Kyle Arbide, Patrick O'connor, Hanna-Grace Rabanes, Qing Zhu | EMSE 6574
Cryptocurrency and S&P 500 Market Trends

1. Introduction

Since the inception of Bitcoin in 2009, decentralized cryptocurrencies have spawned at an astonishing rate. Data suggest that over "10,000 cryptocurrencies [were] in existence as of November 2021;" that number continues to climb.[1] Although Bitcoin remains the heavyweight currency with a market capitalization (which, for cryptocurrencies, amounts to the current price of a single coin multiplied by the total number of coins in circulation) hovering around 1 trillion USD in 2021, the overall crypto market is gaining, reaching a market cap of over 3 trillion USD on 3 November 2021.[2-4]

Such successful growth has prompted increased speculation in cryptocurrencies as not only a viable medium of asset exchange, but as a potential safe haven for large stores of capital: "In September [2021] El Salvador became the world's first nation to adopt bitcoin as legal tender," and continued to demonstrate its trust in the currency when it "bought an additional 100 bitcoins after the currency declined in value" in anticipation that it would recoup their purchase price with interest.[5] Despite dissidence from global financiers toward the use of bitcoin (and by implication other cryptocurrencies) as legal tender, dialogue persists that cryptocurrencies are revolutionary stores of value that are "theoretically immune to government interference or manipulation" due to the decentralized and encrypted nature of the networks across which they are exchanged.

By contrast, centralized markets, such as the New York Stock Exchange, remain the uncontested medium for exchanging relatively stable equities, such as so-called "blue chip" stocks, which are expected to return reasonably steady gains over the long term.[6] For this reason, stock market indexes, and in particular the S&P 500, are viewed as reliable indicators of economic performance. The S&P 500 "is a market-capitalization-weighted index" "of the 500 largest companies in the U.S."[7,8] Because of its weighting conventions, companies with larger market caps have greater influence over the performance of the S&P 500 than smaller firms.[7] Downturns in the equities of the largest U.S. companies may therefore predict dips in the overall market.

As of August-September 2021, 7 of the 10 largest firms on the S&P 500 were technology or internet companies, indicating that the digital sector plays a significant role in driving market performance. Apple Inc.; Microsoft Corp.; Amazon.com, Inc.; Facebook, Inc.; Alphabet Inc. Class A; and Alphabet Inc. Class B comprised the top 6 S&P 500 firms in that order, followed by Nvidia Corp. as the 8th largest firm.[7] Because cryptocurrencies represent the bleeding edge of digital innovation in finance, they may prove to be reliable indicators of economic performance.

This analysis compares three major cryptocurrencies (Bitcoin, Ethereum, and Dogecoin) against the S&P 500 to test the hypothesis that: first, cryptocurrency performance may reliably predict gross market trends in the United States; and second, cryptocurrencies may provide a safe haven for capital, or hedge, against an S&P 500 downturn. To achieve this objective, we assess whether there is evidence of statistical relationships between these assets and develop predictive models of asset performance.

2. Methodology

We utilized the following methodology to conduct our analysis:

1. We established and cleaned a real-time data feed from Yahoo Finance to extract standardized price data for each of the following assets: a) S&P 500 (SP500), b) Bitcoin (BTC), c) Ethereum (ETH), and d) Dogecoin (DOGE).
2. We performed an Exploratory Time Series Analysis to assess individual asset performance over time.
3. We performed Percent Change and Correlation Analyses to compare volatility among and assess for statistically significant relationships between asset classes.
4. We performed a Significant Decline Analysis to assess the performance of cryptocurrency assets during significant downturns in the S&P 500.
5. We developed predictive regression and ARIMA models to forecast asset performance.

2.1 Data Feed Setup

We utilized the **yfinance** library to call real-time price data from Yahoo! Finance (Figure 1). The Yahoo! Finance data set contained the daily opening price, daily closing price and other metrics for BTC, ETH, DOGE, and SP500. Utilizing Yahoo! Finance to extract our raw data allowed us to ensure that the data would be collected over identical time periods and would have consistent formatting with minimal processing. Because Yahoo! Finance is a reputable source for both historical and real-time market performance data, we were also confident that we would be able to extract robust and accurate data for our analysis.

We merged the datasets for each asset onto a single data frame indexed by date. The final data frame contained the daily closing price of each asset as available between the dates of 02 January 2021 through 31 October 2021, inclusive (Figure 2).

Figure 1. Initial Data Feed Setup

```
!pip install yfinance
import pandas_datareader as web
from pandas_datareader import data as pdr
import yfinance as yfin
import pandas as pd
yfin.pdr_override()
```

```
BTC = pdr.get_data_yahoo("BTC-USD", start="2014-01-01", end="2021-10-31")
ETH = pdr.get_data_yahoo("ETH-USD", start="2014-01-01", end="2021-10-31")
DOGE = pdr.get_data_yahoo("DOGE-USD", start="2014-01-01", end="2021-10-31")
SP5 = pdr.get_data_yahoo("^GSPC", start="2014-01-01", end="2021-10-31")
```

```
[*********************100%***********************]  1 of 1 completed
[*********************100%***********************]  1 of 1 completed
[*********************100%***********************]  1 of 1 completed
[*********************100%***********************]  1 of 1 completed
```

Figure 2. Final Merged Dataset

Date	Adj_Close_SP5	Adj_Close_ETH	Adj_Close_DOGE	Adj_Close_BTC
2014-01-02	1831.979980	NaN	NaN	NaN
2014-01-03	1831.369995	NaN	NaN	NaN
2014-01-06	1826.770020	NaN	NaN	NaN
2014-01-07	1837.880005	NaN	NaN	NaN
2014-01-08	1837.489990	NaN	NaN	NaN
...	...	...	...	...
2015-07-19	NaN	NaN	0.000169	273.614014
2015-07-25	NaN	NaN	0.000188	288.696991
2015-07-26	NaN	NaN	0.000187	292.686005
2015-08-01	NaN	NaN	0.000175	281.601013
2015-08-02	NaN	NaN	0.000172	282.614014

2780 rows × 4 columns

2.2 Exploratory Time Series Analysis

We graphed the adjusted daily closing price of each asset over time on a single plot to compare the overall performance of each asset against one another during the period from January 2014 to October 2021 (Figure 3). However, because of the wide disparity in prices between assets, most dramatically between BTC and the remaining three assets, we decided to graph each asset individually to better understand asset volatility (Figure 4).

