Bachelor Thesis

State-dependent evaluation of Random Forest in forecasting the real activity of the US economy using treasury yield implied volatility

Authors:

Karel Räppo

Henry Enno Turu

Supervisor:

Boriss Siliverstovs

April 2021

Riga

COPYRIGHT DECLARATION AND LICENCE

Names of the authors in full: Karel Räppo, Henry Enno Turu

Title of the Thesis: State-dependent evaluation of Random Forest in forecasting the real activity of the US economy using treasury yield implied volatility

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 persons named below are the authors of this thesis.

Pursuant to Article 40 of the Copyright Law the authors hereby agree 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 we are indicated as the authors of the thesis according to Clause 4 Section 1 Article 14 of Copyright Law. We assert our right to be identified as the authors of this thesis whenever it is reproduced in full or in part.

Signed

/Signed/ /Signed/

Karel Räppo Henry Enno Turu

Date

18.05.2021

Contents

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 the GDP growth rate of the US economy. Additionally, we test whether the model's robustness holds 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 - Random Forest. Our results indicate that indeed YIV is a significant predictor of macroeconomic real activity; however, it is crucial 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 Random Forest, 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.

Acknowledgement

We would like to express gratitude to our supervisor Boriss Siliverstovs who has been of great help throughout the whole research process and enabled the work's quality. We especially appreciate the technical guidance with regards to the methodology section of the paper.

1 Introduction

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

One of the most comprehensive papers quantifying the financial market's ability to forecast the macroeconomic real activity was published by Stock & Watson (2003). The researchers reviewed over 93 working papers & articles and documented numerous variables (interest rates, term spreads, returns, exchange rates, etc.) that have been used for forecasting the macroeconomy. Nevertheless, there is no firm agreement on 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, there are even fewer variables that can predict the macroeconomic real activity through more extended periods (i.e. periods of over a year).

In a recent work by Cremers et al. (2021), a new variable for predicting macroeconomic real activity was presented - treasury implied volatility (YIV) - which is calculated using options on treasury bond futures of different maturity periods. YIV can be considered as a measure 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. In the paper, the authors show that YIV is able to forecast the real economy consistently, especially focusing on forecasting 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 the growth rate of 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 (2021), we test whether business cycle phases affect YIV's impact on future GDP growth. In addition, we investigate whether out-of-sample forecast performance is dependent on the business cycle phase - i.e. calculating root mean square forecasting errors (RMSFE) including only recessionary periods or expansionary periods (classification of the US business cycle phases obtained from National Bureau of Economic Research (NBER)).[1]

Secondly, we propose using ML 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 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 RMSFE. While RF's output is difficult to interpret we do not consider it a problem as we are focusing on evaluating the forecasting accuracy of the model (i.e. comparing RMSFE-s).

Thus, our research investigates the following three questions:

1. **Can options on Treasury bond futures (YIV) effectively forecast USA's real activity?**
2. **If so, is this conclusion robust even after considering different business cycle phases and control variables?**
3. **Is Random Forest effective in improving the forecast accuracy?**

We find indeed that YIV is a significant predictor of the US GDP growth rate. Furthermore, we find that it is crucial to take into account the different phases of business cycles in forecasting as there exist significant performance asymmetries between recessionary periods and expansionary periods. Lastly, we show that through using Random Forest the forecasting

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

accuracy (as measured by RMSFE) can be improved significantly when compared to a simple linear model.

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 method used for obtaining it. Subsequently, we offer an in-depth overview of the methodology, which is followed by a description of the results. In the discussion part, we elaborate on how our findings fit in with the existing research. Lastly, we list any possible limitations of the paper, suggest ideas for further research, and present our conclusions.

2 Review of literature

Before digging deeper into the predictive ability of YIV, it is essential to understand 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 Stock & Watson (1989). Ang et al. (2006) follow a similar route and use yield curve to obtain the best maturity short rate for forecasting GDP. Using the process above, they conclude that contrarily to the existing research, short rate is a better predictor compared to any of the term spreads. Furthermore, it is noted that the best form of the slope is the one constructed with maximum maturity difference. Lastly, Gilchrist & Zakrajšek (2012) construct a corporate bond credit spread index which is not only a significant predictor of macroeconomic activity for different variables and time-horizons but also its predictive ability outperforms commonly used BAA-AAA corporate bond spread.

One somewhat 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 cycle phases. 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 a similar 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 the 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 a 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 a 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 the possibility of a recession.

2.2 YIV as a predictor of real activity

Building upon the various aforementioned researches, Cremers et al. (2021) have analyzed over 20 years of data to find out whether YIV could be used to predict different macroeconomic and financial measures such as growth and 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 YIV series using Black's model, which in essence is an adjusted model of the famous Black-Scholes model to value options on futures 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, K is the exercise price, N is cumulative standard normal distribution function. Using the formula and deriving sigma, one can obtain the value of YIV.

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 futures 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 average the time series to obtain monthly data, which is regressed with different macroeconomic and control variables - for example, YIV's effect on the growth rate and volatility of GDP is analyzed. To validate YIV's ability to forecast GDP growth, several aforementioned control variables such as term spreads, credit spreads, stock market volatility, housing starts, etc., are included in the model.

Even though Cremers et al. (2021) possess the data for different maturities (1,5,10,30 years), the researchers 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 a maturity of 5-year, both for cash and futures markets.

2.3 Forecasting asymmetries

In the relatively recent wave of research on forecasting, the problems and limitations regarding full-sample forecasting have become more prevalent. More specifically, it has been found that the business cycle phases 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 the US recessions and expansions from NBER. They conclude that for all different models tested, the GDP growth is significantly harder to forecast during recessions when compared to expansions. 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 et al. (2021) might not be robust throughout the different business cycle phases. 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," Bolhuis & Rayner (2020) bring out the following four key shortcomings of a simple OLS-based forecasting model: collinearity, dimensionality, predictor relevance, and nonlinearity. The latter is also emphasized by Chauvet & Potter (2013), who summarize that the biggest errors for linear models occur near the recessionary periods as the linear relationships tend to break.

To combat some of these shortcomings (mainly collinearity and dimensionality), Siliverstovs & Wochner (2021) assessed the forecasting performance of a dynamic factor model (DFM) compared to a simple autoregressive model. Furthermore, they also included the state-dependent subsamples in 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 how influential observations affect relative forecast accuracy during the Covid-19 crisis. By employing a 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 phase[2]. 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 put it simply, some of the key properties of DFM are its ability to work with large datasets with high dimensional data and predicting comovements of many macroeconomics variables (Stock & Watson (2016)). In our research, however, we are using a limited set of proven control variables (as was done by Cremers et al. (2021)) and focusing on predicting only one variable - the growth rate of GDP. Thus, for our research, using a DFM doesn't serve its purpose.

However, another solution to account for some of the shortcomings of simple OLS based models can be found through implementing 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 important YIV is in predicting macroeconomic real activity and how does it compare against other academically proven predictors.

To sum up, our paper builds upon Cremers et al. (2021) while advancing the methodology in 2 ways:

1. Testing YIV's predictive ability depending on the business cycle phase
2. Introducing ML method in search of more accurate forecasting performance

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

Thus, our hypotheses are the following:

Hypothesis 1. Treasury implied volatility 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. (2021) is inefficient in predicting during turbulent time periods.

