Introduction

Predicting inflation is an important issue in macroeconomics, being paramount to effective monetary and fiscal policy. Throughout history there have been numerous examples of when inflation went out of control and both everyday people and governments faced the severe consequences. The most famous example is likely the Weimar Republic in post-World War 1 Germany, where hyperinflation took hold. This paper takes a practical approach to this issue and utilizes several machine learning algorithms to address this challenge. The methods used include LASSO regression, Random Forest regression, Neural Network Regression, and a Super Learner technique that combines the results of the previous regression techniques. "Despite the fact that these methods already existed in the early 2000s, for a long time they remained almost unnoticed in the professional literature related to the forecasting of inflation in general" (Baybuza 2018). This study builds off the findings of Ivan Baybuza in his paper titled "Inflation Forecasting using Machine Learning Methods" and Sharon Kozicki in her paper "Predicting Real Growth and Inflation with the Yield Spread." The key findings of this paper are that the yield spread is an effective leading indicator of inflation, limiting the dataset to more recent data improves the accuracy of prediction, and that Random Forest and the Super Learner techniques are the most effective inflation prediction algorithms of the techniques tested.

Literature Review

The Phillips curve is a constant inverse relationship between unemployment and inflation. "Although he had precursors, A. W. H. Phillips's study of wage inflation and unemployment in the United Kingdom from 1861 to 1957 is a milestone in the development of macroeconomics" (Hoover et al 2019).

As for the motivation of this study, "Some economists see a puzzle: inflation has not fallen as much as a traditional Phillips curve predicts, given the high level of unemployment" (Ball et all 2011). In short, the reality of the world economy did not match the predictions of existing economic theory as seen in the Philipp's curve. Moreover, one of the most common models is the Phillip's curve which relates inflation to unemployment, but observing data over a 15-year period before 2001, the Phillip's curve was found to be no more accurate than a naïve guess (Atkeson et al. 2001).

Motivating variable selection, interest rates, interest rate spreads, and returns on bonds, stocks, housing, gold, and more have all been considered as leading indicators in this field. Some common indicators include interest rates, term spreads, stock returns, dividend yields, and exchange rates (Stock and Watson 2003). Additionally, in their study, Charles Goodhart and Boris Hofmann found that housing was a strong indicator of inflation, whereas yield spreads were less effective predictors (Goodhart and Hofmann 2002). Lastly, in her 1997 work, Sharon Kozicki extensively discusses the yield spread as a leading indicator of inflation.

In her work, Kozicki finds that spread helps to predict inflation up to three years in advance (Kozicki 1997). While there is extensive evidence of and scholarship on the predictive power of yield spread on real activity, there is only limited scholarship and evidence on the predictive power of the yield spread on inflation, especially on countries outside of the United States (Kozicki 1997). For inflation, Kozicki found that, although the yield spread helps predict inflation at moderate horizons of a few years, the level of yields is a more useful predictor of inflation. “Since 1955, every recession has been preceded by a period in which the yield spread has fallen considerably below its historical average” (Kozicki 1997). However, this is not a foolproof predictor and there have been periods of low spreads not followed by significant economic slowdowns.

In the introduction to his study on using machine learning algorithms to predict inflation in Russia, Ivan Baybuza notes that there is potentially an issue with overfitting resulting from the relatively small number of data points. There are only approximately 700 data points when considering US monthly data on macroeconomic factors excluding prior monthly inflation values. This means that there are relatively few data points relative to the number of explanatory variables leading to the algorithms learning the random movement present in the training data that is not present in general (Baybuza 2018). Baybuza’s work appears much more susceptible to this issue, as he considers 92 macroeconomic indicators on a time span of approximately 170 observations. A potential solution to this issue is the use of machine learning algorithms specializing in this issue. There are numerous techniques including LASSO regression, Ridge regression, Principal Components Analysis, Decision Trees, and Random Forest among many others. Despite their prior existence, there have been relatively few applications of these techniques for predicting macroeconomic trends (Baybuza 2018). In his study, Baybuza selected LASSO regression, Ridge regression, elastic net model, Random Forest, and boosting, and the results of these models were compared with more traditional forecasting methods including random walk, AR (1) or autoregressive model of order one, and AR (P) (Baybuza 2018). Furthermore, Baybuza accounts for the potential non-stationarity of CPI or consumer price index and inflation by calculating the inflation as the first difference of logarithms of CPI. He also evaluates the effectiveness of the prediction algorithms using a variation of root mean squared error or RMSE (Baybuza 2018). Two key results Baybuza found are as follows. First, the use of machine learning algorithms can improve the quality of forecasting of inflation in Russia as compared to the more standard econometric methods that use only lag of inflation as predictors. However, there is a loss of interpretability as compared to the traditional econometric models. Second, Random Forest and Boosting predict inflation better than the base model starting at the two-month time horizon (Baybuza 2018).

Data and Methods

The dataset analyzed consists of monthly data on unemployment, the yield of 6-month treasury notes, the yield on 3-month treasury notes, the yield difference defined as the difference in yield between 3-month treasury notes and 20-year treasury notes, gold price, consumer price index, mean house price, median house price, current inflation, the open, low, high, and close price of the S&P 500 and Dow Jones Industrial Average, the yield of 1-year treasury notes, and the yield of 10-year treasury notes. The dataset consists of data from January of 1963 to October of 2019. The level of the consumer price index was used to calculate the monthly inflation by calculating the percent change in consumer price index. Gold price, all S&P 500 related data, and all Dow Jones Industrial Average related data were manipulated to be in terms of month-to-month net differences rather than absolute levels. The data on mean house price was initially reported as quarterly data rather than monthly data. Therefore, one of the regressions was run on a quarterly basis. For the other regressions, the quarterly mean house price data was changed to monthly house price data through a process of linear interpolation. Lastly, future inflation, the target of our prediction model is the current monthly inflation shifted 12 months forward. The specific data sources can be found in the appendix on the data references page.

As for the techniques used, LASSO regression is like ordinary least squares regression with an additional error term that allows for the coefficients of insignificant data to be reduced to zero. The penalizing term, lambda, is selected through a process of cross validation.

$$\sum_{i=1}^{n}(y_i - \sum_{j} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

Figure 1: LASSO regression formula; OLS with penalizing term (Stephanie Glen)

To account for potential higher order terms and interaction terms all combinations of the explanatory variables up to degree three were combined into a matrix using R's poly function. This interaction matrix was then given to the LASSO regression model. Random Forest regression builds upon the Decision Tree algorithm for regression and classification. Decision trees can be thought of as breaking up the training dataset into smaller and smaller subsections while keeping track of the divisions. Then, suppose we think of the dependent variables falling on the y axis of a graph and the independent or predictor variables as falling on the x axis. If the algorithm is called upon to predict a value in a regression, it will go into the tree and find the closest x value and output the corresponding y value. A Random Forest regression takes this a step further by creating a collection of randomly generated decision trees and basing predictions upon this collection.

