State-dependent forecasting of macroeconomic real activity

Review of Research Findings

Authors:

Karel Räppo

Henry Enno Turu

Supervisor:

Boriss Siliverstovs

Riga

2021

Contents

1 Introduction 3

2 Review of literature 5

2.1 Link between financial markets & real economy . . . 5

2.2 Volatility as a predictor of GDP . . . 5

2.3 Forecasting asymmetries . . . 8

2.4 Advanced methods for macroeconomic forecasting . . . 9

3 Methodology 11

3.1 Linear regressions . . . 11

3.2 Out-of-sample validation . . . 12

3.3 Out of sample validation for different economic cycles . . . 13

3.3.1 Cumulated sum of squared error differential . . . 14

3.4 Machine learning methods . . . 15

3.4.1 Decision Trees . . . 15

3.4.2 Random Forest . . . 16

4 Data & descriptive statistics 18

5 Results 21

5.1 In-sample regressions . . . 22

5.2 Out of sample forecasting . . . 24

5.2.1 Full sample . . . 24

5.2.2 Sub-sample OOS forecasting . . . 25

5.3 Out-of-sample forecasting with Random Forest . . . 25

6 References 27

7 Appendices 30

7.1 Appendix A . . . 30

7.2 Appendix B . . . 31

7.3 Appendix D . . . 32

7.4 Appendix E . . . 33

7.5 Appendix F 35
7.6 Appendix G 36
7.7 Appendix H 38
7.8 Appendix H 39
7.9 Appendix I 40

1 Introduction

Forecasting the real economy has been a widely researched topic as it is crucial for effective policy making. Furthermore, one of the most used channel for forecasting the real economy is through financial markets as these markets (be it equity, fixed income, or commodity markets) incorporate significant amount of forward-looking information with regards to the performance of real economy.

One of the most comprehensive papers quantifying the financial market's ability to forecast macroeconomic real activity was published by Stock & Watson (2003) who reviewed over 93 working papers & articles and documented numerous variables (interest rates, term spreads, returns, exchange rates, etc.) that have been used for forecasting macroeconomy. Nevertheless, there is no strong agreement which can do so consistently, as per the conclusion made: "Some asset prices have been useful predictors of inflation and/or output growth in some countries in some time periods" (Stock & Watson, 2003, p. 822). Furthermore, even less variables are there that can predict macroeconomic real activity through longer time periods (i.e. time periods of over a year).

In a recent work by Cremers et al. (2017), they approach predicting macroeconomic real activity through the treasury implied volatility (YIV) which is derived from options on treasury bond futures of different maturity periods. YIV can be considered as a proxy for interest rate uncertainty as it incorporates market participants' sentiment towards the future outlook for interest rates on which the underlying asset (treasury bond) is dependent on. In the paper the authors show that YIV is able to predict consistently the real economy, especially focusing on predicting the growth of the real GDP as it can be considered as one of the key proxies for the real economy.

Thus, the first part of our research aims to replicate the research by Cremers et al. (2017) - predicting real GDP growth using YIV.

In the second part of the research we build upon the research of Cremers et al. (2017) in two ways.

Firstly, to account for the state-dependency of forecast models, as shown to be relevant by Chauvet & Potter (2013) and Siliverstovs & Wochner (2020), we conduct separate regression analysis for both recessionary periods and expansionary periods (classification of the US business cycles obtain from NBER).

Secondly, we propose using machine learning models to account for the shortcomings of the simple OLS-based models. More specifically, the inability of simple OLS-based models to account for collinearity, dimensionality, predictor relevance and non-linearity (Bolhuis & Rayner, 2020). To elaborate, we apply the framework of a random forest (RF) model as it is able to tackle the aforementioned problems. Thus, the RF model is expected to significantly decrease the root mean square forecasting errors (RMSFE). Nevertheless, it's interpretability is questionable but as it is not the main scope of our research, we do not consider it to be a problem.

Thus, our research question consists of three sub-questions:

1. **Can options on Treasury bond futures effectively predict the USA's macroeconomic and financial activity?**
2. **If so, is this conclusion robust even after taking into account different business cycles.**
3. **Do Machine Learning methods such as Random Forest improve the forecast accuracy?**

The research paper is structured as follows. First we give an overview of existing literature to offer background to our work. As a next step, we describe our dataset and its acquisiton methods, which is followed by methodology. Lastly, we describe the results of our research and connect it with existing research.

2 Review of literature

2.1 Link between financial markets & real economy

Firstly, to dig deeper into various variables that help to predict future macroeconomic conditions, it is essential to understand what is the mechanism that potentially enables using the financial markets for predicting (movements in) the real economy. The base of it being one of the foundations of macroeconomic theory - Fischer theory, according to which the nominal interest rate is equal to the real interest rate and inflation. Hence, in the case of high interest rates and deflation, the outcome should be a recessionary period. Thus, the current stock/bond/commodity market prices include expectations about future performance & earnings - consequently on the aggregate level, the expectations can be generalized to the whole economy.

2.2 Volatility as a predictor of GDP

As the studies about first moments of financial variables have become fairly obsolete, researchers have started focusing on another level - e.g. (implied) volatilities. Fornari & Mele (2019) use the countercyclicality of financial volatility to construct a prediction model and conclude that stock volatility is a significant variable in predicting business cycles. The conclusion is even stronger when combining volatility with term spread - in such a case, their proposed model would have predicted at least 3 of the last recessions. Ferrara et al. (2014) continue the same path - they mix daily financial volatility with monthly industrial production and achieve significant results in predicting GDP; nevertheless their results are limited to the timeframe of Great Recession in 2008-2009.

Similarly to stock prices, option prices reflect future expectations. Taking this into account, David & Veronesi (2014) continue to discover implications of at-the-money (ATM) implied volatility (IV). They find that a positive shock to stock and bonds ATM IV is followed by a decline in future rates and probability of deflation. Thus, there exists a positive relationship between IV shock and possibility of recession.

Building upon the aforementioned various research, Cremers et al. (2017) have analyzed over 20 years of data to find out whether Treasury yield implied volatility can be used to predict different macroeconomic and financial measures such as growth & volatility of GDP, industrial production, employment. Using daily at-the-money option data from 1990 May until 2016 November for different treasury bonds and bills, they calculated daily (treasury) implied volatilities (YIV) using Black's model, which in essence is an adjusted model of the famous Black & Scholes model to value options on future contracts (Black, 1976) - see equation (1).

$$c = e^{-rT} + [FN(d_1) - KN(d_2)] \tag{1}$$

where

$$d_1 = \frac{ln(F/K) + 0.5\sigma_t^2]}{\sqrt{\sigma T}} \qquad d_2 = d_1 - \sqrt{\sigma T} \tag{2}$$

C refers to the price of a call option, F is the price of the underlying future, T is the time to expiration, sigma the volatility of the underlying asset, r is the interest rate. Using the formula and deriving sigma, one can compute the (treasury) implied volatility.

The options are chosen on the basis of exercise price being closest to the price of the underlying bond future - i.e. then its closest to at-the-money. This is done because Ederington & Lee (1993) & (1996) argue that contracts that are closest to at-the-money possess a strong link between spot and future markets. Therefore, these options can be treated as they are options on the bond spot market itself. Furthermore, they argue that those options tend to be the most liquid ones.

After obtaining the daily time series of implied volatility, the authors simply average the daily time series to obtain monthly data, which is regressed with different macroeconomic & control variables.

$$\sum_{j=1}^{j=H} log(1 + GDP_{i,t+j}) = \alpha_H + \beta_H \sigma_{IV,t} + Controls + \varepsilon_{t+H} \tag{3}$$

So, for example, in the main regression described by the formula above, the authors analyze the effect on GDP. In this case, GDP refers to quarterly year-on-year growth rate. It is important to notice that they use the average annualized growth rate that is obtained by dividing the sum of quarterly growth rates with the number of quarters forecasted H.

To validate YIV ability to forecast GDP growth, several control variables such as term spreads, credit spreads, stock market volatility, housing starts, etc. are included in the model. The selection of housing starts (amount of residential property construction projects started) is justified in a research by Ewing & Wang (2005) who find that economic cycles are positively correlated with housing starts. In addition, Cesa-Bianchi et al. (2020) offer a multi-country overview on the connection of realized stock market volatility and real output growth, where they report significant correlation between the two. Even though the authors do not directly use VIX as a measure for volatility, the comovement of VIX and their constructed realized volatility measure is very similar, having correlation of over 90%. Additionally, term spreads have been a key element connected to forecasting recession, which is also verified by Ang et al. (2006) - furthermore, they identify that short rate is the best predictor among all of the term spreads. To complement, Wright (2006) advances the findings further and reports higher performance in forecasting GDP with term spreads and federal funds rates together. Lastly, Gilchrist & Zakrajšek (2012) construct a corporate bond credit spread index which is a significant predictor of macroeconomic activity for different variables & time-horizons.

Even though the authors possess the data for different maturities (1,5,10,30 years). Nevertheless, the research concluded that specifically a 5-year Treasury note significantly predicts most of the aforementioned elements, even after controlling for many other predictors named before. This is in line with the findings of Brandt et al. (2007) who identify that the price discovery tends to mostly happen in contracts with maturity of 5-year, both for cash and future markets.

2.3 Forecasting asymmetries

In the fairly recent wave of research of forecasting, the problems and limitations regarding full-sample forecasting have become more prevalent. More specifically, it has been found that the business cycles have a statistically significant effect on the model's predictive ability. Furthermore, through using state-dependent models, the effects of business cycle asymmetries can be evaluated on the forecasting performance of the model.

For example, in the Handbook of Economics, Chauvet & Potter (2013) evaluate the accuracy of different models with regards to the performance during recessionary periods and expansionary periods using the classifications for US recessions and expansions from National Bureau of Economic Research (NBER). They conclude that for all different models tested, the GDP growth is significantly harder to forecast during recessions when compared to expansionary periods. Furthermore, based on the results they state that although the forecasting ability of some of the models is relatively good during expansions, most of them fail during recessions.

In addition, Siliverstovs (2021) adds to the research by analyzing the impact of influential observations on relative forecast accuracy during Covid-19 crisis. By implementing a newly proposed metric (recursive relative mean squared forecast error) in addition to the cumulative sum of squared forecast error difference (CSSFED) by Welch & Goyal (2008), he concludes that there exist significant differences in relative forecasting error depending on the business cycle.

Therefore, there is reason to believe that the model introduced by Cremers & Fleckenstein & Gandhi might not offer significant information about the future during calm days. Arising from this, we replicate their models while accounting for the effects of different economic phases and see whether it still offers a robust outcome.

