Embracing Nonlinearities:

Further Exploring Machine Learning Applications to Inflation Forecasting

By: Sarah Larino

Honors Thesis
Economics Department
The University of North Carolina at Chapel Hill

Thesis Advisor: Dr. Andrii Babii
Faculty Advisor: Prof. Klara Peter

April 2023

Abstract

In this thesis, I employ a number of machine learning (ML) methods on the inflation forecasting problem space. I utilize macroeconomic indicators alongside textual data and apply ML methods to an updated time horizon. Ultimately, I find that ML methods are a viable alternative to traditional benchmarks under certain time horizon conditions, particularly with the inclusion of textual data. However, in contrast with the previous literature, I demonstrate that some ML models are particularly sensitive to the treatment of outliers. When a full time horizon is employed and outliers are included, certain ML models that performed well in previous analyses are not able to outperform other forecasting methods.

Acknowledgments

This thesis would not have been possible without the inspiration and guidance of my advisor, Dr. Andrii Babii. I am incredibly grateful for his willingness to take on this project with me and for the countless ways in which he has supported me throughout my research. I also want to thank Dr. Klara Peter for her invaluable feedback. Her commitment to excellence has consistently challenged me to construct a better and more universally approachable thesis. Additionally, I want to thank Dr. Mike Aguilar for helping me discover my passion for macroeconomics and sparking my interest in inflation forecasting.

I also want to thank my peers in the thesis program for their advice, feedback, and camaraderie. To my sister Hannah, thank you for your encouragement throughout this process. Finally, Misha, thank you for your patience, your help, and for always being willing to listen to me talk about econometrics.

Table of Contents

List of Tables

List of Figures

1 Introduction

Inflation forecasting is an incredibly difficult task that is simultaneously crucial for a variety of stakeholders. Firms, households, financial market participants, governments, and central banks all rely on inflation forecasting to inform their understanding of future economic conditions, and ultimately, their choices. Forecasts of inflation are of particular import to central banks, and, because of the elevated inflation rates currently being realized in a majority of advanced economies, an accurate understanding of potential future inflation is even more relevant at present.

Historically, inflation forecasts have utilized a variety of different model types, all with varying levels of complexity. Univariate time series models, such as those proposed by Box and Jenkins (1970), Atkeson and Ohanian (2001), and Stock and Watson (2007), are some of the most simplistic, utilizing past inflation levels to understand future inflation. Many theoretical models of macroeconomics, however, would indicate that other variables should serve as drivers of inflation. Recognizing this, econometricians like Stock and Watson (1999) proposed multivariate models based on the Phillips Curve, which relates inflation with unemployment. Other model-based drivers of inflation were explored as well. For example, Anderson et al. (2002) formulated a model based on GDP and M1 money balances. These models often proved to be successful in the short term, particularly in the specific economic conditions of the eras in which they were popularized, but struggled to consistently outperform the simpler univariate models in the long term. Today, dynamic stochastic general equilibrium (DSGE) models, which are based on modern micro-founded macroeconomic theory, are often employed by central banks to predict inflation, but these models have suffered from frequent forecasting errors during the last five years. The univariate methods, while consistent in their performance, are not able to deal with unusual economic conditions and have not been able to be improved substantially in the past decade (Stock and Watson, 2010).

Recently, there have been some investigations into using machine learning (ML) techniques for economic forecasting and nowcasting with very promising results. Machine learning methods have been demonstrated to be successful in both financial forecasting (Gu et al., 2018) and in the forecasting/nowcasting of other macroeconomic indicators, like GDP (Babii et al., 2021). Medeiros et al. (2019) employed machine learning methods to the inflation forecasting problem space, becoming the first to demonstrate that ML could outperform univariate and multivariate benchmarks. However, much remains to be done to demonstrate that machine learning is an effective alternative to more traditional forecasting methods. In this thesis, I attempt to fill in some of the gaps present in the existing literature, and in the remainder of this Introduction, I devote a paragraph to each of these key contributions.

First, I employ textual data, which has not been previously utilized in inflation forecasting. Some past work in economic and financial forecasting has used machine learning methods alongside the introduction of nontraditional data sources and predictors, such as textual data and search data (Babii et al., 2021; Caperna et al., 2020; Kalamara et al., 2020), with much success. This type of data may account for drivers that are missing or underrepresented in typical macroeconomic indicators. In the case of my research, the addition of textual data may serve as a proxy for inflation expectations.

Second, I explore how machine learning methods perform in times of high economic volatility, which is especially salient given that it is under unusual economic conditions that more traditional forecasting methods often flounder. To this end, I apply ML methods to an updated time horizon that encapsulates both the unusual low inflationary period that occurred between 2008 and 2021, where Core Personal Consumption Expenditures (PCE) was consistently below the two percent target of the Federal Reserve, and the period of unusually high inflation that began in 2021 and has persisted to the present. Previous research was limited to a forecast horizon that ended in January of 2016, and only performed a real-time test on data ranging from 2001 to 2015 (Medeiros et al., 2019).

Third, I recognize that much of the value of ML methods is found in their potential ability to perform during unusual inflationary periods. Thus, I do not remove extreme data points during my estimation process. Outlying data points were removed by Medeiros et al., particularly those realized during the Global Financial Crisis in 2008. This was done by the authors to aid the performance of the forecast models, but it does not allow us to as accurately draw a conclusion about the potential for ML methods to succeed in a real-world application. I am able to more clearly assess how all models tested perform during both the 2008 and 2020 recessions.

Finally, I modify the implementation of lags when fitting my LASSO regression model to reduce poor performance caused by over-penalization. Given that machine learning frequently relies on 'big data' datasets, the introduction of lags can lead to a very large number of predictors. For instance, Medeiros et al. introduced four lags to their data, which left them with 508 potential predictors. They also observed poor performance of the LASSO and LASSO model variations that were tested compared to other ML models. Because LASSO shrinks irrelevant predictor coefficients to zero and penalizes a large number of predictors, adding in a multitude of predictors via lags may cause 'crowding out' of otherwise relevant predictors. To solve this, I parameterize my lags according to Legendre polynomials in order to reduce the level of dimensionality present.

I find that with a full time horizon and the inclusion of outliers, the ML models heralded by Medeiros et al., such as the random forest (RF) model, are not able to outperform other forecasting methods. The RF model in particular seems to be sensitive to the treatment of outliers. The LASSO model I fit performs better than the RF model in the full estimation, standing in contrast with previous literature. This indicates that the treatment of lags that I employ is potentially effective in improving the LASSO model's performance. When using textual data, which involves a more limited time horizon due to the nature of the data, ML models are able to outperform standard methods, and the inclusion of textual data appears to improve these models, particularly the RF model.

I organize this thesis as follows. Section 2 provides a more in-depth review of the existing literature, with particular focus given to ML applications. Section 3 lays out the theoretical background of inflation forecasting. An overview of the data and key variables can be found in Section 4. Section 5 details the model specifications and methodology used for my estimations. Results are discussed in Section 6.

2 Literature Review

There is a vast body of literature proposing, critiquing, and refining methods for inflation forecasting. While direct forecasts and real-time forecasts based on survey or market measures remain important in the inflation forecasting field, in the past thirty years, there has been increasing attention given to developing models for inflation forecasting.

Some of the earlier and theoretically more simplistic models for inflation forecasting are univariate time series models, which take time series inflation data and incorporate an element of stochasticity. Univariate methods are very well established in forecasting and include autoregressive methods like those canonized by Box and Jenkins (1970), random walk based models as popularized by Atkeson and Ohanian (2001), and unobserved component stochastic volatility models (Harvey, 1989; Stock and Watson, 2007), alongside many others.

In 1999, Stock and Watson proposed an inflation forecasting model based on the theoretical Phillips Curve, with promising initial results. Ultimately, however, their further research demonstrated that while this model was successful when applied to short term forecasting for the time period between 1970 and 1996, it could not consistently outperform univariate models for later periods or longer-term horizons, even though it did outperform some other multivariate forecasting models. A similar pattern is present with other models that draw on comparable statistical and theoretical ideas, such as the utilization of vector autoregression by

Anderson et al. (2002) and Shoesmith (1992). These models ultimately failed to outperform univariate benchmarks, and simultaneously, research on the univariate benchmarks was finding them impossible to systematically improve (Stock and Watson, 2010).

After the popularization of Phillips Curve motivated models came more complex models based on Neoclassical and New Keynesian macroeconomic theory, where "decision rules of economic agents are derived from assumptions about preferences, technologies, and the prevailing fiscal and monetary policy regime by solving intertemporal optimization problems." The empirical estimation in these dynamic stochastic general equilibrium (DSGE) models is accomplished with Bayesian techniques (Elliot et al., 2006). The most popular DSGE models were developed by the U.S. Federal Reserve and the European Central Bank (ECB).

It was the consistent forecasting errors made by the ECB's DSGE forecast model for inflation that ultimately inspired Medeiros et al. (2019). These authors had the same goal as me: to improve inflation forecasting via machine learning methods. Specifically, they attempted to forecast headline PCE (though core measures and CPI are also discussed). They utilized the FRED-MD dataset—a database of monthly indicators created by the Federal Reserve for the purpose of machine learning and other 'big data' methods. The authors considered the vintage as of January 2016, with their data sample being from January 1960 to December 2015. Ultimately, their models analyzed 508 potential predictors, since they considered four lags of all 122 initial variables, as well as four autoregressive terms. Other adjustments to the data included removing outliers caused by the

2008 Financial Crisis.

The researchers considered a number of models for forecasting inflation, allowing them to compare various ML methods to traditional method benchmarks. Alongside their initial analysis, they also performed a real-time experiment for a subset of the time horizon and potential models.

Ultimately, Medeiros et al. presented a convincing set of results indicating that a number of ML methods could outperform traditional benchmarks. Particularly, they highlighted the random forest (RF) model, which had the smallest errors and consistent variable selection across time horizons. RF focused in on variables primarily from "prices, exchange and interest rates, and the housing and labor markets" and was the best method in periods of both high and low uncertainty. This suggested that the relationship between macroeconomic variables and inflation is nonlinear, and thus, linear methods for forecasting or approximating DSGE models may lead to inaccurate results. The work of Medeiros et al. is distinct from previous literature in that it demonstrated ML methods could beat univariate benchmarks for inflation forecasting, made an effort to highlight the key variables responsible for forecast improvements, and displayed the potential of RF methods in particular.

3 Theoretical Background

Throughout the history of macroeconomics, different schools of economic thought have proposed different theoretical explanations for inflation, laying out a single mechanism or combination of mechanisms that serve as the drivers of inflation. These theoretical drivers are frequently incorporated into the statistical models used to forecast inflation, and fundamentally underlie all of the models that I implement.

