# Credit Spread Forecasting with Machine Learning Models

This paper attempts to investigate the year-to-year variation in credit spread between investment grade corporate bonds and US treasury bonds associated with economic and financial predictors. Methodology is explained in detail, before an account of data preparation steps is given. Inspired by the Merton's credit default risk model (1974), we start with linear regression as a benchmark test, and eventually turn to a host of machine learning techniques including LASSO regression, random forests and boosting algorithm. Ultimately, we aim to forecast the direction and the magnitude of such credit spread. Boosting method, supported by our statistical findings, turns out to be the best performer. Our conclusion is that, at a monthly level, the interest rate spread between AAA and BAA rated cooperate bonds and US Treasury bonds will be increasing at a decreasing rate.

**Contents**



## 1 Introduction

Credit spread in the bond market, commonly perceived as the spread between the yield of a risk-free bond and that of risky bond portfolio, is an important indicator in speculative and trading activities for fixed income investors. Historically, credit spreads have displayed close connection with economic events. From mid-2003 through mid-2007, (Manzoni, ) both investment grade and high yield credit spreads narrowed as recession fears faded. Bouts of tightening took place before the widening in 2008: US investment grade reached its tightest +76bps in March 2005, whereas for high yield credit spreads it did come about until May 2007 at +233bps, shortly before the spreads began to widen drastically. Post-crisis era (Manzoni, ) saw a tightening in credit spread across the board, benefiting from debt refinancing policy and improved liquidity profile by the Fed. These brought about tights of +97bps for investment grade and +323bps for high yield, in the course of 5.5 years from the peaks of the recession through 2014.

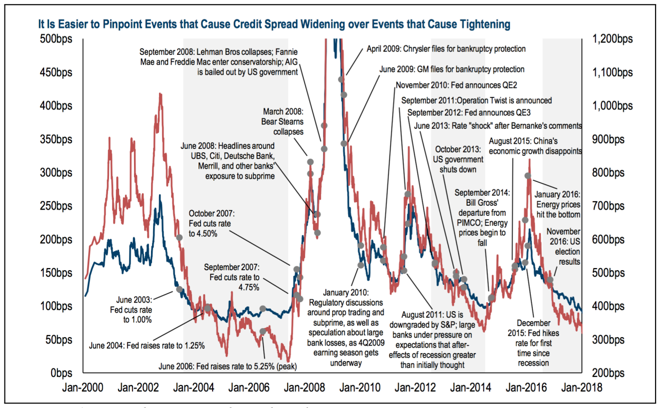

*Figure 1. Credit Spread and Other Investment. Source: (Manzoni, )*

Past literature, additional to economic cycles, has highlighted other factors on credit spread changes, including (i) proxies for the default risk— earnings variability, time of no default (Boardman McEnally, 1981; El-Jahel, 1998; Fisher, 1959; Nielsen, Saa'-Requejo, Santa-Clara, 1993); (ii) bond specific features— callability, marketability, coupon rate, recovery factor (Boardman McEnally, 1981; Duffee, 1998; Fisher, 1959; Silvers, 1973); (iii) returns on the firm's assets (Longstaff Schwartz, 1995); (iv) business cycle and confidence variables—consumer sentiment, returns on stock indices, industrial production, inflation, unemployment (Fons, 1987; Jaffee, 1975; Longstaff Schwartz, 1995); and (viii) interest rate variables—short- and long-term rate, term spread, interest rate volatility and expectations (Duffee, 1998; Fons, 1994; Kim et al., 1993; Mella-Barral Tychon, 1996). We shall see in the following how we work our models with some of these predictors and variables of our choosing. (Gabrielsen, )

More recently, the stringent regulations on risk management resulting from the 2008 crisis also drives the demand for appropriate credit risk models. Alongside stochastic credit risk models, forecasting corporate credit spreads with machine learning based techniques has been a popular research topic. In this specific context, we focus on the credit spreads between average Moody's Baa graded corporate bond yield and U.S. 10-year treasury bond yield. Predictions under random forests, decision tree and boosting algorithms will be presented after careful examination.

The remainder of the paper is structured as follows. In Section 2, we introduce the fundamental methodology for our analysis, namely OLS and machine learning. This is then followed by data preparation, in which data is collected, cleaned and above all, the valuable features are selected using LASSO regression and elastic net. Numerical and graphical results to comparison our models will be presented in Section 4. Python and related libraries were the main source of computational analysis throughout this study.

# 2 Methodology

## 2.1 Benchmark

The spread of yield on corporate bond over government bond results from the default risk of companies. Therefore, we refer to Merton's model (1974) (Gabrielsen, ) which addresses such credit risk and states that

$$E = V_t N(d_1) - Ke^{-r\Delta T} N(d_2) \tag{1}$$

where

$$d_1 = \frac{\ln \frac{V_t}{K} + (r + \frac{\sigma^2}{2})\Delta T}{\sigma \sqrt{\Delta T}} \tag{2}$$

and

$$d_2 = d_1 - \sigma \sqrt{\Delta T} \tag{3}$$

$E$ = theoretical value of a company's equity
$V_t$ = value of company's assets in period t
$K$ = value of company's debt
$r$ = risk free interest rate
$N$ = Cumulative standard normal distribution
$\sigma$ = standard deviation (volatility) of stock returns
$t$ = current time period
$T$ = future time period

Spread between Moody's BAA graded bond yield and the 10-year treasury bond yield is chosen as the subject of prediction. Based on Merton's model, a simple linear regression model is introduced as our benchmark model, regressing credit spread on volatility (VIX index) and risk free rate (10-year treasury bond yield). We strive to improve upon our benchmark in the following.

