Forecasting the price of Bitcoin using neural networks

M Kuyler

orcid.org/0000-0002-1604-3166

Dissertation accepted in fulfilment of the requirements for the degree Master of Commerce in Risk Management at the North-West University

Supervisor: Prof A Heymans

Graduation ceremony: October 2019

Student number: 25066803

ABSTRACT

The evolution of financial technology adds to the complexity of the global financial system and the underlying assets that store its value. This complexity manifests as an adverse market risk profile in assets where fintech can be considered an endogenous variable. A theoretical framework that may contribute toward an improved understanding of this relationship is established. In contrast to the adverse risk profile in these markets, however, the literature still suggests a value proposition in these fintech-endogenous markets. The suggested value proposition is investigated by means of an empirical literature review, and partial recreation of some key findings from previous literature. Subsequently, additional empirical findings are contributed through a comparative set of tests in a controlled environment, with some significant results, specifically in the case where an appropriate trading strategy is back-tested along with some neural network forecasting procedures. The implications for researchers and practitioners are emphasised by a re-contextualisation of how the findings could affect future research in forecasting- and trading methodologies as well as the status quo of portfolio management strategies that risk managers have at their disposal. They key contribution is that risk managers should be able to benefit from the erratic behaviour of fintech-endogenous markets in the form of non-negligible short-term abnormal profit, whilst not having to trade off the diversification properties consistent with the established literature. The junction of forecasting- and trading methodologies used here may result in a “best of both worlds” investment strategy where abnormal profits are possible in the short run, in a simultaneously well-hedged trading environment, which relies on (instead of mitigating) the erratic price-formation phenomena prevalent in fintech-endogenous markets.

Keywords:
Bitcoin, cryptocurrencies, digital assets, forecast, Artificial Neural Networks

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

1.1 Overview and background

Financial technology (fintech) has become a market force that threatens to disrupt traditional financial systems and services, to a material extent, in the near future and beyond (Deloitte, 2018). It is possible to gain a better understanding of the reasonably probable magnitude of such disruption in the future, by considering the developments arising from fintech in the past. Fintech is, possibly, a truly fundamental driving force endogenous to economic activity over time, and since this technology is inherently inseparable from finance, fintech is merely a modern term used to describe relatively recent developments in international financial systems (Schueffel, 2015). As the development of financial technology through time is linked to the overarching problem statement, the investigation may rather become one of how modern instalments of fintech developments may be understood in a quantitative manner. One relatively recent technological development arising from the field of fintech is that of the digital currency, made possible by decentralised electronic ledgers imitating (more efficiently and in a digitised format) an idea that originates from double entry accounting.

Decentralised Ledger Technology (DLT) has attracted widespread attention from various sectors of the economy over recent years and, while DLT has been argued to have much broader applications, its most popular application has been in the creation of decentralised currency (Atsalakis, 2019). Such currencies have solved the well-known double spending problem through the creative use of cryptography (Nakamoto, 2008), which gave rise to their being named cryptocurrencies. By far the most popular implementation of the cryptocurrency is one called Bitcoin, whose origins can be traced back to the 3rd of January 2009 – the day that the Bitcoin network began its activity. The nature of the network and the way it secures value flows and anonymity gave rise to it being called a block chain, as it validates transactions in blocks that are linked to one another through their digital identities. The block chain implementation gained massive popularity from widespread sectors of society and holding ownership of an unspent integer of the fictional currency became a speculative asset to some members of the public and a limited number of institutions. The Bitcoin implementation's popularity and concurrent rise in value caused other implementations to imitate and, as of the 6th of May 2019, there exists 2148 implementations that rely on the same technology that Bitcoin pioneered. As of the same date, the collective of cryptocurrency implementations account for a market cap of $180,795,968,640 of which Bitcoin dominates more than 50% (Coin Market Cap, 2019). Let, however, the importance of these currencies be not over- nor understated. This research takes no position in terms of the future of these implementations other than that similar developments may happen again. In order to support this statement, a short overview of speculative asset bubbles is in order.

There exists considerable support in the literature to the position that Bitcoin and its alternatives may have been nothing more than a speculative bubble (Cheah, 2015). A historical perspective on this case compares Bitcoin with some of the most notorious speculative bubbles in the known modern history of such phenomena. Graphically, this has been represented upon multiple occasions, one of which is presented in Figure 1-1 below.

Figure 1-1: Line Graph Comparing Bitcoin –and NASDAQ Bubble.

(Source: Yahoo Finance, CoinMarketCap)

Similar to the comparison in the graph above, Bitcoin has been compared to other speculative asset bubbles such as the Tulip Mania (Benedetti & Kostovetsky, 2018), the Mississippi Bubble (Chancellor, 2018) and the South Sea Company Bubble (Tirole, 2017). The retrospective evidence makes the case that Bitcoin (at least at its most extreme) was indeed a speculative bubble easy to argue. However, this offers no solution in terms of appropriate action for risk managers given such an occurrence takes place once more in the future. The need for such strategies of appropriate action is even further exacerbated by the regulatory framework within which financial technology operates.

International regulatory activities with regard to fintech exist in various forms across jurisdictions, possibly quite surprisingly in some cases in support of fintech-related ventures. In the United States of America (USA), regulatory support for financial technology came in the form of the Jumpstart Our Business (JOBs) Act signed into law in 2012, allowing crowdfunding ventures to utilise technology toward financing business activities independently of established financial institutions (Stemler, 2013). Similar regulatory support for fintech innovation comes from Saudi Arabia in the form of the SADAD Payment System (SADAD) to aid in a government-regulated fintech system for Electronic Bill Presentment and Payment (EBPP) which should aid in private sector firms' efficiency in payments (Alsudairi & Vasista, 2012). The same theme of government-initiated regulatory support carries over to the South Korean retail banking sector in the form of easier retail banking license acquisition for private firms (Arner, 2016). Similar developments are seen in New Zealand and Australia with regulatory equity crowdfunding support and fintech innovation hubs respectively (Kshetri, 2015).

Much more comprehensive fintech advancement strategies are observed in Singapore (Fan, 2018), Hong Kong (Romanova & Kudinska, 2016) and China (Chen, 2016). These are mere examples of well-known regulatory activities enabling fintech innovation across disparate jurisdictions and the literature presents further evidence of accelerated growth programs in this domain. It is therefore difficult to imagine that these technology-driven processes are to slow down in the near future, and that is precisely why developments in finance arising from fintech may be contributing toward more dynamic and complex interrelationships in financial systems and the underlying assets that store their value.

Financial systems and the underlying assets that serve as value stores within these systems may be exhibiting more complex interrelationships through greater dynamism in their value fluctuations. This is supported by various internationally-influential factors arising from the past relationship between finance and technology (as is discussed at length in chapter 2). These issues of complexity are an established uncertainty in international affairs and, it has been linked to the literature on level two chaotic systems (Patrick & Julia, 2017). The international inter-relatedness of financial markets is also a phenomenon that has been found to increase with time (Cerny, 1994). The implication for risk managers is that more advanced forecasting techniques may provide a value proposition in their ability to aid in the analysis of interrelationships among sets of appropriate variables (Selmi *et al.,* 2018). The reason for this is that the systematic improvements in the machine-learning domain enable the involved techniques to deal with increasingly complex problems. Empirical evidence seems to suggest that some modern machine learning applications are able to learn complex functions between financial variables with significant success, and these findings warrant experimentation in a controlled testing environment.

The international cryptocurrency market therefore offers a unique research opportunity due to some interesting key characteristics. The first of which is the relatively low levels of regulation. The second and perhaps even more significant is the low levels of technical investment-expertise among market participants as illustrated by Lew and Mills (2013). The novelty and idealism behind cryptocurrencies drove their popularity among non-professional investors, and this may have led to erratic human mass psychological phenomena and interesting sentiment-rather than expertise driven price behaviour. The third interesting characteristic of the collective cryptocurrency market is its global accessibility. The technology removed virtually all barriers to international payments, at unprecedentedly low costs as is considered in detail by Bohme (2014). The concurrency of this global payment network may enable geographical factors and cultural inter-dynamics to exhibit price-behaviour patterns. These interesting characteristics of the cryptocurrency market, has caused professional investors, brokers and private investors to find that cryptocurrency could be an investment tool according to Lew and Mills (2013). In this regard, it is necessary to predict future values of cryptocurrency in order to make more accurate investment decisions. Traditional statistical methods such as linear regression have been a popular forecasting tool in the past. However, the fundamental differences between existing asset classes and digital assets like cryptocurrency bring perhaps more uncertainty in market trends. In forecasting problems with significant uncertainty, unknown distributions, lack of existing heuristics and fundamental novelties, computing techniques such as neural networks have become quite popular as Wilamowski (2009) explains. Artificial Neural Networks can adjust their parameters as new information is made available to them, which may give them the ability to capture non-linear trends in financial markets.

It is for this reason that this study hypothesises that more complex, quantitative techniques such as machine learning will provide a better fit for the more complex interrelationships that exist among financial variables. It is concluded then that financial technology has shown throughout historical literature to exercise a significant force upon finance in an iterative manner. This technological force upon finance carries with it the ability to add vastly to the complexity of financial systems and the underlying assets that serve the purpose of storing value. As there are also material signals from international regulatory bodies to improve regulatory support toward fintech development, there is little evidence to suggest that this process of financial technology is expected to slow down in the near future. These circumstances, along with the development of increasingly capable modelling techniques that rely on machine learning, warrant an investigation of machine learning-driven forecasting techniques in order to determine whether they may offer a solution to risk managers that could enable them to improve their quantitative understanding of complex modern developments arising from fintech.

1.2 Problem Statement

The evolution of financial technology correlates positively with the complexity of financial markets. This occurs in two forms, the first of which is the process of technological integration in traditional markets (in which case the effects can be explained relatively effectively as exogenous variables to the existing system). The second form is the rapid demand-driven development of new financial markets for non-traditional, inherently technological assets (in which case the effects manifest significantly more chaotically as they are endogenous to the system). The latter creates the problem/opportunity statement for this research, because even though fintech-associated markets exhibit significantly more adverse risk profiles than traditional asset markets, the empirical literature still suggests a value proposition in these markets both for traders and portfolio managers. This value proposition is firstly in the form of diversification and secondly in the form of a trading strategy-based profit generating technique. The literature suggests several such trading strategies to be effective, however the problem is that it is not known how the underlying forecasting methodologies of these trading strategies compare in a controlled testing environment, nor how their application impact portfolio risk. The gap is therefore that, even though the literature suggests the significance of AI-based trading results in fintech markets, these findings are fragmented, and it is unclear how they may generalise in the greater context of portfolio management amidst rapid global fintech advancement.

1.3 Research Question

Financial technology adds to the complexity of financial markets. Simultaneously, however, the empirical evidence suggests that these fintech-endogenous complexities, even though they manifest adverse risk-return profiles in the markets where they are most prevalent, offer a value proposition to an investment portfolio. This has been found in terms of diversification properties, as well as in a speculative wealth creation framework (each on multiple separate occasions). The latter has enjoyed specific attention in the use of several different neural network-based trading methodologies, which is a combination that seems to be worth exploring in more depth. However, since these findings often present different combinations of trading strategies and neural network specification combinations, the question that is asked now is how the most promising of these trading strategies and neural network specifications compare to one another in a controlled testing environment. Also, how do these techniques impact portfolio risk?

1.4 Research Objectives

1.4.1 General Objective

This paper aims to utilise the established relationship between financial technology and financial market complexity in order to propose a theoretical explanation for recent empirical findings that suggest significant speculative wealth creation implications in the AI-driven Bitcoin investment domain of the literature. The relevant theory is materialised through the application of empirical tests regarding the extent to which market characteristics associated with higher levels of financial technology may be utilised in contributing toward an improved risk-return profile for a given investment portfolio. The theoretical value added will be in the conceptualisation of fintech as a playing field upon which the forces of risk and return interact. The empirical contribution will be to provide a comparison between the various existing “AI-in-Bitcoin trading” methodologies in terms of the extents to which they are able to mitigate the adverse risk profiles in complex, fintech-endogenous markets. The contribution is emphasised by placing the findings back into the context of an investment portfolio.

1.4.2 Specific Objectives

The general objective can be divided up into distinct specific objectives, which are to:

(a) Provide a theoretical framework through which the link(s) between financial technology and different forms of financial market complexities may be understood.
(b) Provide an empirical overview of the real-world market effects of the link between financial technology and financial market complexities, and how these complexities have been utilised toward speculative wealth creation in the past.
(c) Compare the performance of various neural network architectures in the context of a speculative wealth creation problem.
(d) Place the implications back into the context of the risk-return profile of an investment portfolio.

