HIGH SCHOOL EDITION
Journal of Student Research
Volume 13 Issue 3 (2024)

Use of Economic Indicators in Correlation and Prediction Analysis of S&P 500

Nitin Mali

ABSTRACT

This paper covers the correlations of the macroeconomic indicators of GDP, unemployment rates, CPI (Consumer Price Index), interest rates, expected inflation, and home. What was seen was correlations that were both strong and weak, and that correlation never means that they can predict the index, as while using these to predict for the index what was found was that the p-values were exceedingly high, and could thus not justify the indicators in the forecasting of the S&P 500 price, and the model used to predict the index through the indicators was an ARIMA Timeseries forecasting model. The S&P 500 also had positive correlations and negative correlations to indicators that had certain trends occurring, such as Pearson coefficient was positive, but the response and Timeseries model, along with the logistic coefficient was negative, thus showing that correlation does not work as an actual way of showing responses or forecasting for the S&P 500. With Timeseries as well, the Mean Squared Error (MSE) was far too high to show the S&P 500 could be predicted by the indicators. Thus, we can conclude the null hypothesis of the paper, which was that the S&P 500 could not be forecasted or linked to the economic indicators used was proven correct via the high p-values.

Introduction

The S&P 500 is an index of the 500 largest US Corporations worth billions, and according to (Bloomberg, 2024), the S&P covers 80% of all of the market capitalization of the entire stock market, and this index thus has a value in the trillions, above $40 Trillion. This index is also an economic indicator as well, as according to (Duggan, 2023) it is the largest indicator of stocks that investors and governments use to monitor the overall stock performance in the markets and it often is seen as the make or break of the market and indicate it. Thus, what can be inferred is that it is extremely important for economics and will indicate business stock performance and general markets. S&P 500 performance is the main variable that plays the role as that detector. Furthermore, it also serves the purpose of being an income stream due to its average of (Butler, 2020) states to be 10.43% of an average return annually, and it can serve as a stream of income due to that result by each year and what can be seen is that there is often a desire to forecast by investors as that performance can be used either as an asset and as an indicator of the economy and of stocks as well. However, what should be known is that it is extremely problematic and tedious, along with unlikely to predict the S&P 500 because there are certain variables that cannot be quantified of which can predict the S&P 500 and the index does fluctuate to great degree. However, in recent times there have been forecasting models of which can be used to predict the S&P 500 Index using machine learning. According to (Coursera, 2023) machine learning is a subfield of artificial intelligence which can be used to do various tasks like forecasting or analyzing data. However, amidst these capabilities one idea has been developed. Which is to forecast stocks through Machine Learning. As (Rodriguez et al., 2024) has found that decision trees have an 80% accuracy rate while predicting the S&P 500, while rule based classifiers even closer with 88%. Furthermore, for stocks and plotting a specific type of machine learning software exists, of which is called Timeseries. Timeseries is a software of which uses trends and rules and certain thresholds and trends in order to forecast any variable and it also takes in other factors such as correlations and uses p-values through the Dickey Fuller tests as well to test accuracy. And with this model another factor comes in to play of which is the Timeseries using other variables of which can be used to forecast the index correlations

and coefficients as best shown through the Dickey Fuller Test of which uses regression and time series on its lagged level and coefficient form lagged route and distance from 0 of which can be seen through non-stationary shows no unit roots and stationary often shows variance and non-stationary doesn't for which the index is stationary. Furthermore, this paper will use this to predict with other indicators as these are economic indicators that often thrive in the same environment as the S&P 500. As these indicators included will be vast and he used to predict the S&P 500 through timeseries. Correlation and p-value and coefficient tests will evaluate how these factors can predict the index and the timeseries will be analyzedanalyze as well to see how these indicators perform in the prediction of the S&P 500. This study will aim to test the correlation and prediction of the S&P 500 using a set of 5 economic indicators.

Methodology

This methodology will detail the indictors used and rationale, literature collection methods, literature tests to determine scientific validity, statistical models for prediction and correlation, limitations, and hypothesis and null hypothesis.

