# Forecasting US Federal Funds Rate Using Machine Learning Regression

## Table of Contents

## 1. Abstract

This paper is to forecast the United States Federal Funds Rate, one of the most anticipated and recognizable economic indicators in the financial world, by adopting machine learning strategies. The paper introduces seven machine learning methods: Principal Component Analysis (PCA), Linear Regression, Ridge Regression, Lasso Regression, Elastic Net Regression, Extreme Gradient Boosting (XGBoost), and Random Forest as the tools for forecasting a discrete-valued target variable - the United States Federal Funds Rate. We then further compare, analyse and discuss the results from different methods to draw conclusions on the strategies performed.

## 2. Executive Summary

### 2.1 Federal Funds Rate

The federal funds rate is the target interest rate set by the Federal Open Market Committee (FOMC) under the United States Federal Reserve System (FRS) which aims to promote US economic goals by setting monetary policies. According to the FOMC official website, the federal funds rate refers to the interest rate for which depository institutions, i.e., commercial banks, charge by lending cash to other institutions from their reserve balance on an overnight basis. (Federal Reserve, 2022) The FMOC adjusts the rate based on key economic indicators that reflect the economic performance of the country, such as inflation and recession. (James, 2022) The federal funds rate target has varied widely over the years responding to the prevailing economic conditions. For example, it was set as 20% in the 1980s owing to the high inflation, but 0% to 0.25% in the Great Recession in 2007 to encourage growth. The Federal Reserve raised the target rate to 3.75% to 4% during the November 2022 meeting, marking a sixth consecutive rate hike and pushing borrowing costs to a new high since 2008.

### 2.2 The importance of the Federal Funds Rate

Affecting the monetary and financial conditions of a country, the federal funds rate is one of the most important interest rates in the economy of the United States. which in turn have a bearing on critical aspects of the broader economy including employment, growth, and inflation. A chain of events would be triggered by the changes in the federal funds rate. Not only would foreign exchange rates, other short-term and long-term interest rates be affected, but the economic components including unemployment, output, and prices of goods and services would also be impacted. (Federal Reserve, 2022)

Being one of the biggest players in the world's economy, the United States' economy influences the world in different ways. In situations where the United States interest rates are high, it drives up the value of dollars. It is beneficial to citizens in the States as prices for the import of goods would drop owing to the strong domestic currency, however, the opposite for residents located in a weaker domestic currency country. The citizens from a big importing country and a weaker domestic currency country may experience a high inflation due to the high prices for products from the States. This central bank of the country might raise interest rates so as to counter the inflation, support the domestic currency and prevent it from plunging sharply. (Elvis, 2021) Thus, the federal funds rate poses an indirect impact on people who don't live in the United States, such as the interest rate on property mortgage or car loan.

On top of that, the federal funds rate poses direct impacts on global investors investing in the United States financial market. It is because the interest rate directly affects investments. In general, higher interest rates result in a decline in value for the stock market owing to lower future values than investors anticipate. At the same time, in the bond market, bond prices rise when more money flows into this market, however bond yields will fall since it has an inverse relationship with interest rate. Investors always monitor the State's financial policies which substantially affect the market and the whole economy. The Federal Funds Rate is one of the key factors that investors take references on to determine appropriate and corresponding investment strategies.

In order to have a better picture of the global economy as well as prepare ahead for future financial market movements, it is beneficial to predict federal funds rate. By adopting different machine learning strategies, the forecast and analysis of the federal funds rate could be carried out effectively and efficiently.

## 3. <u>Methodologies – Data Preprocessing</u>

To better predict the federal funds rate, we implement various machine learning methods. Typically, macroeconomic, and financial time series data are non-stationary, which are unpredictable and cannot be modelled or forecasted. The federal funds rate is no exception and thus to get a more accurate and reliable prediction, we perform a transformation to convert our raw non-stationary data, which records the monthly federal funds rate from 1970 to 2020, into stationary data with a constant long-term mean and variance over time before training the model.

### <u>3.1 Data cleaning and transformation</u>

A total of 125 variables from different groups which are the possible factors affecting the US fed funds rate including output and income, labour market, housing, consumption, orders and inventories, money and credit, interest and exchange rates, prices, and stock market are used. Since these variables are time series data, transformation of these data is first used through these five methods: first difference, second difference, logarithm, first difference of logarithms, and second difference of logarithms. Since our aim of the project is to predict the rate on a monthly basis, we take a month of time lag for the features by using *iloc[X_lag = 1:]*, where *X_lag = 1*. For instance, features in May 2015 are used to predict the rate in June 2016.