Figure 3. Combined Asset Time Series Analysis

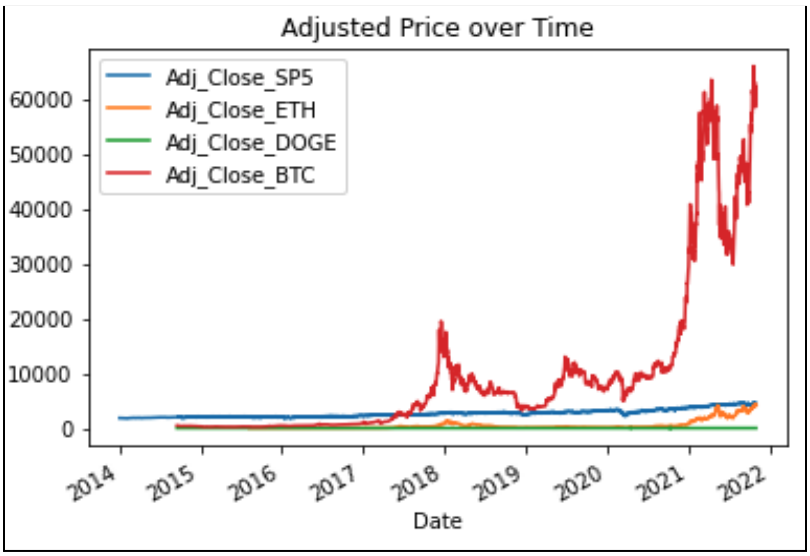
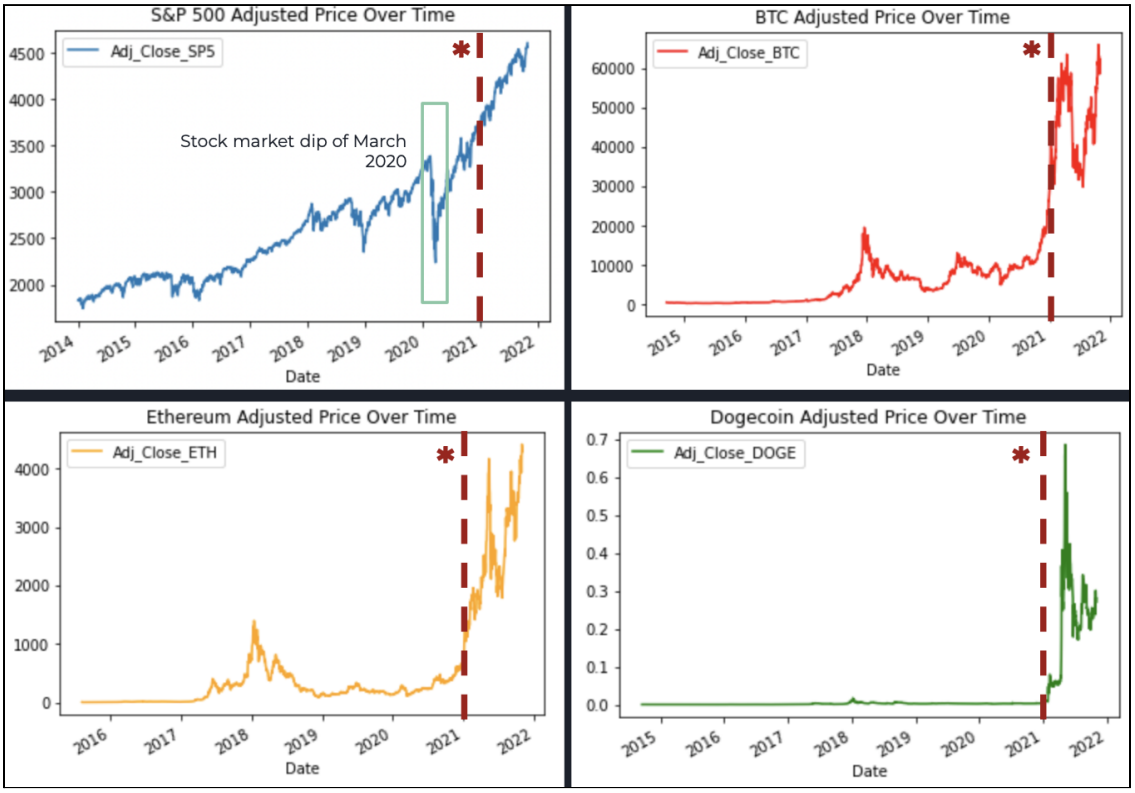


Figure 4. Individual Asset Time Series Analysis



* Dashed line approximates 26 Jan 2021, the date on which Elon Musk tweeted 'Gamestonk!!' on Twitter.

Figure 4 above indicates that BTC, ETH, and DOGE are volatile assets relative to SP500, which has steadily increased in value since 2014, with the exception of a notable dip in the market coinciding with the WHO declaration of COVID-19 as a global pandemic in March 2020. By contrast, BTC and ETH, which existed prior to the COVID-19 pandemic and preceded the creation of DOGE by several years, initially seemed resilient to the March 2020 market downturn. Although an analysis of external market influences is outside the scope of this work, anecdotal evidence suggests that the performance of cryptocurrencies is more readily influenced by trending internet activity. As noted in Figure 4, an uptick in price across all three cryptocurrency assets appears to coincide with viral interest in so-called 'meme stocks.'

2.3 Percent Change Analysis and Significant Decline Analysis

For volatile, non-stop assets, like cryptocurrencies, the day-to-day adjusted closing price may not accurately reflect how much an asset can change between days, months, or years. To better capture this rate of change, our team calculated the percentage change between each days' closing price. Because weekend prices are not available for SP500, we utilized the 'ffill' method in the percent_change function to fill any NA's with the most recent valid integer.

Figure 5. Percent Change Analysis Plot

```
[ ] #Creating a percent change dataframe

    pc1= c1.pct_change(fill_method='ffill')
    pc1=pc1*100

    pc1.plot(title='Daily Percent Change')

    <matplotlib.axes._subplots.AxesSubplot at 0x7f8d2ee1a590>
```

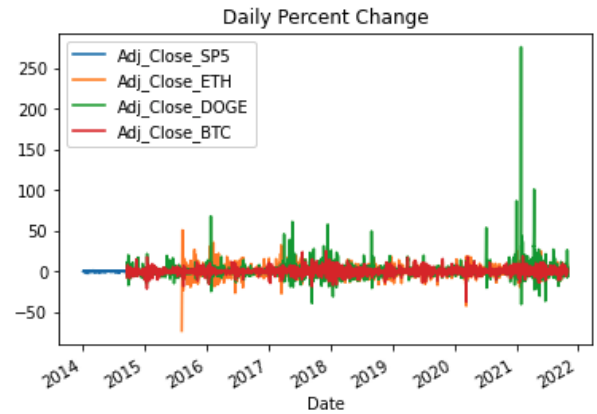
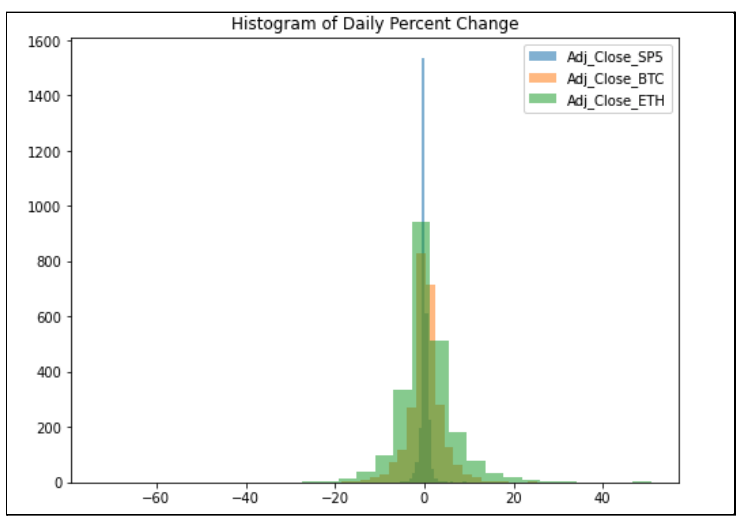


Figure 5 above displays the day-to-day percent change in the price of each asset. It is worth noting that the volatility of cryptocurrencies can be readily viewed in this graphic, as the positive and negative changes in these assets significantly outpace the SP500. Similar conclusions can be drawn from the histograms of SP500, BTC and ETH displayed in Figure 6.

