LUISS

Dipartimento di Economia e Finanza

Cattedra di Asset Pricing

Predictive Modelling of FOMC Decisions and Market Reaction: A Machine Learning Approach

Prof. Borri N.

RELATORE

Prof. Patnaik M.

CORRELATORE

Arturo Marucci
Matr. 745851

CANDIDATO

Anno Accademico 2022/2023

Abstract

This thesis focuses on two primary objectives: developing models to predict Federal Open Market Committee (FOMC) decisions regarding changes in the federal funds target rate and analyzing the stock market reaction to these decisions. Traditional econometric models have been extensively studied in the literature, but this research employs innovative machine learning algorithms to enhance predictive performance. Seven different machine learning models are developed. The results demonstrate the superiority of the machine learning models in terms of accuracy, log-score, and quadratic-score, with the optimal random forest model performing the best. Furthermore, the study employs variable importance methodology and partial dependence plots to identify the macroeconomic variables that significantly influence FOMC decisions. The second part of the thesis focuses on the impact of FOMC decisions on the stock market, distinguishing between expected and unexpected choices. Event study analysis is conducted using S&P 500 data, and regression analysis of CARs demonstrates a significant market response to unexpected changes in the federal funds target rate. Finally, the practical validity of the predictive models is confirmed through a backtesting strategy.

Table of Contents

List of Tables

List of Figures

1. Introduction

Decisions regarding Federal funds target rate have significant implications for the economy, including employment, growth, and inflation. They indirectly affect various short-term interest rates, such as those for loans, as lenders often base their rates on the prime lending rate, which is influenced by the fed funds rate. Changes in the target rate can have a strong impact on the stock market, with even a small decline leading to increased market activity and lower borrowing costs for companies. Therefore, anticipating FOMC decisions is crucial for market participants to adjust their investment strategies. Morover, examining the impact of these decisions on security prices is a topic of great interest to investors and policymakers since it can provide additional insights into the ways in which monetary policy affects financial markets and market expectations. This thesis is situated within this context, encompassing two primary objectives: firstly, to develop models capable of predicting FOMC decisions regarding changes in the Federal funds target rate, and secondly, to analyze the impact of these policy decisions on the stock market by considering market expectations and thus differentiating between expected and unexpected decisions.

Both topics have been extensively studied in the academic literature. Regarding the forecasting of target rate, we first have the seminal work of Taylor (1993) who proposed a policy rule for forecasting Federal fund rate. However, this approach does not account for the discrete nature of FOMC decision-making. Subsequent work, such as that of Hu and Phillips (2004), addressed this limitation by developing a discrete choice approach to predict policy actions. They replicated Dueker's (1999) work, but overcame his problem of nonstationarity of covariates. Their model estimated thresholds that trigger changes in the target rate by the FOMC. These threshold coefficients indicate the size of the gaps between the estimated and actual target rate needed to prompt target adjustments of varying magnitudes. While Hu and Phillips (2004) demonstrated good predictive performance within their estimation sample, they did not provide an estimate of the out-of-sample performance. Pauwels (2012) replicated their model and tested its out-of-sample performance, comparing it with other models developed using combined forecast methodology and introducing metrics like log-score and quadratic-score. Another relevant study by Vasnev (2013) replicated Pauwels' work but focused on the policy actions of the Reserve Bank of Australia.

On the other hand, regarding the academic literature related to the impact of changes in the target federal funds rate on asset prices, one of the most relevant papers is that of Kuttner

(2001). In his research, Kuttner estimated the impact of monetary policy actions on the bond market by distinguishing between expected and unexpected decisions through federal funds futures market data. His results showed a strong response to unanticipated rate changes and a weaker response to anticipated changes. Another important subsequent study is the work of Bernanke and Kuttner (2005), where they analysed the impact of FOMC decisions on the stock market employing Kuttner's (2001) approach to differentiate between the anticipated and unanticipated components of policy decisions. They also discovered evidence supporting a greater stock market reaction to unexpected decisions.

My contribution to the academic literature is twofold. Firstly, I aim to apply innovative techniques, specifically machine learning algorithms, to predict policy choices regarding the Federal funds target rate. I seek to demonstrate that these methodologies exhibit superior predictive performance compared to traditional models. To achieve this, I develop seven different machine learning models, including logit model, support vector machine, decision tree, pruned tree, bagged tree, random forest, and optimal random forest. The dependent variable in my models represents the decision on target rate, with three possible values: hike, no change, and cut. On the other hand, the independent variables encompass various macroeconomic indicators and a variable related to the previous period's target rate decision. These indicators are preprocessed, resulting in a total of 56 regressors. Through a feature selection methodology, I filter these regressors and ultimately utilize 23 in my models. The data primarily come from the Federal Reserve Economic Data (FRED) online database of the Federal Reserve Bank of St. Louis. After developing the models, I compare their performance using three main metrics (accuracy, log-score, and quadratic-score) which are computed employing a 5-fold cross-validation approach.

The optimal random forest emerges as the best model, demonstrating superior performance across all three metrics and surpassing traditional models found in the academic literature. Subsequently, I conduct a detailed analysis of my models' predictive capabilities, examining their accuracy in forecasting each class of the dependent variable. My findings reveal the highest accuracy in predicting no-change rate decisions, while also demonstrating good prediction abilities for the other two classes. This outcome is primarily attributed to the nature of the problem and so of the dataset. Furthermore, I utilize the optimal random forest model to estimate variable importance and partial dependence plots, aiming to identify the variables that contribute most significantly to FOMC policy decisions. The results indicate that prior policy decisions, manufacturing industry trends, consumer behaviour and spending patterns, fluctuations in the bank prime loan rate, and levels of unemployment benefit claims are the variables likely to exert the greatest influence on the FOMC's policy decisions.

Secondly, I aim to analyze the impact of policy decisions on the stock market, distinguishing between expected and unexpected choices. To achieve this, I employ the classic methodology of event study. I utilize daily data from the S&P 500, obtained from WRDS (Wharton Research Data Services). I consider the Federal funds rate decisions as events and designate the dates of FOMC meetings as event dates. Through event study analysis, I examine the Cumulative Abnormal Returns (CAR) of the S&P 500 to assess the impact of these decisions. I perform a regression analysis of CARs, where I include two dummy variables that identify decisions representing positive or negative surprises compared to expectations.

These variables are constructed using the prediction errors from my previously developed machine learning models. This represents the main innovation I intend to introduce to the academic literature—an alternative method for distinguishing between expected and unexpected policy actions. In other words, I aim to verify that my models can incorporate market expectations. The obtained results support my hypothesis, as we observe a positive and significant coefficient for the variable capturing positive surprises during rate-cut decisions and a negative and significant coefficient for negative surprises during rate-hike decisions. This suggests that the market reacts significantly to unexpected changes in the fed funds target rate.

Finally, as a final task, I develop a backtesting strategy to assess the practical validity of my models. This strategy is based on the predictions of my best-performing model and is implemented in both the stock and bond markets. My findings indicate that the strategy generates higher returns compared to a simple buy-and-hold approach, thus confirming the practical validity of my models.

The rest of the thesis is organized as follows. Firstly, I conduct a comprehensive literature review encompassing both the prediction of FOMC decisions regarding the Federal funds target rate and the analysis of the impact of these decisions on asset prices. Subsequently, I present the data utilized in this study and describe the preprocessing methods employed. The following chapter details the hypotheses I aim to test. I then introduce the methodologies utilized for testing these hypotheses, providing an overview of machine learning models and the event study technique. Subsequently, I present the outcomes of my research. Lastly, I draw conclusions based on the obtained results, discussing their relevance to my hypotheses and research questions.

2. Literature Review

2.1 Predicting FOMC decision on Federal fund target rate

Over the past several decades, monetary policy analysts and market participants have increased their attention to the study of the Fed's interest rate policy. The Federal funds rate adjustment process occurs in a discrete manner, both from the point of view of timing, since most of the decisions occur on pre-scheduled meeting days, and from the point of view of the magnitude of the adjustment, since the target rate is changed in multiples of 25 basis points (bp). Despite this, the traditional literature on monetary economics has focused on the development of macroeconomic models for continuous estimate of an "optimal" interest rate for monetary policy. An early striking example is provided by the relevant work of John Taylor, who in 1993 proposed a policy rule tying the change in the Federal fund rate to inflation and economic growth. Specifically, this equation, which would be named Taylor's Rule, expresses the "optimal" federal fund rate as a function of both the gap between the current and desired level of inflation and the gap between current and potential output. The first version of the Taylor's Rule takes the following form assuming a long-run inflation target of 2 percent:

$$r_t^* = \pi_t + 0.5(z_t - z_t^*) + 0.5(\pi_t - 2) + 2$$

where r_t^* refers to the optimal interest rate, π_t to the actual inflation rate, z_t to a measure of the actual output, z_t^* to a measure of the potential output. The key empirical lesson derived from Taylor's rule concern the coefficient associated with inflation, specifically its absolute value greater than one. This means that if inflation rises by one unit the Federal Reserve will raise fed funds by an amount greater than one, and this would lead to an increase in the real interest rate, thereby cushioning inflationary pressures in the economy. However, the Taylor Rule has one obvious limitation: it ignores the discreteness of the policy making process. Predicting the dynamics of FOMC intervention requires a model that takes into account the discrete timing of rate change and the discrete amount of rate change. For this reason, recent literature has focused mainly on the discrete nature of FOMC practices.

One of the first relevant attempts was made by Dueker (1999), who proposed an econometric model, a conditional ordered probit based on stationary data, that examines discrete adjustments in the target Federal funds rate. Its objective is to estimate the thresholds that trigger changes in the target rate by FOMC, as well as the extent of those changes. The threshold coefficients indicate the magnitude of gaps between the optimal and actual target funds rate required to trigger target adjustments of varying sizes. These estimated thresholds are then compared to the actual changes in the target fed funds rate to provide a quantitative measure of the Federal Reserve's willingness or reluctance to initiate changes in the target rate. Dueker (1999) defined five categories based on $\mu,$ a vector of four threshold coefficients. Depending on how the latent desired change in the target fed fund rate ranks in (μ_{j-1}, μ_j), it will belong to a specific category j. Dueker estimated these thresholds coefficients through the Maximum Likelihood technique and obtained results suggesting that to trigger a 25 basis-point increase or decrease in the target rate, the latent target funds rate needs to be approximately 45 basis-points above or below the actual target. These results provide substantial evidence that the discrete nature of the target federal funds rate introduce a level of sluggishness into FOMC policy's response to inflation and output gaps.

However, the Dueker's study suffers a single serious drawback: he assumes that the covariates of his model are stationary, but the explanatory variables consist of aggregate macroeconomics timeseries that are therefore likely to be nonstationary. As a result, the model may exhibit misleading standard errors causing inappropriate statistical inference concerning the covariates. Hu and Phillips (2004) provided a solution to this problem by proposing a nonstationary triple choice model. They replicated the work of Dueker (1999) but providing only two threshold coefficients, μ_{1T} and μ_{2T}. In this way, FOMC policy actions are classified into three categories: “hike”, “no change”, “cut” the interest rate. A difference between this period's optimal interest rate (r^*_{t+1}) and the last period's interest rate (r_t) less than μ_{1T} would indicate that interest rates should be lowered, a difference greater than μ_{2T}would indicate that interest rates should be raised, while a difference between μ_{1T} and μ_{2T}would indicate that there should be no change. Again, the model is assumed to be an ordered probit and is estimated by Maximum Likelihood.

In addition, Hu and Phillips (2004) show that when the explanatory variables are nonstationary the asymptotic of the ML estimators would be different from its stationary counterpart and in particular the threshold values are sample size dependent. Hu and Phillip propose to scale the thresholds by the sample size so that they have the same order of magnitude as the latent variable. In this way the correct standard errors can be obtained. Their study examined the rate-setting behavior of the Federal Reserve using a dataset spanning 8 years, from January 1994 to December 2001, which yielded a total of 64 observations. They employed a simplistic naïve approach, based on the model estimation process, to decide which variables to include in the model. In their initial estimation, they included 11 macroeconomic series and used these covariates in estimating a trichotomous probit model, as mentioned earlier. Through a step-by-step approach, they retained only the variables that exhibited statistical significance in the first stage, namely M2 growth, unemployment claims, consumer confidence, and new orders. Using these variables they re-estimated the model, obtaining as results statistically significant but slightly asymmetric threshold estimates: the estimated threshold for a rate cut is 94 bp while for a rate hike it is 107 bp, thus showing greater weakness for the rate cut. Regarding the goodness of fit, their model accurately predicted 78% (50 out of 64) of the Federal Reserve's federal funds rate decisions within the estimation sample.

Hu and Phillips' work was later criticized by Kim et al. (2009) in two main aspects. Firstly, their approach to model selection is purely empirical and lacks a strong foundation in macroeconomic theory. This leaves it vulnerable to the idiosyncrasies of the dataset used, as different or expanded samples may yield different "best" model specifications. Secondly, they solely evaluate their model's performance based on its ability to predict outcomes within their sample, thus providing no evidence regarding the model's predictive out-of-sample performance.

Based on this critique later papers have extended the Hu and Phillips (2004) analysis. Pauwels (2012) provided a methodology to combine multiple forecasts from discrete choice models for the policy decision on Federal fund target rate. The primary goal of this technique is to improve forecast accuracy. Instead of combining the forecasts with a simple average, Pauwels proposed two intuitive weighting methods based on scoring rules, the log-score and quadratic score rules. He also used these scoring rules as a method for evaluating the performance of discrete models. The log-score and quadratic-score provide a more informative measure than accuracy because it accounts for probabilities, penalizes incorrect predictions, and assesses model calibration. Pauwels tested the out-of-sample predictive performance of these models over 3 different time intervals using a recursive forecasting scenario. He also replicated this test for the models of Hu and Phillips (2004) in order to compare the results. His quadratic weights model tends to have a better accuracy than the H&P models in 2 out of 3 of the time intervals studied, however, the best accuracy (71.9%) is presented by the 4-variable H&P model over the period from 1994-2001 (properly the period covered by the work of Hu and Phillips (2004)). Even in terms of evaluation with scoring rules, the quadratic weighted combination model outperforms the other models most of the time. So Pauwels empirical findings indicate that the combination of forecasted probabilities through the application of quadratic and logarithmic scoring methods generally outperforms both equal weighting of forecasts and predictions derived from multivariate models.

Finally, another relevant paper was produced by Vasnev (2013), who replicated the work done by Pauwels (2012) to predict the monetary policy decisions of the Reserve Bank of Australia (RBA) over a 17-year period from 1993 to 2010. Again, he used forecast combination models and compared them with models developed by Hu and Phillips (2004), evaluating performance through both accuracy and scoring rules. Vasnev obtained very similar results to Pauwels (2011): across all time frames the best model in terms of out-of-sample performance turns out to be the quadratic score weighted model.

