

https://dspace.org/

Testing the adaptive efficiency of bitcoin

Maredi, Maromo

2022
https://hdl.handle.net/10539/35256

Downloaded from a DSpace repository. https://dspace.org/

SCHOOL OF ECONOMICS AND FINANCE
FACULTY OF COMMERCE, LAW AND MANAGEMENT



WITS
UNIVERSITY

Testing The Adaptive Efficiency Of Bitcoin

Maromo Maredi

Supervisor

Prof Yudhvir Seetharam

A research report submitted to the Faculty of Law, Commerce and Management, University of the Witwatersrand in partial fulfilment of the requirements for the degree of Master of Commerce (50%) in Finance.

Johannesburg, South Africa

February 2022

Testing the Adaptive Efficiency of Bitcoin

Abstract

This research aims to investigate an alternative view of market dynamics referred to as the Adaptive Markets Hypothesis which posits that an asset's efficiency will change over time. As such, this research will test whether Bitcoin is time-varyingly efficient. This will be accomplished in three stages. Firstly, whether Bitcoin returns follow a random walk/martingale will be investigated. If they do, that means that they cannot be predicted, thereby providing evidence of the weak-form market efficiency. If they do not follow a random walk, however, the second phase of the investigation turns to whether they can be modelled. The first attempt models the current Bitcoin return as a function of its own lagged values, which is predicated the idea of all relevant information being reflected in historical returns. The inadequacy of this model in its description of the returns generating process, provides evidence that there is private information that historical returns do not reflect which impacts returns. To account for this, the returns generating process is thus modelled using both historical returns and exogenous lagged variables without need to specify the model's functional form. If the model performs better in some periods than in others, it can be inferred thus that Bitcoin is time-varyingly efficient.

Declaration

I, Maromo Maredi, declare that this research paper is my own work and that I have correctly acknowledged the work of others. It is submitted to fulfil the requirements for the degree of Master of Commerce in Finance at the University of the Witwatersrand, Johannesburg. I declare that this research paper has not been submitted for any other degree or examination in this or any other institution.

Maromo Maredi

February 2022

Dedication

To my mothers, Meta, Modiledi and Mohlapa Maredi and my late grandmothers, Shatadi Matlejoane, Pheladi Matlejoane, Pheladi Maredi and Mateadi Maredi.

Acknowledgements

There are several who have supported me throughout the life of this project and have backed me to the hilt. Prof Yudhvir Seetharam, words cannot express how grateful I am to you. You have been a constant source of inspiration and advice throughout what has turned out to be an interesting and wonderful journey. You challenged me to always strive for the best and never rest on my laurels. I would also like to thank Dr Bevan Smith, whose advice and input I am ever thankful for.

I would also like to thank family. My mother, Mokgohloe Maredi, who is my role model and confidante. I am all that I am because you never stopped encouraging me to believe in myself, even (and perhaps, especially) when things became difficult. This research report is the product of the many nights you called and found me working. To my sister, Octavia Maredi, you are the world's best older sister. Thank you for always having my back.

This project would also not have been impossible without the external funding from BankSETA, who have, throughout my masters, ensured that my fees were not a factor in whether I could acquire further training in a field I am deeply passionate about. In the same vein, I would also like to thank Prof Hofmeyr and his team at Study Trust, who have throughout my undergraduate and honours studies, provided me with funding.

Finally, I would like to thank my friends, Kgadile Masemola, Sanele Gumede, Thabisile Mahatlane and Hayley Ncube for their constant support on this journey.

Table of Contents

Declaration ii
Dedication iii
Acknowledgements iv
List of Figures vii
List of Tables viii
Definition of abbreviations, terms and symbols x
1. INTRODUCTION 1
1.1 Background 1
1.2 Research gap and objectives 2
1.2.1 Motivation of the study 2
1.2.2 Research objective(s) and question(s) 3
1.2.3 Potential benefits 3
1.2.4 Hypotheses testing 4
1.3 Structure of the study 4
2 LITERATURE REVIEW 5
2.1 The Efficient Markets Hypothesis 5
2.2 The Non-Assumption Issues with the Efficient Markets Hypothesis 7
2.3 A dynamic approach to efficiency: The AMH 9
2.4 Tests of the AMH in general cryptocurrency markets 11
2.5 Tests of Bitcoin Efficiency 12
2.6 A Review Of Return Forecasting Techniques 14
2.7 The Efficient Markets Hypothesis 18
2.8 Summary of literature review 19
3 DATA AND METHODOLOGY 20
3.1 Data 20
3.2 Methodology 20
3.2.1 Linearity Tests 22
3.2.2 Normality Tests 22
3.2.3 Stationarity Tests 23
3.3 Robustness Checks 24
3.4 Summary 24
4 RESULTS 25
4.1 Normality Tests 25
4.2 Stationarity Tests 26
4.3 Martingale Difference Hypothesis Tests 27

4.4 Linearity Tests 29
4.5 SETAR/ARIMA modelling without external regressors 31
4.6 Fourier Transform 38
4.7 Machine Learning Approaches to Modelling the RGP 48
4.8 Machine Learning Approaches to Modelling the RGP (FT) 53
4.9 NARX Modelling 59
4.9.1 External Regressors 60
4.9.2 Daily and Weekly NARX Models 65
4.9.3 Daily and Weekly NARX Models (FT) 69
4.10 Summary 73
5 CONCLUSION 75
5.1 Limitations 77
5.2 Areas of future research 77
REFERENCES 79
Appendix A: Further Bitcoin Analysis 86

List of Figures

Figure 4.1: Daily SETAR grid search 31
Figure 4.2: Weekly SETAR grid search 31
Figure 4.3: Daily SETAR Model Residuals 35
Figure 4.4: Weekly SETAR Model Residuals 35
Figure 4.5: Daily Model Residuals' ACF 36
Figure 4.6: Weekly Model Residuals' ACF Plot 36
Figure 4.7: Daily AR Model Residuals 46
Figure 4.8: Daily AR(4) Residuals' ACF Plot 46
Figure 4.9: Weekly SARIMA Residuals 47
Figure 4.10: Weekly SARIMA Residuals ACF 47
Figure 4.11: Daily LSTM Diagnostics 49
Figure 4.12: Daily Residuals ACF Plot 50
Figure 4.13: Weekly LSTM Diagnostics 51
Figure 4.14: Weekly LSTM Residuals ACF Plot 51
Figure 4.15: Daily LSTM Diagnostics (FT) 54
Figure 4.16: Daily Residuals Plot 55
Figure 4.17: Daily Residuals ACF Plot 55
Figure 4.18: Weekly LSTM Diagnostics (FT) 57
Figure 4.19: Weekly Residuals Plot 58
Figure 4.20: Weekly Residuals ACF Plot 58
Figure 4.21: MSCI World Index 61
Figure 4.22: Disaggregated MSCI World Index 61
Figure 4.23: Google Trends Monthly Data 62
Figure 4.24: Disaggregated Bitcoin Trend Data 62
Figure 4.25: Federal Funds Rate 63
Figure 4.26: Disaggregated Federal Funds Rate 64
Figure 4.27: US Annual GDP 64
Figure 4.28: Disaggregated US GDP 65
Figure 4.29: Daily NARX Model 66
Figure 4.30: Weekly NARX Residuals 67
Figure 4.31: Daily NARX Residuals ACF 67
Figure 4.32: Weekly NARX Residuals ACF 68
Figure 4.33: Daily NARX Residuals 70
Figure 4.34: Weekly NARX Residuals 71
Figure 4.35: Daily NARX Residuals' ACF 71
Figure 4.36: Weekly NARX Residuals ACF 72

List of Tables

Table 4.1: Daily Returns' Normality Tests 25
Table 4.2: Weekly Returns' Normality Tests 25
Table 4.3: Daily Returns' Stationarity Tests 26
Table 4.4: Weekly Returns' Normality Tests 26
Table 4.5: Martingale Difference Hypothesis Tests - Daily Returns 27
Table 4.6: Hurst Exponent - Daily Returns 28
Table 4.7: Martingale Difference Hypothesis Tests – Weekly Returns 29
Table 4.8: Hurst Exponents – Weekly Returns 29
Table 4.9: BDS Tests – Daily Returns 30
Table 4.10: BDS Test - Weekly Returns 30
Table 4.11: Best Daily SETAR Model Inputs 32
Table 4.12: Best Daily SETAR Model Inputs 32
Table 4.13: Daily SETAR Model 33
Table 4.14: Weekly SETAR Model 34
Table 4.15: Daily Residuals' Normality Tests 37
Table 4.16: Hurst Exponent – Daily Residuals 37
Table 4.17: Weekly Returns' Normality Tests 38
Table 4.18: Weekly Model Residuals Hurst Exponent 38
Table 4.19: Martingale Difference Hypothesis Test – Daily Returns 39
Table 4.20: Hurst Exponent - Daily Returns 40
Table 4.21: BDS Test - Daily Returns 40
Table 4.22: SETAR Test 41
Table 4.23: Martingale Difference Hypothesis - Weekly Returns 41
Table 4.24: Hurst Exponent - Weekly Returns 42
Table 4.25: BDS Test - Weekly Returns 42
Table 4.26: Daily ARIMA Model 43
Table 4.27: Weekly SARIMA Model 43
Table 4.28: Daily Model Residual Normality Tests 44
Table 4.29: Daily Residuals Hurst Exponents 44
Table 4.30: Weekly Model Residuals Normality Tests 45
Table 4.31: Weekly Residuals Hurst Exponents 45
Table 4.32: Daily LSTM Model 48
Table 4.33: Weekly LSTM 49
Table 4.34: Martingale Hypothesis Test - Daily Residuals 52
Table 4.35: Daily Residuals Hurst Exponent 52
Table 4.36: Martingale Hypothesis Test – Weekly Residuals 53
Table 4.37: Weekly Residuals Hurst Exponent 53
Table 4.38: Daily LSTM Model 54
Table 4.39: Martingale Hypothesis Test - Daily Residuals 56
Table 4.40: Daily Residuals Hurst Exponent 56
Table 4.41: Weekly LSTM 57
Table 4.42: Martingale Hypothesis Test – Weekly Residuals 59
Table 4.43: Daily Residuals Hurst Exponent 59

Table 4.44: Daily NARX Model 65
Table 4.45: Weekly NARX Model 66
Table 4.46: Martingale Difference Hypothesis Tests - Daily Residuals 68
Table 4.47: Martingale Difference Hypothesis Test – Weekly Residuals 69
Table 4.48: Daily NARX Model 69
Table 4.49: Weekly NARX Model 70
Table 4.50: Martingale Difference Hypothesis Tests - Daily Residuals 72
Table 4.51: Martingale Difference Hypothesis Tests – Weekly Residuals 73

Definition of abbreviations, terms and symbols

Adaptive Markets Hypothesis: An expanded view of market efficiency that makes provision for changes in the market's underlying conditions.

Bitcoin: An internet-based currency whose authenticity is governed by cryptographic evidence.

Cryptocurrency: A class of internet-based currencies that are backed by cryptographic evidence that is presented to an online ledger. Their exchange is based on near-untraceable addresses.

Adaptive Efficiency: A view of efficiency from Lo (2004) that changes over time in line with changes in the market's conditions, including the level of competition and access to arbitrage opportunity. This is used interchangeably with "time-varying efficiency" and "cyclical efficiency".

Homo economicus: An idealized human being whose decision making is perfectly rational. Homo economicus is often invoked in economic literature to simplify models.

Hurst Exponent: A measure of dependency in a time series. A Hurst Exponent higher than 0.50 indicates that the time series under investigation exhibits long-memory.

Long Short-Term Memory (LSTM): A class of artificial recurrent neural networks that is specialized in processing input sequences and making a contextual prediction about the next step in the sequence. They are often used in time series forecasting.

Martingale: A statistical process where the best predictor of the process' next value is its current value.

Mean Absolute Percent Error (MAPE) criterion: A measure of the predictive accuracy of a (regression) model.

Recurrent Neural Network (RNN): A class of models that use their memory to enable them to process input information and employ it in their prediction of an output.

Satisficing: A process where an individual does not pick the most optimal solution to a problem, but instead chooses a satisfactory solution.

SETAR: A time series approach to modelling a process. It differs from the traditional autoregressive models in its ability to capture regime switches.

1. INTRODUCTION

1.1 Background

Cryptocurrencies have been the subject of intense academic debate: are they a currency, an asset, something that possesses the features of both, or an entirely different category of security altogether? (Böhme *et al*., 2015) In contrast to conventional currencies backed by central authorities, cryptocurrencies are digital currencies whose exchange is backed by cryptographic evidence from a network of linked computers, which are collectively referred to as miners (Selgin, 2015). Despite their introduction to the world as a currency, questions remain about their effectiveness as a store of value and whether they will ever acquire mass acceptance as a form of payment (Bariviera *et al.*, 2017).

In a similar fashion, Bitcoin has attracted plenty of debate and scrutiny. The initial commentary regarding characterised it as either a fad or tool for the exclusive use of those interested in criminal activity (Rose, 2015). Bitcoin has also been accused of being a wholesale fraud that is bound to collapse given its small market relative to other established currencies such as the United States Dollar (USD) (Monaghan, 2017). Over time, that initial spirit of derision has, however, given way to a more measured approach to the study of Bitcoin's properties, such as whether it is informationally efficient (Kristoufek, 2018a).

Fama (1970) defines an informationally efficient market as one in which the return of the asset in the market reflects all available information. Efficiency can take any of the three distinctive forms based on the information the asset incorporates: a weak, semi-strong as well as a strong form. Weak-form efficiency, for example, describes a scenario where a security's historical return and returns data is reflected in the security's current return.

Whether a single security (that exists alongside similar assets) can be tested for "market" efficiency is a relevant point of discussion. While the term "market" is often associated with an assortment of goods and services for sale, there can exist a market for a singular asset, as evidenced by the market for, say, gold. To fully draw out the *reductio ad absurdum*, it can be noted that the gold market is not made any less real by the fact that precious metals (and therefore a market for precious metals) as a real category, exists. In sum, the existence of a market for a category of assets does not preclude the existence of a market for a singular asset.

Moreover, the extant literature is awash in studies of singular assets, such as studies that test whether gold is weak form efficient (Tschoegl, 1980).

In a similar vein, studies have examined whether Bitcoin is weak form efficient (Kristoufek, 2018a). While the conclusion of many of these studies has been that Bitcoin is not weak form efficient in the manner envisaged by Fama (1970), namely that Bitcoin returns do not follow a random walk/martingale (*i.e.*, past information is useless in predicting the future Bitcoin returns,) a new framework for understanding market efficiency has come into focus: the Adaptive Markets Hypothesis (AMH). It posits that efficiency is not a binary state, where an asset is either efficient or not. Instead, it argues that efficiency changes over time in reaction to changes in liquidity, the number of market participants and competition. Therefore, while Bitcoin may not fit the description of efficiency as envisaged by Fama (1970), it may well fit the Lo (2004) view of efficiency. This study aims to investigate whether it does.

1.2 Research gap and objectives

1.2.1 Motivation of the study

This study will approach Bitcoin, the largest cryptocurrency by market capitalization, as an asset. To this end, this study aims to address the market efficiency of Bitcoin from an Adaptive Markets Hypothesis (AMH) perspective. In essence, it will test whether the market efficiency of Bitcoin changes over time. Whilst tests of the informational efficiency of Bitcoin have been conducted, this study aims to investigate whether Bitcoin is cyclically efficient using the robust framework developed in Seetharam (2016).

Furthermore, several events have occurred which could have a significant impact on the efficiency of Bitcoin. The world-wide spread of Covid-19 has seen a marked increase in interest regarding cryptocurrencies, leading to Bitcoin's market capitalization reaching over $700 billion as of May 2021 (Inman, 2020). The resultant financial climate — where relaxed fiscal and monetary policies have been become a mainstay — has led many investors to look to Bitcoin as a "safe haven asset" and a source of exposure to an asset class that is imperfectly correlated with their portfolios (Shahzad *et al*., 2019).

Finally, cryptocurrency exchanges have also grown during the pandemic, with Coinbase (the leading exchange) ultimately listing to extremely positive reception on the New York Stock Exchange (NYSE) in April 2021 (Browning, 2021). The existence of these exchanges has led to the mainstreaming of what was once a "rogue and subculture" asset by opening the door to investors to easily acquire cryptocurrencies (Crabb, 2021).

This provides a ripe opportunity to study whether Bitcoin fits within the view of efficiency as envisaged by the AMH.

1.2.2 Research objective(s) and question(s)

The objective of this research report is to test whether Bitcoin is informationally efficient in a manner consistent with the AMH. The AMH predicts varying degrees of efficiency and inefficiency over time; as such, there will be varying levels of asset return change predictability (Lo, 2004). Due to changes in these factors over time, it is expected that the data will exhibit time-varying levels of predictability. As such, this study will investigate whether Bitcoin has varying levels of efficiency over time.

1.2.3 Potential benefits

While there have been many studies have tested whether the AMH holds with respect to traditional asset such as stocks and futures, much less focus has been devoted to testing whether it is a good explanation of market efficiency in the non-traditional asset classes such as cryptocurrency. As a result, this study is aimed to address that gap and adapt models that have shown success in showing cyclical market efficiency in traditional asset classes (using support vector machines).

Secondly, given the elevated levels of attention and scepticism that Bitcoin and other cryptocurrencies have received in both the academic literature and society in general, it is imperative that further light is shed on Bitcoin. While this study does not delve into the mechanics behind deciding whether to include cryptocurrencies in the portfolio formation process, it nevertheless is aimed at demystifying the cryptocurrency world. Having a clearer picture of the properties of cryptocurrencies allows for a more sober conversation regarding them, particularly as they have gained more significance not just in academia, but in society in

general. El Salvador just became the first sovereign nation to adopt Bitcoin as legal tender, other nations are also investigating whether they should launch virtual versions of their state currencies.

1.2.4 Hypotheses testing

Primary Hypothesis

H_0: Bitcoin efficiency is not cyclical.

H_1: Bitcoin efficiency is cyclical.

Secondary Hypotheses

$H_{0,A}$: Bitcoin returns do not follow a random walk/martingale.

$H_{1,A}$: Bitcoin returns do follow a random walk/martingale.

$H_{0,B}$: Bitcoin returns cannot be modelled via an autoregressive function with no exogenous inputs.

$H_{1,B}$: Bitcoin returns can be modelled by an autoregressive function with no exogenous inputs.

$H_{0,C}$: Bitcoin returns cannot be modelled via an autoregressive function with exogenous inputs.

$H_{1,C}$: Bitcoin returns can be modelled by an autoregressive function with exogenous inputs.

1.3 Structure of the study

This research is structured in five chapters and will proceed as follows: Chapter 2 surveys the efficiency hypothesis literature, from the Efficient Markets Hypothesis (EMH) to its continuation, the Adaptive Markets Hypothesis, and ends with an examination of the asset return forecasting literature. Chapter 3 describes the data, variables, hypotheses and methodology of the research report. This delves into the data and methods that test for whether the Adaptive Markets Hypothesis is a good explanation of Bitcoin return behaviour as well as presenting several models that are used to determine the return generating process. Chapter 4 shows the empirical results of the research report and provides a discussion of the results Chapter 5 concludes the research report.

2 LITERATURE REVIEW

Whether one can predict returns of assets is a question whose answer possesses much value for those asking it. This is because if return prediction is possible, then risk-free profits can be extracted by exploiting one's ability to predict returns. The academic literature is replete with articles that attempt to answer this question, the dominant and novel approaches are discussed below:

2.1 The Efficient Markets Hypothesis

The Efficient Markets Hypothesis (EMH) is a description of market dynamics in a state of equilibrium — an efficient market always fully reflects available information (Fama, 1970). The EMH is subdivided into three forms based on the amount of information that is incorporated in the asset's return: the weak form, semi-strong form as well as its strong form (Lawrence *et al*., 2007). In its weak form, the EMH describes a state in which all the historical returns and returns data of the asset is incorporated into its return. The returns of the security therefore follow a random walk — returns are independent and identically distributed, *i.e.*, a return at one point in the time series does not affect another and the returns are drawn from the same distribution.

This is in line with the argument advanced by Bachelier (1900) that market speculation should be a "fair game": the expected value of engaging in market speculation should be equal to zero, *i.e.*, price changes follow a martingale. An asset's return series follows a martingale when the best predictor of future returns, given the current information set, is the unconditional expectation of today's price (Escanciano & Velasco, 2006). As such, like the random walk view of efficiency, one cannot use past information to predict future returns when the asset's return series follows a martingale. While the random walk and martingale hypotheses are distinct, Fama (1970) argues for their fungibility in discussing weak-form efficiency as, in both cases, they predict that an asset "fully reflects" the information set when the use of trading rules does not yield above buy-and-hold returns. As such, the weak-form of the EMH negates the use of technical analysis, the use of historical data in predicting future asset returns (Malkiel, 2003).

In the semi-strong form, all publicly available information is incorporated into the asset's return (Fama, 1970). As a result, the semi-strong form of the EMH would negate the use of fundamental analysis — the analysis and use of financial information of the assets to predict stock returns (Malkiel, 2003). In its strong form, the EMH describes a state in which, all available information (both private and public) is reflected in the asset's return. This would negate the use of private information in the prediction of future asset returns.

Empirical evidence for and against the EMH spans many asset classes in the literature. Grater and Struweg (2015) examine logarithmic returns of stocks on the Johannesburg Stock Exchange (JSE) and find evidence of a unit root, thereby rejecting the null of weak-form efficiency. In examining the JSE for weak-form efficiency, however, Heymans and Santana (2018) present evidence that contradicts the earlier evidence brought forth by Grater and Struweg (2015). Their findings suggest that while the overall JSE is weak-form efficient, some of its sub-indices are not. This presents *prima facie* evidence that runs counter to Samuelson's dictum: the postulation that the market is "micro-efficient" but not efficient at aggregate stock market level (Jung & Shiller, 2005).

Informational efficiency studies in other asset classes have also produced mixed results. Kristoufek and Vosvrda (2014) find that while energy commodities (such as heating oil) are weak-form efficient, agricultural commodities are not. Arouri *et. al* (2012) examined whether precious metal and energy markets are weak-form efficient. Their results indicated that in the short-run, the null of efficiency is rejected for these markets, while presenting mixed outcomes in the long-run.

While the EMH been widely regarded as the dominant model for describing market dynamics, it has also attracted criticism. Its numerous and overly generous assumptions have been the centre of controversy. It portrays human beings as infallible Bayesian agents, agents with well-calibrated expectations that are updated in line with new information, who act in their own best interests. (Sandroni, 2005) While this description is flattering given that we are human, it is demonstrably wrong. For instance, Simon (1956) shows that instead of engaging in complete optimization, humans instead "satisfice" (opt for solutions which may not be fully optimal but are nevertheless sufficient in order to achieve whatever aims the individual has). Similarly, Tversky and Kahneman (1991) formally demonstrate that humans exhibit biases such as loss aversion. The existence of these

limitations in human computational ability fundamentally undermines a realistic prospect of the EMH holding in the real world (Lo, 2004). It stands to reason thus, that a model built on such faulty assumptions can therefore be discarded.

Besides the computational limits that prohibit us from behaving in a similar fashion to the *homo economicus* that inhabits the economic literature, further limits exist which similarly threaten the EMH paradigm: limits to arbitrage (Thaler, 1999). While not explicitly stated, the EMH precludes persistent arbitrage opportunities since informed speculators can exploit them, thus nullifying them in the process (Alvarez-Ramirez *et al.*, 2012). In practise however, due to lack of access to capital, a clearly risk-free opportunity remains unexploited. As a premier example Thaler (1999) refers to LTCM Capital and their Siamese-Twin (Royal Dutch/Shell) risk-free investment: due to dislocations elsewhere in the market, they imploded before realizing a substantial return on rightfully predicting that the Royal Dutch/Shell share return misalignment would be corrected. Similarly, Khandani and Lo (2007) examine the 2007 hedge fund (statistical arbitrage) crunch and note dislocations in other parts of the market than those arbitrageurs are invested in may severely impact their ability to execute otherwise risk-free and profitable strategies. This shows that sometimes, external factors can inhibit market participants' ability to exploit risk-free opportunities. This should be viewed as a prima facie threat to the viability of the EMH.

2.2 The Non-Assumption Issues with the Efficient Markets Hypothesis

A conundrum facing any proponent of the EMH is the arbitrage paradox. It follows from the weak-form EMH that historical return data should therefore be of no use in forecasting future returns (Lucas, 1978). Indeed, Lo (2004) argues that if the EMH is correct, then there is no room for arbitrage. Given this, then no incentive would exist for investors to gather information that is useful in their quest towards finding risk-free profits. Given that information-gathering agents are a prerequisite of informationally efficient markets, the absence of arbitrageurs, would therefore result in the collapse of the markets (Grossman & Stiglitz, 1976). This paradox (information gatherers are required and yet are not rewarded with risk-free profits and therefore do not exist) illustrates that there is a fundamental problem with the EMH.

In the same paradoxical vein, Lo (2004) jokes that a classical economist walking down a crowded street would be forced to abandon a low-lying $100 because they would reason, in an EMH world, that "if it were real, someone else would have picked it up". If everyone applied this logic, the $100 would remain uncaptured because no one would pick it up, in the same way, markets would not be efficient since arbitrage opportunities would remain unexploited if no arbitrageurs existed. Thus, perfectly efficient markets are an impossibility.

Another fundamental issue that undermines the EMH is excess volatility. While returns should (in a rational world) only occur due to the arrival of news, Shiller (1981) notes that the movement in aggregate stock returns is in excess of what can be explained by the EMH. Thaler (1999) summarized it thus: "bond and stock returns are more volatile than advocates of rational efficient market theory would predict". It is an indictment that a theory whose stated aim is to explain what Fama (1970) termed the "accumulation of evidence" fails to explain evidence before it.

