Predicting the US Treasury Yields using Machine Learning Techniques

ERASMUS UNIVERSITY ROTTERDAM
Erasmus School of Economics

Joshua Oosterlaken (452550)

July 4, 2020

Abstract

In this paper we perform various machine learning techniques to forecast US treasury yields of various maturities for different forecast horizons over the period of March 1992 until July 2016. This research builds further upon the work of Swanson and Xiong (2018). We introduce the gradient boosted tree, elastic net and lasso as new machine learning methods applied to forecasting the yield curve. Additionally, we use the regression tree, random forest and Gaussian process regression methods. From all the machine learning techniques the Gaussion process regressions seems to produce the most accurate forecast. However, none of these machine learning methods is able to produce significantly more accurate forecasts than the ordinary econometric methods applied by Swanson and Xiong (2018).

Supvervisor: Zwan, T. van der
Second assessor: Dijk, D.J.C. van

Contents

1 Introduction **1**

2 Literature Review **2**

3 Data **3**

4 Methodology **3**

4.1 'Ordinary' Econometrics 5

4.1.1 AR and VAR models 5

4.1.2 Dynamic Nelson-Siegel models 5

4.1.3 Principal Component Analysis 6

4.2 Machine Learning 7

4.2.1 Regression Tree 8

4.2.2 Random Forest and Gradient Boosted Tree 8

4.2.3 Regularization 9

4.2.4 Gaussian Process 10

4.3 Forecast Evaluation 11

5 Results **15**

6 Conclusion **17**

1 Introduction

The modelling and forecasting of the term structure of interest rates is an important subject in economic research. Economists believe that the yield curve is a great predictor for economic recessions and has therefore significant influence on monetary policies. Furthermore, the forecasts of yields and interest rates is of great importance for investors and pension funds. Over the years this topic has been analyzed and researched thoroughly for this reason. Many researchers have tried to find a model that is able to accurately predict the yield curve. A popular model is that of Nelson and Siegel (1987). They introduced a three factor model that produces more accurate forecasts than a random walk model unlike most other models at that time. The Nelson-Siegel model and its extensions are therefore widely used in the financial sector. One of these extensions is the dynamic Nelson-Siegel model, which forecasts the latent factors with a autoregressive model. Swanson and Xiong (2018) compare the forecasting ability of the dynamic Nelson-Siegel model with that of simple autoregressive, vector autoregressive and principal components models. However, they state that there is still a lot of room for improvement left. Machine learning algorithms might be one of the possible improvements for predicting the yield curve. Therefore, this paper tries to investigate to what extend machine learning techniques are able to improve the forecasting of the term structure of interest rates.

The social relevance of this paper is the usefulness for investors and policy makers. According to many economists the inversion of the yield curve is a good predictor of recessions and GDP growth (see Haubrich, Dombrosky, et al., 1996; Chauvet and Potter, 2005; Ang, Piazzesi, and Wei, 2006). Therefore, policy makers adjust their monetary policies according to the development of the yield curve. When more accurate forecasts of these yield curves are made, policy makers are able to make these adjustments to the monetary policies earlier and may be able to prevent their country from falling into a great recession. Furthermore, investors are able to make better investments when the yield curve predictions are more accurate, since they can anticipate earlier on a possible recession.

The academic relevance lies in the fact that this paper is an extension to the research of Swanson and Xiong (2018). Their paper compares and extends the models from Nelson and Siegel (1987) and Diebold and Li (2006) with principal components models, autoregressive and vector autoregressive models. Hence, this paper builds even further upon the extensive history of forecasting the term structure of interest rates by applying different machine learning techniques to forecast the term structure of interest rates.

The objective of this paper is to forecast US treasury yields of five different maturities for a 1 month, 3 months and 12 months forecast horizon. We use a 120 months moving window and various models to construct these forecasts. The first part of the paper focuses on replicating the 'ordinary' econometric models as applied by Swanson and Xiong (2018). The first models are the autoregressive and vector autoregressive models with one lag and lags based on the Schwarz information criterion (Schwarz et al., 1978). The next models discussed are the dynamic Nelson-Siegel models as introduced by Diebold and Li (2006). The last 'ordinary' econometric models make use of principal component analysis.

The second part of the paper focuses on the machine learning techniques used for the prediction of the term structure of interest rates. The first machine learning technique we use is the Gaussian process regression as proposed by Sambasivan and Das (2017). Furthermore, we introduce a model where the coefficients are estimated through a lasso and a elastic net regression (see Tibshirani, 1996; Zou and Hastie, 2005). In addition we use a random forest model, since Rahimi (2020) finds that a random forest model (see Breiman, 2001) performs particularly well in the short end of the yield curve. We extend this by introducing a forecast of the yield curve by regression tree models and gradient boosted tree models (see (see Breiman, Friedman, Stone, and Olshen, 1984; Friedman, 2001). All models are evaluated based on their mean squared forecast error relative to that of a benchmark model. The benchmark model used in this research is the autoregressive model with one lag, the same model used as benchmark by Swanson and Xiong (2018). Significant difference in predictive accuracy is tested with the Diebold and Mariano (2002) test.

We find that the machine learning models are not able to make significant more accurate forecasts compared to the simple autoregressive model with one lag. Therefore, predicting the term structure of interest rates does not improve by applying machine learning techniques compared to more ordinary econometric methods.

2 Literature Review

The term structure of interest rates is the relationship between bond yields of different maturities. The yield curve is the graphical representation of this relationship. Estrella and Mishkin (1996) and Dueker (1997) find that the yield curve has significant better performance in predicting recessions than other financial and macroeconomic indicators. Not surprisingly, many economists tried, therefore, to model and forecast this yield curve. The first paper to investigate the modelling of the yield curve was published by Durand (1942). The popular three-factor Nelson-Siegel model for modelling the term structure of interest rates was introduced by Nelson and Siegel (1987). The three factors used in this model can be interpreted as a level, slope and curvature factor. Svensson (1994) extends this Nelson-Siegel model with an extra fourth factor which can be interpreted as an additional slope factor. The Bank of International Settlements reported in 2005 that nine out thirteen central banks applied either the Nelson-Siegel or the Svensson model to estimate their yield curve at that time (BIS,2005). Diebold and Li (2006) introduce the dynamic Nelson-Siegel model to improve the forecasting of the term structure for both short and long horizons. Diebold, Rudebusch, and Aruoba (2006) and Mönch (2008) show that models provide better predictions when macroeconomic variables are taken into consideration. A different approach to model the yield curve, applied by Litterman and Scheinkman (1991) and Knez, Litterman, and Scheinkman (1994), is based on principal component analysis. In recent years machine learning models have become more and more popular in economic modelling and forecasting. Sambasivan and Das (2017) find that a Gaussian process regression shows promising results for forecasting the yield curve for the medium and long term. Rahimi (2020) finds that a random forest model performs particularly well in the short end of the yield curve, mainly due to the ability of the random forest to correct for overfitting.

3 Data

This research builds further upon the work of Swanson and Xiong (2018) and uses the same data to make an improvement of their forecasts. Their monthly term structure data are the by the Federal Reserve Board provided end of the month US zero-coupon yield curve data [1] (Gürkaynak, Sack, and Wright, 2007). The data spans the period January 1982 through July 2016 consisting of 1- through 10-year maturities. Furthermore 103 macro-economic variables are obtained from the FRED-MD as given by the Federal Reserve Bank of St. Louis. McCracken and Ng (2016) provide the adjustments made to this data set[2]. These monthly macro-economic variables span the the same period as the term structure data. Table 1 reports the mean, standard deviation, minimum and maximum for the yield curve data. Furthermore, the table shows the skewness, kurtosis and Jarque-Bera test statistic, where normality is rejected for every maturity. Figure 1 shows the evolution of the yield curve over time.

4 Methodology

The models used to predict the US treasury yields consist of a moving window of 120 months. The 1-step-, 3-step- and 12-step-ahead forecasts are constructed for yields with five different

[1] Available at *quandl.com/data/FED/SVENY-US-Treasury-Zero-Coupon-Yield-Curve*
[2] Available at *research.stlouisfed.org/econ/mccracken/fred-databases*

Table 1: Descriptive statistics of yield data.

Maturity	Mean	Std	Min	Max	Skew	Kurt	JB
1 year	4.508	3.282	0.099	14.009	0.366	2.527	13.152
2 years	4.805	3.310	0.188	14.128	0.386	2.546	13.867
3 years	5.054	3.267	0.306	14.075	0.417	2.580	15.105
4 years	5.271	3.200	0.454	13.978	0.450	2.617	16.540
5 years	5.462	3.128	0.627	13.881	0.478	2.653	17.916
6 years	5.630	3.057	0.815	13.797	0.501	2.686	19.102
7 years	5.778	2.991	1.007	13.729	0.519	2.715	20.055
8 years	5.908	2.931	1.197	13.674	0.533	2.741	20.789
9 years	6.021	2.879	1.380	13.646	0.543	2.764	21.349
10 years	6.120	2.833	1.498	13.651	0.551	2.784	21.787

This table reports the mean, standard deviation, minimum, maximum, skewness, kurtosis and the Jarque-Bera test statistic for the monthly US zero-coupon yield for different maturities over the period January 1982 through July 2016.

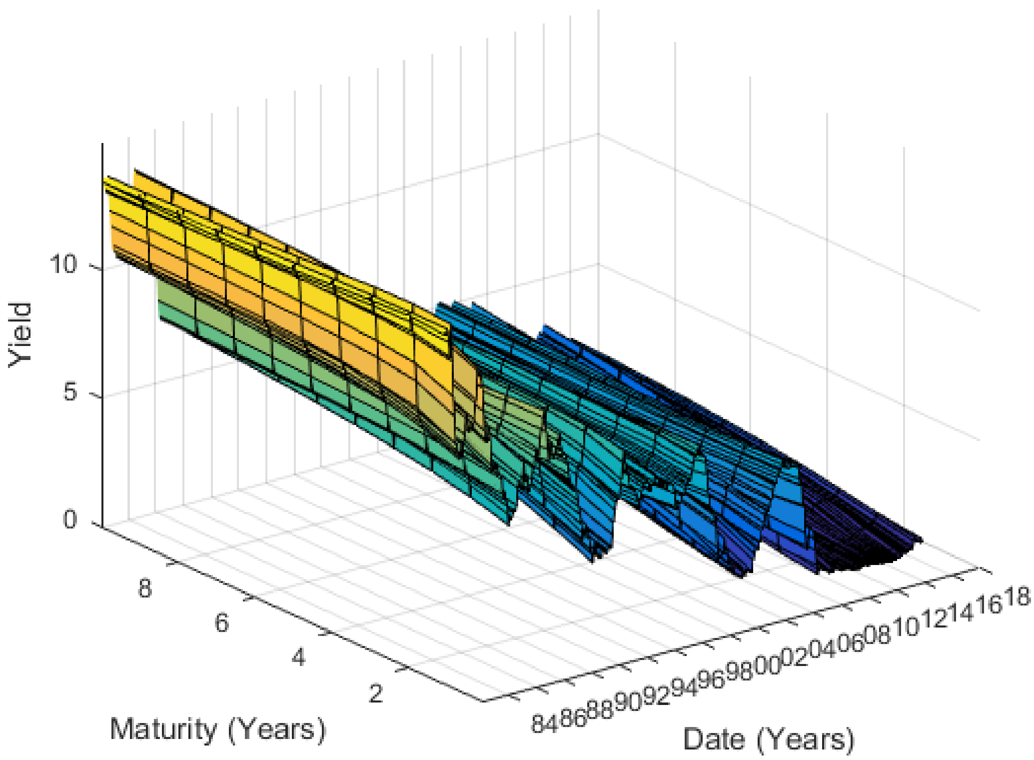

Figure 1: Yield curve over time in monthly frequency.

maturities: 1 year, 2 years, 3 years, 5 years and 10 years. Since this paper tries to compare the machine learning models with the 'ordinary' econometric models as applied by Swanson and Xiong (2018), the same four out-of-sample periods are used as in their paper: 1992:3-1999:12, 2000:1-2007:12, 2008:1-2016:7, 1992:3-2016:7. First the 'ordinary' econometric models are further elaborated on. Furthermore, the machine learning techniques are discussed as well as the forecast evaluation.

4.1 'Ordinary' Econometrics

In this section the methods applied by Swanson and Xiong (2018) are explained[3]. First the autoregressive (AR) and vector autoregressive (VAR) models are summarized. Secondly, the dynamic Nelson-Siegel (DNS) models as introduced by Diebold and Li (2006) are discussed. This section ends with an explanation of the principal component analysis (PCA) based models.

4.1.1 AR and VAR models

First we show the AR model, which is given as follows:

$$y_{t+h}(\tau) = c + \beta' \boldsymbol{W}_t + \epsilon_{t+h}, \tag{1}$$

where $y_{t+h}(\tau)$ is the yield,$\tau \in \{1,2,3,5,10\}$ and denotes the maturity of a bond in years, c is a constant, $\boldsymbol{\beta}$ is the coefficient vector, $\boldsymbol{W}_t$ contains the lags of y_{t+h} and ϵ_{t+h} denotes the error term. The VAR model is given as follows:

$$\boldsymbol{y}_{t+h} = \boldsymbol{c} + \boldsymbol{B}\boldsymbol{W}_t + \boldsymbol{\epsilon}_{t+h}, \tag{2}$$

where $\boldsymbol{y}_{t+1}$ is a vector consisting of the yield with the five different maturities, $\boldsymbol{c}$ is a vector of constants, $\boldsymbol{B}$ is coefficient matrix, $\boldsymbol{W}_t$ is matrix of the lags of $\boldsymbol{y}_{t+h}$ and $\boldsymbol{\epsilon}_{t+h}$ is a vector of error terms. Both an AR and a VAR model with a lag of 1 and both an AR and a VAR model with at most 5 lags, where we choose the lags based on the Schwarz information criterion (SIC), are considered (Schwarz et al., 1978).