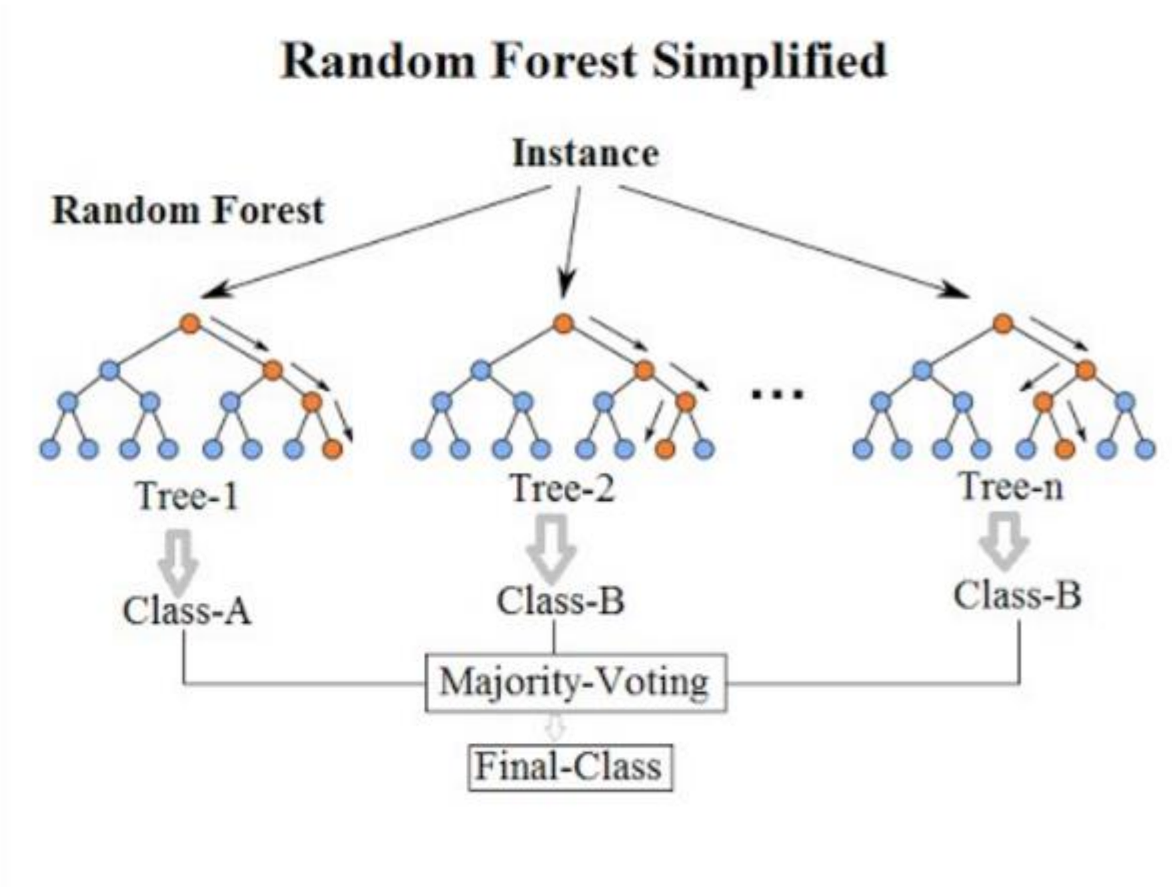

Figure 2: Diagram of Random Forest algorithm's building of random decision trees and selection process (Will Koehrsen)

A neural network can be thought of as a nonlinear combination of its inputs and if a certain threshold is met then a decision is made.

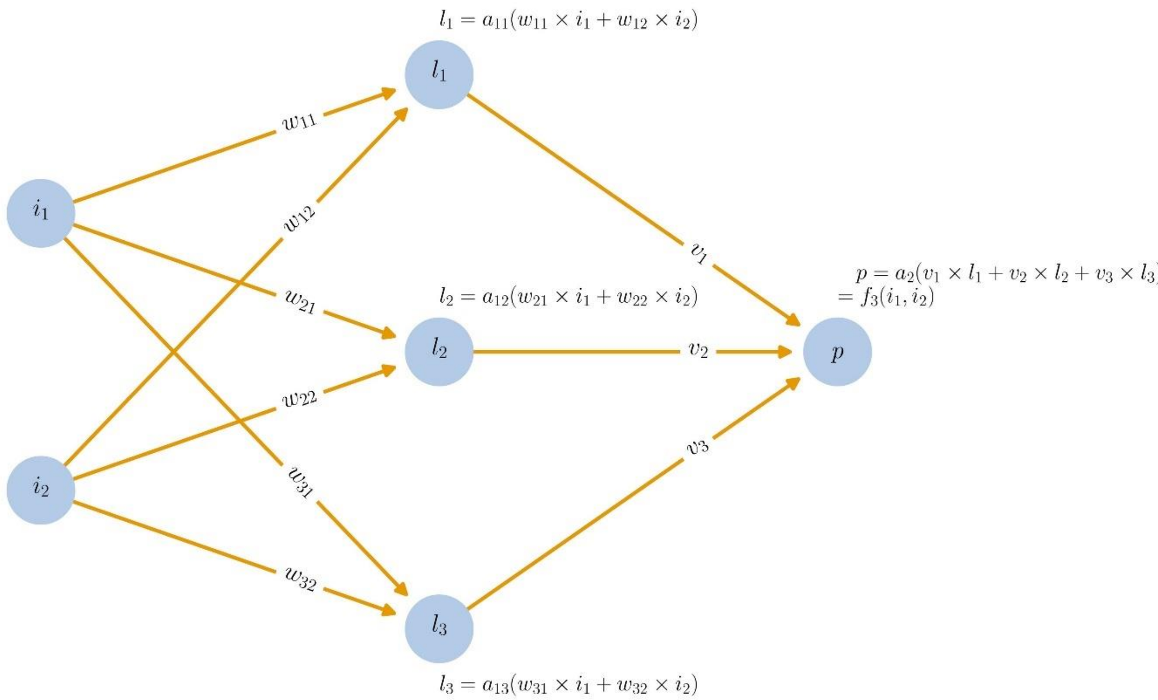

Figure 3: Diagram of Neural Network showing the nonlinear combination of multiple inputs leading to a single output (Joseph Rocca)

For a more in-depth discussion of these techniques, reference The Elements of Statistical Learning by Friedman, Tibshirani, and Hastie. Finally, as is a classical theme in machine learning, the results of the other machine learning algorithms were combined into a super learner. An OLS regression was run on the predicted outputs of the other algorithms and the results of the regression became the basis of the super learner. The following regression was run.

$$Future\widehat{Inflation}_t = \widehat{\beta_1} + \widehat{\beta_2}\widehat{LASSO}_t + \widehat{\beta_3}\widehat{RF}_t + \widehat{\beta_4}\widehat{NN}_t$$

Where:

$Future\widehat{Inflation}_t$ = The predicted month-to-month inflation 12 months in advance of time t as predicted by the super learner

$\widehat{LASSO}_t$ = The predicted month-to-month inflation 12 months in advance of time t as predicted by the LASSO regression model

RF_t = The predicted month-to-month inflation 12 months in advance of time t as predicted by the Random Forest regression model

NN_t = The predicted month-to-month inflation 12 months in advance of time t as predicted by the Neural Network regression model

The effectiveness of each of the regressions will be evaluated through four standard measures of error; MAE or mean absolute error, RMSE or root mean squared error, MSE or mean squared error, and scaled MSE defined as the mean absolute error normalized by the variance of the inflation in the out of sample prediction period. In all cases, the objective of each model is to minimize these measurements of error. For those unfamiliar with these measures, each measure of error will be briefly defined.

$$MSE = \sum (y - \hat{y})^2$$

$$RMSE = \sqrt{\sum (y - \hat{y})^2}$$

$$MAE = \sum |y - \hat{y}|$$

$$sMSE = \frac{mean((y - \hat{y})^2)}{Var(Inflation)}$$

In each case, the end date of the training data set was December of 2014, with the beginning of the dataset, which variables were contained or not contained, and frequency of each dataset being model specific. Thus, the out of sample testing period for each model ran from January of 2015 to the end of the data sampled, in the end of 2019. The metric this paper focuses on is scaled mean squared error.

Results and Findings

In all models, the out of sample testing dataset spanned the time January 2015 to October 2019.

The first model tested set the starting date for the training dataset to January of 1963 and had a quarterly frequency to account for the quarterly frequency of data on mean house prices. This model did not consider the median house price which was also available in the dataset because median house price was viewed as a robustness check for the results from the model with the mean house prices. Moreover, the model only considered the difference in closing prices of the S&P 500, rather than also considering differences in opening price, high price, and low price. Furthermore, this model did not consider the Dow Jones Industrial Average as data for the Dow Jones is not available prior to 1985. One interesting result of this model is that the variables selected for by the LASSO model differed from the variables selected for by the Random Forest model and the Neural Network model. Outside of the constant term in the regression, the LASSO model selected for the following columns: $(unemployment)^2(gold)$, $(unemployment)(yield\ of\ 3\ month\ T-note)(gold)$, $(current\ inflation)$, *(yield difference)(mean house price)(current inflation)*, $(mean\ house\ price)^2(current\ inflation)$, $(yield\ of\ 3\ month\ T-Note)(difference\ in\ closing\ price\ of\ S\&P\ 500)^2$, *(yield of one year T-Note)*, $(mean\ house\ price)^2(yield\ of\ ten\ year\ T-Note)$. Whereas the features selected for by the Random Forest and Neural Network models were *current inflation, yield difference (yield of 20-year bond – yield of 3-month bond), unemployment rate, yield of 6-month bond, yield of 3-month bond, difference in monthly gold prices, and the difference in closing prices of the S&P 500*. Overall, as measured by having the lowest scaled mean squared error, the Random Forest regression model was the most effective of the techniques tested.

