The Faculty of Engineering in Foreign Languages

S&P 500 as an economic indicator of the status of the US economy – ARIMA forecast

Student: Tanase Nicolae-Costin

Bucharest 2022

Contents

Introduction

The S&P500 is an American stock market index based on the market capitalizations of 500 large companies having common stock listed on the NYSE or NASDAQ. The S&P500 index components and their weightings are determined by S&P Dow Jones Indices. It is one of the most commonly followed equity indices, and many consider it to be one of the best representations of the U.S. stock market. The index includes 500 of the top U.S. companies by market capitalization. The S&P500 was first calculated on March 4, 1957. The index reached its first intraday high on November 23, 1963, and its first closing high on January 11, 1964. It then fell sharply in the wake of the assassination of President John F. Kennedy, but quickly recovered and rose to new highs in early 1966. The S&P500 index hit an all-time high of 1,527.46 on January 15, 2000. The S&P500 has had several notable ups and downs in its history. After reaching its peak in 2000, the index fell sharply in the wake of the dot-com bubble burst. It then recovered in the mid-2000s and reached new highs in 2007.

In order to understand how the S&P 500 beats inflation, one must first understand what inflation is. Inflation is the rate at which the prices of goods and services rise over time. The main cause of inflation is the expansion of the money supply. When the money supply grows faster than the economy, prices go up. The S&P 500 is a stock market index that consists of 500 large publicly traded companies. The index is widely regarded as a good indicator of the overall performance of the stock market. The reason the S&P 500 beats inflation is because the prices of the stocks in the index go up at a rate that is higher than the rate of inflation. This is due to the fact that the earnings of the companies in the index grow at a rate that is higher than the rate of inflation. Over the long term, the S&P 500 has outperformed inflation by a wide margin. For example, from December 31, 1970 to December 31, 2016, the Consumer Price Index (CPI) rose by a factor of 11.4. During the same period, the S&P 500 rose by a factor of 36.2. This means that the S&P 500 rose at a rate that was 3.2 times higher than the rate of inflation. There are a number of

reasons why the S&P 500 beats inflation over the long term. One of the main reasons is that the earnings of the companies in the index grow at a rate that is higher than the rate of inflation. This is due to the fact that the companies in the index are able to increase their prices at a rate that is higher than the rate of inflation. Another reason why the S&P 500 beats inflation is that the dividends paid by the companies in the index tend to grow at a rate that is higher than the rate of inflation. This is due to the fact that the companies in the index are able to increase their profits at a rate that is higher than the rate of inflation. Lastly, the S&P 500 beats inflation because the stocks in the index are bought and sold in US dollars. This means that the value of the stocks in the index is not affected by changes in the value of other currencies. The bottom line is that the S&P 500 beats inflation because the prices of the stocks in the index go up at a rate that is higher than the rate of inflation. This is due to the fact that the earnings of the companies in the index grow at a rate that is higher than the rate of inflation.

The S&P 500 is a capitalization-weighted index. This means that the companies in the index are weighted according to their market capitalization (the total value of their outstanding shares). The index is also float-adjusted, meaning that only the shares that are available to the public are included. The S&P 500 was first published in 1957, and it is currently maintained by Standard & Poor's, a division of McGraw Hill Financial. The S&P 500 is widely considered to be a bellwether for the U.S. economy. This is because the companies in the index are leaders in their respective industries, and their stock prices tend to be affected by economic conditions. The S&P 500 is often used as a benchmark for U.S. stock market performance. This is because it is a broad-based index, and it is considered to be a leading indicator of U.S. economic health. The S&P 500 is often used by investors as a gauge of their portfolio's performance. This is because the index is a good representation of the overall U.S. stock market, and it is a leading indicator of U.S. economic health. The S&P 500 is also used by economists as a leading indicator of U.S. economic activity. This is because the companies in the index are leaders in their respective industries, and their stock prices tend to be affected by economic conditions. The S&P 500 is a valuable tool for investors and economists alike. It is a broad-based index that is considered to be a leading indicator of U.S. economic health, and it is a good representation of the overall U.S. stock market.

