MAY 6, 2024

PREDICTING INFLATION IN THE UNITED STATES USING S&P 500 PRICE

CS210 PROJECT STEP II

NUH AL-SHARAFI
32702

Introduction

In a world that is increasingly sensitive to unpredictable change, there's a growing effort to predict as much of the future as possible. Even more so when it comes to money and its derivatives because people's livelihoods depend on it. One of the great enemies to people's livelihoods is inflation. It is the general increase in the price of goods and services in an economy which diminishes people's purchasing power.

This paper is an effort to come up with a reliable predictor for inflation. In order to do so, two datasets will be analyzed. The first one is a stock market dataset (*2019-2024 US Stock Market Data*, 2024) and the second is an inflation dataset (*Bureau of Labor Statistics Data*, n.d.). Both datasets were obtained from Kaggle.com. The stock market dataset will provide a selection of predictors against which inflation will be tested to determine the best predictor.

Data Exploration

Upon opening the stock market dataset, some wrangling was in order. The data types were corrected such that all of them were of type float (which is what they should be). After that the ratio of missing values for each column was observed, but no treatment was necessary. Now that the dataset was ready to be explored, it was time to choose a metric based on which a stock/index will be chosen to be the predictor for inflation. That metric is volatility which is "a statistical measure of the dispersion of returns for a given security or market index." (Hayes, 2024) This choice of metric will allow us to choose a predictor with a very stable price, and that will go a long way in reliably predicting inflation. Upon observation, Figure 1 was obtained. The top 3 contenders are Gold, Berkshire Index, and the S&P 500 index. The S&P 500 stands for the Standard and Poor 500 and it's a "market-capitalization-weighted index of 500 leading publicly traded companies in the U.S." (Kenton, 2023) The S&P 500 provides a wide coverage of companies from different sectors in the United States which makes it a prime candidate for representing movements in the market. In addition to that, it is the 3rd least volatile predictor.

Since the predictor has been chosen to be the S&P_500 index, it is now time to merge the inflation dataset with the S&P 500 price using date as the index. However, there was an issue with dates since inflation data is 1 entry per month and S&P 500 price is 1 entry per day. The remedy was plugging the average price for the S&P 500 for each month corresponding to the inflation dataset. After that, two more features were added to the new dataset. Namely, the percentage change in S&P 500 price and the disparity between the percentage change of S&P 500 price and inflation percentage change. Figure 2 shows 4 boxplots for the 4 variables in the new dataset. The key takeaways are that the S&P 500 percentage change only has a few outlier values, which is good, as is the case for the disparity measure. This indicates that the data behaves nicely and is fit for prediction.