The second model tested also had a training data set that spanned the years 1963 to 2014, but this time the frequency was monthly. To account for the lack of monthly data in the mean house price data, the missing months were filled in with linear interpolation. In this process the quarter-to-quarter differences were assumed to follow a linear pattern, so their differences were divided by three and distributed evenly among the missing months in the data. Like the first model, this model did not consider median house prices nor did it consider the Dow Jones. Moreover, this model only considered the closing prices of the S&P 500. Outside of the constant term in the regression, the LASSO model selected for the following columns: $(unemployment)(yield\ of\ 3\ month\ T-Note)(gold)$, *(yield of 6 month T-Note)(yield difference)(mean house price), current inflation,* $(mean\ house\ price)^2$*(current inflation),* $(current\ inflation)^2$, $(yield\ of\ 1\ year\ T-Note)$, *(current inflation) (yield of 1 year T-Note), and* $(mean\ house\ price)^2(yield\ on\ 10\ year\ T-Note)$. This differs once again from the variables selected for by the Neural Network and Random Forest algorithms, which selected *current inflation, yield difference, unemployment rate, the yield on 6-month treasury notes, and the changes in the closing prices of the S&P 500*. In this model, the LASSO model was less accurate than the Random Forest and Neural Network models as seen by having a higher scaled mean squared error. Overall, as measured by having the lowest scaled mean squared error, the super learner technique was the most effective of the techniques tested.

The third model tested had a training data set that spanned the years 1985 to 2014 and the frequency was monthly. Like the first two models, this model did not consider median house prices. Moreover, this model only considered the closing prices of the S&P 500 and Dow Jones Industrial Average. Outside of the constant term in the regression, the LASSO model selected for the following columns: $(yield\ on\ 3\ month\ T-Notes)^2(mean\ house\ price)$, *(gold)(mean house price)(current inflation),* $(mean\ house\ price)^2(current\ inflation)$, $(gold)(current\ inflation)^2$ *and* $(mean\ house\ price)(current\ inflation)^2$. This once again differed from the variables selected for by the Neural Network and Random Forest models. The Neural Network and Random Forest models selected *current inflation, yield difference, unemployment rate, gold price, changes in the S&P 500 Closing Price,* and *changes in the Dow Jones closing price*. In this model the LASSO model had a higher scaled mean squared error than Random Forest, but a significantly lower scaled mean squared error than the Neural Network. Overall, as measured by having the lowest scaled mean squared error, the super learner technique was the most effective of the techniques tested.

The fourth model tested also had a training data set that spanned the years 1985 to 2014 and a monthly frequency. Like the first model, this model did not consider median house prices nor did it consider the Dow Jones. Moreover, this model only considered the closing prices of the S&P 500. The LASSO model selected none of the columns and simply used the intercept as the prediction, which is a rather unusual result. On the other hand, the Random Forest and Neural Network selected *current inflation, yield difference, unemployment rate, changes in the gold price,* and *changes in the closing price of the S&P 500*. Despite this, the LASSO model had a lower scaled mean squared error than the Neural Network and the scaled mean squared error across all four models tested were rather close falling in the range 0.95 to 1.12. Overall, as measured by having the lowest scaled mean squared error, the super learner technique was the most effective of the techniques tested.

The fifth model tested had a training data set ranging from 1990 to 2014 and had monthly frequency. Like the first model, this model did not consider median house price nor did it consider the Dow Jones Industrial Average. Moreover, this model only considered the closing prices of the S&P 500. Once again, the trend of LASSO selecting different variables from the variables selected by Neural Network and Random Forest continued. A result to note is that the Neural Network was by far the worst prediction method with a scaled mean squared error of nearly double that of the other prediction algorithms. Overall, as measured by having the lowest scaled mean squared error, the super learner technique was the most effective of the techniques tested.

The sixth model tested had a training data set ranging from 1990 to 2014 and had monthly frequency. Like the first model, this model did not consider median house prices. Yet, it did consider the Dow Jones Industrial Average. Moreover, this model only considered the closing prices of the S&P 500. Once again, the trend of LASSO selecting different variables from the variables selected by Neural Network and Random Forest continued. A result to note is that the Neural Network was by far the worst prediction method with a scaled mean squared error of nearly double that of the other prediction algorithms. Overall, as measured by having the lowest scaled mean squared error, the super learner technique was the most effective of the techniques tested.

Conclusions

Comparing results of the same regression technique across testing periods and explanatory variables we arrive at the following conclusions. First, comparing the scaled mean squared error of Random Forest on the monthly data starting in 1963 and the scaled mean squared error of Random Forest on the monthly data starting in 1985 without the Dow Jones, there is a decrease, indicating a more accurate prediction. Further limiting the training data set to begin in 1990, we see a further decrease in scaled mean squared error both relative to the 1963 value and relative to the 1985 value. Moreover, comparing the scaled mean squared error of Random Forest on the monthly data starting in 1985 with the Dow Jones as compared to the scaled mean squared error of Random Forest on the monthly data starting in 1985 without the Dow Jones, we see a decrease from the approximately 0.99 without the Dow Jones to approximately 0.94 with the Dow Jones. Lastly, disregarding the super learner method, we see that in all six of the cases tested, the Random Forest algorithm provided the most accurate prediction.

First, the super learner technique was the best overall predictive method for inflation. Second, limiting the training data to more recent, and hence relevant, data greatly improves the accuracy of prediction. Third, the inclusion of the Dow Jones improves the accuracy of prediction. Lastly, only considering the stand-alone methods of LASSO, Random Forest, and Neural Network, Random Forest is the best prediction algorithm.

Appendix

Figure 4: Graph of predicted versus actual quarterly inflation using the LASSO regression algorithm on data excluding the DJIA starting in 1963 where red points indicate the predicted values and red the actual inflation values.

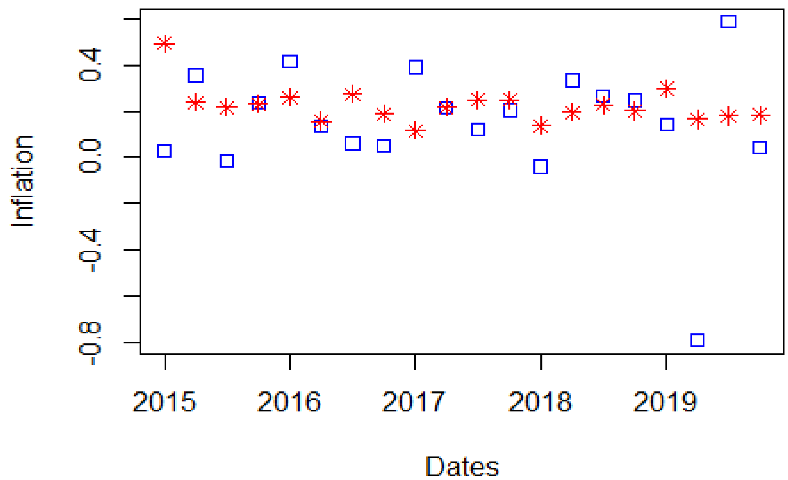

Figure 5: Graph of predicted versus actual quarterly inflation using the Random Forest regression algorithm on data excluding the DJIA starting in 1963 where red points indicate the predicted values and red the actual inflation values.

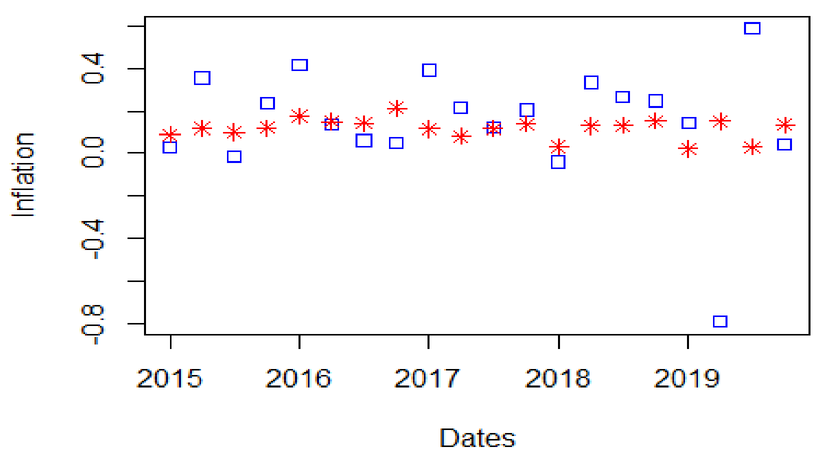

Figure 6: Graph of predicted versus actual quarterly inflation using the Neural Network regression algorithm on data excluding the DJIA starting in 1963 where red points indicate the predicted values and red the actual inflation values.