4.1.2 Dynamic Nelson-Siegel models

First we define the following three vectors of yields for different maturities at time t:
$\boldsymbol{y}_t^4(\tau) = [y_t(12)\ y_t(36)\ y_t(60)\ y_t(120)]$, $\boldsymbol{y}_t^6(\tau) = [y_t(12)\ y_t(24)\ y_t(36)\ y_t(60)\ y_t(84)\ y_t(120)]$ and
$\boldsymbol{y}_t^{10}(\tau) = [y_t(12)\ y_t(24)\ y_t(36)\ y_t(48)\ y_t(60)\ y_t(72)\ y_t(84)\ y_t(96)\ y_t(108)\ y_t(120)]$. Furthermore, the Nelson-Siegel factors are obtained by running the following regression at every period t:

$$\boldsymbol{y}_t(\tau) = \beta_{1,t} + \beta_{2,t}\left[\frac{1-\exp(-\lambda_t\tau)}{\lambda_t\tau}\right] + \beta_{3,t}\left[\frac{1-\exp(-\lambda_t\tau)}{\lambda_t\tau} - \exp(-\lambda_t\tau)\right], \tag{3}$$

where $\boldsymbol{y}_t(\tau) \in \{\boldsymbol{y}_t^4(\tau), \boldsymbol{y}_t^6(\tau), \boldsymbol{y}_t^{10}(\tau)\}$ and λ_t, the rate of decay, is set at 0.0609 as suggested by Diebold and Li (2006). Furthermore they explain that since $\beta_{1,t}$ is constant and can therefore be seen as the 'level' factor, that an increase of the maturity decreases $\beta_{2,t}$ and can therefore

[3] For this section we use the code provided by Swanson: *http://econweb.rutgers.edu/nswanson/comp.htm*

be interpreted as the short-term 'slope' factor and that $\beta_{3,t}$ closely resembles the 'curvature' factor. By running this regression for every point t in the moving window period a time series of estimated betas is created. These betas are used in the following AR regression to make a h-step-ahead forecast:

$$\hat{\beta}_{i,t+h} = c + \gamma_{ii}\hat{\beta}_{i,t}, \qquad \text{for } i = 1,2,3. \tag{4}$$

The betas are forecasted using a VAR(1) model as well:

$$\hat{\boldsymbol{\beta}}_{t+h} = \boldsymbol{c} + \boldsymbol{\gamma}\hat{\boldsymbol{\beta}}_t, \tag{5}$$

where $\hat{\boldsymbol{\beta}}_{t+h} = (\hat{\beta}_{1,t+h}, \hat{\beta}_{2,t+h}, \hat{\beta}_{3,t+h})', \boldsymbol{c}$ is a vector of constants, $\boldsymbol{\gamma} = (\boldsymbol{\gamma_1}, \boldsymbol{\gamma_2}, \boldsymbol{\gamma_3})$, with $\boldsymbol{\gamma_j}$ a vector of coefficients, for $j = 1,2,3$. Additionally, the betas are forecasted using extended versions of these AR and VAR models. Three key macroeconomic variables proposed by Diebold et al. (2006) are added to the models. The extended version of equation (4) is as follows:

$$\hat{\beta}_{i,t+h} = c + \gamma_{ii}\hat{\beta}_{i,t} + \boldsymbol{\alpha}_i'\boldsymbol{M}_t, \qquad \text{for } i = 1,2,3, \tag{6}$$

where $\boldsymbol{\alpha}_i$ is a vector of coefficients and $\boldsymbol{M}_t$ includes the variables annual personal consumption expenditures price deflator, federal funds rate and manufacturing capacity utilization. In addition the extended version of equation (5) becomes:

$$\hat{\boldsymbol{\beta}}_{t+h} = \boldsymbol{c} + \boldsymbol{\gamma}\hat{\boldsymbol{\beta}}_t + \boldsymbol{\alpha}M_t, \tag{7}$$

where $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3)$. Inserting these forecasted betas in the following formula gives the forecasted yield per maturity:

$$\hat{y}_{t+h}(\tau) = \hat{\beta}_{1,t+h} + \hat{\beta}_{2,t+h}\left[\frac{1-\exp(-\lambda_t\tau)}{\lambda_t\tau}\right] + \hat{\beta}_{3,t+1}\left[\frac{1-\exp(-\lambda_t\tau)}{\lambda_t\tau} - \exp(-\lambda_t\tau)\right]. \tag{8}$$

Estimating the betas for the in-sample period using three different yield vectors and forecasting the betas using the two AR models and two VAR models results in twelve different models for forecasting the term structure.

4.1.3 Principal Component Analysis

For the principal component methods again the framework from Swanson and Xiong (2018) is applied. First the set yields with the ten different maturities are standardized before constructing the principal components of this set of yields using the following formula:

$$\boldsymbol{F}_i^{Yield} = \boldsymbol{Y}\boldsymbol{e}_i^{Yield}, \qquad \text{for } i = 1, 2, 3. \tag{9}$$

Here $\boldsymbol{F}_i^{Yield}$ and $\boldsymbol{e}_i^{Yield}$ denote respectively the i^{th} principal component and the i^{th} eigenvector corresponding to the i^{th} largest eigenvalue of of the covariance matrix of $\boldsymbol{Y}$ and $\boldsymbol{Y}$ is the matrix of the set of yields with the ten different maturities. We calculate these eigenvectors using the spectral decomposition of the covariance matrix of $\boldsymbol{Y}$. These principal components are added to the simple AR model in equation (1) , resulting in the following equation:

$$y_{t+h}(\tau) = c + \beta' W_t + \boldsymbol{\alpha}' \boldsymbol{F}_t^{Yield} + \epsilon_{t+h}, \tag{10}$$

where W_t includes one lag and $\boldsymbol{F}_t^{Yield}$ includes either the first, the first and the second or the first, second and third principal components. This results in three different PCA models. In addition, the principal components of the 103 macroeconomic variables are calculated as follows:

$$\boldsymbol{F}_i^{MACRO} = \boldsymbol{X}\boldsymbol{e}_i^{MACRO}, \qquad \text{for } i = 1, 2, 3. \tag{11}$$

Here $\boldsymbol{F}_i^{MACRO}$ and $\boldsymbol{e}_i^{MACRO}$ denote respectively the i^{th} principal component and the i^{th} eigenvector corresponding to the i^{th} largest eigenvalue of of the covariance matrix of $\boldsymbol{X}$ and $\boldsymbol{X}$ is the matrix of the 103 macroeconomic variables. These principal components are added to the simple AR model with one lag as well which gives equation (10) only with $\boldsymbol{F}_t^{Yield}$ replaced by $\boldsymbol{F}_t^{MACRO}$ and therefore resulting in three different models as well. Furthermore, both the first principal component and the first and second principal component of the 103 macroeconomic variables are added to the two VAR models, the six DNS models without the three macroeconomic variables and the three models with the principal components of the yields, resulting in an additional 22 models.

4.2 Machine Learning

In the following section we explain the machine learning models used in this paper. All models use three different sets of explanatory variables. The first set consists of the yields with the 10 different maturities, the second set consists of the three key macroeconomic variables proposed by Diebold et al. (2006) and the third set is the combination of the first two sets.

4.2.1 Regression Tree

The first machine learning method being considered is the regression tree model as introduced by Breiman et al. (1984). A regression tree consists of decision nodes and leaves. In every decision node a input variable is tested against a certain threshold. One goes, starting from the top of the tree, down on a different branch depending on the outcome. At every node visited this process is repeated and ends at a leaf at the bottom of the tree. This leaf provides the predicted value of the target variable. The regression tree itself is constructed by, starting with the whole predictive space, top-down splitting this predictive space in two different regions at each node. These regions are chosen such that the residual sum of squares (RSS) is minimized, where the RSS is defined in the following way:

$$RSS = \sum_{t:x_t \in R_1(j,s)} (y_t - \hat{y}_{R_1})^2 + \sum_{t:x_t \in R_2(j,s)} (y_t - \hat{y}_{R_2})^2, \tag{12}$$

where $R_1(j,s) = \{X|X_j \leq s\}$, $R_2(j,s) = \{X|X_j > s\}$ and $\hat{y}_{R_i}$ is mean response for the training set of the i^{th} region. The input variables considered at each node are chosen through minimizing the RSS as well. An example of a regression tree is given in Figure 2. For the creation of the regression tree models the RegressionTree class available via Matlab is used.

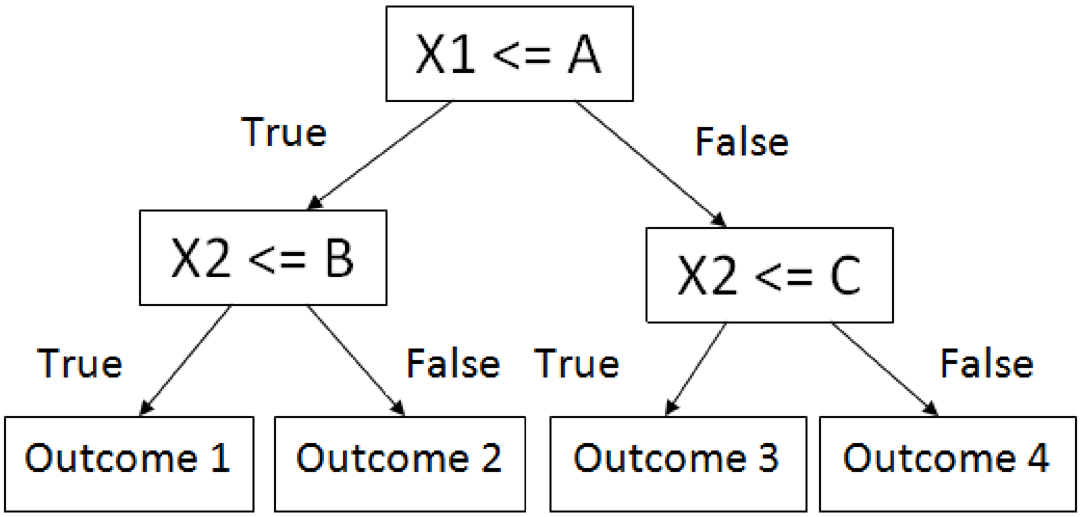

Figure 2: Regression tree example. When, for example, X1 is smaller than threshold A, you continou to the decision node to the left and if then X2 is larger than B you end up at the leaf with Outcome 2

4.2.2 Random Forest and Gradient Boosted Tree

The next machine learning method is the random forest as proposed by Breiman (2001). Random Forest is an ensemble technique that uses bootstrap aggregation. For the random forest a 100 regression trees are constructed by training them on a random sample of the input observations. Furthermore, these samples are drawn with replacement, such that certain data points can be chosen multiple times in a single regression tree. In addition it randomly selects a sample of variables to evaluate at every node. The prediction provided by the random forest is the average of the results of the 100 single regression trees. The random forest is performed using the RegressionEnsemble class in Matlab.

Additionally, we apply the gradient boosted tree from Friedman (2001). The gradient boosted tree is a ensemble technique as well. It builds the regression trees one at a time, while using the errors of the previous constructed regression tree. The errors are calculated using the least squares criteria. The algorithm for the gradient boosted tree is given in Algorithm 1. This model is constructed via the RegressionEnsemble class in Matlab.

Algorithm 1: Gradient boosted tree

1 $F_0(x) = \bar{y}$
for $m = 1$ to 100 **do**
  (2) $\tilde{y}_i = y_i - F_{m-1}(\boldsymbol{x}_i)$, for $i = 1, 2, \ldots, N$
  (3)$(\rho_m, \boldsymbol{a}_m) = \arg\min_{\boldsymbol{a},\rho} \sum_{i=1}^{N} [\tilde{y}_i - \rho h(\boldsymbol{x}_i; \boldsymbol{a})]^2$
  (4) $F_m(\boldsymbol{x}) = F_{m-1}(\boldsymbol{x} + \rho_m h(\boldsymbol{x}; \boldsymbol{a}_m)$
end for
Result $F_M(x)$;

4.2.3 Regularization

The next machine learning methods make use of regularization. They extend the standard least squares function with an extra penalty term. The first regularization method applied is the least absolute shrinkage and selection operator (lasso)(Tibshirani, 1996). This method performs both variable selection and shrinkage by adding the sum of absolute values of the regression coefficients as a penalty to the least squares loss function. The objective function of the lasso is given by the following:

$$\min_{c,\boldsymbol{\beta}} \left(\frac{1}{2N} \sum_{i=1}^{N} (y_i - c - \boldsymbol{x}_i' \boldsymbol{\beta})^2 + \lambda \sum_{j=1}^{P} |\beta_j| \right), \tag{13}$$

where λ is the weight given to the penalty term. The lasso function is solved through the coordinate descent algorithm (Friedman, Hastie, and Tibshirani, 2010). λ is chosen as the largest value of λ such that the mean squared error (MSE) is minimized. The lasso model is estimated using the lasso function in Matlab. The second regularization method is the elastic net (enet) (Zou and Hastie, 2005). The enet extends equation (13) with a ridge penalty term.