In the following pages, I aim to analyze and make a forecast on the S&P 500 ETF and also to analyze what this signifies to the economic status of the United States.

Data description

In the analysis, the shares of the S&P 500 will be studied. The start point will be 01.02.1993 and the endpoint will be 29.05.2022. The source of the data is Yahoo Finance and the information contained for each share is:

- Date
- High, the highest price for the share for the day
- Low, the lowest price for the share for the day
- Open, the opening price for the share for the day
- Close, the closing price for the share for the day
- Volume, the number of shares transacted for the day

The data isn’t continuous, having skip dates when the stock exchange is closed, for example on the weekends.

Candlestick Analysis

For starters, the candlestick graph will be analyzed for the stated period. A candlestick chart (also called Japanese candlestick chart) is a style of financial chart used to describe price movements of a security, derivative, or currency. Each "candlestick" typically shows one day, thus a one-month chart may show the 20 trading days as 20 "candlesticks".

FIGURE 1 CANDLESTICK CHART OF S&P 500 ETF

As seen in the chart above, the shares of S&P shares have fluctuated throughout the period while maintaining a upper trend that has constantly beat inflation, experiencing a sharp incline after the 2008 crisis. The highest value was reached on 2021-01-04, that being 477.73 USD , and the lowest was reached in 1994-07-01, that being 26.96 USD. Of course, this doesn't tell us much since the inflation on this timeframe is quite significant. The chart also shows that the fluctuations happen even throughout the day, as you'd expect from a such important ETF.