Figure 6. Histogram of Daily Percent Change



Our team employed a number of methodologies to better understand the statistical relationship between the performance of the SP500 and cryptocurrencies and to address the question of whether Crypto Assets Provide a hedge against broader market downturns.

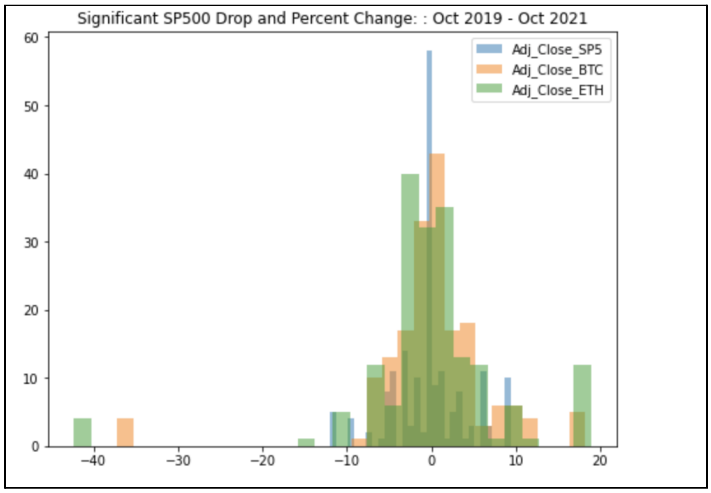
To evaluate these questions, we performed numerous iterations of re-creating the graphs above with various slices of our original dataset. We also found stronger evidence of a statistical relationship between all of our assets utilizing more recent data from the past two years and creating a subset of data for the twenty days following a significant drop in the SP500. The charts below display the correlation of assets in the past two years following a significant drop in the SP500. The correlation matrix (Figure 7)

displays that there is evidence of a statistical relationship between BTC, ETH and DOGE following a significant downturn in the SP500. Additionally, based on the histogram displayed in Figure 8 below, there is evidently a possibility of a massive drop in BTC and ETH, approximately 40%, following an SP500 decline. These factors lead us to conclude that currently, the major cryptocurrencies are correlated with the SP500 and do not currently represent a safe haven from SP500 corrections and broader market downturns.

Figure 7. Correlation Matrix between Assets

	Adj_Close_SP5	Adj_Close_ETH	Adj_Close_DOGE	Adj_Close_BTC
Adj_Close_SP5	1.000000	0.551959	0.506111	0.501037
Adj_Close_ETH	0.551959	1.000000	0.957943	0.964946
Adj_Close_DOGE	0.506111	0.957943	1.000000	0.966486
Adj_Close_BTC	0.501037	0.964946	0.966486	1.000000

Figure 8. Significant Decline Analysis



Our team even created functions to calculate the correlations of each crypto asset and the SP500 in every following a 3 standard deviation drop in the SP500. A portion of these iterative correlations are below, and display that evidence of statistical relationship between the SP500 and the crypto stays consistent across the days following a large drop in the SP500. Our team utilized numerous timeframes for the days following a drop in the SP500 and saw consistent evidence of correlation between the SP500 and the crypto assets.

Figure 9. Sample of Iterative Correlation Analysis

	Day 1:Adj_Close_SP5	Day 2:Adj_Close_SP5	Day 3:Adj_Close_SP5	Day 4:Adj_Close_SP5	Day 5:Adj_Close_SP5	Day 6:Adj_Close_SP5	Day 7:Adj_Close_SP5	8
Adj_Close_SP5	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	
Adj_Close_ETH	0.550860	0.520619	0.510777	0.514657	0.507647	0.488956	0.501028	
Adj_Close_DOGE	0.477501	0.473434	0.480955	0.488159	0.456752	0.421906	0.433174	
Adj_Close_BTC	0.529461	0.495172	0.491731	0.502168	0.479841	0.469774	0.478522	

2.4 Regression Model

Selecting Dependent and Independent Variables

To expand our correlation and exploratory analysis, our team wanted to fit our regression model with variables that would produce relevant conclusions towards understanding the relationship between the S&P and cryptocurrencies. On top of this, we wanted to develop our analysis on how the cryptocurrencies interact. With the goal of maximizing the accuracy of this model, we made a few key decisions on variable characteristics.

The variables we selected for the regression model are a calculation of the weekly percent change for each stock. Using weekly percent change helps us overcome three potential issues in our model. First, using percent change eliminates the large difference in value between our assets and lets us compare the result coefficients. Second, a weekly report will help combat the increased volatility of the cryptocurrencies. Finally, we no longer have to adjust for weekends when the market is closed but cryptos are free to trade.

Model Fitting

Using these characteristics we were able to fit four models. The first two attack the first goal of understanding the relationship between our cryptos and the S&P. Model 1 uses the S&P as the dependent variable and model 2 uses BTC. The results are as follows:

Figure 10. Results from Model 1, S&P Prediction

IVs	Coef
BTC % Change	0.0206
ETH % Change	0.0065
DOGE % Change	-0.0009
R^2 Value	0.0179
n	233

Figure 11. Results from Model 2, BTC Prediction

IVs	Coef
SP5 % Change	0.3858
ETH % Change	0.2907
DOGE % Change	0.0262
R^2 Value	0.1901
n	233

The low R^2 values in Figure 10 and Figure 11 confirm the results of our correlation analysis above. The weekly change in our cryptos is a poor predictor of the S&P 500 and vice versa. This score is especially poor in comparison to our 3rd model, where we bring in Litecoin (LTC), Binance Coin (BNB), and Stellar, and our 4th model where we use only recent data (after 2019).

Figure 12. Results from Model 3, BTC Prediction

IVs	Coef
LTC % Change	0.3608
XLM % Change	0.0205
BNB % Change	0.0289
ETH % Change	0.2397
DOGE % Change	-0.0172
R^2 Value	0.6841
n	160

Figure 13. Results from Model 4, After 2019

IVs	Coef
LTC % Change	0.2919
BNB % Change	0.1170
ETH % Change	0.2825
R^2 Value	0.7862
n	222

Our 3rd and 4th model results in Figure 12 and Figure 13 that the relationship between the crypto markets is much stronger than with the S&P. Specifically, the result of our 4th model shows a strong relationship between weekly performance of Litecoin and Ethereum with performance of Bitcoin, especially recently as cryptocurrencies have become more common.

2.5 Time Series Theory, ARIMA Theory and ADCT test

After our implementation of the regression model, we investigated the potential to predict cryptocurrency asset price with the asset's own previous price data.

To investigate this capability, we need to understand the theories of Time Series and the ARIMA model.

In Time Series, data has 2 variables: time and the variable we want to predict. We observe data in equal intervals, like days, weeks, or months. Our cryptocurrencies satisfy these requirements, and they also have trends – upward and/or downward movement of data. They could be Time Series for our prediction analysis. However, Time Series has to be stationary before being applied in models.