Hypothesis 3. Using Random Forest, it is possible to significantly reduce the RMFSE of the forecast, mainly thanks to its ability to take into account nonlinearity.

3 Methodology

3.1 Linear regressions

As already mentioned before, the first part of our methodology consists of replicating the research 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, $\text{GDP}_{\text{i,t+j}}$ 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 next 4 quarters.

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

where $GDP_{i,t+j}$ refers to forward-looking GDP, $\sigma_{IV,t}$ to YIV at time t.

In the linear models, we compare the predictive ability of YIV within different time periods (H=1,2,3,..,12). We also check whether there exist any asymmetries regarding economic cycle phases - 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 NBER.

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

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.

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

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.

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).

$$\begin{gathered} SFE = \sum_{j=1}^{j=H} (log(1+GDP_{i,t+j}) - log(1+\widetilde{GDP}_{i,t+j})^2 \\ RMSFE = \sqrt{mean(SFE)} \end{gathered} \tag{6}$$

where $log(1+GDP_{i,t+j)}$ refer to actual values and $log(1+\widetilde{GDP}_{i,t+j})$ to predicted values.

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 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 (see Equation (6)). Having the SFE-s, we take the mean from the values and then take the square root to obtain RMSFEs..

3.3 Out of sample validation for different economic cycle phases

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 (2021). 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 RMSFEs 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 A 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 (see Appendix B. 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.

$$rRMSFE = \frac{RMSFE_{subset}}{RMSFE_{full}} \tag{7}$$

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 Cumulative sum of squared error difference

As the recessionary periods in comparison with expansionary periods 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 CSSFED 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:

$$CSSFED = \sum_{t=1}^{T}[(e_{benchmark,t})^2 - (e_{advanced,t})^2] \tag{8}$$

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 relative forecasting performance.

3.4 Machine learning methods

As 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 economic crises, we expect to find forecasting gains by utilizing RF. In our research, we focus specifically on RF model due to its advantages discussed in the literature review.

3.4.1 Decision Trees

In order to dig deeper into the methodology of RF, it is crucial to understand one of its core elements - decision trees. Decision tree is a rather straightforward 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 divides (splits) the whole sample into two subsets by user-defined criteria. In our case, trees are used for regressions and as the dependent variable is continuous, the selected criteria for each split is to reduce mean squared error (MSE). Thus, the algorithm takes the whole dataset and picks the variables that possess the biggest influence on the dependent variable, and then splits the dataset further until stopping criteria is met. Consequently, it arrives at the final (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, there are no strong assumptions about the underlying data and its form, which in return enables capturing 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 RF was proposed by Breiman (2001). RF itself is an ensemble based supervised learning method. As the name suggests there are two main elements behind it.

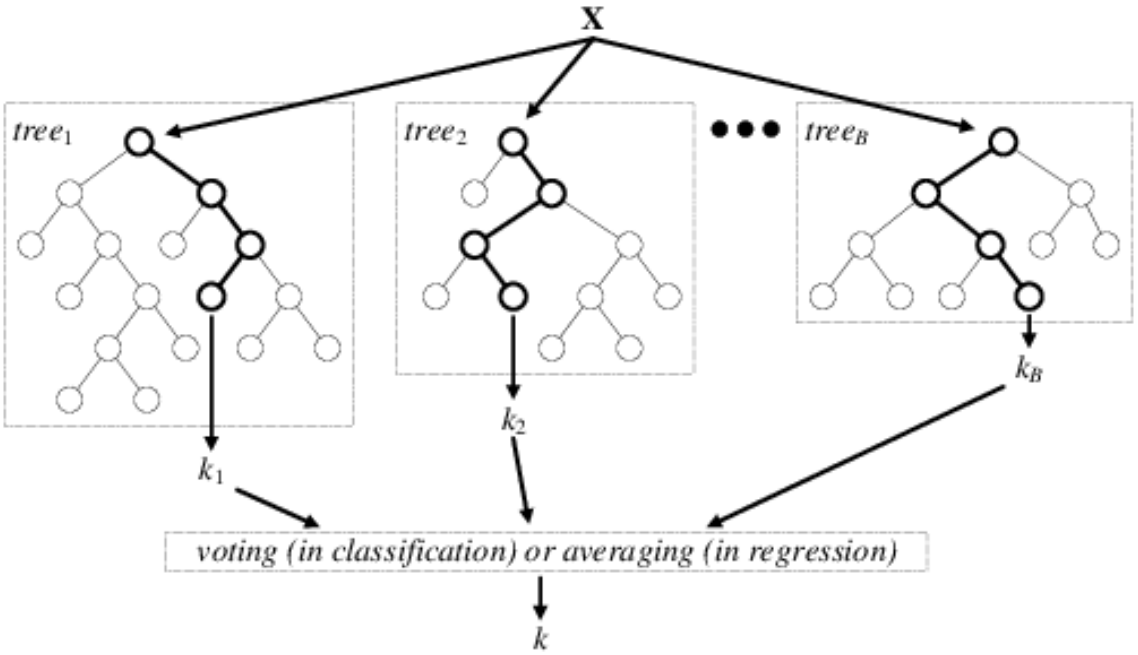


Figure 3.1: Random forest. Figure from (Verikas et al., 2016)

First is the randomness part - RF uses bootstrap aggregation (bagging) to construct random samples of the initial training dataset. Furthermore, the randomness is also included in variable selection - it randomly selects n number of variables of the total set N used for splitting at each node. Secondly, forest, which refers to the aggregation means that 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 and 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 X is created.
2) Then the model randomly selects n number variables amongst the total set N.
3) Subsequently, the best variable and splitting criterion is selected, on the basis of which the current node is splitted into two sub-nodes. 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) and then model prediction k is made.
5) The output k is achieved by averaging the estimations of each tree(k1..kb) in the model.

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 A)
- Control Variables:
  - US treasury interest rates to construct term spreads
  - credit spreads
  - credit spread index
  - 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 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 Appendix C).

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). Based on the obtained risk free rates we calculate term spreads which are defined as the differences

- 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, in addition to the 3-month treasury note, we also construct (SRT03M) which is defined as its change compared to the previous quarter. 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 4.1 below. 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 analysis, we exclude 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.

All the variables and the respective descriptions have been brought out in the Appendix D.

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 in Figure 5.1, 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 considerable 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.