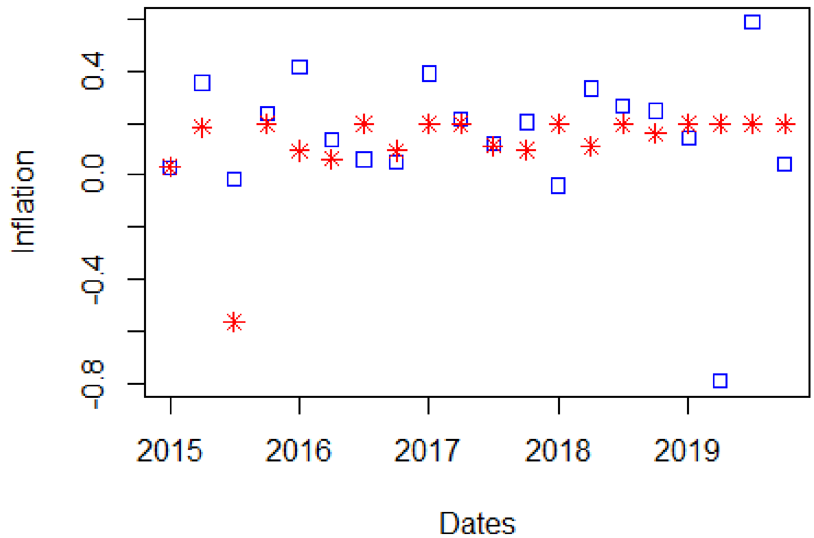

Figure 7: Graph of predicted versus actual quarterly inflation using the Super Learner regression algorithm on data excluding the DJIA starting in 1963 where red points indicate the predicted values and red the actual inflation values.

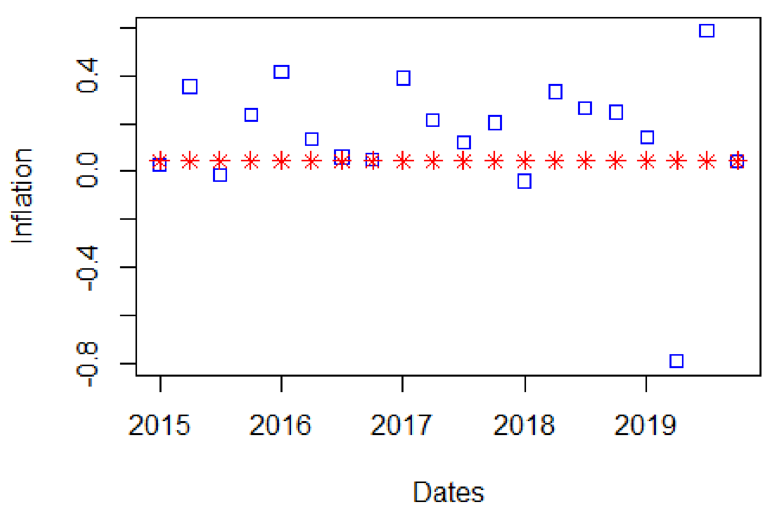

Figure 8: Graph of predicted versus actual monthly inflation using the LASSO regression algorithm on data excluding the DJIA starting in 1963 where red points indicate the predicted values and red the actual inflation values.

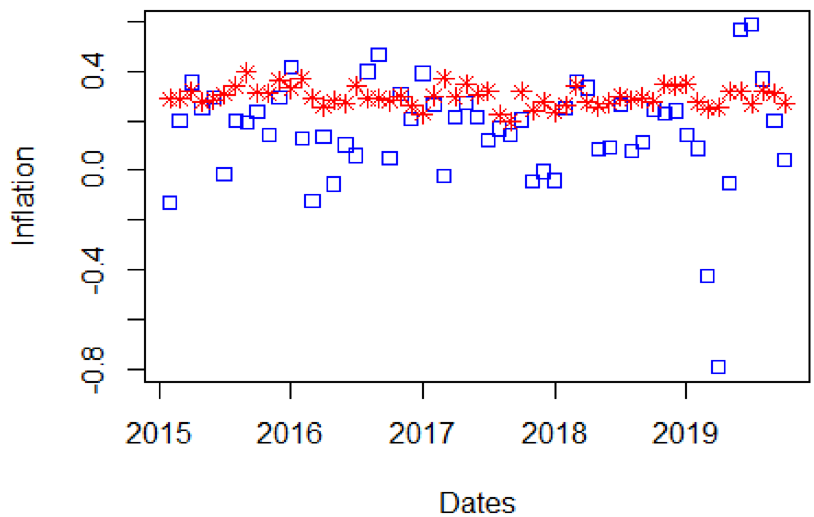

Figure 9: Graph of predicted versus actual monthly inflation using the Random Forest regression algorithm on data excluding the DJIA starting in 1963 where red points indicate the predicted values and red the actual inflation values.

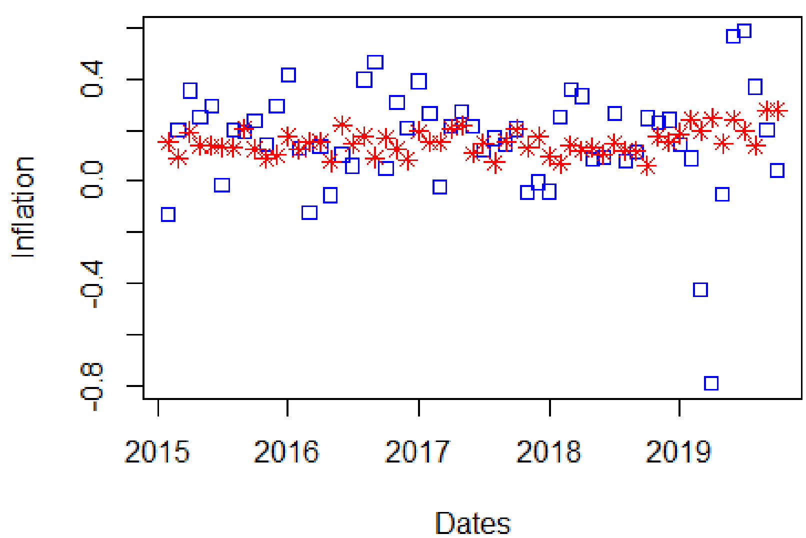

Figure 10: Graph of predicted versus actual monthly inflation using the Neural Network regression algorithm on data excluding the DJIA starting in 1963 where red points indicate the predicted values and red the actual inflation values.

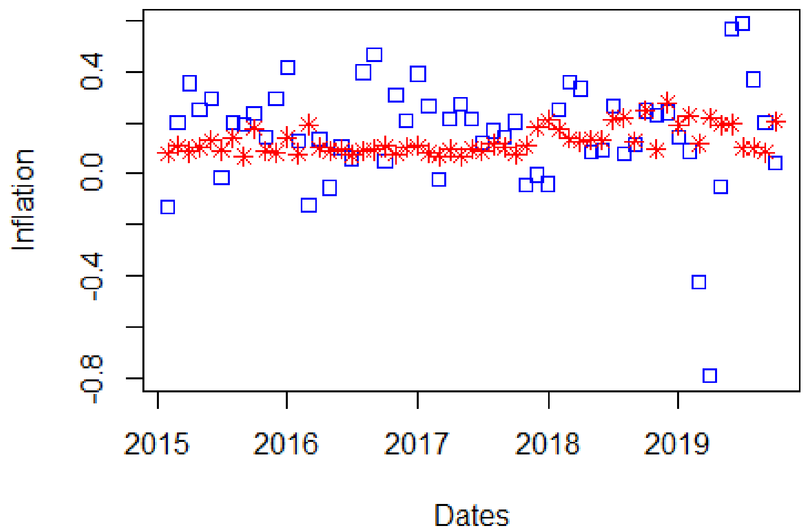

Figure 11: Graph of predicted versus actual monthly inflation using the Super Learner regression algorithm on data excluding the DJIA starting in 1963 where red points indicate the predicted values and red the actual inflation values.

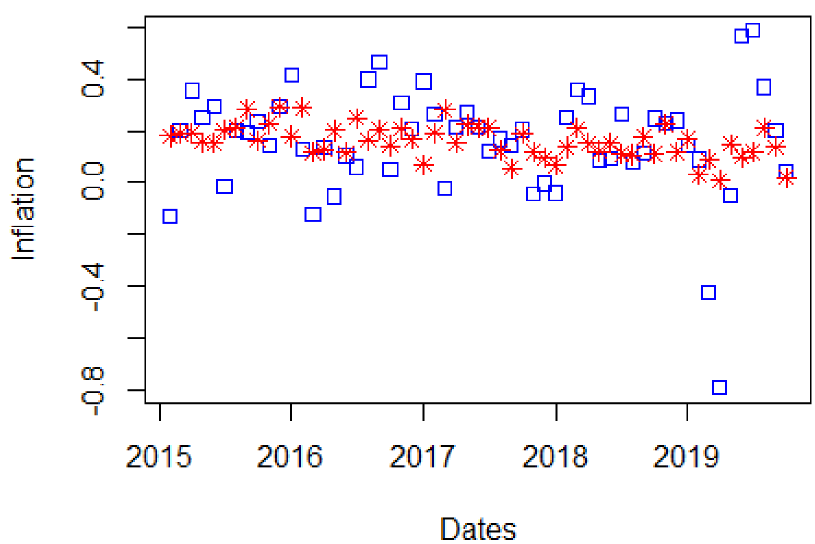

Figure 12: Graph of predicted versus actual monthly inflation using the LASSO regression algorithm on data including the DJIA starting in 1985 where red points indicate the predicted values and red the actual inflation values.

