Distribution Agreement

Predicting GDP growth across different Quantiles

by

Qinwen Wang

Rho Seunghwa, Ph.D.
Advisor

Quantitative Theory & Methods

Rho Seunghwa, Ph.D.
Adviser

Ruoxuan Xiong, Ph.D.
Committee Member

Bree Ettinger, Ph.D.
Committee Member

2022

Predicting GDP growth across different Quantiles

by

Qinwen Wang

Rho Seunghwa, Ph.D.
Advisor

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Quantitative Theory & Methods

2022

Abstract

Predicting GDP growth across different Quantiles

by Qinwen Wang

We study the relationship between real GDP growth and economic and financial conditions across different quantiles. To overcome the invalidity of quantile regression models rising from persistent regressors, we apply IVXQR method that handles size distortion while preserving discriminatory powers. Increasingly severe financial conditions are associated with an increase in conditional volatility, indicating that financial conditions are informational for vulnerability predictions. The GDP growth forecast for a tight economy becomes more conservative after addressing the invalidity for QR methods. To compare out of sample forecast performance across models, we define final prediction error (FPE) using the quantile loss comparison. Generalized Random Forest has the best performance among different methods, implying that splitting the data into splitting and estimation sets would help GDP growth forecast.

Predicting GDP growth across different Quantiles

by

Qinwen Wang

Rho Seunghwa, Ph.D.
Advisor

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Quantitative Theory & Methods

2022

Acknowledgements

I am extremely grateful to my faculty advisor, Dr. Rho Seunghwa, for her consistent support and encouragement through the whole research process. This thesis and the research would not have been possible without the guidance from my advisor. I would also like to thank Dr. Ruoxuan Xiong and Dr. Bree Ettinger for their kind membership, and to my parents for their constant love and support.

Contents

1 Introduction **1**

2 Data **3**

3 Method **4**

3.1 Quantile Regression 4

3.2 Methods robust to the existence of persistent regressor 5

3.3 Tree Based Methods 7

3.3.1 Quantile Regression Forest 7

3.3.2 Generalized Quantile Random Forest 8

4 Results **9**

4.1 Univariate Analysis 9

4.2 Bivariate Analysis 11

5 Out-of-Sample Prediction **19**

5.1 Visualizations 22

6 Data Simulation **27**

7 Conclusion **29**

References **29**

Appendix **31**

1 Introduction

The ability to measure and forecast GDP growth is of great importance because it indicates the general health of the economy and helps policymakers make decisions. Due to its desirable application, many researches have been done to investigate new models and different covariates for GDP estimation. [Marjanović et al., 2016] proposed a model that predicts GDP growth based on carbon dioxide (CO2) emissions, and [Yoon, 2021] presents GDP growth forecasts based on machine learning method using gradient boosting and random forest models. They mostly often focus on point estimates for conditional mean of GDP growth along with other economic variables, which can overly generalize the information given by the data. It is of equal interest for us to understand the relationship between independent variables and a dependent variable at different quantiles, i.e. the behaviors of economies at different economic conditions.

The same idea applies not only to economic forecasting but also to analysis in other fields. For example, we would like to study the factors influencing total medical expenditures for people with low-, medium-, and high- expenditures; or the factors influencing students' grades for those who come from a low-, medium-, and high- income family. To achieve this goal, we apply quantile regression models. Different from traditional linear regressions, quantile regression models the relationship between independent variable and a set of dependent variables across quantiles rather than just the conditional mean.

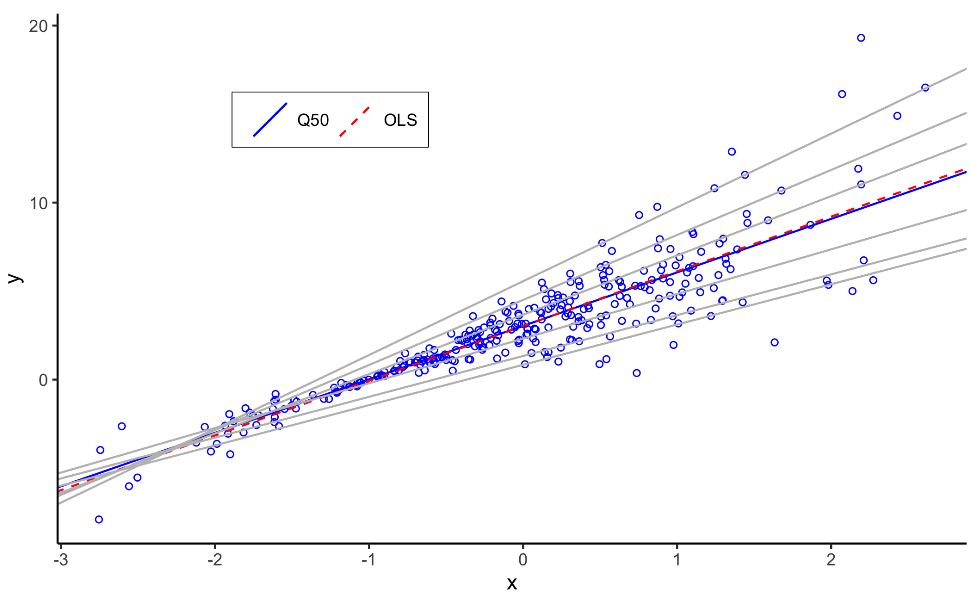


Figure 1: Comparison of OLS and quantile regression

In Figure 1, we visualizes a simple simulation where the independent variables have nonlinear relationships with predictor variables. Gray lines represent different quantile regression lines, and we can see that each of them are very different from the least squares estimate of the conditional mean function (red dashed line). Therefore, it is shown that when outcomes are non-normally distributed, focusing merely on conditional mean may omit valuable information about the data, and thus leading to a poor characterization of the results and an unreliable description of possible impact.

In fact, research into quantile regression has gained pace over the past few years, given its ability of delivering a more comprehensive picture of the effect of the independent variables on the dependent variable. Research in GDP growth forecast using QR methods has found that lower quantiles of GDP growth varies with financial conditions and the upper quantiles is stable over time (see [Adrian et al., 2019]). However, QR methods can be invalid if one or more predictors are persistent because it will lead to size distortion when performing inference. To address the invalidity, we also apply IVX-QR methods proposed by [Lee, 2016] to correct the distortion coming through persistent predictors while preserving discriminatory power. We also generate out of sample estimates using expand window forecast using different models. To compare the performance across different models, we defined final prediction error using out of sample quantile loss comparison.

In this paper, we reassess the relationship between GDP growth and economic and financial conditions across quantiles by (1) using a more recent dataset that has an extended time period, and (2) applying quantile regression, tree based methods and methods robust to persistent regressors. Among all the machine learning methods, we choose to use tree based method because it can capture the non-linear relationship between economies under different conditions. For example, during the recession versus the booming periods, the relationship between financial conditions and the GDP growth can change. Then it is like having different regimes, and such conditions can be better captured by tree based methods because tree based methods divide data into pieces.

Estimates of future GDP growth using different models show asymmetry between upper and lower conditional quantiles using QR and IVXQR methods, which aligns with the main finding in [Adrian et al., 2019] that increasingly severe financial conditions are associated with an increase in conditional volatility. Quantile random forest methods, on the other hand, fails to present the differences of volatility of GDP prediction between upper and lower quantiles. We also find that future forecasts of GDP growth in lower quantiles are more conservative or pessimistic using IVXQR methods. Calculating the final prediction errors across models for different quantiles, we find that generalized random forest seems to have the best performance among all the methods. We then build a data simulation to analyze what factors of dependent and independent variables affect the prediction accuracy.

The rest of the paper is organized as follows. Section 2 explains the data. Section 3 introduces all the methods that we used for analysis. Section 4 presents our results. Section 5 compares the out of sample forecasting performance. Section 6 presents a data simulation. Section 7 concludes.

2 Data

To test whether the general conclusion in [Adrian et al., 2019] still holds with extended period and different methods, we use data (real GDP growth and National Financial Conditions Index (NFCI)) from the same source [a]. NFCI provides a comprehensive weekly estimates on U.S. financial conditions in money markets, debt and equity markets and the traditional and shadow banking systems. The index is constructed as a weighted average of 105 measures of financial activity [b] each expressed relative to their sample averages and scaled by their sample standard deviations. Historically, positive values for NFCI have been associated with tighter-than-average financial conditions, while negative values have been associated with looser-than-average financial conditions. We set the starting point to be January 1973, which is identical to the paper, and extended the ending period from 2015 Q4 to 2019 Q4, which is before the pandemic outbreak. To obtain quarterly NFCI, we also follow the convention of Federal Reserve Economic Data, where weeks that start in one quarter and end in the next one are fully assigned to the latter quarter.

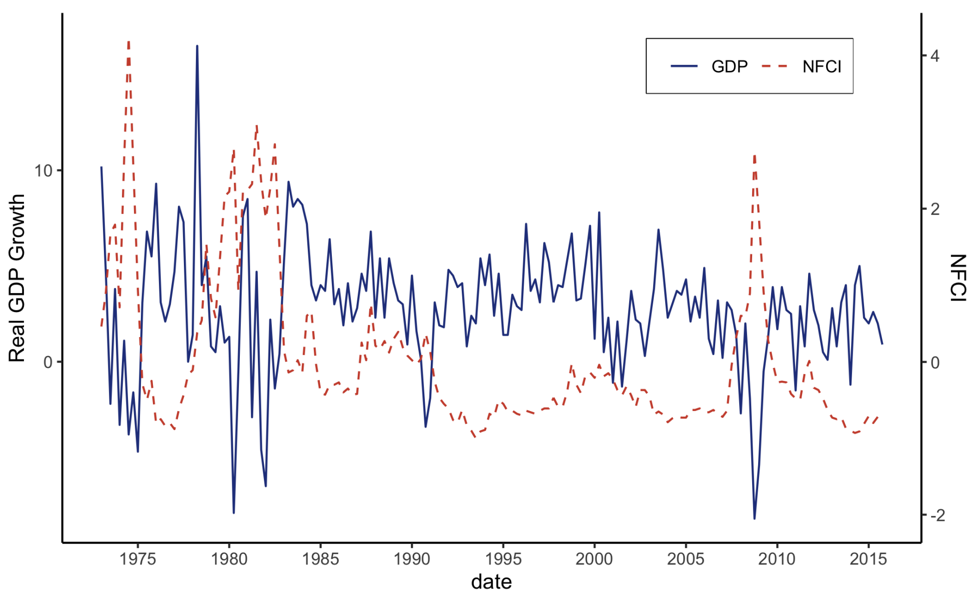


Figure 2: Raw Data

[a]Real GDP is downloaded from FRED `https://fred.stlouisfed.org/series/A191RL1Q225SBEA`. The NFCI data is available at `https://www.chicagofed.org/publications/nfci/index`.

[b]A detailed list of indicators is downloadable at **this link**

Figure 2 show the time series plot for quarterly Real GDP growth and quarterly NFCI. On average, real GDP growth has a higher volatility than that of NFCI. It is also suggested that extreme low values of real GDP growth coincide with high values of NFCI. This makes sense because tight financial conditions can imply low GDP growth. Based on this observation, predictions of future GDP growth using NFCI and past GDP values are meaningful.

3 Method

3.1 Quantile Regression

In this paper, we will be interested in learning different performances across quantiles. In a time series setting, when modeling an h-step ahead dependent variable, the objective function for solving the τ^{th} quantile of y is described by the following equation:

$$Q_{y_{t+h}|x_t}(\tau) = x_t'\beta_\tau \tag{1}$$

Traditional linear regressions estimate the conditional mean by solving the following objective function:

$$\hat{\beta} = argmin \sum_{i=1}^{n} (y_{t+h} - x_t'\beta)^2$$

we want to find β_τ, the τ^{th} quantile of y. To find β_τ, we can start with median ($\beta_{0.5}$). Median can be defined as the solution to a minimization of sum of absolute residuals problem:

$$\hat{\beta}_{0.5} = argmin \sum_{i=1}^{n} |y_{t+h} - x_t'\beta_{0.5}|$$

The symmetry nature of the function means that the minimization of sum of residuals must have equal number of positive and negative residuals, thus making sure that the numbers of observations on both sides of median are the same. Similarly, we can find other quantiles by solving a problem of minimizing sum of absolute residuals but with asymmetry penalties depending on whether residuals are positive or negative.

$$\begin{aligned}\hat{\beta}_\tau &= argmin \sum_{t=1}^{T-h} |y_{t+h} - x_t'\beta_\tau| \cdot \{\tau \cdot \mathbb{1}[y_{t+h} - x_t'\beta_\tau \geq 0] + (1-\tau) \cdot \mathbb{1}[y_{t+h} - x_t'\beta_\tau < 0]\} \\ &\equiv \rho_\tau(y_{t+h} - x_t'\beta_\tau) \\ \text{where} \quad \rho_\tau(u) &= \sum_{i=1}^{n} u_i(\tau - \mathbb{1}[u_i < 0])\end{aligned}$$

In general, the quantile regression penalizes with $\tau|y_i - x_i'\beta_\tau|$ for underpredictions, and penalizes with $(1-\tau)|y_i - x_i'\beta_\tau|$ for overpredictions.

