UNIVERSITÀ CATTOLICA DEL SACRO CUORE

Campus of Milan

Faculty of Banking, Finance and Insurance

Bachelor program in Finance

Predictive modeling of bitcoin log returns: an ARIMA and SARIMA approach

Supervisor,

Professor - Andrea Monticini

Report by

Immanuel Berger

515532

Academic Year 2023/2024

Table of contents

Chapter 1: Introduction

1.1 Background of Cryptocurrencies

Cryptocurrencies are decentralized digital financial assets that differs from conventional currencies and banking systems. They offer key advantages, such as the absence of any centralized authority, providing facilitating transfers of value without the need of a centralized power, lower transaction costs, improved security, and more financial inclusion. (Coinbase, 2023)

Bitcoin was the first cryptocurrency that emerged in 2007 with an online whitepaper, by a yet unknown person or group, who is known under the name of Satoshi Nakamoto.(Coinbase, 2023)

Following the financial crisis in 2008, the idea and aim of Bitcoin was to create a currency that would not be backed by governments or financial institutions and function as a decentralized store of value for investments.

With the most recent ETF launch of Bitcoin that links cryptocurrencies with traditional finance and retail investors, as well as other key events over time, cryptocurrencies have garnered significant attention and mass adoption to this day. (Official, 2023) These changes marked a critical turning point in the legitimacy and acceptability of cryptocurrencies within the international financial system. As of March 2024, the combined market capitalization of cryptocurrencies is 2.33 trillion dollars, with Bitcoin leading the pack with a market capitalization of 1.3 trillion dollars. (CoinMarketCap, n.d.)

1.2 Volatility and nonlinearity of cryptocurrency markets

The price of an asset rapidly rising or falling is known as volatility in financial markets; higher volatility indicates larger and more frequent price fluctuations, while lower volatility demonstrates the opposite. (Hayes, 2024) This means that the greater the volatility, the more rapid the movement of an asset can be in both directions, therefore increasing the risk of potential greater losses or greater returns over shorter periods of time.

Since cryptocurrencies are still a relatively new asset class, they are riskier than more established asset classes due to their lower market capitalization and their tendency for volatility, nonlinearity, and complexity. (Tong, 2022) One example is Bitcoin has over eight 50% price declines since inception, however, with multiple continued recoveries now reaching an all-time high of 73,000$ in March of 2024. These patterns show Bitcoin’s significant volatility in both directions. (Caleb & Brown, n.d.)

1.3 Time Series Forecasting in Financial Markets

Time series analysis is focused on analyzing data across time by using traditional statistical methods. It is very commonly used in forecasting within financial markets to discuss asset prices and market trends. However, statistical models that will be used in this investigation, such as the autoregressive model (AR), moving average model (MA), autoregressive moving average model (ARMA), and other models, assume stationarity and assumptions like linearity. However, financial markets can still exhibit complex and dynamic patterns, like nonlinearity that do not correspond to these assumptions, therefore distorting models' forecast accuracy and fit. However, with data processing, transformation and other methods it is possible to fit captured data correctly. This is because the data would comply more with the necessary statistical assumptions enhancing the model's efficiency in capturing the time series data correctly for modeling. (Tableau, n.d.)

1.4 Objectives of the Study

The purpose of this research is to examine the predictive performance of ARIMA and SARIMA models on Bitcoin, considering its inherently volatile nature. The results will discuss the efficiency of ARIMA and SARIMA models on forecasting accuracy and whether the extension of ARIMA to SARIMA models changes accuracy, hence, the following research question is formulated:

How effective are ARIMA and SARIMA models in predicting Bitcoin log returns, and does the extension from ARIMA to SARIMA improve forecast accuracy, given the volatile nature of the cryptocurrency market?

1.5 Structure of the content

The structure of the content is organized as follows: In the literature review, the theoretical underpinnings of forecast modeling in specific ARIMA and SARIMA models as well as the predictability of the cryptocurrency market will be covered. Following that, the methodology will be explained, such as data processing and modeling, and last the discussion of results following a conclusion section with limitations and suggested further research.

Chapter 2: Literature Review

2.1 Introduction

Models, such as ARIMA and SARIMA, are commonly and widely used for forecasting various types of datasets. Over time, studies have explored the application of models like the ARIMA and SARIMA models in applications within economics, finance, and capital markets. However, the unique dynamics within the cryptocurrency market pose a significant challenge within the realm of forecasting, and since the goal of this paper is to examine how effectively the ARIMA and SARIMA models can forecast the logarithmic returns of Bitcoin, by comparing the predicted values to the actual values, it is mostly relevant to focus on studies that apply these models, particularly within capital markets that may have similar characteristics to cryptocurrencies.

2.2 ARIMA and SARIMA Modeling

The approach of investigating the efficiency of ARIMA models for forecasting capital markets has also been made in a study by Shakir Khan, who has analyzed the efficiency of an ARIMA model on forecasting the Netflix stock price using a 5-year historical dataset to forecast 100 days (Khan, & Alghulaiakh, 2020). With the use of the auto ARIMA function, as well as building customized ARIMA models on the Netflix stock data, Shakir Khan was able to conclude that the ARIMA model was able to provide accurate results that closely reflected the actual values of the stock price performance, showing the potential of using ARIMA models on time series data to forecast the stock market. In this case, the forecasted values showed a horizontal price action, which was similar to how the stock price performed. (Khan, & Alghulaiakh, 2020) Literature on stock price prediction using ARIMA models based on this research paper reports positive outcomes. The research paper of Yang Si revolved around the use of the ARIMA model for predicting the Bitcoin price. The methodology was oriented towards providing smaller timeframe forecasts, in specific 5 and 10 days of the price of Bitcoin due to its inherent volatility (PDF, 2022). The conclusion was that the 5-day forecast provided more accurate results based on error metrics, such as root mean squared error (RMSE), mean absolute percentage error (MAPE), and mean absolute error (MAE). Furthermore, Yang Si noted that the greater the forecast span is, the greater the errors. The key takeaway was that while the ARIMA model seems suitable for forecasting assets like Bitcoin, longer forecasting periods can significantly increase the likelihood of errors. (PDF, 2022)