1.5 Research Method

The research method is broken up into distinct parts that serve separately to reach the objectives. The first research objective is the focus of section 2.2 wherein the developments of financial technology are considered chronologically, alongside which the relevant theoretical links to market complexities are explored. This undertaking serves to create a suitable background and historical overview of some key developments that led to the point where research such as this can be undertaken, whilst still constructing a robust theoretical framework through which the links between financial technology and financial market complexity may be understood. This process entails both a broader discussion of key concepts in financial technology (section 2.2), followed by a detailed overview of a suitable modern application thereof i.e. Bitcoin (section 2.3). The theoretical framework is therefore established in section 2.2, and section 2.3 serves as the background to a case study within the conceptualised framework. During this procedure, as theoretical links between fintech and various financial market complexities are developed from the theory, some relevant empirical evidence of the manifestations of financial market complexities are mentioned. However, the in-depth analysis of the empirical phenomena in fintech-endogenous financial markets is reserved for consideration in the subsequent section as it pertains to investment strategies based on the theoretical links. In each instance, therefore, where the development of a significant theoretical link is the central focus, some of the supporting empirical findings that pertain to it may be referenced in section 2.2, but the empirical evidence-driven analysis of market phenomena (the second research objective) is attended to in section 2.4 where it is more relevant to forecasting strategies, and thereby categorised by empirical evidence sourced from the real-world market. This categorisation serves dually to emphasise the chronological development of the key theoretical concepts (section 2.2; objective a), whilst still maintaining rigorous empirical categorisation of the relevant market complexities (arising from fintech) as they pertain to asset valuation methodologies (section 2.4; objective b). Seeing as section 2.4 aims to link the literature on asset valuation methodologies to fintech-associated market phenomena, the relevant concepts are attended to in the form of modelling techniques that have been used in the literature to generate the findings upon which this research contribution is based. As the focus here is on asset valuation and how appropriate techniques have been found to successfully deal with increasing prevalence of market complexity, the discussion of the asset valuation methodology section begins with some relevant background and categorisation, followed by some of the simplest techniques that arise from classical linear statistics. This is used as a baseline from which more powerful models are described, and ensures consistency in the build-up of asset valuation methodology complexity coincidental to market complexity. Finally, these topics culminate into a summarising review of the empirical literature of complex modelling techniques in complex financial markets in section 2.5, after which the empirical contributions of this study are set to follow in the subsequent chapters.

The third and fourth objectives are the empirical contributions. These objectives aim to add to the body of knowledge through the process of experimentation. The third objective aims to compare the performance of various neural network architectures in the context of a speculative wealth creation problem. The data originate from the online database CoinMarketCap and comes in the form of a univariate time series containing 2189 observations of daily price (in USD) data for Bitcoin. The dataset's first entry is on 28 April 2013 and the last 25 April 2019. The first couple of steps in the methodology include simple and widely-known descriptive measures, which are considered to be standard steps in data analysis. These steps include visual inspection of a line graph and, more formally, an Augmented Dickey-Fuller (ADF) test combined with log-differencing procedures to achieve stationarity. Once stationarity is obtained and the order of integration is known and accounted for, the data will be checked for meaningful lag structures with the use of an Autocorrelation Function (ACF) and a Partial Autocorrelation Function (PACF). Furthermore, an autocorrelation matrix will be calculated to confirm or deny what is suggested by the ACF and PACF, after which lag plots will be visually presented to further establish whether all of the metrics' results coincide. Finally, a decomposition plot is to be visually presented in order to investigate patterns in the data pertaining to trend, seasonality and the frequency of these phenomena. This is essentially a preliminary exercise in order to ensure that there indeed exists the necessary autoregressive properties in the data that warrant the use of the methodologies that involve using past data to minimise forecasting error (in this case, neural networks).

The next step is to estimate a group of models originating from empirical findings in the literature. In the execution of this step, the model accuracy will be compared to what had been found in the previous studies that warranted the involved model's inclusion to ensure that the appropriate empirical results have been reproduced with integrity. The models to be tested include 3 neural networks of increasing architectural complexity. These include a convolutional neural network, a recurrent neural network and a long short-term memory neural network. In order to guarantee the integrity of the comparison between the neural networks, the prediction methodology and the trading strategy methodology are kept the same for all of the involved models in this section.

Every network in this comparison is tasked with using the first 50 entries of the dataset as input to predict the 51st entry, which forms a single iteration. The second iteration would consider entries 2 to 51 of the dataset as input and attempt to predict observation number 52. The third iteration estimates the model parameters using observations 3 to 52 as input and attempts to predict the 53rd observation. This is what is known as a rolling window forecast. Thusly, for every model, a daily predictions file will be created in the format:

t	Predicted t+1	Real t+1
xxxx	Xxxx	xxxx

After such a file is created for all the involved networks, a single, independent, unifying, profit maximising agent is given the first two columns of one row at a time. This agent is kept constant for all the models used in this section, so as to establish integrity in the comparison of the results. The trading program then makes the investment decision given the first two columns of a single row, and the last column of the given row is used to generate a profit (or loss) given the subtraction of Predicted t+1 from Real t+1. For every iteration (row in the dataset after the 50th row), the profit (loss) is added (subtracted) to (from) a variable "profit". After all of the rows have been iterated over, the value of the profit variable equals the profit/loss that the given model would have made if the agent were made to trade off of the model's predictions as input. This process ensures that the results are comparable, as every model's output is treated exactly the same. This will provide a foundational benchmark in the form of a back-testing result based on how the model would have fared on the historical data. The appropriate error metrics (RMSE, RMSLE, MAPE) are to be reported as a more general measure of likely performance in the case of controlled and reasonable extrapolation.

This brings us to the fourth research objective, which is to place the obtained results back into the context of an investment portfolio. The core contribution that this objective attempts to make is to contextualise the extent to which an AI-based trading strategy improves the risk-return ratio of Bitcoin and thus assets that exhibit similar characteristics. This is done to assess the implications of the findings in the previous objective for fund managers specifically, because if these contributions are found to improve the risk-return ratio of Bitcoin, they may hold significant value to fund managers in terms of portfolio risk. This is particularly handy, as the trading strategy employed to reach the third objective ensures that there will always be some exposure to both assets that are traded (in this case, USD to Bitcoin), and that the trading profit is realised in the margins of exposure to the respective assets. This is explained in more detail during the report on the trading strategy (chapter 4). The results are obtained by considering the risk-return characteristics of Bitcoin alone, and comparing these characteristics to the situation where Bitcoin is traded daily with a specifically balanced AI-based trading strategy. This will answer the question of whether including an AI-based trading strategy to Bitcoin will improve its risk-ratio, whilst the investor may still benefit from its diversification properties as has been proven useful in previous research. The procedure serves primarily as a contextualisation of the implications of the results, in order to establish the value of the insights to the fund managers-portion of the intended audience of this research.

1.6 Chapter Outline

The remaining chapters of this dissertation delineate as follows:

Chapter 2: Provides a theoretical framework through which the link(s) between financial technology and different forms of financial market complexities may be understood, as well as an empirical overview of the real-world market effects of these links and how their corresponding financial market complexities have been utilised toward speculative wealth creation in the past.

Chapter 3: Is the data and methodology chapter, and thus explains the detailed semantics of the methodology as well as how its implementation will reach the specific empirical objectives.

Chapter 4: Provides the results obtained during the implementation of the research methodology and a discussion thereof.

Chapter 5: Provides a conclusive discussion of how the objectives were met, in order to answer the research questions, in order finally to solve the problem statement of this research.

CHAPTER 2 LITERATURE REVIEW

This chapter serves to introduce and review the first concept at the core of this study, namely fintech. The chapter is divided into two sections, which are firstly a historical overview of financial technology, followed by a case-specific overview of one of the field's more recent developments: Bitcoin. This will provide a logical run-up toward the subsequent discussion which will focus the attention on analytical instruments through which rapid fintech developments may be understood.

2.1 The Background on and disambiguation of financial technology and fintech

The purpose of this section is to provide an overview of the theoretical capabilities of financial technology as a concept. It is an investigation into any form of technology enforcing change upon any form of finance, with the purpose of discovering financial technology holistically through time. This section will be followed by a narrower discussion of technology's influence on money, and finally a case study thereof. First, however, a disambiguation between financial technology and fintech is suitable.

The term *fintech* is a portmanteau of 'financial technology', and a formal, scientific definition of the term is given, through extensive scientific research, by Schueffel (2016) as a *new* industry that applies technology to improve financial activities. The difference between fintech and financial technology as a concept, however, must be addressed, because technology has been influencing finance long before the word fintech has had its first recorded use (Thomas & Morse, 2017). What's more, Schueffel (2016) performs scientific trend analysis on the term fintech and insists that it refers to modern developments in the field of financial technology exclusively of such developments, say, in the 1800s. It is therefore not justified to refer to any technological influence on finance as fintech, as this would oppose the conclusion to which Schueffel (2016) arrived in such a meticulously scientific fashion. This means that there is a question to find an answer to before this discussion of fintech can continue. The question is: "When did financial technology become fintech?"

2.1.1 When did financial technology become fintech?

The difference between financial technology and fintech is most clearly distinguished upon the consideration that financial technology is no coined term, but merely two words pertaining to a concept, for which the etymology of each shall suffice in exploration of their joined meaning i.e. finance (n.) "settlement of a debt" and technology (n.) "a discourse or treatise on an art" (Online Etymology Dictionary, 2019). The word fintech, contrarily, is a coined term referring to a specific area within modern financial technology as explained by Schueffel (2016). According to Trebacz

(2019), the term fintech was first used in 1980 by Peter Knight, who was at the time the news editor of the Sunday Times and used the term to describe a computer program that changed the parameters of an e-mail in his inbox. This particular use of the word does not quite satisfy Schueffel (2016)'s definition, and the search for the moment in history when financial technology became fintech continues to 1993 when Citicorp officially coined the term in a context that satisfies the definition laid out by Schueffel (2016). The term fintech was used by Citicorp as the name for its Financial Services Technology Consortium called "Fintech (the word, that is) Evolves" (Herrera, 2017). Seeing as this use of the word is contextually satisfactory for how fintech is discussed here, 1993 is then concluded as the year of birth for the term fintech.

Given then, the disambiguation that financial technology and fintech are indeed distinguished phenomena, and that technological developments in finance over time morphed into a subsector of the financial services industry known as fintech, the discussion of the history of financial technology can be considered given that the following underlying premise is posited: all fintech is financial technology, but not all financial technology is necessarily fintech. This satisfies the definition presented by Schueffel (2016) and allows the discourse to continue with a logical and scientific framework that is unhampered by ambiguity. Given then, that the section concluded here distinguishes between the terms financial technology and fintech, the term fintech is further used according to the disambiguation presented here.

2.2 A theoretical framework investigating the links between fintech and financial market complexity over time

A well-founded case exists for the argument that financial technology is truly ancient, and possibly even that technology is inherently indistinguishable from finance – from the logic that all developments onward from barter itself are examples of technological advances in finance (Davies, 2010). According to this logic, money itself is an application of financial technology, which is a thought process that resonates with one of the most recent developments in the fintech industry – Bitcoin and other cryptocurrencies. This premise provides consistency in the view that finance and technology are hard to conceptually separate without digging back to a time before money was invented, and barter was the way of transferring value. However, the consistency provided by the consideration of the ancientness of financial technology, comes at a price in terms of the mechanics of reasoning thereabout. That price is the deductive logic that either the entirety of the history of money is to be considered, or a time in history must be suitably chosen to begin the discourse of "modern" financial technology. Therefore, consider the history of all finance over all time, or define a timeframe which would constitute modern financial technology of relevance to the problem statement. Arner *et al.* (2016) provides a solution to this problem in the form of a categorisation of fundamentally different eras of fintech developments over time, which begins in the 1800s.

The modern history of financial technology can be categorised into three epochs, described suitably by Arner *et al.* (2016) as Fintech 1.0, 2.0 and 3.0. This categorisation is used here as well, as it delineates developments of distinct underlying nature into three suitable groups. It is noted that the author in reference recognises the novelty of the term fintech itself, and posits that it is a new term that can be used to describe an old relationship between finance and technology (which agrees with what is described in the preliminary disambiguation section above). In keeping with Arner *et al.* (2015)'s categorisation, "Fintech 1.0" refers to the period during which financial technology solutions replaced analogue systems with digital ones, but very little industry disruption took place as a by-product of this replacement. "Fintech 2.0" refers to the period when digitisation of traditional financial services took to market i.e. the solutions from the Fintech 1.0 epoch resulted in industry disruption. What is left then is that "Fintech 3.0" refers to recent developments of financial technology, that satisfy Schueffel (2016)'s definition of the term fintech. The modern fintech landscape is also investigated separately with regard to the certain force fields (economic, public perception, political and regulatory) within which it performs its functions (finance & investment, payment system infrastructure, data security & monetisation, operations & compliance and consumer interfacing). After this discussion, attention is shifted toward the utmost recent developments with specific focus on probable future outcomes predicted by cutting-edge industry research. The scene is set, then, for the discovery of the earliest forms of financial technology – Fintech 1.0.