There are other predictive failures that have hampered the EMH: the random walk model predicts that asset returns cannot be predicted using publicly available data. However, much evidence has been adduced which shows that some widely known seasonal patterns persist in asset returns. Among them is the (in)famous January effect where, returns of companies that previously performed poorly in the previous 11 months drastically improve in January (Thaler, 1987).

There have been some attempts to explain the persistence of the January effect including the tax-loss selling hypothesis - the notion that money managers sell-off "loser" stocks in order to realize their losses on them with the intent of minimizing their tax burden (Haug & Hirschey, 2006). In other academic circles, the "window dressing" hypothesis — money managers sell off "losers" in order to have a portfolio of "winners" in time for their reporting period and pick up those assets after their reporting period, causing improved demand and thus, superior returns — has also found favour (Maxwell, 1998). There is evidence that the January effect is a persistent, global anomaly (Shen *et al.*, 2020). Without litigating the validity of each of its proposed explanations, one may recognize that its mere existence threatens the EMH. It is a barefaced counterexample to the EMH: one can predict, based on publicly available data that poorly performing shares will perform better in January.

Finally, many have argued that the EMH is not testable, and thereby falsifiable. Karl (2013) argues that the line between science and pseudoscience is in whether the models can be falsified; and the EMH given our inability to falsify it, is therefore relegated to the leagues of pseudoscience. To remedy this, a more dynamic model that is inspired by complex systems, evolutionary psychology and behavioural ecology is developed that harmonizes the EMH with these and other issues and is (crucially) falsifiable: the AMH (Lo, 2004).

2.3 A dynamic approach to efficiency: The AMH

Instead of treating the frailties of the EMH as deviations from otherwise efficient markets, or irrelevances to be rationalized away, Lo (2004) argues for a model of efficiency that also accounts for the EMH anomalies, termed the Adaptive Markets Hypothesis (AMH). This model incorporates various subjects inter alia; Behavioural Ecology, Complex Systems, Sociology and Psychology (Seetharam, 2016).

At its point of departure, the AMH argues that humans have evolved to make decisions that maximise the survival probability of their genetic material (Lo, 2004). In contrast, the current orthodox view in Economics is that humans make decisions which are rationally optimal (Cox & Hayne, 2006). While those views are not inherently mutually exclusive (the survival imperative is not manifestly opposed in any way to rational optimization) there is some degree of difference between the choice architecture in each model. This distinction is encapsulated in the idea that instead of making fully optimized choices, humans are instead engaged in "satisficing" — optimization subject to constraints such as time or computational ability (Simon, 1956). This can be distilled to "trying as best as humanly possible to make the best decision" (Seetharam, 2016).

An evolutionary answer to how an individual begins to "satisfice" is converged upon through the processes of random mutation and natural selection: The process begins when an individual attempts to solve a problem by implementing a random solution, if it works (survives the selection process), it is then kept, if it fails, then it is rejected (Lo, 2004). As a result, several permutations of random solutions with refinements are selected for (or against) based on environmental pressures are implemented until a satisfactory solution is converged upon. This solution, given a stable

environment, propagates and becomes entrenched. The individual, faced with a similar-enough problem will utilize the satisfactory solution. Once the terrain is sufficiently distinct such that the current solution is not applicable, a new solution needs to be found. The above description captures how human beings (and agents in financial markets) learn.

Lo (2004) further argues that not changing tack in the face of ecological changes — continuing to apply a previously-successful solution to an environment for which that solution is not satisfactory — is tantamount to "irrationality". In the financial context, this would emerge in such a context as, say, the application of a trading rule past its useful shelf-life. Black (1986) argues that in the financial jungle, funds are distributed away from those who mistake noise for a signal and trade on it (so-called "irrational traders") to information gatherers who trade on genuine signals. This creates an incentive for information gathering and ensures that as much arbitrage is mopped up as possible, thus reducing overall sub-optimal behaviour within the markets (Lo, 2004). It therefore stands to reason that survival will be arrogated by those whose financial frameworks are "fit", or in this case, "satisfactory".

In sum, the AMH holds that market efficiency is a function of the environment — the number of market participants, the levels of competition among those participants and the availability of risk-free opportunities (Lo, 2004). As such, the AMH conceives of efficiency in a dynamic fashion and argues that it instead follows a cyclical process in line with changes in the environment. When opportunities for profit are plentiful and competition is low, there is low efficiency and when harsher conditions set in, market efficiency ticks up (Noda, 2016).

While the AMH and EMH both employ the term "efficiency", what each hypothesis refers to is unique to the specific hypothesis (Auret & Vivian, forthcoming). Thus, while Bitcoin may not be weak form efficient in the manner described by the EMH (its returns follow a random walk), it may nevertheless follow a cyclical process of efficiency, where at some points, it is efficient, and at others, it is not.

2.4 Tests of the AMH in general cryptocurrency markets

Given the novelty of cryptocurrencies, the literature that is aimed at determining its efficiency is also recent. The initial set of studies aimed to directly test the Fama (1970) conception of market efficiency by testing whether a cryptocurrency's returns follow a random walk or martingale. The random walk model consists of two hypotheses; firstly, that returns are identically distributed and that they are independent. This is represented as: $f(r_{j,t+1}|\phi_t) = f(r_{j,t+1})$. In this form, the random walk model shows that the marginal, $f(r_{j,t+1}|\phi_t)$, and conditional probability, $f(r_{j,t+1})$, distributions of an independent random variable are identical (Fama, 1970).

Given this formulation, the extant literature has tested whether an asset's (either logarithmic or differential) returns follows a random walk by examining whether its returns show significant autocorrelation: if the returns process displays significant autocorrelation, the null of a random walk can be rejected (Frennberg & Hansson, 1993). An asset whose returns do not follow a random walk is therefore regarded as not being weak-form informationally efficient. In addition, the adaptive markets hypothesis posits that assets have periods where they are efficient and periods in which they are not (Lo, 2004). Whether cryptocurrencies are efficient in the manner envisaged by the AMH hinges on whether they exhibit this time-varying efficiency.

In this regard, several multi-cryptocurrency studies have emerged. Kristoufek and Vosvrda (2019) examined whether a collection of prominent cryptocurrencies are efficient by employing the econophysics-based Efficiency Index which among other things, tests for, long range dependence and entropy. Their findings suggested that all cryptocurrencies under investigation, including Litecoin, Ripple and Bitcoin, were not weak-form efficient. They also found that there are periods in which the cryptocurrencies were efficient, hence making a clear interpretation in favour of adaptive efficiency.

Similar multi-currency studies have also uncovered evidence that there are periods of efficiency and non-efficiency in the cryptocurrency markets. Noda (2021) examined Bitcoin and Ethereum for time-varying efficiency by employing a generalized least squares-based time-varying model and found that they indeed exhibited varying degrees of efficiency over time. In the same vein, by examining the

results of a battery of tests that capture time-varying efficiency, Khursheed *et al.* (2020) found that Litecoin, Bitcoin, Monero and Steller were indeed time-varyingly efficient. Interestingly, both Noda (2020) and Khursheed et. al (2020) agreed that Bitcoin has, relative to its peer cryptocurrencies, longer periods of efficiency.

2.5 Tests of Bitcoin Efficiency

The initial evidence has overwhelmingly been in favour of the conclusion that Bitcoin is not informationally efficient. Urquhart (2016) found that Bitcoin is not weakly efficient through a set of tests including the Ljung-Box and Variance-Ratio tests (these test whether there is autocorrelation in the returns, and thereby test whether the returns follow a random walk). Similarly, Bariviera (2017) examined Bitcoin returns' Hurst exponent, this would be an indication of whether there is long term memory in the returns, which would indicate that Bitcoin is not weak-form efficient. Prior to 2014, the Hurst exponent often was above 0.5 — indicating that there is persistence in the returns — while post-2014, its value creeped closer to 0.5, showing a lack of memory in the returns. Two conclusions flow from these findings; firstly, prior to 2014, Bitcoin was not weak-form efficient and, crucially, Bitcoin (in)efficiency is time-varying.

There has also been counterevidence provided on whether Bitcoin is efficient. In response to Urquhart (2016), Nadarajah and Chu (2017) investigate whether Bitcoin is really weak-form inefficient. They show that using an odd power transformation on Bitcoin returns, R_t^m, where m is an odd integer, does not lead to distortion or information loss. Thus, instead of examining Bitcoin returns, they examine an odd power transformation of the returns to determine whether Bitcoin is weak-form efficient. Adopting a similar level of methodological rigour, they employ several tests of efficiency to develop a rich view of Bitcoin dynamics. Their results indicate that Bitcoin is, in fact, informationally efficient.

It is worth reiterating that while the Nadarajah and Chu (2017) findings run counter to the current consensus, further examination of Bitcoin weak-form efficiency by other academics did not vindicate their results. Kristoufek (2018) employed the relatively novel Efficiency Index, developed by Kristoufek and Vosvrda (2013) to examine whether Bitcoin was weak-form efficient. It captures both local herding (through the fractal dimension) and the correlation structure of returns (through long

and short-term memory). The results indicated that between 2010 and 2017, Bitcoin was not weak-form efficient.

While Lo (2004) outlines an alternative view of efficiency, no mathematical definition against which assets can be tested against to determine whether they are 'adaptively efficient' is provided. Despite this deficiency, several first-generation tests have been devised to test for whether assets are efficient in a manner consistent with the EMH. These tests aim to examine whether efficiency is non-static. (Alvarez-Ramirez *et al.*, 2018) In essence, these tests examine whether a measure of efficiency (such as the Hurst exponent) changes over time in line with evolving market dynamics.

The current consensus is that Bitcoin has periods where it is informationally efficient and periods where it is not; a nod to the AMH concept of efficiency. In line with this consensus, Brauneis and Mestel (2018) argue that the EMH is an inappropriate explanation of cryptocurrency returns. Indeed, Kristoufek (2018) argues that the informational efficiency of Bitcoin is connected to the environmental conditions of the market. Similarly, while the results from Urquhart (2016) indicated that Bitcoin is not weakly efficient, they indicated that it was it was, however, "moving towards" efficiency. This would support the Lo (2004) notion of efficiency, where efficiency can be moved to and from based on the prevailing market dynamics.

Further evidence of time-varying efficiency has been uncovered outside the cryptocurrency markets. By employing time-varying autoregressive (TV-AR) processes to model Japanese stock markets' (TOPIX and TSE2) returns, Noda (2016) demonstrated that they exhibit various degrees of informational efficiency over time. This is in line with the earlier work of Ito *et al.* (2014). They employed a non-Bayesian time-varying vector autoregressive model (TV-VAR) to estimate the joint degree of efficiency of the stock markets of the highly integrated G7 countries. Their results indicate that there are times in which international stock markets are efficient and others in which they are not.

Seetharam (2016) further outlines a rigorous framework for determining whether an asset is, instead, time-varyingly efficient, in other words, is efficient in a dynamic manner consistent with the AMH. It can be outlined as follows:

1) Test whether the returns follow a random walk. If they do, then the analysis stops here as the asset is weak-form efficient. If they do not, we proceed to modelling the return-generating process.

2) Test whether an autoregressive (AR) data generating process with no additional lagged variables, save for lagged variables is suitable for modelling the returns generating process. The Brock-Dechert-Scheinkman (BDS) test to determine whether to use an ARIMA-family or STAR family to model the returns generating process. The STAR family of models are preferred over ARIMA models where the BDS test finds that there is non-linear dependence in the returns. If the chosen model is found to be suitable but still produces significant constant or error terms, it implies that additional factors over and above historic returns affect contemporaneous returns.

3) The return generating process (RGP) is modelled without specifying the functional form of the model.

If this does not adequately describe the returns the returns-generating process, it implies that there is some private information that affects the contemporaneous returns. A machine learning approach is once more used to model the RGP with additional variables.

2.6 A Review Of Return Forecasting Techniques

To test whether a security is weak-form efficient, its return series is examined. If it follows a random walk or Martingale (which, similar to a random walk, means that past information cannot be used to predict future returns) process, it is deemed to be weak-form efficient. Failing which, an attempt at modelling the return generating process can be made. From a statistical perspective, the Box-Jenkins methodology is often used (Makridakis & Hibon, 1997). It is typified by the use of an autocorrelation and partial autocorrelation plots of the returns and iteratively estimating an Autoregressive Integrated Moving Average — ARIMA (p,d,q) — model based on each successive model's diagnostics.

The process can be described as follows:

1) Generate an AR (1) model (the simplest of the ARIMA family) to attempt to model the time series:

$$X_t = c + \varphi X_{t-1} + \varepsilon_t \quad (1)$$

Where ε_t represents a white-noise process with a mean = 0 and a constant variance = σ_ε^2.

2) Examine the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots of the residuals for serial correlation.

3) If any serial correlation exists, use the ACF and PACF plots as guides for the ARMA order of the model, differencing if necessary.

4) Once the new ARMA (p,q) model is fitted to the data, a new examination of the ACF and PACF along with other model diagnostics are used to determine whether this model is appropriate for forecasting. If it is, then it is adopted.

5) If not, then the process returns to step 3), where a new ARMA (p,q) model is generated until one sufficiently describes the return generating process and thereby leaves no significant autocorrelation in the residuals.

While this approach is often preferred in time series modelling, it has been shown to deliver inferior performance to Smooth Transition Autoregressive (STAR) models when there is non-linear dependence (Feng & Liu, 2003). As such, the decision tool regarding whether a STAR model is to be used instead of the ARMA (p,q) model is the BDS test for nonlinear dependence (Baum *et al.*, 2021). Moreover, Crawford and Fratantoni (2003) argue that regime-switching models are a better fit for markets that have exhibit boom-bust cycles and are non-linearly dependent. In concert, Seetharam and Britten (2015) show that market cycles can be modelled using a regimes-switching approach such as a STAR model. The process of finding the best STAR model to use entails using a grid search approach, by typically homing in on a model with minimal pooled Akaike Information Criterion (AIC).

While these approaches serve as a good forecasting baseline, they often have been shown to be inferior in performance to machine learning approaches to forecasting: in the extant literature, there has been a notable improvement from a model diagnostics perspective when using machine learning approaches over purely statistical approaches in forecasting returns (Ding & Qin, 2020). Furthermore,

machine learning approaches have the advantage that they do not need to be pre-specified in a similar fashion to the Box-Jenkins approach (Seetharam, 2016).

Despite their numerous advantages, machine learning approaches are also riddled with a rather severe drawback: they are often "black-boxes" as they are not interpretable due to their structural complexity (Rudin, 2019). While the field of interpretable machine learning algorithms is certainly blooming, some evidence remains that the best (and therefore, most used) algorithms are often uninterpretable. In fact, Adadi and Berrada (2018) argue that their high predictive accuracy is what ultimately leads to their inability to be interpreted. As such, there may be some trade-off between accuracy and interpretability.

It is to be noted, however, that machine learning approaches are not made equal, with respect to the nature of the task at hand. The K-Nearest Neighbours (kNN) algorithm is almost exclusively employed in classification tasks (Weinberger & Saul, 2009). In contrast, while Support Vector Machines (SVMs) used in classification tasks, they have been demonstrated to be equally excellent in time series forecasting. (Aggarwal *et al.*, 2020)

Another often-used family of deep learning algorithms in time series forecasting, especially in finance, are recurrent neural networks (RNNs). They, however, suffer from a "vanishing gradient" problem that results in difficulties in the learning process. (Hochreiter, 1998) The "vanishing gradient" problem refers to a scenario in which the gradients of a loss function tend to zero as more layers (which often use a sigmoid activation function) are added thus increasing the difficulty of training the network. (Li *et al.*, 2019) Long-Short Term Memory (LSTM) are an augmented RNN that does not suffer from this problem. (Bahad *et al.*, 2019) Their ability to restrain gradients from exploding or vanishing is attributable to their internal memory state cell's ability to sustain a continual error flow. Cao *et al.* (2019)

Moreover, unlike traditional RNNs, LSTMs filter a time series' historical data through their three gates and thereby extract useful information (Hochreiter & Schmidhuber, 1997). It is unsurprising thus that LSTMs performance in time series prediction outshines that of their vanilla RNN counterparts. McNally (2016) demonstrates that an LSTM network outperforms both a Bayesian

optimised RNN network and ARIMA model in time series prediction. In addition, LSTMs can be used in combination with other methods in order to improve their predictive capabilities. Cao *et al.* (2019) demonstrate that a CEEMDAN-LSTM — where the original time series is decomposed into a finite number of intrinsic mode functions (IMF) prior to reconstructing the signal — outperforms a vanilla LSTM in financial time series prediction. Li *et al.* (2019) also show that an evolutionary attention-based LSTM (EA-LSTM) similarly outperforms a vanilla LSTM in time series prediction. It avoids local minima traps by employing a competitive random search (CRS) instead of a gradient-based approach to solve the focus layer weights. The R packages, "*keras*" and "*tensorflow*" have machine learning capabilities and thus enable the use of Recurrent Neural Networks, including LSTMs.

Seetharam (2016) found that non-linear autoregressive (NARX) models can both capture non-linear dependency in asset returns. NARX models have found favour in engineering and other physical sciences. Cadenas *et al.* (2016) show that NARX models are useful in predicting physical phenomena such as wind speed. Similarly, Pisoni *et al.* (2009) demonstrate that air pollution (specifically above-threshold pollution in cities) can be forecasted with the assistance of a NARX model. It is represented as:

$$y_t = F(y_{t-1},\ y_{t-2},\ y_{t-3}, \dots, u_t, u_{t-1}, u_{t-2}, \dots) + \varepsilon_t \qquad (2)$$

Where F is a neural network, such as a Support Vector Machine (SVM), y_t represents the return generating process and u_{t-k} and y_{t-k-1} represent the exogenous series and lagged values of the returns, respectively.

The use of Support Vector Machines (SVMs) in predicting financial time series is well documented. Kim (2003) demonstrates that SVMs outperform back-propagated neural networks in forecasting financial index' returns. He argues that this is attributable to SVMs being more generalizable due to their implementation of the structural risk minimization principle. Cao and Tay (2001) also show that SVMs outperform multi-layer perceptron networks in forecasting financial returns. The "*mltsp*" and "*e1071*" packages have NARX model-building capabilities that allow one to use a NARX model powered by an SVM.

Given the demonstrated relative advantages of SVM and LSTM approaches to forecasting financial time series, they are employed in the machine learning section of Bitcoin return forecasting. This is done in sequential order; the LSTM approach is employed in univariate forecasting whereas the SVM is used in the forecasting Bitcoin return with the help of external regressors.

2.7 The Efficient Markets Hypothesis

As with most field marred by controversy, there has been a clash of findings regarding assets' efficiency. For example, while most studies such as Kristoufek (2018) reject the view that Bitcoin is weak-form efficient, Nadarajah and Chu (2017) find that an odd-power transformation of Bitcoin returns, R_t^m, is weak-form efficient. Given the divergent findings in the literature, it is imperative thus that attempts to examine the efficiency of Bitcoin returns be run on a parallel twin set to have results that are robust to noise; a return series that has been transformed as well as a non-transformed Bitcoin return series.

An often-used transformation is the Fourier Transformation. It is defined as:

$$FT(\omega) = \int_{-\infty}^{\infty} s(t)e^{-j\omega t}dt \tag{3}$$

Where $FT(\omega)$ is the Fourier Transform of the time series, $s(t)$, which itself has been decomposed for the purposes of transforming it. (Alrumaih & Al-Fawzan, 2002) Applying a Fourier Transform to a time series has an overall impact of "de-noising" it. (Jansen van Rensburg & van Vuuren, 2020) It is thus regarded in extant literature as "cleaning" the data with the aim of having a better calibrated model of the data-generating process. Having a filtered data set to compare against a non-transformed data set also allows for a more robust set of results.

In cases where the return series is shown not to follow a martingale or random walk, an attempt to model it can be done. Given that additional factors outside of historical returns can affect returns,

their use as exogenous regressors in modelling returns may prove to be of use. Vo *et al.* (2021) demonstrate that US GDP data has a positive impact on Bitcoin returns. Similarly, Panagiotidis et. al (2019) found that traditional stock markets have a strong relationship with Bitcoin returns. Their results, however, contrast with those of Vo *et al.* (2021) in that they show that there is a weak relationship between Bitcoin returns and the macroeconomy. Li *et al.* (2021) show that there is a strong link between social media coverage and Bitcoin returns. In a similar vein of connecting the public consciousness to Bitcoin returns, Dastgir et al. (2019) show that there is a bi-directional causal link between Bitcoin returns and Google Trends searches. Given that these factors are consistently shown to affect returns, Seetharam (2016) argues that they need to be incorporated into the return-generating process model in the spirit of the Arbitrage Pricing Theory (APT).

2.8 Summary of literature review

This chapter walked through both the early views of efficiency, delineated between the random walk and martingale views of weak-form efficiency, and delved into the issues that plague the Efficient Markets Hypothesis. The focus was then shifted onto the Adaptive Markets Hypothesis and the evidence in its favour in extant literature. A discussion on whether the general cryptocurrency market's dynamics are described better by the EMH or the AMH followed, followed by a focused overview of the evidence of the AMH's explanatory power with regards to Bitcoin dynamics. When an asset is weak-form efficient (and by corollary, follow a martingale or random walk, their returns cannot be forecasted). As such, literature regarding techniques used in return forecasting for assets that are not weak-form efficient is entertained. To round off, a discussion on best practices in forecasting was explored.

3 DATA AND METHODOLOGY

The data and methodology that are used to test the main and sub-hypotheses are described below:

3.1 Data

This study will use the daily closing return data of Bitcoin from Coindesk from Oct 2013 to Oct 2021. This selection of dates is due to the availability of Bitcoin return data from Coindesk. The closing return data is then transformed into a log return using the familiar formula:

$$R_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \quad (4)$$

Where P_t and P_{t-1} represent prices at time t and t-1 closing prices, respectively. The data was collected at three different frequencies, *viz.*, daily, weekly and monthly.

Macroeconomic, fundamental, and behavioural data was also obtained that will be included in the exogenous model in a similar fashion to the Arbitrage Pricing Theory (APT). These variables include US GDP data, the MSCI World Index, US Fed Funds rates, Bitcoin Google Trends Data. Like the Bitcoin return data, the exogenous variable data will also be collected at three different frequencies — daily, weekly and monthly. These variables were selected based on previous literature which has demonstrated their capability in predicting Bitcoin returns. Vo *et al.* (2021) demonstrate that US GDP data affects Bitcoin returns. Similarly, Demir *et al.* (2018) also showed that economic uncertainty has a relationship with Bitcoin returns.

3.2 Methodology

The methodology is outlined (broadly) in the three steps below:

1) Test whether the returns follow a random walk using a battery of tests that includes the Automatic Portmanteau (AQ), Generalized Spectral (GS) and Dominguez-Lobato (DL) tests. If they do, then that is evidence in favour of the weak form of the EMH. If not, it can thus be established whether modelling the returns is feasible.
2) A BDS test is then used to determine whether the returns follow a linear process, if they do an autoregressive integrated moving average (ARIMA) process is used to model the return

generating process. If not, then the return generating process is modelled using a regime-switching autoregressive (AR) process with no additional (exogenous) variables, save for the lagged dependent variable (SETAR) is used. The general form of Threshold Autoregressive Model (TAR) family of models, which SETAR models form a part of, are represented as:

$$x(t) = \varphi_j(0) + \sum_{i=1}^{nj} \varphi_j + (i)x(t-i) + \ a_j(t) \tag{5}$$

Where *x(t)*, the variable's current value is determined based on its own lagged values, *x(t-i)*.

3) If this successful, this is evidence of the semi-strong form of the EMH (public information is incorporated into the security's return). If the model contains a significant constant or error terms, that implies that significant additional factors over and above the lagged dependent variable have an impact on the contemporaneous returns.
4) If there are significant additional factors that affect the returns contemporaneously, a Long Short Term Memory network (LSTM) is utilized to model the data generating process. It is used without additional variables. An LSTM has three sigmoid states that are known as the forget gate, input gate and the output gate (Hochreiter & Schmidhuber, 1997). These are represented as:

$$f_t = \ \sigma\big(x_t * U_f + H_{t-1} * W_f\big) \tag{6}$$

$$i_t = \ \sigma(x_t * U_i + H_{t-1} * W_i) \tag{7}$$

$$o_t = \ \sigma(x_t * U_o + H_{t-1} * W_o) \tag{8}$$

Where each gate receives, processes and passes the signal on to the other gates in the following order: input*(i)*, forget*(f)*, which regulates which data to throw out and retain, and output *(o)*.

5) If the model is inadequate to describe the return generating process, the enquiry moves to the final step.

The second neural network, a non-linear autoregressive exogenous model (NARX), is employed. It models the returns generating process through using both lagged returns and additional exogenous variables that are incorporated in line with the Arbitrage Pricing Theory (APT). It is represented as,

$$y_t = F(y_{t-1},\ y_{t-2},\ y_{t-3}, \dots, u_t, u_{t-1}, u_{t-2}, \dots) + \ \varepsilon_t \tag{9}$$

where F is a neural network and y_t represents the return generating process. u_{t-k} and y_{t-k-1} represent the exogenous series and lagged values of the returns, respectively. A neural network approach in both regards is advantageous given that it allows for modelling of the DGP without the requirement that the functional form of the model be specified in advance. It has been shown to be applicable to modelling cyclical efficiency (Seetharam, 2016).

3.2.1 Linearity Tests

The Brock-Dechert-Scheinkman (BDS) test is used to examine whether a time series is linear. The test statistic is given by:

$$V_m = \sqrt{T}\frac{C_{m,\epsilon} - C_{1,\epsilon}^m}{S_{m,\epsilon}} \tag{10}$$

Where the correlational integral, $C_{m,\epsilon}$, captures the of repetition of a temporal pattern and $S_{m,\epsilon}$ is the standard deviation of $\sqrt{T}C_{m,\epsilon} - C_{1,\epsilon}^m$.