Indicators Used

For this experiment, the macroeconomic indicators used will include interest rates, expected inflation rates, Gross Domestic Product (GDP), unemployment rates, Consumer Price Index (CPI), and home sales. Inflation is being used to possibly predict the S&P 500 as according to (Roncaglia de Carvalho et al., 2017), economic development often correlates positively to inflation is often inhibited via inflation due to higher costs and thus reduced demands. Furthermore, according to (Asta, 2023) inflation trends can be used to forecast the S&P 500 because interest rates and inflation can be used through cyclical-based forecasting to predict via scholastic trends. Thus, this presents interest rates and inflation via cyclical components being able to be able to forecast the index. Furthermore, through this method, CPI works as well as a forecast model due to it being similar to inflation. Additionally, unemployment works as well due to (Bilello, 2018) stating that unemployment and the S&P 500 have inverse reactions as S&P 500 on average has a 5.6% return on unemployment years and 12.7% on average showing how unemployment is a factor which could possibly be used. Next off is home sales and there is little literature supporting these two however, according to (Bhutta & Ringo, 2020) interest rates influence home buying and rate cuts led to a 14% increase in sales, thus showing the common variable of interest rates of which both cause the S&P 500 index to become higher once low due to Tobin-Q model and same applies to homes and thus this variable allows for a correlation for the two.

Literature Collection and Tests

This paper will collect literature from academic journals and universities. Additional credible sources like Investopedia used. Literature will be tests by the CRAAP test. According to (Kurpiel, 2024), the CRAAP test is a reliability test to show source credibility and the "C" in it stands for "currency", the "R", which is "relevant" as in if it relates to the topic and age level and discussion are appropriate to audience, "A" is for "authority" as in publisher qualifications and organization and if the URL provides insight on the author, "A" for "accuracy" such as correctness/verification and errors grammatically and conceptually and last is "P" which is "purpose" such as what the literature is meant to do and the author's intention.

Statistical Models

The models used include Time series forecast w/automated Dickey fuller test, and logistic regression (coefficient, correlation, and p-value) of which the equation and justification follow:

Timeseries (ARMIA)

$$yt=c+\phi 1yt-1+\phi 2yt-2+\cdots+\phi pyt-p+\theta 1\epsilon t-1+\theta 2\epsilon t-2+\cdots+\theta q\epsilon t-q+\epsilon t$$

The ARMIA model, which stands for Autoregressive Moving Average, is a combination of two statistical models used to analyze and forecast time series data. The model incorporates both autoregression (AR) and moving average (MA) components to capture patterns in the data.

Mean Squared Error (MSE)

$$n1\sum i=1n(yi-y\hat{}i)\hat{}2$$

This will use difference between predicted and actual values and measure average difference between them and then use the difference squared to show a rate of means.

Logistic Regression

$$\log(1-P(y=1)P(y=1))=\beta 0+\beta 1x1+\beta 2x2+\cdots+\beta pxp$$

Logistic regression is a widely used statistical method for binary classification problems, where the outcome variable is dichotomous, representing two possible states. It models the probability of the outcome using a logistic function, transforming a linear combination of predictor variables into a probability score between 0 and 1. The model estimates the coefficients of the predictor variables through maximum likelihood estimation, providing insights into the relationship between predictors and the log-odds of the outcome.

Correlation Coefficient Equation

The correlation coefficient, denoted as

$$\sum (xi-x^-)\,(yi-y^-)/(\sqrt{\sum (xi-x^-)\,(xi-x^-)\ \sum (yi-y^-)\,(yi-y^-)})$$

r, is a statistical measure that quantifies the strength and direction of the linear relationship between two continuous variables. It ranges from -1 to 1, where values close to 1 indicate a strong positive linear relationship, values close to -1 signify a strong negative linear relationship, and values around 0 suggest no linear relationship. The coefficient is calculated as the covariance of the two variables divided by the product of their standard deviations, thus normalizing the measure to ensure it is dimensionless.

P-value: The p-value is a statistical metric used to determine the significance of results in hypothesis testing. It quantifies the probability of obtaining an observed effect, or one more extreme, assuming the null hypothesis is true. A low p-value, typically less than a chosen significance level (e.g., 0.05), indicates that the observed data is unlikely under the null hypothesis, leading to its rejection. Conversely, a high p-value suggests insufficient evidence to reject the null hypothesis.

Z score: $\beta_1/SE(\beta_1)$

The Z-score in the context of hypothesis testing for regression coefficients is calculated as the coefficient (β) divided by its standard error (SE).

Two Tailed Test

$$p=2(1-\Phi(|z|))$$

The two-tailed test is a statistical method used to determine whether there is a significant difference between a sample statistic and a population parameter, or between two sample statistics, in either direction. Unlike a one-tailed test, which only considers deviations in one direction, a two-tailed test assesses the possibility of an effect in both directions, thus providing a more comprehensive evaluation. This test calculates the probability of observing a value as extreme or more extreme than the sample statistic under the null hypothesis, considering both tails of the probability distribution. If the resulting p-value is less than the chosen significance level (e.g., 0.05), the null hypothesis is rejected, indicating a statistically significant difference.