2.2.1 Fintech 1.0

For the sake of logical flow, the Fintech 1.0 epoch is distinguished in terms of the analogue era and digital era with regard to its underlying technological developments. The analogue era is discussed first, with the telegraph considered its inception. Then, the focus is shifted to the digitisation of previously analogue methods during the 1960s, which lay the foundation for fintech as we know it today. This is in the form satisfactory of Schueffel's (2016) formal scientific definition of fintech. First to be considered then, is the analogue era of the Fintech 1.0 epoch.

2.2.1.1 The analogue era of Fintech 1.0

The term "Fintech 1.0", as first introduced by Arner *et al.* (2016) describes the period from 1866 to 1967, a time during which contemporary technology was already enforcing change upon finance, but the underlying systems were analogue as opposed to digital. Even though these systems were comparatively simple to what is seen in industry today, it still enabled global connectivity in finance for the first time throughout history, and set the foundations for the modern digital disruption trends that are evidently influential today (Wulan, 2017). One, and perhaps the most prominent, such a foundational development is the invention of the telegraph.

The telegraph (first commercial use in 1838), coupled with the laying of transatlantic cable in 1866 is widely is considered the point in history where modern fintech had its infrastructural roots (Arner *et al.,* 2017). The conjoined use of these two technologies allowed reliable analogue transmission of handwriting, signatures or drawings across international oversea borders. While these signal transmissions were confined to a small surface (150x100mm), and took a long time (approx. 108 seconds for 25 handwritten words), they allowed reliable signature verification in banking transactions across vast oceans. This was the most prominent use of the pantelegraph (Sabine, 1869). During the telegraph's era, other analogue technologies also played an important role in financial interlinkage across borders or vast geographical spaces. Such technologies included railroads, canals and steamships (Schoonover, 2013). While steamships and railroads are hardly fintech, this was during the same era that the development of the Fedwire Funds Service occurred. This was a dedicated national funds transfer network featuring a Morse code system that directly connected the Central Banks, Financial Services Board and Treasury of the United States Federal Government (Gilbert *et al.,* 1997). The technologies of the time, therefore, albeit not digital, allowed for the first interconnected financial systems. It is this same era that J.M. Keynes (1920) was referring to when he wrote:

> "*The inhabitant of London could order by telephone, sipping his morning tea in bed, the various products of the whole earth, in such quantity as he might see fit, and reasonably expect their early delivery upon his doorstep; he could at the same moment and by the same means adventure his wealth in natural resources and new enterprises of any quarter of the world, and share, without exertion or even trouble."*

It is writings such as this that illuminate the extent to which financial globalisation had already been achieved by the turn of the 18^{th} century. However, the streamlining of the globalisation process was not allowed to continue unhampered indefinitely. In fact, the international risks and general economic disarray associated with World War I significantly constrained the developments of financial globalisation, and thereby fintech applications also (Broadberry & Harrison, 2005). This generally observed decline in the rate of globalisation due to war, however, also proved to deliver its own positive externalities in the years to come, prime of which as pertains to the current context, is strategic technological advance manifested through warfare. It has been found that engaging in the act of warfare has economic effects on the sovereignties involved (Coccia, 2018). One such an effect is significantly impactful on capital stock through the process of capital flight, which is the economic abstraction of the process where the nation concentrates vast amounts of its resources on war-related activities (Collier, 1999). Conjoined with this are the findings of (among many others), Roland (1995) who investigated the effects of war on technological advancements and consequently found that wartime has a significant impact on scientific and technological invention. This conjunction aligns with the empirical investigation of fintech development before, during and after the period in which World War I occurred. Contemporarily modern fintech was prevalent, yet analogue, simple and mechanical before the war, virtually non-existent during the war, and completely reinvented after the war, with the direct influence of wartime-developed technologies enacting this reinvention. Examples of this include the private company International Business Machines (IBM) capitalising on technology developed during World War 1, by introducing digital computer tools into private companies commercially. Also, the commercialisation of the first hand-held calculator by Texas Instruments in 1967, and perhaps much more systemic to an economy, the commercialisation of consumer credit products by Diners' Club, Bank of America and American Express all in the 1950s. All of these breakthroughs were acts of making commercial the scientific inventions and developments that had been invented during World War I. Collectively these events are described by Arner *et al.,* (2017) as the consumer revolution pertaining to financial technology. Perhaps the most notorious of the events encompassed by the fintech consumer revolution (concluding the first era of fintech 1.0), was the establishment of a global telex network which was in place by 1966. This network provided the same utility as is earlier embodied by the J.M. Keynes quote of globalisation in the 1920s. However the conjunction of a network of pantelelegraphs, transatlantic cables, railroads and steamships, provide their utility in an analogue fashion. The global telex network of 1966 provided its utility based on underlying digital technology, and digital inventions, as history unfolded, were able to fundamentally capitalise on the coming years of digital revolution and the exponential nature of digitisation characterised exemplarily by Moore's Law (Moore, 1965). These developments, specifically the handheld digital calculator, marked the start of the era of digitisation in fintech.

Given then that the historical discussion so far, considers the developments in the Fintech 1.0 epoch pertaining to underlying analogue technologies, it is logical to consider next the process through which these technologies were made digital over time.

2.2.1.2 The digital era of Fintech 1.0

The creation of the first handheld calculator ushered in the era during which digitisation of existing financial technology solutions took place. This era stretches from 1967 to 1987, and is described by Arner *et al.,* (2015) as a time when financial services transitioned from an analogue to a digital industry. Certain key developments set the foundations for the second period of financial globalisation, first among which was the previously mentioned handheld calculator, and more specifically among its relevant financial applications, the Automatic Teller Machines (ATM). The first ATM was put into commercial use by the Enfield branch of Barclays

Bank in London, on the 27th of June in 1967 (McNeil, 2002). The ATM Industry Association (ATMIA) recorded over three million ATM's in use on the 50th anniversary of the ATM in 2017 and listed among their insights that the ATM laid the groundwork for modern day 24hr self-service culture in digital retail financial services. Of course, the invention of the ATM is merely one of many, solving the problem of more efficient public interfaces through which financial services are delivered, and other contemporary innovations should be considered along with it.

There was another revolution taking place around the same time and pertained to the area of payments, specifically interbank settlements, a landscape that has grown exponentially in size and complexity as an inherent support mechanism for international finance (Gai *et al.,* 2011; Arinaminpathy *et al.,* 2012). The world's first cybernetic interbank settlements hub was formed in 1968 in the UK and was named the Inter-Computer Bureau, forming the basis of what is known today as the Bankers' Automated Clearing Services (BACS) (Welch, 1999). The technological equivalent of the BACS in the US was the Clearing House Interbank Payments System (CHIPS), established in 1970 (Lingl, 1981). Concurrently, the aforementioned Fedwire (previously an analogue telegraphic system) underwent digitisation. Considering that Fedwire was established in 1918, and yet still was subject to digitisation along with domestic interbank clearances, reflects the need for interconnection between domestic settlements and cross-border settlements, and although the now digitised Fedwire did enable marginally more efficient cross border clearing of payments, the borders in question were still only that of states in the USA. Logically, the next step will have been to digitise existing analogue international payment clearance systems such as the pantelegraph, and so it was done, in the form of the establishment of the Society of Worldwide Interbank Financial Telecommunications (SWIFT) in 1973. The digitisation of a network of such worldwide-concurrency nature is assumed to lead to greater interconnection among banks, a rationale to which the collapse of Herstatt Bank in 1974 has been attributed (Benston & Kaufman, 1995). This phenomenon, previously of mere cross-border interaction among banks, which was now transformed into a concurrent, efficient and digital global interbank market, gave rise to previously unknown implications for all economies involved (and some uninvolved yet implicated through spill-over effects) and called for international regulatory reform (Schenk, 2014). So much so, that Mourlon-Druol (2015) specifically refers to the fall of Bankhaus Herstatt as one of the landmarks of post-war financial history, and further investigates international regulatory reform that may have arisen from these events surrounding 1974. Among such regulatory reform of the time was a series of soft law agreements on developing regulation toward robust payment systems (Schooner & Taylor, 2009). Since the first regulatory acts were signed into law through the respective political and legal mechanisms of all countries involved, regulation has continued to play an ongoing role in global finance in a perpetual evolutionary interaction – a statement proven by the sheer size and complexity of international foreign exchange markets and accompanying regulation

(Saunders *et al.,* 2006). As regulation and foreign exchange market size can be contrasted in terms of its activity between the early 1970s and the modern day, the area of securities is quite different. This is especially true when the greater context of the comparatively much more vast history of securities trading is, in turn, contrasted with the relatively short history of foreign exchange markets and their regulatory structures. This comparison also establishes that the history of securities trading is well recorded and, that activities in these markets have been well regulated for a comparatively longer time than foreign exchange markets (Banner, 1997).

The consideration of the digital era of Fintech 1.0 in terms of how it affected securities trading is then a question of how securities trading activity flowed over to digital systems from physical trade floors dating back to the 1600s and not a question of how digital securities trading products enabled securities trading in the first place. This is, once again, a key consideration as to why the era in question is investigated from the perspective of a digitisation process of technologies that were already available and in use in an analogue fashion. It is emphasised therefore that the creation of digital securities trading and all regulation to follow from this, are once again instances of the effects of digitisation, not of invention.

In order to further investigate this, the consideration of the National Association of Securities Dealers Automated Quotations (NASDAQ), along with the National Market System is a key factor. The NASDAQ was established in 1971 and its establishment initiated the decline in the activities of fixed securities commissions. These declines, along with various political pressures, led to deregulation of securities trading through legislature proposed and enacted by the Securities Exchange Commission in 1975 (Jarrell, 1984). This process of deregulatory events coupled with digitisation played a central role in the establishment of a National Market System in the US (Hamilton, 1978). Any national market system, however, has to take into account and serve all of the participants that serve, in turn, its growth or functioning toward the manifestation of prosperity. Therefore, these national market systems have to cater for interbank markets, markets for professional investors and, of course, consumers. In terms of consumers, the contemporary developments of the analogue-to-digital era of Fintech 1.0 included attempts to introduce online forms of banking. This attempt in the US was, however abandoned in 1983 - three years after its inception (Choron & Choron, 2011), and a similar attempt toward online banking emerged in the UK under the Nottingham Building Society (Daniel, 1999). The case in the UK, however, was not so quickly abandoned, and created demand for banks to improve their Information Technology (IT) systems, which complemented internal operations' efficiency, continuously and gradually replacing legacy (paper based) systems with computerised procedures over time. This, however, contributed toward supporting developments in risk management technology aimed toward internal risks in a bank. One such an invention is that of Innovation Market Solutions (IMS) in 1981, called the Bloomberg terminal.

The Bloomberg terminal also serves as good foundation to the reasoning that fintech is inherently an invention of versatile banks, and not necessarily some form of independent exogenous force that threatens to penetrate banking sector market share. This reasoning can also be interpolated with metrics from today's banking sector – specifically upon consideration that Goldman Sachs, for example, employ more software engineers than LinkedIn, Twitter or Facebook (Marino, 2015), evidently competing sufficiently well for talents with fintech start-ups and tech companies. This duality of fintech – coming both from small tech companies and big banks – is investigated in great depth in the modern fintech trends section.

This point in time serves as conclusion to the era in which previously existing analogue solutions to global financial technology underwent digitisation, and simultaneously builds the foundational framework within which to consider the next epoch of fintech history – Fintech 2.0. The developments that built on and followed the digitisation age described above cannot simply be viewed in the same light, as they begin to consider the risks of globalisation and the digitisation thereof more explicitly. The section that follows clearly considers developments that are associated with digitisation and the risk management of these developments.