2.4 Advanced methods for macroeconomic forecasting

Based on the recent research on forecasting, there is significant evidence on the poor performance of simple autoregressive models when forecasting macroeconomic real activity. In an IMF working paper "Deus ex Machina? A Framework for Macro Forecasting with Machine Learning" published in February 2020, Bolhuis and Rayner bring out the following four key shortcomings of a simple OLS-based forecasting model: Collinearity, Dimensionality, Predictor relevance, and Non-linearity.

To combat some of these shortcomings (collinearity and dimensionality), Siliverstovs & Wochner (2020) assessed the forecasting performance of a dynamic factor model (DFM) compared to a simple autoregressive model. Furthermore, they also included the state-dependent subsamples to their research. The main finding is that there is a significant performance improvement in forecasting capability during the recessionary periods when using the dynamic factor model. Thus, it further illustrates the need for accounting the state-dependency and the importance of using more advanced models (especially during the recessionary periods) over simple autoregressive models to account for the shortcomings of simple OLS-based models.

However, for our research, using a DFM doesn't serve its purpose as we are using a limited set of proven control variables which diminish the problems of collinearity and dimensionality significantly. Furthermore, as brought out by Bolhuis & Rayner (2020), while the DFM-s account for collinearity and dimensional shortcomings, they are still unable to account for the predictor relevance and non-linearity. This in turn increases the root mean square forecast errors (RMSFE) and thus, makes the predictions less accurate.

A solution for this can be found through implementing machine learning (ML) models which have been increasingly taken into use also in applied economics research. Researches have been drawn more and more towards the ML methods in forecasting mainly due to its ability to take into account nonlinearity and its emphasis on out-of-sample forecasting to avoid overfitting which in turn improves the performance with regards to the forecasting accuracy and robustness (Carrasco & Rossi, 2016). Furthermore, through ML methods we will be able to test for predictor relevance to get an overview of how significant variable YIV is in predicting macroeconomic real activity and how does it compare against other academically proven predictors.

To take it together, our paper builds upon Fleckenstein et al while improving the methodology in 2 ways:

1. Addressing the topic of forecast model's state dependency
2. Introducing machine learning method in search of more accurate forecasting performance

Proceedingly, our hypotheses are the following:

Hypothesis 1. The treasury bond implied volatility (YIV) is a significant predictor of future macroeconomic real activity.

Hypothesis 2. Due to the business-cycle related asymmetries, the model based on the full-sample forecasts is inefficient in predicting during turbulent time periods.

Hypothesis 3. Using machine learning methods, it is possible to significantly reduce the RMFSE of the forecast thanks to its ability to take into account collinearity, predictor relevance and non-linearity while also maintaining the interpretability.

3 Methodology

3.1 Linear regressions

The first part of our methodology consists of replicating the methodology conducted by Cremers et al. (2017). In other words, we test whether the treasury implied volatility can be used to predict the future macroeconomic real activity.

To quantify the predictability of macroeconomic real activity using the treasury implied volatility, we run an ordinary least squares (OLS) regression. In the first regression, we take the 5-year YIV and use it to predict the forward-looking GDP growth. To specify, GDP i,t+j refers to logarithmic values of year-on-year quarterly growth rate of the real GDP. H is equal to the periods predicted - e.g. if H=4, it means that we are taking the rolling overlapping average of GDP growth over the 4 quarters.

$$\sum_{j=1}^{j=H} log(1 + GDP_{i,t+j})/H = \alpha_H + \beta_H \sigma_{IV,t} + \varepsilon_{t+H} \tag{4}$$

In the linear models, we compare the predictive ability of YIV within different time periods (H=1,2,3,..,12). Furthermore, to validate YIVs predictive ability, we construct different models (including autoregressive models) by adding in various financial and economical control variables to see whether the significance of our main variable persists. The control variables include term spreads, credit spreads, stock market implied volatility (VIX), number of new residential construction starts (HOUSNG). In order to enable the comparison of the variables ,we standardize all of the independent variables so the mean is equal to 1 and standard deviation to 0. Furthermore, all of our reported coefficients as well as standard errors are adjusted for heteroskedasticity and also autocorrelation (HAC) - to do so, we use the Newey-West methodology with automatic bandwidth selection process. Nevertheless, it is important to mention that the errors are not calculated manually, instead the R package 'sandwich'[1] will be used.

[1] The documentation is available from https://cran.r-project.org/web/packages/sandwich/sandwich.pdf

$$\sum_{j=1}^{j=H} log(1 + GDP_{i,t+j})/H = \alpha_H + \beta_H \sigma_{IV,t} + Controls + \varepsilon_{t+H} \tag{5}$$

Next, to determine the direction of the causality, we run a Vector Autoregressive model (VAR) Granger Causality test on every variable to make sure that e.g. YIV indeed granger-causes movements in GDP.

We also check whether there exist any asymmetries regarding economic cycles - i.e. if YIV exerts a bigger/smaller effect on GDP when the current state is a recession or an expansion. To do so, we introduce a model with a dummy variable that is equal to 1(0) during a recessionary(expansionary) period. The business cycle dating (i.e. recessionary and expansionary periods) for the US economy is taken from the National Bureau of Economic Research (NBER)[2].

$$\sum_{j=1}^{j=H} log(1 + GDP_{i,t+j})/H = \alpha_H + \beta_H \sigma_{IV,t} + Dummy + \varepsilon_{t+H} \tag{6}$$

3.2 Out-of-sample validation

After having done the regression with YIV, subsample dummy and controls as independent variables, we proceed with conducting an out-of-sample test to validate the robustness of the model. To evaluate the robustness of the out-of-sample forecast we use the root mean square forecasting error (RMSFE).

$$\begin{gathered} SFE = \sum_{j=1}^{j=H} (log(1 + GDP_{i,t+j}) - log(1 + \widetilde{GDP}_{i,t+j})^2 \\ RMSFE = \sqrt{mean(SFE)} \end{gathered} \tag{7}$$

To calculate the full model RMSFE we first have to obtain square forecasting errors (SFE).

[2] https://www.nber.org/research/data/us-business-cycle-expansions-and-contractions

For that we construct a predictive model with a 5-year rolling estimation window. Thus, we use the previous 20 quarters to predict the next H quarters (note: here H denotes lead values not rolling averages).The reason behind opting for rolling estimation window is that in this way the predicted datapoint is continously moving; thus the forecast error is not so much dependent on the selected forecast interval (i.e. there would be big differences if one used 85 datapoints to predict 20 datapoints, or 70 datapoints to predict 35 datapoints). From that regression we obtain the predicted values and through incorporating the actual values we can then compute the SFE-s. Having the SFE-s, we take the mean from the values and then take the square root.

3.3 Out of sample validation for different economic cycles

In the second part of our research we build upon the research conducted by Cremers et. al. We test whether the model holds when accounting also for the possible business cycle performance asymmetries as suggested by Siliverstovs & Wochner (2020).

More specifically, we compare the root-mean-square-forecasting-error (RMSFE) of the predictive model during recession and expansion with full sample forecast to find out whether full sample forecast's RMSFE forecasts are robust during the expansionary and recessionary subsamples (in other words, we check whether a distributional shift occurs during the recessionary and/or expansionary periods).

To calculate the recessionary subsample RMSFE, we first obtain SFE-s through doing the full regression as described before, however, in calculating RMSFE we take only the SFE-s related to recessions (according to NBER classification). Based on those recessionary SFE-s we calculated the RMSFE. Similarly, in expansionary-only RMSFE calculation we excluded those recessionary SFE-s.

$$rRMSFE = \frac{RMSFE_{subset}}{RMSFE_{full}} \tag{8}$$

To offer a better comparison of the models forecasting performance, we assess the sub-models with expansionary & recessionary data points compared to benchmark model by calculating RMSFE. In the case of the formula having a value smaller than 1 indicates that the subset model does have a superior performance over the fullsample model.

3.3.1 Cumulated sum of squared error differential

As the recessions in comparison with expansionary periods do have less observations, one might start to doubt the conclusion drawn from such few observations. To tackle this, we use the cumulative sum of squared forecast error difference (CSS-FED) proposed by Welch & Goyal (2008). Even though the main use-case for the model is to compare the forecast performance of different models, it has another benefit evn more applicable to our needs - it helps to dissect the forecast error and hence see whether the (dis)improvement of relative performance is due to actual continuous (dis)improvement or it is dependent on few influential observations. Thus, the CSS-FED is applied in order to test how the few observations of recessions impact the whole forecasting error. As mentioned in the literature review, there is reason to believe their influence on the whole sample forecast error is crucial. In other words, recessionary periods RMSFE is more influential and outweighs the importance of expansionary periods RMSFE. The CS-SFED can be calculated according to the formula below:

$$CSSFED = \sum_{t=1}^{T} [(e_{benchmark,t})^2 - (e_{advanced,t})^2] \tag{9}$$

where $\mathrm{e}_{\mathrm{benchmark,t}}$ refers to forecast errors of benchmark model, and $\mathrm{e}_{\mathrm{advanced,t}}$ of the advanced model at time t.

The resulting figure can be plotted over time to visually identify how each observation contributes to the cumulative forecasting error difference - i.e. if each datapoint increases the forecast error difference, the graph shows an upwards trend. This means that the advanced model continuously outperforms the benchmark model. Contrarily, if little difference exists between each observation's error differential, the graph will stay fairly smooth. In the case when one observation's sum of forecasting error difference(SFED) is significantly bigger from the previous ones, big jumps can be identified on the graph - these jumps mark the observations with largest influence on the overall forecasting performance.

3.4 Machine learning methods

As Machine Learning (ML) methods successfully tackle nonlinearities that cannot be accounted for in simple linear models, the forecasting gains tend to be the highest during times of high economic uncertainty Coulombe et al. (2020). As our research includes many different economic crises, we expect to find forecasting gains by utilizing the Macroeconomic Random Forest method. In our research, we focus specifically on a (modified) random forest based method due to its advantages over other ML methods, which will be discussed in the following paragraphs.

3.4.1 Decision Trees

In order to dig deeper into the methodology of random forest, it is crucial to understand one of its core elements - decision trees. Decision tree is a rather straightforward non-parametric algorithm that can be used for both regressions and classifications. The name 'decision tree' derives from the fact that the algorithm is built in a tree-like structure - it recursively splits the whole sample into subsets by following predefined criteria. In our case when the dependent variable is continuous, reducing variance is used as the selection criteria. Thus, the algorithm takes the whole dataset and picks the variables that possess the biggest influence on the dependent variable, splitting the dataset further until stopping criteria is met; so it arrives at the leaf node where the decision is made. The stopping criteria are hyper-parameters defined by the user such as maximum depth, minimum leaf size (no of observations in the leaf node), minimum number of samples, etc (Zhang & Trubey, 2019).