2.2 The impact of policy changes in target Federal funds rate on asset prices

The literature aimed at analyzing the impact of FOMC decisions about target interest rate on asset prices is quite extensive. An early example is provided by the paper by Cook and Hahn (1989), who attempted to measure the effect of changing the federal fund target rate on Treasury yields. They performed a regression of the change in the bill or bond rate the day of the FOMC decision on the change in the target funds rate. Running this regression for bond rates at different maturities, they found that an increase in the target rate corresponds to a positive and large movement in short-term interest rates, and a moderate but significant movement in intermediate- and long-term rates. Specifically, a one percentage point increase in the target fund rate causes a 50-basis point increase in bill rates and a 10-basis point increase in 20-year bond rates, with both coefficients significant at the 1% level. These findings corroborate the commonly held belief among participants in financial markets that the Federal Reserve exerts significant control over market interest rates by managing the funds rate.

Nonetheless, there is a potential limitation in using fluctuations in the Federal funds target rate as an explanatory factor, as it fails to account for the forward-looking nature of financial markets. Consequently, the expected element of monetary policy decisions should not have a substantial impact on asset prices. To address this issue, Kuttner (2001) makes use of Fed fund futures rates, allowing for the differentiation of changes in the target funds rate into anticipated and unanticipated components. In particular, the surprise element of any change in the target fund rates can be determined by examining the difference in the rate implied by the current-month futures contract compared to the rate on the day prior to the policy decision. At this point Kuttnner (2001) wanted to inquire whether there are any discrepancies in the reactions of bill and bond rates to these two components. Following Cook and Hahn's style of analysis, he regressed the change in the bond rate for different maturities on the expected and unexpected components. He obtained coefficients that are very different: the response to the surprise changes is large and significant; while the response to the anticipated component is small and statistically insignificant, thus being consistent with the expectations hypothesis of the term structure. Overall, the response of interest rates to surprising changes in the target is considerably higher than the response to "raw" changes. This is evident for short-term securities, but even more for long-term securities.

Later Bernanke and Kuttner (2005) analyzed the impact of FOM decisions on stock prices. They applied the event-study approach previously developed by Kuttner (2001), distinguishing between expected and unexpected policy action through the use of Federal fund futures data. Therefore, they regressed the CRSP value-weighted return on the anticipated and unanticipated components of the policy decision. As a result, they obtained that the stock market response to a surprise change is negative and significant. Specifically, a 25-basis-point rate cut would lead to a 1-day stocks return of 1%. However, this method of estimates of the reaction of asset prices to changes in target rate suffers a weakness; the event-study results rely on the assumption that the error term and fund rate changes are orthogonal. But there are at least two cases where this assumption is violated. First, there could be a contemporaneous effect of the stock prices movement on short-term rates. Second, monetary policy and the stock market could respond simultaneously to new information, such as macroeconomic news. In both cases, there could be bias in the estimation of the impact of monetary policy. Subsequent studies have attempted to address these endogeneity and joint-response issues: Gurkaynak, Sack and Swanson (2004) used intraday, rather than daily, data to estimate the impact of FOMC decisions on asset prices, in this way they isolated the impact of the policy change from the influence of other news occurring before or after the announcement; Rigobon and Sack (2002) proposed an estimator that produces consistent estimates of market response by exploiting the heteroskedasticity introduced by exogenous monetary policy actions. In the end, both studies present very similar results to those obtained by Bernanke and Kuttnner (2005) using the event-study technique. Consequently, "it seems reasonable to proceed with the event-study approach, while recognizing that it may provide slightly conservative estimates of the stock market's response to monetary policy"[1].

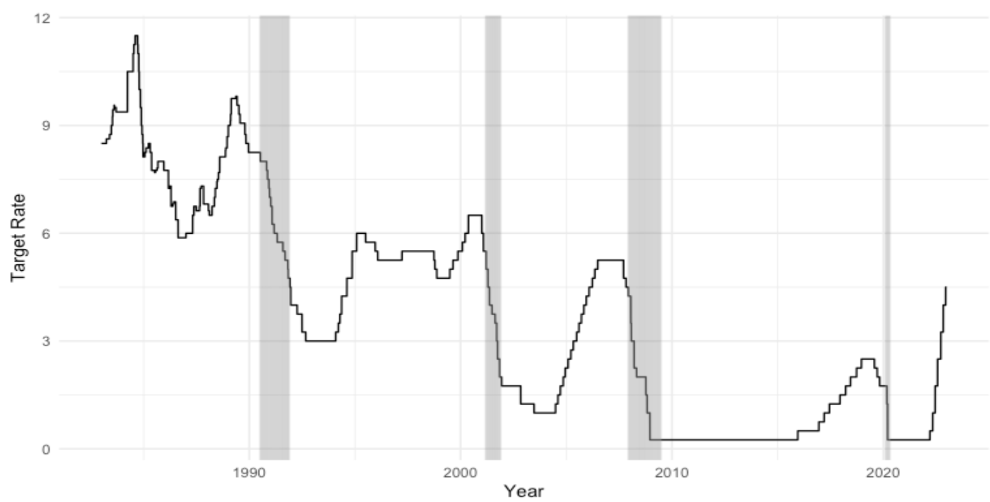
[1] Bernanke, Ben S. and Kuttner, Kenneth N. and Kuttner, Kenneth N. (2005), "What Explains the Stock Market's Reaction to Federal Reserve Policy? ".

3. Data

3.1 Data selection

The sample data includes monthly observations of the Federal funds target rate and other economic variables from February 1983 to December 2022. Specifically, the data I'm interested in are the dates of FOMC decisions regarding target rate. Considering both scheduled meetings, which occur regularly eight times a month, and extraordinary meetings, I obtained a total of 399 observations over this period. Since the frequency of my data is monthly and in some months the FOMC met more than once, I considered as target rate decision for a given month the one related to the last meeting in that month. In this way, the total number of observations becomes 352. For each meeting date I associated the old and new target rate and the magnitude of the rate change, which generally equals multiples of 25 basis-points. Among these 352 observations, the target rate was hiked 70 times, cut 57 times and kept constant 225. As can be seen, the classes are slightly unbalanced; this distribution reflects the nature of FOMC policy decision, where rate adjustments are often implemented in response to significant shifts in key macroeconomic indicators or emerging risks. The larger number of instances where the rate remained unchanged suggests that the FOMC considers a steady policy stance appropriate during periods of relative stability or when the economic environment does not necessitate immediate intervention. Hence, this class imbalance highlights the careful deliberation undertaken by the FOMC in shaping monetary policy. In Figure 1, the graph depicts Federal fund target rate levels from 1983 to 2022. Both FOMC meeting dates and target interest rate data were downloaded from the Federal Reserve Economic Data (FRED) online database of the Federal Reserve Bank of St. Louis.

Figure 1: Federal funds target rate (1983-2022)



Concerning the selection of macroeconomic indicators for inclusion in empirical models, there is no uniform standard in the academic community as we could see in the literature review. In my case, I initially included 22 economic series, Table 1 shows the description of the indicators.

Table 1: Macroeconomic indicator description

Consumer Price Index for All Urban Consumers: is a price index of a basket of goods and services paid by urban consumers
Inflation: percent change from one year ago in the consumer price index measure
Core CPI: price index of a basket of goods and services that excludes volatile items such as food and energy
Core Inflation: percent change from one year ago in the Core CPI
Unemployment rate: number of unemployed as a percentage of the labor force
Personal Consumption Expenditures: is another measure of the spending on goods and services by people of the United States
Total Vehicle Sales: measures the annualized number of new vehicles sold in the reported month
Bank Prime Loan Rate: interest rate that commercial banks charge creditworthy customers
M1: measure of the money supply that includes currency, demand deposits, and other liquid deposits, including savings deposits
M2: measure of the money supply that includes M1 plus savings accounts, money market accounts and small time deposits
M3: measure of the money supply that includes M2 money as well as large time deposits, institutional money market funds, short-term repurchase agreements, and larger liquid funds
Initial Claim: measure of claims filed by unemployed individuals after a separation from an employer
Consumer Confidence Index: monthly survey of how consumers feel about the economy, personal finances, business conditions, and buying conditions conducted by the University of Michigan
Average Working Hours: measure of the average weekly hours per worker for which pay was received
Capacity Utilization: is equal to an output index divided by a capacity index for all the industries
Industrial Production: measures the real output of all relevant establishments located in the United States
Producer Price Index: measures the change in the prices paid to U.S. producers of goods and services
Exports: measures the value of exported goods and services for the U.S.
S&P 500: is a market-capitalization-weighted index of 500 leading publicly traded companies in the U.S
Manufacturing PMI: is a monthly index of U.S. economic activity based on a survey conducted by Institute for Supply Management (ISM) of purchasing managers at more than 300 manufacturing firms
GDP: measures the market value of goods and services produced by labor and property located in the U.S.
Potential GDP: estimate of the output the economy would produce with a high rate of use of its capital and labor resources

All the observations range from February 1983 to December 2022. The frequency is monthly for all data series except GDP and Potential GDP, for which the frequency is quarterly. Most of the data were retrieved using FRED's API on Phyton. Only data relative to the S&P 500 and Manufacturing PMI were obtained from WRDS (Wharton Research Data Services) and the Institute for Supply Management website, respectively.

Then, in order to carry out the event study to analyze the impact of policy decisions about target rate on the stock market, I collected daily frequency data related to market factors. So, I re-downloaded the observations for the S&P 500 from WRDS but at daily level. Finally for the development of the back-testing strategy, I downloaded data on U.S. Zero Coupon Bond yields at different maturities (10,20,30 years) from FRED online database and data on Nasdaq Financial-100 Index and S&P United States REIT Index from Refinitiv.

3.2 Data Preprocessing

Once retrieved, I preprocessed the raw data in such a way as to create the variables for my forecast models.

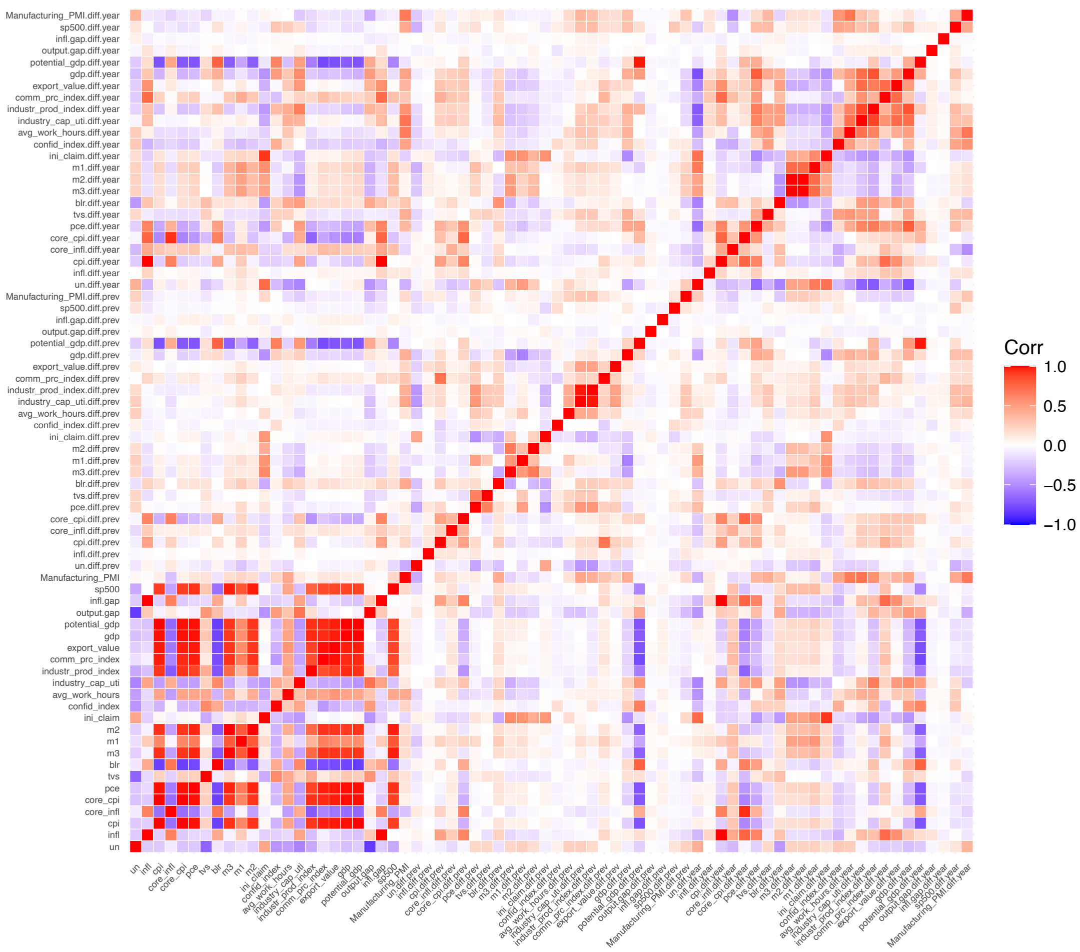
First, recall that my models are discrete triple-choice models, whose dependent variable must represent the FOMC's decisions, which can be to hike, cut, or hold constant the target interest rate. For this purpose, I have generated a new variable that takes, respectively: value 1 if the change in the Federal funds target rate resulting from the FOMC meeting is positive (hike); value -1 if the change is negative (cut); and value 0 if the change is zero (hold constant). Then, I used the macroeconomic series to generate the independent variables of my models. Remembering that the frequency of my dependent variable is monthly, I first transformed the quarterly series (GDP and Potential GDP) into monthly series. To do this, I used the "tempdisagg" package in R, which allowed us to disaggregate and interpolate a low-frequency time series into a higher-frequency series, keeping the sum, mean, first or last value of the resulting high-frequency series consistent with the low-frequency series[2]. Once I obtained monthly GDP and Potential GDP series, I used them to create a new variable, the Output gap. This indicator is calculated as the percentage difference between GDP and Potential GDP and measures the efficiency of the economy, specifically the extent to which an economy is operating below or above its productive capacity. In addition, I added an inflation gap variable, calculated as the difference between the current inflation rate and the target inflation rate (set at 2%). The decision to add these two indicators was inspired by most of the econometric models from the academic literature, which tend to include these two variables as regressors.

[2] Specifically, "tempdisagg" uses the Chow-Lin Method for Temporal Disaggregation which is part of the methodological handbooks of the DIME, a director group of Eurostat - one of the European Commission Directorates-General.

