Time Series Analysis of Monero Price

Aishwarya Dev
Department of Computer Science
UCLA
aishwaryadev30@g.ucla.edu
305515097

Abstract

Cryptocurrencies are known widely for their anonymity and privacy. But it has been extensively shown that even with its reputation of untraceability, the Blockchain is not entirely secure. Bitcoin, the most widespread cryptocurrency as of now has been under skepticism over its security. The security vulnerabilities of Bitcoins have been outlined prominently by a pseudonymous author under the name Nicholas Van Saberhagen in their research paper. van Saberhagen (2018). Such security vulnerabilities are addressed primarily by privacy coins which add an extra layer of security on top of a transaction. Monero belongs to the class of privacy coins or Advanced Encryption Cryptocurrencies (AECs) and is in fact one of the most popular AECs. Known for its security and anonymity, Monero is a controversial currency due to its growing popularity among users of the dark web. While some users leverage the capabilities of Monero to seek privacy, there are a good number of nefarious users who exploit Monero to mask their illicit activities on the dark web. One disturbing example of Monero's notoriety is implicit from the increasing trend in Monero as a preferred cryptocurrency on ransomware sites on the dark web. A lot of these sites have increasingly started to transact in Monero over Bitcoin which is surprising and alarming at the same time. There are instances of hackers incentivizing Monero transactions through discounts on ransom amount over Bitcoin. Due to its notoriety, Monero is currently not supported on platforms like Coinbase to avoid scrutiny from law enforcement agencies. Needless to say, this shrouds Monero in controversy and keeps the law enforcement agencies on high alert over its increased use and a predictive time series analysis of this controversial currency could help us potentially discover trends to model the volatile behavior of cryptocurrencies in general and Monero in particular.

1 Introduction

Cryptocurrencies are a type of digital currency that have been increasingly emerging as an alternative form of payment. One of the most critical features of a cryptocurrency is its decentralization and absence of any regulating bodies such as banks or other financial institutions. Cryptocurrencies use a secure distributed ledger known as a Blockchain to keep track of transaction records through a process known as Mining. Mining allows users to reach secure, robust consensus for each transaction. Mining also introduces wealth in the form of new units of currency. Cryptocurrencies lack a central authority to mediate transactions because they are inherently peer-to-peer systems that reply on miners to validate transactions. Mukhopadhyay et al. (2016)

Bitcoin is the most popular and widely used cryptocurrency as of now. The popularity of Bitcoin is evident from its synonymous usage with the term cryptocurrency among a lot of people. The bitcoin is essentially a network of users that communicate with each other using the bitcoin protocol over the Internet. The bitcoin protocol is available in the form of an open source software application and allows users to store and transfer Bitcoins for selling and buying goods, or to exchange bitcoins with other currencies / cryptocurrencies. Bitcoins are issued through a process of handling transactions over the network known as Bitcoin mining. Vranken (2017)

Despite its massive popularity, Bitcoin is not immune to security vulnerabilites. Several works have highlighted these vulnerabilities including Urquhart (2016), Nadarajah & Chu (2017) and van Saberhagen (2018). van Saberhagen (2018) written by a pseudonymous author under the name Nicholas Van Saberhagen, prominently outlines Bitcoin's security issues and introduces CryptoNote, a novel cryptocurrency protocol that led to the creation of Monero.

Monero belongs to a class of privacy-centric cryptocurrencies known as Advanced Encryption Cryptocurrencies (AECs) and is in fact one of the most popular AECs. Monero was built on the foundation of addressing Bitcoin's traceability issue which was called a "critical flaw" by the author in van Saberhagen (2018). With its roots in anonymity and security, Monero has been known for being practically untraceable. Monero uses chaff coins called "mixins" along with the actual coins which allows the users to privately and securely transact while obscuring their transaction details. In addition, it also uses technologies like stealth addressing and ring signatures for maximum anonymity. While there are works citing the loopholes within Monero's mixin sampling strategies such as Möser et al. (2017) and other potential breach strategies, it is still one of the most valued privacy currencies today, especially among dark web users. It also has the third largest community of developers after Bitcoin and Ethereum.

While some users leverage the capabilities of Monero to seek privacy, there are a good number of nefarious users who exploit Monero to mask their illicit activities on the dark web. One disturbing example of Monero's notoriety is implicit from the increasing trend in Monero as a preferred cryptocurrency on ransomware sites on the dark web. A lot of these sites have increasingly started to transact in Monero over Bitcoin which is surprising and alarming at the same time. Roughly 10 - 20 per-cent of all ransom paid on the dark web is in Monero and this number is estimated to rise to over 50 per-cent by the year 2023. There are instances of hackers incentivizing Monero transactions through discounts on ransom amount over Bitcoin. Due to its notoriety, Monero is currently not supported on platforms like Coinbase beacuse of the intensive scrutiny and discretion from law enforcement agencies.