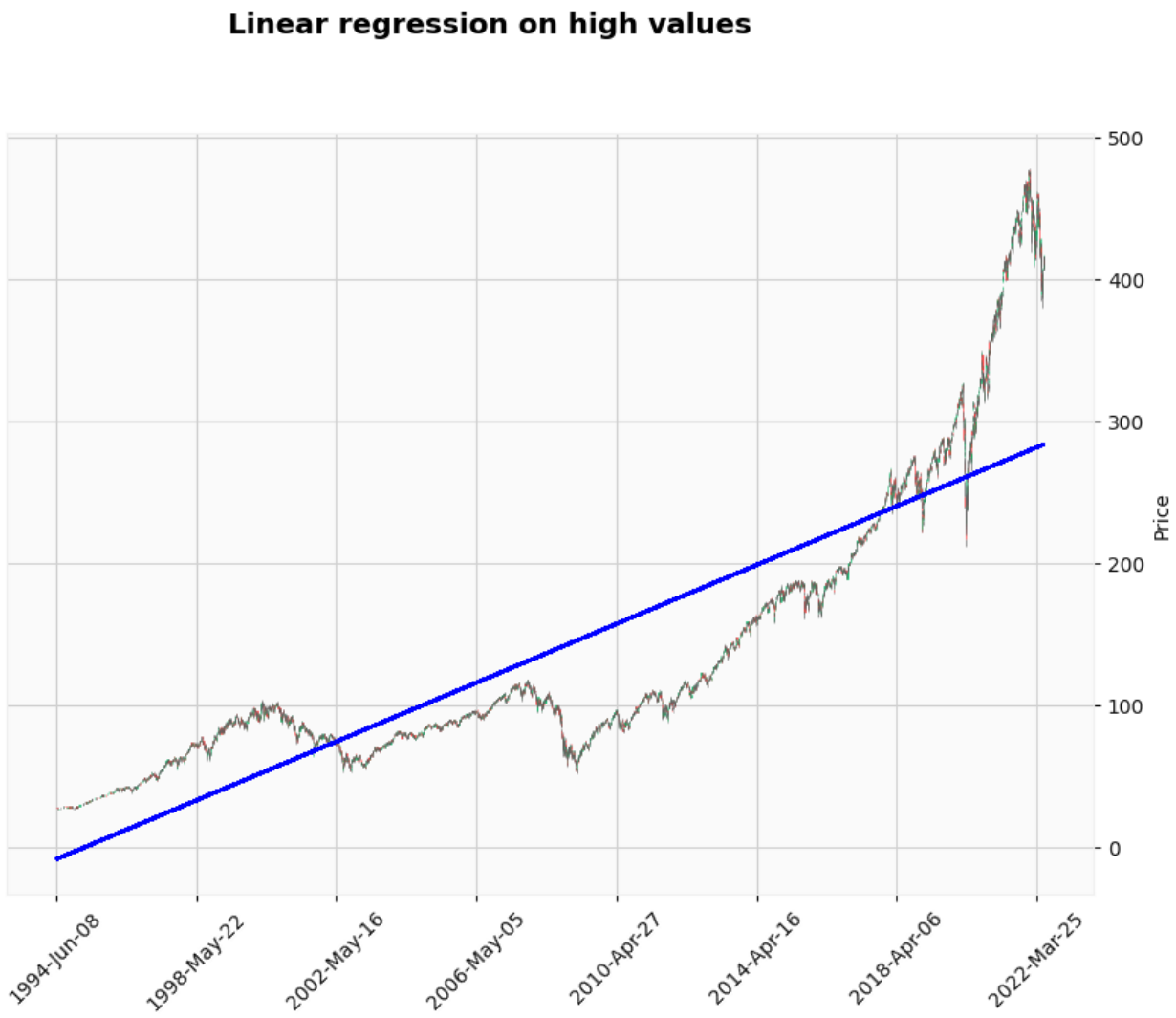

FIGURE 2 CANDLESTICK CHART WITH THE LINEAR REGRESSION OF THE HIGH VALUES

The regression line of the High series shows the upward trend of the share value for the period and it also shows that non-stationarity exists.

count	7052.000000
mean	137.461749
std	99.316990
min	26.967525
25%	75.695551
50%	96.511663
75%	181.142402
max	478.493512

The High series can be described as having:

- A mean of 137.46
- A standard deviation of 99.31
- A minimum value of 26.96
- A maximum value of 478.49
- The first quartile of 75.69
- The median or second quartile of 96.51
- The third quartile of 181.14

FIGURE 3 DESCRIPTIVE STATISTICS FOR HIGH SERIES

Next, we will be looking at the visual representation of the distribution of values:

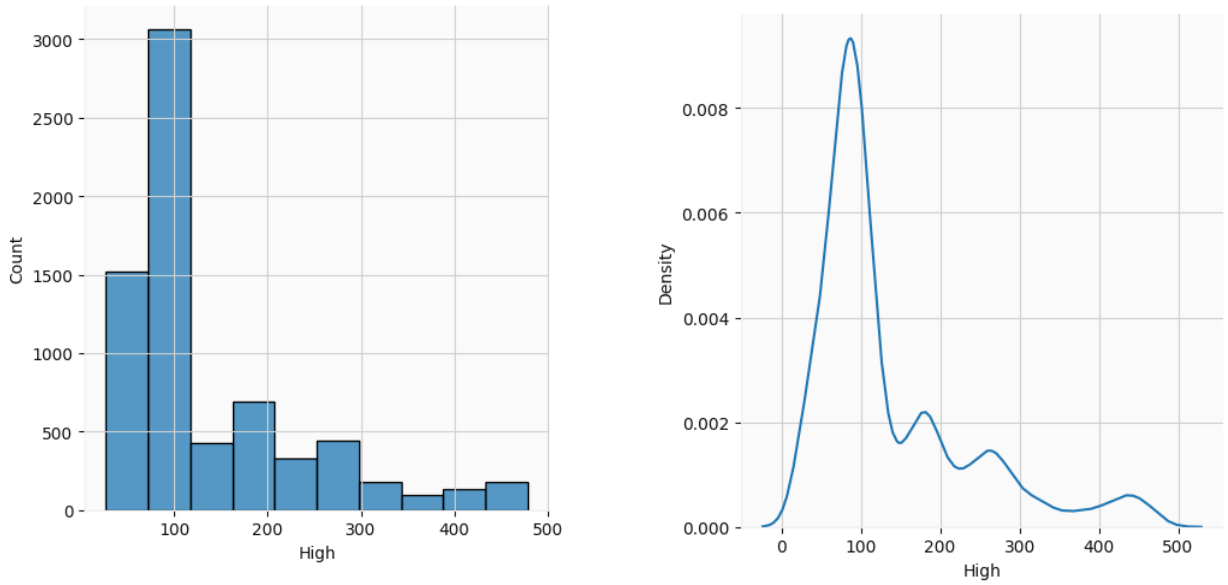

FIGURE 4 DISTRIBUTION OF HIGH SERIES

It can be seen from the previous chart that the series doesn't follow a Gaussian distribution, far from it actually. This means that in order to be analyzed the series will most likely have to be differentiated.

Stationarity analysis

Firstly, the autocorrelation coefficient must be studied using a autocorrelation function:

```
array([1.        , 0.99924438, 0.99846879, 0.99769298, 0.99693343,
       0.99615785, 0.99538034, 0.99460703, 0.99383723, 0.99309139,
       0.99236579, 0.99165552, 0.99093225, 0.99020816, 0.98949661,
       0.98875142, 0.98799484, 0.98724849, 0.98651198, 0.98580855,
       0.98507582, 0.98432978, 0.98358899, 0.98281992, 0.9820144 ,
       0.9811884 , 0.98040837, 0.97964711, 0.97884554, 0.97803318,
       0.97724413, 0.97644021, 0.97562915, 0.97478306, 0.97388616,
       0.97299585, 0.97212269, 0.97126752, 0.97039116, 0.96951013,
       0.96861828])
```

FIGURE 5 AUTOCORRELATION FUNCTION

It is kind of clear that the trend is not very strong (not having the oscillation and movements of a single stock), having a serious upward trend only in recent years (post 2008).