Hypothesis

The economic indicators can predict the S&P 500 Index Price

Null-Hypothesis

The economic indicators cannot be used to predict for the S&P 500 index Prices

Limits

These are aiming to test the predictability of the S&P 500 index via certain economic indicators, and this possibly may not account for certain factors such as generalities and is only being numerical and Timeseries scan possibly make mistakes through prediction by using normality to predict of which can be detrimental to the S&P 500 predictions. And correlation cannot predict successfully. And correlation and coefficient does not provide causation.

Literature Review

In this literature review, several topics will be discussed using one or multiple pieces of academic literature and websites, mainly academic articles. For the review the indicators and outside factors on the index will be discussed as to how they can correlate to, predict, and possibly affect the index.

GDP

First off is GDP, and this is a large-scale economic indicator of which has the ability to either impact or predict even for the S&P 500. According to (Ball & French, 2021), the S&P 500 and the GDP do have a light correlation

with one another the correlation of which is 0.08, and thus can be seen that the index and GDP have a correlation. Of which this can be explained via another piece of literature which has found that there is causative relations between the two. According to (Mora Ros, 2019), in a bachelor thesis detailing the relationship and analysis of why this works that way the source has found through VAR regression what can be seen is the positive relationship attributed to investor confidence through GDP rises from the source and often the S&P 500 also when it goes up leads to positive shock effect in the GDP in 1-2 quarters due to higher investments and purchases. Next off, via the Granger casualty test what can be seen is that as a study limit, the index can only predict the GDP but not the other way around. Furthermore, in this relationship, it is important to understand direct and indirect causation. GDP has an indirect causation as GDP can be attributed to investor confidence and this is the bridge of variables of which the S&P 500 is affected by as investor confidence is what takes the S&P 50 fluctuations. Also, (Parnes, 2020) this paper says another topic of anomalies of which are important to consider due to deviations and what can be seen is that GDP and S&P 500 do have anomalies together based on seasonal prices and interest rates and trade balance as outside factors of which influence these variables. What can be inferred from this research is that correlation and predictions will be influenced via trends and anomalies, which can set correlations possible values to change. However, conclusively speaking what can be seen is that GDP and the S&P 50 could have some correlation, which could be used to indicate a possible ability to infer the S&P 500 direction via the GDP rates. However, correlation does not directly mean prediction, it often remains a staple of time-series and logistic regression which can be tried to use to predict values.

Inflation and CPI

Next is the inflation and CPI of the economy connected to the index. As inflation and the stock market are correlated, if inflation is overly high the S&P 500 will often go down in value. The academic literature review here will explain concepts related to the index and inflation. According to (Bouri et al., 2023), what can be seen is that inflation and expectations of it and CPI will decrease the real wealth of households and thus as a result often lower interest rates which are inversely correlated to the S&P 500, and the stock market and thus the S&P 500 goes up through inflation expectations. Furthermore, it is important to know that the inflation and S&P 500 have been positively correlated in recent years. However, correlation does not imply causation. What can be observed is that after deflationary periods, often quantitative easing will rescue the stock price, and the reason is that GDP and stocks go down due to the lack of consumption of stocks and goods. Since stocks go down due to this as a reason from that. However, what can be observed is that overly high inflation harms the index as according to (Nagy et al., 2024), there is a statistically significant relationship of which is negative between the S&P 500 and the CPI/inflation as well and this implies that the index and the inflation will be negatively correlated as the study even goes on to state that a 1% increase in inflation can be seen to associate with a 9.14 decrease in the S&P 500 of which would be nearly 2% of the index and thus showing a strong correlation between the index returns and inflation. According to (Sathyanarayana & Gargesa, 2018) as well between these factors what can be observed is contrary from Bouri of which shows negative association with inflation as Pearson coefficient shows a -0.28 correlation and similar trends through Canada GSPTSE and France FCHI with -0.1861 and -0.1785 correlations between inflation and stock performance. Thus, showing an inverse and negative correlation between inflation, also applicable to CPI as well and the S&P 500. Furthermore, this analysis applies to CPI growth, and the CPI growth/CPI will be measured with the index to show a correlation between the two. However, what should be known is generalities of which often inflation expectations and inflation and CPI growth will often if overly high evidently harm the index as that is well known in economics. Furthermore, the inflation expectation and CPI can be used to forecast the index possibly due to this research as these do connection to one another and thus can show possible ability to move together and predict the index as a result of a combination-based model.