3.2.2 Normality Tests

As a measure to ensure that the results are robust, normality tests, which examine whether the data exhibits normality, are presented below:

3.2.2.1 The Jarque-Bera Test

The Jarque-Bera test for normality is a goodness of fit test (Jarque & Bera, 1980). The test statistic, JB, is a function of the measures of kurtosis (K) and skewness (S) computed from the sample. Where the Jarque-Bera test statistic is given by:

$$JB = \frac{n}{6} * \left(S^2 + \frac{(K-3)^2}{4}\right) \tag{10}$$

Where the sample kurtosis is represented by $K = \frac{\hat{u}_4}{\hat{u}_2^2}$, and the sample skewness, $S = \frac{\hat{u}_3}{\hat{u}_2{}^{1.5}}$. Where $\widehat{u_2}$ and $\widehat{u_3}$ are the theoretical second and third theoretical moments of a chi (χ^2) distribution. Under normality, the theoretical values of S = 0 and K = 3 (Thadewald & Büning, 2007). The null hypothesis

of this test is that the population is normally distributed. The null is rejected when the results of the Jarque-Bera test in show a large χ^2 value and near-zero p-value.

3.2.2.2 The Shapiro-Wilk Test

Similarly, the null hypothesis of the Shapiro and Wilk (1965) test is that the population is normally distributed. The Shapiro-Wilk test statistic is given by:

$$SW = \frac{\left(\sum_{x=1}^{n} a_i X_{(i)}\right)^2}{\sum_{x=1}^{n}(X_i - \overline{X_i})^2}. \quad (11)$$

The null is rejected when the test's associated p-value is near-zero.

3.2.3 Stationarity Tests

Two stationarity tests are presented below:

3.2.3.1 The Kwiatkowski-Phillips-Schmidt-Shin (KPSS)

The null hypothesis for the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test is that the data is stationary around a deterministic trend. Its test statistic is given by:

$$KPSS = T^{-2} \frac{\sum_{t=1}^{T} \hat{S}_t^2}{\hat{\lambda}^2} \quad (12)$$

Where the test statistic, under a null of stationarity, follows a normal distribution. The null of stationarity is rejected when the test's associated p-value is less than 0.05.

3.2.3.2 The Augmented Dickey-Fuller (ADF) Test

The null hypothesis for the Augmented Dickey-Fuller test is that the data is not stationary. Its test statistic is given by:

$$DF_\tau = \frac{\hat{\gamma}}{SE(\hat{\gamma})} \quad (13)$$

The test statistic is compared against a critical value, where, if the computed value is more negative than the critical value, then the null of a unit root is rejected.

3.2.3.3 The Phillips-Perron Test

The Phillips-Perron (PP) test is a non-parametric test that may also be used to examine whether a time series has a unit root. Similar to the ADF test, the null hypothesis for the Phillips-Perron test is that the data is not stationary. All PP test types used the model below to test for the null of a unit root:

$$y_t = c + \delta t + a y_{t-1} + e(t) \quad (14)$$

Where y_t, the time series' current value is predicted based on its previous lagged value, y_{t-1} and an innovation term, $e(t)$.

3.3 Robustness Checks

In order to ensure that the results of this research report are robust to noise, a Fast Fourier Transform, a variant of the Fourier Transform that requires much less computational costs than the traditional version, is employed. Its use ultimately results in the de-noising of the data, effectively resulting in an ability to compare an untransformed data set against a transformed one. Furthermore, several stationary tests, normality and Martingale Difference Hypothesis tests are employed to ensure that results are robust to the idiosyncrasies of the various tests. Finally, two different sets of machine learning techniques are employed (with the caveat that they sit at different stages of the methodology,) while not a precautionary measure per se, it nevertheless ensures that, given the black-box nature of neural networks, that an alternative view of the return generating process can be entertained.

3.4 Summary

The methodology will be tested with four distinct steps on transformed and untransformed data:

1) Determine, via a battery of tests, whether returns follow a random walk/martingale, if they do, then the enquiry stops at this level as they cannot be forecasted.
2) If they do not, they are subjected to linearity tests (beginning with the BDS test, and if its results are inconclusive, a SETAR test) to test whether a linear (ARIMA-family) or non-linear (SETAR) model should be used to model the return generating process. Once that decision is made, the model is estimated via the various R packages and its residuals are tested to examine model suitability, following which, the Hurst Exponent is used as an acid test to decide whether to continue attempting to model the return generating process or end the enquiry at this level.
3) If the Hurst Exponent results prompt further modelling, a machine learning approach (LSTM) is used to attempt to model the return generating process. Its residuals are, similar to step 2,

analyzed and the Hurst Exponent is also used as a decision criterion to determine whether to use a machine learning approach that incorporates exogenous regressors.

4) In the event that the computation of the Hurst Exponent calls for the incorporation of exogenous regressors in the machine learning model, a NARX model is used. Its residuals are analyzed and if found to be a good fit, this is taken as *prima facie* evidence of time varying evidence, which machine learning approaches can capture, unlike their statistical approach counterparts.

4 RESULTS

4.1 Normality Tests

The Jarque-Bera test for normality is a goodness of fit test (Jarque & Bera, 1980). The test statistic, JB, is a function of the measures of kurtosis (K) and skewness (S) computed from the sample. The null hypothesis of this test is that the population is normally distributed. The results of the Jarque-Bera test in Table 1 show a large χ^2 value and near-zero p-value, indicating that the null of normality can be rejected. Similarly, the null hypothesis of the Shapiro and Wilk (1965) test is that the population is normally distributed. Its associated p-value is near-zero, also indicating that the null of normality can be rejected.

Table 4.1: Daily Returns' Normality Tests

Test	Jarque-Bera Test	Shapiro-Wilk Test
Test Statistic	5673.60	0.91
P-Value	0***	0***

Note: * Significant at the 10% level, ** Significant at the 5% level, *** Significant at the 1% level.

The results of the Jarque-Bera test in Table 2 show a large χ^2 value and near-zero p-value, indicating that the null of normality can be rejected. Similarly, the Shapiro-Wilk test results indicate that since the p-value is near-zero, the null of normality can be rejected.

Table 4.2: Weekly Returns' Normality Tests

Test	Jarque-Bera Test	Shapiro-Wilk Test
Test Statistic	54.97	0.97
P-Value	0***	0***

Note: * Significant at the 10% level, ** Significant at the 5% level, *** Significant at the 1% level

4.2 Stationarity Tests

Table 3 below displays the results of three stationarity tests for daily returns. The null hypothesis for the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test is that the data is stationary. The p-value is greater than 0.05 (0.1) and thus, we fail to reject the null of stationarity. The Phillips-Peron and Augmented Dickey-Fuller (ADF) tests' null hypotheses is that the data is not stationary. The p-value is less than 0.05 (0.01) and thus, we reject the null of non-stationarity in both instances. The three tests converge on the conclusion that the data can be considered as stationary.

Table 4.3: Daily Returns' Stationarity Tests

Test	KPSS	Phillips-Perron	ADF
Test Statistic	0.07	-3202	-12.25
p-value	0.1*	0.01***	0.01***

Note: * Significant at the 10% level, ** Significant at the 5% level, *** Significant at the 1% level.

Table 4 below displays the results of three stationarity tests for weekly returns. The null hypothesis for the KPSS is that the data is stationary. The p-value is greater than 0.05 (0.1) and thus, we fail to reject the null of stationarity. The Phillips-Peron and ADF test's null hypotheses is that the data is not stationary. The p-value is less than 0.05 (0.01) and thus, we reject the null of non-stationarity in both instances. The three tests agree that the data can be considered to be stationary.

Table 4.4: Weekly Returns' Normality Tests

Test	KPSS	Phillips-Perron	ADF
Test Statistic	0.06	-387.79	-8.38
p-value	0.1*	0.01***	0.01***

Note: * Significant at the 10% level, ** Significant at the 5% level, *** Significant at the 1% level.

4.3 Martingale Difference Hypothesis Tests

Whether the daily return series follows a martingale is tested via a battery of tests whose results are displayed in Table 5 below. The automatic portmanteau (AQ) statistic asymptotically follows the χ^2 distribution with one degree of freedom. The null hypothesis of no return autocorrelation is rejected at the 5% level of significance if the value of AQ is greater than 3.84. The AQ test results (0.36) thus fail to reject the null of no return autocorrelation.

While the Automatic Portmanteau test only tests for linear dependence, the Generalized Spectral Test (GS) investigates both linear and non-linear dependence. Both the Dominguez and Lobato test and the Generalized Spectral test can detect non-linear dependence, however, the Generalized Spectral Test has been shown to have more power against non-linear dependence (Salisu *et al*., 2016). The Dominguez-Lobato test results have p-values of p = 0 and p = 0.1. This means that the null of a martingale difference sequence (efficiency) cannot be rejected in one test but can be in the other, indicating a conflicting view. The GS test returns a p-value = 0.2, also indicating that the null of weak-form efficiency cannot be rejected. Taken together, the results of the AQ test, the GS and the Dominguez-Lobato tests seem to reject both linear and non-linear dependence.

Table 4.5: Martingale Difference Hypothesis Tests - Daily Returns

Test	AQ Test	DL Test – Cp	DL Test – Kp	GS Test
t-statistic	0.37	0.36	1.11	0.2
p-value	0.54	0.1*	0***	NA

Note: * Significant at the 10% level, ** Significant at the 5% level, *** Significant at the 1% level.

The Hurst Exponent is a long-range dependence estimator. Couillard and Davison (2005) argue that the true value of the Hurst exponent for a finite time series cannot be computed. Furthermore, each Hurst exponent computation method offers some mix of drawbacks and advantages; the rescaled range analysis (R/S) Hurst, for example, was demonstrated by Davies and Harte (1987) to falsely lead to the conclusion that there was long-term dependence when the time series had short-term correlation.

Annis and Lloyd (1976) derived the Empirical Hurst Exponent, whose rescaled range's expected value was given by:

$$E\left(\frac{R}{S}\right)_n = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{n}{2}\right)} \sum_{r=1}^{n-1} \sqrt{\frac{n-r}{r}} \tag{15}$$

In response, Peters (1994) proposed the Corrected Empirical Hurst Exponent which avoided its uncorrected predecessors' overestimation of the rescaled range for small *n* values. Couillard and Davison (2005) find, however, that the Empirical Hurst Exponent is useful in determining whether a financial time series exhibited long-term memory. To get around the aforementioned problems, several methods of calculating the Hurst exponent are used to get a more robust answer regarding whether the time series has long-term memory.

The Hurst Exponents for the daily returns are in table 6 below. In contrast to the other inconclusive Martingale Difference Hypothesis test results, however, the numerous methods of calculating the Hurst exponent arrive at the same conclusion: they strongly indicate that there is persistence (long-term memory) in the data. This would undermine weak form efficiency. This therefore means that there is room to continue modelling the return generating process. The critical value for the Hurst Exponent is 0.50, *i.e.*, if the Hurst Exponent is greater than 0.5, then it can be concluded that there is long-term dependence in the returns.

Table 4.6: Hurst Exponent - Daily Returns

Hurst Exponent Type	Simple R/S Hurst Exponent	Corrected R/S Hurst Exponent	Empirical Hurst Exponent	Corrected Empirical Hurst Exponent	Theoretical Hurst Exponent
Statistic Value	0.54	0.61	0.64	0.61	0.53

Table 7 displays the results of the Martingale Difference Hypothesis Test results for the weekly log returns. The automatic portmanteau (AQ) test's null hypothesis of no return autocorrelation is rejected

at the 5% level of significance if the value of AQ is greater than 3.84. The AQ test results (3.33) thus fail to reject the null of no return autocorrelation. The Dominguez-Lobato test results have p-values = 0. This means that the null of martingale difference sequence (efficiency) can be rejected. Similarly, the GS test returns a p-value = 0, also indicating that the null of weak-form efficiency can be rejected. The results of the AQ test, the GS and the Dominguez-Lobato tests indicate that there is non-linear dependence.

Table 4.7: Martingale Difference Hypothesis Tests – Weekly Returns

Test	AQ Test	DL Test – Cp	DL Test – Kp	GS Test
t-statistic	3.33	0.80	1.60	0
p-value	0.07*	0***	0***	NA

Note: * Significant at the 10% level, ** Significant at the 5% level, *** Significant at the 1% level.

Similar to the daily returns' outcomes, Table 8 shows that the weekly returns' different Hurst exponents arrive at an overall conclusion of persistence (long-memory) in the data. This would undermine weak form efficiency.

Table 4.8: Hurst Exponents – Weekly Returns

Hurst Exponent Type	Simple R/S Hurst Exponent	Corrected R/S Hurst Exponent	Empirical Hurst Exponent	Corrected Empirical Hurst Exponent	Theoretical Hurst Exponent
Statistic Value	0.55	0.61	0.79	0.73	0.55

4.4 Linearity Tests

Finally, tables 9 and 10 display the results of the BDS test for the daily and weekly returns respectively. The p-values of the BDS test for both daily returns are 0. This indicates that the data is not linear. As a result, a non-linear model (such as a SETAR model) can be used to model the daily

return-generating process. Seetharam and Britten (2015) recommend using a grid-search to find the optimal parameter values of the non-linear model to be used.

Table 4.9: BDS Tests – Daily Returns

Standard Normal =				
	[0.02]	[0.04]	[0.06]	[0.08]
[2]	14.49	13.58	12.64	12.47
[3]	20.22	17.22	15.92	15.63
P-Value =				
	[0.02]	[0.04]	[0.06]	[0.08]
[2]	0***	0***	0***	0***
[3]	0***	0***	0***	0***

Note: * Significant at the 10% level, ** Significant at the 5% level, *** Significant at the 1% level.

Similar to the daily return results, the BDS test results for the weekly returns in Table 10 below indicate that the return-generating process is non-linear as the p-values for the test are 0. To model the return-generating process therefore, a grid-search approach is employed.

Table 4.10: BDS Test - Weekly Returns

Standard Normal =				
	[0.06]	[0.011]	[0.17]	[0.23]
[2]	4.4	5.20	5.16	5.75
[3]	6.43	7.47	6.98	7.18
P-Value =				
	[0.06]	[0.011]	[0.17]	[0.23]
[2]	0***	0***	0***	0***
[3]	0***	0***	0***	0***

Note: * Significant at the 10% level, ** Significant at the 5% level, *** Significant at the 1% level.

4.5 SETAR/ARIMA modelling without external regressors

The grid search obtained the best SETAR parameters by minimizing the pooled Akaike Information Criterion. This criterion attempts to create a balance between model-fit (by having as many parameters as possible that are explain the data in the model) and parsimony (ensuring that the minimal parameters necessary to model the data is included). The results of the grid search then used as the inputs for the SETAR model. Figures 1 and 2 display a graphical representation of the grid search.

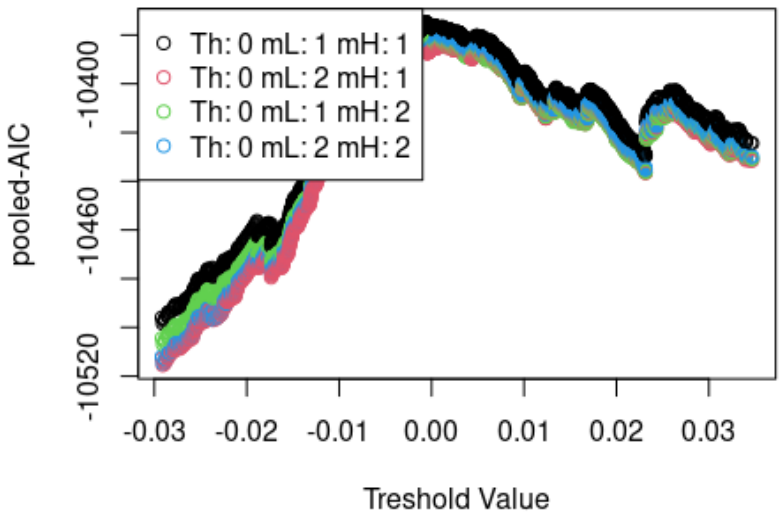


Figure 4.1: Daily SETAR grid search

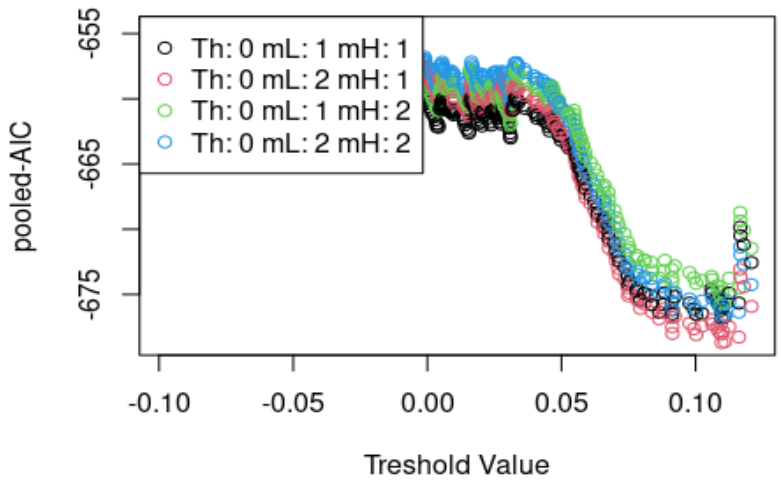


Figure 4.2: Weekly SETAR grid search

Tables 11 and 12 contain the saved SETAR model inputs for the daily and weekly return-generating processes. Notably both daily and weekly SETAR grid searches resulted in 2 coefficients on the low regime (mL) and a single high regime coefficient (mH).

Table 4.11: Best Daily SETAR Model Inputs

thDelay	Th	pooled-AIC	mL	mH
0	-0.03	-10515.48	2	1

Table 4.12: Best Daily SETAR Model Inputs

thDelay	Th	pooled-AIC	mL	mH
0	-0.11	-678.69	2	1

Table 13 captures the results of the daily model. The generated model is a SETAR with two regimes (a high and low regime) where most of the points (84,99%) are captured by the high regime. Notably, ϕ_{L1} and ϕ_{L2} and the constant are statistically significant. To examine whether the model is a good fit, the residuals are examined. In general, autocorrelation in the residuals indicates that the model may not be a good fit.

Table 4.13: Daily SETAR Model

SETAR Model – Daily Returns		
Low-Regime Coefficient	Value	P-Value
Constant – L	-0.01	0.09*
ϕ_{L1}	-0.17	0***
ϕ_{L2}	-0.16	0***
Constant – H	0.01	0.23
ϕ_{H1}	0.04	0.12
Model Fit and Diagnostics		
AIC	MAPE	Threshold
-18874	144.80%	-0.03
Low-Regime Point Proportion	High-Regime Point Proportion	
15%	85%	

Note: * Significant at the 10% level, ** Significant at the 5% level, *** Significant at the 1% level.

Table 14 displays the results the weekly return SETAR model. The generated model is a SETAR with two regimes (a high and low regime) where most of the points (72,88%) are captured by the low regime. Notably, ϕ_{L1} and ϕ_{H1} and the constant are statistically significant. To examine whether the model is a good fit, the residuals are examined.

Table 4.14: Weekly SETAR Model

SETAR Model – Weekly Returns		
Low-Regime Coefficient	Value	P-Value
Constant – L	-0.03	0.10
ϕ_{L1}	-0.28	0.04**
ϕ_{L2}	0.03	0.74
Constant – H	0.002	0.73
ϕ_{H1}	0.33	0***
Model Fit and Diagnostics		
AIC	MAPE	Threshold
-1874	161.10%	-0.03
Low-Regime Point Proportion	High-Regime Point Proportion	
27.12%	72.88%	

Note: * Significant at the 10% level, ** Significant at the 5% level, *** Significant at the 1% level.

A time series plot of the residuals seems to show volatility bunching (non-constant standard deviation) but a zero mean in Figures 3 and 4 below. This indicates that the residuals approximate, but do not follow a white noise process, more formal tests are required to arrive at a conclusion, however.

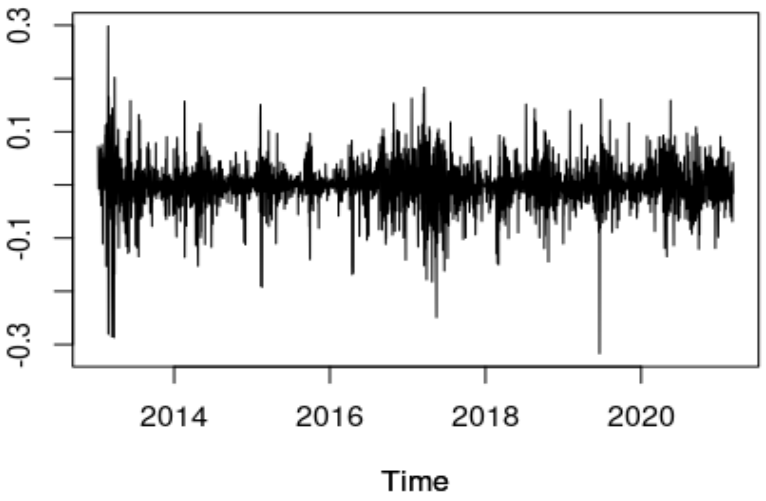


Figure 4.3: Daily SETAR Model Residuals

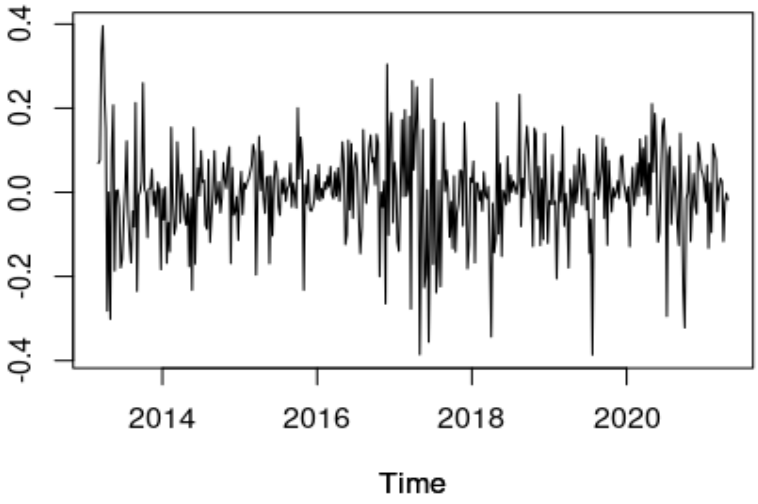


Figure 4.4: Weekly SETAR Model Residuals

An examination of the residuals' ACF will aid in helping decide whether there is any serial correlation in the model's residuals. The correlogram is show in Figure 5 below. The ACF plot of the daily SETAR model's residuals indicates that there are a few edge-cases of statistically significant autocorrelation in the residuals. There are also at least two lags at which there is clearly significant autocorrelation. This may indicate that there is some further information that needs to be added into the model (and thus, this may not be the best "final" model of the return generating process).

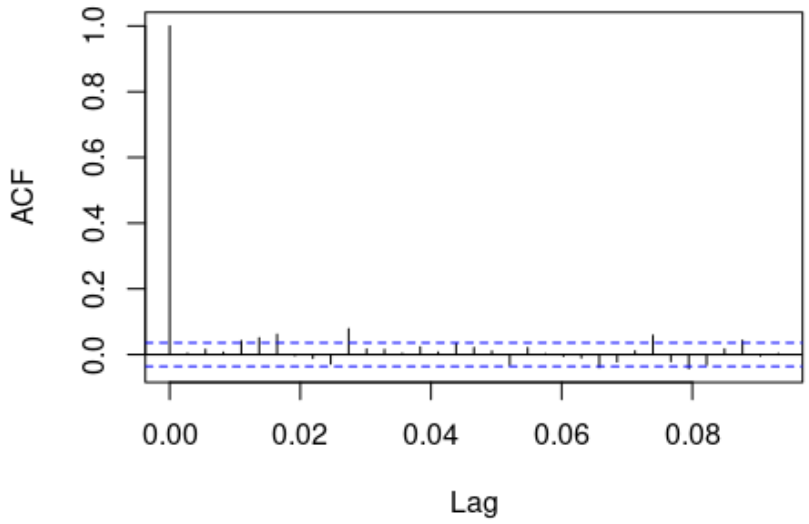
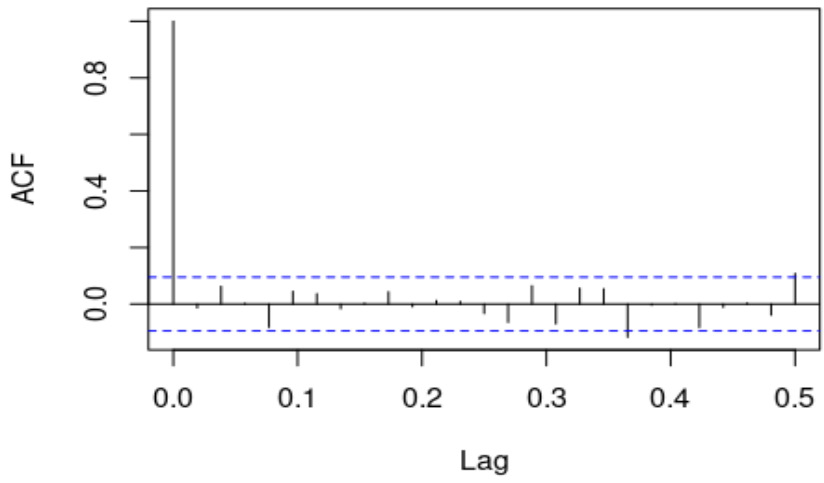


Figure 4.5: Daily Model Residuals' ACF

While the weekly ACF plot in Figure 6 shows even fewer edge-cases of statistically significant autocorrelation in the residuals, there is at least one lag at which there is clearly significant autocorrelation. This may indicate that there is some further information that needs to be added into the model. In addition to a visual inspection of the ACF plots, formal statistical are employed to test for normality in the residuals.



Note: The blue bands indicate the line of significant, at lags where it is crossed, it can be inferred that there is significant autocorrelation in the time series under investigation.