2.2.2 Fintech 2.0 (1987-2008): Further developments of digital financial services

The term Fintech 2.0 is in keeping with the framework of fintech history proposed by Arner *et al.,* (2015). It is a short-hand descriptive of the relevant events that took place in financial technology between the years 1987 and 2008, and bear very little fundamental relation to the process of digitisation of previously analogue technologies (Fintech 1.0) as described above. The Fintech 2.0 epoch's end is considered to be the global financial crisis of 2008, for the reason that an investigatory approach is taken here as to whether or not and to what extent fintech developments could have contributed toward the crisis. The Fintech 2.0 epoch is considered first from the perspective of the contemporarily novel attention that events and undertakings encompassed by the term fintech enjoyed in the light of the after effects of the discussion pertaining to digitisation above. Then, thereafter, it is considered from a perspective where regulatory action is the prime area of inquiry (a mechanism through which to describe the interactions between financial innovation and regulatory response thereto). This regulatory perspective is considered after the effects of the digitisation process are reviewed below.

2.2.2.1 Fintech 2.0 begins with international regulatory attention

The events and undertakings during and after the digitisation era described in the previous section, brought about a stern reinvestigation of cross-border financial interconnections from both a risk management and a regulatory point of view. The life and economic role of the investment banker was romanticised to an extent by film (Wall Street, 1987), and global financial interlinkage coupled with stronger, more ambitious, more sentiment-driven digital investment banking practices marked the start of the Fintech 2.0 era with the “Black Monday” stock market crash of 1987 (Waldrop, 1987). While Black Monday is not the specific matter under investigation here, the effects that it had on regulation are too relevant to omit. The Black Monday crash evidenced that international interlinkage in payments cannot be causally uncoupled from international interlinkage in financial risk, and even today, approximately thirty years after the events in question, there is still no clear consensus on the causes of the crash. Nonetheless, a significant share of focus from regulatory and investigatory entities was paid to the use of computerised trading systems that automate trading decisions by financial institutions (Waldrop, 1987). The lessons learned from such programmed trading led to various regulatory mechanisms specifically in electronic markets. Among these mechanisms were “circuit breakers”, which are sets of software protocols that control the speed of the reflection of price changes in electronic markets. The regulatory mechanisms were, however, not limited to electronic markets and the speed of price formation, and also focused on cooperative frameworks, similar in spirit to regulatory attempts after the 1974 Herstatt collapse, which aimed to improve cooperation across borders in order to avoid such failure in the future. Additionally, from a regulatory viewpoint, the European Union (EU) was forming across the Atlantic Ocean, a process that would, by the coming of the year 1992, establish a framework for a future where EU member countries would converge to the use of a single financial market. This process entailed many sub-processes, including the Big Bang financial liberalisation process in the UK, the 1992 Maastricht Treaty and the establishment of numerous financial services Directives and Regulatory bodies from the late 1980s, which would later finally bring to light the establishment of a single financial market for all of the EU. Such regulatory cooperation seemed to contribute positively toward the manner in which financial systems sustain rapid digitisation, as by the late 1980s, the global financial services industry was largely supported by electronic transactions between financial institutions, market participants and to a limited extent end users of financial products (Sassen, 2005), whilst simultaneously engaging in unprecedented interconnected behaviour and the globalisation of news (Palmer *et al.,* 1998). This led to some considering the global financial services industry as the first ever digital industry (Engelen, 2010), which drove demand for greater risk management protocols from which, in turn, Value at Risk (VaR) procedures took the foreground from analysis employing Long Term Capital Management (LTCM), and while the Asian and Russian financial crises of 1997-1998 undoubtedly contributed to such regulatory phenomena, it was not until the invention (or more specifically, commercialised use) of the internet, that regulation really had to adapt fast.

The emergence of internet based software applications once again posed a challenge to regulation to keep up with unprecedentedly rapid developments in fintech related banking practices. Some perspective of the speed at which banking practices incorporated the use of the World Wide Web (WWW) is gained upon consideration that, in 1995, Wells Fargo was the first bank in the world to provide an online banking application (in the form of online account checking), while six years later, eight banks in the world had at least one million clients active online (Riggs, 2015). The vast and rapid increase in the number of active users online required unprecedented investment into IT systems by financial institutions and by the beginning of the 21st century, financial service providers and regulators alike were pushed toward digital solutions. These effects of underlying IT systems development, coupled with the emergence of virtual banks (ING Direct, HSBC Direct) in the UK required more effective regulatory action in a virtual, information-driven financial services landscape.

These developments in regulation may make it beneficial to consider the relationship between innovation and regulation more closely, specifically in the context of the early half of the Fintech 2.0 developmental epoch discussed above. Another reason why such consideration may be beneficial is toward gaining insight into the relationship between financial innovation and the 2008 financial crisis. As has been stated previously, there exists a lack of consensus as to the causes of the crisis, yet a look at the relevant conjecture will provide possibly a certain perspective relevant to the discussion of Fintech 3.0 which is to follow thereafter.

2.2.2.2 Regulatory responses to the early events of Fintech 2.0

The above consideration of the early half of the Fintech 2.0 developmental epoch shows that developments in digitisation of financial services began to draw significantly more attention from regulatory bodies internationally. It evidences reactive legislature from the EU and the Russian and Chinese markets which unfolded later in time – during latter half Fintech 2.0. This section attempts to consider these phenomena in the light of developments leading up more closely to the 2008 global financial crisis, and to what extent the inter-dynamics among innovators and regulators contributed to causes and remedies of the intermarket dissonance that drove market forces toward the events that transpired during and after the 2008 crisis. The important questions with relevance to the interaction between innovation and regulation that are investigated here are firstly whether there exists reflexivity between the two market agencies (innovators and regulators). Also, what time frame responses from regulation take on as it reacts to financial innovation and finally whether legislation has ever been considered redundant and thus removed, and if so through which structures such deregulatory activity has taken place. The discussion can therefore continue on to a more formal regulatory perspective of Fintech 2.0.

The first consideration is that of reflexivity between two relevant groups of market participants. These are innovators and regulators. The reason for this is that as innovations arising from early Fintech 2.0 developments began to draw more attention from international regulatory bodies (as is discussed in the previous section), and the responses from such regulatory agents are expected to have had a reflexive correlation with innovations later on in the Fintech 2.0 developmental epoch. A good example of concrete (legislative) regulation attempts arising from the former half of Fintech 2.0 came from the far-East of the world's economies. More specifically, in the form of a keynote address from the then Deputy Chief Executive of the Hong Kong Monetary Authority (HKMA), David Carse, in 1999. The aims of the endeavour were to re-establish growth policy in the aftermath of the 1997 Asian financial crisis, and therein to consider a proposed regulatory framework needed to react to e-banking (Carse, 1999). It is noted at this stage that e-banking had been around since 1980 (Choron & Choron, 2011). Upon contextual reconsideration of the proposed logic of reflexivity between innovation and regulation, it is evident that if any feedback exists between the involved entities, the frequency of such feedback is relatively low. Such a vast lag in the reaction time of a discrete process (such as regulation) that is essentially reactive to innovation in the nature of their interaction, places the relative speed of continuous technological advance into perspective. The reaction time lag that exists between innovation and its appropriate regulatory response is sometimes considered to be beneficial to the way that efficient markets evolve. This rationale is founded in the sense that there exists little benefit in attempts of regulating all new innovations as they occur, and that such behaviour will negatively influence efficiency and growth. Arner *et al.* (2017) compare the macroeconomic policy pertaining to fintech of Singapore with that of Hong Kong to establish such reasoning, and thereby deduce that such "pre-emptive" regulation would increase the regulatory workload and simultaneously stifle innovation. Considering the conjunction of such logic with the distinguished forms of productive and non-productive labour in an economy (Smith, 1776), results in the conclusion that pre-emptive regulation would decrease activities probable to result in increased marginal productivity of labour (innovation) and increase expenditure on activities that are classified as non-productive labour (regulation). The severe undesirability of such an outcome makes the reactive nature of regulation naturally more acceptable and preferable to the alternative of pre-emption. However, reactive regulation – albeit superior to the alternative – carries its own risks. During the latter half of Fintech 2.0, e-banking exacerbated old risks that were of relevance to regulators above and beyond the level to which they were accustomed to maintaining in traditional brick-and-mortar banking models. One of the simplest risks exacerbated by e-banking is that of a virtual run on the bank. By providing direct and unlimited access to accounts, consumers need no longer be physically present in order to withdraw funds and this could have implications for liquidity risk. In the same 1999 keynote address by David Carse as is referred to earlier, it is mentioned that internet-based banks face the same type of risks as their traditional counterparts, but that the internet may heighten such risks (Carse, 1999).

Apart from the exacerbation of traditional risks, e-banking also introduces fundamentally novel risks to the banking industry, chief among which is credit risk (Sathye, 2005). Credit risk however, and more specifically, the way that financial technology changes credit risk profiles, has two sides to it (both a positive and a negative market effect are measurable). Firstly, when one removes the consumer's physical barrier to credit acquisition, it is expected that competition for credit would increase, that a larger number of borrowers and lenders could be connected to each other and therefore that the market for credit would gain overall efficiency. This is good for consumers upon first glance. However, when one considers market interconnection and contagion of risk, the US provides an example of how similar causes (market-driven interest rates instantiated by deregulation in 1980) could adversely affect market stability (Millon-Cornett & Tehranian, 1989). The benefits and detriments that e-banking brings in terms of broader credit markets can also be considered in terms of credit models and their efficiency. On the one side, the loan officer in an arbitrary town does not know all or even nearly most of the credit applicants in such a broad system as erected by e-banking practices. This takes the human gut feel of trusting and knowing an individual's reputation out of the credit decision making process, and the reader is encouraged to consider the effects of this subjectively. Upon further consideration of a broader credit market instantiated by e-banking, it also becomes clear that this absence of relationships with loan officers now require much more granular and mechanical big data solutions that contain information on credit applicants, which banks may use to facilitate the credit decision. These considerations are, however, of comparatively little scope when one considers the more fundamentally pressing matters of the time during which these occurrences took place. This is because, at the time of the international regulation discussion around these issues, it seemed to have been widely assumed that financial technology would continue to be an instrument applied by large, centralised, regulated, state-transparent and generally supervised financial institutions (Carse, 1999). This is the key concept that separates the developments of Fintech 2.0 with those of Fintech 3.0 (discussed in depth in the following section), as at the time during which regulatory reform was made (reactive to innovative fintech developments prior to the late 1990s), comparatively few small private companies were engaging in market activities associated with formally supervised financial institutions.

Using the UK as an example among many, it is learned that incorporating a firm with the word "bank", "banc" or "banking" is only allowed by regulators to companies who agree to be supervised financial services providers. Such legislation is a common phenomenon the world over in countries' attempts to protect financial markets from unknown risks associated with unsupervised financial service providers. However, the later developments of Fintech 3.0 would prove that the financial service provision may not so easily be limited to willingly regulated and supervised firms, placing the emphasis on consumer education of risk rather than government regulation thereof. Still, later in time it would be found that consumers trust tech firms more than they trust banks, a damning judgment to the effectiveness of consumer risk education (LTP Team, 2015).

This idea sets a satisfactory background to the consideration of the developmental epoch of Fintech 2.0 wherein the international discussion of the regulation of financial technology was a key factor. It is now possible to consider the consequent developmental epoch which is referred to as Fintech 3.0.

2.2.3 Fintech 3.0 (2008 – now) from the 2008 aftermath to cutting edge trends

We begin now to approach the true centrepiece of this work, which is the modern development epoch of fintech. In keeping with the stratification of temporal development as it is used in preceding sections, it is called Fintech 3.0 and characterises the stages of fintech development after the 2008 financial crisis. It is stated at this point that the developments discussed in this section are the first which truly satisfy the scientific definition concluded upon by Schueffel (2016).

This section is discussed in terms of the core factors that impact the developments of fintech in the modern day. The aim is thus to set a backdrop of market forces and then thereafter to consider fintech within this environment. These market forces begin with the public perception of finance in general.

2.2.3.1 Fintech 3.0 and the public perception of the financial services industry

As the Fintech 3.0 discussion begins with the period after the 2008 financial crisis, the public perception of the financial services industry is an important consideration to make. Public perception is the start of this background to later regulatory discussions, yet it is critical that the public perception of financial services becomes more and more important to consider and protect, the more fintech threatens to privatise business functions originally performed by regulated banks.

In the years following the 2008 financial crisis, a lot of research was done in order to understand the origins and causes of such wealth destruction. In this process, as the origins became more widely understood, the public perception of the people in the banking sector deteriorated (Agarwal *et al*., 2014). The effects, however, were much more widespread than public perception issues with regard to financial service providers. In the broader economy, it is estimated that approximately 8.7 million US workers lost their jobs as a direct economic cause to the crisis of 2008.

