SSE RIGA

State-dependent forecasting of macroeconomic real activity

Bachelor Thesis

Authors:

Karel Räppo

Henry Enno Turu

Supervisor:

Boriss Siliverstovs

Riga

2021

COPYRIGHT DECLARATION AND LICENCE

Name(s) of the author(s) in full: Karel Räppo, Henry Enno Turu

Title of the Thesis: **….**

I (We) hereby certify that the above-named thesis is entirely the work of the persons named below, and that all materials, sources and data used in the thesis have been duly referenced. This thesis – in its entirety or in any part thereof – has never been submitted to any other degree commission or published.

In accordance with Section 1 of the Copyright Law of Latvia, the person(s) named below is (are) the author(s) of this thesis.

Pursuant to Article 40 of the Copyright Law the author(s) hereby agree(s) and give an explicit licence to SSE Riga to deposit one digital copy of this thesis in the digital catalogue and data base at SSE Riga Library for an unlimited time and without royalty. The licence permits SSE Riga to grant access to the duly deposited thesis to all users of the catalogue and data base without royalty and limitations to downloading, copying and printing of the digital thesis in whole or in part provided I am (we are) indicated as the author(s) of the thesis according to Clause 4 Section 1 Article 14 of Copyright Law. I (We) assert my (our) right to be identified as the author(s) of this thesis whenever it is reproduced in full or in part.

Signed

___________________ ___________________

/Karel Räppo/ /Henry Enno Turu/

Date

07.04.2021

Contents

Abstract **5**

1 Introduction **6**

2 Review of literature **9**

2.1 Historically recognized predictors of GDP 9

2.2 Forecasting asymmetries 12

2.3 Advanced methods for macroeconomic forecasting 12

3 Methodology **15**

3.1 Linear regressions 15

3.2 Out-of-sample validation 16

3.3 Out of sample validation for different economic cycles 17

3.3.1 Cumulated sum of squared error differential 18

3.4 Machine learning methods 19

3.4.1 Decision Trees 20

3.4.2 Random Forest 21

4 Data & descriptive statistics **23**

5 Results **26**

5.1 In-sample regressions 27

5.2 Out of sample forecasting 29

5.2.1 Full sample 29

5.2.2 Sub-sample OOS forecasting 30

5.3 Out-of-sample forecasting with Random Forest 30

6 Discussion **33**

6.1 Discussion of results 33

6.2 Limitations & further research 35

7 Conclusion **36**

8 References **37**

9 Appendices **40**

9.1 Appendix A 40
9.2 Appendix B 41
9.3 Appendix D 42
9.4 Appendix E 43
9.5 Appendix F 44
9.6 Appendix G 46
9.7 Appendix H 47
9.8 Appendix I 48
9.9 Appendix J 49

Abstract

Despite being a widely researched topic, forecasting macroeconomic real activity especially using financial markets is still a focus of many researchers. Historically, many researchers have set out to discover new and novel variables, however, there is no strong agreement on which variables could predict the macroeconomic real activity consistently. In this research, we examine a variable called YIV (treasury implied volatility) and its predictive ability on different measures of GDP growth. Additionally, we test whether the model's robustness holds also during different subsample periods (recessionary and expansionary periods) as there is significant evidence from the latest literature that performance asymmetries exist. Lastly, we set out to discover whether the model's performance could be improved by using a machine learning (ML) based method that accounts for the non-linearities. Our results indicate that indeed YIV is a significant predictor of macroeconomic real activity, however, it is important to account for a model's performance asymmetries as few points around the recessionary periods have a substantial contribution to the model's accuracy. Furthermore, we find that through using an ML-based method the model's accuracy can be improved significantly thanks to the method's ability to account for nonlinearities.

Keywords: Predicting real economy, treasury implied volatility, performance asymmetries, ML-based methods.

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 James H. 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" (James H. Stock & Watson, 2003, p. 822). Furthermore, there are even less variables 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. (2021), 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.

As our paper builds on the initial discoveries of Cremers et al. (2021), the first part of our paper focuses on replication their results to test the validity - i.e whether real GDP can be predicted using YIV.

In the second part of the research we advanced their paper from two perspectives.

Firstly, to account for the state-dependency of forecast models as shown to be relevant by Chauvet & Potter (2013) and Siliverstovs & Wochner (2020), we test whether business cycles affect the size of YIV's impact on future GDP growth. In addition, we investigate whether out-of-sample forecast performance is dependent on the business cycle - i.e calculating RMSFE including only recessionary periods or expansionary periods (classification of the US business cycles obtained 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 (Coulombe, 2020). Thus, the RF model is expected to significantly decrease the root mean square forecasting errors (RMSFE). Nevertheless, it's output is difficult to interpret 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 USA's macroeconomic activity?**
2. **If so, is this conclusion robust even after taking into account different business cycles and control variables.**
3. **Do Machine Learning methods such as Random Forest improve the forecast accuracy?**

The research paper is structured as follows. First, we offer context to our work by describing various research already done in the same field. As the next step, we describe the data and the acquisition method used for obtaining it. Subsequently, we offer an in-depth overview of the methods used, which is followed by description of results. In the discussion part we elaborate how our findings fit in with the existing research. Lastly, we list any possible limitations of the paper and suggest ideas for further research.

2 Review of literature

Before digging deeper into the predictive ability of treasury yield implied volatility it is essential to understand what are the other variables that can explain the connection between financial markets and the real economy.

2.1 Historically recognized predictors of GDP

One of the most recognized predictors of recessions and GDP growth are term spreads - moreover, term spreads have been included as one of the most important variables in the business cycle indicator index by James H. Stock & Watson (1989). Ang et al. (2006) use yield curve to obtain the best maturity short rate for forecasting GDP - furthermore, it is noted that the best form of slope is the one constructed with maximum maturity difference. Using the aforementioned process, they conclude that contrarily to the existing research, short rate is a better predictor compared to any of the term spreads. Lastly, Gilchrist & Zakrajšek (2012) construct a corporate bond credit spread index which not only is a significant predictor of macroeconomic activity for different variables & time-horizons but also its predictive ability outperforms commonly used BAA-AAA corporate bond spread.

One rather different variable that can be potentially used for macroeconomic forecasting is housing starts (amount of residential property construction projects started) as it has been found to be positively correlated with economic cycles (Ewing & Wang, 2005).

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. 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. Regardless of not directly using VIX as a measure for volatility, the conclusion should hold for VIX as well - the comovement of VIX and their constructed realized volatility measure is very similar, having correlation of over 90%.

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 shock to stock and bonds ATM IV is followed by a decline in future real rates. Thus, there exists a positive relationship between IV shock and possibility of recession.

Building upon the aforementioned various research, Cremers et al. (2021) 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 L. H. 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 afpremertioned control variables such as term spreads, credit spreads, stock market volatility, housing starts, etc. are included in the model.

Even though the authors possess the data for different maturities (1,5,10,30 years), 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.

2.2 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 accounting for state-dependency, 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.

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.

2.3 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. The latter is also emphasized by Chauvet & Potter (2013), who summarize that the biggest errors of linear models happen during the recessionary periods as in the beggining or the end of the period linear relationships break. Hence, non-linear models compared to linear models can improve the forecasting performance and output additional information.

Also, 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. In addition, Siliverstovs (2021) adds to the research by analyzing the impact of influential observations on relative forecast accuracy during Covid-19 crisis. By employing a newly proposed metric (recursive relative mean squared forecast error) together with the cumulative sum of squared forecast error difference (CSSFED) presented in Welch & Goyal (2008), he concludes that there exist significant differences in relative forecasting error depending on the business cycle.[1]. 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

[1] Note:a more detailed description can be found in the section "Methodology."

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 Cremers et al. (2021) while improving the methodology in 2 ways:

1. Checking whether YIV helps to predict GDP growth controlling for 3 different definitions
2. Testing YIV's predictive ability depending on the business cycle
3. 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 proposed by Cremers et al. 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 mainly thanks to its ability to take into account non-linearity.

3 Methodology

3.1 Linear regressions

As already mentioned before, the first part of our methodology consists of replicating the methodology conducted by Cremers et al. (2021). 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, $\mathrm{GDP_{i,t+j}}$ 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 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.

