 93(1):279–288. 50

Elliott, G., Gargano, A., and Timmermann, A. (2013). Complete subset regressions. *Journal of Econometrics*. 12

Elliott, G. and Timmermann, A. (2016). *Economic Forecasting*. princeton. 24, 25

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, 33(1):1. 37

Gaglianone, W. P. and Lima, L. R. (2012). Constructing density forecasts from quantile regressions. *Journal of Money, Credit and Banking*, 44(8):1589–1607. 4, 28

Gaglianone, W. P. and Lima, L. R. (2014). Constructing optimal density forecasts from point forecast combinations. *Journal of Applied Econometrics*, 29(5):736–757. 4, 28

Gaglianone, W. P., Lima, L. R., Linton, O., and Smith, D. R. (2011). Evaluating value-at-risk models via quantile regression. *Journal of Business & Economic Statistics*, 29(1). 5

Ghysels, E., Santa-Clara, P., and Valkanov, R. (2004). The midas touch: Mixed data sampling regression models. *Finance*. 27, 31

Ghysels, E., Santa-Clara, P., and Valkanov, R. (2005). There is a risk-return trade-off after all. *Journal of Financial Economics*, 76(3):509–548. 31

Goyal, A. and Welch, I. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4):1455–1508. 2, 3, 11, 12, 14

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, pages 424–438. 5

Granger, C. W. and Newbold, P. (1986). *Forecasting time series*. Academic Press, 2nd edition. 5

Heckman, J. J., Lochner, L. J., and Todd, P. E. (2006). Earnings functions, rates of return and treatment effects: The mincer equation and beyond. *Handbook of the Economics of Education*, 1:307–458. 50

Issler, J. V. and Lima, L. R. (2009). A panel data approach to economic forecasting: The bias-corrected average forecast. *Journal of Econometrics*, 152(2):153–164. 9, 15, 30

Jarque, C. M. and Bera, A. K. (1987). A test for normality of observations and regression residuals. *International Statistical Review/Revue Internationale de Statistique*, pages 163–172. 39

Jepsen, C., Troske, K., and Coomes, P. (2014). The labor-market returns to community college degrees, diplomas, and certificates. *Journal of Labor Economics*, 32(1):95–121. 51, 55

Judge, G. G., Hill, R. C., Griffiths, W., Lutkepohl, H., and Lee, T.-C. (1988). Introduction to the theory and practice of econometrics. 2, 7, 25

Jurado, K., Ludvigson, S. C., and Ng, S. (2015). Measuring uncertainty. *The American Economic Review*, 105(3):1177–1216. 40

Kane, T. J. and Rouse, C. E. (1995). Labor-market returns to two-and four-year college. *The American Economic Review*, 85(3):600–614. 50

Koenker, R. (2005). *Quantile regression*. Number 38. Cambridge university press. 6, 29

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. 9

Kuzin, V., Marcellino, M., and Schumacher, C. (2011). Midas vs. mixed-frequency var: Nowcasting gdp in the euro area. *International Journal of Forecasting*, 27(2):529–542. 31

Kwiatkowski, D., Phillips, P. C., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics*, 54(1-3):159–178. 39, 78

Lamarche, C. (2010). Robust penalized quantile regression estimation for panel data. *Journal of Econometrics*, 157(2):396–408. 51

Leigh, D. E. and Gill, A. M. (1997). Labor market returns to community colleges: Evidence for returning adults. *Journal of Human Resources*, pages 334–353. 50, 51, 53

Lima and Meng (2016). Out-of-sample equity-premium prediction: a quantile combination approach. *Journal of Applied Econometrics*. 25, 28, 29

Lovenheim, M. F. and Owens, E. G. (2014). Does federal financial aid affect college enrollment? evidence from drug offenders and the higher education act of 1998. *Journal of Urban Economics*, 81:1–13. 50

Lu, X. and Su, L. (2016). Shrinkage estimation of dynamic panel data models with interactive fixed effects. *Journal of Econometrics*, 190(1):148–175. 51

Ludvigson, S. and Ng, S. (2010). A factor analysis of bond risk premia. *Handbook of Empirical Economics and Finance*, pages 313–372. 40

Ludvigson, S. C. and Ng, S. (2009). A factor analysis of bond risk premia. Technical report, National Bureau of Economic Research. 35

Ma, L. and Pohlman, L. (2008). Return forecasts and optimal portfolio construction: a quantile regression approach. *The European Journal of Finance*, 14(5):409–425. 2, 7, 25

Manzan, S. (2015). Forecasting the distribution of economic variables in a data-rich environment. *Journal of Business & Economic Statistics*, 33(1):144–164. 8

Mariano, R. S. and Murasawa, Y. (2003). A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics*, 18(4):427–443. 31

Meligkotsidou, L., Panopoulou, E., Vrontos, I. D., and Vrontos, S. D. (2014). A quantile regression approach to equity premium prediction. *Journal of Forecasting*, 33(7):558–576. 2, 7, 20, 22, 25

Pascarella, E. T., Terenzini, P. T., and Feldman, K. A. (2005). *How college affects students*, volume 2. Jossey-Bass San Francisco, CA. 50

Patton, A. J. and Timmermann, A. (2007). Properties of optimal forecasts under asymmetric loss and nonlinearity. *Journal of Econometrics*, 140(2):884–918. 4, 5, 28

Pesaran, M. H. and Timmermann, A. (1995). Predictability of stock returns: Robustness and economic significance. *The Journal of Finance*, 50(4):1201–1228. 19

Pettenuzzo, D., Timmermann, A., and Valkanov, R. (2014). Forecasting stock returns under economic constraints. *Journal of Financial Economics*, 114(3):517–553. 40, 42

Pettenuzzo, D., Timmermann, A., and Valkanov, R. (2016). A midas approach to modeling first and second moment dynamics. *Journal of Econometrics*. 25, 27, 29, 31, 32, 42, 45, 49

Qian, J. and Su, L. (2016). Shrinkage estimation of common breaks in panel data models via adaptive group fused lasso. *Journal of Econometrics*, 191(1):86–109. 51

Rapach, D. and Zhou, G. (2013). Forecasting stock returns. *Handbook of Economic Forecasting*, 2(Part A):328–383. 1

Rapach, D. E., Strauss, J. K., and Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *The Review of Financial Studies*, pages 821–862. 11, 12, 13, 14

Rossi, B. (2013). Advances in forecasting under instability. *Handbook of Economic Forecasting*, 2:1203–1324. 47

Seftor, N. S. and Turner, S. E. (2002). Back to school: Federal student aid policy and adult college enrollment. *Journal of human resources*, pages 336–352. 50

Shapiro, S. S., Wilk, M. B., and Chen, H. J. (1968). A comparative study of various tests for normality. *Journal of the American Statistical Association*, 63(324):1343–1372. 39

Stenberg, A. and Westerlund, O. (2008). Does comprehensive education work for the long-term unemployed? *Labour Economics*, 15(1):54–67. 50

Taylor, J. W. (2007). Forecasting daily supermarket sales using exponentially weighted quantile regression. *European Journal of Operational Research*, 178(1):154–167. 2, 7, 25