As decision trees are non-parametric, it means that no strong assumptions about the underlying data and its form is made, which in return enables capturing different forms such as non-linearities in the data. However, this makes them subject to overfitting - it might be the case that the algorithm might start capturing random movements (noise) instead of actual meaningful patterns. There are three main ways to tackle this

1) Tuning the hyperparameters
2) Pruning - growing the full tree and then eliminating decision nodes so that the general accuracy preserves
3) Random forest

3.4.2 Random Forest

Decision Trees are not a robust method as the results are extremely dependent on the dataset, even a small change in the initial training data can yield different outcomes. This is the very reason why Random Forest was proposed by Breiman (2001). Random forest itself is an ensemble based supervised learning method. As the name suggests there are two main elements behind it.

First is the randomness part - random forest uses bootstrap aggregation (bagging) to construct random samples of the initial dataset. Furthermore, the randomness is also included in variable selection - it randomly selects variables used for splitting at each node. Secondly, forest, which refers to the aggregation. The conclusion of the final model is reached by aggregating and averaging the output of individual decision trees. These two elements help to tackle overfitting as RF randomly creates a high number of combinations on the basis of which to create splits, which reduces the correlations betweens samples (Zhang & Trubey, 2019). Hence, the final outcome is much more robust & less likely to be subject to overfitting.

To take this process together in algorithmical terms:

1) Through bootstrap aggregation a sample set out of the predefined training data is

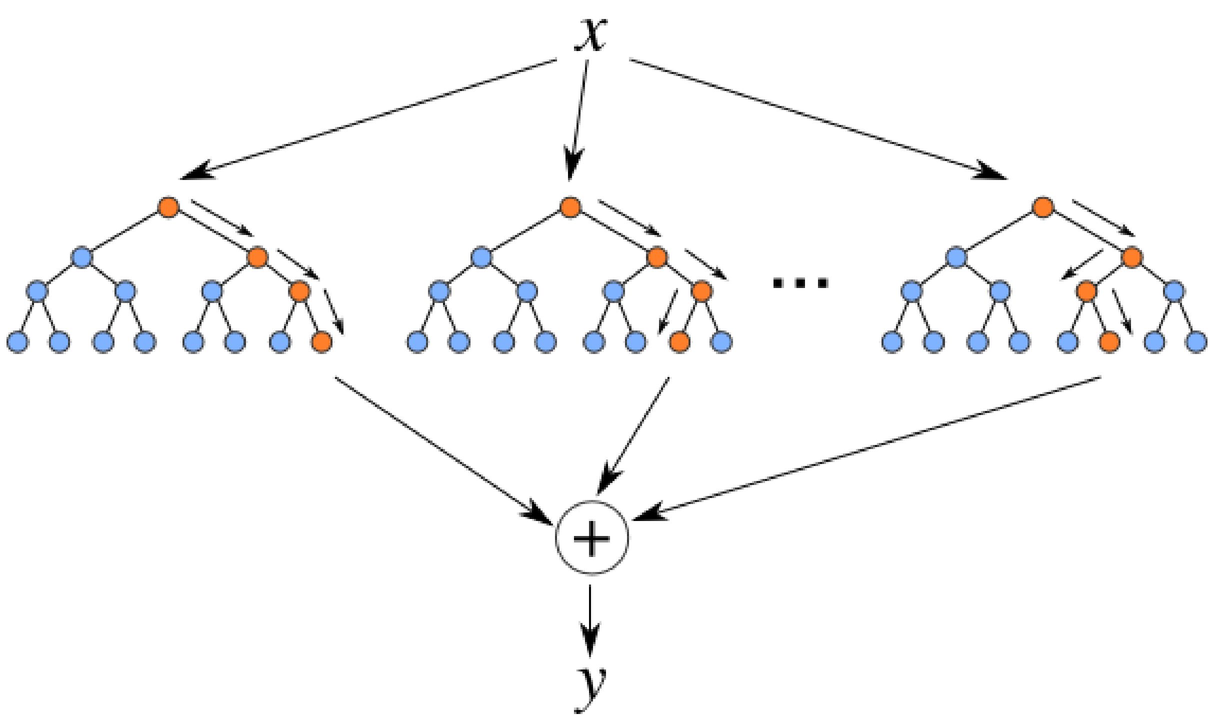


Figure 3.1: Random forest example

created.

2) Then the model randomly selects x number variables amongst the total set X.

3) Subsequently, the best variable and splitting criterion is selected, on the basis of which the current node is splitted into two sub-nodes. 4) More specifically, the choice is made on the basis of mean squared error (MSE) - MSE is minimized at each split.

4) This process is repeated until each terminal node reaches minimum size (by default 5).

5) The output is achieved by averaging the estimation of each tree in the model.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \gamma)^2 \tag{10}$$

4 Data & descriptive statistics

For the regression analysis, we have extracted the following data:

- Quarterly 5-year Treasury Implied Volatility
- Quarterly GDP year-on-year growth rate
- Control Variables

5. US treasury interest rates to construct term spreads
6. credit spreads
7. credit spread index
8. stock returns (SPY)
9. stock market implied volatility (VIX)
10. Residential market construction starts

As mentioned in the review of literature, YIV is constructed using Black model & deriving implied volatility through option prices, time-to-maturity, etc. The data regarding options on treasury bond futures, however, is collected by CME and available only through accessing their database.

Due to the data price being behind the paywall, we had to resort to other measures to access the data. Firstly, we contacted prof. Cremers, Gandhi & Fleckenstein, who were ready to share their data with us if the CME group gives their consent. Nevertheless, as the work is still in the publication process, we could not rely solely on that possibility, and thus resort to other options - extracting data via plot digitizers (read more in 7.1).

The data regarding GDP year-on-year growth rate was extracted on a quarterly basis from Archival FRED (2020). Furthermore, the vintage 2019-12-20 was extracted as it coincided the best with the data used in the research paper by Cremers et al. (2017). In addition, most of the data for constructing control variables were also extracted from the Fred database, such as:

- Quarterly risk free rates (Treasury constant maturity rates for 3 month, 6 months, 1 year, 5 year and 10 year)
- Quarterly corporate bond yields
  - Moody's seasoned AAA corporate bond
  - Moody's seasoned BAA corporate bond
- Quarterly housing (new residential property construction) starts

In addition to individual corporate bonds yields, we have also added corporate bond credit spread index by Gilchrist and Zakrajsek (2012), which was obtained from the official replication dataset OPENCSR(Gilchrist & Zakrajšek, 2019). As per authors, the index Lastly, the daily data for stock market implied volatility (VIX) is taken from CBOE, which is later aggregated into quarterly data (CBOE, 2020). The monthly SPY historical data is obtained from Yahoo Finance, which again is aggregated into quarterly data (NASDAQ, 2020).

We use the risk free rates to compute different term spreads which is defined as the difference

- between 10 year and 12 month treasury constant maturity rate (variable TRM1012).
- between 10 year and 6 month treasury constant maturity rate (variable TRM1006).
- between 10 year and 3month treasury constant maturity rate (variable TRM1003)
- between 5 year and 6 month treasury constant maturity rate (variable TRM0506)
- between 5 year and 3 month treasury constant maturity rate (variable TRM0503)

Furthermore, we construct changes in the short rate that is defined as the change in the rate of treasury 3 month note (SRT03M). In addition to the credit spread index(CRSZGI) and individual corporate yields, we calculate yield spread between AAA and BAA corporate bonds, defined as baa_aaa. All the used variables prior to standardization can be seen in the table below 4.1. It includes summary statistics for main variables used in our research. Statistics include mean, standard deviation,, min, 1st quartile, median, 3rd quartile, max & number of valid data points.

Table 4.1: Summary Statistics

Variable	Mean	Std.Dev	Min	Q1	Median	Q3	Max
Panel A: YIV & GDP							
YIV	3.34	1.31	1.39	2.60	3.00	3.62	9.21
GDP	2.50	1.78	-3.92	1.71	2.61	3.98	5.30
Panel B: Control Variables							
AAA	6.22	1.52	3.46	5.20	6.00	7.43	9.40
DBAA	7.18	1.47	4.50	6.18	7.25	8.22	10.61
baa_aaa	0.96	0.40	0.56	0.70	0.89	1.06	3.00
VIX	19.81	7.35	11.03	14.17	17.56	24.01	58.74
spy_logreturn	6.88	17.42	-53.43	1.02	10.92	18.24	34.95
housng	3.18	51.49	-151.80	-16.80	14.10	36.10	117.70
gz_spr	1.98	1.23	0.79	1.09	1.64	2.59	7.66
TRM1003	1.86	1.13	-0.63	0.84	2.03	2.74	3.61
TRM1006	1.73	1.14	-0.63	0.73	1.88	2.61	3.53
TRM1012	1.59	1.06	-0.36	0.66	1.74	2.52	3.35
TRM0503	1.28	0.83	-0.64	0.61	1.38	1.96	2.88
TRM0506	1.14	0.81	-0.64	0.53	1.25	1.75	2.72
SRT03M	-0.32	1.34	-4.25	-0.77	-0.07	0.10	2.58

Note:
The variables are shown prior to the standardization process.

5 Results

Before diving to details, we plot quarterly YIV with quarterly GDP to see whether visual patterns arise. As it can be seen, the YIV seems to have a negative correlation with GDP. This relationship is especially profound during Global Financial Crisis (around 2008-2009) when YIV surges up a bit before where a big decline in GDP happens. This is confirms our base for analysis as suggested by the literature review that indeed the treasury could have predictive ability over GDP.