Unemployment

Unemployment is also another indicator which affects the S&P 500 and has a correlation to it and can serve as a possible device to help predict the S&P 500 as it often has been on several occasions associated with heightening when the S&P 500 is going down. According to (Gonzalo & Taamouti, 2017), what can be seen is that often that unemployment often if anticipated leads to an increase in stock market and that when 1% of unemployment goes up thus the stock market and S&P 500 go up by 0.25%, and for quantiles 0.35 to 0.8 increase in anticipated unemployment grows stock rises. However, through Granger Causality Test, this is nonlinear as there are quantiles of which stock price will and will not be affected by the unemployment rates. However, what should be kept in mind is that often the unemployment causes this S&P 500 shift as the paper states that it shows statistically significant p-values and coefficients of the unemployment to fed funds as the correlation is -0.896 and thus the fed funds go down of which cause the S&P 500 to go up as a result of low interest rates. And thus, through some unemployment shift of which is not overly high the index goes up. However, unanticipated unemployment has no effect on the index according to the paper. Next off is (Rahman et al., 1970) of which used a test of correlation matrices and VAR modeling to show changes and correlations of the S&P 500 to unemployment and what was seen for correlation matrix was a 0.1761 correlation between the S&P 500 and the unemployment of which is a slight positive. However, what should be considered is normality in correlations as these can occur due to unconnected variables having trends going on at the same time. And the paper has also found that unemployment rates also can impact investor confidence as well and confidence about the economy which can thus cause the index to go down subsequently. However, these 2 are not very connected, as these have only a slight correlation and thus shows this to be an insignificant variable in predicting the S&P 500 index.

Home Sales

Home sales and the S&P 500 are often correlated and these can be seen most likely to have a positive or negative relationship based on the way people see them as they either serve as going up during similar conditions or either being as different and alternative investments of one another, and of which this literature review of multiple articles aims to conduct and analyze the relationship between the index and the housing sales and house market conditions. According to (Liow et al., 2019) the US stock market and its real estate market do have a relationship of which several studies have explained. And this paper introduces the concept of “Wealth effect” and that is when the stock market goes up and thus the people spend more on their homes and thus boosting home sales and prices and they buy homes as well. And along with that what can be seen is a level of causation due to this effect. Additionally, what can be known is that these two end up having greater correlations as the house market grows thus the index grows and these can go either way as well, and these 2 through regression models can be seen to bilaterally grow. Furthermore, according to (Dieci et al., 2018), what has been seen among markets is a possible inverse level of relationship as often what has been seen is when the stock market went down 2000-2003 it often promoted those in stocks switch investment to homes. Furthermore, what has been seen is that often investor invests in stocks more than homes in stock boom times and thus can show a comparative relationship of homes being bought less than stocks during this time. However, often the paper brings up a concept similar to the wealth effect as stocks often go up home sales will go up and this is due to the increase in buying of homes due to investor profits. However, they still see home buying often as a stronger asset when stocks are underperforming suggesting as a possible causative relationship that often can be either sided. And the previous claim of stocks outperforming house and in that time investors gravitating towards stocks is proven by (Tsai et al., 2011), of which states that often what can be seen is that rapid mean reversion occurs when the stock outperforms housing price. However, the inverse is not true. And thus showing that the

stock market will often be preferred by investors relative to homes when the index goes up. However, what can be seen is that the wealth effect comes into play and thus can suggest for possible positive relationships as well. However, this correlation is evidenced via causation of which has been studied and proven. And this is an exception to correlation not meaning causation.

Interest Rates

Interest rates are another variable which can be correlated to and have been used in several studies to predict the S&P 500. They often have a commonly predictable relationship with the S&P 500 as they often are attributed to negative stock market responses. Furthermore, according to (Kim, 2023), what can be seen is that high market cap stocks often will end up outperforming the interest rates stock as they have strong investor confidence attributed to them. However, often the flight to quality effect occurs during interest rate shocks as investors will often switch to safe assets such as bonds during interest rate shocks. Furthermore, what can also be seen from the study is that interest rates reduce firm flexibility and thus affect investors. And what can also be seen is that Tobin Q finds that interest rates often will have an inverse relationship with the stock market. And there are other sources which detail the stock market and interest rates of which according to (Gu et al., 2021) shows an important theory of interest rates being unable during high economic development and GDP periods not being able to surpass stock development and thus showing the stock market to resist even increasing interest rates often and thus interest rates not affecting the index much. However, what should be kept in mind is that during slow economic times the market will go down during heightening interest rates. And thus shows interest rates as a weak inhibitor of stocks. However, still having a negative association with them. And due to this they often have rather positive correlation coefficients for some experiments according to the source. Next off, according to (Alzoubi, 2022), what can be seen is a large impact on the S&P 500 and general stocks as a result of interest rates as what can first be seen is that 1% increase in interest rates are associated with a 5% decrease in stocks value, and this has been seen according to the source as a result of increasing consumption and investment encouragement which causes higher activity and this is due to the lower prices and supply and demand law. Next off is that short run model shows causative effects of interest rates on the S&P 500 index what can be seen is that ARDL model error correction term shows that long run equilibrium at a speed of 83.8% is adjusted to through short term shocks from previous years and thus shows the interest rate impact on the stock market to be strong. Thus, what can be seen is that the S&P 500 will go down often due to interest rates because of this. However, interest rates will be outpaced over a short term, but long term can affect the index and take away its growth and even cause it to decline and what has been seen is ARDL model showing causative effects such as interest rates linking to investor confidence and sentiment of which connects to the S&P 500.