A study that provides also an additional contribution to this research published in IIETA focused on the performance of ARIMA modeling and SARIMA on certain datasets, such as sales data, temperature dataset, and stock prices, to compare their forecasting efficiency. The study concluded that in general, both models performed relatively well, however, for each dataset based on performance error metrics including mean squared error (MSE), MAPE, RMSE, the SARIMA model outperformed the ARIMA model. For the stock prices dataset, the SARIMA model demonstrated a lower MSE of 0.028, while the ARIMA model had an MSE of 0.041, and a MAPE of 7.32% versus ARIMA's 9.21% and a lower RMSE of 0.167 versus the 0.202, hence, providing more accurate results. This means that the SARIMA model consistently outperformed the ARIMA model in each dataset. (IIETA, 2023)

2.3 Predictability of Cryptocurrencies

Studies like the research from Helder Sebastião and Pedro Godinho also build on this research. Cryptocurrency predictability remains a significant challenge due to high volatility and deviations from fundamental indicators. Helder Sebastião and Pedro Godinho both examine the predictability of major cryptocurrencies, like Bitcoin and Ethereum, to investigate if models can be used to provide profitable strategies in trading despite the market inefficiencies. (Sebasti & Godinho, 2021). Under this research, machine models have been treated, however, the inherent concept of unpredictability would apply to any quantitative model to forecast within this market. The study has been able to conclude that there were strategies that did beat the market, with win rates above 60%. Nonetheless, these strategies generated lower returns compared to the general minima and maxima returns of cryptocurrencies, as well as being more exposed to high tail risks, with CVaRs, at 1% meaning the worst 1% of scenarios has an expected average loss between 3.88% and 13.40% and maximum drawdowns meaning the maximum observed loss from a peak of 11.15% to 48.06%. (Sebasti & Godinho, 2021) This means that there were greater risks underpinning a lower return potential. Essentially, this concludes that there is a possibility of using models to predict cryptocurrencies presenting positive return strategies with correct data partitioning, parameter setting, attribute space, and other methods; however, it can still be argued that these models have a relatively lower forecasting performance than traditional capital markets.

Chapter 3: Methodology

3.1 Theoretical Foundation

3.1.1 Stationarity in Time Series Models

A fundamental assumption for time series data based on time series models is that the data must be stationary. (Forecasting: Principles and practice, 2024) This implies that the mean, variance, and autocorrelation structure of time series data do not change over time and are constant. Since financial markets in general and especially cryptocurrencies exhibit high volatility with major deviations that challenges the assumption of stationarity, the model's reliability and forecast accuracy will be heavily distorted since the changing statistical properties over time influence the estimated parameters of a model, therefore leading to biased and unreliable results. To change the time series data structure to be stationary, transformations must be conducted, such as using the methods of differencing, which computes differences between the consecutive observations, or logarithmic transformations that change the data into logarithms. Both methods help stabilize its statistical properties to ensure underlying assumptions of stationarity are met. (Forecasting: Principles and practice, 2024)

3.1.2 Autoregressive Model

An autoregressive model is based on values of the past regressing its own values. This can be expressed in order p as (Forecasting: Principles and practice, 2024):

$$y_t = c + \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \ldots + \emptyset_p y_{t-p} + \epsilon_t$$

The variable y_t refers to the predictor variable that is being regressed, with c being a constant, $\emptyset$ the parameters of the predictor variables, and ϵ_t being the error term or also corresponding to white noise. (Forecasting: Principles and practice, 2024)

This model focuses on forecasting the future of time series data using the past values, hence meaning that the model would focus depending on how much you regress back to forecast what the future value will be. (Fernando, n.d.)

3.1.3 Moving Average Model

A moving average model is commonly used as a stock indicator in technical analysis. It is the simple arithmetic average of time series data helping level data over a specific period and providing a simplified overview of trend and pattern. (Fernando, n.d.) This can be expressed in order q as:

$$y_t = c + \in_t + \emptyset_1 \in_{t-1} + \emptyset_2 \in_{t-2} \ldots + \emptyset_q \in_{t-q}$$

This model looks at values depending on the amount of lags and averages them out to find the general trend direction of time series data. The greater the lag, the greater the model captures the general trend over a longer period. (Fernando, n.d.)

3.1.5 ARIMA Model

The ARIMA model combines the autoregressive and moving average models with a third component known as integrated that represents the differencing of the time series data to ensure modeling stationary time series data. The mathematical formula is the combined formula of the AR and MA models, which is expressed as follows:

$$y'_t = c + \emptyset_1 y'_{t-1} + \cdots + \emptyset_p y'_{t-p} + \in_t + \emptyset_1 \in_{t-1} + \cdots + \emptyset_q \in_{t-q}$$

The combination of both models as well as the time series data being differenced for stationarity expressed as y'_t the model ARIMA with order (*p, d, q)* integrates autoregressive and moving average components to effectively capture and forecast stationary time series data. (Forecasting: Principles and practice, 2024)

3.1.6 SARIMA Model

The SARIMA model, also known as the Seasonal Autoregressive Integrated Moving Average Model, is an extension of the ARIMA model, capturing also the seasonality of the data. This is done by extending the ARIMA model to: ARIMA (*p, d, q) (P, D, Q)m.* The extension of the parameters denoted by the P, D, and Q parameters in capital letters provides the seasonal components to the model. This is based on the process of backshifts denoted by *B*, which is used to represent lagged values of a time series. This allows the model to identify and account for underlying repeating patterns in the data. (Forecasting: Principles and practice 2024)

3.2 Process of Research

3.2.1 Data Collection and Transformation

The historical bitcoin price data is collected through the Yahoo Finance API. The training dataset will span from the 1st of October 2014 to the 10th of February 2024. The datapoints beyond the training dataset are excluded to avoid bias within the forecasting models, allowing for proper comparison between the 30-day forecasted values from the 10th of February and the actual values from the testing dataset. Using the 10-year span of historical data, the data will then be processed using Python. Since the models ARIMA and SARIMA will be utilized for this research, and they assume stationarity of time series data transformations of data are computed. In this research the logarithmic transformations to the returns of Bitcoin will be computed to stabilize the statistical properties of the data.