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

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). This is done by constructing a predictive model with a 5-year rolling estimation window. This means that we use the previous 20 quarters to predict the next H quarters (note: here H denotes average quarterly year-on-year growth rates H quarters ahead as in the original paper). The reason behind opting for rolling estimation window is that in this way the

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

predicted datapoint is continuously 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 which together with the actual values can be used to 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. (2021). 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.

For this part of the research we shift our dependent variable from being average quarterly year-on-year growth rates (denoted as H1, H2, and etc. as in the original paper) to quarterly growth rates of GDP h-quarters ahead (denoted in our paper as F1, F2, and etc.). The reason for that lies in the underlying formulas of these dependent variables. In appendix 9.9 we have dissected three formulas for the dependent variables used in the academia. The first is average quarterly year-on-year growth rates (as is used by Cremers et al. (2021)), the second is h-step ahead average quarterly growth rate and the last one is h-step ahead quarterly growth rate. We have taken H4 as a comparison point for these three variables. As it can be see from the average quarterly year-on-year growth rates 4 quarters ahead, the variable relies more heavily on the point near the current time period (e.g. more weight on t, t+1, and t+2 growth rates while relying less on t+3 and t+4). The next variable i.e. the h-step ahead average quarterly growth rate weighs the next four quarters' growth rates equally. However, the pitfall is that the actual and predicted values are smoothed significantly due to the averaging so as the period increases, the squared errors become smaller. Thus, the model using this growth rate variable would produce smaller RMSFEs in the future periods as the extremes are averaged out. The last variable, h-step ahead quarterly growth rate, is calculated only based on the GDP growth rate 4 quarters ahead.

Based on the aforementioned, we decided to use the h-step ahead quarterly growth rate for RMSFE predictions as it focuses strictly on the forecasted quarter's growth rate and eliminates the problem with unequal weighting and unproportionate RMSFE in the further forecasting horizons.

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 relative root mean square forecasting error (rRMSFE). 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, there might be a reason to doubt the conclusion drawn from such few observations. To tackle this, we use the cumulative sum of squared forecast error difference (CSSFED) 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 feature even more useful to our research - 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 CSSFED is applied in order to test how the few observations of recessions impact the whole forecasting error. As mentioned by Siliverstovs (2021), 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 CSSFED can be calculated according to the formula below:

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

where $e_{benchmark,t}$ refers to forecast errors of benchmark model, and $e_{advanced,t}$ 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 Random Forest method.

In our research, we focus specifically on a 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 (Y. 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.

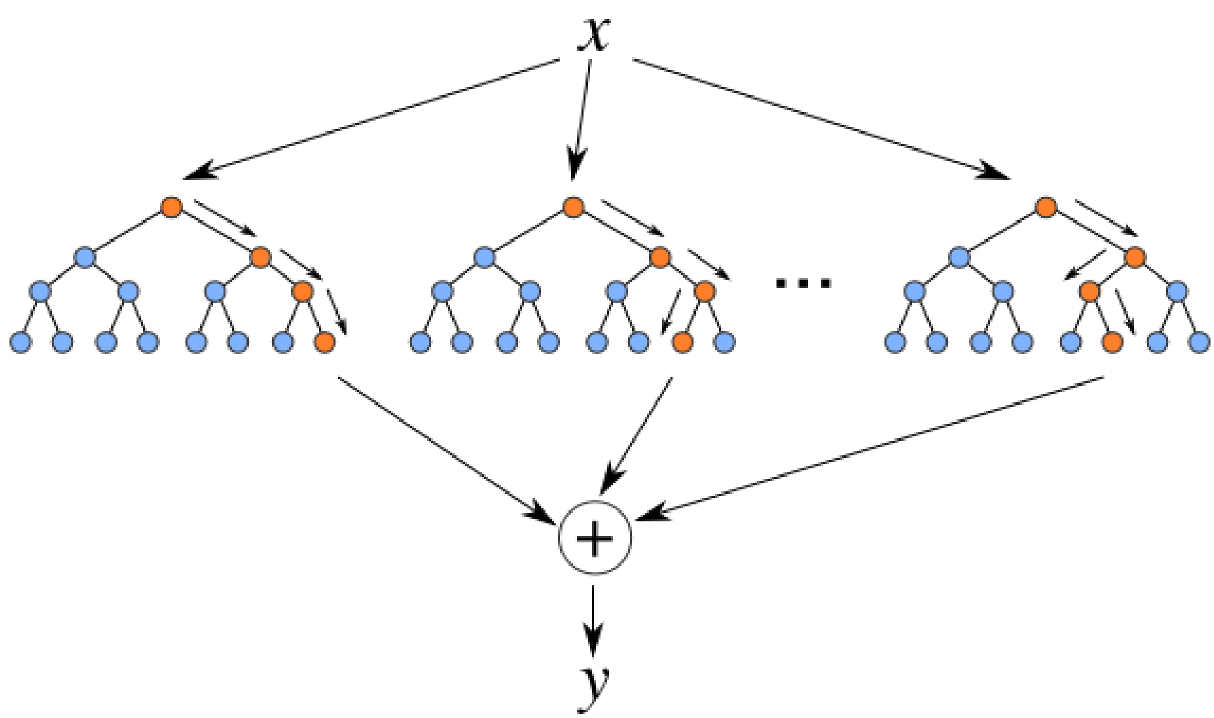


Figure 3.1: Random forest example

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 (Y. 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 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 observations).
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 data (used for calculating respective dependent variables for which the equations can be found in the appendix 9.9
- Control Variables:
  - US treasury interest rates to construct term spreads
  - credit spreads
  - credit spread index
  - stock returns (SPY)
  - stock market implied volatility (VIX)
  - 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 9.1).

The data regarding GDP 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 included corporate bond credit spread index, which was obtained from the official replication dataset from OPENCSR (Gilchrist & Zakrajšek, 2019). As per authors, the dataset is constructed "using the prices of corporate bonds trading in the secondary market" (Gilchrist & Zakrajšek, 2012, p. 1693). The daily data for stock market implied volatility (VIX) is taken from CBOE, which is later aggregated into quarterly data (CBOE, 2020). Lastly, 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.

For the out-of-sample tests, we exclude SPY & GZ_SPR from the regressions as their dataset is limited (NAs present). Due to the random selection in RF NA's among predictor variables are not allowed; hence, if we want to have comparable results with linear model exclusion is required.

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
housng	3.18	51.49	-151.80	-16.80	14.10	36.10	117.70
DGS3MO	2.95	2.32	0.01	0.16	3.14	5.11	8.01
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	-0.08	0.42	-1.39	-0.16	-0.01	0.08	0.83

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

5 Results

For preliminary analysis, we plot quarterly YIV with quarterly GDP growth rates 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 while a big decline in GDP growth happens. This confirms our base for analysis as suggested by the literature review that indeed the treasury could have the predictive ability over GDP growth rate.