At this point for each economic indicator I calculated the difference from the previous period and the same period in the previous year. This approach allowed us to capture changes and trends in macroeconomic variables over time and to focus on relative changes rather than absolute values. Using both these differences and the indicators themselves, I obtained a total of 72 independent variables to include in the first estimation stage. To these I then decided to add another variable related to the target rate change undertaken by the FOMC at the previous meeting. In this way, my models are able to take into account the historical influence of the past FOMC's decisions (i.e., the fact that past FOMC decisions may have a lasting impact on their future decisions) as well as the continuity of FOMC's policy (i.e., the fact that it maintains a consistent policy stance over multiple meetings).

To avoid multicollinearity problems, I calculated the correlation between all macroeconomic variables. The results are shown in Figure 2.

Figure 2: Correlation Matrix Macroeconomic Variables



As we can see in the lower left part of the graph there is a lot of correlation between some raw economic indicators. A few variables obviously exhibit perfect multicollinearity such as inflation and inflation gap. Also other macroeconomic indicators show very high correlations such as CPI, Core CPI, PCE, M1, M2, M3, Industrial Production, Producer Price Index, Exports value, S&P 500, GDP and Potential GDP; these correlations can be explained by their inherent interdependencies and common underlying factors. These indicators are related because of economic interconnectedness, supply and demand dynamics, business cycle effects, and political and external factors. For this reason I decided to exclude the most correlated variables, taking a correlation of 0.7 as the decision threshold. So, among this group of highly interrelated variables, I decided to keep only the Consumer Price Index, M1 and Inflation. Regarding the correlations of the other pre-processed variables, we note only a few single very high correlations and a high aggregate interconnectedness in the upper right part of the graph between some of the year difference variables. I therefore decided to remove the following variables: cpi.diff.year, core_cpi.diff.year, industry_cap_uti.diff.prev, potential_gdp.diff.year, industr_prod_index.diff.year, prc_index.diff.year, m2.diff.year, export_value.diff.year. In this way I obtained a total of 56 regressors to use in the development of my models.

Finally, to prevent any potential look-ahead bias, all macroeconomic regressors have been lagged by one period, ensuring that observations on economic indicators precede the FOMC's decision on the target rate.

4. Hypotheses Development

The primary objective of this paper is to develop machine learning models capable of accurately predicting FOMC policy decisions regarding the Federal funds target rate. Furthermore, I aim to utilize these models to conduct a comprehensive analysis of the impact of these decisions on financial markets. Consequently, I have formulated two main hypotheses that I seek to test:

Hypothesis 1: My machine learning models are expected to outperform the conventional econometric models developed in the academic literature in terms of accuracy and log/quadratic-score.

Hypothesis 2: The developed machine learning models are capable of capturing market participants' expectations about future FOMC decisions.

To test the first hypothesis, I will conduct a direct comparison between the performance metrics of my models and those documented in the literature review. On the other hand, the second hypothesis will be examined through regression analysis of the market Cumulative Abnormal Returns (CARs) obtained from an event study. Specifically, I will regress the CARs on the variables associated with negative or positive surprises of the FOMC decision, which are created on the results of my models. Statistically significant coefficients with economically meaningful signs would support my second hypothesis. By investigating these hypotheses, I aim to shed light on the predictive power of my machine learning models and their ability to provide insights into the impact of FOMC decisions on the financial markets.

5. Methodology

Forecasting FOMC decisions on target Federal funds rate can be accomplished through either a continuous approach, as in Taylor (1993), or a discrete approach, as addressed by the rest of the academic literature. I chose the discrete approach for the development of my forecasting models since it best reflects the nature of the FOMC's policy actions and allows us to carry out comparisons with the other models developed in the past. Regarding the choice of models for predicting FOMC decisions, I decided to use a logistic regression as a baseline model, as extensively done in the academic literature. I then decided to develop several machine learning models, which represent the real contribution I want to make to the literature. In fact, machine learning models allow us to benefit from several advantages over classical econometric models, such as handling complex data, detecting nonlinear relationships, generalizing well, and incorporating a wide range of informative variables. In addition, unlike econometric models, which are primarily focused on obtaining unbiased estimations of parameters, machine learning algorithms allow us to test model results on a new portion of the data thus obtaining a measure of predictive performance. On the other hand, machine learning models suffer from some drawbacks: they become difficult to interpret and to derive economic intuitions. However, these issues can be overcome through tools such as the variable importance and partial dependence plot. Finally, I developed again my models but using a continuous approach to conduct inference on the Event Study results, particularly to distinguish between expected and unexpected policy decisions. I used the models' predictive errors to identify surprises in target rate changes, thus proposing an alternative method to that of Kuttner (2001).

5.1 Forward Selection

Before starting to develop the predictive models, I used a method for subset selection of the predictors. This allows us to identify and select a subset of relevant variables from my original feature set. The goal is to choose the most informative and discriminating features that contribute most to the predictive performance of a machine learning model.

The method I have used is Forward Stepwise Selection which appears to be the most computationally efficient compared to other approaches such as best subset selection. While best subset selection considers all possible models containing subsets of predictors, forward stepwise selection considers a narrower set of models. It starts with an empty model and gradually adds predictors, one at a time, until all of them are included. The added predictor is chosen based on the greatest improvement in model fit, which is determined in terms highest R2 or highest accuracy. Unlike best subset selection, which requires fitting 2^p models, where p denotes the number of predictors; forward stepwise selection involves fitting a null model and $p - k$ models in the kth iteration, $k = 0, \ldots p - 1$, for a total of $1 + \sum_{k=0}^{p-1}(p - k) = 1 + p(p + 1)/2$ models. This generates a clear computational advantage, especially in high-dimensional scenarios. However, forward stepwise selection may not always yield the best model among all possible models. Despite its limitations, forward stepwise selection tends to perform well in practice. Table 2 shows the forward stepwise selection algorithm.

Table 2: Forward stepwise selection algorithm

1. Start with a null model M_0, which contains no predictors
2. For $k = 0, \ldots p - 1$:
    a) Consider all the $p - k$ models that augment the predictor in M_k with on additional predictor.
    b) Chose the best model, in terms of highest R2 or highest Accuracy, among these models and call it M_{k+1}
3. Select a single best model from among $M_0, \ldots, M_p$ using cross-validated prediction error

5.2 Forecasting Models

5.2.1 Logistic Model

The logistic model is a statistical method for handling a classification problem, i.e., a problem in which the dependent variable is composed of categories or classes. This model allows us to predict the probability that an observation belongs to each of the categories of the qualitative variable. Specifically, the logistic regression model estimates the relationship between the independent variables and the log-odds of the dependent variable, which is then transformed into predicted probabilities using the logistic function. The logistic function maps any real-valued number to a value between 0 and 1, which represents the probability of belonging to a specific class. If the response variable has more than two classes, as in my case, it is necessary to use an extension of the logistic model called the multinomial logistic regression model. This model uses the following logistic function for estimating probabilities:

$$\Pr(Y = k|X = x) = \frac{e^{\beta_{k0}+\beta_{k1}x_1+\cdots+\beta_{kp}x_p}}{1 + \sum_{l=1}^{K-1} e^{\beta_{l0}+\beta_{l1}x_1+\cdots+\beta_{lp}x_p}} \tag{1}$$

for $k = 1, \ldots, K-1$,

where K is the number of classes of dependent variable. Function 1 is fitted using a method called maximum likelihood. This approach ensures that the model estimates are chosen to maximize the probability of observing the given data. At this point we can obtain for all observation the probability of belonging to each class by substituting within 1 the estimated coefficients. Through a few steps we can rewrite equation 1 as follows:

$$log\left(\frac{\Pr(Y = k|X = x)}{\Pr(Y = K|X = x)}\right) = \beta_{k0} + \beta_{k1}x_1 + \cdots + \beta_{kp}x_p \tag{2}$$

The left-hand side is called log odds or logit, and we can notice that is linear in X for the logistic regression model. While in a linear regression model, β represents the average change in Y associated with a one-unit increase in X, in logistic regression, increasing X by one unit changes the log odds by β. The relationship between $\Pr(Y = k|X = x)$ and X is not linear, so β does not correspond to the change in p(X) associated with a one-unit increase in X. So, the direction and magnitude of $\Pr(Y = k|X = x)$ change due to a one-unit change in X depend on the current value of X.

5.2.2 Support Vector Machine

The Support Vector Machine (SVM) is a powerful classification algorithm that builds upon the concepts of hyperplanes and the maximal margin classifier. In machine learning, a hyperplane is a flat subspace in a *p*-dimensional space, where *p* represents the number of dimensions. For example, in two dimensions, a hyperplane is a line, while in three dimensions, it is a plane. In general, a hyperplane can be defined by an equation of the form:

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots .. + \beta_p X_p = 0 \tag{3}$$

where $\beta_0, \beta_1, \ldots, \beta_p$ are parameters. This equation divides the space into two halves, and the sign of the left-hand side determines on which side a point lies. In classification tasks, the goal is to develop a classifier that can correctly classify test observations based on their features. One approach is to use a separating hyperplane, which perfectly separates the training observations according to their class labels. The separating hyperplane satisfies the condition

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots .. + \beta_p X_p > 0$$

for observations labeled as one class and

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots .. + \beta_p X_p < 0$$

for observations labeled as the other class. The maximal margin classifier is a natural choice when constructing a classifier based on a separating hyperplane. It is the separating hyperplane that has the farthest minimum distance from the training observations, known as the margin. The maximal margin classifier classifies a test observation based on its location relative to the maximal margin hyperplane.

Support Vector Machines build upon the concept of the maximal margin classifier. SVMs aim to find the optimal hyperplane that not only separates the classes but also maximizes the margin while allowing for some misclassification. This is achieved by introducing slack variables that permit observations to fall within the margin or even on the wrong side of the hyperplane. The optimization problem associated with SVMs involves finding the hyperplane coefficients that maximize a specific objective function while satisfying certain constraints:

$$max_{\beta,\varepsilon,M} \quad M \tag{4}$$

subject to $\quad y_i\left(\beta_0 + \sum_{j=1}^{n} \alpha_j \langle x, x_j \rangle\right) \geq M(1 - \varepsilon_i),$

$$\sum_{i=1}^{n} \varepsilon_i \leq C, \quad \varepsilon_i \geq 0, \quad \sum_{j=1}^{p} \sum_{k=1}^{2} \beta_{jk}^2 = 1$$

where M is the width of the margin, $\varepsilon_i \ldots, \varepsilon_n$ are the slack variables that allow individual observation to be on the wrong side of the margin or hyperplane, C is a tuning parameter that determines the number and severity of the violation to the margin and hyperplane that the model will tolerate. Instead, $\langle x, x_j \rangle$ refers to the inner product between a new point x and each of the training points x_j and is defined as *kernel* in SVMs.

A *kernel* quantifies the similarity between observations and can be chosen based on the desired behaviour. The advantage of using kernels in SVMs, which also distinguishes it from the support vector classifier, is the possibility of addressing the problem of nonlinear boundaries between classes by expanding the feature space with kernels. Using a non-linear kernel in the SVM algorithm allows for a more flexible decision boundary, capturing complex relationships in the data. In my case I decided to use a radial kernel. A radial kernel is characterized by high flexibility in modeling intricate patterns in the data. It quantifies the similarity between observations based on their Euclidean distance. By using the radial kernel, the SVM can assign higher weights to observations that are closer to the decision boundary, allowing for a more accurate and precise classification. Additionally, the radial kernel's local behavior ensures that only nearby training observations have a significant impact on the class labels of test observations. This means that the radial kernel focuses on the most relevant observations, resulting in improved classification performance.

SVMs are suited for two-class classification problems, however in my current setting I have three classes. To address this problem, I decided to use an extension of SVMs: the *one-versus-all* approach. With this approach, I train *K* SVMs, where each SVM compares one of the *K* classes to the remaining *K*-1 classes. The SVM for the kth class is trained to distinguish that class (coded as +1) from the others (coded as -1). When given a test observation I assign it to the class with the highest confidence score based on the SVM model's decision function. This decision is made because it indicates a high level of confidence that the test observation belongs to the *k*th class rather than any of the other classes.

The value of tuning parameter *C* that I used in the model was found through the grid search method. I tested a grid of values ranging from 1 to 50 and select the best choice using a cross validation approach.

5.2.3 Tree-Based Models

5.2.3.1 Decision Tree

The decision tree is a machine learning algorithm that can be applied to both regression and classification problems. It has the advantage of being simple and useful for interpretation. The process of building a decision tree follows two steps:

1. The set of possible values for $X_1, X_2, \dots, X_p$, called predictor space, is divided into *J* distinct and non-overlapping regions, $R_1, R_2, \dots, R_J$.
2. For each observation that falls into the region R_J, the algorithm makes the same prediction, which, in the case of classification tree, is the most commonly occurring class of training observations in R_J to which it belongs.

In the classification setting, the criterion used to divide the predictor space into J regions differs from that of the regression tree. Instead of minimizing the RSS, the focus is on minimizing the *Gini index*, which is represented by the following equation:

$$G = \sum_{k=1}^{K} \hat{p}_{mk}(1 - \hat{p}_{mk}) \tag{5}$$

where $\hat{p}_{mk}$ is the proportion of the training data in the *m*th region of the *k*th class. The *Gini index* is a measure of node *purity*, the lower its value the higher the *purity* of the node, defined by the number of observations in a node from a single class.

However, considering every possible partition of the feature space into *J* boxes is computationally infeasible. Hence, the decision tree algorithm uses a top-down, greedy approach known as recursive binary splitting. Recursive binary splitting starts at the top of the tree with all observations in a single region and proceeds by successively splitting the predictor space. At each step, the best split is determined by selecting a predictor X_J and a cutpoint *s* that maximizes the reduction in *Gini index*. The predictor space is then divided into two regions: $\{X|X_J < s\}$ and $\{X|X_J \geq s\}$. This splitting process is repeated iteratively, searching for the best predictor and cutpoint to further minimize the *Gini index* within each resulting region. The splitting continues until a stopping criterion is met, such as reaching a maximum depth or a minimum number of observations in a region. Once the regions $R_1, R_2, \dots, R_J$ are created, predictions for new test observations are made according to the *majority vote* rule, that is, the most commonly recurring class among all predictions.

Decision trees are constructed by recursively partitioning the data space from the root node to the terminal nodes, capturing complex relationships and interactions between variables. The importance of variables is reflected in the selection of the root node and subsequent nodes based on the smallest Gini index. However, decision trees tend to suffer from high variance, making them sensitive to data changes and prone to overfitting. To address this, a pruning technique can be applied.