3.2 Methods robust to the existence of persistent regressor

The usual QR method is not valid if one or more predictors are persistent for it leads to size distortion when performing inference. When the NFCI process is approximated with an AR(1) process, the AR parameter estimate for the period of 2013 - 2019 is around 0.88, indicating that NFCI is highly correlated across time. In this section, we apply two newly developed methods proposed by [Lee, 2016] and [Cai et al., 2022], which are robust to the presence of persistent regressors to find out whether the results found in [Adrian et al., 2019] is still maintained. In addition, we compare the out of sample forecasting performance across different models in section 5.

To address the invalidity coming through persistent regressors, we adopt the recent IVX-QR method proposed by [Lee, 2016]. IVX-QR adopts IVX filtering method proposed by [Magdalinos & Phillips, 2009] to the quantile regression setting. The IVX filtering method is designed to correct the size distortion without sacrificing much predictive power using an instrumental variable that is computed by taking the middle ground between the first difference and using the original variable without transformation. Specifically, the IVX filtering defines the instrument variable, $\tilde{z}$, in the following way:

$$\tilde{z}_t = R_{nz}\tilde{z}_{t-1} + \Delta x_t, \quad R_{nz} = I_k + \frac{C_z}{n^\delta}, \quad \delta \in (0,1).$$

Here, x_t is a $k \times 1$ vector of persistent regressors, $C_z = c_z I_k$, $c_z < 0$, and $\tilde{z}_0 = 0$. Note that $R_{nz} = 0_k$ implies IV boiling down to the first difference transformation, and $R_{nz} = I_k$, implies IV being the level of the variable without transformation. The R_{nz} computed from data by bounding the Type I error lies between O_k and $-I_k$ and hence the transformation results in a intermediate degree of persistence. We chose the transformation as recommended by [Lee, 2016]: (1) we fix $c_z = -5$, (2) estimate the QR endogeneity $\lambda(\tau) = -cor(\infty[u_{t,\tau} < 0], u_{tx})$ in the sample, and (3) pick δ using the Table A.3 in [Lee, 2014] which is obtained through simulation to achieve the Type I error bounded by 0.075 with maximum power. One of the advantages of IVX-QR method is that the researcher need not distinguish between near unit root and unit processes because both cases would compute the IVs by transforming the predictors such that the IVs have intermediate persistence.

For the model we consider, the relevant moment conditions would be

$$\sum_{t=1}^{n}\left(\tau - \mathbb{1}[y_{t+h} - \hat{\beta}_{0,\tau} - \hat{\beta}_{1,\tau} ggdp_t - \hat{\beta}_{2,\tau} NFCI_t < 0]\right) = o_p(1)$$
$$\sum_{t=1}^{n} ggdp_t \left(\tau - \mathbb{1}[y_{t+h} - \hat{\beta}_{0,\tau} - \hat{\beta}_{1,\tau} ggdp_t - \hat{\beta}_{2,\tau} NFCI_t < 0]\right) = o_p(1)$$
$$\sum_{t=1}^{n} \tilde{z}_t \left(\tau - \mathbb{1}[y_{t+h} - \hat{\beta}_{0,\tau} - \hat{\beta}_{1,\tau} ggdp_t - \hat{\beta}_{2,\tau} NFCI_t < 0]\right) = o_p(1)$$
$$\tilde{z}_t = (1 - 5/n^{\delta})\tilde{z}_{t-1} + \Delta NFCI_t$$

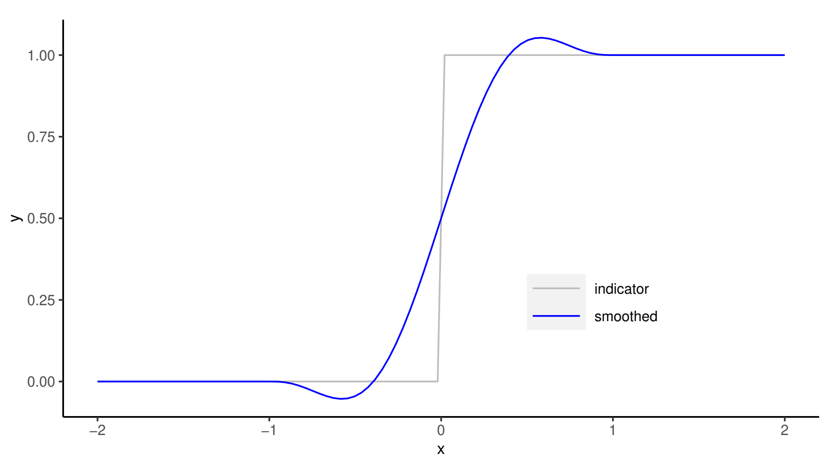
To obtain the $\hat{\beta}_\tau$ that satisfies the above moment conditions, we use [Kaplan & Sun, 2017] (SEE-IVXQR) which smoothes out the moment condition by replacing the indicator function $\mathbb{1}[y_{t+h} - \hat{\beta}_{0,\tau} - \hat{\beta}_{1,\tau} ggdp_t - \hat{\beta}_{2,\tau} NFCI_t < 0]$ with

$$G\left(\frac{\hat{\beta}_{0,\tau} + \hat{\beta}_{1,\tau} ggdp_t + \hat{\beta}_{2,\tau} NFCI_t - y_{t+h}}{h}\right).$$

For example, the indicator function can be approximated by a $G()$ function,

$$G(\nu) = \begin{cases} 0 & \nu \leq -1 \\ 0.5 + \frac{105}{64}\left(\nu - \frac{5}{3}\nu^3 + \frac{7}{5}\nu^5 - \frac{3}{7}\nu^7\right) & u \in [-1,1] \\ 1 & \nu \geq 1 \end{cases}$$

which is plotted below. How smooth would be defined by the bandwidth h. The larger bandwidth h would lead to smoother function which would in turn lead to larger bias and smaller variance. For the details related to the conditions of the $G()$ function, please see [Kaplan & Sun, 2017].



For the current model the smoothed moment conditions would be

$$\sum_{t=1}^{n}\begin{pmatrix}1\\ ggdp_t\\ \tilde{z}_t\end{pmatrix}\left(\tau - G\left(\frac{\hat{\beta}_{0,\tau}+\hat{\beta}_{1,\tau}ggdp_t+\hat{\beta}_{2,\tau}NFCI_t-y_{t+h}}{h}\right)\right)=0.$$

For the bandwidth h, we use the plug-in bandwidth which minimizes the mean squared error of the SEE defined above.

3.3 Tree Based Methods

In this section we describe the Quantile Regression Forest and the Generalized Quantile Random Forest that are used for forecasting. For the current bivariate model, $Y = ggdp_{t+h}$ and $X = (ggdp_t, NFCI_t)$. Since there are only two variables, our case would be identical to bagging rather than the random forest but other than the same variables are considered for each split, the remaining part of the following discussion stays the same.

3.3.1 Quantile Regression Forest

Quantile Regression Forest, proposed by [Meinshausen, 2006], is similar to random forest but would obtain the distribution $F(y|X=x) = Pr[Y \le y|X=x]$ instead of the mean only. To estimate for a new unknown sample, each target value in the training data is given a weight $w_i(x,\theta)$ which has a sum of one. In other words, if it is in the same leaf as the new sample, then the weight is the fraction of samples in the same leaf. If not, then the weight is zero. Formally, the weight can be defined as

$$w_i(x,\theta) = \frac{\mathbb{1}[x_i \in l(x,\theta)]}{\#l(x,\theta)}$$

where $l(x,\theta)$ denotes the leaf that x falls into for tree $T(\theta)$ and $\#l(x,\theta)$ is the total number of observations in the node. Let $w_i(x)$ be the average of $w_i(x,\theta_t)$ over the collection of trees $T(\theta_t)$ in a random forest, $t = 1, \cdots, k$,

$$w_i(x) = k^{-1}\sum_{t=1}^{k} w_i(x,\theta_t)$$

In stead of calculating the mean value of the target variable in each tree leaf, quantile regression forest records all observed responses in the leaf, making percentile calculation feasible. In other words, we can traverse the tree given an input x, and estimate for the predictive distribution by analyzing the distribution of target values present in that leaf. Specifically the distribution would be computed in the following way:

$$\hat{F}\left[y|X=x\right]=\sum_{i=1}^{n}w_i(x)\mathbb{1}[Y_i\leq y]$$

The τ quantile would be obtained by

$$\hat{Q}_{Y|X}(\tau)=min\{y:\hat{F}\left[y|X=x\right]\geq\tau\}.$$

Since our model is in the time series context, it would make more sense to perform the subsampling based on the block sampling method. The performance based on different block sizes are compared using the out of sample quantile loss comparison.

3.3.2 Generalized Quantile Random Forest

The generalized random forest (grf) proposed by [Athey et al., 2016] uses random forest to obtain the neighborhood weight when estimating the moment conditions locally. In the current bivariate model, we are interested in quantile regression model and the relevant moment conditions would be

$$E\left[\left.\begin{pmatrix}1\\ ggdp_t\\ \tilde{z}_t\end{pmatrix}\left(\tau-\mathbb{1}\left[\hat{\beta}_{0,\tau}+\hat{\beta}_{1,\tau}ggdp_t+\hat{\beta}_{2,\tau}NFCI_t-y_{t+h}>0\right]\right)\right|NFCI_t=c_1,ggdp_t=c_2\right]=0$$

To obtain the parameter estimate, the model should be estimated locally around $(NFCI_t = c_1, ggdp_t = c_2)$. Hence when computing the parameter estimates that solve the above moment conditions we should be able to define the weight, $w(NFCI_t, ggdp_t; c_1, c_2)$ which defines the local area of c_1, c_2 This weight would be larger when $(NFCI_t, ggdp_t)$ lies closer to (c_1, c_2). The grf defines whether $(NFCI_t, ggdp_t)$ lies closer (c_1, c_2) or not using random forest. Hence if more trees define $(NFCI_t, ggdp_t)$ to be neighborhood of (c_1, c_2) then the weight would be larger. Once the weight is obtained, the problem would boil down to solving the following sample moment conditions:

$$\frac{1}{n}\sum_{t=1}^{T}w(NFCI_t,ggdp_t;c_1,c_2)\begin{pmatrix}1\\ ggdp_t\\ \tilde{z}_t\end{pmatrix}\left(\tau-\mathbb{1}\left[\hat{\beta}_{0,\tau}+\hat{\beta}_{1,\tau}ggdp_t+\hat{\beta}_{2,\tau}NFCI_t-y_{t+h}>0\right]\right)=0$$

The grf is similar to random forest but it would have separate the training set into splitting and estimation sets. The splitting set would be used to split the trees, and the estimation set would be used to obtain the weights of the neighbors. By separating the splitting and estimation sets, the estimates obtained from grf have asymptotic normality and hence statistical significance can be obtained. The splitting rule of the grf seeks to maximize the heterogeneity in parameter estimates, where the estimate of the child node would be approximated using the estimate and the gradient of its parent node for computational efficiency. This approximation leads to the splitting based on the pseudo-outcomes.

4 Results

In this section, we use quarterly NFCI, real GDP growth, and the combination of them to predict future GDP growth for different quantiles using the methods we introduced in the previous section.

4.1 Univariate Analysis

First, we will be considering the univariate situation where future GDP growth is modeled using only economic or financial conditions. Estimation of τ^{th} quantile of $ggdp_{t+h}$ is defined as:

$$\hat{Q}_{ggdp_{t+h}|ggdp_t}(\tau) = \hat{\beta}_{0,\tau} + \hat{\beta}_{1,\tau} ggdp_t, \text{ and } \hat{Q}_{ggdp_{t+h}|NFCI_t}(\tau) = \hat{\beta}_{0,\tau} + \hat{\beta}_{1,\tau} NFCI_t$$

The following figures show the predictions for one- and four-quarter GDP growth using QR and IVXQR methods with the scatter plot of true value. Figure 3 shows that the slopes are quite similar across quantiles for GDP growth, showing almost parallelism. For NFCI, the slopes differ significantly across different quantiles and from the OLS. The lower quantile is especially steep, whereas the upper quantile has a comparatively small positive slope. This figure suggests that NFCI is more informative than current GDP growth when predicting GDP at different quantiles, especially for a tight economy. Results from IVXQR methods suggest similar conclusions. Different from QR method, Figure 4 suggests that slopes for upper and lower quantiles do not show significant difference from each other when predicting 1 quarter ahead; but show a greater discrepancy from each other when predicting 1 year ahead. Moreover, upper quantile slopes change from positive to negative when replacing QR method with IVXQR. After correcting size distortion for QR methods, the asymmetry between upper and lower quantile for one-quarter GDP growth estimation is reduced, but the differences of slopes across quantiles still remain, if not greater, for one-year ahead forecast.