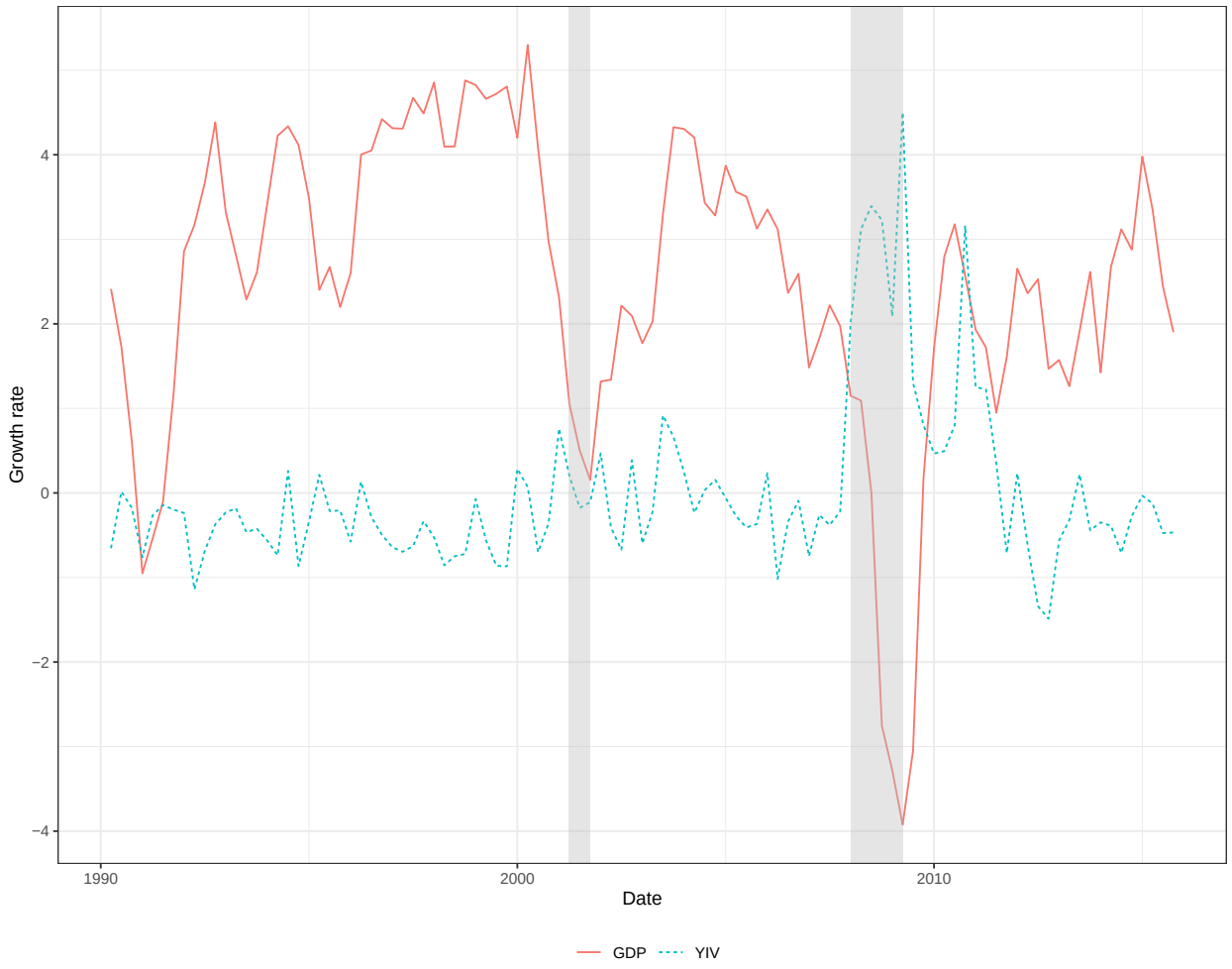


Figure 5.1: GDP Growth(%) vs 5-year Treasury Implied Volatility

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 Table 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 one standard deviation increase in YIV results in a -1.03/4 * 1.31% = 0.34% decrease in GDP growth within the next quarter. For predicting 8 quarters ahead (2 years), one standard deviation increase in YIV results in 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.

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, the inclusion of dummy decreased the significance of YIV while improving R-squared - 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.

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

Additionally, we constructed a regression with YIV and control variables as dependent variables - as it can be seen from the Appendix E YIV is still highly significant throughout all of the forecasting horizons. Finally, we combined the previous variables with all of the available control variables in Cremers et al. (2021) into one regression equation (Appendix F. After including all the relevant control variables, the YIV remains significant at 5% in H4 and at 10% in H8. In this model, the regression results implicate that one standard deviation increase in YIV should result in a -0.35 * 1.31% = -0.46 % decrease in GDP growth within the next 4 quarters. Additionally, when predicting 8 quarters ahead, a one standard deviation increase in YIV results in 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 (OOS) 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 a 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 G), 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 H) - i.e RMSFE's tend to increase with the forecast horizon, except for F2. In other words, the model's accuracy in predicting GDP growth got worse in predicting further time periods. Looking at performance of the full-model, it can be seen that it yields the highest forecast errors compared to individual panels.

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 (H), 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

Table 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 RF model to see whether the forecasting accuracy can be improved. As it can be seen from Figure 5.2 the overall performance of the RF model is better when compared to OLS across all of the prediction periods as indicated by the upward sloping graph (meaning that OLS's squared errors are larger than the respective squared errors of the RF model). Furthermore, the same conclusion is derived when looking at the Appendix I where it can be seen that the RF model significantly outperforms during the full sample, and also separately in both subset periods. Additionally, these results are consistent with Appendix G and Appendix J. The latter can be used to visually validate the difference between the accuracy of RF 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 RF model. Also, this can be noted that on the CSSFED graph below since the bigger jumps tend to happen around recessionary periods.

In Figure 5.3, we have plotted the importance of different variables as extracted from the RF model. The figure shows on which variables the RF 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

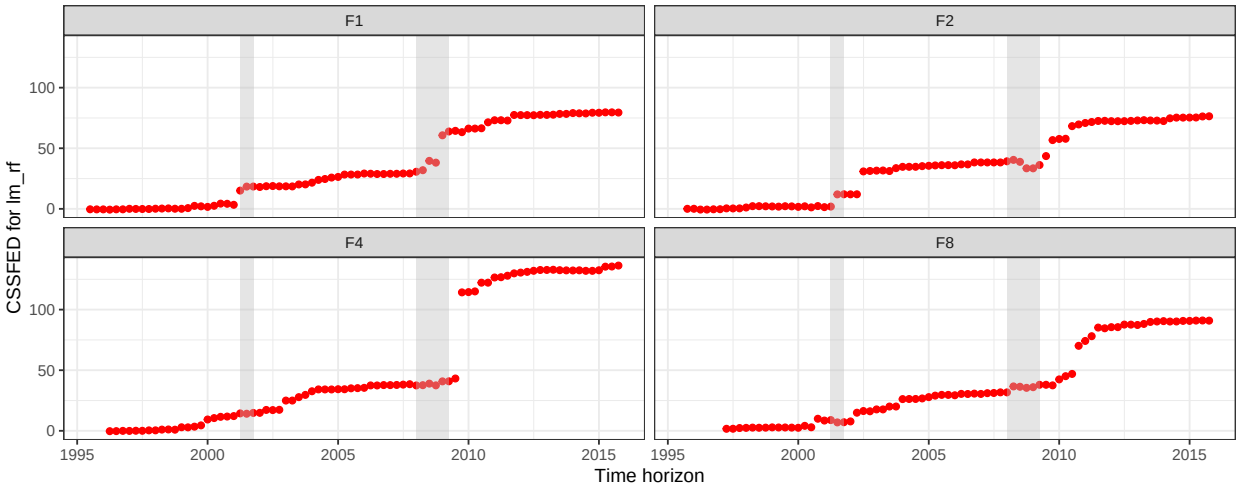


Figure 5.2: Cumulative sum of squared forecast error differential (Linear model - Random Forest)

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 RF model, except for 1 quarter ahead of predictions.