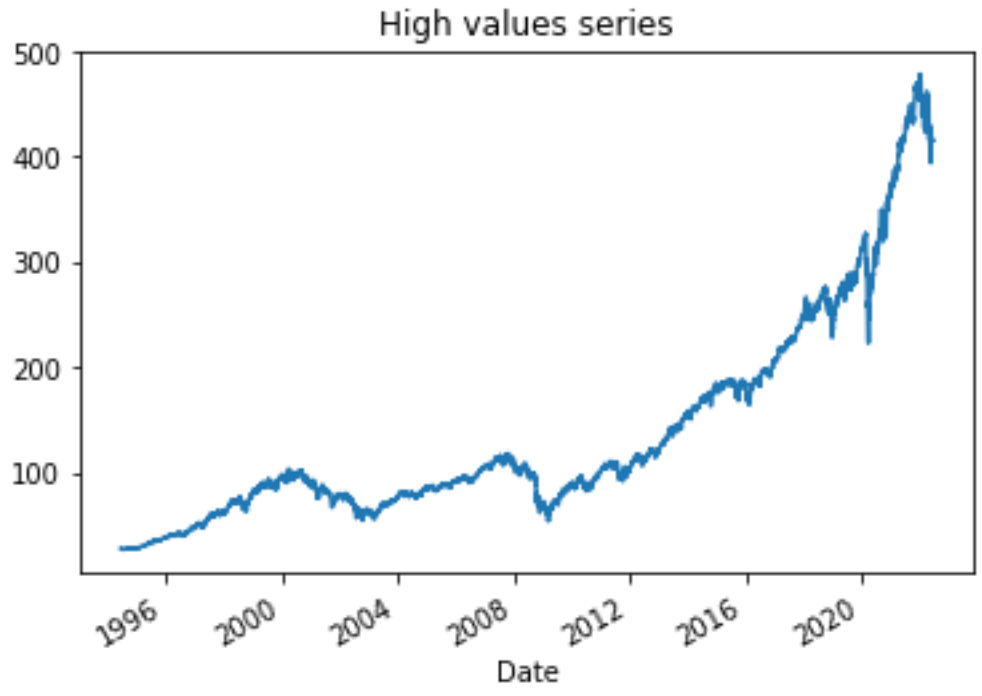

FIGURE 6 HIGH VALUES SERIES

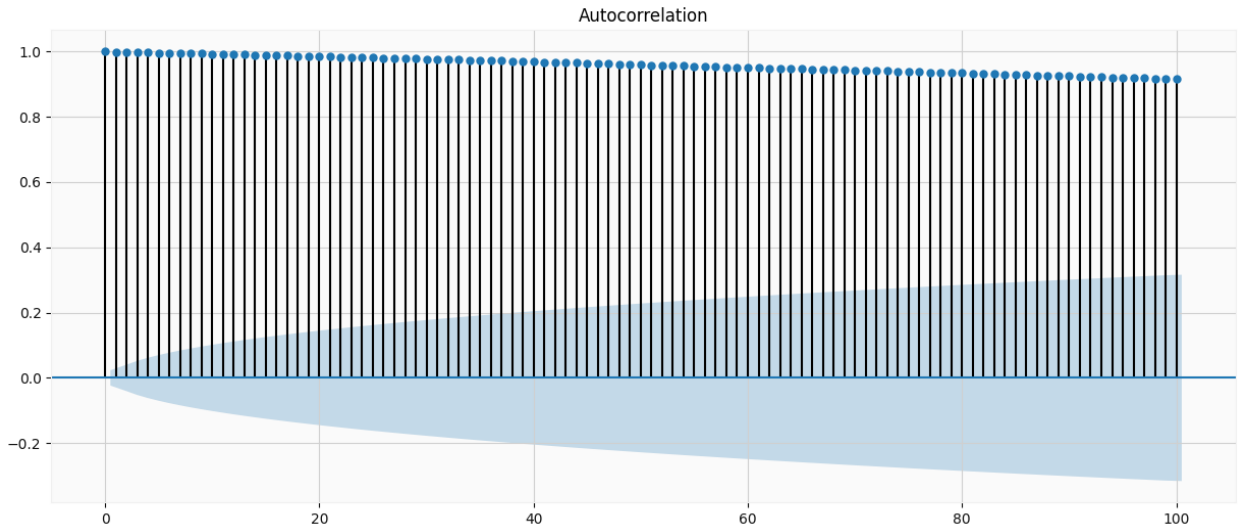

FIGURE 7 HIGH VALUES ACF CHART

For a more detailed view of the autocorrelation plot, the notebook shows more clearly that the series is not stationary and must be differentiated.