Therefore, the enet combines both the strong variable selection property from the lasso and the strong shrinkage property from the ridge. The enet objective function is given by the following:

$$\min_{c,\boldsymbol{\beta}} \left(\frac{1}{2N} \sum_{i=1}^{N} (y_i - c - \boldsymbol{x}_i' \boldsymbol{\beta})^2 + \lambda \sum_{j=1}^{P} \left(\alpha |\beta_j| + \frac{(1-\alpha)}{2} \boldsymbol{\beta}_j^2 \right) \right), \tag{14}$$

where λ is given by the same specification as in the lasso function and α is set to 0.5 such that the ridge penalty and the lasso penalty term get equal weight. The objective function is solved through the coordinate descent algorithm as well. The Matlab function lasso is used for the enet model.

4.2.4 Gaussian Process

In this section the Gaussian process as suggested by Williams and Rasmussen (2006) is further elaborated on. Consider the following standard regression model:

$$y = f(\boldsymbol{x}) + \epsilon, \tag{15}$$

where y is the observed value, f is the regression function, $\boldsymbol{x}$ is a vector of inputs and $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$. From equation (15) follows that:

$$\boldsymbol{y} \sim \mathcal{N}(\mu(X), K + \sigma_\epsilon^2 I), \tag{16}$$

where $\mu(X)$ and $K = k(X,X)$ are respectively the expected value and the covariance function (kernel) of $f(X)$. The joint distribution of the observed data $\boldsymbol{y}$ and the regression function $f(X_*)$, where X_* is the out of sample data, is given by the following:

$$\begin{bmatrix} \boldsymbol{y} \\ f(X_*) \end{bmatrix} = \mathcal{N}\left(\begin{bmatrix} \mu(X) \\ \mu(X_*) \end{bmatrix}, \begin{bmatrix} K + \sigma_\epsilon^2 I & K_* \\ K_*' & K_{**} \end{bmatrix} \right), \tag{17}$$

where $K_* = k(X, X_*)$ and $K_{**} = k(X_*, X_*)$. Williams and Rasmussen (2006) show the following:

$$f(X_*)|X, \boldsymbol{y}, X_* \sim \mathcal{N}(\boldsymbol{\mu}_*, \boldsymbol{\sigma}_*^2), \tag{18}$$

where

$$\boldsymbol{\mu}_* = \mu(X_*) + K_*'(K + \sigma_\epsilon^2 I)^{-1}(\boldsymbol{y} - \mu(X)), \tag{19}$$

and

$$\boldsymbol{\sigma}_*^2 = K_{**} - K_*'(K + \sigma_\epsilon^2 I)^{-1} K_*. \tag{20}$$

After every prediction the model is trained again using the 120 month moving window. Thus, there is only one observation x_* and we can rewrite equation (19) into:

$$\boldsymbol{\mu}_* = \mu(\boldsymbol{x}_*) + \boldsymbol{k}_*'(K + \sigma_\epsilon^2 I)^{-1}(\boldsymbol{y} - \mu(\boldsymbol{x}_*)), \tag{21}$$

where $\boldsymbol{k}_*$ is a vector of covariances between the forecast observation x_* and the n training points $\boldsymbol{x}$ such that:

$$\boldsymbol{k}_* = [k(\boldsymbol{x}_1, \boldsymbol{x}_*, \ldots, k(\boldsymbol{x}_n, \boldsymbol{x}_*)]'. \tag{22}$$

Therefore equation (21) can be written as a linear combination of kernel functions:

$$\boldsymbol{\mu}_* = \mu(\boldsymbol{x}_*) + \boldsymbol{k}_*'\boldsymbol{\alpha} = \mu(\boldsymbol{x}_*) + \sum_{i=1}^{n} \alpha_i k(\boldsymbol{x}_i, \boldsymbol{x}_*). \tag{23}$$

Sambasivan and Das (2017) find that a combination of a linear and squared exponential kernel produce good results. Therefore, the following kernel is used in the Gaussian process regression:

$$k(\boldsymbol{x}, \boldsymbol{x}_*) = \sigma_f^2 \left(\boldsymbol{x}'\boldsymbol{x}_* + \exp\left(-\frac{(\boldsymbol{x} - \boldsymbol{x}_*)^2}{2\ell^2} \right) \right). \tag{24}$$

This Gaussian process regression now has three hyperparameters that have to be estimated: σ_f^2, ℓ^2 and σ_ϵ^2, where ℓ is the length scale that defines how smooth the kernel function is. These hyperparameters can be obtained by maximizing the following log marginal likelihood function (Williams and Rasmussen, 2006):

$$\log\, p(\boldsymbol{y}|X, \sigma_f^2, \ell, \sigma_\epsilon^2) = -\frac{1}{2}\boldsymbol{y}'(K + \sigma_\epsilon^2 I)\boldsymbol{y} + \frac{1}{2}\log(|K + \sigma_\epsilon^2 I|) + \frac{n}{2}\log(2\pi), \tag{25}$$

where n is the length of the input data, which is 120. For the training of the Gaussian process regression the Matlab class RegressionGP is used.

4.3 Forecast Evaluation

The benchmark model is the same as in the research of Swanson and Xiong (2018): the simple AR model with one lag. For every model the mean square forecast error (MSFE) over the out-of-sample period is calculated:

$$MSFE_i = \frac{\sum(y_t - \hat{y}_{i,t})^2}{L}, \tag{26}$$

where $MSFE_i$ is the MSFE for model i and L is the number of test sample observations. The MSFEs are used to calculate the relative mean squared forecast errors (rMSFE) using the following formula:

$$rMSFE_i = \frac{MSFE_i}{MSFE_{Benchmark}}, \tag{27}$$

where $rMSFE_i$ is the rMSFE for model i and $MSFE_{Benchmark}$ is the MSFE of the simple AR model. The Diebold-Mariano (DM) test is used to test whether a model has a significant different MSFE than the benchmark model (Diebold and Mariano, 2002). The DM test statistic is given as follows:

$$DM = \frac{\bar{d}_t}{\hat{\sigma}_{\bar{d}_t}}, \tag{28}$$

where $\bar{d}_t = \frac{1}{L}\sum d_t$, $d_t = (y_t - \hat{y}_{1,t})^2 - (y_t - \hat{y}_{2,t})^2$, $\hat{\sigma}_{\bar{d}_t} = \frac{\hat{\sigma}_{d_t}}{\sqrt{L}}$ and $DM \sim N(0,1)$

TABLE 2: Models used to forecast the yield curve	
Model	Description
AR(1)	Autoregressive model with one lag
VAR(1)	Five-dimensional vector autoregressive model with one lag
VAR(1) + FB1	VAR(1) model with one principle components added, principle component analysis based on all 103 macroeconomic variables
VAR(1) + FB2	VAR(1) model with two principle components added, principle component analysis based on all 103 macroeconomic variables
AR(SIC)	Autoregressive model with lag(s) selected by the Schwarz information criterion
VAR(SIC)	Five-dimensional vector autoregressive model with lag(s) selected by the Schwarz information criterion
VAR(SIC) + FB1	VAR(SIC) model with one principle components added, principle component analysis based on all 103 macroeconomic variables
VAR(SIC) + FB2	VAR(SIC) model with two principle components added, principle component analysis based on all 103 macroeconomic variables
DNS(1)	Dynamic Nelson–Siegel (DNS) model with underlying AR(1) factor specifications fitted with 10-dimensional yields: maturity $\tau = 12, 24, 36, 48, 60, 72, 84, 96, 108, 120 month$
DNS(2)	DNS model with underlying AR(1) factor specifications fitted with six-dimensional yields: maturity $\tau = 12, 24, 36, 60, 84, 120 months$
DNS(3)	DNS model with underlying AR(1) factor specifications fitted with four-dimensional yields: maturity $\tau = 12, 36, 60, 120 months$
DNS(4)	DNS model with underlying VAR(1) factor specifications fitted with 10-dimensional yields: maturity $\tau = 12, 24, 36, 48, 60, 72, 84, 96, 108,$ 120 months
DNS(5)	DNS model with underlying VAR(1) factor specifications fitted with six-dimensional yields: maturity $\tau = 12, 24, 36, 60, 84, 120 months$
DNS(6)	DNS model with underlying VAR(1) factor specifications fitted with four-dimensional yields: maturity $\tau = 12, 36, 60, 120 months$
DNS(1) + FB1	DNS(1) model with one principle component added, principle component analysis based on all 103 macroeconomic variables
DNS(2) + FB1	DNS(2) model with two principle component added, principle component analysis based on all 103 macroeconomic variables
DNS(3) + FB1	DNS(3) model with one principle component added, principle component analysis based on all 103 macroeconomic variables
DNS(4) + FB1	DNS(4) model with one principle component added, principle component analysis based on all 103 macroeconomic variables
DNS(5) + FB1	DNS(5) model with one principle component added, principle component analysis based on all 103 macroeconomic variables
DNS(6) + FB1	DNS(6) model with one principle component added, principle component analysis based on all 103 macroeconomic variables
DNS(1) + FB2	DNS(1) model with two principle component added, principle component analysis based on all 103 macroeconomic variables
DNS(2) + FB2	DNS(2) model with two principle component added, principle component analysis based on all 103 macroeconomic variables
DNS(3) + FB2	DNS(3) model with two principle component added, principle component analysis based on all 103 macroeconomic variables
DNS(4) + FB2	DNS(4) model with two principle component added, principle component analysis based on all 103 macroeconomic variables

TABLE 2 (Continued)

Model	Description
DNS(5) + FB2	DNS(5) model with two principle component added, principle component analysis based on all 103 macroeconomic variables
DNS(6) + FB2	DNS(6) model with two principle component added, principle component analysis based on all 103 macroeconomic variables
DNS(1)+MAC	DNS(1) model with three key macroeconomic variables added: manufacturing capacity utilization, the federal funds rate and annual price inflation
DNS(2)+MAC	DNS(2) model with three key macroeconomic variables added: manufacturing capacity utilization, the federal funds rate and annual price inflation
DNS(3)+MAC	DNS(3) model with three key macroeconomic variables added: manufacturing capacity utilization, the federal funds rate and annual price inflation
DNS(4)+MAC	DNS(4) model with three key macroeconomic variables added: manufacturing capacity utilization, the federal funds rate and annual price inflation
DNS(5)+MAC	DNS(5) model with three key macroeconomic variables added: manufacturing capacity utilization, the federal funds rate and annual price inflation
DNS(6)+MAC	DNS(6) model with three key macroeconomic variables added: manufacturing capacity utilization, the federal funds rate and annual price inflation
DIF(1)	Diffusion index model with one principle component estimator based on all 10-dimensional yields
DIF(2)	Diffusion index model with two principle component estimator based on all 10-dimensional yields
DIF(3)	Diffusion index model with three principle component estimator based on all 10-dimensional yields
DIF(4)	Diffusion index model with one principle component estimator based on all 103 macroeconomic variables
DIF(5)	Diffusion index model with two principle component estimator based on all 103 macroeconomic variables
DIF(6)	Diffusion index model with three principle component estimator based on all 103 macroeconomic variables
DIF(1)+FB1	DIF(1) model with one principle component added, principle component analysis based on all 103 macroeconomic variables
DIF(2)+FB1	DIF(2) model with one principle component added, principle component analysis based on all 103 macroeconomic variables
DIF(3)+FB1	DIF(3) model with one principle component added, principle component analysis based on all 103 macroeconomic variables
DIF(1)+FB2	DIF(1) model with two principle components added, principle component analysis based on all 103 macroeconomic variables
DIF(2)+FB2	DIF(2) model with two principle components added, principle component analysis based on all 103 macroeconomic variables
DIF(3)+FB2	DIF(3) model with two principle components added, principle component analysis based on all 103 macroeconomic variables
TRE(1)	Regression tree model with all 10-dimensional yields as input variables

TABLE 2 (Continued)

Model	Description
TRE(2)	Regression tree model with the three key macroeconomic variables as input variables
TRE(3)	Regression tree model with all 10-dimensional yields and the three key macroeconomic variables as input variables
FOR(1)	Random forest model with all 10-dimensional yields as input variables
FOR(2)	Random forest model with the three key macroeconomic variables as input variables
FOR(3)	Random forest model with all 10-dimensional yields and the three key macroeconomic variables as input variables
GBT(1)	Gradient boosted tree model with all 10-dimensional yields as input variables
GBT(2)	Gradient boosted tree model with the three key macroeconomic variables as input variables
GBT(3)	Gradient boosted tree model with all 10-dimensional yields and the three key macroeconomic variables as input variables
LAS(1)	Lasso regression model with all 10-dimensional yields as input variables
LAS(2)	Lasso regression model with the three key macroeconomic variables as input variables
LAS(3)	Lasso regression model with all 10-dimensional yields and the three key macroeconomic variables as input variables
ENE(1)	Elastic net regression model with all 10-dimensional yields as input variables
ENE(2)	Elastic net regression model with the 3 key macroeconomic variables as input variables
ENE(3)	Elastic net regression model with all 10-dimensional yields and the three key macroeconomic variables as input variables
GPR(1)	Gaussian process regression model with all 10-dimensional yields as input variables
GPR(2)	Gaussian process regression model with the three key macroeconomic variables as input variables
GPR(3)	Gaussian process regression model with all 10-dimensional yields and the three key macroeconomic variables as input variables

5 Results

Tables 3-6 report the relative MSFEs for the term structure of interest rates forecasts of the models from Table 2 for a 1 month forecast horizon for the 1-, 2-, 3-, 5- and 10-years maturities. The four different forecast subsamples are: 1992:3-1999:12, 2000:1-2007:12, 2008:1-2016:7 and 1992:3-2016:7. The AR(1) model is used as a benchmark for the construction of the relative MSFEs. Rejection of the null hypothesis of equal predictive accuracy is indicated by asterisks after the rMSFEs. The best performing models during the first two subsamples given in Table 3 and 4 are the DNS+FB type models. However, these models are not able to produce significant more accurate forecasts during the most recent subsample as given in Table 5. Nevertheless, when taking the whole prediction period into consideration, the DNS+FB type models do provide lower MSFEs than the AR(1) model and for the 1 year maturity they are able to perform significantly better. When focusing on the machine learning models, it stands out that in none of the subsamples they give significant more accurate forecasts than the simple AR(1) model and that only the GPR(1) and GPR(3) models have rMSFEs lower than 1 for subsamples 2, 3 and 4 for some of the maturities. Furthermore, the addition of the three key macroeconomic variables doesn't seem to produce lower rMSFEs in contrast to the models without those variables, whereas the models with only the three key macroeconomic variables are the worst performing models.

