A Time Series and Time Lag Analysis of Bitcoin (BTC), Ethereum (ETH) and Ripple (XRP)

JOSEPH ALLEN

University of Southampton, SO17 1BJ, United Kingdom

ja7g19@soton.ac.uk

Student ID: 31500145

The objectives for this report are to assess to what degree are Bitcoin, Ethereum and Ripple are correlated and if there is an obvious time lag between the cryptocurrencies. The report uses a combination of differencing techniques, moving average, augmented-dickey-fuller analysis, and ARIMA modelling to produce a plot forecasting the price for Bitcoin, Ethereum and Ripple for the next month and the next 6 months.

1. WHAT IS CRYPTOCURRENCY?

A cryptocurrency is a digital currency that is solely virtual and is not issued or regulated by a central bank [1]. The system relies on decentralized technologies, often utilizing blockchain, to ensure a transparent ledger for all users. Cryptography is used to secure transactions and regulate the creation of new units. The emergence of cryptocurrencies began with the introduction of Bitcoin in 2009 by an anonymous entity known as Satoshi Nakamoto [2]. Since Bitcoin's inception, numerous alternative cryptocurrencies such as Ethereum and Ripple developed. These alternative coins use similar principles of decentralization and cryptography to offer their own innovative features and functionalities [3]. Cryptocurrencies provide several advantages over traditional financial systems: They enable direct transactions between parties without the need for intermediaries like banks and users have greater control over their funds using digital wallets and private keys. Furthermore, cryptocurrencies have the potential to promote financial inclusivity by providing access to financial services for individuals who may not have access to traditional banking infrastructure [4].

2. WHAT IS TIME SERIES AND TIME LAG ANALYSIS?

Time series analysis is a statistical technique used to interpret data collected over time, typically at regular intervals. It involves studying patterns, trends, and relationships within the data to gain insights into the underlying dynamics of the system being studied [5]. Time lag analysis, on the other hand, focuses on exploring the relationships between variables with a time delay, where the effect of a change in one variable is observed after a certain period [6]. When studying cryptocurrencies, time series analysis plays a crucial role in understanding their price movements and volatility. Cryptocurrencies like BTC and ETH are known for their highly unpredictable and complex price patterns, making traditional analysis methods less effective [7]. Information gathered from time series studies can be valuable for developing trading strategies and risk management in cryptocurrency markets [8]. In cryptocurrency research, time series analysis helps researchers to forecast future price movements and assess market risk, models used consider factors like seasonality, trends, and random volatility that influence cryptocurrency prices. Cryptocurrency markets are highly interconnected, and time lag analysis can identity changes in the price of one cryptocurrency that can influence the prices of others, but with a time delay [9]. By studying time lags, lead-lag relationships can be identified and determine how changes in one cryptocurrency's price affect other cryptocurrencies after a specific period. This analysis helps uncover dependencies between cryptocurrencies, which can be useful for portfolio diversification and identifying potential buy/sell opportunities [10].

Bitcoin, Ethereum, and Ripple are chosen for this report as they stand out in terms of market dominance, technological innovation, and widespread adoption. Bitcoin, often referred to as 'digital gold' is the pioneer of cryptocurrency. Its market dominance has consistently remained high, as of September 2021 Bitcoin accounted for about 40% of the total market capitalization of all cryptocurrencies [11]. This dominance positions it as a crucial element for a time series and time lag analysis on cryptocurrency. Ethereum is the preferred platform for developers and projects due to its robustness and versatility. Additionally, ETH demonstrates a higher number of active addresses and daily transactions compared to other cryptocurrencies [12]. Ripple has partnerships with numerous financial institutions, including banks and payment providers, aiming to improve the speed and affordability of cross-border transactions [13].

3. DATA COLLECTION AND PRE-PROCESSING

Daily closing prices for Bitcoin, Ethereum and Ripple as a ratio of the coin to the US dollar, from 1st January 2017 to 7th May 2023 were downloaded using the yfinance package. These prices were concatenated, with the dates of the closing prices set as the index of the data frame using the pandas module in order to be used for a time series analysis. To clean the data of infinite or missing values, infinite values were replaced with NaN values to be dropped from the dataset shortly after. Missing values were dropped and replaced with the closing price for the next day.