Turner, L. J. (2016). The returns to higher education for marginal students: Evidence from colorado welfare recipients. *Economics of Education Review*, 51:169–184. 50

Urzúa, C. M. (1996). On the correct use of omnibus tests for normality. *Economics Letters*, 53(3):247–251. 39

Van de Geer, S. A. (2008). High-dimensional generalized linear models and the lasso. *The Annals of Statistics*, pages 614–645. 8

West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica: Journal of the Econometric Society*, pages 1067–1084. 13

Xiao, Z. and Lima, L. R. (2007). Testing covariance stationarity. *Econometric Reviews*, 26(6):643–667. 39, 78

Zhao, Z. and Xiao, Z. (2014). Efficient regressions via optimally combining quantile information. *Econometric theory*, 30(06):1272–1314. 24, 29, 39

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320. 35, 37

Appendix

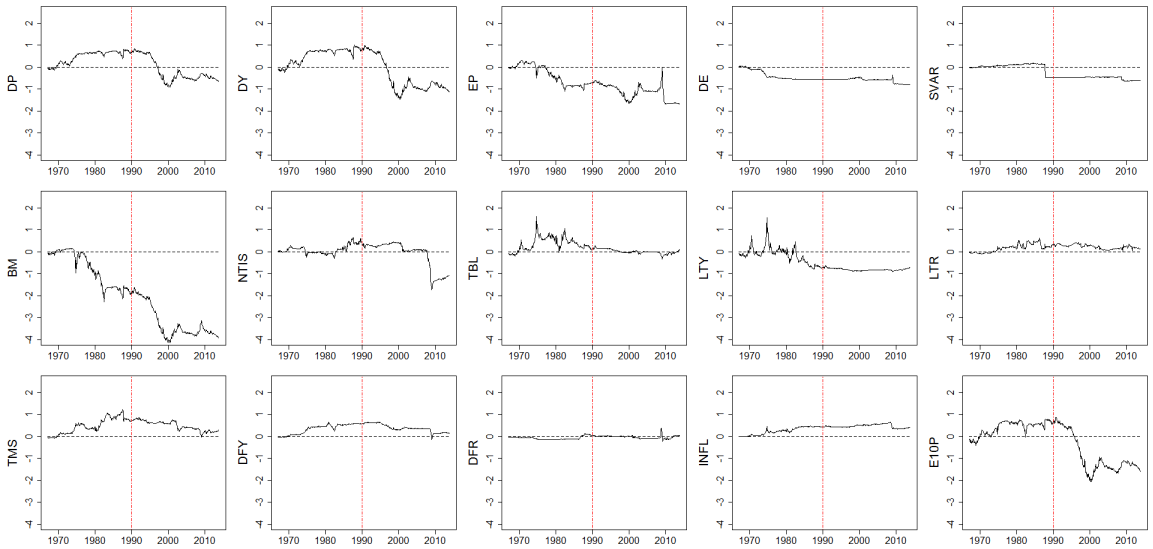

Figure A.1: Cumulative squared prediction error for the benchmark model minus the cumulative squared prediction errors for the single-predictor regression forecasting models, 1967.1-2013.12

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.

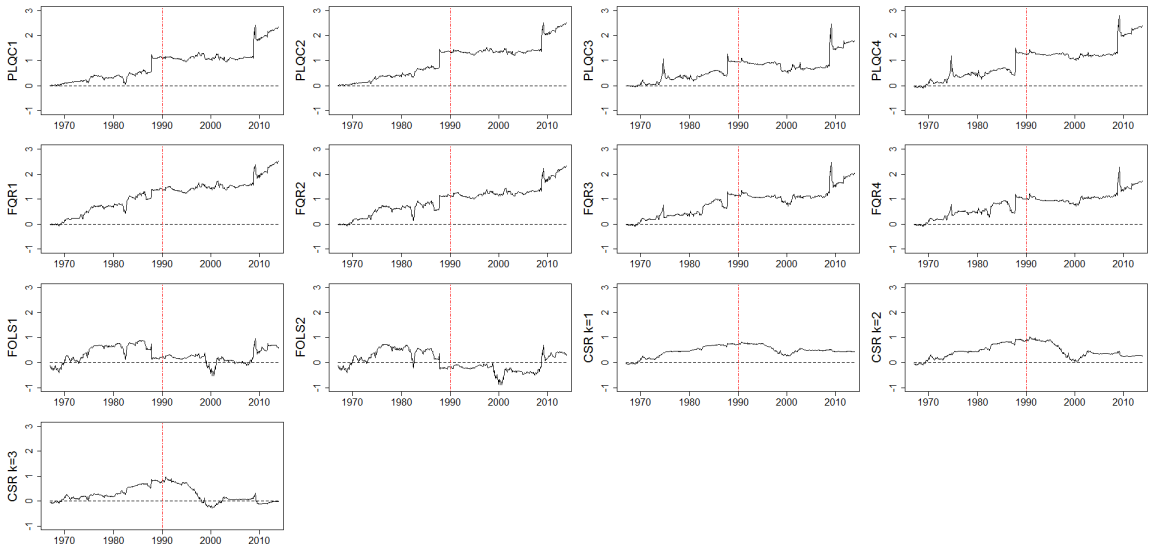

Figure A.2: Cumulative squared prediction error for the benchmark model minus the cumulative squared prediction errors for the *FQR*, *FOLS*, *CSR* and *PLQC* models, 1967.1-2013.12

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.

Table A.1: 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																
*FOLS*1	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
*FOLS*2	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
*FQR*1	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
*FQR*2	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
*FQR*3	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
*FQR*4	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
*PLQC*1	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
*PLQC*2	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
*PLQC*3	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
*PLQC*4	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

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-Mariano (2012) 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.

Table A.2: 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%

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.

Table A.3: Frequency of variables selected over OOS : Jan 1967 − Dec 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%