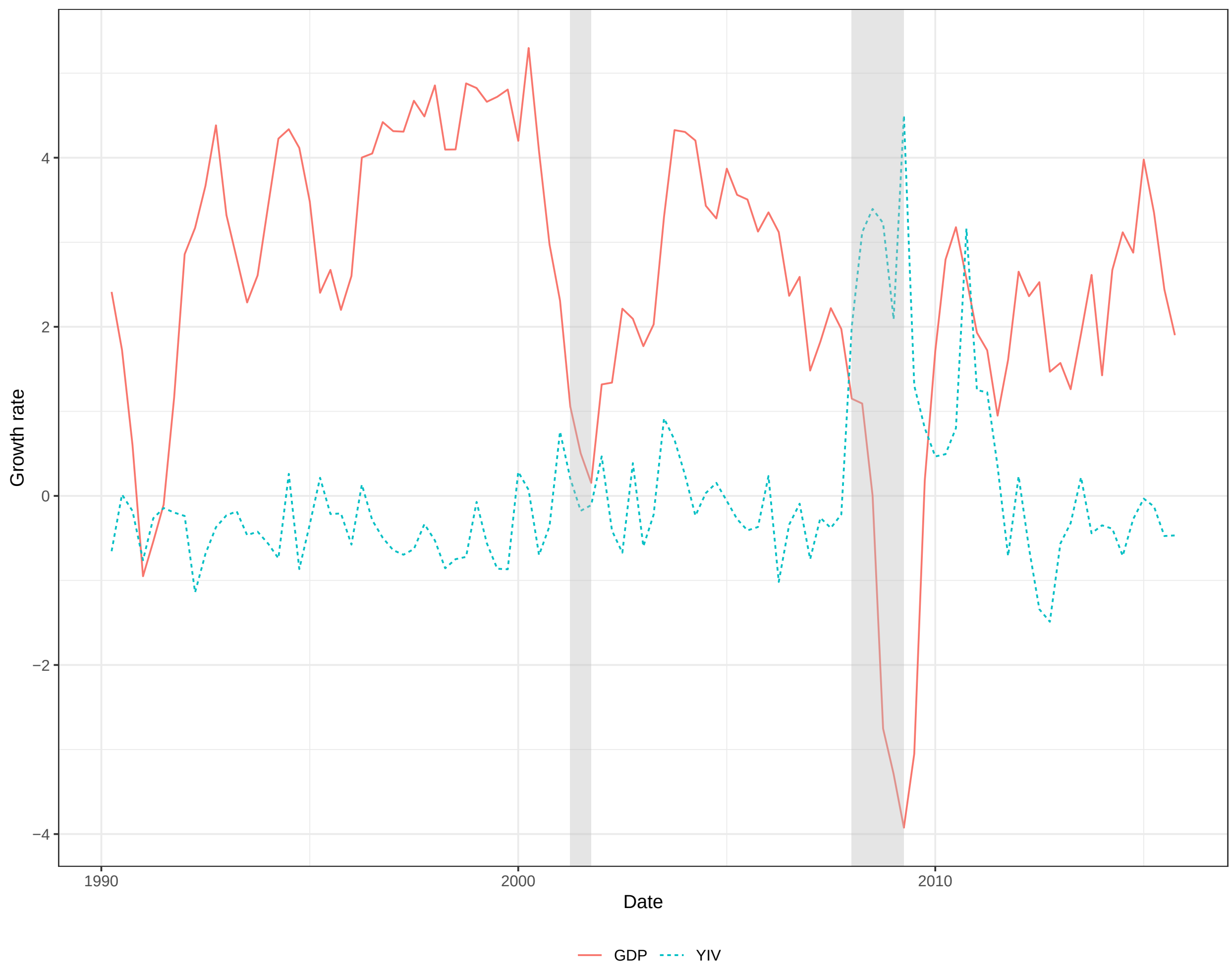


Figure 5.1: GDP Growth(%) vs 5-year Treasury Implied Volatility

5.1 In-sample regressions

We start our analysis by replicating the regressions in the paper by Cremers et al. (2021) starting with regressing YIV to GDP growth throughout different rolling periods - i.e. from H=1 to H=8 quarterly GDP growth rolling averages (the regression formula and summary can be seen in 5.1). For interpreting the regression results, the coefficient is multiplied by the standard deviation and the result obtained is the impact on yearly GDP growth. Hence, as an example of H=1, the results implicate that 1 standard deviation increase in YIV is associated with a -1.03/4 * 1.31% = 0.34% decrease in GDP growth within the next quarter. For predicting 8 quarters ahead (2 years), 1 standard deviation increase in YIV is associated with a -0.6 * 2 * 1.31% = 1.57% decrease in GDP growth within the next 8 quarters. 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% of confidence level. The model's R-squared varies within the range of 36% for predicting 2 quarters ahead down to 20% for predicting 2 years ahead. Lastly, 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

	H1	H2	H4	H8
YIV_estimate	-1.03 ***	-1.03 ***	-0.93 ***	-0.6 ***
YIV_std.error	0.27	0.25	0.23	0.17
r.squared	0.35	0.36	0.34	0.2
adj.r.squared	0.34	0.36	0.33	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. As it can be seen from the table 5.2), the dummy's coefficient is negative and significant at a 1% confidence level throughout all predicted time spans - nevertheless, the interaction term between YIV and dummy variable is insignificant throughout all forecast horizons. Hence, this means that the recessionary period only influences the intercept, not the slope of the variable. However, adding the dummy improved R-squared significantly - 54% in predicting ahead GDP growth 4 quarters' rolling averages when compared to 34% of respective YIV-only model. The latter indicates that there is a structural break in data during recessions and that using the full model to predict during recessionary periods could result in worse prediction accuracy.

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

Table 5.2: Regression with state-dependency

	H1	H2	H4	H8
YIV_estimate	-0.41	-0.42 **	-0.36 *	-0.29
YIV_std.error	0.25	0.15	0.2	0.22
dum_estimate	-2.77 ***	-2.85 ***	-2.39 ***	-1.2 ***
dum_std.error	0.4	0.31	0.36	0.26
YIV:dum_estimate	-0.25	-0.22	-0.27	-0.18
YIV:dum_std.error	0.28	0.19	0.21	0.26
r.squared	0.56	0.6	0.54	0.28
adj.r.squared	0.55	0.59	0.53	0.25

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

Finally, we combined the previous variables with all of the available control variables in Cremers et al. (2021) into one regression equation (Table 9.1). After including all the relevant control variables, the YIV loses some of its significance and remains significant at 5% in H4 and at 10% in H8. In this model, the regression results implicate that 1 standard deviation increase in YIV is associated with a -0.35 * 1.31% = -0.46 % decrease in GDP growth within the next 4 quarters. Additionally, when predicting 8 quarters ahead a 1 standard deviation increase in YIV is associated with a -0.34* 1.31% * 2 = 0.89% decrease in the next 8 quarters' average annualized growth rate.

5.2 Out of sample forecasting

5.2.1 Full sample

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. It is important to note that from here on out we use h-step forward-looking quarterly growth rates as dependent variable (denoted as F1, F2, etc.) instead of the year-on-year averaged growth rates (denoted H1, H2, etc.). For a more detailed explanation on changing the dependent variable please refer to the methodology section. Thus, for the out-of-sample regressions, we use 5-year rolling windows to predict 1-8 quarters ahead. Looking at the full model's predicted values and actual observations (see appendix 9.5), it can be seen that the predicted values tend to differ more as the forecast horizon increases. While forecasting 1 period ahead, the predicted graph is fairly similar to the actual observations, for F4, two big forecast errors can be noticed.

This is consistent with the RMSFEs calculated (see appendix 9.4). It is that the RMSFE's tend to increase with the forecast horizon, except for F2. In other words, the interpretation of it was that the model's accuracy in predicting GDP growth got worse in predicting further time periods. Comparing the full-model with only YIV or only individual control variable models, it can be seen that the full model yields the highest forecast errors as in addition to the other control variables, the full model includes the following independent variables: dummy variable, housing market, VIX, 1-month treasury yield and changes in short rate.

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 from the same figure (9.4), during the recessionary period the RMSFE is significantly higher when compared to the full-sample RMSFE - furthermore, this holds for all the different models.

Table 5.3: Relative RMSFEs

	H1	H2	H4	H8
rRMSE_recess	1.73	1.52	2.07	1.97
rRMSE_expans	0.87	0.91	0.75	0.78

Figure 5.3 describes subsample relative RMSFE within different forecasting periods. 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.

5.3 Out-of-sample forecasting with Random Forest

Lastly, we construct a random forest model to see whether the forecasting accuracy can be improved by introducing non-linearity. As it can be seen from the figure (5.2) the overall performance of the random forest model is better when compared to OLS across all of the prediction periods as indicated by the upward sloping graph (meaning that OLS squared errors are larger than the respective squared errors of the random forest model). Furthermore, the same conclusion is derived when looking at the 9.6) where it can be seen that the RF model significantly outperforms during full sample, and also separately in both subset periods. Additionally, these results are consistent with (9.5) and (9.7). The latter can be used to visually validate the difference between accuracy of random forest and linear model as during H4 and H8 forecasts as linear model's errors spike especially high after the 2008-2009 Financial crisis which is not the case for the random forest model. Also, this can be noted that on the CSSFED graph below since the bigger jumps tend to happen around recessionary periods.