3.2.2 Outlier Detection and Handling

Outlier detection and handling is key to modeling time series data, since they can distort the accuracy of forecasting models. Outliers have several different causes, such as rare events or any mistakenly recorded values (FastCapital, n.d.). In this research, the method of trimming using the z-score is used. Z-score refers to the standardized scores within data based on how far the data is from the mean through the multiples of its standard deviation (Prabhakaran, 2023). By using the following formula, 99.7% of the data points are being considered, while the remaining data points outside are being trimmed down and removed from the dataset:

$$Z = \mu \pm 3\sigma$$

Z is denoted by the z-score, μ the mean of the data, and σ the standard deviation. This means that datapoints falling outside of the mean plus three times the standard deviation is disregarded, helping remove any outliers and capture the more necessary datapoints.

3.2.3 Stationarity Testing

There are several methods to test stationarity for time series data, whether it is through graphical interpretation or statistical testing. The method utilized is the Augmented Dickey-Fuller test, also known as the unit root test. Unit root is a characteristic that refers to data being stationary. (Prabhakaran, 2022) The Augmented Dickey-Fuller test is an extension of the Dickey-Fuller test, that investigates for stationarity at greater depth (Prabhakaran, 2022), since it expands on the test equation that follows the following hypothesis:

$$H0 = Unit\ Root\ (Non - stationary) \qquad H1 = No\ Unit\ Root(\ Stationary)$$

The p-value is used to interpret the hypothesis test of the ADF test. Usually a p-value with the critical region of 0.05 is used, which implies if the ADF test result results with a p-value of below 0.05, the null hypothesis is rejected, meaning that, there is no unit root present and, therefore, presents a stationary time-series data.

3.2.4 Model Selection

With data being processed for modeling time series data, the most appropriate model that provides goodness of fit and has a balance with model complexity must be estimated. This is done through evaluation of the models fit through the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) scores. Both are widely used criteria to evaluate the quality of a model in relation to model complexity, while AIC puts emphasis on the maximum likelihood of data, BIC emphasizes greater penalties for model complexity (Kumar, 2023). Essentially, the lower the AIC and BIC values are compared to other estimated parameters of a model, the greater the goodness of fit there is, including a favorable balance with the complexity of a model. The aim is to find the best AIC and BIC, which is presented with the lowest AIC and BIC scores.

3.2.5 Residual Analysis

After the model selection, it is empirical to analyze the residuals of the data, which correspond to the difference between the actual values of the data and the fitted values. Analyzing residuals supports and strengthens the conclusion of how well the model captured the data's structure and dynamics. This hypothesis is supported under the condition that residuals have no autocorrelation and constant statistical properties with a mean of 0. If there is autocorrelation or a mean deviating away from zero, it would imply bias within the forecast, therefore making it inefficient for actual forecasts. (Forecasting: Principles and practice, 2024)

3.2.6 Forecast & Performance Evaluation

Lastly, when producing the forecasted values for both models, a comparative analysis is conducted to strengthen the assessment of the estimated fitted models. Three different metrics will be used to evaluate the model's performance. One of them being the mean absolute error also known as MAE, that measures the average magnitude of errors in the set of forecasts. The root mean squared error (RMSE) also helps in evaluating the quadratic magnitude measure of error that puts more weight on greater errors within the forecast. Since both MAE and RMSE metrics provide only information on errors across the dataset, the mean absolute percentage error (MAPE) is used since it evaluates a model's performance in a percentage in correspondence to the errors suiting it for easier comparative analysis and evaluation. The lower the percentage is of MAPE, the better the models fit. Along with the researched models, more simpler models, such as a simple AR (1), EMA, and SMA models, will be used to further strengthen a comparative analysis. (Serre, 2024). In addition to this, forecasted values for both ARIMA and SARIMA models will be graphed out along with the actual values to visually interpret their performance.

Chapter 4: Results and Discussion

4.1 Descriptive Statistics & Data Transformation

Figure 1 presents the graph of returns of Bitcoin over time from 2014 October to February 2024, and Figure 2 represents a seasonal decomposition on the returns of Bitcoin. Over time, statistical properties, especially after 2020, have high variability implying non-stationarity. The seasonal decomposition also provides further evidence on non-stationarity, as residuals are not in accordance with constant statistical properties and high variability, along with the presence of seasonal patterns within the data.