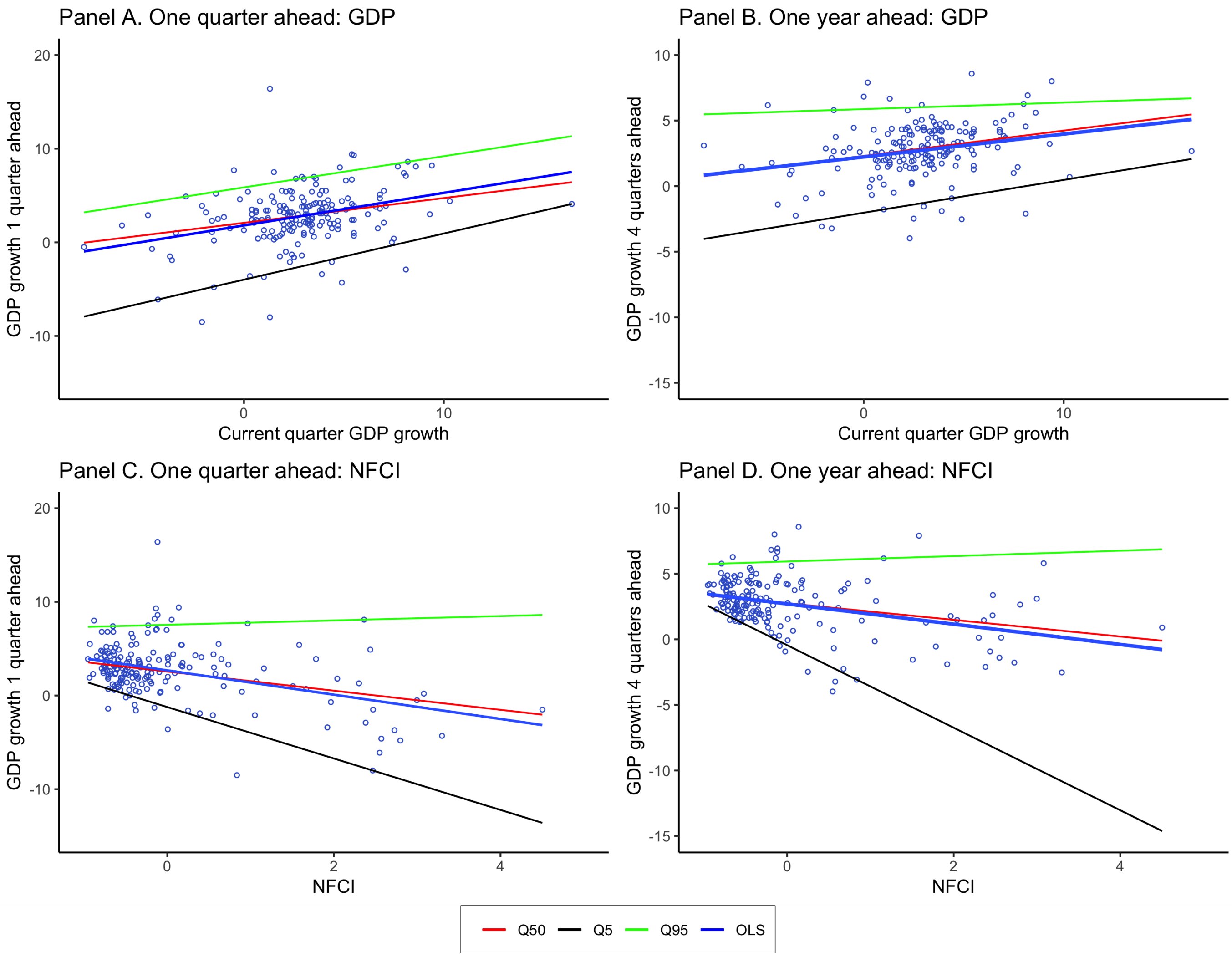
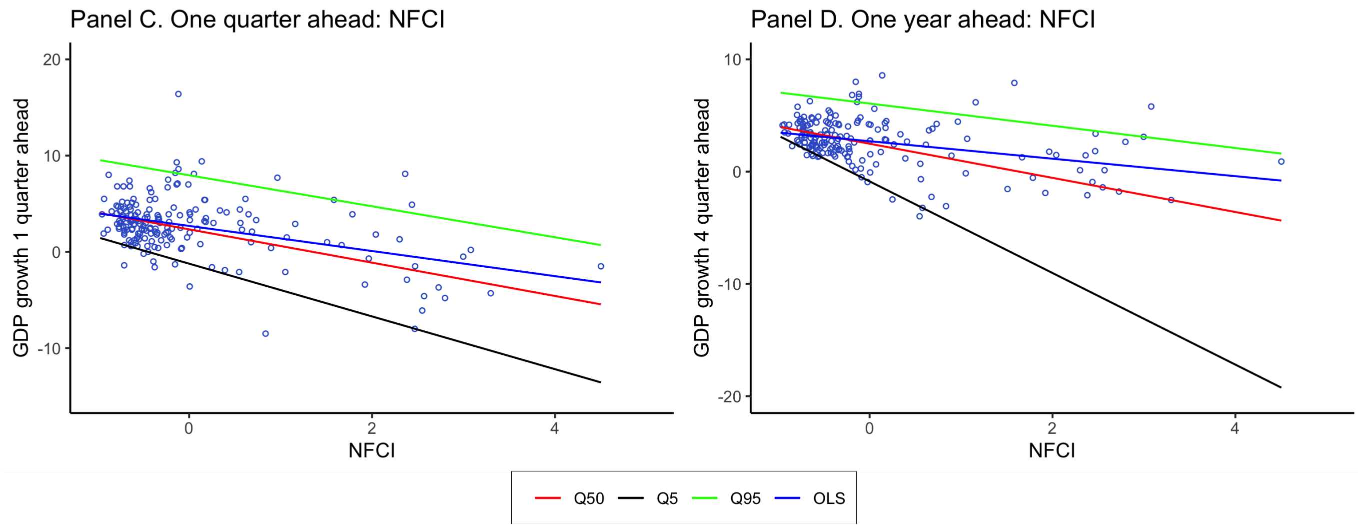
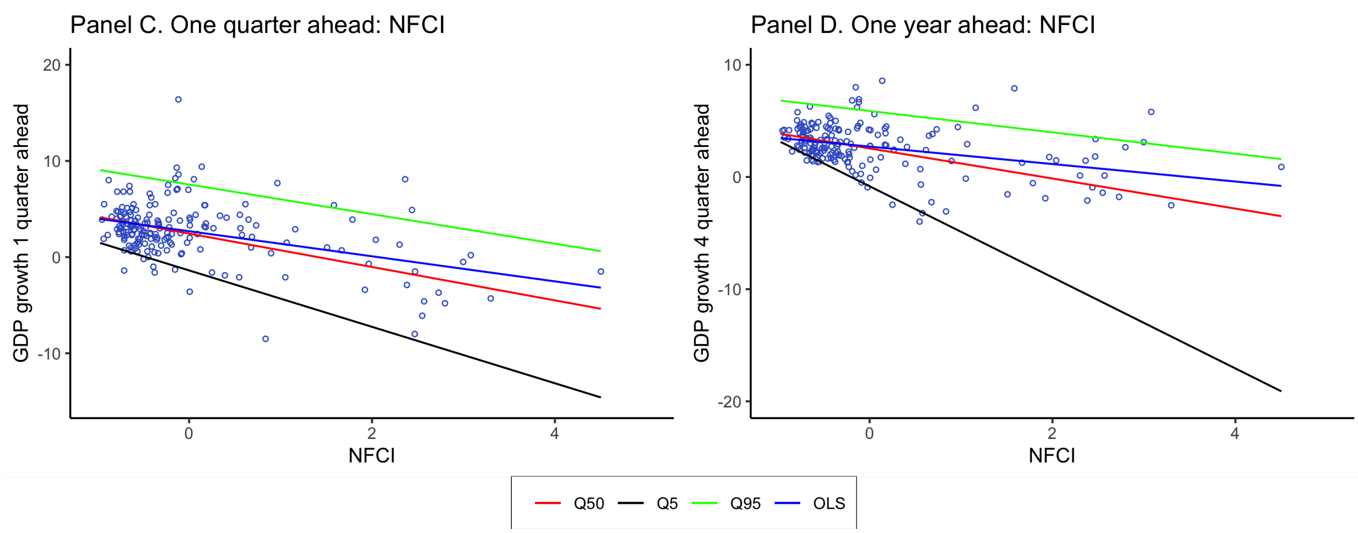


Figure 3: Quantile Regression



(a) ivqr



(b) see

Figure 4: IVXQR

4.2 Bivariate Analysis

Estimation of τ^{th} quantile of $ggdp_{t+h}$ under bivariate condition can be defined as:

$$\hat{Q}_{ggdp_{t+h}|ggdp_t+NFCI_t}(\tau) = \hat{\beta}_{0,\tau} + \hat{\beta}_{1,\tau} ggdp_t + \hat{\beta}_{2,\tau} NFCI_t$$

We start with an analysis on whether estimates for different quantiles are significantly from each other or not. We choose to use 4-order VAR to capture the dynamical relationship of NFCI and GDP under linear relationship assumption. Vector Autoregression with up to four steps jointly is able to obtain not only the variance but also covariance between variables. The model is specified by the following equation:

$$\begin{pmatrix} NFCI_t \\ ggdp_t \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \end{pmatrix} \boldsymbol{X}' + \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

where $\boldsymbol{X} = \begin{pmatrix} 1 & NFCI_{t-1} & ggdp_{t-1} & NFCI_{t-2} & ggdp_{t-2} & NFCI_{t-3} & ggdp_{t-3} & NFCI_{t-4} & ggdp_{t-4} \end{pmatrix}$

Then we are able to calculate the residuals ($\binom{u_1}{u_2}$) and get the covariance matrix($\hat{\Sigma}$) between them. Now we can recover NFCI and GDP growth ($\binom{N\hat{FCI}_t}{gg\hat{d}p_t}$) with linear structure with randomly generated new residuals($\binom{\hat{u_1}}{\hat{u_2}}$) by plugging-in the first four observations of NFCI and GDP growth.

$$\begin{pmatrix} N\hat{FCI}_t \\ gg\hat{d}p_t \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \end{pmatrix} \hat{\boldsymbol{X}}' + \begin{pmatrix} \hat{u_1} \\ \hat{u_2} \end{pmatrix}, \begin{pmatrix} \hat{u_1} \\ \hat{u_2} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \hat{\Sigma}\right)$$

We fit the recovered data into a quantile regression model across 0.05, 0.1, 0.15 ... 0.9 and 0.95 quantiles. Repeat the process for 1000 times and record the coefficents of $ggdp_t, NFCI_t$ for each iterations. For each 1000 recorded $\beta_{j,\tau}$, where $j \in (1, 2)$, we then find their 2.5, 5, 16, 50, 84, 95, and 97.5 percent quantiles to create the confidence interval or the band shown in Figure 5. Median is represented by the black dashed line, OLS is the blue dashed line, and the red dotted line is in-sample fit coefficients.

It is easy to see that in-sample fit coefficients (red line) is stable across quantiles when predicting with current GDP growth, meaning that GDP growth is not informative in predicting future GDP growth at different quantiles. Moreover, the values are mostly positive, which makes sense because we would expect future GDP growth to increase if GDP is growing in the past time periods. Panel C & Panel D, on the other hand, show that the in-sample coefficients are statistically significantly different across quantiles. The in-sample fit coefficients are smaller at lower quantiles and clustered around 0 when $\tau > 0.7$, meaning that financial conditions have more explanatory power when predicting GDP growth for a tight economy than a loose economy. This observation also confirms that the quantile regression is meaningful under current model and estimation from different quantiles are informative.