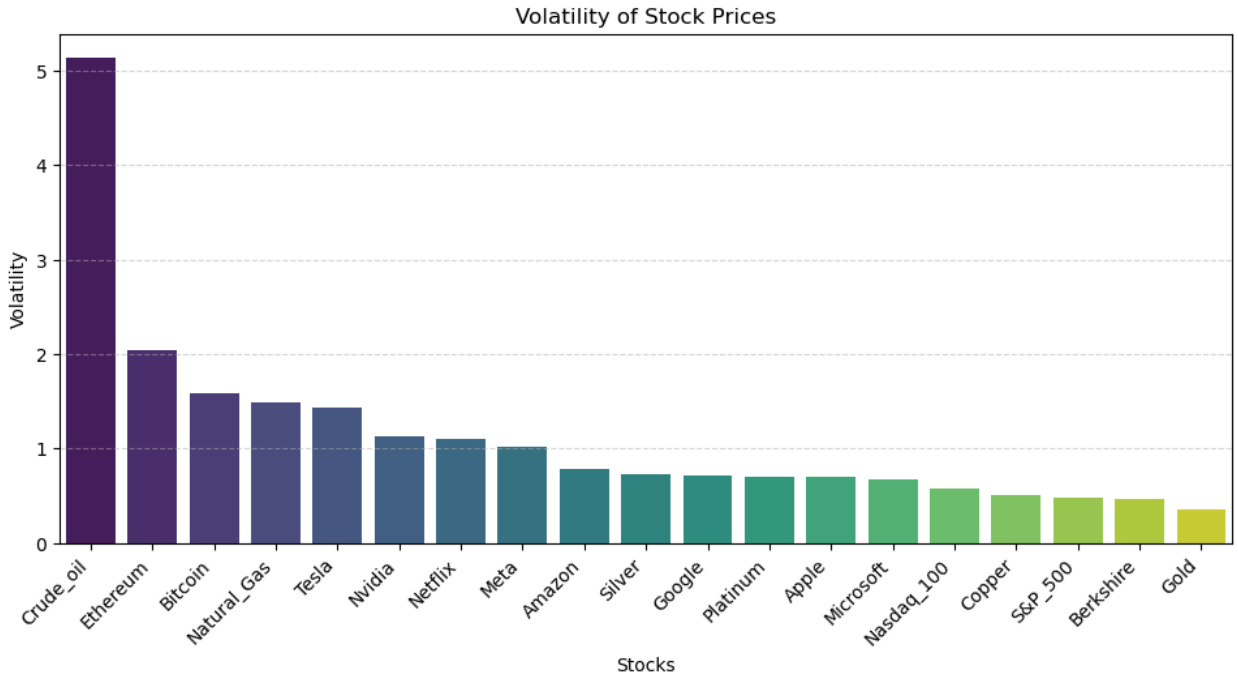

Figure 1 Bar chart of stock/index price volatility (lower is better).

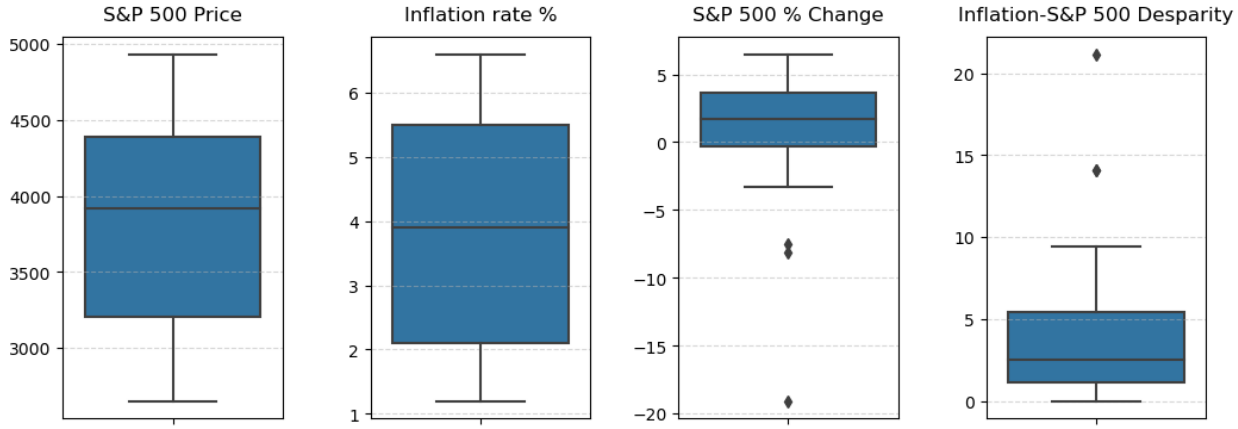

Figure 2 Boxplots of the variables in the new dataset

Data Analysis

Then, after concluding that the data is fit for prediction, it is only fitting to have a look at the histograms and respective densities for the S&P 500 and inflation rate. Figures 3 and 4 show the S&P 500 and inflation rate histogram/density respectively.