Inflation is often linked to the money supply via the Quantity Theory of Money, which describes a relationship between inflation and money supply that is directly proportional. (Specifically, $MV = PY$, where V is velocity of money, M is money supply, P is the price level, and Y is output.) Ceteris paribus, when the level of currency in circulation is increased, each individual unit of currency loses value. This leads to a higher overall price level (Friedman and Schwartz, 1963).

Phillips (1958) proposed an inverse relationship between wage inflation and unemployment, canonized as the Phillips Curve. Subsequent research building on this relationship solidified another potential driver of inflation: inflation expectations (Fisher, 1911; Friedman, 1968). Thus, the expectations augmented Phillips Curve was developed. It became a popular tool for inflation forecasting and is traditionally described as $\pi = \pi^e - \epsilon(U - U_N)$. This equation states that inflation is determined by inflation expectations as well as the gap between the unemployment rate and the "natural" rate of unemployment. (Here ϵ is assumed to be some positive constant that is fixed.)

The Phillips Curve also ties into another proposed driver of inflation, slack. Theories surrounding slack suggest that when firms utilize resources (both capital and labor) intensively, production costs rise, which causes firms to raise prices in order to maintain profit margins (Peach et al., 2011).

Similarly, cost-push inflation is an inflation driver that occurs when production costs rise. This can happen due to supply side shocks (such as a jump in a commodity prices due to an exogenous event) or because laborers demand increased wages (Samuelson and Solow, 1960). The idea that wage increases may lead to inflation increases in a cyclical nature is often referred to as the 'wage-price spiral.'

Structural inflation is related to the same type of shocks that may cause cost-push inflation. Essentially, theories of structural inflation state that market frictions result in goods shortages in certain economic sectors, which drives up prices (Olivera, 1964).

There is also demand-pull inflation, which highlights shifts in aggregate demand as a key inflation driver (Barth and Bennett, 1975). When aggregate demand increases and aggregate supply does not, this results in an increase in the overall price level.

Finally, there is the fiscal theory of inflation, which has garnered more attention recently. This theory asserts that inflation is primarily determined by a government's fiscal policy, rather than their monetary policy (Cochrane, 2023). The conceit is that if a government's fiscal policy does not allow it to pay off its debts out of future tax revenue, then, in order to fulfill its obligations, the government will allow a high level of inflation to reduce the value of its debts and close the gap in its inability to pay.

4 Data

4.1 Inflationary Trends

Following the 2008 financial crisis, Core PCE inflation, the measure of inflation used by the U.S. Federal Reserve, was consistently below its two percent target. This remained the case even as the labor market recovered from the recession and the unemployment rate began to drop. In response to persistently low inflation, in 2019, the Fed changed its inflation targeting scheme to Average Inflation Targeting, allowing inflation to float above the two percent target as long as the target was achieved on average. This change was largely to combat the risk of inflation expectations becoming anchored below the Fed's target (Martínez-García et al., 2021).

In 2020, there were significant supply side disruptions as a result of the COVID-19 pandemic. Simultaneously, to stimulate the economy, the Federal Reserve enacted a number of expansionary policies, dropping interest rates and engaging in Large Scale Asset Purchases. A number of fiscal policy measures were also taken, including the passing of the the Coronavirus Aid, Relief, and Economic Security Act (CARES Act) which provided stimulus checks to qualifying individuals in order to relieve financial distress.

Unsurprisingly, when world economies began to reopen, pent-up demand, increased currency in circulation, persistent supply shocks, and the effects of fiscal stimulus led to a steep increase in inflation. Between March of 2021 and February of 2022, Core PCE inflation climbed from 2.05% to a peak of 5.42%. The inflation realized by consumers was even more dramatic. The Headline Consumer Price Index (CPI) reached 9.00% in June of 2022. Since 2022, all measures of inflation have dropped. This is largely due to the Fed's hiking of interest rates and the depletion of the consumer wealth that built up during the pandemic as a result of fiscal policies. However, inflation remains elevated across the board. As of January 2023, Core PCE inflation was at 4.71%, and Core CPI was at 5.5%.

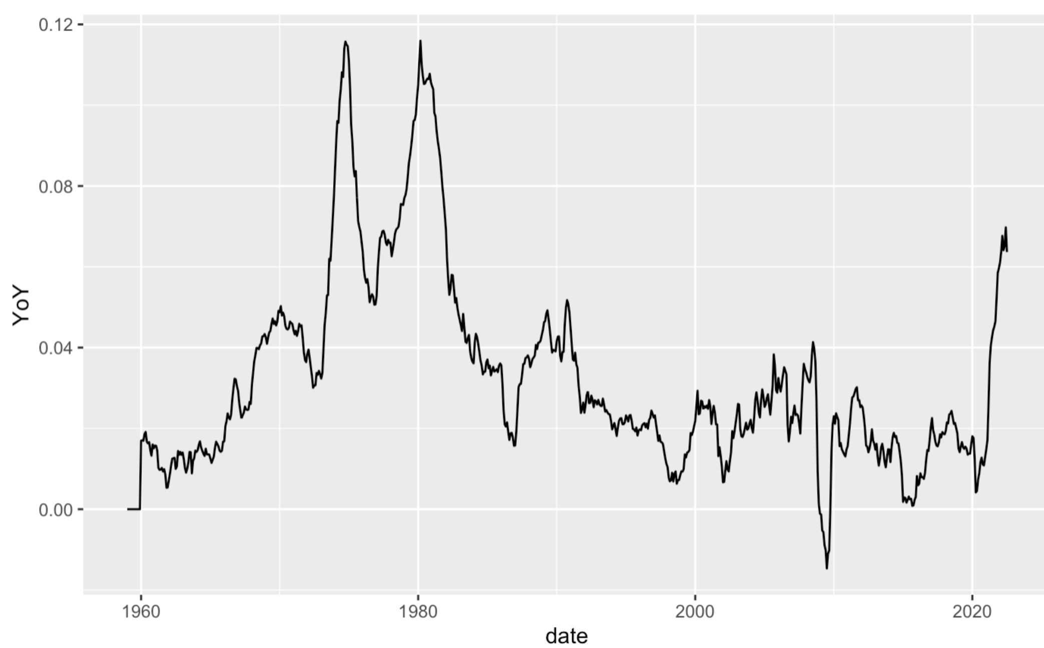


Figure 4.1: Personal Consumption Expenditures, Percent Change from a Year Ago

4.2 Key Variable and Summary Statistics

The Personal Consumption Expenditures Price Index (PCEPI) will serve as the key response variable, inflation, in my model. PCEPI measures the prices of goods and services paid by consumers in the U.S. and is typically chained to a base year, 2012. PCEPI is the preferred inflation measure of the Federal Reserve and is released by the Bureau of Economic Analysis.

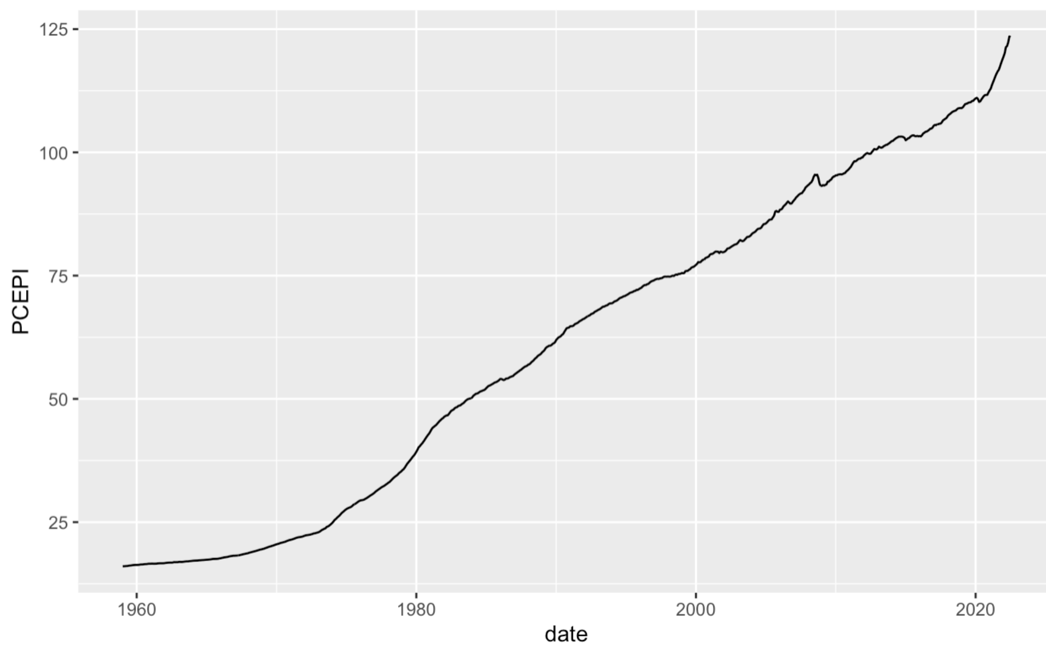


Figure 4.2: PCEPI, Index

There are several potential alternative variables. One of the most obvious alternatives would be the Consumer Price Index (CPI). Almost all research on inflation forecasting focuses on one of these four variables: CPI, Core CPI, PCE, or Core PCE, though there are additional ways of measuring inflation, such as the

GDP Deflator. Notably, PCE considers a broader range of expenditures than CPI and is calculated using data from firm surveys, as opposed to consumer surveys. I transform PCEPI to be stationary by taking the monthly log difference. Explicitly, this means that in my models, inflation (π) is given by

$$\pi_t = 100 \ln(\frac{P_t}{P_{t-1}}) \tag{4.1}$$

An overview of the appearance of this variable in each form is given in the table below.