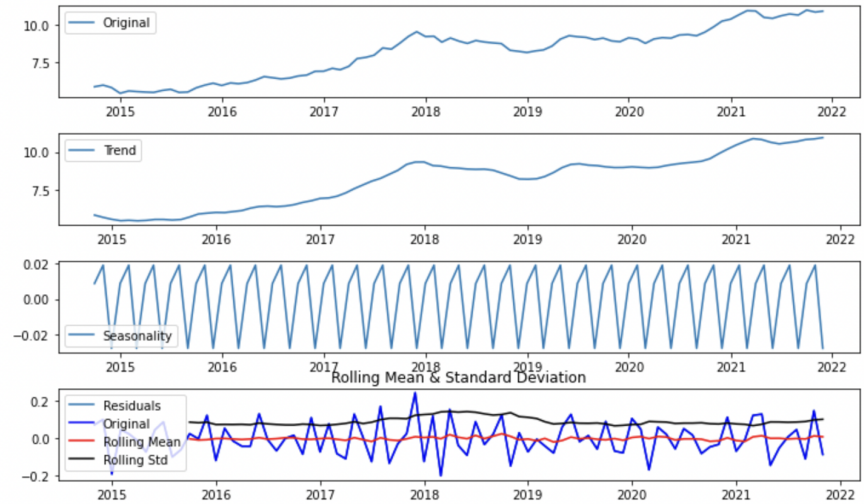
After several tests and trials, we chose to utilize the following data to conduct our analysis and develop our ARIMA model:

- Bitcoin monthly data from 2014-10-01 to 2021-11-28 (12 months as lags and rolling window),
- Ethereum monthly data from 2015-09-01 to 2021-12-06 (12 months as lags and rolling window),
- Dogecoin weekly from 2012-09-15 to 2021-12-06 (12 weeks as lags and rolling window)

'Rolling' means we use every continuous 12 months or 12 weeks data to do the average calculation.

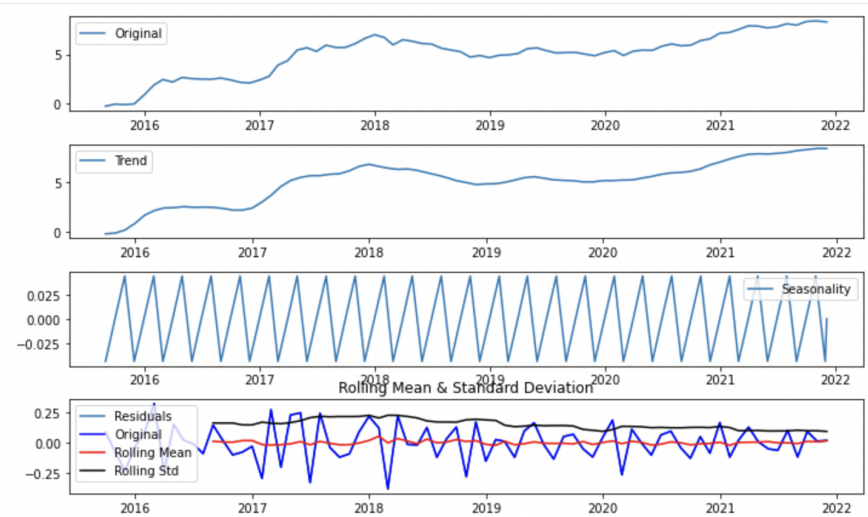
Figures 14-16 are the last test of stationary of Bitcoin, Ethereum and Dogecoin after using 3 methods of transformation (log, exponential decay and time shift):

Figure 14. ADCT Test for Stationary of Bitcoin (BTC)



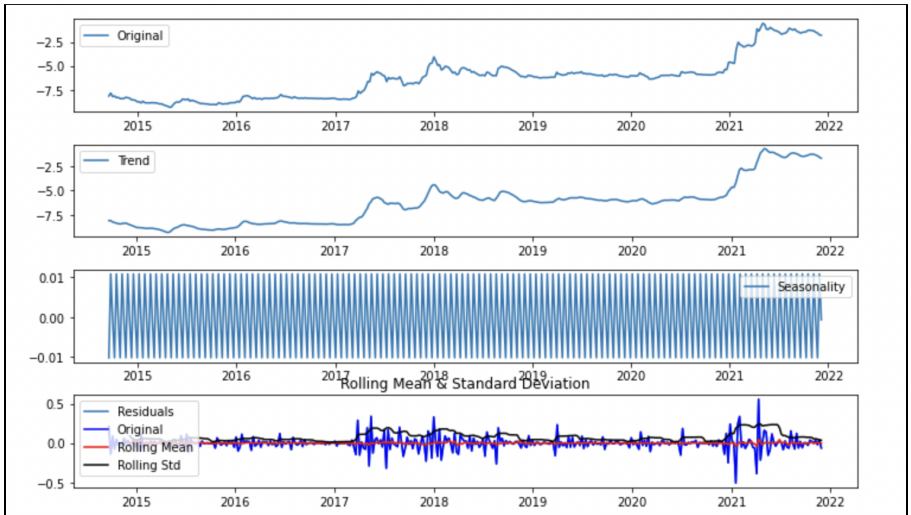

```
Results of Dickey Fuller Test:
Test Statistic                  -6.110299e+00
p-value                          9.367983e-08
#Lags Used                       6.000000e+00
Number of Observations Used      7.800000e+01
Critical Value (1%)             -3.517114e+00
Critical Value (5%)             -2.899375e+00
Critical Value (10%)            -2.586955e+00
dtype: float64
```

Figure 15. ADCT Test for Stationary of Ethereum (ETH)




```
Results of Dickey Fuller Test:
Test Statistic                  -5.607116
p-value                          0.000001
#Lags Used                       6.000000
Number of Observations Used     68.000000
Critical Value (1%)             -3.530399
Critical Value (5%)             -2.905087
Critical Value (10%)            -2.590001
dtype: float64
```

Figure 16: ADCT Test for Stationary of Dogecoin




```
Results of Dickey Fuller Test:
Test Statistic                  -7.941579e+00
p-value                          3.303011e-12
#Lags Used                       1.600000e+01
Number of Observations Used      3.590000e+02
Critical Value (1%)             -3.448697e+00
Critical Value (5%)             -2.869625e+00
Critical Value (10%)            -2.571077e+00
dtype: float64
```

Conclusions from the ADCT Tests for Stationary of BTC, ETH and Dogecoin

We assume the null hypothesis that the time series is not stationary. The p-values are all very low, and the rolling mean and rolling standard deviation lines are parallel to x-axis, which means that we can reject the null hypothesis that the time series is not stationary. In another aspect, if the Test statistics is less than the critical values, we can also reject the null hypothesis and say that the series is stationary. We can still try utilizing the ARIMA model.

Theory of ARIMA

ARIMA is short for Auto-Regressive Integrated Moving Average. 'Auto-Regressive' refers to the mean lags of the stationaried series. 'I' means 'Integrated', a series which needs to be differenced to be made stationary. 'Moving Average' refers to the mean lags of the forecast errors.

In the ARIMA model, we need to determine 3 key parameters: p, d, q.

- p = the number of autoregressive terms
- d = the number of nonseasonal differences
- q = the number of moving-average terms

When d = 0, there is no difference, the differenced (stationarized) series is the original series. When d = 1, we use the original series minus the data of yesterday (or last week, or last month, it depends on the data interval you choose) to create the new differenced series. Here, we use d = 1. That is, we calculate the difference between last month's price and this month's price. Then, we calculate the difference between the price from two months ago and last month's price. This calculation continues to be iterated, as illustrated in Figure 17 below.