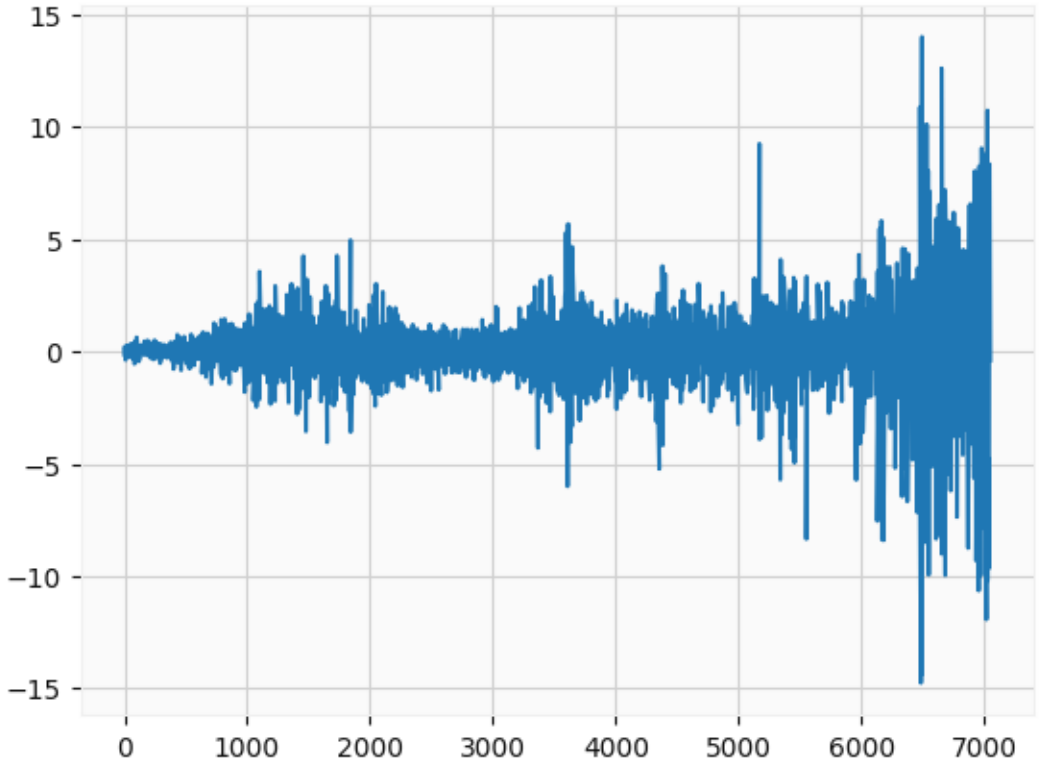

FIGURE 8 HIGH VALUES DIFFERENCED

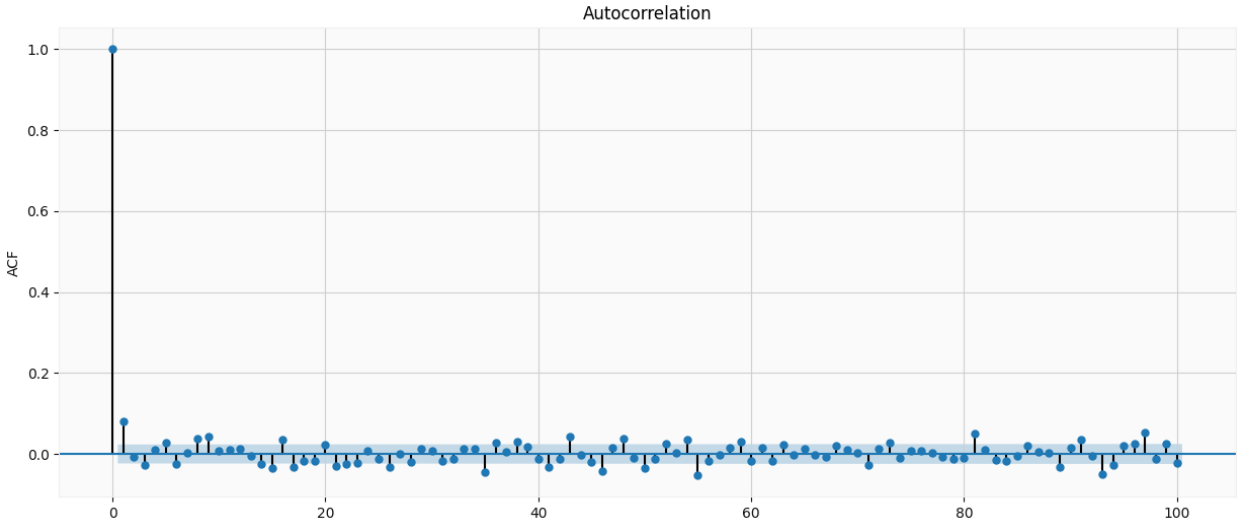

FIGURE 9 ACF CHART FOR DIFFRENCED VALUES

From the previous graphs it is clear that the series needs no more differencing in order to become stationary.