5.2.3.2 Tree Pruning

The goal of the Pruning technique is to select a smaller subtree that generalizes well to unseen data while maintaining interpretability. Rather than growing a complex tree, a large tree T_0 is first constructed, which is then pruned to obtain a subtree by minimizing the total misclassification error rate. Estimating the error for every possible subtree is impractical due to their large number. Instead, cost complexity pruning, also known as weakest link pruning, offers a solution. Cost complexity pruning involves a sequence of trees indexed by a nonnegative tuning parameter α. For each value of α, a subtree $T \subset T_0$ is selected to minimize a criterion represented by the following equation:

$$R_\alpha(T) = E(T) + \alpha|T| \tag{6}$$

where $|T|$ indicates the number of terminal nodes of the tree T and $E(T)$ is the misclassification rate of the tree T, computed as the sum of the of misclassification errors at each note of the tree T. $R_\alpha(T)$ represents the cost-complexity measure and serves as a penalized version of the resubstitution error rate. It balances the subtree's complexity (number of terminal nodes) and its fit to the training data. If α is small, the largest tree is chosen as the complexity penalty term becomes negligible. Conversely, as α increases, a smaller subtree is favored to prevent overfitting. The process of obtaining the sequence of subtrees as α increases is straightforward. The value of α can be chosen using a validation set or cross-validation. Once determined, the corresponding subtree is obtained from the full dataset. This pruning approach is summarized in the following algorithm:

Table 3: Pruning algorithm

1. Grow a large tree on the training data stopping when each terminal node has fewer than some minimum number of observations
2. Apply cost complexity pruning to the large tree in order to obtain a sequence of best subtrees, as function of α
3. Use K-fold cross-validation to choose α. That is, divide the training observations into K folds. For each $k = 1, \dots, K$:
    a. Repeat Steps 1 and 2 on all but the kth fold of the training data.
    b. Evaluate the misclassification error on the data in the left-out kth fold, as a function of α. Average the results for each value of α, and pick α to minimize the average error.
4. Return the subtree that corresponds to the chosen value of α.

5.2.3.3 Bagging

Bootstrap aggregation, or bagging, is a technique used to reduce the variance of statistical learning methods, particularly decision trees. The fundamental idea behind bagging is that averaging multiple prediction models built on different training sets can decrease variance and enhance test set accuracy. However, obtaining multiple training sets is usually impractical. Instead, bootstrap samples are generated by resampling from the original training set. By constructing B different bootstrapped training sets, we can train the method on each set separately to obtain B prediction models ($\hat{f}^1(x)$, $\hat{f}^2(x), \dots, \hat{f}^B(x)$). Averaging these predictions results in a single low-variance statistical learning model, a process known as bagging:

$$\hat{f}_{bag}(x) = \frac{1}{B}\sum_{b=1}^{B} \hat{f}^b(x) \tag{7}$$

In the case of classification problems, where the outcome variable is qualitative, the averaging of predictions is carried out according to the majority vote approach: for each test observation, the class predicted by each of the B trees is recorded, and the overall prediction is determined by selecting the most commonly occurring class among the B predictions. In my case I decided to set the value of B to 100, since beyond this number adding additional bags does not progressively improve model estimates but only computational time.

Bagging has proven to be highly effective in improving prediction accuracy by combining numerous trees into a single procedure.

5.2.3.4 Random Forest

Random forests offer an enhanced version of bagged trees by introducing a modification that reduces the correlation between the trees. Instead of considering all predictors at each split, random forests randomly select a subset of predictors, m, from the total p predictors. Typically, m is chosen to be approximately equal to the square root of p ($m \approx \sqrt{p}$).

This modification is aimed at decorrelating the trees to avoid high correlation among predictions. In bagging, if there is a strong predictor in the dataset, it tends to be used in the top split of most or all trees, resulting in highly correlated predictions. However, averaging highly correlated quantities does not lead to a significant reduction in variance compared to averaging uncorrelated quantities. Random forests address this issue by restricting each split to only consider a subset of predictors, making the resulting trees less variable and more reliable. This process decorrelates the trees and increases the chance for other predictors to have an impact. The key distinction between bagging and random forests lies in the selection of the predictor subset size, m. If m equals p, random forests essentially become bagging. However, using a smaller value of m, especially when dealing with a large number of correlated predictors, can be beneficial. For the same reason mentioned earlier regarding bagging, I set the number of trees (*ntrees*) used by the model to reduce variance to 100.

5.2.3.5 Optimal Random Forest

Finally, I developed a random forest model using the optimal number of regressors, m. I set up a grid ranging from 1 to p, representing the potential values for m. Then, I implemented a loop that fitted a random forest with a given number of trees for each value of m. The selection of the optimal m is based on minimizing the OOB estimate of test error.

The Out-of-Bag (OOB) test error is a metric used to estimate the performance of a bagged tree or random forest model. It is calculated by evaluating each individual tree in the ensemble on the training data that were not used during its construction. During the training process, bagged trees or random forest models are built by randomly selecting subsets of the original training data, allowing for the creation of multiple trees. Each tree is constructed using a bootstrap sampling technique, where some observations are included in the training set multiple times, while others are left out (out-of-bag observations). The OOB test error is computed by averaging the classification error for each of the *i*th training observation, using trees in which that observation was OOB.

So, I stored the OOB estimate of test error rate for each of the *p* random forest models, and selected as the optimal *m*, the one with which the lowest OOB error is associated. With this optimal value of *m*, I re-estimated a random forest model, thus obtaining the best tree model.

5.2.3.6 Variable Importance

Bagging and random forest are known to enhance prediction accuracy compared to using a single classification tree. However, it comes at the cost of interpretability. While decision trees provide a clear and understandable diagram, representing a bagged model or a random forest model with numerous trees becomes challenging, and identifying the most important variables becomes less apparent. Consequently, they sacrifice interpretability for improved prediction accuracy. To address this limitation, an overall summary of the importance of variables for bagged and random forest classification trees can be obtained using two different methods, according to Breiman (2001) and Friedman (2001). The first approach is based on node impurity and involves the Gini index. By calculating the total decrease in the Gini index (3) resulting from splits over a specific predictor and averaging it across all the trees, the importance of each predictor can be determined. A higher value indicates a more influential predictor in the classification trees. Similarly, for regression trees, the total decrease in the residual sum of squares (RSS). The second method, on the other hand, is based on accuracy. The importance of the predictor is measured by calculating the decrease in accuracy in predictions on the out of bag samples when a given variable is removed. Even in this case, a higher value indicates a more influential predictor.

5.2.3.7 Partial dependence plot

The partial dependence plot is a valuable tool for analyzing the marginal effect of a feature on the predicted outcome of a model (Friedman 2001). By fixing the prediction function at specific values of the selected features and averaging over the remaining features, it provides insights into the relationship between the target variable and a particular feature. One of the key advantages of partial dependence plots is their ability to reveal the nature of the relationship between the target variable and a feature, whether it is linear, monotonic, or more complex. The partial dependence function for regression is defined as:

$$\hat{f}x_s(x_s) = Ex_c[\hat{f}(x_s, x_c)] = \int \hat{f}(x_s, x_c)dP(x_c) \tag{8}$$

Here, x_s represents the set of features for which the partial dependence function is being plotted, while x_c refers to the other features used in the machine learning model. By marginalizing the output of the machine learning model $\hat{f}$ over the distribution of features x_c, the resulting function represents the relationship between the x_s features of interest and the predicted outcome. To estimate the partial function $\hat{f}x_s$ along x_s, we employ the Monte Carlo method, calculating averages by using training data:

$$\hat{f}x_s(x_s) = \frac{1}{N}\sum_{i=1}^{n} \hat{f}(x_s, x_{c_i}) \tag{9}$$

In this formula, x_{c_i} represents the actual feature values from the dataset for the features that are not of interest, while n denotes the number of instances in the dataset. In my case of classification tasks, where the machine learning model outputs probabilities, the partial dependence function illustrates the probability associated with a specific class given various values of the x_s features. Multi-class problems can be handled by plotting individual lines or separate plots for each class.

5.2.4 Evaluation Metrics

In machine learning analysis, evaluation metrics play a crucial role in assessing the performance of models, particularly their ability to generalize to new data. They also aid in model selection by identifying the best-performing model from a set of trained models. Accuracy or error rate (1-accuracy) is a commonly used evaluation metric in classification problems. Specifically, it is the most widely used evaluation metric in the academic literature related to forecasting FOMC decisions on target rate. It measures the number of correct predictions made by a model and is calculated based on the values present in the confusion matrix, which is a tabular representation of the model's predicted labels compared to the actual labels. The confusion matrix is typically a square matrix with rows and columns representing the different classes or labels in the classification problem. The diagonal elements of the matrix correspond to the correctly classified instances, while the off-diagonal elements represent the misclassifications. The accuracy is computed by dividing the sum of correctly classified instances by the total number of instances in the dataset. In addition, by looking at the confusion matrix we are able to calculate the accuracy of the model for individual class predictions.

However, my machine learning models also produce class prediction probabilities, which are often disregarded when computing accuracy. Accuracy relies on a single decision threshold, such as 0.5 in binary classification, where observations with prediction probabilities above 0.5 are classified as positive and the rest as negative. This approach fails to fully leverage the informative results that the model provides, limiting the precision of the analysis. To address this, I followed Pauwels (2012) by introducing two additional evaluation metrics, namely the log-score and the quadratic-score, to provide a more comprehensive assessment of model performance.

Log-score takes into account the logarithm of the predicted probabilities, allowing for a more nuanced analysis. It captures the probabilistic nature of the predictions and provides a measure of the model's confidence in its predictions. Higher log-score values indicate more accurate and confident predictions. In my current setting in which the models produce three probabilities for each prediction ($\hat{P}_{-1}, \hat{P}_0, \hat{P}_1$), the log-score has the following form:

$$S^l = \log(\hat{P}_j) \tag{10}$$

where $\hat{P}_j$ is the probability predicted by the model for the state that actually happens. Similarly, quadratic-score considers both the predicted probabilities and their proximity to the true class labels. It quantifies the discrepancy between predicted probabilities and actual outcomes, rewarding predictions that are close to the true values. In my specific case the quadratic-score rule is given by:

$$S^q = 2\hat{P}_j - (\hat{P}_{-1}^2 + \hat{P}_0^2 + \hat{P}_1^2) \tag{11}$$

These scores are calculated for each out-of-sample predictions and then are averaged to obtain a single final score. By utilizing log-score and quadratic-score alongside accuracy, this research aims to gain deeper insights into the performance of the models in predicting FOMC decisions. These metrics offer a more comprehensive understanding of the model's reliability, precision, and confidence in its predictions.

5.2.5 K-fold Cross Validation

In all the models I applied a 5-fold cross-validation approach. This technique provides several benefits for model generalization compared to a single train-test split. Firstly, cross-validation allows for a more comprehensive assessment of the model's performance by dividing the dataset into multiple subsets or folds. This approach ensures that each data point has the opportunity to be part of the test set, providing a more reliable estimation of the model's ability to generalize to unseen data. By averaging the performance metrics across multiple folds, I obtain a more robust evaluation of the model's predictive power. Additionally, cross-validation helps in mitigating the potential bias that could arise from a particular division of the dataset. It allows for a more balanced representation of the data across different folds, reducing the impact of any specific subset's characteristics on the model's performance. This is particularly relevant in the case of predicting FOMC decisions, where the dataset may contain various factors that could introduce temporal dependencies or specific trends. By adopting a cross-sectional approach instead of a chronological one, I can mitigate the influence of time-related patterns and focus on capturing the underlying relationships between macroeconomic indicators and decision outcomes. Choosing a cross-sectional design over a temporal one allows for the utilization of the entire available dataset, maximizing the information used for model training and evaluation. Thus, by leveraging the benefits of cross-validation, I can enhance the robustness and generalizability of the machine learning models developed for predicting FOMC target Federal funds rate.

5.3 Event Study

Once the models for predicting Fed policy decisions have been developed, the goal of this paper is to analyze the impact these decisions have on financial markets. Specifically, I want to study how unexpected decisions regarding a change in Federal fund target rate impacts the stock market. To do this, I used the Event Study methodology introduced by Craig MacKinlay (1997). The event study is a technique to measure the effect of economic event on the value of firms.

The first step in conducting an event study is to identify the event of interest, which in this case is the announcement of a change in the Federal funds target rate by the Federal Reserve. Getting the timing right is crucial to an event study analysis. I therefore considered as event dates the FOMC meeting dates previously used for model development. However, some policy decisions are not communicated on the same day that the FOMC meets. For this reason, I filtered out all those events in which the day of the FOMC meeting and the day of the effective market change of the target fed funds rate (thus the day of the public announcement) did not coincide and were far apart. In this way I obtained 343 total events from the original number of 352.

At this point, my interest lies in distinguishing between expected and unexpected FOMC policy decisions. To achieve this, I have revised my machine learning models, adopting a continuous approach instead of a discrete one. Consequently, my new continuous variable represents the newly determined Federal fund target rate at each FOMC meeting. I made this decision to allow the calculation of errors for each observation, which entails finding the difference between the actual observed value of the target rate and the value estimated by the model. To compute these errors, I employed the random forest model with 5-fold cross-validation. This approach allowed us to treat each observation as a test set, providing us with an estimate of the forecast error for each individual data point. I opted against calculating squared errors as my objective is to differentiate between positive and negative surprises based on the sign of estimation errors. Specifically, I categorized events with a positive error as negative surprises since the FOMC's actual target rate exceeded public expectations according to the model. Conversely, negative errors were associated with positive surprises as the actual target rate fell below expectations. Subsequently, I arranged the errors in descending order and generated two new dummy variables: one denoting negative surprises, assigned a value of 1 for the top 20 observations with the highest errors (corresponding to the 95th percentile); the other representing positive surprises, assigned a value of 1 for the last 20 observations with the lowest errors (corresponding to the 5th percentile). As a result, each event was associated with three variables: the two surprise-related dummy variables and one indicating the sign of the change in the federal fund target rate. My analysis focuses on examining the impact of policy decisions on the overall U.S. stock market. Therefore, I selected the S&P 500 as a proxy variable for studying market movements.

The next step in conducting an event study involves defining the event window and estimation window. The event window encompasses a specific time period surrounding the event date. In my case, I chose to extend it from 5 days before the event to 5 days after. This decision was driven by my desire to capture early market reactions prior to the announcement of the target rate decision. On the other hand, the estimation window is set several days prior to the event to allow for the estimation of Normal Returns. In my study, I selected an estimation window ranging from 90 days before the event to 6 days before.

To measure the impact of the event, the actual stock price movements during the event window are compared with the expected or normal return. Several methods can be employed to estimate the expected return, such as the mean adjusted model, market adjusted model, or market model. Given that my study variable is the market index itself, which is influenced by the event, the only method I could use is the mean adjusted mode. This approach involves calculating the normal return as the average of the returns of my study variable during the estimation window.