Figure 4.6: Weekly Model Residuals' ACF Plot

The normality test results of the daily model residuals are shown in Table 15. The Jarque-Bera test returns a near-zero p-value, indicating that the null of normality in the residuals can be rejected. The Shapiro-Wilk test also has a near-zero p-value, also indicating that the null of normality in the residuals can similarly be rejected. The errors are tested for normality to examine whether the tentative model provides a good explanation of the underlying data generating process. In the same vein, the examination of the stationarity of the residuals is important insofar as it aids in the avoidance of spurious regressions. (Wooldridge, 2012)

Table 4.15: Daily Residuals' Normality Tests

Test	Jarque-Bera	Shapiro-Wilk
Test Statistic	5626.20	0.91
p-value	0***	0***

Note: * Significant at the 10% level, ** Significant at the 5% level, *** Significant at the 1% level.

The Hurst exponent results in Table 16, can aid in deciding whether to use more advanced approaches to model the return generating process. Since the Hurst Exponent > 0.5, it therefore is rational to more advanced approaches to model the data generating process.

Table 4.16: Hurst Exponent – Daily Residuals

Hurst Exponent Type	Simple R/S Hurst Exponent	Corrected R/S Hurst Exponent	Empirical Hurst Exponent	Corrected Empirical Hurst Exponent	Theoretical Hurst Exponent
Statistic Value	0.55	0.62	0.65	0.61	0.53

Table 17 displays the results of the normality tests of the weekly model residuals. In similar fashion to the daily residuals' test results, the Jarque-Bera test returns a near-zero p-value, indicating that the null of normality in the residuals can be rejected. The Shapiro-Wilk test also has a p-value less than 5%, also indicating that the null of normality in the residuals can similarly be rejected.

Table 4.17: Weekly Returns' Normality Tests

Test	Jarque-Bera	Shapiro-Wilk
Test Statistic	43.36	0.97
p-value	0***	0***

Note: * Significant at the 10% level, ** Significant at the 5% level, *** Significant at the 1% level.

Table 18 displays the weekly residuals' Hurst exponents. While the weekly residuals' Hurst exponent is closer to 0.5 than the daily residuals' Hurst exponent, it nevertheless is greater than 0.5. Taking the Hurst exponent as a decision criterion regarding whether to continue modelling, the conclusion is therefore that for both the daily and weekly time series, more advanced attempts to model the return generating process must be employed.

Table 4.18: Weekly Model Residuals Hurst Exponent

Hurst Exponent Type	Simple R/S Hurst Exponent	Corrected R/S Hurst Exponent	Empirical Hurst Exponent	Corrected Empirical Hurst Exponent	Theoretical Hurst Exponent
Statistic Value	0.53	0.61	0.80	0.74	0.55

Before that, an examination of whether an application of a Fourier Transform to the data will yield a different conclusion (regarding whether to use more advanced approaches to model the return generating process) is undertaken. A Fourier Transform may be regarded as "cleaning" the data with the aim of having a better calibrated model of the data-generating process. Having a filtered data set to compare against a non-transformed data set also allows for a more robust set of results.

4.6 Fourier Transform

Prior to applying a Fourier Transform, Jansen van Rensburg and van Vuuren (2020) note that the data has to firstly be differenced in order to remove any upward/downward trend. This is because the discrete Fourier Transform assumes that the input signal has a non-time varying mean (stationary). The log-returns have indeed been differenced, which is evidenced by the lack of a trend. The differencing procedure is carried out and the transformed data is plotted, and the mean is indeed

constant (equal to 0 over time). This therefore allows for a Fourier transform. While there is indeed no trend in the data, it nevertheless exhibits volatility clustering.

A Fast Fourier Transform (FFT) is applied to the data. It reduces the traditional Fourier Transform's computational expense by requiring $O(nlogn)$ instead of the standard $O(n)^2$ operations to compute the Fourier coefficients. A Fourier Transform would therefore be excessively difficult given the length of the time series under consideration. Furthermore, given that *R* has trouble dealing with complex numbers, the real component of the Fourier Transformed series is extracted instead. This real series is used as an input into the tests and models that follow.

Martingale Difference Hypothesis (MDH) Test

The Martingale Difference Hypothesis results for the transformed daily returns are displayed in Table 19. The automatic portmanteau (AQ) test's null hypothesis of no return autocorrelation is rejected at the 5% level of significance if the value of AQ is greater than 3.84. The AQ test results (434) reject the null of no return autocorrelation. The Dominguez-Lobato test results have p-values = 0. This means that the null of martingale difference sequence (efficiency) can be rejected. Similarly, the GS test returns a p-value = 0, also indicating that the null of weak-form efficiency can be rejected. The results of the AQ test, the GS and the Dominguez-Lobato tests indicate that there is both linear and non-linear dependence.

Table 4.19: Martingale Difference Hypothesis Test – Daily Returns

Test	AQ Test	DL Test – Cp	DL Test – Kp	GS Test
t-statistic	434.95	0.64	1.45	0
p-value	0***	0***	0***	NA

Note: * Significant at the 10% level, ** Significant at the 5% level, *** Significant at the 1% level.

Table 20 contains the results of the different Hurst exponent calculation methods. While they arrive at different values, their overall conclusion is that there is persistence (long-term memory) in the data. This would undermine weak form efficiency.

Table 4.20: Hurst Exponent - Daily Returns

Hurst Exponent Type	Simple R/S Hurst Exponent	Corrected R/S Hurst Exponent	Empirical Hurst Exponent	Corrected Empirical Hurst Exponent	Theoretical Hurst Exponent
Statistic Value	0.62	0.62	0.52	0.50	0.53

Finally, the data is then tested for non-linearity using the BDS test prior to deciding between an ARIMA or SETAR model. The results of the BDS test in Table 21 are inconclusive given that when using 3 as an embedding dimension, the results point to non-linearity, while an embedding dimension = 2 points to linearity. The SETAR test is therefore used as a robustness check at 2 and 3 embedding dimensions.

Table 4.21: BDS Test - Daily Returns

Standard Normal =				
	[0.81]	[1.62]	[2.42]	[3.23]
[2]	-1.08	-1.01	-0.84	-0.30
[3]	2.53	2.72	3.19	3.99
P-Value =				
	[0.81]	[1.62]	[2.42]	[3.23]
[2]	0.28	0.31	0.40	0.76
[3]	0.01**	0.01**	0***	0***

Note: * Significant at the 10% level, ** Significant at the 5% level, *** Significant at the 1% level.

The SETAR test results are displayed in table 22 below. They show that an ARIMA model is favoured over a SETAR model. A tentative ARIMA model is auto-generated to explore the Akaike Information Criterion (AIC), Log-Likelihood as well as possible MA and AR lag orders. This is accomplished with the help of the "auto.arima" function in the *forecast* package in R.

Table 4.22: SETAR Test

Test of linearity against setar (2) and setar (3)		
	Test	P-value
1vs2	5.83	0.80
1vs3	13.27	0.80
Test of linearity against setar (2) and setar (3)		
	Test	P-value
1vs2	7.51	0.70
1vs3	13.01	1

Note: * Significant at the 10% level, ** Significant at the 5% level, *** Significant at the 1% level.

The Martingale Difference Hypothesis test results are shown in Table 23. The automatic portmanteau (AQ) test's null hypothesis of no return autocorrelation is rejected at the 5% level of significance if the value of AQ is greater than 3.84. The AQ test results (83) reject the null of no return autocorrelation. The Dominguez-Lobato test results both have p-values > 0.05. This means that the null of martingale difference sequence (efficiency) cannot be rejected. In contrast, the GS test returns a p-value = 0, also indicating that the null of weak-form efficiency can be rejected. Overall, the results of the AQ test, the GS and the Dominguez-Lobato tests indicate that there is both linear and non-linear dependence.

Table 4.23: Martingale Difference Hypothesis - Weekly Returns

Test	AQ Test	DL Test – Cp	DL Test – Kp	GS Test
t-statistic	83.62	0.13	0.95	0
p-value	0***	0.4	0.2	NA

Note: * Significant at the 10% level, ** Significant at the 5% level, *** Significant at the 1% level.

The Hurst exponents of the weekly returns are in table 24. The different Hurst exponent calculation methods arrive at different values, their overall conclusion is that there is persistence (long-term

memory) in the data. In concert with the other Martingale Difference Hypothesis test, this would undermine weak form efficiency.

Table 4.24: Hurst Exponent - Weekly Returns

Hurst Exponent Type	Simple R/S Hurst Exponent	Corrected R/S Hurst Exponent	Empirical Hurst Exponent	Corrected Empirical Hurst Exponent	Theoretical Hurst Exponent
Statistic Value	0.57	0.62	0.44	0.41	0.55

The BDS test results in Table 25 conclusively show that the real component of the Fourier Transformed weekly return series is best represented by a linear process; this therefore favours an ARIMA model over a SETAR model.

Table 4.25: BDS Test - Weekly Returns

Standard Normal =				
	[0.82]	[1.64]	[2.46]	[3.27]
[2]	-1.50	-1.56	-1.35	-0.67
[3]	0.84	0.67	0.02	0.42
P-Value =				
	[0.82]	[1.64]	[2.46]	[3.27]
[2]	0.13	0.12	0.18	0.50
[3]	0.40	0.51	0.98	0.68

Note: * Significant at the 10% level, ** Significant at the 5% level, *** Significant at the 1% level.

A tentative ARIMA model is auto-generated to explore the Akaike Information Criterion (AIC), Log-Likelihood as well as possible MA and AR lag orders. A tentative ARIMA model with the help of the "auto.arima" function in the *R* package "*forecast"* is entertained. The coefficients of the model,

an AR(4), and its associated diagnostic measures are displayed in Table 26. Notably the MAPE for the Fourier Transformed daily model vs the SETAR model using daily, untransformed, return data is 179% vs 144%. This means that the Fourier Transformed daily model is outperformed by the non-transformed model.

Table 4.26: Daily ARIMA Model

Coefficients	AR(1)	AR(2)	AR(3)	AR(4)
	0.05	0.23	-0.03	0.19
Standard Error	0.02	0.02	0.02	0.02
Error Measures	ME	RMSE	MAPE	MASE
	0.01	1.52	179.61	0.67
Diagnostics	σ^2	Log-likelihood	AIC	BIC
	2.31	-5480.61	10971.22	11001.22

The generated model is a SARIMA (2,0,2) (2,0,0) with a non-zero mean. the MAPE for the Fourier Transformed weekly model is 265% vs 161%. This means that the Fourier Transformed weekly model is outperformed by the non-transformed model. The coefficients of the model and its associated diagnostic measures are displayed in Table 27.

Table 4.27: Weekly SARIMA Model

Coefficients	AR(1)	AR(2)	MA(1)	MA(2)	SA(1)	SAR(2)	Mean
	-0.10	0.71	0.14	-0.44	0.14	-0.01	0.13
Standard Error	0.10	0.10	0.13	0.12	0.01	0.1	0.15
Error Measures and Diagnostics							
Error Measures	ME	RMSE	MAPE	MASE			
	-0.011	1.52	265.02	0.70			
Diagnostics	σ^2	Log-likelihood	AIC	BIC			
	2.40	-782.61	1581.23	1613.66			

The normality test results for daily model's residuals are displayed in Table 28. The Jarque-Bera test returns a p-value > 0.05, indicating that the null of normality in the residuals cannot be rejected. Similarly, the Shapiro-Wilk test also has a p-value > 0.05, also indicating that the null of normality in the residuals cannot be rejected.

Table 4.28: Daily Model Residual Normality Tests

Test	Jarque-Bera	Shapiro-Wilk
Test Statistic	3.15	1
p-value	0.21	0.38

Note: * Significant at the 10% level, ** Significant at the 5% level, *** Significant at the 1% level.

As shown by Table 29, the majority of the Hurst exponents lie above 0.50. This indicates that there is long term dependence in the residuals. Taking the Hurst exponent as a decision criterion regarding whether to continue modelling, the conclusion is therefore that more advanced attempts to model the return generating process are required.

Table 4.29: Daily Residuals Hurst Exponents

Hurst Exponent Type	Simple R/S Hurst Exponent	Corrected R/S Hurst Exponent	Empirical Hurst Exponent	Corrected Empirical Hurst Exponent	Theoretical Hurst Exponent
Statistic Value	0.56	0.54	0.45	0.42	0.53

The weekly model's normality test results are contained Table 30 below. The Jarque-Bera test returns a p-value > 0.05, indicating that the null of normality in the residuals cannot be rejected. Similarly, the Shapiro-Wilk test also has a p-value > 0.05, also indicating that the null of normality in the residuals cannot be rejected.

Table 4.30: Weekly Model Residuals Normality Tests

Test	Jarque-Bera	Shapiro-Wilk
Test Statistic	0.96	0.62
p-value	0.62	0.64

Note: * Significant at the 10% level, ** Significant at the 5% level, *** Significant at the 1% level.

The weekly residuals' corrected empirical and Empirical Hurst exponents are lower than 0.5, indicating no need to continue modelling the return series. However, most of the Hurst exponents in Table 31 point to further need to continue modeling the weekly return generating process.

Table 4.31: Weekly Residuals Hurst Exponents

Hurst Exponent Type	Simple R/S Hurst Exponent	Corrected R/S Hurst Exponent	Empirical Hurst Exponent	Corrected Empirical Hurst Exponent	Theoretical Hurst Exponent
Statistic Value	0.51	0.53	0.38	0.34	0.55

Figure 7 shows that the AR(4) model's residuals exhibit heavy bunching around the mean (0). Whether there is autocorrelation in the residuals is examined via the ACF plot.

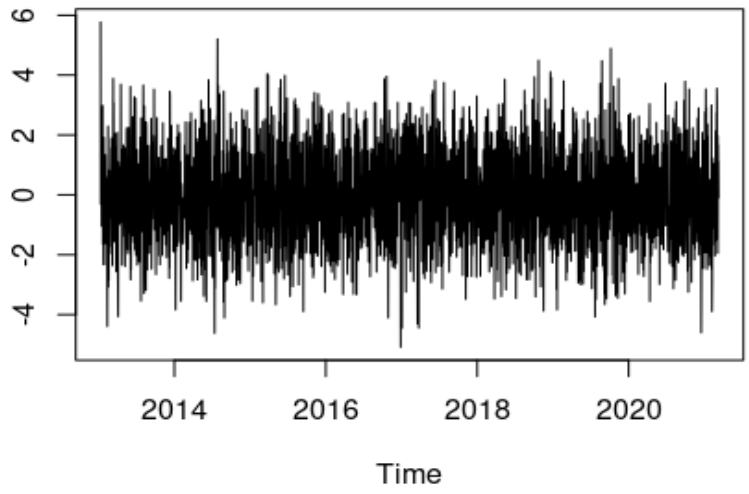
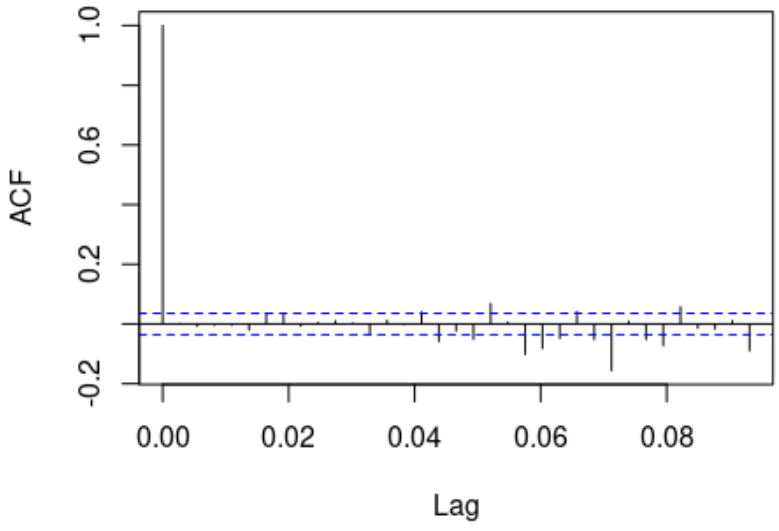


Figure 4.7: Daily AR Model Residuals

The ACF plot in Figure 8, shows that there is significant autocorrelation in the transformed daily model's residuals. This indicates that the model may not be a good fit for the data and may create room for further modelling.



Note: The blue bands indicate the line of significant, at lags where it is crossed, it can be inferred that there is significant autocorrelation in the time series under investigation.

Figure 4.8: Daily AR(4) Residuals' ACF Plot

The plot of the weekly model's residuals in Figure 9 below. The residuals exhibit heavy bunching around the mean (0). Whether there is autocorrelation in the returns is examined via the ACF plot.

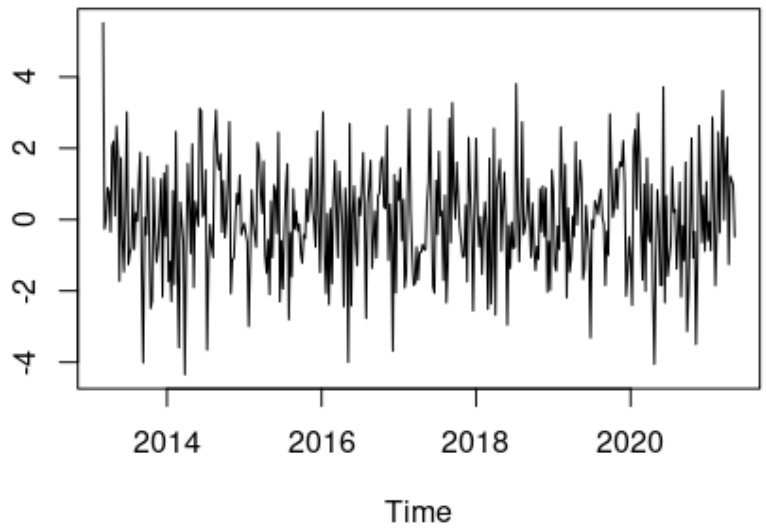
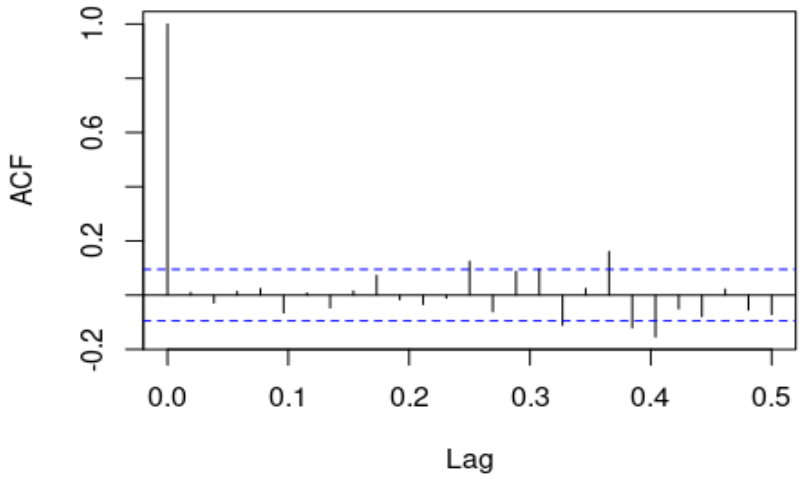


Figure 4.9: Weekly SARIMA Residuals

The ACF plot in Figure 10, shows that there is significant autocorrelation in the transformed daily model's residuals at a few lags. This indicates that the model may not be a good fit for the data and may create room for further modelling.



Note: The blue bands indicate the line of significant, at lags where it is crossed, it can be inferred that there is significant autocorrelation in the time series under investigation.

Figure 4.10: Weekly SARIMA Residuals ACF

The Fourier Transformed and non-transformed data thus lead us to the same conclusion. We can therefore proceed to utilising machine learning approaches to model the return series.

4.7 Machine Learning Approaches to Modelling the RGP

The data is partitioned along a 70-30 line, where 70% of the data is set aside for training and 30% for validation. The LSTM model uses a dense, dropout and LSTM layer (with a "sigmoid" activation function). It is also stateful, meaning that it imports the experience that is gleaned from prior sequences to the current sequence. This is useful when dealing with asset returns, where the current return is affected by historical data. The Adaptive Moment Estimation (Adam) optimization algorithm is used which provide is adept at handling gradients on noisy problems. There is also a need to re-scale the input data to fit the LSTM's activation function's range. By default, the activation function for an LSTM is a sigmoid function, whose range is [-1, 1] and as such, the normalization of the data needs to ensure that the data fits within this range. The minimum and maximum values of the training data set are used as the scaling coefficients to scale both the training and testing data sets. They are also used to scale the predicted values. This effectively ensures therefore that the minimum and maximum values of the test data do not influence the model.

The defined model is a sequential model with 12 950 parameters, all of which are trainable. This model is then fit to the training data with 20 epochs. The daily model is displayed in Table 32.

Table 4.32: Daily LSTM Model

Layer (type)	Output Shape	Parameter No.
LSTM layer	(1,50)	10400
Dense layer	(1,50)	2550
Model Parameters		
Total parameters		12950
Trainable parameters		12950
Non-trainable parameters		0

The weekly LSTM is outlined in Table 33. The defined model is a sequential model with 12 950 parameters, all of which are trainable. This model is then fit to the training data with 20 epochs.

Table 4.33: Weekly LSTM

Layer (type)	Output Shape	Parameter No.
LSTM layer	(1,50)	10400
Dense layer	(1,50)	2550
Model Parameters		
Total parameters		12950
Trainable parameters		12950
Non-trainable parameters		0

Figure 11 indicates that the model is performing well as the loss (Root Mean Squared Error) is declining as training occurs. There is evidence of diminishing marginal returns post the fifth epoch, however.

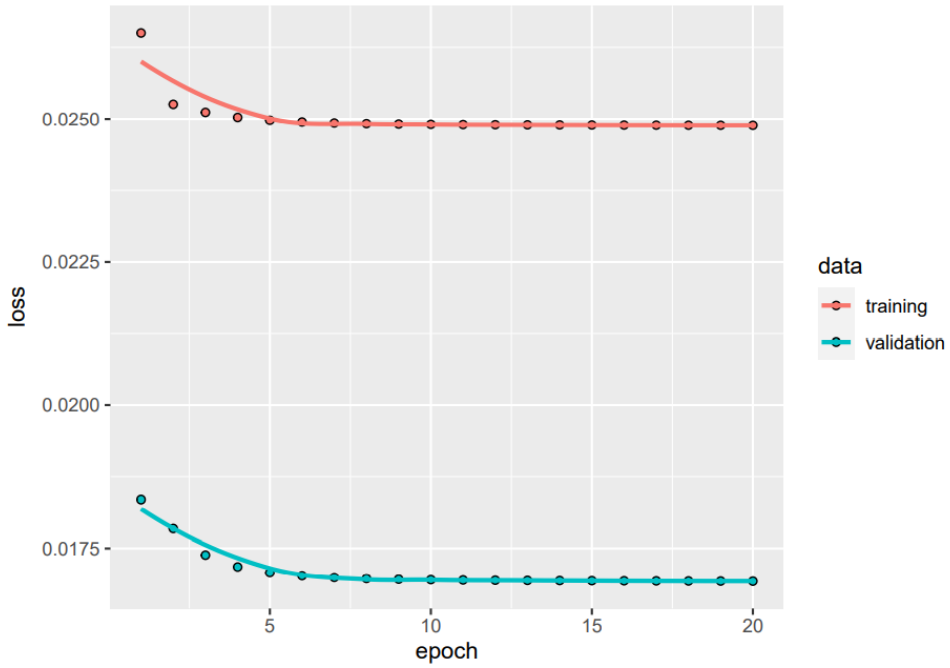


Figure 4.11: Daily LSTM Diagnostics

The daily LSTM's residuals are displayed in Figure 12 below, notably, for all lags, there is no statistically significant autocorrelation. This indicates that the model is a good fit for the data but whether further modelling (with external regressors) is required hinges on further formal statistical tests.

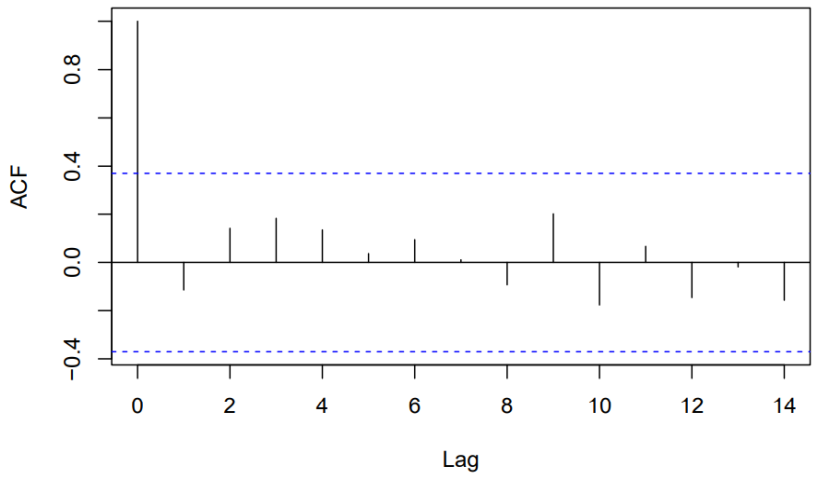


Figure 4.12: Daily Residuals ACF Plot

The weekly model in Figure 13 is performing much better during training than it does during validation stages. This may indicate that the model (especially after the fifth epoch, where the validation loss is not declining and is in fact, increasing) is overfitting, *i.e.*, it has learnt noise and thus does not generalize well to validation data.