Table 4.1: PCE Summary Statistics

	Index	Year/Year	Stationary
Min	16.04	-0.01467	-0.0110197
1st Quartile	27.37	0.01576	-0.0010963
Median	64.33	0.02478	-0.0000215
Mean	61.02	0.03229	-0.0000028
3rd Quartile	89.70	0.04206	0.0011613
Max	123.51	0.11594	0.0078035

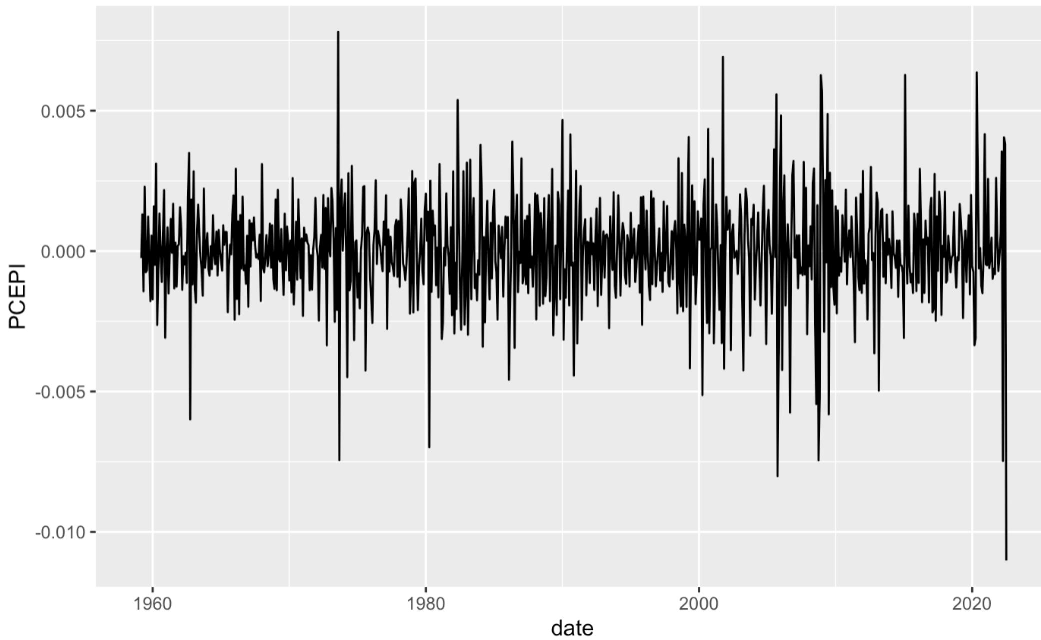


Figure 4.3: PCEPI, Stationary

4.3 Data Description

The primary data source for my thesis is FRED-MD, a database of macroeconomic variables that is updated and curated by the data desk at the Federal Reserve Bank of St. Louis. It consists of 134 monthly U.S. indicators and is intended to be used as a 'big data' source for macroeconomic analysis and research. The time series indicators included in FRED-MD extend from January 1959 to August 2022. The chosen indicators and their grouping are built to mimic the popular Stock-Watson dataset, while removing the need for researchers to accommodate data revisions, as well as updating in real time. The unit of observation, type of data, and sample design all vary within the database. For instance, Real Personal Income observes the individual, whereas Industrial Production: Consumer Goods observes relevant plants. The Civilian Unemployment Rate is based off of a household survey, while the 10-Year Treasury Rate is financial market data. While the data comes from a variety of sources—the U.S. Bureau of Labor Statistics, Moody's Corporation, and the U.S. Energy Information Administration, to name a few—the data is all of high quality, with very well-documented gathering procedures. A full list of the included indicators can be found in the Appendix.

FRED-MD as a data source has the strength of being high quality and of uniform frequency. Similar data sources have been used not only for forecasting with machine learning, but also for other methods of inflation forecasting as well. The Stock-Watson dataset that inspired FRED-MD was notably developed by the same researchers who developed the Phillips Curve model for inflation forecasting. One potential limitation of this data is that, given its purely time series nature, there is little nuance to provide regarding mean differences by characteristics. For instance, unemployment varies by race, gender, and geographic location, but I—and my data—will not be considering this nuance. Additionally, its monthly frequency is unhelpful for very short-term forecasts or nowcasts of inflation. That is why, if I wished to extend my models to apply to a near-term forecasting horizon, I would likely have to introduce higher frequency data, such as daily or hourly indicators (e.g., daily crude oil prices). Unfortunately, these higher frequency time series are relatively short, since historically, this type of data was not possible to track/gather. This severely limits their forecasting usefulness.

In addition to the macroeconomic indicators pulled from FRED-MD, for one of my estimations, I also introduce additional textual data. This data is sourced from Structure of News, a dataset crafted by financial economists at the Yale School of Management that summarizes Wall Street Journal articles from 1984 to 2017 into monthly topical themes. The 800,000 articles studied are simplified to a "bag of words" using Latent Dirichlet Allocation to reduce dimensionality; a vocabulary of 18,432 unique terms is transformed into 180 topics that are sorted into a taxonomy using hierarchical agglomerative clustering. To create monthly topic attention allocations, attention estimates of the 180 topics are summed over all articles published in the same calendar month (Bybee et al., 2021).

Ideally, this textual data serves as a proxy for inflation expectations and catches nuances that are not represented in the more traditional data sources. Notably, in utilizing this textual data, I am limited in time horizon, since the data is only available between 1984 and 2017. Once again, I also am limited in terms of geographic nuance, since my textual data is also nationwide.

5 Empirical Models

5.1 Benchmark Models

I estimate three benchmark models. These models represent the classic univariate time series models and the theoretical based models of inflation. I primarily use them to understand the performance of the ML methods employed. The first is a random walk (RW) model, which is the naive model described by Atkeson and Ohanian (2001). The simplest form of this model has a forecast equation for inflation (π) that is specified according to

$$\hat{\pi}_{t+h|t} = \pi_t \tag{5.1}$$

Here $h = 1, ..., 12$, where h is the forecasting horizon.

Next, I estimate an autoregressive (AR) model. I utilize a direct AR model rather than an iterative one. In this model, p is the number of autoregressive terms as determined by BIC, and the parameters ϕ are estimated by OLS (Henry and Pesaran, 1993).

The formal statement of the model is

$$\pi_{t+h|t} = \phi_{0,h}\pi_t + \phi_{1,h}\pi_t + ... + \phi_{p,h}\pi_{t-p+1} + u_t \tag{5.2}$$

Thus, my forecast equation is given as

$$\hat{\pi}_{t+h|t} = \hat{\phi}_{0,h}\pi_t + \hat{\phi}_{1,h}\pi_t + ... + \hat{\phi}_{p,h}\pi_{t-p+1} \tag{5.3}$$

Both of these models are standard univariate methods for inflation forecasting. In the previous literature, these models have also served as performance benchmarks and have been shown to be incredibly difficult to beat, especially over very long term forecast horizons.

I also estimate a theory based model for forecasting inflation, a Phillips Curve motivated model (PC). This multivariate model is based on the work of Gordon (1982) and describes future inflation as depending on lagged inflation and the unemployment rate U_t. (I have not included supply shock variables for simplicity's sake.)

$$\pi_{t+h|t} = \pi_t + U_t + u_t \tag{5.4}$$

Estimation is accomplished using the contemporaneous value plus four lags of both the U3 unemployment rate and PCE inflation.

5.2 Comparison Models

I then estimate a number of models that will be compared with the benchmarks.

The first is a factor model. In a factor model, to reduce the dimension of a model based on all possible predictors, I decompose a given indicator x_{it} to identify certain factors f_t that are present across all variables, but produce a certain effect. In the simplest form, the model can be described as:

$$x_{it} = \lambda_i' f_t + u_i t \tag{5.5}$$

Where λ_i represents the factor loadings for factors f_t. These latent factors are extracted and π_t is regressed on f_t to produce the forecast equation

$$\hat{\pi}_{t+h} = \hat{f}_t' \hat{\beta} \tag{5.6}$$

This method is particularly salient when dealing with economic data, since economists often think of there being driving forces or factors behind many realized changes in economic indicators.

I base my estimation on the methods described by Bai and Ng (2008). In accordance with this, I utilize a preselection step, wherein I only compute the principal components for variables that have passed a certain threshold t-statistic. By only selecting and estimating principle components for variables that have high predictive power for inflation, I reduce computational complexity significantly.

I then estimate a bagging model. Bagging, or bootstrap aggregating, involves aggregating a number of bootstrap samples. (A bootstrap sample is a replicate dataset that is drawn at random with replacement from the existing dataset's observations.) To create my bagging model, I estimate an OLS model for each bootstrap sample in order to select potential variables, using t-statistic as a gauge for predictive power. Then, I estimate a new regression with the variables that are selected. These coefficients are applied to compute the forecast on the actual sample. This is repeated for all bootstrap samples, with the final forecast being an average of all of the bootstrap forecasts. This approach is directly based on the work of Breiman (1996).

Next is a model based on complete subset regression (CSR). With complete subset regression, the goal is to average the forecasts created by all possible models, but this becomes difficult with high dimensions of data when there are many possible predictive variables (Elliott et al., 2013). To solve this issue, I estimate an OLS regression for each potential variable, and, similar to the bagging model, I select variables based on their t-statistic. This preselection step is necessary to reduce computational complexity. I then can calculate the CSR forecast to be the average of all the selected forecasts.

For my model utilizing shrinkage, I estimate a Least Absolute Shrinkage and Selection Operator (LASSO) model. With LASSO, I perform regularization on an OLS model to penalize the abundance of predictors. Depending on the specification, this process results in predictors being eliminated from the model, as their coefficients shrink to zero. This allows me to produce a sparse model. For LASSO, the objective is to solve

$$\min \left[\sum_{t=1}^{T-h} (\pi_{t+h} - \beta_h x_t)^2 + \lambda \sum_{i=1}^{n} |\beta_{h,i}| \right] \tag{5.7}$$

β is the vector of coefficients that maps to all covariates x. The penalty function, in accordance with the work of Tibshirani (1996), is $\lambda \sum_{i=1}^{n} |\beta_{h,i}|$.

Next, I produce a random forest (RF) forecast. Similar to bagging, a random forest model involves averaging over bootstrap samples. However, now I also employ decision trees. These trees are constructed from bootstrap samples, and the final forecast is an average of the forecasts produced by each tree, hence the moniker 'forest'. This is the crux of the method originally proposed by Breiman (2001).

Since random forest models are based on decision trees, not OLS, they are much better at handling nonlinear relationships. OLS assumes that there is a single relationship between regressors and that this relationship holds over all observations. On the contrary, decision trees are formed via recursive partitioning, which involves subsetting the dataspace into smaller and smaller regions. Each of these partitions represents a 'leaf' on the decision tree. For each leaf on the tree, a very simple model is fit. Specifically, it is assumed that the dependent variable is predicted by the elements of the leaf according to some constant.