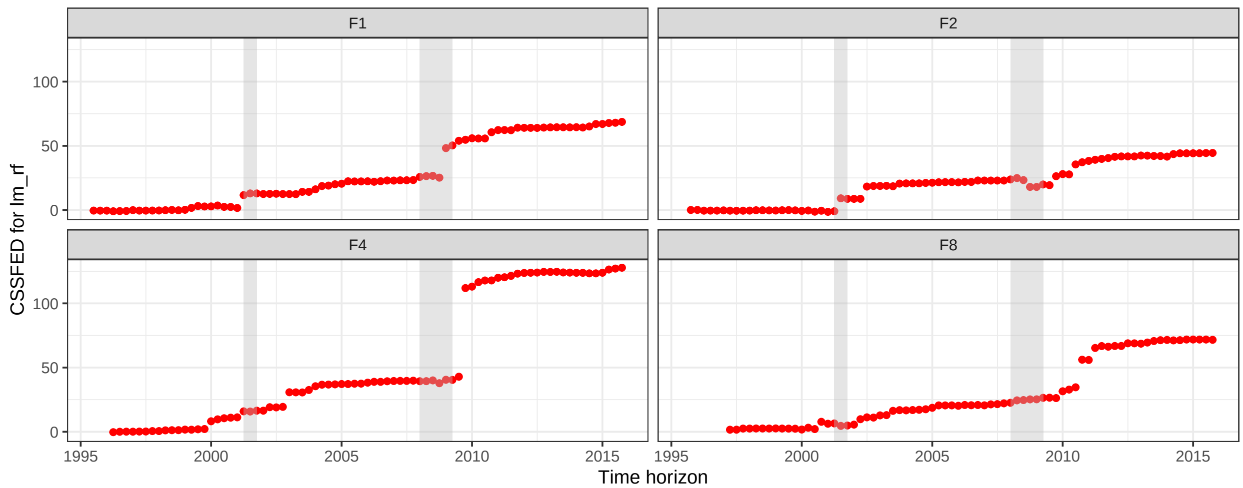


Figure 5.2: Cumulative sum of squared forecast error differential (CSSFED = Linear model - Random forest)

On 5.3) we also plot the importance of different variables as extracted from the random forest model. The figure shows on which variables the random forest model relies the most to make predictions i.e. which variables are the most important for the model's accuracy. As it can be seen, throughout different forecast horizons the importance of the variables change - e.g. for H1 predictions the most important variable is baa_aaa (yield spread between BAA and AAA yields) while for H8 the most important variable is TRM1006 (10 years and 6-month treasury yield spread). During H1 and H2 YIV is 2nd and 1st respectively, however, during H4 and H8 the importance still exists but is among the lowest. Lastly, the dummy variable displays no importance for the random forest model, except for 1 quarter ahead of predictions.

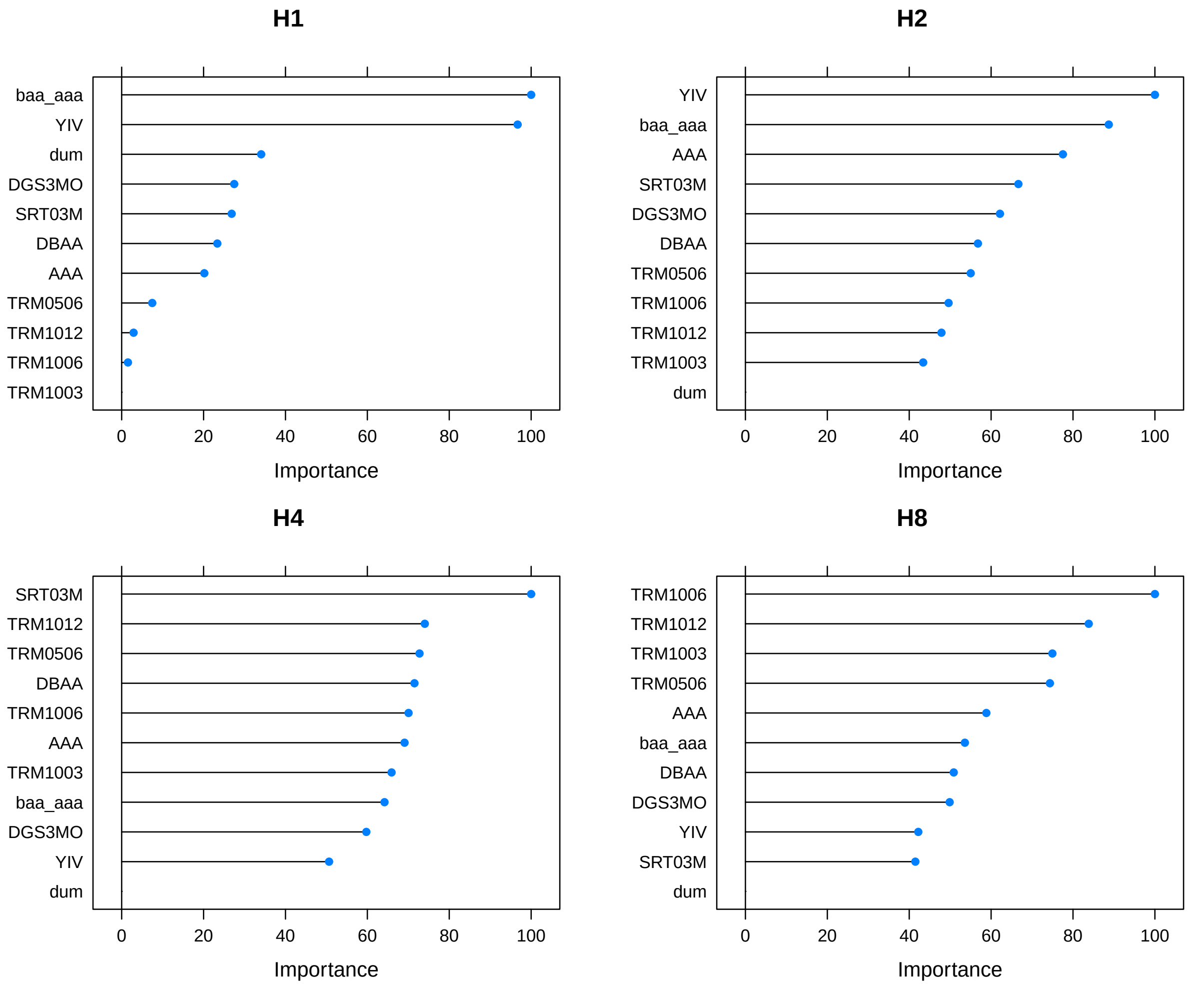


Figure 5.3: Variable importance

6 Discussion

In the following section, we offer context from the literature review to our findings. The goal is to analyze how our findings fit in with the existing research by other academia. Secondly, we give an overview of the limitations of our research and additionally, suggest opportunities for further research on this topic.

6.1 Discussion of results

On the basis of the results stated previously, we can accept the stated hypothesis 1 that treasury implied volatility is indeed a significant predictor throughout the periods from 1 to 8 quarters ahead. This outcome is robust even after including various control variables that have been historically important predictors of real output such as term spreads, credit spreads, stock market implied volatility, and several new residential construction starts. The latter indicates that although accounting for many of the most significant variables, the YIV's predictive ability still persists meaning that YIV can be considered as a solid indicator of GDP growth. Our results are consistent with the paper by Cremers et al. (2021) while the main difference that arises is from std. error. Thus, this verifies the validity of our methodology and that the extracted data can be used for other parts of the research.

Next, the aim was to test the robustness of the model and variable with respect to subsample periods (expansionary and recessionary periods) as from recent literature there exists significant evidence that there are performance asymmetries concerning subsample periods. For example, Chauvet & Potter (2013) conclude that GDP growth is significantly more difficult to predict during recessionary periods. To test it, we include a dummy variable to the regression taking a value of 1 during recessions and 0 during expansions (business cycle classification is taken from NBER). We find that indeed there exist significant performance asymmetries - during recessionary periods the model's accuracy is far worse when compared to the expansionary periods. To further illustrate the point, we plot the root mean squared forecasting errors (RMSFEs) in figure 9.4. As it can be seen from the graph the linear model predicts significantly worse during the recessionary periods when compared to expansionary and full sample periods. Hence, the results confirm our second hypothesis i.e. the full-sample is not robust in making predictions during turbulent periods as the accuracy of the model suffers tremendously.