Figure 17. ARIMA Forecasting Equation

ARIMA forecasting equation

- Let Y denote the *original* series
- Let y denote the *differenced* (stationarized) series

No difference (d=0): $y_t = Y_t$

First difference (d=1): $y_t = Y_t - Y_{t-1}$

Second difference (d=2): $y_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2})$

$= Y_t - 2Y_{t-1} + Y_{t-2}$

Figure 18. Forecasting Equation for y

Forecasting equation for *y*

$$\hat{y}_t = \underbrace{\mu}_{\text{constant}} + \underbrace{\phi_1 y_{t-1} + \ldots + \phi_p y_{t-p}}_{\text{AR terms (lagged values of } y)} \underbrace{- \theta_1 e_{t-1} \ldots - \theta_q e_{t-q}}_{\text{MA terms (lagged errors)}}$$

By convention, the AR terms are + and the MA terms are –

Not as bad as it looks! Usually $p+q \leq 2$ and either p=0 or q=0 (pure AR or pure MA model)

In the forecasting phase, we transform and get a forecasting equation using p, d, q and the differenced series in Figure 17.

Figure 18 contains the equation interpreted by Robert Nau, which is very similar to our multi-regression functions.[9]

Our team created graphs to estimate p, d, and q based on the visualization's content and compared the Residual Sum of Squares to compare different model versions.

From Figure 19 on the following page, which displays plots of the autocorrelation function (ACF) and partial autocorrelation function for BTC, we see that the curve touches y=0 and it crosses the first line at x=2 or 1. Thus, from theory, Q = 2 or 1 From the PACF graph, we see that the curve touches the y=0.0 line and crosses the first line at x=2 or 1. Thus, from theory, P = 2 or 1. We use the same theory to determine the Q and P for ETH and DOGE (Figures 20 and 21).

Figure 19: Use ACF to Find q and Use PACF to find p for BTC

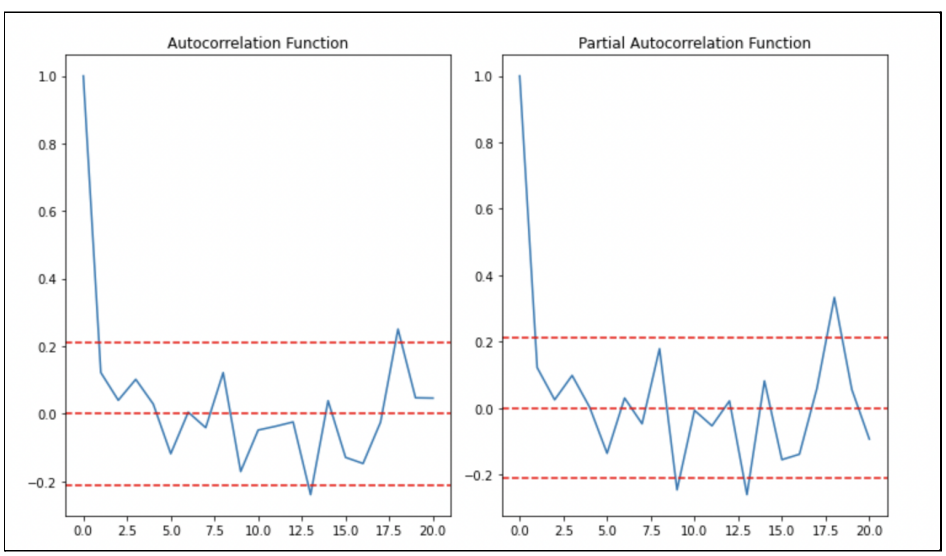


Figure 20. Use ACF to Find q and Use PACF to Find p for ETH

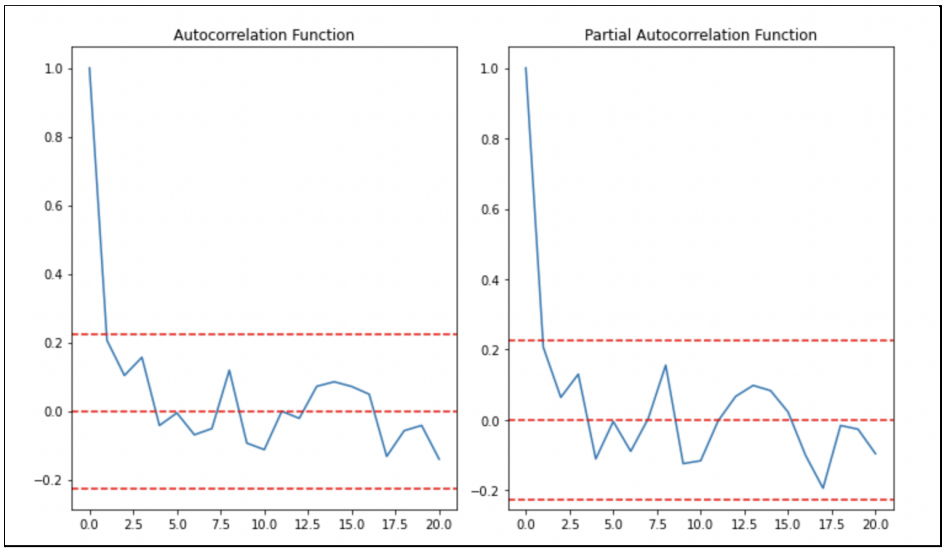
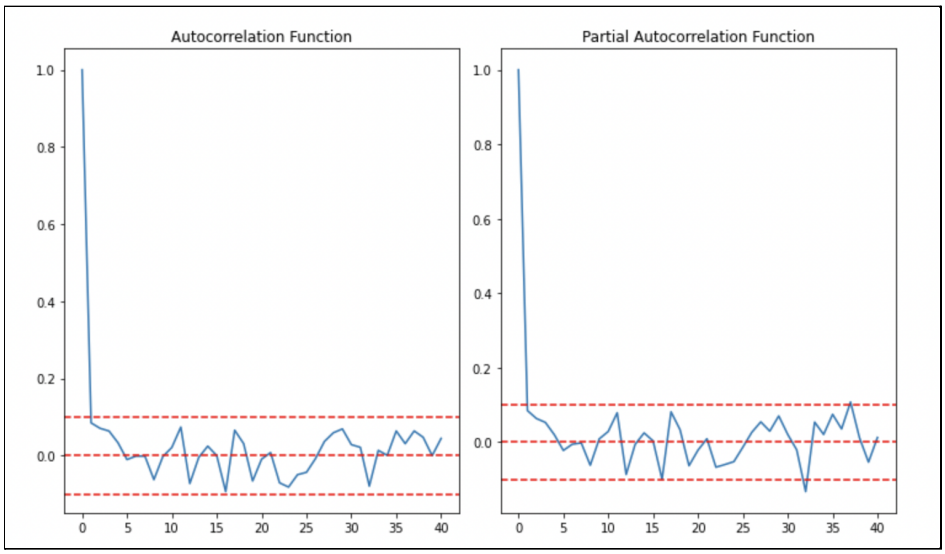


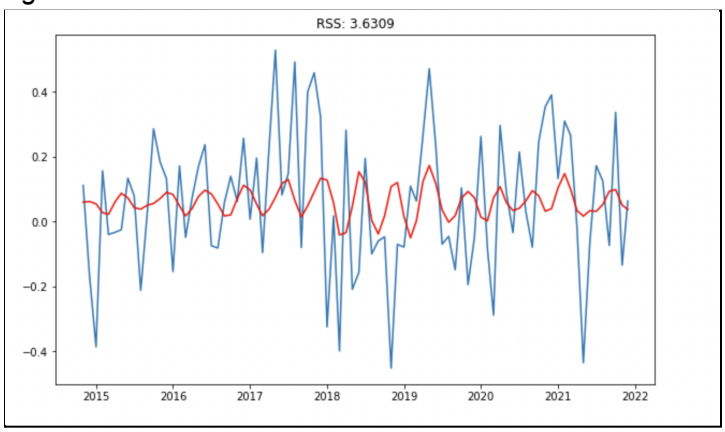
Figure 21. Use ACF to Find q and Use PACF to Find p for DOGE



2.7 ARIMA Model Fit

ARIMA is AR + I + MA. Before we utilize an ARIMA model, we will first utilize AR and MA models and assess their fit. For every cryptocurrency, the ARIMA model yielded smaller RSS than individual AR & MA models, so our Time Series will fit better on an ARIMA model.