Subsequently, I calculate Abnormal Returns, a commonly used metric in event studies that represents the difference between the actual return of a security during the event window and the expected return. Abnormal Returns help isolate the specific impact of the event on stock prices, controlling for other factors. For each event, I retain the Abnormal Returns specifically related to the event window and calculate the Cumulative Abnormal Returns (CARs) by summing the ARs cumulatively. This process yields 343 CARs, which I use for regression analysis with the three previously mentioned variables. Specifically, I conduct two regressions, differentiating them based on the direction of the decision regarding changes in target rate. The regression equations take the following form in both cases:

$$CAR_i = \alpha + \beta PosSurprise_i + \gamma NegSurprise_i + \varepsilon_i \quad (12)$$

In the first regression, only CARs associated with a FOMC decision to raise rate are considered, while in the second regression, only CARs associated with a FOMC decision to lower rate are considered.

The event study methodology is based on 4 main assumptions: abnormal returns and cumulative abnormal returns are uncorrelated across events (A1-A2); the model is linear in parameters (A3); the variance of the error term is equal between events and does not depend on covariates (A4). In my case, Assumptions 1 and 2 may be violated as the Federal Reserve's policy decisions could be correlated, especially those made in close proximity or within the same period. This could lead to a wrong computation of standard errors for the coefficients and consequently to an incorrect statistical significance analysis. To address this issue, one solution is to assume the presence of clusters where events can be correlated, while ensuring that different clusters are uncorrelated. By employing the cluster formula, I can obtain the correct standard errors for the parameters. In my study, I consider clusters based on the year, assuming that target rate decisions within the same year are correlated. This decision is motivated by the understanding that the Federal Reserve's monetary policy actions are often driven by broader economic conditions and objectives that persist over a certain time frame.

In conclusion, through the event study analysis, I aim to examine the impact of target Federal funds rate decisions on the market index, differentiating between the direction of the decision and whether it represents a negative or positive surprise.

6. Results

In the upcoming chapter, I will present the outcomes of the methodologies described previously. The results will be presented in three separate sections. Firstly, I will assess the predictive performance of my machine learning models. Secondly, I will delve into the findings of the event study. Lastly, I will outline a straightforward investment strategy based on the predictive capabilities of my models.

6.1 Forecasting Results

Before delving into the analysis of the predictive performance of my machine learning models, it is important to present the outcomes of the forward selection process I implemented. This selection process aimed to identify a subset of the most relevant regressors from the original set. Initially, I conducted forward selection by setting the maximum subset size to my total number of regressors, which was 56. This approach allowed us to find the optimal number of regressors (23) to include, associated with the highest level of performance metrics. At this point, I performed forward selection again, with 23 as the maximum subset size. The outcome of this process is displayed in Table 4, presenting the selected regressors to be incorporated into my models.

Table 4: Regressors selected by forward stepwise selection

Previous Target rate decision	Initial Claim previous period diff.
Consumer Price Index	Confidence Index previous period diff.
Core Inflation	Average working hours previous period diff.
Total Vehicle Sales	Exports previous period diff.
Bank Prime Loan Rate	Potential GDP previous period diff.
M1	PCE previous year diff.
Initial Claim	Bank Prime Loan Rate previous year diff.
Average working hours	M3 previous year diff.
Capacity Utilization	Initial Claim previous year diff.
Manufacturing PMI	Average working hours previous year diff.
Core Inflation previous period diff.	Capacity Utilization previous year diff.
Total Vehicle Sales previous period diff.	

Using the aforementioned regressors, I constructed my machine learning models as outlined in the preceding chapter. The Table 5 presents the results, facilitating a comparison of the performance among the different models. Accuracy, log-score, and quadratic-score were employed as evaluation metrics for both in-sample and out-of-sample performance. All model metrics were computed using 5-fold cross-validation, so averaging the performance across each fold. Furthermore, I included the performance metrics of historical models discussed in the literature review, specifically those developed by Hu and Phillips (2004), Pauwels (2012), and Vasnev (2013). Since these studies considered three distinct time periods, I opted for a fairer comparison by presenting an average of the performance across these periods, aligning it with my models developed using K-fold cross-validation.

Table 5: Model predictive performance

	In-Sample Forecast			Out-of-sample Forecast		
Models | Scores	Accuracy (%)	Log-score	Quadratic-score	Accuracy (%)	Log-score	Quadratic-score
Logit	78.9	-0.49	0.71	70.4	-0.65	0.61
SVM	97.5	-0.68	0.64	72.7	-0.96	0.47
Decision Tree	87.6	-0.51	0.71	68.7	-Inf	0.50
Pruned Tree	83.3	-0.50	0.71	68.7	-Inf	0.50
Bagging	1	-0.15	0.94	72.9	-Inf	0.64
Random Forest	1	-0.15	0.95	73.8	-0.59	0.65
Optimal RF	1	-0.15	0.95	**74.9**	**-0.58**	**0.65**
			Past Models			
Hu&Phillips	78.0	-0.63	0.59	65.6	-1.31	0.45
Pauwels	-	-	-	64	-0.86	0.47
Vasnev	-	-	-	66.7	-0.63	0.65

Notes: the results of Hu and Phillips (2004) reported here were calculated by Pauwels (2012), because as I anticipated in the literature review Hu and Phillips had originally tested their model only within their sample. For the models of Pauwels (2012) and Vasnev (2013), unfortunately, no in-sample performance is available. The log score for some models has the value -Inf, this indicates that the model produced a predicted probability of zero for one or more observations in the test dataset. This could suggest that the model struggled to correctly generalize those particular observations.

In terms of in-sample accuracy, the Bagging, Random Forest, and optimal Random Forest models achieved the highest scores, all at 100%. They were closely followed by the SVM model with a score of 97.5% and the decision tree model with a score of 87.6%. Furthermore, the first three models also exhibited strong performance in terms of log-score and quadratic score. However, it is important to note that such high accuracy scores can be attributed to overfitting, whereby the models have learned the specific patterns and noise present in the training data. Moving to the out-of-sample performance, the optimal random forest model demonstrated the highest accuracy at 74.9%, followed by normal random forest model at 73.8% and bagging at 72.9%. Random Forest models also displayed better log-score and quadratic score compared to other models. Comparing these results with past models, the Hu & Phillips model achieved an in-sample accuracy of 78.0% and an out-of-sample accuracy of 65.6%. Pauwels' model showed an out-of-sample accuracy of 64%, while Vasnev's model had an out-of-sample accuracy of 66.7%. Overall, the results indicate that the optimal random forest model outperformed all other models developed in this thesis, as well as the models found in the existing literature. With higher accuracy, log-score, and quadratic-score values in both in-sample and out-of-sample forecasts, the random forest model consistently demonstrates a superior performance in my dataset. Its ability to capture complex relationships and handle high-dimensional data makes it the most effective among the various supervised learning models considered in this study. My *Hypothesis 1* therefore cannot be rejected.

To provide a more detailed analysis of the predictive performance of the models, out-of-sample accuracy was calculated for each of the three classes of the dependent variable (hike, no change, cut). The Table 6 below presents the results, calculated using 5-fold cross-validation. It also includes the accuracy of the past models for comparison:

Table 6: Model accuracy at class level

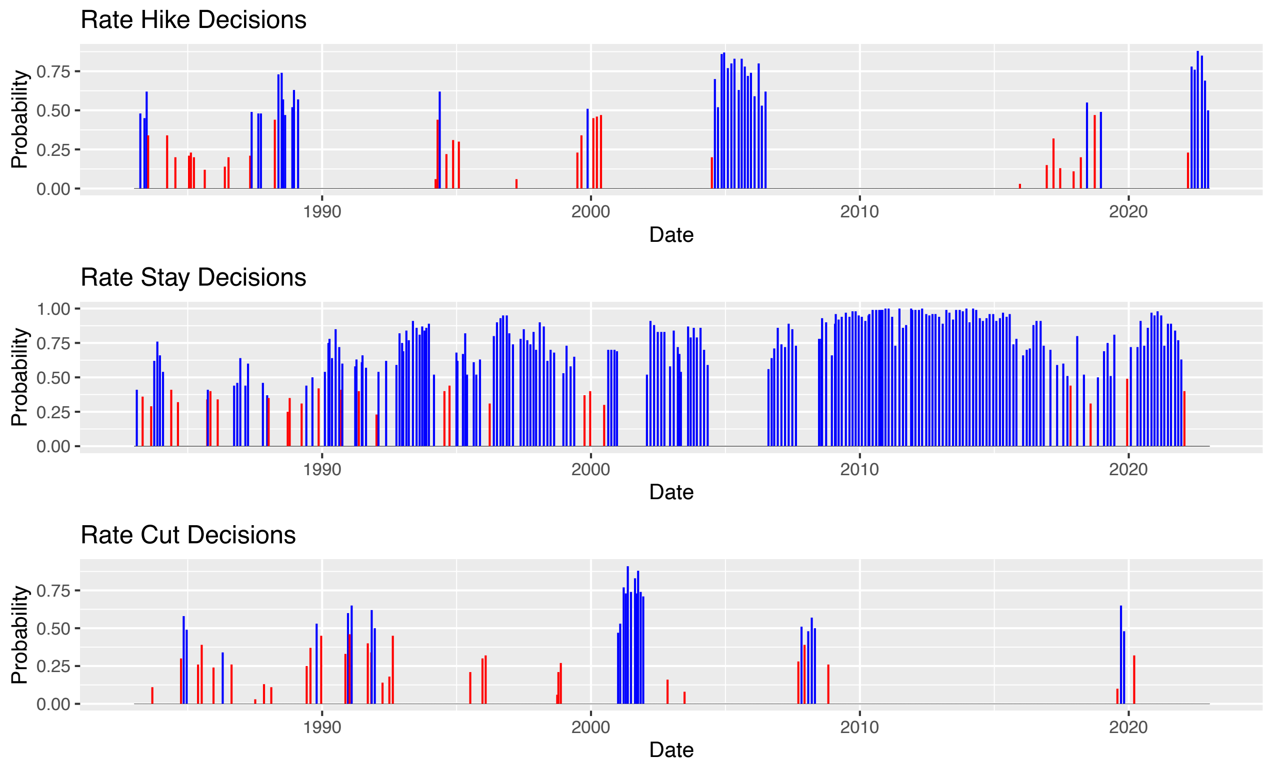
Out-of-sample Correct prediction (%)

Models | Classes	Cut	No Change	Hike	Total
Logit	45.6	83.0	53.75	70.4
SVM	39.3	84.9	**60.7**	72.7
Decision Tree	**49.0**	77.2	56.9	68.7
Pruned Tree	40.8	82.2	50.4	68.7
Bagging	45.3	86.4	54.0	72.9
Random Forest	39.3	**89.7**	54.0	73.8
Optimal RF	47.5	88.6	55.2	**74.9**
Hu&Phillips	42.8	73.7	60.5	65.6
Pauwels	33.6	73.9	62.7	64
Vasnev	44.4	78.3	38.9	67.6

Notes: again, the results reported for past models are calculated as an average of the accuracy over the three separate time periods considered in the original works.

As observed, the class with the highest accuracy among all the models is the "no change" class. This can be attributed to the slight data imbalance, as explained in the Data chapter. Since the FOMC decisions not to change target rate are more frequent, the models have been trained to predict this class with greater accuracy. Nonetheless, the models also demonstrate a good ability to predict rate hike and rate cut decisions. In fact, all models exhibit a higher probability than the 33.3 percent probability achieved by random class selection. Notably, the models exhibit a higher predictive ability for rate-raising decisions compared to rate-lowering decisions. The SVM model, which demonstrates the highest accuracy for rate-raising decisions, achieves an accuracy of 60.7 percent, while the Decision Tree model, which performs better for rate-lowering decisions, achieves an accuracy of 49 percent. Furthermore, the random forest model demonstrates the highest accuracy in predicting decisions to maintain rate unchanged. To provide a visual representation of the results, Figure 3 illustrates the (out-of-sample) predicted probabilities for each policy intervention using my best model, the optimal random forest.

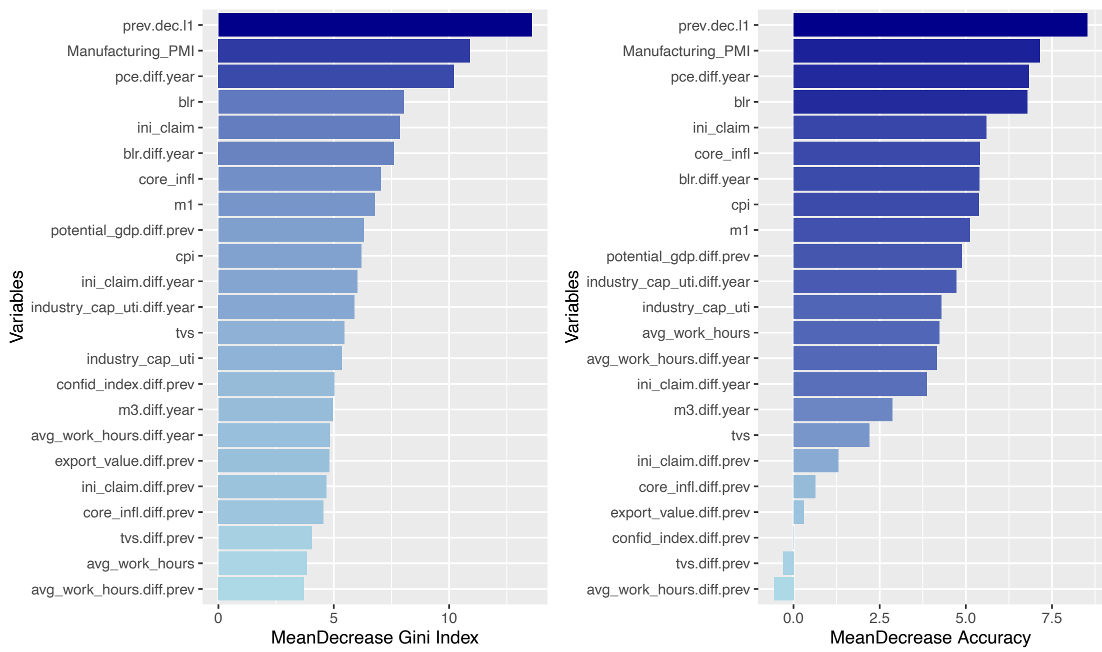
Figure 3: Optimal random forest out-of-sample predicted probabilities



The results encompass the entire time period from the beginning of 1983 to the end of 2022. Each graph within the figure corresponds to one of the three FOMC decisions regarding target rate, and the vertical rectangles depict the probabilities estimated by the optimal random forest for these decisions. Within this graphical representation, correct predictions are highlighted in blue, while incorrect predictions are marked in red. Also the graphical representation demonstrates the slight imbalance among the classes and the subsequent ability of the models to accurately predict decisions to maintain stable rate.