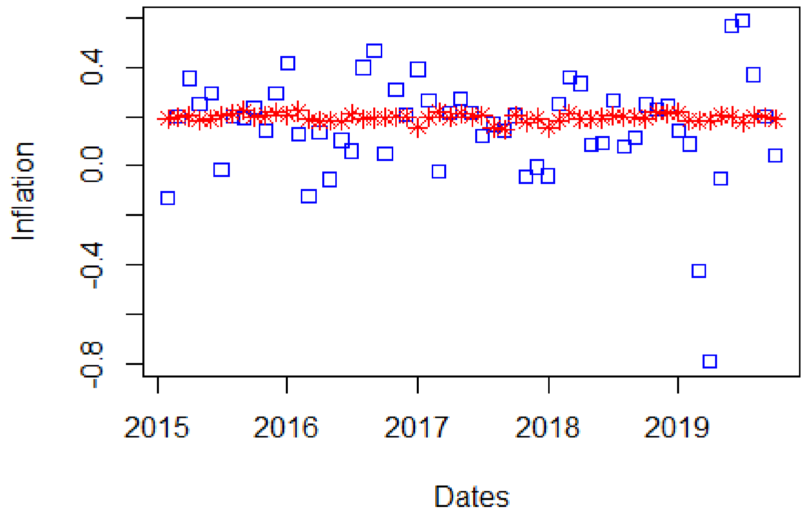

Figure 13: Graph of predicted versus actual monthly inflation using the Random Forest regression algorithm on data including the DJIA starting in 1985 where red points indicate the predicted values and red the actual inflation values.

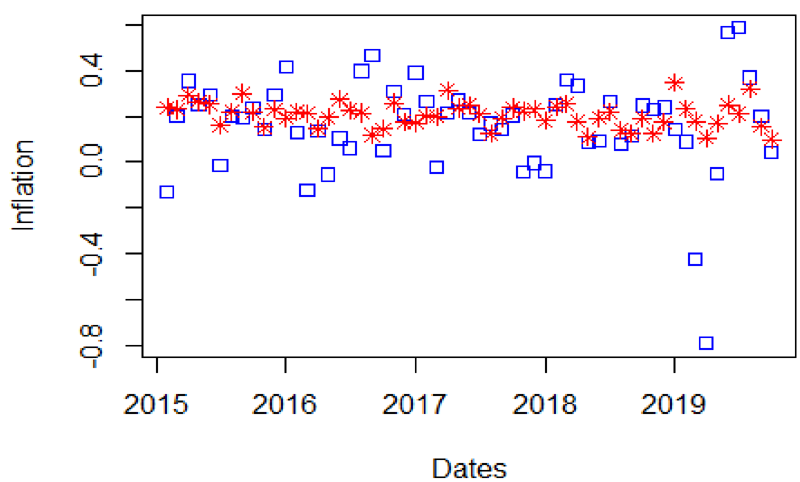

Figure 14: Graph of predicted versus actual monthly inflation using the Neural Network regression algorithm on data including the DJIA starting in 1985 where red points indicate the predicted values and red the actual inflation values.

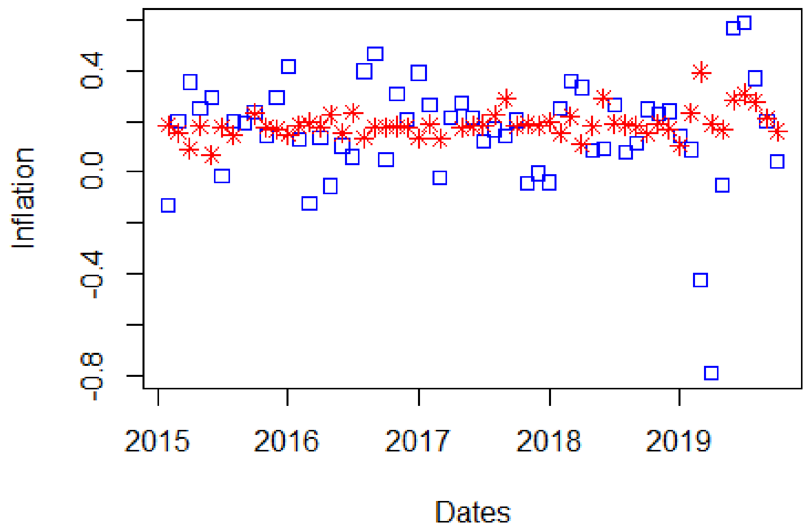

Figure 15: Graph of predicted versus actual monthly inflation using the Super Learner regression algorithm on data including the DJIA starting in 1985 where red points indicate the predicted values and red the actual inflation values.

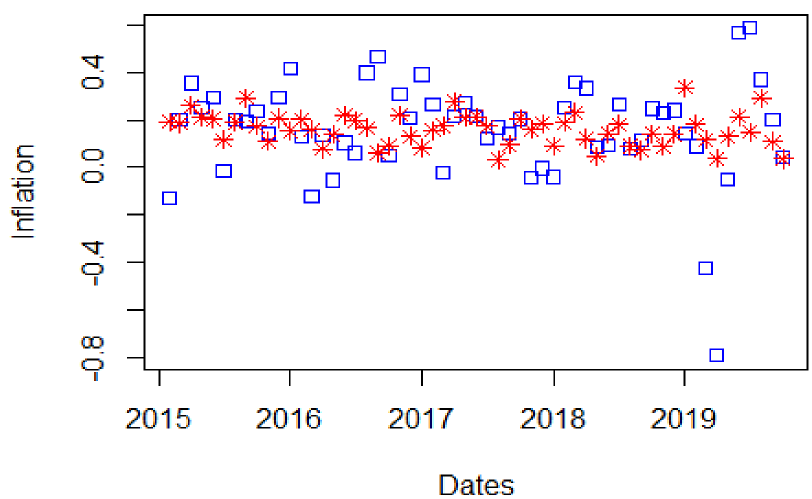

Figure 16: Graph of predicted versus actual monthly inflation using the LASSO regression algorithm on data excluding the DJIA starting in 1985 where red points indicate the predicted values and red the actual inflation values.

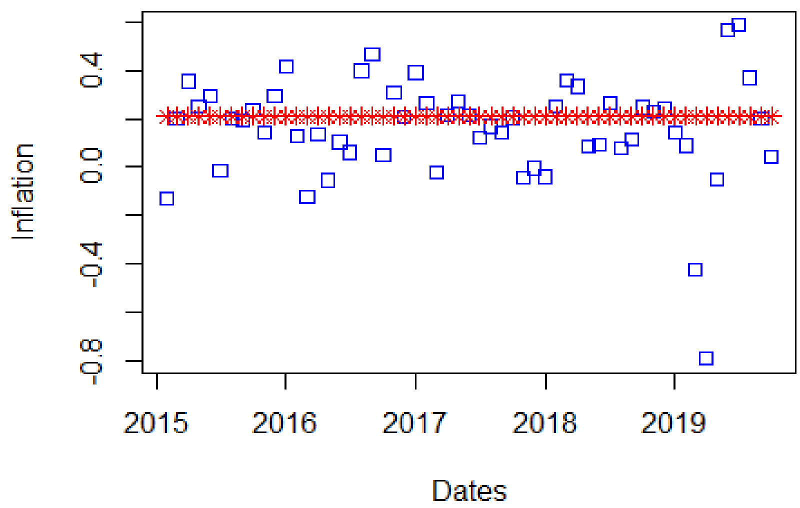

Figure 17: Graph of predicted versus actual monthly inflation using the Random Forest regression algorithm on data excluding the DJIA starting in 1985 where red points indicate the predicted values and red the actual inflation values.