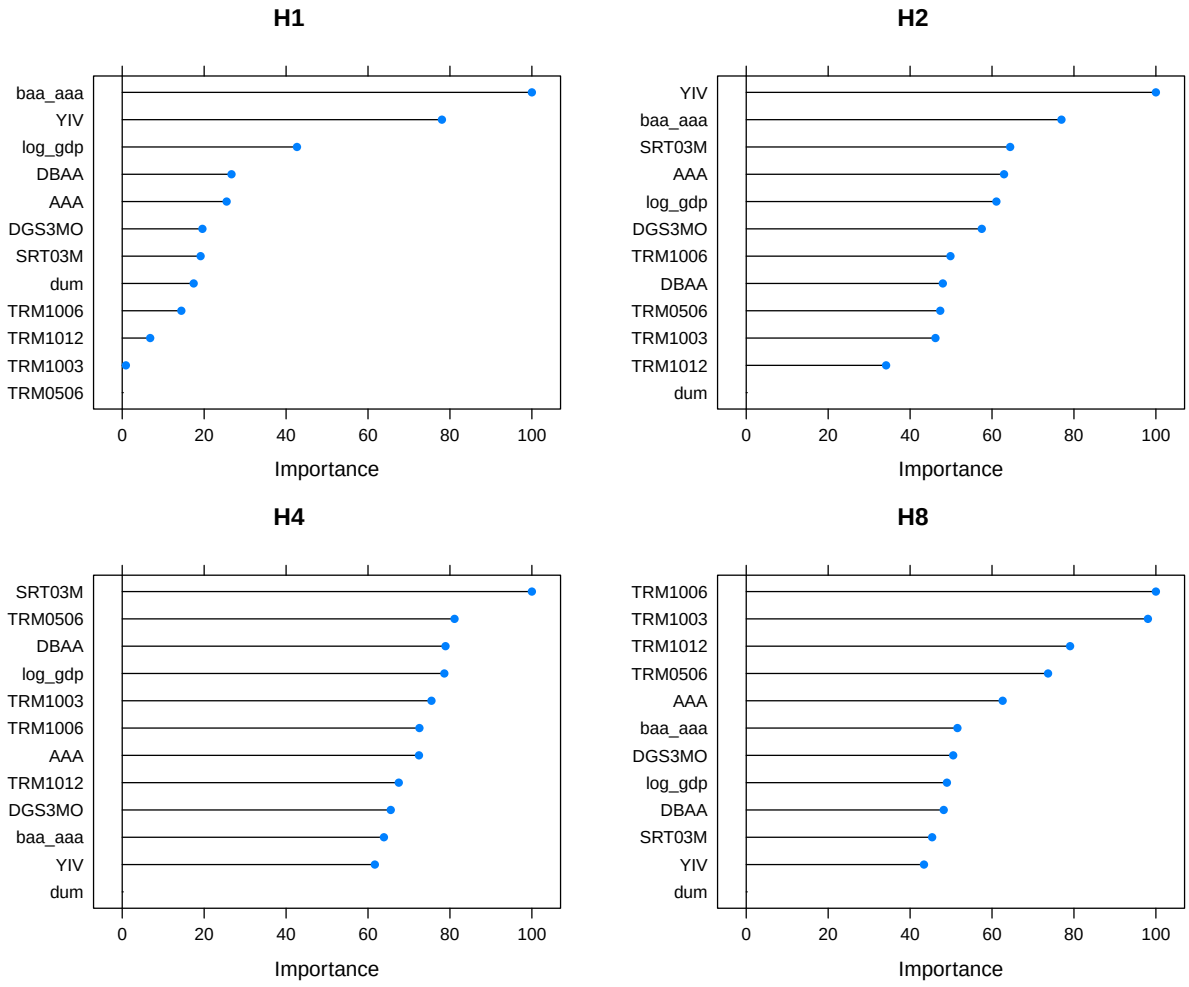


Figure 5.3: Variable importance

6 Discussion

In the following section, we offer context from the literature review to our findings. The goal is to analyze how our results 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

Based on 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 new residential construction starts. The latter indicates that although accounting for many of the most significant variables, the YIV's predictive ability persists meaning that YIV can be considered as a solid indicator of GDP growth. In general, our results are consistent with the paper by Cremers et al. (2021) verifying 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 is significant evidence that performance asymmetries concerning subsample periods exist. 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 in the regression, taking a value of 1 during recessions and 0 during expansions (business cycle phase classification is taken from NBER).

Firstly, from the in-sample regressions results, we can see that after the inclusion of dummy variable, YIV and controls lose most of their significance. This is due to the fact that they all possess similar informational content regarding future recessions but especially for the

closer time periods, dummy can do so with less error - i.e if the current quarter is a recession, there is a high probability that the next period is also a recession. Nevertheless, as the forecast period increases, the significance and also coefficients decrease, whereas the same figures for YIV increase. Also, as the coefficient for dummy variable is significant throughout the whole forecast horizon, it indicates that structural break in the data exists (implying non-linearities).

Next, we construct out of sample tests. By plotting the root mean squared forecasting errors (RMSFES) in Appendix H, we find that there indeed exist significant performance asymmetries - during recessionary periods, the model's accuracy is worse when compared to the expansionary periods. To further illustrate the point, we can analyze predicted vs actual values on Appendix G. The linear model predicts significantly worse around the recessionary periods when compared to expansionary 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 the RF model is effective in improving the forecasting 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 the RF model has a better forecasting accuracy compared to the linear model. This conclusion can be equated to Siliverstovs & Wochner (2021) & 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 RF model over the linear model tend to be around market turndowns. On the CSSFED plot, it can be seen that big jumps primarily occur around the recessionary periods. This means that the RF model has superior performance around recessionary periods. To analyze the matter in-depth, the models' predicted values compared to the actual GDP growth can be examined (Appendix G).

Comparing the two models' predicted and actual values, it can be noticed that RF fails to fully estimate the declines in GDP growth rates during recessions, i.e. it underestimates compared to the linear model. Nevertheless, as the linear model has significant spikes, especially during and after recessions, its overall error is higher. Hence, our results are aligned with the claims of Chauvet & Potter (2013), according to whom advanced methods that tackle non-linearities are expected to outperform simple models during turmoils. Also, the fact that RF fails to reach the depths of recessions with its predictions is consistent with G. Zhang & Lu (2012), who states the underestimation during recessionary period is a common problem with RF. As it can be seen from Figure G.2, RF is not able to reach the depths during recessionary periods when compared to the linear model on Figure G.1. This stems from the fact that RF cannot extrapolate values outside of the range of values in the training dataset. This, however, isn't a problem for the linear model.

6.2 Limitations and further research

The main limitations regarding our work lie in the quality of data - as mentioned in the Data section, the YIV 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 is available only on a quarterly basis).