Then we compare the in-sample fit coefficients between QR and IVXQR methods. Figure 6 shows that the coefficients calculated by using IVXQR methods are close to and generally lower than the coefficients extrapolated from QR method. The conclusion that NFCI is more informative in estimating across quantiles than GDP growth still holds after addressing the invalidity rising from persistent regressor. For NFCI, $\beta\tau$ is more cluster to 0 when $\tau > 0,7$, which supports our argument that changes in financial conditions have relatively less information in predicting upper quantiles of GDP growth, comparing to forecasting for lower quantiles of GDP growth.

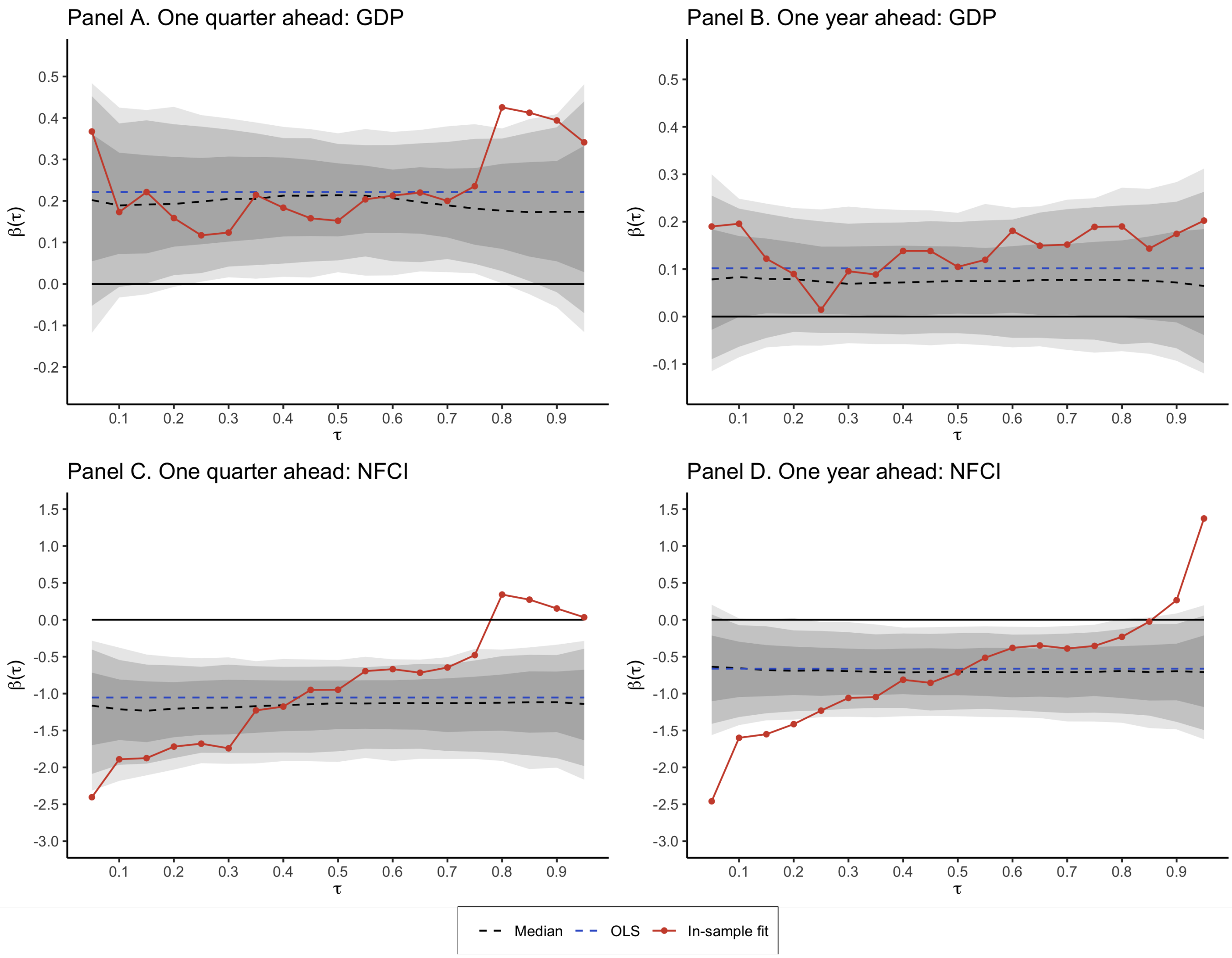


Figure 5: Estimated Quantile Regression Coefficients

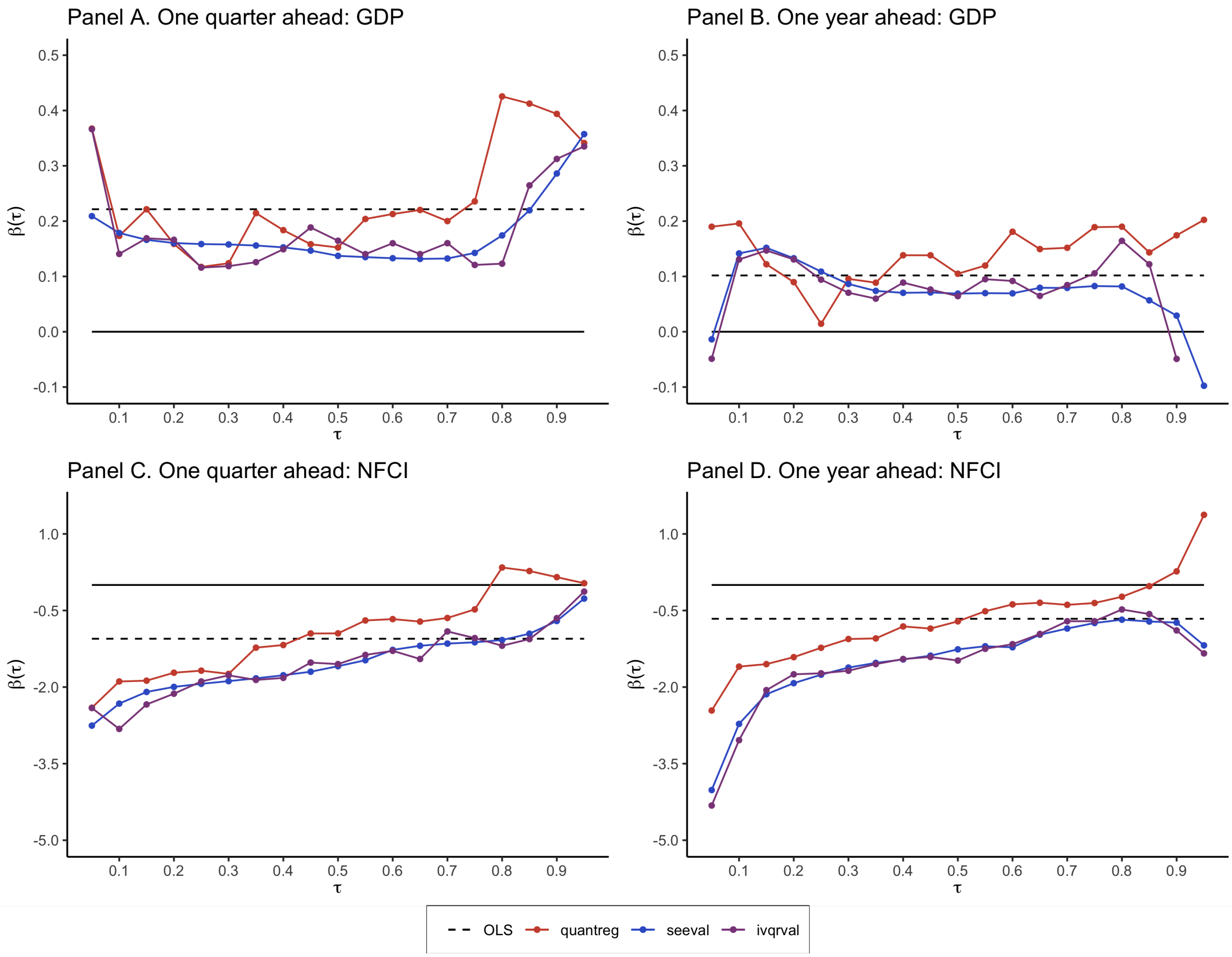


Figure 6: Compare across models

Figure 7 to Figure 11 visualize the predicted distributions of 5, 10, 25, 75, 90 and 95 percent quantile for one- and four-quarter GDP growth using different models. After including extended period, the asymmetry between upper and lower quantiles still holds when using QR methods. Both subfigures in Figure 7, especially one-year-ahead prediction, indicate that the lower conditional quantile predictions have a higher volatility. Figure 8 and Figure 9 imply that Quantile Random Forest does not show difference of volatility in prediction across quantiles. For IVXQR methods, one-year ahead forecast shows a greater asymmetry between upper and lower quantiles than one quarter forecast, which aligns with our finding in Figure 4. The predicted distribution are generally lower in Figure 10 and Figure 11 than that from QR method across quantile, meaning that IVXQR methods provide a more conservative estimates for risk analysis.

