ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Bachelor Thesis Econometrics, Quantitative Finance

July 5, 2020

Forecasting real GDP with information on government spending using quantile methods

Author:

Laura Nieberg

Student ID number: 429886

Supervisor:

dr. A.M. Schnucker

Second Assessor:

dr. Y. Sun

Abstract

Much research has been done in the non-linearity of government spending and its impact on the real economy. This paper provides new evidence on the non-linearity of fiscal shocks through quantile regression. In particular, quantile regression is used to construct quantile specific impulse response functions. Consequently, quantile methods are employed to forecast different quantiles of output (real GDP). This paper uses two data sets: a baseline set and an extended one. The baseline set is employed to prove the non-linearity of fiscal shocks. It is then used to construct forecasts and extended with additional variables to enhance the forecasting performance. From this research, it can be concluded that the impact of fiscal shocks is indeed non-linear and differs per quantile in which output is estimated. Quantile methods using only the baseline variables result in accurate forecasts. The forecasting performance does not improve with adding additional variables.

Contents

1 Introduction

Government spending and its impact on the real economy is a topic of broad discussion. Policy makers often have the objective to stimulate the economy during a recession, and slow down the economy during an expansion. To adequately respond to the real economy, policy makers should understand the impact of government spending. Much research has been done in this direction. Early contributions of the effect of fiscal policy came from Blanchard & Perotti (2002). They used linear structural vector autoregressions (VAR) to estimate the impact of government spending shocks. More recent research has focused on non-linear models to estimate the effect of government spending. Auerbach & Gorodnickenko (2012) provide evidence of this non-linearity of government spending shocks. They show that the impact of government spending shocks on output is systematically larger in recessions compared to expansions. Linnemann & Winkler (2016) present new evidence on the non-linearity of fiscal shocks by using quantile regression to estimate a VAR model. As Ramey (2011) argues in her research, there is some delay in the impact of government spending on the real economy. If one increases government spending now, what would be the impact tomorrow? Therefore, policy makers should also be aware of the future possible states of output to implement fiscal policy well.

The aim of this paper is to see if non-linear models using information on government spending can forecast output (real GDP) adequately. In particular, three quantile regression methods are evaluated regarding their forecasting performance. The use of quantile regression in this research is beneficial for two reasons. First, quantile regression allows the parameters to differ across different quantiles of output. Since the effect of government spending shocks are non-linear, quantile regression is an adequate method to use. Moreover, quantile regression gives a more complete picture about the future states of output compared to a linear estimation methods, which only gives information about the mean forecast. In this research I first provide evidence on the non-linearity of fiscal shocks by computing the impulse response functions of output relative to a government spending shock. Consequently, I use three different methods to forecast different states of future output and evaluate the three methods.

The first method that is used for both providing evidence on the non-linearity of government spending and forecasting different states of output, is quantile vector autoregression (QVAR). This methodology is based on Cecchetti & Li (2008), who compute quantile-specific impulse response function to estimate the non-linear effect of asset price booms and crashes. Linnemann & Winkler (2016) follow this research and estimate the impact of government spending on the real economy by applying quantile regression to a VAR model. They argue that this approach has two advantages relatively to traditional linear estimation methods and to smooth transition estimated used by Auerbach & Gorodnickenko (2012). First, quantile regression allows the effect of government spending on the real output to differ across conditional quantiles. Second, relatively to the smooth transition methods, there is no need to characterize the data to pre-specified recession or expansion periods. Hence, my research follows the paper of Linnemann & Winkler (2016) for providing evidence on the non-linearity of government spending shocks.

Next to the VAR model, two machine learning techniques are evaluated to forecast output. The advantage of machine learning techniques relative to traditional forecasting methods (such as VAR models) is that they are able to capture non-linear relations between response and explanatory variables. Traditional methods rely on strict mathematical relations between inputs and outputs which does not allow non-linear relations. Quantile regression allows the parameters to differ across quantiles, but within the quantiles the relation between inputs and output is linear. Machine learning methods are much more flexible and able to capture complex relations. Therefore, they are in general powerful tools and widely applied in the forecasting area. Machine learning methods are in particular useful for this research as the impact of government spending is non-linear.

The first methods that is applied, is the quantile regression forest by Meinshausen (2006). This quantile regression forest is derived from the random forest algorithm by Breiman (2001). Where the random forest algorithm gives an approximation of the conditional mean, the quantile regression forest gives an approximation of the full conditional distribution. The quantiles can subsequently be derived from the conditional distribution. Meinshausen (2006) shows that this method can construct prediction intervals that cover new observations with a high probability. This method is in particular power full with high-dimensional data. Therefore, the data set as used in Linnemann & Winkler (2016) is extended with additional variables. Random forest is an ensemble technique which does not need any optimization in contrast to other machine learning methods. Instead, it grows an ensemble of trees. Therefore, the computation time of this method is much less compared to other machines learning techniques which is why this method is evaluated.

The second machine learning methods that I will use in this research is quantile regression neural network (QRNN), which is a neural network approach to quantile regression. It is based on artificial neural networks (ANNs). ANNs have attractive features for forecasting and are therefore extensively used in the forecasting area (Zhang et al., 1998). First, they do not require prior information about the distribution of the data, they only require enough observations. Instead of guessing the underlying distribution of the data, ANNs learn from experience. Another advantage of ANNs is, that they can generalize well. Most of the time they can predict the unknown part of the population correctly. Therefore, ANNs are widely applied in forecasting, as forecasts try to estimate the unknown part of the population. The third advantage of ANNs that makes them suitable for forecasting is that they can approximate underlying functions between response and explanatory variables well, whereas traditional statistical methods (like VAR models) have limitations due to the complexity of some relations. The final advantage of ANNs in the area of forecasting is that the models are nonlinear. Traditional forecasting methods often assume that time series studied are generated from linear processes. In the real world, this is usually not the case.

QRNN regression is based on a single hidden-layer feedforward network network and is similar to traditional ANNs. QRNN is therefore also a highly suitable method for forecasting. The major advantage that makes it in particular useful for this research, is that is able to estimate potentially non linear models without the need to specify the precise function in advance (Taylor, 2000). Therefore, the forecasting performance of this method is evaluated as well.

This paper uses the same variables as the paper of Linnemann & Winkler (2016) to provide evidence on the non-linearity of fiscal shocks. I use US quarterly data ranging from 1955Q1 to 2019Q4 which is the longest sample on which the data are available. The variables used are: government spending, output, net taxes, real interest rate, and the ratio of government debt held by the public to GDP. The data are cleaned from seasonal effects and the trends in the data are removed by log-quadratic detrending. Additional variables are then added to the originial data set since the random quantile forest algorithm is particularly powerful with high-dimensional data. This set is called the extended set and consists of 26 variables ranging from 1992Q2 to 2019Q4.

The main results are as follows. The impulse response functions and fiscal multipliers show that the response of output to a government spending shocks differs across quantiles. The impact of fiscal shocks are much more persistent when output is relatively low. Opposite effects occur when output is in the highest quantile of its conditional distribution. The variable set used to prove the non-linearity provide accurate forecasts on output. The QVAR model yields most precise forecasts on short horizons, whereas the machine learning methods, in particular QRNN, work well for longer horizons. Adding additional variables does not improve the forecast quality. The QVAR forecasts become inaccurate due to the fact of dimensionality, whilst the machine learning methods still provide accurate forecasts. The forecasts are however poorer compared to the baseline data set, as the in-sample period contains less observations.

In what follows, I first present the data in section 2. Section 3 elaborates on the different methods used, both for proving the non-linearity and forecasting. Section 4 demonstrates the most important results and a discussion is provided. Finally, Section 5 summarized the most important results and highlights ideas for future research.

2 Data

This paper uses two data sets. The first data set is used for both providing evidence on the non-linearity of government spending. This data set is then extended with additional variables to see if the forecasting performance can be enhanced.

2.1 Baseline Data Set

The baseline data is the set similar to the research of Linnemann & Winkler (2016). The set of variables used is government spending G_t, real GDP Y_t, real net taxes ψ_t, the short-run real interest rate R_t and the ratio of government debt held by public relative to GDP D_t. The data are retrieved from the FRED database on a quarterly basis ranging from 1955Q1 to 2019Q4 which is the longest sample over which the data are available. Only US data is used in this research. The data are seasonally adjusted. The level variables G_t, Y_t, ψ_t, and D_t are measured as log-deviations from quadratic time trends as these variables contain strong upward trends. In the research of Linnemann & Winkler (2016), quadratic detrending was also used to remove the trend in the level variables. To facilitate comparisons, quadratic detrending is used to remove the trend in variables here as well. The ratio of government debt held by public relative to GDP D_t is only available on an annual basis. Therefore, this series is interpolated to obtain quarterly data. The transformations that are applied to the data, both to the baseline and extended data set, are summarized in Table 3 in Appendix A.

This paper only takes real GDP (output) as response variable as it adequately measures the economic activity. To ordering of the variable matters in the computation of the impulse response function since a Cholesky decomposition is used to obtain orthogonal shocks. Keeping up with literature, the government spending G_t variable is ordered first. The variables have the following ordering: $(G_t, Y_t, \psi_t, R_t, D_t)$. Summary statistics and correlations of the baseline data data set are displayed in Appendix A, Table 4 and Table 5 respectively. Visual representations of the transformed data are given in Appendix A, Figure 10.

2.2 Extended Data Set

Quantile regression forest gives accurate predictions with high-dimensional data (Meinshausen, 2006). Therefore, the baseline data set is extended with some additional variables that correlate with output. The additional variables are based on the research of Stock & Watson (1999), who evaluated several models in their macroeconomic forecasting performance. They split up U.S. Macro-economic time series in 5 different categories, namely: (A) Income, output, sales, and capacity utilization, (B) Employment and unemployment, (C) Construction, inventories and orders, (D) Interest rates and asset prices, and (E) Nominal, prices, wages and money. To enrich the baseline variable set, several variables per category are added to the baseline variable set. The additional variables are retrieved from the FRED database on a quarterly basis ranging from 1992Q2 to 2019Q4 since this is the longest time period over which the variables are available. There are 21 additional variables, so the total data set (which includes the baseline set), contains 26 variables. The variables and the transformations are displayed in Table 3 in Appendix A. The level variables are log-quadratic detrended in this data set as well. A descriptive statistics of the extended transformed data is given in Appendix A, in Table 6. Visual representations of the transformed data are displayed in Appendix A, Figure 11.

3 Methods

In the following section, the methods for both estimating the model as well as for forecasting are elaborated on. First, quantile regression is used to prove the non-linearity of output relative to government spending. Afterwards, quantile regression is used to estimate the future states of output using an quantile vector autoregressive model (QVAR) and two machine learning approaches: quantile regression forest and QRNN.

3.1 Quantile Regression

First I will eleborate on quantile regression in general since this method is used for both estimating and forecasting. To illustrate the working of quantile regression, consider the following simple regression model:

$$y_t = x_t'\beta + \varepsilon_t, \tag{1}$$

where x_t is a $k \times 1$ vector of regressors, β a vector of coefficients, and $E(\varepsilon|x) = 0$. The OLS estimator of Equation 1 can be written as:

$$\hat{\beta}_{ols} = \arg\min_{\beta} \sum_{t=1}^{T} (y_t - x_t'\beta)^2. \tag{2}$$

By applying OLS, the error function is evaluated at its mean. The mean of the variable y_t conditional on the explanatory variables is defined as: $E(y_t|x_t) = x_t'\beta$. It is of interest to see the impact of fiscal policy shocks during recessions and expansions separately, therefore the conditional distribution of y_t is evaluated at different quantiles. Koenker & Basset (1978) came up with the concept of quantile regression. It measures the impact of a change in a variable x_t on the q^{th} quantile of y_t. The function in Equation 2 can be generalized by writing the function as follows:

$$\hat{\beta} = \arg\min_{\beta} \sum_{t=1}^{T} \rho(y_t - x_t'\beta), \tag{3}$$

where $\rho(\cdot)$ is a weighting function. In order to obtain the quantile weight function, consider the following weighting function:

$$\rho_q(y_t - x_t'\beta) = \begin{cases} q(y_t - x_t'\beta) & \text{if } (y_t - x_t'\beta) \geq 0 \\ (1-q)|y_t - x_t'\beta| & \text{otherwise,} \end{cases} \tag{4}$$

where q is the quantile where the function is evaluated. The errors get a positive weight q ,when the estimation error exceeds 0, and a weight $(1 - q)$ when the estimation error is smaller than 0. In the case of $q = 0.5$, the function in Equation 4 minimizes the sum of absolute deviations, which coincides with the estimation of the conditional distribution of y_t at its median. The estimator $\beta(q)$, which measures the conditional effect of x_t on y_t in the q^{th} quantile, is defined as:

$$\hat{\beta}(q) = \arg\min_{\beta(q)} \sum_{t=1}^{T} \rho_q(y_t - x_t'\beta(q)), \tag{5}$$

with ρ_q as defined in equation 4. When Equation 5 is evaluated in different quantiles q, one can determine the impact of x_t on the whole conditional distribution of y_t.

3.1.1 Quantile Vector Autoregression

The baseline data contain quarterly observations of five different variables. Since there are multiple variables, I model the data as a vector autoregression model (VAR). The VAR model is estimated using quantile regression. A VAR(p) model with k variables and p lags is defined as follows:

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + u_t, \tag{6}$$

where y_t is a $k \times 1$ vector of variables, c a $k \times 1$ vector of intercepts, u_t a $k \times 1$ vector of errors, and A_i are $k \times k$ matrices of coefficients.

The VAR model in this research is estimated by quantile regression. The general QVAR(p) model in the case of p lags and k variables can be written as follows:

$$y_t = c(q) + A_1(q)y_{t-1} + A_2(q)y_{t-2} + \cdots + A_p(q)y_{t-p} + u_t(q), \tag{7}$$

where y_t is a $k \times 1$ vector of variables, $c(q)$ a $k \times 1$ vector of intercept, $u_t(q)$ a $k \times 1$ vector of errors, and A_i are $k \times k$ fixed coefficient matrices. k is equal to 5 (variables) for the baseline data set and 26 for the extended data set. The parenthesis indicate that the parameters are dependent on the quantile in which they are estimated. The quantiles in which the conditional distribution of the variables are estimated can differ per variable, so the coefficients in each equation are dependent on the quantiles in which the variables are estimated.

The number of lags that are included in the model is determined by the information criteria. The values of the information criteria should be minimized when choosing the correct VAR order. The values of the information criteria at different lags for both data sets are displayed in Table 8 and 9 in the Appendix B.1 respectively. Table 7 in Appendix B.1 displays the correct VAR order according to the information criteria. For robustness, in the baseline set, forecasts are constructed based on both AIC and BIC criteria (so with 5 and 3 lags respectively). The impulse response functions are computed with the VAR model with 5 lags. Since the criteria agree on the number of lags in the extended set, only one VAR model is estimated with 3 lags. The information criteria give infinite values for lags larger than three, which indicates that the extended data set is too large for the VAR to give proper estimations. The model suffers from dimensionality.

In this paper, following the research of Linnemann & Winkler (2016), the output equation (y) is estimated at 3 different quantiles of its conditional distribution, $q = \{0.1, 0.5, 0.9\}$. When $q = 0.1$, the output is at the lowest quantile of its conditional distribution which coincides with a recession. Similarly, estimating output at $q = 0.9$ coincides with an expansion. When $q = 0.5$, the output is estimated at its median which is close to the OLS estimate (since OLS estimates the system at the mean). The other variables are estimated at the median of their conditional distribution. The model is estimated by applying equation-by-equation quantile regression.

3.1.2 Fitted Values

First the fitted values of the lower and upper quantile will be compared to the actual output realizations. In this way, one can observe when the actual output is close to the 0.1 quantile or 0.9 quantile forecast. Hence, displaying when shocks occur that push output far below or above its conditional mean forecast. At these time, the conditional mean will give a poor estimate of the actual output.