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

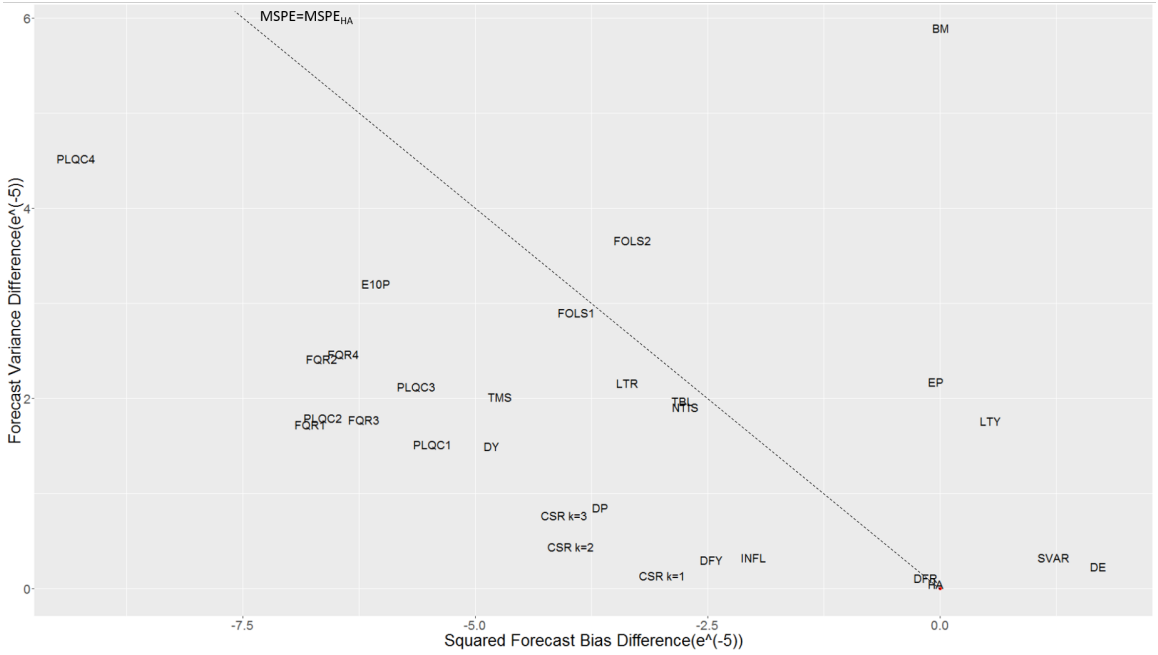

Figure A.3: Scatterplot of forecast variance and squared forecast bias relative to historical average, 1967.1 - 1990.12

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*.

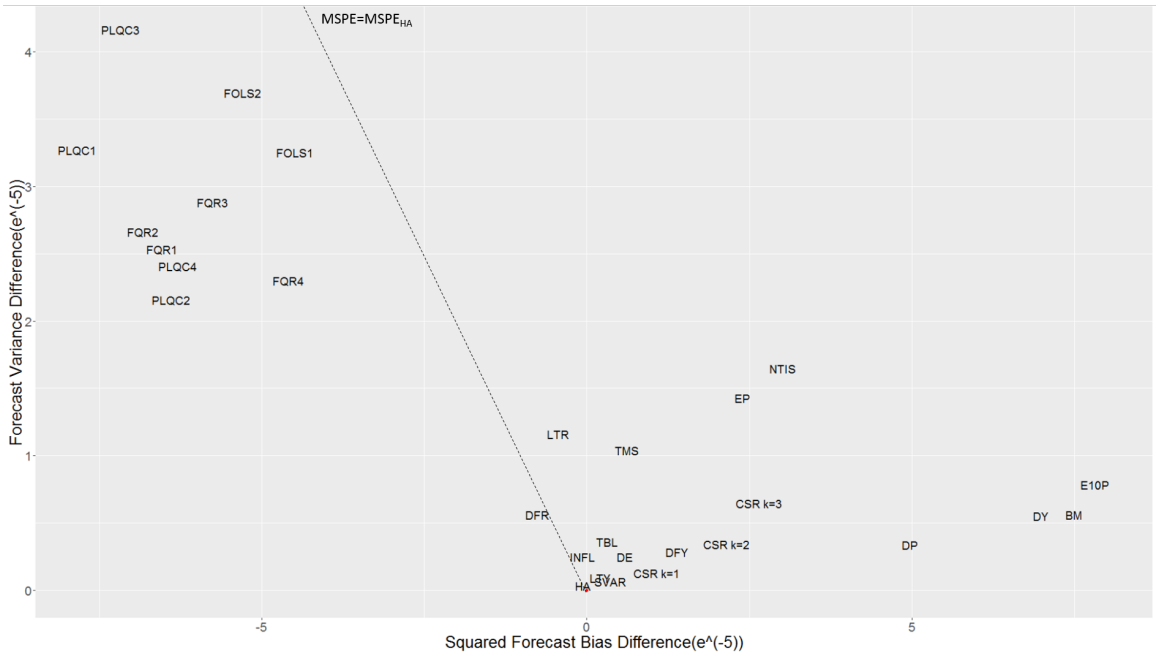

Figure A.4: Scatterplot of forecast variance and squared forecast bias relative to historical average, 1991.1 - 2013.12

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*.

Figure A.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.

Table A.4: Quantile Scores

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

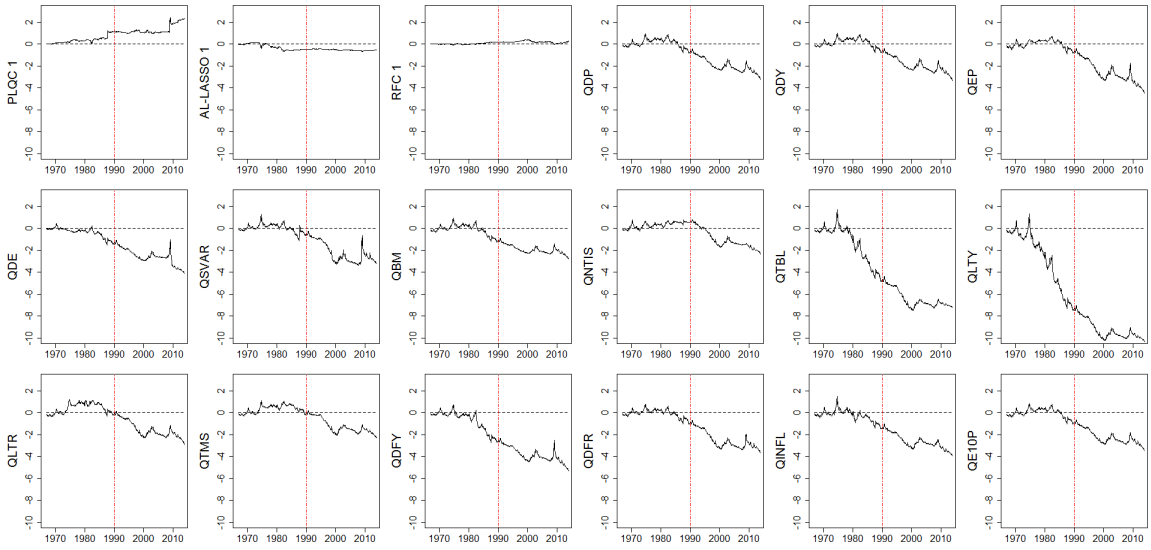

Figure A.6: Cumulative squared prediction error for the benchmark model minus the cumulative squared prediction errors for the *PLQC* 1, AL-LASSO 1, *RFC* 1, and singe-predictor quantile forecasting models, 1967.1-2013.12

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.