2 Data

The project data has been collected for the daily price of Monero (in USD) over the last 6 months from investing.com. `https://www.investing.com/crypto/monero/historical-data`. There are roughly 184 data points ranging from May 29, 2022 to Nov 29, 2022. This data will be used for the time-series analysis of Monero Price and help us forecast the future price from the theoretical analysis techniques.

Figure 1 shows the source data for the project. On the y-axis is the Monero price (in USD) and on the x-axis are the dates between May 29, 2022 and Nov 29, 2022 sampled daily.

3 Analysis

3.1 Stationarity

A stationary time series is a process for which the statistical properties or moments (e.g., mean and variance) do not vary in time. A time series is stationary if it does not have any **trends** or **seasonal effects**. If it does have a trend and is not stationary, it needs to be detrended before further analysis. Generally speaking, for predictive or forecasting purposes, it is important to assume independence between data points.

For my analysis, I used two approaches for stationarity analysis.

1. Splitting the dataset into two equal groups and comparing their respective means and variances. Figure 2 shows how the variance varies widely across the two groups. This suggests the possibility of non-stationarity at the most fundamental level of analysis.
2. To solidify the stationarity hypothesis, I used the Augmented Dickey-Fuller Test to check the P-value. The ADF test statistic was -2.713149 and P-value is 0.071790. This P-value is pretty close to the threshold for null hypothesis which is 0.05. However, this closeness does not approximate to statistical significance. Therefore, it is safe to say that the time series is non-stationary as the p-value is not small enough to reject the null hypothesis.

3.2 Data Cleanup

There are some outliers in the data which represent the natural variation. For the most accurate statistical representation of the data, it is ideal to eliminate outliers. However, removing outliers tampers with the pristineness of data hence an evidently better alternative is to clip the outliers outside some minimum and maximum bounds.

Figure 3 shows the box plot of outliers in my data. The minimum and maximum bounds are respectively Q1 - 1.5*IQR and Q3 + 1.5*IQR where Q1 and Q3 have been set to the 25th and 75th percentile of the data and IQR represents the inter-quartile range. There are around 15 outliers in this range and they are not too extreme. Given the limited number of data points, I decided to clip the data outside the 5th and 95th percentile. This effectively reduced the standard deviation from 18.57 to 17.08. Figure 4 and Figure 5 show the price histograms before and after the clipping.

3.3 ACF and PACF of data

The autocorrelation function (ACF) and partial autocorrelation function (PACF) of the time series help analyze the ARMA model/s selection for forecasting through parameter determination.

Figure 6 shows the ACF and PACF plots of the original time series. the ACF plot shows a decaying approximately sinusoidal pattern while the PACF plot shows a significant spike at lag 1 and none thereafter. This appears to be an AR process.

3.4 Regression Curve Fitting

After establishing the non-stationarity of the time series, it is important to detrend it to make it stationary. Regression analysis helps detrend the data by fitting a curve that best approximates it and subtracting the differences from the original data to generate a residual series which is detrended.

I used Ordinary Least Squares (OLS) regression analysis to fit a linear curve through the series. Then I generated the residual series which is seemingly smoother. Figure 7 and Figure 8 show the regression curve fitting and residual series respectively. The AIC and BIC values were 1603 and 1609 respectively.

3.5 Spectral Analysis

After detrending the linear trend, I plotted the periodogram of the time series to identify any seasonal patterns in the data. Figure 9 shows the periodogram of the detrended data and it appears to be pretty smooth, indicating no seasonal trends or cycles in the data.

3.6 First Differencing

While regression modeling for detrending is clearly superior for linearly increasing data, it becomes pretty wobbly when the data becomes stochastic as is the case with financial data. For such data, there is a multi-variate dependency that requires more complex regression analysis. For preliminary modeling, detrending followed by differencing evidently does a better job.

Figure 10 shows the first differenced series. After differencing the P-value becomes 20e-27 which is sufficient to pass the ADF stationarity test. The mean becomes approximately 0. At this pint, the series is pretty stationary so we plot the ACF and PACF plots of the first differenced series.