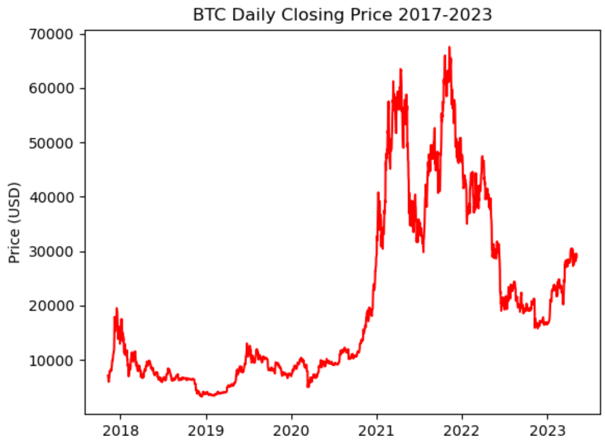

Fig. 1

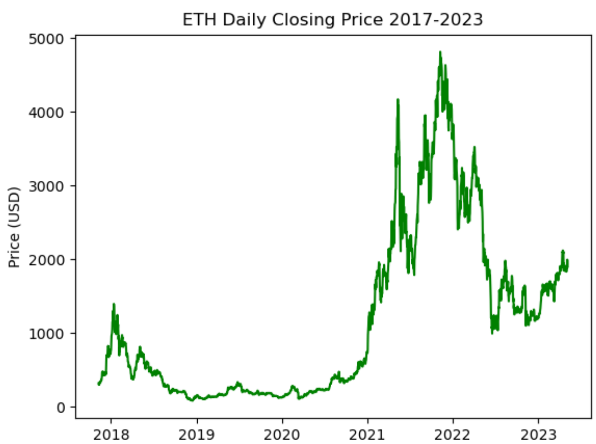

Fig. 2

Fig. 3. The daily closing prices for Bitcoin, Ethereum and Ripple from the 1st January 2017 to 7th May 2023 were plotted using publicly available data from Yahoo Finance. All three cryptocurrencies had elevated prices during the start of 2018, however falling due to a market bubble burst [14]. BTC, ETH and XRP all experienced the most volatility in price from the start of 2021 to midway through 2022.

4. METHODOLOGY

A. Rolling Mean and Rolling Standard Deviation

The rolling mean (moving average) was taken to smooth price data over a specified time window allowing myself to remove the noise generated by short term volatility and identify the direction of the trend in the crypto coin. The rolling mean was calculated by taking the mean of closing prices over a window, and then moving the window forward one step and recalculating the average [14]. Similarly, the rolling standard deviation was taken to highlights periods of volatility and stability within the markets. This involved computing the standard deviation of the closing price over a specified window and shifting the time window, recalculating the standard deviation at each step. This process is repeated throughout the entire dataset, creating a series of mean and standard deviation values. The choice of the window size for the two techniques is important. A monthly window (30 closing prices), provides more sensitive and responsive fluctuations, capturing recent price movements [15]. On the other hand, the biannual window (180 closing prices) offers smoother fluctuations that reflect longer-term trends. BTC as well as ETH showed tremendous growth during 2021 whilst XRP remained sideways in movement, as indicated by the 180-day rolling mean and the 30-day rolling mean. Due to the high volatility in standard deviation in the 30-day rolling mean, the 180-day rolling mean for BTC, and ETH began to level out before a strong descent due to conflict in Ukraine and the federal reserve raising interest rates to combat inflation [16].

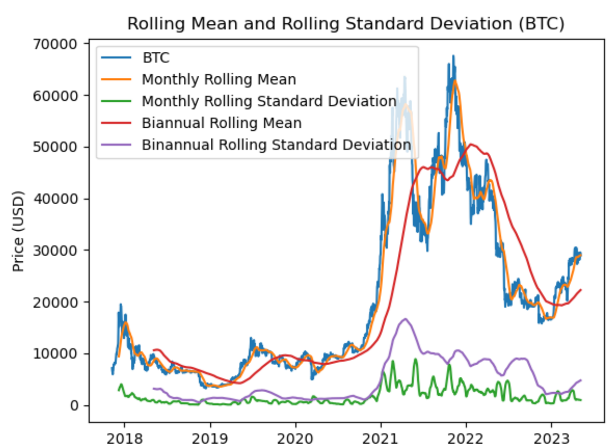

Fig. 4

In industry as well as retail applications, analysts may compare the current price to its rolling mean. When the price crosses above the rolling mean, it may indicate a bullish signal, suggesting potential buying opportunities [17]. Examples of this