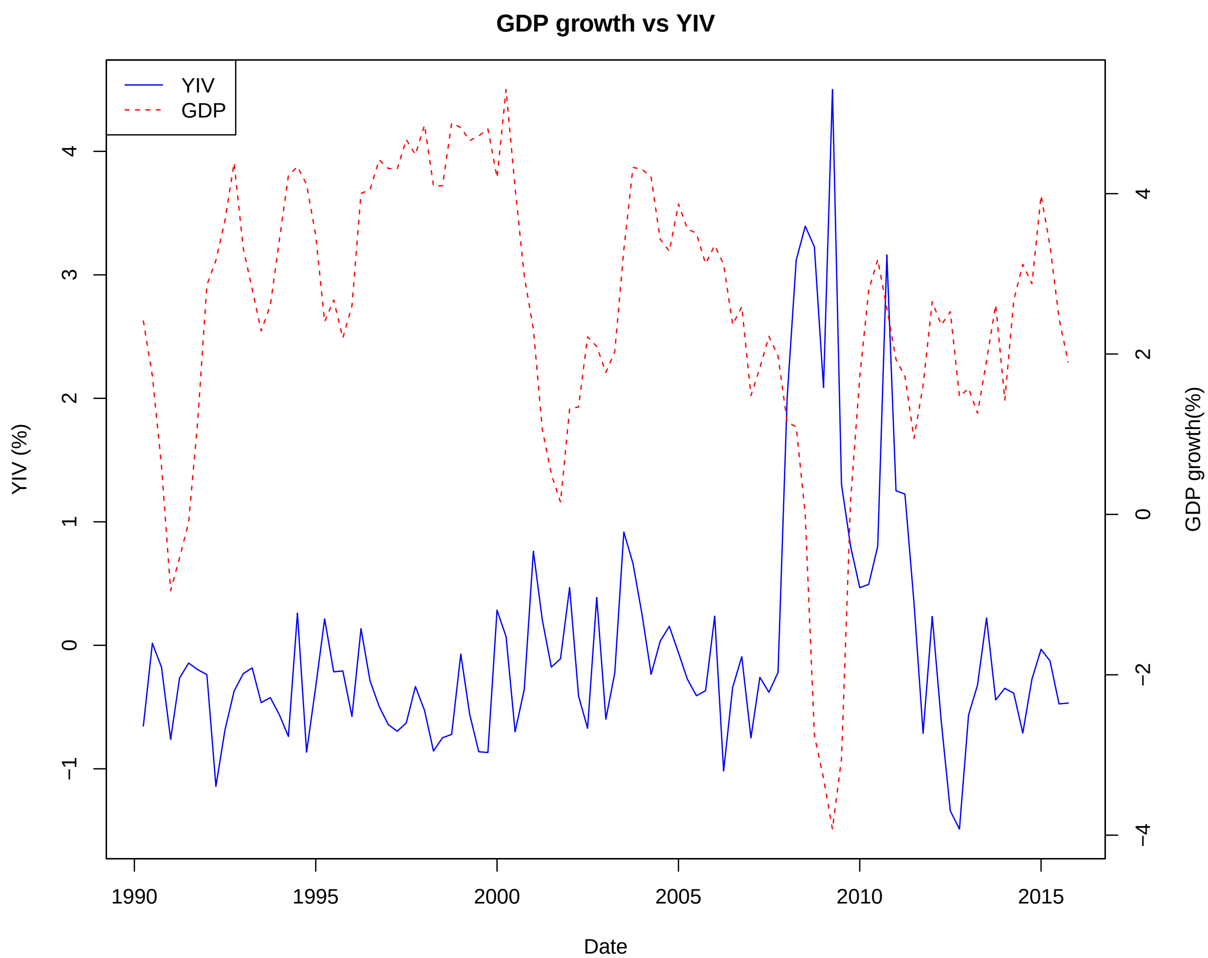


Figure 5.1: GDP Growth(%) vs 5-year Treasury Implied Volatility

5.1 In-sample regressions

We start our analysis by regressing YIV to GDP growth throughout different rolling periods - i.e. 4, 6, 8, 10 and 12 quarter GDP growth rolling averages (the regression formula and summary can be seen in table 5.1). Our regression results implicate that 1 standard deviation increase in YIV is associated with a -1 * 1.31% = 1.31% decrease in gdp growth within next 4 quarters. For predicting 12 quarters ahead (3 years), 1 standard deviation increase in YIV is associated with a -0.54 * 1.31% * 3 = 2.11% decrease in GDP growth. To illustrate the magnitude of this reduction, one should note that the average year-on-year growth rate in our sample is 2,5%.

Furthermore, it can be seen, YIV's coefficients are significant throughout the prediction periods within 1% confidence level. The model's R-squared gradually decreases from 39% in predicting ahead GDP growth 4 quarters' rolling averages to 19% in predicting ahead 12 quarters' rolling averages. These results are consistent with the paper by Cremers et al. (2017) while the only difference is from std. error. Thus, this verifies the validity of our methodology and that the extracted data can be used to proceed with the research process. Nevertheless, as an additional robustness check to ensure that YIV predicts GDP growth and not the other way around, we run the Granger causality test by which we indeed conclude that YIV granger-caused GDP growth.

$$\sum_{j=1}^{j=H} log(1 + GDP_{i,t+j})/H = \alpha_H + \beta_H \sigma_{IV,t} + \varepsilon_{t+H} \tag{11}$$

Table 5.1: Regression output

	H4	H6	H8	H10	H12
YIV_estimate	-1.00	-0.87	-0.71	-0.61	-0.54
YIV_std.error	0.24	0.22	0.18	0.14	0.13
YIV_p.value	0.00	0.00	0.00	0.00	0.00
r.squared	0.39	0.35	0.28	0.23	0.20
adj.r.squared	0.39	0.34	0.27	0.22	0.19

Note:
*** - p<0.01, ** - p<0.05, * - p<0.1. Reported standard error is adjusted for heteroskedasticity

Next, we add a dummy variable representing recessionary periods (according to the NBER classification) i.e. the dummy takes a value of 1 during recession and 0 during expansion table 5.2). As it can be seen from the table, the dummy's coefficient is negative and significant at 1% confidence level throughout all predicted timespans. Furthermore, adding the dummy improved R-squared significantly - 61% in predicting ahead GDP growth 4 quarters' rolling averages when compared to 39% of respective YIV only model. The latter indicates that there is a structural break in data during recessions and that full model predictions during recessionary periods can result in poor prediction accuracy. Nevertheless, the interaction term of the dummy variable & YIV yield insignificant results. Hence, this means that recessionary period only influences the intercept, not slope of the variable.

$$\sum_{j=1}^{j=H} log(1 + GDP_{i,t+j})/H = \alpha_H + \beta_H \sigma_{IV,t} + Dummy + \varepsilon_{t+H} \tag{12}$$

Table 5.2: Regression with state-dependency

	H4	H6	H8	H10	H12
YIV_estimate	-0.54	-0.48	-0.41	-0.40	-0.38
YIV_std.error	0.14	0.13	0.10	0.10	0.10
YIV_p.value	0.00	0.00	0.00	0.00	0.00
dum_estimate	-2.72	-2.31	-1.75	-1.22	-0.93
dum_std.error	0.40	0.34	0.22	0.26	0.28
dum_p.value	0.00	0.00	0.00	0.00	0.00
r.squared	0.61	0.54	0.40	0.30	0.25
adj.r.squared	0.60	0.53	0.39	0.29	0.23

Note:
*** - p<0.01, ** - p<0.05, * - p<0.1. Reported standard error is adjusted for heteroskedasticity

Additionally, we also constructed a model with controls as regressors (Table 7.1). The controls variables were selected among the variables described in the data section. More specifically, the following controls were selected initially: DGS1, DGS5, DGS10, DGS3MO, DGS6MO, TRM0503, TRM0506, TRM1003, TRM1006, TRM1012, STR03M, BAA-AAA, VIX, HOUSNG and GZ_SPR. However, due to multicollinearity problems among DSG and TRM controls we constructed a model for variable selection and selected the best predictor among those groups. For selection criteria, both R-squared and Mallow's Cp were used and according to this we chose DGS10 and TRM1012 as the best predictors out of DGS and TRM control groups. Thus, in the final regression model the following controls were included: DGS1, TRM1012, STR03M, BAA-AAA, VIX, HOUSNG and GZ_SPR.

Finally, we included all the aforementioned independent variables to one regression equation (Table 7.2). As it can be seen, even when including the all the relevant control variables, the YIV still remains a significant predictor throughout all the periods. In the final model, the regression results implicate that 1 standard deviation increase in YIV is associated with a -0.59 * 1.31% = 0.77% decrease in gdp growth within next 4 quarters (1% confidence level). Additionally, when predicting 12 quarters ahead a 1 standard deviation increase in YIV is associated with a -0.33* 1.31% * 3 = 1.30% decrease in the next 12 quarters' average annualized growth rate within a 5% confidence level. The latter regression indicates that, although accounting for many of the most significant variables with regards to the GDP growth prediction, the YIV's predictive ability still persists meaning that YIV can be considered as a solid indicator of GDP growth.

5.2 Out of sample forecasting

5.2.1 Full sample

Lastly, we were not only interested in the in-sample performance of the variable. Hence, we constructed out-of-sample regressions to compute RMSFE-s of full-sample and subsample models (results can be seen in table **??**). For the out-of-sample regressions we used 5 year rolling windows and predicted ahead 1 to 12 quarters. As it can be seen from the table and graph the full-sample out-of-sample RMSFE-s rose steadily up until predicting 9 quarters ahead and then slightly dropped. In other words, the interpretation of it was that YIV-s accuracy in predicting GDP growth got worse in predicting further time periods with a slight improvement from the 10th quarter.

5.2.2 Sub-sample OOS forecasting

Next, we wanted to compare the full-sample OOS RMSFE-s with the subsample OOS RMSFE-s. As it can be seen, during the recessionary period the RMSFE is significantly higher when compared to the full-sample RMSFE. This confirms our second hypothesis i.e. the full-sample is not robust in making predictions during turbulent time periods as the accuracy of the model suffers tremendously. During the expansionary periods the OOS RMSFE improves slightly which is in line with the hypothesis as the inaccuracies of the expansionary periods are filtered out.

Figure 5.2 describes subsample relative RMSFE within different forecasting period. If the value is over 1, it indicates that the benchmark model(full sample model) has superior forecasting performance compared to the corresponding subsample model. Over most of the horizons, the results verify our hypothesis and are consistent with the findings of Siliverstovs & Wochner (2020) - the model with reccessionary subsample has much lower forecasting performance (bigger error) compared to full-sample model, and vice versa for expansionary.

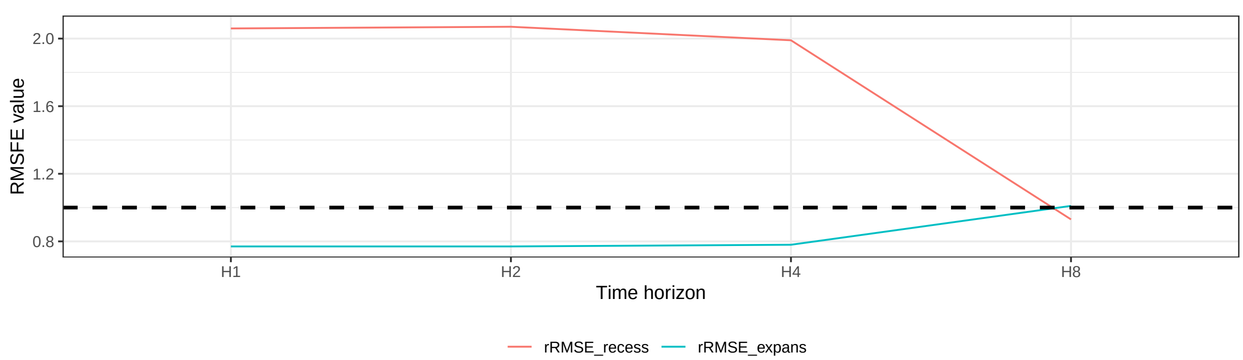


Figure 5.2: Relative Recessionary & Expansionary RMSE

5.3 Out-of-sample forecasting with Random Forest

Lastly, we erected a random forest model to try to improve the forecasting accuracy through having non-linearity. As it can be seen from the graphs (**??**) the performance in some cases is slightly better than OLS (e.g. during recessionary periods), however, is some cases (e.g. during expansionary periods) the model's accuracy is slightly worse. This holds also after tuning the model for number of trees.