Results and Analysis

This section will include and analyze the results of the experiment and derive insights based on what can be seen from these results. Furthermore, will conclude with what the results of the study mean for the economic indicators predicting and being correlated to the S&P 500.

Logistic Regression

Predictor	Description	Coefficient	P-value	Correlation
T10YIE	10 Year Expected Inflation	-0.285498	0.770633	0.682
UNRATE	Unemployment Rates	-0.099086	0.605322	-0.218
CPIAUCSL	CPI	-0.002980	0.833332	0.932
HOUST	Home Sales	-0.000133	0.942379	0.825
GDP	Gross Domestic Product	-0.000019	0.865580	0.946
REAINTRATRATEAR10Y	Real Interest Rate	-0.235604	0.684703	0.315

Figure 1.

What can be seen in this data is that a large contingent of these economic indicators have correlations that some could be perceived as strong while others could be seen as weak and some have coefficients of which are significant. However, none of these through using logistic regression to predict S&P values can successfully predict the index values, and this is due to the index having more of consumer sentiment on stocks to be the largest factor as sentiment is the largest predictor and these have very little to do with those fluctuations in sentiment and the amount of change they can be attributed to is low, along with that logistic regression itself alone cannot be used as a prediction model for the index as logarithms often cannot show a real prediction, for which time series for the predictions was used for. These economic indicators must be evaluated individually. What can be observed with the first of which is 10 year inflation expected is that the correlation is positive. However, with logistic regression different can be seen, and there is normality likely contributing to the correlation coefficient. However, through logistic regression aiming to factor this into a logarithmic equation rather than this goes down, stocks will often inversely go with inflation heightening. However, what can also be seen is that p-value is 0.77 and thus what can be finalized is that through inflation expectations the S&P 500 index cannot be predicted. However, the S&P 500 will go down when inflation goes up. However, correlation and coefficients cannot attribute to causation. However, the previous academic literature illustrates a slight causative relationship as investor sentiment can change via inflation rates such as overly high inflation can lower investor confidence. Next off is unemployment, and the correlation is -0.218 with a negative coefficient that is -0.1 and that shows the S&P 500 will respond negatively to unemployment shocks. However, unemployment alone cannot predict the S&P 500 index due to complexity of other factors. Next off is CPI of which the S&P 500 has a -0.003 coefficient to and a 0.932 correlation of which correlation can be attributed to normalities due to CPI due to inflation consistently going up and along with the S&P 500 of which goes up as well, and thus can be seen that normality attributes to this. However, the coefficient is negative, and along with the negative value shows no statistical significance as (coefficient<-0.01) and thus cannot state significant response of the index to the variable. Of which the p-value is used for and p-value has found that (p=0.83) and thus it is extremely likely to fail as a predictor of the index and thus cannot be used as logistic regression cannot be used. However, this variable cannot forecast the index. Next off is Home Sales and this has a Correlation Coefficient of 0.825 to the Index and a negative coefficient. And this correlation can be attributed to house sales going up as according to (National Association of Realtors, 2021) of which states that Home Sales have gone up from 1,040,000 to 1,210,000 over the past year of which correlation has used and the S&P 500 has gone up as well over a year and thus shows as to why the index and home sales correlate together. Furthermore, home sales and the index have an insignificant coefficient as it is -0.0001 and it shows that home sales and the S&P 500 have no significant relationship. The p-value also demonstrates inability to measure changes in the S&P 500 through home sales due to a high p-value of (p=0.94) and thus showing the home sales to have no predictability with the S&P