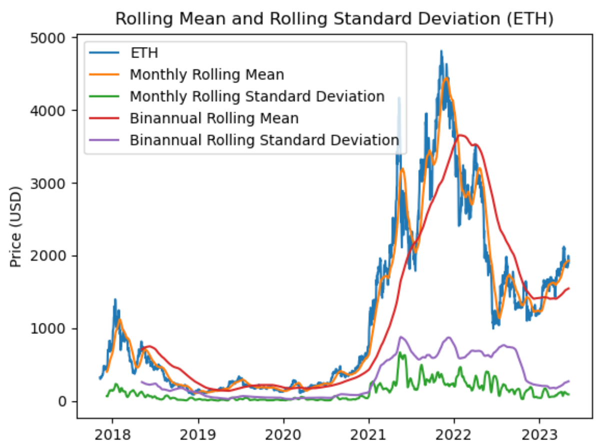

Fig. 5

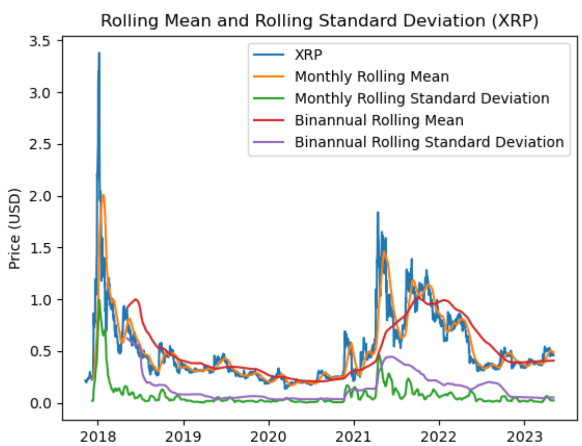

Fig. 6

positive crossover are in early 2021 for BTC, ETH and XRP. Conversely, when the price crosses below the rolling mean, it may signal a bearish trend, indicating potential selling opportunities. Examples of this are towards the middle portion of 2022 for BTC, ETH and XRP. Additionally, analysts often use multiple rolling averages simultaneously, to identify longer-term trends and potential support and resistance levels [18]. The crossover of these moving averages, known as the "golden cross" (50-day moving average crossing above the 200-day moving average) or the "death cross" (50-day moving average crossing below the 200-day moving average), is considered significant by many traders and can influence trading decisions [19]. Currently the 30 day and 180 day rolling mean indicate that BTC and ETH are in an upwards trend, however XRP has a neutral sentiment.

B. Are the Series Stationary?

The Dickey-Fuller test was implemented as a starting check if the dataset is stationary or non-stationary. It uses the augmented Dickey-Fuller (ADF) regression model, which incorporates lagged differences of the variable under examination [20]. The test check for the presence of a unit root, a characteristic suggesting non-stationarity. The test starts off with the null hypothesis H_0. This assumes the data has a unit root, indicating non-stationarity. A dataset with a unit root is inappropriate for a time series analysis as this exhibit random walk behaviour where the future values are only determined by the previous value so therefore there is no systematic structure to the data. The alternative hypothesis, H_1 presumes there is no unit root, indicating stationarity [21]. The test was run with the raw data for BTC, ETH and XRP leading to p-values for BTC and ETH being greater than 0.05. To combat this, the data is backwards differenced to remove trends and systematic patterns in the data. The test is run again with this new data resulting in p-values under 0.05 and ADF statistical numbers lower than their respective critical values for each significance level: (1%, 5% and 10%). Details for the models used within this hypothesis testing are in the appendix for this report. The differenced critical values at

Table 1. Raw Test Statistic and P-Value

Coin	ADF Statistic	P-Value
Bitcoin (BTC)	-1.4280998565540866	0.5687330426160875
Ethereum (ETH)	-1.3952216619159907	0.5845647578426763
Ripple (XRP)	-3.849637746150579	0.002438388119502058

Table 2. Differenced Test Statistic and P-Value

Coin	Differenced ADF Statistic	Differenced P-Value
Bitcoin (BTC)	-7.810040199862575	$7.110003919560612\times10$
Ethereum (ETH)	-11.21715308069265	$2.0560629595122946\times10$
Ripple (XRP)	-9.054030211508406	$4.754730434146849\times10$

1%, 5% and 10% are -3.43, -2.862, -2.567 respectively. All ADF values for BTC, ETH and XRP are under these, therefore we can reject the null hypothesis and say the series is stationary according to the Dickey-Fuller test.

C. Are the Series Correlated?

The Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) were used to examine the correlation between the BTC, ETH and XRP time series and their lagged values. ACF examines the relationship between a time series and its past values. Correlation coefficients are calculated for different time lags, revealing the strength and direction of the linear trend between observations at different time points. PCF, however, checks the correlation between observations at different lags while considering the indirect correlation through intermediate lags [22]. It specifically measures the correlation between two observations at a given lag, excluding the influence of shorter lags. The ACF and PACF plots together yielded strong spikes at lag 1 in the ACF and PACF plot. From the ACF plot, this indicated strong autocorrelation in the first order differencing of the data, the lack of significant spikes at other lags indicated that there was no seasonality or cyclic pattern in the data. The strong spike at lag 1 for the PACF plot implied an autoregressive relationship in the data. Beyond the first lag, most points fall within the confidence interval suggesting that there is no significant autocorrelation beyond lag.
With the ACF and PACF plots being identical (including and

after lag 1), there is no clear evidence of any rolling mean or autoregressive patterns, the series are independent and are influenced by random, unrelated factors at further lags. This pattern of ACF and PACF plot are most seen in stationary time series and simple random walks [23]. Details of how the ACF and PACF plots are made is in the appendix.