Figure 11 shows the ACF and PACF plots of the first differenced series. There are no evident spikes outside the cut-offs which makes it difficult to say anything definitively so I took the second differences to be more conclusive.

3.7 Second Differencing

Figure 12 shows the second differenced series. Figure 13 shows the ACF and PACF plots of the second differenced series. From the ACF and PACF of the second differenced TS, it can be seen that ACF has a spike at lag 1 outside the cutoff and no lags thereafter whereas PACF has 3 lags outside cutoff and no lags thereafter which means the model that should give the best fit should be an ARIMA(3, 2, 1).

4 Results

4.1 Evaluation Metric

In order to select the best ARIMA model for forecasting, I used Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) as evaluation criteria. Figure 14 shows the results obtained for different projected ARIMA models in terms of AIC and BIC values. ARIMA(3, 2, 1) supersedes all other models in our evaluation which validates my hypothesis for model selection.

4.2 Selected Best Fit Model

ARIMA(3, 2, 1) was selected as the best fit model. Figure 15 shows the standardized residuals, ACF of residuals, Normal Q-Q plot and p-values of Ljung box statistic for ARIMA(3, 2, 1). The ACF of residuals is like white noise. The Q-Q plot indicates an almost accurate fit. The p-values in the Ljung box statistic are pretty significant. Overall, the selected model is theoretically a good fit for the data.

4.3 Forecasting

Forecasting was done on 6 months data from Apr 29, 2022 to Oct 29, 2022 leaving November out as a test set. Here too, ARIMA(3, 2, 1) gave the best fit. Figure 16 shows the forecast. The predictions looked promising but were entirely off when compared against the actual price data for the month of November. This difference is illustrated in Figure 17 and Figure 18.

The pink line in Figure 17 shows the part of the actual data that was split into training and test for forecasting. The left portion of the split is the data common between both datasets, the right portion is the actual data for November.

The left portion of the split in Figure 18 shows April data which may be ignored for the current analysis. The right portion of the split shows the predicted price data for the month of November.

Clearly, the predicted data does not model the actual data. This is expected given the highly stochastic and volatile nature of financial data and the plethora of factors influencing it.

5 Conclusion and Future Work

I performed time series analysis to forecast Monero prices using various analysis techniques. These techniques helped me arrive at a theoretically accurate ARIMA model for prediction as evidenced by the results. However, the prediction did not model the actual data because of the extremely volatile and stochastic nature of financial data. The problem gets even more compounded considering that cryptocurrency prices are influenced by a multitude of social, political, economical and global factors which are difficult to model through preliminary analysis. This suggests that the existing theoretical analysis techniques are not equipped to approximate this stochasticity at the level of univariate time series analysis. A more rigorous and thorough analysis involving multivariate data, complex regression analysis and more sophisticated modeling could potentially provide more valuable insights and better predictions.

6 Figures

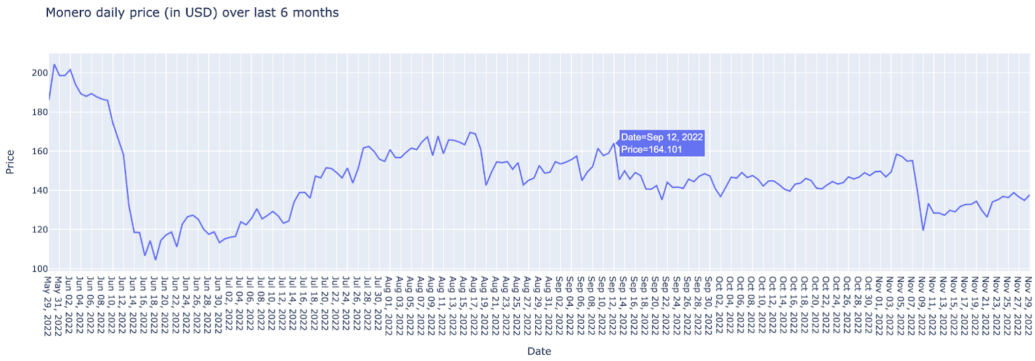

Figure 1: Monero price data (USD) from May 29, 2022 to Nov 29, 2022

Mean 1: 143.699500	Mean 2: 148.402108
Variance 1: 71.889429	Variance 2: 607.694940

Figure 2: Group mean and variance (1:1 split)