500. Next off is GDP of which has a correlation of 0.946 and what can be seen is that these have strong correlations due to the GDP having gone up 64% over Q1 2014 to Q1 2024, of which is a 10 year period and the S&P since then has gone up 124.35% and thus proving a positive correlation explanation. And the coefficient is low and likely due to large fluctuations and the S&P consistently going down while GDP goes up and thus attributing it to the S&P variability. Furthermore, another point of logistic regression serving as an improper model to predict as well is important as the p-value is of (p=0.85) of which is unable to produce an accurate forecast of the index through the indicator. And the coefficient is -0.00002 and this is due to that fluctuation count of the S&P 500 consistently fluctuating from high and low prices and thus showing possible explanations. However, these variables have correlation and often the S&P 500 rises when the GDP is high as GDP can influence investor confidence about the economy as well as that is a large scale factor of which is included. In confidence of investors. Last off is interest rates of which these have a 0.315 correlation to the index, of which is contradictory to the Tobin Q model that shows negative correlation. However, what can be seen is a -0.23 coefficient of which shows the index having negative response to the rise in the interest rates. And interest rates also can shape investor confidence as seen in the literature above and can often cause changes in investor confidence of which can attribute the index to going down when interest rates go up. Furthermore, the p-value is insignificant and supports null hypothesis as (p=0.68) due to being overly high. And thus what can be seen is that this variable cannot be used to predict the S&P 500 and logistic regression as well cannot be used to forecast.

Timeseries Graphs and Analysis

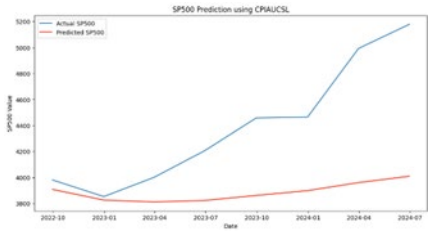

Figure 2.

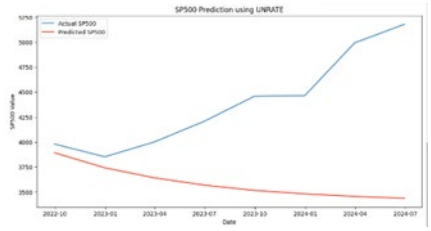

Figure 3.

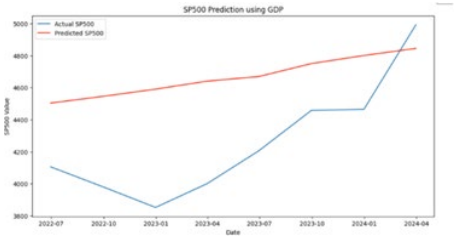

Figure 4.

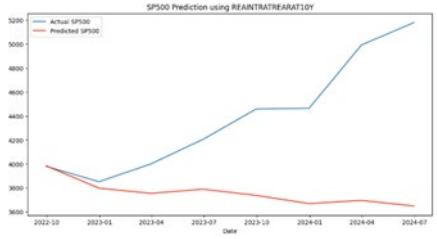

Figure 5.

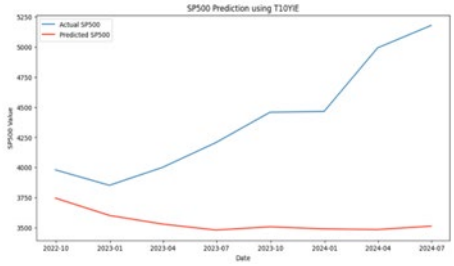

Figure 6.

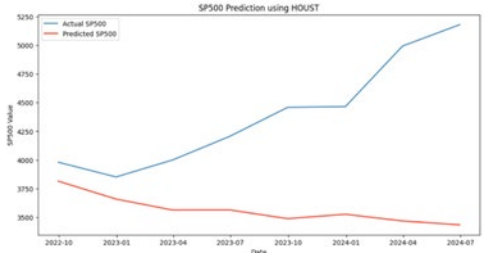

Figure 7.

Predictor	MSE
GDP	82,727.54
CPIAUCSL (CPI)	505,763.97
T10YIE (Inflation)	50,952.96
UNRATE (Unemployment)	237,665.52
REAINTRATREARAT10Y (Interest Rate)	164,493.77
HOUST (Housing Starts)	233,802.51

Figure 8.

What can be seen via time series analysis is that the factors were used to predict the index through trend-based analysis and training the data and then testing it and testing the S&P 500 via using trends in movements. This section will evaluate each in correlation. First off is the Housing Sales and what can be seen is that the index could not be predicted via Housing Sales factored in and shows the near opposite of the direction of the index and a mean squared error of (MSE=233,802.51) and thus showing housing unable to predict the S&P 500. Next