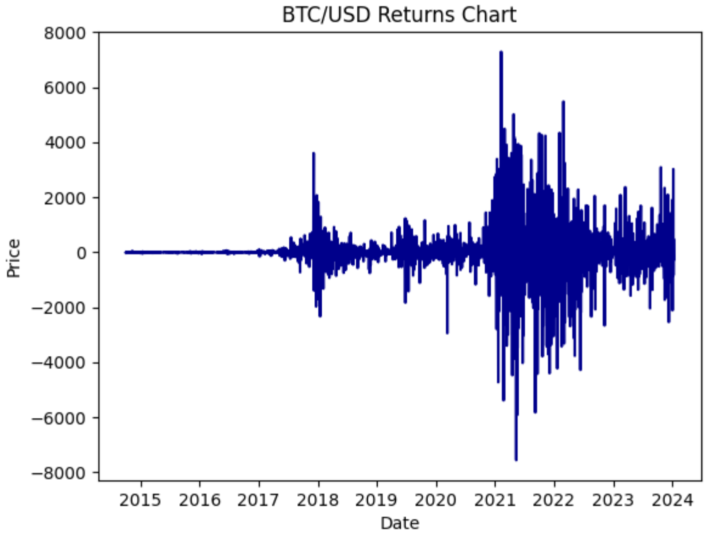

Figure 1: Chart of Bitcoin/USD Returns

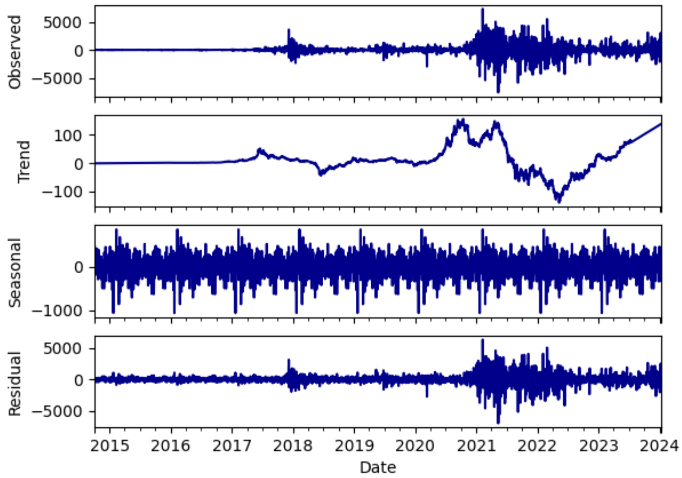

Figure 2: Seasonal Decomposition of Bitcoin/USD Returns

Computing logarithmic returns of Bitcoin helps stabilize the statistical properties of data within the time-series data, and trimming the data helps remove any inefficiencies for forecasting time series data. Figure 3 being the logarithmic returns of Bitcoin over time and Figure 4 being the logarithmic returns with outliers being trimmed out both portray more stable statistical properties, deeming it more fit to model and interpret a stationary time series data.

Figure 4 presents the closest based on visual interpretation to be a stationary time series data for applying ARIMA and SARIMA models effectively. Table 1 provides an overview of descriptive statistics between the non-transformed time series data and the time-series data after computing logarithmic returns and trimming the outliers. With Table 1, it is evident that the time series data after logarithmic computation as well as trimming the outliers, has the lowest difference between maxima and minima being -0.110 and 0.113 and standard deviation being 0.031, proving to be more stationary than the non-transformed dataset.

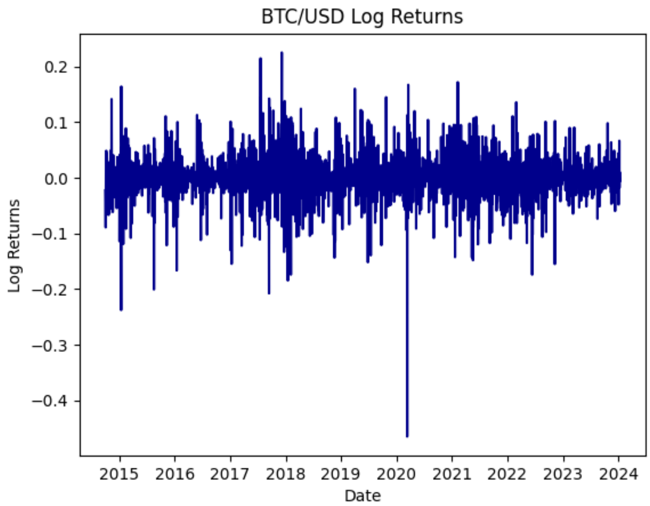

Figure 3: Logarithmic Returns of Bitcoin/USD over time

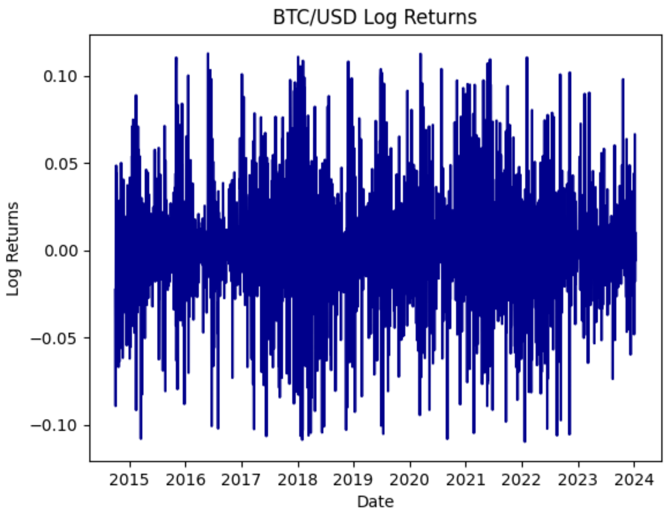

Figure 4: Logarithmic Returns of Bitcoin/USD with outliers trimmed

Variable	Number of Observations	Mean	Min	Max	STD
Bitcoin Returns	3389	13.56	-7554.04	7293.02	785.49
Bitcoin Log Returns	3389	0.001	-0.465	0.225	0.037
Bitcoin Log Returns (Trimmed Outliers)	3334	0.002	-0.110	0.113	0.031

Table 1: Descriptive Statistics of Bitcoin Returns and their respective transformations

4.2 ARIMA and SARIMA Model Selection

Figures 6 and 7 present the ACF and PACF plot, of the transformed time series data. In both figures, it is evident that there is no autocorrelation within the data, which is crucial for applying ARIMA and SARIMA models. This is shown by each lag of the ACF and PACF after lag 0 having no or little deviation outside of the confidence intervals.