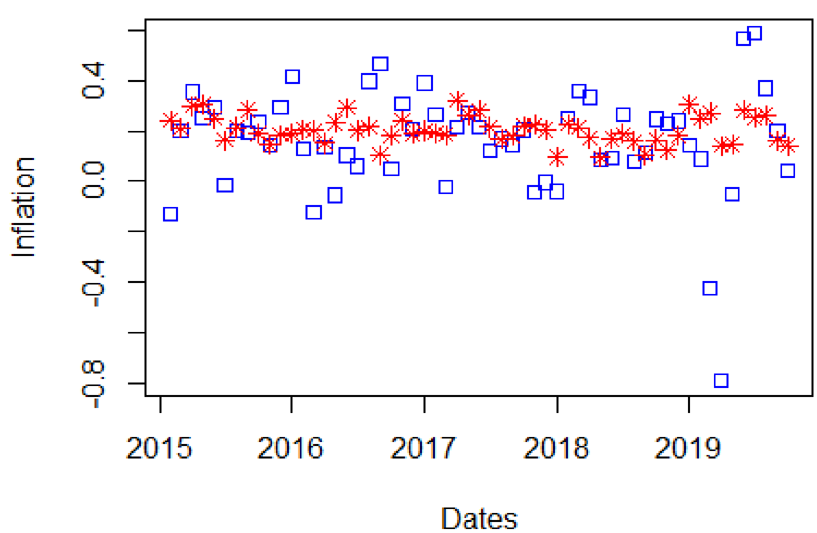

Figure 18: Graph of predicted versus actual monthly inflation using the Neural Network regression algorithm on data excluding the DJIA starting in 1985 where red points indicate the predicted values and red the actual inflation values.

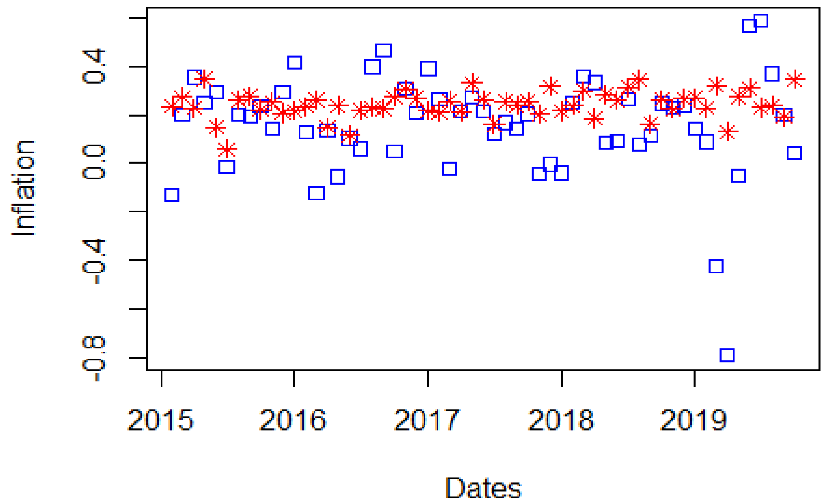

Figure 19: Graph of predicted versus actual monthly inflation using the Super Learner regression algorithm on data excluding the DJIA starting in 1985 where red points indicate the predicted values and red the actual inflation values.

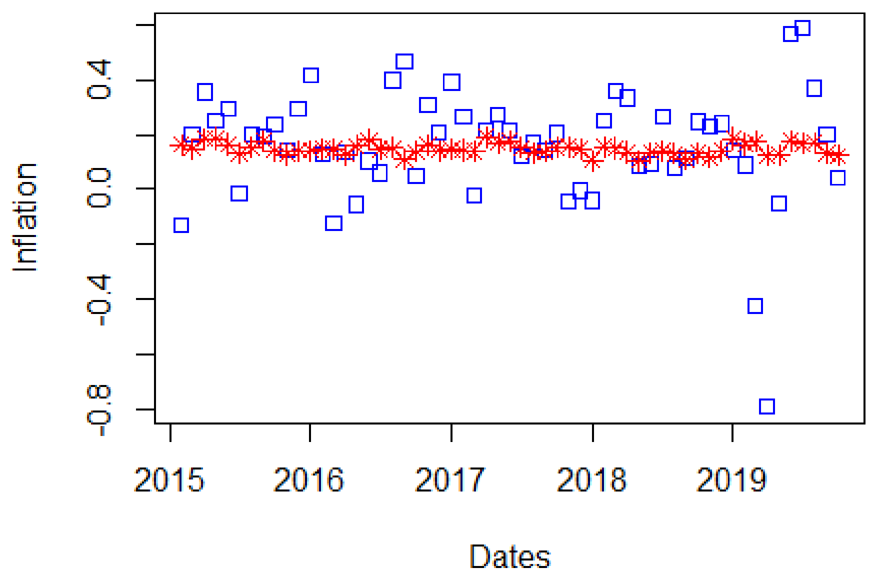

Figure 20: Graph of predicted versus actual monthly inflation using the LASSO regression algorithm on data excluding the DJIA starting in 1990 where red points indicate the predicted values and red the actual inflation values.

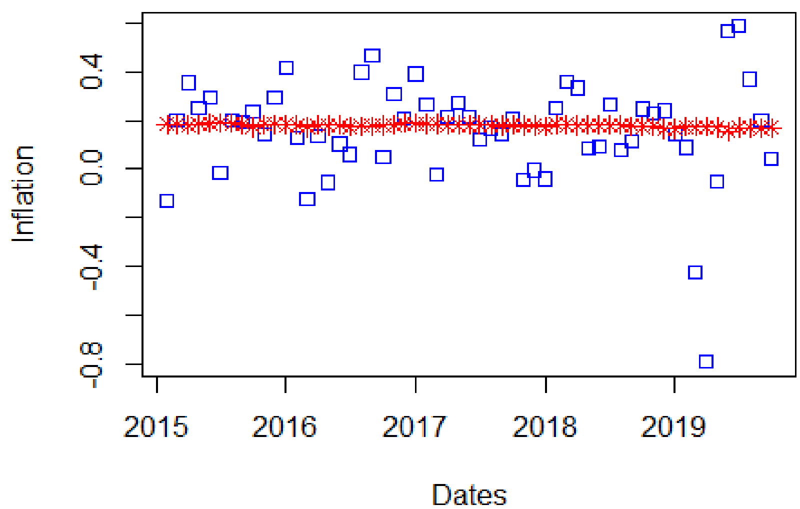

Figure 21: Graph of predicted versus actual monthly inflation using the Random Forest regression algorithm on data excluding the DJIA starting in 1990 where red points indicate the predicted values and red the actual inflation values.

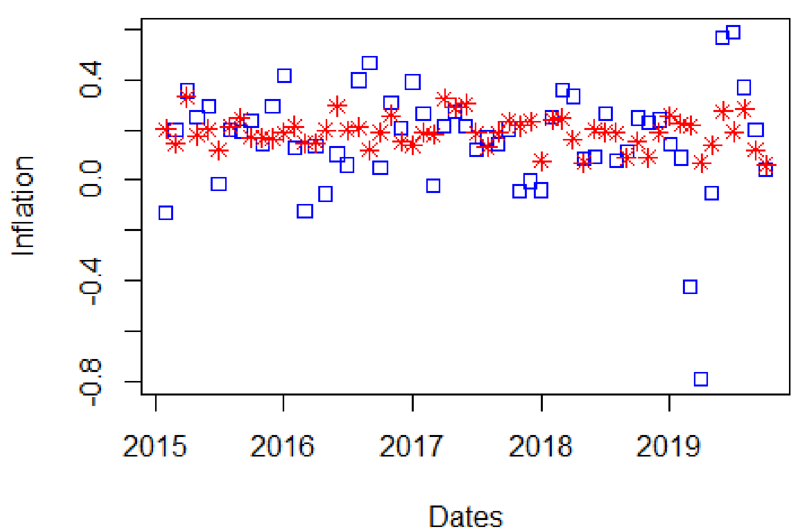

Figure 22: Graph of predicted versus actual monthly inflation using the Neural Network regression algorithm on data excluding the DJIA starting in 1990 where red points indicate the predicted values and red the actual inflation values.

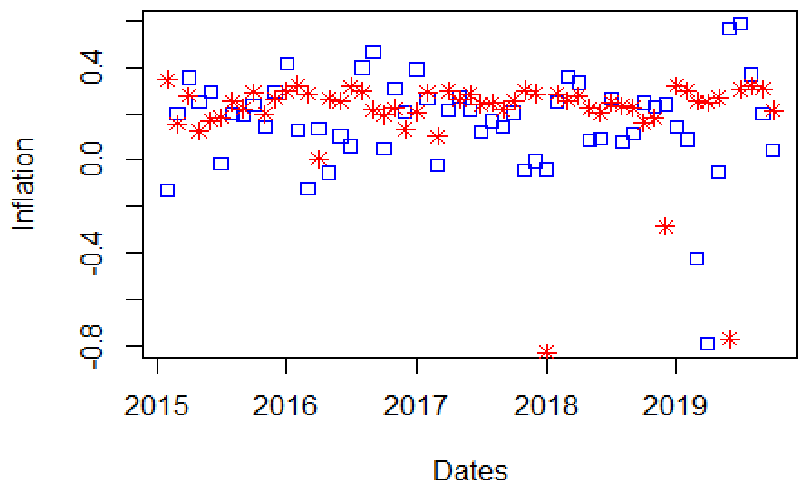

Figure 23: Graph of predicted versus actual monthly inflation using the Super Learner regression algorithm on data excluding the DJIA starting in 1990 where red points indicate the predicted values and red the actual inflation values.