Lastly, we also accept our third hypothesis that machine learning models, more specifically Random Forest model, can improve the forecast performance. In comparison with the linear model, we were able to yield smaller RMSFE's, most likely due to its ability to take into account non-linearities. This conclusion can be easily identified by utilizing cumulative sum of squared error differences (Figure: 5.2) - as the graphs are upward sloping, the interpretation is that on average for each observation the random forest model has a better forecast accuracy compared to the linear model. This conclusion can be equated to Siliverstovs & Wochner (2020) & Siliverstovs (2021) who conclude that the advanced models, in their case DFM, outperform simple linear models.

Also, it was concluded that the biggest forecasting gains of advanced models tend to be around market turndowns. On the CSSFED plot, it can be seen that big jumps mostly exist occur around the recessionary periods. This means that random forest model has superior performance around recessionary periods. To analyze the matter in-depth, one can examine the models' predicted values compared to the actual GDP growth(RF: 9.5, LM: 9.4). Comparing the two models' predictions to actuals, it can be noticed that random forest fails to fully estimate the declines in GDP during recessions - i.e. it underestimates compared to the linear model. Nevertheless, as the linear models have random spikes, especially during and after recessions, its error is higher. Hence, our results are aligned with the claims of Siliverstovs & Wochner (2020) & Siliverstovs (2021) and Chauvet & Potter (2013) according to whom advanced methods that tackle non-linearities are expected to outperform simple models during turmoils. Also, the fact that random forest fails to reach the depths of recessions with its predictions is consistent with G. Zhang & Lu (2012) who states the underestimation during such recessionary period is a common problem with random forest. To elaborate, this happens due to RF algorithms as it uses averaging when making predictions - thus, the extreme values tend to be underestimated and smaller values overestimated.

6.2 Limitations & further research

The main limitations regarding our work lie in the data quality - as mentioned in the Data section the data was behind the paywall, which drastically limited our options with the research (shorter timeframe & only quarterly data). Therefore, as the first step it would be necessary to obtain the exact data elements required for Black's formula, and then use it to calculate the treasury implied volatility. This would enhance the research in two ways:

1) the data quality is much more reliable and hence it could improve the outcome of results
2) the scope of the research can be broadened drastically

By having the underlying data, the granularity of the work can be improved - for example, one could analyze shorter-term forecasting performance/effects by calculating monthly YIV time series and selecting another proxy for macroeconomic growth that exists on a monthly basis (GDP available only quarterly).

Secondly, as the data quality is much more robust, an in-depth explanatory analysis could be done. In this research, we rather focus on the "if"-part, but not how - i.e. can YIV predict GDP taking into account different circumstances. As an extension of the work, one could analyze how exactly YIV affects GDP in the upcoming periods - since random forest itself is very difficult to interpret and is considered to be rather a black-box model, we recommend looking into a method named macroeconomic random forest (MRF) developed by Coulombe (2020)[3]. The main advance in comparison to the random forest is that MRF further improves the former by adding a linear component - not only does this help with overfitting of RF but it also enables the interpretation of the outcome.

[3]Author has provided also R package that can be downloaded on request from https://philippegouletcoulombe.com/code

7 Conclusion

The research sought to discover whether another financial variable, treasury implied volatility (YIV), can predict macroeconomic real activity (more specifically the growth rate of GDP). Through replicating the research conducted by Cremers et al. (2021), we validate our first hypothesis that YIV indeed is a significant predictor of macroeconomic real activity. Furthermore, we validated that the variable is robust even after controlling for many existing relevant predictors. Secondly, we accept our second hypothesis that the model based on full sample is inefficient in predicting GDP during turbulent periods - this is because business cycle-related asymmetries exist. As mentioned in the results, having calculated 2 business cycle dependent RMSFE's (in addition to the general RMSFE), we can clearly note that the RMSFE is higher during the recessionary period, no matter the model used. Finally, we proceed to combat the shortcomings of the linear model with the aim to improve forecast accuracy. For that, we turn our focus on the machine learning (ML) based method, more specifically the random forest (RF) model, which has proven to combat one of the key shortcomings of the linear model – the ability to account for non-linearities. Through implementing the random forest model we indeed find confirmation to our third hypothesis that using the model we are able to improve the forecasting accuracy measured through the root mean square forecasting error (RMSFE) - furthermore, this result can be generalized to both expansionary and recessionary subsamples. Furthermore, plotting the RF model against the benchmark linear model in terms of CSSFED yields an upward sloping graph, validating our results that the RF model consistently outperforms the linear model.

8 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. (2020). *The Macroeconomy as a Random Forest* [PhD thesis]. https://doi.org/10.2139/ssrn.3633110

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

Cremers, M., Fleckenstein, M., & Gandhi, P. (2021). Treasury yield implied volatility and real activity. *Journal of Financial Economics*, *xxxx*. https://doi.org/10.1016/j.jfineco.2020.12.009

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, James H., & Watson, M. W. (1989). *New Indexes of Coincident and Leading Economic Indicators* (Vol. 4, p. 351). https://doi.org/10.2307/3584985

Stock, James 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

Zhang, G., & Lu, Y. (2012). Bias-corrected random forests in regression. *Journal of Applied Statistics*, *39*(1), 151–160. https://doi.org/10.1080/02664763.2011.578621

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

9 Appendices

9.1 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.