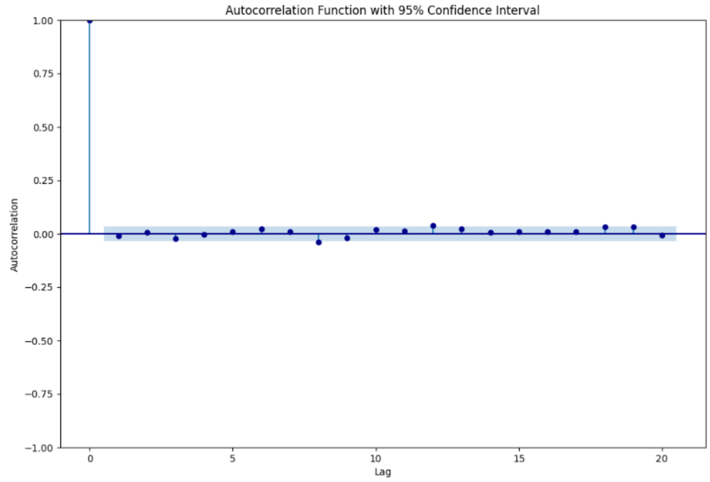

Figure 6: Autocorrelation Function (ACF) plot

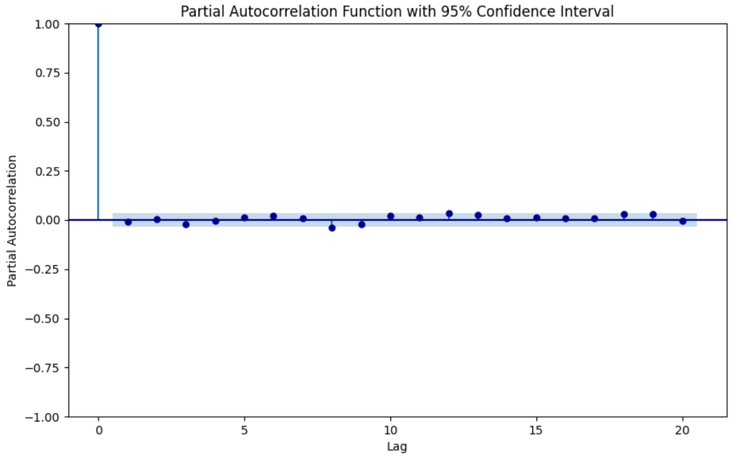

Figure 7: Partial Autocorrelation Function (PACF) plot

To further strengthen the hypothesis of stationarity within the data, the ADF test is computed as seen in Table 2, and with a result below the p-value of 0.05, therefore rejecting the null hypothesis, it is possible to imply stationarity within the data. The ADF test reported a p-value below 0.05, therefore further confirming that the data is stationary. This means that the ARIMA and SARIMA models do not require any further differencing of the data to have stationarity to model, implying that the parameter for the ARIMA and SARIMA models of *d* is equal to 0.

ADF:	**-58.262575**
P-Value:	0.0
Num of Lags	0
Observations for ADF Regression and Critical Values Calculation:	3333
Critical Values:	
1%:	-3.4323134993
5%:	-2.8624075589
10%:	-2.5672318190

Table 2: ADF Test

With the auto ARIMA and SARIMA functions on Python, parameters were estimated based on the AIC and BIC metrics. The ARIMA model with order (1,0,1) and the SARIMA model with order (0,0,0) (2,0,0,6) were found to have the lowest AIC and BIC values, as seen in Table 3. This implies that the ARIMA model incorporates an autoregressive order of 1 and a moving average order of 1 without any further differencing as interpreted in the ACF, PACF plots, and ADF test above. The SARIMA model only incorporates a seasonal order P equal to 2, including a seasonal component of 6. This means that the model contains no non-seasonal parameters, while it includes two seasonal AR terms that occur every 6-time steps and in this case every 6 months**.** When comparing both models, the SARIMA model has lower metrics in AIC and BIC, possibly suggesting a better fit.

Model	**AIC**	**BIC**
ARIMA 10,0,1)	-13602.784	-13578.336
SARIMA (0,0,0) (2,0,0,6)	-13608.821	-13584.373

Table 3: AIC and BIC Values of best-fitted models

4.3 Model Diagnostics

Looking at the residuals of the ARIMA (1,0,1) model on Figure 8, the residuals relatively act in accordance with white noise, showing no trends in pattern and constant statistical properties. Furthermore, the residuals of the model show no autocorrelation as well, and the Q-Q plot and histogram both show properties corresponding relatively to a normal distribution with minor differences to the distribution indicating a relatively good-fit on the time series data.

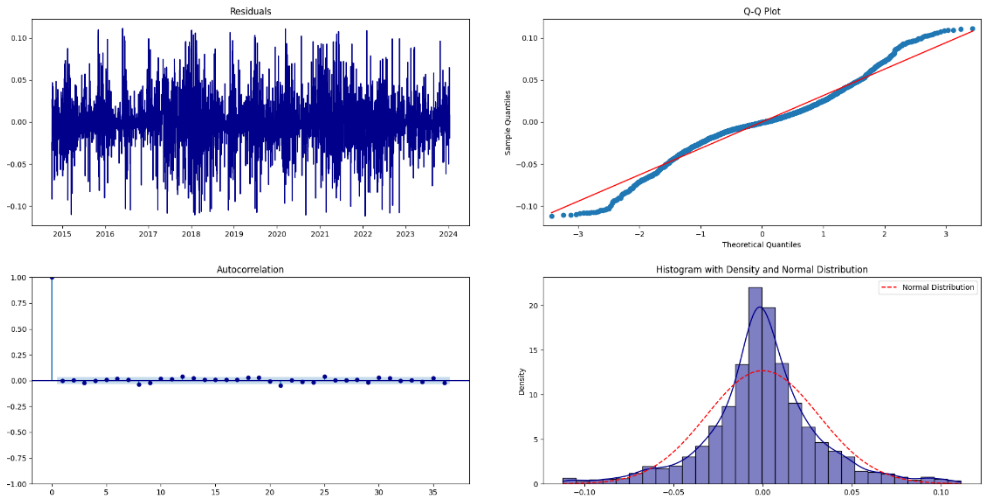

Figure 8: Residual Diagnostics of ARIMA (1,0,1)

Figure 9 shows the visual presentation of the residuals of the SARIMA model with order (0,0,0) (2,0,0,6). Figure 9 also shows properties that align with a relatively good model fit with no autocorrelation or any major deviations from the conditional properties of time series data to fit for forecast.