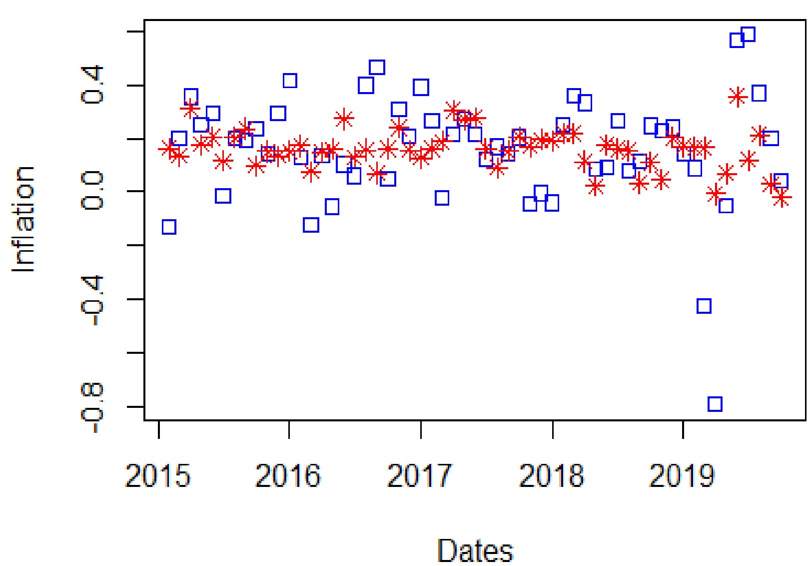

Figure 24: Graph of predicted versus actual monthly inflation using the LASSO regression algorithm on data including the DJIA starting in 1990 where red points indicate the predicted values and red the actual inflation values.

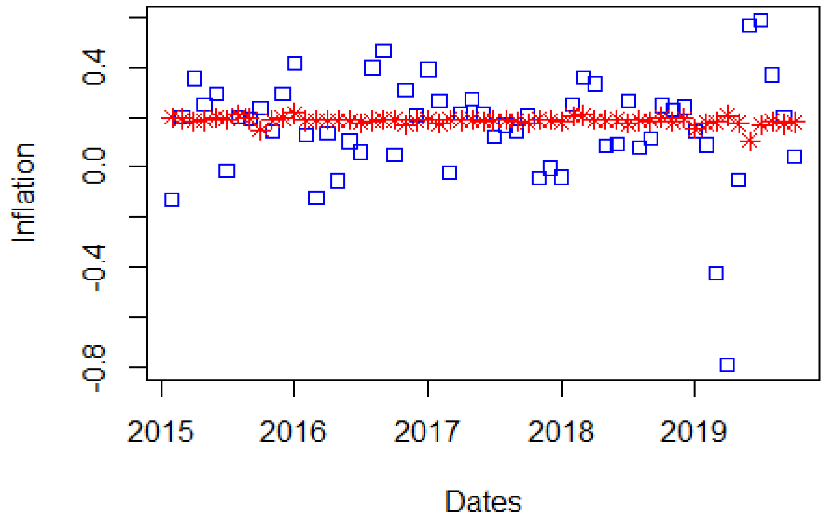

Figure 25: Graph of predicted versus actual monthly inflation using the Random Forest regression algorithm on data including the DJIA starting in 1990 where red points indicate the predicted values and red the actual inflation values.

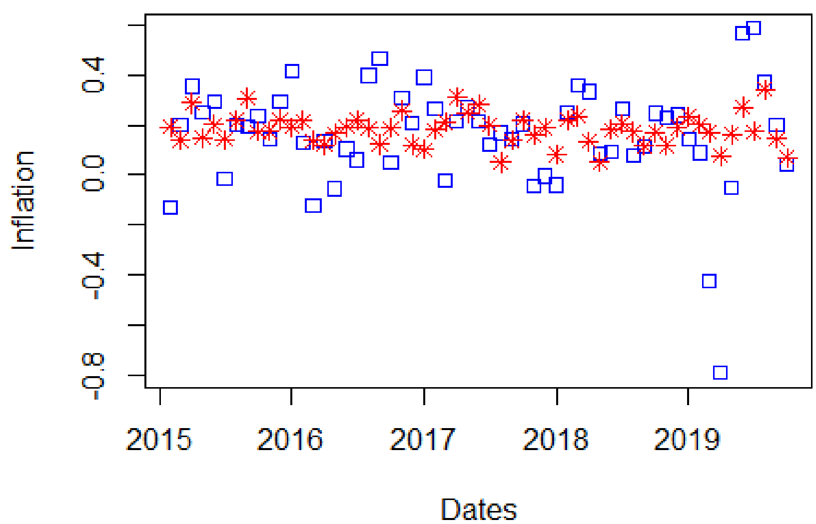

Figure 26: Graph of predicted versus actual monthly inflation using the Neural Network regression algorithm on data including the DJIA starting in 1990 where red points indicate the predicted values and red the actual inflation values.

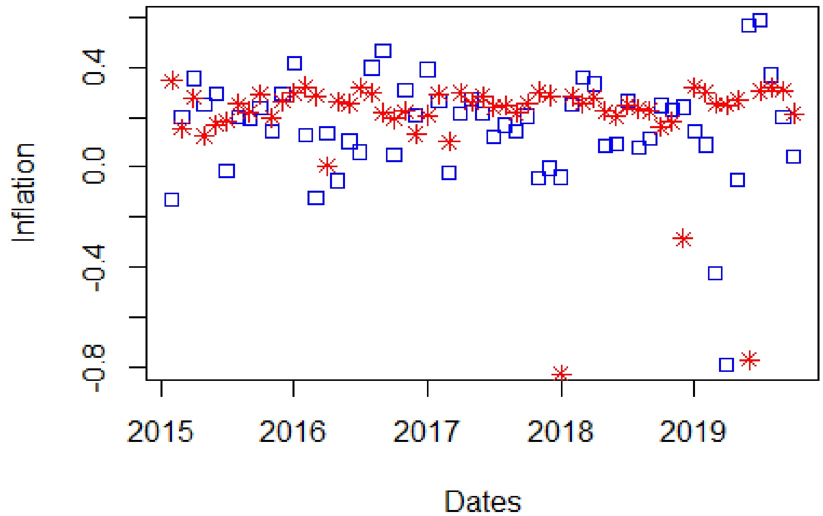

Figure 27: Graph of predicted versus actual monthly inflation using the Super Learner regression algorithm on data including the DJIA starting in 1990 where red points indicate the predicted values and red the actual inflation values.

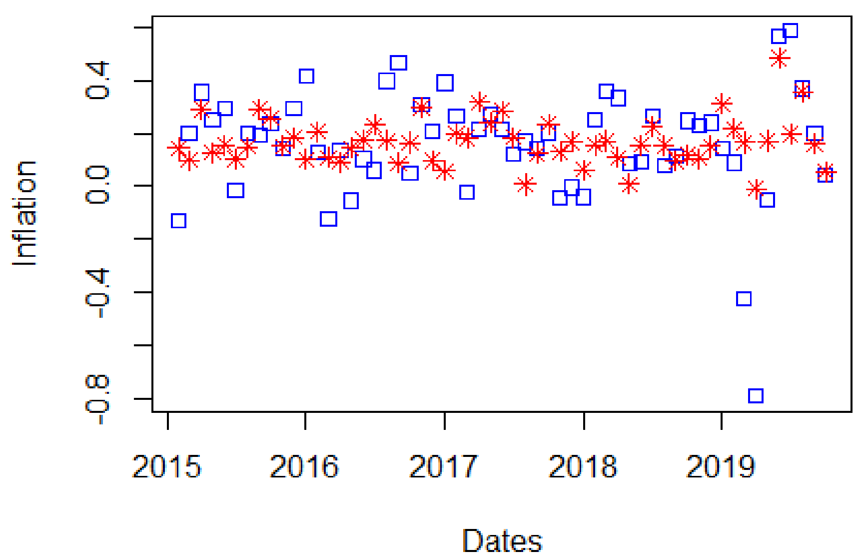

Table 1: Mean Squared Error Summary

Quarterly LASSO No DJIA	0.083
Quarterly Random Forest No DJIA	0.079
Quarterly Neural Network No DJIA	0.090
Quarterly Super Learner No DJIA	0.083
Monthly LASSO No DJIA	0.064
Monthly Random Forest No DJIA	0.052
Monthly Neural Network No DJIA	0.055
Monthly Super Learner No DJIA	0.043
Monthly LASSO With DJIA	0.047
Monthly Random Forest With DJIA	0.045
Monthly Neural Network With DJIA	0.067
Monthly Super Learner With DJIA	0.042
Monthly LASSO without DJIA after 1985	0.049
Monthly Random Forest without DJIA After 1985	0.047
Monthly Neural Network without DJIA After 1985	0.054
Monthly Super Learner without DJIA After 1985	0.046
Monthly LASSO Without DJIA After 1990	0.047

