Data Science Final

Caden Finley

May 10, 2023

1 Introduction

Inflation is an important tool can be manipulated by the federal reserve to control the United States Economy and by extension the money supply. Inflation is an important economic indicator because it affects the purchasing power of consumers and the profitability of businesses. Depending on the level, it can be beneficial or detrimental to growth in the United States. When interest rates are manipulated at moderate levels, effects of inflation are seen like the strength of the dollar going up and the cost of goods going down. On the other hand, with high or volatile inflation, it can destabilize an economy and erode savings and investments. The time value of money is something most people look when trying to understand inflation. That concept being what will your money today be worth in the future based on economic conditions. The paper here will involve using various financial packages in R to look at Macro-level data about inflation, the Consumer Price Index, and Personal Consumption Expenditure Index. These packages will look to visualize financial market data to see trends involving inflation. Data from the St. Louis Federal Reserve will be scrapped by using an API to get accurate data involving inflation. Once the data is acquired, it will be tested with several functions to see if inflation has a correlation with certain variables. One way of doing this is with understanding conditional distribution and using various R functions. The paper looks to understand the effects of certain variables in a linear regression using various weights. These will contain tables showing the data I looked at as well as charts to visualize the change in the financial markets. With a combination of machine learning and traditional financial analysis, the paper serves as an alternate way of predicting inflation.

2 Literature Review

There have been various approaches by Economists to use Machine Learning in order to better predict inflation. A paper by Emanuel Kohlscheen titled ”What does machine learning say about the drivers of inflation?” discusses some interesting ideas regarding this topic. For example, in the paper it goes over a random forest or tree model that looks to make more sense of inflation drivers that are apparent in various economies. With inflation it can be difficult to understand what should and should not be considered, but this model shown in this paper isolates ”out of sample results”, a similar idea to what is looked at in this paper. Inflation can have a potential bias in it with all of the different goods that are included in the Consumer Price Index. A result of looking out of sample for results shows in the paper that ”expectations” and ”past inflation” were the top variables for helping better predict inflation (Kohlscheen, 2021). When looking at the expectation variable for inflation prediction, the paper discusses a method of analyzing the relationship between inflation expectations and actual inflation. The method involves looking at predicted inflation rates based on different levels of expected inflation, while holding all other variables constant at their mean values. This analysis can provide insight into the effects of inflation expectations on actual inflation, with the results interpreted similarly to regression coefficients in econometrics. However, one of the weaknesses of this paper is that it does not offer a structured econometric theory with this idea. Expected inflation is based off of what the Federal Reserve estimates it to be, thus you have to weight for that number to come out first. Essentially, the results of the paper show that a out of sample root mean square errors (RMSE) beat the in-sample benchmark econometric models. Taking an alternative approach to understand how inflation can really affect the market is a key that is discussed in my own analysis as well.

Another paper titled "Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods" by **?** discusses how alternative models can be used to more accurately predict inflation than traditional models. One of the prevalent models in this paper is the auto regressive integrated moving average "ARIMA" model. This model looks to account for seasonality adjustments and complex time series data similar to what you would look for to model inflation. The author finds that machine learning methods generally outperform traditional econometric models in terms of forecasting accuracy, particularly when using a larger number of variables and incorporating data from different sources. The paper also examines the importance of feature selection and regularization techniques to improve the performance of machine learning models. From this paper, machine learning could be a possible better approach for predicting inflation in the future. Overall the findings from these papers showed interesting results. **?** found that machine learning methods such as Random Forest and Boosting can accurately forecast inflation in Russia. Medeiros 2019 also found that machine learning methods, particularly the Random Forest model, outperform traditional models in forecasting U.S. inflation. Kohlscheen 2022 used a machine learning technique to predict inflation across 20 advanced countries and found that inflation expectations play an important role in inflation outcomes. Overall, these papers suggest that machine learning methods can improve the accuracy of inflation forecasting.

3 Data

The majority of the data used for this analysis comes from the St. Louis Federal Reserve and the Bureau of Labor Statistics regarding accurate information on inflation. When breaking down inflation from the Bureau of Labor Statistics, you see the consumer price index is a combination of good and services tracked throughout the years in the US. Some of these goods include: electricity, clothes, food, gas, etc. For this analysis I looked at variables that would have relevancy to middle-class consumers. This include goods in the CPI index like ”electricity, shelter, and food”. Higher weights were assigned to these values in order see if the results for an estimated inflation rate would be different than what is reported in the CPI. Collecting the data required finding the different indexes and making sure the years and variables lined up. It was important to pay attention to the code that was listed at the end of subject line because this was what was going to be what was referenced in R when building the charts. Once this data was assigned to a variable, looking at the ”value” column was crucial as it was the main indicator for what the Consumer Price Index was going to read. Observing the differences across variables and averaging these values in order to make a more relevant value was the idea here. However, this process is discussed in more detail during the empirical method section. The data for this paper required some intuition with finding the different codes and cleaning it to be able to use multiple indexes at once. It was important to get the data separately from each variable to understand the differences between them and how they affect the normal CPI index readings.