3.1.3 Impulse Response

To provide evidence on the non-linearity of fiscal shocks, quantile specific impulse response functions are computed for the baseline data set. The impulse response function shows the response of one variable to an exogenous shock in the another variable. Suppose I have a bivariate ($k = 2$) model with 1 lag ($p = 1$) and there is an exogenous shock in the first variable. This can be written as follows:

$$u_0^* = \begin{pmatrix} u_{1,0}^* \\ u_{2,0}^* \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \tag{8}$$

where $u_{i,0}^*$ is the shock in the i^{th} variable at time 0. The impulse response functions track the effect of this shock through estimating the VAR system first. The effect of this shock can be traced as follows:

$$y_0 = \begin{pmatrix} y_{1,0} \\ y_{2,0} \end{pmatrix} = \begin{pmatrix} u_{1,0}^* \\ u_{2,0}^* \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{9}$$

$$y_1 = \begin{pmatrix} y_{1,1} \\ y_{2,1} \end{pmatrix} = A_1 y_0 = A_1 u_0^* \tag{10}$$

$$\vdots$$

$$y_h = \begin{pmatrix} y_{1,h} \\ y_{2,h} \end{pmatrix} = A_1 y_{h_1} = A_1^h u_0^*, \tag{11}$$

where $y_{i,t}$ is the response of variable i at time t, and A_1 a 2×2 fixed coefficient matrix. In this paper, the response of y_t and g_t to a positive one percent shock in the median of government spending is evaluated up to 20 periods ($h = 20$). To compute the impulse response function, the covariance matrix of residuals is orthogonalized first through a Cholesky decomposition as shocks are likely to be correlated. In this way, one can obtain orthogonal shocks that are not correlated. In this transformation, u_0^* in Equation 8 is replaced by $e_0^* = P^{-1}u_0^*$, where P is defined by $\Sigma_u = PP'$ which is the Cholesky decomposition of the covariance matrix of shocks. The effect of the orthogonalized shock can be traced by replacing u_0^* by $P^{-1}u_0^*$ in Equation 9, 10, and 11 respectively.

The response of a variable to a shock in another variable can be tracked with the estimated coefficients of the VAR model. The impulse response functions are computed for the 3 different quantiles $q = \{0.1, 0.5, 0.9\}$. The ordering of variables matters for the Cholesky decomposition and are ordered as follows: $(G_t, Y_t, \psi_t, R_t, D_t)$, which keeps up with earlier research. 90% confidence intervals on the impulse response functions are then computed through the bootstrapping of residuals.

The impulse responses are only computed for the baseline data set since interpretation and ordering of the variables becomes too difficult for the extended data set. The baseline data set alone succeeds in proving the non-linearity of government spending shocks. Evaluating the impulse response functions on the other data set is therefore out of the scope of this research.

3.1.4 Fiscal Multipliers

To facilitate comparisons between the effects of fiscal policy at different quantiles and strengthen the evidence of the non-linearity of fiscal shocks, I compute fiscal multipliers. Fiscal multipliers normalize the output response to a change in the government spending variable. In particular, I compute two ratio's: the maximum point-to-point ratio and the cumulative ratio. The point-to-point ratio is defined as: $R_h(q) = \hat{y}_h(q)/\hat{g}_h(q)$, where $\hat{y}_h(q)$ and $\hat{g}_h(q)$ are the impulse responses of output and government spending respectively, h periods after the shock with the variables estimated at the q^{th} quantile. The maximum point-to-point ratio is the maximum of these values, $MR_q = max_{h\in\{0,\cdots,12\}} R_h(q)$. To compute this ratio, all quantiles between 0.05 and 0.95 are estimated in steps of 5% and the largest ratio within 12 periods (h) is then stored. So the maximum-point-to-point ratio displays the largest normalized response per quantile.

The cumulative ratio is defined as: $CR_h(q) = [\sum_{h=0,\cdots,12} \hat{y}_h(q)]/[\sum_{h=0,\cdots,12} \hat{g}_h(q)]$. It measures the normalized total response of output in the first 12 periods after the shock relative to government spending and gives a more complete picture on the impact of fiscal shocks compared to the point-to-point multiplier.

3.2 Forecasting

After proving the non-linearity of government spending shocks, I will see if the variables included in the research of Linnemann & Winkler (2016) yield accurate forecasts on the output variable y. First, I give a brief overview of the general set-up of the forecast. Consequently, I elaborate on the different methods used for forecasting. To end with, I discuss forecasting performance measures used in the evaluation of the forecasts.

Both the baseline as well as the extended data set are used to construct forecasts. The sets are both divided up into two subsamples: the in-sample period and the out-sample period. In the baseline data set, the in-sample period is set to 1955Q1 to 2014Q4 and the out-sample period, for which the forecasts are constructed, is 2015Q1 to 2019Q4. Hence, the in-sample period contains 240 observations (N_{in}) and the out-of sample period 20 observations (N_{out}). The extended data set is used to see if the forecasting performance can be enhanced using additional variables. The in-sample period in this set is 1992Q2 to 2014Q4, and the out-sample period is 2015Q1 to 2019Q4. Thus, the in-sample set contains 91 observations (N_{in}) and the out-sample set 20 observations (N_{out}). The data sets have the same out-of-sample set to facilitate comparisons. The forecasts are constructed using 3 different forecast horizons, namely $h = \{1, 4, 12\}$. Where $h = 1$ is the one-quarter ahead forecast, $h = 4$ the one-year ahead forecasts , and $h = 12$ the three year ahead forecast. All forecasts are constructed using an expanding window. This means that the entire in-sample period (observation 1 to N_{in}) is used to construct the first sequence of forecasts for horizon 1 to 12. Hence, the first observation that is forecasted with horizon 12 is observation 252 in the baseline variable set ($N_{in} + h$). Consequently, the models are re-estimated using observations 1 tot $N_{in} + 1$ to produce a new sequence of forecasts for horizon 1 to 12. Finally, for each forecasts horizon h, $N_{out} - h + 1$ forecasts are created. So for the baseline set, with horizon 12, there are 9 forecasted values for observation 252 up to and including 260. An overview of the in-and out sample period for both data sets are given in Appendix B.2, Table 10.

In this research, quantile methods are used to forecast the output at three different quantiles, namely $q = \{0.1, 0.5, 0.9\}$. The 0.1 and 0.9 quantile forecasts are used to construct confidence intervals around the median forecast. Using quantile methods, I obtain information about different future quantiles, whereas normal estimation methods usually only give information about the mean. The median forecast is the leading forecast which is used to evaluate the forecasting performance of the forecasts since this forecast is closest to the mean forecast.

3.2.1 Quantile Vector Autoregression Model (QVAR)

To begin with, the QVAR model defined in Section 3.1.1 is employed to construct forecasts. In literature, VAR models are popular for macro-economic time series forecasting. They can adequately capture the dynamic behaviour of both economic and financial time series (Zivot & Wang, 2006).

The general h-step ahead forecast of a VAR(p) model with constant estimated with quantile regression is defined as:

$$\hat{y}_{T+h|T} = \hat{c}(q) + \hat{A}_1(q)\hat{y}_{T+h-1|T} + \cdots \hat{A}_h(q)y_T + \cdots + \hat{A}_p(q)y_{T+h-p}, \tag{12}$$

where $y_{T+h|T}$ is a $k \times 1$ vector of h-step ahead forecasts and where $\hat{y}_{T+h-1|T}, \cdots \hat{y}_{T+1|T}$ are forecasted using similar schemes. The forecasts are depended on the conditional quantiles in which the variables are forecasted. Since the output variable is forecasted at three different quantiles, there will be three different forecast vectors corresponding to the different quantiles.

3.2.2 Quantile Regression Forest

Quantile regression forest was first introduced by Meinshausen (2006). Quantile forest is a machine learning technique that is based on the random forest algorithm by Breiman (2001). The algorithm behind quantile forest resembles the random forest algorithm. The quantile forest algorithm gives information about the entire conditional distribution of a variable, whereas the random forest algorithm only gives information about the mean. First I will first briefly explain the intuition behind the random forest algorithm.

The random forest algorithm grows an ensemble of decision trees using independent observations. At each tree, the tree makes a class prediction of the data. The class prediction that has the most votes becomes the model's prediction. Random forest keeps the mean of the observations for each node in each tree. On the contrary, quantile regression forest stores the value of all observations in each node. The model is usually divided in a training and test sample. The model is trained using the training sample and predictions are made based on the test sample. The response and prediction variables are separated to establish the relation between these variables in the training sample.

In contrast to the recursive iterated method which is used to construct h-step ahead forecasts in the QVAR model, quantile regression forest creates direct forecasts. The advantage of the direct forecasts is that is does not have to predict the predictor variables in the test set. Normally, for constructing predictions on the response variables, data on the prediction variables are used from the test set. Since I try to forecasts future states of output for an out-of-sample period, information on the prediction variables will not be available which is why direct forecasting is useful for this purpose.

Again, forecasts are constructed for 3 different horizons using an expanding window. Direct forecasts for horizon 1 to 12 are constructed using all in-sample observations. The in-sample observations are then split up in different training samples, to overcome the problem of overfitting. The forecasts are thus constructed using information on different training samples. The size of the training sample is dependent on the window size. If the window size is for instance 5 years, the in-sample period will be split up in 12 training samples (in-sample observations/ window size in quarters) in the baseline data set. For robustness, forecasts are created using different window sizes, namely 3, 5, and 10 years. The number of forecasts available per time period is equal to the number of different training samples. To come up with one forecast per time period, I take the mean of the forecasts. This process is repeated using an expanding window to create new forecasts. An overview of the different forecasting methods, windows sizes, and abbreviations used in the result section is given in Appendix B.2, Table 11.

3.2.3 Quantile Regression Neural Network (QRNN)

Finally, QRNN is employed to produce forecasts on output. QRNN is widely applied in the forecasting area due to its attractive features. It is particularly useful in this research since it is able to estimate non-linear relationships adequately without a prior specification of the data or model.

The QRNN is displayed in Figure 1. The left or the first layer is the input layer which receives the information on the variables and consists of m input neurons for the predictors. The most right layer is the output layer were one output neuron yields the prediction. The layers between the input and the output layers are called the hidden layers. In the case of forecasting, the single hidden-layer feedforward network is the most common used (Zhang et al., 1998) which is illustrated in Figure 1.

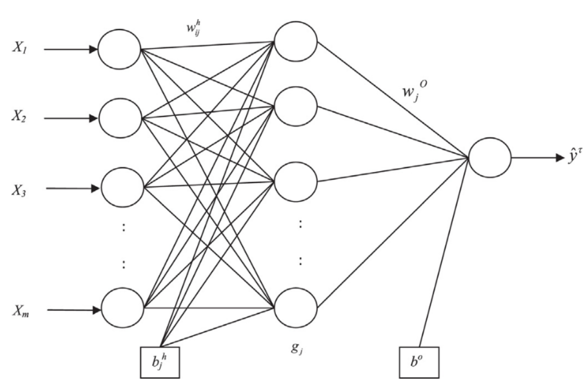


Figure 1: Quantile Regression Neural Network

The process of training makes the QRNN different from traditional ANNs. If the model is trained using Equation 13 as cost function, then the output is estimated at the condtional regression quantiles (Taylor, 2000).

$$E(q) = \frac{1}{T}\sum_{t=1}^{T}\rho(y_t - \hat{y}_t) \tag{13}$$

$$\rho_q(y_t - \hat{y}_t) = \begin{cases} q(y_t - \hat{y}_t) & if(y_t - \hat{y}_t) \geq 0 \\ (1-q)|(y_t - \hat{y}_t)| & \text{otherwise,} \end{cases} \tag{14}$$

where y_t is the value of the response variable at time t, $\hat{y}_t$ the predicted value at time t, T the number of observations in the training sample, q the quantile in which the equation is estimated, and $\rho(\cdot)$ the weighing function.

The direct forecasts are constructed in a similar way as the quantile random forest algorithm. The training sets are trained with the QRNN algorithm, instead of the quantile random forest algorithm. Due to heavily computations, only one window size is used in each data set. In the baseline data set the window size is set as 10 years and in the extended data set as 5 years since both sets differ in length. The number of hidden nodes is 2, which allows the model to look for non-linear decision boundaries. An overview of the forecasting methods are given in B.2, Table 11.

3.2.4 Forecast Evaluation

The forecast performance is measured using 5 different criteria. Namely the RMSE, MAE, Theil's U, the ratio of observations that the quantile intervals covers, and the Diebold Mariano Test statistics. The measures compare the median ($q = 0.5$) forecasts with the actual values, since the median forecasts is closest to the OLS estimate. I will briefly elaborate on the forecasts accuracy measures in this section.

Root Mean Squared Error The Root Mean Squared Error is a scale dependent measure of the dataset and is measures the scaled squared error. The RMSE is computed as follows:

$$RMSE = \sqrt{\sum_{t=1}^{T}\frac{(\hat{y}_t - y_t)^2}{T}}, \tag{15}$$

where $\hat{y}_t$ is the median forecast of y_t at time t, y_t the actual value, and T the total number of forecasts. The advantage of the RMSE is that it is on the same scale of the data and therefore a widely used measure. The disadvantage of this measure is that it is not robust to outliers. Therefore, also the Mean Absolute Error is evaluated.

Mean Absolute Error The Mean Absolute Error measures the absolute average error of the forecast. The advantage of this criteria relative to the RMSE is that this measure is robust to outliers. It is computed in the following way:

$$MAE = \frac{1}{T}\sum_{t=1}^{T} \mid (\hat{y}_t - y_t) \mid, \tag{16}$$

where $\hat{y}_t$ is the median forecast of y_t at time t, y_t the actual value, and T the total number of forecasts.

Theil's U Theil's U (Theil, 1966) is a relative measure of forecast quality and measures the forecast error relative to forecasting with minimal data. A Theil's U value close to zero indicates a good forecast, whereas a value close or larger then one indicates a poor forecast. For a value larger then one, the forecasting method applied should be rejected as it not able to beat forecasts resulting from simple extrapolation (Bliemel, 1973). The measure is easy to interpret and understand which is why also this criteria is evaluated.

$$U = \sqrt{\frac{\frac{1}{T}\sum_{t=1}^{T-1}(\frac{\hat{y}_{t+1}-y_{t+1}}{y_t})^2}{\frac{1}{T}\sum_{t=1}^{T-1}(\frac{y_{t+1}-y_t}{y_t})^2}}, \tag{17}$$

where $\hat{y}_t$ is the median forecast of y_t at time t, y_t the actual value, and T the total number of forecasts.

Prediction Ratio The prediction ratio is used to test whether the forecasting methods are able to construct confidence intervals which cover all future observation. The intervals are created with the 0.1 and 0.9 quantile forecasts. The prediction ratio is the number of actual output realizations that lies within the interval divided by the total number of observations (for the out-sample). This measure is evaluated to see which models can give accurate prediction intervals. Next to the prediction ratio, graphs of the intervals and actual output are created to evaluate the accuracy of the intervals.

Diebold-Mariano Test Statistics The Diebold-Mariano test statistic is used to test whether two forecast methods differ significantly. From the above criteria, one can define what forecasting method performs best, but it also of interest to see if the methods differ significantly in performance. The Diebold-Mariano test statistics indicate whether the methods differ significantly. Suppose I compare two different forecasting methods. The forecasts errors are defined as:

$$\varepsilon_{it} = \hat{y}_{it} - y_t, \text{ for i=1,2}, \tag{18}$$

where $\hat{y}_{it}$ is the median forecast of y_t at time t constructed with method i. The loss differential is then defined as:

$$d_t = g(\varepsilon_{1t}) - g(\varepsilon_{2t}), \tag{19}$$

where $g(\cdot)$ is a squared loss function ($g(\varepsilon_{it}) = \varepsilon_{it}^2$). The forecasts have equal accuracy if and only if the expected value of the differential is equal to zero for all t. The Diebold-Mariano statistics test the hypothesis $H_0 : E(d_t) = 0$ versus the alternative hypothesis $H_1 : E(d_t) \neq 0$.