6 References

Ang, A., Piazzesi, M., & Wei, M. (2006). What does the yield curve tell us about GDP growth? *Journal of Econometrics*, *131*(1-2), 359–403. https://doi.org/10.1016/j.jeconom.2005.01.032

Archival FRED. (2020). *Dataset. (Vintage 2020-20-12)*. https://alfred.stlouisfed.org/

Black, F. (1976). The pricing of commodity contracts. *Journal of Financial Economics*, *3*(1-2), 167–179. https://doi.org/10.1016/0304-405X(76)90024-6

Bolhuis, M. A., & Rayner, B. (2020). *Deus ex Machina? A Framework for Macro Forecasting with Machine Learning, WP/20/45, February 2020*.

Brandt, M. W., Kavajecz, K. A., & Underwood, S. E. (2007). Price discovery in the treasury futures market. *The Journal of Futures Markets*, *27*(11), 1021–1051. https://doi.org/10.1002/fut

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32. https://doi.org/10.1201/9780429469275-8

Burda, B. U., O'Connor, E. A., Webber, E. M., Redmond, N., & Perdue, L. A. (2017). Estimating data from figures with a Web-based program: Considerations for a systematic review. *Research Synthesis Methods*, *8*(3), 258–262. https://doi.org/10.1002/jrsm.1232

Carrasco, M., & Rossi, B. (2016). In-Sample Inference and Forecasting in Misspecified Factor Models. *Journal of Business and Economic Statistics*, *34*(3), 313–338. https://doi.org/10.1080/07350015.2016.1186029

CBOE. (2020). *Vix Dataset*. https://ww2.cboe.com/products/vix-index-volatility/vix-options-and-futures/vix-index/vix-historical-data

Cesa-Bianchi, A., Pesaran, M. H., Rebucci, A., Chudik, A., Diebold, F., Elenev, V., Giannone, D., Fusari, N., Lenza, M., Noual, P., Primiceri, G., Rossi, B., Smith, R., Song, Z., Timmermann, A., & Zaffaroni, P. (2020). Uncertainty and Economic Activity: A Multicountry Perspective. *The Review of Financial Studies*, *33*(8), 3393–3445. http://www.nber.org/papers/w24325

Chauvet, M., & Potter, S. (2013). Forecasting output. *Handbook of Economic Forecasting*, *2*, 141–194. https://doi.org/10.1016/B978-0-444-53683-9.00003-7

Coulombe, P. G., Leroux, M., Stevanovic, D., & Surprenant, S. (2020). How is machine learning useful for macroeconomic forecasting? *arXiv*. http://arxiv.org/abs/2008.12477

Cremers, M., Fleckenstein, M., & Gandhi, P. (2017). Treasury Yield Implied Volatility and Real Activity. *Journal of Financial Economics (JFE), Forthcoming*. https://doi.org/10.2139/ssrn.3006473

David, A., & Veronesi, P. (2014). Investors' and Central Bank's uncertainty embedded in index options. *Review of Financial Studies*, *27*(6), 1661–1716. https://doi.org/10.1093/rfs/hhu024

Drevon, D., Fursa, S. R., & Malcolm, A. L. (2017). Intercoder Reliability and Validity of WebPlotDigitizer in Extracting Graphed Data. *Behavior Modification*, *41*(2), 323–339. https://doi.org/10.1177/0145445516673998

Ederington, L. H., & Lee, J. H. (1993). *How Markets Process Information: News Releases and Volatility* (Vol. 48, pp. 1161–1191). https://doi.org/10.1111/j.1540-6261.1993.tb04750.x

Ederington, L., & Lee, J. (1996). The Creation and Resolution of Market Uncertainty: The Impact of Information Releases on Implied Volatility. *The Journal of Financial and Quantitative Analysis*, *31*(4), 513–539. http://journals.cambridge.org/abstract%7B/_%7DS0022109000023784

Ewing, B. T., & Wang, Y. (2005). Single housing starts and macroeconomic activity: An application of generalized impulse response analysis. *Applied Economics Letters*, *12*(3), 187–190. https://doi.org/10.1080/1350485052000337806

Ferrara, L., Marsilli, C., & Ortega, J. P. (2014). Forecasting growth during the Great Recession: Is financial volatility the missing ingredient? *Economic Modelling*, *36*, 44–50. https://doi.org/10.1016/j.econmod.2013.08.042

Fornari, F., & Mele, A. (2019). Financial volatility and real economic activity. *Journal of Financial Management*, *1*(2), 155–196. https://doi.org/10.4324/9780429456572

Gilchrist, S., & Zakrajšek, E. (2012). Credit spreads and business cycle fluctuations. *American Economic Review*, *102*(4), 1692–1720. https://doi.org/10.1257/aer.102.4.1692

Gilchrist, S., & Zakrajšek, E. (2019). *Replication data for: Credit Spreads and Business Cycle Fluctuations.* American Economic Association. https://www.openicpsr.org/openicpsr/project/112536/version/V1/view

NASDAQ. (2020). *SPDR S&P 500 ETF Trust (SPY)*. https://finance.yahoo.com/quote/SPY/history/

Siliverstovs, B. (2021). *Gauging the Effect of Influential Observations on Measures of Relative Forecast Accuracy in a Post-COVID-19 Era: Application to Nowcasting Euro Area GDP Growth*.

Siliverstovs, B., & Wochner, D. (2020). State-Dependent Evaluation of Predictive Ability †. *Journal of Forecasting*. https://doi.org/10.1002/for.2715

Stock, J. H., & Watson, M. W. (2003). Forecasting Output and Inflation: The Role of Asset Prices. *Journal of Economic Literature*, *XLI*, 788–829.

Welch, I., & Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, *21*(4), 1455–1508. https://doi.org/10.1093/rfs/hhm014

Wright, J. H. (2006). The Yield Curve and Predicting Recessions. *Finance and Economics*

Discussion Series, *2006*(07), 1–19. https://doi.org/10.17016/feds.2006.07

Zhang, Y., & Trubey, P. (2019). Machine Learning and Sampling Scheme: An Empirical Study of Money Laundering Detection. *Computational Economics*, *54*(3), 1043–1063. https://doi.org/10.1007/s10614-018-9864-z

7 Appendices

7.1 Appendix A

As researchers typically do not typically post underlying data with their research, various plot digitizers have seen an exponential increase in use. Drevon et al. (2017) researched intercoder reliability, during which over 3500 data points were extracted with WebPlotDigitizer from 36 different graphs. Nevertheless, they controlled the validity of the results and concluded that there was a near perfect correlation (r=0.989 with p-value <0.01) between extracted and actual data. Nevertheless, the limitations mentioned highlight coders previous experience with plot-digitizing tools.

Furthermore, Burda et al. (2017) also highlight that systematic reviewers often tend to have data constraints which is why plot digitizers are of a great help. They estimated data using WebPlotDigitizer and conclude that the extraction done by different coders was consistent; nevertheless, in the case of continuous data (compared to event data), the distribution varied more. Whatsoever, the intreclass coefficient for both types of plots was over 95%.

We also used the WebPlotDigitzer in our research and as validity test extracted GDP from the same graph as YIV time series & plotted it with actuals - see the graph below.