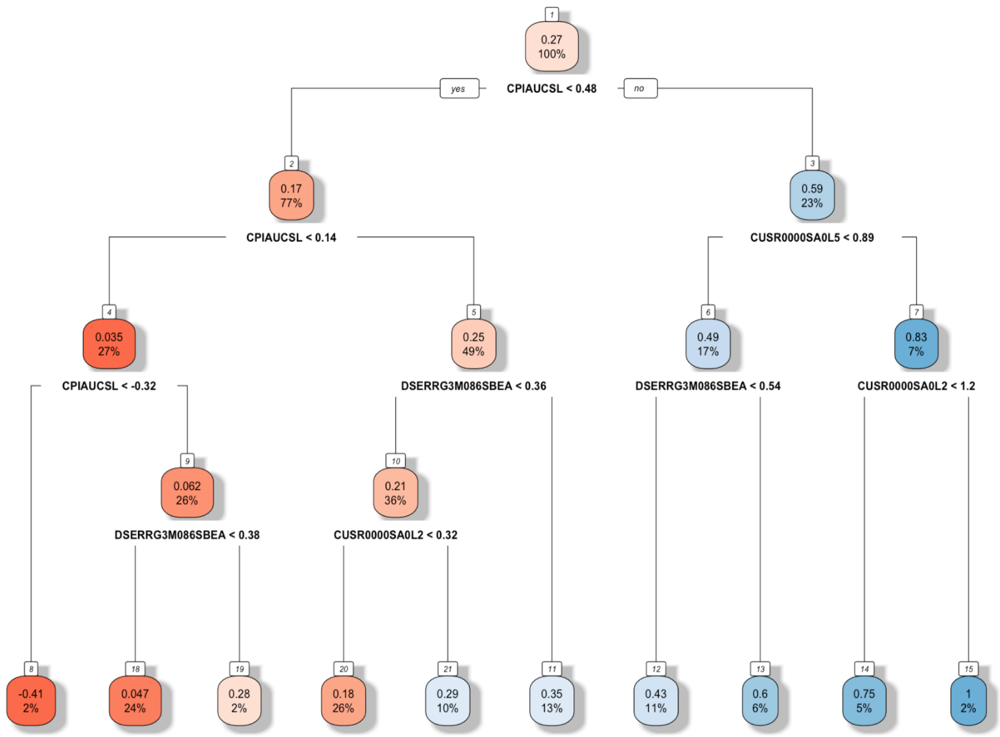


Figure 5.1: A Sample Decision Tree for Predicting PCEPI

Figure 5.1 illustrates a simplified decision tree for predicting PCEPI, based on my data. While this is considerably less complex than the decision trees utilized by my random forest model, it provides an idea of the basic structure of these trees.

In my case, then, the forecast equation for my RF model is given by

$$\hat{\pi}_{t+h} = \frac{1}{B}\sum_{b=1}^{B}\left[\sum_{k=1}^{Kb}\hat{c}_{k,b}I_{k,b}(x_t;\hat{\theta}_{k,b})\right] \tag{5.8}$$

B is the number of bootstrap samples. K_b are the terminal nodes or 'leaves.'

c is a constant that describes the relationship between all x and π. It is estimated by taking the sample average of realizations of the response variable within a given region, so, specifically, $\hat{c}_{k,b}$ is the average inflation in region k for a tree b. I is the product of the indicator function defining region splits. Specifically, it indicates if an x falls within a given region that is defined by a set of parameters θ.

$$I_k(x_t; \theta_k) = \begin{cases} 1, \text{if } x_t \in R_k(\theta_k) \\ 0, \text{otherwise} \end{cases} \tag{5.9}$$

The final forecast, then, estimates a relationship for each leaf of a tree and sums each of these leaves before finally averaging all trees together.

Finally, I use a number of methods to combine forecasts. Forecast combinations allow for 'hedging' of the risks present in any single model, potentially reducing forecast errors. For instance, if one model is susceptible to poor accuracy under certain economic conditions that another model performs well under, but under other conditions the opposite is true, theoretically, a forecast combination can result in improved accuracy (Weiss et al., 2019).

5.3 Methodology

5.3.1 Rolling Window

When creating my forecasts, I employ a rolling window. This means that I select a certain window size and estimate an h-step ahead forecast using the data present in that window. Next, I 'step forward' by one time increment, add that new observation to my dataset, and remove the oldest observation to then create another h-step ahead forecast. This approach makes my forecasts more robust to structural changes. All of the models are run through this rolling window mechanism.

5.3.2 Lags

To improve upon the issues that introducing lags may create with certain ML models, particularly LASSO, I parameterize the lags I introduce to my data to reduce the dimension of slope coefficients. The theory behind this process is as follows:

Suppose there is a time series $(y_t, x_t) \in R^2, 1 \leq t \leq T$ that has the following forecasting model with horizon h.

$$y_{t+h} = \sum_{j=0}^{m-1} \beta_j x_{t-j} + u_{t+h} \tag{5.10}$$

We can parameterize this model according to

$$\beta_j = w(\frac{j}{m}; \theta^h), \forall j = 0, ..., m-1 \tag{5.11}$$

where $w(s, \theta) = \sum_{l=1}^{L} \theta_l f_l(s), \forall s \in [0, 1]$, $\theta = (\theta_1, ..., \theta_L)^T, (f_l)_{l=1}^{L}$ is a set of simple functions, and L is small.

This allows us to achieve the following equivalencies:

$$y_{t+h} = \sum_{j=0}^{m-1} w(\frac{j}{m}; \theta^h) x_{t-j} + u_{t+h} \tag{5.12}$$

$$y_{t+h} = \sum_{j=0}^{m-1} \left[\sum_{l=1}^{L} \theta_l f_l(\frac{j}{m}) \right] x_{t-j} + u_{t+h} \tag{5.13}$$

$$y_{t+h} = \sum_{j=0}^{m-1} \theta_l \sum_{j=0}^{m-1} f_l(\frac{j}{m}) x_{t-j} + u_{t+h} \tag{5.14}$$

Thus, $y_{t+h} = \sum_{l=1}^{L} \theta_l z_{l,t} + u_{t+h}$, where $z_{l,t} = \sum_{j=0}^{m-1} f_l(\frac{j}{m}) x_{t-j}$.

If we align our data in vectors and matrices, we can express this as $y = \mathbf{XW}\theta + u$, where y is a response vector, u is an error vector, and $\mathbf{X}$ is a matrix of observations. This gives us a vector θ, as described previously, and a matrix of $\mathbf{W}$, which is composed of weights that are formed from the function $z_{l,t}$.

In practice, I parameterize the lags in my model according to Legendre polynomials, which serve as weights on all predictor and lagged predictor coefficients to help shrink dimensionality.

5.3.3 Textual Data Preselection

In order to reduce the number of predictors in my models, when working with the textual data, I perform a preselection step. I create a correlation matrix between all textual topics and inflation and only select topics that demonstrate predictive power. The threshold for correlation when selecting this data is 0.2. That is, if the absolute value of the correlation between a given predictor and inflation is greater than 0.2, it is selected. A table of the selected textual variables can be found in the Appendix.

6 Results

I compare all of my models via a number of methods, primarily root mean squared error (RMSE) and mean absolute error (MAE).

I also utilize the Diebold-Mariano test to compare predictive accuracy between models and determine if differences in accuracy are statistically significant.

6.1 Full Estimation

First, I compute results for the full dataset time horizon. The forecasts are computed for the length of the set of rolling windows, which consists of the 2007 to 2022 period. This estimation does not incorporate the utilization of textual data. Table 6.1 reports the RMSE for a selection of models relative to the random walk (RW) for the 1-, 3-, 6-, 9-, and 12-month forecasting horizons. Table 6.2 reports the MAE for the same models and forecast horizons. An unabbreviated version of these tables, including all horizons and models, can be found in the Appendix. The lowest error in each horizon is bolded.

For both RMSE and MAE, the best performing model for the short term horizons is the factor model. Specifically, the factor model outperforms the AR model in the 1-, 2-, and 3-month step ahead forecasts. For the mid-term forecasts, the AR model outperforms all of the other models. In the year ahead forecast, however, the factor model is once again the most accurate. Following the AR and factor

Table 6.1: Full Estimation Root Mean Squared Error

	t+1	t+3	t+6	t+9	t+12	Mean
AR	0.10496	0.17114	**0.15622**	**0.15042**	0.17611	**0.15542**
CSR	0.11165	0.18857	0.17475	0.17213	0.19521	0.17120
Factor	**0.09626**	**0.16762**	0.18144	0.19139	**0.15387**	0.16592
Bagging	0.24622	0.34992	0.31635	1.01308	0.35077	0.47153
Lasso	0.10587	0.17964	0.16897	0.17164	0.18567	0.16696
RF	0.13379	0.20604	0.18369	0.18587	0.21960	0.19093

The table shows the root mean squared error (RMSE) for a selection of models relative to the random walk (RW). The error measures were calculated from 180 rolling windows covering the 2007 to 2022 period. Columns 1-5 show a selection of the forecasting horizons for readability. The lowest error for each horizon is bolded. Column 6 displays the mean RMSE for each model over all forecasting horizons.

Table 6.2: Full Estimation Mean Absolute Error

	t+1	t+3	t+6	t+9	t+12	Mean
AR	0.07430	0.12018	**0.11199**	**0.10689**	0.12639	**0.11041**
CSR	0.08007	0.12875	0.12331	0.12154	0.14139	0.12011
Factor	**0.05507**	**0.11799**	0.12721	0.13023	**0.11343**	0.11500
Bagging	0.13428	0.19943	0.18820	0.28876	0.23740	0.21981
Lasso	0.07718	0.12716	0.11857	0.11852	0.13440	0.11825
RF	0.09271	0.14042	0.12550	0.12703	0.15076	0.12993

The table shows the mean absolute error (MAE) for a selection of models relative to the random walk (RW). The error measures were calculated from 180 rolling windows covering the 2007 to 2022 period. Columns 1-5 show a selection of the forecasting horizons for readability. The lowest error for each horizon is bolded. Column 6 displays the mean MAE for each model over all forecasting horizons.

models, the next best performing model is the LASSO model. This can be further seen in the last columns of Table 6.1 and Table 6.3, which report the mean RMSE and MAE values for select models over all forecasting horizons. (An expanded version of these tables can be found in the Appendix.) For both MAE and RMSE, the order of model performance is AR, Factor, LASSO, CSR, RF, then Bagging.

To further understand the performance of the models, I graph the squared error of each model over the 180 window horizon. This allows me to see what time periods different models struggled to forecast, as well as to more directly compare errors in a way that is not as reductive as an average or median. Figure 6.1 and 6.2 display these errors for the AR, Factor, LASSO, and RF models on the t+1 and t+12 forecasting horizons respectively. The large spikes in errors seen both early and late in these figures correspond to the 2008 and 2020 recessions respectively.

For the one month ahead forecast, as shown in Figure 6.1, we see that during the 2008 financial crisis, both the factor and LASSO models are better than the AR model at handling the shock, realizing a smaller spike in errors during this period. The factor model performs incredibly well for this horizon and time period, avoiding the spike in errors from which all of the other models suffer. However, during the Covid-19 pandemic, the one month ahead factor forecast sees a monumental spike in forecasting error, while the other models handle the shock with relative ease. During the 2020 to 2022 period, the RF model realizes less errors than the LASSO model and is more on par with the AR model.

Some similar patterns are present in the one year ahead forecast. The factor model vastly outperforms all others during the 2008 crisis period, but underperforms during the 2020 recession. The LASSO model sees smaller spikes in error than the AR model during both periods of uncertainty, but only outperforms AR during the 2020 period.