Tables 7-10 report the relative MSFEs for the term structure of interest rates forecasts of the models from Table 2 for a 3 months forecast horizon for the 1-, 2-, 3-, 5- and 10-years maturities. The four different forecast subsamples are: 1992:7-1999:12, 2000:1-2007:12, 2008:1-2016:7 and 1992:7-2016:7. The results for the three step forecast horizon are pretty similar to that of the one step forecast horizon. Again the DNS + FB type models produce the best results, except for the most recent subsample given in Table 9. In this subsample the AR model where the lags are chosen by the SIC gives the most accurate predictions. This model performs well across all subsamples for the three step forecast horizon. The machine learning methods, on the other hand, all have rMSFEs above one, except the GPR(3) model for the one year maturity in subsample 2 with an rMSFE of 0.974 given in Table 8. Again the addition of the three key macroeconomic variables does not seem to be an improvement compared to the models without those three key macroeconomic variables.

Tables 11-14 report the relative MSFEs for the term structure of interest rates forecasts of the models from Table 2 for a 12 months forecast horizon for the 1-, 2-, 3-, 5- and 10-years maturities. The four different forecast subsamples are: 1994:1-1999:12, 2000:1-2007:12, 2008:1-2016:7 and 1994:1-2016:7. For the 12 step forecast horizon the best performing models are the (V)AR type models and the DNS type models. This is different from the one month and three months forecast horizon, where clearly the best models where the DNS + FB type models. The machine learning methods with the lowest rMSFEs are the regression tree models, random forest models and the Gaussian process regression models. However, in the most recent subsample, given in Table 13, the lasso models and the elastic net models perform surprisingly well. Only for the 10 years maturity none of these models makes significant more accurate predictions than the AR(1) model. Furthermore, it seems that for the 12 step forecast horizon the addition of the three key macroeconomic variable does slightly improve the results of the machine learning techniques.

When we focus on all the tables at the same time, it seems that the machine learning models perform better in the third more recent subsample, whereas the ordinary econometric models have their worst performance during this period. Furthermore, the models with only the principal components as explanatory variables have rMSFEs of bigger than one for all maturities and all forecast horizons.

6 Conclusion

In this paper we try to answer to what extend machine learning techniques are able to improve the forecasting of the term structure of interest rates. We performed various machine learning methods and evaluated them based on their MSFE relative to the MSFE of a simple autoregressive model with one lag. However, these models perform rather disappointing. First, we replicate the econometric models proposed by Swanson and Xiong (2018). Thereafter, we use regression trees, random forests and gradient boosted trees to forecast the yield for five different maturities and three different forecast horizons. Next we apply the lasso and elastic net models, which shrink the number of variables used in the forecasts. The last machine learning method used to forecast the different yields we apply is the Gaussian process regression. For these machine learning models we use three sets of explanatory variables. The first set consists of 10 yields with different maturities, the second set of the three key macroeconomic variables suggested by Diebold et al. (2006) and the third set is the combination of the first two sets.

The regression tree based models are not able to significantly beat the benchmark AR(1) model in predicting the term structure of interest rates. This seems to contradict the findings of Rahimi (2020). However, Rahimi does not compare his random forest model with an AR(1) model, but only with a dynamic Nelson-Siegel model where the latent factors are forecasted using a AR model. When we compare our random forest model with the DNS(1), DNS(2) and DNS(3) we find that the the random forest model only performs better during the third most recent subsample. This difference with the findings of Rahimi is possibly due to the use of a different data set.

The lasso and elastic net model are not able to produce more significant more accurate forecasts of the yield curve than the AR(1) model as well. A reason for this could be the amount of explanatory variables used in the models. Increasing the amount of explanatory variables may improve the predictive performance of these models.

Table 3: One-step-ahead relative MSFEs of all forecasting models (subsample 1: 1992:3–1999:12)

Model	rMSFE					Model	rMSFE				
Maturity	1 year	2 years	3 years	5 years	10 years	Maturity	1 year	2 years	3 years	5 years	10 years
AR(1)	1.000	1.000	1.000	1.000	1.000	DNS(6)+MAC	1.119	1.165	1.130	1.161	1.188
VAR(1)	1.099	1.108	1.103	1.098	1.141	DIF(1)	3.048	2.655	1.926	0.919**	2.245
VAR(1)+FB1	0.819**	0.868*	0.893*	0.927	1.045	DIF(2)	1.274	1.067	1.038	1.029	1.199
VAR(1)+FB2	0.844	0.874	0.897	0.940	1.106	DIF(3)	0.973	1.046	1.044	1.049	1.128
AR(SIC)	0.864**	0.942*	0.958	0.974	0.972**	DIF(4)	2.238	2.303	2.337	2.382	2.438
VAR(SIC)	1.099	1.108	1.103	1.098	1.141	DIF(5)	2.253	2.338	2.386	2.455	2.588
VAR(SIC)+FB1	0.819**	0.868*	0.893*	0.927	1.045	DIF(6)	2.236	2.320	2.359	2.410	2.514
VAR(SIC)+FB2	0.844	0.874	0.897	0.940	1.106	DIF(1)+FB1	2.208	2.182	1.717	0.950	2.239
DNS(1)	1.032	1.097	1.061	1.039	1.067	DIF(2)+FB1	1.340	1.074	1.026	1.039	1.254
DNS(2)	1.036	1.088	1.053	1.046	1.064	DIF(3)+FB1	0.958	1.006	1.021	1.060	1.164
DNS(3)	1.040	1.123	1.066	1.045	1.037	DIF(1)+FB2	2.002	1.933	1.489	0.969	2.065
DNS(4)	1.088	1.160	1.104	1.070	1.102	DIF(2)+FB2	1.269	1.052	1.016	1.029	1.247
DNS(5)	1.095	1.147	1.095	1.081	1.098	DIF(3)+FB2	0.947	1.007	1.022	1.057	1.177
DNS(6)	1.094	1.190	1.107	1.065	1.071	TRE(1)	1.898	2.118	1.736	2.140	2.203
DNS(1)+FB1	0.900	0.862*	0.895	0.981	0.981	TRE(2)	90.607	62.668	53.187	43.335	63.626
DNS(2)+FB1	0.891	0.865*	0.903	1.000	0.980	TRE(3)	2.345	2.225	2.181	2.575	2.499
DNS(3)+FB1	0.876	0.868*	0.896	1.006	0.990	FOR(1)	2.179	1.581	1.458	1.591	2.074
DNS(4)+FB1	0.784**	0.861**	0.870**	0.922	0.990	FOR(2)	64.868	46.561	38.998	37.432	43.655
DNS(5)+FB1	0.785**	0.854**	0.867**	0.934	0.987	FOR(3)	2.396	1.608	1.521	1.635	2.121
DNS(6)+FB1	0.775***	0.882**	0.872**	0.930	0.985	GBT(1)	2.113	2.590	2.145	2.455	2.610
DNS(1)+FB2	0.960	0.908	0.948	1.053	1.053	GBT(2)	93.773	60.341	54.303	55.356	65.699
DNS(2)+FB2	0.948	0.911	0.957	1.074	1.051	GBT(3)	2.653	3.102	2.273	2.435	3.049
DNS(3)+FB2	0.933	0.911	0.948	1.081	1.073	LAS(1)	1.822	1.872	2.005	2.259	4.494
DNS(4)+FB2	0.789**	0.844**	0.858**	0.920	0.988	LAS(2)	446.429	375.295	370.567	425.594	682.249
DNS(5)+FB2	0.790**	0.840**	0.857**	0.934	0.985	LAS(3)	1.744	2.008	2.392	2.671	6.000
DNS(6)+FB2	0.775**	0.863**	0.860**	0.929	0.987	ENE(1)	2.108	1.705	1.724	1.582	4.737
DNS(1)+MAC	1.028	1.099	1.073	1.056	1.095	ENE(2)	444.809	374.603	368.661	425.032	683.862
DNS(2)+MAC	1.029	1.089	1.065	1.063	1.091	ENE(3)	1.938	1.803	2.043	1.881	6.318
DNS(3)+MAC	1.032	1.123	1.079	1.062	1.063	GPR(1)	1.023	**1.067**	**1.078**	**1.092**	**1.193**
DNS(4)+MAC	1.132	1.147	1.129	1.154	1.191	GPR(2)	64.573	46.084	39.960	38.229	45.884
DNS(5)+MAC	1.130	1.140	1.125	1.164	1.184	GPR(3)	**1.004**	1.087	1.106	1.119	1.201

This table reports the relative mean squared forecast error (rMSFE) for predictions of the monthly US treasury bond yields of various maturities for a one month forecast horizon, where the AR(1) model is used as a benchmark. Entries superscripted with ***, ** and * denote rejections of the null of equal predictive accuracy at 0.01, 0.05 and 0.10 significance levels, respectively, based on application of the Diebold–Mariano test and indicate that the listed model produces more accurate forecasts compared to the AR(1) benchmark, based on MSFE loss. Bold entries denote the best performing machine learning models for each maturity.

Table 4: One-step-ahead relative MSFEs of all forecasting models (subsample 2: 2000:1–2007:12)

Model	rMSFE					Model	rMSFE				
Maturity	1 year	2 years	3 years	5 years	10 years	Maturity	1 year	2 years	3 years	5 years	10 years
AR(1)	1.000	1.000	1.000	1.000	1.000	DNS(6)+MAC	0.949	1.057	1.056	1.190	1.097
VAR(1)	0.970	1.029	1.032	1.045	1.110	DIF(1)	2.474	2.046	1.688	1.062	1.788
VAR(1)+FB1	0.733**	0.858*	0.906	0.971	1.084	DIF(2)	1.288	1.112	1.104	1.061	1.214
VAR(1)+FB2	0.810	0.899	0.936	1.003	1.157	DIF(3)	1.029	1.128	1.114	1.073	1.121
AR(SIC)	0.939	1.033	1.033	1.035	1.015	DIF(4)	1.566	1.733	1.830	1.930	1.961
VAR(SIC)	0.970	1.029	1.032	1.045	1.110	DIF(5)	1.349	1.688	1.805	1.884	1.937
VAR(SIC)+FB1	0.733**	0.858*	0.906	0.971	1.084	DIF(6)	1.389	1.697	1.804	1.868	1.919
VAR(SIC)+FB2	0.810	0.899	0.936	1.003	1.157	DIF(1)+FB1	1.575	1.633	1.468	1.045	1.794
DNS(1)	1.211	1.015	1.000	1.094	0.959*	DIF(2)+FB1	1.093	1.001	1.027	1.038	1.227
DNS(2)	1.182	1.016	1.012	1.121	0.958**	DIF(3)+FB1	0.892	1.021	1.049	1.053	1.115
DNS(3)	1.150	1.015	0.998	1.126	0.983	DIF(1)+FB2	1.435	1.673	1.521	1.039	1.667
DNS(4)	1.017	1.067	1.031	1.082	1.026	DIF(2)+FB2	1.117	1.024	1.039	1.046	1.184
DNS(5)	1.014	1.058	1.034	1.110	1.027	DIF(3)+FB2	0.875	1.023	1.058	1.059	1.122
DNS(6)	1.021	1.099	1.037	1.099	1.032	TRE(1)	1.728	1.562	1.724	1.758	1.548
DNS(1)+FB1	0.780*	0.851	0.860	0.947	0.947	TRE(2)	75.620	50.094	41.182	34.342	23.738
DNS(2)+FB1	0.773*	0.842	0.859*	0.966	0.944	TRE(3)	1.862	1.811	1.962	1.908	1.907
DNS(3)+FB1	0.770*	0.873	0.863	0.966	0.944	FOR(1)	2.784	1.907	1.600	1.359	1.144
DNS(4)+FB1	0.708***	0.853**	0.866**	0.962	0.959	FOR(2)	65.510	40.243	30.064	21.867	17.613
DNS(5)+FB1	0.703***	0.840**	0.865**	0.987	0.960	FOR(3)	3.006	2.026	1.668	1.359	1.131
DNS(6)+FB1	0.713***	0.884*	0.872**	0.979	0.965	GBT(1)	1.835	2.244	2.161	1.895	1.646
DNS(1)+FB2	0.717**	0.741**	0.763**	0.887	0.855**	GBT(2)	92.456	65.436	47.425	38.136	27.007
DNS(2)+FB2	0.707**	0.734**	0.766**	0.912	0.854**	GBT(3)	2.078	1.898	2.069	2.493	1.674
DNS(3)+FB2	0.697**	0.756**	0.765**	0.915	0.877**	LAS(1)	2.392	1.864	1.851	1.980	2.465
DNS(4)+FB2	0.727***	0.793***	0.824***	0.961	0.933	LAS(2)	240.590	189.181	182.306	211.990	353.721
DNS(5)+FB2	0.721***	0.791***	0.832**	0.991	0.935	LAS(3)	1.625	1.584	1.869	2.317	4.046
DNS(6)+FB2	0.703***	0.810***	0.824***	0.983	0.960	ENE(1)	2.602	2.095	1.976	1.436	2.895
DNS(1)+MAC	1.065	0.982	1.002	1.099	0.979	ENE(2)	240.377	190.009	182.504	212.072	354.666
DNS(2)+MAC	1.037	0.983	1.011	1.125	0.977	ENE(3)	2.033	1.737	1.880	1.406	5.264
DNS(3)+MAC	1.000	0.983	1.000	1.129	0.997	GPR(1)	0.962	**1.013**	**1.023**	**1.052**	**1.126**
DNS(4)+MAC	0.972	1.040	1.056	1.165	1.064	GPR(2)	63.757	40.299	30.444	23.207	18.417
DNS(5)+MAC	0.960	1.037	1.065	1.197	1.065	GPR(3)	**0.895**	1.047	1.039	1.078	1.179

See notes to table 3.