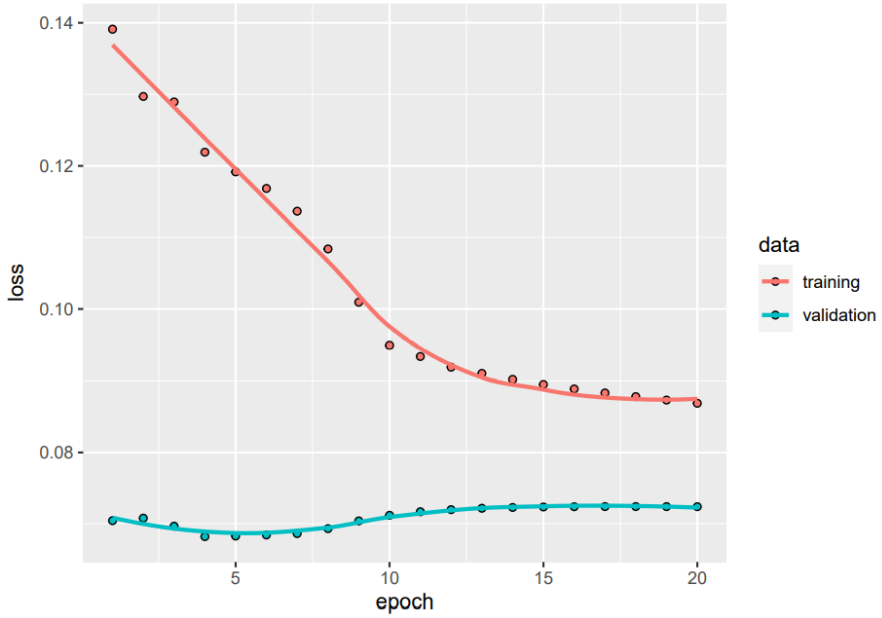


Figure 4.13: Weekly LSTM Diagnostics

The weekly LSTM's ACF plot is displayed in Figure 14. There is significant autocorrelation on lag 20. This indicates that the weekly LSTM model may benefit from the addition of external regressors. Prior to that, however, formal residual analysis is undertaken to determine if this is indeed true.

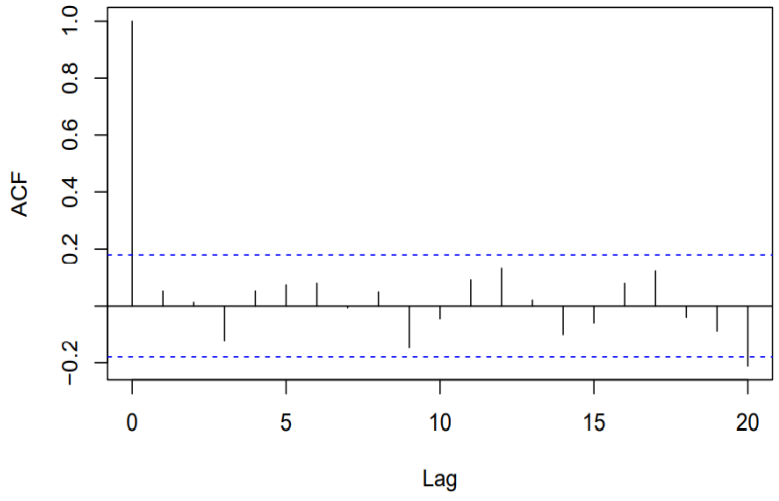


Figure 4.14: Weekly LSTM Residuals ACF Plot

Daily Model Residual Analysis

The results of the Martingale Difference Hypothesis are displayed in Table 34 below. The AQ test statistic is less than 3.84, as such there is failure to reject the null hypothesis of no autocorrelation in

the residuals. The Dominguez-Lobato test results have p-values > 0.05. This means that the null that the residuals are a Martingale Difference Sequence (efficiency) cannot be rejected. Similarly, the GS test returns a p-value = 0.1, indicating that the null that the residuals follow a random walk cannot be rejected. This, in addition to a visual inspection of the ACF plot of the residuals indicates that there is no significant autocorrelation in the residuals. On its own, this is an important milestone over the traditional statistical approach. (SETAR)

Table 4.34: Martingale Hypothesis Test - Daily Residuals

Test	AQ Test	DL Test – Cp	DL Test – Kp	GS Test
t-statistic	0.49	0.22	0.99	0.1
p-value	0.49	0.8	0.4	NA

Note: * Significant at the 10% level, ** Significant at the 5% level, *** Significant at the 1% level.

The acid test for whether to continue to model with external regressors is based on the outcome of the Hurst exponent in Table 35. The various Hurst exponents overwhelmingly agree that the residuals can be further modelled since they exceed 0.5. This means that a NAR with external regressors (NARX) may be employed as an advanced form of modelling the return generating process.

Table 4.35: Daily Residuals Hurst Exponent

Hurst Exponent Type	Simple R/S Hurst Exponent	Corrected R/S Hurst Exponent	Empirical Hurst Exponent	Corrected Empirical Hurst Exponent	Theoretical Hurst Exponent
Statistic Value	0.58	0.99	0.99	0.82	0.62

Weekly Model Data

The results of the Martingale Difference Hypothesis are displayed in Table 36 below. The null hypothesis of no return autocorrelation is rejected at the 5% level of significance if the value of AQ is greater than 3.84. The AQ test results cannot reject the null hypothesis of no autocorrelation in the residuals. The Dominguez-Lobato test results have p-values > 0.05. This means that the null that the residuals are a Martingale Difference Sequence (efficiency) cannot be rejected. Similarly, the GS test

returns a p-value = 0.65, also indicating that the null that the residuals follow a random walk cannot be rejected. This, in addition to a visual inspection of the ACF plot of the residuals indicates that there is no significant autocorrelation in the residuals. Similar to the results of the daily residuals, this is an important milestone over the traditional statistical approach. (SETAR)

Table 4.36: Martingale Hypothesis Test – Weekly Residuals

Test	AQ Test	DL Test – Cp	DL Test – Kp	GS Test
t-statistic	0.30	0.11	0.4	0.65
p-value	0.58	0.40	0.60	NA

Note: * Significant at the 10% level, ** Significant at the 5% level, *** Significant at the 1% level.

The Hurst Exponent results are displayed in Table 37 below. The various Hurst exponents overwhelmingly agree that the residuals can be further modelled since they exceed 0.5. This means that a NAR model with external regressors (i.e., NARX) may be employed.

Table 4.37: Weekly Residuals Hurst Exponent

Hurst Exponent Type	Simple R/S Hurst Exponent	Corrected R/S Hurst Exponent	Empirical Hurst Exponent	Corrected Empirical Hurst Exponent	Theoretical Hurst Exponent
Statistic Value	0.54	0.67	0.54	0.57	0.52

4.8 Machine Learning Approaches to Modelling the RGP (FT)

The same process is repeated with a Fourier Transformed data set. The defined model is a sequential model with 12 950 parameters, all of which are trainable. This model is then fit to the training data with 20 epochs. The daily model is displayed in Table 38.

Table 4.38: Daily LSTM Model

Layer (type)	Output Shape	Parameter No.
LSTM layer	(1,50)	10400
Dense layer	(1,50)	2550
Model Parameters		
Total parameters		12950
Trainable parameters		12950
Non-trainable parameters		0

Figure 15 indicates that the model is performing well as the loss (Root Mean Squared Error) is declining as training occurs. Interestingly, relative to the untransformed results, training continues to provide benefits even at later epochs.

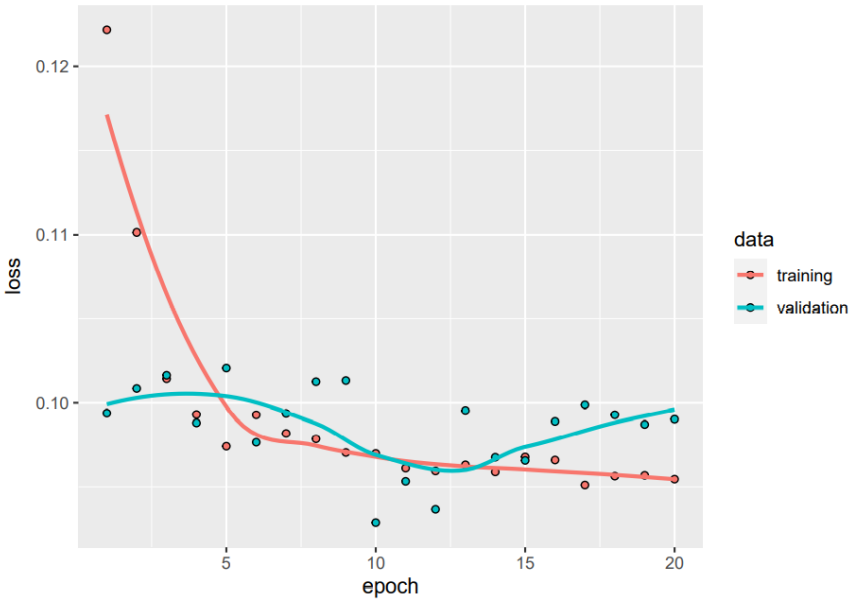


Figure 4.15: Daily LSTM Diagnostics (FT)

The plot of the daily model's residuals in Figure 16 below. The residuals have a wide dispersion around a slightly negative mean. Whether there is autocorrelation in the returns is examined via the ACF plot.

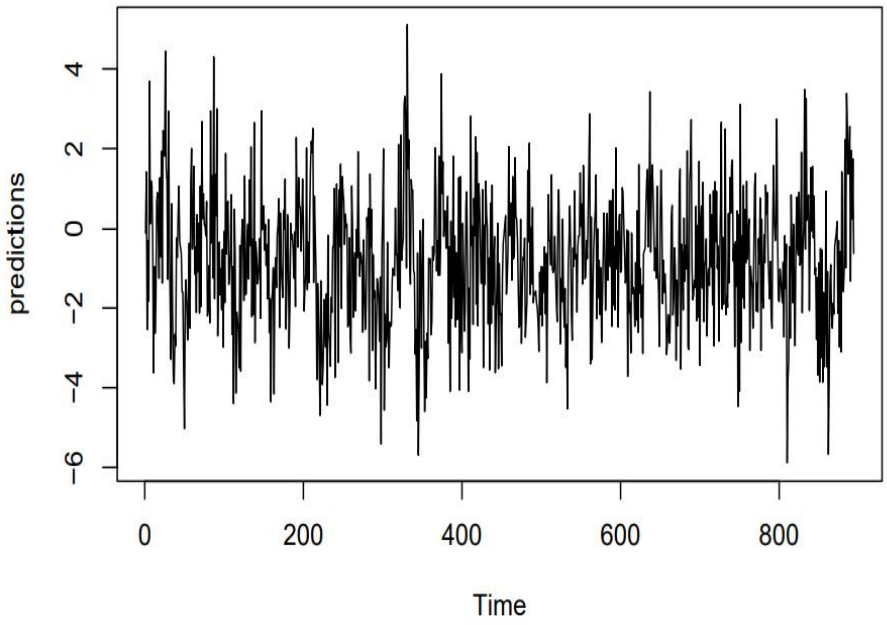
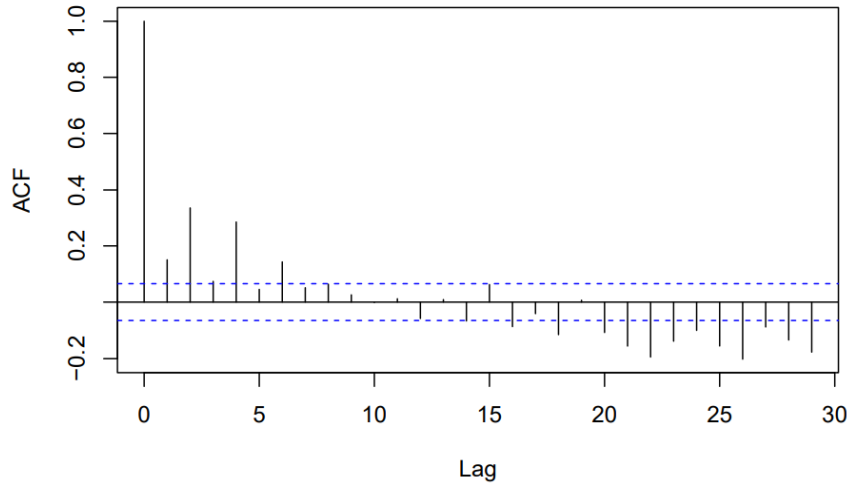


Figure 4.16: Daily Residuals Plot

The daily LSTM's ACF plot is displayed in Figure 17. There is significant autocorrelation on several lags. This indicates that the daily LSTM model may benefit from the addition of external regressors. This is further examined via formal residual analysis.



Note: The blue bands indicate the line of significant, at lags where it is crossed, it can be inferred that there is significant autocorrelation in the time series under investigation.

Figure 4.17: Daily Residuals ACF Plot

Daily Model Residual Analysis

The results of the Martingale Difference Hypothesis are displayed in Table 39 below. The AQ test statistic is greater than 3.84, as such, the null hypothesis of no autocorrelation in the residuals is rejected. This is in concert with a visual inspection of the ACF plot. The Dominguez-Lobato test results have p-values < 0.05. This means that the null that the residuals are a Martingale Difference Sequence (efficiency) can be rejected. Similarly, the GS test returns a p-value = 0, indicating that the null that the residuals follow a random walk is be rejected. All these are signs that the Fourier Transformation (at this stage, vis-à-vis the non-transformed sample) produces worse results when modelling is performed.

Table 4.39: Martingale Hypothesis Test - Daily Residuals

Test	AQ Test	DL Test – Cp	DL Test – Kp	GS Test
t-statistic	187.23	1.65	1.82	0
p-value	0***	0***	0***	NA

Note: * Significant at the 10% level, ** Significant at the 5% level, *** Significant at the 1% level.

The acid test for whether to continue to model with external regressors is based on the outcome of the Hurst exponent in Table 40. The various Hurst exponents overwhelmingly agree that the residuals can be further modelled since they exceed 0.5. This means that a NAR with external regressors (NARX) may be employed as an advanced form of modelling the return generating process.

Table 4.40: Daily Residuals Hurst Exponent

Hurst Exponent Type	Simple R/S Hurst Exponent	Corrected R/S Hurst Exponent	Empirical Hurst Exponent	Corrected Empirical Hurst Exponent	Theoretical Hurst Exponent
Statistic Value	0.56	0.57	0.43	0.41	0.54

Weekly Model

The weekly LSTM is outlined in Table 41. The defined model is a sequential model with 12 950 parameters, all of which are trainable. This model is then fit to the training data with 20 epochs.

Table 4.41: Weekly LSTM

Layer (type)	Output Shape	Parameter No.
LSTM layer	(1,50)	10400
Dense layer	(1,50)	2550
Model Parameters		
Total parameters		12950
Trainable parameters		12950
Non-trainable parameters		0

Figure 18 indicates that the model is performing well as the loss (Root Mean Squared Error) is declining as training occurs. Similar to the daily results, relative to the untransformed results, training continues to provide benefits even at later epochs, with the best epoch being epoch 10.

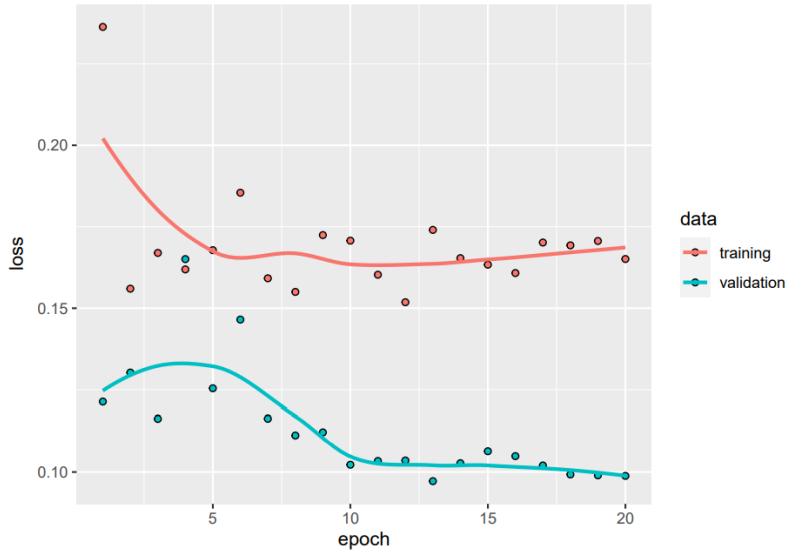


Figure 4.18: Weekly LSTM Diagnostics (FT)

The plot of the weekly model's residuals in Figure 19 below. The residuals have a wide dispersion around a near-zero mean. Whether there is autocorrelation in the returns is examined via the ACF plot.

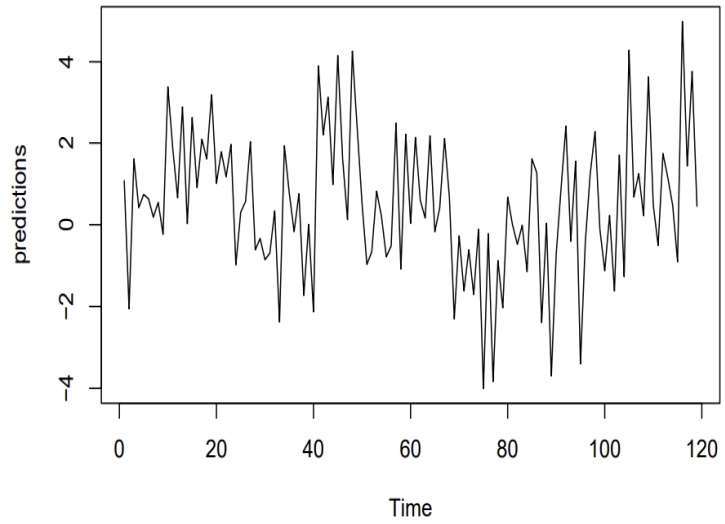
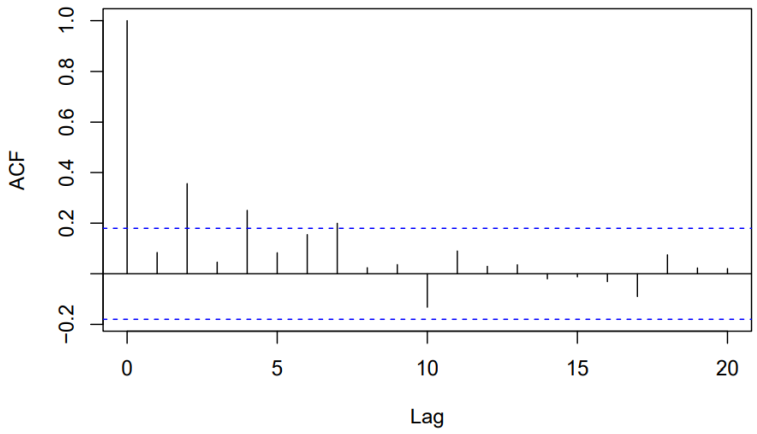


Figure 4.19: Weekly Residuals Plot

The weekly LSTM's ACF plot is displayed in Figure 20. There is significant autocorrelation on several lags. This indicates that the weekly LSTM model may benefit from the addition of external regressors. This is further examined via formal residual analysis.



Note: The blue bands indicate the line of significant, at lags where it is crossed, it can be inferred that there is significant autocorrelation in the time series under investigation.

Figure 4.20: Weekly Residuals ACF Plot

Weekly Model Residual Analysis

The results of the Martingale Difference Hypothesis are displayed in Table 42 below. The AQ test statistic is greater than 3.84, as such, the null hypothesis of no autocorrelation in the residuals is rejected. This is in concert with a visual inspection of the ACF plot. The Dominguez-Lobato test results have p-values > 0.05. This means that the null that the residuals are a Martingale Difference Sequence (efficiency) cannot be rejected. Similarly, the GS test returns a p-value = 0.15, indicating that the null that the residuals follow a random walk cannot be rejected. The weekly results differ significantly from their daily counterparts.

Table 4.42: Martingale Hypothesis Test – Weekly Residuals

Test	AQ Test	DL Test – Cp	DL Test – Kp	GS Test
t-statistic	11.71	0.11	0.68	0.15
p-value	0***	0.2	0.8	NA

Note: * Significant at the 10% level, ** Significant at the 5% level, *** Significant at the 1% level.

To decide whether to model with external regressors, the results of the Hurst exponent in Table 43 are analysed. The various Hurst exponents overwhelmingly agree that the residuals can be further modelled since they exceed 0.5. This means that a NAR with external regressors (NARX) may be employed as an advanced form of modelling the return generating process.

Table 4.43: Daily Residuals Hurst Exponent

Hurst Exponent Type	Simple R/S Hurst Exponent	Corrected R/S Hurst Exponent	Empirical Hurst Exponent	Corrected Empirical Hurst Exponent	Theoretical Hurst Exponent
Statistic Value	0.65	0.76	0.63	0.65	0.52

4.9 NARX Modelling

Seetharam (2016) argues that where a machine learning approach that only utilises historical information fails, a non-linear autoregressive model with exogenous inputs (NARX) can be turned to. The model is represented mathematically as:

$$y_t = F(y_{t-1}, y_{t-2}, y_{t-3}, \ldots, u_t, u_{t-1}, u_{t-2}, u_{t-3}, \ldots) + \epsilon_t \quad (16)$$

Where the function F is a nonlinear function (specifically, a Support Vector Machine), y represents returns and the lagged values of y as well as information about u, the exogenous input, helps in the prediction of y. The NARX model is produced from the R packages "*mltsp*" and "*e1071*" that have capabilities for building SVMs into NARX models that are used for regression. Furthermore, similar to a linear ARIMA-family model, the NARX model will use the (p,q)(P,Q) structure to denote its internal make-up, with the difference being that the SVM in this instance dictates the makeup of the NARX model (instead of it being built on the ACF and PACF structures).

4.9.1 External Regressors

External regressor data is uploaded and the "*tempdisagg*" package is used to temporally disaggregate them where necessary. This effectively converts time series data from a lower frequency (such as quarterly/monthly/weekly) to higher frequency data (daily/weekly) as needed. The Denton-Cholette method performs an interpolation that meets the temporal additivity constraint. Loss of information is confirmed via visual inspection of the disaggregated and original time series: if no visible distortions are there, the series is accepted as an external regressor.

4.9.1.1 MSCI World Index

The MSCI World Index possesses an upward trend. There is a sharp decline in early 2020 that corresponds to the global proliferation of the novel coronavirus, Covid-19. MSCI publishes a daily index, which unfortunately had several missing days in its data set. As such, the weekly index is used as a close substitute and is then temporally disaggregated into a daily dataset using the Denton-Cholette method. Its output is graphically displayed in Figure 21.

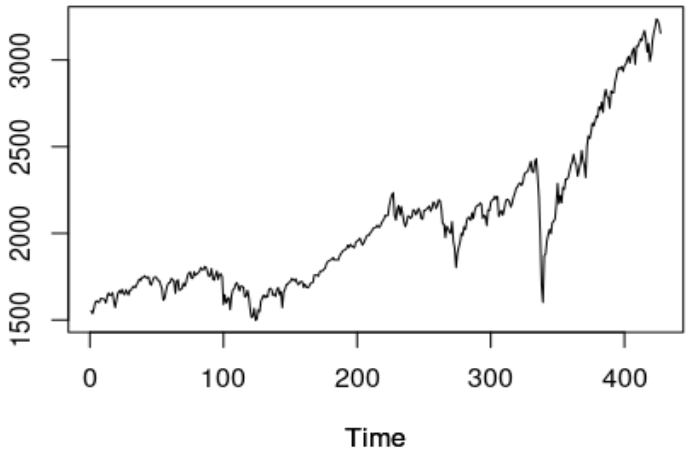


Figure 4.21: MSCI World Index

Figure 22 displays the disaggregated MSCI World Index. The temporally disaggregated index closely matches the original weekly data set. This indicates that the Denton-Cholette method was successful in the conversion and resulted in minimal informational loss. This disaggregation should therefore have negligible impact on the final NARX model. The disaggregated index is thus included as an external regressor.

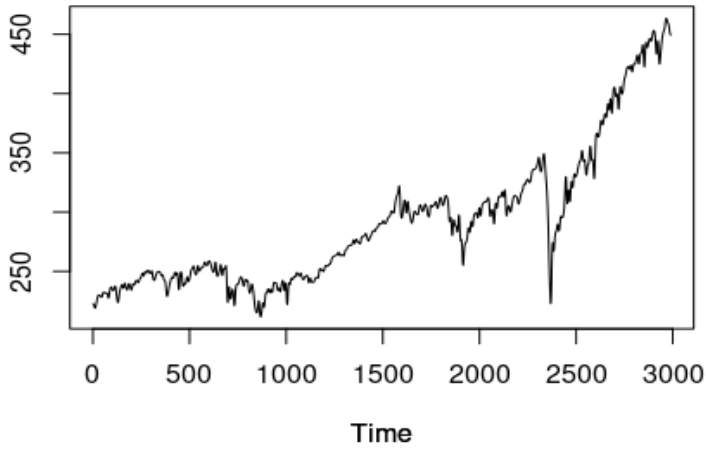


Figure 4.22: Disaggregated MSCI World Index

4.9.1.2 Google Trends

Google Trends provides search trends data on a "0 – 100" scale. The scale can be regarded as a proxy for worldwide interest in the search item. A "0" on the scale indicates virtually no interest in the item

while a “100” indicates that the item garnered a high level of interest around the world. Google provides Trend data on a monthly frequency. The data visualized in Figure 23 is collected from October 2013 to October 2021. While Bitcoin garnered little attention in its early years, it later experienced patches of exceedingly large amounts of attention.