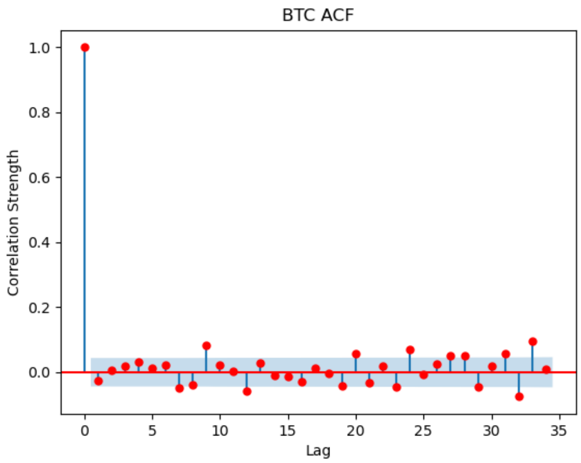

Fig. 7

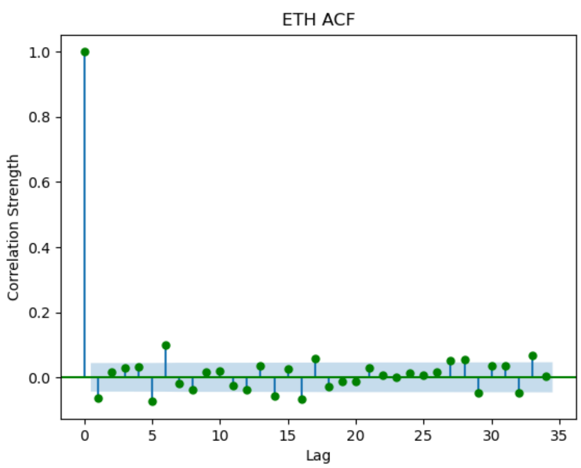

Fig. 8

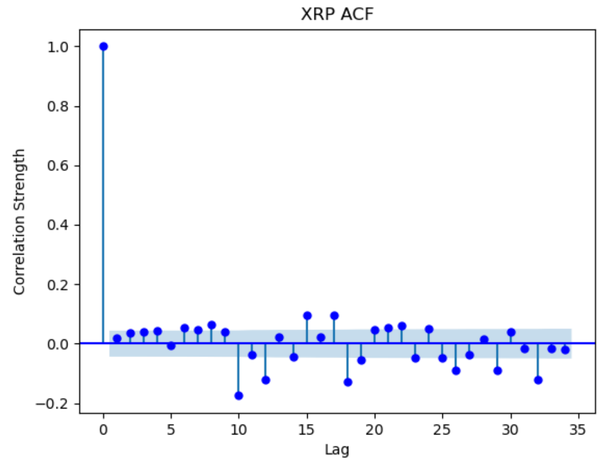

Fig. 9

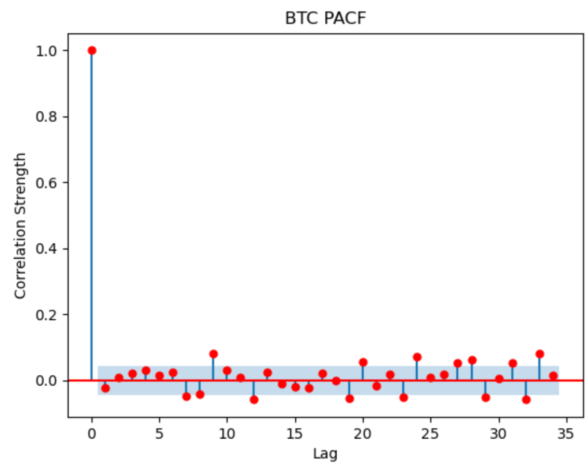

Fig. 10

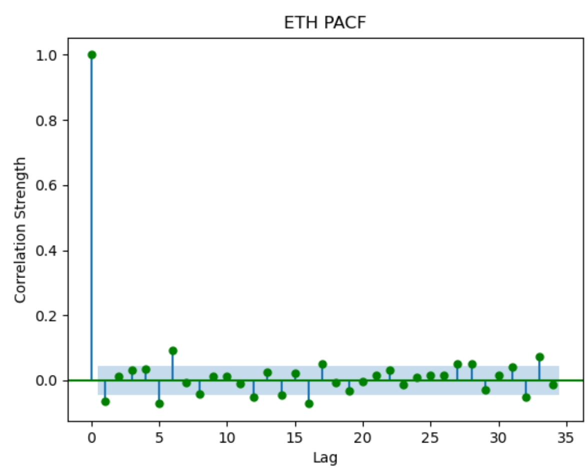

Fig. 11

D. Forecasting Bitcoin, Ethereum and Ripple

The final component of this report is the ARIMA function. It provides a forecast of BTC, ETH and XRP for the next month and 6 months. The Autoregressive Integrated Moving Average model combines autoregression, order of differencing, and the moving average to make a forecast. The autoregressive component derives the current value of the series from a linear combination of its previous values. P in the ARIMA function is the order of the autoregression, indicating how many past values are considered in the model [24]. In this case, using the ACF plots we have seen that there is strong evidence for autoregression at the first lag, therefore P=1. In terms of differencing, we have seen that first order differencing produces a series that is stationary, therefore D=1. The moving average is the linear combination of previous error terms in the forecast equation. Q in ARIMA, is the order