The sample mean and autocorrelation function are used as input for the test statistics and are defined in the following equations:

$$\overline{d} = \frac{1}{T}\sum_{t=1}^{T} d_t, \tag{20}$$

$$\gamma_k = \frac{1}{T}\sum_{t=k+1}^{T}(d_t - \overline{d})(d_{t-k} - \overline{d}), \tag{21}$$

where d_t denotes the loss function at time t, T the total number of forecasts, and k the order of the autocorrelation. The Diebold-Mariano test statistics now can be calculated as:

$$DM = \frac{\overline{d}}{\sqrt{(\gamma_0 + 2\sum_{k=1}^{h-1} \gamma_k / T)}}. \tag{22}$$

Under the null-hypothesis that the forecasting methods do not differ significantly, the Diebold-Mariano statistic follows a standard normal distribution (e.g. $DM \sim N(0,1)$).

4 Results

In this section, the result are presented. First evidence is provided on the non-linearity of fiscal shocks. Consequently, the forecasts are evaluated using the different methods.

4.1 Non-linearity government spending

4.1.1 Fitted Values

To provide evidence on the non-linearity of fiscal shocks, it is of interest to see where the actual output is close to the highest or lowest decile of its conditional distribution. Here I follow the research of Linnemann & Winkler (2016), who also compare the lowest and highest decile with the actual output. The fitted (predicted) values of the lowest and highest quantile together with the actual output are presented in Figure 2 and resemble the results of Linnemann & Winkler (2016). Just as in their research, it can be observed when the actual output is close to an expansion or to a recession, so when linear methods will not provide good estimations (as the actual output is close to either an expansion or to an recession and not to the mean). For instance, one can note that the output in the years 2004-2008 was close to the 0.9 conditional quantile forecasts, which pushed the output far above its mean. During these years, linear methods will fail to adequately estimate output.

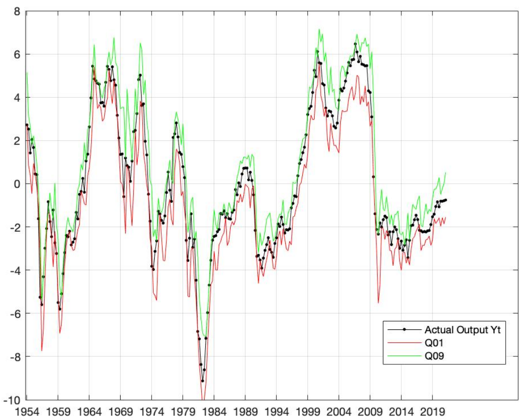


Figure 2: Fitted values baseline data set estimated with QVAR where the green and red line present the one step ahead forecast (fitted values) for the 0.9 and 0.1 quantile respectively. Actual output is given in black.

4.1.2 Impulse Response

The quantile specific orthogonalized impulse response functions are displayed in Figure 3. In the upper 3 figures, the response of output relative to a positive 1% shock in the median of government spending is displayed, with 90% confidence intervals given in dashed lines. When output is estimated at its lowest conditional quantile (during a recession) the effect of a government spending shock peaks after 9 to 11 quarters. Afterwards the impact decreases and even becomes negative. It can be seen that the impact of the shock is much more persistent when output is estimated at it lowest decile. When output is estimated at its median ($q = 0.5$), the impact of a government spending shock is only small and declines slowly towards zero. The effect of a government spending shock during an expansion, when $q = 0.9$, is first negative and becomes positive after approximately 10 periods. The different impulse response functions prove that the impact of government spending is non-linear and differs across quantiles. Less evidence on the non-linearity of government spending in response to a fiscal shock can be derived from the second row of Figure 3. Here, the responses do not differ significantly.

The response of the output relative to a government spending shock, when output is estimated at the lowest decile, has the same hump-shaped recovery as in Linnemann & Winkler (2016). The output peaks between 7 and 9 periods after the shock. After these periods, the response function declines at a much smaller rate than in this research. There is no clear difference between the impulse response functions estimated at the median in this research compared to the research of Linnemann & Winkler (2016). At the highest decile, the response becomes negative after some periods, similar to Linnemann & Winkler (2016). However, the response increases after some periods in this research which does not happen in the research of Linnemann & Winkler (2016). The response of government spending relative to a government spending shock resemble the results of Linnemann & Winkler (2016), except for the lowest decile, where the effect of a shock increases instead of decreases. The overall conclusion is similar to Linnemann & Winkler (2016): the effect of fiscal policy is indeed non-linear and differs per quantile. Small differences may be caused by small dissimilarities between the data sets and the number of lags included in the research. I interpolated some variables that were not available on a quarterly basis and did not know the exact serie source of the ratio of government debt held by the public to GDP d_t series. In this research, the lags are chosen by the information criteria and the impulse response functions are based on 5 lags, whilst Linnemann & Winkler (2016) work with 4 lags.

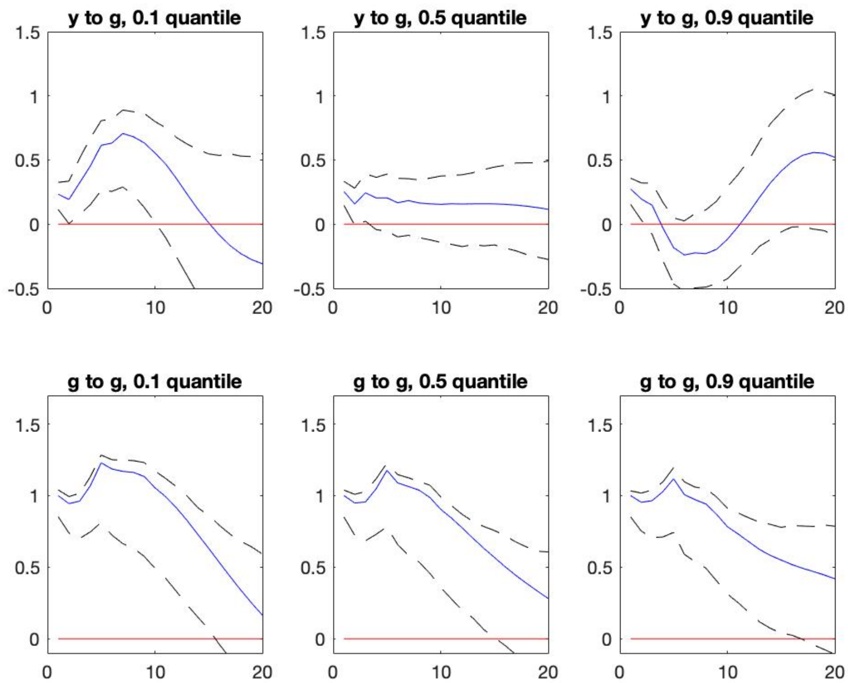


Figure 3: Impulse response functions of output (real GDP) and government spending relative to a 1% positive shock in the median of government spending, when the parameters are estimated at the 0.1, 0.5, and 0.9 output quantiles

4.1.3 Fiscal Multipliers

Fiscal multipliers are computed to give a better overview of the normalized response of output relative to government spending at different quantiles. Figure 4 and 5 display the maximum point-to-point and cumulative ratio respectively. The normalized response are shown on the y-axis and the quantiles in which the ratios are estimated on the x-axis. One can clearly see in Figure 4 that the maximum normalized response of output to a government spending shock is higher when output is at its lowest quantiles. The ratio varies strongly across the quantiles in which it is estimated. The maximum point-to-point ratio is the highest when output is estimated at $q = 0.15$. The graphs show that fiscal policy shocks are strong and persistent when they occur in phases were output is low.

The cumulative ratio (CR) gives a more complete picture than the maximum point-to-point ratio as it gives not just information on one point, but on the entire response. The results look quite similar: fiscal policy shocks are much more persistent when they occur in phases were output is low.

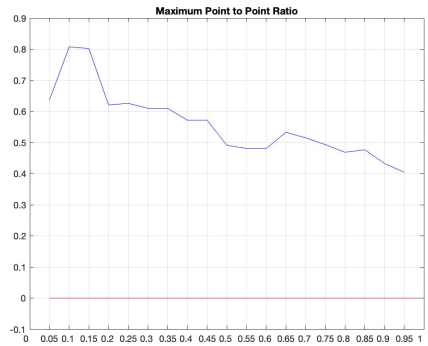


Figure 4: Maximum point-to-point multiplier with normalized response on the y-axis and the quantiles in which the equations are estimated on the x-axis

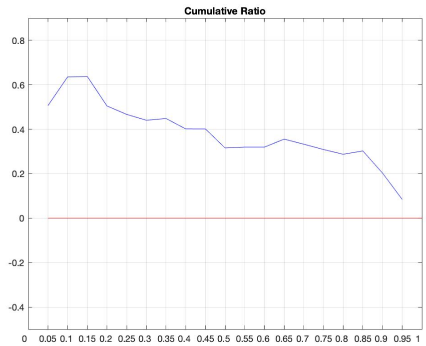


Figure 5: Cumulative multiplier with normalized response on the y-axis and the quantiles in which the equations are estimated on the x-axis

4.2 Forecasting performance evaluation

In this section, the forecasting performance of the three models on both data sets are compared. I will first start by evaluating the forecasting performance on the baseline data set. Consequently, the extended data set is evaluated.

4.2.1 Baseline Data Set

Performance measures Table 1 presents the RMSE, MAE, Theil's U, and the interval ratio obtained by using the different forecasting models. The Diebold-Mariano (DM) test statistics are displayed in Table 2. The best performing method per horizon is highlighted in grey. A positive value in Table 2 indicates that the "row" method performs better than the "column" method. It obvious to see that the forecasting performances deteriorates for almost all models when the forecasting horizon increases. All models except for QRNN and QVAR-AIC yield accurate one-step-ahead forecasts. The QVAR model with the number of lags chosen by the BIC information criteria performs best on horizon 1, but not significantly better than the forest forecasts. On horizon 4, QRNN performs best according to the performance measures, but performs slightly worse according to the DM test statistics. This difference is however not significant. On horizon 12, the forest methods with an estimation window of 10 yields the most accurate forecasts, but does not perform significantly better than the QRNN algorithm. Both QVAR models perform poor on this horizon and are significantly outperformed by the random forest algorithm with estimation windows of 5 and 10 years.

In short, I can not draw a clear conclusion about what method is superior in forecasting for the baseline data set. The QVAR model with the number of lags chosen by the AIC information criteria is significantly outperformed on every forecast horizon. The QVAR-BIC model performs well on short horizons but is outperformed on longer horizons, whilst both machine learning methods succeed in providing accurate forecasts on longer horizons. In general, one can say that the data set used Linnemann & Winkler (2016) with only 5 variables, can provide accurate forecasts on the output variable.

Table 1: Forecasting performance measures

	Baseline Data Set				Extended Data Set			
	RMSE	**MAE**	**Theil U**	**Interval**	**RMSE**	**MAE**	**Theil U**	**Interval**
				Horizon=1				
QVAR-AIC	0.350	0.287	0.211	0.950				
QVAR-BIC	0.251	0.192	0.151	1.000	0.696	0.552	1.397	1.000
Forest, 3yrs	0.292	0.222	0.177	1.000	0.527	0.476	0.238	0.800
Forest, 5yrs	0.299	0.229	0.183	1.000	0.555	0.499	0.251	0.800
Forest, 10 yrs	0.301	0.231	0.183	1.000	0.595	0.535	0.273	0.800
QRNN	0.402	0.336	0.243	0.900	0.542	0.461	0.244	0.550
				Horizon=4				
QVAR-AIC	0.964	0.746	1.140	1.000				
QVAR-BIC	0.811	0.619	0.487	1.000	16.894	13.736	1.474	1.000
Forest, 3yrs	0.835	0.702	0.578	1.000	1.359	1.150	0.568	0.647
Forest, 5yrs	0.837	0.713	0.582	1.000	1.416	1.234	0.592	0.706
Forest, 10 yrs	0.790	0.667	0.529	1.000	1.552	1.409	0.649	0.588
QRNN	0.714	0.570	0.428	1.000	1.410	1.257	0.590	0.059
				Horizon=12				
QVAR-AIC	0.879	0.702	1.161	1.000				
QVAR-BIC	1.010	0.749	0.967	1.000	2891.580	1485.689	0.508	1.000
Forest, 3yrs	0.651	0.562	0.623	1.000	1.511	1.334	0.466	0.555
Forest, 5yrs	0.608	0.508	0.582	1.000	1.425	1.220	0.439	0.555
Forest, 10 yrs	0.442	0.377	0.423	1.000	1.519	1.449	0.442	0.556
QRNN	0.478	0.368	0.457	1.000	2.060	1.815	0.635	0.444

Notes: This table reports the forecasting performance per forecasting method for different horizons for both data sets. The best performing method per horizon and per data set is highlighted. The names in the left column resemble the forecasting methods. Since only one QVAR model with 3 lags is used in the extended data set, the QVAR-AIC row is left blank.

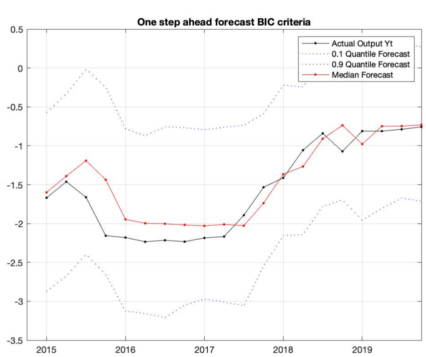


Figure 6: Median, 0.1 quantile and 0.9 quantile one-step ahead forecasts BIC criteria with the median forecast given in red, actual output in black, and 0.1-0.9 interval in blue dashed lines

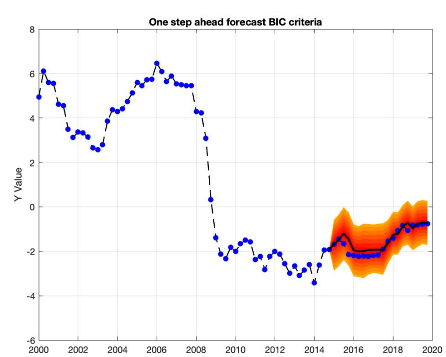


Figure 7: Median, 0.1 quantile and 0.9 quantile one-step ahead forecasts BIC criteria with the 0.1-0.9 interval in orange, median forecast in black, and the actual observations given in blue dots

Table 2: Diebold-Mariano Test Statistics

	Baseline Data Set					Extended Data Set			
	Horizon 1								
	QVAR-AIC	**QVAR-BIC**	Forest, 3yrs	**Forest, 5yrs**	**Forest, 10 yrs**	**QVAR**	**Forest, 3 yrs**	**Forest, 5yrs**	Forest, 10 yrs
QVAR-AIC									
QVAR-BIC	3.058***								
Forest, 3yrs	0.844	-0.896				3.947***			
Forest, 5yrs	0.769	-1.068	-1.417			3.946***	-2.609**		
Forest, 10 yrs	0.832	-1.236	-0.647	-0.068		3.945***	-2.852**	-2.177**	
QRNN	-2.012*	-3.081***	-1.796*	-1.770*	-1.906*	2.191**	0.734	1.564	2.264***
	Horizon 4								
	QVAR-AIC	**QVAR-BIC**	Forest, 3yrs	**Forest, 5yrs**	**Forest, 10 yrs**	**QVAR**	**Forest, 3 yrs**	**Forest, 5yrs**	Forest, 10 yrs
QVAR-AIC									
QVAR-BIC	6.575***								
Forest, 3yrs	5.167***	-1.503				3.471***			
Forest, 5yrs	5.309***	-1.182	1.511			3.479***	-2.380**		
Forest, 10 yrs	5.372***	-1.001	2.420**	0.897		3.466***	-3.913***	-4.371***	
QRNN	5.793***	-0.170	0.921	0.657	0.549	3.465***	0.184	0.535	1.018
	Horizon 12								
	QVAR-AIC	**QVAR-BIC**	Forest, 3yrs	**Forest, 5yrs**	**Forest, 10 yrs**	**QVAR**	**Forest, 3 yrs**	**Forest, 5yrs**	Forest, 10 yrs
QVAR-AIC									
QVAR-BIC	2.059*								
Forest, 3yrs	5.602***	1.621				4.132***			
Forest, 5yrs	6.771***	1.920*	2.300			4.042***	4.253***		
Forest, 10 yrs	5.775***	2.027*	2.945**	2.260*		4.134***	0.742	-0.981	
QRNN	-0.422	-1.067	0.921	-4.962***	-7.189	4.014***	-2.142*	-2.370**	-1.998*

*This table reports t-statistics for Diebold-Mariano tests for both data sets. A positive value is reported when a "row" model is superior to a "column" model. *, **, *** mean statistical significance at 10%, 5%, and 1% respectively. The best performing method per horizon per model and per data set is highlighted. Since only one QVAR model with 3 lags is used in the extended data set, the QVAR-AIC row is left blank.*

Forecast Graphs Since all methods provide confidence intervals that cover all new observations, I will evaluate the graphs of the forecast interval to see how accurate the intervals are.