Monthly Random Forest Without DJIA After 1990	0.043
Monthly Neural Network Without DJIA After 1990	0.101
Monthly Super Learner Without DJIA After 1990	0.041
Monthly LASSO With DJIA After 1990	0.047
Monthly Random Forest With DJIA After 1990	0.043
Monthly Neural Network With DJIA After 1990	0.126
Monthly Super Learner With DJIA After 1990	0.040

Table 2: Root Mean Squared Error Summary

Quarterly LASSO No DJIA	0.288
Quarterly Random Forest No DJIA	0.281
Quarterly Neural Network No DJIA	0.300
Quarterly Super Learner No DJIA	0.288
Monthly LASSO No DJIA	0.252
Monthly Random Forest No DJIA	0.227
Monthly Neural Network No DJIA	0.234
Monthly Super Learner No DJIA	0.207
Monthly LASSO With DJIA	0.218

Monthly Random Forest With DJIA	0.212
Monthly Neural Network With DJIA	0.259
Monthly Super Learner With DJIA	0.206
Monthly LASSO without DJIA after 1985	0.222
Monthly Random Forest without DJIA After 1985	0.217
Monthly Neural Network without DJIA After 1985	0.232
Monthly Super Learner without DJIA After 1985	0.214
Monthly LASSO Without DJIA After 1990	0.217
Monthly Random Forest Without DJIA After 1990	0.208
Monthly Neural Network Without DJIA After 1990	0.318
Monthly Super Learner Without DJIA After 1990	0.203
Monthly LASSO With DJIA After 1990	0.217
Monthly Random Forest With DJIA After 1990	0.208
Monthly Neural Network With DJIA After 1990	0.355
Monthly Super Learner With DJIA After 1990	0.199

Table 3: Mean Absolute Error Summary

Quarterly LASSO No DJIA	0.193
Quarterly Random Forest No DJIA	0.187
Quarterly Neural Network No DJIA	0.194
Quarterly Super Learner No DJIA	0.205
Monthly LASSO No DJIA	0.183
Monthly Random Forest No DJIA	0.159
Monthly Neural Network No DJIA	0.169
Monthly Super Learner No DJIA	0.145
Monthly LASSO With DJIA	0.146
Monthly Random Forest With DJIA	0.143
Monthly Neural Network With DJIA	0.181
Monthly Super Learner With DJIA	0.145
Monthly LASSO without DJIA after 1985	0.149
Monthly Random Forest without DJIA After 1985	0.145
Monthly Neural Network without DJIA After 1985	0.160
Monthly Super Learner without DJIA After 1985	0.148
Monthly LASSO Without DJIA After 1990	0.149

Monthly Random Forest Without DJIA After 1990	0.145
Monthly Neural Network Without DJIA After 1990	0.202
Monthly Super Learner Without DJIA After 1990	0.144
Monthly LASSO With DJIA After 1990	0.149
Monthly Random Forest With DJIA After 1990	0.145
Monthly Neural Network With DJIA After 1990	0.236
Monthly Super Learner With DJIA After 1990	0.143

Table 4: Scaled Mean Squared Error Summary

Quarterly LASSO No DJIA	1.095
Quarterly Random Forest No DJIA	1.047
Quarterly Neural Network No DJIA	1.188
Quarterly Super Learner No DJIA	1.097
Monthly LASSO No DJIA	1.333
Monthly Random Forest No DJIA	1.083
Monthly Neural Network No DJIA	1.151
Monthly Super Learner No DJIA	0.896
Monthly LASSO With DJIA	0.994

Monthly Random Forest With DJIA	0.941
Monthly Neural Network With DJIA	1.412
Monthly Super Learner With DJIA	0.888
Monthly LASSO without DJIA after 1985	1.035
Monthly Random Forest without DJIA After 1985	0.991
Monthly Neural Network without DJIA After 1985	1.129
Monthly Super Learner without DJIA After 1985	0.957
Monthly LASSO Without DJIA After 1990	0.990
Monthly Random Forest Without DJIA After 1990	0.901
Monthly Neural Network Without DJIA After 1990	2.126
Monthly Super Learner Without DJIA After 1990	0.864
Monthly LASSO With DJIA After 1990	0.990
Monthly Random Forest With DJIA After 1990	0.901
Monthly Neural Network With DJIA After 1990	2.638
Monthly Super Learner With DJIA After 1990	0.833

References

Atkeson, Andrew and Ohanian, Lee, (2001), Are Phillips curves useful for forecasting inflation?, Quarterly Review, 25, issue Win, p. 2-11.

Baybuza, I. (2018). Infation Forecasting Using Machine Learning Methods. Russian Journal of Money and Finance, 77(4), pp. 42–59.

Goodhart, Charles & Boris Hofmann, 2002. "Asset Prices and the Conduct of Monetary Policy," Royal Economic Society Annual Conference 2002 88, Royal Economic Society.

Hoover, Kevin D., Related Content, Steven Horwitz, Joel Slemrod, Lawrence H. Summers, and Kenneth A. Small. “Phillips Curve.” Econlib, 2019. https://www.econlib.org/library/Enc/PhillipsCurve.html.

Inflation Dynamics and the Great Recession Laurence M. Ball and Sandeep Mazumder NBER Working Paper No. 17044 May 2011 JEL No. E31

Koehrsen, Will. “Random Forest Simple Explanation.” Medium. Medium, August 18, 2020. https://williamkoehrsen.medium.com/random-forest-simple-explanation-377895a60d2d.

Rocca, Joseph. “A Gentle Journey from Linear Regression to Neural Networks.” Medium. Towards Data Science, March 21, 2021. https://towardsdatascience.com/a-gentle-journey-from-linear-regression-to-neural-networks-68881590760e.

Sharon Kozicki, 1997. "Predicting real growth and inflation with the yield spread," Economic Review, Federal Reserve Bank of Kansas City, vol. 82(Q IV), pages 39-57.

Stephanie Glen. "Lasso Regression: Simple Definition" From StatisticsHowTo.com: Elementary Statistics for the rest of us! https://www.statisticshowto.com/lasso-regression/

Stock, James, H., and Mark WWatson. 2003. "Forecasting Output and Inflation: The Role of Asset Prices." Journal of Economic Literature, 41 (3): 788-829.

Data References

Board of Governors of the Federal Reserve System (US), 1-Year Treasury Constant Maturity Rate [DGS1], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/DGS1, May 31, 2021.

Board of Governors of the Federal Reserve System (US), 10-Year Treasury Constant Maturity Rate [GS10], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/GS10, May 31, 2021.

Board of Governors of the Federal Reserve System (US), 20-Year Treasury Constant Maturity Rate [GS20], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/GS20, June 1, 2021.

Board of Governors of the Federal Reserve System (US), 3-Month Treasury Bill: Secondary Market Rate [TB3MS], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/TB3MS, May 31, 2021.

Board of Governors of the Federal Reserve System (US), 6-Month Treasury Bill: Secondary Market Rate [TB6MS], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/TB6MS, June 1, 2021.

Bundesbank Gold Prices. "Monthly Gold Prices in the US from December 1950 to December 2019." https://datahub.io/core/gold-prices/r/1.html: https://datahub.io/core/gold-prices/r/1.html, 2021.

"Dow Jones Industrial Average Historical Rates (DJI)." Investing.com. Fusion Media Limited, 2021. https://www.investing.com/indices/us-30-historical-data.

"S&P 500 Historical Rates (SPX)." Investing.com. Fusion Media Limited, 2021. https://www.investing.com/indices/us-spx-500-historical-data.

U.S. Bureau of Labor Statistics, Unemployment Rate [UNRATENSA], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/UNRATENSA, May 31, 2021.

U.S. Census Bureau and U.S. Department of Housing and Urban Development, Average Sales Price of Houses Sold for the United States [ASPUS], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/ASPUS, May 31, 2021.

U.S. Census Bureau and U.S. Department of Housing and Urban Development, Median Sales Price for New Houses Sold in the United States [MSPNHSUS], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/MSPNHSUS, May 31, 2021.