The effects are undoubtedly further reaching than this and become more difficult to quantify as they grow so. However, the reasoning can easily be made with the available relevant research that there are two categories of economic agents that were directly impacted by this (the latter of which pertains directly to fintech). The first category of economic agent directly impacted by the crisis is, of course, the abovementioned general public (the now jobless 8.7 million individuals). Among the effects of the crisis on the general public is their perception of the sector that caused their financial strife, and thusly a civil distrust of the financial sector was developed. The second category of economic agents directly impacted by the crisis is the group of people formerly employed by the now much smaller financial sector. This left an under-utilised or - at the very least - underpaid group of financially educated individuals looking for work in which to apply their skills. Couple this with a new generation of hopeful graduates aiming to secure a livelihood in the financial services industry (after years of training for it), and the result is an oversupply of highly technical skills pertaining to the financial industry, but no sufficient demand for such skills in order to absorb them into the financial services labour market. Fintech 3.0 results as a necessary side effect of such market mechanisms.

This brings a conclusion to the considerations of the effects of the 2008 financial crisis on public perception of financial institutions and thereby the topic is primed for reconsideration later on as it pertains to modern fintech and possible future outcomes thereof. The discussion can now continue to the effects of the crisis on regulation, in such nature that regulation may also be primed for later consideration of fintech in a similar way.

2.2.3.2 Fintech 3.0 and regulation of the financial services industry

In the years after the 2008 financial crisis, the investigation of possible causes of its wealth destruction led to a great deal of research being done in terms of these causes. It is logical that after a market collapse with such vast implications, people will want to know what happened to cause it, and avoid such behaviour in the future. This gave rise to many new regulatory activities to better manage the risks that probably caused the negative effects. Among such regulatory activities were new sets of compliance obligations.

The compliance obligations that originated in the attempts toward proper management of the risks that probably caused the 2008 financial crisis came in the form of altered commercial incentives and business structures for financial institutions. More plainly put, the methods through which the universal banking model enacted its business functions were widely challenged (Ferrari, 2015). Among such challenges came, for example, regulatory legislature in terms of ring-fencing obligations and increased regulated capital, which were meant to shrink a given financial institution’s capacity to originate low value-to-risk loans. More specifically and discretely, the abuse of financial innovations such as Collateralised Debt Obligations (CDOs) also became more strenuously regulated as they gave financial institutions the ability to detach credit risk from its underlying instruments i.e. the loans upon which they are built. The literature of post 2008 regulatory activity also includes many contributions in terms of plans through which "orderly failure" of banks can be achieved as opposed to disorderly failure (Anabtawi & Schwarcz, 2013). Some of this proposed regulation generated by the research sector made it to the real financial industry and was applied, for example, in the form of Recovery Resolution Plans (RRP's) and supervised stress testing procedures in order to reduce systemic risks posed by large banks (Barberis, 2012). The increased regulatory activity resulted in some fundamental changes to the financial service provision landscape, and fintech is expected to have been among the sectors that capitalised on these changes.

The mechanism through which regulation may have aided in the proliferation of fintech solutions deserves more attention, because the regulatory legislature arising from lessons learned during the 2008 financial crisis (among others, the Dodd Frank Act and Basel 3) were surely most welcome to consumers and stakeholders as such regulation is expected to prevent failures of the same constructs that could impact the financial sector in similar ways (Buckley, 2016). Yet the very same post-crisis regulatory reforms had unintended consequences arising from the nature in which they limit bank mobility and competitiveness in efficient service delivery. A suitable case study for this is the capital reserve requirements imposed by Basel III. The *prima facie* conclusion of the effects that Basel III's capital requirements are expected to have on the financial sector is that it should improve market stability and an arbitrary Basel III-compliant bank's capacity to absorb risk. However, upon deeper consideration, these same capital reserve requirements can divert capital away from Small and Medium Enterprises, start-ups, individuals and generally higher risk clients. The effects of this must be considered. If a regulated institution is, through regulation, not allowed to provide credit to borrowers above and beyond a certain point of risk, does this make the credit need of these borrowers disappear? Or does it rather create need for a completely unregulated, pseudo-black market for credit? The latter is the evident answer and if this were not the case, peer-to-peer (P2P) lending solutions and laws like the US JOBs Act will never have existed in the first place. This is not an opinion as to whether regulated financial institutions should or should not lend to borrowers with adverse risk profiles, but rather an iteration of the fact that the economy will privatise the provision of needs which are not catered for in the regulated public sector through the establishment of black markets (Reuter, 1985). Fintech, therefore, can be expected to step in where banks are not allowed to serve due to their regulation.

This concludes the discussion in which the inter-relationship between regulation and financial technology is primed for the later considerations around the probable futures of fintech. It is evident from this discussion that whilst regulation seems to improve the risk profiles of the entities which are regulated, it does not solve the problem of high risk lending, but merely disintermediates the regulated entities through the establishment of non-regulated entities. The further effects of this may lead toward dead weight losses and negative externalities which are not accounted for in the assessment of the risk profiles of the collective supervised market of regulated institutions. This will become a key theme in the discussion of cutting edge fintech and probable futures of fintech later on, but first, Fintech 3.0 must be considered in terms of the political landscape, so that insights arising therefrom, too, can be considered in the later section, bringing together all of the necessary background primed for investigation.

2.2.3.3 Fintech 3.0 and the political landscape

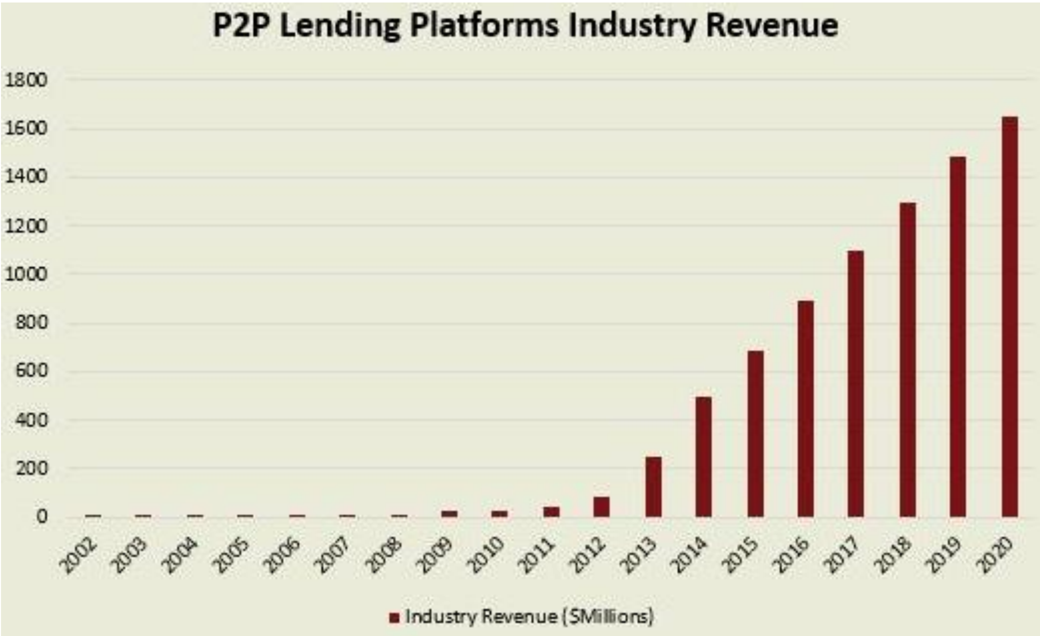
From the discussion so far, it is possible to consider the way that fintech (satisfactory of Schueffel (2015)'s definition) evolved from an economic perspective and the relationship that its development has had with regulation throughout this development. There is, however, one domain still unaccounted for, which is the political landscape and how it shaped modern fintech. The relationship between Fintech 3.0 and politics will therefore be discussed in this section, in order to prime the political discussion for later use in a more holistic evaluation of the fintech industry and its probable future outcomes.

As is argued by Demertzis *et al.* (2018), the political discussion around fintech should be much wider than legislation affirming regulation of capital markets. Instead, the "fintech opportunity" of the modern day is one that includes opportunity for politicians as well, as both benefits upon its inclusion in- and detriments upon its exclusion from political decision making, exist. Topfer (2018) argues, too, that the Global Financial Network (GFN) initiative deserves more political attention than it seems to be getting at this stage and Gabor and Brooks (2017) focus on the fintech opportunity in the sense that it offers potential to manifest greater financial inclusion in the country that considers it as a political tool. It is agreed that there exists an opportunity in improving political structures through fintech and some such attempts have even been made in the past.

A suitable example of political attempts to embrace fintech is the Jumpstart Our Business Startups (JOBS) act signed into law by Barack Obama on April 5th, 2012. The preamble of the act states that it is an act toward increasing American job creation and economic growth by improving access to the public capital markets for emerging growth companies. Title III of the act, the CROWDFUND Act, enabled entrepreneurs and small businesses to sell limited amounts of equity in their companies to a large number of public investors via social networks and other internet platforms – a process previously illegal under the United States securities laws (Stemler, 2013). From a political point of view, it is difficult to argue that such a law would be unpopular, as it is expected to raise capital for small firms and thereby aid in job creation, whilst simultaneously creating a legal market for credit provision that is not suitable in its risk profile for highly regulated and supervised financial institutions. On the other hand, high levels of unemployment and low levels of credit may challenge the effectiveness and job performance of elected representatives.

The creation of jobs and stimulation of higher risk credit is therefore a politically incentivised decision that can be enabled through fintech e.g. P2P internet banking. Simultaneously, it is quite difficult to argue that political support for entrepreneurship will have a downside from a policy perspective. These political circumstances, when assessed jointly, formed an ideal environment for fintech solutions to develop in order to cater for the now-legal flows of capital from individuals to small firms, start-ups and entrepreneurs. The effect of the JOBS act can be summarised in the following figure:

Figure 2-1: A bar chart depicting the industry revenue of P2P lending platforms.



(Source: IBisWorld)

from whence it must be stated that the JOBS act did not have its main aim set toward developing fintech directly (as Demertzis *et al.,* (2018) suggest) or even indirectly for that matter. The payment channels that were created by fintech were demands from the market after the legalisation of P2P finance and serve as an illustration of the political impact on fintech innovation.

This concludes the discussion of the politics of fintech and should provide a reasonable idea of how political decisions can impact fintech developments. With the politics of fintech concluded, it is now possible to consider the various impacts from economic, public perception, regulatory and political origin in the Fintech 3.0 era, and thereby create a more holistic consideration of the modern fintech field, through which possible future outcomes may be deducted.

2.2.3.4 Fintech 3.0 recent developments and trends

The purpose of this section is to review recent (latter half of the post 2008 era) fintech developments within the framework described above. In the preceding sections, the empirical inter-relationships between fintech and its economic, political, public perception and regulatory environment fields were investigated. This has been with the purpose of priming these aspects of fintech for a more global discussion of recent trends in fintech business functions on firm level and how they are probable to extrapolate to future developments of the industry. The idea is not to forecast numerically (yet), but rather to consider the likelihood of future firm competition outcomes within the specified fields which have been found thus far to influence the growth path of the fintech industry. It is thus emphasised here (having established the influence of the relevant fields in which fintech operates) that fintech has fundamentally different interactions with variable industry functions such as finance & investment, customer interfacing, payment system infrastructure, operations & risk management and data security & monetisation. The discussion following here attempts therefore to consider recent industry research in terms of these functions of fintech, whilst having posited that these functions are performed in dynamic dependency upon the relevant environments (economic, political, public perception and regulatory) discussed in the preceding sections.

Fintech has had fundamental and formative success in the field of finance & investment as far back as the dematerialisation of securities trading and the inception of NASDAQ. However, with the required political support e.g. JOBS Act, more modern developments have also been made, such as the development of crowdfunding and P2P lending platforms (Bruton *et al.,* 2015). Looking forward, recent trends in fintech as they pertain to finance & investment seem to indicate the likelihood of robotic advisory systems (robo-advisors) gaining market traction in the near future (Wright, 2017). These robo-advisor systems, however they may be packaged, branded and sold, are essentially algorithmically commanded allocation, management and optimisation of client assets. Examples of such systems in use today are, among others, E-Trade’s “Adaptive Portfolio”, Fidelity’s “Go” and TD Ameritrade’s “Essential Portfolios”. The same technology is observed in earlier, more specialised phases of financial services supply chain integration, specifically in highly specialised enterprise data solutions such as the Sequential Quantum Reduction and Extraction Model (SQREEM). SQREEM maps the digital footprints of over 300 million US individuals as of 2017 and has clients such as Deutsche Bank, UBS and Wells Fargo (Wright, 2017).