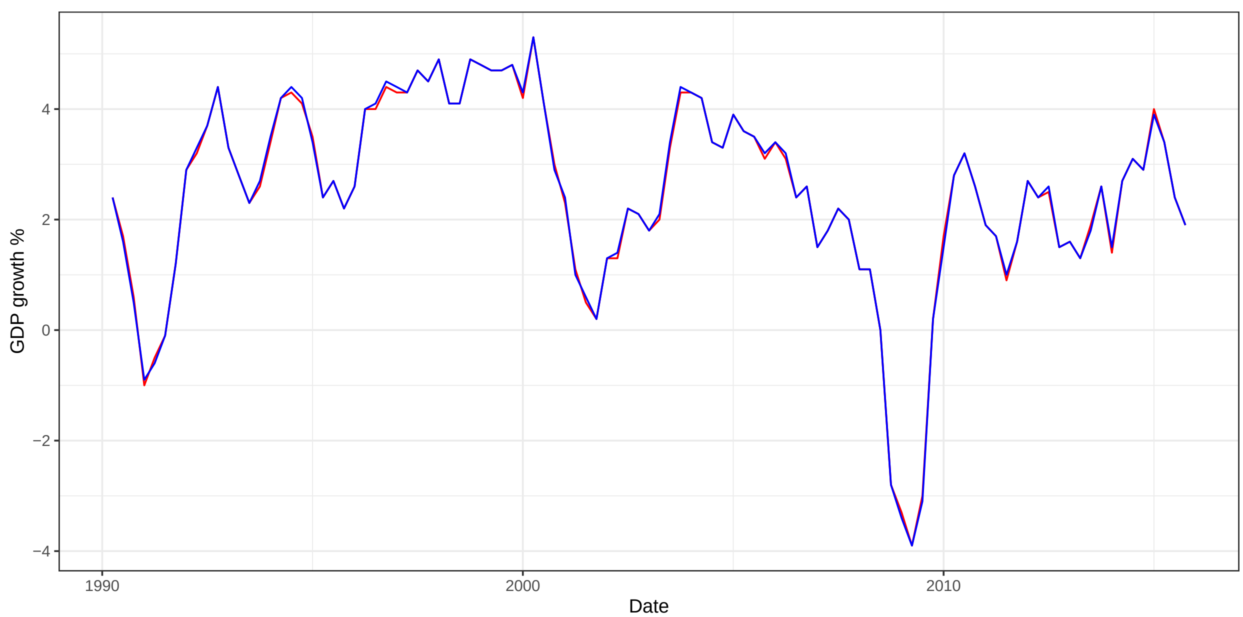


Figure 7.1: Actual vs Extracted GDP growth rate in %

7.2 Appendix B

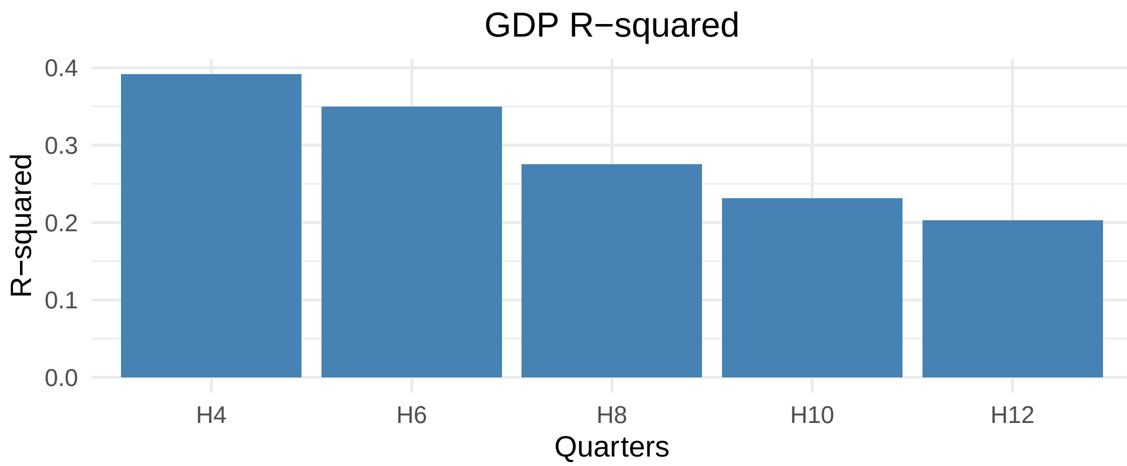


Figure 7.2: Regressions' R-squared

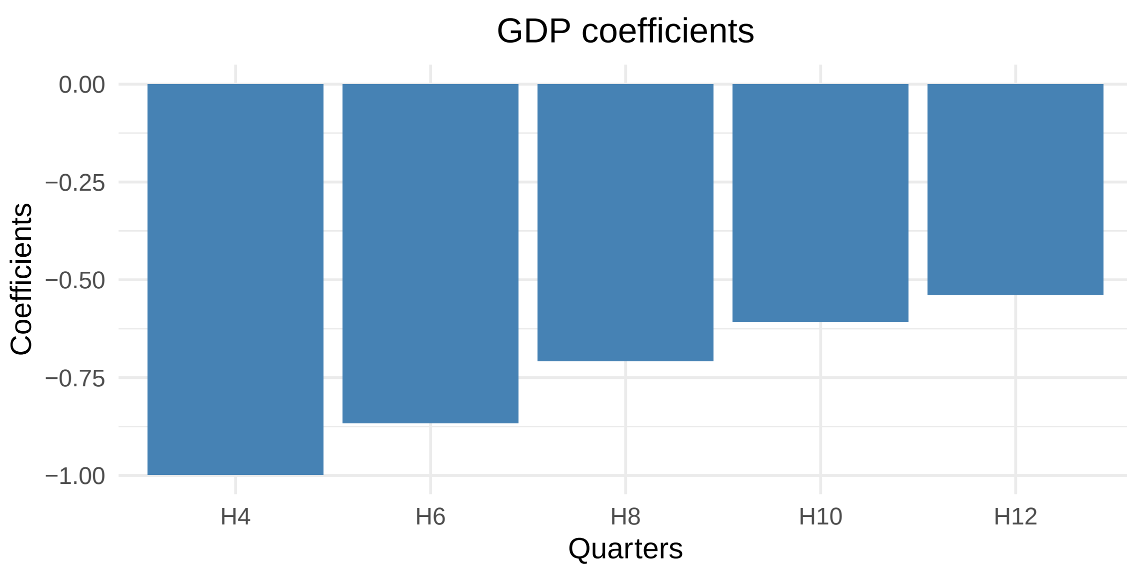


Figure 7.3: Regressions' coefficients

7.3 Appendix D

Notes: This table includes regression with controls. The equation for the regression is following:

$$\sum_{j=1}^{j=H} log(1 + GDP_{i,t+j})/H = \alpha_H + \beta_H \sigma_{IV,t}^{INT} + \varepsilon_{t+H} \tag{13}$$

Table 7.1: Regression with state-dependency

	H4	H6	H8	H10	H12
DGS1__estimate	0.31	0.51	0.87	1.23	1.52
DGS1__std.error	0.46	0.51	0.53	0.55	0.55
DGS1__p.value	0.49	0.32	0.10	0.03	0.01
TRM1012__estimate	0.42	0.57	0.82	1.12	1.34
TRM1012__std.error	0.36	0.44	0.48	0.51	0.51
TRM1012__p.value	0.25	0.20	0.09	0.03	0.01
SRT03M__estimate	0.23	0.23	0.18	0.09	-0.01
SRT03M__std.error	0.21	0.28	0.29	0.25	0.23
SRT03M__p.value	0.28	0.42	0.53	0.74	0.95
baa__aaa__estimate	-0.44	-0.16	-0.09	-0.19	-0.28
baa__aaa__std.error	0.24	0.29	0.30	0.30	0.29
baa__aaa__p.value	0.07	0.60	0.78	0.52	0.34
VIX__estimate	-0.12	-0.22	-0.19	-0.12	-0.10
VIX__std.error	0.21	0.25	0.24	0.22	0.21
VIX__p.value	0.57	0.39	0.43	0.61	0.63
housng__estimate	0.08	0.18	0.33	0.38	0.29
housng__std.error	0.15	0.17	0.20	0.20	0.16
housng__p.value	0.59	0.28	0.11	0.07	0.08
gz__spr__estimate	-0.53	-0.80	-0.76	-0.50	-0.28
gz__spr__std.error	0.33	0.33	0.39	0.46	0.48
gz__spr__p.value	0.12	0.02	0.06	0.29	0.56
spy__logreturn__estimate	0.02	0.01	0.00	0.00	0.00
spy__logreturn__std.error	0.02	0.02	0.02	0.02	0.02
spy__logreturn__p.value	0.20	0.66	0.96	0.96	0.90
r.squared	0.75	0.64	0.58	0.57	0.60
adj.r.squared	0.72	0.59	0.52	0.52	0.54

Note:

*** - p<0.01, ** - p<0.05, * - p<0.1. Reported standard error is adjusted for heteroskedasticity

7.4 Appendix E

Notes: YIV, dummy and controls as independent variables. The equation for the regression is following:

$$\sum_{j=1}^{j=H} log(1 + GDP_{i,t+j})/H = \alpha_H + \beta_H \sigma_{IV,t}^{INT} + \varepsilon_{t+H} \tag{14}$$

Table 7.2: Regression with state-dependency

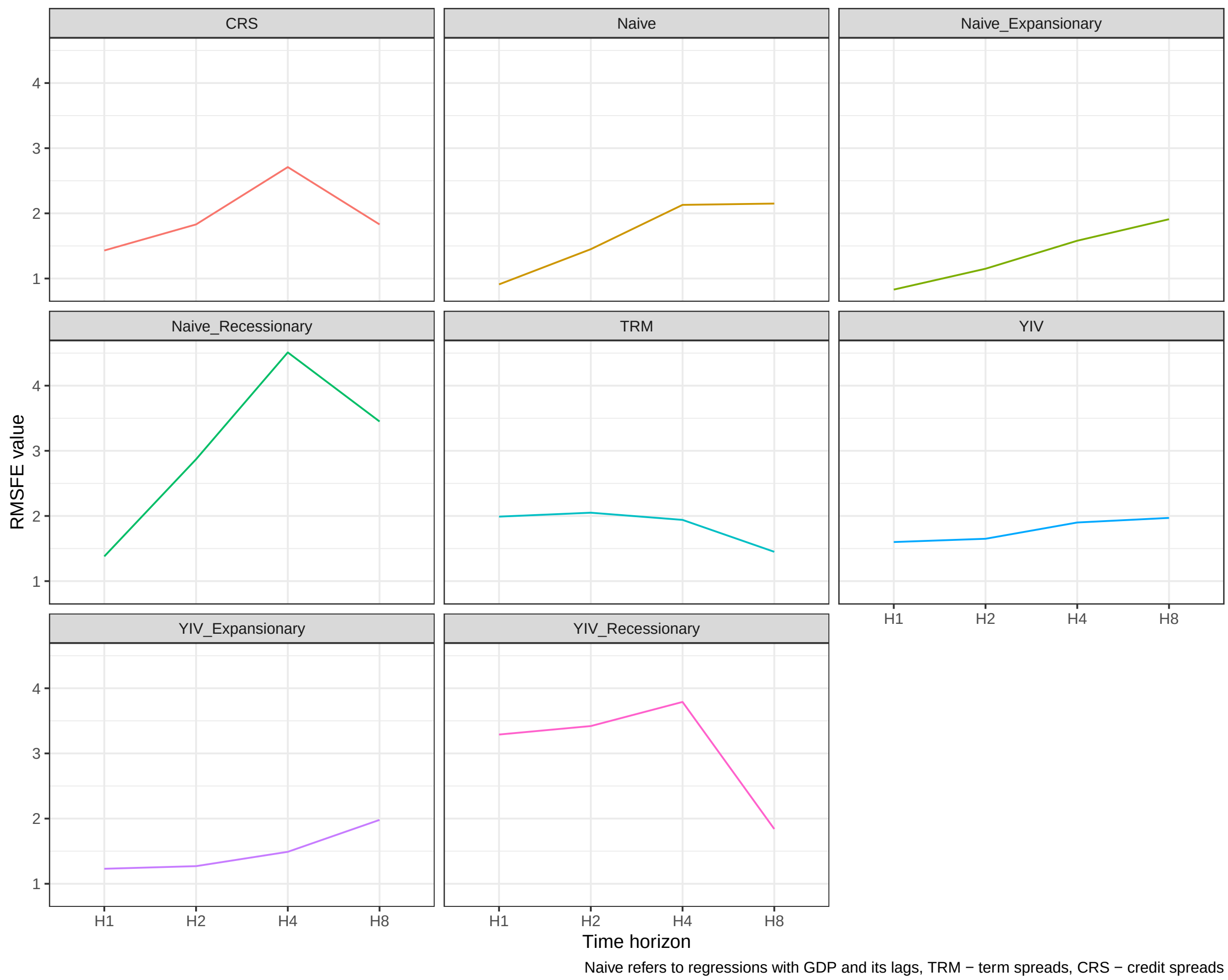
	H4	H6	H8	H10	H12
YIV__estimate	-0.59	-0.68	-0.57	-0.47	-0.33
YIV__std.error	0.14	0.19	0.18	0.15	0.15
YIV__p.value	0.00	0.00	0.00	0.00	0.03
dum__estimate	-1.11	-0.99	-0.88	-0.63	-0.63
dum__std.error	0.51	0.68	0.66	0.56	0.50
dum__p.value	0.04	0.15	0.18	0.27	0.21
DGS1__estimate	0.43	0.63	0.97	1.30	1.58
DGS1__std.error	0.39	0.46	0.49	0.53	0.53
DGS1__p.value	0.27	0.18	0.05	0.02	0.00
TRM1012__estimate	0.51	0.67	0.91	1.19	1.38
TRM1012__std.error	0.32	0.42	0.48	0.52	0.50
TRM1012__p.value	0.12	0.12	0.06	0.03	0.01
SRT03M__estimate	0.02	0.03	0.01	-0.04	-0.13
SRT03M__std.error	0.20	0.28	0.31	0.28	0.25
SRT03M__p.value	0.92	0.91	0.99	0.88	0.59
baa__aaa__estimate	-0.03	0.29	0.29	0.11	-0.05
baa__aaa__std.error	0.22	0.30	0.32	0.32	0.31
baa__aaa__p.value	0.88	0.33	0.37	0.73	0.87
VIX__estimate	0.09	0.00	0.00	0.03	0.02
VIX__std.error	0.17	0.18	0.18	0.18	0.18
VIX__p.value	0.60	1.00	0.99	0.88	0.93
housng__estimate	0.18	0.28	0.41	0.44	0.34
housng__std.error	0.15	0.19	0.24	0.23	0.19
housng__p.value	0.24	0.14	0.09	0.07	0.08
gz__spr__estimate	-0.32	-0.58	-0.57	-0.35	-0.16
gz__spr__std.error	0.35	0.36	0.40	0.47	0.49
gz__spr__p.value	0.38	0.11	0.16	0.46	0.74
spy__logreturn__estimate	0.03	0.01	0.00	0.00	0.00
spy__logreturn__std.error	0.01	0.02	0.02	0.02	0.02
spy__logreturn__p.value	0.06	0.40	0.79	0.80	0.82

r.squared	0.84	0.75	0.67	0.64	0.64
adj.r.squared	0.81	0.71	0.61	0.58	0.58

Note:
*** - p<0.01, ** - p<0.05, * - p<0.1. Reported standard error is adjusted for heteroskedasticity