Table B.1: Correlation coefficients among randomly selected lag predictors

Lag predictors of macro variables						
	lag1PI	lag2MTinvent	lag4Retailsales	lag5Currency	lag7IPconsgds	lag8Reservesnonbor
lag1PI	1.000	0.994	0.996	0.970	0.979	0.384
lag2MTinvent	0.994	1.000	0.988	0.942	0.993	0.325
lag4Retailsales	0.996	0.988	1.000	0.976	0.973	0.385
lag5Currency	0.970	0.942	0.976	1.000	0.908	0.495
lag7IPconsgds	0.979	0.993	0.973	0.908	1.000	0.263
lag8Reservesnonbor	0.384	0.325	0.385	0.495	0.263	1.000
Lag predictors of a financial variable *IRS*						
Lag	1	41	81	121	161	201
1	1.000	0.825	0.726	0.543	0.596	0.572
41	0.825	1.000	0.892	0.612	0.684	0.644
81	0.726	0.892	1.000	0.757	0.782	0.733
121	0.543	0.612	0.757	1.000	0.742	0.666
161	0.596	0.684	0.782	0.742	1.000	0.901
201	0.572	0.644	0.733	0.666	0.901	1.000

This table demonstrates the correlation coefficients among 6 randomly selected lag predictors of monthly macroeconomic predictors and financial variable *IRS*. As we can see, some of the lag predictors are highly correlated with correlation coefficients close to 1.

Table B.2: Stationarity test results

Out-of-sample periods	Second order (covariance) stationary test		First order (mean) stationary test	
	XL Stats	Critical Value (5%)	KPSS Stats	Critical Value (5%)
$1966Q1 - 2008Q4$	1.976	2.07	0.106	0.463
$1966Q1 - 2012Q1$	**2.129**	2.07	0.194	0.463

Table B.2 shows the results of stationarity test for two periods. We apply Xiao and Lima (2007) to check whether the scale of the distribution is stationary or not. For the mean, we apply the Kwiatkowski et al. (1992) test. Test statistics and critical values at 5% level are reported. As we can see there, mean stationary is not rejected at both periods. But scale (second-order) stationarity is rejected during $1966Q1 - 2012Q1$, but not at $1966Q1 - 2008Q4$

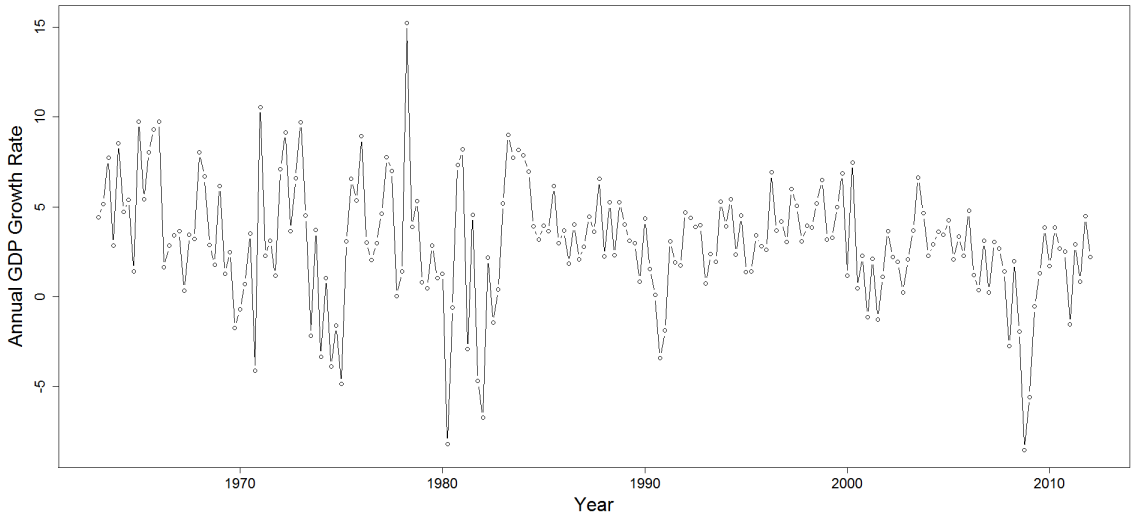

Figure B.1: *GDP* growth rate over $1963Q1 - 2012Q1$

Figure B.1 displays the annualized quarterly GDP growth rate over the period $1963Q1 - 2012Q1$. As we can see, extreme observations appear in 1971, 1978, 1980 and 2008.

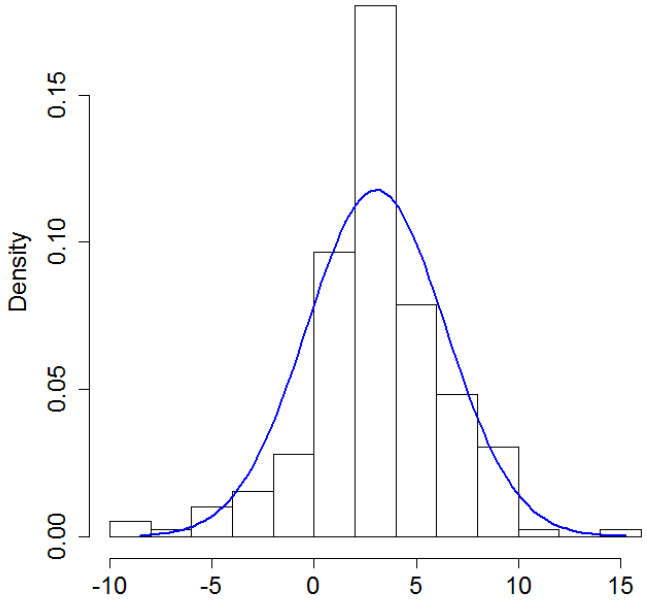

Figure B.2: Histogram of *GDP* growth rate over $1963Q1 - 2012Q1$

Figure B.2 displays the histogram plot of the annual *GDP* growth rate, compared to a normal distribution curve represented by the blue curve.

Table B.3: Description of variables

Variable(s)	Frequency	Range	# of lags	Description	Data sources
GDP growth rate (y_t)	Quarterly	1962Q1 - 2012 Q1	4	Annualized log change of real GDP	Federal Reserve Bank of St. Louis
132 Macro series (X_t^M)	Monthly	1960m1 - 2011m12	12	"output and income", "labor market", "housing", "consumption, orders and inventories", "money and credit" and "bond and exchange rates"	Jurado, Ludvigson, and Ng (2013)
RET	Daily	Jan. 1st,1962 - Dec. 31st, 2011	222	Excess return on the market ($Rm - Rf$)	Kenneth French's website
SMB	Daily	Jan. 1st,1962 - Dec. 31st, 2011	224	The average return difference between 3 small portfolios and 3 big ones	Kenneth French's website
HML	Daily	Jan. 1st,1962 - Dec. 31st, 2011	224	The average return difference between 2 value portfolios and 2 growth ones	Kenneth French's website
MOM	Daily	Jan. 1st,1962 - Dec. 31st, 2011	224	The average return difference between 2 high prior return portfolios and 2 low prior return ones	Kenneth French's website
IRS	Daily	Jan. 1st,1962 - Dec. 31st, 2011	246	10-Year Treasury Constant Maturity minus Federal Funds rate	Federal Reserve Bank of St. Louis
DFF	Daily	Jan. 1st,1962 - Dec. 31st, 2011	365	Effective Federal Funds rate	Federal Reserve Bank of St. Louis

This table describes each variable in details, including its frequency, data range, number of lags included in the forecasting model, data description and sources.