Returning more strictly to markets and investment, technologies such as Dataminr and Kensho forecast financial market movements with unprecedented success (Laurent *et al.*, 2015), and given that wealth management remains one of the best-performing business activities for banks globally (Deloitte, 2018), the forecasting element of market risk management is expected to show significant growth in the years to come. This notion is also supported by demographic trends of maturing holders of wealth in advanced economies which results in greater demand for wealth preservation and intergenerational wealth transfer. Among these markets, however, the robo-advisor threat is vanishing and the same principles of algorithmic asset management are being incorporated in activity layers further away from end users. This begins to pertain more to the disparity of technological penetration over demographical strata.

Consolidating variation over such demographical strata, four per cent Compound Annual Growth Rate (CAGR) growth in revenue for the largest global wealth managers is estimated between Q4 2018 and Q1 2020 (JPMorgan, 2018). The likely future business model for wealth managers is the combination of a low-cost trading platform, equipped with savings tools and an interactive advice section (Deloitte, 2018). Also, customer experience, expansion of product suites, goal-based financial advice, tax strategies and estate planning are expected to be the domains of stern competition (EY, 2018). Other product offering areas seem to be expected to change significantly with the introduction of fintech solutions such as REACH: a mobile application framework for closing transactions remotely through solving various problems of practicality in the process. Even though specific examples such as REACH are solving specialised problems, there seems to be relatively promising convergence on a larger (industry wide) scale toward deeply personalised financial service delivery solutions that are built on highly automated foundational systems (Palle, 2019). It must, however, be remembered that any such highly automated underlying systems will share to an extent the efficiency profile of the payment system upon which it is based. Payment systems infrastructure is the second fintech function up for discussion, for in which there is wide firm-level industry consensus that disruptive and fintech-enabled change is imminent.

In terms of payment systems and the infrastructure thereof, various public sector researchers as well as leading private sector firms in financial services industries are in agreement that technological improvement of payment system infrastructure is both necessary and imminent (Qian, 2019; PWC, 2016; EY, 2018). Given that payments systems are expected to change in the future, it is logical to consider the state of some current payment systems. Global legacy payment systems suffer from friction (Rucker *et al.,* 2019), yet the technology to remove such friction is already available. It is for this reason that Hengeveld (2019) describes the current situation of global payments as humans “adding friction” to inherently frictionless processes. Given such conditions it is strange to imagine that global (and even some domestic) payments are burdened with so much friction, and through rationale such as this it becomes more understandable that the likelihood of international payment system efficiency improvement is not far from impacting the real world's markets. Considering then the current state of alternative-to-legacy payment systems, it is found that this transition to infrastructures with dramatically lower friction is already underway, just rather divided among competitive participants. PayPal exceeded the 250 million active-users-worldwide milestone in 2018. Amazon continues to disrupt with native payment solutions (Berg & Knights, 2019). All whilst ground breaking solutions are rising up from the East (specifically China) as well, especially from competitors such as AliPay, Tencent and iQiyi (Sethi, 2019). It is thus clear that the market for frictionless digital payment infrastructure is divided among early competitors. However, the division has had little effect on hampering industry growth, considering that contactless payments alone are expected by some to represent a third of in-store transactions as soon as 2020 (Smith, 2018). Market insights that forecast such high rates of change in the payment system innovation market topology are more understandable upon consideration also that initiatives like the revised Payment Services Directive (PSD2) are regulatory attempts to proliferate payment system innovation and foster competition (Kosse & Hielkema, 2019).

Following the money trail in the light of initiatives like PSD2 also shows that firms are capitalising on regulatory support for such activities. This is evident from Mergers and Acquisitions (M&A) data pertaining to payments processing solutions, which dramatically increased from $31.2 billion in the 2017 financial year to $43.2 billion in the first half of the 2018 financial year (Deloitte, 2018). Other prominent examples of such capitalisation on payment infrastructure reform include record breaking deals like PayPal's acquisition of iZettle and Adyen's $1 billion public offering in Europe (Lunden, 2018). Looking forward, the first realisation that will make itself evident is that traditional card issuers will face difficulty increasing fee income from payments in a landscape so rich with non-traditional, low-cost, start-up culture fintech firms (PWC, 2017). The focus on finding the competitive edge in the domain of card issuers is expected to shift from income-from-card fees to predictable interest income on cards, which will have a dramatic effect on the bottom line of some card service providers. Furthermore, as these payment service providers search elsewhere for a competitive edge that might come in the sector for business to business (B2B) channels, a rationale made evident in Visa's recent acquisition of Fraedom, which is software focused on the B2B sector. Other, slightly less obvious areas of competition arising from payment infrastructure innovation include the value of the data acquired in the role of the overseer of a large payment channel. Turning these data into insights can prove to become a valuable competitive edge. Another area of competition could be interoperability between payment infrastructure and a wider acceptance of payment methods from a given service provider. The use of biometrics is also expected to play a more prominent role in the future of payments (Proenca, 2019). These variables are all part of the current reality faced by payment infrastructure providers, and unlocking the full potential of efficient payment systems may depend on joint efforts from disparate industry role players to consciously move toward a more efficient system that is interoperable, low cost, secure and frictionless. Such a payment system, however, as described above, will push the domain of competition from providing a payment infrastructure to other areas such as data and security.

In terms of data security and the monetisation of data, fintech is essentially guaranteed to be disruptive if not revolutionary. The research into the topic is, however, less quantitative than that of the two preceding sections and thus based more on conjecture and extrapolation from what is theoretically possible to what is realistically ideal in implementation. This discussion will not attempt to extrapolate conjecture to the future in a non-quantitative fashion, and therefore is limited to the most recent publications of possible frameworks through which data can be communicated more securely and also value channels through which data can be monetised. It must be emphasised at this stage that the discussion around the monetisation of data is not all that new, as innovative strategies from Fintech 2.0 and earlier Fintech 3.0 had already been exploiting data flows as value flows. This trend has just become much more emphasised, and data monetisation is expected to morph into a domain where it will become a key competition area for financial institutions once profit margins on more traditional services are marginalised across competitors. Innovations that have been part of ordinary human life, such as social networking applications, have been monetising data for many years, and more recent data market innovations such as the abovementioned SQREEM have begun to specialise in consumers' digital footprints alone.

The matter of data monetisation in fintech's near future is not one of whether or not it will happen, but rather how the data market topology is expected to change, e.g. many specialised competitors or large tech companies integrating data flows into their value propositions. As data becomes nothing short of a commodity, adding value wherever it flows, the securitisation thereof will also become more important. This is especially evident following the sets of circumstances in earlier fintech innovations, through which it has become clear that security in the financial system is a national security discussion much more than it is a personal privacy issue. The rapid digitisation of the financial industry means cybercrime and cyber-espionage become all the more valuable activities in geopolitical interactions between sovereignties. The risks associated with data security are underestimated across governments, individuals and private industries alike (Moody's, 2015). This matter of data security and monetisation is then an interesting discussion, although difficult to quantitatively measure, seeing as there are disparate possible outcomes across the two domains of data security and accompanying risks, and data monetisation and the accompanying efficiency. The consideration of these arguments makes it natural to consider next these risks and benefits from a regulatory perspective. Regulation has been able to provide stability toward financial activities in the past, and if global markets wish to gain more efficiency than they have to take on risk, then regulation must be part of the discussion on all levels of industry.

Considering recent developments, and the resulting likely near-future outcomes, of modern fintech from a regulatory perspective can be beneficial in reaping efficiency gains from technology, whilst ideally keeping financial stability intact. The discussion of regulation of modern fintech, however, is intuitively divisible into two sub categories of risk management in terms of operations on firm level, and risk management in terms of industry governance and inter-industry cooperation toward some predefined goal. The first to consider is firm-level risk management and regulation that pertains to the operations of a given firm, and the challenges faced by those affected by fintech in this domain. Considering first the way that financial operations evolved through time, it is discovered that during Fintech 2.0 in the 1990s and more recently post-2008 the development of quantitative finance and its application in firm operations have reformed managerial and operations structures, particularly during the adaptations toward structures such as the techniques under VaR models discussed in previous sections. The developments in these areas are understandable, considering the demand for systems to monitor compliance especially after the regulations brought forward from 2008. The development and maintenance of such systems in turn created demand for improved IT infrastructure and evolved digitisation thereby.

The digitisation and IT infrastructure in turn promoted innovation within regulation itself, giving rise to structured compliance efficiency improvement attempts called regulation technology (RegTech) (Butler & O'Brien, 2019). Looking forward, RegTech is expected to play a more pronounced role in industry attempts toward efficient compliance and improved stability (Larsen & Gilani, 2017). However, considering such phenomena on a much wider (industry and inter-industry) scale yields insights on more collective attempts toward comprehensive financial system regulatory frameworks as they pertain to financial technology improvements. This is a discussion where a global perspective is necessary, as some of the same relevant phenomena seem to have occurred in different areas around the world. From the EU, the path of Europe toward digitisation and "datafication" rests very strongly on regulatory practices (Zetzsche *et al.,* 2019). Globally, investments in RegTech in the first half of 2018 reached $1.4 billion, which exceeds the same metric for the whole of 2017 (Mondres, 2019). In the other global fintech superpower, the Chinese market, there seems to be leading research in incorporating industry-standard government-controlled RegTech with high levels of efficiency (Yao, 2018). Looking forward, with such developments taking place the world over, there seems to be a probable set of events in the financial services industry that may lead, through certainly an extent of variation, toward unparalleled efficiency in global financial technology infrastructure coupled with relatively high levels of financial stability and regulatory efficiency. The only remaining area in which to consider the financial system of the future is how users are expected to interface with finance.

Considering all of the areas discussed before in this section, the probability of significant efficiency gains in the global financial system seems material. Given such likely improvements across the four domains of finance & investment, payment system infrastructure, data security & monetisation and regulation technology, a set of circumstances is made possible where humans create a technological system of global finance boasting unparalleled efficiency coupled with unparalleled relative stability. Such an ideological (and even a much less ideological) argument for global finance, however, leaves serious questions about how consumers of financial services will interface with such a system.

This topic has been lightly touched upon with regard to the likelihood of biometrics being included somewhere in the value chain, but a more comprehensive discussion is needed. Considering briskly the past of consumers interfacing with finance, we know from preceding sections that more efficient and convenient interfacing solutions contribute significantly to market penetration and user adoption (e.g. the sections on ATM's and e-banking). The current state of conditions in this regard is also worth considering, where banking usage trends seem to be converging toward majority mobile and thus already heavily reliant on telecommunications infrastructure (Shareef *et al.,* 2018). Looking forward, the area of convenient and efficient customer interfacing is expected to become an increasingly major competitive area for service providers, whether they be traditional and supervised or fintech and agile (PWC, 2016; KPMG, 2018; Arner, 2017). An interesting situation to monitor will be how large tech firms utilise their market capitalisation and digital footprint data to penetrate the financial service interfacing models of the future (Van Loo, 2018).

The future fintech guarantees only one constant: change. The recent trends in the discussion above prove that the global financial industry is evolving at a faster pace than ever before, and that even the rate of change seems to be increasing. Considering the developments of such changes over time may lend insight into the relative likelihood of certain sets of circumstances ever becoming part of reality, but if there is one thing that can truly be said without hesitation, it is that the future of fintech is more uncertain than ever. The discussion of the fields (economic, public perception, political and regulatory) within which the global fintech machine performs its functions (finance & investment, payment system infrastructure, operations & compliance, data security & monetisation and consumer interfacing) gives a holistic overview of the possible and probable future outcomes of fintech.

Such a broad, holistic consideration ensures, however, only one thing: uncertainty. It is therefore a reasonable course of action to attempt to investigate fintech phenomena at a more granular level, in order to generate a finer understanding of a more focused area. Eventually, it will be useful to quantitatively express such a development of fintech, in order to shed light on how the financial professionals of the future may generate an *a priori* understanding of likely future developments. Better yet, to determine quantitatively, what a likely set of outcomes may be. However, before it is possible to begin with methods of quantitatively understanding any data, the scope of the discussion must be narrowed to a single, significant development in financial technology, preferably originating in recent years and that exhibits a meaningful dataset to analyse. This will enable all of the preceding discussions, considerations and conjecture to culminate in a single quantitative case study of fintech developments. A fintech development satisfying all of the requirements for such a case study to be meaningful is the invention of the cryptocurrency Bitcoin.