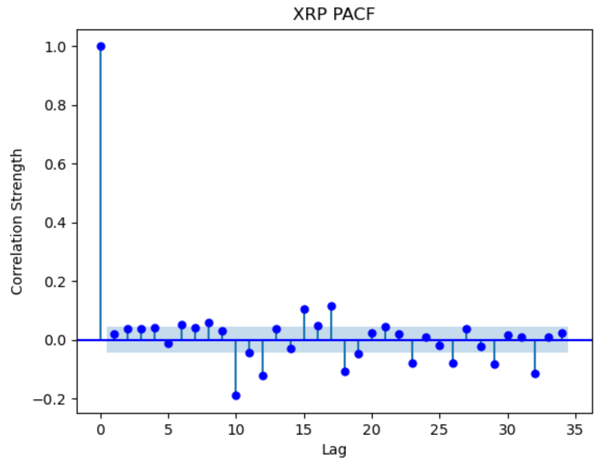

Fig. 12

of the moving average, indicating how many previous error terms are used in the model. Using the PACF plot, we set Q=1 as the significant spike is at the first lag. The PACF considers the intermediate lags and is used when measuring the number of error terms [25]. To forecast the data, ARIMA uses previous values of the time series and the estimated parameters to predict the future values. The forecasts are generated by propagating the model forwards, using the three components of ARIMA. The results of the forecast show that BTC and ETH are set to rise slightly, however the forecast for XRP is neutral. BTC is set to rise to $29,249.81 in a month and $30,898.23 in the next 6 months. ETH is set rise to $1,929.88 in a month and $2,048.62 in the next 6 months. XRP is set to remain around $0.46 in a month and $0.48 in the next 6 months.

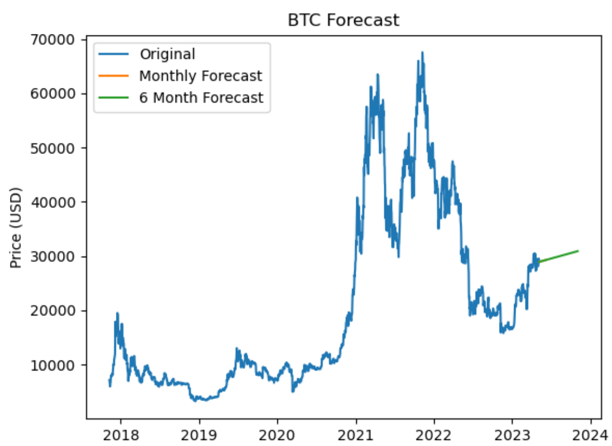

Fig. 13

E. Evaluating ARIMA

To evaluate the ARIMA model, the residuals of the ARIMA model fits for each cryptocurrency are taken to see as the reader, the goodness-of-fit of the ARIMA models to their coins. The BTC residuals appear to fluctuate around zero without any discernible patterns or trends. The ETH residuals, like the BTC

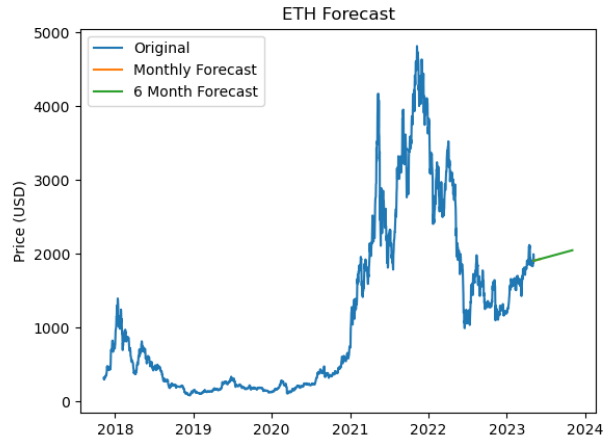

Fig. 14

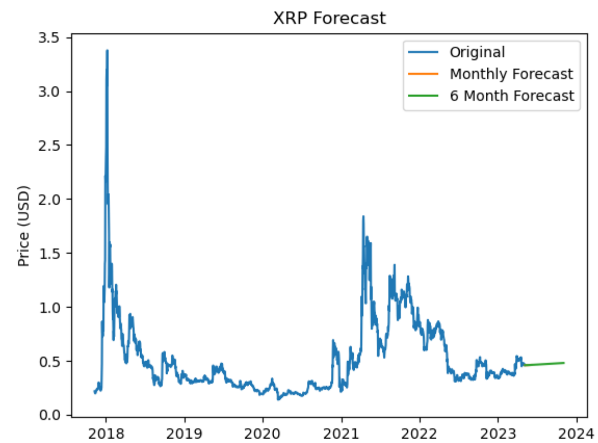

Fig. 15

residuals, they seem to oscillate around zero without any distinct patterns. Again, the XRP residuals oscillate around zero without any clear patterns. Plotting the residuals and checking for any systematic fluctuations ensures that the ARIMA models capture the underlying patterns and volatility in the data. The residuals should appear random, indicating that the model has captured the relevant information in the time series [26].