Table B.4: Out-of-sample forecast performance for *GDP* growth over $2001Q1 - 2008Q4$: *RMSFE* and *CW*

	$h=1$		$h=3$		$h=6$	
Models:	*RMSFE*	*CW*	*RMSFE*	*CW*	*RMSFE*	*CW*
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Forecasts based on distributional and high-frequency information						
$f_{t+h,t}$-LASSO-MAC	0.954	**0.08**	0.817	**0.04**	0.801	**0.01**
$f_{t+h,t}$-LASSO-RET	0.919	**0.04**	0.860	**0.04**	0.846	**0.02**
$f_{t+h,t}$-LASSO-SMB	0.925	**0.07**	0.845	**0.04**	0.865	**0.02**
$f_{t+h,t}$-LASSO-HML	0.926	**0.08**	0.843	**0.04**	0.809	**0.01**
$f_{t+h,t}$-LASSO-MOM	0.980	0.15	0.870	**0.06**	0.802	**0.01**
$f_{t+h,t}$-LASSO-IRS	1.059	**0.10**	0.871	**0.03**	0.815	**0.01**
$f_{t+h,t}$-LASSO-DFF	1.079	0.21	0.896	**0.03**	0.802	**0.01**
$f_{t+h,t}$-EN-MAC	0.935	**0.08**	0.797	**0.04**	0.823	**0.02**
$f_{t+h,t}$-EN-RET	0.946	**0.06**	0.801	**0.04**	0.846	**0.02**
$f_{t+h,t}$-EN-SMB	0.923	**0.08**	0.801	**0.04**	0.829	**0.02**
$f_{t+h,t}$-EN-HML	0.944	**0.05**	0.813	**0.03**	0.821	**0.01**
$f_{t+h,t}$-EN-MOM	0.980	**0.10**	0.851	**0.05**	0.840	**0.02**
$f_{t+h,t}$-EN-IRS	1.038	**0.09**	0.849	**0.03**	0.814	**0.02**
$f_{t+h,t}$-EN-DFF	0.987	0.11	0.867	**0.03**	0.822	**0.01**
$f^{MIDAS}_{t+h,t}$-RET	0.905	**0.01**	0.924	**0.04**	0.982	0.17
$f^{MIDAS}_{t+h,t}$-SMB	1.112	0.95	0.993	0.29	0.957	**0.02**
$f^{MIDAS}_{t+h,t}$-HML	1.043	0.66	0.937	**0.05**	1.051	0.22
$f^{MIDAS}_{t+h,t}$-MOM	1.008	0.11	0.939	**0.02**	0.999	0.13
$f^{MIDAS}_{t+h,t}$-IRS	1.056	0.31	0.868	**0.00**	0.867	**0.01**
$f^{MIDAS}_{t+h,t}$-DFF	1.067	0.59	0.929	**0.01**	0.872	**0.01**
Panel B: Forecasts based on high-frequency information only						
$y_{t+h,t}$-LASSO-MAC	1.142	0.15	1.067	0.24	0.901	**0.00**
$y_{t+h,t}$-LASSO-RET	1.073	**0.06**	1.112	0.27	0.982	**0.02**
$y_{t+h,t}$-LASSO-SMB	1.153	0.2	1.073	0.22	0.912	**0.00**
$y_{t+h,t}$-LASSO-HML	1.173	0.19	1.179	0.34	0.969	**0.02**
$y_{t+h,t}$-LASSO-MOM	1.111	**0.10**	1.149	0.32	0.985	**0.00**
$y_{t+h,t}$-LASSO-IRS	1.139	0.17	1.041	0.17	0.886	**0.00**
$y_{t+h,t}$-LASSO-DFF	1.071	**0.09**	1.013	0.17	0.903	**0.00**
$y_{t+h,t}$-EN-MAC	1.013	**0.05**	0.985	**0.06**	1.049	0.13
$y_{t+h,t}$-EN-RET	0.970	**0.01**	1.133	0.13	1.137	0.24
$y_{t+h,t}$-EN-SMB	1.018	**0.06**	0.967	**0.05**	1.206	0.45
$y_{t+h,t}$-EN-HML	0.968	**0.04**	1.064	0.11	1.050	0.11
$y_{t+h,t}$-EN-MOM	0.920	**0.03**	1.129	0.16	1.236	**0.09**
$y_{t+h,t}$-EN-IRS	1.053	**0.08**	0.964	**0.07**	1.033	**0.10**
$y_{t+h,t}$-EN-DFF	1.014	**0.06**	0.978	**0.10**	1.007	**0.09**
$y^{MIDAS}_{t+h,t}$-RET	0.919	**0.01**	0.964	**0.10**	0.976	0.16
$y^{MIDAS}_{t+h,t}$-SMB	1.174	0.94	1.028	0.76	0.946	**0.01**
$y^{MIDAS}_{t+h,t}$-HML	1.080	0.84	0.959	**0.06**	1.065	0.21
$y^{MIDAS}_{t+h,t}$-MOM	1.058	0.29	1.010	0.18	1.074	0.46
$y^{MIDAS}_{t+h,t}$-IRS	1.060	0.27	0.876	**0.00**	0.846	**0.01**
$y^{MIDAS}_{t+h,t}$-DFF	1.067	0.49	0.911	**0.00**	0.846	**0.01**

In this table, odd columns $1, 3, 5$ show the root mean squared forecast error (*RMSFE*) of each conditional model relative to the benchmark model *AR(4)* over the period $2001 - 2008$. If the value of *RMSFE* is less than 1, it indicates that the underlying model outperforms the benchmark during the out-of-sample period. To test whether a conditional model can produce significantly better forecasts than the benchmark model, we rely on the Clark and West (2007) test. P-values of one-side test are reported in even columns $2, 4, 6$. A p-value lower than 0.1 highlights that the underlying model produces a significantly better forecast of *GDP* growth rates over the $AR(4)$ model at a 10% significance level.

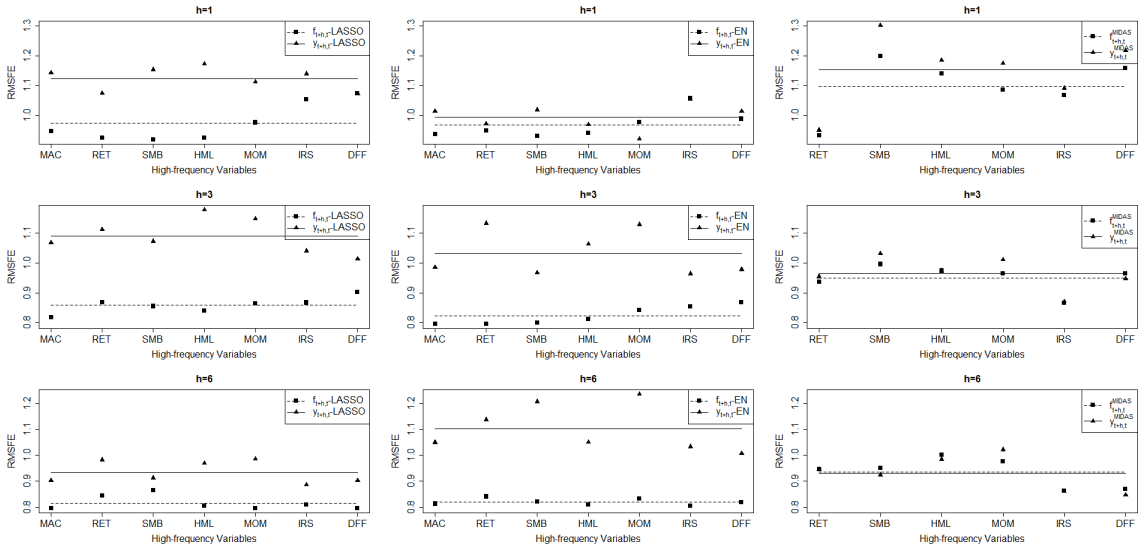

Figure B.3: The *RMSFE* comparison for multi-step ahead forecasts over $2001Q1 - 2008Q4$.