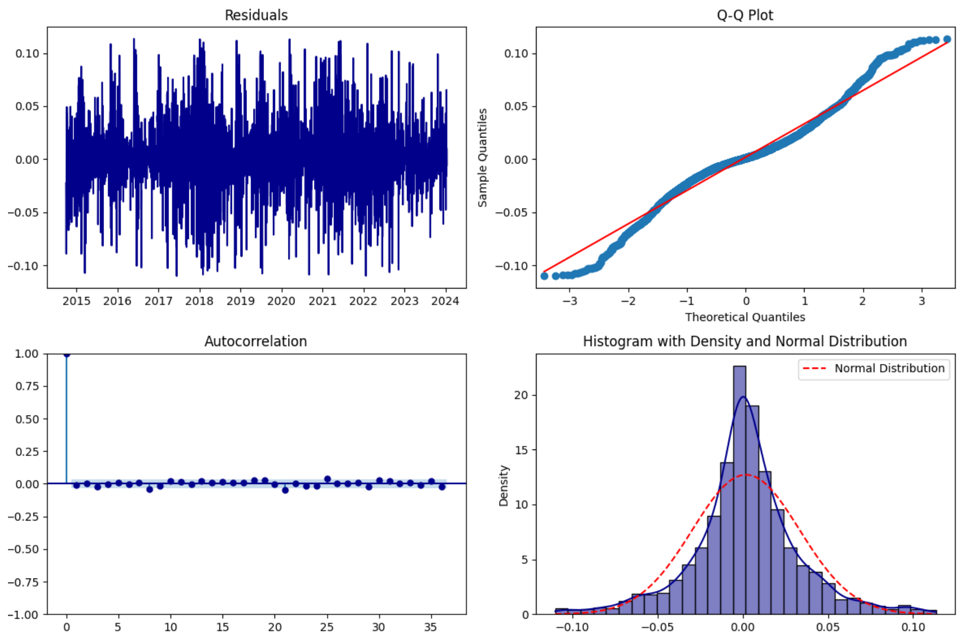

Figure 9: Residual Diagnostics of SARIMA (0,0,0) (2,0,0,6)

The p-values of the Ljung-Box Test for the both proposed models, which are seen in Table 4, also confirm within 10 lags that each p-value is above the critical region of 0.05, therefore not rejecting the null hypothesis, which means that for both proposed models the residuals are independently distributed and have no autocorrelation that would have distorted the models forecast accuracy.

Model	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 6	Lag 7	Lag 8	Lag 9	Lag 10
ARIMA (1,0,1)	1.0	0.95	0.60	0.75	0.80	0.68	0.75	0.32	0.29	0.27
SARIMA (0,0,0) (2,0,0,6)	0.51	0.77	0.55	0.71	0.78	0.86	0.90	0.46	0.46	0.42

Table 4: Ljung-Box Test (p-value)

4.4 Model Evaluation

As mentioned in the methodology section, simpler models are also used as a comparative analysis in terms of analyzing the ARIMA and SARIMA model approaches. In this research, a autoregressive model with order 1 (AR (1)), a simple moving average with lag of 10 (SMA 10) and an exponential moving average with lag of 10 (EMA (10)) are utilized. Table 5 provides an overview based on the mean absolute error, root mean squared error and mean absolute percentage error. Looking at Table 5, based on the MAPE values, the proposed ARIMA model with order (1,0,1) had the lowest percentage, slightly below the SARIMA model and the other simpler models. This means that the ARIMA model has performed well in terms of accuracy, capturing the closest forecasted values to the actual values of the logarithmic returns of Bitcoin. Although the Simple Moving Average model has the lowest MAE and RMSE, it produced the highest MAPE with a value of 1.73%, meaning that the SMA model captured lower average errors, however, its percentage error relative to the actual values was higher. All models were able to have relatively similar outputs based on error metrics, but ARIMA projected overall based on all 3-error metrics the lowest error deviation to the actual values, meaning that the model was able to capture relatively the forecast the best, however, showing not too big marginal differences with the other models.

Model	MAE	RMSE	MAPE
ARIMA (1,0,1)	0.01655	0.02444	1.28%
SARIMA (0,0,0) (2,0,0,6)	0.01656	0.02445	1.33%
AR (1)	0.01655	0.02444	1.29 %
SMA (10)	0.01635	0.02420	1.73%
EMA (10)	0.01719	0.02499	1.32%

Table 5: Model Evaluation Metrics Table

4.5 Forecast

A visual representation of the forecasts versus the actual values of the bitcoin logarithmic returns is presented for both models. Figures 10 and 11 present both visualizations of the ARIMA models forecast with the comparison of the actual values. It can be visually evaluated that the general trend of the logarithmic returns of bitcoin has been captured, as well as that the 95% confidence intervals of the model were mostly respected with one slight deviation at the beginning of the forecast. However, it can be noted that short-term fluctuations that are clearly deviating away from the forecasted values were not fully accounted for but only the general trend.

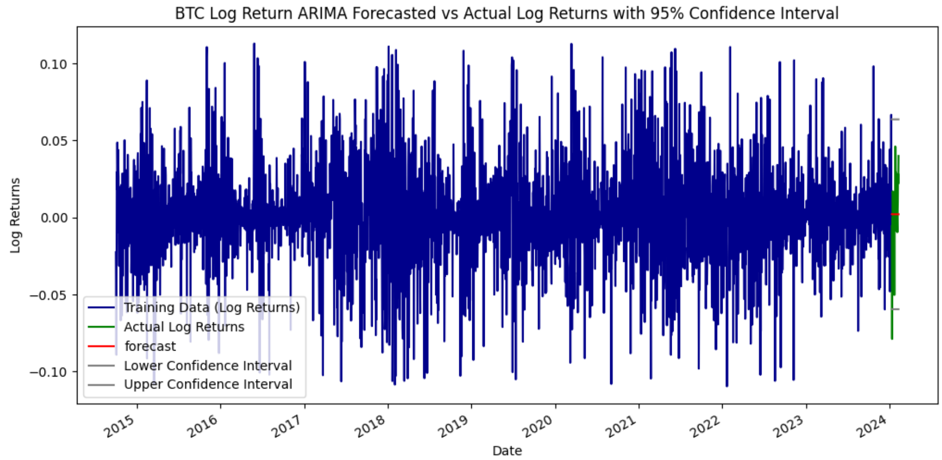

Figure 10: ARIMA (1,0,1) Forecast on Bitcoin Logarithmic Returns vs Actual Values