Table 5: One-step-ahead relative MSFEs of all forecasting models (subsample 3: 2008:1–2016:7)

Model	rMSFE					Model	rMSFE				
Maturity	1 year	2 years	3 years	5 years	10 years	Maturity	1 year	2 years	3 years	5 years	10 years
AR(1)	1.000	1.000	1.000	1.000	1.000	DNS(6)+MAC	1.144	1.199	1.132	1.437	1.077
VAR(1)	1.177	1.242	1.222	1.180	1.158	DIF(1)	2.521	2.326	2.076	1.245	1.621
VAR(1)+FB1	1.249	1.311	1.293	1.281	1.297	DIF(2)	1.581	1.336	1.315	1.215	1.196
VAR(1)+FB2	1.369	1.451	1.406	1.345	1.320	DIF(3)	1.058	1.392	1.388	1.259	1.257
AR(SIC)	0.920	1.028	1.005	0.999	1.009	DIF(4)	4.145	3.409	2.884	2.440	2.219
VAR(SIC)	1.177	1.242	1.222	1.180	1.158	DIF(5)	4.718	3.707	2.960	2.305	1.980
VAR(SIC)+FB1	1.249	1.311	1.293	1.281	1.297	DIF(6)	4.699	3.805	3.151	2.573	2.192
VAR(SIC)+FB2	1.369	1.451	1.406	1.345	1.320	DIF(1)+FB1	4.053	3.207	2.411	1.280	1.575
DNS(1)	2.108	1.044	1.137	1.444	0.932*	DIF(2)+FB1	2.127	1.637	1.439	1.270	1.226
DNS(2)	1.966	1.085	1.223	1.542	0.928*	DIF(3)+FB1	1.367	1.666	1.532	1.326	1.320
DNS(3)	1.710	1.003	1.098	1.521	0.980	DIF(1)+FB2	4.438	3.232	2.257	1.179	1.452
DNS(4)	1.396	1.156	1.141	1.387	1.042	DIF(2)+FB2	2.100	1.615	1.403	1.182	1.106
DNS(5)	1.317	1.127	1.167	1.459	1.030	DIF(3)+FB2	1.341	1.623	1.448	1.190	1.147
DNS(6)	1.231	1.220	1.137	1.450	1.071	TRE(1)	1.372	1.801	1.743	2.029	2.452
DNS(1)+FB1	2.132	1.536	1.390	1.511	1.078	TRE(2)	139.756	100.203	75.180	46.067	21.543
DNS(2)+FB1	2.030	1.522	1.421	1.583	1.072	TRE(3)	1.559	2.070	2.055	2.459	2.585
DNS(3)+FB1	1.914	1.577	1.379	1.582	1.110	FOR(1)	1.456	1.677	1.730	1.601	1.526
DNS(4)+FB1	1.420	1.315	1.212	1.387	1.132	FOR(2)	103.158	76.155	56.616	36.700	20.745
DNS(5)+FB1	1.373	1.275	1.223	1.452	1.123	FOR(3)	1.502	1.739	1.777	1.660	1.595
DNS(6)+FB1	1.306	1.403	1.213	1.437	1.138	GBT(1)	2.089	2.476	2.371	2.316	2.017
DNS(1)+FB2	2.259	1.661	1.469	1.523	1.075	GBT(2)	200.279	125.565	94.255	55.715	23.052
DNS(2)+FB2	2.149	1.645	1.497	1.591	1.068	GBT(3)	2.051	2.857	2.470	2.458	2.084
DNS(3)+FB2	2.044	1.707	1.462	1.595	1.106	LAS(1)	9.156	6.381	4.609	3.396	4.911
DNS(4)+FB2	1.553	1.467	1.327	1.454	1.177	LAS(2)	22.994	24.709	30.986	50.870	96.424
DNS(5)+FB2	1.503	1.423	1.332	1.513	1.166	LAS(3)	3.940	3.164	2.920	2.528	4.866
DNS(6)+FB2	1.442	1.552	1.325	1.501	1.180	ENE(1)	9.407	6.573	4.720	3.394	5.114
DNS(1)+MAC	1.720	1.051	1.094	1.331	0.943	ENE(2)	22.565	24.054	30.774	50.362	96.263
DNS(2)+MAC	1.604	1.064	1.149	1.413	0.939	ENE(3)	5.346	3.730	2.920	2.434	5.270
DNS(3)+MAC	1.429	1.060	1.078	1.406	0.966	GPR(1)	**0.819**	0.966	0.968	0.984	1.078
DNS(4)+MAC	1.316	1.137	1.141	1.382	1.056	GPR(2)	104.178	76.496	56.695	37.218	20.790
DNS(5)+MAC	1.228	1.108	1.162	1.447	1.041	GPR(3)	0.910	**0.922**	**0.910**	**0.928**	**0.983**

See notes to table 3.

Table 6: One-step-ahead relative MSFEs of all forecasting models (subsample 4: 1992:3–2016:7)

Model	rMSFE					Model	rMSFE				
Maturity	1 year	2 years	3 years	5 years	10 years	Maturity	1 year	2 years	3 years	5 years	10 years
AR(1)	1.000	1.000	1.000	1.000	1.000	DNS(6)+MAC	1.055	1.127	1.100	1.252	1.112
VAR(1)	1.063	1.103	1.101	1.102	1.139	DIF(1)	2.702	2.334	1.863	1.067	1.838
VAR(1)+FB1	0.874*	0.955	0.990	1.046	1.165	DIF(2)	1.344	1.141	1.128	1.095	1.203
VAR(1)+FB2	0.940	1.003	1.029	1.081	1.213	DIF(3)	1.014	1.151	1.151	1.119	1.181
AR(SIC)	0.906*	0.998	0.999	1.004	1.001	DIF(4)	2.361	2.293	2.256	2.229	2.198
VAR(SIC)	1.063	1.103	1.101	1.102	1.139	DIF(5)	2.397	2.349	2.281	2.197	2.128
VAR(SIC)+FB1	0.874*	0.955	0.990	1.046	1.165	DIF(6)	2.404	2.366	2.314	2.253	2.193
VAR(SIC)+FB2	0.940	1.003	1.029	1.081	1.213	DIF(1)+FB1	2.334	2.165	1.774	1.081	1.818
DNS(1)	1.331	1.052	1.053	1.177	0.976	DIF(2)+FB1	1.403	1.159	1.120	1.105	1.234
DNS(2)	1.291	1.058	1.075	1.218	0.973	DIF(3)+FB1	1.016	1.147	1.149	1.134	1.215
DNS(3)	1.226	1.053	1.046	1.213	0.996	DIF(1)+FB2	2.279	2.092	1.677	1.056	1.681
DNS(4)	1.123	1.120	1.083	1.166	1.053	DIF(2)+FB2	1.381	1.156	1.114	1.080	1.167
DNS(5)	1.108	1.106	1.086	1.201	1.047	DIF(3)+FB2	1.000	1.140	1.134	1.096	1.147
DNS(6)	1.093	1.158	1.085	1.189	1.059	TRE(1)	1.718	1.822	1.733	1.964	2.108
DNS(1)+FB1	1.109	0.996	0.994	1.122	1.012	TRE(2)	94.738	65.150	53.318	40.744	33.366
DNS(2)+FB1	1.081	0.991	1.003	1.155	1.008	TRE(3)	1.982	2.021	2.063	2.291	2.354
DNS(3)+FB1	1.050	1.016	0.993	1.157	1.027	FOR(1)	2.276	1.736	1.578	1.507	1.553
DNS(4)+FB1	0.886*	0.951	0.946	1.071	1.041	FOR(2)	73.151	50.013	39.381	31.369	25.849
DNS(5)+FB1	0.874*	0.935	0.947	1.104	1.037	FOR(3)	2.460	1.809	1.639	1.538	1.592
DNS(6)+FB1	0.861**	0.990	0.950	1.095	1.045	GBT(1)	1.994	2.423	2.203	2.204	2.060
DNS(1)+FB2	1.132	0.993	0.992	1.127	1.001	GBT(2)	115.537	75.853	60.613	48.986	35.566
DNS(2)+FB2	1.100	0.988	1.003	1.162	0.998	GBT(3)	2.291	2.551	2.235	2.464	2.213
DNS(3)+FB2	1.069	1.010	0.991	1.167	1.027	LAS(1)	3.592	2.794	2.536	2.483	4.047
DNS(4)+FB2	0.924	0.951	0.951	1.090	1.052	LAS(2)	273.118	225.951	216.608	237.001	330.809
DNS(5)+FB2	0.911	0.939	0.955	1.123	1.047	LAS(3)	2.155	2.069	2.300	2.497	4.914
DNS(6)+FB2	0.885	0.982	0.951	1.115	1.061	ENE(1)	3.840	2.867	2.509	2.051	4.331
DNS(1)+MAC	1.188	1.040	1.049	1.152	0.995	ENE(2)	272.326	225.899	215.944	236.697	331.458
DNS(2)+MAC	1.153	1.040	1.063	1.187	0.991	ENE(3)	2.691	2.171	2.177	1.862	5.546
DNS(3)+MAC	1.102	1.052	1.047	1.187	1.001	GPR(1)	0.955	**1.024**	**1.031**	**1.046**	1.124
DNS(4)+MAC	1.105	1.101	1.102	1.224	1.094	GPR(2)	72.532	49.925	39.905	32.290	26.706
DNS(5)+MAC	1.081	1.091	1.109	1.258	1.086	GPR(3)	**0.940**	1.036	1.034	1.049	**1.101**

See notes to table 3.

Table 7: Three-step-ahead relative MSFEs of all forecasting models (subsample 1: 1992:7–1999:12)

Model	rMSFE					Model	rMSFE				
Maturity	1 year	2 years	3 years	5 years	10 years	Maturity	1 year	2 years	3 years	5 years	10 years
AR(1)	1.000	1.000	1.000	1.000	1.000	DNS(6)+MAC	0.950	1.053	1.054	1.038	1.025
VAR(1)	1.014	1.062	1.074	1.063	1.035	DIF(1)	1.678	1.558	1.285	0.987	1.381
VAR(1)+FB1	0.981	1.037	1.055	1.050	1.030	DIF(2)	1.272	1.252	1.227	1.207	1.248
VAR(1)+FB2	1.070	1.119	1.134	1.130	1.114	DIF(3)	1.209	1.241	1.218	1.181	1.195
AR(SIC)	0.885**	0.963	0.974	0.969	0.933**	DIF(4)	1.217	1.278	1.312	1.355	1.430
VAR(SIC)	1.014	1.062	1.074	1.063	1.035	DIF(5)	1.465	1.533	1.563	1.568	1.541
VAR(SIC)+FB1	0.981	1.037	1.055	1.050	1.030	DIF(6)	1.489	1.558	1.587	1.587	1.538
VAR(SIC)+FB2	1.070	1.119	1.134	1.130	1.114	DIF(1)+FB1	1.272	1.340	1.196	0.998	1.406
DNS(1)	1.061	1.079	1.065	1.047	1.029	DIF(2)+FB1	1.216	1.199	1.178	1.176	1.240
DNS(2)	1.067	1.082	1.069	1.054	1.031	DIF(3)+FB1	1.050	1.100	1.111	1.124	1.184
DNS(3)	1.065	1.080	1.062	1.046	1.028	DIF(1)+FB2	1.395	1.513	1.388	1.200	1.408
DNS(4)	0.990	1.075	1.063	1.024	1.003	DIF(2)+FB2	1.316	1.305	1.276	1.244	1.266
DNS(5)	0.996	1.071	1.058	1.024	1.002	DIF(3)+FB2	1.140	1.203	1.209	1.199	1.229
DNS(6)	1.000	1.086	1.063	1.017	0.989	TRE(1)	1.770	2.068	1.864	1.889	2.472
DNS(1)+FB1	0.922	0.914	0.954	1.013	1.000	TRE(2)	19.027	11.414	10.116	9.741	11.119
DNS(2)+FB1	0.923	0.924	0.964	1.024	1.004	TRE(3)	1.742	1.965	2.174	2.056	2.364
DNS(3)+FB1	0.918	0.914	0.955	1.025	1.022	FOR(1)	1.442	1.427	1.469	1.677	1.988
DNS(4)+FB1	0.855**	0.961	0.973	0.966	0.959	FOR(2)	13.180	10.002	9.058	8.895	10.386
DNS(5)+FB1	0.861**	0.957	0.969	0.966	0.956	FOR(3)	1.547	1.460	1.475	1.606	1.950
DNS(6)+FB1	0.860**	0.968	0.972	0.961	0.951	GBT(1)	1.939	2.109	1.993	2.378	2.487
DNS(1)+FB2	1.033	1.019	1.039	1.063	1.008	GBT(2)	19.532	12.835	9.541	10.664	13.289
DNS(2)+FB2	1.032	1.027	1.048	1.074	1.013	GBT(3)	1.782	2.367	2.271	2.635	2.999
DNS(3)+FB2	1.034	1.024	1.045	1.080	1.035	LAS(1)	2.416	2.493	2.171	2.299	5.076
DNS(4)+FB2	0.920*	1.016	1.025	1.014	0.998	LAS(2)	98.778	88.212	89.804	104.342	160.987
DNS(5)+FB2	0.930	1.017	1.025	1.019	0.999	LAS(3)	2.380	3.024	2.864	3.051	6.409
DNS(6)+FB2	0.919*	1.018	1.021	1.008	0.990	ENE(1)	2.517	2.389	2.184	2.258	5.169
DNS(1)+MAC	1.056	1.099	1.091	1.063	1.025	ENE(2)	99.304	88.253	89.678	104.289	161.298
DNS(2)+MAC	1.063	1.102	1.093	1.068	1.028	ENE(3)	2.456	2.631	2.689	3.026	6.749
DNS(3)+MAC	1.060	1.102	1.089	1.061	1.020	GPR(1)	**1.122**	1.307	1.283	1.332	1.401
DNS(4)+MAC	0.942	1.043	1.052	1.040	1.031	GPR(2)	13.671	10.586	9.687	9.635	11.394
DNS(5)+MAC	0.945	1.037	1.045	1.036	1.026	GPR(3)	1.187	**1.227**	**1.219**	**1.260**	**1.304**

See notes to table 3.