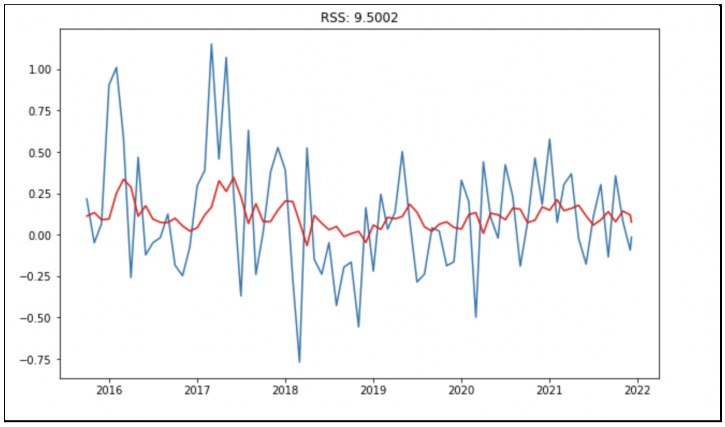
Figure 22. ARIMA Model Fit for BTC



In the BTC ARIMA model, we take order = (2,1,2)*, and it has RSS = 3.6309, suggesting that the fit between our BTC ARIMA model and BTC time series model are closely fitted. (Smaller RSS values indicate closer fit.)

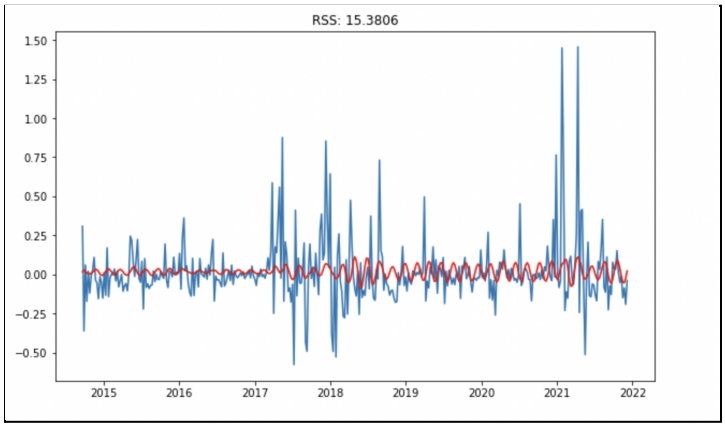
*(order = (q,d,q))

Figure 23: ARIMA Model Fit for ETH



In the ETH ARIMA model, we take order = (2,1,1), and it has RSS = 9.5002, which is sufficient for our time series model.

Figure 24. ARIMA Model Fit for DOGE



In the DOGE ARIMA model, we take order = (2,1,2), and it has RSS = 15.3809, which is large, but we still utilize a full ARIMA model for our time series forecasting.

2.8 ARIMA Model Prediction

We developed ARIMA models for each of our cryptocurrency assets (Figures 25-27). The BTC and ETH ARIMA models fit their respective time series plots closely; the DOGE ARIMA model, however, is less accurate. This may be due to the fact that DOGE has brief periods of very large price movements but is largely stable outside of these spikes.

Figure 25. ARIMA Model for BTC (Transformed Back from Log)

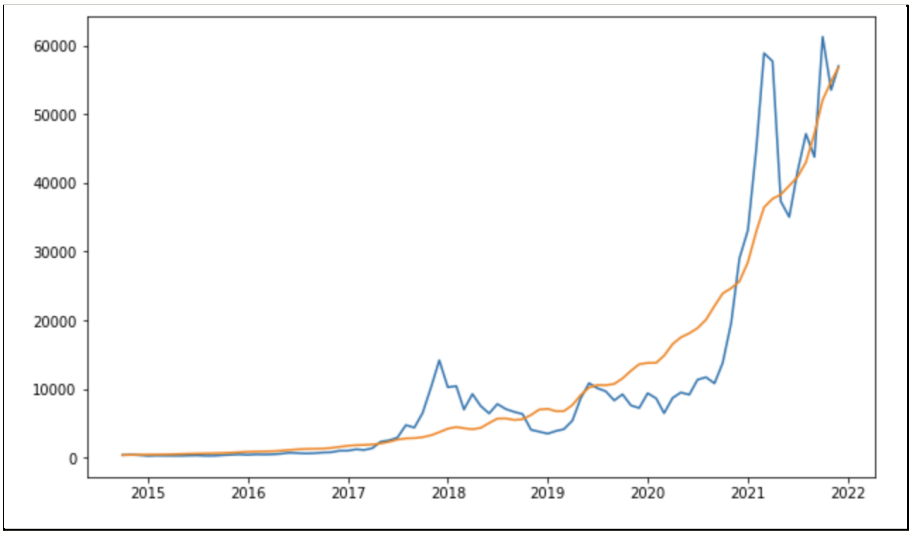


Figure 26. ARIMA Model for ETH (Transformed Back from Log)