The ratio of observations that lies within the 0.1 quantile - 0.9 quantile interval is 1 for almost all forecasting models on the different horizons (which is displayed in the last column of Table 1). To give an idea about the interval size and the observations that lie within it, plots of the forecasts are created. The graphs of the best performing method on horizon 1 are displayed above, whilst the complete set of plots (for all horizons and methods) can be found in Appendix C.

Figure 6 displays the 0.1 and 0.9 quantile forecasts in the blue dashed lines, the median forecast in red, and the actual output in black. Figure 7 presents the forecasted median values and the interval of the 0.1 and 0.9 quantile in orange. The blue rounds display the observations of the actual output whereas the forecasted median is visible in black. It can be observed that the one-step ahead QVAR-BIC forecast displayed in Figure 6 and 7 respectively are quite accurate and resemble the actual output. The black and the red line in Figure 7 deviate only small in the beginning and converge to each other towards the end. The interval (in dashed lines in Figure 7 and in red in Figure 6) is fairly small and accurate for future states of forecasts. When the forecast horizon increases, which can be seen in Appendix C.2 and C.3 respectively, the median forecasts become more imprecise. The interval in which the forecasts lies increases substantial which makes it useless to indicate an accurate range for future observations.

4.2.2 Extended Data Set

Extra variables are added to the baseline data set to see if the forecasting performance can be enhanced using information on new variables. Table 1 and 2 display the results of the extended data set. If these results are compared to the results of the baseline data set, one can observe that the forecasting performance actually deteriorates for all forecasting models.

The QVAR model is significantly outperformed on all horizons by all different methods. The QVAR forecasts on longer horizons are highly imprecise. The forecast values become large due to the fact that past forecasts are recursively involved in the next forecast. The QVAR model performs particularly poor due to the fact of dimensionality. Unimportant or redundant variables cause biased and poor estimations what again causes poor forecasts (Uematsu & Tanaka, 2018). Therefore, QVAR forecasting with high dimensional variables without variable selection results in inaccurate forecasts.

Both quantile regression and QRNN do not suffer from dimensionality and work particular well with high-dimensional data. The QRNN model provides the most accurate forecast on a one-step horizon, but does not perform significantly better than the quantile forest algorithms. The quantile forest algorithm with an estimation window of 10 years performs significantly poorer then the quantile forest algorithm with estimation window 3 and 5 years respectively. This can be explained by the fact that the in-sample period for the extended data set is only approximately 22 years. The estimation window of 10 years trains the quantile forest on two subsamples, whereas the estimation window of 3 and 5 years are trained on 7 and 4 subsamples respectively, what causes more inaccurate forecasts.

From these results, it can be concluded that machine learning methods work well with high dimensional data. Since the in-sample period is only small, the estimation window for the machine learning methods should be small as well. In this way, one can still obtain quite accurate forecasts. The forecasts are more inaccurate for the extended data set compared to the baseline data set. One should however note that the in-sample set of the extended data contains less observations than the in-sample set of the baseline data, which causes more inaccurate forecasts.

Forecast graphs The best performing one-step ahead forecast are again displayed to visualize the 0.1 and 0.9 quantile forecasts. The complete set of forecast plots are displayed in Appendix D. One can observe that the one-step ahead forecasts are quite accurate and that the 0.1-0.9 quantile interval covers new observations with a high probability. The QVAR model is not able to construct a interval that contains future observations. The forecasts become more inaccurate as the horizon increases. The width of the intervals also increases which decreases the usefulness of the intervals.

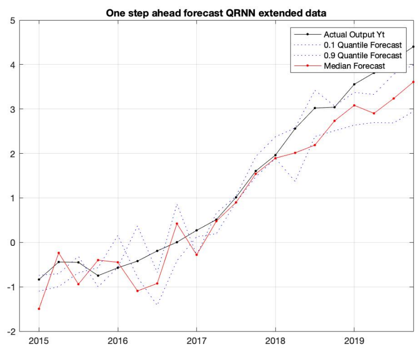


Figure 8: Median, 0.1 quantile and 0.9 quantile one-step ahead QRNN forecasts with the median forecast given in red, actual output in black, and 0.1-0.9 interval in blue dashed lines

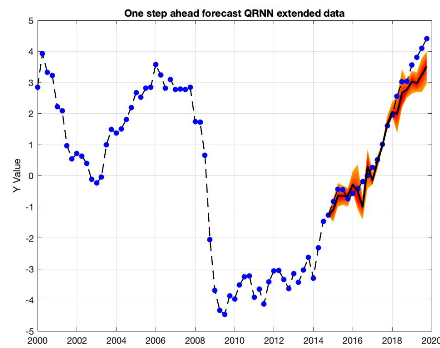


Figure 9: Median, 0.1 quantile and 0.9 quantile one-step ahead forecasts BIC criteria with the 0.1-0.9 interval in orange, median forecast in black, and the actual observations given in blue dots

5 Conclusion

In this research, the non-linearity of government spending shocks is proven through the computation of impulse response functions. These impulse response functions were constructed by applying quantile regression to a VAR model, where the output variable y was estimated in three different quantiles of its conditional distribution. The impulse response functions show that the response of the output variables y indeed differs across quantiles and that shocks are much more persistent when output is at its lowest decile. Information on government spending and other macro-economic variables were then used to forecast the output (real GDP).

Output is forecasted using three forecasting methods: QVAR, quantile regression forest, and QRNN. The advantage of the quantile random forest relative to the VAR-model is, that is works well with high-dimensional data. For that purpose, additional variables are added to the set that was used to construct the quantile specific impulse response functions. The forecast were constructed using an expanding window on forecasts horizons $h = \{1, 4, 12\}$ which resemble the one quarter, one year and three year ahead forecasts.

For the baseline data set, no clear conclusion can be drawn in what method is superior in forecasting. The QVAR model with the number of lags choosen by the BIC criteria performs well on shorter horizon, wheras machine learning methods perform better on longer horizons. The QVAR model with the number of lags chosen by the AIC criteria was significantly outperformed by all forecasting methods. In general, the forecasts as well as the confidence intervals are quite accurate, and one can say that the information on government spending used in Linnemann & Winkler (2016) are suitable to construct forecasts on output.

Additional variables were added to the baseline data set to see if the forecasting performance can be enhanced. The new data set has a shorter time period due to the availability of some variables. The QVAR model performs poor on all forecast horizons. This is due to the fact of dimensionality. Unimportant or redundant variables cause biased and poor estimations what causes again poor forecasts. The quantile forest algorithm and QRNN do not suffer from this issue and provide better forecasts than the QVAR model on this data set. It can be concluded that machine learning methods work well with high dimensional data and are a reliable and easy way to construct forecasts. The estimation should however be chosen carefully, as there should be enough different training samples to train the models on.

In general, one can say that the government spending variable G_t, net taxes ψ_t, interest R_t ,and ratio of government debt held by the public relative to GDP dD_t are able to construct accurate forecasts on the output variable Y_t. Adding extra variables in this research does not improve the forecast performance of the models. This is because of the dimensionality issue in the QVAR model and of less training samples in the extended set. The forecasting performance deteriorates substantially for the QVAR model on the extended data set. A good idea for further research would be to perform forecasts with the QVAR model with performing variable selection first. An idea for variable selection would be LASSO variable selection. Next to LASSO variable selection, one can also think about using factor models to construct forecasts.

Another thing that might be interesting for further research is to see how well the methods perform on a different out-of-sample period. In this research, the years 2015-2020 were taken as out-of-sample period. It would be of interest to see how well the models can perform on another out-of-sample period. Since the models are not compared on different out-of-sample period, it can be coincidentally that the one and four step-ahead forecasts are quite accurate. Further investigation into different out-of-sample periods can make models robust to other periods.

References

Auerbach, A., & Gorodnickenko, Y. (2012). Measuring the output responses to fiscal policy. *American Economic Journal: Economic Policy*, 1-27.

Blanchard, O., & Perotti, R. (2002). An empirical characterization of the dynamic effects of changes in government spending and taxes on output. *Quarterly Journal of Economics*, 1329-68.

Bliemel, F. (1973). Theil's forecast accuracy coefficient: A clarification. *Journal of Marketing Research.*

Breiman, L. (2001). Random forests. *Machine Learning*, 45:5-32.

Cecchetti, S., & Li, H. (2008). Measuring the impact of asset price booms using quantile vector autoregressions.

Koenker, R., & Basset, G. (1978). Regression quantiles. *Econometrica*, 33-51.

Linnemann, L., & Winkler, R. (2016). Estimating nonlinear effects of fiscal policy using quantile regression methods. *Oxford Economic Papers*, 1120-1145.

Meinshausen, N. (2006). Quantile regression forests. *Journal of Machine learning*, 983-999.

Ramey, V. (2011). Identifying government spending shocks: it's all in the timing. *Journal of Economics*, 1-50.

Stock, J., & Watson, M. (1999). Comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. *Cointegration, Causality and Forecasting*.

Taylor, J. (2000). A quantile regression neural network approach to estimating the conditional density of multiperiod returns. *Journal of Forecasting*.

Theil, H. (1966). *Applied economic forecasting.* Chicago: Rand McNally.

Uematsu, Y., & Tanaka, S. (2018). High-dimensional macroeconomic forecasting and variable selection via penalized regression. *The econometric journal*.

Zhang, G., Eddy Patuwo, B., & Hu, M. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*.

Zivot, E., & Wang, J. (2006). *Vector autoregressive models for multivariate time series.* New York, NY: Springer.

Appendices

A Data

Table 3: Transformation and ID's data series, with QD as log-quadratic detrending

Short	Series Title	Construction	Series ID
G_t	Government Spending	QD	GDPC1
Y_t	Output	QD	GCEC96
ψ_t	Net Taxes	QD($\frac{W54+W782-A084}{GDPDEF}$)	W054RC1Q027SBEA, A084RC1Q027SBEA, W782RC1Q027SBEA
R_t	Real Interest Rate	$\frac{FEDFUNDS}{100} - log(\frac{GDPDEF}{GDPDEF(-1)}) * 4$	FEDFUNDS, GDPDEF
D_t	Government Debt held by Public to GDP Ratio	QD, Interpolated	FYPUGDA188S
IPT_t	Industrial Production: Total	QD	IPB50001N
IPM_t	Industrial Production: Manufacturing	QD	IPMAN
$IPEG_t$	Industrial Production: Electric and Gas Utitilies	QD	IPUTIL
$IPCG_t$	Industrial Production: Consumer Goods	QD	IPCONGD
PCE_t	Personal Consumption Expenditures	QD	PCEPI
UN_t	Unemployment Rate	UNRATE/100	UNRATE
AVH_t	Average Weekly Hours Worked	AWHMAN/100	AWHMAN
$MANIS_t$	Manufacturing: Inventories to Sales Ratio	MNFCTRIRSA/100	MNFCTRIRSA
$MANDG_t$	Manufacturing: Durable Goods	QD	DGORDER
$MANNDG_t$	Manufacturing: Non-Durable Goods	QD	IPG311A2S
$MERW_t$	Merchant: Wholesales	QD	WHLSLRSMSA
$MERSDG_t$	Merchant: Sales Durable	QD	S423SMM144SCEN
$MERSNDG_t$	Merchant: Non-Durable Sales	QD	S4248SM144SCEN
$RETT_t$	Total Retail Trades	QD	USASARTMISMEI
BIT_t	Total Business Inventories	QD	BUSINV
PI_t	Private Inventories	QD	A371RX1Q020SBEA
PPI_t	Producer Price Index	QD	PPIACO
$M1_t$	Monetary aggregate M1	QD	M1
$M2_t$	Monetary aggregate M2	QD	M2
$M3_t$	Monetary aggregate M3	QD	M3
$Loans_t$	Commercial and Industrial Loans Outstanding	QD	TOTCI

Table 4: Descriptive statistics of the baseline data set ranging from 1995Q1 to 2019Q4

Row	Mean	Var	Std	Min	Max
Government Spending G_t	-0.012	26.316	5.130	-8.989	13.772
Output Y_t	-0.004	10.637	3.261	-9.125	6.465
Taxes ψ_t	0.006	182.038	13.492	-46.420	31.530
Interest R_t	1.648	6.569	2.563	-2.681	10.120
Public government debt ratio D_t	-0.062	361.058	19.002	-34.701	36.558

Table 5: Correlation coefficients of the baseline data set ranging from 1955Q1 to 2019Q4

	G_t	Y_t	ψ_t	R_t	D_t
Government Spending G_t	1.000	0.501	0.124	-0.134	-0.081
Output Y_t	0.501	1.000	0.694	-0.095	-0.503
Taxes ψ_t	0.124	0.694	1.000	0.156	-0.214
Interest R_t	-0.134	-0.095	0.156	1.000	0.141
Public government debt ratio D_t	-0.081	-0.503	-0.214	0.141	1.000

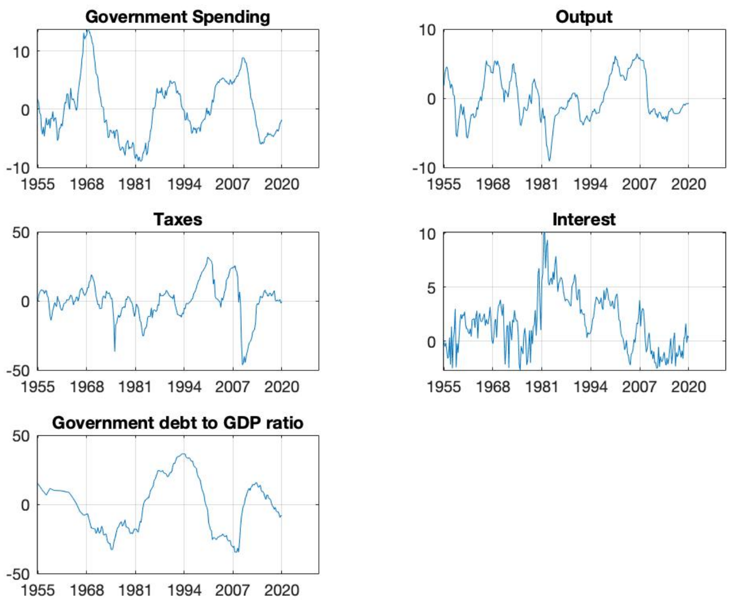


Figure 10: Time series plot of baseline variables ranging from 1955Q1 to 2019Q4

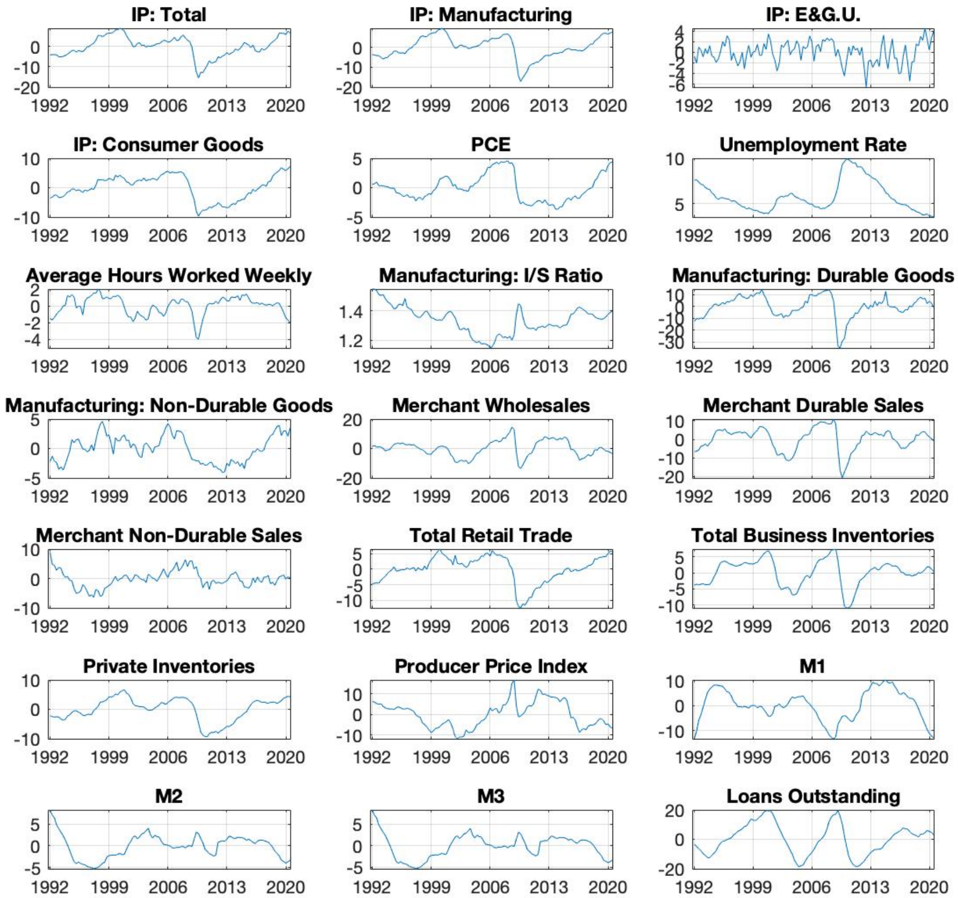


Figure 11: Time series plot of the different variables ranging from 1992Q2 to 2019Q4