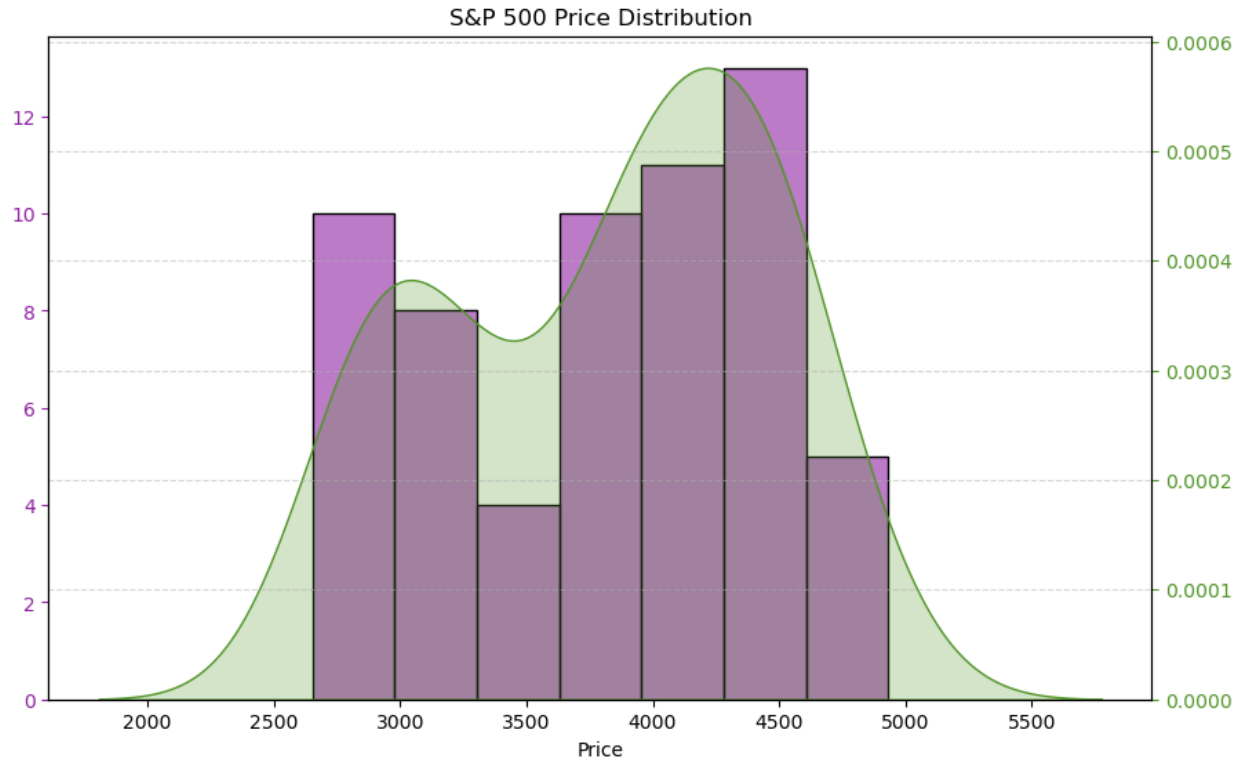

Figure 3 S&P 500 distribution and density.