5. CONCLUSION

Overall, this report investigates the time series and time lag analysis of Bitcoin, Ethereum and Ripple price data. From taking the forecast of BTC, ETH and XRP; BTC and ETH are set to rise within the next 6 months, however XRP is set to remain neutral in price evolution. To create the forecast, the time series data needed to be backwards differenced to ensure its stationarity as well as making ACF and PACF plots to assess the correlation between the respective cryptocurrencies and their lagged values. This provided us the with the optimal parameters to run an ARIMA model on the time series data to forecast price. As per the stated objectives, the cryptocurrencies are correlated as the ACF plots showed that any price point shows a strong

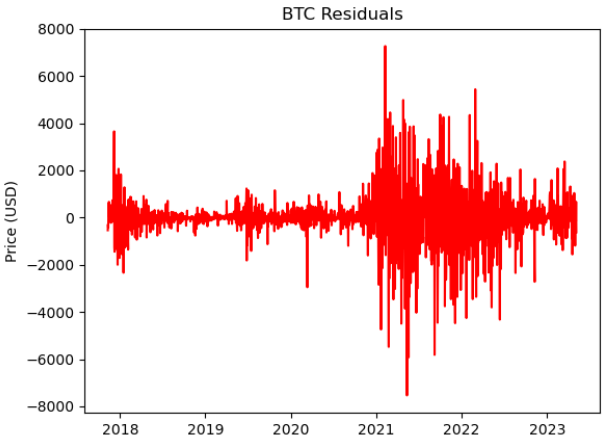

Fig. 16

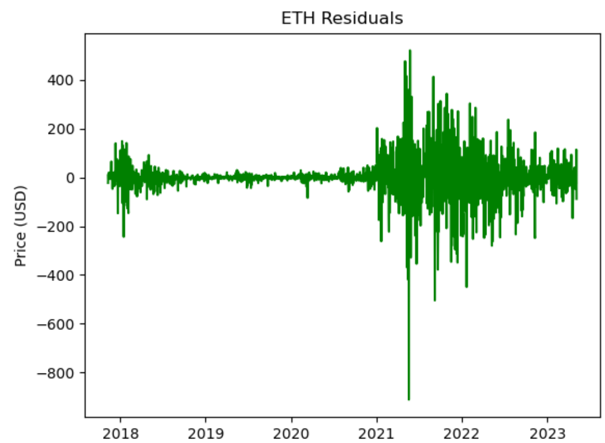

Fig. 17

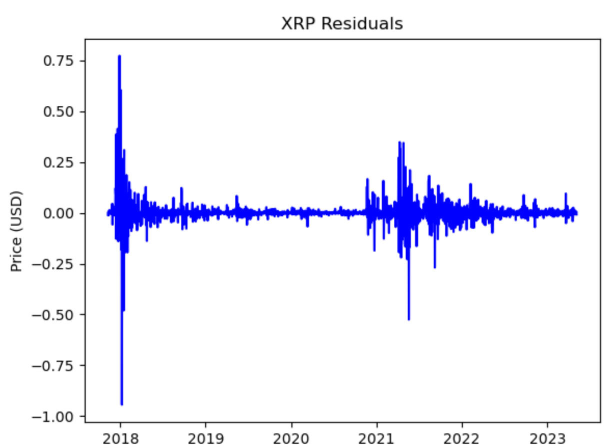

Fig. 18

dependency on the previous price point. This describes a first order autoregressive nature which was implemented into ARIMA. There is an obvious lag between the cryptocurrencies as PACF plots gave strong positive spikes for the first lag, indicating that the price of the cryptocurrencies is affected by a linear combination of the previous price points.