In Figure B.3, each row corresponds to a forecast horizon. In each row there are three charts, each representing a different forecasting method. The horizontal axis in each chart displays the high-frequency predictors included by the forecasting models. The dotted (solid) line depicts the average *RMSFE* for the group of $f_{t+h,t}$ ($y_{t+h,t}$) forecasts. For instance, at $h = 1$, the average *RMSFE* for $f_{t+h,t} : EN$ models across all high-frequency predictors is 0.84, so the dotted line starts at 0.84 on the vertical axis of that chart. The lower the *RMSFE*, the better a model forecasts the *GDP* growth.

Table B.5: Out-of-sample forecasts of *GDP* growth over $2001Q1 - 2008Q4$: p-values of *CW* tests

	h=1	h=3	h=6
LASSO Models			
$f_{t+h,t}$-LASSO-MAC v.s. $y_{t+h,t}$-LASSO-MAC	**0.00**	**0.00**	**0.08**
$f_{t+h,t}$-LASSO-RET v.s. $y_{t+h,t}$-LASSO-RET	**0.00**	**0.00**	**0.01**
$f_{t+h,t}$-LASSO-SMB v.s. $y_{t+h,t}$-LASSO-SMB	**0.00**	**0.00**	**0.03**
$f_{t+h,t}$-LASSO-HML v.s. $y_{t+h,t}$-LASSO-HML	**0.00**	**0.00**	**0.01**
$f_{t+h,t}$-LASSO-MOM v.s. $y_{t+h,t}$-LASSO-MOM	**0.01**	**0.00**	**0.01**
$f_{t+h,t}$-LASSO-IRS v.s. $y_{t+h,t}$-LASSO-IRS	**0.00**	**0.00**	**0.07**
$f_{t+h,t}$-LASSO-DFF v.s. $y_{t+h,t}$-LASSO-DFF	**0.01**	**0.00**	**0.06**
EN Models			
$f_{t+h,t}$-EN-MAC v.s. $y_{t+h,t}$-EN-MAC	**0.00**	**0.00**	**0.00**
$f_{t+h,t}$-EN-RET v.s. $y_{t+h,t}$-EN-RET	**0.00**	**0.00**	**0.00**
$f_{t+h,t}$-EN-SMB v.s. $y_{t+h,t}$-EN-SMB	**0.02**	**0.00**	**0.00**
$f_{t+h,t}$-EN-HML v.s. $y_{t+h,t}$-EN-HML	**0.01**	**0.00**	**0.00**
$f_{t+h,t}$-EN-MOM v.s. $y_{t+h,t}$-EN-MOM	**0.01**	**0.00**	**0.00**
$f_{t+h,t}$-EN-IRS v.s. $y_{t+h,t}$-EN-IRS	**0.01**	**0.02**	**0.00**
$f_{t+h,t}$-EN-DFF v.s. $y_{t+h,t}$-EN-DFF	**0.00**	**0.00**	**0.00**
MIDAS Models			
$f^{MIDAS}_{t+h,t}$-RET v.s. $y^{MIDAS}_{t+h,t}$-RET	**0.03**	**0.02**	0.6
$f^{MIDAS}_{t+h,t}$-SMB v.s. $y^{MIDAS}_{t+h,t}$-SMB	**0.01**	**0.00**	0.79
$f^{MIDAS}_{t+h,t}$-HML v.s. $y^{MIDAS}_{t+h,t}$-HML	**0.01**	**0.06**	0.14
$f^{MIDAS}_{t+h,t}$-MOM v.s. $y^{MIDAS}_{t+h,t}$-MOM	**0.00**	**0.00**	**0.00**
$f^{MIDAS}_{t+h,t}$-IRS v.s. $y^{MIDAS}_{t+h,t}$-IRS	0.25	0.24	0.96
$f^{MIDAS}_{t+h,t}$-DFF v.s. $y^{MIDAS}_{t+h,t}$-DFF	0.42	0.76	1

Table B.5 displays the p-values of one-sided Clark and West (2007) test, under the null hypothesis of equal predictability of $f_{t+h,t}$ and $y_{t+h,t}$ models across the three major methods.

Table B.6: Out-of-sample forecast performance for model combination over $2001Q1 - 2008Q4$

	$h=1$		$h=3$		$h=6$	
Models:	*RMSFE*	*DM*	*RMSFE*	*DM*	*RMSFE*	*DM*
FC-both	0.929	0.21	0.884	0.11	0.875	**0.03**
FC-$f_{t+h,t}$	0.944	0.28	0.843	**0.07**	0.841	**0.03**
FC-$y_{t+h,t}$	0.955	0.31	0.954	0.30	0.920	**0.06**

Table B.6 shows the *RMSFE* of equal-weighted forecast combinations relative to the *AR(4)*. A value less than 1 suggests that the combined forecasts outperform *AR(4)*. We also report the p-values of Diebold and Mariano (2012) test for the null hypothesis that the benchmark model *AR(4)* does at least as well as the forecast combination models. If the p-value is lower than 0.1, we conclude that the forecast combination model produced significantly better forecasts than benchmark during this period.

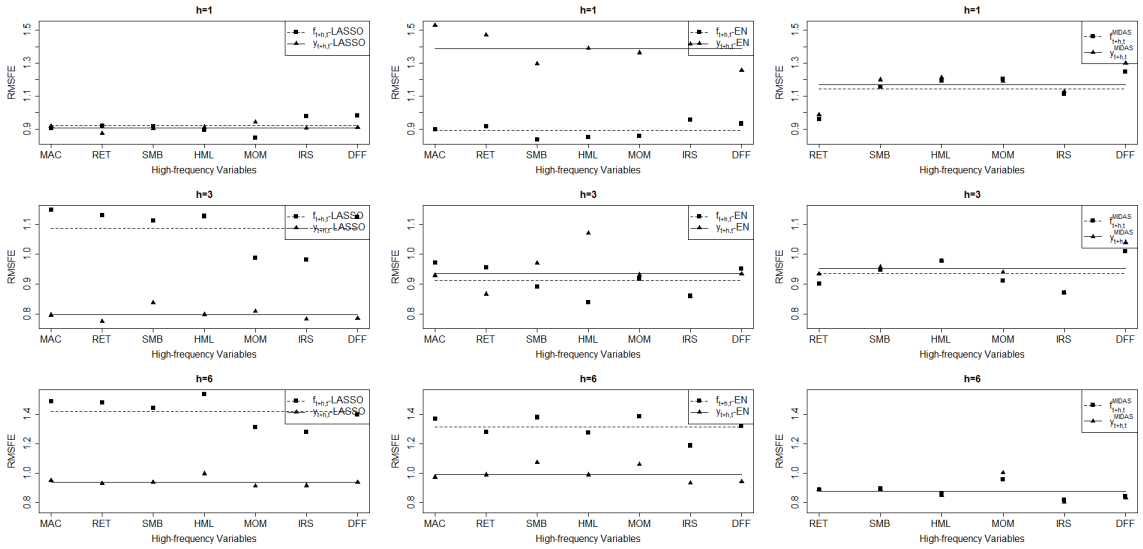

Figure B.4: The *RMSFE* comparison for multi-step ahead forecasts over $2008Q1 - 2012Q1$.