## 2.2 Step-wise Selection

To build on our benchmark regression model, we first perform a step-wise selection based on VIF (Variance Inflation Factor). The step-wise selection algorithm is then performed again based on AIC (Akaike information criterion) on remaining variables to obtain an efficient model.

## 2.3 Machine Learning

Compared to regular regression, regularized regression models weigh more penalty on large coefficients and therefore is a naturally preferred solution to the collinearity problem.

Lasso regression is a useful tool of feature selection since it tends to drive coefficients close to zero due to the nature of the $L1$ norm. Here we perform Lasso regression as the initial attempt for improvement beyond our ordinary linear regression model.

$$l(\beta) = \|X\beta - y\|_2^2 + \lambda \|W\beta\|_1 \tag{4}$$

Then, we used elastic net which also includes $L2$ penalty compared to Lasso.

$$l(\beta) = \|X\beta - y\|_2^2 + \lambda_1 \|W\beta\|_1 + \lambda_2 \|W\beta\|_2^2 \tag{5}$$

To capture non-linear features, it is necessary to incorporate tree-based machine learning algorithms. There are three underlying reasons for the choice of these particular algorithms. First, our predictors are on various scales. Unlike Lasso regression, tree-based models do not require normalization and hence are suitable for our analysis. Second, it is statistically legitimate to feed stationarized data into tree based models. Another advantage of choosing tree-based model is that these models do not require advanced parameter tuning. (Han, Subrahmanyam, & Zhou, )

As an aside, since we emphasize on the explanatory power of our models, Principal Component Analysis (PCA) will not be carried out, despite the fact that it is generally considered a useful tool in extracting predictive factors.

To compare different machine learning based models, we define our scoring criterion as the average Euclidean distance between prediction results and true values:

$$\frac{\sum_{i=1}^{n} \epsilon^2}{N} \tag{6}$$

where $\epsilon$ is the residual for each point and $N$ is the number of total points.

Furthermore, our data set is partitioned into three subsets: training set (for training, 1985-2016), cross-validation set (for model and parameter selection, 2016-2018) (Amato & Luisi, ) and test set (for testing, 2018-2019). While it is desirable to obtain the model that generates the lowest possible absolute scores, ideally we would want a slightly higher score for cross-validation set than that for training data. Specifically, we wish to avoid the situation of over-fitting where cross-validation scores are significant higher than training scores, as well as under-fitting where we observe high scores for both data sets.

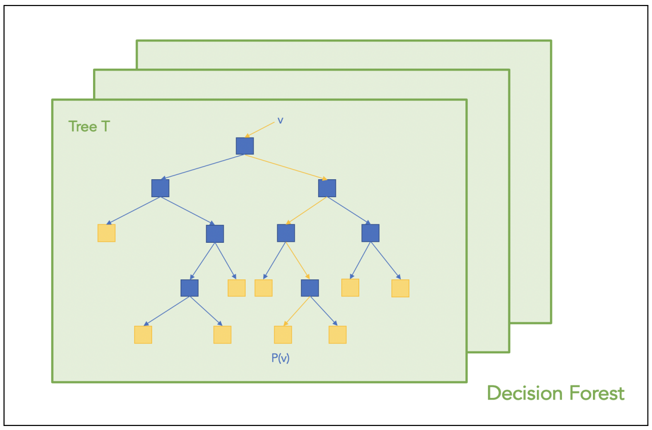


*Figure 2. Random forests or random decision forests are an ensemble learning method for regression by constructing a multitude of decision trees at training time and outputting the class that is the mean prediction (regression) of the individual trees. Source: (Litterman & Iben, )*

## 3 Data Preparation

### 3.1 Data Collection

An important aspect to consider in our study is the spectrum of data types to work with. In this case, 50 features of financial and economic significance were pre-selected for further validation. Examples of variables from the financial sector include stock returns, volatility indices, while economic indicators can be CPI and unemployment rate amongst others, as shown in Table 1 below. In order to minimize idiosyncratic external impact on the predictive power of our models and forecast results, data from 1986 to 2019 was used to avoid the interest rate and inflation rate hike during the late 1970s and early 1980s.

In the following, the data is categorized as credit spread, market specific, and macroeconomic, and the reasons for choosing these data are described in details in the following section. The main data source is Federal Reserve Economic Data.