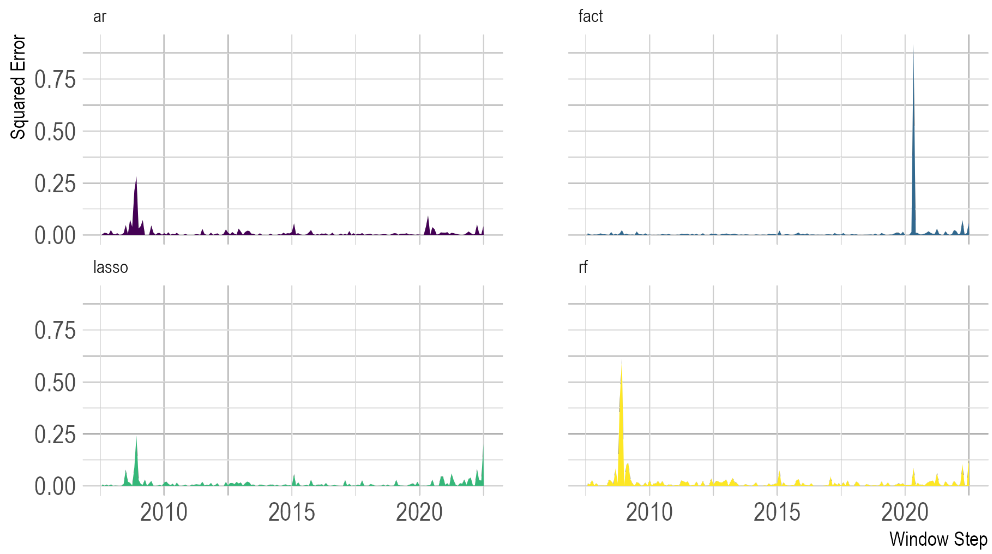


Figure 6.1: Error for selected models over time, t+1 forecasting horizon

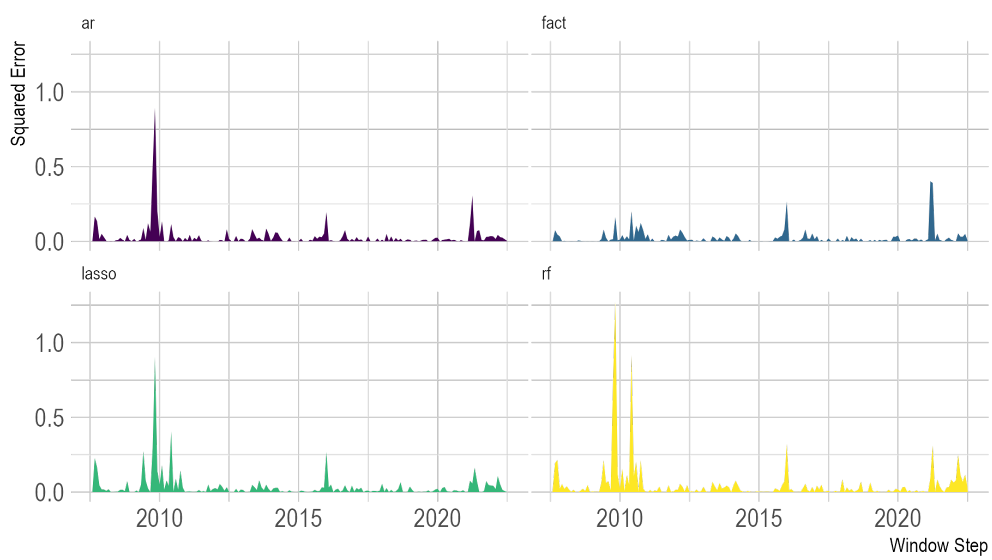


Figure 6.2: Error for selected models over time, t+12 forecasting horizon

For all models, I run the Diebold-Mariano test. With Diebold-Mariano, I compare the errors (RMSE or MAE) of two forecast models to see if the predictive accuracies of the models are statistically significantly different. The null hypothesis is that the predictive accuracies of the two forecasts are the same; if a small enough p-value (usually <0.05) is seen, then we can reject the null hypothesis and conclude that the accuracies of the two models are actually different. I find that the AR model is significantly different in predictive accuracy than all models except for the factor model. For the AR/factor model comparison, I am unable to reject the null hypothesis.

In this estimation, only the factor model can beat out AR in RMSE, but in looking at the squared errors over time, we do see that there are times when the ML models perform well. There is a very disparate effect between how models handle the 2008 and 2020 recessions. This stark difference in performance indicates that there may be potential for combination forecasts to improve accuracy, with Factor + LASSO or Factor + RF seeming particularly promising.

Using a variety of methods, I produce a combination forecast for both Factor + LASSO and Factor + RF. (A table of these various methods and their performance can be found in the Appendix.) After comparing combination methods for the one step ahead forecast and selecting the best performing method, I see that the Factor + LASSO model shows the most promise. I then produce a Factor + LASSO combination model for all forecasting horizons. The performance of this model is reported in Table 6.3.

Ultimately, for both RMSE and MAE this combination forecast is not able to

Table 6.3: Factor/LASSO Combination RMSE and MAE

	RMSE	MAE
t+1	**0.16898**	**0.12370**
t+2	0.20099	0.14677
t+3	0.20129	0.14560
t+4	0.20279	0.14586
t+5	0.20379	0.14898
t+6	0.21584	0.15729
t+7	0.21486	0.15370
t+8	0.22364	0.15939
t+9	0.21251	0.15336
t+10	0.21604	0.15324
t+11	0.21292	0.15162
t+12	0.21891	0.15700
Mean	0.20771	0.14971

The table shows the mean absolute error (MAE) and root mean squared error (RMSE) for the factor + LASSO combination model. The error measures were calculated from 180 rolling windows covering the 2007 to 2022 period. The combination method used is Least Absolute Deviation regression with rolling weights. The forecast horizon with the lowest error is bolded.

reduce errors significantly enough to outperform the other models.

6.2 Textual Data Estimation

I then perform the same estimations and analysis but with the inclusion of textual data. Because of the nature of my textual dataset, I am limited to the time period between 1984 and 2017. For the sake of comparison, I run all models on this time horizon without the textual data as well. Excitingly, in this setting, the RF model is able to outperform the AR model, especially in longer forecast horizons. (The factor model is the best performing for the one month ahead forecast.)

Table 6.4: Textual Estimation Root Mean Squared Error

	t+1	t+2	t+3
AR	0.15373	0.22775	0.22952
CSR	0.14424	0.22961	0.24061
Factor	**0.11393**	0.20349	0.21851
Lasso	0.17002	0.21178	0.26694
RF	0.16145	**0.19924**	**0.19900**

The table shows the root mean squared error (RMSE) for a selection of models relative to the random walk (RW), in an estimation including textual data. The lowest error for each forecast horizon is bolded.

In Figure 6.3, we see that the factor, LASSO, and RF models outperform the AR model, and they particularly perform well during the 2008 financial crisis, which is the large spike present in the center of these figures.

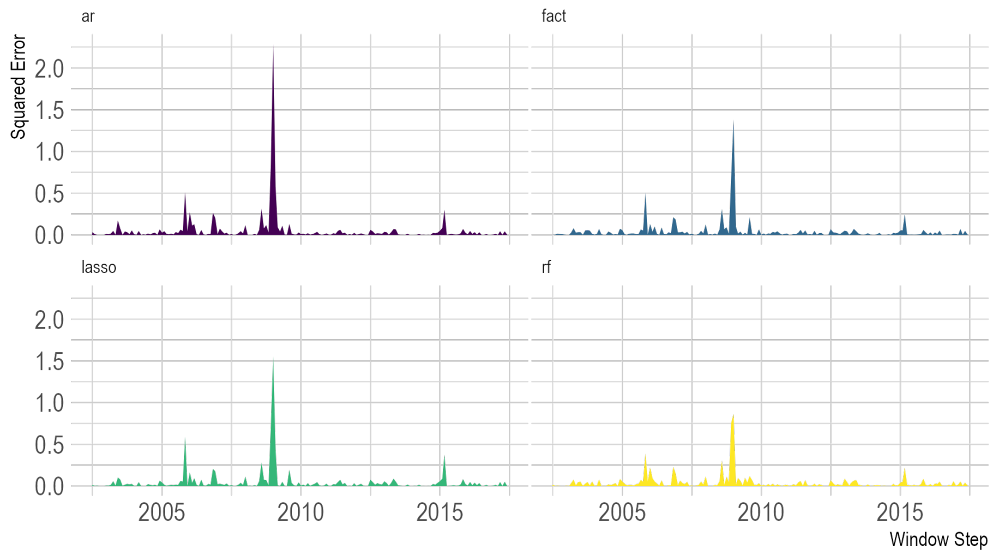


Figure 6.3: Error with textual data for selected models over time, t+2 forecasting horizon

However, when I use the Diebold-Mariano test to check for statistically significant difference in accuracy between the models that use textual data and the models fitted to the same horizon without textual data, I am unable to reject the null hypothesis.

Adding in textual data improves the performance of the ML models, especially the RF model, though it is not enough to prove a statistically significant difference in accuracy. Additionally, since I do not have textual data past 2017, I am unable to see how the textual data might have affected the performance of the ML models during the Covid-19 pandemic. This is one of the major limitations of my analysis.

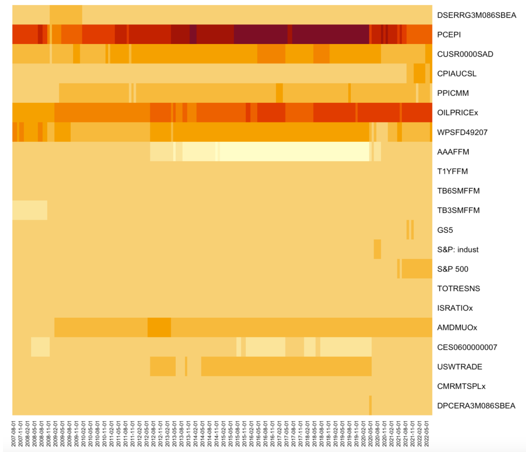
6.3 Variable Selection

In order to better understand the performance of my ML models, I explore the variable selection of the LASSO and RF specifications.

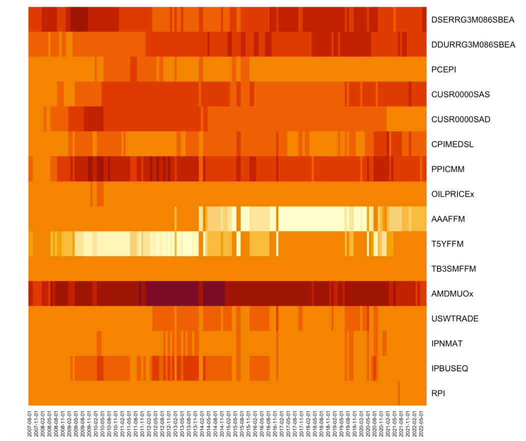
Figures 6.4 and 6.5 report the variable importance for the LASSO and RF models respectively for four different forecasting horizons. In order to improve legibility, for the LASSO model, variables that were not selected (i.e. variables that were shrunk to zero for the entire forecasting period) are not included in the figure. Similarly, lags of variables are not displayed.