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 growth rate taking into account different circumstances. As an extension of the work, one could analyze how exactly YIV affects GDP growth rate in the upcoming periods - since RF 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 RF 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 growth around turbulent periods - this is because business cycle-related asymmetries exist. As mentioned in the results, having calculated two business cycle phase 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 Random Forest, 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*. IMF.

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

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. (2021). State-Dependent Evaluation of Predictive Ability. *Journal of Forecasting*. https://doi.org/10.1002/for.2715

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

Stock, J. H., & Watson, M. W. (2003). Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature*, *41*(3), 788–829. https://doi.org/10.1257/jel.41.3.788

Stock, J. H., & Watson, M. W. (2016). *Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics* (1st ed., Vol. 2, pp. 415–525). Elsevier B.V. https://doi.org/10.1016/bs.hesmac.2016.04.002

Verikas, A., Vaiciukynas, E., Gelzinis, A., Parker, J., & Charlotte Olsson, M. (2016). Electromyographic patterns during golf swing: Activation sequence profiling and prediction of shot effectiveness. *Sensors (Switzerland)*, *16*(4). https://doi.org/10.3390/s16040592

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

A Appendix A

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

$$\frac{1}{h}\sum_{j=1}^{j=h}\ln(1+GDP_{t+j}) \tag{9}$$

For $h=1$

$$\ln(1+GDP_{t+1}) = \ln\left(1+\frac{y_{t+1}-y_{t-3}}{y_{t-3}}\right) = \tag{10}$$

$$= \ln\left(1+\frac{y_{t+1}}{y_{t-3}-1}\right) = \ln\left(\frac{y_{t+1}}{y_{t-3}}\right) = \tag{11}$$

$$= \ln y_{t+1} - \ln y_{t-3} = \tag{12}$$

$$= \ln y_{t+1} - \ln y_t + \ln y_t - \ln y_{t-1} + \ln y_{t-1} - \ln y_{t-2} + \ln y_{t-2} - \ln y_{t-3} = \tag{13}$$

$$= (\ln y_{t+1} - \ln y_t) + (\ln y_t - \ln y_{t-1}) + (\ln y_{t-1} - \ln y_{t-2}) + (\ln y_{t-2} - \ln y_{t-3}) = \tag{14}$$

$$= \Delta \ln y_{t+1} + \Delta \ln y_t + \Delta \ln y_{t-1} + \Delta \ln y_{t-2} \tag{15}$$

For $h=4$

$$\frac{1}{4}(\ln(1+GDP_{t+4}) + \ln(1+GDP_{t+3}) + \ln(1+GDP_{t+2}) + \ln(1+GDP_{t+1})) = \tag{16}$$

$$= \frac{1}{4}(\Delta \ln y_{t+4} + \Delta \ln y_{t+3} + \Delta \ln y_{t+2} + \Delta \ln y_{t+1} + \tag{17}$$

$$+ \Delta \ln y_{t+3} + \Delta \ln y_{t+2} + \Delta \ln y_{t+1} + \Delta \ln y_t + \tag{18}$$

$$+ \Delta \ln y_{t+2} + \Delta \ln y_{t+1} + \Delta \ln y_t + \Delta \ln y_{t-1} + \tag{19}$$

$$+ \Delta \ln y_{t+1} + \Delta \ln y_t + \Delta \ln y_{t-1} + \Delta \ln y_{t-2} +) = \tag{20}$$

$$= \frac{1}{4}(\Delta \ln y_{t+4} + 2\Delta \ln y_{t+3} + 3\Delta \ln y_{t+2} + 4\Delta \ln y_{t+1} + 3\Delta \ln y_t + 2\Delta \ln y_{t-1} + \Delta \ln y_{t-2}) \tag{21}$$

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

$$\frac{1}{4}(\ln(1+GDP_{t+4})) = \tag{22}$$

$$= \frac{1}{4}(\Delta \ln y_{t+4} + \Delta \ln y_{t+3} + \Delta \ln y_{t+2} + \Delta \ln y_{t+1}) \tag{23}$$

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

$$\Delta \ln y_{t+4} \tag{24}$$

B Appendix B

Following figures describe the effect of averaging with the alternative variables of GDP growth. As it can be seen, the variance declines drastically between the two forecast periods H1 (N1) and H8 (N8), which consequently affects the RMSFEs.

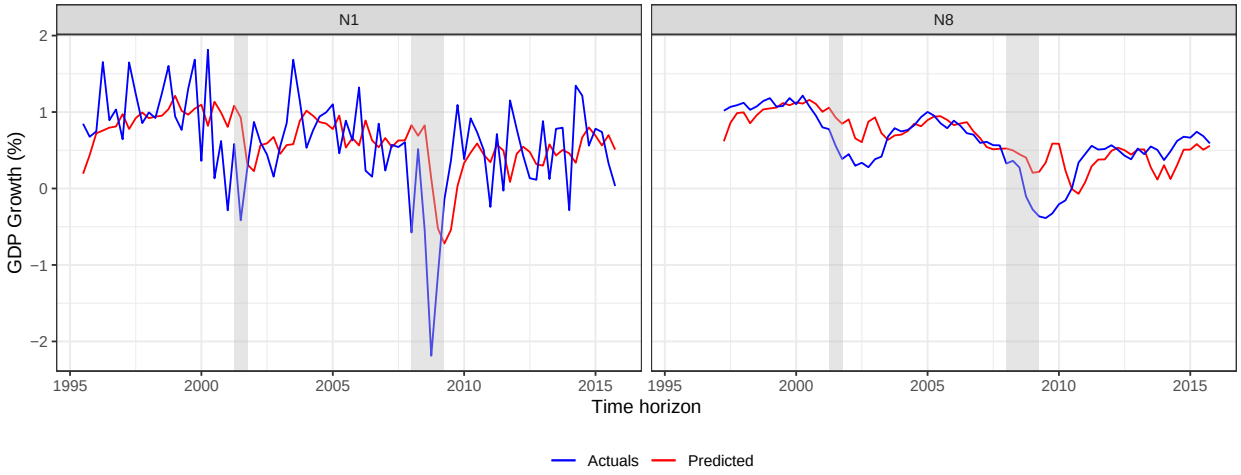


Figure B.1: Average quarterly growth rates of GDP h-quarters ahead

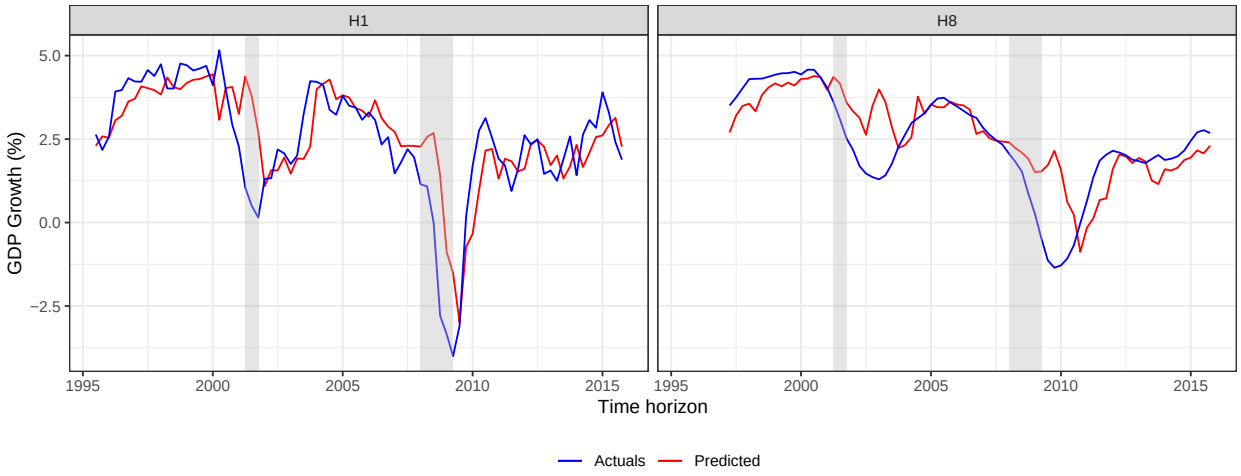


Figure B.2: Average quarterly year-on-year growth rates (as in the replicated paper)

C Appendix C

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 interclass 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.