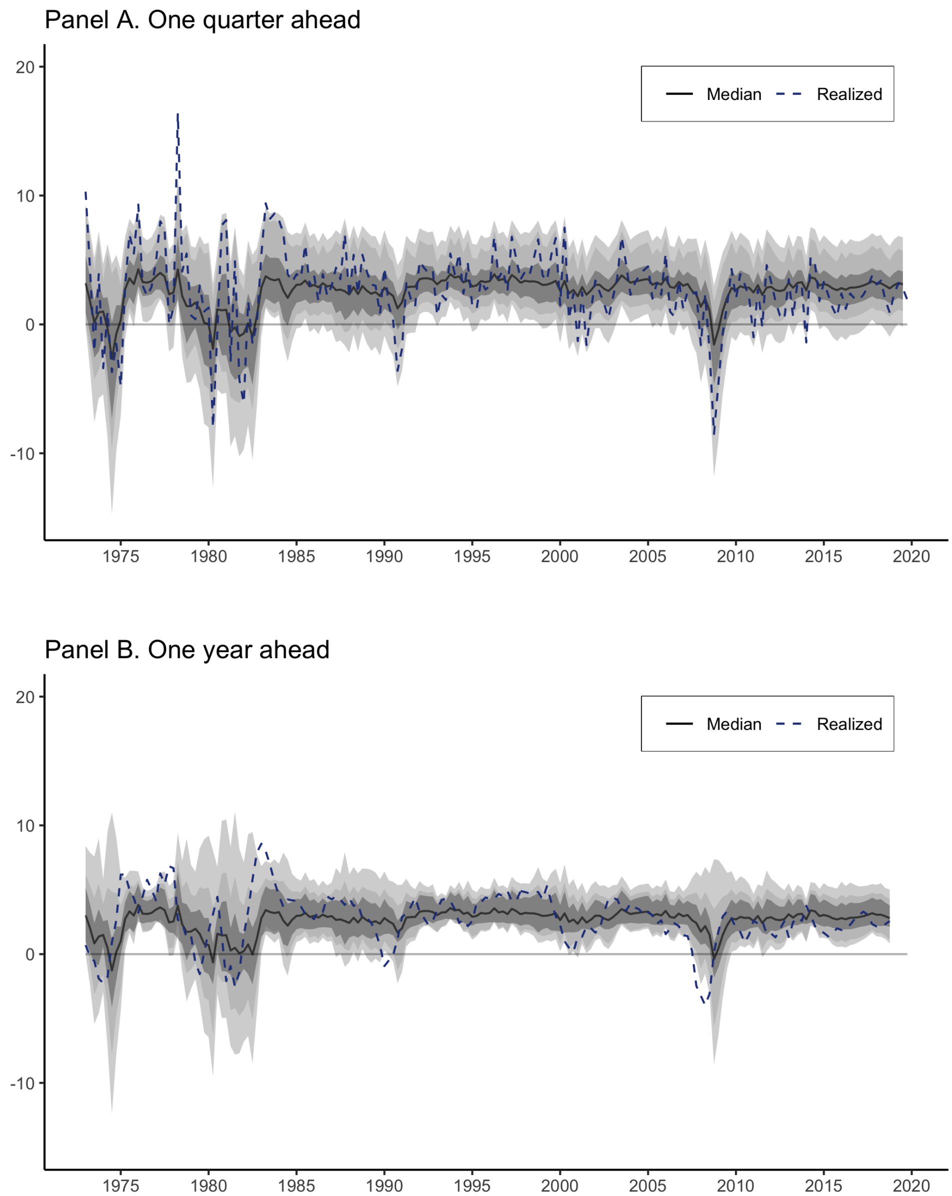


Figure 7: Predicted Distributions using QR

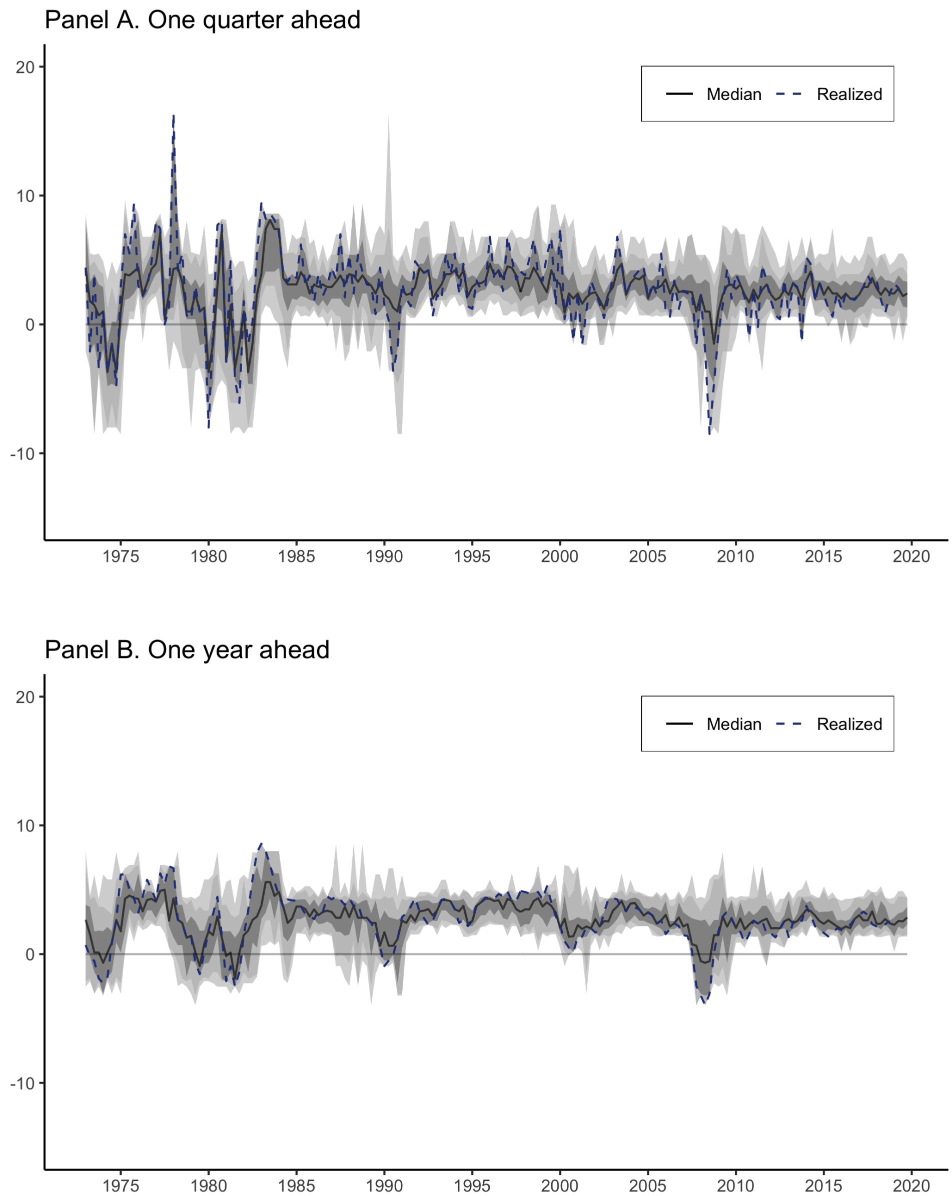


Figure 8: Predicted Distributions using Quantile Random Forest

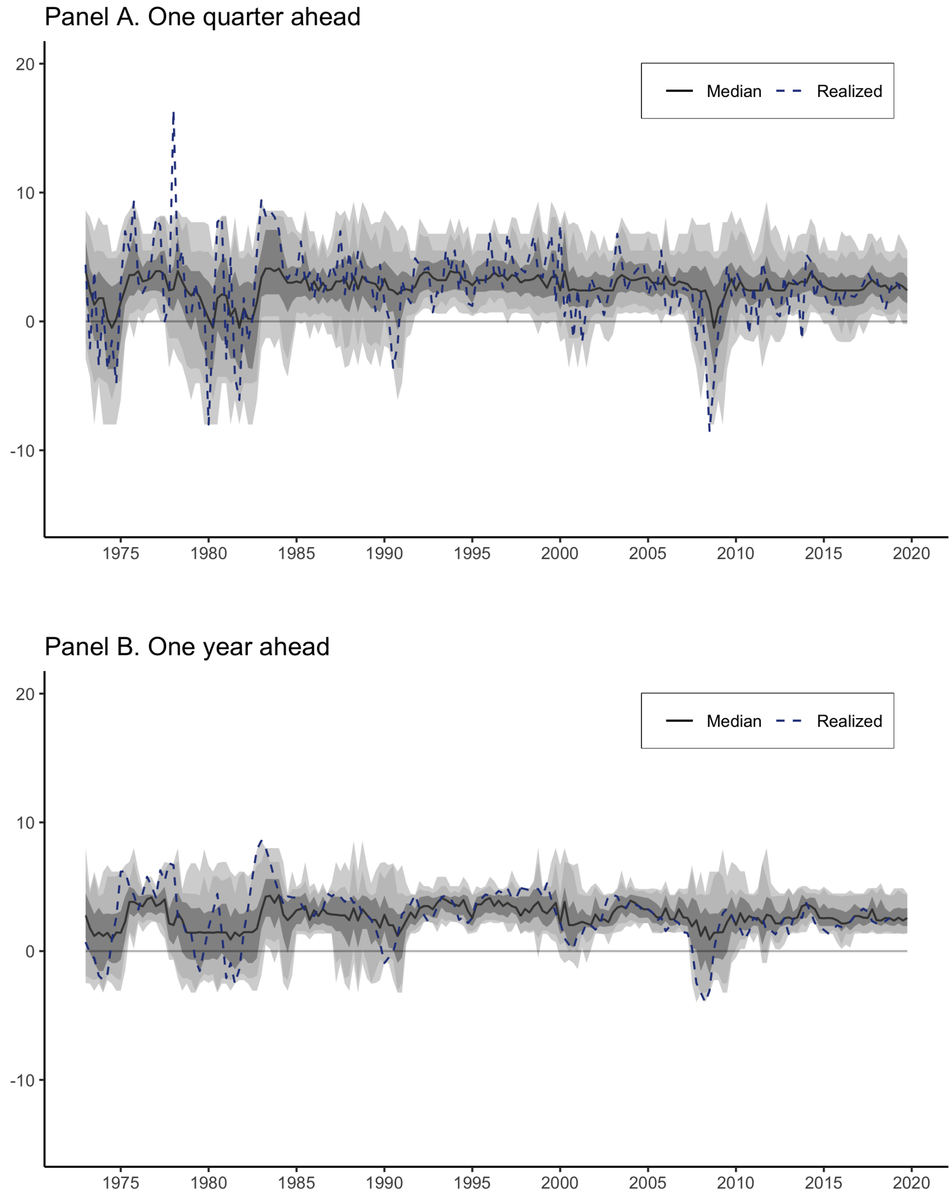


Figure 9: Predicted Distributions using Generalized Quantile Forest

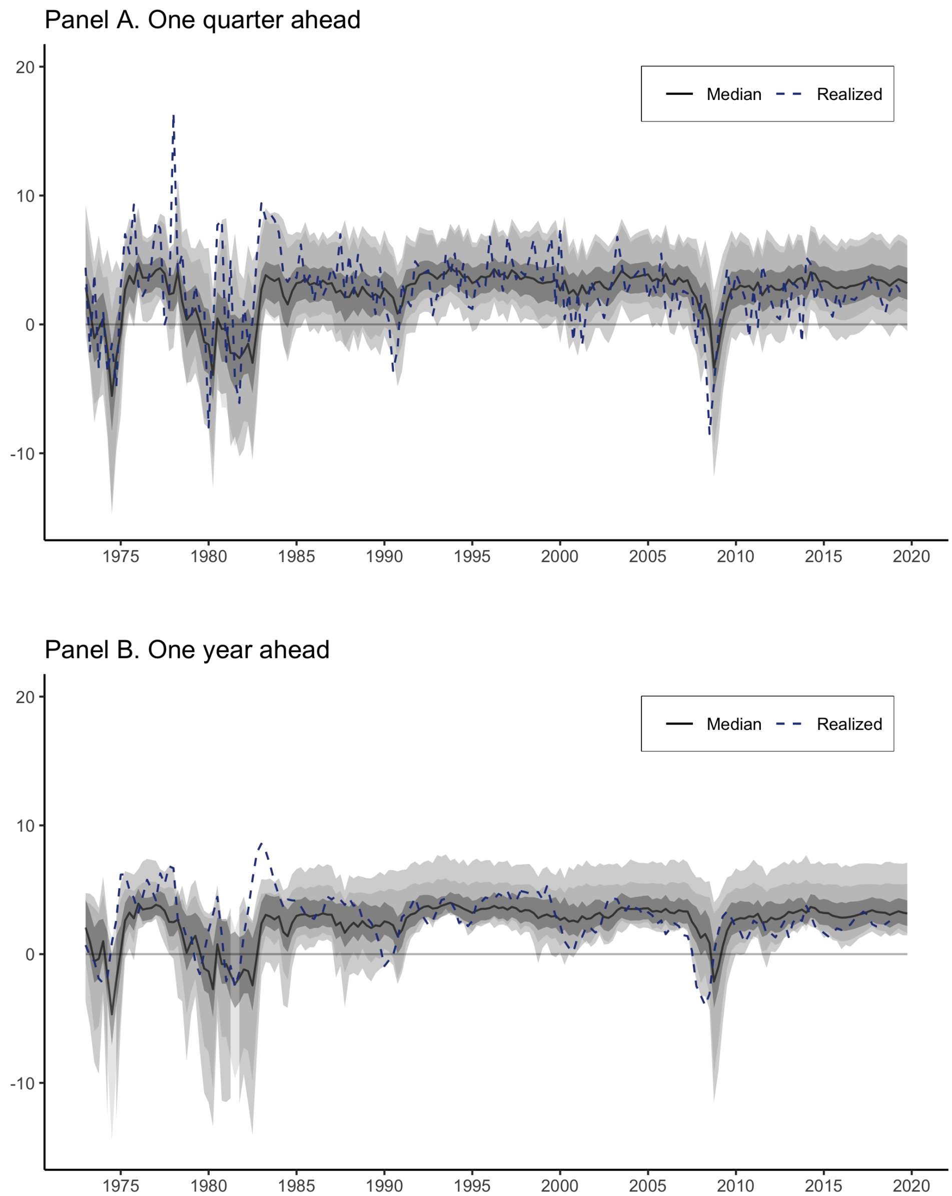


Figure 10: Predicted Distributions using IVXQR-ivqr

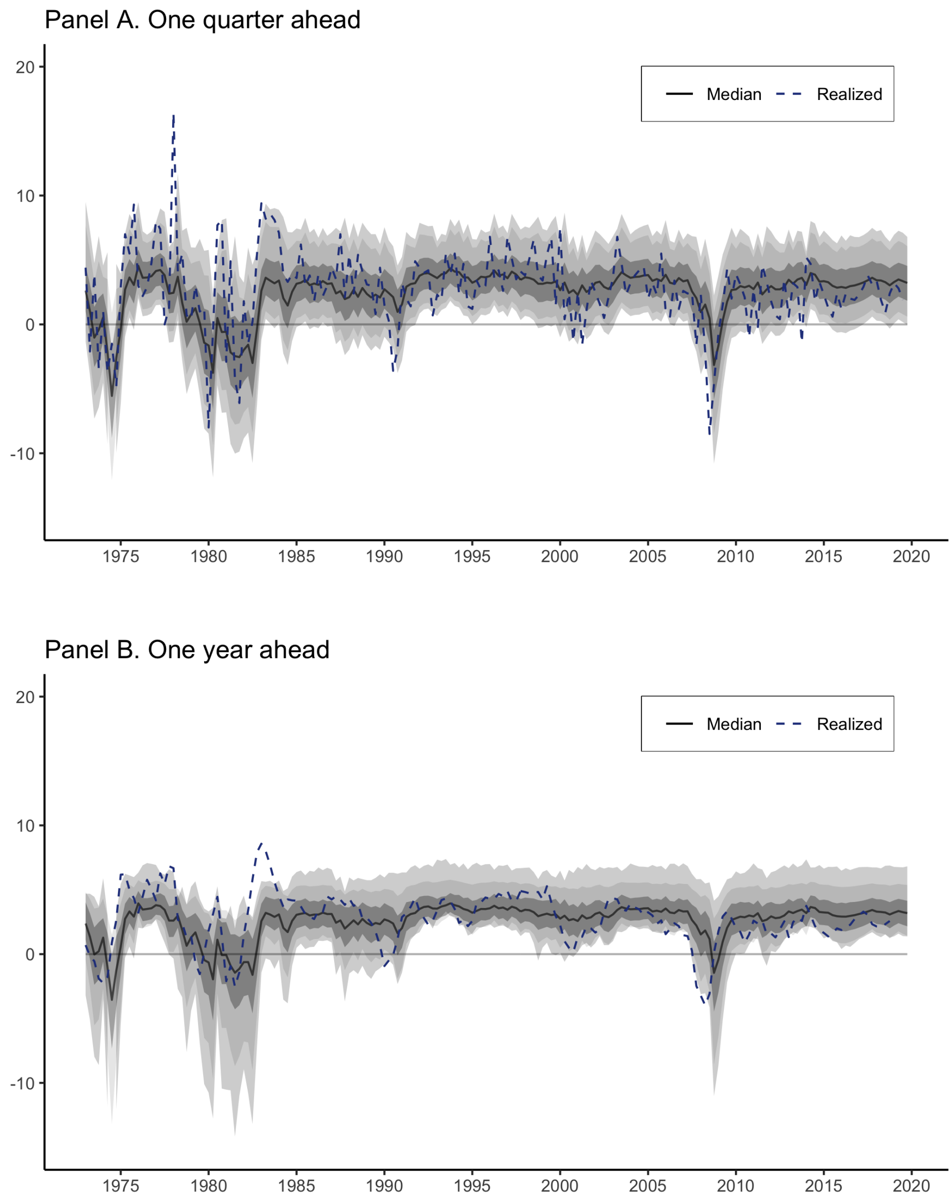


Figure 11: Predicted Distributions using IVXQR-see

5 Out-of-Sample Prediction

In this section, we conduct out-of-sample predictions across different methods, and compare their performances based on the out of sample quantile loss comparison. First, we use Expanding Window Forecast to calculate the prediction. Figure 12 provides a visualization to help make the idea more intuitive. Let's consider the starting point for estimation to be k (our out-of-sample predictions start from 1997Q4), and the total number of observations is T. In general, we use the first k periods of observations as our training set to predict for $k+1$ period. Then we use the first $k+1$ periods of data to make a prediction for $k+2$ period.

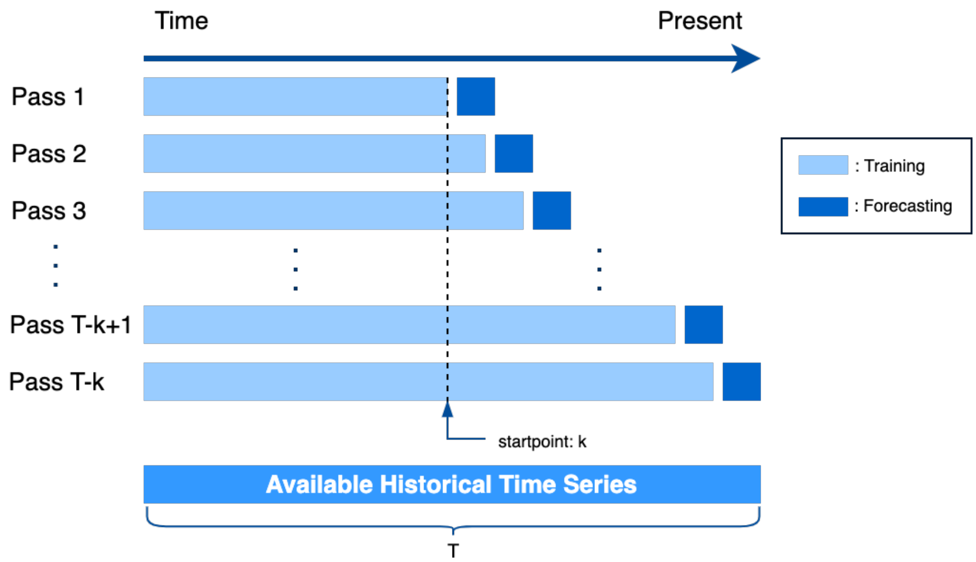


Figure 12: Expand Window Forecast

To compare performances for different models, we defind the Final Prediction Error (FPE) using the following equation:

$$\frac{1}{T-k}\sum_{s=k}^{T}\rho_\tau(y_s-\hat{y}_s)$$

where $\hat{y}_s$ is prediction for τ. We then determine the optimal combination of hyperparameters by selecting configuration with the smallest FPE. The following tables present the comparison of FPE for different packages and the optimal configurations chosen for them [c]. After fixing maxnode=11 and minnode=10, we present FPE of different block sizes in Table 2. Both 1-quarter and 4-quarter forecast show that when block-size increases, FPE first decreases and then increases at certain point.

`rangerts`	1-quarter ahead	maxnode=11, minnode=10, blocksize=20
	4-quarter ahead	maxnode=12, minnode=10, blocksize=12
`grf`	1-quarter ahead	minnode=10
	4-quarter ahead	minnode=3

Table 1: Optimal Configurations for packages

[c]FPE for different configurations can be found in Appendix Table 6

	blocksize=0	blocksize=4	blocksize=8	blocksize=12	blocksize=16
1-quarter ahead	0.5866	0.5520	0.5589	0.5608	0.5585
4-quarter ahead	0.4263	0.4251	0.4237	0.4231	0.4186
	blocksize=20	blocksize=24	blocksize=28		
1-quarter ahead	0.5503	0.5570	0.5593		
4-quarter ahead	0.4205	0.4157	0.4233		

Table 2: Final Prediction Error for different block sizes

Visualizations for out-of-sample predictions across quantiles for different method using their optimal configurations are presented as follows.