### <u>3.2 Principal component analysis (PCA)</u>

We clean the data by using *dropna()* before training the model. In this study, we decide to perform PCA to reduce the dimensionality through extracting the “common factors” from the original large set of predictors. Compared with the other two feature selection methods SelectKBest and RFE, in which the hyperparameter is the number of top features to select, we think PCA is the most suitable method for selecting the features based on our raw data and it performs at a low cost of model accuracy. We can see the run time of PCA is much shorter than RFE, thus a higher cost-efficiency to perform.

### 3.3 Rolling window

To split the sample into a training and test data set, the rolling window is used to train the model to predict the next time step of the federal funds rate. We suspect there have been structural changes in predictors affecting the federal funds rate through these decades. For instance, the financial market and the rates space have been tremendously impacted by the 2008 financial crisis when the unconventional monetary policy - negative interest rate was adopted for the first time. Compared to the Great Moderation, the Fed has been intentionally keeping the rate at a much lower level since 2008 to stimulate the economy. Until recently, the jumbo rate hikes have been delivered to combat the highest inflation in 40 years. Clearly, factors affecting the federal funds rate have been fundamentally changing, which causes us to use the rolling window. The prediction is stored or evaluated against the known rate. Then, the window is modified by dropping the most distant sample and adding the most recent sample and the process is repeated. We decide to use both 80% and 70% of the data sample as the size of the rolling window, thus verifying the robustness of the model and observing the best fitted one.



## 4. Methodologies – Data Analysis

### 4.1 Grid Search for optimal hyperparameters

Instead of randomly choosing the number of principal components, grid search is used to figure out the optimal number by using the hyperparameter defined in PCA. The RMSE is the lowest when the number of principal components is 30 and 11 when the train set is 80% of the original data and 70% of the original data respectively. It reflects the lowest validation error.

While PCA is the major hyperparameter in Lasso, Ridge and Elastic regression, there are more hyperparameters in Random Forest and Extreme Gradient Boosting. Besides the principal components, the number of trees in the forest [100, 200, 300] should be considered before building the Random Forest model. The optimal number of trees we found was still 100, which validated the fact that increasing number of trees have no correlation with bettering the performance of the forest model (Oshiro, T.M., Perez et al, 2012). The principal component was 6. For Extreme Gradient Boosting, we primarily found the optimal learning rate and principal components, which were 0.2 and 9 respectively. Learning rate is important here because it determines the size of update, hence the performance of learning the previous predictors.

### 4.2 Pipeline

We use pipeline chain for standardisation, PCA, and the regression estimators. It is a convenient way which enables us to predict once on the data to fit a whole sequence of estimators as we are using various machine learning methods to train the model. At the same time, pipeline helps to avoid leaking statistics from test data into the trained model to ensure that the same samples from our data are used to train the transformers and predictors to enhance accuracy.

### 4.3 Linear, Ridge, Lasso and Elastic Net Regression Machine Learning Methods

Apart from vanilla linear regression, we decide to implement other regularization methods including ridge regression, lasso regression, and elastic net to reduce overfitting issues. Ridge regression adds a regularization term that penalizes large weights and lasso regression adds another regularization term that adds a penalty to the absolute value of the magnitude of coefficients. Elastic net is simply a mix of both ridge and lasso. The parameters for ridge, lasso are alpha = 1000 and alpha = 30000 respectively. The parameters for elastic net are alpha = 100 and L1 ratio = 0.5. To compare all regularizations, we decide to use MAE and RMSE as the performance standards.

### 4.4 Random Forest and Extreme Gradient Boosting

Besides regressions, ensemble methods are the alternative algorithms by aggregating various weak learners to achieve a stronger learner which can be harnessed to handle regression and classification tasks. The random forest model randomly samples features and subsets of the training dataset, then decorrelates the weak learners. With various decision trees possessing random features and subsets, predictions are made using bootstrap aggregation. A lower variance estimator than decision trees is expected, and each decision tree has less correlation with one another. Apart from bagging, the boosting algorithm is also one of the popular ensemble methods. To further reduce overfitting, extreme gradient boosting is used in our attempts. The fundamental gradient boosting method trains predictors sequentially and correct its previous predictor every time to achieve the results. Extreme gradient boosting is built on this concept with additional regularization terms to penalize models with high complexity. To reduce the run time of searching for the optimal hyperparameters, we set the number of trees to be 100 by default setting. Also, we set the L2 regularization term (lambda) and L1 regularization term (alpha) to be 1 and 0.5 to make the model more conservative. For consistency, we compare the results with the three regressions by using MAE and RMSE.