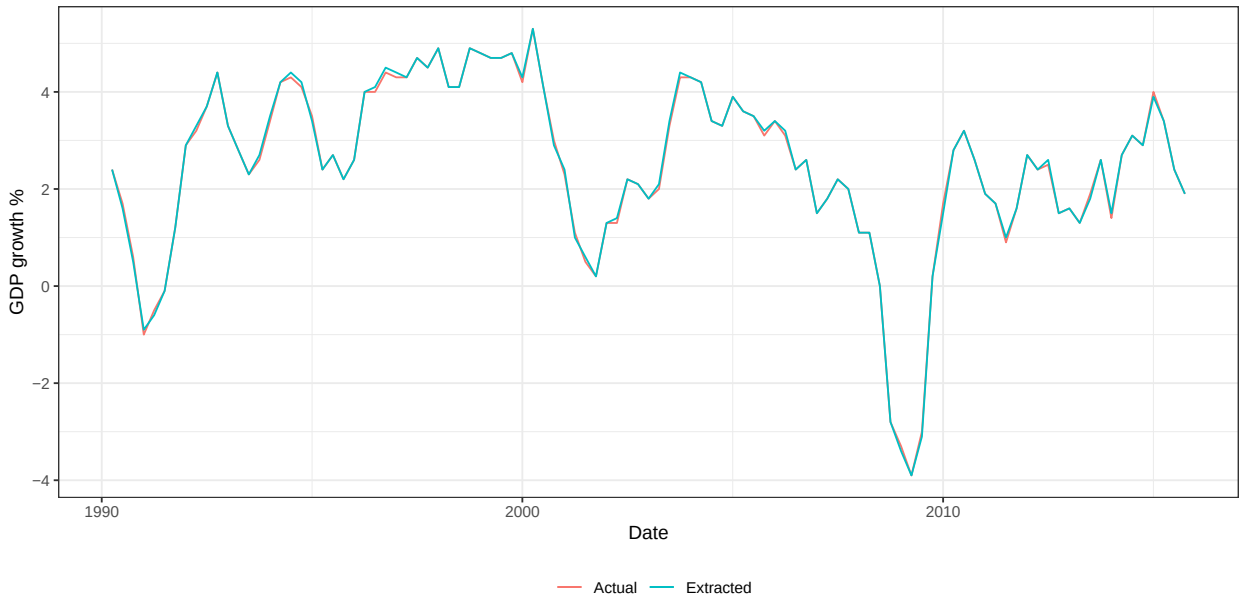


Figure C.1: Actual vs Extracted GDP growth rate in %

D Appendix D

Table D1: Overview of main variables

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

E Appendix E

Notes: The following regressions includes YIV and controls as independent variables. For the description of the variables in the output, please refer to Appendix D. The specification for the regression is the following:

$$\frac{1}{H}\sum_{j=1}^{j=H} log(1+GDP_{i,t+j}) = \alpha_H + \beta_H \sigma_{IV,t}^{INT} + Controls + \varepsilon_{t+H} \tag{25}$$

Table E1: Regression with state-dependency

	H1	H2	H4	H8
YIV_estimate	-0.28 **	-0.41 **	-0.57 **	-0.46 ***
YIV_std.error	0.12	0.19	0.24	0.15
log_gdp_estimate	0.8 ***	0.66 ***	0.41 ***	0.28 **
log_gdp_std.error	0.08	0.1	0.13	0.12
TRM0503_estimate	0.54 ***	0.58 **	0.4	-0.07
TRM0503_std.error	0.18	0.25	0.38	0.45
DGS3MO_estimate	0.6 **	0.56	0.09	-0.98
DGS3MO_std.error	0.3	0.41	0.66	0.92
SRT03M_estimate	-0.02	0.07	0.18	0.23
SRT03M_std.error	0.09	0.11	0.2	0.26
VIX_estimate	-0.03	-0.05	-0.09	-0.09
VIX_std.error	0.08	0.1	0.12	0.1
AAA_estimate	-0.61 **	-0.55	-0.07	1.07
AAA_std.error	0.29	0.39	0.64	0.88
housng_estimate	-0.07	-0.03	0.04	0.22
housng_std.error	0.07	0.1	0.12	0.13
r.squared	0.83	0.75	0.59	0.51
adj.r.squared	0.82	0.73	0.55	0.47

Note:

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

F Appendix F

Notes: The following regressions includes YIV, dummy and controls as independent variables. For the description of the variables in the output, please refer to Appendix D. The specification for the regression is the following:

$$\frac{1}{H}\sum_{j=1}^{j=H} log(1+GDP_{i,t+j}) = \alpha_H + \beta_H \sigma_{IV,t}^{INT} + Controls + Dummy + \varepsilon_{t+H} \tag{26}$$

Table F1: 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

G Appendix G

The following figures display predicted vs actual values for 1) linear model, 2) Random Forest. For the estimations, the full model has been used, variables: YIV, dummy, log_gdp, TRM1003, TRM1006, TRM1012, TRM0506, AAA, DBAA, baa_aaa, DGS3MO, SRT03M.