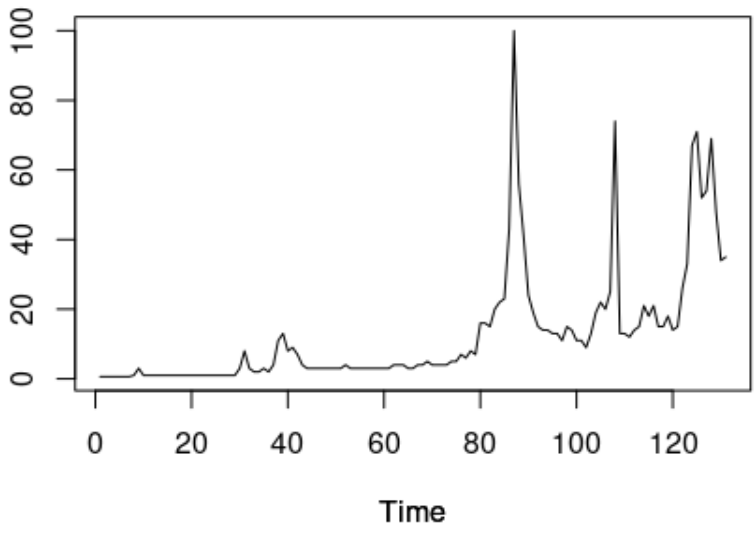


Figure 4.23: Google Trends Monthly Data

The disaggregated daily data in Figure 24 is visually similar to the original monthly data. This means that the disaggregation was successful. Using it as an external regressor thus will not negatively impact the model.

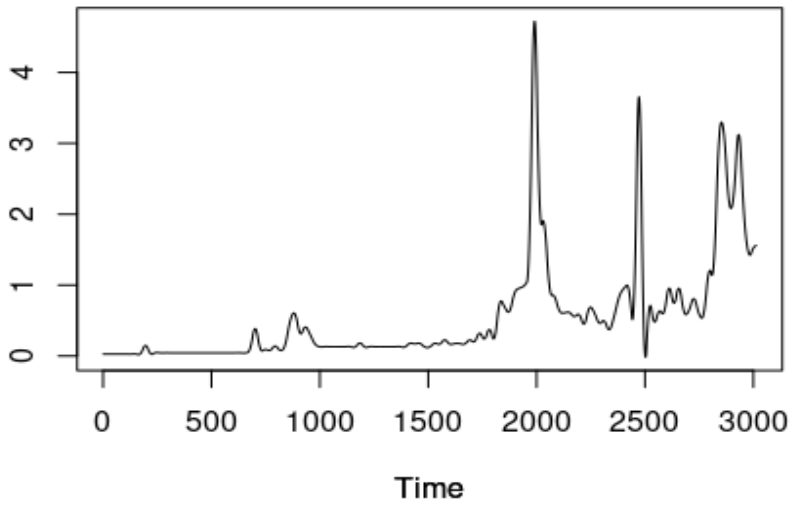


Figure 4.24: Disaggregated Bitcoin Trend Data

4.9.1.3 Federal Funds Rate

The US Federal Funds Rate is the rate at which depository institutions lend each other money at on the overnight market. This rate is set biquarterly (eight times in a year) when the Federal Open Market Committee (FOMC) meets. As evident in Figure 25, while the Federal Funds rate had an upward trajectory prior to Covid-19, it was sharply lowered in response to the pandemic as the Federal Reserve aimed to stimulate the US economy. Given that this data is supplied on a biquarterly basis, it is thus temporally disaggregated to convert it into daily data.

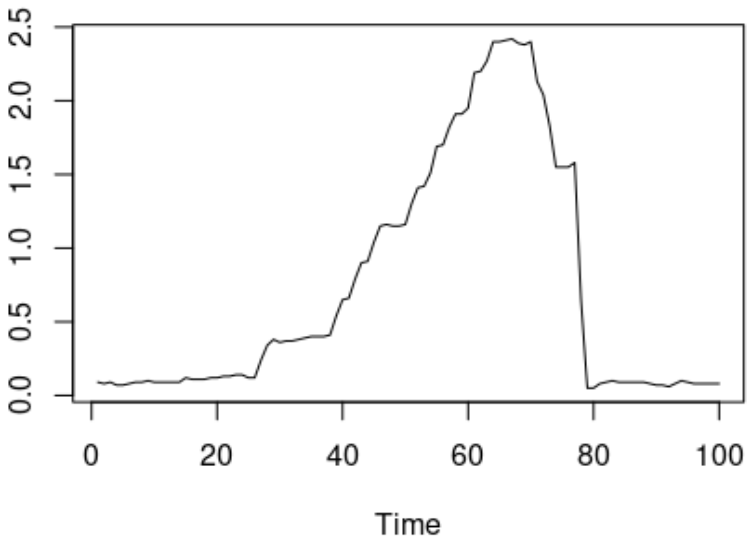


Figure 4.25: Federal Funds Rate

The disaggregated daily data in Figure 26 is visually similar to the original biquarterly data. This means that the disaggregation was successful. Using it as an external regressor thus will not negatively impact the model.

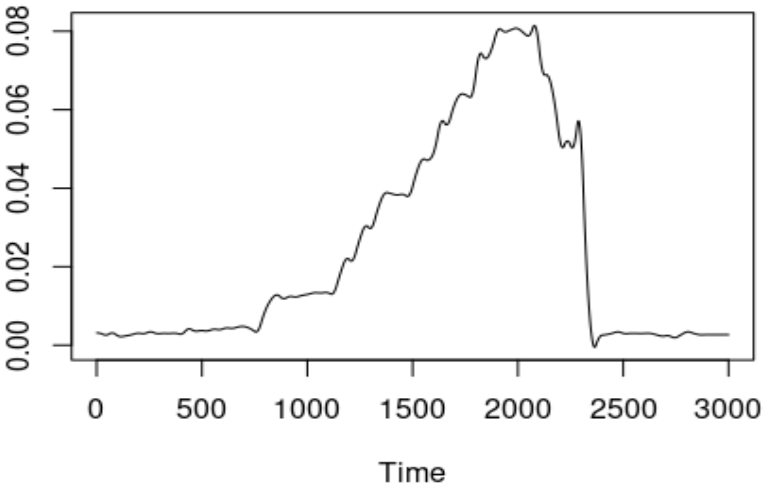


Figure 4.26: Disaggregated Federal Funds Rate

4.9.1.4 United States' Gross Domestic Product

The US GDP is taken as a proxy for the size of the US economy. This data is provided on an annual basis and is graphically represented by Figure 27.

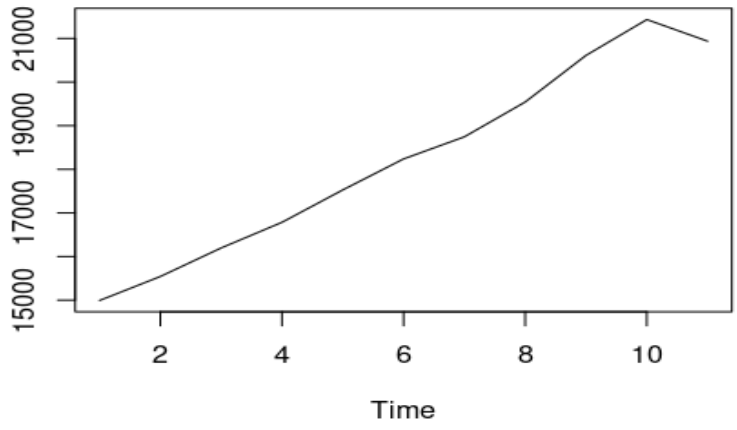


Figure 4.27: US Annual GDP

While appearing visually smoother, the disaggregated data in Figure 28 captured two important items; the Covid-19 induced contraction of the economy in 2020 and the US GDP's overall upward trajectory. This means that the disaggregation was successful. Using it as an external regressor thus will not negatively impact the model.

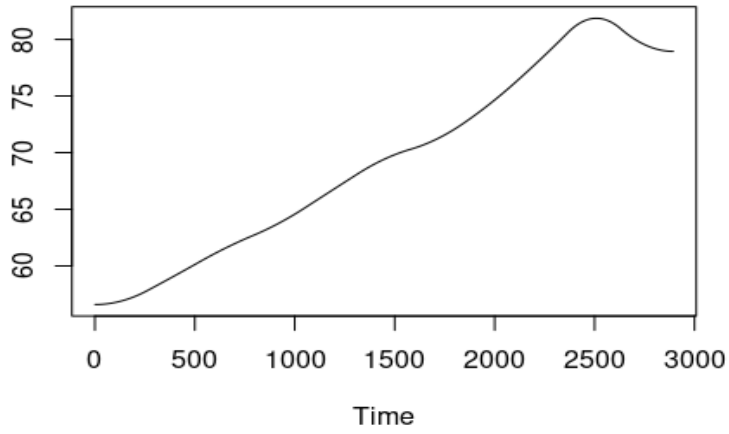


Figure 4.28: Disaggregated US GDP

4.9.2 Daily and Weekly NARX Models

In the daily and weekly NARX models, we similarly employ a 70-30 partition of the data. Table 44 displays the chosen NARX model. The NARX (4,0) (0,0) model uses an SVM learner with a radial kernel. It has 1682 support vectors.

Table 4.44: Daily NARX Model

SVM.default	(x = input, y = target)
Parameters	
SVM Type	Eps-Regression
SVM Kernel	Radial
Cost	1
Epsilon	0.1
Number of Support Vectors	1682

Table 45 displays the chosen NARX model, which is a NARX (2,0) (0,0) that uses an SVM learner along with a radial kernel. It has 261 support vectors.

Table 4.45: Weekly NARX Model

SVM.default	(x = input, y = target)
Parameters	
SVM Type	Eps-Regression
SVM Kernel	Radial
Cost	1
Gamma	0.33
Epsilon	0.1
Number of Support Vectors	261

The Daily NARX model's residuals in Figure 29 have a zero mean. They however exhibit volatility clustering, which is not typical for a white noise process. The test for whether there is autocorrelation in the residuals (a mark of a model that is ill-fitted to the data) is a visual inspection of the ACF plot.

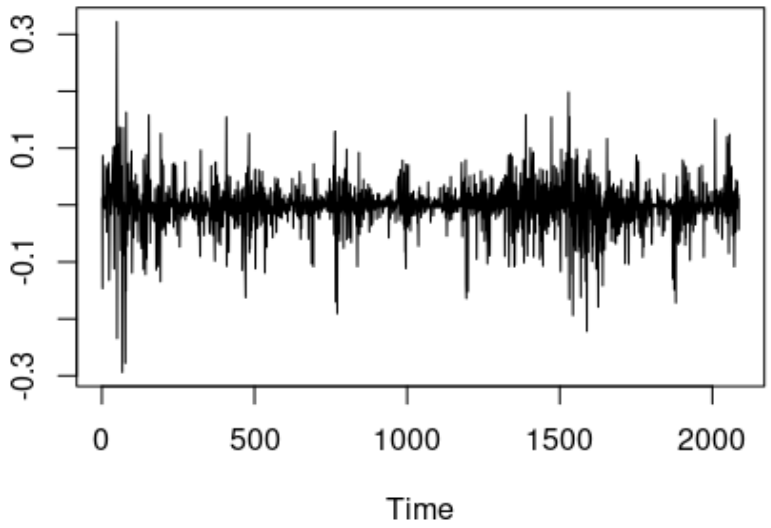


Figure 4.29: Daily NARX Model

Similarly, the Weekly NARX model's residuals in Table 30 below have a zero mean. The residual series also exhibits volatility bunching.

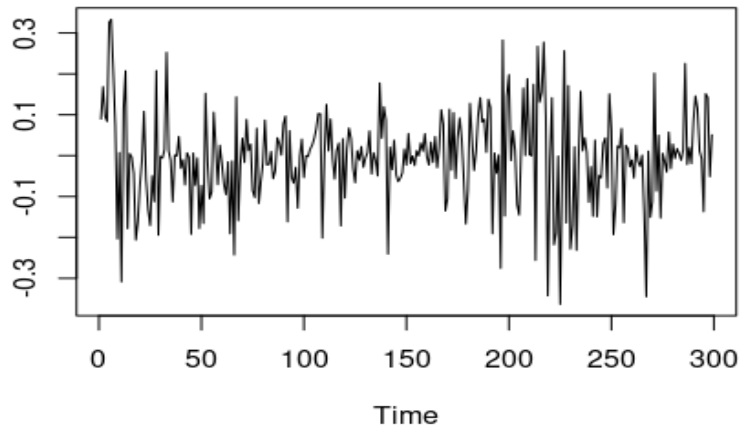
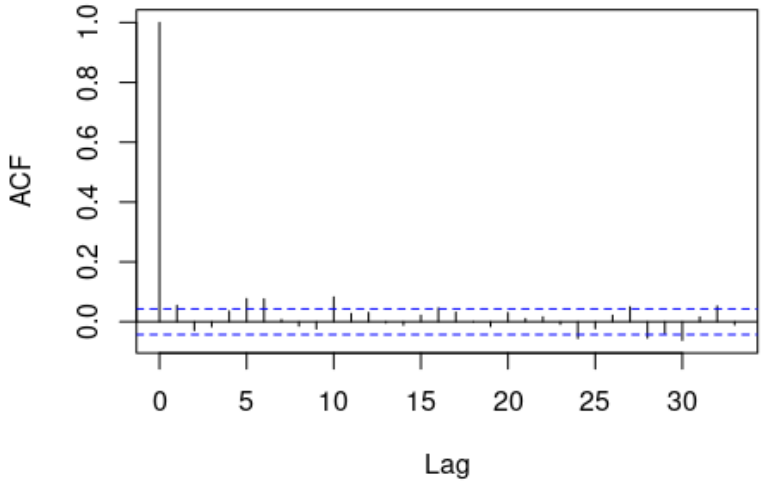


Figure 4.30: Weekly NARX Residuals

The ACF plot of the NARX model's residuals in Figure 31 shows signs of significant autocorrelation at several lags. This means that there is further information that is not reflected in Bitcoin daily returns.

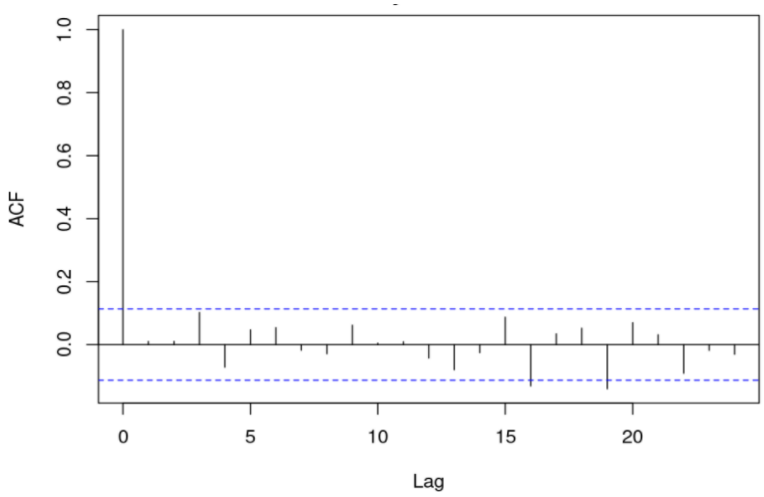


Note: The blue bands indicate the line of significant, at lags where it is crossed, it can be inferred that there is significant autocorrelation in the time series under investigation.

Figure 4.31: Daily NARX Residuals ACF

In the weekly NARX model's residuals ACF plot in Figure 32, there are two borderline cases of significant autocorrelation at lag 16 and 19. However, for the most part, there is no autocorrelation in

the weekly NARX's residuals which may indicate model suitability which can be confirmed with formal residual analysis.



Note: The blue bands indicate the line of significant, at lags where it is crossed, it can be inferred that there is significant autocorrelation in the time series under investigation.

Figure 4.32: Weekly NARX Residuals ACF

The Martingale Difference Hypothesis Test results are presented in Table 46. The null hypothesis of no return autocorrelation is rejected at the 5% level of significance if the value of AQ is greater than 3.84. The AQ test results cannot reject the null hypothesis of no return autocorrelation. The Dominguez-Lobato test results have p-values < 0.05. This means that the null that the residuals are a Martingale Difference Sequence (efficiency) can be rejected. In contrast, the GS test returns a p-value = 0.25, also indicating that the null that the residuals follow a random walk cannot be rejected. This, in addition to a visual inspection of the ACF plot of the residuals indicates sparse significant autocorrelation in the residuals, indicating that the model can be accepted.

Table 4.46: Martingale Difference Hypothesis Tests - Daily Residuals

Test	AQ Test	DL Test – Cp	DL Test – Kp	GS Test
t-statistic	1.79	2.10	2.15	0.25
p-value	0.18	0***	0***	NA

Note: * Significant at the 10% level, ** Significant at the 5% level, *** Significant at the 1% level.

The weekly Martingale Difference Hypothesis Test results are presented in Table 47. The null hypothesis of no return autocorrelation is rejected at the 5% level of significance if the value of AQ is greater than 3.84. The AQ test results cannot reject the null hypothesis of no return autocorrelation. The Dominguez-Lobato test results have p-values > 0.05. This means that the null that the residuals are a Martingale Difference Sequence (efficiency) cannot be rejected. Similarly, the GS test returns a p-value = 0.25, also indicating that the null that the residuals follow a random walk cannot be rejected. This, in addition to a visual inspection of the ACF plot of the residuals indicates that the model can be taken as a good representation of the return generating process.

Table 4.47: Martingale Difference Hypothesis Test – Weekly Residuals

Test	AQ Test	DL Test – Cp	DL Test – Kp	GS Test
t-statistic	0.01	0.04	0.49	0.25
p-value	0.89	1	1	NA

Note: * Significant at the 10% level, ** Significant at the 5% level, *** Significant at the 1% level.

4.9.3 Daily and Weekly NARX Models (FT)

Table 48 displays the chosen NARX model. The NARX (4,0) (0,0) model uses an SVM learner with a radial kernel. It has 1916 support vectors.

Table 4.48: Daily NARX Model

SVM.default	(x = input, y = target)
Parameters	
SVM Type	Eps-Regression
SVM Kernel	Radial
Cost	1
Gamma	0.2
Epsilon	0.1
Number of Support Vectors	1916

Table 49 displays the chosen NARX model, which is a NARX (2,0) (0,0) that uses an SVM learner along with a radial kernel. It has 270 support vectors.

Table 4.49: Weekly NARX Model

SVM.default	(x = input, y = target)
Parameters	
SVM Type	Eps-Regression
SVM Kernel	Radial
Cost	1
Gamma	0.33
Epsilon	0.1
Number of Support Vectors	270

The Daily NARX model's residuals in Figure 33 have a zero mean. They however exhibit volatility clustering, which is not typical for a white noise process. The test for whether there is autocorrelation in the residuals (a mark of a model that is ill-fitted to the data) is a visual inspection of the ACF plot.

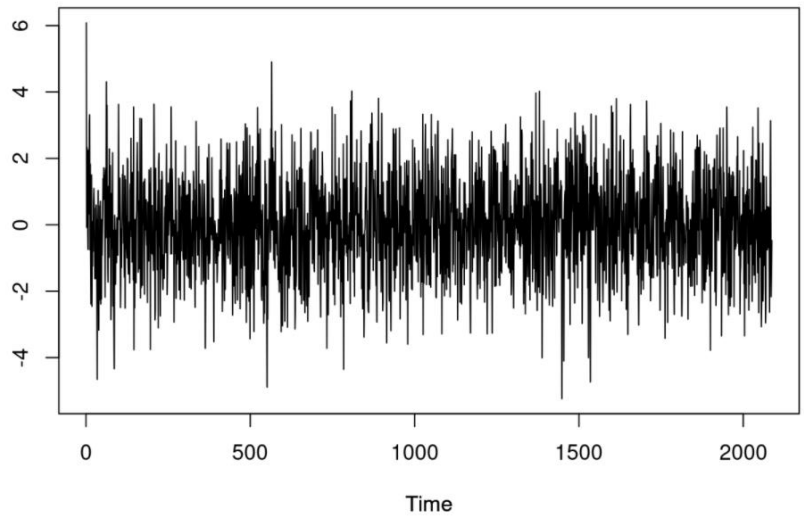


Figure 4.33: Daily NARX Residuals

Similarly, the Weekly NARX model's residuals in Table 34 below have a zero mean. The residual series also exhibits volatility bunching.

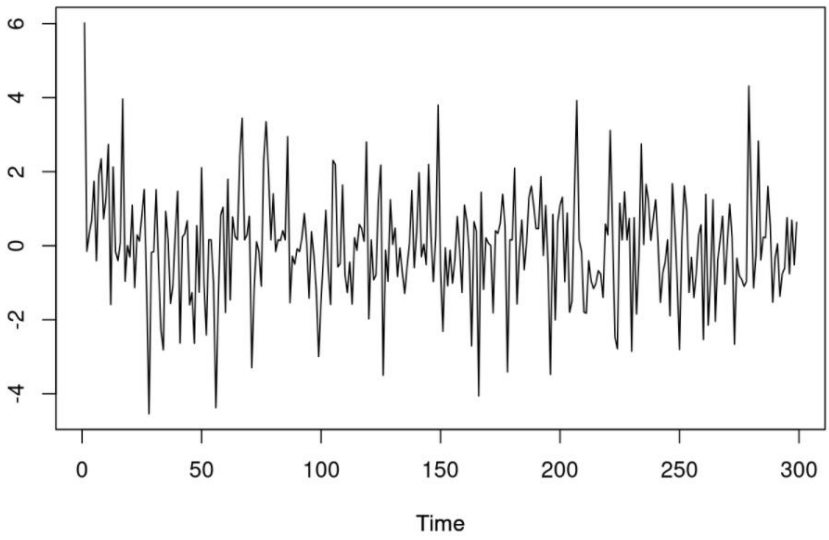
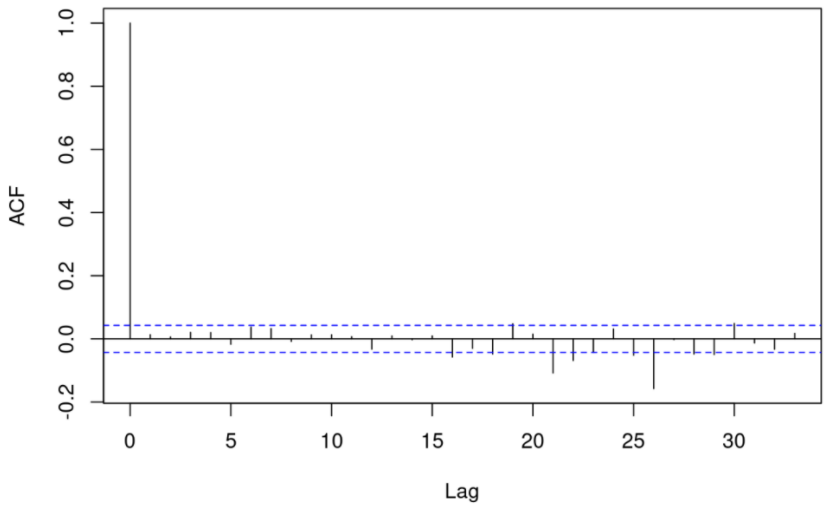


Figure 4.34: Weekly NARX Residuals

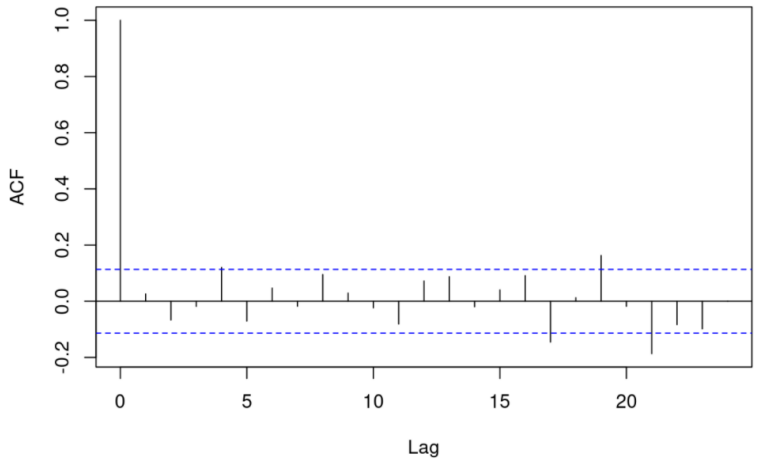
The ACF plot of the NARX model's residuals in Figure 35 shows signs of significant autocorrelation at three lags. This means that there is further information that is not reflected in Bitcoin daily returns.



Note: The blue bands indicate the line of significant, at lags where it is crossed, it can be inferred that there is significant autocorrelation in the time series under investigation.

Figure 4.35: Daily NARX Residuals' ACF

In the weekly NARX model's residuals ACF plot in Figure 36, there are two borderline cases of significant autocorrelation at lag 16 and 19. However, for the most part, there is no autocorrelation in the weekly NARX's residuals which may indicate potential model suitability.



Note: The blue bands indicate the line of significant, at lags where it is crossed, it can be inferred that there is significant autocorrelation in the time series under investigation.

Figure 4.36: Weekly NARX Residuals ACF

The Martingale Difference Hypothesis Test results are presented in Table 50. The null hypothesis of no return autocorrelation is rejected at the 5% level of significance if the value of AQ is greater than 3.84. The AQ test results cannot reject the null hypothesis of no return autocorrelation. The Dominguez-Lobato test results have p-values > 0.05. This means that the null that the residuals are a Martingale Difference Sequence (efficiency) cannot be rejected. In contrast, the GS test returns a p-value = 0.95, also indicating that the null that the residuals follow a random walk cannot be rejected. This is evidence that the model can be taken as a good representative of the return generating process.

Table 4.50: Martingale Difference Hypothesis Tests - Daily Residuals

Test	AQ Test	DL Test – Cp	DL Test – Kp	GS Test
t-statistic	0.38	0.09	0.72	0.95
p-value	0.54	0.2	0.4	NA

Note: * Significant at the 10% level, ** Significant at the 5% level, *** Significant at the 1% level.

The weekly Martingale Difference Hypothesis Test results are presented in Table 51. The null hypothesis of no return autocorrelation is rejected at the 5% level of significance if the value of AQ

is greater than 3.84. The AQ test results cannot reject the null hypothesis of no return autocorrelation. The Dominguez-Lobato test results have p-values > 0.05. This means that the null that the residuals are a Martingale Difference Sequence (efficiency) cannot be rejected. Similarly, the GS test returns a p-value = 0.25, also indicating that the null that the residuals follow a random walk cannot be rejected. This, in addition to a visual inspection of the ACF plot of the residuals indicates that the model can be viewed as a good representation of the return generating process.