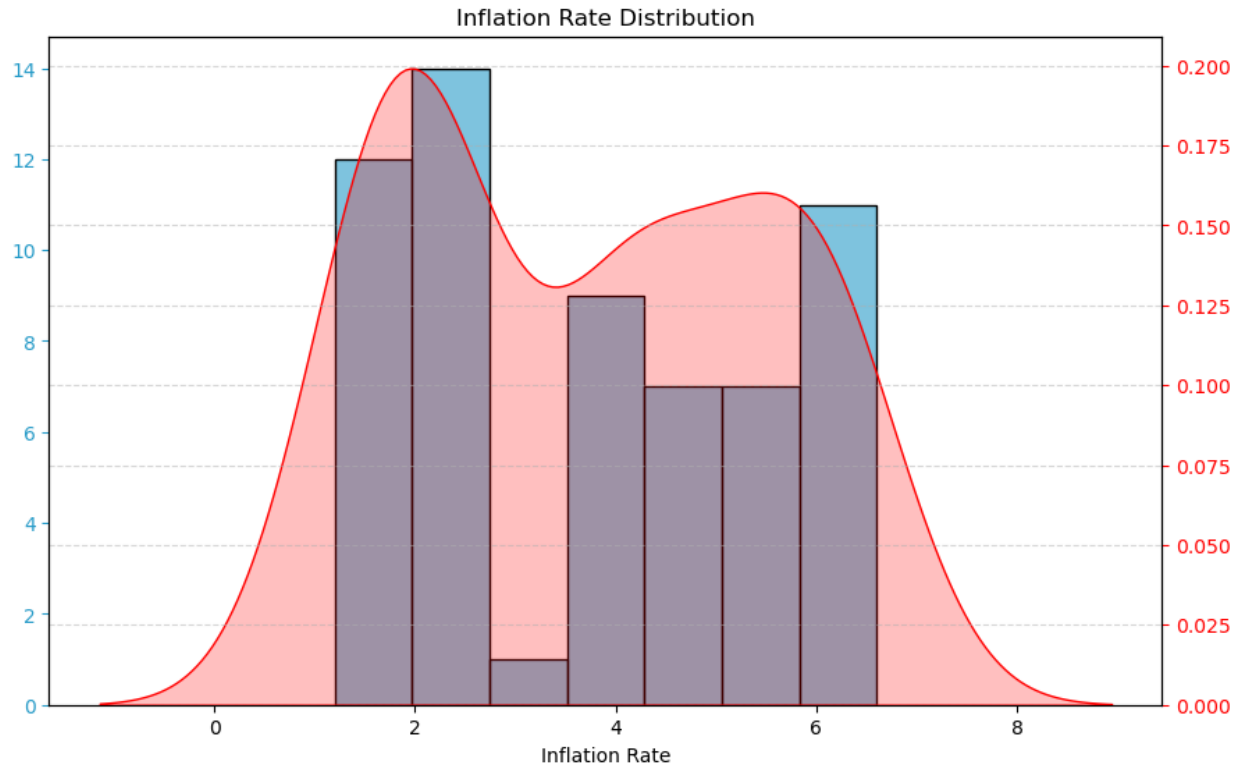

Figure 4 Inflation rate distribution and density.

It is very clear that the densities are similar. So, the next step is looking at a scatter plot of the 2 variables. Figure 5 showcases the scatterplot. Furthermore, similarity can be observed when looking at the changes over time via a time-series. Figure 6 shows a time series showcasing both S&P 500 and inflation changes over time.

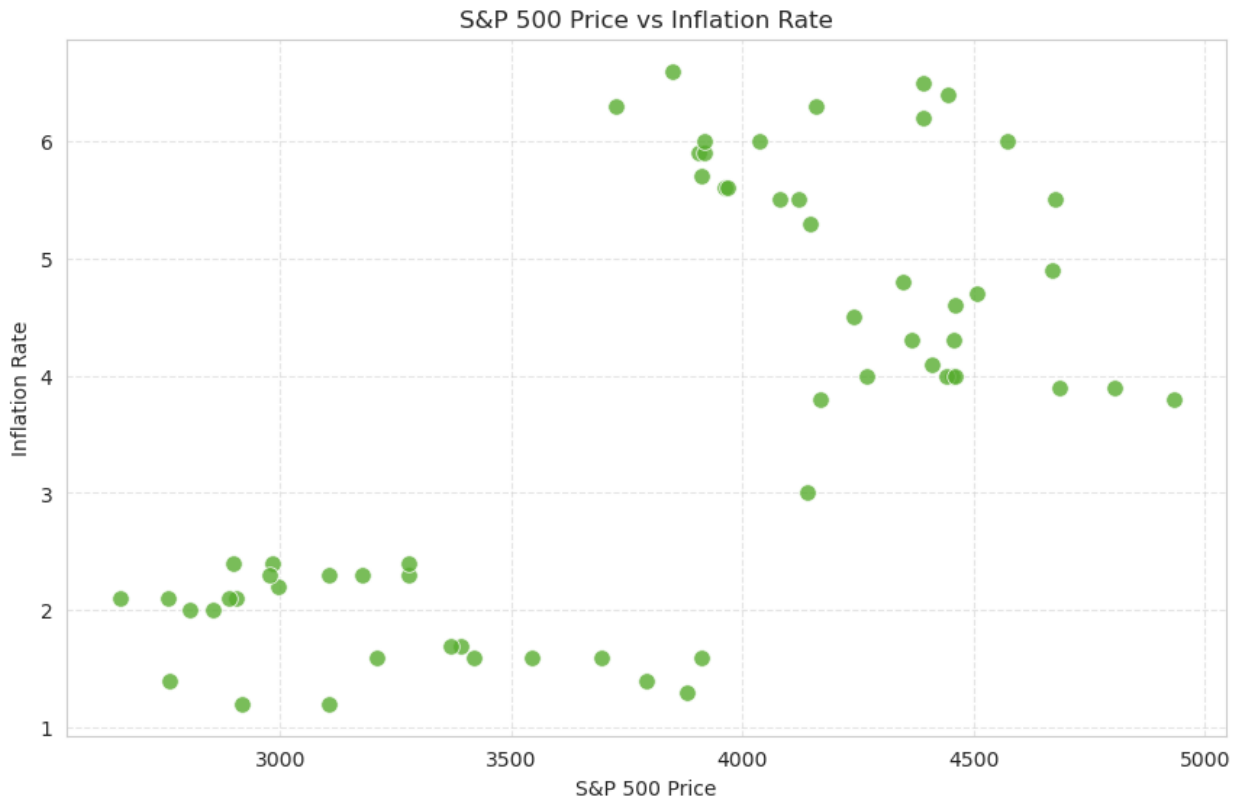

Figure 5 S&P 500 Price vs. Inflation rate scatterplot.