These heat maps immediately make apparent the sparsity of the LASSO model in comparison to the RF model. Additionally, we can easily see that not only does the variable selection vary between the two models, but it also varies over forecasting horizons. In general, for both models, the level of model complexity tends to be less for very short term forecasts. This complexity increases as the forecasting horizon increases, before eventually becoming more simplistic again for the longer term horizons. This matches with what we would expect given the previously discussed prior forecasting literature and theories of inflation. In a one month ahead forecast, inflation in the current month is likely a very reliable predictor, and we see that both the RF and LASSO models favor PCEPI (the current inflation level) very heavily in their one step ahead forecast. Conversely, in the long run, inflation tends to return to a given norm (recall that not only do central banks target inflation levels to maintain steady inflation over time, but also very simple models, such as AR or RW, tend to perform well in the very long term because they often hover around and assume a return to these historical norm levels). In general, we can think of long term inflationary trends as being less driven by exogenous shocks/noise and more driven by the ‘underlying forces’ of inflation. We see that for the LASSO model, the year ahead forecast focuses on durable and nondurable goods expenditures, unfilled orders for durable goods, housing starts in the Midwest, and wholesale trade.

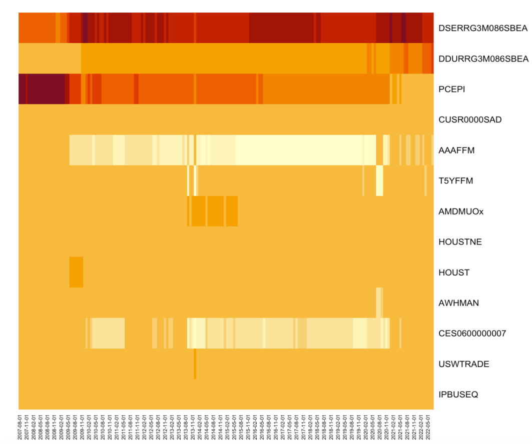
The t+3 forecast is the most complex of those compared in the figures for both of the models, but we see that while the LASSO model is primarily focused on price variables, the RF model incorporates more complexity, such as Moody’s Aaa Corporate Bond Yields Minus FEDFUNDS and business inventories. Interestingly, the variables selected lean heavily towards prices and output categories, but focus much less on employment than we might expect based off of the theories of inflation discussed in Chapter 3.



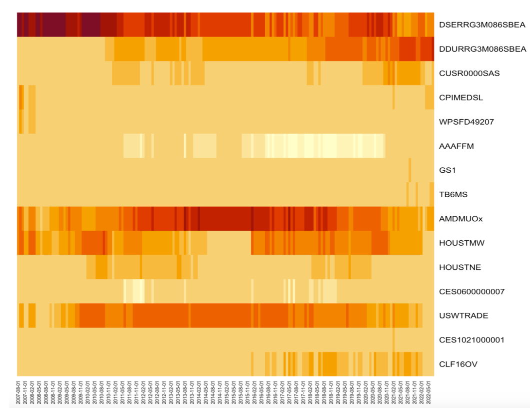
(a) t+1



(b) t+3

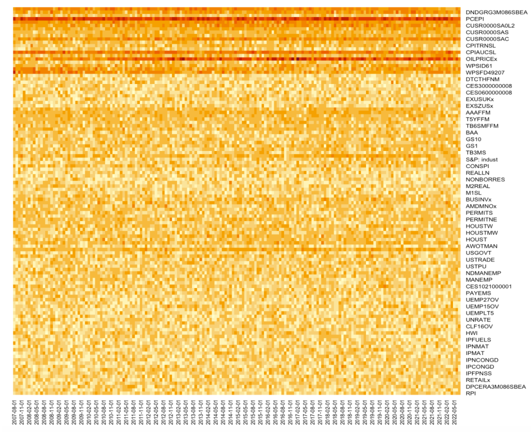


(c) t+6

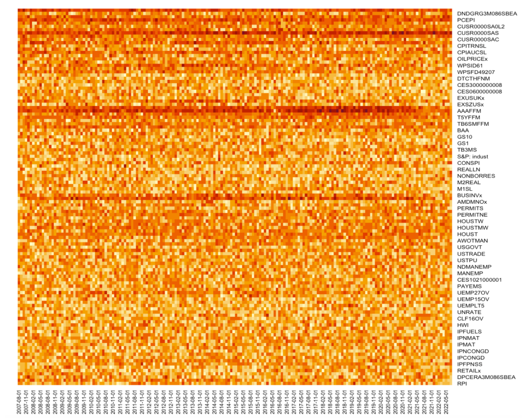


(d) t+12

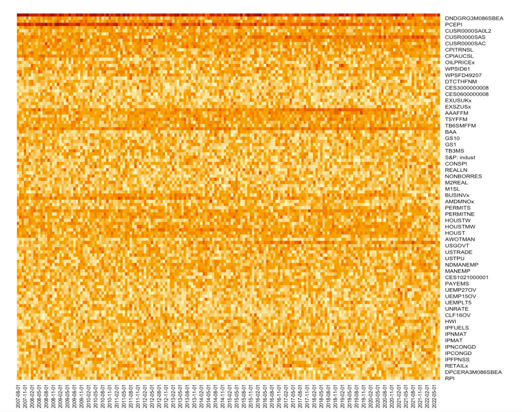
Figure 6.4: LASSO full estimation variable importance



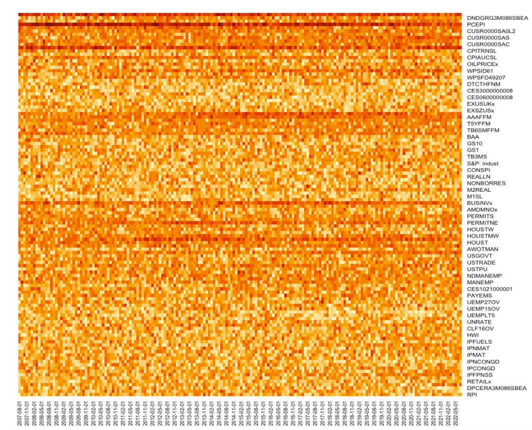
(a) t+1



(b) t+3



(c) t+6



(d) t+12

Figure 6.5: Random Forest full estimation variable importance

7 Conclusions

I demonstrate that the specificities of ML models are incredibly important to their ultimate efficacy as forecasting tools. When including the outlying data that previous literature did not, ML models are not able to outperform other forecasting methods, with the RF model appearing to be the most sensitive to these outlying data points. Similarly, the treatment of lags can significantly affect ML model performance. In the case of the LASSO model, by changing the treatment of lags from that which was utilized in previous literature, I am able to improve the model's performance. Via the inclusion of textual data, I am able to improve the performance of ML models, particularly the RF model, such that they outperform standard benchmarks in most forecasting horizons. However, this estimation is limited by the time horizon of my available dataset.

While I am not able to replicate the results of Medeiros et al., ML models in my research do show promise in their usefulness as forecasting tools. They accommodate well for nonlinearities and do not rely as heavily on underlying assumptions about theoretical inflation drivers, making them more flexible during times of unusual economic activity. (For instance, the factor model struggled to forecast during the COVID-19 pandemic, while the LASSO model was much more flexible.) The LASSO model particularly realized less errors during economic crises than any other model (though this did not balance out its higher realized errors during periods of higher certainty). This ability to cope with economic crisis could be useful for the policy making of central banks during periods when other forecasting models flounder. Both the LASSO and RF models also showed stability in their variable selection mechanisms.

Notably, there appears to be promise in the incorporation of textual data. This is an area of inflation forecasting that is novel to my research and definitely should receive further attention and testing going forward, especially as this type of data becomes more readily accessible. While textual data, because of its inherently large datasets, is unwieldy for more traditional statistical models, it is utilized to great effect by ML models.

Given that ML models performed better during unusual economic conditions, namely the LASSO model (or the RF model, when textual data is available), these models may be particularly useful for countries that experience more unusual inflationary conditions, or where the typical inflation drivers do not as strongly hold.

Ultimately, further research on inflation forecasting should consider ML models, but more attention needs to be given to the tuning of these ML tools in order to better understand how they could best be employed.

Bibliography

R. G. Anderson, D. L. Hoffman, and R. H. Rasche. A vector error-correction forecasting model of the us economy. *Journal of Macroeconomics*, 24(4):569–598, 2002. doi: 10.1016/s0164-0704(02)00067-8.

A. Atkeson and L. E. Ohanian. Are phillips curves useful for forecasting inflation? *Quarterly Review*, 25(1), 2001. doi: 10.21034/qr.2511.

A. Babii, E. Ghysels, and J. Striaukas. Machine learning time series regressions with an application to nowcasting. *Journal of Business and Economic Statistics*, 40(3):1094–1106, 2021. doi: 10.1080/07350015.2021.1899933.

J. Bai and S. Ng. Forecasting economic time series using targeted predictors. *Journal of Econometrics*, 146(2):304–317, 2008. doi: 10.1016/j.jeconom.2008.08.010.

J. R. Barth and J. T. Bennett. Cost-push versus demand-pull inflation: Some empirical evidence: Comment. *Journal of Money, Credit and Banking*, 7(3):391, 1975. doi: 10.2307/1991632.

Box and Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-Day, 1970.

L. Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, 1996. doi: 10.1007/bf00058655.

L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001. doi: 10.1023/a:1010933404324.

L. Bybee, B. Kelly, A. Manela, and D. Xiu. Business news and business cycles. *SSRN*, 2021. doi: 10.3386/w29344.

G. Caperna, M. Colagrossi, A. Geraci, and G. Mazzarella. Googling unemployment during the pandemic: Inference and nowcast using search data. *SSRN Electronic Journal*, 2020. doi: 10.2139/ssrn.3627754.

J. H. Cochrane. *Fiscal theory of the price level*. Princeton University Press, 2023.

G. Elliot, C. W. J. Granger, and A. Timmerman. *Handbook of Economic Forecasting*. Elsevier North-Holland, 1st edition, 2006.

G. Elliott, A. Gargano, and A. Timmermann. Complete subset regressions. *Journal of Econometrics*, 177(2):357–373, 2013. doi: 10.1016/j.jeconom.2013.04.017.

I. Fisher. *The purchasing power of money: Its determination and relation to credit interest and crises*. The MacMillan Company, 1911.

M. Friedman. The role of monetary policy. *The American Economic Review*, 58 (1), Mar 1968.

M. Friedman and A. J. Schwartz. *A monetary history of the United States: 1867-1960*. Princeton Univ. Press, 1963.

R. J. Gordon. Inflation, flexible exchange rates, and the natural rate of unemployment. *Workers, Jobs and Inflation,*, page 89–158, 1982. doi: 10.3386/w0708.