Figure 3: Box plot of Outliers

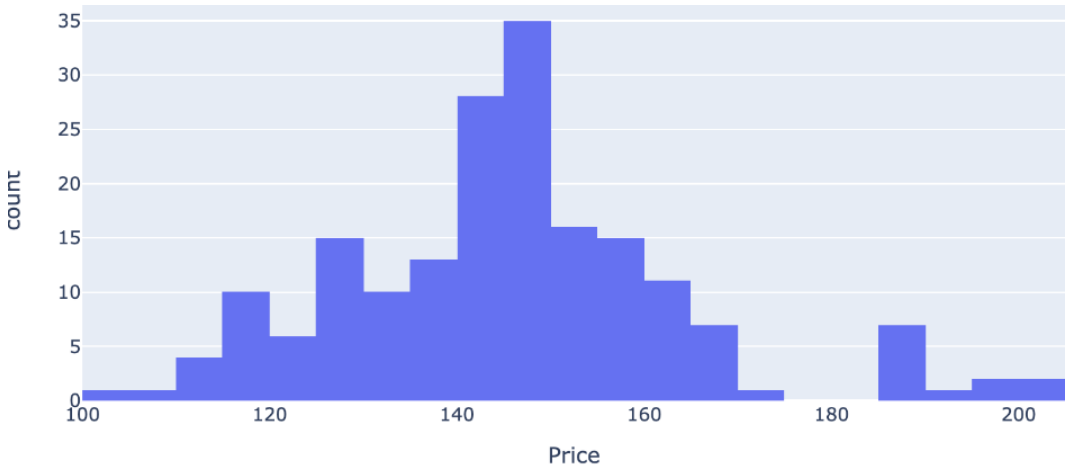

Figure 4: Histogram before clipping outliers

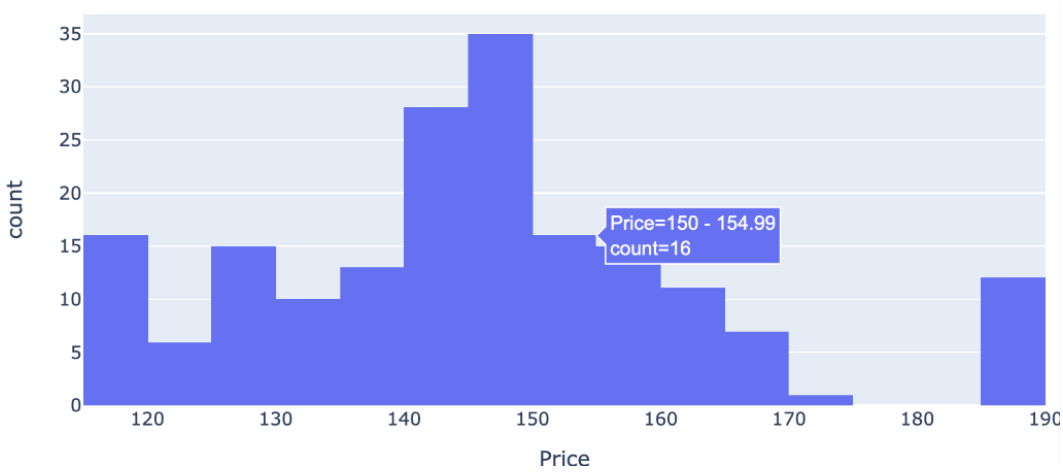

Figure 5: Histogram after clipping outliers

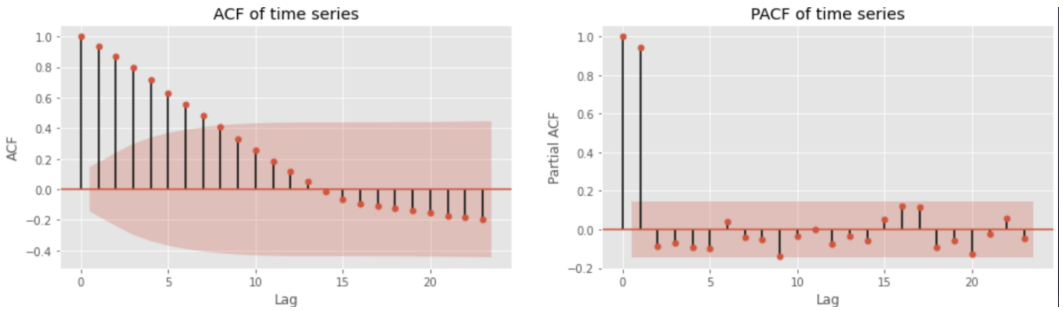

Figure 6: ACF and PACF of the data

Figure 7: OLS Regression curve fitting