*Table 1*
*Variables. Data Source: Federal Reserve Economic Data*

| Financial Sector | Variable Example |
|---|---|
| Corporate Bond | Moody's Aaa, Baa |
| Government Bond | 2,5,7,10 Year T Bond |
| Commodity | Gold, Copper |
| Volatility | VIX, VXO |
| Stock | Russell 1000, 2000, Midcap, NASDAQ |
| Real Estate | REIT |
| **Economic Sector** | |
| Unemployment | Unemployment Rate for Each Sector |
| CPI | CPI for Each Sector |
| Credit | Credit for Each Sector |
| GDP | (Real) GDP |
| Debt | Federal, Household Debt |
| Industry | Industrial Production Index, Manufacturing |

#### 3.1.1 Credit Spread.

Yields on the US Treasury bond with 2, 5, 7, 10 years to maturity were collected as the default-free rate to construct our credit spread. The US Treasury bonds were selected based primarily on their individual levels of liquidity and economic effects. The time series of yields on US Treasury bonds are presented in figure 4. As one can see, the yields on US Treasury bonds with different maturities share a rather similar trend. It is also worth pointing out that the 10-year bond yields the highest return while the 2- year bond offers the least, which reflects the liquidity premium.(Bedendo, Cathcart, & El-Jahel, )

As far as risky bond is concerned, we chose yields on Moody's Baa and Aaa rated corporate bonds as representative. Specific to our analysis, the yield on an aggregate credit rating is calculated as the average yield of all companies with the same credit rating. The time series of yields on Moody's Baa and Aaa bonds are presented in figure 3. Evidently, average yield of Baa bonds is higher than that of Aaa bonds, due to higher default risk.

Recall that we defined the credit spread as the difference between the yields of corporate bonds and government bonds. The 10-year Treasury bond is selected as the baseline risk-free bond since it is generally the most susceptible of all to economic conditions and investors' views on investments. When there is an economic downturn, investors switch to 10-year treasury bonds from riskier trading activities for hedging purposes, which in turn leads to an increase in the 10-year yield. In figure 5, we notice a similar trend in the Aaa/10-year Treasury yield spread and the Baa/10-year Treasury yield spread, with the latter being more prominent. To that end, we went ahead with our analysis, quoting the Baa rated corporate yield minus US Treasury yield as the credit spread. In general, credit spreads are wider in times of economic uncertainty and narrower when there is less perceived risk. This fact is also reflected in the credit spread spike occurred during the financial crisis from 2008 to 2010, as shown in the figure.

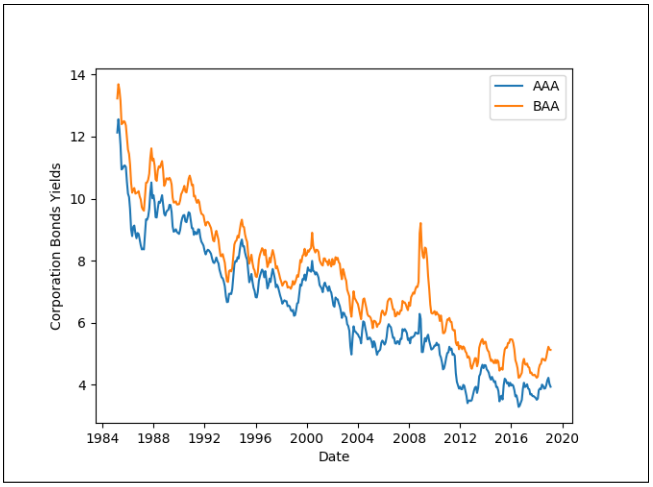


*Figure 3. Corporate bond yield from 1986 to 2019. Yields on Moody's Aaa and Baa grade bonds. Data Source: Federal Reserve Economic Data*

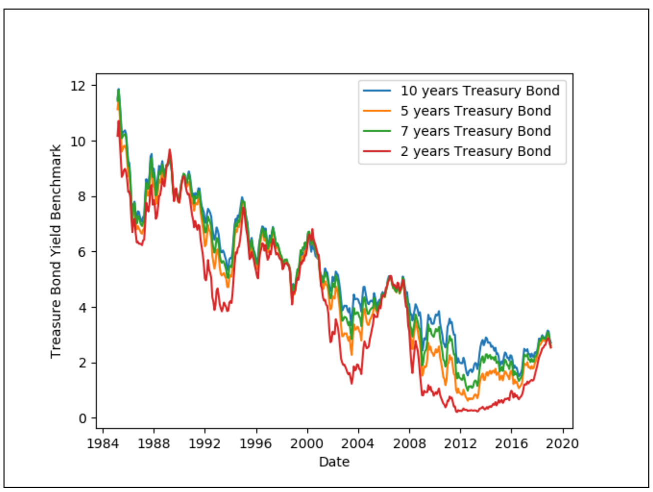


*Figure 4. Treasury bond yield from 1986 to 2019. Yields on 2 year, 5 year, 7 year, 10 year Constant Maturity Treasury bond. Data Source: Federal Reserve Economic Data*

**3.1.2 Financial Factors.** Financial factors are believed to be good predictors of credit spread (Amato & Luisi, ). When investors sense an economic downturn, they act on it right away, causing dramatic changes in the financial markets. For instance, when investors sell off stocks and herd to bonds for return security, stock prices drop as well as bond yields. Our main financial indicators consist of government bond, slope of the yield curve, curvature of the yield curve, term spread, volatility, commodity, stock, and real estate. Examples of features are displayed in table 2. Government bond is a measure of default-free interest rate, slope of yield curve and term spread reflects investors' view about the short and long term economic condition, commodity could be a measure of the inflation and and investors' confidence in the economy, and similarly, stock market can also show investors' confidence in the economy.