Figure B.4 Each row corresponds to a forecast horizon. Within each row there are three charts, each one showing a different forecasting method. The horizontal axis in each chart displays the high-frequency predictors included by the forecasting models. The dotted (solid) line depicts the average *RMSFE* for the group of $f_{t+h,t}$ ($y_{t+h,t}$) forecasts. The lower the *RMSFE*, the better a model forecasts the *GDP* growth.

Table B.7: Performance of forecast combination models over $2008Q1 - 2012Q1$

	$h = 1$		$h = 3$		$h = 6$	
Models:	*RMSFE*	*DM*	*RMSFE*	*DM*	*RMSFE*	*DM*
FC-both	0.87	0.11	0.77	**0.05**	0.923	0.26
FC-$f_{t+h,t}$	0.87	0.12	0.82	0.12	1.102	0.63
FC-$y_{t+h,t}$	0.92	0.18	0.80	**0.06**	0.878	**0.04**

Table B.7 displays the evaluation results for forecast combination models at the post-crisis periods. It shows results for equal-weighted forecast combinations. These results support the conclusion that forecast combination can be used to overcome the loss of forecasting accuracy caused by structural breaks in the quantile function.

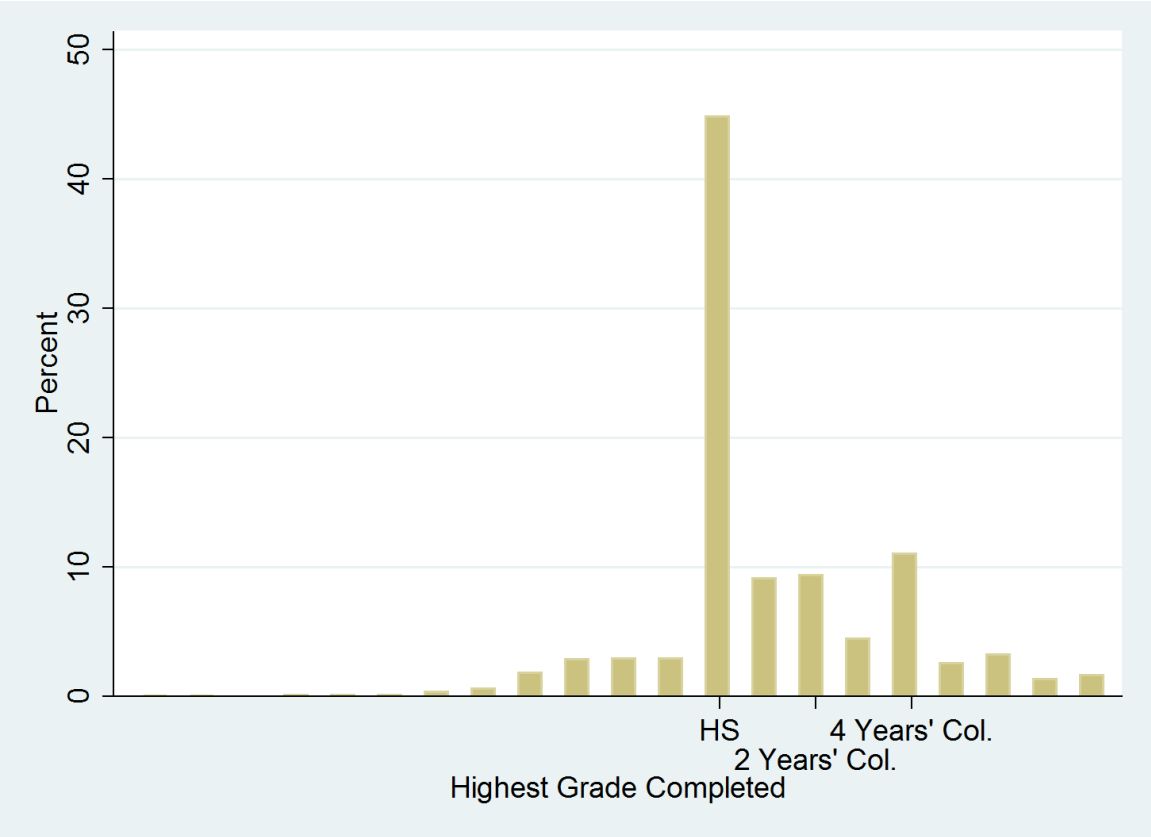

Figure C.1: Histogram for highest grade completed for the whole sample

The histogram Figure C.1 displays the percentage of individuals that fall into each category of highest grades completed for the whole sample with 12,686 individuals. About 5% individuals never finish high school. Over 40% individuals receive the 12^{th} grade as the highest grade completed. About 11% individuals completed two years' college education, while about 15% completed 4 years' college.

Table C.1: Summary statistics of *MSFE*

Stats	DS	FE	% reduction
Mean	4.03	4.84	16.8%
1st-Quantile	3.95	4.75	16.9%
Median	4.01	4.84	17.2%
3rd-Quantile	4.08	4.92	17.0%

Table C.1 shows a detailed summary of statistics based on the *MSFE* from double-selection model and fixed-effect model without selection. For example, the average value of *MSFE* based on *DS* model is about 4.03, while it is 4.84 for *FE* model without selection. In other words, on average, the *DS* model reduces *MSFE* by 16.8% compared to the *FE* model.