off is inflation of which was used through using as a prediction factor via training with the S&P 500 and Inflation and what was seen was a full opposite trend and best quantified by (MSE=50,952.96) and thus showing immense error in the difference predicted values and the index real value through using the performance comparison and forecasting via correlation and performance. Next off is CPI of which also demonstrates a parabolic function almost and used CPI trends to forecast the index and what can be seen is that this does not work to forecast the index as well and has the highest MSE of (MSE=505,763.97) and thus through analyzing CPI trends and training it cannot be forecasted for the index. Next is Unemployment of which has an opposite almost trend to the index and uses the index and unemployment to predict and fails to successfully forecast the index and thus through ARIMA cannot predict the index. Furthermore, the MSE is (MSE=237,665.52) and thus the ARIMA trends and unemployment with S&P 500 cannot be used to predict the index. Next off is interest rates, and this had opposite direction and mainly used trends such as coefficients to predict the index through these trends. However, despite both of the S&P 500 and interest rate trends could not combined predict the S&P 500 and thus what could be seen was that the index could not be predicted via the interest rates and the MSE value for these is (MSE=164,493.77) and the interest rates cannot predict. Last off is GDP of which what can be seen is that it predicted the S&P at the start to be overly high but now the real S&P 500 is much higher now compared and this cannot predict right values however what can be seen is that it shows possible trends in the index through showing general positive trend. However, it does not show full success rather than generalities and MSE is (MSE=82,727.54) and thus showing an inaccuracy in the prediction. This model used the trends and used stationary testing and the patterns through the growth and loss of the index and the indicator and thus predict the index through that. However, cannot be attributed to successfully predicting the index through these combined trends.

Conclusion

What can be concluded from this study is that the economic indicators of CPI, Inflation, Interest Rates, Housing Sales, GDP, and unemployment fail to predict the index through logistic regression due to overly high p-values, which shows that a logistic based regression cannot predict the index, along with a timeseries based approach cannot predict the S&P 500 despite using timeseries based modeling of which co-movements were used to predict the index through the indicators. They often were seen to move differently from the index even via timeseries. However, GDP had some direction of which could predict the index via direction. However, it had a high MSE, which showed inaccuracy. And thus the economic indicators are concluded to not be able to predict the index, thus proving the null hypothesis. Furthermore, what can be seen is that Logistic regression is unable to be used as a prediction model due to it being based on powers, exponentiations, and linearity, and stocks are not linear and thus cannot be a certified model. However, timeseries can be due to the ability to understand variability. Furthermore, what was learned is that the S&P 500 is a prediction model of these other indicators such as GDP and inflation. Along with it, these have levels of co-movement to one another as well via testing through logistic regression of which several indicators have co-movements with the S&P 500. The literature review shows the S&P 500 as a forecasting model for some of these. Furthermore, what can be seen is that inflation, interest rates, and unemployment will have negative coefficients with the index of which can show that the S&P 500 will respond negatively through changes. And what must be known from this is that correlation serves as insignificant as it negates normality, and trends and other factors as normality can come into changing the index as correlation between the index and the interest rates were positive, however, the coefficient was negative and thus showing correlation to be a bad model as Tobin Q suggests that the S&P 500 will respond negatively to the interest of which the coefficient showed and thus showing how correlation fails to account for increasing and non-correlated data moving in the same direction. Lastly is that the S&P 500 is too variable to be predicted via trend analysis and economic indicators and the economic indicators cannot forecast the index at all due to the variability of the index and the possible trends coming into play via timeseries. And the p-

values showing that as well. And thus what can be concluded is that the economic indicators are not sufficient enough to predict the S&P 500.

Acknowledgments

I would like to thank my advisor for the valuable insight provided to me on this topic.

References

Alqaralleh, H., Canepa, A., & Salah Uddin, G. (2023). Dynamic relations between housing Markets, stock Markets, and uncertainty in global Cities: A Time-Frequency approach. The North American Journal of Economics and Finance, 68(101950), 101950. https://doi.org/10.1016/j.najef.2023.101950

Alzoubi, M. (2022). Stock market performance: Reaction to interest rates and inflation rates. Banks and Bank Systems, 17(2), 189–198. https://doi.org/10.21511/bbs.17(2).2022.16

Ball, C., & French, J. (2021). Exploring What Stock Markets Tell Us About GDP In Theory and Practice. Research in Economics, 75(4). https://doi.org/10.1016/j.rie.2021.09.002

Bhutta, N., & Ringo, D. (2020). The effect of interest rates on home buying: Evidence from a shock to mortgage insurance premiums. Journal of Monetary Economics. https://doi.org/10.1016/j.jmoneco.2020.10.001

Bilello, C. (2018, May 7). The Unemployment Rate And The Stock Market | Seeking Alpha. Seekingalpha.com; Seeking Alpha. https://seekingalpha.com/article/4170913-unemployment-rate-and-stock-market

Bloomberg. (2024, June 10). SPX Quote - S&P 500 Index. Bloomberg.com. https://www.bloomberg.com/quote/SPX:IND

Bouri, E., Nekhili, R., Kinateder, H., & Choudhury, T. (2023). Expected inflation and U.S. stock sector indices: A dynamic time-scale tale from inflationary and deflationary crisis periods. Finance Research Letters, 55(103845), 103845. https://doi.org/10.1016/j.frl.2023.103845