S. Gu, B. T. Kelly, and D. Xiu. Empirical asset pricing via machine learning. *SSRN Electronic Journal*, 2018. doi: 10.2139/ssrn.3159577.

A. C. Harvey. *Forecasting, structural time series models and the Kalman filter*. Cambridge Univ. Press, 1989.

S. G. B. Henry and B. Pesaran. Var models of inflation. *Bank of England Quarterly Bulletin*, page 231–239, May 1993.

E. Kalamara, A. Turrell, C. Redl, G. Kapetanios, and S. Kapadia. Making text count: Economic forecasting using newspaper text. *SSRN Electronic Journal*, Aug 2020. doi: 10.2139/ssrn.3610770.

E. Martínez-García, J. Coulter, and V. Grossman. Fed's new inflation targeting policy seeks to maintain well-anchored inflation expectations, 2021. URL https://www.dallasfed.org/research/economics/2021/0406.

M. C. Medeiros, G. F. Vasconcelos, Veiga, and E. Zilberman. Forecasting inflation in a data-rich environment: The benefits of machine learning methods. *Journal of Business and Economic Statistics*, 39(1):98–119, 2019. doi: 10.1080/07350015.2019.1637745.

J. H. Olivera. On structural inflation and latin-american 'structuralism'. *Oxford Economic Papers*, 16(3):321–332, 1964. doi: 10.1093/oxfordjournals.oep.a040958.

R. W. Peach, R. W. Rich, and A. Cororaton. How does slack influence inflation? *SSRN Electronic Journal*, 2011. doi: 10.2139/ssrn.1895526.

A. W. Phillips. The relation between unemployment and the rate of change of money wage rates in the united kingdom, 1861-1957. *Economica*, 25(100):283, 1958. doi: 10.2307/2550759.

P. A. Samuelson and R. M. Solow. Analytical aspects of anti-inflation policy. *The American Economic Review*, 50(2):177–194, 1960. doi: http://www.jstor.org/stable/1815021.

G. L. Shoesmith. Co-integration, error correction and improved medium-term regional var forecasting. *Journal of Forecasting*, 11(2):91–109, 1992. doi: 10.1002/for.3980110202.

J. Stock and M. Watson. Forecasting inflation. *Journal of Monetary Economics*, 44(2):293–335, 1999. doi: 10.3386/w7023.

J. Stock and M. Watson. Modeling inflation after the crisis. *National Bureau of Economic Research*, 2010. doi: 10.3386/w16488.

J. H. Stock and M. W. Watson. Why has u.s. inflation become harder to forecast? *Journal of Money, Credit and Banking*, 39:3–33, 2007. doi: 10.1111/j.1538-4616.2007.00014.x.

R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996. doi: 10.1111/j.2517-6161.1996.tb02080.x.

E. Weiss, Christoph, E. Raviv, and G. Roetzer. Forecast combinations in r using the forecastcomb package. *The R Journal*, 10(2):262, 2019. doi: 10.32614/rj-2018-052.

8 Appendix

8.1 FRED-MD Data

Table 8.1 displays the indicators included in the FRED-MD dataset. Column 1 is the indicator number, Column 2 is the transformation code, Column 3 is the indicator name in FRED, Column 4 is a descriptive name, and Column 5 is the group code number. The FRED-MD data groups (in numerical order) are Output and Income; Labor Market; Housing; Consumption, orders, and inventories; Money and Credit; Interest and Exchange Rates; Prices; and Stock Market.

Table 8.1: FRED-MD Indicators

ID	TC	FRED ID	Name	GC
1	5	RPI	Real Personal Income	1
2	5	W875RX1	Real personal income ex transfer receipts	1
3	5	DPCERA3M086SBEA	Real personal consumption expenditures	4
4	5	CMRMTSPLx	Real Manu. and Trade Industries Sales	4
5	5	RETAILx	Retail and Food Services Sales	4
6	5	INDPRO	IP Index	1
7	5	IPFPNSS	IP: Final Products and Nonindustrial Supplies	1
8	5	IPFINAL	IP: Final Products (Market Group)	1
9	5	IPCONGD	IP: Consumer Goods	1
10	5	IPDCONGD	IP: Durable Consumer Goods	1
11	5	IPNCONGD	IP: Nondurable Consumer Goods	1
12	5	IPBUSEQ	IP: Business Equipment	1
13	5	IPMAT	IP: Materials	1
14	5	IPDMAT	IP: Durable Materials	1
15	5	IPNMAT	IP: Nondurable Materials	1
16	5	IPMANSICS	IP: Manufacturing (SIC)	1
17	5	IPB51222s	IP: Residential Utilities	1
18	5	IPFUELS	IP: Fuels	1
19	1	NAPMPI	ISM Manufacturing: Production Index	1
20	2	CUMFNS	Capacity Utilization: Manufacturing	1
21	2	HWI	Help-Wanted Index for United States	2
22	2	HWIURATIO	Ratio of Help Wanted/No. Unemployed	2
23	5	CLF16OV	Civilian Labor Force	2
24	5	CE16OV	Civilian Employment	2
25	2	UNRATE	Civilian Unemployment Rate	2
26	2	UEMPMEAN	Average Duration of Unemployment (Weeks)	2
27	5	UEMPLT5	Civilians Unemployed - Less Than 5 Weeks	2
28	5	UEMP5TO14	Civilians Unemployed for 5-14 Weeks	2
29	5	UEMP15OV	Civilians Unemployed - 15 Weeks & Over	2
30	5	UEMP15T26	Civilians Unemployed for 15-26 Weeks	2
31	5	UEMP27OV	Civilians Unemployed for 27 Weeks and Over	2
32	5	CLAIMSx	Initial Claims	2
33	5	PAYEMS	All Employees: Total nonfarm	2
34	5	USGOOD	All Employees: Goods-Producing Industries	2
35	5	CES1021000001	All Employees: Mining and Logging: Mining	2
36	5	USCONS	All Employees: Construction	2
37	5	MANEMP	All Employees: Manufacturing	2
38	5	DMANEMP	All Employees: Durable goods	2
39	5	NDMANEMP	All Employees: Nondurable goods	2

Continued on next page

Table 8.1 – continued from previous page

ID	TC	FRED ID	Name	GC
40	5	SRVPRD	All Employees: Service-Providing Industries	2
41	5	USTPU	All Employees: Trade, Transportation & Utilities	2
42	5	USWTRADE	All Employees: Wholesale Trade	2
43	5	USTRADE	All Employees: Retail Trade	2
44	5	USFIRE	All Employees: Financial Activities	2
45	5	USGOVT	All Employees: Government	2
46	1	CES0600000007	Avg Weekly Hours : Goods-Producing	2
47	2	AWOTMAN	Avg Weekly Overtime Hours : Manufacturing	2
48	1	AWHMAN	Avg Weekly Hours : Manufacturing	2
49	1	NAPMEI	ISM Manufacturing: Employment Index	2
50	4	HOUST	Housing Starts: Total New Privately Owned	3
51	4	HOUSTNE	Housing Starts, Northeast	3
52	4	HOUSTMW	Housing Starts, Midwest	3
53	4	HOUSTS	Housing Starts, South	3
54	4	HOUSTW	Housing Starts, West	3
55	4	PERMIT	New Private Housing Permits (SAAR)	3
56	4	PERMITNE	New Private Housing Permits, Northeast (SAAR)	3
57	4	PERMITMW	New Private Housing Permits, Midwest (SAAR)	3
58	4	PERMITS	New Private Housing Permits, South (SAAR)	3
59	4	PERMITW	New Private Housing Permits, West (SAAR)	3
60	4	NAPM	ISM Manufacturing: PMI Composite Index	4
61	4	NAPMNOI	ISM Manufacturing: New Orders Index	4
62	4	NAPMSDI	ISM Manufacturing: Supplier Deliveries Index	4
63	4	NAPMII	ISM Manufacturing: Inventories Index	4
64	5	ACOGNO	New Orders for Consumer Goods	4
65	5	AMDMNOx	New Orders for Durable Goods	4
66	5	ANDENOx	New Orders for Nondefense Capital Goods	4
67	5	AMDMUOx	Unfilled Orders for Durable Goods	4
68	5	BUSINVx	Total Business Inventories	4
69	2	ISRATIOx	Total Business: Inventories to Sales Ratio	4
70	6	M1SL	M1 Money Stock	5
71	6	M2SL	M2 Money Stock	5
72	5	M2REAL	Real M2 Money Stock	5
73	6	BOGMBASE	Monetary Base	5
74	6	TOTRESNS	Total Reserves of Depository Institutions	5
75	7	NONBORRES	Reserves Of Depository Institutions	5
76	6	BUSLOANS	Commercial and Industrial Loans	5
77	6	REALLN	Real Estate Loans at All Commercial Banks	5
78	6	NONREVSL	Total Nonrevolving Credit	5
79	2	CONSPI	Nonrevolving consumer credit to Personal Income	5
80	5	S&P 500	S&P's Common Stock Price Index: Composite	8
81	5	S&P: indust	S&P's Common Stock Price Index: Industrials	8
82	2	S&P div yield	S&P's Composite Common Stock: Dividend Yield	8
83	5	S&P PE ratio	S&P's Composite Common Stock: Price-Earnings Ratio	8
84	2	FEDFUNDS	Effective Federal Funds Rate	6
85	2	CP3Mx	3-Month AA Financial Commercial Paper Rate	6
86	2	TB3MS	3-Month Treasury Bill:	6
87	2	TB6MS	6-Month Treasury Bill:	6
88	2	GS1	1-Year Treasury Rate	6
89	2	GS5	5-Year Treasury Rate	6
90	2	GS10	10-Year Treasury Rate	6
91	2	AAA	Moody's Seasoned Aaa Corporate Bond Yield	6
92	2	BAA	Moody's Seasoned Baa Corporate Bond Yield	6
93	1	COMPAPFFx	3-Month Commercial Paper Minus FEDFUNDS	6
94	1	TB3SMFFM	3-Month Treasury C Minus FEDFUNDS	6
95	1	TB6SMFFM	6-Month Treasury C Minus FEDFUNDS	6
96	1	T1YFFM	1-Year Treasury C Minus FEDFUNDS	6
97	1	T5YFFM	5-Year Treasury C Minus FEDFUNDS	6
98	1	T10YFFM	10-Year Treasury C Minus FEDFUNDS	6
99	1	AAAFFM	Moody's Aaa Corporate Bond Minus FEDFUNDS	6
100	1	BAAFFM	Moody's Baa Corporate Bond Minus FEDFUNDS	6
101	5	TWEXAFEGSMTHx	Trade Weighted U.S. Dollar Index	6
102	5	EXSZUSx	Switzerland / U.S. Foreign Exchange Rate	6
103	5	EXJPUSx	Japan / U.S. Foreign Exchange Rate	6
104	5	EXUSUKx	U.S. / U.K. Foreign Exchange Rate	6
105	5	EXCAUSx	Canada / U.S. Foreign Exchange Rate	6
106	6	WPSFD49207	PPI: Finished Goods	7
107	6	WPSFD49502	PPI: Finished Consumer Goods	7
108	6	WPSID61	PPI: Intermediate Materials	7
109	6	WPSID62	PPI: Crude Materials	7
110	6	OILPRICEx	Crude Oil, spliced WTI and Cushing	7
111	6	PPICMM	PPI: Metals and metal products:	7
112	1	NAPMPRI	ISM Manufacturing: Prices Index	7
113	6	CPIAUCSL	CPI : All Items	7
114	6	CPIAPPSL	CPI : Apparel	7