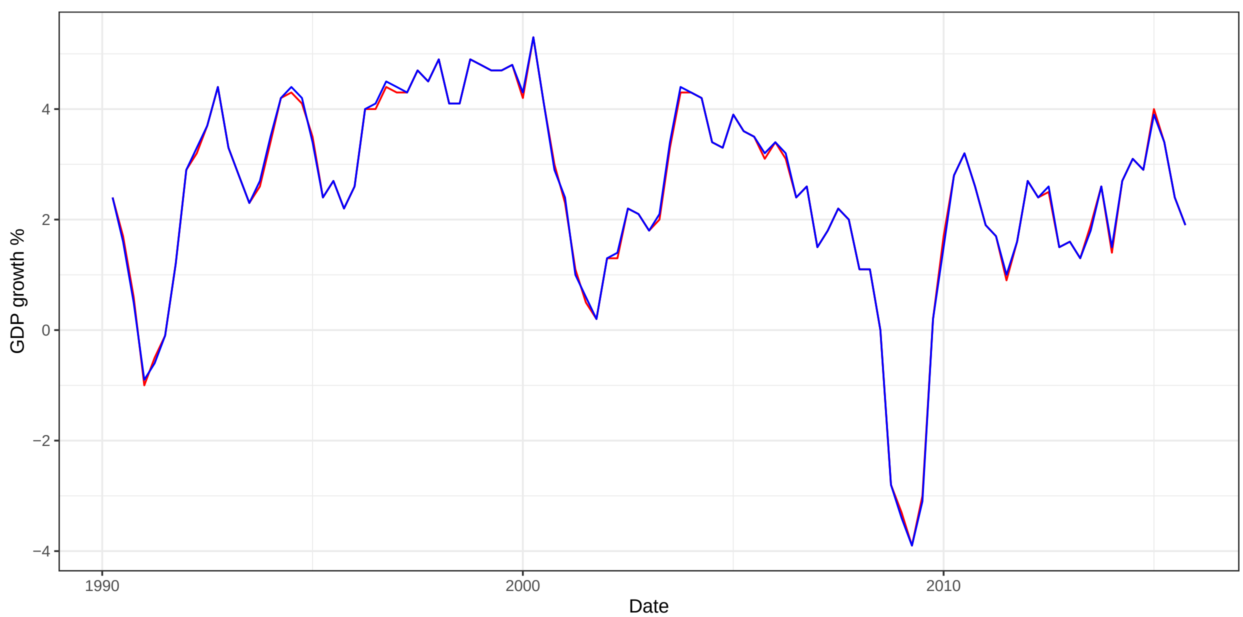


Figure 9.1: Actual vs Extracted GDP growth rate in %

9.2 Appendix B

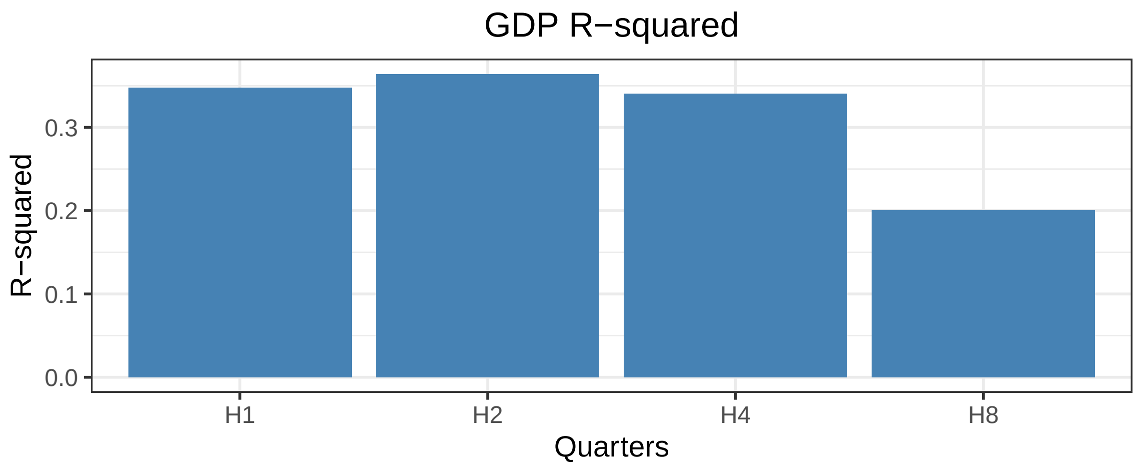


Figure 9.2: Regressions' R-squared

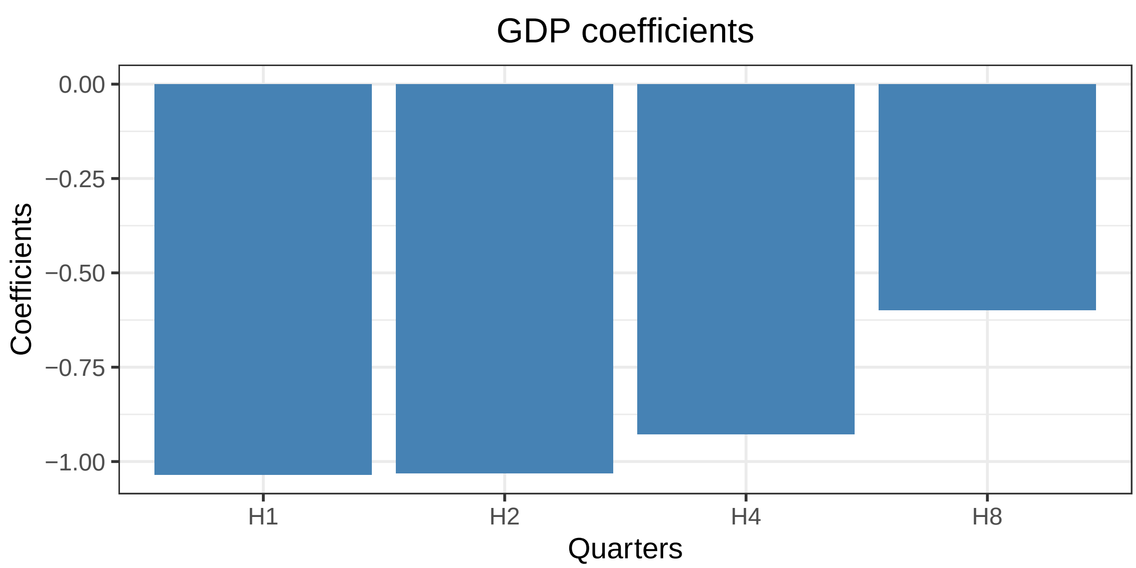


Figure 9.3: Regressions' coefficients

9.3 Appendix D

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

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

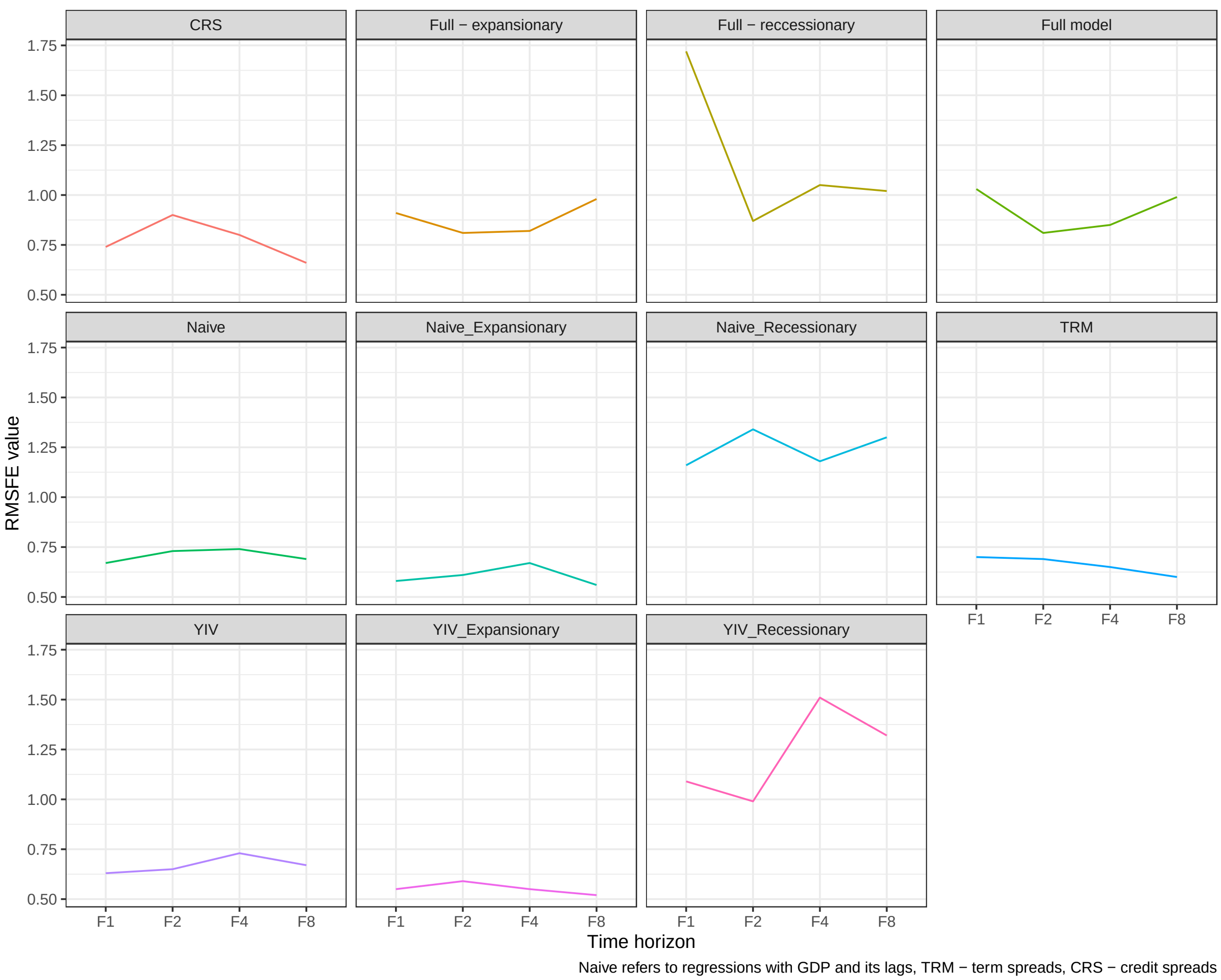
Table 9.1: Regression with state-dependency

	H1	H2	H4	H8
YIV_estimate	-0.13	-0.22	-0.35 *	-0.34 **
YIV_std.error	0.1	0.15	0.18	0.12
dum_estimate	-1.37 ***	-1.81 ***	-2.02 ***	-1.06 **
dum_std.error	0.45	0.55	0.65	0.42
log_gdp_estimate	0.71 ***	0.54 ***	0.29 **	0.21 *
log_gdp_std.error	0.08	0.1	0.14	0.12
TRM0503_estimate	0.48 ***	0.5 **	0.31	-0.12
TRM0503_std.error	0.17	0.23	0.37	0.45
DGS3MO_estimate	0.6 **	0.55	0.08	-0.99
DGS3MO_std.error	0.29	0.39	0.64	0.95
SRT03M_estimate	-0.14	-0.09	0.01	0.14
SRT03M_std.error	0.08	0.1	0.18	0.26
VIX_estimate	0.03	0.03	0	-0.04
VIX_std.error	0.08	0.09	0.11	0.1
AAA_estimate	-0.49 *	-0.39	0.11	1.17
AAA_std.error	0.28	0.38	0.63	0.9
housng_estimate	0	0.07	0.15	0.28 *
housng_std.error	0.06	0.07	0.11	0.14
r.squared	0.86	0.8	0.66	0.54
adj.r.squared	0.85	0.78	0.62	0.49