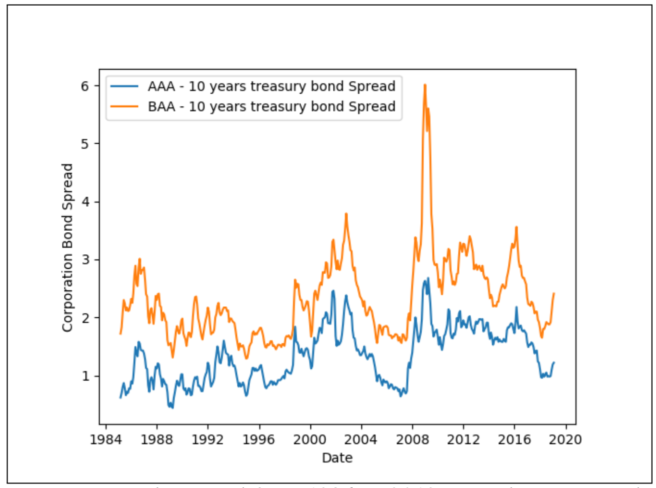


*Figure 5. Credit spread from 1986 to 2019. Moody's Aaa grade bond yield minus 10 year Treasury bond and Baa grade bond yield minus 10 year Treasury bond yield. Data Source: Federal Reserve Economic Data*

*Table 2*
*Financial Factors. Data Source: Federal Reserve Economic Data*

| | |
|---|---|
| Government Bond | 10 Year Treasury Bond |
| Slope of Yield Curve | Zero-coupon Yield Curve |
| Term Spread | 10 Year minus 2 Year Treasury Bond |
| Commodity | Gold, Copper, Corn |
| Volatility | VIX |
| Stock | Russell 1000 |
| Real Estate | REIT |

**3.1.3 Macroeconomic Factors.** On a macro level, credit spread is reflective of economic conditions. When economy falls short, default risk pertaining to corporate bonds increases, giving rise to a higher yield for corporate bonds demanded by investors. On the contrary, government bond yield drops due to higher demand. Credit spread widens as a result of economic slowdown, as seen in figure 5 above. (ANTON SCHOLIN, )

For this reason, we hand-picked some typical economic indicators to forecast credit spread. As can be seen in Table 3, the list contains unemployment rate, inflation rate, gross domestic product, debt, and industrial production, along with some example features from each sector. Unemployment, CPI, and GDP are typical macroeconomic indicators, credit and debt reflects the amount of debts and may indicates creditworthiness, and industrial production is affected by economic condition.

*Table 3*
*Macroeconomic Factors. Data Source: Federal Reserve Economic Data*

| | |
|---|---|
| Unemployment | Civilian Unemployment Rate |
| CPI | CPI for All Urban Consumers |
| Credit | Total Consumer Credit Owned and Securitized |
| GDP | Real Gross Domestic Product |
| Debt | Federal Debt held by Federal Reserve Banks |
| Industry Production | Industrial Production: Manufacturing |

### 3.2 Data Cleaning

The data set is divided into training data and test data. The training data is from 1986 to 2018, and is cleaned using the steps discussed below.

**3.2.1 Interpolation and Sampling.** The next step involves improvements on data consistency, which we achieve using interpolation and sampling due to the limitation of the availability of the original data set. Precisely, our data requires processing in line with monthly frequency in order to forecast the monthly credit spread.

Sampling was applied to data where only daily records are available, such as gold prices in the commodity category. For other variables that are not tracked on a monthly basis, such as quarterly Natural Rate of Unemployment, we implemented linear interpolation. We could improve the interpolation method such as using Brownian bridge; however, due to limit of the scope of our work, this could be considered in future work.

**3.2.2 Stationarity.** As in most financial contexts, our data exhibits time series property. In order to be eligible to be fed into our proposed models, the raw data requires processing. As a consequence, we introduced augmented Dickey–Fuller (ADF) test with p-value of 0.05 as threshold to test whether our return adjusted series are stationary. In statistics, alternative hypothesis of ADF test usually means that the data is stationary. For market and economic data that develops in time such as Russel 2000 Index and Consumer Price Index, returns were computed before entering the models. For those percentage based indicator such as VIX and bond yield, we took the one lag difference and used difference as inputs.