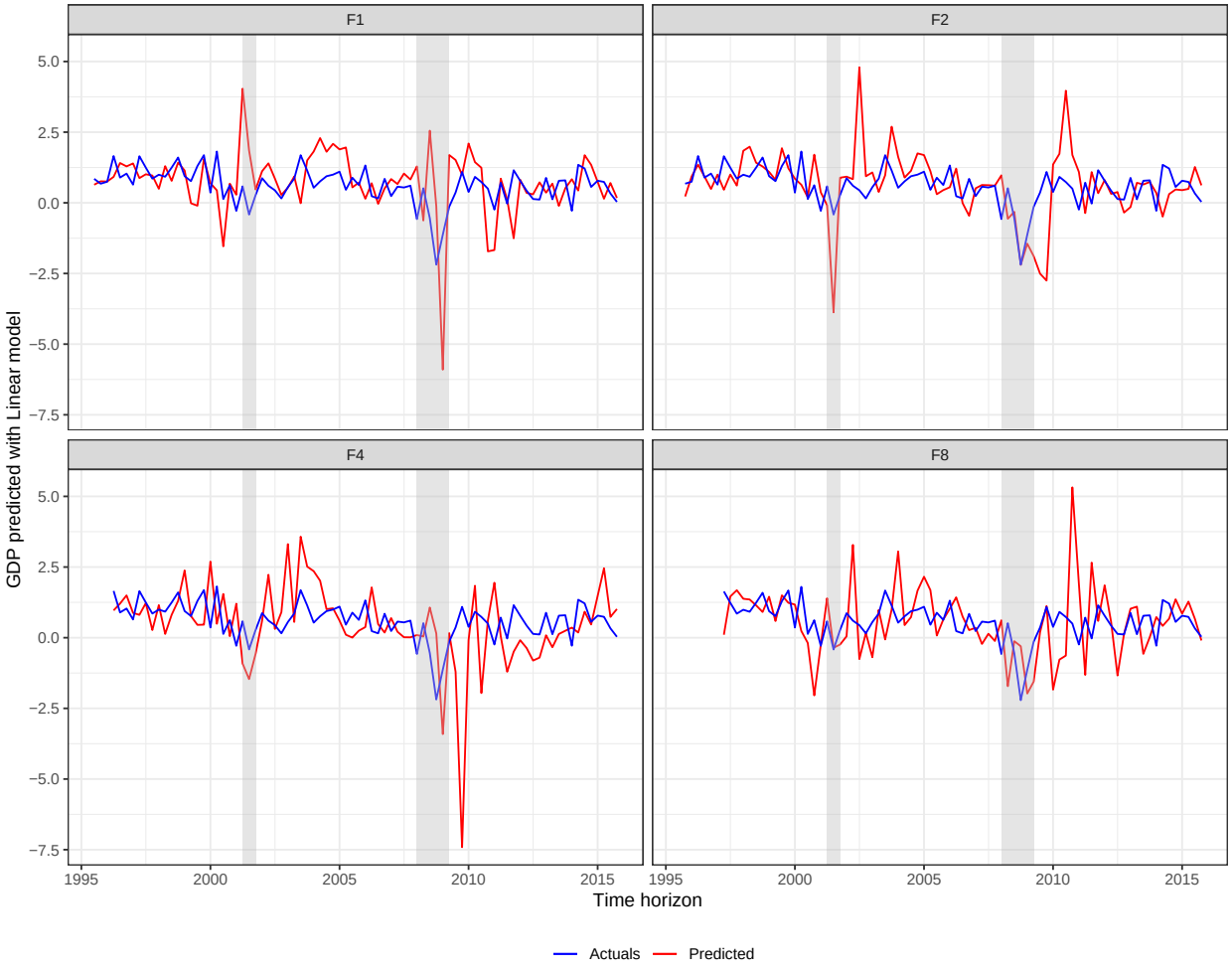


Figure G.1: Predicted vs actual results (Linear model)

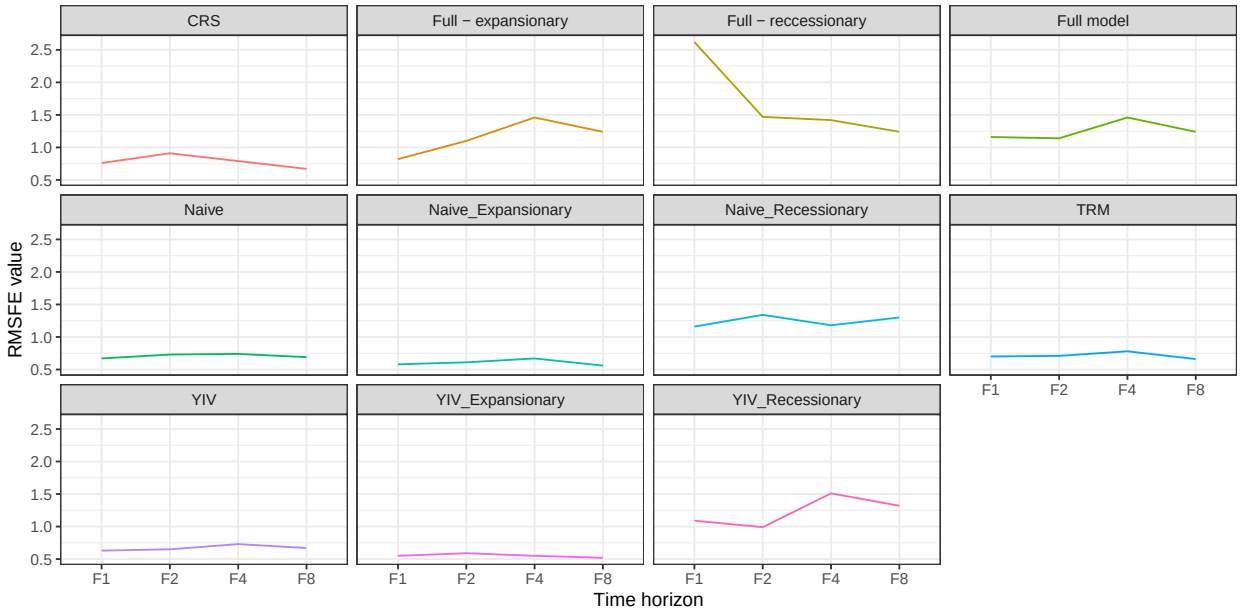
Figure G.2: Predicted vs actual results (Random Forest)

H Appendix H

The following table displays out-of-sample RMSFEs for various estimations. Naive is a panel with only log_gdp, TRM is a panel with all of the term spreads, CRS is a panel with credit spreads and corporate bond yields. Full model includes all of the variables: YIV, dummy, log_gdp, TRM1003, TRM1006, TRM1012, TRM0506, AAA, DBAA, baa_aaa, DGS3MO, SRT03M.

Table H1: Out-of-sample RMSFE for various estimations

	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	0.71	0.78	0.66
CRS	0.76	0.91	0.79	0.67
full	1.16	1.14	1.46	1.24
full_rec	2.62	1.47	1.42	1.24
full_exp	0.82	1.10	1.46	1.24



I Appendix I

The figure below displays RMSFEs for both linear model and Random Forest in 3 different panels: 1) with only expansionary observations, 2) with only recessionary observations as defined by NBER recessions and 3) the full-model.