Table 8:Three-step-ahead relative MSFEs of all forecasting models (subsample 2: 2000:1–2007:12)

Model	rMSFE					Model	rMSFE				
Maturity	1 year	2 years	3 years	5 years	10 years	Maturity	1 year	2 years	3 years	5 years	10 years
AR(1)	1.000	1.000	1.000	1.000	1.000	DNS(6)+MAC	0.854***	0.901*	0.898*	0.959	0.996
VAR(1)	0.832***	0.880***	0.885**	0.909*	1.013	DIF(1)	1.641	1.460	1.331	1.146	1.286
VAR(1)+FB1	0.837***	0.880***	0.882**	0.905**	1.009	DIF(2)	1.234	1.217	1.191	1.188	1.354
VAR(1)+FB2	0.839***	0.889**	0.895**	0.924*	1.043	DIF(3)	1.186	1.280	1.261	1.241	1.340
AR(SIC)	0.819***	0.877**	0.873**	0.881**	0.937	DIF(4)	0.946	1.084	1.188	1.330	1.358
VAR(SIC)	0.832***	0.880***	0.885**	0.909*	1.013	DIF(5)	0.970	1.143	1.210	1.305	1.488
VAR(SIC)+FB1	0.837***	0.880***	0.882**	0.905**	1.009	DIF(6)	1.007	1.182	1.247	1.331	1.511
VAR(SIC)+FB2	0.839***	0.889**	0.895**	0.924*	1.043	DIF(1)+FB1	0.943	1.057	1.096	1.196	1.428
DNS(1)	1.250	1.068	1.009	1.051	0.913***	DIF(2)+FB1	0.915	1.018	1.082	1.184	1.462
DNS(2)	1.233	1.069	1.019	1.066	0.913***	DIF(3)+FB1	0.911	1.081	1.149	1.227	1.413
DNS(3)	1.232	1.063	1.009	1.078	0.977	DIF(1)+FB2	1.090	1.273	1.266	1.189	1.728
DNS(4)	0.905***	0.919**	0.895**	0.915**	0.929	DIF(2)+FB2	0.913	1.059	1.124	1.196	1.414
DNS(5)	0.900***	0.916**	0.897**	0.926**	0.929	DIF(3)+FB2	0.905	1.098	1.171	1.242	1.432
DNS(6)	0.915***	0.930*	0.897**	0.920**	0.942	TRE(1)	2.160	1.477	1.568	1.751	1.856
DNS(1)+FB1	0.676**	0.705**	0.724**	0.827**	0.846**	TRE(2)	16.525	12.579	11.322	11.637	11.576
DNS(2)+FB1	0.674**	0.706**	0.730**	0.840*	0.846**	TRE(3)	2.189	1.712	1.462	1.745	1.884
DNS(3)+FB1	0.672**	0.703**	0.720**	0.842*	0.891*	FOR(1)	1.773	1.502	1.315	**1.183**	1.254
DNS(4)+FB1	0.830***	0.857***	0.846***	0.884**	0.909*	FOR(2)	12.944	10.366	8.693	7.467	7.435
DNS(5)+FB1	0.830***	0.857***	0.851***	0.898**	0.909*	FOR(3)	1.792	1.489	1.347	1.214	**1.197**
DNS(6)+FB1	0.833***	0.863**	0.845***	0.889**	0.925*	GBT(1)	2.165	1.767	1.863	1.864	2.267
DNS(1)+FB2	0.794*	0.755**	0.773**	0.898	0.921	GBT(2)	19.769	14.682	12.059	11.824	11.774
DNS(2)+FB2	0.784*	0.754**	0.779**	0.912	0.920	GBT(3)	1.997	1.979	1.788	1.548	2.142
DNS(3)+FB2	0.793	0.757**	0.777**	0.929	0.998	LAS(1)	3.464	2.531	2.362	1.775	3.780
DNS(4)+FB2	0.833***	0.860***	0.849***	0.887**	0.915*	LAS(2)	45.880	45.727	48.777	65.766	134.252
DNS(5)+FB2	0.833***	0.860***	0.854***	0.901**	0.916*	LAS(3)	2.410	2.358	2.806	2.509	6.657
DNS(6)+FB2	0.836***	0.866**	0.848***	0.892**	0.931*	ENE(1)	3.587	2.770	2.099	1.799	4.255
DNS(1)+MAC	1.073	1.026	1.026	1.097	0.963*	ENE(2)	46.119	45.850	48.888	65.759	133.963
DNS(2)+MAC	1.055	1.028	1.036	1.111	0.962*	ENE(3)	2.372	2.279	2.116	2.074	7.859
DNS(3)+MAC	1.049	1.018	1.025	1.122	1.020	GPR(1)	1.146	1.255	1.295	1.403	1.897
DNS(4)+MAC	0.853***	0.897*	0.900*	0.952	0.972	GPR(2)	12.814	10.320	8.849	7.916	7.964
DNS(5)+MAC	0.849***	0.896*	0.903*	0.964	0.970	GPR(3)	**0.974**	**1.176**	**1.237**	1.296	1.448

See notes to table 3.

Table 9: Three-step-ahead relative MSFEs of all forecasting models (subsample 3: 2008:1–2016:7)

Model	rMSFE					Model	rMSFE				
Maturity	1 year	2 years	3 years	5 years	10 years	Maturity	1 year	2 years	3 years	5 years	10 years
AR(1)	1.000	1.000	1.000	1.000	1.000	DNS(6)+MAC	0.988	0.981	0.987	1.061	0.911**
VAR(1)	0.975	0.998	0.974	0.927*	0.926*	DIF(1)	1.367	1.366	1.315	1.124	1.230
VAR(1)+FB1	0.937	0.976	0.961	0.924	0.935	DIF(2)	1.514	1.527	1.428	1.249	1.170
VAR(1)+FB2	0.942	0.978	0.963	0.926	0.936	DIF(3)	1.483	1.602	1.535	1.369	1.219
AR(SIC)	0.975	0.971	0.946**	0.921**	0.917**	DIF(4)	2.757	2.332	2.023	1.677	1.430
VAR(SIC)	0.975	0.998	0.974	0.927*	0.926*	DIF(5)	2.918	2.444	2.107	1.728	1.429
VAR(SIC)+FB1	0.937	0.976	0.961	0.924	0.935	DIF(6)	3.131	2.713	2.456	2.145	1.729
VAR(SIC)+FB2	0.942	0.978	0.963	0.926	0.936	DIF(1)+FB1	2.756	2.260	1.817	1.282	1.270
DNS(1)	1.825	1.278	1.304	1.407	0.971	DIF(2)+FB1	2.146	1.940	1.674	1.376	1.230
DNS(2)	1.762	1.310	1.357	1.453	0.968	DIF(3)+FB1	1.867	1.948	1.741	1.460	1.272
DNS(3)	1.652	1.203	1.260	1.431	1.022	DIF(1)+FB2	2.826	2.300	1.852	1.328	1.252
DNS(4)	1.046	0.974	0.990	1.050	0.907*	DIF(2)+FB2	2.111	1.912	1.662	1.377	1.244
DNS(5)	1.025	0.973	1.001	1.071	0.899*	DIF(3)+FB2	1.810	1.895	1.702	1.437	1.256
DNS(6)	0.997	0.979	0.987	1.070	0.919*	TRE(1)	1.867	1.649	1.815	2.238	1.821
DNS(1)+FB1	2.186	1.848	1.660	1.545	1.093	TRE(2)	32.547	33.425	25.600	16.914	9.243
DNS(2)+FB1	2.154	1.846	1.675	1.571	1.089	TRE(3)	1.234	1.730	1.617	1.860	1.658
DNS(3)+FB1	2.118	1.823	1.633	1.563	1.136	FOR(1)	1.258	1.546	1.539	1.401	1.348
DNS(4)+FB1	0.957	0.929	0.948	1.014	0.914	FOR(2)	20.402	18.229	15.393	11.961	8.182
DNS(5)+FB1	0.939	0.927	0.957	1.035	0.907	FOR(3)	**1.108**	1.414	1.460	1.436	1.384
DNS(6)+FB1	0.915	0.939	0.947	1.032	0.920*	GBT(1)	1.244	1.850	2.026	2.157	2.225
DNS(1)+FB2	2.215	1.871	1.643	1.478	1.041	GBT(2)	38.157	30.898	26.558	15.911	12.274
DNS(2)+FB2	2.184	1.863	1.651	1.499	1.036	GBT(3)	1.198	1.814	2.019	2.398	2.515
DNS(3)+FB2	2.158	1.853	1.619	1.493	1.074	LAS(1)	8.242	5.012	3.330	2.671	4.209
DNS(4)+FB2	0.984	0.948	0.964	1.027	0.923	LAS(2)	5.728	6.372	8.200	14.847	35.004
DNS(5)+FB2	0.966	0.945	0.972	1.047	0.916	LAS(3)	5.813	4.365	3.087	3.110	5.662
DNS(6)+FB2	0.943	0.960	0.964	1.045	0.930	ENE(1)	8.669	5.201	3.493	2.557	4.357
DNS(1)+MAC	1.864	1.393	1.372	1.423	1.006	ENE(2)	5.407	6.199	8.085	15.032	34.983
DNS(2)+MAC	1.811	1.417	1.416	1.467	1.003	ENE(3)	6.237	4.576	3.550	2.744	6.052
DNS(3)+MAC	1.721	1.337	1.339	1.451	1.056	GPR(1)	1.132	1.594	1.477	1.346	1.156
DNS(4)+MAC	1.039	0.977	0.992	1.045	0.903*	GPR(2)	19.460	17.003	14.508	11.058	7.933
DNS(5)+MAC	1.017	0.978	1.004	1.065	0.895*	GPR(3)	1.168	**1.312**	**1.335**	**1.235**	**1.130**

See notes to table 3.

Table 10: Three-step-ahead relative MSFEs of all forecasting models (subsample 4: 1992:7–2016:7)