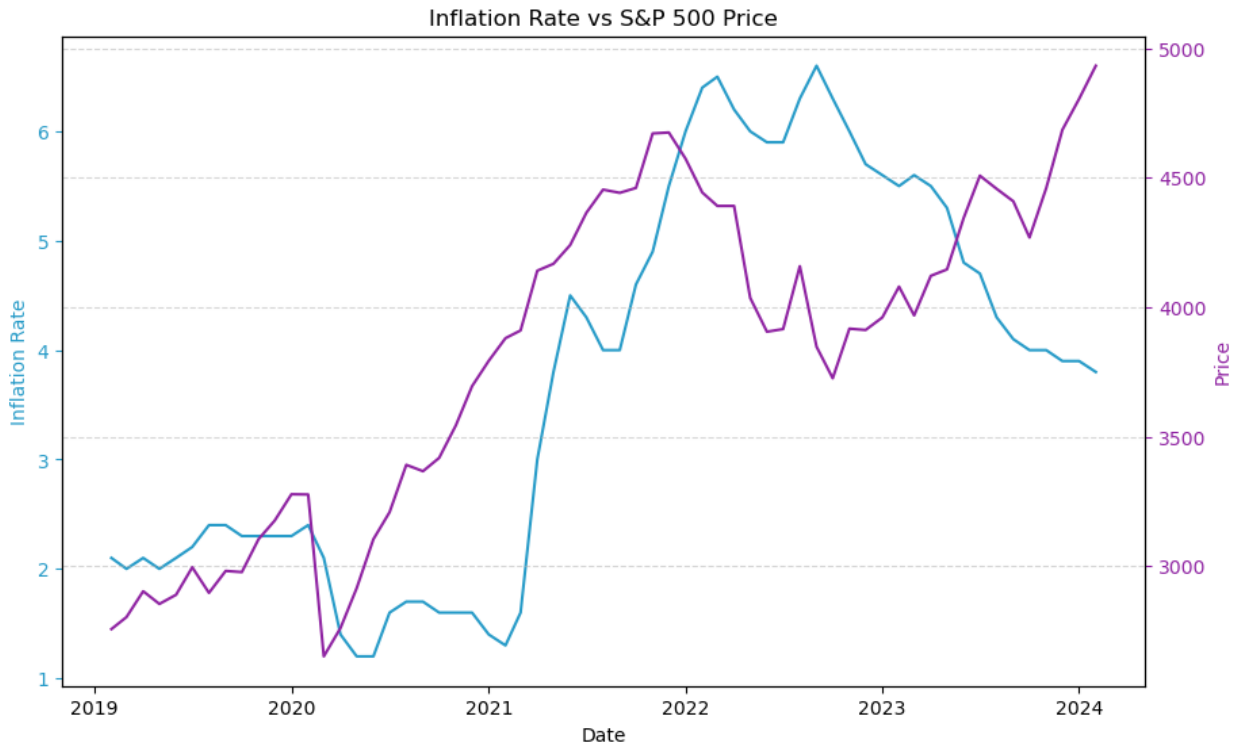

Figure 6 time-series of inflation rate and S&P 500 price.

In addition to the clear correlation, it seems to be the case that the inflation rate is copying the S&P 500 movements but with a delay. So, the question becomes how much delay? The answer to this question will tell how far ahead the S&P 500 price movements today will be reflected in the inflation rate.

To answer this question, the S&P 500 price was shifted almost a year ahead, and for each month, the correlation was recorded, the result is shown in the following table:

Table 1 correlation values (S&P 500 vs. Inflation) for each month shift.

Correlation	P-value	shift
0.675436	2.355623e-09	0
0.725902	5.295923e-11	1
0.763252	2.096243e-12	2
0.794029	1.036548e-13	3
0.817315	8.806770e-15	4
0.839533	6.295722e-16	5
0.862595	2.597116e-17	6
0.885620	5.906922e-19	7
0.908078	6.435731e-21	8
0.915386	2.135743e-21	9
0.907051	4.853067e-20	10
0.882724	2.326572e-17	11

Table 1 shows that a 9-month shift results in the highest correlation (0.915) between the S&P 500 and inflation. Therefore, the answer to the question is that the S&P 500 movements today will most likely be reflected by inflation in 9 months.