Butler, D. (2020, May 5). Historical S&P 500 Returns. TheStreet; TheStreet. https://www.thestreet.com/investing/annual-sp-500-returns-in-history

Časta, M. (2023). Inflation, interest rates and the predictability of stock returns. Finance Research Letters, 58, 104380–104380. https://doi.org/10.1016/j.frl.2023.104380

Coursera. (2023, September 11). What Is Machine Learning? Definition, Types, and Examples. Coursera; Coursera. https://www.coursera.org/articles/what-is-machine-learning

Dieci, R., Schmitt, N., & Westerhoff, F. (2018). Interactions between stock, bond and housing markets. Journal of Economic Dynamics and Control, 91, 43–70. https://doi.org/10.1016/j.jedc.2018.05.001

Duggan, W. (2023, November 12). S&P 500 Definition. US News Money; US News. https://money.usnews.com/investing/term/sp500#:~:text=The%20S%26P%20500%20is%20a%20market-capitalization-weighted%20stock%20market,market%20and%20the%20U.S.%20economy%20as%20a%20whole.

Gonzalo, J., & Taamouti, A. (2017). The reaction of stock market returns to unemployment. Studies in Nonlinear Dynamics & Econometrics, 21(4). https://doi.org/10.1515/snde-2015-0078

Gu, G., Zhu, W., & Wang, C. (2021). Time-varying influence of interest rates on stock returns: evidence from China. Economic Research-Ekonomska Istraživanja, 35(1), 1–20. https://doi.org/10.1080/1331677x.2021.1966639

Kim, J. (2023). Stock market reaction to US interest rate hike: evidence from an emerging market. Heliyon, 9(5), e15758. https://doi.org/10.1016/j.heliyon.2023.e15758

Kurpiel, S. (2024, April 25). Research Guides: Evaluating Sources: The CRAAP Test. Benedictine University; Benedictine University. https://researchguides.ben.edu/source-evaluation
Liow, K. H., Huang, Y., & Song, J. (2019). Relationship between the United States housing and stock markets: Some evidence from wavelet analysis. The North American Journal of Economics and Finance, 50(101033), 101033. https://doi.org/10.1016/j.najef.2019.101033
Mora Ros, J. (2019). U.S. GDP and S&P 500 : an inquiry into the nature and causes of the econometric relation between GDP and stock market in the U.S. Repositori.upf.edu, 31. e-Repositori upf. http://hdl.handle.net/10230/42502
Nagy, M., Valaskova, K., Kovalova, E., & Macura, M. (2024). Drivers of S&P 500's Profitability: Implications for Investment Strategy and Risk Management. Economies, 12(4), 77. https://doi.org/10.3390/economies12040077
National Association of Realtors. (2021, July 1). Existing Home Sales: Housing Inventory. FRED, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/HOSINVUSM495N
Parnes, D. (2020). Exploring economic anomalies in the S&P500 index. The Quarterly Review of Economics and Finance, 76(76), 292–309. https://doi.org/10.1016/j.qref.2019.09.012
Rahman, M., Mustafa, M., & Caples, S. (1970). Influences of Unemployment Rates and S&P 500 Movements On Eight Selected U.S. Casino Stock Performances. Journal of Business Strategies, 31(1), 221–240. https://doi.org/10.54155/jbs.31.1.221-240
Rodriguez, F. S., P. Norouzzadeh, Anwar, Z., E. Snir, & Rahmani, B. (2024). A machine learning approach to predict the S&P 500 absolute percent change. Discover Artificial Intelligence, 4(1). https://doi.org/10.1007/s44163-024-00104-9
Roncaglia de Carvalho, A., Ribeiro, R. S. M., & Marques, A. M. (2017). Economic Development and inflation: a Theoretical and Empirical Analysis. International Review of Applied Economics, 32(4), 546–565. https://doi.org/10.1080/02692171.2017.1351531
Sathyanarayana, S., & Gargesa, S. (2018). An Analytical Study of the Effect of Inflation on Stock Market Returns. IRA-International Journal of Management & Social Sciences (ISSN 2455-2267), 13(2), 48. https://doi.org/10.21013/jmss.v13.n2.p3
Tsai, I-Chun., Lee, C.-F., & Chiang, M.-C. (2011). The Asymmetric Wealth Effect in the US Housing and Stock Markets: Evidence from the Threshold Cointegration Model. The Journal of Real Estate Finance and Economics, 45(4), 1005–1020. https://doi.org/10.1007/s11146-011-9304-5
U.S. Bureau of Economic Analysis. (2023). Gross Domestic Product. Stlouisfed.org. https://fred.stlouisfed.org/series/GDP