Model	rMSFE					Model	rMSFE				
Maturity	1 year	2 years	3 years	5 years	10 years	Maturity	1 year	2 years	3 years	5 years	10 years
AR(1)	1.000	1.000	1.000	1.000	1.000	DNS(6)+MAC	0.918**	0.976	0.978	1.017	0.974
VAR(1)	0.927**	0.974	0.977	0.972	0.987	DIF(1)	1.589	1.475	1.310	1.081	1.298
VAR(1)+FB1	0.909***	0.960	0.966	0.965	0.988	DIF(2)	1.312	1.300	1.261	1.212	1.246
VAR(1)+FB2	0.941**	0.994	1.001	1.002	1.027	DIF(3)	1.263	1.339	1.310	1.255	1.243
AR(SIC)	0.878***	0.930**	0.928***	0.925***	0.928***	DIF(4)	1.460	1.438	1.435	1.438	1.411
VAR(SIC)	0.927**	0.974	0.977	0.972	0.987	DIF(5)	1.591	1.582	1.557	1.523	1.484
VAR(SIC)+FB1	0.909***	0.960	0.966	0.965	0.988	DIF(6)	1.665	1.668	1.664	1.658	1.604
VAR(SIC)+FB2	0.941**	0.994	1.001	1.002	1.027	DIF(1)+FB1	1.477	1.434	1.306	1.147	1.359
DNS(1)	1.320	1.119	1.101	1.150	0.976	DIF(2)+FB1	1.304	1.294	1.260	1.235	1.295
DNS(2)	1.300	1.129	1.118	1.171	0.976	DIF(3)+FB1	1.181	1.284	1.276	1.254	1.279
DNS(3)	1.273	1.101	1.089	1.166	1.012	DIF(1)+FB2	1.598	1.594	1.452	1.233	1.433
DNS(4)	0.967*	0.989	0.981	0.994	0.946*	DIF(2)+FB2	1.329	1.343	1.310	1.265	1.297
DNS(5)	0.962*	0.986	0.982	1.004	0.943*	DIF(3)+FB2	1.195	1.318	1.312	1.281	1.293
DNS(6)	0.963*	0.999	0.981	0.999	0.949**	TRE(1)	1.960	1.735	1.738	1.941	2.058
DNS(1)+FB1	1.111	1.041	1.034	1.100	0.995	TRE(2)	21.105	16.866	14.288	12.423	10.517
DNS(2)+FB1	1.103	1.045	1.044	1.116	0.995	TRE(3)	1.815	1.809	1.766	1.894	1.964
DNS(3)+FB1	1.092	1.035	1.027	1.114	1.031	FOR(1)	1.549	1.486	1.425	1.426	1.542
DNS(4)+FB1	0.868***	0.912**	0.918**	0.952**	0.928**	FOR(2)	14.761	12.011	10.434	9.273	8.754
DNS(5)+FB1	0.866***	0.910***	0.920***	0.962*	0.925**	FOR(3)	1.549	1.461	1.422	1.423	1.532
DNS(6)+FB1	0.861***	0.919**	0.917**	0.956**	0.932**	GBT(1)	1.874	1.913	1.951	2.139	2.328
DNS(1)+FB2	1.206	1.106	1.081	1.124	0.997	GBT(2)	23.973	17.669	14.585	12.548	12.496
DNS(2)+FB2	1.194	1.106	1.089	1.139	0.997	GBT(3)	1.738	2.085	2.025	2.195	2.585
DNS(3)+FB2	1.193	1.104	1.079	1.145	1.040	LAS(1)	4.222	3.078	2.522	2.225	4.398
DNS(4)+FB2	0.898***	0.938*	0.943**	0.974	0.947*	LAS(2)	54.446	52.564	54.460	65.760	105.315
DNS(5)+FB2	0.897***	0.938**	0.947**	0.987	0.945*	LAS(3)	3.192	3.059	2.895	2.882	6.186
DNS(6)+FB2	0.889***	0.944*	0.940**	0.979	0.951**	ENE(1)	4.409	3.179	2.465	2.185	4.614
DNS(1)+MAC	1.251	1.136	1.133	1.177	1.001	ENE(2)	54.652	52.590	54.428	65.790	105.339
DNS(2)+MAC	1.234	1.144	1.148	1.196	1.001	ENE(3)	3.301	2.929	2.675	2.620	6.774
DNS(3)+MAC	1.209	1.121	1.124	1.192	1.034	GPR(1)	1.135	1.351	1.334	1.360	1.438
DNS(4)+MAC	0.927**	0.969	0.979	1.011	0.966	GPR(2)	14.653	11.931	10.519	9.449	9.151
DNS(5)+MAC	0.921**	0.967	0.981	1.020	0.961	GPR(3)	**1.092**	**1.225**	**1.254**	**1.265**	**1.275**

See notes to table 3.

Table 11: 12-step-ahead relative MSFEs of all forecasting models (subsample 1: 1994:1–1999:12)

Model	rMSFE					Model	rMSFE				
Maturity	1 year	2 years	3 years	5 years	10 years	Maturity	1 year	2 years	3 years	5 years	10 years
AR(1)	1.000	1.000	1.000	1.000	1.000	DNS(6)+MAC	1.237	1.280	1.253	1.176	1.045
VAR(1)	1.312	1.302	1.280	1.207	1.049	DIF(1)	0.984	0.925*	0.838***	1.124	1.829
VAR(1)+FB1	1.299	1.290	1.267	1.194	1.036	DIF(2)	1.328	1.346	1.493	1.748	1.905
VAR(1)+FB2	1.292	1.282	1.260	1.189	1.031	DIF(3)	1.254	1.224	1.348	1.586	1.812
AR(SIC)	1.208	1.217	1.206	1.132	0.967	DIF(4)	1.122	1.156	1.184	1.225	1.292
VAR(SIC)	1.312	1.302	1.280	1.207	1.049	DIF(5)	1.619	1.610	1.647	1.689	1.690
VAR(SIC)+FB1	1.299	1.290	1.267	1.194	1.036	DIF(6)	1.718	1.695	1.723	1.747	1.712
VAR(SIC)+FB2	1.292	1.282	1.260	1.189	1.031	DIF(1)+FB1	1.340	1.283	1.158	1.371	2.125
DNS(1)	0.635***	0.682***	0.730***	0.846**	0.954	DIF(2)+FB1	1.622	1.749	1.871	2.054	2.170
DNS(2)	0.640***	0.688***	0.737***	0.853**	0.956	DIF(3)+FB1	1.459	1.542	1.653	1.850	2.084
DNS(3)	0.624***	0.669***	0.718***	0.845**	0.973	DIF(1)+FB2	1.487	1.503	1.548	1.960	2.324
DNS(4)	1.276	1.298	1.259	1.165	1.022	DIF(2)+FB2	1.883	1.908	2.002	2.166	2.262
DNS(5)	1.283	1.298	1.258	1.168	1.025	DIF(3)+FB2	1.593	1.637	1.752	1.961	2.176
DNS(6)	1.284	1.301	1.256	1.157	1.013	TRE(1)	1.756	1.783	1.510	1.767	1.890
DNS(1)+FB1	0.905	0.889	0.952	1.097	1.159	TRE(2)	2.387	2.066	1.905	2.136	2.782
DNS(2)+FB1	0.902	0.895	0.960	1.104	1.157	TRE(3)	2.119	2.204	1.498	1.984	1.871
DNS(3)+FB1	0.897	0.882	0.950	1.113	1.200	FOR(1)	1.392	1.406	**1.418**	**1.523**	1.672
DNS(4)+FB1	1.225	1.265	1.237	1.157	1.024	FOR(2)	**1.304**	**1.360**	1.517	1.939	2.459
DNS(5)+FB1	1.232	1.265	1.236	1.158	1.027	FOR(3)	1.402	1.416	1.441	1.564	**1.637**
DNS(6)+FB1	1.234	1.270	1.236	1.150	1.017	GBT(1)	2.669	2.287	2.275	2.018	2.092
DNS(1)+FB2	1.055	0.980	1.049	1.226	1.295	GBT(2)	2.716	2.201	1.714	2.486	2.746
DNS(2)+FB2	1.047	0.984	1.056	1.233	1.293	GBT(3)	2.469	2.430	2.319	2.071	2.250
DNS(3)+FB2	1.059	0.984	1.060	1.256	1.351	LAS(1)	5.481	6.650	7.571	8.878	12.598
DNS(4)+FB2	1.210	1.252	1.227	1.152	1.020	LAS(2)	25.858	24.790	26.050	29.047	36.006
DNS(5)+FB2	1.216	1.250	1.225	1.153	1.022	LAS(3)	6.138	7.285	7.885	9.363	12.919
DNS(6)+FB2	1.220	1.257	1.226	1.147	1.014	ENE(1)	5.473	6.645	7.438	8.836	12.292
DNS(1)+MAC	0.685**	0.729**	0.776**	0.889	0.977	ENE(2)	25.888	24.911	26.070	29.097	36.018
DNS(2)+MAC	0.689**	0.734**	0.782**	0.895	0.978	ENE(3)	6.296	7.226	7.913	9.324	12.601
DNS(3)+MAC	0.672**	0.716**	0.765**	0.889	0.998	GPR(1)	4.372	4.263	3.833	3.400	3.340
DNS(4)+MAC	1.228	1.275	1.253	1.181	1.050	GPR(2)	1.484	1.530	1.722	2.187	2.766
DNS(5)+MAC	1.233	1.273	1.250	1.180	1.050	GPR(3)	2.063	1.864	1.635	1.693	2.017

See notes to table 3.

Table 12: 12-step-ahead relative MSFEs of all Forecasting models (subsample 2: 2000:1–2007:12)

Model	rMSFE					Model	rMSFE				
Maturity	1 year	2 years	3 years	5 years	10 years	Maturity	1 year	2 years	3 years	5 years	10 years
AR(1)	1.000	1.000	1.000	1.000	1.000	DNS(6)+MAC	0.583***	0.497***	0.450***	0.447***	0.568***
VAR(1)	0.567***	0.475***	0.432***	0.431***	0.557***	DIF(1)	1.244	1.157	1.124	1.142	1.537
VAR(1)+FB1	0.583***	0.488***	0.444***	0.441***	0.561***	DIF(2)	1.895	1.456	1.321	1.372	1.801
VAR(1)+FB2	0.587***	0.493***	0.451***	0.452***	0.584***	DIF(3)	2.284	1.721	1.565	1.569	1.959
AR(SIC)	0.574***	0.485***	0.444***	0.445***	0.547***	DIF(4)	0.842***	0.966	1.112	1.448	1.910
VAR(SIC)	0.567***	0.475***	0.432***	0.431***	0.557***	DIF(5)	1.019	1.109	1.243	1.765	3.530
VAR(SIC)+FB1	0.583***	0.488***	0.444***	0.441***	0.561***	DIF(6)	1.037	1.146	1.299	1.857	3.729
VAR(SIC)+FB2	0.587***	0.493***	0.451***	0.452***	0.584***	DIF(1)+FB1	1.001	1.056	1.129	1.539	2.403
DNS(1)	0.708***	0.602***	0.571***	0.631***	0.770***	DIF(2)+FB1	1.544	1.364	1.412	1.711	2.703
DNS(2)	0.707***	0.605***	0.576***	0.638***	0.771***	DIF(3)+FB1	1.798	1.664	1.720	2.028	2.985
DNS(3)	0.702***	0.599***	0.570***	0.639***	0.811***	DIF(1)+FB2	1.032	1.138	1.272	1.842	3.941
DNS(4)	0.593***	0.507***	0.460***	0.454***	0.564***	DIF(2)+FB2	1.850	1.804	1.980	2.589	4.305
DNS(5)	0.593***	0.506***	0.459***	0.456***	0.564***	DIF(3)+FB2	2.098	2.053	2.226	2.832	4.471
DNS(6)	0.599***	0.510***	0.460***	0.452***	0.563***	TRE(1)	1.153	1.191	1.332	1.522	2.233
DNS(1)+FB1	0.548***	0.507***	0.520***	0.658***	0.999	TRE(2)	1.491	1.370	1.268	1.630	2.815
DNS(2)+FB1	0.547***	0.508***	0.523***	0.662***	0.996	TRE(3)	1.219	1.456	1.485	1.728	2.423
DNS(3)+FB1	0.543***	0.502***	0.516***	0.661***	1.036	FOR(1)	1.357	1.306	1.304	1.426	**1.969**
DNS(4)+FB1	0.595***	0.509***	0.461***	0.455***	0.564***	FOR(2)	**1.094**	**1.069**	**1.091**	**1.304**	2.269
DNS(5)+FB1	0.595***	0.508***	0.461***	0.457***	0.565***	FOR(3)	1.479	1.427	1.393	1.499	1.985
DNS(6)+FB1	0.598***	0.511***	0.461***	0.452***	0.562***	GBT(1)	1.367	1.334	1.398	1.676	2.523
DNS(1)+FB2	1.023	1.049	1.143	1.530	2.724	GBT(2)	1.438	1.426	1.604	1.786	3.045
DNS(2)+FB2	1.023	1.055	1.152	1.541	2.725	GBT(3)	1.254	1.607	1.556	1.962	2.466
DNS(3)+FB2	1.010	1.036	1.130	1.521	2.739	LAS(1)	2.128	2.049	1.921	1.918	4.485
DNS(4)+FB2	0.594***	0.510***	0.464***	0.460***	0.574***	LAS(2)	3.392	3.946	4.959	8.677	27.737
DNS(5)+FB2	0.594***	0.510***	0.464***	0.462***	0.575***	LAS(3)	1.830	1.844	1.786	2.076	4.978
DNS(6)+FB2	0.597***	0.511***	0.463***	0.456***	0.571***	ENE(1)	2.022	1.920	1.793	1.907	4.494
DNS(1)+MAC	0.671***	0.609***	0.604***	0.700***	0.884***	ENE(2)	3.395	3.957	4.930	8.665	27.743
DNS(2)+MAC	0.669***	0.611***	0.609***	0.706***	0.884***	ENE(3)	1.693	1.633	1.617	2.088	5.128
DNS(3)+MAC	0.665***	0.606***	0.604***	0.708***	0.922**	GPR(1)	1.837	1.786	1.711	1.894	3.491
DNS(4)+MAC	0.579***	0.495***	0.451***	0.449***	0.567***	GPR(2)	1.134	1.114	1.151	1.403	2.372
DNS(5)+MAC	0.580***	0.495***	0.450***	0.451***	0.566***	GPR(3)	1.373	1.460	1.575	1.951	3.321

See notes to table 3.