At this point I decided to use my best model, optimal random forest, to understand well the relationship between the regressors and my dependent variable. By utilizing the variable importance methodology outlined in the preceding chapter, my objective is to discern the economic indicators that exert the greatest influence on FOMC policy decisions.

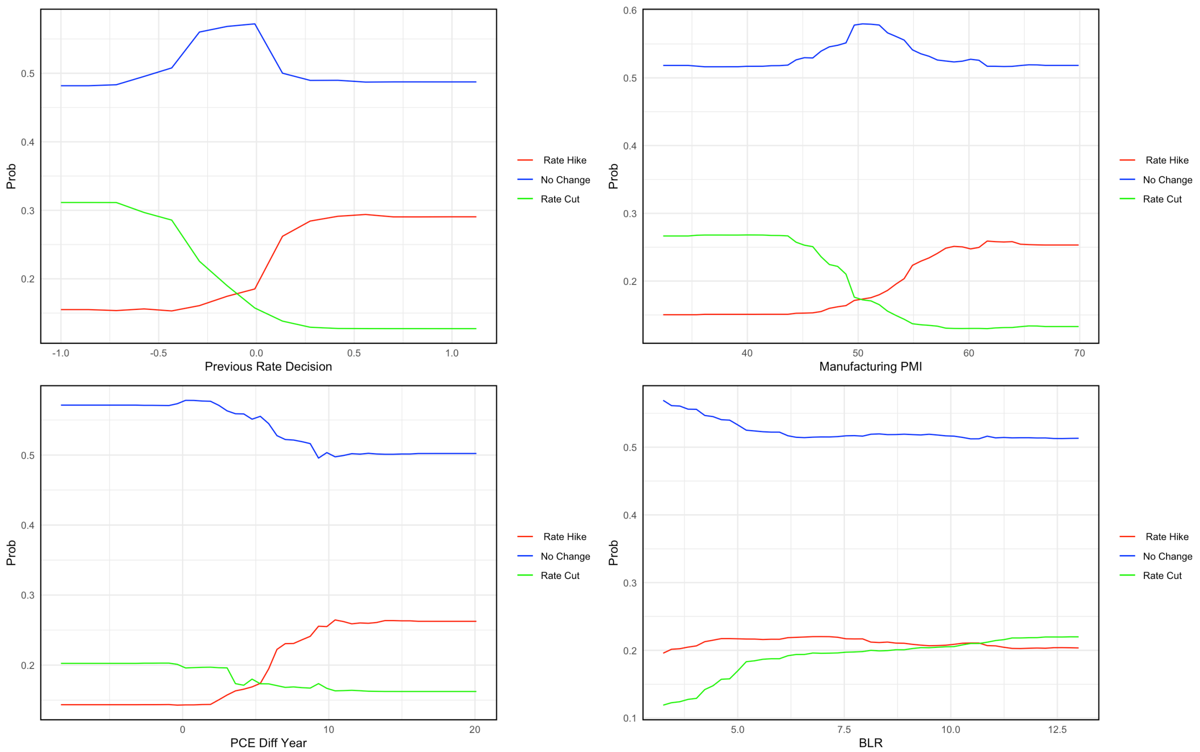
Figure 4: Variable Importance



From the Figure 4, we observe that the previous period's decision (prev.dec.l1) is the most important variable in terms of both decreasing accuracy and Gini index. This finding confirms the observed continuity in FOMC policy decisions and underscores the influence of past FOMC actions on the current decision-making process. Additionally, we observe that the Manufacturing Purchasing Managers' Index (Manufacturing_PMI), the year-on-year difference in personal consumption expenditures (pce.diff.year), the bank prime loan rate (blr), and initial jobless claims (ini_claim) for both evaluation methods remain among the top five most important variables This suggests that performance of the manufacturing industry, consumer behavior and spending trends, fluctuations in the bank loan rate and the level of jobless claims are likely to influence the FOMC policy decision.

Finally, I conducted a detailed analysis to examine how these significant variables impact the estimations of my top-performing model, the optimal random forest. Specifically, I utilized the partial dependence plot methodology discussed earlier to analyze the four most important variables. Figure 5 displays the partial dependence plots calculated using 5-fold cross-validation.

Figure 5: Partial Dependence Plots



These plots illustrate the probability predicted by the model for each of the three classes, as function of the values of the economic indicators. It is evident that the probability of decisions to keep rates unchanged consistently outweighs that of the other classes, once again emphasizing the impact of the slight dataset imbalance. Notably, for three of the variables, the relationship between the probabilities of rate hikes and rate cuts is inverse. Particularly, when considering the previous decision regarding the target rate, we observe the prevailing continuity effect of FOMC policy. If the previous change in the target rate is negative, the probability of a rate cut is higher, whereas if the previous change is positive, the probability of a rate hike is higher. Moreover, we observe that the probability of maintaining rates constant increases if there had been no previous change. Regarding the Manufacturing PMI, we observe that at lower levels, the probability of a rate cut is higher, which diminishes as economic activity in the manufacturing industry improves, while the probability of a rate hike increases. This outcome could be attributed to the Fed's intention to moderate the economy during robust growth and stimulate it during slowdowns. Once again, we notice that the probability of keeping rates steady peaks around the median value of the Manufacturing PMI. As for the year-on-year difference in personal consumption expenditures, we observe that at negative values of this difference, indicating a decrease in PCE, the probability of rate cuts is higher to stimulate the economy. Conversely, at positive and high values of this difference, the probability of rate cuts and maintaining rates constant decreases in favour of a higher probability of rate hikes, aligning with the objective of slowing down the economy. Finally, concerning the bank loan rate, a clear inverse relationship between the probability of rate cuts and rate hikes is not evident. However, it can be observed that both probabilities tend to increase as the bank loan rate increases, while the probability of keeping rates constant decreases.

6.2 Event Study Results

In this section, I will present the results of the event study conducted following the methodology outlined in the previous chapter. The first objective is to examine the impact of FOMC target rate change decisions on the stock market. To begin, I will analyze the results of a regression of the S&P 500 Cumulative Abnormal Returns (CARs) on the intercept, differentiating between events associated with rate hikes and rate cuts. For rate hikes, I anticipate obtaining a negative and statistically significant coefficient on the intercept. This is because an increase in the target federal funds rate immediately raises short-term borrowing costs for financial institutions, which subsequently affects borrowing costs for companies and consumers, resulting in a negative impact on the market index. Conversely, for rate cuts, I expect a positive and significant coefficient.

The results are presented in the Table 7. We observe a negative intercept coefficient of -0.75, significant at the 1 percent confidence level, for rate hikes. In contrast, we find a positive intercept coefficient of 1.65, significant at the 1 percent confidence level, for rate cuts. This implies that, on average, the S&P 500 experiences a negative cumulative abnormal return of 0.75 percent between five days before and after a rate hike decision, while it shows a positive cumulative abnormal return of 1.65 percent in the case of a rate cut decision. These results corroborate the prevailing theory that rate hikes exert a negative influence on the stock market, while rate cuts have a positive impact. Additionally, they highlight how the market reacts more significantly to a decrease in the target rate.

At this point I aim to test the *Hypothesis 2* of my study, which focuses on evaluating the ability of my best model to capture market participants' expectations regarding future decisions. To accomplish this, I performed again the previous regression of CARs by incorporating two dummy variables that identify decisions representing positive or negative surprises, as explained in Formula 12. My hypothesis cannot be rejected if I obtain a positive and statistically significant coefficient on the Pos_Surprise dummy for decisions to lower target rate, and a negative and significant coefficient on the Neg_Surprise dummy for decisions to raise target rate. This expectation is grounded in the belief that in the case of rate increases, a negative surprise - indicating a decision to raise the target rate to a higher level than expected - should result in a more pronounced negative impact on the market index. On the other hand, when rates are decreased, a positive surprise - indicating a decision to lower the target rate to a level lower than expected - is anticipated to have a more positive effect on the market index. The results are shown in the Table 7.

Table 7: S&P 500 CARs regressions

	Rate Hike		Rate Cut	
Regressor	(a)	(b)	(c)	(d)
Intercept	-.7648*** (-2.99)	-.6659** (-2.10)	1.6511*** (3.31)	1.0950** (1.89)
Pos_Surprise	-	2.6101 (0.56)	-	3.7163*** (5.19)
Neg_Surprise	-	-1.9589*** (-3.10)	-	.6171 (1.06)

*Notes: Columns (a) and (c) report the results of regressions of CARs on the intercept, and columns (b) and (d) report the results of regressions that include surprise components. All variables are expressed in percentage terms. In the case of (a) and (b) the sample consists of 66 observations related to different rate hike decisions, while in the case of (c) and (d) the sample is 49 observations related to different rate cut decisions. The parentheses contain the t-statistics. The t-statistics are calculated with standard errors estimated using a cluster formula to account for the correlations across the events, as previously stated. ***,**,* denote 1%, 10%, 20% significance level, respectively.*

For rate hike decisions, I obtained a negative intercept coefficient of -0.6659, which is statistically significant at the 10 percent level. Additionally, the coefficient for the Neg_Surprise dummy variable is -1.9589, also significant at the 1 percent level. These results indicate that when the Fed decides to raise rate, the S&P 500 experiences, on average, negative cumulative abnormal returns of 0.6659% over the five days before and after the decision. Furthermore, if the decision represents a negative surprise, the CARs become even more negative by 1.95%. The coefficient for the Pos_Surprise dummy variable is positive but not statistically significant. This suggests that a positive surprise may potentially mitigate the negative impact on the market index, but further analysis is required to confirm its significance. On the other hand, for decisions to cut rate, I obtained a positive intercept coefficient that is statistically significant at the 10 percent level. Additionally, the coefficient for the Pos_Surprise dummy variable is highly positive and significant at the 1 percent level. Specifically, in the case of a rate cut decision, the S&P 500 would experience an average CAR of 1.09%. If the decision represents a positive surprise, the cumulative return on the index would increase by an additional 3.71%. Again we observe higher coefficients (in absolute terms) for rate cut decisions, indicating a stronger market reaction to surprises associated with these policy decisions.

These findings align with Bernanke and Kuttner's results, suggesting that my model demonstrates the ability to capture the element of surprise associated with changes in the Federal funds target rate. Consequently, it seems to incorporate market participants' expectations into its predictions, providing valuable insights into the relationship between policy decisions and market behavior.

6.3 Backtesting strategy

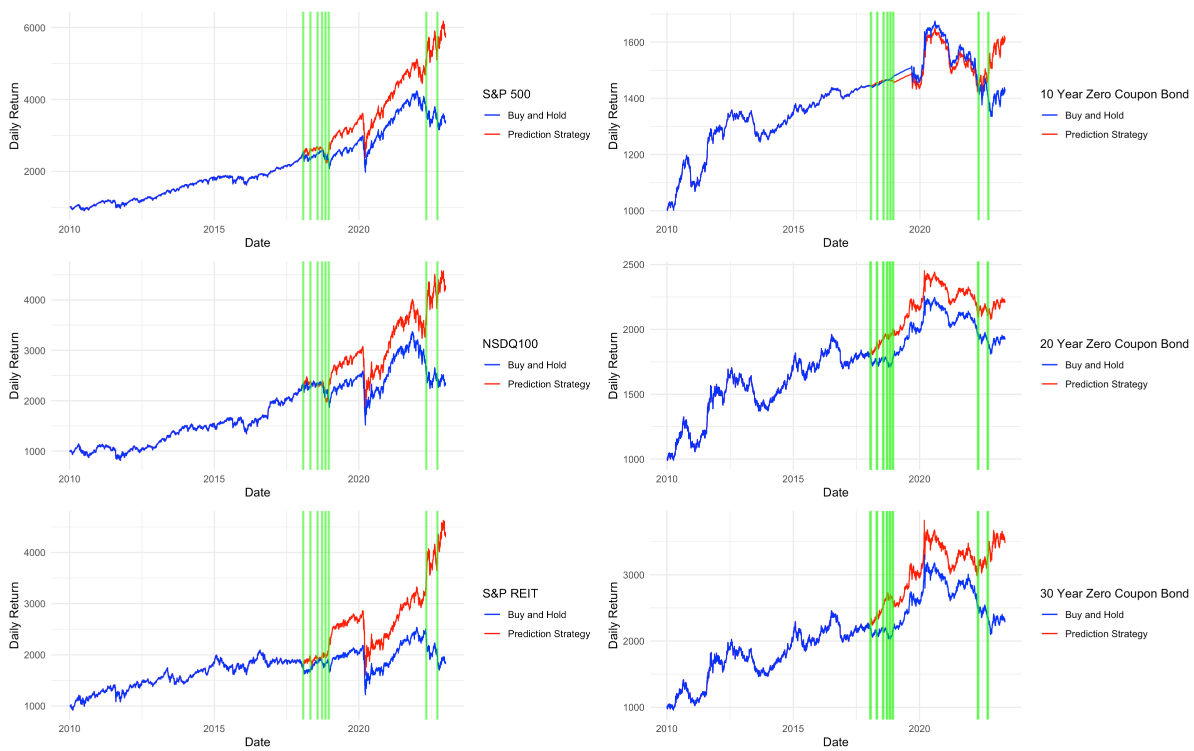
In the following section, I implemented a backtesting strategy to assess the practical validity of my models. Throughout this strategy, I consistently utilized my best model, the optimal random forest. I trained this model with observations prior to 2010, and attempted to predict the FOMC's decisions regarding the Federal funds target rate from the beginning of 2010 until the end of 2022. Subsequently, I developed a straightforward investment strategy focusing on forecasting rate hike decisions in both the stock and bond markets. For the stock market, my strategy concentrated on three market indices highly sensitive to changes in the Federal funds target rate: S&P 500, Nasdaq Financial-100, and S&P United States REIT. The approach involved taking short positions on these indices when my model predicted a rate increase decision. I expected such decisions to exert a negative impact on these indices. Specifically, I initiated the short position 15 days before the FOMC meeting and closed it 5 days later. In remaining days, I adopt a long position on the indices, as for a buy-and-hold strategy. Regarding the bond market, I developed my strategy on U.S. Zero Coupon Bonds with maturities of 10, 20, and 30 years. I selected long-term ZCB bonds due to their higher price sensitivity to interest rate changes and as a matter of computational simplicity. Historical fitted yield data were used to calculate the prices of these bonds, assuming I purchased each bond at the beginning of the strategy as if it were newly issued. For bonds that reached maturity, I reinvested in new bonds with the same maturity at the current price determined by the current fitted yield. Like the stock market strategy, I took short positions from 15 days before to 5 days after a Fed decision to raise rate, while maintaining long positions for the remainder of the time. The results of this strategy for both the stock and bond portfolios are presented in the Table 8. Additionally, I included the returns of a simple buy-and-hold strategy for each security to facilitate comparison. The reported returns represent the cumulative sum of the daily returns of the strategies.