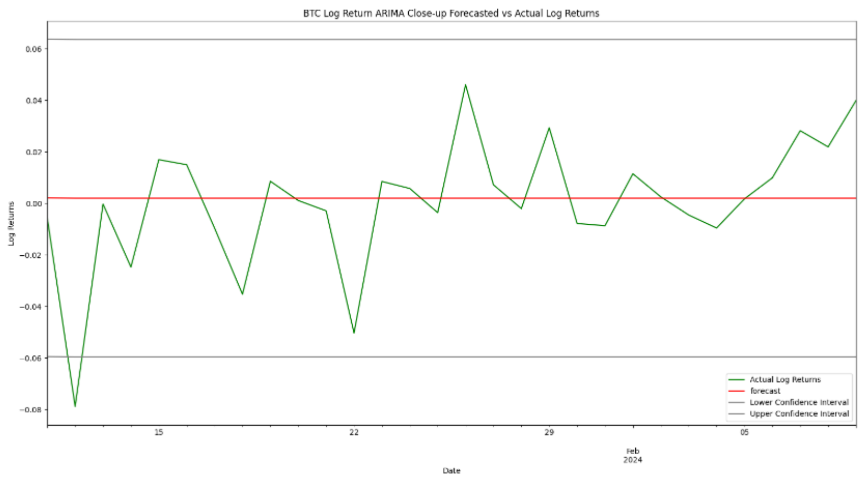

Figure 11: Close-up of Forecasted values of ARIMA (1,0,1) on Bitcoin Logarithmic Returns vs Actual Values

Figures 12 and 13 both show the visualizations of the SARIMA model. The SARIMA model was also able to capture the general trend of the forecast relatively well. Comparing it to the visualization of the figures above, the SARIMA model in comparison also forecasted small fluctuations in the first 10 days of the forecast. However, these fluctuations were not as extreme as the actual values.

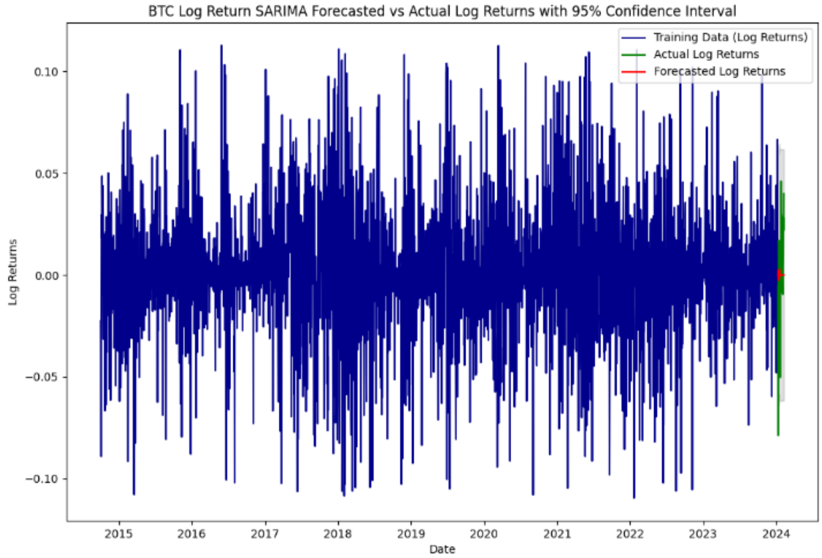

Figure 12: SARIMA (0,0,0) (2,0,0,6) Forecast on Bitcoin Logarithmic Returns vs Actual Values

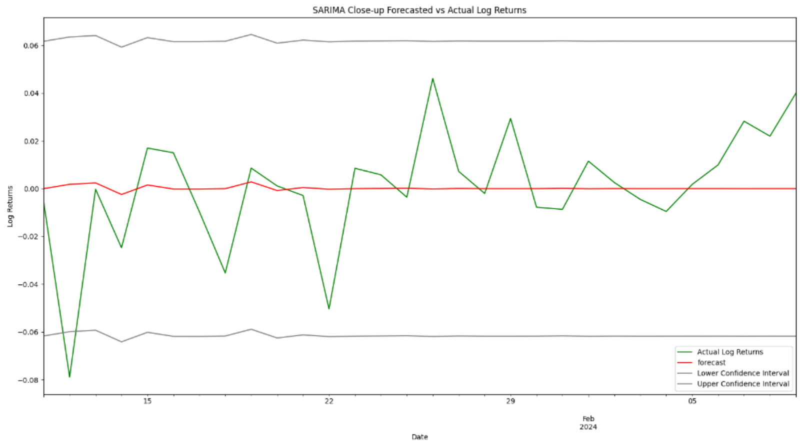

Figure 13: Close-up of Forecasted values of SARIMA (0,0,0) (2,0,0,6) on Bitcoin Logarithmic Returns vs Actual Values

Both models may not match exactly the actual values but capture the general trend well. Additionally, most of the actual values lie within the confidence intervals of 95% besides one deviation, making the model statistically reasonable for forecasting. Even if SARIMA was able to account for some fluctuations within the first period of the forecast, both models' forecast was less responsive relative to the actual movements, showing little to no fluctuations and more of a smooth line.

Chapter 5: Conclusion

5.1 Conclusion

In this thesis, the best-fit ARIMA and SARIMA models were estimated to forecast 30 days of the logarithmic returns of Bitcoin. To validate the models and find the most appropriate model for forecasting, certain methods and criteria were utilized. First, determining the stationarity of the time series data as well as detecting any extreme outliers to the data to make necessary data transformations and handling. Following that, the PACF and ACF models were used to strengthen the validation, and the best-fit parameters for both models were found according to the AIC and BIC values as benchmarks. The analysis of the models' residuals was visualized as well as statistically analyzed using the Ljung-box test for further validation of the model-fit. After the validation, both models provided a 30-day forecast, which was statistically compared using the three-error metrics and compared to simpler models as other benchmarks for comparison. Furthermore, the forecasted values of both models have been visually presented along with the actual values.

With the visual and statistical interpretation, the results indicate, both ARIMA and SARIMA models performed relatively well. The three-error metrics reported low values, with ARIMA having the lowest mean absolute percentage error of 1.28% and lower mean absolute errors as well as root mean squared errors in comparison to the SARIMA model. This means that the ARIMA model has outperformed the SARIMA model across all metrics. Nonetheless, both reported relatively low error metrics, similarly to the simpler models. Based on the visual interpretation it is conclusive that the forecast of both models has been able to capture the general trend well besides the volatile movements within the confidence bands and the data point that deviated out of the confidence intervals. Despite ARIMA's marginally better performance the differences based on the error metrics between the other models including SARIMA and the simpler alternatives (AR, SMA, EMA), it can be also concluded, that for short-term forecasting, even simpler models may provide comparable results, though ARIMA still demonstrated the better performance.