Table 13: 12-step-ahead relative MSFEs of all forecasting models (subsample 3: 2008:1–2016:7)

Model	rMSFE					Model	rMSFE				
Maturity	1 year	2 years	3 years	5 years	10 years	Maturity	1 year	2 years	3 years	5 years	10 years
AR(1)	1.000	1.000	1.000	1.000	1.000	DNS(6)+MAC	0.739***	0.675***	0.689***	0.747***	0.711***
VAR(1)	0.737***	0.716***	0.695***	0.663***	0.709***	DIF(1)	1.190	1.280	1.277	1.357	1.155
VAR(1)+FB1	0.778***	0.756***	0.731***	0.686***	0.707***	DIF(2)	2.405	2.453	2.228	1.789	1.155
VAR(1)+FB2	0.796***	0.773***	0.747***	0.697***	0.713***	DIF(3)	2.787	2.711	2.641	2.260	1.472
AR(SIC)	0.729***	0.714***	0.687***	0.674***	0.737***	DIF(4)	1.978	1.719	1.539	1.291	1.086
VAR(SIC)	0.737***	0.716***	0.695***	0.663***	0.709***	DIF(5)	1.859	1.677	1.611	1.564	1.457
VAR(SIC)+FB1	0.778***	0.756***	0.731***	0.686***	0.707***	DIF(6)	2.332	2.278	2.257	2.110	1.711
VAR(SIC)+FB2	0.796***	0.773***	0.747***	0.697***	0.713***	DIF(1)+FB1	2.036	1.789	1.563	1.316	1.167
DNS(1)	1.468	1.389	1.483	1.559	1.118	DIF(2)+FB1	2.280	2.164	1.982	1.663	1.171
DNS(2)	1.451	1.412	1.520	1.591	1.118	DIF(3)+FB1	2.490	2.487	2.410	2.139	1.569
DNS(3)	1.423	1.356	1.467	1.583	1.179	DIF(1)+FB2	1.886	1.726	1.614	1.618	1.462
DNS(4)	0.765***	0.689***	0.697***	0.737***	0.687***	DIF(2)+FB2	2.289	2.228	2.123	1.907	1.430
DNS(5)	0.759***	0.694***	0.708***	0.748***	0.683***	DIF(3)+FB2	2.503	2.530	2.505	2.297	1.737
DNS(6)	0.747***	0.679***	0.692***	0.747***	0.706***	TRE(1)	1.585	2.081	1.965	1.875	1.513
DNS(1)+FB1	1.744	1.554	1.511	1.534	1.209	TRE(2)	3.190	3.932	3.967	3.894	3.058
DNS(2)+FB1	1.734	1.557	1.524	1.552	1.206	TRE(3)	1.973	2.059	2.354	2.298	1.556
DNS(3)+FB1	1.736	1.542	1.502	1.554	1.274	FOR(1)	1.341	1.576	1.666	1.544	**1.189**
DNS(4)+FB1	0.813***	0.730***	0.737***	0.771***	0.703***	FOR(2)	2.568	3.125	3.427	3.322	2.625
DNS(5)+FB1	0.804***	0.735***	0.746***	0.781***	0.699***	FOR(3)	1.232	1.553	1.679	1.633	1.201
DNS(6)+FB1	0.791***	0.718***	0.730***	0.780***	0.723***	GBT(1)	1.559	1.727	2.102	1.953	1.485
DNS(1)+FB2	1.895	1.779	1.792	1.842	1.427	GBT(2)	3.397	4.071	4.294	3.788	3.254
DNS(2)+FB2	1.887	1.785	1.809	1.863	1.425	GBT(3)	1.474	2.194	2.450	2.063	1.458
DNS(3)+FB2	1.877	1.757	1.771	1.851	1.486	LAS(1)	1.621	0.851***	0.648***	0.519***	1.501
DNS(4)+FB2	0.827***	0.742***	0.747***	0.777***	0.703***	LAS(2)	**0.679*** **	0.932***	1.536	3.487	9.576
DNS(5)+FB2	0.816***	0.745***	0.755***	0.786***	0.699***	LAS(3)	0.884***	0.972***	1.034	0.727***	1.620
DNS(6)+FB2	0.804***	0.730***	0.740***	0.786***	0.723***	ENE(1)	1.789	0.794***	**0.527*** **	**0.489*** **	1.670
DNS(1)+MAC	1.579	1.583	1.729	1.824	1.274	ENE(2)	0.646***	0.910***	1.495	3.553	9.549
DNS(2)+MAC	1.555	1.600	1.761	1.850	1.269	ENE(3)	0.711***	**0.655*** **	0.701***	0.561***	1.797
DNS(3)+MAC	1.543	1.557	1.723	1.865	1.353	GPR(1)	1.306	1.513	1.533	1.733	2.386
DNS(4)+MAC	0.758***	0.685***	0.695***	0.739***	0.694***	GPR(2)	2.736	3.301	3.576	3.441	2.840
DNS(5)+MAC	0.751***	0.690***	0.705***	0.749***	0.691***	GPR(3)	1.359	1.592	1.620	1.430	1.225

See notes to table 3.

Table 14: 12-step-ahead relative MSFEs of all forecasting models (subsample 4: 1994:1–2016:7)

Model	rMSFE					Model	rMSFE				
Maturity	1 year	2 years	3 years	5 years	10 years	Maturity	1 year	2 years	3 years	5 years	10 years
AR(1)	1.000	1.000	1.000	1.000	1.000	DNS(6)+MAC	0.707***	0.661***	0.647***	0.687***	0.770***
VAR(1)	0.706***	0.660***	0.643***	0.665***	0.767***	DIF(1)	1.197	1.146	1.104	1.191	1.492
VAR(1)+FB1	0.724***	0.674***	0.655***	0.673***	0.763***	DIF(2)	1.941	1.653	1.546	1.561	1.602
VAR(1)+FB2	0.730***	0.680***	0.661***	0.680***	0.771***	DIF(3)	2.267	1.854	1.755	1.744	1.738
AR(SIC)	0.695***	0.652***	0.635***	0.657***	0.748***	DIF(4)	1.149	1.159	1.216	1.358	1.419
VAR(SIC)	0.706***	0.660***	0.643***	0.665***	0.767***	DIF(5)	1.299	1.312	1.395	1.698	2.204
VAR(SIC)+FB1	0.724***	0.674***	0.655***	0.673***	0.763***	DIF(6)	1.435	1.478	1.581	1.895	2.367
VAR(SIC)+FB2	0.730***	0.680***	0.661***	0.680***	0.771***	DIF(1)+FB1	1.292	1.250	1.227	1.445	1.871
DNS(1)	0.879**	0.784***	0.794***	0.910**	0.953	DIF(2)+FB1	1.729	1.598	1.617	1.777	1.984
DNS(2)	0.875**	0.792***	0.806***	0.923*	0.954*	DIF(3)+FB1	1.917	1.821	1.855	2.015	2.192
DNS(3)	0.863***	0.773***	0.788***	0.919*	0.994	DIF(1)+FB2	1.295	1.323	1.395	1.813	2.540
DNS(4)	0.725***	0.673***	0.656***	0.686***	0.753***	DIF(2)+FB2	1.959	1.912	2.014	2.324	2.627
DNS(5)	0.724***	0.674***	0.658***	0.690***	0.753***	DIF(3)+FB2	2.127	2.088	2.199	2.502	2.764
DNS(6)	0.725***	0.674***	0.655***	0.685***	0.756***	TRE(1)	**1,336**	1,478	1,500	1,665	1,866
DNS(1)+FB1	0.880	0.794**	0.810**	0.975	1.125	TRE(2)	2,014	2,034	1,960	2,306	2,892
DNS(2)+FB1	0.876	0.796**	0.816**	0.983	1.122	TRE(3)	1,518	1,706	1,673	1,928	1,937
DNS(3)+FB1	0.873	0.787**	0.805**	0.985	1.173	FOR(1)	1,358	**1,380**	**1,402**	**1,477**	1,595
DNS(4)+FB1	0.730***	0.678***	0.661***	0.692***	0.759***	FOR(2)	1,472	1,558	1,667	1,948	2,457
DNS(5)+FB1	0.729***	0.679***	0.663***	0.696***	0.759***	FOR(3)	1,410	1,452	1,463	1,547	**1,593**
DNS(6)+FB1	0.729***	0.677***	0.659***	0.692***	0.764***	GBT(1)	1,586	1,572	1,708	1,823	2,014
DNS(1)+FB2	1.234	1.195	1.264	1.539	1.807	GBT(2)	2,073	2,120	2,198	2,441	3,025
DNS(2)+FB2	1.232	1.201	1.274	1.551	1.806	GBT(3)	1,468	1,866	1,885	2,012	2,035
DNS(3)+FB2	1.222	1.183	1.254	1.543	1.850	LAS(1)	2,453	2,532	2,679	3,151	5,974
DNS(4)+FB2	0.731***	0.679***	0.663***	0.695***	0.761***	LAS(2)	5,731	6,652	8,072	12,013	23,821
DNS(5)+FB2	0.730***	0.680***	0.665***	0.699***	0.760***	LAS(3)	2,177	2,532	2,737	3,395	6,279
DNS(6)+FB2	0.729***	0.678***	0.661***	0.695***	0.766***	ENE(1)	2,425	2,438	2,552	3,128	5,941
DNS(1)+MAC	0.889*	0.838**	0.875**	1.022	1.054	ENE(2)	5,729	6,674	8,049	12,034	23,817
DNS(2)+MAC	0.882**	0.844**	0.886*	1.032	1.052	ENE(3)	2,071	2,323	2,569	3,351	6,290
DNS(3)+MAC	0.875**	0.828***	0.872**	1.036	1.101	GPR(1)	2,047	2,126	2,060	2,196	3,046
DNS(4)+MAC	0.708***	0.662***	0.649***	0.687***	0.765***	GPR(2)	1,562	1,651	1,772	2,086	2,665
DNS(5)+MAC	0.707***	0.662***	0.651***	0.690***	0.764***	GPR(3)	1,461	1,554	1,595	1,763	2,156

See notes to table 3.

The final machine learning method, the Gaussian process regression, does produce rMSFEs smaller than one for some subsamples. Nevertheless, in none of these cases the Gaussian process regression models makes predictions that are significant more accurate than that of an AR(1) model. Sambasivan and Das (2017) find that in the medium and long term the Gaussian progress regression models do well compared to (V)AR models. However, they only report the root mean squared error of both models and not whether their is a significant difference between. Since the differences between the root mean squared errors are rather small, this corresponds to the the slightly smaller than one rMSFEs found for the third more recent subsample, which closely resembles the period investigated by Sambasivan and Das.

Furthermore, we find that the addition of the three key macroeconomic variables proposed by Diebold et al. (2006) does not substantially increase the predictive accuracy of the machine learning methods. All in all, the machine learning techniques are not able to provide significant more accurate forecasts than the AR(1) model and are therefore not an improvement compared to the more ordinary econometric models.

One of the limitations of this paper is of course that not all machine learning techniques are being considered in this research. Therefore, future research could focus on other machine learning methods and whether they produce better results than the machine learning methods proposed in this paper. Another extension one could make is the application of forecast combinations. Future research could make a combination of the different machine learning forecasts and check whether this yields any improvement compared to the single forecasts, since, according to Swanson and Xiong (2018), the inclusion of forecast combinations is useful for time series forecasts. Furthermore, as mentioned earlier, the addition of extra explanatory variables may improve the predictive accuracy of the machine learning methods.

References

A. Ang, M. Piazzesi, and M. Wei. What does the yield curve tell us about gdp growth? *Journal of econometrics*, 131(1-2):359–403, 2006.

L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen. *Classification and regression trees.* CRC press, 1984.

M. Chauvet and S. Potter. Forecasting recessions using the yield curve. *Journal of forecasting*, 24(2):77–103, 2005.

F. X. Diebold and C. Li. Forecasting the term structure of government bond yields. *Journal of econometrics*, 130(2):337–364, 2006.

F. X. Diebold and R. S. Mariano. Comparing predictive accuracy. *Journal of Business & economic statistics*, 20(1):134–144, 2002.

F. X. Diebold, G. D. Rudebusch, and S. B. Aruoba. The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of econometrics*, 131(1-2):309–338, 2006.

M. J. Dueker. Strengthening the case for the yield curve as a predictor of us recessions. *Federal Reserve Bank of St. Louis Review*, 79(2):41, 1997.

D. Durand. Basic yields of corporate bonds, 1900-1942. In *Basic Yields of Corporate Bonds, 1900-1942*, pages 1–40. NBER, 1942.

A. Estrella and F. S. Mishkin. The yield curve as a predictor of us recessions. *Current issues in economics and finance*, 2(7), 1996.

J. Friedman, T. Hastie, and R. Tibshirani. Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, 33(1):1, 2010.

J. H. Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.

R. S. Gürkaynak, B. Sack, and J. H. Wright. The us treasury yield curve: 1961 to the present. *Journal of monetary Economics*, 54(8):2291–2304, 2007.

J. G. Haubrich, A. M. Dombrosky, et al. Predicting real growth using the yield curve. *Economic Review*, 32(1):26–35, 1996.

P. J. Knez, R. Litterman, and J. Scheinkman. Explorations into factors explaining money market returns. *The Journal of Finance*, 49(5):1861–1882, 1994.

R. Litterman and J. Scheinkman. Common factors affecting bond returns. *Journal of fixed income*, 1(1):54–61, 1991.

M. W. McCracken and S. Ng. Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34(4):574–589, 2016.

E. Mönch. Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented var approach. *Journal of Econometrics*, 146(1):26–43, 2008.

C. R. Nelson and A. F. Siegel. Parsimonious modeling of yield curves. *Journal of business*, pages 473–489, 1987.

K. Rahimi. *Forecast comparison of financial models.* PhD thesis, University of Surrey, 2020.

R. Sambasivan and S. Das. A statistical machine learning approach to yield curve forecasting. In *2017 International Conference on Computational Intelligence in Data Science (ICCIDS)*, pages 1–6. IEEE, 2017.

G. Schwarz et al. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978.

L. E. Svensson. Estimating and interpreting forward interest rates: Sweden 1992-1994. Technical report, National bureau of economic research, 1994.

N. R. Swanson and W. Xiong. Big data analytics in economics: What have we learned so far, and where should we go from here? *Canadian Journal of Economics/Revue canadienne d'économique*, 51(3):695–746, 2018.

R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

C. K. Williams and C. E. Rasmussen. *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA, 2006.

H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, 67(2):301–320, 2005.