## 5. Result Analysis

Based on Machine Learning methods of Principal Component Analysis, Linear Regression, Ridge Regression, Lasso Regression, Elastic Net Regression, Extreme Gradient Boosting and Random Forest, we concluded the following result analysis.

## 5.1 Result Analysis of Linear, Ridge, Lasso and Elastic Net Regression

Firstly, we use 80% of the total data as the training set (i.e. 477) with PCA and standardisation in the pipeline to train the model. The table below shows the final result of regression models:

| | MAE | RMSE |
|---|---|---|
| **lin_reg** | 0.144021 | 0.220194 |
| **Ridge** | 0.121308 | 0.301486 |
| **Lasso** | 0.048979 | 0.119687 |
| **Elastic Net** | 0.048979 | 0.119687 |

It is shown that both Lasso Regression and Elastic Net Regression model obtained the lowest mean absolute error (MAE) and root mean square error (RMSE). It reflects that with a regularisation included in the regression model can help us get a better performance of the model by reducing the validation error.

Secondly, we try to set the training set proportion to be 70% of the total data (i.e 417) with PCA and standardisation in the pipeline to train the model to observe the differences. The table below shows the final result of regression models.

| | MAE | RMSE |
|---|---|---|
| **Linear** | 0.121518 | 0.186553 |
| **Ridge** | 0.099905 | 0.187317 |
| **Lasso** | 0.063514 | 0.121844 |
| **Elastic Net** | 0.063514 | 0.121844 |

From the result, the Lasso Regression and Elastic Net regression still performs the best under a different ratio of the train-test set split. This shows the robustness of the model as different ratios of the train-test set split show similar results. At the same time, we also observe that the RMSE for both Linear and Ridge Regression reduces significantly.

## 5.2 Result Analysis of Random Forest and Extreme Gradient Boosting

We use 80% of the total data to train the model with the two ensemble methods, Random Forest and Extreme Gradient Boosting.

The result reveals that Extreme Gradient Boosting achieves a slightly better performance than Random Forest because of its capability of resolving the overfitting issues.

| | MAE | MSE | RMSE |
|---|---|---|---|
| Random Forest | 0.079650 | 0.029809 | 0.172655 |
| Extreme Gradient Boosting | 0.102028 | 0.029251 | 0.171029 |



## 6. Conclusion

| | RMSE | Running Time |
|---|---|---|
| Linear Regression | 0.220194 | 3.334002 |
| Ridge | 0.301486 | 3.221865 |
| Lasso | 0.119687 | 3.420271 |
| Elastic Net | 0.119687 | 3.176772 |
| XGBoost | 0.171029 | 12.851920 |
| Random Forest | 0.172655 | 50.908632 |

After adopting the above six machine learning methods, we find that Elastic Net is the best forecasting strategy, with the lowest RMSE.

To begin with, although Linear Regression is the simplest and easiest model to implement, it tends to overfit the training data when more features are added, with a low variance but a high bias. The problem of overfitting will lead to a low training error but high test error, resulting in poor generalization of the model. As generalization is the goal of machine learning, we need to make a trade-off between bias and variance, so that our model can perform well on new, unseen data. To constrain our models by using regularization, we implement Ridge Regression, Lasso Regression, and Elastic Net Regression.

From the results, we can conclude that Elastic Net and Lasso outperform Ridge Regression, with lower RMSE. It may be because Ridge does not reduce the number of variables in a model even if they are irrelevant, so it will not perform well in feature reduction. (Feline, 2022) When it comes to Elastic Net and Lasso, although they produce the same RMSE, Elastic Net is the most preferred model as Lasso does not perform well with multicollinearity. For example, if variables are highly correlated, Lasso will randomly select one without understanding the context. However, it may eliminate relevant features, negatively affecting our interpretation of the data. (Vijay, 2018) Since Elastic Net combines characteristics of both Ridge and Lasso, it will not randomly drop variables with high collinearity, and thus is a better model.

By comparing Random Forest and Extreme Gradient Boosting (XGBoost), Random Forest decorrelates weak learners by randomly sampling features and subsets of a training dataset. On the other hand, XGBoost trains predictors sequentially, each one trying to correct its previous predictor. Besides, it contains a regularization term to control the complexity of the model, mitigating the problem of overfitting.