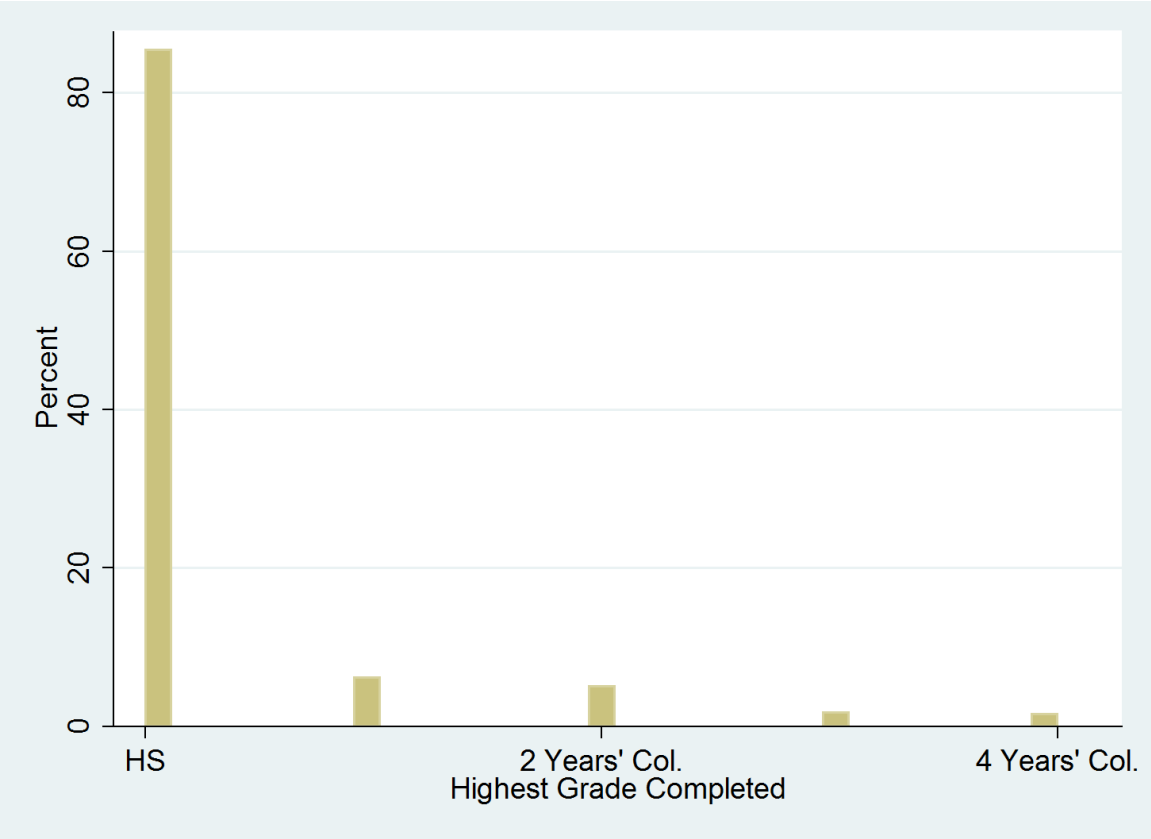

Figure C.2: Histogram for highest grade completed for the restricted sample

Figure C.2 displays the percentage of individuals falling into each category of highest grades completed for the restricted sample with 3,742 individuals. Over 80% of the non-continuing high school graduates never completed even one year's college education. In other words, less than 20% non-continuing students ever returned to college. For those returned and completed some college, they are most likely to complete one year's or two years' college.

Table C.2: Summary statistics of wage effects

Model	DS	FE
Mean	1.12	1.04
1st-Quantile	0.93	0.84
Median	1.13	1.01
3rd-Quantile	1.31	1.22

Table C.2 displays the summary statistics for the estimated wage effects from double-selection and fixed-effect models. On average, the adult education can increase future hourly wages by \$1.04 to \$1.12. Specifically, the estimated average return of college enrollment based on double-selection model is about 7.7% higher than that from the fixed-effect model.

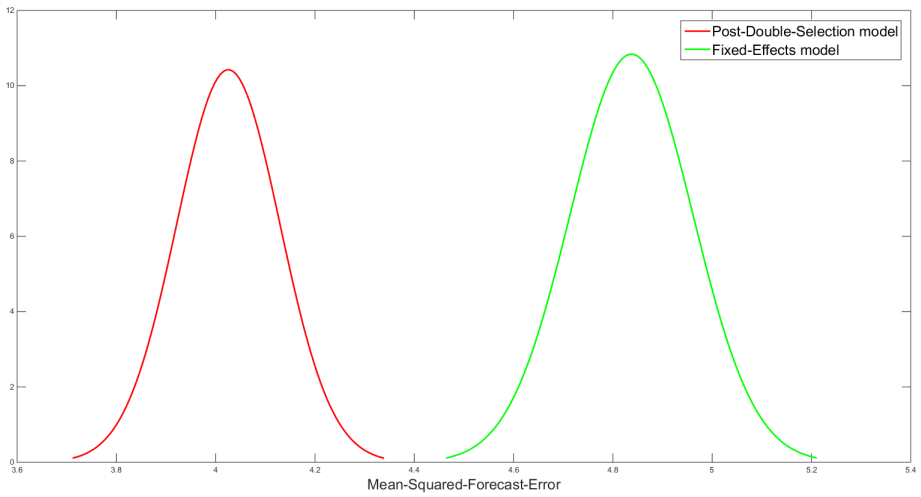

Figure C.3: The fitted distribution of *MSFE* for *DS* and *FE* models

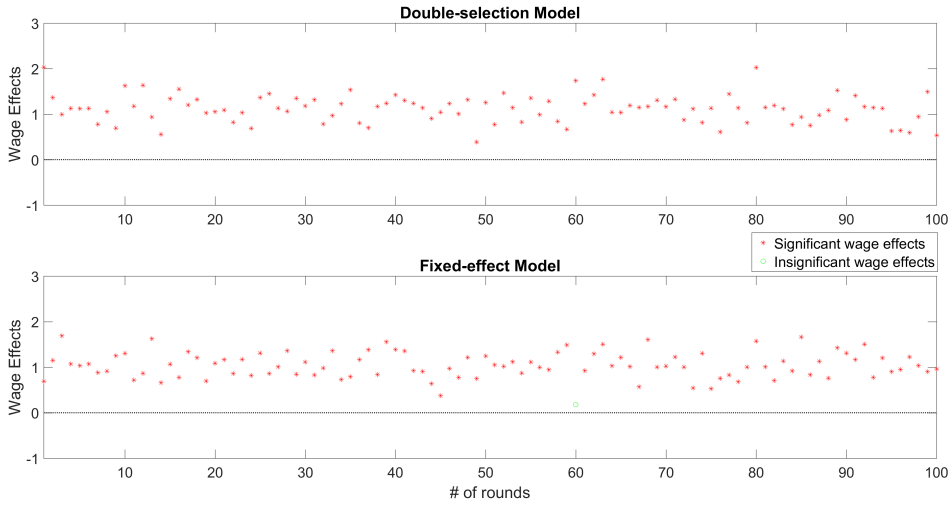

Figure C.4: Wage effects of adult education

Figure C.4 displays shows the coefficient estimates for the treatment variable based on two estimation models, double-selection or fixed-effect models. A red star point corresponds to a significant coefficient estimate, indicating a significant wage effect, whereas green circle points represent insignificant wage effects.

Vita

Fanning Meng was born in Jinzhong, Shanxi, a small town in China. She spent her first 19 years there before she attended the Central University of Finance and Economics in Beijing, China, in Fall 2007. As a undergraduate, her major was international Finance. Within four years of college, she met several influential professors who encouraged her to go abroad for more advanced education. After obtaining the Bachelors of Arts degree in Finance, she went to University of Tennessee, Knoxville and became one of the Economics Ph.D. students in 2011. Since then, she started her 5 years' Ph.D. study in Economics, specifically focusing on International Economics and Economic Forecasting. She is now ready to move forward and apply what she have learnt these years in her future working place.