7.5 Appendix F

	H1	H2	H4	H8
YIV	1.60	1.65	1.90	1.97
YIV-Recess.	3.29	3.42	3.79	1.84
YIV-Expans.	1.23	1.27	1.49	1.98
Naive	0.91	1.45	2.13	2.15
Naive-Recess.	1.38	2.87	4.51	3.45
Naive-Expans.	0.83	1.15	1.58	1.91
TRM	1.99	2.05	1.94	1.45
CRS	1.43	1.83	2.71	1.83



7.6 Appendix G

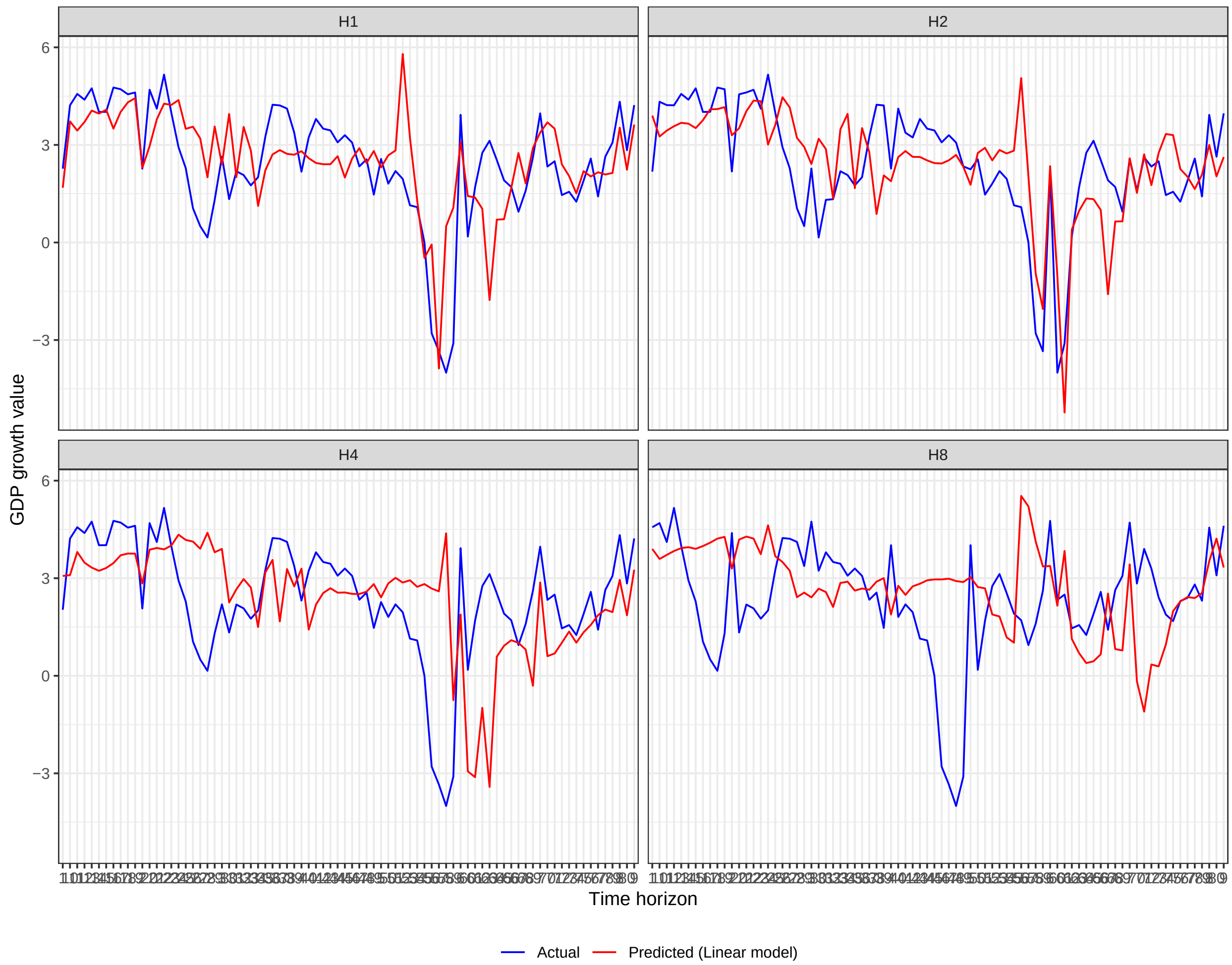


Figure 7.4: Predicted vs actual results (Linear model)

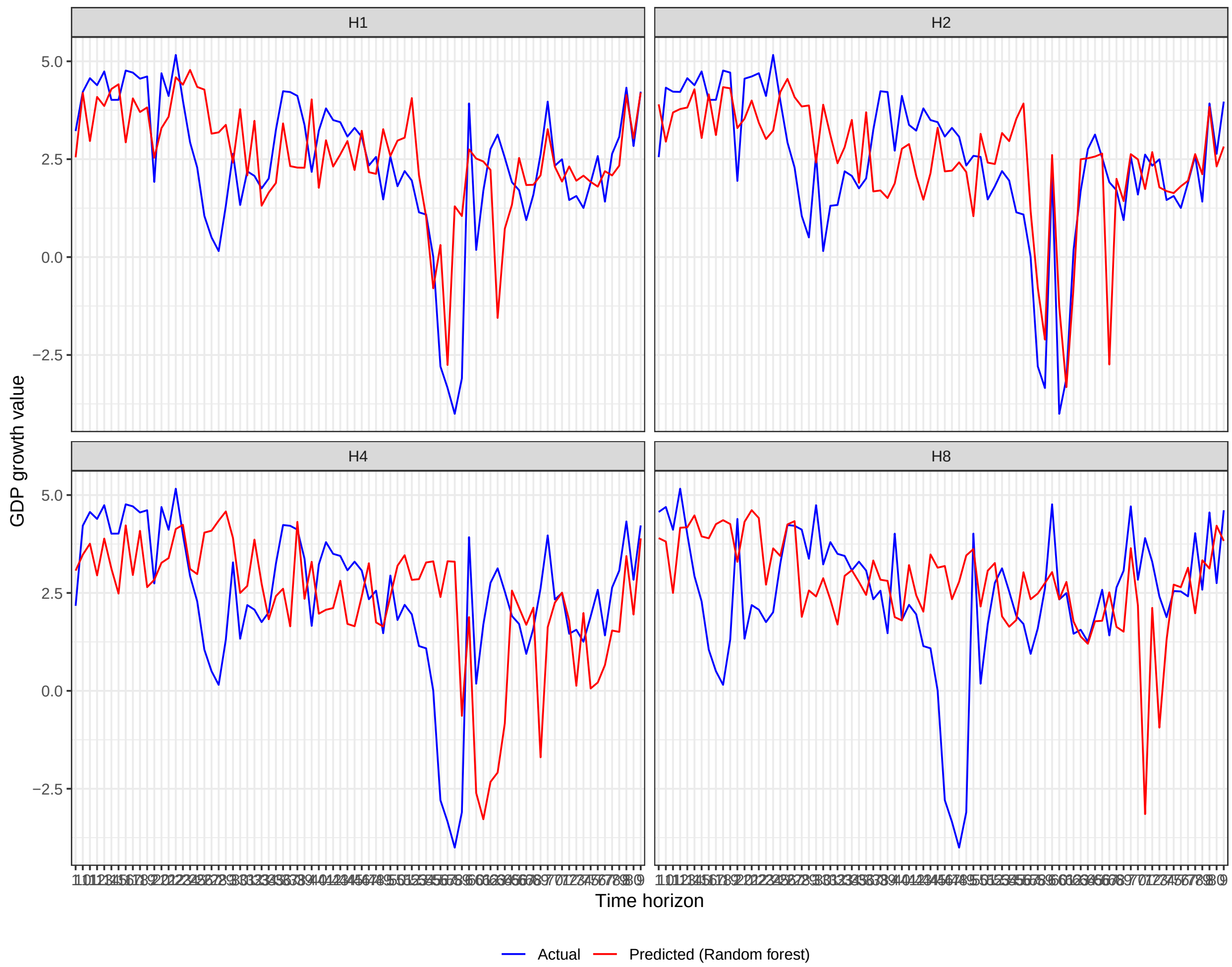


Figure 7.5: Predicted vs actual results (Random forest)