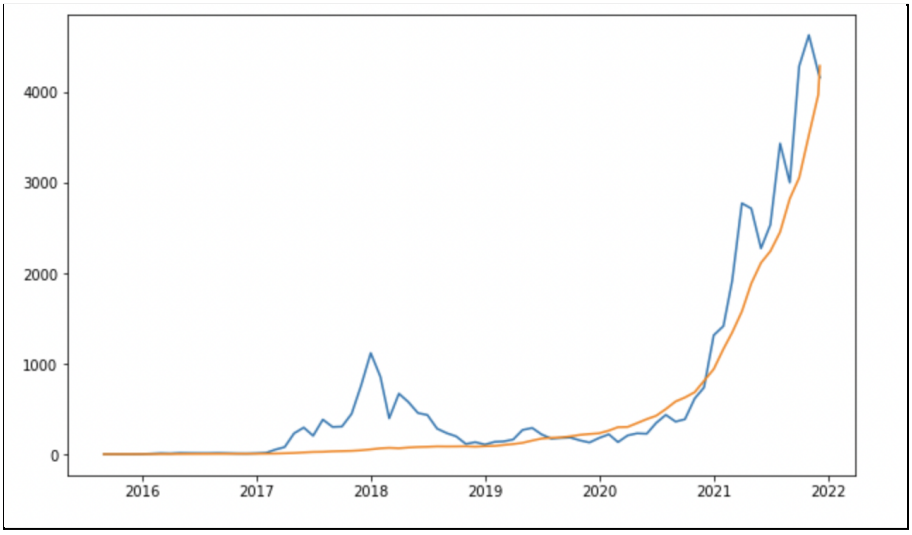
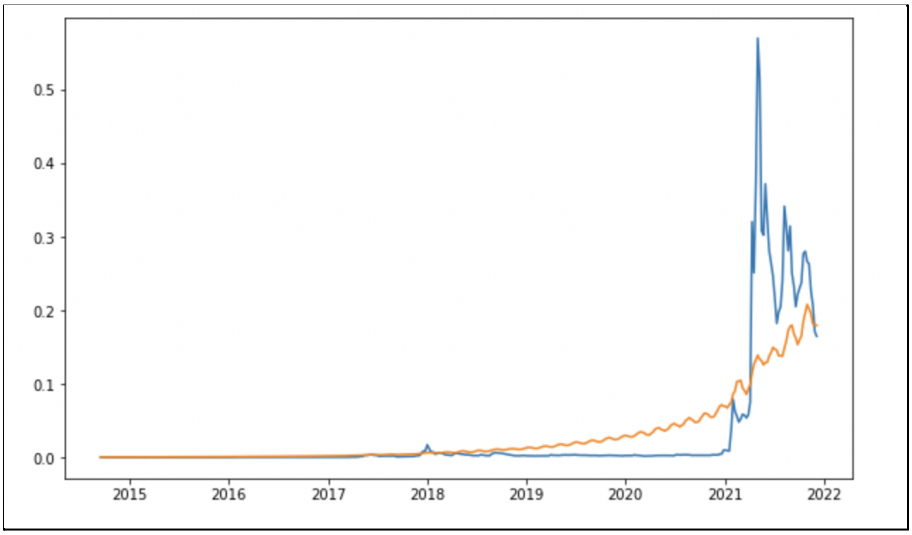


Figure 27. ARIMA Model for DOGE (Transformed Back from Log)



2.9 Asset Performance Forecasting with ARIMA Models

We extrapolated our ARIMA models for each cryptocurrency asset to forecast asset performance over the next four and a half years. Based on our ARIMA models for each asset, our forecasts suggest that all of these assets will likely continue to attain higher prices, but our 95% confidence intervals suggest that there is still a possibility of a decline in the price of all of these assets.