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

9.4 Appendix E

	F1	F2	F4	F8
YIV	0.63	0.65	0.73	0.67
YIV-Recess.	1.09	0.99	1.51	1.32
YIV-Expans.	0.55	0.59	0.55	0.52
Naive	0.67	0.73	0.74	0.69
Naive-Recess.	1.16	1.34	1.18	1.30
Naive-Expans.	0.58	0.61	0.67	0.56
TRM	0.70	0.69	0.65	0.60
CRS	0.74	0.90	0.80	0.66
full	1.03	0.81	0.85	0.99
full_rec	1.72	0.87	1.05	1.02
full_exp	0.91	0.81	0.82	0.98



9.5 Appendix F

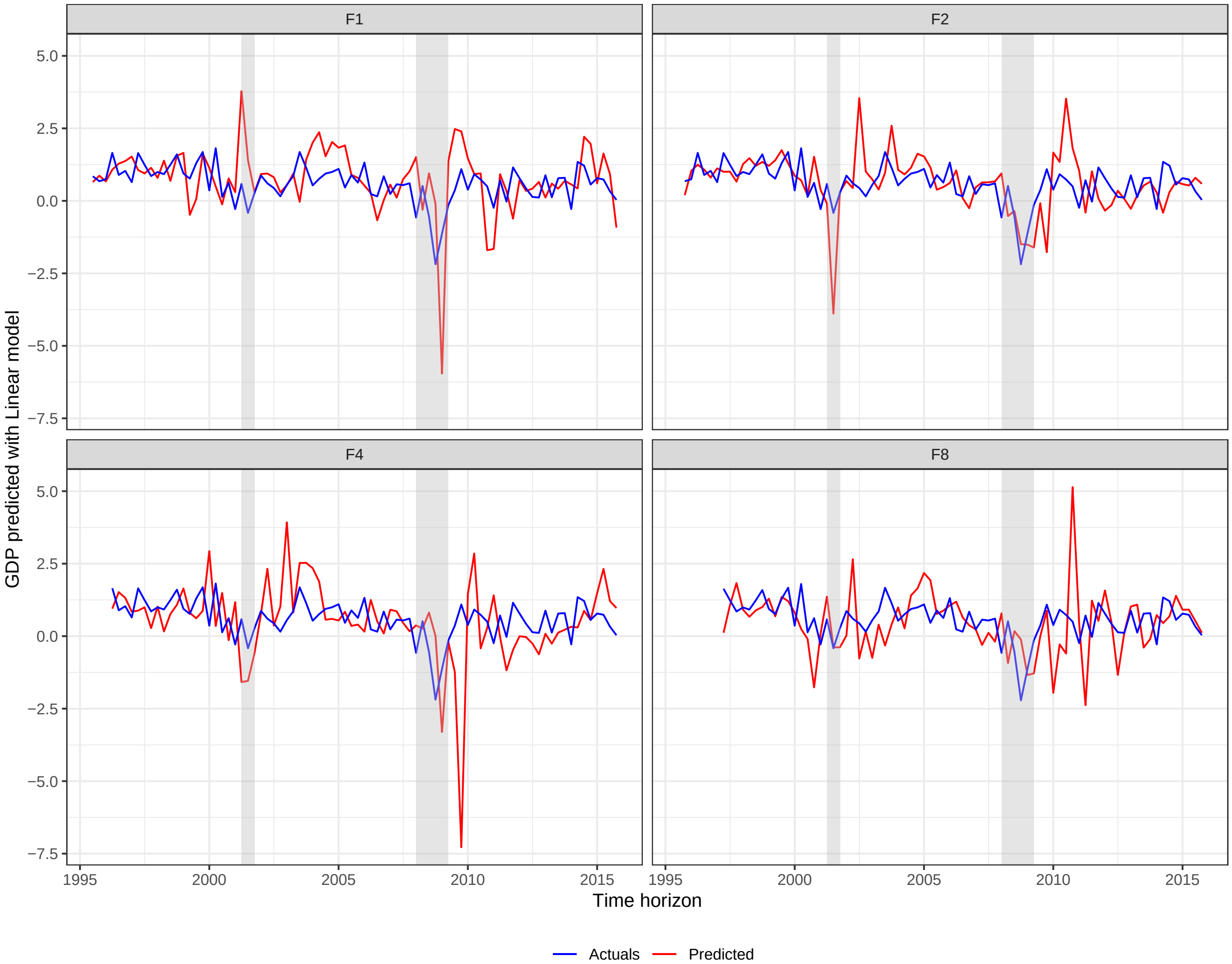


Figure 9.4: Predicted vs actual results (Linear model)

Figure 9.5: Predicted vs actual results (Random forest)

9.6 Appendix G[4]

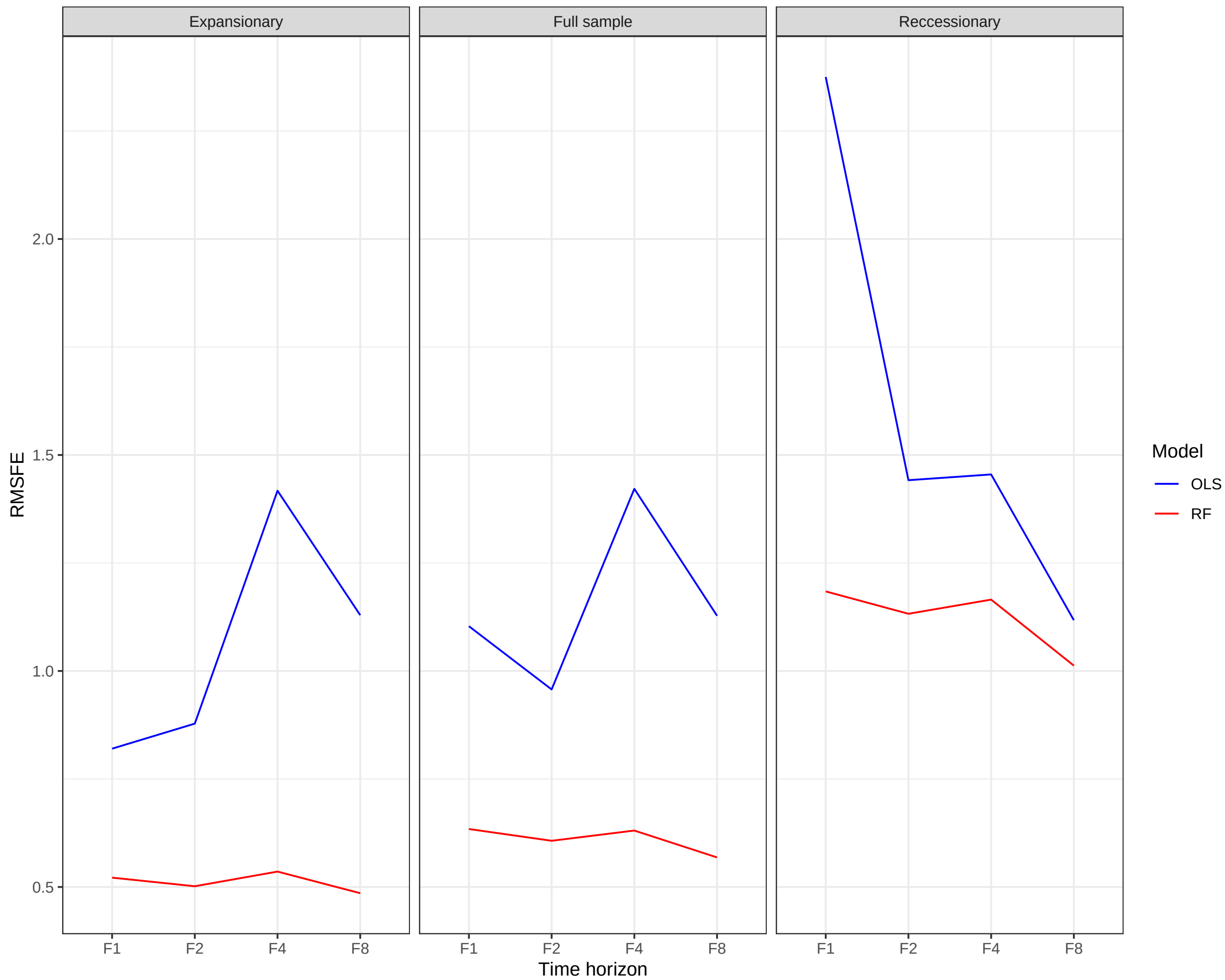


Figure 9.6: RMSFE-s of linear model and random forest

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

9.7 Appendix H

Variable	Description
YIV	5 - year Treasury Implied Volatility
GDP	Real gross domestic product
VIX	Returns of VIX index
DBAA	BAA corporate bond yields
AAA	AAA corporate bond yields
baa_aaa	Yield spread between BAA and AAA yields
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
DGS3MO	Three month corporate bond yield