Table 4.51: Martingale Difference Hypothesis Tests – Weekly Residuals

Test	AQ Test	DL Test – Cp	DL Test – Kp	GS Test
t-statistic	0.27	0.16	0.91	0.55
p-value	0.60	0.6	0.6	NA

Note: * Significant at the 10% level, ** Significant at the 5% level, *** Significant at the 1% level.

In line with the findings in Seetharam (2016), the NARX models performed better from a residuals analysis perspective than their associated NAR models. In turn, the NAR models performed better than either the SETAR or ARIMA models prior to them. While Seetharam (2016) did not include a Fourier Transformation of the data, the outcome of the results of this study indicates that by order of model fit, the following ranking holds: NARX > NAR > SETAR/ARIMA. In the case of this research report, the rankings are robust to the data transformation. Furthermore, the results of this research report are consistent with Seetharam (2016) with regards to the SETAR model's inability to model the return generating process.

4.10 Summary

In testing out the methodology laid out in Seetharam (2016), several results have emerged. To test whether the returns follow a random walk/martingale, several Martingale Difference Hypothesis tests are employed. Their results indicated, regardless of transformation, that Bitcoin returns do not follow a martingale. An attempt is thereafter made to model the return, the model (whether a linear or non-linear model) is decided upon based on the BDS test. If the results of the BDS test are inconclusive, a SETAR test is used as an extra data point to decide on the choice of model. In the overwhelming majority of cases, however, the BDS test results were conclusive. Many cases, given the non-linearity of the returns, resolved in favour of a non-linear model. In all cases, the formal residual analysis

called for further modelling, which opened the door to machine learning approaches, which can capture cyclical efficiency.

The results indicated that while the LSTM model was a good fit for the data, it was nevertheless outperformed by the NARX model. In this respect, how good a model was at modelling the data is dependent on the frequency of the data. On a daily frequency, there seemed to be more noise, thus resulting in more autocorrelation in the models' residuals. The weekly frequency models that exhibited less noise in the residuals are the models overall better fit the data than their daily frequency counterparts.

Interesting patterns emerged, the Fourier Transformed data resulted in a more noticeable pattern of bunching in their residuals than their non-transformed counterparts. The transformation also led to overall better validation performance, with longer lasting and more noticeable declines in loss metrics, indicating that the models derived from non-transformed samples may be overfitting and not learning, thus leading to worse generalization results.

5 CONCLUSION

This study touches two hotly contested topics in the finance literature: Bitcoin and asset return efficiency. Chapter 1 introduces and provides background regarding Bitcoin. Bitcoin suffers from employing several technical concepts such as hash keys and hot/cold wallets and being shrouded in an air of mystique. Moreover, it has been shrouded for much of its lifetime in secrecy. For instance, it is unclear who its founder, Satoshi Nakamoto, really is. Its infamy has also not helped matters: a new wave of international crime syndicates has used it to escape know-your-customer (KYC) laws and launder money around the world. In addition, many famous "ransomware" hacks where the hackers demanded payment in Bitcoin have also firmly fixed a negative spotlight on its anonymity features.

It is thus unsurprising that some countries, mostly led by China, have enacted or mulled nationwide bans on not just cryptocurrency and decentralized financial institutions that deal in them. Similarly, many captains of the banking industry have voiced largely negative opinions regarding Bitcoin. For example, Buckham *et al.* (2021) argue that while the growing trust deficit in the mainstream monetary system is somewhat warranted, the growth in cryptocurrencies and decentralized financial institutions is a worrying trend.

In the same controversial vein, efficiency has become a contentious topic. The literature review covered in Chapter 2 goes over the spirited arguments over it. Proponents of classical economics and the EMH have lobbed several criticisms at behavioural finance in general and the AMH specifically. Chief among these is the commonly held philosophy of science adage that it "takes a theory to beat a theory". (Farber, 2001) Thus, by extension, it has often then argued that given that the EMH is the one-eyed efficiency hypothesis in the land of the blind, it thus deserves to be the king of the hill. It is thus tempting to see the case being brought forth by the Adaptive Markets Hypothesis as being an attempt at usurpation. What it is, however (and this study goes to great pains to point out,) in Seetharam (2016)'s words, is "a more encompassing hypothesis" that better explains efficiency with the added benefit of being falsifiable. Indeed, it would be remiss to see this study as an attempt to dethrone the EMH.

The success of this hypothesis does not spell the end for the Efficient Markets Hypothesis. Instead, it should be seen as part of an extension of a rich history stretching back from the earliest works in

describing market dynamics. It aims to create a systematic approach where the question of whether an asset is efficient is not purely down to whether its returns follow a random walk, but only the first of a four-part enquiry, covered in Chapter 3, that goes thus:

1) Do the returns follow a random walk? If they do not, that means that the asset in question is not weak-form efficient. This therefore means can attempt to model the return generating process.
2) This thus opens the door for the use formal statistical methods to decide whether to use an ARIMA model (if the returns are only linearly dependent) or a regime-switching model (if the returns are also non-linearly dependent). Once the model is assembled, examine its residuals: is there significant autocorrelation in the residuals? If yes, continue to model the return generating process without pre-specifying its functional form. If not, discontinue the enquiry.
3) Input the returns series into a machine learning model such as a NAR model without external regressors. If this model fits, that means that all public information is reflected in the return (this is an analogue to semi-strong form efficiency). If there is significant autocorrelation in the residuals, then proceed to modelling with external regressors.
4) Input the return series into a machine learning model with external regressors, such as a NARX model. If this model fits, that means that public and private information is reflected in the return (which fits with the strong-form efficiency).

Chapter 4 provides an analysis of the results of the study. Both daily and weekly returns data was found not to follow a random walk. Following this finding, the BDS and SETAR tests are applied to the returns data to determine its linearity, and therefore whether an ARIMA or SETAR test was suitable for modelling the data. A Fast Fourier Transformation was also applied to the data to "clean" it. Once it was clear in both the untransformed and transformed cases that ARIMA and SETAR approaches were not suitable for modelling the return generating process, the data was fed to a machine learning algorithm to model the return generating process without pre-specifying its functional form. While this approach marginally performed better than the traditional statistical approach, autocorrelation in its residuals was nevertheless significant (more so for the daily than weekly data). This meant that the modelling could continue to a machine learning approach that also includes external regressors.

At each step, residuals of the models were subjected to a battery of tests to ensure that their analysis is robust. While the AQ test cannot pick up non-linear dependency in the data, Generalized Spectral

and Dominguez-Lobato tests can. By employing this medley of tests, whether the model residuals exhibit dependency (linear or otherwise) can be answered satisfactorily. This ensures that modelling does not continue where the extant model is a sufficient representative of the return generating process.

The acid test for whether Bitcoin exhibits what this study terms "adaptive" efficiency is whether the machine learning approaches, which capture time-varying efficiency, are accepted as good candidate models for the return generating process. This is akin to a rejection of the secondary hypotheses, $H_{0,C}$ and $H_{0,B}$: Bitcoin returns cannot be modelled via an autoregressive function (with exogenous inputs). The varying levels of machine learning prediction success over time are an indication that the market is adaptively efficient. Moreover, if the prediction does of the machine learning algorithm does not vary over time, then Seetharam (2016) argues that this indicates that the assimilation of new/private information into asset returns is not instantaneous. In our case, both conditions hold: the NARX model's ability to forecast changes over time, as evidenced by varied levels of residual levels over time and the machine learning algorithm is accepted as a good model of the return generating process. This therefore is evidence that Bitcoin is adaptively efficient in the manner envisaged by the AMH.

5.1 Limitations

The use of Support Vector Machines (SVMs) and neural networks in research has presented some pressing problems; post ingesting large input data sets, they in turn require a large length of training time before producing outputs. An excessively large training set will also result in overfitting and will therefore have to be guarded against. Like neural networks, SVMs have also been marked out as "black boxes" — they are hard to interpret, this is especially important given that the interpretation would have enabled a "sense-check" of the weights and network topology.

5.2 Areas of future research

Future research regarding the use of machine learning approaches to forecast asset returns presents opportunities for the use of genetic and other algorithms for choosing the hyperparameters/number of layers of the machine learning models instead of them being chosen by the researcher. Similarly, given the single-asset nature of this study, comparative studies that span multiple currencies could

also enrich the literature on the Adaptive Markets Hypothesis using the methodology derived in Seetharam (2016).

REFERENCES

Adadi, A., & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, *6*, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

Aggarwal, D., Chandrasekaran, S., & Annamalai, B. (2020). A complete empirical ensemble mode decomposition and support vector machine-based approach to predict Bitcoin prices. *Journal of Behavioral and Experimental Finance*, *27*, 100335. https://doi.org/10.1016/j.jbef.2020.100335

Alrumaih, R. M., & Al-Fawzan, M. A. (2002). Time Series Forecasting Using Wavelet Denoising an Application to Saudi Stock Index. *Journal of King Saud University - Engineering Sciences*, *14*(2), 221–233. https://doi.org/10.1016/S1018-3639(18)30755-4

Alvarez-Ramirez, J., Rodriguez, E., & Alvarez, J. (2012). A multiscale entropy approach for market efficiency. *International Review of Financial Analysis*, *21*, 64–69.

Alvarez-Ramirez, J., Rodriguez, E., & Ibarra-Valdez, C. (2018). Long-range correlations and asymmetry in the Bitcoin market. *Physica A: Statistical Mechanics and Its Applications*, *492*, 948–955. https://doi.org/10.1016/j.physa.2017.11.025

Annis, A. A., & Lloyd, E. H. (1976). The expected value of the adjusted rescaled Hurst range of independent normal summands. *Biometrika*, *63*(1), 111–116.

Arouri, M. E. H., Hammoudeh, S., Lahiani, A., & Nguyen, D. K. (2012). Long memory and structural breaks in modeling the return and volatility dynamics of precious metals. *The Quarterly Review of Economics and Finance*, *52*(2), 207–218. https://doi.org/10.1016/j.qref.2012.04.004

Bahad, P., Saxena, P., & Kamal, R. (2019). Fake News Detection using Bi-directional LSTM-Recurrent Neural Network. *Procedia Computer Science*, *165*, 74–82. https://doi.org/10.1016/j.procs.2020.01.072

Bariviera, A. F. (2017). The inefficiency of Bitcoin revisited: A dynamic approach. *Economics Letters*, *161*, 1–4. https://doi.org/10.1016/j.econlet.2017.09.013

Bariviera, A. F., Basgall, M. J., Hasperué, W., & Naiouf, M. (2017). Some stylized facts of the Bitcoin market. *Physica A: Statistical Mechanics and Its Applications*, *484*, 82–90. https://doi.org/10.1016/j.physa.2017.04.159

Baum, C., Hurn, S., & Linday, K. (2021). The BDS test of independence. *The Stata Journal*, *21*(2), 279–294. https://doi.org/10.1177/1536867X211025796

Böhme, R., Christin, N., Edelman, B., & Moore, T. (2015). Bitcoin: Economics, Technology, and Governance. *Journal of Economic Perspectives*, *29*(2), 213–238. https://doi.org/10.1257/jep.29.2.213

Brauneis, A., & Mestel, R. (2018). Price discovery of cryptocurrencies: Bitcoin and beyond. *Economics Letters*, *165*, 58–61. https://doi.org/10.1016/j.econlet.2018.02.001

Browning, K. (2021, May 13). Coinbase made $771 million in profit in the first quarter, benefiting from crypto mania. *The New York Times*. Retrieved from https://www.nytimes.com/2021/05/13/technology/coinbase-profit.html

Buckham, D., Robyn, W., & Christiaan, S. (2021). *The End Of Money: The Great Erosion of Trust in Banking, China's Minsky Moment and the Fallacy of Cryptocurrency*. Mercury.

Cadenas, E., Rivera, W., Campos-Amezcua, R., & Cadenas, R. (2016). Wind speed forecasting using the NARX model, case: La Mata, Oaxaca, México. *Neural Computing and Applications*, *27*(8), 2417–2428. https://doi.org/10.1007/s00521-015-2012-y

Cao, J., Li, Z., & Li, J. (2019). Financial time series forecasting model based on CEEMDAN and LSTM. *Physica A: Statistical Mechanics and Its Applications*, *519*, 127–139. https://doi.org/10.1016/j.physa.2018.11.061

Cao, L., & Tay, F. E. H. (2001). Financial Forecasting Using Support Vector Machines. *Neural Computing & Applications*, *10*(2), 184–192. https://doi.org/10.1007/s005210170010

Couillard, M., & Davison, M. (2005). A comment on measuring the Hurst exponent of financial time series. *Physica A: Statistical Mechanics and Its Applications*, *348*, 404–418. https://doi.org/10.1016/j.physa.2004.09.035

Crabb, J. (2021). Opinion: Coinbase IPO will be a turning point. *International Financial Law Review*. Retrieved from https://www.proquest.com/docview/2511011385/abstract/CBA92C06E49749BBPQ/1

Crawford, G. W., & Fratantoni, M. C. (2003). Assessing the Forecasting Performance of Regime-Switching, ARIMA and GARCH Models of House Prices. *Real Estate Economics*, *31*(2), 223–243. https://doi.org/10.1111/1540-6229.00064

Dastgir, S., Demir, E., Downing, G., Gozgor, G., & Lau, C. K. M. (2019). The causal relationship between Bitcoin attention and Bitcoin returns: Evidence from the Copula-based Granger causality test. *Finance Research Letters*, *28*, 160–164. https://doi.org/10.1016/j.frl.2018.04.019

Davies, R. B., & Harte, D. S. (1987). Tests for Hurst effect. *Biometrika*, *74*(1), 95–101.

Demir, E., Gozgor, G., Lau, C. K. M., & Vigne, S. A. (2018). Does economic policy uncertainty predict the Bitcoin returns? An empirical investigation. *Finance Research Letters*, *26*, 145–149. https://doi.org/10.1016/j.frl.2018.01.005

Ding, G., & Qin, L. (2020). Study on the prediction of stock price based on the associated network model of LSTM. *International Journal of Machine Learning and Cybernetics*, *11*(6), 1307–1317. https://doi.org/10.1007/s13042-019-01041-1

Dominguez M.A., & Lobato, I. N. (2003). Testing the Martingale Difference Hypothesis. *Econometrics Reviews*, 22 (4), 351-377.

Escanciano, J. C., & Velasco, C. (2006). Generalized spectral tests for the martingale difference hypothesis. *Journal of Econometrics*, *134*(1), 151–185. https://doi.org/10.1016/j.jeconom.2005.06.019

Escanciano, J. C., & Lobato, I. N. (2009). An automatic portmanteau test for serial correlation. *Journal of Econometrics*, *151*(2), 140-149.

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, *25*(2), 383–417. https://doi.org/10.2307/2325486

Farber, D. A. (2001). Toward a New Legal Realism Reviews. *University of Chicago Law Review*, *68*(1), 279–304.

Feng, H., & Liu, J. (2003). A SETAR model for Canadian GDP: Non-linearities and forecast comparisons. *Applied Economics*, *35*(18), 1957–1964. https://doi.org/10.1080/0003684032000160674

Frennberg, P., & Hansson, B. (1993). Testing the random walk hypothesis on Swedish stock prices: 1919–1990. *Journal of Banking & Finance*, *17*(1), 175–191. https://doi.org/10.1016/0378-4266(93)90087-T

Grater, E., & Struweg, J. (2015). Testing weak form efficiency in the South African market. *Journal of Economic and Financial Sciences*, *8*(2), 621–632. https://doi.org/10.10520/EJC177406

Grossman, S. J., & Stiglitz, J. E. (1976). Information and Competitive Price Systems. *The American Economic Review*, *66*(2), 246–253.

Haug, M., & Hirschey, M. (2006). The January Effect. *Financial Analysts Journal*, *62*(5), 78–88. https://doi.org/10.2469/faj.v62.n5.4284

Heymans, A., & Santana, L. (2018). How efficient is the Johannesburg Stock Exchange really? *South African Journal of Economic and Management Sciences*, *21*(1), 1–14. https://doi.org/10.4102/sajems.v21i1.1968

Hochreiter, S. (1998). The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions | International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *6*(2), 107–116.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.

Inman, P. (2020, November 17). Bitcoin jumps to three-year high as Covid crisis changes investor outlook. *The Guardian*. Retrieved from

http://www.theguardian.com/technology/2020/nov/17/bitcoin-jumps-to-three-year-high-as-covid-crisis-changes-investor-outlook

Ito, M., Noda, A., & Wada, T. (2014). International stock market efficiency: A non-Bayesian time-varying model approach. *Applied Economics*, *46*(23), 2744–2754. https://doi.org/10.1080/00036846.2014.909579

Jansen van Rensburg, J., & van Vuuren, G. (2020). Evaluating investment decisions based on the business cycle: A South African sector approach. *Cogent Economics & Finance*, *8*(1), 1852729. https://doi.org/10.1080/23322039.2020.1852729

Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, *6*(3), 255–259. https://doi.org/10.1016/0165-1765(80)90024-5

Jung, J., & Shiller, R. J. (2005). Samuelson's Dictum and the Stock Market. *Economic Inquiry*, *43*(2), 221–228. https://doi.org/10.1093/ei/cbi015

Karl, P. (2013). Logik der Forschung. In *Karl Popper: Logik der Forschung*. Akademie Verlag. Retrieved from https://www.degruyter.com/document/doi/10.1524/9783050063782/html

Khandani, A. E., & Lo, A. W. (2007). *What Happened To The Quants In August 2007?* 58.

Khursheed, A., Naeem, M., Ahmed, S., & Mustafa, F. (2020). Adaptive market hypothesis: An empirical analysis of time –varying market efficiency of cryptocurrencies. *Cogent Economics & Finance*, *8*(1), 1719574. https://doi.org/10.1080/23322039.2020.1719574

Kim, K. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, *55*(1), 307–319. https://doi.org/10.1016/S0925-2312(03)00372-2

Kristoufek, L. (2018a). On Bitcoin markets (in)efficiency and its evolution. *Physica A: Statistical Mechanics and Its Applications*, *503*, 257–262. https://doi.org/10.1016/j.physa.2018.02.161

Kristoufek, L. (2018b). On Bitcoin markets (in)efficiency and its evolution. *Physica A: Statistical Mechanics and Its Applications*, *503*, 257–262. https://doi.org/10.1016/j.physa.2018.02.161

Kristoufek, L., & Vosvrda, M. (2013). Measuring capital market efficiency: Global and local correlations structure. *Physica A: Statistical Mechanics and Its Applications*, *392*(1), 184–193. https://doi.org/10.1016/j.physa.2012.08.003

Kristoufek, L., & Vosvrda, M. (2014). Commodity futures and market efficiency. *Energy Economics*, *42*, 50–57. https://doi.org/10.1016/j.eneco.2013.12.001

Kristoufek, L., & Vosvrda, M. (2019). Cryptocurrencies market efficiency ranking: Not so straightforward. *Physica A: Statistical Mechanics and Its Applications*, *531*, 120853. https://doi.org/10.1016/j.physa.2019.04.089

Lawrence, E. R., McCabe, G., & Prakash, A. J. (2007). Answering Financial Anomalies: Sentiment-Based Stock Pricing. *Journal of Behavioral Finance*, *8*(3), 161–171. https://doi.org/10.1080/15427560701547248

Li, Youru, Zhu, Z., Kong, D., Han, H., & Zhao, Y. (2019). EA-LSTM: Evolutionary attention-based LSTM for time series prediction. *Knowledge-Based Systems*, *181*, 104785. https://doi.org/10.1016/j.knosys.2019.05.028

Li, Yue, Goodell, J. W., & Shen, D. (2021). Comparing search-engine and social-media attentions in finance research: Evidence from cryptocurrencies. *International Review of Economics & Finance*, *75*, 723–746. https://doi.org/10.1016/j.iref.2021.05.003

Lo, A. W. (2004). The Adaptive Markets Hypothesis. *The Journal of Portfolio Management*, *30*(5), 15–29. https://doi.org/10.3905/jpm.2004.442611

Lucas, R. (1978). Asset prices in an exchange economy. *Econometrica*, *46*, 1429–1446.

Makridakis, S., & Hibon, M. (1997). ARMA Models and the Box–Jenkins Methodology. *Journal of Forecasting*, *16*(3), 147–163. https://doi.org/10.1002/(SICI)1099-131X(199705)16:3<147::AID-FOR652>3.0.CO;2-X

Malkiel, B. G. (2003). The Efficient Market Hypothesis and Its Critics. *Journal of Economic Perspectives*, *17*(1), 59–82.

Maxwell, W. F. (1998). The January Effect in the Corporate Bond Market: A Systematic Examination. *Financial Management*, *27*(2), 18–30. https://doi.org/10.2307/3666290

McNally, S. (2016). *Predicting the price of Bitcoin using Machine Learning* (National College of Ireland, Dublin). National College of Ireland, Dublin. Retrieved from http://norma.ncirl.ie/2496/

Monaghan, A. (2017, November 30). Bitcoin is a vehicle for fraudsters, warns Goldman Sachs boss. *The Guardian*. Retrieved from http://www.theguardian.com/business/2017/nov/30/bitcoin-is-a-vehicle-for-fraudsters-warns-goldman-sachs-boss

Nadarajah, S., & Chu, J. (2017). On the inefficiency of Bitcoin. *Economics Letters*, *150*, 6–9. https://doi.org/10.1016/j.econlet.2016.10.033

Noda, A. (2016). A Test of the Adaptive Market Hypothesis using a Time-Varying AR Model in Japan. *Finance Research Letters*, *17*, 66–71. https://doi.org/10.1016/j.frl.2016.01.004

Noda, A. (2021). On the evolution of cryptocurrency market efficiency. *Applied Economics Letters*, *28*(6), 433–439. https://doi.org/10.1080/13504851.2020.1758617

Panagiotidis, T., Stengos, T., & Vravosinos, O. (2019). The effects of markets, uncertainty and search intensity on bitcoin returns. *International Review of Financial Analysis*, *63*, 220–242. https://doi.org/10.1016/j.irfa.2018.11.002

Peters, E. E. (1994). *Fractal Market Analysis: Applying Chaos Theory to Investment and Economics*. John Wiley & Sons.

Pisoni, E., Farina, M., Carnevale, C., & Piroddi, L. (2009). Forecasting peak air pollution levels using NARX models. *Engineering Applications of Artificial Intelligence*, *22*(4), 593–602. https://doi.org/10.1016/j.engappai.2009.04.002

Rose, C. (2015). The Evolution Of Digital Currencies: Bitcoin, A Cryptocurrency Causing A Monetary Revolution. *International Business & Economics Research Journal (IBER)*, *14*(4), 617–622. https://doi.org/10.19030/iber.v14i4.9353

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

Salisu, A. A., Oloko, T. F., & Oyewole, O. J. (2016). Testing for martingale difference hypothesis with structural breaks: Evidence from Asia–Pacific foreign exchange markets. *Borsa Istanbul Review*, *16*(4), 210–218. https://doi.org/10.1016/j.bir.2016.09.001

Sandroni, A. (2005). Efficient Markets and Bayes' Rule. *Economic Theory*, *26*(4), 741–764.

Seetharam, Y. (2016). *The dynamics of market efficiency: Testing the adaptive market hypothesis in South Africa* (Thesis). Retrieved from http://wiredspace.wits.ac.za/handle/10539/21982

Seetharam, Y., & Britten, J. (2015). Non-linear modelling of market cycles in South Africa. *International Journal of Emerging Markets*, *10*, 670–683. https://doi.org/10.1108/IJoEM-05-2013-0079

Selgin, G. (2015). Synthetic commodity money. *Journal of Financial Stability*, *17*, 92–99. https://doi.org/10.1016/j.jfs.2014.07.002

Shahzad, S. J. H., Bouri, E., Roubaud, D., Kristoufek, L., & Lucey, B. (2019). Is Bitcoin a better safe-haven investment than gold and commodities? *International Review of Financial Analysis*, *63*, 322–330. https://doi.org/10.1016/j.irfa.2019.01.002

Shapiro, S. S., & Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, *52*(3/4), 591–611. https://doi.org/10.2307/2333709

Shen, Y., Hung, C., Chiou, J., & Shen, K. (2020). The January Effect and Prospect Theory in Taiwan. *Emerging Markets Finance and Trade*, *56*(5), 1113–1123. https://doi.org/10.1080/1540496X.2019.1598367

Shiller, R. J. (1981). Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends? *The American Economic Review*, *71*(3), 421–436.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*(2), 129–138.

Thadewald, T., & Büning, H. (2007). Jarque–Bera Test and its Competitors for Testing Normality – A Power Comparison. *Journal of Applied Statistics*, *34*(1), 87–105. https://doi.org/10.1080/02664760600994539

Thaler, R. H. (1987). Anomalies: The January Effect. *Journal of Economic Perspectives*, *1*(1), 197–201. https://doi.org/10.1257/jep.1.1.197

Thaler, R. H. (1999). The End of Behavioral Finance. *Financial Analysts Journal*, *55*(6), 12–17. https://doi.org/10.2469/faj.v55.n6.2310

Tschoegl, A. E. (1980). Efficiency in the gold market—A note. *Journal of Banking & Finance*, *4*(4), 371–379. https://doi.org/10.1016/0378-4266(80)90015-1

Tversky, A., & Kahneman, D. (1991). Loss Aversion in Riskless Choice: A Reference-Dependent Model*. *The Quarterly Journal of Economics*, *106*(4), 1039–1061. https://doi.org/10.2307/2937956

Urquhart, A. (2016). The inefficiency of Bitcoin. *Economics Letters*, *148*, 80–82. https://doi.org/10.1016/j.econlet.2016.09.019

Vo, A., Chapman, T. A., & Lee, Y.-S. (2021). Examining Bitcoin and Economic Determinants: An Evolutionary Perspective. *Journal of Computer Information Systems*, *0*(0), 1–15. https://doi.org/10.1080/08874417.2020.1865851

Weinberger, K. Q., & Saul, L. K. (2009). *Distance Metric Learning for Large Margin Nearest Neighbor Classification*. *10*, 207–244.