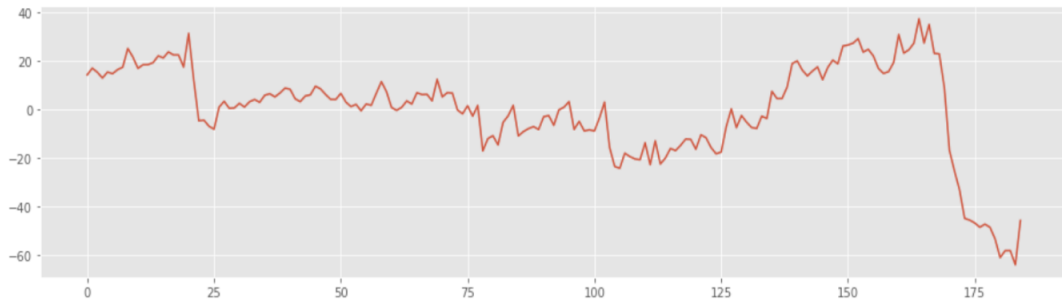

Figure 8: Detrended data

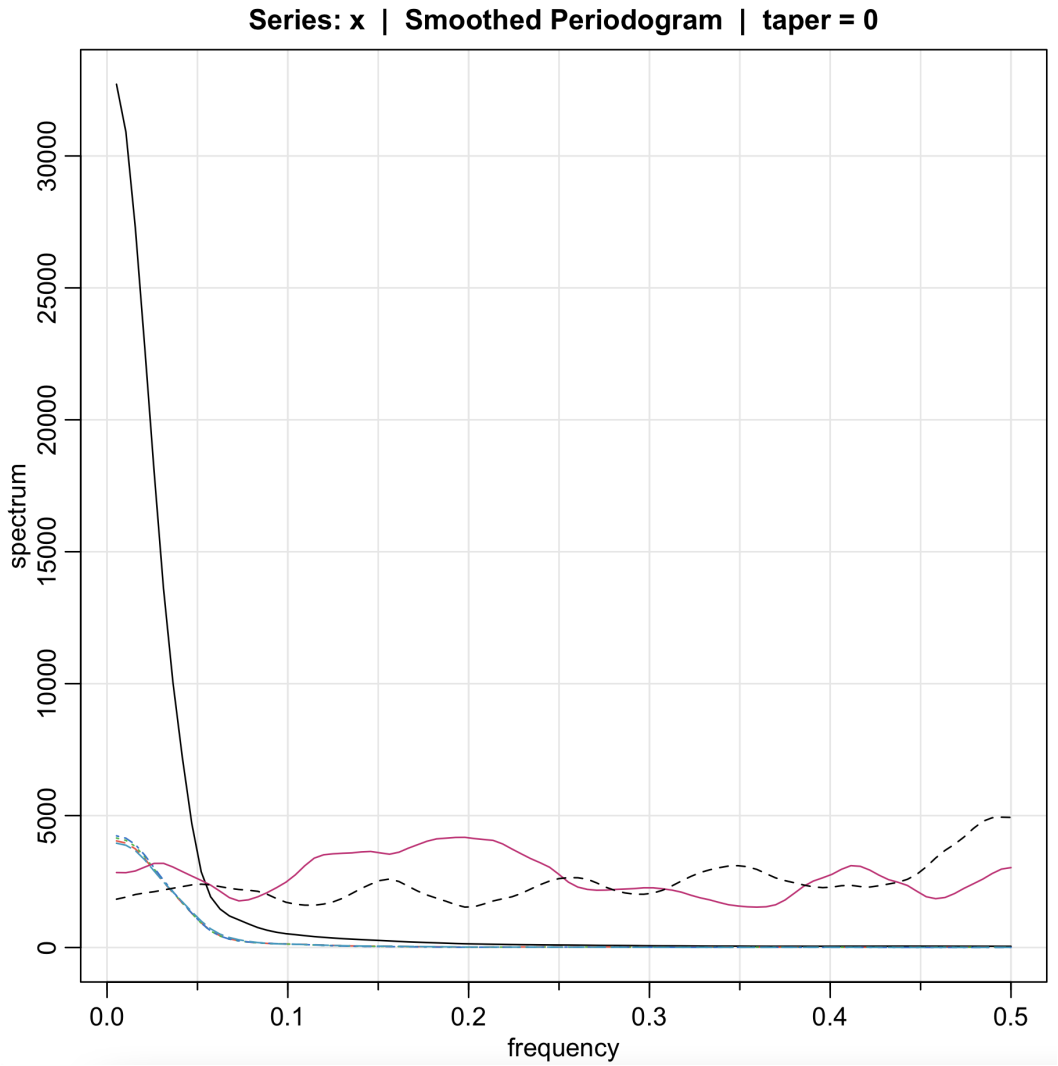

Figure 9: Smoothed Periodogram of detrended data

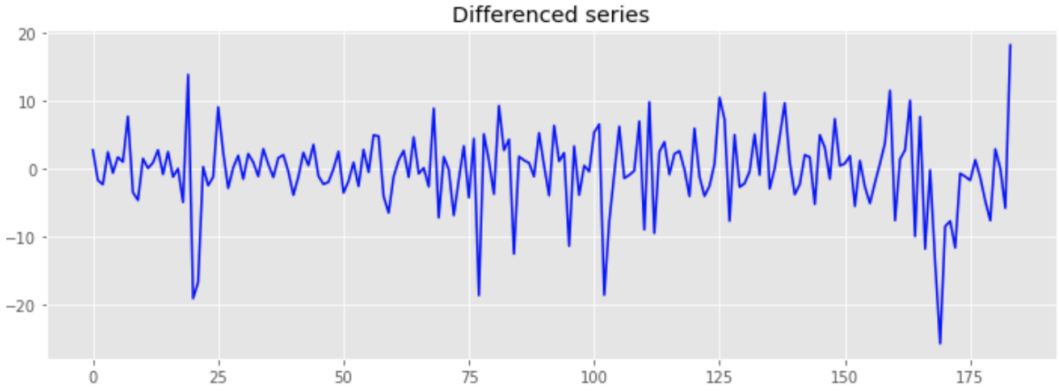

Figure 10: First Differenced Series

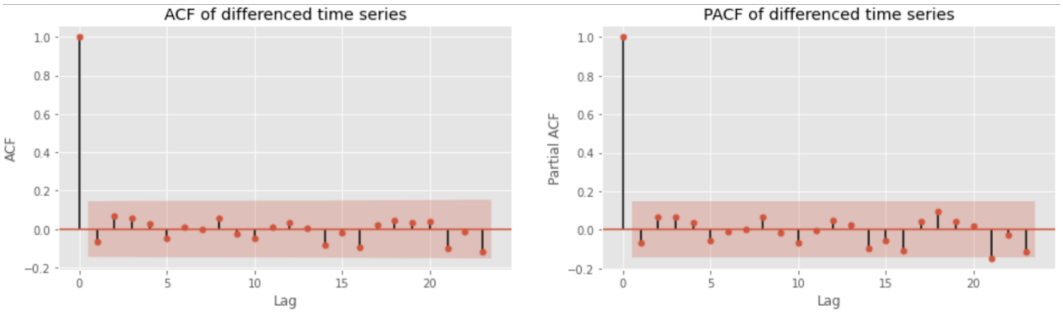

Figure 11: ACF and PACF of first differenced series