4 Empirical Method

The process of changing the variables around to get a more realistic inflation rate required manipulation of data using machine learning techniques. One of these ways was to assign different weights to the variables that are more prevalent in the average consumer’s life. For example, the cost of housing or shelter as it is referred to on the St. Louis Federal Reserve’s website, was something was assigned a larger weight for the purpose of seeing how inflation responds in reaction to this variable. For this process, I used the VAR() function from the "vars" package to fit a Vector Autoregressive Model. Using the variables specified in the previous section was crucial with this model's design as it is a great model for dealing with multiple time series variables. The equation for implementing this regression model is as follows:

$$Y_t = \alpha + \beta_1 Y_t -_1 + \beta_2 Y_t -_2 + .. + \epsilon_t \tag{1}$$

The equation is interpreted as follows: where Yt represents the dependent variable at time t Yt-1 in this case represents the independent variable at time t-1, Yt-2 represents the independent variable at time t-2, and so on. The coefficients Beta1, Beta2, etc. represent the weights assigned to the lagged independent variables Yt-1, Yt-2, etc. These coefficients represent how much impact the previous values of the independent variable have on the current value of the dependent variable. This is important to understand when using the different variables in the CPI basket and assigning different weights to them. The epsilon term serves as the error term or the random variation in the dependent variable that cannot be explained by the independent variables. Overall, this equation represents a vector autoregressive model that can be used to understand how the previous values of a variable can impact its current value. This concept is very useful with inflation as expected inflation or interest rates play a big part in what the next rate could be.

5 Findings

Manipulating the basket of goods present in the Consumer Price Index can help reveal potential biases that are masked. When the return or index number of certain variables largely outweighs are more important group, it can cause the score of the CPI (inflation) to be misleading. This concept is known as Geometric Mean Formula Bias. This is when the CPI uses a geometric mean formula to calculate price changes, which tends to give more weight to goods and services that have experienced price increases, while giving less weight to goods and services that have experienced price decreases. Many times you will see essentials like food and the cost of living have dramatically increased due in part by inflation. However, these increases might not seem as large when they are brought down by other variables like gasoline, which include numerous types. The unclear inclusion of multiple types of gas results in negative monthly percentage change, which is normally a good thing, but when the groupings of what are included in gas skew the results it becomes a problem. The model that is shown in this paper tries to offset for the equal weight that is normally assigned to the consumer price index. By isolating some of the more important variables, it can lead to a more realistic figure.

6 Conclusion

Machine Learning Techniques can be Incorporated in various ways to better enhance the predictability of inflation. Whether it is with a random forest model or a vector autoregressive model, it is important to look beyond what is just reported in the consumer price index. The VAR model shown in this paper is something that can be used outside of this study as it depends on the consumer's preferences and needs. If a consumer values clothing more and will let that affect their purchasing power, then you are able to find the code and plug that into the model with an assigned weight to understand how the price of apparel going up really does affect you. This can be used for multiple items or variables depending on what is important to the user. Inflation is a powerful tool, making it essential to understand as not just as a consumer but for businesses and producers as well. Machine learning is here to stay and it is only a matter of time before more and more firms use machine learning to forecast inflation even better than this model attempts to. Knowing inflation means knowing when to pull back or scale up and operation, and the same thing goes for consumer spending an saving. Overall, there seems to be many machine learning techniques that can better predict inflation, again showing why it is crucial to know about this topic.

7 References

Medeiros, M.C., Vasconcelos, G.F., Veiga, Á., Zilberman, E. (2019). Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods. Journal of Business Economic Statistics, 39, 98 - 119.

Pfajfar, D., Žakelj, B. (2014). Experimental evidence on inflation expectation formation. Journal of Economic Dynamics and Control, 44, 147-168.

Kohlscheen, E. (2022). What does machine learning say about the drivers of inflation? SSRN Electronic Journal.

Ülke, V., Sahin, A., Subasi, A. (2016). A comparison of time series and machine learning models for inflation forecasting: empirical evidence from the USA. Neural Computing and Applications, 30, 1519 - 1527.

Lin, C., Wang, C. (2013). Forecasting China's inflation in a data-rich environment. Applied Economics, 45, 3049 - 3057.