### 3.3 Exploratory Feature Analysis

First round of data cleaning narrowed our working range down to 32 variables.The correlation matrix among these variables is plotted in the figure below. As one can see, the typical positive correlations among stock index, beginning with RU, is expected. Also, there is evident negative correlation between volatitily, VIXCLS and VXOCLS, and stock index, and this is also expected. When volatility increases, market panics and in turn causes selling of stocks.(Chen, Cheng, & Wu, )

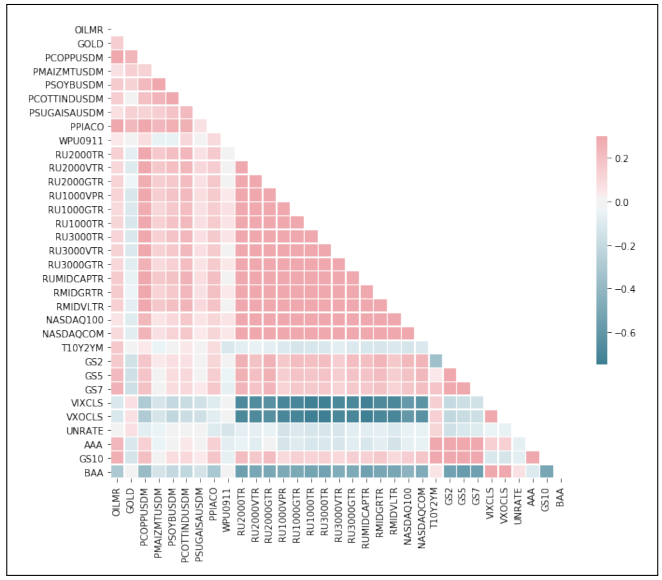


*Figure 6. Heatmap. From top to bottom, OILMAR to WPU0911 are commodity such as oil pice, RU2000TR to NASDAQCOM are stock index, T10Y2YM is term spread 10 year over 2 year government bond yield, GS2,5,7,10 are 2,5,7,10 year Treasury bonds, VIXCSL and VXOCSL are volatility index, UNIRATE is unemployment, and BAA and AAA are Moody's BAA and AAA grade bond yield.*

Below are some typical scatter plots between capturing the relation between our response and predictors.

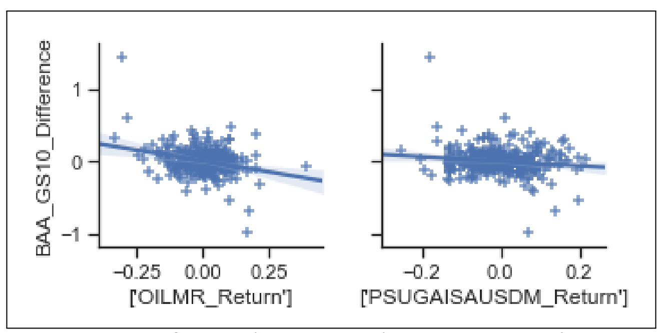


*Figure 7. Significant relation example. OILMAR - oil, PSUGAISAUSDM - sugar.*

For example, as shown above, oil and sugar monthly return appears to exhibit some linear relationship with our response at this stage. We hence keep these variables for further analysis.

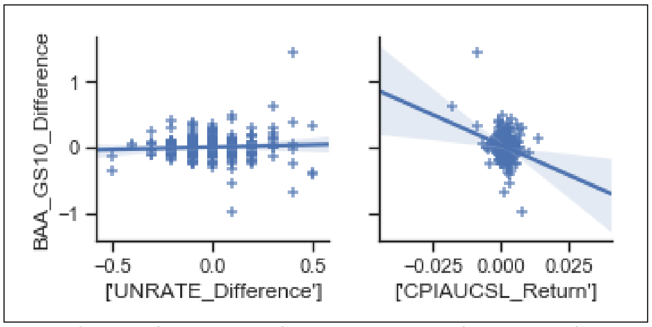


*Figure 8. No relation example. UNRATE - Civilian Unemployment Rate, CPIAUCSL - CPI for All Urban Consumers*

On the contrary, unemployment rate and CPI appear to have little contribution to changes in credit spread. We remove these variables.

## 4 Computational Steps and Results

### 4.1 Benchmark

As mentioned before, we set out on simple regression analysis based on Merton's model with predictor features: VIXCLS (volatility), GS10 (10-year government bonds), T10Y2YM (term spread) and response feature: difference between Baa and GS10 yields as benchmark reference. In the coming analysis, we aim to examine whether the proposed methods outperform our benchmark model.

```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       0.052008   0.019152   2.716  0.02009 *
VIXCLS_Difference 0.002916   0.004540   0.642  0.53382
GS10_Difference  -0.497487   0.148551  -3.349  0.00649 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.07068 on 11 degrees of freedom
Multiple R-squared:  0.5113,    Adjusted R-squared:  0.4224
F-statistic: 5.754 on 2 and 11 DF,  p-value: 0.01949
```

*Figure 9. Benchmark*

It turns out that the the VIX represented volatility is not significant here. So we dropped VIX term, rending another benchmark model with only 1 predictor.

```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      0.05243    0.01867   2.809  0.01578 *
GS10_Difference -0.46472    0.13606  -3.416  0.00512 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06893 on 12 degrees of freedom
Multiple R-squared:  0.4929,    Adjusted R-squared:  0.4507
F-statistic: 11.67 on 1 and 12 DF,  p-value: 0.00512
```

*Figure 10. Benchmark*

Moving forward, we employed a step-wise selection algorithm based on VIF with a threshold of 5 to remove potential collinearity in predictors. Step-wise selection was then