Table 6: Descriptive statistics of extended data set ranging from 1992Q2 to 2019Q4

	Mean	**Var**	**Std**	**Min**	**Max**
G_t	-0.066	10.391	3.223	-5.652	6.434
Y_t	0.004	5.632	2.373	-4.470	4.402
ψ_t	0.123	301.304	17.358	-48.733	25.864
R_t	0.708	4.592	2.143	-2.529	4.938
D_t	0.087	213.638	14.616	-26.585	25.595
INT_t	0.042	25.317	5.032	-15.510	9.043
$INDM_t$	0.040	30.580	5.530	-17.170	9.667
$INDEG_t$	0.009	4.244	2.060	-6.556	4.509
$INDCG_t$	0.046	17.727	4.210	-9.710	7.193
PCE_t	-0.007	5.244	2.290	-3.678	4.586
UN_t	5.799	2.696	1.642	3.533	9.933
AVH_t	0.018	1.273	1.128	-3.954	2.021
$MANIS_t$	1.338	0.009	0.093	1.147	1.550
$MANDG_t$	0.083	87.440	9.351	-35.130	14.721
$MANNDG_t$	0.026	5.141	2.267	-4.097	4.612
$MERW_t$	-0.021	30.518	5.524	-13.628	14.485
$MERSDG_t$	0.068	38.777	6.227	-20.809	10.897
$MERSNDGT_t$	-0.106	8.185	2.861	-6.393	8.931
$RETT_t$	0.041	18.618	4.315	-12.470	6.497
BIT_t	0.035	16.678	4.084	-10.906	7.369
PI_t	0.017	15.468	3.933	-9.267	6.697
PPI_t	-0.052	39.344	6.272	-11.789	16.595
$M1_t$	0.146	37.426	6.118	-13.183	10.246
$M2_t$	-0.086	7.507	2.740	-5.291	8.328
$M3_t$	-0.086	7.473	2.734	-5.328	8.306
$Loans_t$	0.008	109.650	10.471	-18.729	19.598

B Methods

B.1 Lag Selection

Table 7: Number of lags to include according to the information criteria

Information Criteria	baseline set	extended set
AIC(n)	5	3
HQ(n)	2	3
BIC(n)	1	3
FPE(n)	5	3

Table 8: Values information criteria baseline data set

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
AIC(n)	3.508	3.394	3.338	3.406	3.132	3.174	3.282	3.415	3.358	3.430
HQ(n)	3.692	3.731	3.828	4.049	3.928	4.123	4.385	4.671	4.767	4.993
BIC(n)	3.964	4.229	4.553	5.000	5.105	5.527	6.015	6.527	6.850	7.302
FPE(n)	33.387	29.802	28.194	30.221	23.029	24.105	26.996	31.046	29.582	32.138

Table 9: Values information criteria extended set

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
AIC(n)	-42.019	-62.861	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
HQ(n)	-33.693	-46.517	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
SC(n)	-21.267	-22.126	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
FPE(n)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

B.2 Forecasting Details

Table 10: Overview of the in and out-sample periods for both data sets

	In-Sample	**Out-Sample**	$\mathbf{N}_{in}$	$\mathbf{N}_{out}$	**Variables**
Baseline	1955Q1-2014Q4	2015Q1-2019Q4	240	20	5
Extended	1992Q2-2014Q4	2015Q1-2019Q4	91	20	26

Table 11: Overview of the different forecasting methods

Abbreviations	Methods	Lags	Window Size	Number of training samples
QVAR AIC	QVAR	5	n/a	n/a
QVAR BIC	QVAR	2	n/a	n/a
Forest, 3. yrs	Quantile Forest	5	3 years	20
Forest, 5. yrs	Quantile Forest	5	5 years	12
Forest, 10. yrs	Quantile Forest	5	10 years	6
QRNN	QRNN	5	10 years	6
QVAR	QVAR	3	n/a	n/a
Forest, 3. yrs	Quantile Forest	3	3 years	8
Forest, 5. yrs	Quantile Forest	3	5 years	5
Forest, 10. yrs	Quantile Forest	3	10 years	2
QRNN	QRNN	3	5 years	5

C Forecasts Baseline set

C.1 One-step ahead forecasts

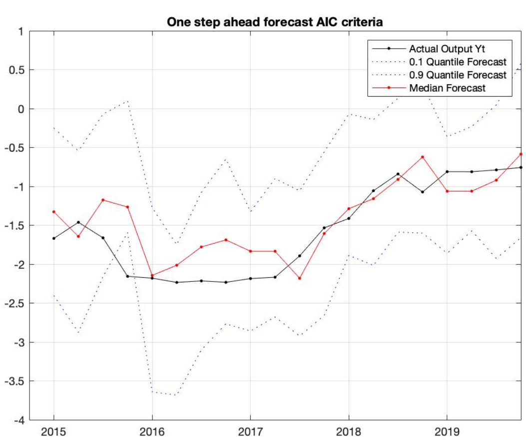


Figure 12: Median, 0.1th quantile and 0.9th quantile one-step ahead forecasts AIC criteria

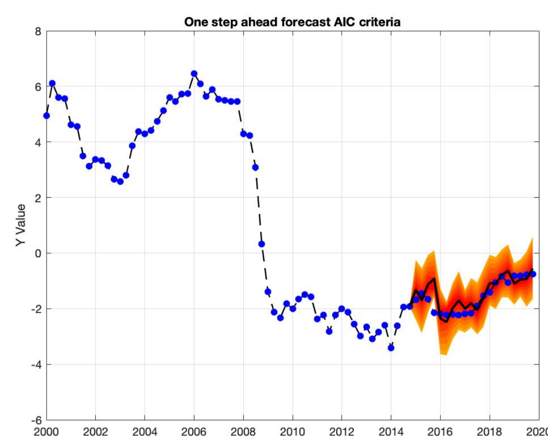


Figure 13: Median, 0.1th quantile and 0.9th quantile one-step ahead forecasts AIC criteria

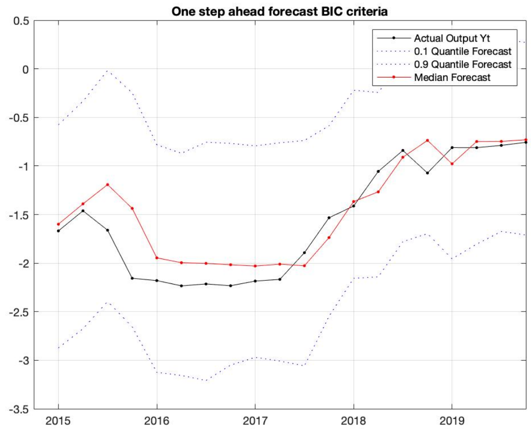


Figure 14: Median, 0.1th quantile and 0.9th quantile one-step ahead forecasts BIC criteria

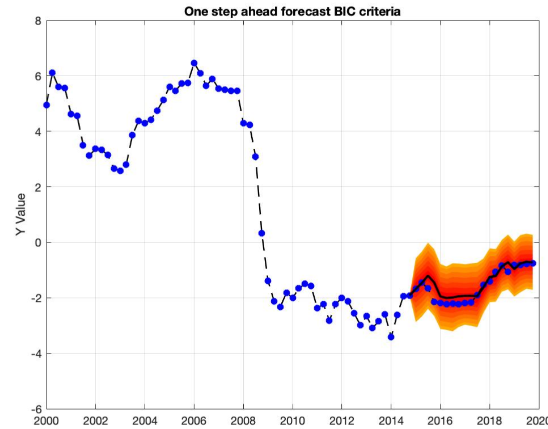


Figure 15: Median, 0.1th quantile and 0.9th quantile one-step ahead forecasts BIC criteria

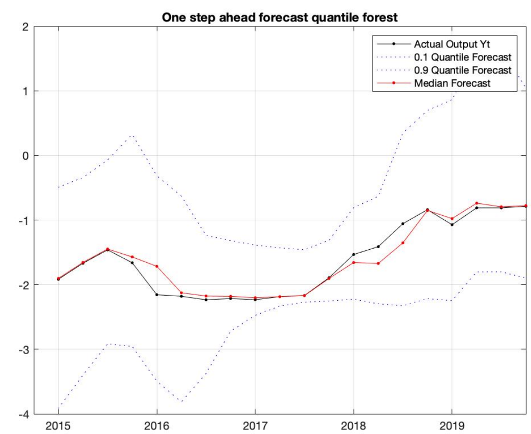


Figure 16: Median, 0.1th quantile and 0.9th quantile one-step ahead quantile forest forecasts, w=3 yrs

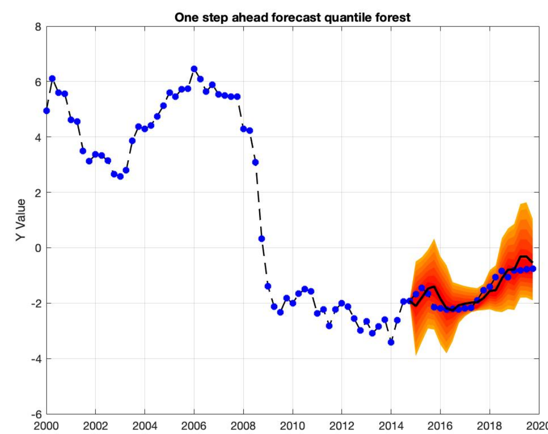


Figure 17: Median, 0.1th quantile and 0.9th quantile one-step ahead quantile forest forecasts, w=3 yrs

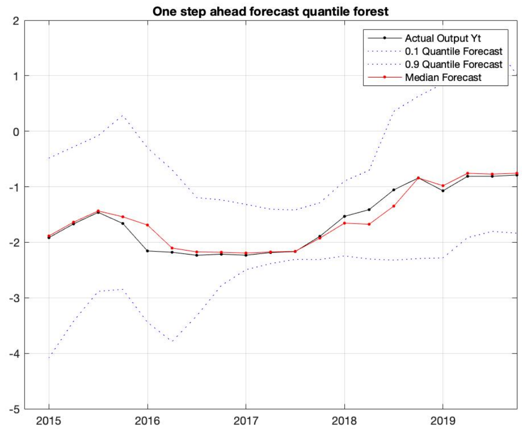


Figure 18: Median, 0.1th quantile and 0.9th quantile one-step ahead quantile forest forecasts, w=5 yrs

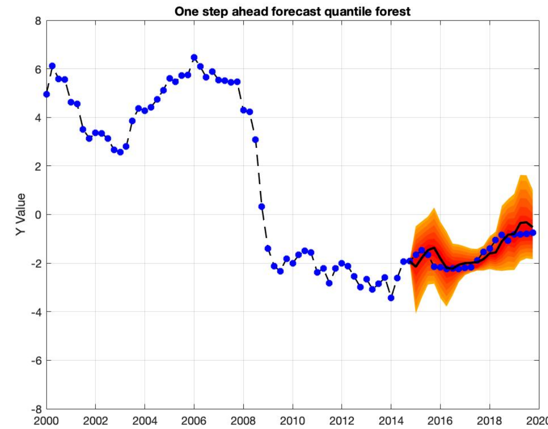


Figure 19: Median, 0.1th quantile and 0.9th quantile one-step ahead quantile forest forecasts, w=5 yrs

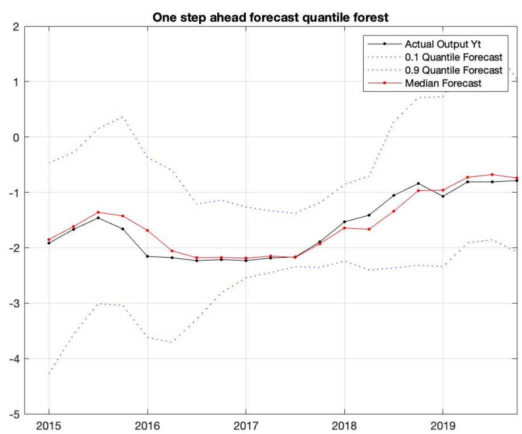


Figure 20: Median, 0.1th quantile and 0.9th quantile one-step ahead quantile forest forecasts, w=10 yrs

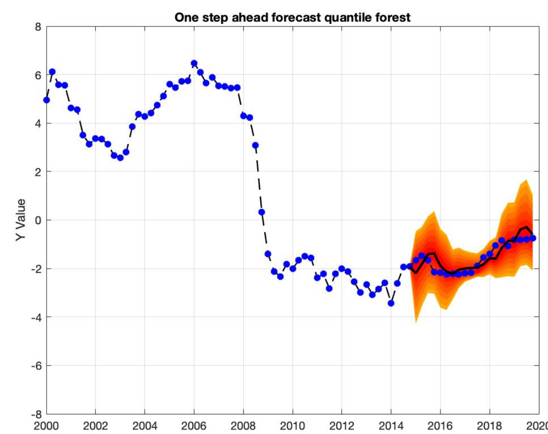


Figure 21: Median, 0.1th quantile and 0.9th quantile one-step ahead quantile forest forecasts, w=10 yrs

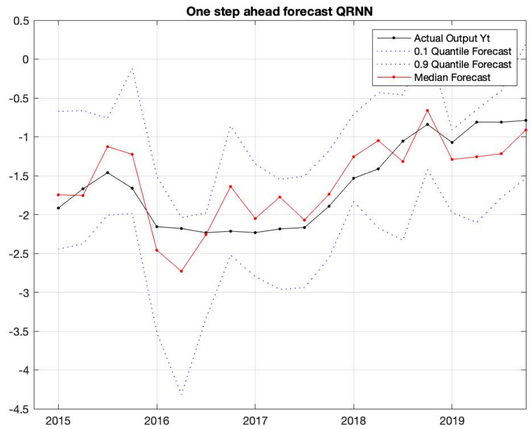


Figure 22: Median, 0.1th quantile and 0.9th quantile one-step ahead QRNN

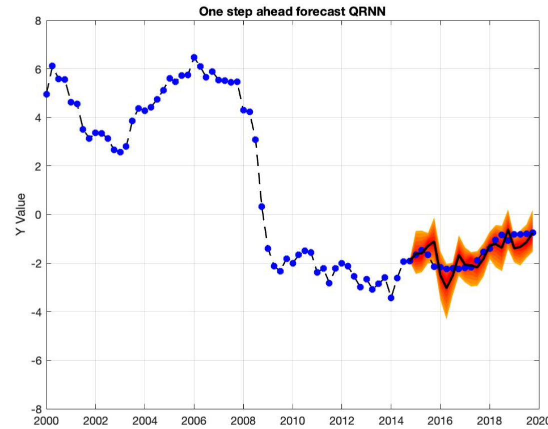


Figure 23: Median, 0.1th quantile and 0.9th quantile one-step ahead QRNN

C.2 Four-step ahead forecasts

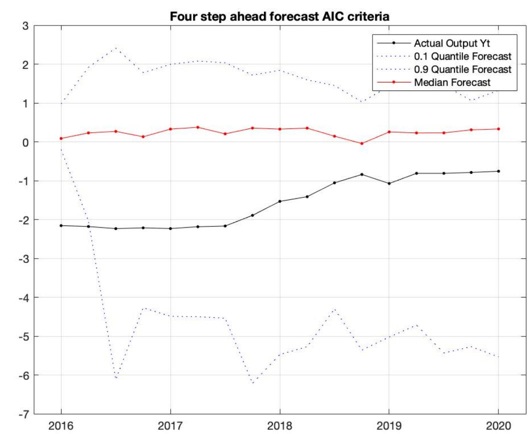


Figure 24: Median, 0.1th quantile and 0.9th quantile four-step ahead quantile forecasts AIC criteria