For the sake of looking at the correlation, figures 7 and 8 show a new scatterplot and a new time-series respectively.

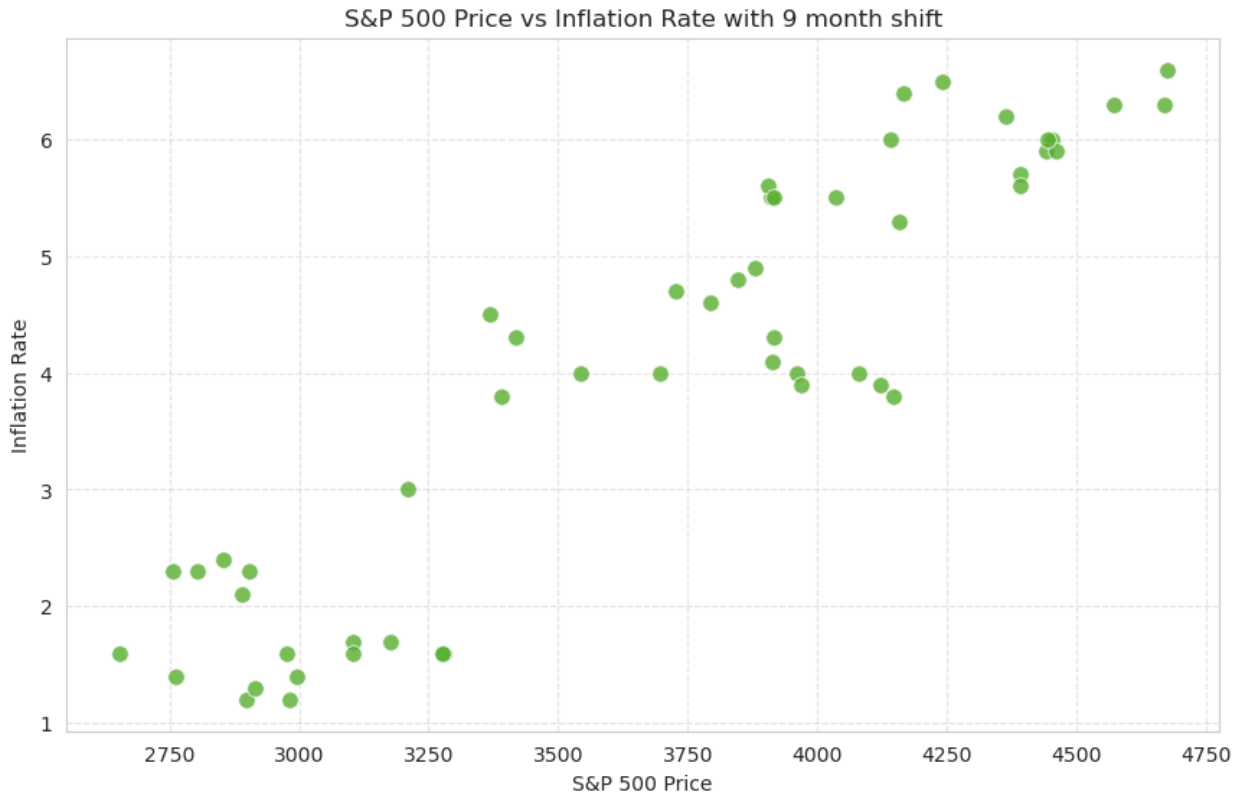

Figure 7 Scatterplot with 9-month shift

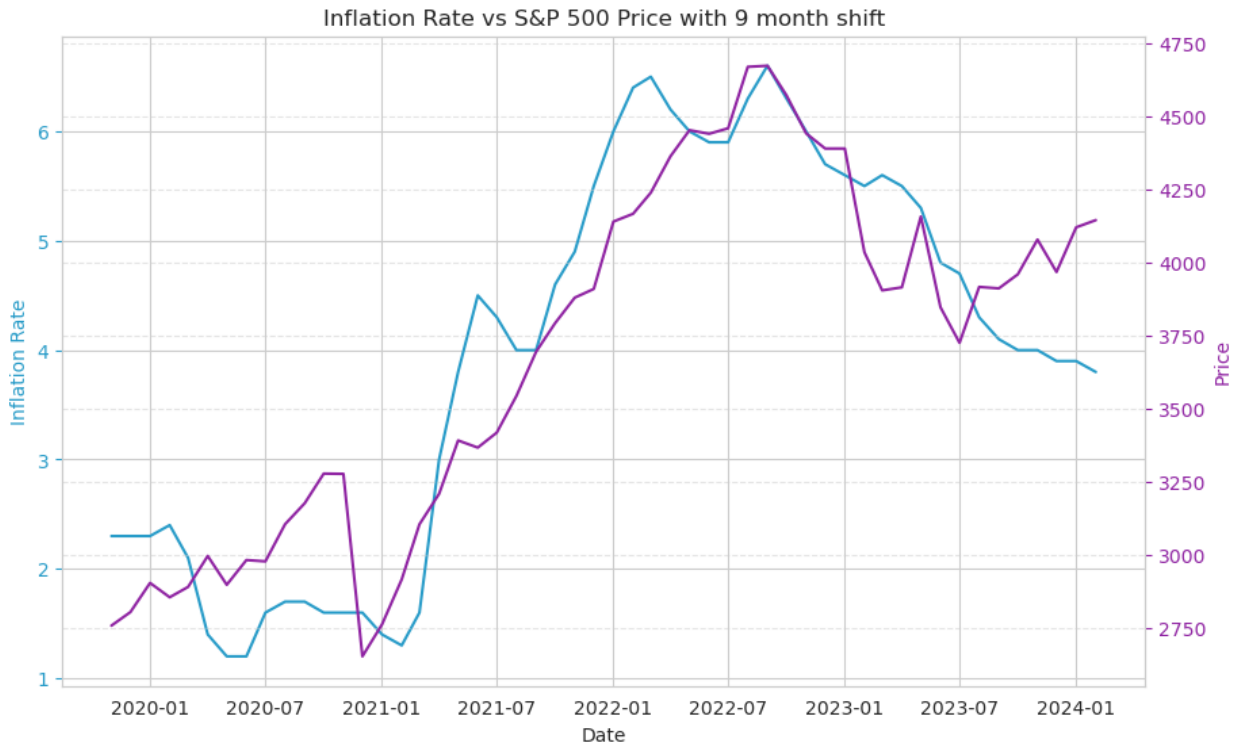

Figure 8 Time-series with 9-month shift

Hypothesis Testing

H0: The S&P 500 **is not** a reliable predictor of inflation in the United States

H1: The S&P 500 **is** a reliable predictor of inflation in the United States

Alpha: 0.1

After all the exploration, it's time to make a decision regarding the hypothesis. The test statistic that will be used to test the hypothesis is the Pearson-Correlation. For alpha choice 0.1, if the Pearson Correlation gives a p-value for the correlation less than 0.1, that would mean that we're in the critical region, therefore the Null hypothesis will be rejected in favor of the Alternate hypothesis. Accepting the alternate hypothesis means that there exists a relationship between the S&P 500 price and inflation that is more than pure chance.

Upon carrying out the computations that were presented in Table 1, the p-values and correlations for each shift can be observed. Then, as a result of choosing a 9-month shift, we look at the 9-month shift P-value (2.135743e-21) to determine the result of the hypothesis test. Since the P-value is much smaller than the pre-determined alpha, the Null hypothesis is rejected for the alternative.

This means that the S&P 500 and inflation rate 9 months later exhibit a tight and highly correlated relationship, which further implies that we can predict inflation using the Price of the S&P 500.

Linear Regression Model

So, upon proving that there's a relationship between the S&P 500 and inflation, how can this relationship be exploited?

One obvious way would be to construct a linear regression model that could be used to predict inflation using the S&P 500. For this model, 20% of the data was reserved for testing while the remaining 80% was used for training the model. Then, upon training, the R^2 value was used to evaluate the model's performance and that came out to **0.84**.

The model came up with the following equation:

Inflation_rate = (S&P_500P - 2450.2) / 310.68

Beyond that, Figure 9 shows a scatterplot of Inflation and S&P 500 price with the model's equation on top.