6. APPENDIX

The specific model used for this report was the Augmented Dickey-Fuller regression model. It estimates the presence of a unit root by considering lagged differences. The model equation is: $\Delta y(t) = \alpha + \beta t + \gamma y(t-1) + \delta_1 \Delta y(t-1) + \delta_2 \Delta y(t-2) + ... + \delta_p \Delta y(t-p) + \epsilon(t)$. $\Delta y(t)$ is the differenced BTC, ETH and XRP data. α is the constant term. β is the linear trend coefficient. γ is the lagged variable level coefficient. δ_1, δ_2, ..., δ_p are the lagged differences of the variable coefficients up to order p. $\epsilon(t)$ is the error term [27]. To estimate the ADF Regression Model, Ordinary least squares (OLS) regression is used. The t-statistic is computed for γ, which tests the null hypothesis that $\gamma = 0$. If gamma significantly deviates from zero, it suggests the absence of a unit root and supports stationarity. The ADF t-statistic gauges how strong the evidence is against the null hypothesis; it is cross-referenced against critical values derived from the Dickey-Fuller distribution to assess statistical significance. If the test statistic is lower than the critical value, the null hypothesis is rejected, indicating data stationarity. Conversely, if the test statistic exceeds the critical value, the null hypothesis is not rejected, implying non-stationarity.
The ACF is calculated as follows:

The average μ and standard deviation σ of a given time series is found. For each lag value k, the correlation coefficient is computed between the original series and the series shifted by k units of time. Each observation is multiplied at time t by the corresponding observation at time t-k, products are summed, and divided by (n-k)σ^2, where n represents the total number of observations. The resulting autocorrelation function (ACF) values range from -1 to 1 [28]. A value of 1 indicates a perfect positive correlation, 0 suggests no correlation, and -1 signifies a perfect negative correlation. The partial autocorrelation function (PACF) is obtained using the Yule-Walker equations, which involve a recursive process. For a given lag k, the PACF is determined by fitting an autoregressive (AR) model of order k to the data and extracting the coefficient associated with lag k.

7. TIME SERIES IN ACADEMIA AND INDUSTRY

Time series analysis is extensively employed in the finance and investment sectors to forecast stock prices, market trends, exchange rates, and other financial indicators. By examining historical time series data, financial institutions can identify recurring patterns and trends, enabling them to make informed decisions on investments, risk management, portfolio optimization, and trading strategies. Techniques like autoregressive integrated moving average (ARIMA) models and exponential smoothing methods are commonly utilized for financial forecasting. Time series analysis provides valuable insights for demand forecasting and inventory management in industries such as retail, manufacturing, and supply chain management. By analyzing historical sales data and other relevant factors, businesses can predict future demand, identify seasonal patterns, and make well-informed decisions regarding production planning, inventory levels, and supply chain optimization. Accurate demand forecasting aids in minimizing stockouts, reducing inventory carrying costs, and enhancing customer satisfaction. Time series analysis plays a pivotal role in predictive maintenance, particularly in industries like manufacturing, energy, and transportation. Through the monitoring and analysis of time-stamped sensor data, companies can detect anomalies and patterns indicative of equipment failures or maintenance requirements. Time series analysis techniques, such as ARIMA, exponential smoothing, or machine learning algorithms like recurrent neural networks (RNNs), facilitate the prediction of remaining useful life of machinery, optimization of maintenance schedules, reduction of downtime, and prevention of costly equipment failures [29].

Time series analysis is extensively used in the field of econometrics and macroeconomics. Researchers analyze time series data to understand and model economic variables such as GDP, inflation rates, interest rates, and unemployment rates. Time series analysis techniques help in identifying economic trends, estimating parameters of econometric models, and testing economic hypotheses. Econometric models such as autoregressive integrated moving average (ARIMA), vector autoregression (VAR), and cointegration analysis are commonly employed in academic research to study the dynamics of economic variables over time. Time series analysis is crucial in the field of environmental science and climate research. Researchers analyze time series data of environmental variables such as temperature, precipitation, atmospheric composition, and sea levels to study long-term trends, detect periodic patterns (e.g., seasonal variations), and understand climate change. Time series analysis helps in identifying and quantifying climate patterns and anomalies, assessing the impact of human activities on the environment, and making predictions about future climate scenarios. Statistical techniques like Fourier analysis, wavelet analysis, and autoregressive modeling are commonly used in climate research. Time series analysis plays a vital role in health and biomedical research. Researchers analyze time series data related to patient vital signs, disease progression, treatment outcomes, and epidemiological trends. Time series analysis enables the detection of temporal patterns in health data, identification of risk factors, prediction of disease outcomes, and evaluation of treatment effectiveness. Techniques such as autoregressive integrated moving average (ARIMA), state space modeling, and spectral analysis are frequently employed in health and biomedical research to analyze time-dependent data [30].