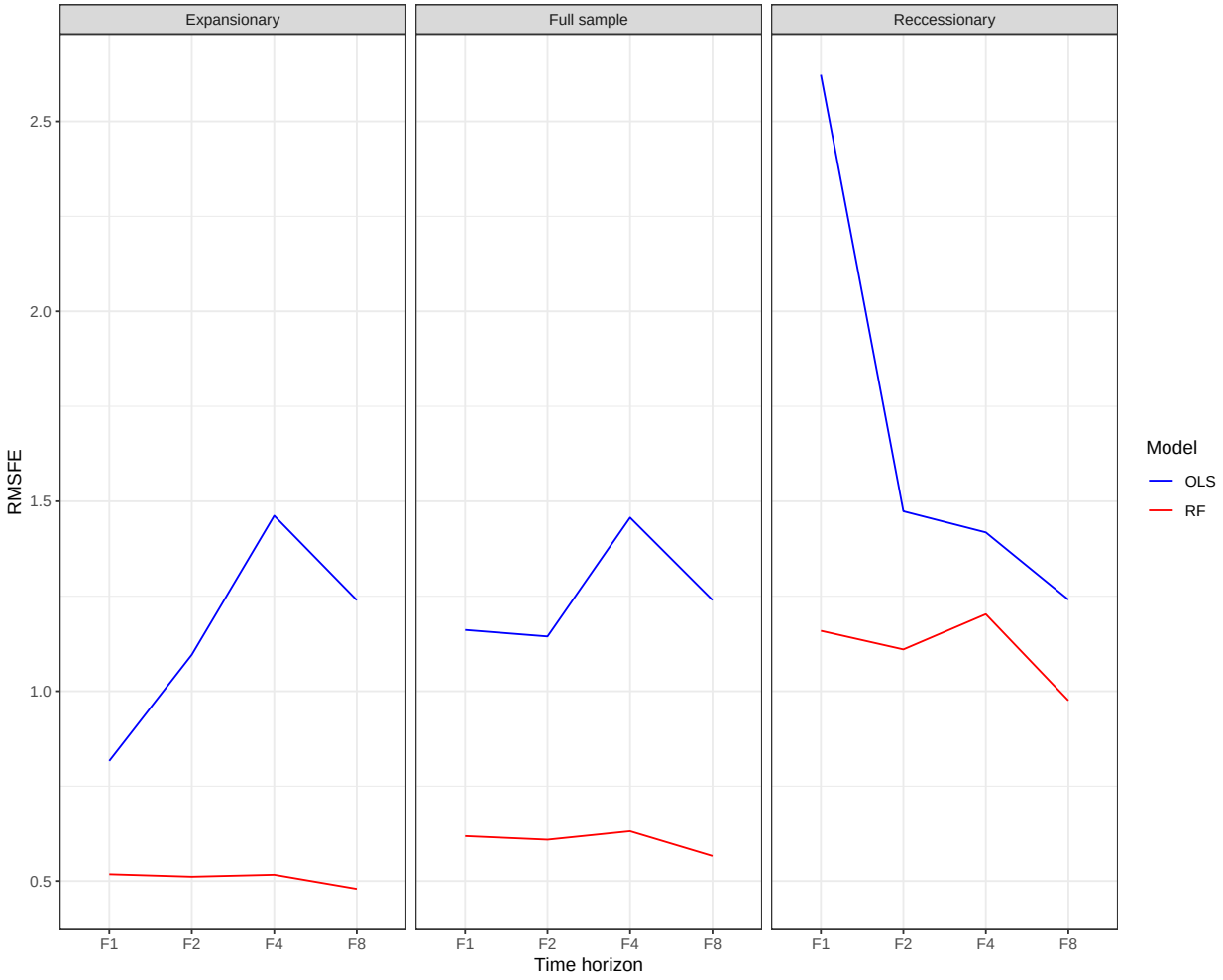


Figure I.1: RMSFE-s of linear model and Random Forest

J Appendix J

The figure below depicts squared errors for both models across all of the forecast horizons. Squared errors have been calculated as the squared differences of actual and predicted values. The variables are interpreted as follows: OOS_error_lm - squared errors for linear model; OOS_error_rf - squared errors for Random Forest.

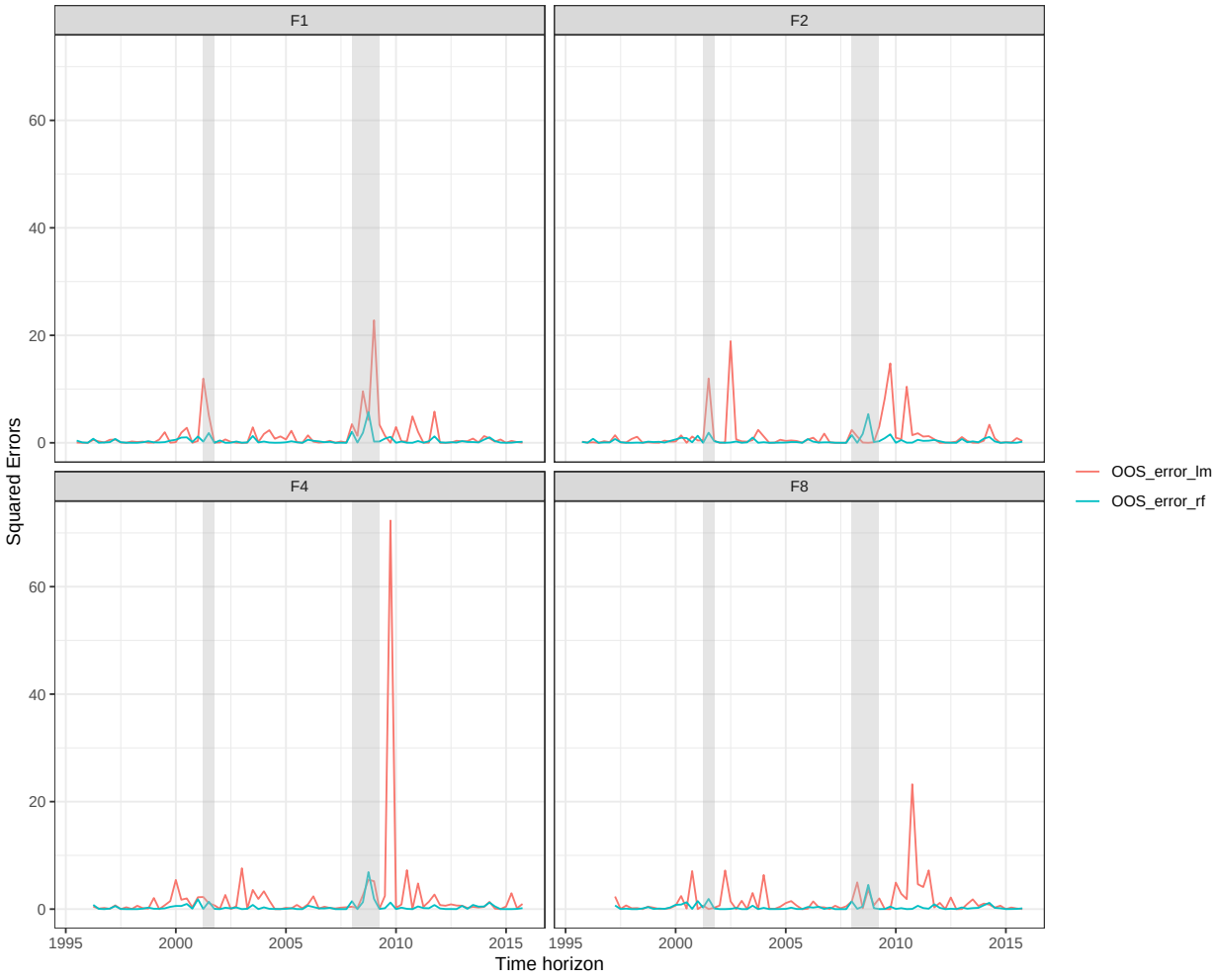


Figure J.1: Squared errors