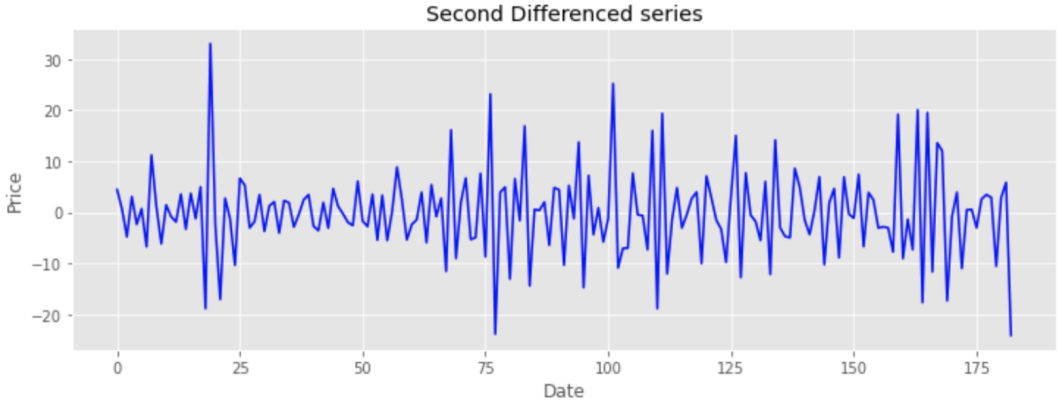

Figure 12: Second Differenced series

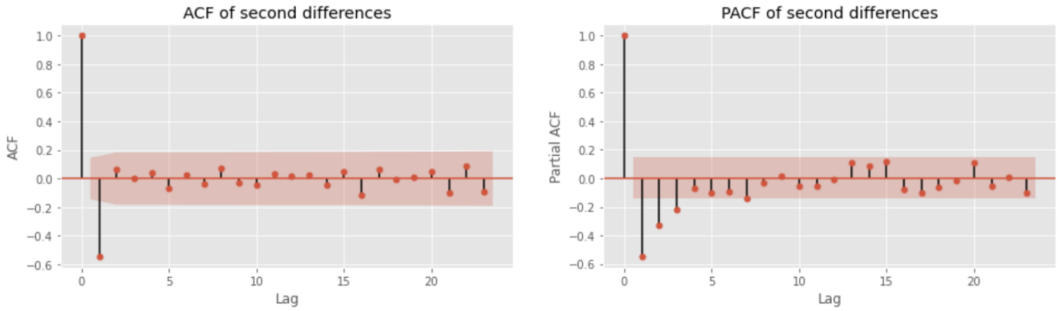

Figure 13: ACF and PACF of second differenced series

ARIMA(p, d, q)	AIC	BIC
ARIMA(3, 2, 1)	1184.35	1203.61
ARIMA(3, 1, 1)	1184.98	1204.27
ARIMA(1, 0, 7)	1195.53	1227.74
ARIMA(1, 0, 8)	1195.22	1230.65