5.1 Visualizations

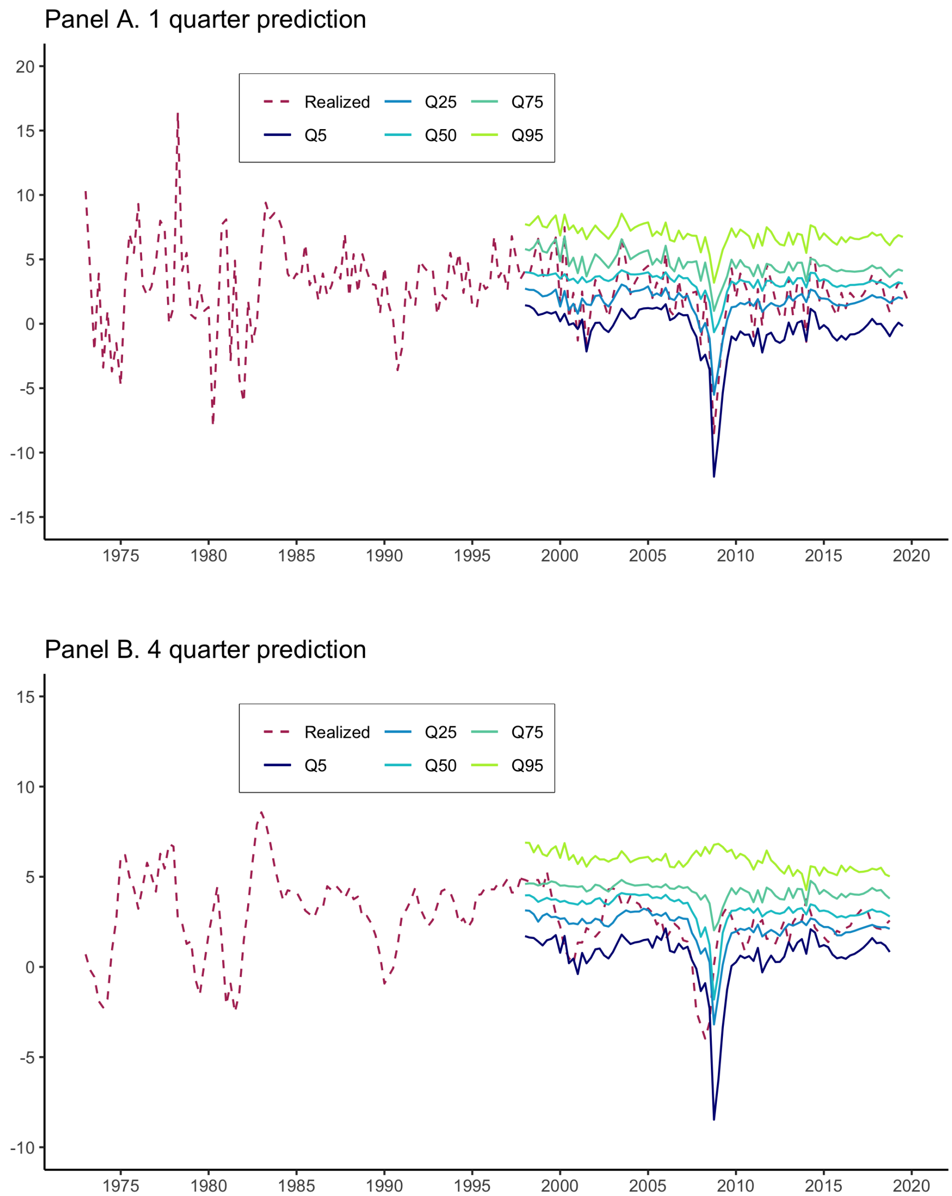


Figure 13: GDP growth Prediction using QR

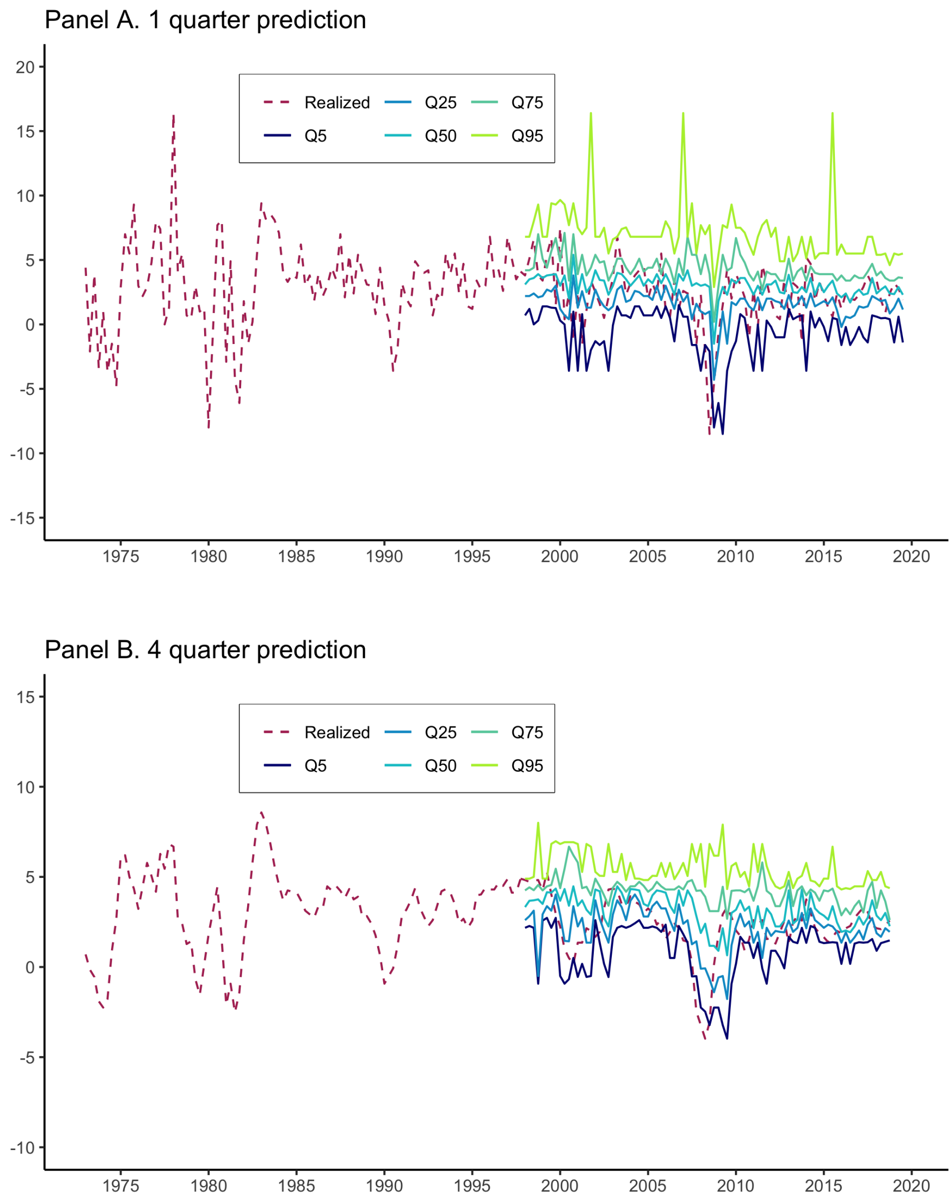


Figure 14: GDP growth Prediction using Quantile Random Forest

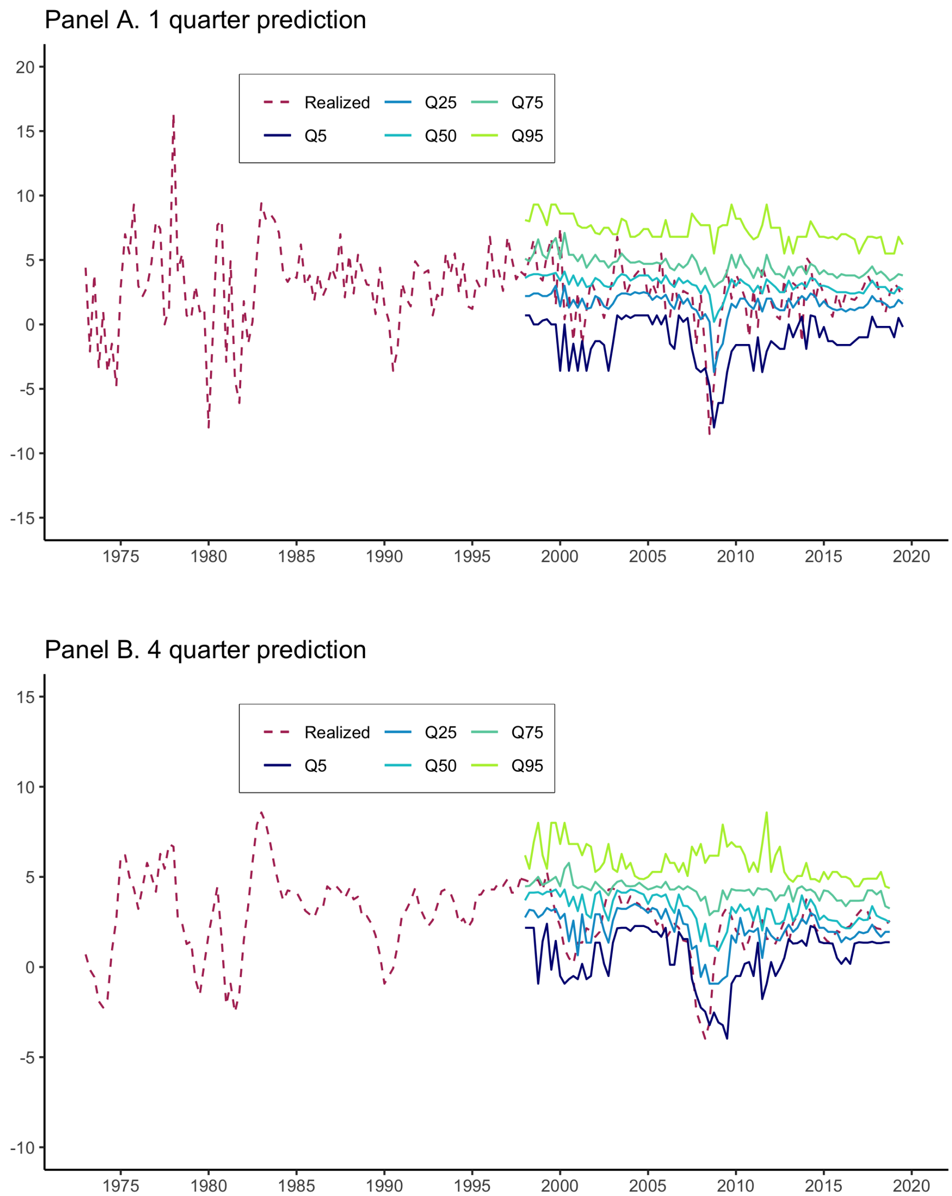


Figure 15: GDP growth Prediction using General Random Forest

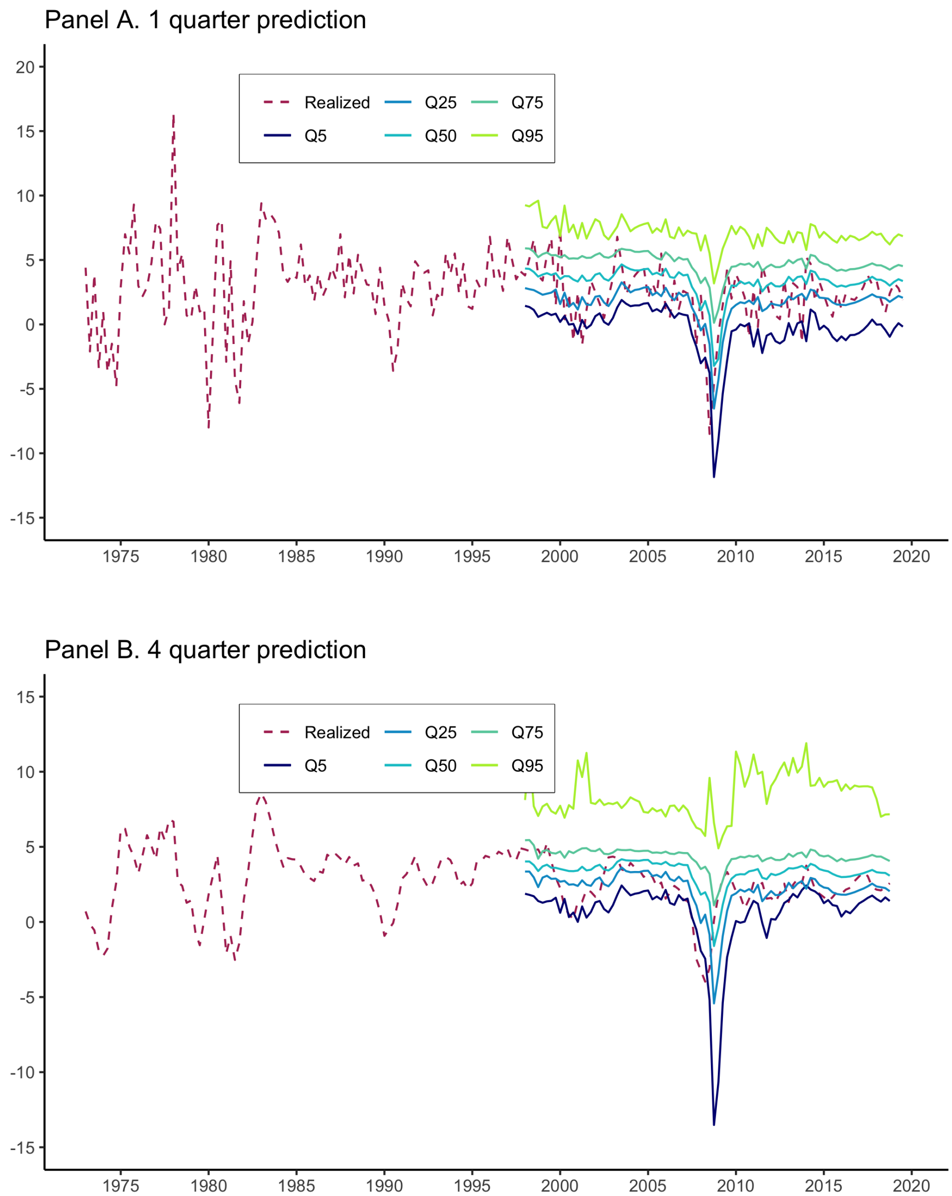


Figure 16: GDP growth Prediction using IVXQR-ivqr

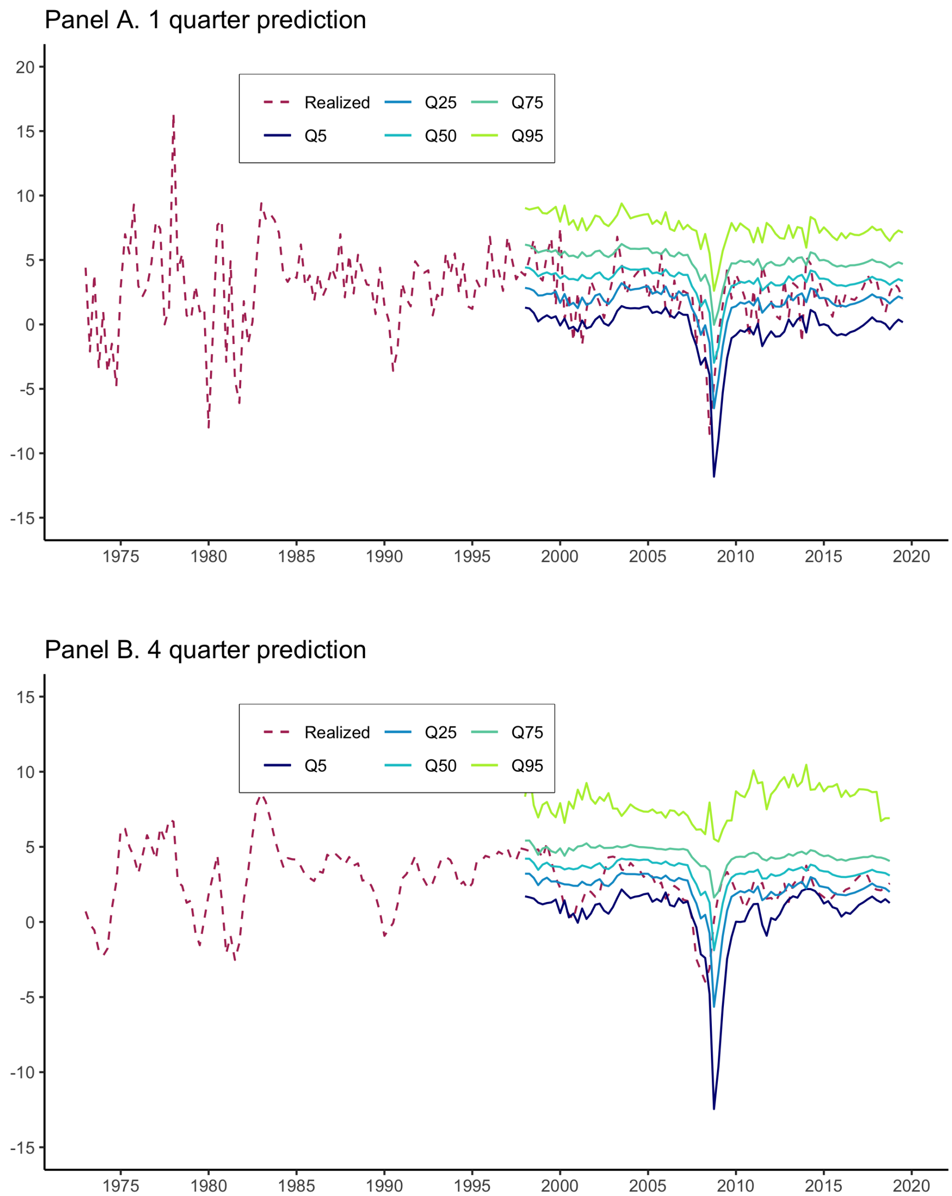


Figure 17: GDP growth Prediction using IVXQR-see

Predictions in lower quantiles appear to be more volatile than those in upper quantiles, which aligns with [Adrian et al., 2019]'s conclusion that financial conditions is correlated with GDP growth vulnerablity. In fact, illustrations using IVQR methods shows a higher volatility at the lower quantile. Panle B in Figure 16 and Figure 17 shows that the blue line reaches almost to 15 while the rest have a lowest point $\in (-5, -10)$.