Wooldridge, J. M. (2012). Introductory econometrics: a modern approach. *Southwestern College Publishing, Nashville, T ATN*, *41*, 673-690.

Appendix A: Further Bitcoin Analysis

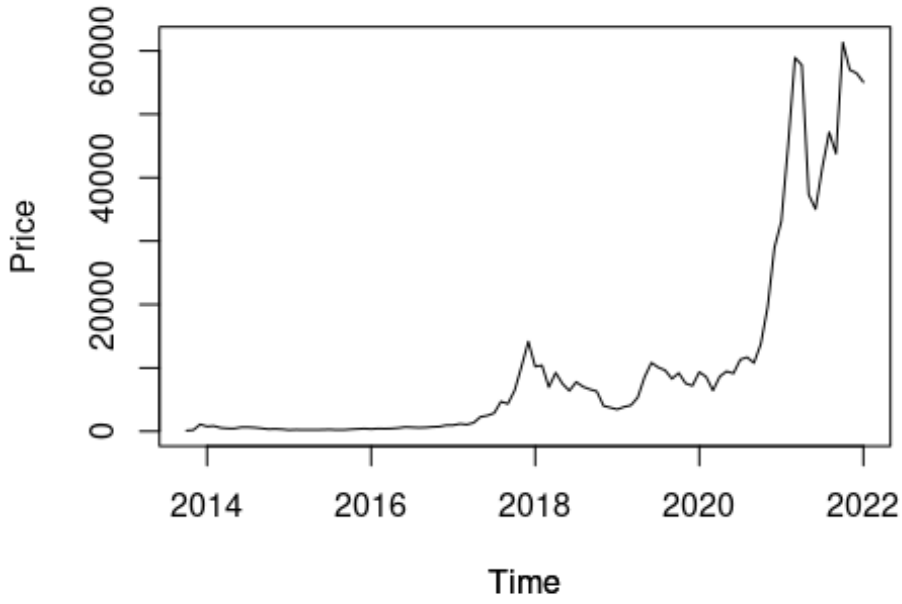


Figure 0.1: Monthly Bitcoin Price

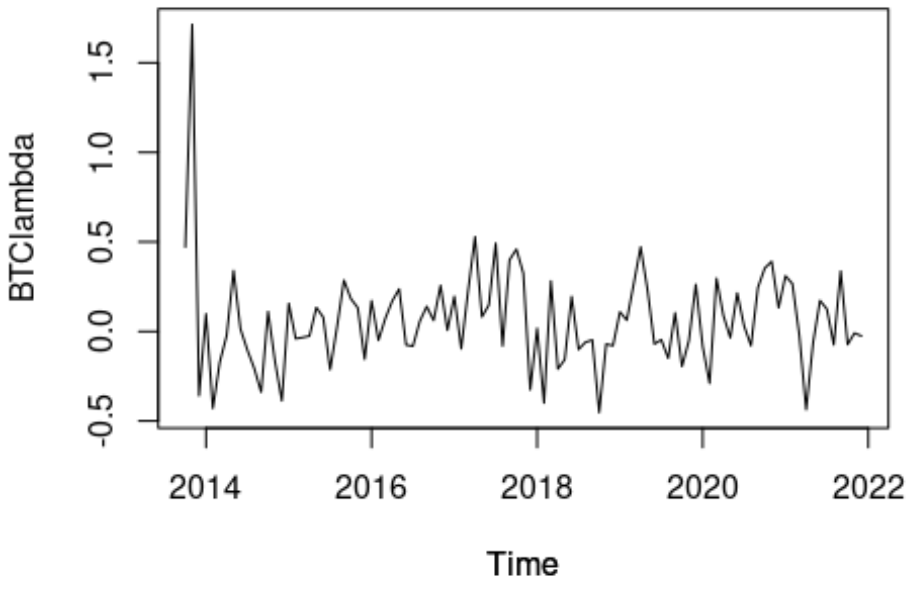


Figure 0.2: Monthly Bitcoin Return

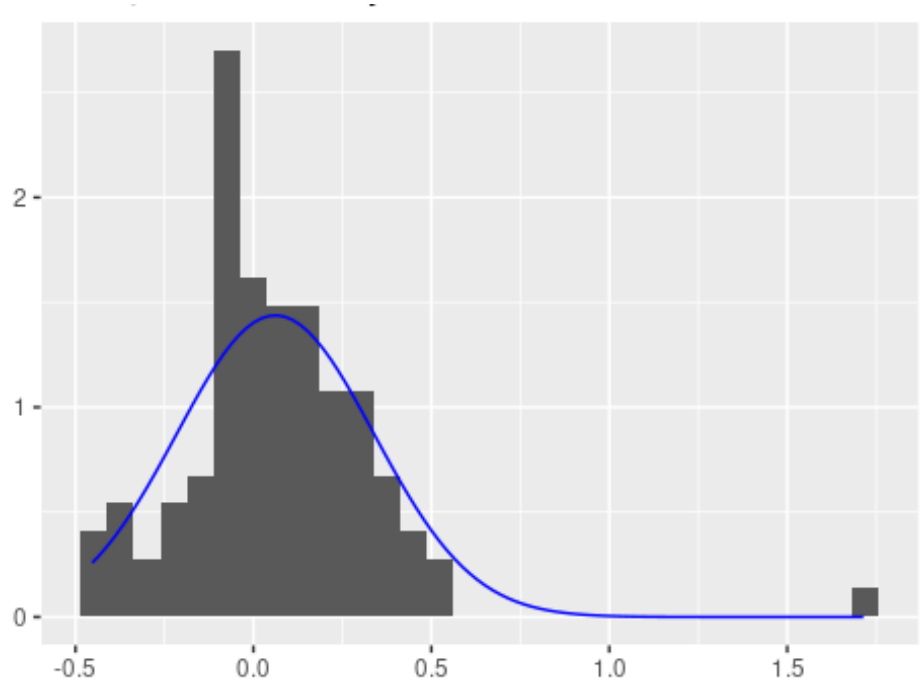
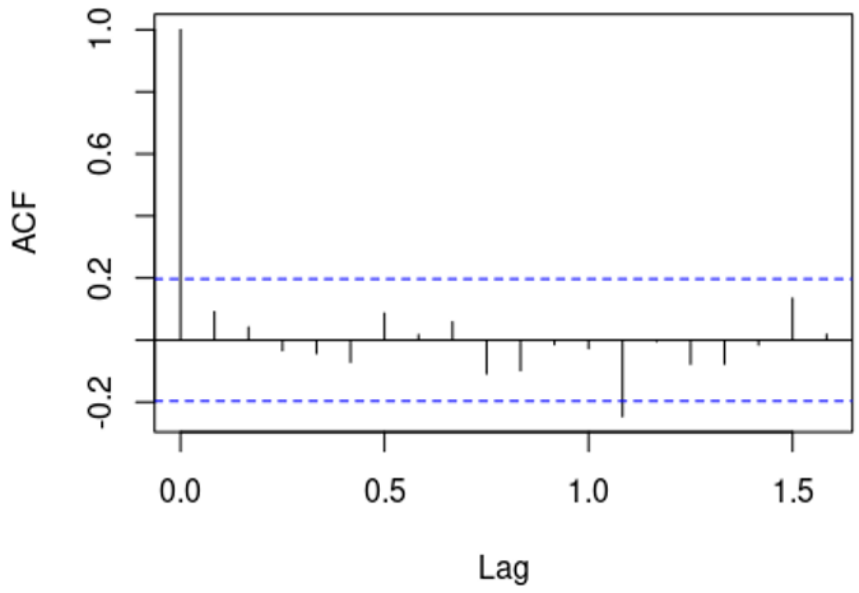


Figure 0.3: Monthly Bitcoin Returns Histogram



Note: The blue bands indicate the line of significant, at lags where it is crossed, it can be inferred that there is significant autocorrelation in the time series under investigation.

Figure 0.4: Monthly Bitcoin Returns ACF

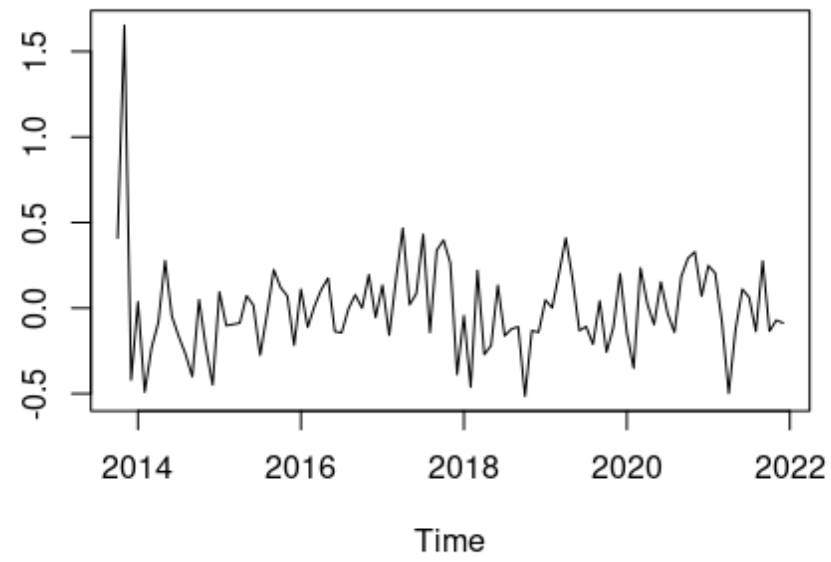


Figure 0.5: Monthly ARIMA Model Residuals

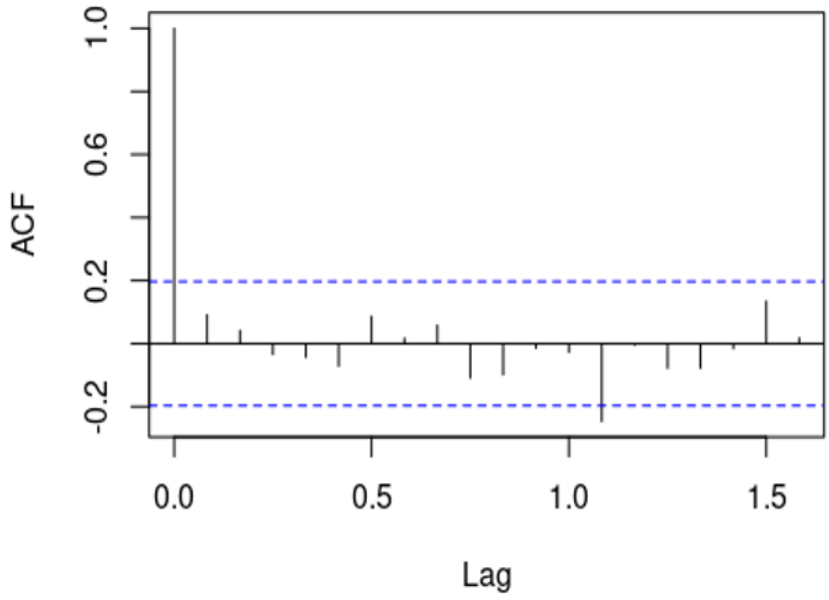


Figure 0.6: Monthly ARIMA Model Residual ACF

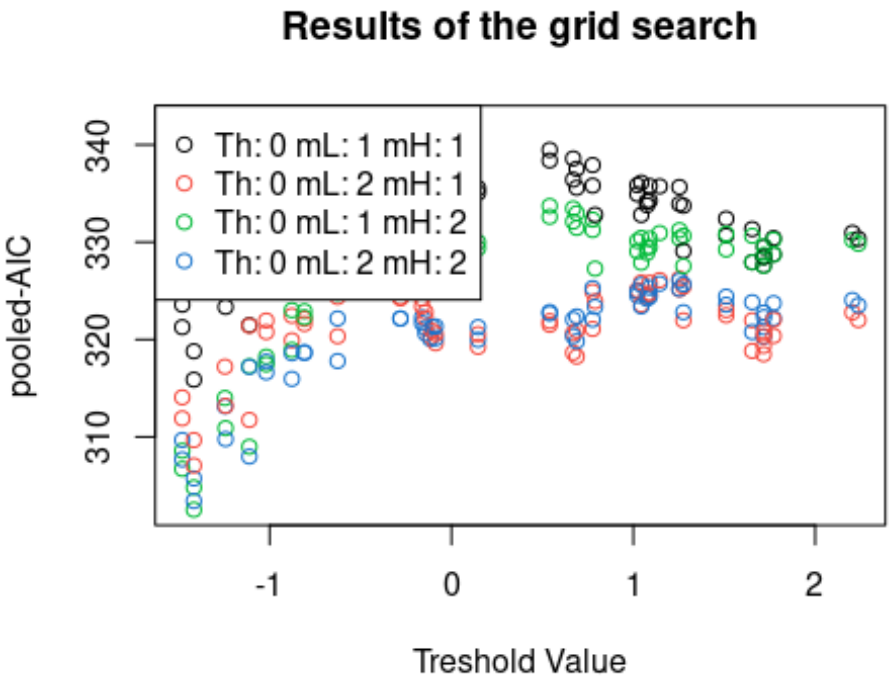


Figure 0.7: Fourier Transformed Monthly SETAR Grid Search

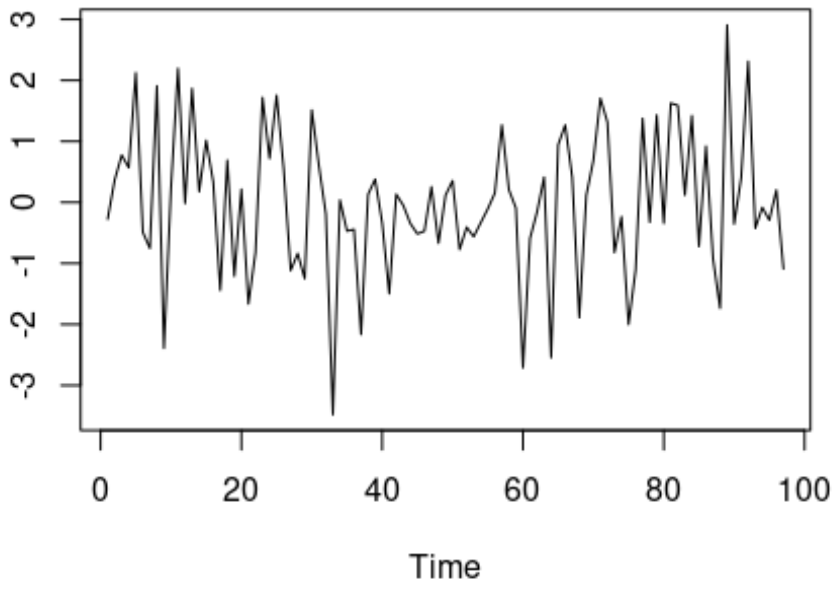
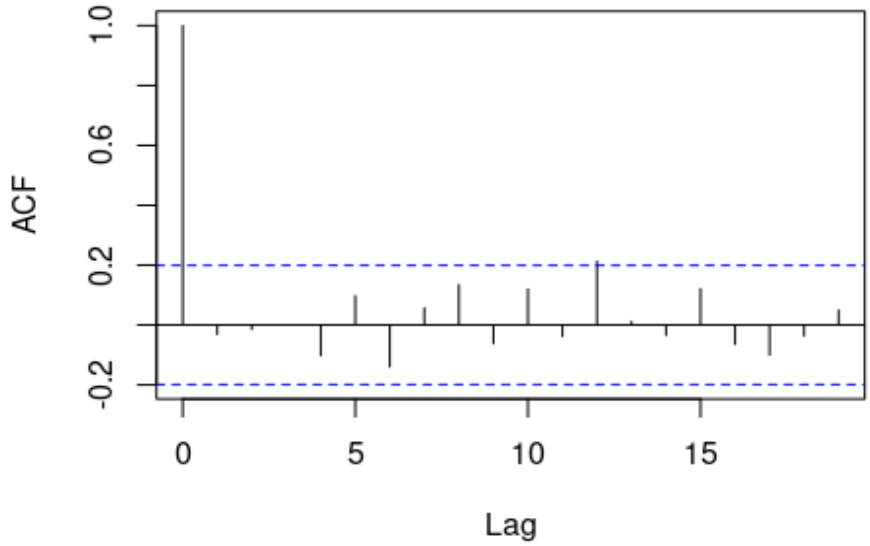


Figure 0.8: Fourier Transformed SETAR Residuals



Note: The blue bands indicate the line of significant, at lags where it is crossed, it can be inferred that there is significant autocorrelation in the time series under investigation.

Figure 0.9: Fourier Transformed SETAR Residuals ACF

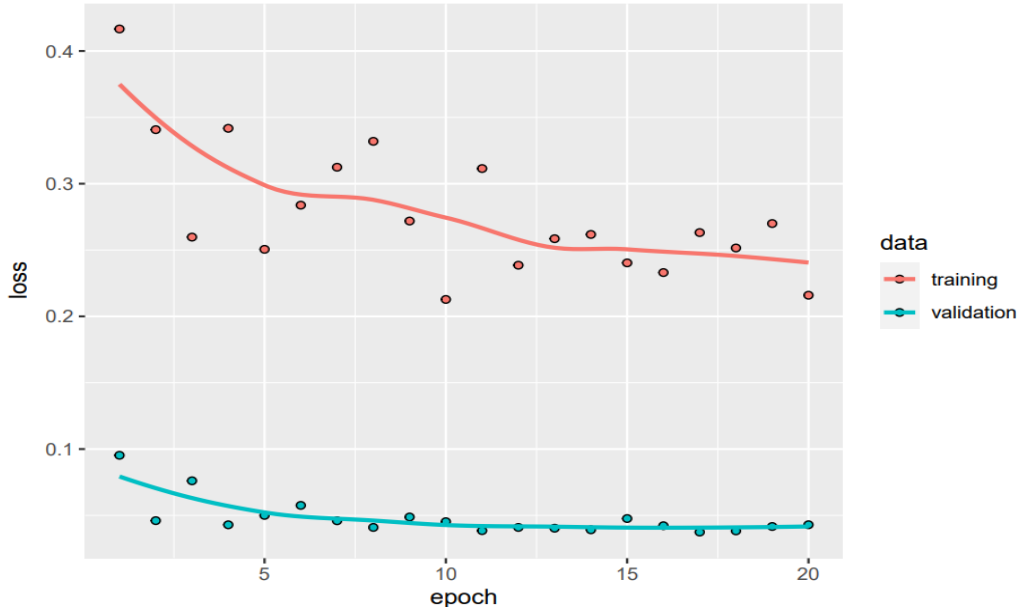


Figure 0.10: Monthly LSTM Diagnostics

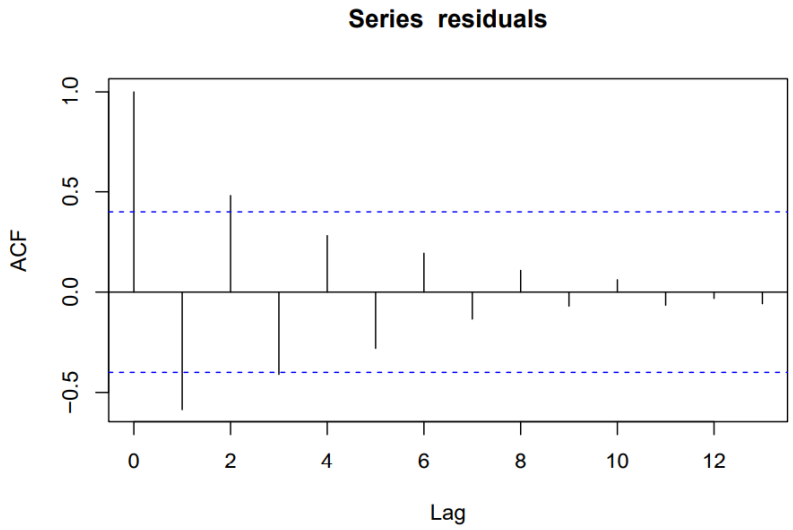


Figure 0.11: Monthly LSTM Residuals ACF

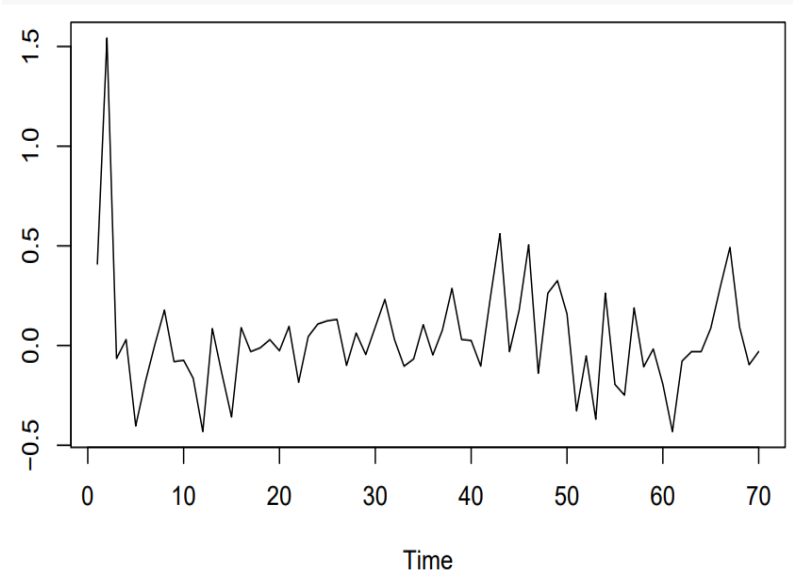


Figure 0.12: Monthly NARX Model Residuals

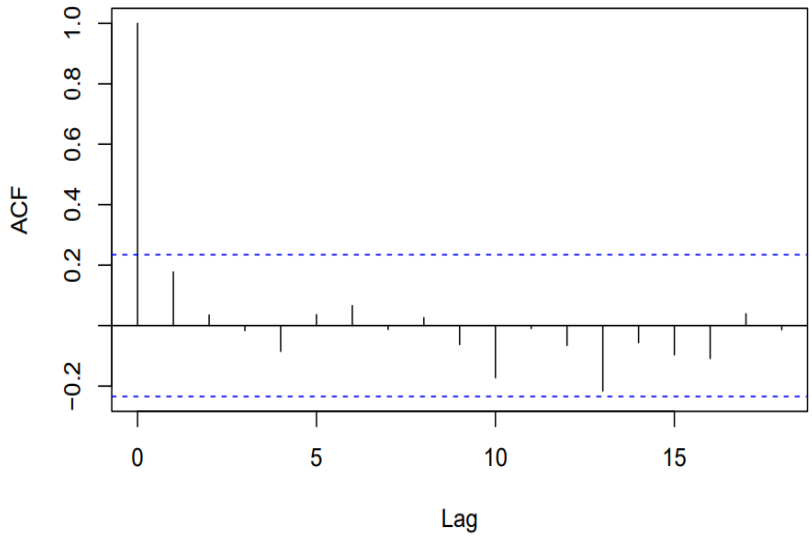


Figure 0.13: Monthly NARX Residuals ACF Plot

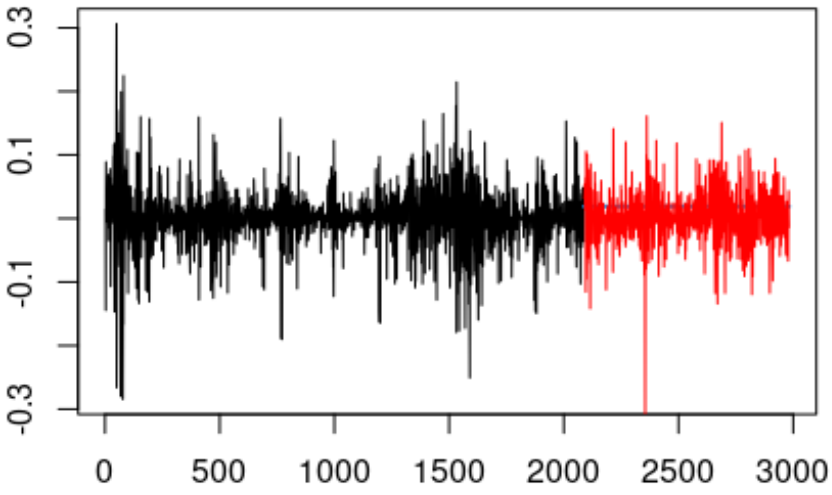


Figure 0.14: Daily NARX Forecast

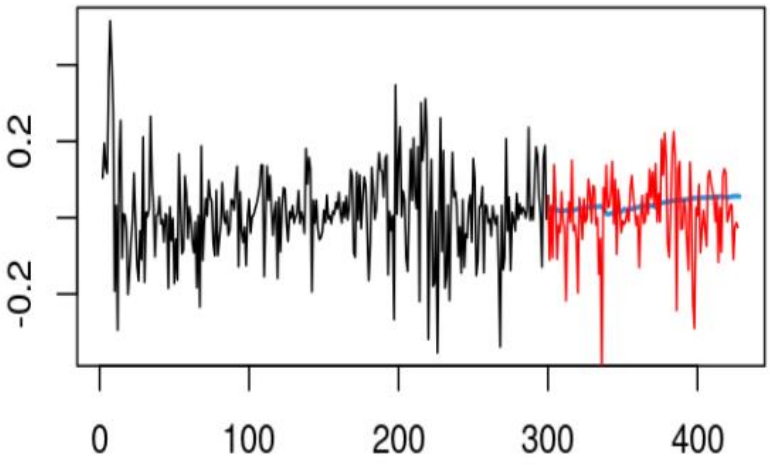


Figure 0.15: Weekly NARX Forecast