Table 8: Backtesting strategy results

	Stock Market			Bond Market		
	S&P 500	NSDQ100	S&P REIT	10y ZCB	20y ZCB	30y ZCB
Prediction Strategy	196.76%	175.25%	178.53%	48.98%	87.62%	146.81%
Buy-and-hold Strategy	142.73%	115.43%	92.87%	36.95%	73.84%	104.65%
Difference	54.03%	59.82%	**85.66%**	12.03%	13.78%	**43.16%**

The strategy based on my models' forecasts demonstrates better performance compared to the buy-and-hold strategy for each individual security. The difference in returns between the two strategies over the 12-year investment period is not substantial, this is primarily due to the limited number of rate-cutting decisions predicted by the model during this timeframe, which amounts to only 8. We can see that the largest return in the case of the stock market were obtained by the S&P United States REIT. This outcome is reasonable considering that the real estate sector is one of the stock market sectors most exposed to changes in Federal funds rate. Conversely, in the bond market, the 30-year Zero Coupon Bond (ZCB) records the highest returns. This finding aligns with expectations as bonds with longer maturities exhibit higher duration and therefore greater sensitivity to rate changes. Figure 6 presents a graphical representation of the different strategies. The graph depicts an index with an initial value of 1000, showing variations based on the daily returns of each strategy. The vertical green bands represent periods when the prediction strategy takes a short position on stocks.

Figure 6: Graphical representation of the backtesting Strategy



7. Conclusion

This paper is divided into two main parts. The first part focuses on analyzing the behavior of the Federal Reserve during policy decisions. Specifically, I aim to develop models capable of predicting the FOMC's decisions regarding the Federal funds target rate. While traditional econometric models have been extensively explored in the academic literature for this purpose, my contribution lies in the application of new techniques, such as machine learning algorithms, to achieve better performance in terms of metrics like accuracy and log/quadratic-score. To accomplish this, I developed seven different machine learning models and assessed their performance using 5-fold cross-validation. The results consistently support my hypothesis, indicating that my models outperform traditional econometric models across all metrics. Notably, the random forest model with optimized predictor selection demonstrates the highest predictive precision. I further analyze the models' specific accuracy in predicting the three types of target rate decisions (hike, no change, cut), and find that they excel in recognizing unchanged rates due to dataset characteristics. However, the models also demonstrate good performance in predicting rate hike and rate cut decisions, with better accuracy for the former. Additionally, I employ the variable importance methodology to identify the macroeconomic variables that have the greatest influence on the Fed's policy decisions, and the partial dependence plot approach to gain insights into the nature of their influence.

The second part aims to analyze the impact of FOMC decisions on financial markets. Here, I leverage the models developed in the first part, particularly their predictive errors, to distinguish between expected and unexpected monetary policy actions. This approach represents another valuable contribution to the academic literature, which has traditionally relied on Federal funds futures data to gauge policy expectations. To test this hypothesis, I first conduct an event study to investigate the effects of target rate decisions on the market index. Subsequently, I perform regression analysis on the Cumulative Abnormal Returns (CARs) derived from the event study, incorporating dummy variables that represent unexpected positive and negative policy decisions. The obtained results support my second hypothesis, with significant positive coefficients for positive surprises during rate cuts, indicating a stronger positive impact on the stock market. Conversely, significant negative coefficients for negative surprises during rate hikes imply a greater negative impact.

Finally, I implement a short backtesting strategy that provides supporting evidence for the practical validity of my predictive models.

However, it is important to acknowledge a limitation of this work. The dataset exhibits a slight imbalance due to the greater frequency of decisions to maintain rates unchanged compared to rate cut or hike decisions. As these decisions significantly affect financial markets and the overall economy, the Fed exercises caution when making such choices. This slight imbalance, combined with the limited number of observations, may lead machine learning models to overfit the training set. This is evident in my models' strong out-of-sample performance in recognizing decisions to maintain rates constant. One potential solution to address this issue is the utilization of resampling techniques like random oversampling or Synthetic Minority Oversampling Technique (SMOTE). However, even when employing these techniques, I obtained similar results. Therefore, I ultimately decided to utilize the original dataset. Future research avenues to tackle this problem could involve exploring more efficient resampling methodologies or utilizing larger, less imbalanced datasets along with more advanced machine learning models.

In conclusion, this study demonstrates the effectiveness of machine learning models in predicting Federal Reserve's target rate decisions and analyzing their impact on financial markets. The findings support the superiority of these models over traditional econometric approaches. Overall, this research contributes to improving predictive models and understanding the relationship between monetary policy decisions and financial markets.

References

Bernanke, Ben S. and Kuttner, Kenneth N. and Kuttner, Kenneth N. (2005), “What Explains the Stock Market's Reaction to Federal Reserve Policy? “.

Bernanke, Ben S., and Mark Gertler, (1999), “Monetary policy and asset price volatility”, *Federal Reserve Bank of Kansas City Economic Review 84*, 17–51.

Breiman, L., (2001), “Random Forests”, Machine Learning 45, 5–32.

Chow, Lin, (2014), “Chow-Lin Method for Temporal Disaggregation”, *Memobust Handbook on Methodology of Modern Business Statistics,* Eurostat.

Cook T., Hahn T. (1988), “The effect of changes in the federal funds rate target on market interest rates in the 1970s”, *Journal of Monetary Economics*, Volume 24,Pages 331-351.

Dueker M.J. (1999), "Measuring monetary policy inertia in target Fed funds rate changes," Review, Federal Reserve Bank of St. Louis, vol. 81(Sep), pages 3-10.

Friedman, Jerome H (2001), “Greedy function approximation: A gradient boosting machine.” Annals of statistics:1189-1232.

Gurkaynak, Refet S., Brian P. Sack, and Eric T. Swanson, (2004), “The effects of monetary policy on asset prices: An intraday event–study analysis”, Working paper, Board of Governors of the Federal Reserve System.

Hamilton, J. D., Jorda, O., (2002), “A model of the federal funds rate target. Journal of Political Economy”, 110(5):1135–1167.

Hu, Ling, and Peter C. B. Phillips (2004), “Dynamics of the Federal Funds Target Rate: A Nonstationary Discrete Choice Approach.”, *Journal of Applied Econometrics*, vol. 19, no. 7, pp. 851–67.

James G., Witten D., Hastie T., Tibshirani R. (2013), “An Introduction to Statistical Learning: with Applications in R.”, New York: Springer.

Kim H., Jackson J., Saba R., (2009), "Forecasting the FOMC's interest rate setting behavior: a further analysis*," Journal of Forecasting*, vol. 28(2), pages 145-165.

Kuttner, Kenneth N., (2001), "Monetary policy surprises and interest rates: Evidence from the Fed funds futures market," *Journal of Monetary Economics*, vol. 47(3), pages 523-544.

Kuttner, Kenneth N., (2003), "Dating changes in the Federal funds rate", 1989–92, Working paper, Oberlin College.

MacKinlay A.C., (1997), "Event Studies in Economics and Finance," *Journal of Economic Literature*, American Economic Association, vol. 35(1), pages 13-39.

Pauwels, Laurent L. and Vasnev, Andrey L., (2012), "Forecast Combination for Discrete Choice Models: Predicting FOMC Monetary Policy Decisions".

Rigobon R., and Sack B., (2003), "Measuring the reaction of monetary policy to the stock market", *Quarterly Journal of Economics*,118, 639–670.

Rigobon R., Sack B., (2004) "The impact of monetary policy on asset prices", *Journal of Monetary Economics*, vol. 51(8), pages 1553-1575.

Rosa C., (2019), "The Impact of Federal Reserve Monetary Policy Announcements on Asset Prices: A Review", *Rivista Internazionale di Scienze Sociali*, vol. 127(1), pages 3-24.

Taylor, John B., (1993), "Discretion versus policy rules in practice*", Carnegie-Rochester Conference Series on Public Policy*, vol. 39(1), pages 195-214.

Vasnev A., Skirtun, Pauwels, (2013), "Forecasting Monetary Policy Decisions in Australia: A Forecast Combinations Approach", *Journal of Forecasting,* vol. 32(2), pages 151-166.

Winkler, R. L. (1996), "Scoring rules and the evaluation of probabilities (with discussion). Test", 5(1):1–60.

Winkler, R. L., (1993), "Evaluating probabilities: asymmetric scoring. Management Science", 4:1395–1405.

Summary

Predictive Modelling of FOMC Decisions and Market Reaction: A Machine Learning Approach

Abstract

This paper focuses on two primary objectives: developing models to predict Federal Open Market Committee (FOMC) decisions regarding changes in the Federal funds target rate and analyzing the stock market reaction to these decisions. Traditional econometric models have been extensively studied in the literature, but this research employs innovative machine learning algorithms to enhance predictive performance. Seven different machine learning models are developed. The results demonstrate the superiority of the machine learning models in terms of accuracy, log-score, and quadratic-score, with the optimal random forest model performing the best. Furthermore, the study employs variable importance methodology and partial dependence plots to identify the macroeconomic variables that significantly influence FOMC decisions. The second part of the work focuses on the impact of FOMC decisions on the stock market, distinguishing between expected and unexpected choices. Event study analysis is conducted using S&P 500 data, and regression analysis of CARs demonstrates a significant market response to unexpected changes in the Federal funds target rate. Finally, the practical validity of the predictive models is confirmed through a backtesting strategy.

1. Introduction

Decisions regarding Federal funds target rate have significant implications for the economy, including employment, growth, and inflation. They indirectly affect various short-term interest rates, such as those for loans, as lenders often base their rates on the prime lending rate, which is influenced by the fed funds rate. Changes in the target rate can have a strong impact on the stock market, with even a small decline leading to increased market activity and lower borrowing costs for companies. Therefore, anticipating FOMC decisions is crucial for market participants to adjust their investment strategies. Morover, examining the impact of these decisions on security prices is a topic of great interest to investors and policymakers since it can provide additional insights into the ways in which monetary policy affects financial markets and market expectations. This paper is situated within this context, aiming to address the following research questions: “Can machine learning models accurately predict FOMC decisions regarding changes in the Federal funds target rate?”; “What is the impact of these policy decisions on the stock market when considering market expectations?”. To this end, the structure of this paper is divided into two parts: firstly, development of models capable of predicting FOMC decisions regarding changes in the Federal funds target rate, and secondly, analysis of the impact of these policy decisions on the stock market by considering market expectations and thus differentiating between expected and unexpected decisions.

2. Literature

Both topics have been extensively studied in the academic literature. Regarding the forecasting of target rate, we first have the seminal work of Taylor (1993) who proposed a policy rule for forecasting Federal fund rate. However, this approach does not account for the discrete nature of FOMC decision-making. Subsequent work, such as that of Hu and Phillips (2004), addressed this limitation by developing a discrete choice approach to predict policy actions. Their model calculated threshold coefficients representing the gaps between estimated and actual target rates that trigger adjustments by the FOMC. While Hu and Phillips (2004) demonstrated good predictive performance within their estimation sample, they did not provide an estimate of the out-of-sample performance. Pauwels (2012) replicated their model and tested its out-of-sample performance, comparing it with other models developed using combined forecast methodology and introducing metrics like log-score and quadratic-score. Another relevant study by Vasnev (2013) replicated Pauwels' work but focused on the policy actions of the Reserve Bank of Australia.

On the other hand, regarding the academic literature related to the impact of changes in the target Federal funds rate on asset prices, a first very relevant paper is that of Kuttner (2001). In his research, Kuttner estimated the impact of monetary policy actions on the bond market by distinguishing between expected and unexpected decisions through federal funds futures market data. His results showed a strong response to unanticipated rate changes and a weaker response to anticipated changes. Another important subsequent study that is a key reference for this work is that of Bernanke and Kuttner (2005). They analyzed the impact of FOMC decisions on the stock market using Kuttner's (2001) previous approach to distinguish between anticipated and unanticipated components of policy decisions. As a result, they obtained that the stock market response to a surprise increase in target rate is negative and significant. Specifically, a 25-basis-point rate cut would lead to a 1-day stocks return of 1%. Therefore, they also discovered evidence supporting a greater stock market reaction to unexpected decisions.

3. Hypotheses and Contribution

Based on the previous research questions and the literature review conducted, I formulated two main research hypotheses:

- Hypothesis 1: *My machine learning models are expected to outperform the conventional econometric models developed in the academic literature in terms of accuracy and log/quadratic-score*.
- Hypothesis 2: *The developed machine learning models are capable of capturing market participants' expectations about future FOMC decisions*.

The second Hypothesis implicitly assumes that market participants are rational individuals who base their expectations on macroeconomic variables. By testing these hypotheses, I would make the following contributions to the literature: firstly, apply innovative techniques, specifically machine learning algorithms, to predict policy choices regarding the fed funds target rate; secondly, analyze the impact of policy decisions on the stock market, distinguishing between expected and unexpected choices through an alternative method based on my ML models.

4. Data

This section covers the dataset used for analysis, encompassing monthly observations from February 1983 to December 2022. To construct machine learning predictive models, the dependent variable and regressors were defined. The dependent variable represents FOMC policy decisions on the target rate, resulting in 399 observations, later reduced to 352 by considering the latest decision in months with multiple meetings. The dataset includes decision dates and target rate changes (hike, cut, no change). It exhibits a distribution of 70 rate hikes, 57 rate cuts, and 225 no changes. This slightly imbalanced distribution is representative of the nature of FOMC policy decisions: as rate hike/cut decisions significantly affect financial markets and the economy as a whole, the Fed exercises some caution in making these choices. Concerning the selection of macroeconomic indicators for inclusion in empirical models as independent variables, I initially included 22 economic series, Table 1 shows the description of this indicators. The frequency was monthly, except for GDP and Potential GDP, which were quarterly, transformed into monthly series for consistency using a method for temporal disaggregation in R. Key indicators like the Output Gap and Inflation Gap were calculated from GDP, Potential GDP, and inflation rates. Then, in order to carry out the event study to analyze the impact of policy decisions about target rate on the stock market, I collected daily S&P 500 data as proxy for the market. Most of the data were downloaded from the Federal Reserve Economic Data (FRED) online database of the Federal Reserve Bank of St. Louis. Only data relative to the S&P 500 and Manufacturing PMI were obtained from WRDS (Wharton Research Data Services) and the Institute for Supply Management website, respectively.