Continued on next page

Table 8.1 – continued from previous page

ID	TC	FRED ID	Name	GC
115	6	CPITRNSL	CPI : Transportation	7
116	6	CPIMEDSL	CPI : Medical Care	7
117	6	CUSR0000SAC	CPI : Commodities	7
118	6	CUSR0000SAD	CPI : Durables	7
119	6	CUSR0000SAS	CPI : Services	7
120	6	CPIULFSL	CPI : All Items Less Food	7
121	6	CUSR0000SA0L2	CPI : All items less shelter	7
122	6	CUSR0000SA0L5	CPI : All items less medical care	7
123	6	PCEPI	Personal Cons. Expend.: Chain Index	7
124	6	DDURRG3M086SBEA	Personal Cons. Exp: Durable goods	7
125	6	DNDGRG3M086SBEA	Personal Cons. Exp: Nondurable goods	7
126	6	DSERRG3M086SBEA	Personal Cons. Exp: Services	7
127	6	CES0600000008	Avg Hourly Earnings : Goods-Producing	2
128	6	CES2000000008	Avg Hourly Earnings : Construction	2
129	6	CES3000000008	Avg Hourly Earnings : Manufacturing	2
130	2	UMCSENTx	Consumer Sentiment Index	4
131	6	MZMSL	MZM Money Stock	5
132	6	DTCOLNVHFNM	Consumer Motor Vehicle Loans Outstanding	5
133	6	DTCTHFNM	Total Consumer Loans and Leases Outstanding	5
134	6	INVEST	Securities in Bank Credit at All Commercial Banks	5
135	1	VIXCLSx	VIX	8

8.2 Textual Data

Table 8.2: Selected Textual Variables

Term	Correlation (PCEPI)
Changes	-0.2498790
Economic.ideology	-0.2426341
Savings...loans	0.2457363
Connecticut	0.2272212
Steel	0.2288302
Bond.yields	-0.2123466
Activists	0.2346383
Nonperforming.loans	0.2003337
Revised.estimate	0.2759805
Economic.growth	-0.2624416
Chemicals.paper	0.2292829
European.sovereign.debt	-0.2915158
Programs.initiatives	-0.2383538
Drexel	0.2840041
Treasury.bonds	0.2589611
Challenges	-0.2332520
People.familiar	-0.2455434
Publishing	0.2091472

Continued on next page

Table 8.2 – continued from previous page

Term	Correlation (PCEPI)
Financial.crisis	-0.2987179
Aerospace.defense	0.3093890
Recession	-0.3674969
Computers	0.2069321
Small.changes	0.2783810
Small.possibility	-0.2633905
Agreement.reached	0.2325187
Canada.South.Africa	0.2150613
Investment.banking	-0.2350150
Spring.summer	-0.2337073
Mid.level.executives	0.2452292
Agriculture	0.2062711
Takeovers	0.2518234
Southeast.Asia	-0.2101884
Corrections.amplifications	-0.2599948
Currencies.metals	0.2730645
Germany	0.2223704
Rental.properties	-0.2268973
Committees	0.2255680
Subsidiaries	0.2250796
Terrorism	-0.2488386
Commodities	0.2880102
Convertible.preferred	0.2611357
Currencies.metals	0.2730645
Germany	0.2223704
Rental.properties	-0.2268973
Committees	0.2255680
Subsidiaries	0.2250796
Terrorism	-0.2488386
Commodities	0.2880102
Convertible.preferred	0.2611357
Macroeconomic.data	0.2637487
Reagan	0.2261628
Trading.activity	0.2390607
National.security	-0.2010512
Continued on next page	

Table 8.2 – continued from previous page

Term	Correlation (PCEPI)
Private.equity.hedge.funds	-0.2159399
Size	-0.2433178
Retail	-0.2150240
Long.short.term	-0.2536075
Lawsuits	0.2210968
Revenue.growth	-0.2166038

8.3 Estimation Results

Table 8.3: Full Estimation RMSE

	t+1	t+2	t+3	t+4	t+5	t+6	t+7	t+8	t+9	t+10	t+11	t+12
AR	0.10496	0.15276	0.17114	0.16763	0.16176	0.15622	0.16813	0.17674	0.15042	0.13978	0.13942	0.17611
CSR	0.11165	0.16668	0.18857	0.17201	0.17073	0.17475	0.18574	0.19366	0.17213	0.16080	0.16244	0.19521
Factor	0.09626	0.13270	0.16762	0.19683	0.17409	0.18144	0.18416	0.20107	0.19139	0.16643	0.14519	0.15387
Bagging	0.24622	0.39889	0.34992	0.42035	0.34872	0.31635	0.40580	0.74155	1.01308	0.72055	0.34619	0.35077
Lasso	0.10587	0.15599	0.17964	0.17456	0.16861	0.16897	0.17325	0.19743	0.17164	0.16165	0.16026	0.18567
RF	0.13379	0.18859	0.20604	0.21063	0.19711	0.18369	0.19224	0.19588	0.18587	0.18407	0.19370	0.21960

Table 8.4: Full Estimation MAE

	t+1	t+2	t+3	t+4	t+5	t+6	t+7	t+8	t+9	t+10	t+11	t+12
AR	0.07430	0.10769	0.12018	0.11595	0.11430	0.11199	0.11854	0.12558	0.10689	0.10179	0.10137	0.12639
CSR	0.08007	0.11393	0.12875	0.11660	0.11799	0.12331	0.13102	0.13532	0.12154	0.11468	0.11675	0.14139
Factor	0.05507	0.08493	0.11799	0.12969	0.12446	0.12721	0.13120	0.14306	0.13023	0.12045	0.10223	0.11343
Bagging	0.13428	0.20296	0.19943	0.21433	0.20770	0.18820	0.21150	0.27068	0.28876	0.26285	0.21958	0.23740
Lasso	0.07718	0.11350	0.12716	0.11949	0.11719	0.11857	0.12477	0.13871	0.11852	0.11429	0.11521	0.13440
RF	0.09271	0.13208	0.14042	0.14027	0.13147	0.12550	0.13577	0.13374	0.12703	0.12096	0.12849	0.15076

Table 8.5: Factor + LASSO Combination Methods Performance, t+1 Horizon

	ME	**RMSE**	**MAE**	**MPE**	**MAPE**
WA	-0.00644	0.20904	0.14133	-0.97653	297.85570
TA	-0.00644	0.20904	0.14133	-0.97653	297.85570
SA	-0.00644	0.20904	0.14133	-0.97653	297.85570
rollOLS	0.01080	0.17418	0.12547	-57.44807	206.63250
rollLAD	-0.00979	0.16898	0.12370	-80.24449	236.20190
OLS	-0.00000	0.18860	0.13212	48.70956	278.21330
NG	-0.00729	0.18887	0.13145	47.50011	289.94860
MED	-0.00644	0.20904	0.14133	-0.97653	297.85570
LAD	-0.01404	0.18914	0.13114	32.86808	293.74060
InvW	-0.00660	0.20257	0.13860	7.86769	293.84820
Eig4	0.00000	0.19305	0.13473	33.19326	277.73810
Eig3	-0.00691	0.19318	0.13420	25.55612	287.87190
Eig2	0.00000	0.21121	0.14220	3.29767	289.74400
Eig1	-0.00639	0.21130	0.14230	-3.76748	299.32080
CLS	-0.00691	0.19318	0.13420	25.55612	287.87190
BG	-0.00653	0.20521	0.13969	4.07917	295.26220
Auto	-0.00000	0.18860	0.13212	48.70956	278.21330

Table 8.6: Factor + RF Combination Methods Performance, t+1 Horizon

	ME	**RMSE**	**MAE**	**MPE**	**MAPE**
WA	-0.00577	0.21082	0.14134	-46.47720	319.94010
TA	-0.00577	0.21082	0.14134	-46.47720	319.94010
SA	-0.00577	0.21082	0.14134	-46.47720	319.94010
rollOLS	0.01295	0.20805	0.14131	-76.44009	247.10400
rollLAD	-0.00787	0.18703	0.13234	-120.66310	287.38940
OLS	-0.00000	0.20348	0.13958	-54.57574	309.90930
NG	-0.00564	0.20670	0.13850	-57.92172	328.23940
MED	-0.00577	0.21082	0.14134	-46.47720	319.94010
LAD	-0.01453	0.20719	0.13824	-71.09694	344.11280
InvW	-0.00570	0.20753	0.13928	-52.79987	323.73650
Eig4	0.00000	0.20842	0.13996	-59.30095	329.10360
Eig3	-0.00556	0.20849	0.13947	-65.44522	336.84300
Eig2	0.00000	0.21177	0.14196	-38.72638	310.72380
Eig1	-0.00578	0.21185	0.14187	-45.12100	319.18480
CLS	-0.00564	0.20670	0.13850	-57.92172	328.23940
BG	-0.00574	0.20940	0.14061	-48.67902	321.21460
Auto	-0.00000	0.20348	0.13958	-54.57574	309.90930

Table 8.7: Factor + LASSO Combination Model Errors

	ME	**RMSE**	**MAE**	**MPE**	**MAPE**
t+1	-0.009785549	0.16898	0.12370	-80.24449	236.2019
t+2	-0.001792205	0.20099	0.14677	-168.8634	365.0719
t+3	-0.002907416	0.20129	0.14560	-165.4643	366.1160
t+4	-0.009641443	0.20279	0.14586	-146.224163	354.6082
t+5	-0.004511879	0.20379	0.14898	-172.64072	382.8772
t+6	-0.009711331	0.21584	0.15729	-181.0373	400.0626
t+7	0.001169535	0.21486	0.15370	-141.2866	354.1472
t+8	0.006465232	0.22364	0.15939	-157.5009	370.2831
t+9	0.01009231	0.21251	0.15336	-122.1995	327.1258
t+10	0.02186981	0.21604	0.15324	-144.52005	339.3081
t+11	0.02638063	0.21292	0.15162	-97.47263	286.3800
t+12	0.02235237	0.21891	0.15700	-108.37928	309.3059
Mean	0.00417	0.20771	0.14971	-140.48611	340.95733