2.3 The background to Bitcoin

The Bitcoin transaction network is an autonomous computer application which programmatically preserves the integrity of value flows. It can therefore be defined as an open-source, peer-to-peer digital currency (Brito & Castillo, 2013).

However, Bitcoin is not the first protocol to solve the problem of providing digital assets with verifiable scarcity. In fact, more specifically, cryptographically-secured pseudonymous digital cash was already formally proposed by academia more than 30 years before the creation of Bitcoin (Chaum & David, 1983). Similar developments pertaining to the concept of digital cash included “Hashcash” by Adam Back (Back, 2002), “b-money” by Wei Dai (Dai, 1998), and “bit gold” by Nick Szabo (Szabo, 2008).

2.3.1 Satoshi Nakamoto

“Satoshi Nakamoto” is presumed to be a pseudonym for the individual or group of individuals who are responsible for the creation of the Bitcoin software. From the communication style of the Satoshi Nakamoto entity it does seem however that it might be a single individual. Satoshi acted as an individual even to the extent of actively engaging with the public on the Bitcoin forum with regard to technical information and modifications involving the original software protocol, as shown in Wallace (2011).

Soon after Bitcoin gained widespread prominence in the media, people began to speculate about the true identity of the Nakamoto character. The first round of speculation involved multiple suspects based merely on their previous engagements in such technology and the aforementioned high profile cryptography-for-digital-cash experts Wei Dai, Nick Szabo and Hal

Finney were directly confronted upon occasion, each denying categorically to have been involved in the creation of the Bitcoin protocol.

The investigations, however, did not cease. News companies like *The New Yorker* and *Fast Company* launched their own respective investigations into the identity of the Nakamoto character. Possible candidates (excluding Dai, Szabo and Finney) were, among others, Michael Clear, Vili Lehdonvirta, Neal King, Vladimir Oksman and Charles Bry. The most notorious turn up of these investigations was a matter of what is at best circumstantial evidence linking King, Oksman and Bry to the Bitcoin protocol. This came in the form of an encryption patent application filed by the three men a mere 72 hours before the aforementioned World Wide Web domain name “bitcoin.org” was registered. The technology filed in the patent application and that of Bitcoin were similar, and later-performed text analysis found a common phrase “… computationally impractical to reverse” in both of the documents concerned. All three of the abovementioned individuals denied involvement with Bitcoin.

Other notable individuals to have been accused of/accredited with the creation of Bitcoin include a well-known Japanese mathematician by the name of Shinichi Mochizuki (Nelson, 2013), the “Dread-Pirate Roberts”: Silk Road dark-net market regulator Ross William, and even a group of European financial technology industry experts.

The only conclusion to be drawn from a wealth of speculation ranging widely in absurdity is that the true identity of Satoshi Nakamoto remains a mystery, which is a thread that is more comprehensively discussed by O’Hagan (2016).

2.3.2 Creation

The first evidence of Bitcoin’s existence dates back to the 18th of August 2008, when the World Wide Web domain of “bitcoin.org” was registered, according to the Internet Corporation for Assigned Names and Numbers (ICANN). The domain is under “WhoisGuard Protection”, which means that the identity of the person who registered the domain is not public information. On the 31st of October of the same year, all subscribers to a mailing list called “*Cryptography – The Cryptography and Cryptography Policy Mailing List*” received a link to a paper authored by Satoshi Nakamoto titled “*Bitcoin: A Peer-to-Peer Electronic Cash System*”. In this paper, the author claimed to have generated what was then described as a “system for electronic transactions without relying on trust” (Nakamoto, 2008). On the 3rd of January 2009, the so-called “genesis block” of the Bitcoin transaction network was mined. The first open source Bitcoin client was released on the 9th of January 2009 which was hosted by the website “SourceForge.com”. This Bitcoin client will have allowed anyone in the world with an internet connection to become a participant in the Bitcoin network of nodes.

The “launch” of Bitcoin is discussed in depth by Cohan (2017), for which a brief account is given here. Satoshi Nakamoto mined the Bitcoin genesis block himself, which contained a block reward equal to 50 Bitcoins. The next step was to involve more nodes, so that the network could become self-sustaining. This involved Nakamoto uploading the software which enables any computer with internet access to become a Bitcoin node – referred to as a “Bitcoin Client” – to a website called *SourcForge.com.* The first person to download and run this software was the aforementioned Hal Finney. Shortly after, 10 Bitcoins were signed over from Nakamoto to Finney presumably as a reward for his participation. This was the first ever Bitcoin transaction (Cohan, 2017).

After the initiation of the network and the first transaction, the next big issue in the development was to determine a fair market value – or at least exchange rate - for it. Seeing as there had, at that stage, only ever been one transaction, and no market for buyers nor sellers of the “not-yet-a-currency”, this was a rather peculiar challenge. When they opened the world’s first crypto-asset exchange, the New Liberty Standard became the only entity who tried to solve the problem of finding a market value for Bitcoin (Filippi & Loveluck, 2016). They based their exchange rate on the amount of electricity it would take to run a computer capable of mining Bitcoin, and the fact that there will have been an ever-decreasing supply of Bitcoin. The combination of these factors allowed the New Liberty Standard to mathematically determine a market value as 1$ = 1309 BTC. This was, however, only a mathematical derivation of the value of Bitcoin, based on the amount of time it takes a computer to “mine”/unlock/create Bitcoins. From an economic perspective there had not at this stage of Bitcoin’s developments been any value flows through which to determine a relative value for the protocol.

On the 22nd of May 2010, a day now widely commemorated as “Bitcoin Pizza Day”, a man by the name of Laszlo Hanyecz made the first purchase of a good with Bitcoin when he bought two Papa John’s pizzas for 10 000 Bitcoins (Benjamin, 2011). According to various news sources including - but not limited to - Wilmoth (2018), the first documented purchase using Bitcoin happened like this: Hanyecz, who was at the time a Florida-based computer programmer, posted a message on the *bitcointalk* forum reading*: “I’ll pay 10,000 bitcoins for a couple of pizzas.. like maybe 2 large ones…*”. The full post can be found on the forum and is also preserved authentically by (Wilmoth, 2018). A British man took up Hanyecz’s offer and had two Papa John’s pizzas delivered to him in exchange for the 10 000 Bitcoins. This amount of Bitcoin at the height of its value would have exceeded R3 billion.

Throughout Bitcoin’s years of gaining followers (or at least awareness), some events were more significant than others. One of the first material value-jumps that Bitcoin underwent accompanied its feature on a moderately popular news and technology website called *Slashdot.org*. The Slashdot website had, at the time, a quite significant audience of technophiles. Concurrent with the article featuring Bitcoin, (over the 5 days following the event) the value of one Bitcoin increased ten-fold in comparison to the USD. The exact date of Bitcoin's feature on Slashdot was the 11th of July 2010.

Even though the New Liberty Standard was the first exchange to offer trade in Bitcoin, it was not the vehicle to take Bitcoin to the height of its relatively mainstream adoption. An exchange that did play an important role in the early days of Bitcoin value hikes, though, was Mt. Gox. Jed McCaleb, a computer programmer previously famous for the creation of the successful e-Donkey peer-to-peer network in the year 2000, was the man responsible for the Mt. Gox exchange. McCaleb was previously involved with multiple projects of a similar nature, which includes the online exchange for '*Magic: The Gathering'* playing cards. Due to his previous experience with the technology required to set up an exchange, it was presumably relatively easy for McCaleb to erect Mt. Gox for Bitcoin. After McCaleb created Mt. Gox, however, he soon found himself incapable of keeping up with the business's demands, and sold the 'mtgox.com' domain to Mark Karpeles on March 6, 2011. The following three years saw Mt. Gox grow to dominate the world of Bitcoin trading. Even though the Mt.Gox exchange was rising in value, the volatility of the Bitcoin price is something that cannot go unquestioned.

Looking back in time, it is evident that Bitcoin can be classified as a comparatively volatile asset (Dyhrberg, 2016). One of the first significant value drops (which is discussed here due to its influence upon the adoption and number of market participants) came shortly after the Bitcoin pizza purchase event. In August of 2010, vulnerability in the Bitcoin protocol was discovered and the first "crypto-hack" in a line of many was performed. The Bitcoin mining system was exploited to the value of 184 billion Bitcoins. This amount is absurdly high. In fact, it is approximately nine thousand times as many as the Bitcoin protocol will ever allow. Needless to say, it did not take long for developers and community members to realise what caused the vulnerability, and the integrity of the system was recovered by the next day. Although Bitcoin was back in action, the system had been provably compromised for the first time.

Despite the protocol's integrity being compromised as is explained above, the Bitcoin price (and market cap) continued to rise steadily. By November 2010, the BTC total market capitalisation exceeded $1 million for the first time in history. With the amount of Bitcoins in circulating supply at that point in time, this market cap figure meant that Bitcoin was valued around the $0.50/BTC mark.

The next major instalment in the Bitcoin growth story is the infamous *Silk Road* connection. Silk Road is an anonymous online drug market place which was first revealed in June 2011 (Barratt, 2012). The Silk Road marketplace was only accessible to people who make use of the *Tor* anonymising software, which is an application that hides the IP address of any computer through encryption. The Silk Road website, while no longer active today, was reported to have borne striking similarities to eBay. Goods and services were categorised on the website, and a wide range of illicit goods and services were openly available under the following categories: ecstasy, cannabis, dissociative, psychedelics, opioids, stimulants, benzodiazepines etc. The structure was similar to the likes of Amazon today, where buyers and sellers were each assigned reputation points by other participants in the market. These reputation ratings included points awarded based on the quality of their products, how fast they ship and their level of professionalism and discretion during the completion of the transactions. Silk Road and its accompanying dynamics are discussed in more detail by Van Hout & Birmingham (2014). The reason Silk Road is so relevant here is of course its payment methods. It is difficult to measure how much of Silk Road payment activity involved Bitcoin, but a wide range of sources seem to report evidence that a significant amount of these transactions were in fact performed on the Bitcoin blockchain (Barratt, 2012; Van Hout & Birmingham, 2013). It is, at this stage of the Bitcoin growth chronology, a matter that is equal parts difficult to prove and easy to justify. However, upon the arrest of the previously mentioned Ross Ulbricht a.k.a. the *Dread Pirate*, a full investigation was launched by, among other institutions, The Federal Bureau of Investigation (FBI) of the United States of America (USA). The investigation saw $28.5 million worth of Bitcoin seized from Ulbricht, revealed perhaps some idea of the scope of Silk Road trade and concluded with a life-long prison sentence for Ulbricht (Greenberg, 2013).

However immoral and detrimental to society the Silk Road drug trade industry was, during its development, Bitcoin saw a doubling in market value from the $0.50/BTC mark in November 2010 to $1/BTC in February 2011. The subsequent two months saw three new digital exchanges opening their virtual doors, supporting trade in both Fiat currency and Bitcoin (Moore *et al.*, 2018).

While the Silk Road, coupled with payments in Bitcoin, made a comparatively superior system of illicit substance trade possible, the issue of exchanging Bitcoins back to spendable Fiat currency aided the creation of multiple exchanges over a short period of time (Brandvold *et al.*, 2015). With exchanges that allow the transfer of fiat currencies to Bitcoin and back, and a large number of Bitcoin transactions implicated in illicit drug trade, Bitcoin exchanges soon elicited commentary such as "…the next virtual havens for money laundering and tax evasion" (Gruber, 2013). The unregulated nature of the Bitcoin exchange industry, however, carried its own wealth of risks with it, most notably in the form of "crypto-hacks". The first Bitcoin exchange hack had as its victim the aforementioned Mt. Gox. According to Trautman (2014), a hacker gained the credentials of an official auditor working for the exchange. This allowed the hacker access to the website's user database with corresponding email addresses and passwords. The hacker then used their access to place false offers to sell Bitcoins, driving the price down to just $0.01/BTC.

Mt. Gox reacted by reversing fraudulent transactions, halting trading and re-securing their systems. Attacks against specific user accounts were also launched to withdraw funds – using the credentials obtained in the initial hack. The resulting thefts amounted to over 4019 BTC. While subsequent crypto-hacks were to become an important theme, the Mt. Gox hack discussed here was the first significant hack that shook the then-nascent landscape. Even though software attacks may seem to significantly increase the risks associated with an asset, it did not stop other entities from proposing similar value propositions.

Around this time, other technologies started to emerge based on the Bitcoin open source computer code. This contributed to immense growth from a single software protocol to an international industry in the years to come.