4.1. Data Preprocessing

Once retrieved, I preprocessed the raw data in such a way as to create the variables for my forecast models. For each economic indicator I calculated the difference from the previous period and the same period in the previous year. This approach allowed us to capture changes and trends in macroeconomic variables over time and to focus on relative changes rather than absolute values. Using both these differences and the indicators themselves, I obtained a total of 72 independent variables to include in the first estimation stage. I introduced an additional variable linked to the previous FOMC meeting's target rate change, enabling my models to consider both the historical impact of prior decisions and the continuity of FOMC's policy stance. To avoid multicollinearity problems, I calculated the correlation between all macroeconomic variables. The results are shown in Figure 2. Highly correlated variables, with a threshold set at 0.7, were pruned to retain 56 relevant regressors. Additionally, to prevent any potential look-ahead bias, all macroeconomic regressors have been lagged by one period, ensuring that observations on economic indicators precede the FOMC's decision on the target rate. Finally, through a stepwise forward selection, I filtered the most informative and relevant features that contribute most to the predictive performance of a machine learning model. This resulted in a final selection of 23 regressors for inclusion in the models, shown in Table 4.

5. Methodology

5.1 Forecasting FOMC decisions

Forecasting FOMC decisions on target Federal funds rate can be accomplished through either a continuous approach or a discrete approach. I opted for the discrete approach, consistent with the academic literature and the nature of FOMC policy actions.

Regarding the choice of models for predicting FOMC decisions, I decided to use a logistic regression as a baseline model, as extensively done in the academic literature. The logistic model employs a log-odds transformation to estimate the probabilities of the three classes in my specific case: rate hike, rate cut, or no change. I then decided to develop several machine learning models, which offer distinct advantages over traditional econometric models such as handling complex data and detecting nonlinear relationships. Specifically, I developed 6 different machine learning models: Support Vector Machine, Decision Tree, Pruned Tree, Bagged Tree, Random Forest, Optimal Random Forest.

SVM model, is a powerful classification algorithm that builds upon the concepts of hyperplanes and the maximal margin classifier. SVMs aim to find the optimal hyperplane that not only separates the classes but also maximizes the margin while allowing for some misclassification. Subsequently, I delved into Tree-based Models, commencing with a Decision Tree tailored for classification problems. This method follows an algorithm that involves partitioning the predictor space into distinct regions based on the minimization of the Gini index, a measure of node purity, which is defined by the number of observations in a node from a single class. However, decision trees tend to suffer from high variance, making them sensitive to data changes and prone to overfitting. To address this, I applied the pruning technique. It involves constructing a large tree and then trimming it based on minimizing misclassification errors using cost complexity pruning. Next, Bagging is another technique focused on reducing the variability of statistical learning methods, particularly decision trees. It involves creating numerous prediction models using bootstrapped training sets and averaging their predictions to enhance predictive accuracy. Random Forests build upon the concept of bagging by introducing a strategy to mitigate the inter-tree correlation issue. Rather than considering all predictors at each split, Random Forests select a random subset of predictors, m, for each split. This approach aims to make the resulting trees less correlated and more dependable. The Optimal Random Forest Model further refines this approach by identifying the best number of regressors, m. This is achieved through a systematic process that seeks to minimize the Out-of-Bag (OOB) estimate of test error. OOB error measures model performance by evaluating each tree on the training data not used during its creation. The optimal m, which leads to the lowest associated OOB error, is identified and used to develop the Optimal Random Forest.

These algorithms are powerful for enhancing prediction accuracy, but they often sacrifice interpretability due to the complexity of models. To address this, assessing variable importance becomes crucial. Breiman (2001) and Friedman (2001) propose methods for this purpose. The first method employs the Gini index, calculating the decrease in Gini index resulting from splits over a predictor across all trees. This average decrease indicates the importance of each predictor. Alternatively, the second approach uses the decrease in accuracy. Another significant tool is the partial dependence plot, for analyzing feature effects on model outcomes. It establishes the relationship between the target variable and a specific feature.

5.1.2 Evaluation Metrics

Accuracy or error rate is a commonly used metric in classification, gauging the percentage of correct predictions by a model. It is calculated simply as the ratio of the number of correct predictions of the model to the total number of predictions. Yet, accuracy often overlooks class prediction probabilities, reducing analysis precision. To address this, Pauwels (2012) introduced two additional metrics: the log-score and the quadratic-score. The log-score considers predicted probabilities' logarithms, capturing the model's confidence in predictions:

$$S^l = \log(\hat{P}_j) \tag{1}$$

where $\hat{P}_j$ is the probability predicted by the model for the state that actually happens. Higher log-score values indicate more accurate and confident predictions. Similarly, quadratic-score considers both the predicted probabilities and their proximity to the true class labels. It quantifies the discrepancy between predicted probabilities and actual outcomes, rewarding predictions that are close to the true values. In my specific case the quadratic-score rule is given by:

$$S^q = 2\hat{P}_j - (\hat{P}_{-1}^2 + \hat{P}_0^2 + \hat{P}_1^2) \tag{2}$$

These metrics offer a more comprehensive understanding of the model's reliability, precision, and confidence in its predictions.

5.1.3 K-fold Cross Validation

A 5-fold cross-validation strategy was employed across all models. This technique offers superior model generalization compared to single train-test splits. By dividing the dataset into folds, each data point is included in the test set, ensuring reliable assessment of model generalization. Averaging performance metrics across folds provides a robust evaluation of predictive power. Cross-validation also diminishes bias from dataset divisions, enhancing model performance consistency.

5.2 Analysis of the impact of policy decision on stock market

To study how unexpected decisions regarding a change in Federal fund target rate impacts the stock market, I employed the Event Study methodology introduced by Craig MacKinlay (1997). The event study is a technique to measure the effect of economic event on the value of firms. The first step in conducting an event study is to identify the event of interest, which in this case is the announcement of a change in the Federal funds target rate by the Federal Reserve.

At this point, my interest lies in distinguishing between expected and unexpected FOMC policy decisions. To achieve this, I have revised my machine learning models, adopting a continuous approach instead of a discrete one. Consequently, my new dependent variable represents the newly determined Federal fund target rate at each FOMC meeting. This alteration enabled error calculation by comparing predicted and actual target rates. Employing a random forest model with 5-fold cross-validation facilitated error estimation for each data point. I categorized positive errors as negative surprises, where the actual rate exceeded expectations, and vice versa for negative errors. Subsequently, I arranged the errors in descending order and generated two new dummy variables: one denoting negative surprises, assigned a value of 1 for the top 20 observations with the highest errors (corresponding to the 95th percentile); the other representing positive surprises, assigned a value of 1 for the last 20 observations with the lowest errors (corresponding to the 5th percentile). As a result, each event was associated with three variables: the two surprise-related dummy variables and one indicating the sign of the change in the federal fund target rate. My analysis focuses on examining the impact of policy decisions on the overall U.S. stock market. Therefore, I selected the S&P 500 as a proxy variable for studying market movements.

The next step in conducting an event study involves defining the event window and estimation window. The event window spans from 5 days before to 5 days after the event, capturing early market responses. The estimation window ranges from 90 days before to 6 days before the event, used for calculating Normal Returns. Employing a mean adjusted, I determine the expected return as the average during the estimation window. I then compute Abnormal Returns (ARs) in the event window, which represent the difference between actual and expected returns, aiding in isolating event impacts. Calculating Cumulative Abnormal Returns (CARs) by summing ARs for each event, I conduct regression analysis using two regressions based on rate change direction. The regression equations take the following form in both cases:

$$CAR_i = \alpha + \beta PosSurprise_i + \gamma NegSurprise_i + \varepsilon_i \tag{3}$$

In the first regression, only CARs associated with a FOMC decision to raise rate are considered, while in the second regression, only CARs associated with a FOMC decision to lower rate are considered. Ensuring the method's assumptions are met, I consider cluster-based standard errors to accommodate correlated Federal Reserve policy decisions.

6. Results

In the forthcoming chapter, I will present the results of the previously outlined methodologies, categorizing them into two sections: the first section will evaluate the predictive performance of my machine learning models, and the second section will explore the outcomes of the event study. Lastly, I will outline a straightforward investment strategy based on the predictive capabilities of my models.

6.1 Forecasting Results

Table 5 presents a comprehensive overview of the performance of the 7 different models described and of historical models discussed in the literature review, using evaluation metrics like accuracy, log-score, and quadratic-score for both in-sample and out-of-sample predictions. Notably, Bagging, Random Forest, and Optimal Random Forest models achieved the highest in-sample accuracy scores at 100%, followed by SVM and decision tree models. However, such accuracy scores might result from overfitting. For out-of-sample performance, the Optimal Random Forest model displayed the highest accuracy at 74.9%, followed closely by the normal Random Forest and Bagging models. Random Forest models also displayed better log-score and quadratic score compared to other models. Comparing with past models, the Optimal Random Forest model outperformed all others, exhibiting superior accuracy, log-score, and quadratic-score values, affirming its efficacy in capturing complex relationships and handling high-dimensional data. This supports the verification of Hypothesis 1.

Table 6 presents out-of-sample accuracy results for each specific class of the dependent variable (hike, no change, cut) across the models using 5-fold cross-validation, including historical models for comparison. Notably, the "no change" class achieves the highest accuracy across models due to data imbalance, as elaborated in the Data chapter. Since the FOMC decisions not to change target rate are more frequent, the models have been trained to predict this class with greater accuracy. However, the models also exhibit good ability in predicting rate hike and cut decisions, surpassing the random class selection probability of 33.3%. Particularly, the models perform better in predicting rate-raising decisions than rate-lowering ones, with SVM achieving 60.7% accuracy for hikes and Decision Tree reaching 49% for cuts. The random forest model excels in predicting unchanged rate decisions.To provide a visual representation of the results, Figure 3 illustrates the (out-of-sample) predicted probabilities for each policy intervention using my best model, the optimal random forest. Also the graphical representation demonstrates the slight imbalance among the classes and the subsequent ability of the models to accurately predict decisions to maintain stable rate.

Using the optimal random forest model, I employed the variable importance methodology to uncover the key economic indicators influencing FOMC policy decisions. Figure 4 highlights that the previous period's decision (prev.dec.l1) holds the most significant impact, emphasizing the continuity in FOMC actions. Moreover, variables such as Manufacturing Purchasing Managers' Index (Manufacturing_PMI), year-on-year difference in personal consumption expenditures (pce.diff.year), bank prime loan rate (blr), and initial jobless claims (ini_claim) also ranked high in importance. This suggests that performance of the manufacturing industry, consumer behavior and spending trends, fluctuations in the bank loan rate and the level of jobless claims are likely to influence the FOMC policy decision.

Furthermore, I conducted a detailed assessment of the impact of the two most influential variables on the optimal random forest predictions using partial dependence plots. Figure 5 depicts these plots, illustrating predicted probabilities for each class as a function of economic indicator values. The plots demonstrated that probability of decisions to keep rates unchanged consistently outweighs other classes due to dataset imbalance. Notably, the relationship between rate hikes and cuts is inverse for the previous period's decision variable, indicating a continuity effect in FOMC policy. Regarding the Manufacturing PMI, we observe that at lower levels, the probability of a rate cut is higher, which diminishes as economic activity in the manufacturing industry improves, while the probability of a rate hike increases. This outcome could be attributed to the Fed's intention to moderate the economy during robust growth and stimulate it during slowdowns.

6.2 Event Study Results

This section presents the key outcomes of the event study conducted as outlined in the previous chapter. The results are summarized in Table 7.

The primary aim was to assess the impact of FOMC target rate change decisions on the stock market. The regression analysis of the S&P 500 Cumulative Abnormal Returns (CARs) on the intercept was executed for both rate hike and rate cut decisions, revealing significant insights. For rate hikes, a notable negative intercept coefficient of -0.76 was observed, signifying a 0.76% average negative cumulative abnormal return around the decision period. Conversely, a positive intercept coefficient of 1.65 was found for rate cuts, indicating a 1.65% average positive cumulative abnormal return. These results align with the common understanding that rate hikes negatively affect the stock market, while rate cuts yield a positive impact, underlining the market's heightened response to rate cuts.

Furthermore, to test Hypothesis 2, the regression was extended by incorporating dummy variables representing positive and negative surprises. For rate hikes, a statistically significant negative intercept coefficient of -0.66 was achieved, alongside a significant coefficient of -1.95 for the Neg_Surprise dummy variable. These outcomes indicate that during rate hikes, the S&P 500 experiences an average negative cumulative abnormal return of 0.66%, increasing by 1.95% in the case of a negative surprise decision. In contrast, decisions to lower rates yielded a positive and significant intercept coefficient, along with a highly positive coefficient for the Pos_Surprise dummy. Specifically, in rate cut decisions, the S&P 500 witnessed an average CAR of 1.09%, increasing by an additional 3.71% with a positive surprise decision. Notably, rate cut decisions generated higher coefficients, underlining a greater market response to surprises linked with these policy choices.

These results suggest that the market reacts significantly to unexpected changes in target fed funds rate, aligning with Bernanke and Kuttner's (2005) findings. So, Hypothesis 2 cannot be rejected as the method based on the prediction errors of my machine learning models are likely to capture market expectations about changes in the Federal funds target rate.

6.3 Backtesting strategy

In this section, a backtesting strategy was executed to evaluate the practical validity of the optimal random forest model. This model, trained on pre-2010 data, was employed to predict FOMC decisions on the Federal funds target rate from 2010 to 2022. Two investment strategies were formulated: one focusing on rate hike predictions in the stock market and the other in the bond market. For stocks, a short position was taken on indices like S&P 500, Nasdaq Financial-100, and S&P United States REIT from 15 days before to 5 days after a predicted rate increase, while long positions were maintained during other times. In the bond market, a similar approach was adopted with U.S. Zero Coupon Bonds of varying maturities. From Table 8, results showed the prediction-based strategy outperformed the buy-and-hold approach, with the most significant returns observed in the S&P United States REIT and 30-year Zero Coupon Bond. This demonstrated the effectiveness of the model-based strategy. A visual representation of the strategies was presented in Figure 6.

7. Conclusion

This paper is structured into two main sections. The first part focuses on analyzing the Federal Reserve's behavior during policy decisions, aiming to develop models for predicting FOMC decisions on the Federal funds target rate. The results show that the application of innovative machine learning algorithms provides a more accurate and reliable prediction of FOMC decisions regarding the Federal funds target rate compared to traditional models found in the academic literature. The second part analyzes FOMC decision impacts on financial markets, leveraging predictive errors of models to distinguish expected and unexpected actions. I found that by attempting to capture market expectations through a machine learning methodology based on macroeconomic variables, we observe a significant reaction of market participants to unexpected changes in the fed funds target rate. These results support the conclusion that most market participants seem to be rational individuals who primarily base their monetary policy expectations on macroeconomic information.

A mio papà.