From the results, XGBoost outperforms Random Forest with a lower RMSE and significantly lower running time. Although both are decision tree algorithms, the training data is processed differently. XGBoost is better than Random Forest in terms of running time and accuracy.

The calculation in Random Forest takes a much longer time than in XGBoost. It may be due to a large number of trees in Random Forest, making the algorithm so slow. Furthermore, the results of XGBoost are more accurate than that of Random Forest. Random Forest constructs the trees and trains the data independently, whereas XGBoost trains each data sequentially, correcting errors of previous predictors each time. (Milos, 2022) Hence, XGBoost is a better model.

Of all the models, Ridge Regression, Lasso Regression, and Elastic Net Regression generated lower RMSE and much shorter running time than XGBoost and Random Forest. It may be because XGBoost and Random Forest tend to fit non-linear associations, but the relationships of our data tend to be linear. Therefore, we can simply transform the data to linearity and adopt a regularized linear regression model. Based on the results, Elastic Net Regression is the best predictor of federal funds rate.

## 7. Future Work and Recommended Actions

Through this model, players in the financial sector can predict the trend of the federal funds rate and make corresponding moves to strategically capture the opportunity.

For banks, especially commercial banks, most of the profits are attributed to the net interest income (NII), which is the difference between the revenue generated from a bank's interest-bearing assets and the expenses associated with paying on its interest-bearing liabilities (Alicia, 2022). Typically, commercial banks take customers' deposits as liabilities and use them to provide loans and mortgages to retail customers and corporations. A higher rate will lead to a higher NII for banks, which are reflected in recent earnings of these corporations such as JPMorgan Chase, Bank of America Corp. in the United States, and HSBC Holdings in Hong Kong. The interest rates are usually determined by the London Interbank Offer Rate (LIBOR) or the newly used Secured Overnight Financing Rate (SOFR), which are usually positively affected by the federal funds rate. Therefore, banks can expect a higher profit when they predict a higher rate which incentivizes them to lend more and fully utilize their excess reserves.

On the other hand, the federal funds rate can also be regarded as a leading indicator of the financial markets for other players to make investment or trading decisions. Historically, the stock market tends to fall in each rate reduction cycle and rise in each rate hike cycle since 1995, apart from the rate hike cycle that we are experiencing. For instance, when the federal funds rate decreased drastically in September 2007, the S&P 500 index plummeted 42.65% until the rate reduction stopped in December 2008. Traders or portfolio management can therefore foresee a risk-on or risk-off market based on the prediction of the federal funds rate, thus making profitable decisions. They could sell risky assets including cryptocurrencies, real estates, equities when they foresee a higher rate in the future and manage their positions and portfolio according to their view based on the prediction of the federal funds rate.

Therefore, the model is a useful tool to facilitate decision-making in the finance industry and implement corresponding strategies to maximize one's profit.

## 8. References


1. Federal Reserve (2022) Federal Open Market Committee - About the FOMC. Retrieved from: https://www.federalreserve.gov/monetarypolicy/fomc.htm
2. James Chen (2022) Investopedia - Federal Funds Rate: What It Is, How It's Determined, and Why It's Important. Retrieved from: https://www.investopedia.com/terms/f/federalfundsrate.asp
3. Elvis Picardo (2021) Investopedia - Effects of Exchange Rate. Retrieved from: https://www.investopedia.com/articles/forex/053115/understand-indirect-effects-exchange-rates.asp
4. Feline Wang (2022) Lasso vs Ridge vs elastic net: ML. GeeksforGeeks. Retrieved from: https://www.geeksforgeeks.org/lasso-vs-ridge-vs-elastic-net-ml/
5. Vijay Swamy (2018). Lasso versus Ridge versus elastic net. Medium. Retrieved from: https://medium.com/@vijay.swamy1/lasso-versus-ridge-versus-elastic-net-1d57cfc64b58
6. Milos Simic (2022). Gradient boosting trees vs. random forests. Baeldung on Computer Science. Retrieved from: https://www.baeldung.com/cs/gradient-boosting-trees-vs-random-forests
7. Oshiro, T. M., Perez, P. S., & Baranauskas, J. A. (2012, July). How many trees in a random forest?. In International workshop on machine learning and data mining in pattern recognition (pp. 154-168). Springer, Berlin, Heidelberg.
8. Tuovila, A. (2022, May 17). *Net interest income: What it is, how it's calculated, examples*. Investopedia. Retrieved from https://www.investopedia.com/terms/n/net-interest-income.asp