7.7 Appendix H[3]

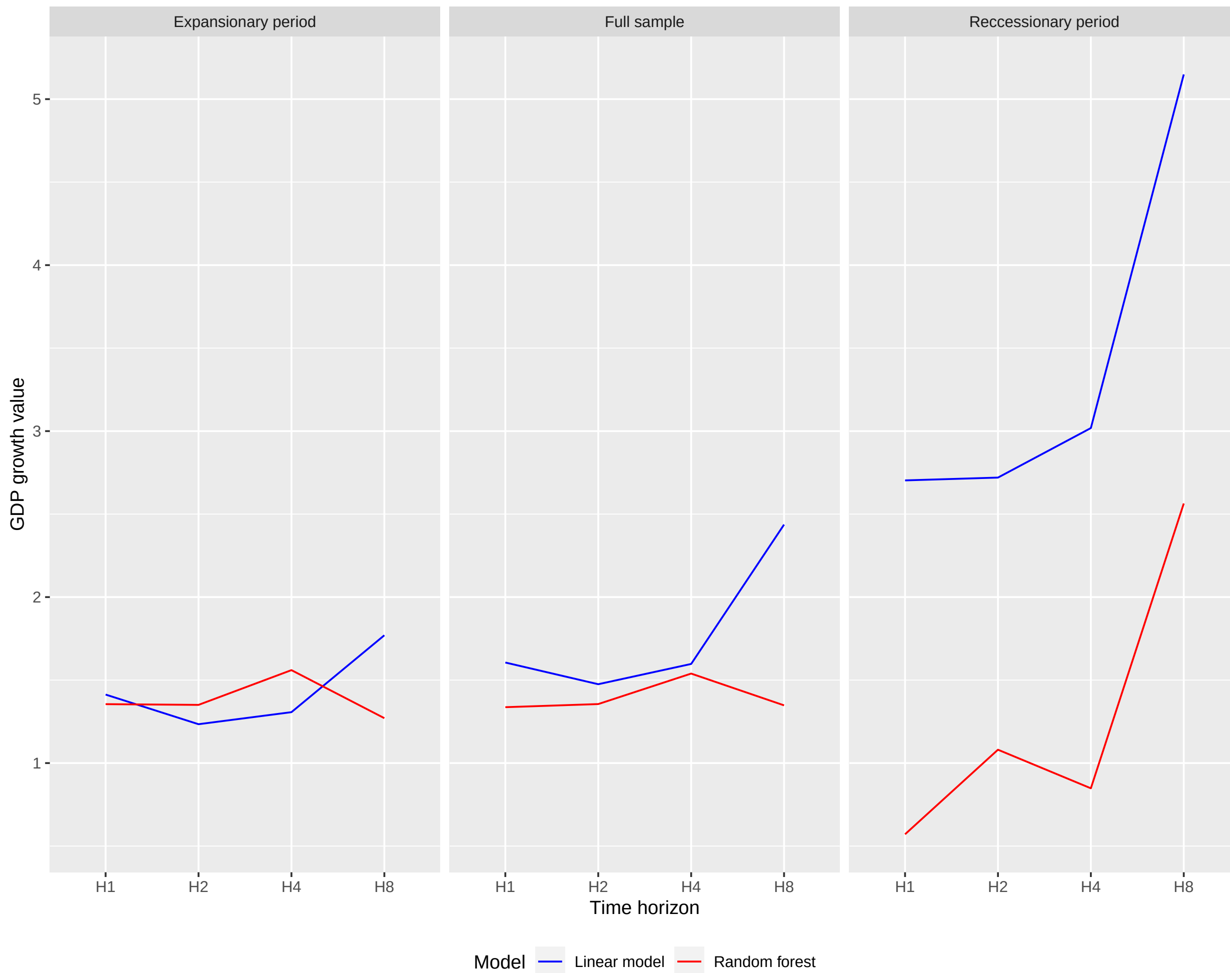


Figure 7.6: RMSFE-s of linear model and random forest

[3]Full R-Code with data is available upon request in github repository: https://github.com/karelrappo/thesis2020

7.8 Appendix H

Variable	Description
YIV	5 - year Treasury Implied Volatility
GDP	Year-on-year change in real gross domestic product
spy_logreturn	logarithmic yearly returns of SPY index
VIX	Returns of VIX index
DBAA	BAA corporate bond yields
AAA	AAA corporate bond yields
baa_aaa	Yield spread between BAA and AAA yields
gz_spr	Credit risk spread index
housng	New housing market starts
SRT03M	Changes in 3 month treasury yield
TRM1003	TRM1003 - 10 year and 3 month treasury yield spread
TRM1006	TRM1006 - 10 year and 6 month treasury yield spread
TRM1012	TRM1012 - 10 year and 1 year treasury yield spread
TRM0503	TRM0503 - 5 year and 3 month treasury yield spread
TRM0506	TRM0506 - 5 year and 6 month treasury yield spread

7.9 Appendix I

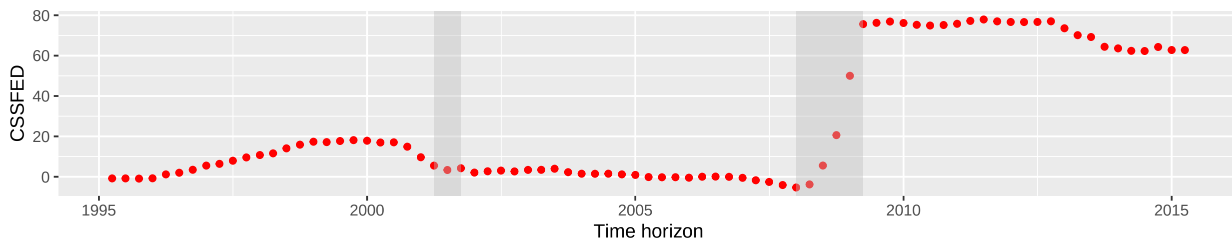


Figure 7.7: Cumulative sum of squared forecast error differential (Linear model vs benchmark historic mean model

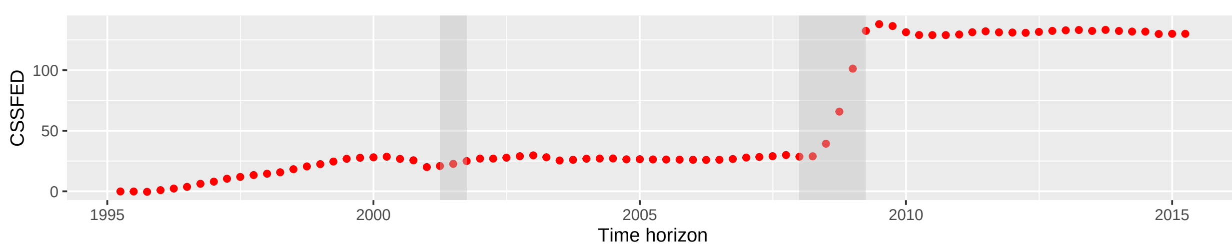


Figure 7.8: Cumulative sum of squared forecast error differential (Random forest model vs benchmark historic mean model

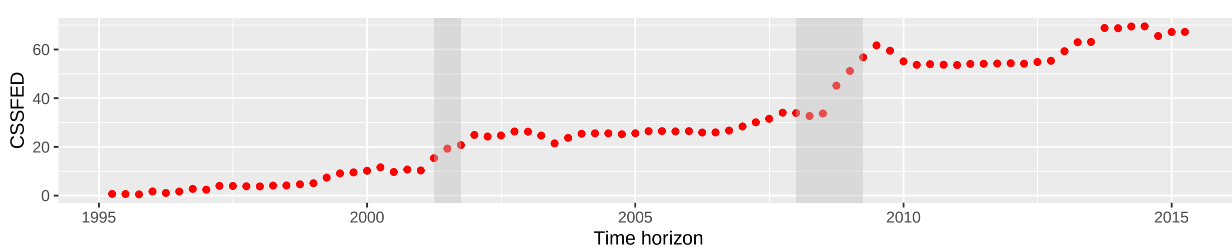


Figure 7.9: Cumulative sum of squared forecast error differential (Random forest model vs benchmark linear model

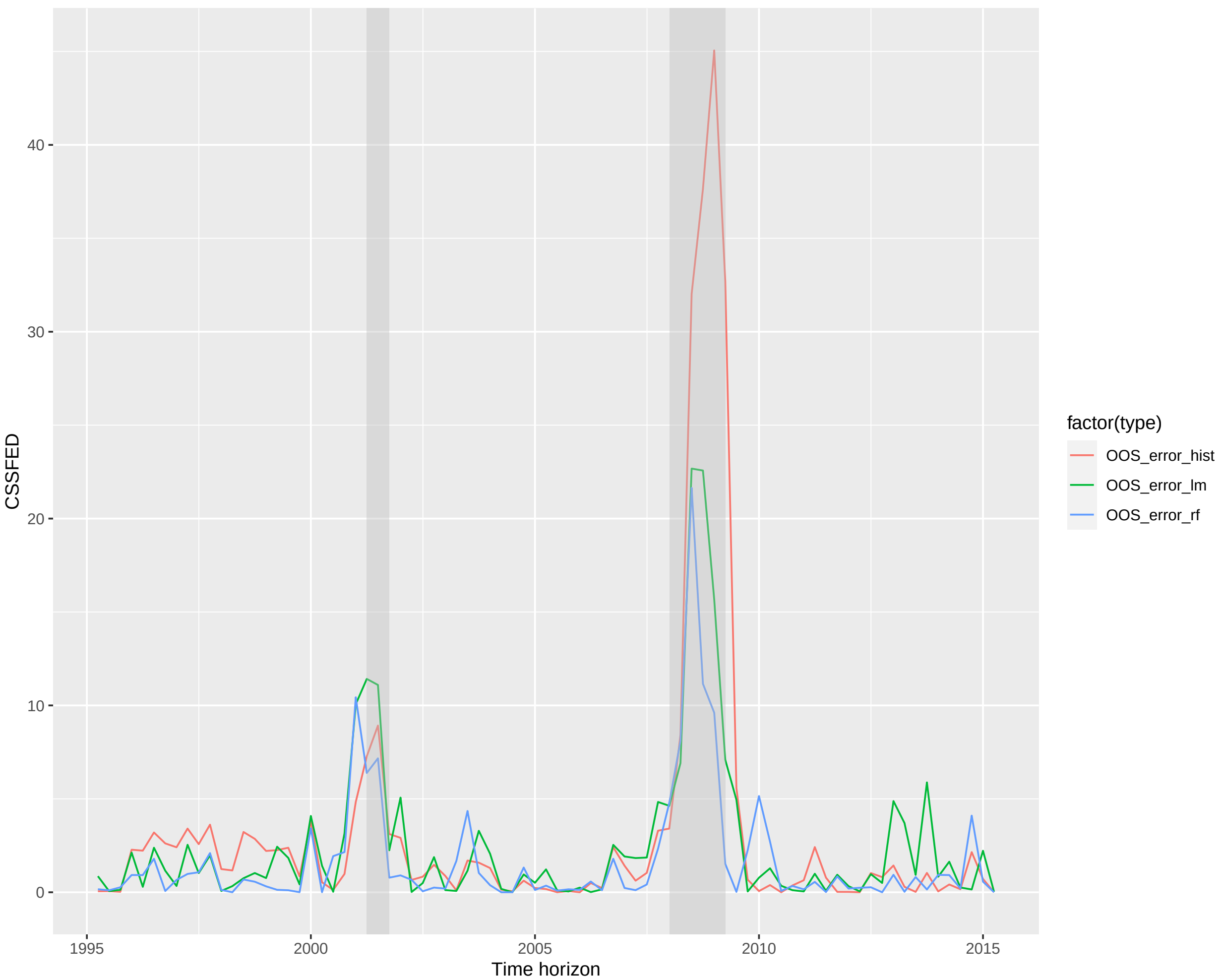


Figure 7.10: Squared errors