9.8 Appendix I

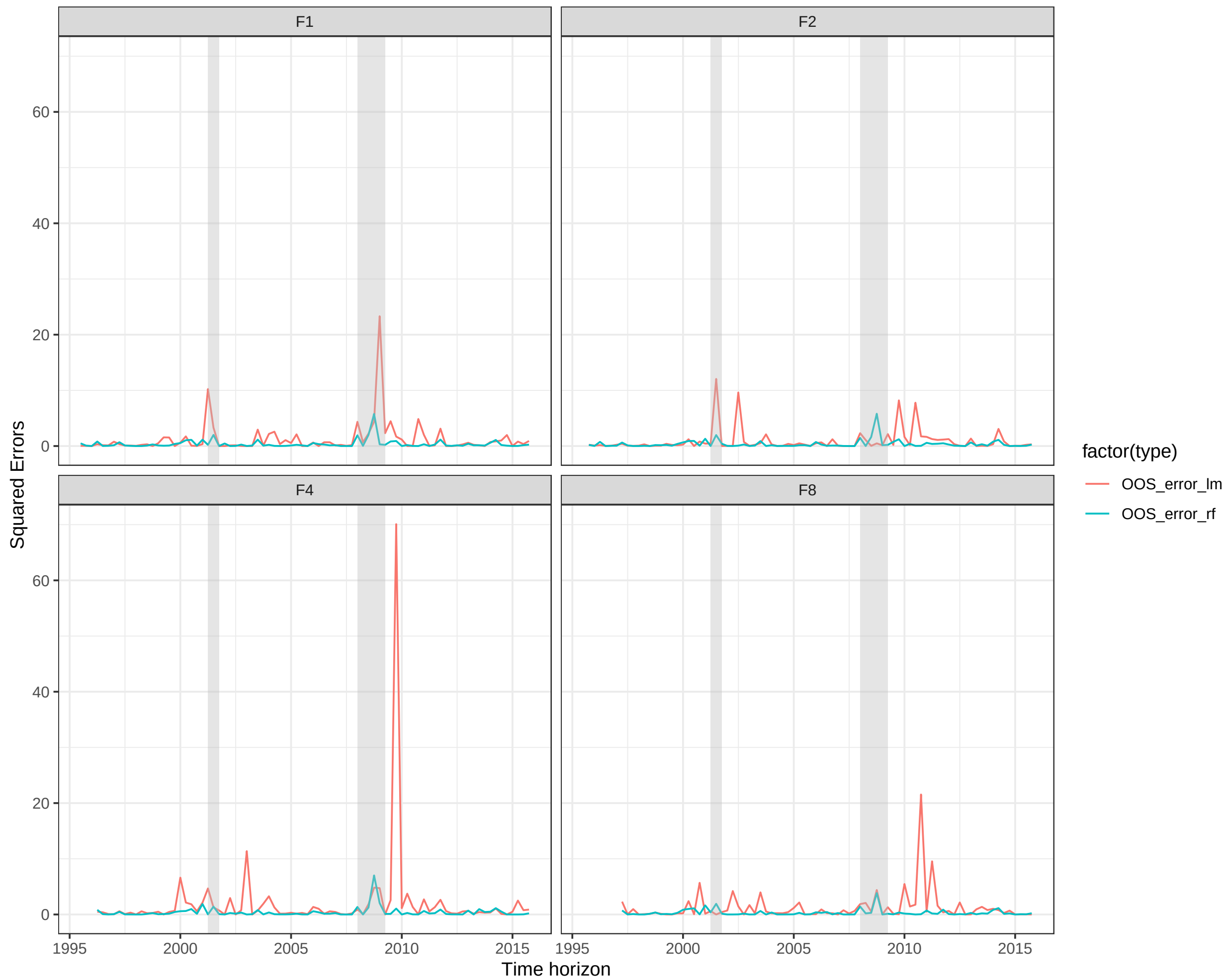


Figure 9.7: Squared errors

9.9 Appendix J

Formula for dependent variable as used in the paper where GDP_{t+j} is the GDP year-on-year growth rate in quarter $t+j$.

$$\frac{1}{h}\sum_{j=1}^{j=h}\ln(1+GDP_{t+j})(\#eq:BS_note) \tag{14}$$

For $h=1$

$$\begin{aligned}
\ln(1+GDP_{t+1}) &= \ln\left(1+\frac{y_{t+1}-y_{t-3}}{y_{t-3}}\right)= && (15)\\
&= \ln\left(1+\frac{y_{t+1}}{y_{t-3}-1}\right)=\ln\left(\frac{y_{t+1}}{y_{t-3}}\right)= && (16)\\
&= \ln y_{t+1}-\ln y_{t-3}= && (17)\\
&= \ln y_{t+1}-\ln y_t+\ln y_t-\ln y_{t-1}+\ln y_{t-1}-\ln y_{t-2}+\ln y_{t-2}-\ln y_{t-3}= && (18)\\
&= (\ln y_{t+1}-\ln y_t)+(\ln y_t-\ln y_{t-1})+(\ln y_{t-1}-\ln y_{t-2})+(\ln y_{t-2}-\ln y_{t-3})= && (19)\\
&= \Delta\ln y_{t+1}+\Delta\ln y_t+\Delta\ln y_{t-1}+\Delta\ln y_{t-2} && (20)
\end{aligned}$$

For $h=4$

$$\begin{aligned}
&\frac{1}{4}(\ln(1+GDP_{t+4})+\ln(1+GDP_{t+3})+\ln(1+GDP_{t+2})+\ln(1+GDP_{t+1}))= && (21)\\
&=\frac{1}{4}(\Delta\ln y_{t+4}+\Delta\ln y_{t+3}+\Delta\ln y_{t+2}+\Delta\ln y_{t+1}+ && (22)\\
&+\Delta\ln y_{t+3}+\Delta\ln y_{t+2}+\Delta\ln y_{t+1}+\Delta\ln y_t+ && (23)\\
&+\Delta\ln y_{t+2}+\Delta\ln y_{t+1}+\Delta\ln y_t+\Delta\ln y_{t-1}+ && (24)\\
&+\Delta\ln y_{t+1}+\Delta\ln y_t+\Delta\ln y_{t-1}+\Delta\ln y_{t-2}+)= && (25)\\
&=\frac{1}{4}(\Delta\ln y_{t+4}+2\Delta\ln y_{t+3}+3\Delta\ln y_{t+2}+4\Delta\ln y_{t+1}+3\Delta\ln y_t+2\Delta\ln y_{t-1}+\Delta\ln y_{t-2}) && (26)
\end{aligned}$$

This formula is for average quarterly growth rate for $h=4$**:**

$$\begin{aligned}
&\frac{1}{4}(\ln(1+GDP_{t+4}))= && (27)\\
&=\frac{1}{4}(\Delta\ln y_{t+4}+\Delta\ln y_{t+3}+\Delta\ln y_{t+2}+\Delta\ln y_{t+1}) && (28)
\end{aligned}$$

Formula for quarterly growth rate for $h = 4$:

$$\Delta \ln y_{t+4} \tag{29}$$