Figure 14: Comparison between different projected ARIMA models

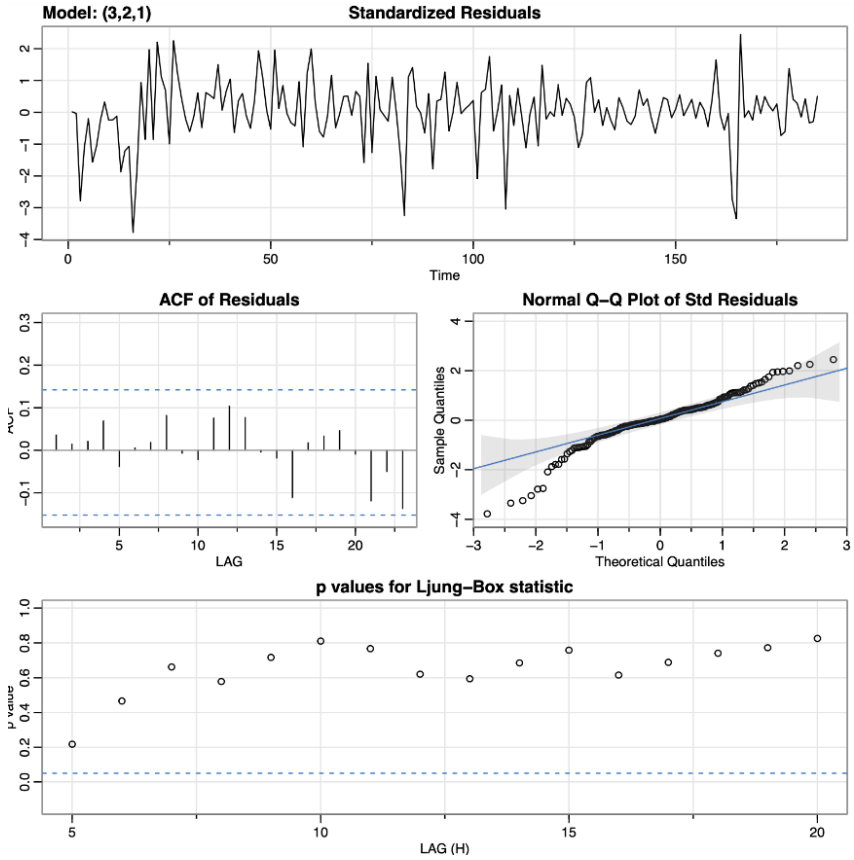

Figure 15: ARIMA(3, 2, 1) results

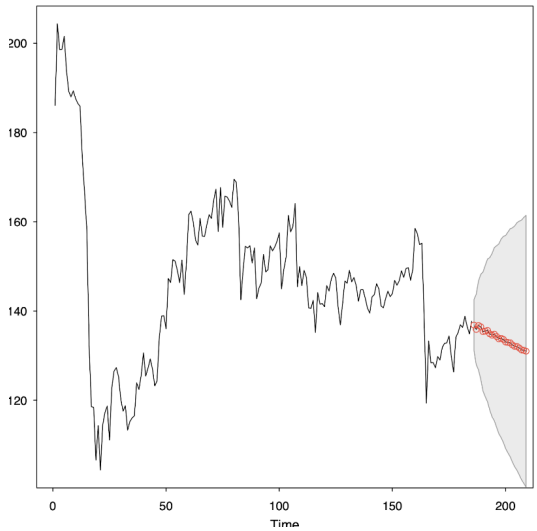

Figure 16: ARIMA(3, 2, 1) Forecast

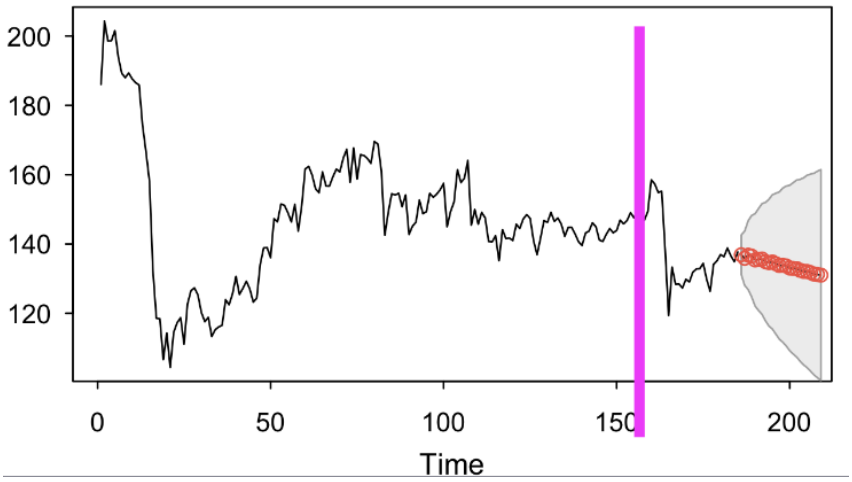

Figure 17: May 29, 2022 to Nov 29, 2022 forecast

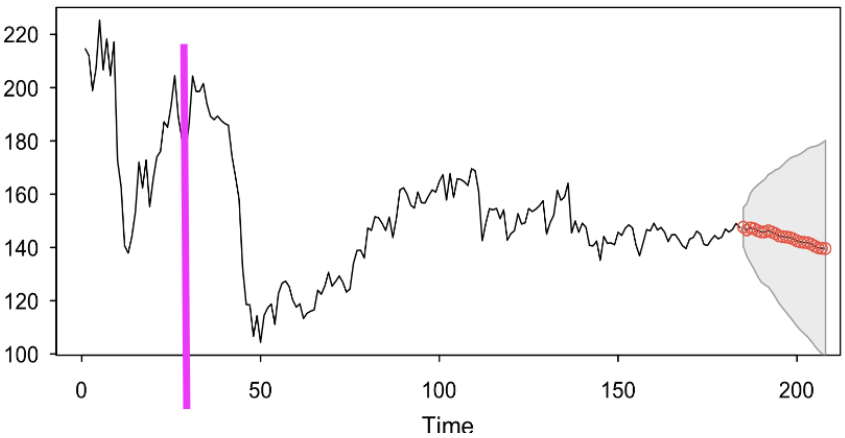

Figure 18: Apr 29, 2022 to Oct 29, 2022 forecast

7 Code

The colab notebook and R file for this project are available at `https://github.com/Aishwarya3011/STATS-221---Monero-Price-TSA.git`

References

Malte Möser, Kyle Soska, Ethan Heilman, Kevin Lee, Henry Heffan, Shashvat Srivastava, Kyle Hogan, Jason Hennessey, Andrew Miller, Arvind Narayanan, et al. An empirical analysis of traceability in the monero blockchain. *arXiv preprint arXiv:1704.04299*, 2017.

Ujan Mukhopadhyay, Anthony Skjellum, Oluwakemi Hambolu, Jon Oakley, Lu Yu, and Richard Brooks. A brief survey of cryptocurrency systems. In *2016 14th annual conference on privacy, security and trust (PST)*, pp. 745–752. IEEE, 2016.

Saralees Nadarajah and Jeffrey Chu. On the inefficiency of bitcoin. *Economics Letters*, 150:6–9, 2017.

Andrew Urquhart. The inefficiency of bitcoin. *Economics Letters*, 148:80–82, 2016.

Nicolas van Saberhagen. Cryptonote v 2.0. 2018.

Harald Vranken. Sustainability of bitcoin and blockchains. *Current opinion in environmental sustainability*, 28:1–9, 2017.