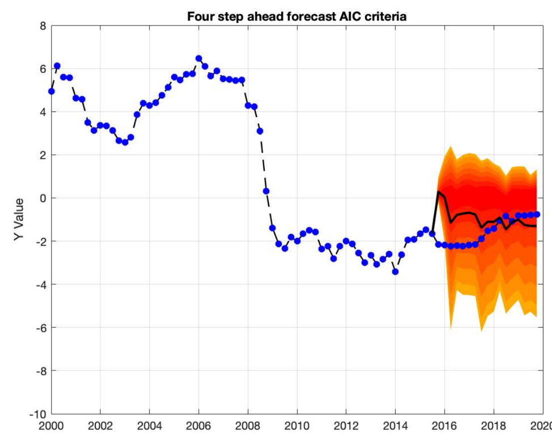
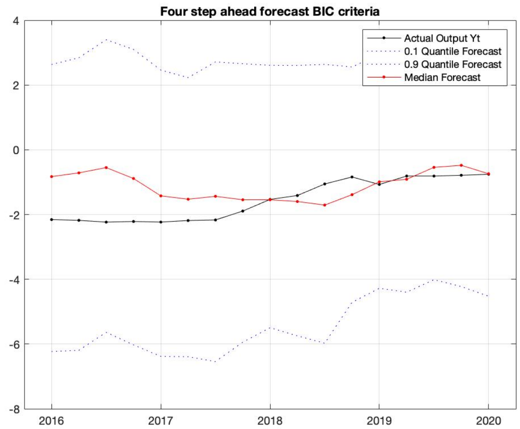


Figure 25: Median, 0.1th quantile and 0.9th quantile four-step ahead quantile forecasts AIC criteria



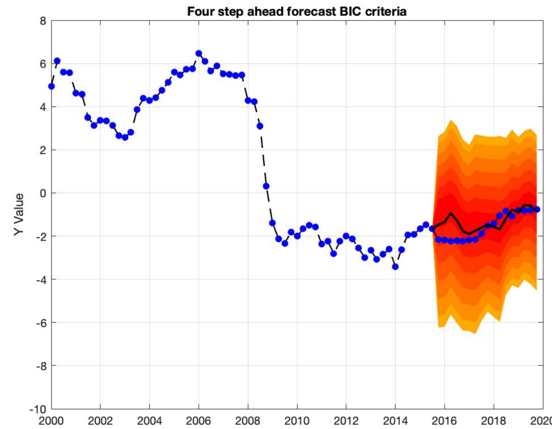
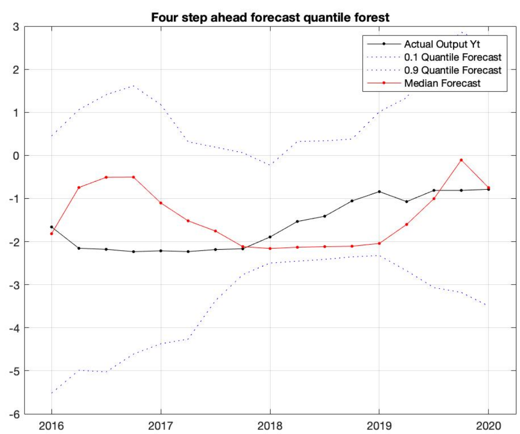


Figure 26: Median, 0.1th quantile and 0.9th quantile four-step ahead quantile forecasts BIC criteria

Figure 27: Median, 0.1th quantile and 0.9th quantile four-step ahead quantile forecasts BIC criteria



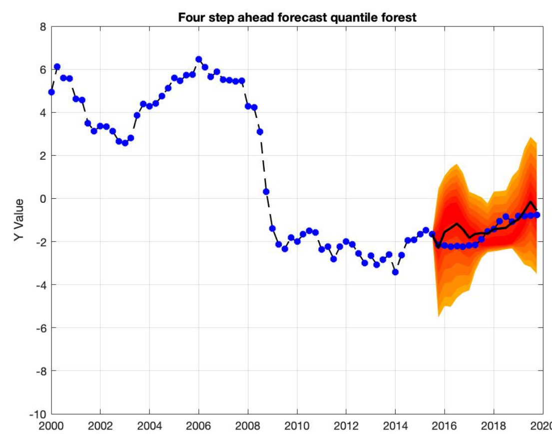


Figure 28: Median, 0.1th quantile and 0.9th quantile four-step ahead quantile forest forecasts, w=3 yrs

Figure 29: Median, 0.1th quantile and 0.9th quantile four-step ahead quantile forest forecasts, w=3 yrs

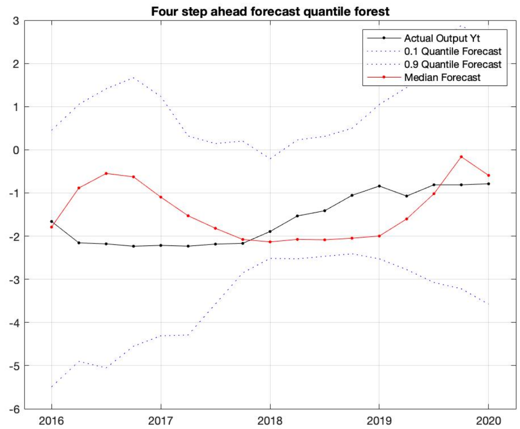


Figure 30: Median, 0.1th quantile and 0.9th quantile four-step ahead quantile forest forecasts, w=5 yrs

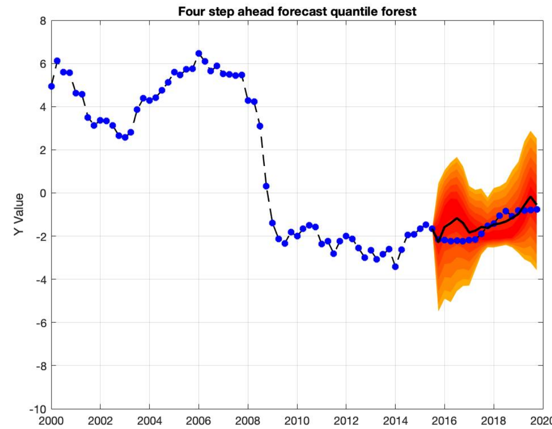


Figure 31: Median, 0.1th quantile and 0.9th quantile four-step ahead quantile forest forecasts, w=5 yrs

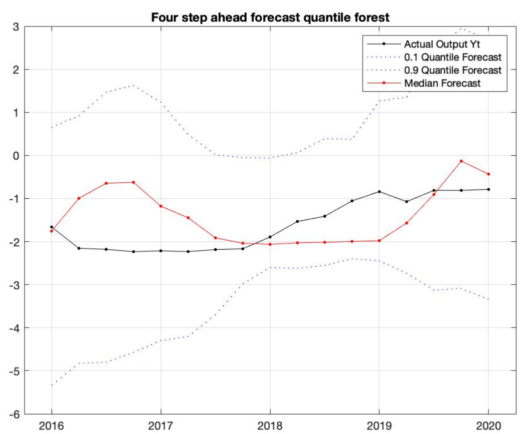


Figure 32: Median, 0.1th quantile and 0.9th quantile four-step ahead quantile forest forecasts, w=10 yrs

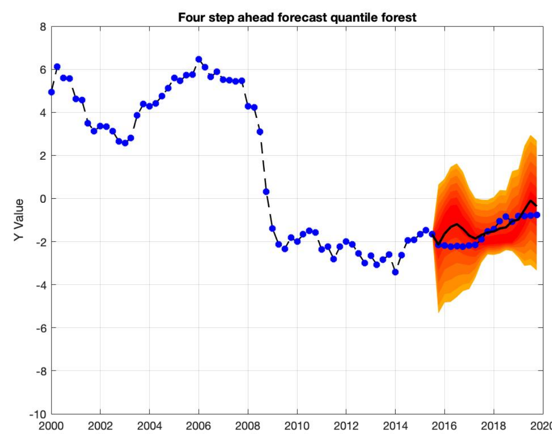
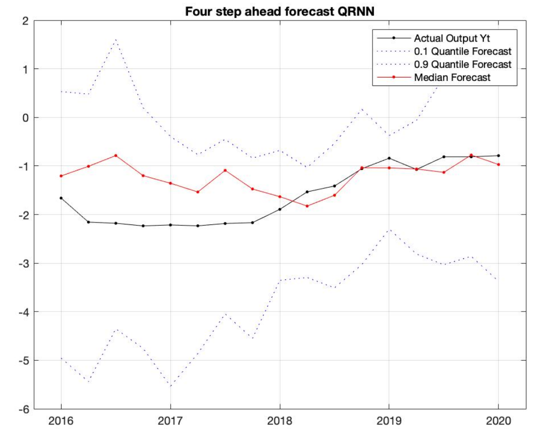


Figure 33: Median, 0.1th quantile and 0.9th quantile four-step ahead quantile forest forecasts, w=10 yrs



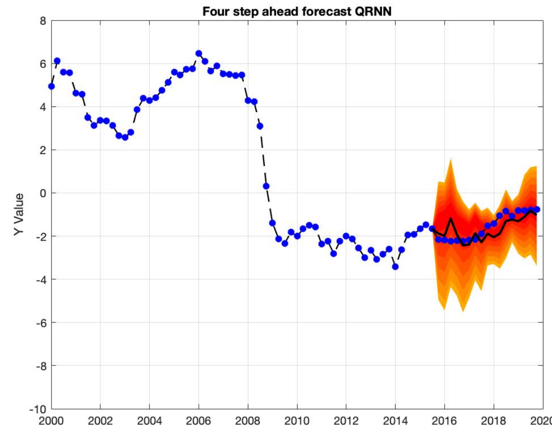
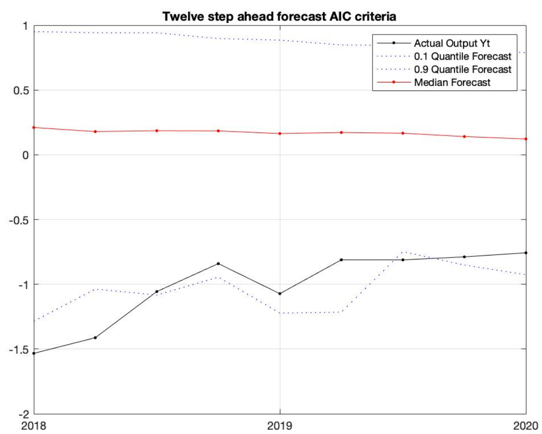


Figure 34: Median, 0.1th quantile and 0.9th quantile four-step ahead QRNN

Figure 35: Median, 0.1th quantile and 0.9th quantile four-step ahead QRNN

C.3 Twelve-step ahead forecasts



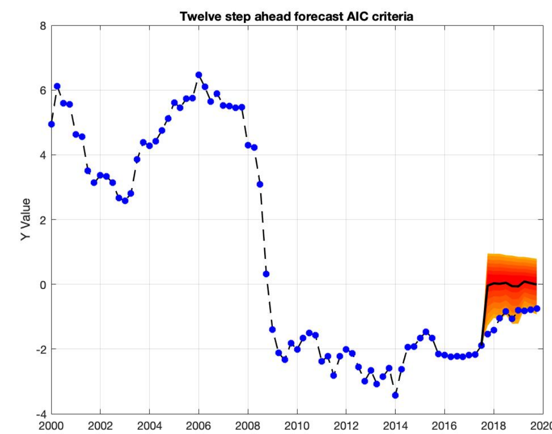


Figure 36: Median, 0.1th quantile and 0.9th quantile twelve-step ahead quantile forecasts AIC criteria

Figure 37: Median, 0.1th quantile and 0.9th quantile twelve-step ahead quantile forecasts AIC criteria

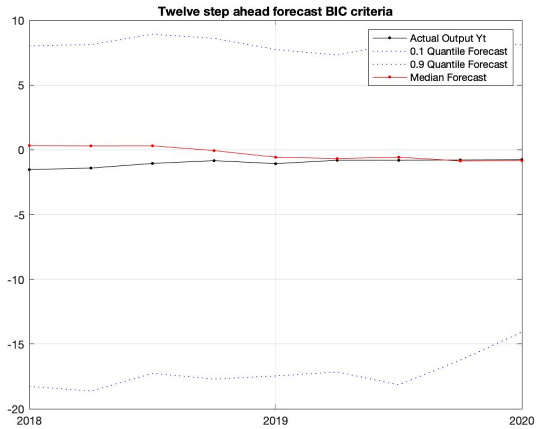


Figure 38: Median, 0.1th quantile and 0.9th quantile twelve-step ahead quantile forecasts BIC criteria

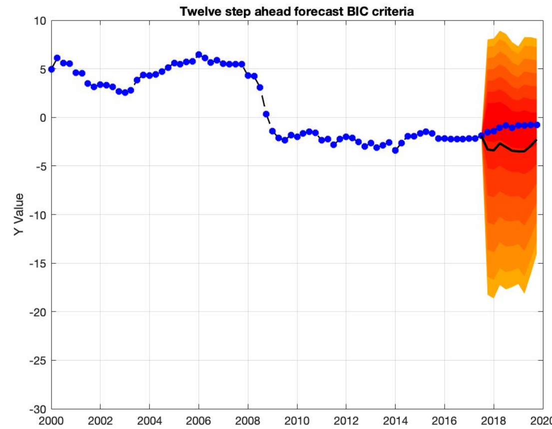


Figure 39: Median, 0.1th quantile and 0.9th quantile twelve-step ahead quantile forecasts BIC criteria

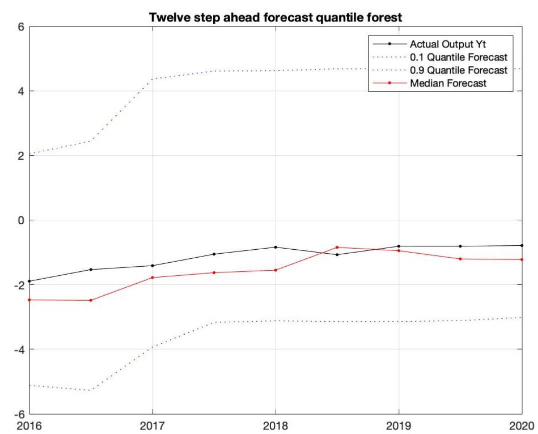


Figure 40: Median, 0.1th quantile and 0.9th quantile four-step ahead quantile forest forecasts, w=3yrs

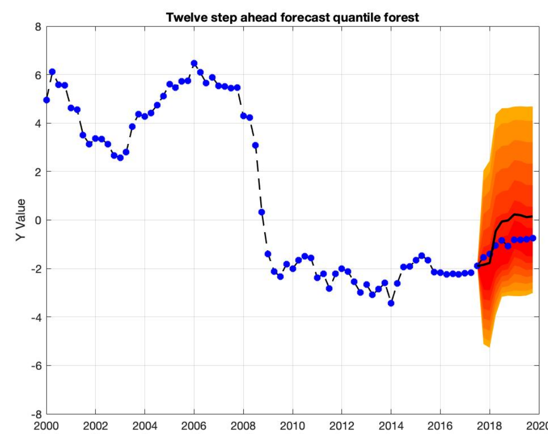


Figure 41: Median, 0.1th quantile and 0.9th quantile twelve-step ahead quantile forest forecasts, w=3yrs