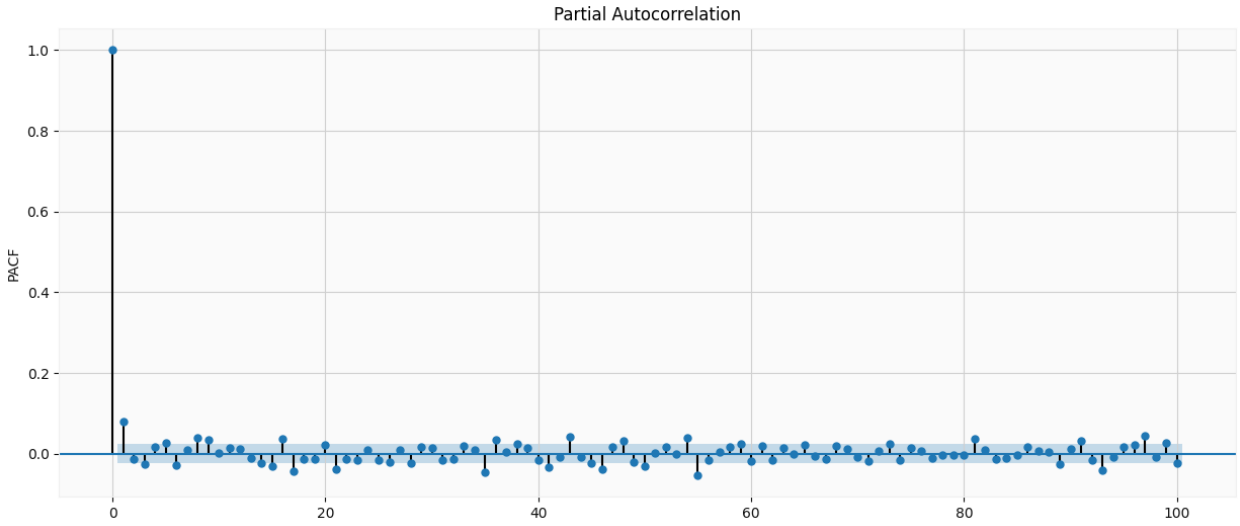

FIGURE 10 PACF FOR HIGH DIFFERENCED VALUES

After being differentiated once the series becomes stationary. There are no significant values present in the ACF and PACF charts. Therefore, an ARIMA model can now be built for the series. Now, the choice of the three parameters, P, D and Q remains. D is clearly 1, since it only

took 1 differencing to achieve stationarity. I have chosen P and Q to be 1 since in both the ACF and PACF charts for 100 lags there are very few outliers and are all close to the limit. Also, the ARIMA summary gives me a poorer result when increasing P and Q.

Arima model and forecasting

After the model is built, this is the summary:

ARIMA Model Results

Dep. Variable:	D.High	No. Observations:	7051
Model:	ARIMA(1, 1, 1)	Log Likelihood	-12773.205
Method:	css-mle	S.D. of innovations	1.481
Date:	Wed, 08 Jun 2022	AIC	25554.410
Time:	17:46:04	BIC	25581.854
Sample:	1	HQIC	25563.863

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.0550	0.019	2.891	0.004	0.018	0.092
ar.L1.D.High	-0.0139	0.112	-0.124	0.902	-0.234	0.206
ma.L1.D.High	0.0945	0.111	0.847	0.397	-0.124	0.313

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	-72.0391	+0.0000j	72.0391	0.5000
MA.1	-10.5841	+0.0000j	10.5841	0.5000

This shows us that only one order of differencing worked very well.

FIGURE 11 ARIMA MODEL RESULTS

For the final part of this paper, a forecast will be made for the built model. The original series will be split into two, a training series and a testing series. 95% of the original series will be used for training and 5% for testing the predictions. This is unusual, but the sharp increase in 2008 caused the ARIMA model to perform poorly if we take pre-2008 data for training. Another alternative would be to take a train/test split on post-2008 data, but I feel that it is too small a sample to be used for this purpose.