8. BIBLIOGRAPHY

1. Nakamoto S. Bitcoin: a peer-to-peer electronic cash system. In 2008 [cited 2023 May 15]. Available from: https://www.semanticscholar.org/paper/Bitcoin

2. Eddy TTM, Georges BB, Salomon NEP. Towards a new model for the production of civil status records using blockchain. JIS [Internet]. 2023 [cited 2023 May 15];14(01):52–75. Available from: https://www.scirp.org/journal/doi.aspx?doi=10.4236/jis.2023.141005

3. Bitcoin and cryptocurrency technologies [Internet]. 2016 [cited 2023 May 15]. Available from: https://press.princeton.edu/books/hardcover/9780691171692/bitcoin-and-cryptocurrency-technologies

4. He S. Impact of blockchain applications on trust in business. IB [Internet]. 2020 [cited 2023 May 15];12(03):103–12. Available from: https://www.scirp.org/journal/doi.aspx?doi=10.4236/ib.2020.123007

5. Brockwell PJ, Davis RA. Introduction to time series and forecasting [Internet]. Cham: Springer International Publishing; 2016 [cited 2023 May 15]. (Springer Texts in Statistics). Available from: http://link.springer.com/10.1007/978-3-319-29854-2

6. Time series analysis: forecasting and control, 5th edition | wiley [Internet]. Wiley.com. [cited 2023 May 15]. Available from: https://www.wiley.com/en-gb/Time+Series+Analysis

7. Forecasting: principles and practice(2nd ed) [Internet]. [cited 2023 May 15]. Available from: https://otexts.com/fpp2/

8. Bouri E, Gupta R, Lau CKM, Roubaud D, Wang S. Bitcoin and global financial stress: A copula-based approach to dependence and causality in the quantiles. The Quarterly Review of Economics and Finance [Internet]. 2018 [cited 2023 May 15];69(C):297–307. Available from: https://ideas.repec.org//a/eee/quaeco/v69y2018icp297-307.html

9. Kjærland F, Khazal A, Krogstad EA, Nordstrøm FBG, Oust A. An analysis of bitcoin's price dynamics. Journal of Risk and Financial Management [Internet]. 2018 Dec [cited 2023 May 15];11(4):63. Available from: https://www.mdpi.com/1911-8074/11/4/63

10. Cryptocurrency prices, charts and market capitalizations [Internet]. CoinMarketCap. [cited 2023 May 15]. Available from: https://coinmarketcap.com/

11. Introduction [Internet]. [cited 2023 May 15]. Available from: https://docs.etherscan.io/

12. Crypto solutions for business | ripple [Internet]. [cited 2023 May 15]. Available from: https://ripple.com/

13. Pabuçcu H, Ongan S, Ongan A, Pabuçcu H, Ongan S, Ongan A. Forecasting the movements of Bitcoin prices: an application of machine learning algorithms. QFE [Internet]. 2020 [cited 2023 May 15];4(4):679–92. Available from: http://www.aimspress.com/rticle/doi/10.3934/QFE.2020031

14. Fracassi C, Kogan S. Pure momentum in cryptocurrency markets [Internet]. Rochester, NY; 2022 [cited 2023 May 15]. Available from: https://papers.ssrn.com/abstract=4138685

15. Gupta H, Chaudhary R. An empirical study of volatility in cryptocurrency market. Journal of Risk and Financial Management [Internet]. 2022 Nov [cited 2023 May 15];15(11):513. Available from: https://www.mdpi.com/1911-8074/15/11/513

16. Singh V, Chen SS, Singhania M, Nanavati B, kar A kumar, Gupta A. How are reinforcement learning and deep learning algorithms used for big data based decision making in financial industries–A review and research agenda. International Journal of Information Management Data Insights [Internet]. 2022 Nov 1 [cited 2023 May 15];2(2):100094. Available from: https://www.sciencedirect.com/science/article/pii/S2667096822000374

17. Bhardwaj, V., Kumari, N. (2021). Cryptocurrency Price Volatility: Determinants and an Empirical Study. Journal of Economic Structures, 10(1), 1-17. DOI: 10.1186/s40008-021-00284-0. 18. Tsantekidis, A., Passalis, N., Tefas, A., Kanniainen, J. (2019). Forecasting Cryptocurrency Value by Sentiment Analysis: An Heterogeneous Ensemble Learning Approach. Expert Systems with Applications, 118, 533-543. DOI: 10.1016/j.eswa.2018.09.047. 19. Mehta, N., Yadav, P., Kumar, A. (2021). Forecasting Cryptocurrency Prices Using Time Series Analysis and Long Short-Term Memory Networks. SN Computer Science, 2(6), 1-11. DOI: 10.1007/s42979-021-00832-4. 20. Dickey, D. A., Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association, 74(366a), 427-431. 21. Hamilton, J. D. (1994). Time series analysis. Princeton university press.

22. Brockwell, P. J., Davis, R. A. (2016). Introduction to time series and forecasting. Springer. 23. Cryer, J. D., Chan, K. S. (2008). Time series analysis: with applications in R (2nd ed.). Springer. 24. Cryer, J. D., Chan, K. S. (2008). Time series analysis: with applications in R (2nd ed.). Springer. 25. Brockwell, P. J., Davis, R. A. (2016). Introduction to time series and forecasting (3rd ed.). Springer. 26. Hyndman, R. J., Athanasopoulos, G. (2018). Forecasting: Principles and Practice (2nd ed.) 27. Hamilton, J. D. (1994). Time series analysis. Princeton university press.