*Table 4*
*Mean Square error of Benchmark model*

| Train MSE | Test MSE |
|---|---|
| 0.0223 | 0.0075 |

applied again to the remaining predictors based on AIC and a linear regression model with R-square 0.7672 (Figure 11) was achieved.

```
Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)          0.017347   0.005323   3.259 0.001240 **
PCOPPUSDM_Return    -0.339898   0.088443  -3.843 0.000147 ***
PSUGAISAUSDM_Return -0.169936   0.062242  -2.730 0.006682 **
PPIACO_Return       -2.115638   0.522895  -4.046 6.55e-05 ***
RU2000VTR_Return     1.187436   0.357017   3.326 0.000984 ***
RU1000VPR_Return     1.849343   0.773449   2.391 0.017382 *
RU1000TR_Return     -3.147924   0.787608  -3.997 7.98e-05 ***
RU3000VTR_Return     1.628300   0.702413   2.318 0.021075 *
RUMIDCAPTR_Return   -5.171389   1.007955  -5.131 5.03e-07 ***
RMIDGRTR_Return      2.679183   0.652122   4.108 5.07e-05 ***
NASDAQ100_Return     1.278618   0.421026   3.037 0.002588 **
NASDAQCOM_Return    -1.350999   0.569344  -2.373 0.018243 *
T10Y2YM_Difference  -0.767014   0.116144  -6.604 1.68e-10 ***
GS2_Difference      -0.654822   0.162431  -4.031 6.95e-05 ***
GS5_Difference      -0.274550   0.132654  -2.070 0.039290 *
AAA_Difference       0.894964   0.060575  14.774  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.08422 on 318 degrees of freedom
Multiple R-squared:  0.7777,    Adjusted R-squared:  0.7672
F-statistic: 74.17 on 15 and 318 DF,  p-value: < 2.2e-16
```

*Figure 11. Step-Wise*

*Table 5*
*Mean Square error of step-wise selected model*

| Train MSE | Test MSE |
|---|---|
| 0.0068 | 0.0060 |

### 4.2 LASSO Regression

Introduction of LASSO regression serves two purposes. First, we would like to refine our research by only selecting significant features. Second, it effectively eliminates multicollinearity between predictors. Data was split in such a way that we ended up with a training set (80%), and a test set (20%). Recognizing that it is a special case of elastic net, we used $l1$ ratio equal to 1 and applied $L1$ norm as penalty measure. Under the assumption of $l1 = 1$, and 5-fold cross validation, regularization parameter was chosen by comparing the fit scores of training set and test set, from a sequence of 60 elements which begins at $10^{-5}$ and ends with 10. Our finding gives the optimal regularization parameter for lasso regression of approximately $2.019 \times 10^{-5}$.

*Table 6*
*Lasso Result. Variable names are described in Figure 6.*

| Feature | coefficient |
|---|---|
| OILMR Return | -0.179838 |
| PCOPPUSDM Return | -0.343926 |
| PSUGAISAUSDM Return | -0.052171 |
| RUMIDCAPTR Return | -0.111703 |
| T10Y2YM Difference | -0.032636 |
| GS5 Difference | -0.223092 |
| VIXCLS Difference | -0.009558 |
| VXOCLS Difference | 0.017253 |
| AAA Difference | 0.781327 |
| GS10 Difference | -0.623206 |

### 4.3 Elastic Net

Subsequent to Lasso regression analysis, we advanced our research by incorporating an elastic net model. Instead of fixing $l1$ ratio as 1, we set up an interval $[0, 1]$ for possible values of the $l1$ ratio. In other words, the penalty is a combination of $L1$ norm and $L2$ norm, where an $l1$ ratio equal to 0 means keeping $L2$ norm as penalty coefficient. However, only part of the set was tested as it is not realistic to loop over every ratio between 0 and 1. Under the assumption of $l1 = [.1, .5, .7, .9, .95, .99, 1]$ and 5-fold cross validation, regularization parameter was chosen by comparing the fit scores of training set and test set, from a sequence of 60 elements which begins at $10^{-5}$ and ends with 10. The optimal regularization parameter turns out to be $2.019 \times e^{-5}$, which is the same as the regularization parameter of the LASSO case when we set $l1$ to be 0.7 as a trade-off between the Ridge and Lasso.

### 4.4 Tree based machine learning model

Finally, we experimented with tree-based machine learning models, including decision tree, random forests and boosting.

Unlike most network-based machine learning tools such as RNN and LSTM, tree-based models are readily tractable since one can visualize model performance by plotting significance of each feature. In addition, the decision tree- based algorithms are unaffected by the scale of the data because each feature is subject to independent analysis and thus data division is independent of the scale. As a result, data preprocessing, e.g. normalization or standardization is not a requirement. These models typically show better performance when there are multiple scales and binary features.