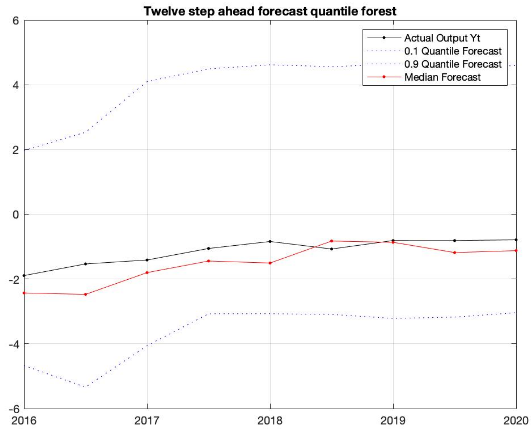


Figure 42: Median, 0.1th quantile and 0.9th quantile twelve-step ahead quantile forest forecasts, w=5yrs

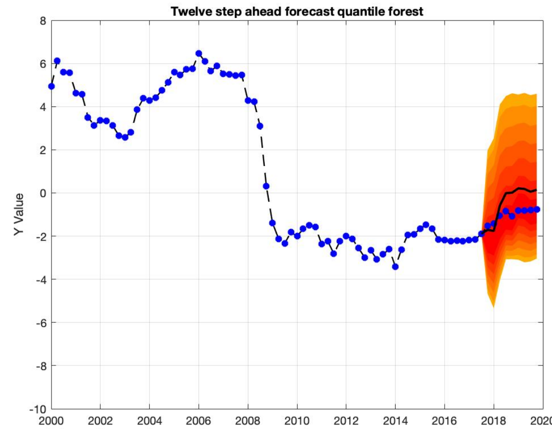


Figure 43: Median, 0.1th quantile and 0.9th quantile twelve-step ahead quantile forest forecasts, w=5yrs

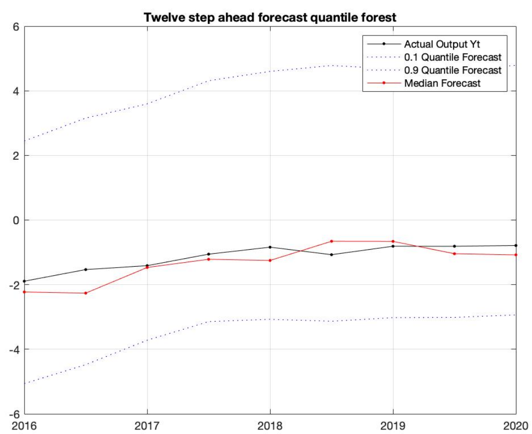


Figure 44: Median, 0.1th quantile and 0.9th quantile twelve-step ahead quantile forest forecasts, w=10yrs

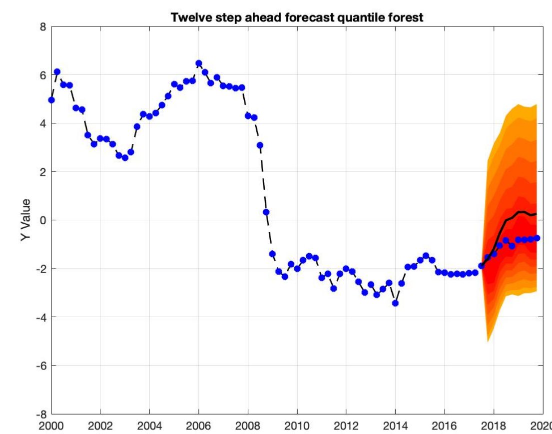


Figure 45: Median, 0.1th quantile and 0.9th quantile twelve-step ahead quantile forest forecasts, w=10yrs

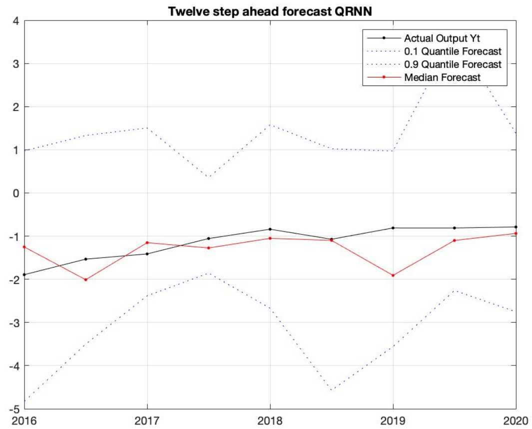


Figure 46: Median, 0.1th quantile and 0.9th quantile twelve-step ahead QRNN

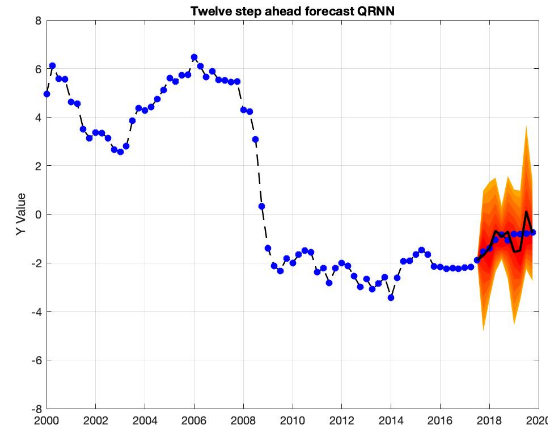


Figure 47: Median, 0.1th quantile and 0.9th quantile twelve-step ahead QRNN

D Forecasts Extended set

D.1 One-step ahead forecasts

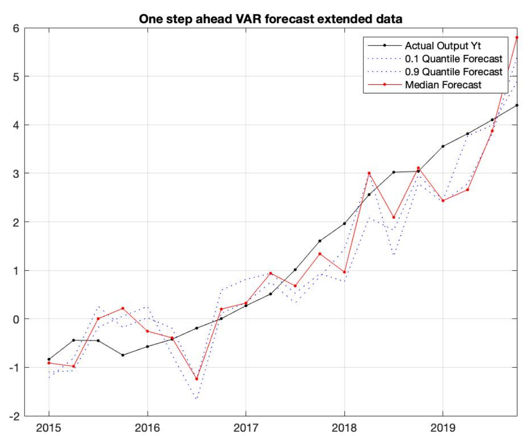


Figure 48: Median, 0.1th quantile and 0.9th quantile one-step ahead quantile forecasts VAR Model

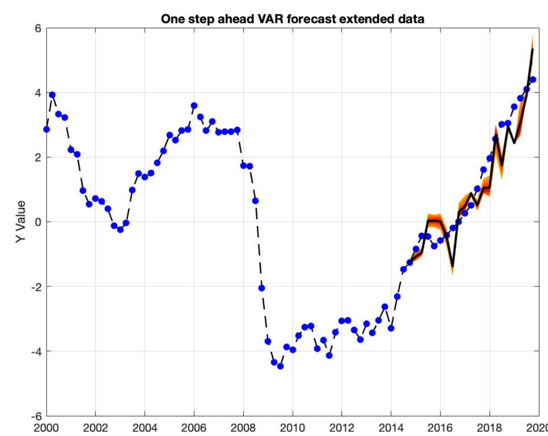


Figure 49: Median, 0.1th quantile and 0.9th quantile one-step ahead quantile forecasts VAR Model

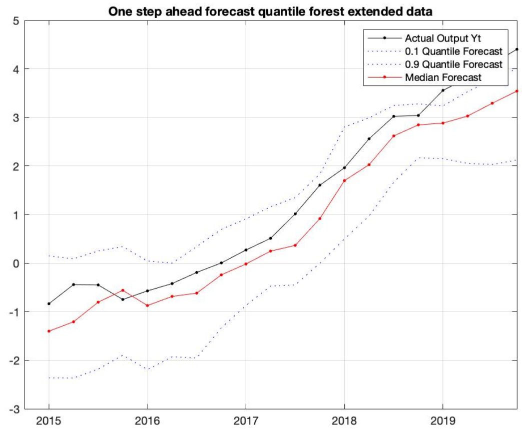


Figure 50: Median, 0.1th quantile and 0.9th quantile one-step ahead quantile forest forecasts, w=3yrs

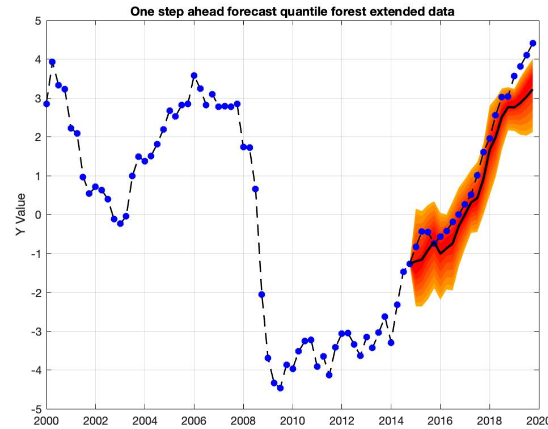


Figure 51: Median, 0.1th quantile and 0.9th quantile one-step ahead quantile forest forecasts, w=3yrs

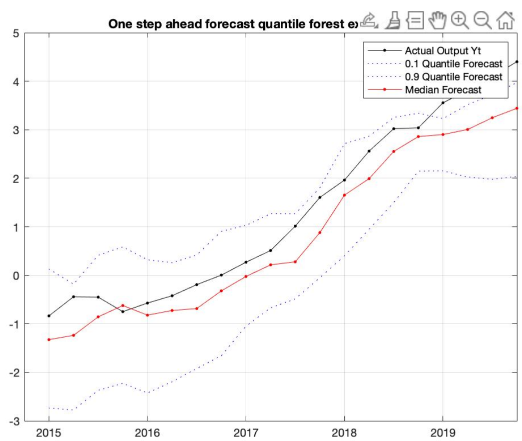


Figure 52: Median, 0.1th quantile and 0.9th quantile one-step ahead quantile forest forecasts, w=5yrs

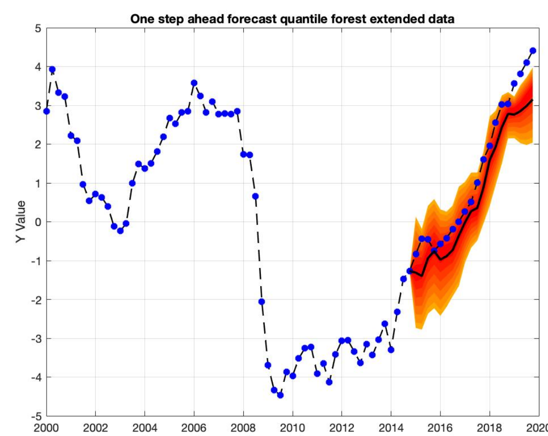


Figure 53: Median, 0.1th quantile and 0.9th quantile one-step ahead quantile forest forecasts, w=5yrs

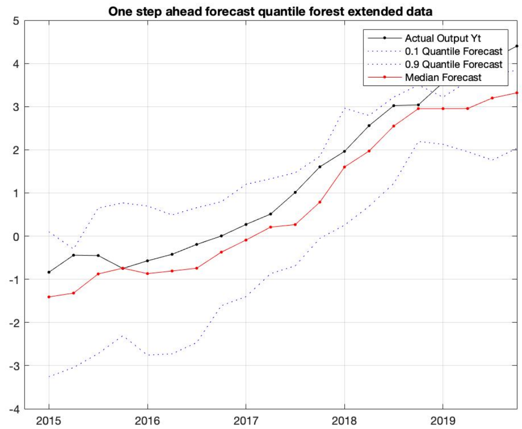


Figure 54: Median, 0.1th quantile and 0.9th quantile one-step ahead quantile forest forecasts, w=10yrs

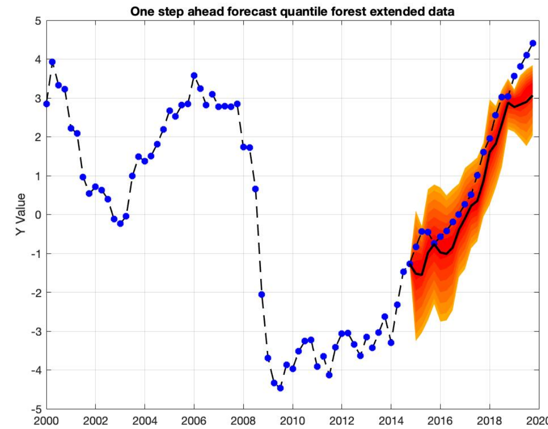


Figure 55: Median, 0.1th quantile and 0.9th quantile one-step ahead quantile forest forecasts, w=10yrs

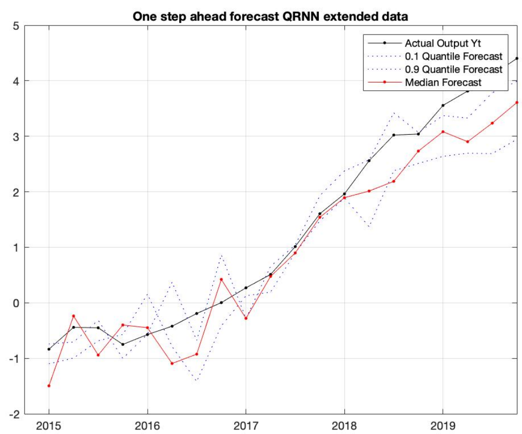


Figure 56: Median, 0.1th quantile and 0.9th quantile one-step ahead QRNN

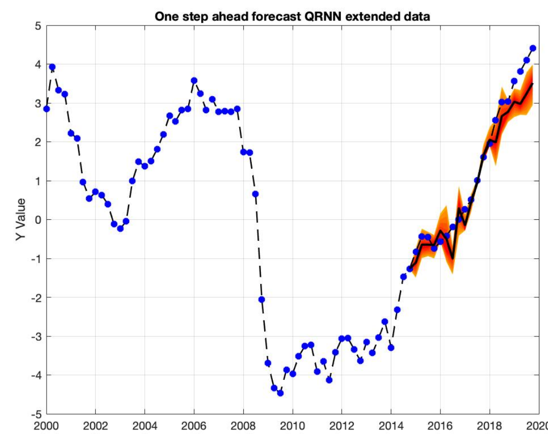


Figure 57: Median, 0.1th quantile and 0.9th quantile one-step ahead QRNN

D.2 Four-step ahead forecasts

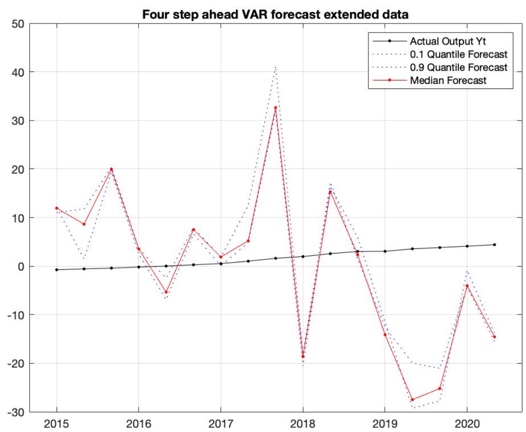


Figure 58: Median, 0.1th quantile and 0.9th quantile four-step ahead quantile forecasts VAR Model

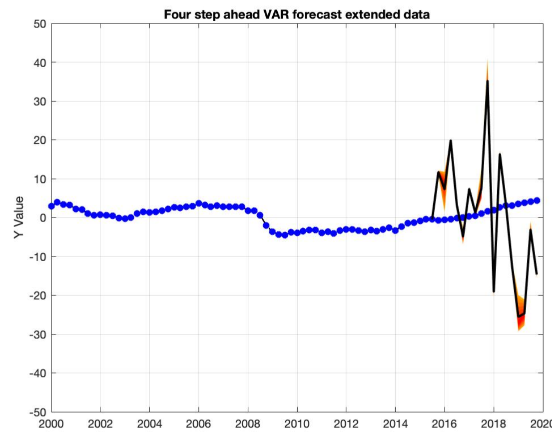


Figure 59: Median, 0.1th quantile and 0.9th quantile four-step ahead quantile forecasts VAR Model