5.2 Implications

Literature provides theoretical explanations for the occurrence of these results. The analogous work of Yang Si in predicting Bitcoins' price using the ARIMA model discussed that the price volatility as well as the rapid changes make the forecasting more difficult especially the longer the term for forecasting is needed. In Yang Si's case, a 5-day and 10-day forecasting period has been used, while in this research, a 30-day forecast has been established proving to have marginally greater errors. (PDF, 2022) Furthermore, cryptocurrencies have nonlinear structure characteristics as well as trends in cycles and unpredictability in the long-term, which is driven by many aspects. (Zhongwen et al. 2022) These findings highlight that, while data transformation and validation methods help stabilize certain statistical characteristics of the time series data to fit for model the nonlinearity, volatility or sudden structural changes may still not be fully eliminated, hence, making the models not capture the more complex movements within the forecasting period.

5.3 Limitations and further research

This research is not without limitations. Both ARIMA and SARIMA have limitations and weaknesses, since the models cannot account for all sudden structural changes, nonlinearity and volatility that is inherent in cryptocurrencies, even after data transformations are made. Furthermore, due to the inherent characteristics, the greater the forecasting period is, the more errors there will be and decline in accuracy. This paper has also only accessed one period data of Bitcoin. More periods can be used to compare if the forecasting accuracy of both models provide similar results and help identify any anomalies. A further investigation could also be made into more complex types of models or methods that consider such volatility, such as the ARIMA-GARCH or machine learning models that could possibly forecast more accurate results. More in-depth research can also be made considering if there is a model that outperforms other models and if a more complex model would outperform simpler ones or vice versa. Additionally, when looking at the characteristics of cryptocurrencies it can be investigated on the certain key indicators that would provide the highest forecast accuracy between data manipulation or model selection.

Bibliography

Journal Articles

1. Khan, S., & Alghulaiakh, H. (2020). Arima model for accurate time series stocks forecasting. *International Journal of Advanced Computer Science and Applications (IJACSA)*. https://thesai.org/Publications/ViewPaper?Volume=11&Issue=7&Code=IJACSA&SerialNo=65
2. Sebastião, H., & Godinho, P. (2021, January 6). Forecasting and trading cryptocurrencies with machine learning under changing market conditions - financial innovation. *SpringerOpen*. https://jfin-swufe.springeropen.com/articles/10.1186/s40854-020-00217-x
3. (PDF, 2022) Using Arima model to analyse and predict bitcoin price. (n.d.). https://www.researchgate.net/publication/366296638_Using_ARIMA_model_to_analyse_and_predict_bitcoin_price
4. IIETA. (2023). Time series modeling and forecasting using autoregressive integrated moving average and seasonal autoregressive integrated moving average models. https://iieta.org/journals/i2m/paper/10.18280/i2m.220404

Book Chapters

1. Forecasting: Principles and practice (2024). *(2nd ed)*. 8.3 Autoregressive models. https://otexts.com/fpp2/AR.html

---

Websites

1. Ajitesh Kumar. (2023, November 10). AIC & BIC for selecting regression models: Formula, examples. *Analytics Yogi*. https://vitalflux.com/aic-vs-bic-for-regression-models-formula-examples
2. Coinbase.(2024). What is cryptocurrency?. *Coinbase*. https://www.coinbase.com/learn/crypto-basics/what-is-cryptocurrency
3. CoinMarketCap. (n.d.). *Kryptowährungspreise, diagramme und marktkapitalisierungen*. https://coinmarketcap.com/de/
4. David Serre. (2024, June 11). Time series forecasting model performance evaluation in predictive planning. *SAP Community*. https://community.sap.com/t5/technology-blogs-by-sap/time-series-forecasting-model-performance-evaluation-in-predictive-planning/ba-p/13725388#:~:text=MAPE%20is%20proposed%20in%20Predictive,different%20models%20(or%20entities).
5. Official C. (2024, April 22). Bitcoin and Ethereum ETF explained: Definition, how it works and recent institution approvals. *Medium*. https://medium.com/coda-official/bitcoin-and-

ethereum-etf-explained-definition-how-it-works-and-recent-institution-approvals-dbeb19250eba

6. FasterCapital. (n.d.). Outliers: Handling anomalies: Addressing outliers in Arima forecasting. https://fastercapital.com/content/Outliers--Handling-Anomalies--Addressing-Outliers-in-ARIMA-Forecasting.html
7. Fernando, J. (n.d.). What are autoregressive models? How they work and example. *Investopedia*. https://www.investopedia.com/terms/a/autoregressive.asp#:~:text=Autoregressive%20models%20are%20statistical#:~:text=Autoregressive%20models%20are%20statistical
8. Fernando, J. (n.d.). Moving average (MA): Purpose, uses, formula, and examples. *Investopedia*. https://www.investopedia.com/terms/m/movingaverage.asp
9. Hayes, A. (2024, July 3). *Volatility: meaning in finance and how it works with stocks*. Investopedia. https://www.investopedia.com/terms/v/volatility.asp
10. Prabhakaran, S. (2022, April 4). Augmented Dickey-Fuller (ADF) test - must read guide - ml+. *Machine Learning Plus*. https://www.machinelearningplus.com/time-series/augmented-dickey-fuller-test/
11. Prabhakaran, S. (2023, August 5). How to detect outliers with Z-score. *Machine Learning Plus*. https://www.machinelearningplus.com/machine-learning/how-to-detect-outliers-with-z-score/
12. Tableau. (n.d.). *Time series analysis: Definition, types, techniques, and when it's used*. https://www.tableau.com/learn/articles/time-series-analysis
13. Caleb & Brown, Why is crypto so volatile? Understanding market movements: *Caleb & Brown Blogs RSS*. (n.d.). https://calebandbrown.com/blog/crypto-volatility/
14. Zhongwen Tong, Apopo, N., Cheah, E. T., Corbet, S., Dyhrberg, A. H., Lin, X., Luo, W., Urquhart, A., Yang, L., & Zhao, Z. Y. (2022, March 19). Nonlinear Dynamics Analysis of cryptocurrency price fluctuations based on Bitcoin. Finance Research Letters. https://www.sciencedirect.com/science/article/abs/pii/S1544612322001155