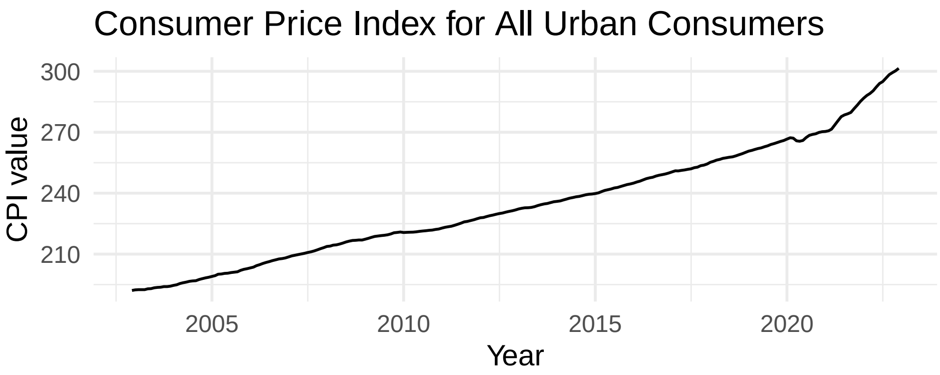

Figure 1: CPIAUC

8 Charts and Figures

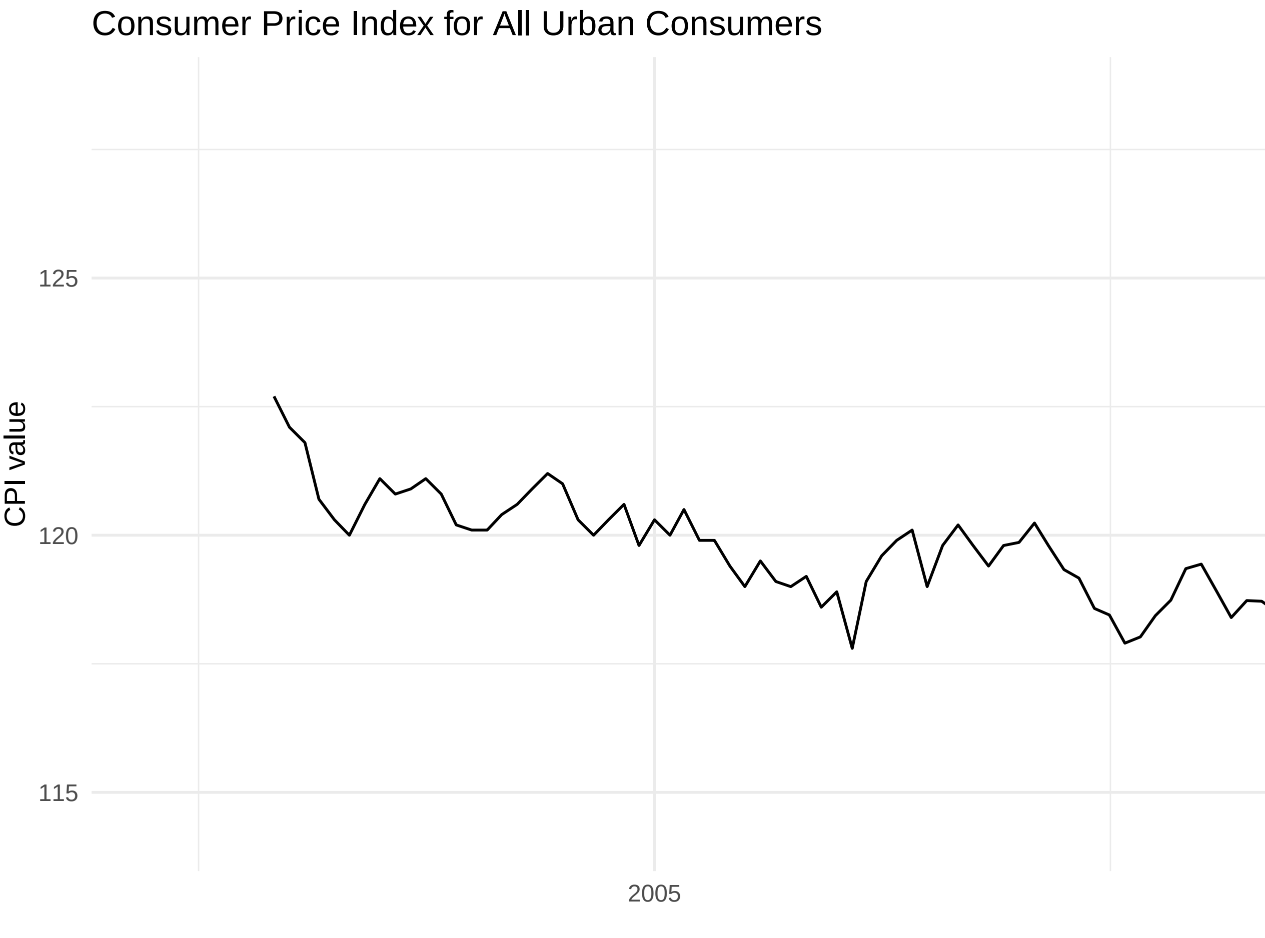

Figure 2: CPIAPPSL