Figure 29: Forecasting for BTC by ARIMA Model

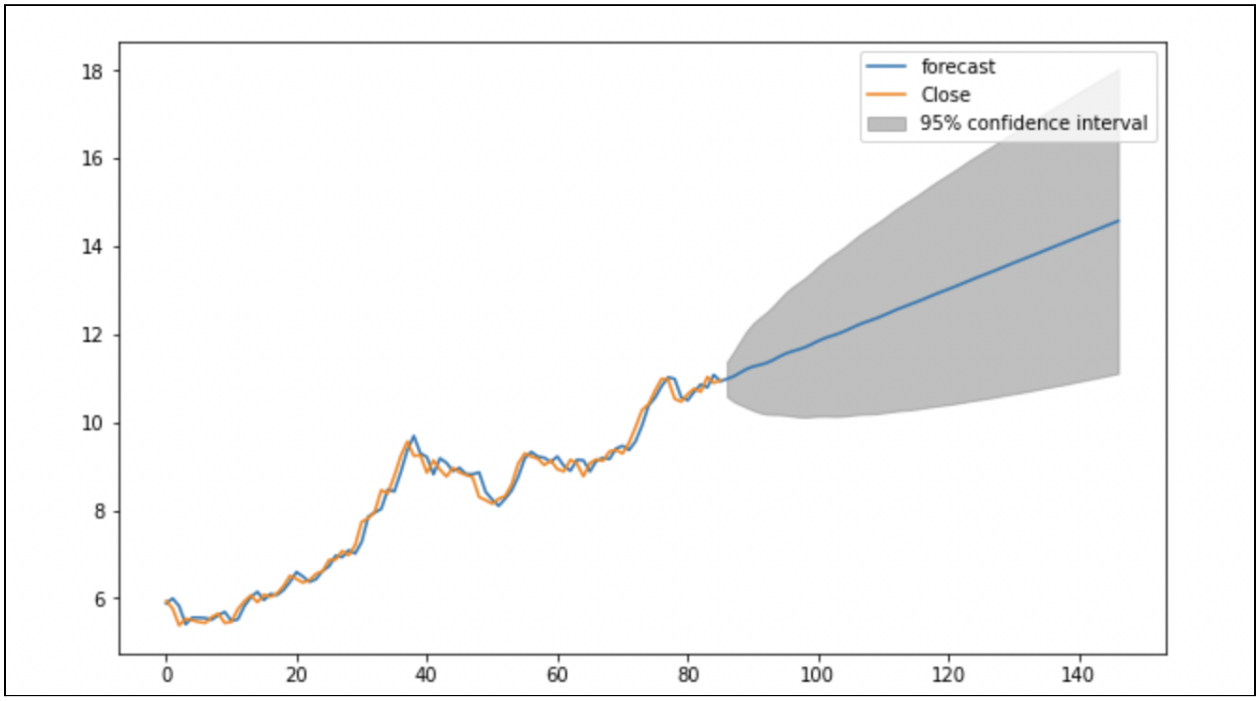


Figure 30: Forecasting for ETH by ARIMA Model

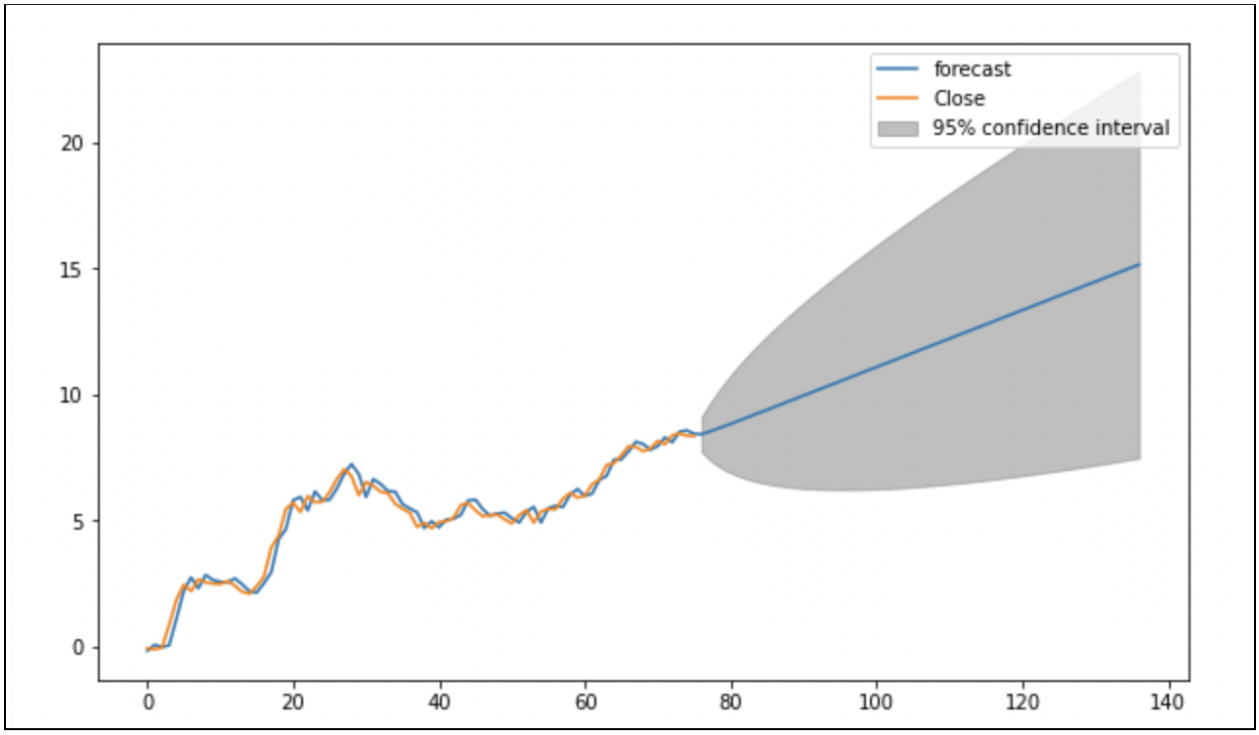
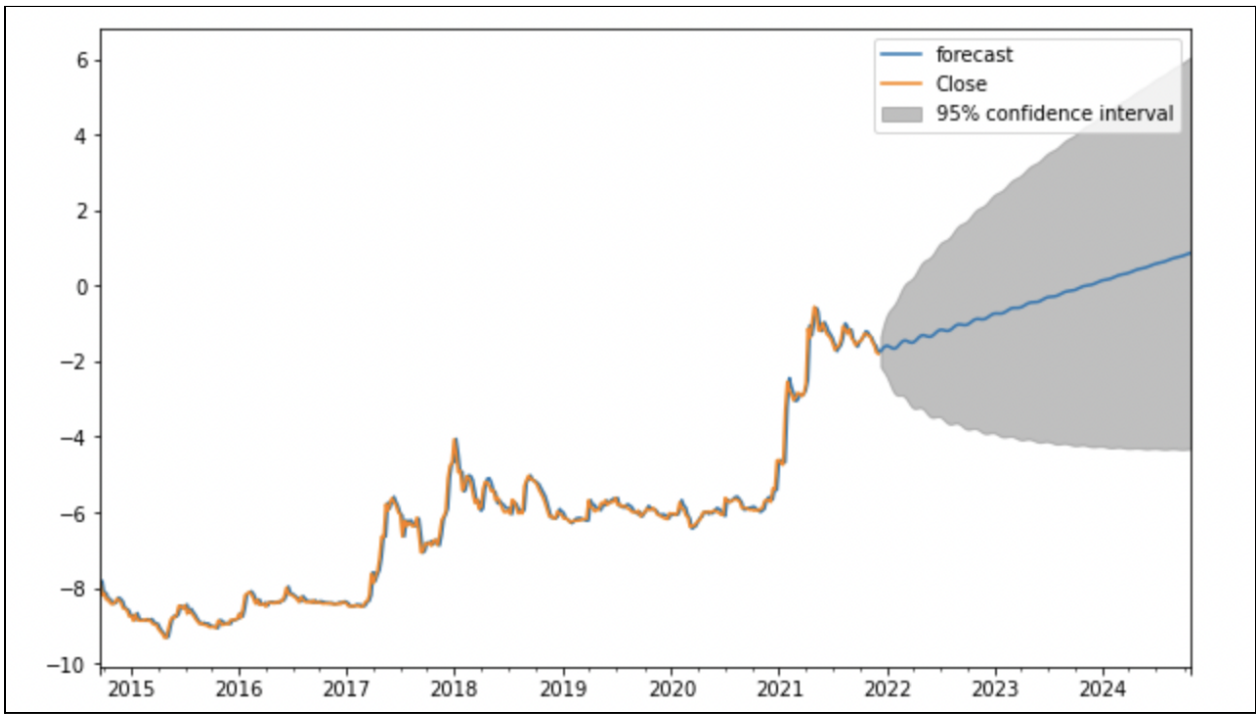


Figure 31. Forecasting for DOGE by ARIMA Model



3. Conclusion

Our analysis demonstrates that there is greater volatility within cryptocurrency assets than within the market as a whole. Our time series analysis of each asset illustrates a clear upward trend in asset performance. However, our percent change analysis reveals larger-magnitude price fluctuations among newer coins ETH (between -30% and +30%) and DOGE relative to BTC (between -20% and +20%) and SP500 (between -10% and +10%). Despite disparities in asset stability, our correlation analyses suggest that there is evidence of a weak statistical relationship over time between the performance of the S&P 500 and large cryptocurrencies. However, this statistical relationship appears to *strengthen* in the aftermath of an SP500 downturn. Thus, because cryptocurrency assets tend to be more strongly correlated with SP500 performance during a downturn, cryptocurrency assets would *not* provide an effective hedge against broader market slumps, as they might be expected to decline in value in parallel with the market.

Our regression analysis confirms that, broadly speaking, the performance of cryptocurrency assets are much more strongly correlated with the performance of *other cryptocurrency assets* than with the S&P 500, despite the tendency of cryptocurrency assets to fall alongside the larger economy. Our ARIMA models were most accurate when applied to older cryptocurrencies, such as Bitcoin, and decreased in accuracy for younger assets. However, the ARIMA model is promising, and may provide accurate forecasts of cryptocurrency market performance in the aggregate.

References

1. Conway L. "10 Important Cryptocurrencies Other Than Bitcoin." Investopedia. https://www.investopedia.com/tech/most-important-cryptocurrencies-other-than-bitcoin/#citation-1 . 30 Nov 2021. Accessed 05 Nov 2021.
2. "What is market cap?" Coinbase. https://www.coinbase.com/learn/crypto-basics/what-is-market-cap. Accessed 05 Dec 2021.
3. d Best R. "Dailiy Bitcoin (BTC) market cap history up until November 22, 2021." Statista. https://www.statista.com/statistics/377382/bitcoin-market-capitalization/. 23 Nov 2021. Accessed 05 Dec 2021.
4. Ossinger J. "Crypto World Hits #3 Trillion Market Cap as Ether, Bitcoin Gain." Bloomber. https://www.bloomberg.com/news/articles/2021-11-08/crypto-world-hits-3-trillion-market-cap-as-ether-bitcoin-gain. 08 Nov 2021. Accessed 05 Nov 2021.
5. "Bukele says El Salvador 'bought the dip' and purchased 100 extra bitcoins." Reuters. https://www.reuters.com/markets/us/bukele-says-el-salvador-bought-dip-purchased-100-extra-bitcoins-2021-11-26. 26 Nov 2021. Accessed 05 Dec 2021.
6. Folger J. "Investing in Forex vs. Stocks." Investopedia. https://www.investopedia.com/articles/forex/11/forex-or-stocks.asp. 24 Aug 2021. Accessed 05 Dec 2021.
7. Alpert G. "Top S&P 500 Stocks by Index Weight." Investopedia. https://www.investopedia.com/top-10-s-and-p-500-stocks-by-index-weight-4843111 03 Sep 2021. Accessed 05 Dec 2021.
8. Swenson Sam. "The S&P 500 Index." The Motley Fool. https://www.fool.com/investing/stock-market/indexes/sp-500/. 12 Nov 2021. Accessed 05 Dec 2021.
9. Nau R. "Lecture notes on forecasting" https://people.duke.edu/~rnau/Slides_on_ARIMA_models--Robert_Nau.pdf. 2014.
10. Chatterjee A."Time Series For beginners with ARIMA" Ahttps://www.kaggle.com/freespirit08/time-series-for-beginners-with-arima. July 2018.