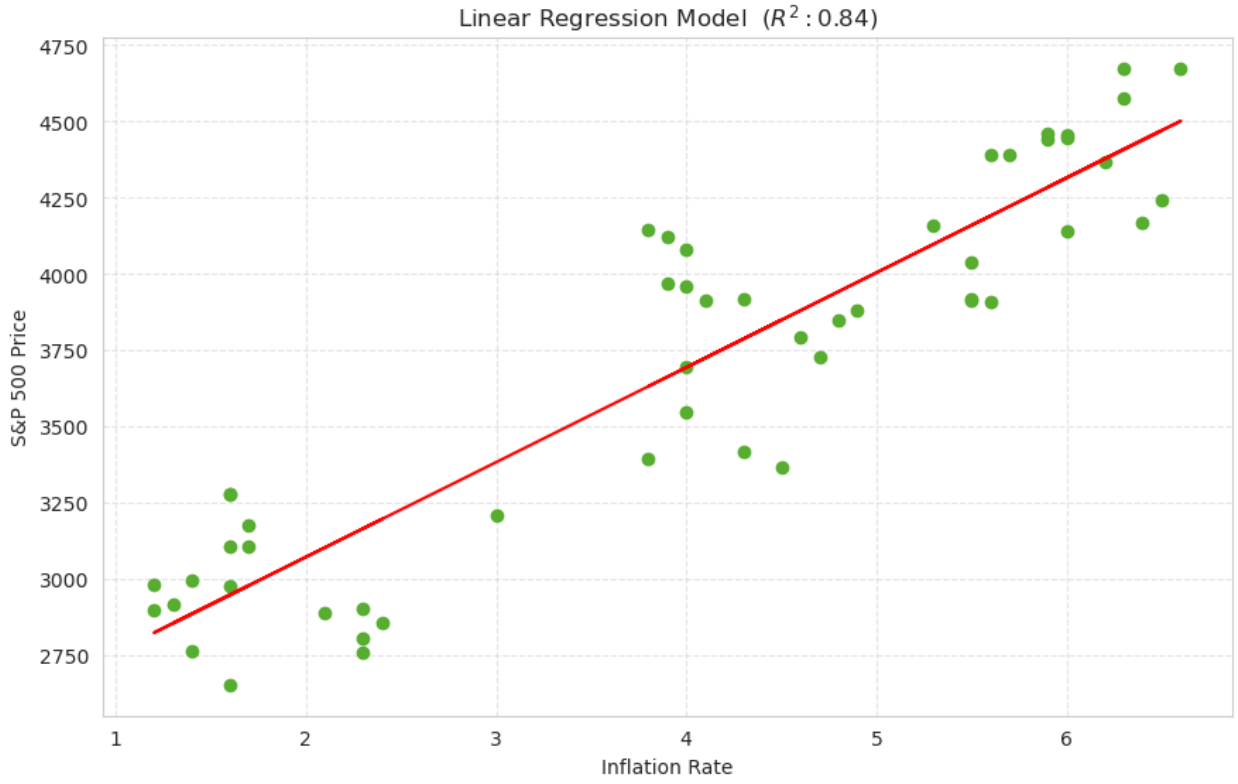

Figure 9 Linear Regression on ScatterPlot

Conclusion

To conclude, while predicting change can be beneficial, it is important to consider that this will put us into more rigid and comfortable positions which will make the aftermath of an unpredictable

event or a black swan even more catastrophic. This is not to say prediction should be banished and data scientists are witches, but it casts a light on a dangerous drawback that can be avoided. In fact, being able to predict inflation 9 months ahead of time as was discussed above can allow legislators to design monetary policy that will curb the effects of inflation, therefore benefiting citizens. On the other hand, predicting inflation can allow investors and capitalists to time their loan applications in such a way that minimizes the interest. Finally, the U.S. is not the only country in the world and this paper hopes to inspire such models for other countries.

References:

2019-2024 US stock market data. (2024, February 4). Kaggle. https://www.kaggle.com/datasets/saketk511/2019-2024-us-stock-market-data?resource=download

Bureau of Labor Statistics Data. (n.d.). Bureau of Labor Statistics. https://data.bls.gov/timeseries/CUUR0000SA0L1E?output_view=pct_12mths

Hayes, A. (2024, April 4). *Volatility: Meaning In Finance and How it Works with Stocks*. Investopedia. https://www.investopedia.com/terms/v/volatility.asp

Kenton, W. (2023, September 22). *S&P 500 Index: What It's for and Why It's Important in Investing*. Investopedia. https://www.investopedia.com/terms/s/sp500.asp