FIGURE 12 FORECASTING FOR EVERY VALUE

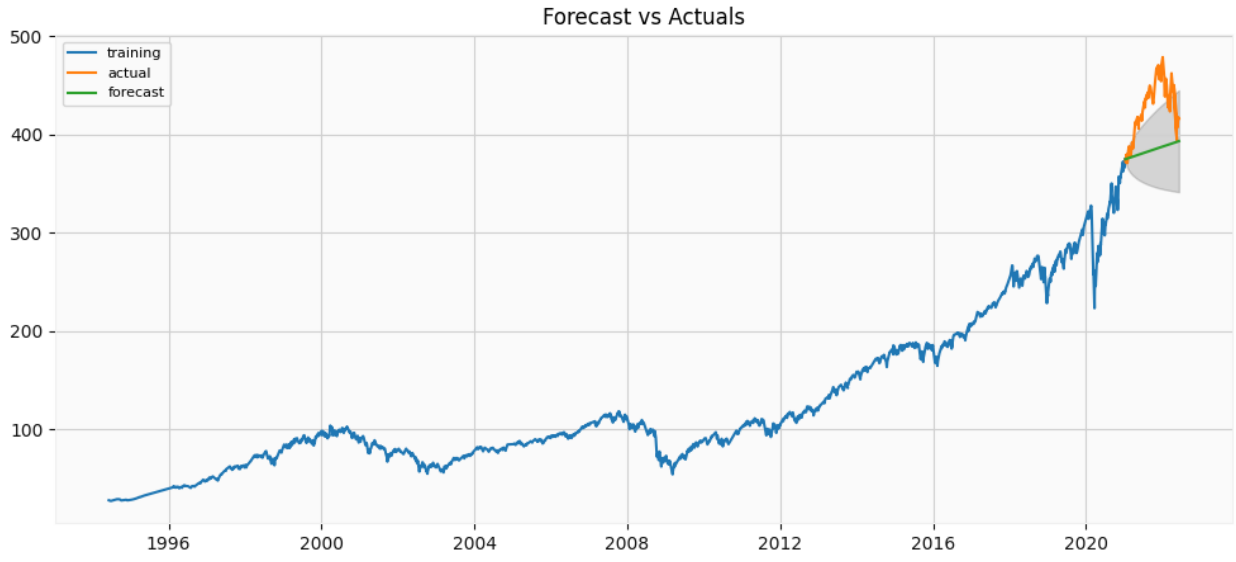

FIGURE 13 FORECASTING WITH A 95/5 TRAIN/TEST SPLIT

```
[30] from sklearn.metrics import mean_squared_error
     from math import sqrt

     mse = mean_squared_error(test, fc_series)
     RMSE = sqrt(mse)
     print(RMSE)

     51.1675625656885
```

FIGURE 14 ROOT MEAN SQUARED ERROR

From the chart above we can conclude that the prediction the model offered is satisfactory. The RMSE is 51.16, which is unexpectedly good considering the timeframe chosen. It's this good not because of the method, but probably just because the data doesn't have that much seasonality and trend for a long time.

Conclusions

In conclusion, given the data, the S&P 500 is highly correlated with the US economy. It crashes when the economy is in recession, and it increases along with economic expansion. It is always slightly in front of inflation, and it usually is a great indicator of the current status of US economics. Using ARIMA forecasting, we can get the trend not only of the S&P 500, but to a high extent the whole US economy. A notable mention is that the ARIMA model assumes a little too much, and since its existence has experienced a lot of financial anomalies ultimately caused by the US economy. During its existence it has seen some artificial growth and some crashes caused by burst bubbles, and also has been raised by the somewhat unhealthy money printing that the US has been doing for the last decade. The upward trend signifies more the S&P 500 value in contrast to the USD rather than what the USD is worth in the economy, which might be leading to a correlation decrease between this ETF and the economy in the future.