On the other hand, one disadvantage of utilizing such tree-based models could be laborious work on processing nonstationary data sets since extrapolation to explore unknown fields is not easily achievable. Furthermore, over-fitting persists even with pre-pruning or post-pruning, for which we resort to an ensemble model, meaning a mix of models, namely

*Table 7*
*Elastic Net Result. Variable names are described in Figure 6.*

| Feature | coefficient |
|---|---|
| OILMR Return | -0.064110 |
| PCOPPUSDM Return | -0.340542 |
| PSOYBUSDM Return | 0.091073 |
| PCOTTINDUSDM Return | -0.023160 |
| PSUGAISAUSDM Return | -0.152340 |
| PPIACO Return | -1.914882 |
| WPU0911 Return | 0.193763 |
| RU2000VTR Return | 0.719010 |
| RU2000GTR Return | 0.248529 |
| RU1000VPR Return | 1.085843 |
| RU1000TR Return | -1.407746 |
| RU3000TR Return | -0.946958 |
| RU3000VTR Return | 1.678075 |
| RUMIDCAPTR Return | -1.253693 |
| RMIDGRTR Return | 0.526218 |
| RMIDVLTR Return | -1.567084 |
| NASDAQ100 Return | 0.839924 |
| NASDAQCOM Return | -0.810591 |
| T10Y2YM Difference | -0.165744 |
| GS2 Difference | -0.043896 |
| GS5 Difference | -0.298856 |
| VIXCLS Difference | -0.003916 |
| VXOCLS Difference | 0.005034 |
| AAA Difference | 0.868052 |
| GS10 Difference | -0.555685 |

random forest and boosting is adopted instead of a single decision tree. However, most of the drawbacks were tackled in our data preparation step. (Litterman & Iben, )

Random forests or random decision forests is an ensemble learning method for regression by constructing a multitude of decision trees at training time and outputting the class that is the mean prediction (regression) of the individual trees. Random decision forests improve upon decision trees in that the problem with over-fitting can be avoided and parameter tuning can be largely saved. Moreover, here n_jobs is an available operation for computational efficiency. The tuned parameters from cross-validation are max features, which determines the randomness of decision tree and a small max features value would reduce over-fitting, In general, the greater n_estimators value is the better the model performance, leading up to more computational resources consumption. In this case, the optimal result was achieved by selecting the optimal values of n_estimators and max_features and further improvement was seen through parallel computing.(Manzoni, )

In contrast to random forest, gradient boosting is a continuous method to evolve trees, which means trees interact and correct each other simultaneously, yielding a robust prediction model in the form of an ensemble of weak prediction models. It builds the model in a stage-wise fashion like other boosting methods do, and it generalizes them by allowing optimization of an arbitrary differentiable loss function.(Van Landschoot, )

Again, mean squared error was the criterion for model performance. Our data set is partitioned into three subsets: training set (for training, 1985-2016), cross-validation set (for model and parameter selection, 2016-2018) (Amato & Luisi, ) and test set (for testing, 2018-2019). For monthly-level prediction, the interest rate spread between cooperate bonds and US Treasury bonds will increase, with a slowing second-order rate.

As the scores suggested, decision tree is an over-fitting model which fits training set perfectly but not the testing set. In comparison, random forests and boosting models avoid over-fitting in some sense and achieve better performance in the prediction of interest rate spread. In our analysis, most of the trend was captured by observing the mean of prediction variables but the problem with non-stationary variance persists. We could have resolved the issue of time-dependent variances by taking the logarithm. However, it would diminish the explanatory power of features to a great extent. For example, the logarithm of the variance of the Treasury bond yield would be a meaningless factor as the variance is time-dependent. Therefore, we decided to trade predictive ability of our models for explanatory authenticity.

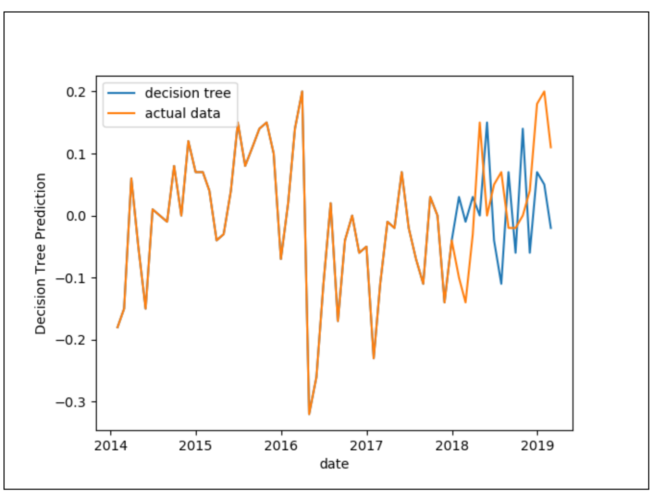


*Figure 12. Decision Tree is an over-fitting model which fits training set perfectly but not in testing set. Testing set is after 2018.*

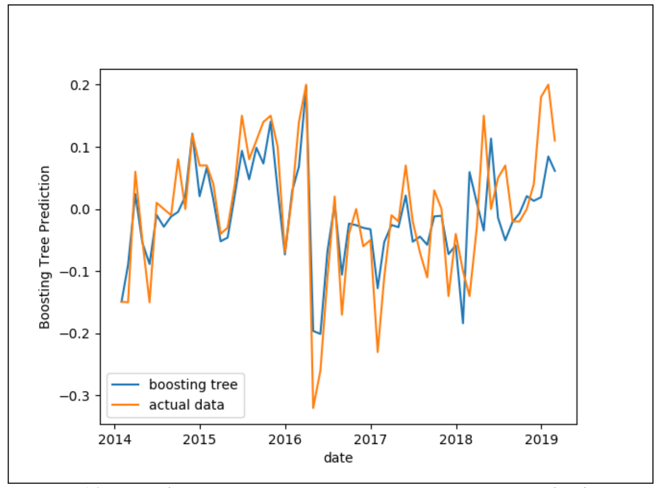


*Figure 13. Gradient Boosting Tree uses continuous method to improve trees, which means each tree would correct the error of other trees and produces a prediction model in the form of an ensemble of weak prediction models. Testing set is after 2018.*