Table 3 summarizes the one- and four-quarter ahead predictions' FPE for different methods across quantiles. All the models seem to perform better at the end tails and have a higher FPE around 0.5 quantiles. four-quarter ahead forecast also generally have lower FPE comparing to one-quarter ahead forecast. Compare FPE across different models, generalized random forest has the best performance for out of sample prediction. This indicates that splitting the data into splitting and estimation sets would help forecast GDP growth .

	quantreg		rangerts		grf		ivqr		see	
	h1	h4	h1	h4	h1	h4	h1	h4	h1	h4
$\tau = 0.05$	0.2706	0.1977	0.3195	0.1615	0.2430	0.1379	0.2692	0.1625	0.2565	0.1599
$\tau = 0.25$	0.6752	0.5128	0.7432	0.5610	0.6848	0.4752	0.6772	0.5269	0.6779	0.5200
$\tau = 0.50$	0.8236	0.6649	0.8471	0.7000	0.8029	0.6345	0.8428	0.7159	0.8497	0.7178
$\tau = 0.75$	0.6939	0.5184	0.6966	0.5468	0.6736	0.5202	0.7143	0.5496	0.7479	0.5755
$\tau = 0.95$	0.2392	0.1823	0.2392	0.1648	0.2518	0.1796	0.2493	0.3087	0.2619	0.2882

Table 3: Final Prediction Error summary

6 Data Simulation

In this section, we build a data simulation to analyze the factors that influence performance of GDP growth prediction using tree based methods. We consider the data generating process below:

$$y_t = (1 + \beta \cdot x_{t-1})(u_t + 3),\ x_t = \rho x_{t-1} + v_t$$

where

$$\begin{pmatrix} u \\ v \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \phi \\ \phi & 1 \end{pmatrix}\right)$$

Let $Q_{y_t|x_{t-1}}(\tau)$ denote the τ^{th} quantile of $y_t|x_{t-1}$. Then

$$\begin{aligned} \tau &= Pr[y_t \leq Q_{y_t|x_{t-1}}(\tau)] \\ &= Pr[(1 + \beta x_{t-1})(u_t + 3) \leq Q_{y_t|x_{t-1}}(\tau)] \\ &= Pr\left[u_t \leq \frac{Q_{y_t|x_{t-1}}(\tau)}{1 + \beta x_t} - 3\right] \end{aligned}$$

Since $\tau = Pr\left[u_t \leq \frac{Q_y(\tau)}{1+\beta x_t} - 3\right]$, we can denote $\frac{Q_y(\tau)}{1+\beta x_i} - 3$ as $Q_u(\tau)$.

$$
\begin{aligned}
\frac{Q_{y_t|x_{t-1}}(\tau)}{1+\beta x_{t-1}} - 3 &= Q_u(\tau) \\
\Rightarrow Q_{y_t|x_{t-1}}(\tau) &= (Q_u(\tau)+3)(1+\beta x_{t-1}) \\
&= (Q_u(\tau)+3) + (Q_u(\tau)+3)\,\beta x_{t-1} \\
&= \beta_{0,\tau} + \beta_{1,\tau}\, x_{t-1}
\end{aligned}
$$

where $\beta_{0,\tau} = Q_u(\tau) + 3$, and $\beta_{1,\tau} = (Q_u(\tau) + 3)\beta$. Then we test combinations of $\rho = 0, 0.3, 0.9$ and $\phi = 0, -0.05, -0.95$ to find out their impact on predictions using Quantile Regression Forest and Generalized Quantile Random Forest across quantiles. The following tables summarize mean square error for quantile predictions using two different methods across different combinations of ρ and ϕ. Both tables show that predictions become more inaccurate at upper and lower quantiles. Compare between methods, Generalized Quantile Random Forest generally has a lower MSE than Quantile Regression, which aligns with our conclusion in the previous section. When x is highly persistent ($\rho = 0.9$), high correlation between u and v ($\phi = -0.95$) will lead to a worse prediction.

Configuration	Quantile 10%	Quantile 30%	Quantile 50%	Quantile 70%	Quantile 90%
$\rho\, 0.3\ \phi - 0.05$	0.1266	0.0983	0.0925	0.0991	0.1263
$\rho\, 0.3\ \phi - 0.95$	0.1266	0.0958	0.0867	0.0926	0.1248
$\rho\, 0.3\ \phi\ \ 0$	0.1252	0.0985	0.0931	0.0985	0.1259
$\rho\, 0.9\ \phi - 0.05$	0.1176	0.0898	0.0831	0.0896	0.1173
$\rho\, 0.9\ \phi - 0.95$	0.1186	0.0903	0.0842	0.0924	0.1205
$\rho\, 0.9\ \phi\ \ 0$	0.1179	0.0884	0.0820	0.0897	0.1190
$\rho\, 0\ \ \phi - 0.05$	0.1284	0.0995	0.0939	0.0997	0.1287
$\rho\, 0\ \ \phi - 0.95$	0.1265	0.0964	0.0895	0.0974	0.1266
$\rho\, 0\ \ \phi\ \ 0$	0.1290	0.1005	0.0961	0.1019	0.1289

Table 4: Mean Square Error for Quantile Regression Forest

Configuration	Quantile 10%	Quantile 30%	Quantile 50%	Quantile 70%	Quantile 90%
$\rho\, 0.3\, \phi \quad -0.05$	0.0715	0.0428	0.0388	0.0429	0.0725
$\rho\, 0.3\, \phi \quad -0.95$	0.0713	0.0423	0.0383	0.0424	0.0718
$\rho\, 0.3\, \phi \quad 0$	0.0718	0.0426	0.0383	0.0426	0.0721
$\rho\, 0.9\, \phi - 0.05$	0.0711	0.0429	0.0390	0.0429	0.0725
$\rho\, 0.9\, \phi - 0.95$	0.0734	0.0433	0.0396	0.0430	0.0712
$\rho\, 0.9\, \phi \quad 0$	0.0709	0.0429	0.0391	0.0429	0.0730
$\rho\, 0 \quad \phi - 0.05$	0.0730	0.0423	0.0389	0.0431	0.0728
$\rho\, 0 \quad \phi - 0.95$	0.0722	0.0430	0.0386	0.0430	0.0724
$\rho\, 0 \quad \phi \quad 0$	0.0718	0.0421	0.0386	0.0429	0.0723

Table 5: Mean Square Error for Generalized Quantile Random Forest

7 Conclusion

Analysis of the relationship between real GDP growth and economic and financial conditions across different quantiles gives us a more comprehensive understanding of economies under different conditions. The recent pandemic reaffirms the importance to analyze growth vulnerability. In this paper, we apply different models on a recent dataset to study their relationships across quantiles.

Our finding supports the argument that increasingly severe financial conditions are associated with an increase in conditional volatility. After addressing the invalidity of QR methods rising from persistent regressors, the GDP growth forecast is more conservative. Results from out-of-sample predictions also suggest that financial conditions are informational for vulnerability predictions. Comparing out-of-sample quantile loss comparisons across different models, we conclude that generalized random forest has the best best performance. The estimate logic behind grf, separating training set into splitting and estimation sets, is helpful to capture the non-linear relationship between economies under different conditions and thus predict future GDP growth.

References

[Adrian et al., 2019] Adrian, T., Boyarchenko, N., & Giannone, D. (2019). Vulnerable growth. *American Economic Review*, 109(4), 1263–89.

[Athey et al., 2016] Athey, S., Tibshirani, J., & Wager, S. (2016). Generalized random forests.

[Cai et al., 2022] Cai, Z., Chen, H., & Liao, X. (2022). A new robust inference for predictive quantile regression. *Journal of Econometrics*.

[Kaplan & Sun, 2017] Kaplan, D. M. & Sun, Y. (2017). Smoothed estimating equations for instrumental variables quantile regression. *Econometric Theory*, 33(1), 105–157.

[Lee, 2014] Lee, J. H. (2014). Online supplement to predictive quantile regression with persistent covariates: Ivx-qr approach.

[Lee, 2016] Lee, J. H. (2016). Predictive quantile regression with persistent covariates: IVX-QR approach. *Journal of Econometrics*, 192(1), 105–118.

[Magdalinos & Phillips, 2009] Magdalinos, T. & Phillips, P. C. B. (2009). Limit theory for cointegrated systems with moderately integrated and moderately explosive regressors. *Econometric Theory*, 25(2), 482–526.

[Marjanović et al., 2016] Marjanović, V., Milovančević, M., & Mladenović, I. (2016). Prediction of gdp growth rate based on carbon dioxide (co2) emissions. *Journal of CO2 Utilization*, 16, 212–217.

[Meinshausen, 2006] Meinshausen, N. (2006). Quantile regression forests. *journal of machine learning research*, 7, 983–999.

[Yoon, 2021] Yoon, J. (2021). Forecasting of real gdp growth using machine learning models: Gradient boosting and random forest approach. *Computational Economics*, 57(1), 247–265.

Appendix

A. Illustration

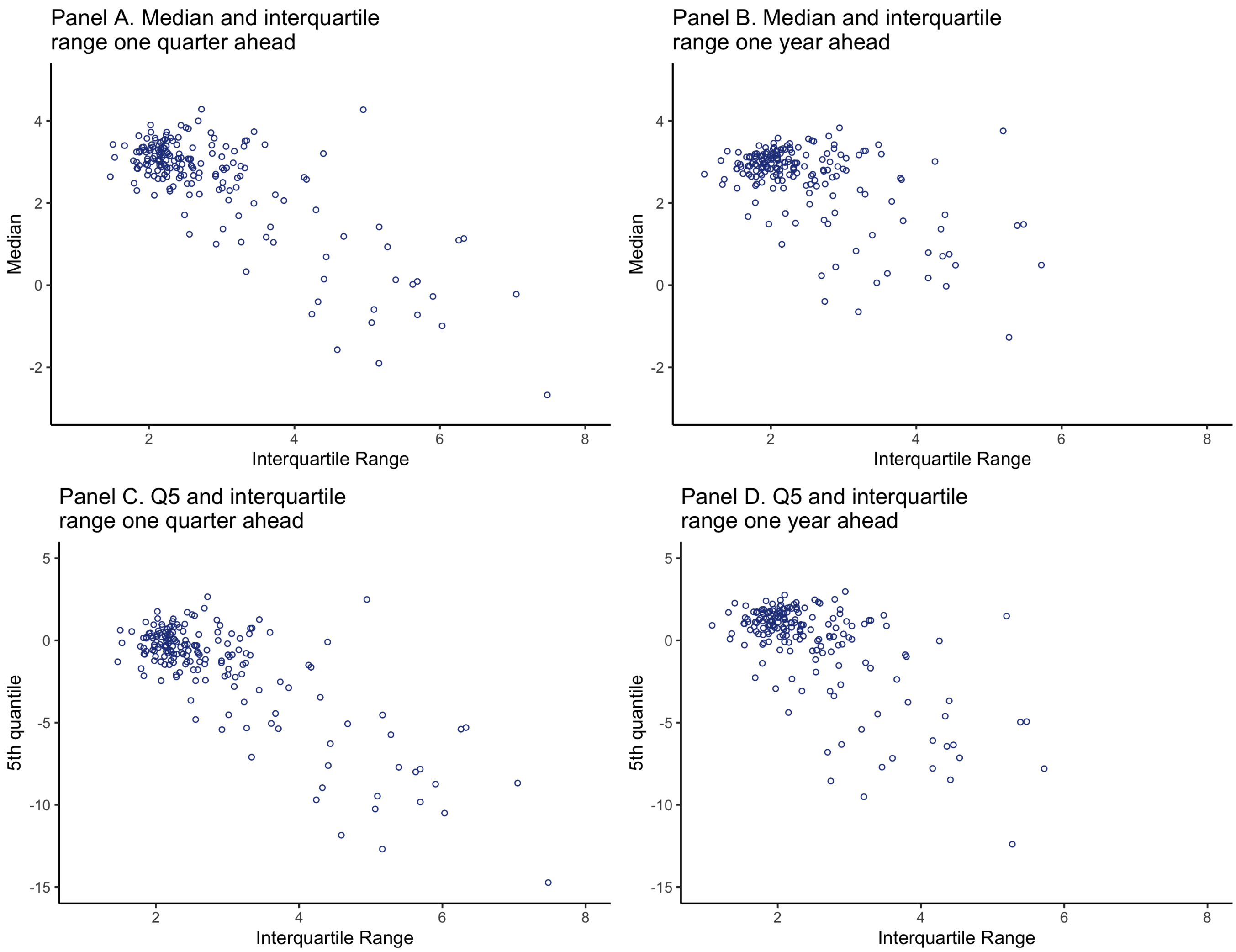


Figure 18: Median, Interquartile Range, and 5 Percent Quantile of the Predicted Distribution

B. FPE Table

	minnode=3	minnode=4	minnode=5	minnode=6	minnode=7	minnode=8	minnode=9	minnode=10
h1 pred	0.53767	0.54138	0.53837	0.53733	0.53623	0.53464	0.53497	0.53121
h4 pred	0.38950	0.38958	0.39018	0.38976	0.39059	0.39103	0.39257	0.39471
	minnode=11	minnode=12	minnode=13	minnode=14	minnode=15	minnode=16	minnode=17	
h1 pred	0.53284	0.53659	0.53649	0.53772	0.54011	0.54097	0.54075	
h4 pred	0.39569	0.39641	0.39764	0.39960	0.40035	0.40111	0.40202	

Table 6: `grf` FPE for each configurations