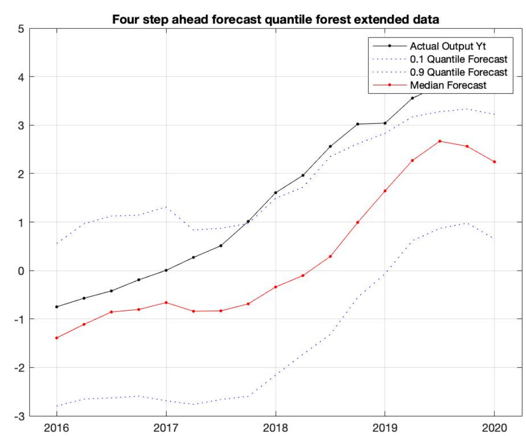


Figure 60: Median, 0.1th quantile and 0.9th quantile four-step ahead quantile forest forecasts, w=3yrs

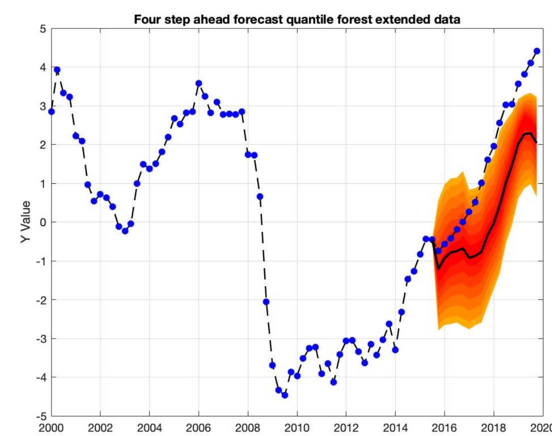


Figure 61: Median, 0.1th quantile and 0.9th quantile four-step ahead quantile forest forecasts, w=3yrs

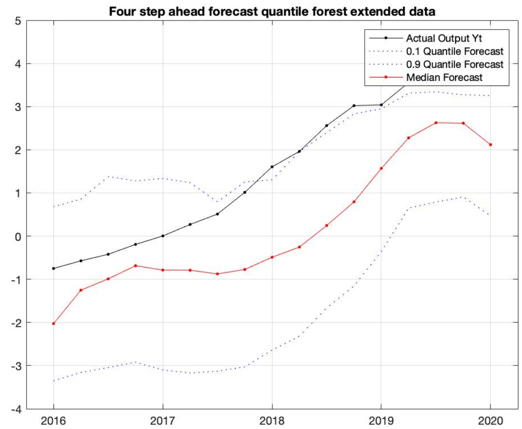


Figure 62: Median, 0.1th quantile and 0.9th quantile four-step ahead quantile forest forecasts, w=5yrs

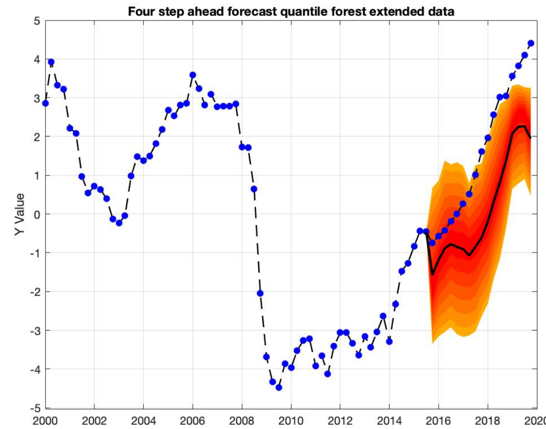


Figure 63: Median, 0.1th quantile and 0.9th quantile four-step ahead quantile forest forecasts, w=5yrs

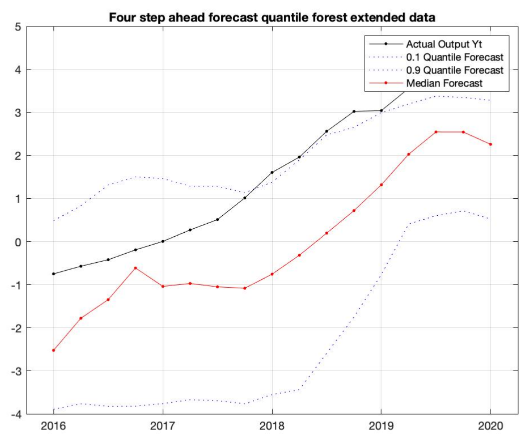


Figure 64: Median, 0.1th quantile and 0.9th quantile four-step ahead quantile forest forecasts, w=10yrs

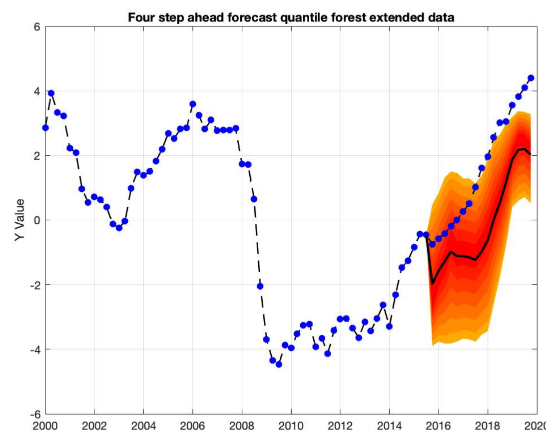


Figure 65: Median, 0.1th quantile and 0.9th quantile four-step ahead quantile forest forecasts, w=10yrs

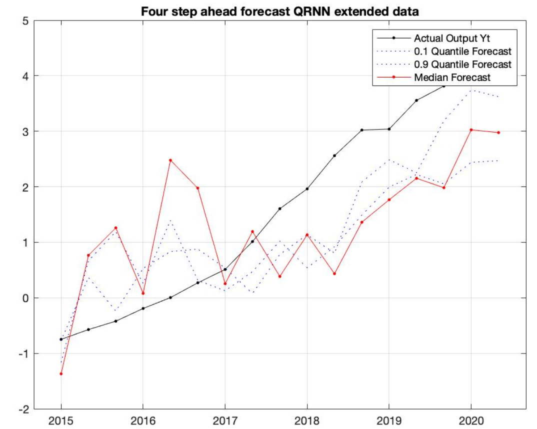


Figure 66: Median, 0.1th quantile and 0.9th quantile four-step ahead QRNN

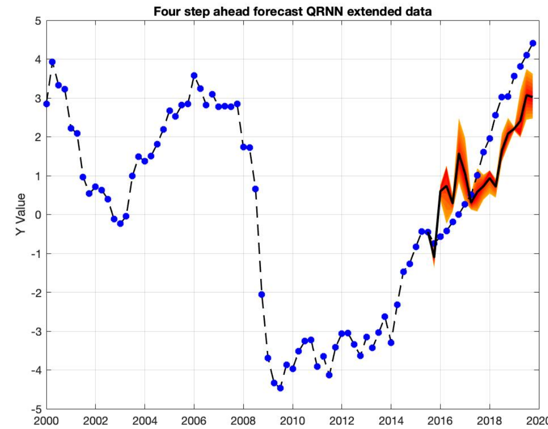


Figure 67: Median, 0.1th quantile and 0.9th quantile four-step ahead QRNN

D.3 Twelve-step ahead forecasts

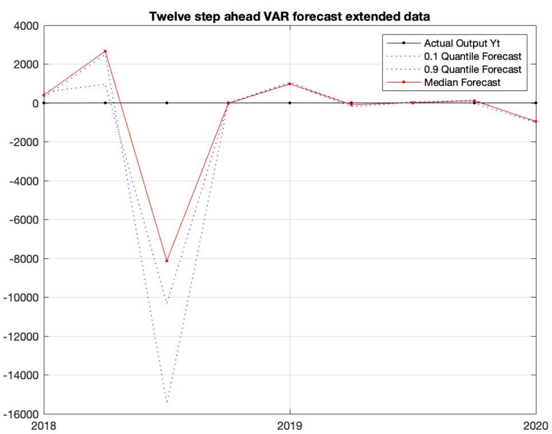


Figure 68: Median, 0.1th quantile and 0.9th quantile twelve-step ahead quantile forecasts VAR Model

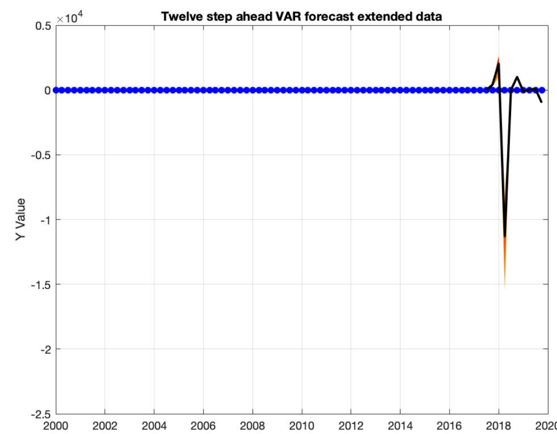
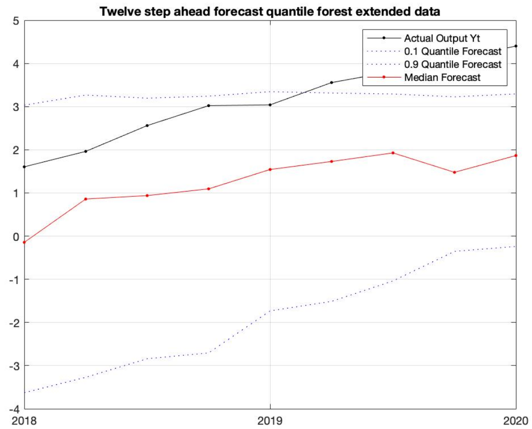


Figure 69: Median, 0.1th quantile and 0.9th quantile twelve-step ahead quantile forecasts VAR Model



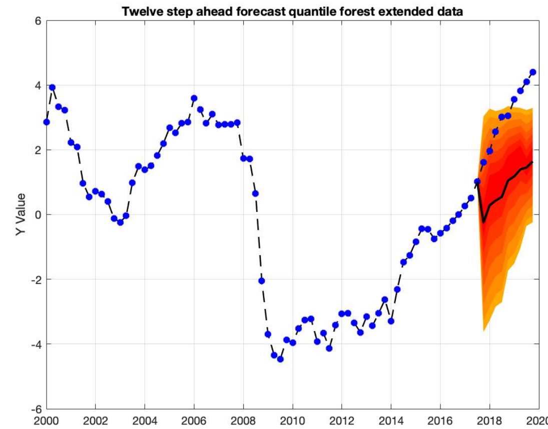
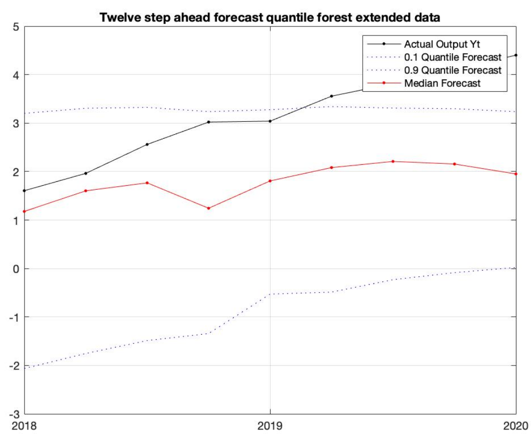


Figure 70: Median, 0.1th quantile and 0.9th quantile twelve-step ahead quantile forest forecasts, w=3yrs

Figure 71: Median, 0.1th quantile and 0.9th quantile four-step ahead quantile forest forecasts, w=3yrs



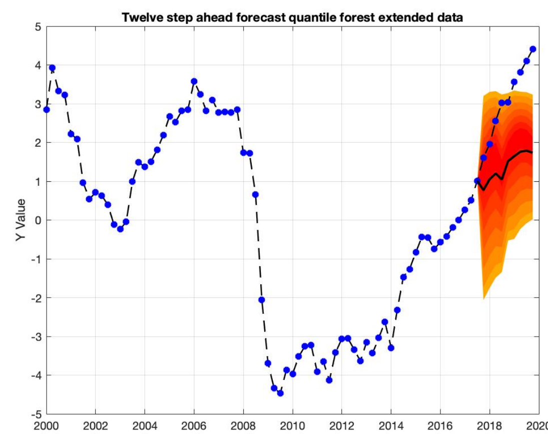


Figure 72: Median, 0.1th quantile and 0.9th quantile twelve-step ahead quantile forest forecasts, w=5yrs

Figure 73: Median, 0.1th quantile and 0.9th quantile twelve-step ahead quantile forest forecasts, w=5yrs

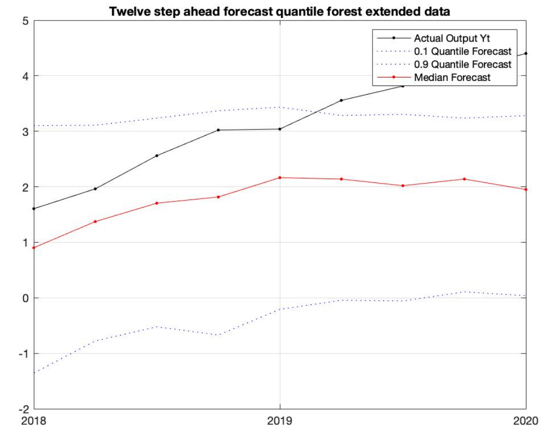


Figure 74: Median, 0.1th quantile and 0.9th quantile twelve-step ahead quantile forest forecasts, w=10yrs

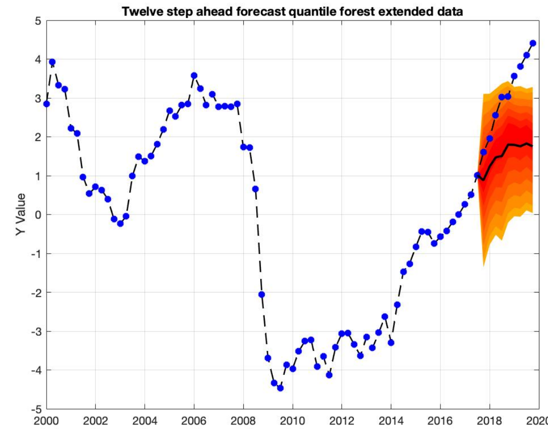


Figure 75: Median, 0.1th quantile and 0.9th quantile twelve-step ahead quantile forest forecasts, w=10yrs

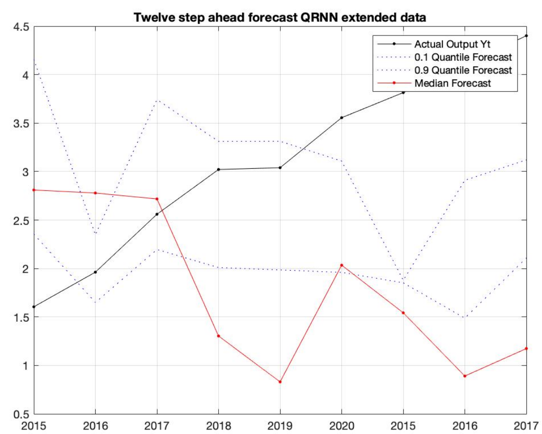


Figure 76: Median, 0.1th quantile and 0.9th quantile twelve-step ahead QRNN

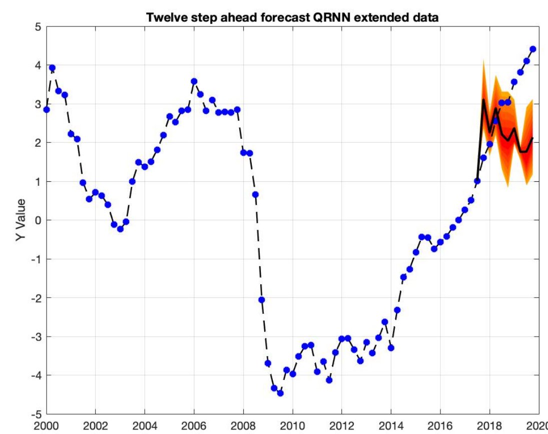


Figure 77: Median, 0.1th quantile and 0.9th quantile twelve-step ahead QRNN