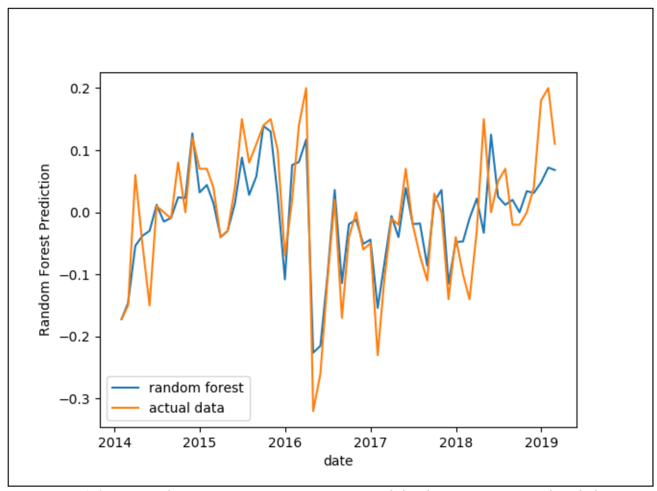


*Figure 14. Random Forest is an ensemble learning method for regression by constructing a multitude of decision trees at training time and outputting the class that is the mean prediction (regression) of the individual trees. Testing set is after 2018.*

### 4.5 Comparison of models

Table 8 summarizes the comparison of model performance. Overall, the non-linear models lead to slightly more accurate results than the linear models. This could be explained by the non-linear relationship between these features and credit spread captured by the complex mechanism of machine learning models. Among the non-linear models, boosting predicts the most accurately.

## 5 Conclusion

Both random forests and gradient boosting methods resolve the over-fitting issue seen in decision trees. In partic-

*Table 8*
*Comparison Between Models*

| Model | Train MSE | Test MSE |
|---|---|---|
| Benchmark | 0.0223 | 0.0075 |
| Stepwise | 0.0068 | 0.0060 |
| Decision Tree | $5.54e^{-32}$ | 0.0059 |
| Random Forest | 0.0039 | 0.0056 |
| Boosting | 0.0021 | 0.0054 |

ular, gradient boosting outperforms the rest after advanced parameter tuning. According to our tree- based machine learning models, the interest rate spread between AAA and BAA rated cooperate bonds and US Treasury bonds will be increasing at a decreasing rate for the monthly-level prediction.

## 6 References


Amato, J. D., Luisi, M. (2005). Macro factors in the term structure of credit spreads ∗..

ANTON SCHOLIN, F. P. M. (2017). On credit spreads: An autoregressive model approach.

Bedendo, M., Cathcart, L., El-Jahel, L. (2004, 11). The shape of the term structure of credit spreads: An empirical investigation.

Chen, R.-r., Cheng, X., Wu, L. (2005, 03). The term structure of credit spreads: Theory and evidence on credit default swaps. *SSRN Electronic Journal*. doi: 10.2139/ssrn.686970

Gabrielsen, A. (2010, July). *Modelling the dynamics of credit spreads of european corporate bond indices.* Retrieved from http://openaccess.city.ac.uk/1116/

Han, B., Subrahmanyam, A., Zhou, Y. (2017). The term structure of credit spreads, firm fundamentals, and expected stock returns. *Journal of Financial Economics*, *124*(1), 147 - 171. Retrieved from http://www.sciencedirect.com/science/article/pii/S0304405X17300028 doi: https://doi.org/10.1016/j.jfineco.2017.01.002

Litterman, R. B., Iben, T. (1991). Corporate bond valuation and the term structure of credit spreads. *The Journal of Portfolio Management*, *17*(3), 52–64. Retrieved from https://jpm.iijournals.com/content/17/3/52 doi: 10.3905/jpm.1991.409331

Manzoni, K. (2002a). Clash of the tightens: History of credit spreads. *Investable Theme*, *11*(2), 183 - 218. Retrieved from http://www.sciencedirect.com/science/article/pii/S1057521902000741 (Special Issue: Credit Derivatives) doi: https://doi.org/10.1016/S1057-5219(02)00074-1

Manzoni, K. (2002b). Modeling credit spreads: An application to the sterling eurobond market. *International Review of Financial Analysis*, *11*(2), 183 - 218. Retrieved from http://www.sciencedirect.com/science/article/pii/S1057521902000741 (Special Issue: Credit Derivatives) doi: https://doi.org/10.1016/S1057-5219(02)00074-1

Van Landschoot, A. (2004). *Determinants of euro term structure of credit spreads*. ECB Working Paper No. 397.