2.3.3 Why Bitcoin was started

The reason for the creation of Bitcoin remains a subject of debate until today, a debate of which the main arguments will be summarised here. The fact that Satoshi Nakamoto is an anonymous entity, while creating a certain amount of curiosity amongst new users, could also raise some red flags for others. The anonymity may decrease trust in the protocol, as there is no way to trace the origins of hypothetical inconsistencies in the development of the system. While the effects of Nakamoto's anonymity are difficult to determine, it is certain that it contributes to the uncertainty of the protocol.

Satoshi Nakamoto purposefully intended to create a "trust-less" cash system. Nakamoto explicitly stated that the reason for his creation of the Bitcoin protocol is to remove the third party intermediaries from digital money transfers (Nakamoto, 2008). Financial intermediaries are responsible for many duties in an economy – one of which includes the establishment and ongoing protection of trust, according to Hurley *et al.* (2014). Naturally, therefore, if they are removed from the value flows of a digital transaction, some other concept must be relied upon to establish this trust. In the case of Bitcoin, it is consensus among nodes in the network which are incentivised in such a way that their actions to benefit themselves contribute to the proliferation of the network. Of course, in exchange for the functions that banks perform in the economy, they are also paid a premium which is then, in turn, used to pay for their expenses. Seeing as Bitcoin's expense profile is very different from that of a bank, the transactions on the Bitcoin network are facilitated through the mining process, and the cost to maintain the network is shared among participants and nodes who engage in mining blocks in exchange for block rewards (Nakamoto, 2008). The incentives for creating such a self-sustaining network without a central authority are oblique, which is why the convictions that lead Nakamoto to create the Bitcoin network are easy to question. To that end, Nakamoto published a post on a public forum of which the following is a short excerpt:

"*The root problem with conventional currency is all the trust that's required to make it work. The central bank must be trusted not to debase the currency, but the history of fiat currencies is full of breaches of that trust. Banks must be trusted to hold our money and transfer it electronically, but they lend it out in waves of credit bubbles with barely a fraction in reserve."* – Nakamoto (P2Pfoundation, 2009).

This forum post, in conjunction with Nakamoto (2008), may shed some light on Nakamoto's personal convictions for the creation of the Bitcoin protocol and concludes this section of chapter 2. The focus can now be shifted toward a logical build-up to the empirical evidence of fintech-associated market complexities in the market for Bitcoin. First however, a background to asset valuation will be provided, so that these market complexities can be discussed concurrently with relevant empirical evidence of how they have been utilised toward speculative wealth creation.

2.4 Background on asset valuation

The historical origins of investment go as far back as barter itself, with evidence of a legal framework for establishing collateral having been found in the Code of Hammurabi, a document dating back to about 1754 BCE (Hooker & King, 1996). The framework has been described by Dyneley (1904) to lay out the rules for establishing the means for a pledge of collateral by codifying debtor and creditor rights to pledged land. These are surprisingly sophisticated asset management techniques considering contemporary technologies. Investment and asset valuation should, therefore, be seen as two things that developed together, because asset valuation is an integral part of investment. The investment decision in its most basic form is broken down into two factors known as *risk* and *return* (Farrar, 1962). In order for the investor to consider risk vs. return, however, each must first be identified and measured. This is where a multitude of techniques can be used to determine the value of an asset. As described by Myers (1984), the process of asset valuation informs the investment decision, and is therefore an integral part of the financial world.

2.4.1 Non-speculative asset valuation

According to Damodaran (2012), there are three broad classifications of approaches to the valuation process. These are: firstly, to consider the value of an asset a function of its discounted cash flows; secondly, to consider the value of an asset a function of the value of other assets (relative valuation); and thirdly, to consider the value of an asset a function of a contingent claim. This is a problem when one considers the valuation of Bitcoin, as it has very volatile and relatively unpredictable cash flows (Dyhrberg, 2016), low- or no correlation with traditional assets (Baur *et al.*, 2018) and no market for contingent claims (Chen *et al.,* 2018).

Results obtained with these otherwise-highly effective methods that Damodaran (2012) so meticulously reviews, are therefore – in the case of Bitcoin – experimental at best. However, Bitcoin's unique lack of properties with which to perform accurate mathematical valuations does not stop us from experimenting with the literature for speculative valuation.

2.4.2 Speculative asset valuation

In the realm of active investment and speculation, the valuation techniques employed by speculators are a little different from those used in investment banking and wealth creation. In fact, when trying to predict the future of the stock market, the challenge of valuation changes to *forecasting* (Penman, 2010). The Efficient Market Hypothesis (Fama & Malkiel, 1970) becomes the central theory, and the main goal is no longer to accurately value assets, but to forecast probability of certain price movements in the future.

When attempting to forecast future prices, it is possible to vary the attention paid toward the underlying nature of an asset. If a given time series is used as input to an arbitrary forecasting technique, the model in use can either be set up to attach a weight to the fundamental nature or not. The methods used in speculative wealth creation and active investment can therefore be intuitively divided into two broad categories: financial data analysis and time series data analysis, both of which are discussed below in terms of their viability to act as a solution to the problem statement of this research. It should be clarified that this is not a formal categorisation among academia, but rather a conceptual split to aid in the logistics of this discussion.

2.4.2.1 Financial data analysis

This section is concerned with the process of analysing data from the principle standpoint that it is of a financial nature, as opposed to merely considering data a 'raw' time series whilst paying no concern to its financial nature (which is discussed in the next section).

When performing speculative analysis in an active investment context, the methodologies available to the investor can be classified into two broad categories. These are fundamental- and technical analysis (Mpofu *et al.*, 2016). Both of these topics are treated here in terms of their viability to act as solutions toward the problem statement of this research.

2.4.2.1.1 Fundamental analysis

Fundamental analysis is aimed toward determining the intrinsic value of corporate securities by a careful examination of key value-drivers, such as earnings, risk, growth and competitive position (Lev & Thiagarajan, 1993). The true origin of fundamental analysis is a difficult concept to define, for merely "thinking of the stock market in a way that pertains fundamentally to value" perhaps precedes even the structured, regulated stock market itself, but the formal publication of fundamental analysis as a speculative trading strategy (the context in which it is discussed here) traces back to an individual who Kaza (2000) calls the “father of fundamental analysis”: Benjamin Graham (Spooner, 1984). The work that Kaza (2000) refers to when calling Graham the father of fundamental analysis is actually a joint effort by Graham and Dodd (1934), and while Graham was an active investment fund manager and a lecturer at Columbia University at the time, Dodd was an Assistant Professor at the same institution. It is, however, still Graham alone who is attributed the title of “father of fundamental analysis” by Kaza (2000) and Spooner (1984). The reason for this is a single section in the book by Graham & Dodd (1934) that describes a model of arbitrage which was developed solely by Graham. The “general formula for analysing potential arbitrage situations” is described in Graham and Dodd (1934) as follows:

$$\frac{C \times G - L(100\% - C)}{Y \times P}, \tag{2.1}$$

where G is the expected gain in the event of success; L is the expected loss in the event of failure; C is the expected chance of success expressed as a percentage; Y is the expected time of holding in years; and P is the current price of the security. The formula equates to the “annual potential gain from an arbitrage” according to Kaza (2000).

This investigation of arbitrage (as a means to create speculative wealth for the investor) led to further inquiries concerning intrinsic value for many years to come. Consequent inquiries arising from intrinsic value perceptions served as direct competition to the Efficient Market Hypothesis, according to Buffet (1988). . In a letter Warren Buffet wrote to the shareholders of Berkshire-Hathaway as chairman of the private wealth management company (in 1988), he stated that the Efficient Market Theory (EMT) had become an “almost-holy scripture” in academic circles during the 1970s. He went on to state that the market was “frequently efficient” but that it would be a mistake to think that it is *always* efficient.

The inter-relational dynamic between the EMT and fundamental analysis continued on for many years to come, with two opposing schools of thought emerging. From this 50 year period of dissonant conjecture, during which neither school-of-thought was able to prove the other wrong, it must be concluded that the field of fundamental analysis relies on a significant amount of subjective reasoning. While this is no problem for the active investor seeking to gain profit from speculation based on subjective reasoning, it renders the field of fundamental analysis incapable of solving the problem statement of this research. However, fundamental analysis has been proven beyond arbitrage.

Fundamental analysis extends far beyond the anecdotal (or even scientific) success of arbitrage vs. the EMH. It also branches, as per the definition by Lev and Thiagarajan (1993), into the process of considering financial statements as input for trade decisions. Financial statements offer significant decision-making inputs for investors, allowing for the calculation of a range of measures. These measures range from ratio and cash flow measures to earning predictions and equity valuation (Subramanyam, 2009). Financial statement analysis also enables the analyst to “reconstruct the economic reality embedded in financial statements” (Subramanyam, 2009).

While this may be true and is an absolutely significant factor for the speculative investor who invests in traditional assets, it may be a premise of adding value that could be challenged in the future. The reason for this is, of course, that technological advances may continue to reinvent financial systems to such an extent that financial statements are no longer available for ‘modern’ assets, or that are change so dynamic and change so rapidly that inputs to investment decision-making models become corrupt in a traditional sense. If, for example, a price-earnings ratio is considered as an arbitrary financial ratio, what does the fundamental analyst do if no earnings data is published, or if the definition of earnings is modified by some technological leap of the future? The mere fact that these financial ratios are not even transferable across traditional asset classes already proves that they are too subjective and asset-specific to be considered a viable solution to this research.

Fundamental analysis and the effectiveness thereof is, therefore, not questioned nor challenged here. It is just considered a method of analysis that fails to perform robustly across asset classes, not objectively measurable and depends too heavily on the subjective opinion of the investor. Fundamental analysis is however not the only tool used by analysts to assess the relative value of assets. The other major form of analysis performed by analysts is technical analysis. Technical analysis is considered in the next section.

2.4.2.1.2 Technical analysis

Forecasting techniques belonging to the field of technical analysis are, like fundamental analysis, among those used in speculative investment (Mpofu *et al*., 2016). The field has a recorded modern history of longer than a hundred years, which is discussed below after the introduction and a formal definition is given. These sections will be followed by a review of the success rates of some of these techniques, after which the topic is placed back in the context of the current research question. Finally, technical analysis is considered for its viability to act as a solution to the problem statement of this research.

2.4.2.1.2.1 Introduction and definition

A simple, broad and modern definition of technical analysis, according to Kirkpatrick and Dahlquist, (2010), is that it is a study of prices in freely traded markets with the intent of making profitable trading or investment decisions. The authors continue to describe technical analysis as an instrument that has its roots in basic economic theory, referring to the underpinnings of Edwards *et al.*, (2007), who delineates the assumptions of technical analysis as:

- stock prices are determined solely by the interaction of demand and supply;
- stock prices tend to move in trends;
- shifts in demand and supply cause reversals in trends;
- shifts in demand and supply can be detected in charts, and;
- chart patterns tend to repeat themselves.

While this gives a satisfactory understanding of the process and perhaps even scope of technical analysis, a more formal definition is suitably given by Scott *et al.* (2016) as the study of data generated by the action of markets and by the behaviour and psychology of market participants and observers. The authors elaborate that such study is usually applied to forecasting – that is, estimating the probabilities for the future course of prices for a market, investment, or speculation by interpreting the data in the context of precedent. In order to consider technical analysis as a possible solution to the problem statement of this research, the history and origin thereof must first be considered.

2.4.2.1.2.2 History and origin

The modern documented history of technical analysis can be traced back to the turn of the 18th century, with Charles Henry Dow the founder of its underlying premise: *Dow Theory* (Brown *et al.*, 2002). It should be qualified that earlier hints of something preceding modern documented technical analysis appear in Joseph de la Vega's accounts of Dutch markets in the early 17th century (Corzo, 2014). Early proof of what is today known as the *candle stick* method was also found in the writings of Homma Munehisa, a rice merchant of 18th century Asia, who is closely associated with the emergence of a futures market for rice in 1710 (Kamo & Dagli, 2009).

The collective and more comprehensive body of work accredited as the "origin of modern technical analysis", was however originally only intended to be editorial notes to the *Wall Street Journal*. It is for this journal that Charles Dow was the editor for a total of 13 years (1889-1902), and as such wrote notes for the journal. In his weekly writings, Dow never once used the term "Dow Theory", and no intention is clear that he aimed for his work to become the comprehensive field of forecasting techniques that it is today. After Dow's death in 1902, his successor as *Wall Street Journal* editor William Peter Hamilton continued his work, which was at the time still only known as a series of *Review and Outlook* columns in their newspaper. As Edwards *et al.* (2018) explains, most of what is known as Dow Theory of stock market movements today, comes not from the founding editor of *The Wall Street Journal* (Charles H. Dow), but from his successor, William Peter Hamilton.

In 1922, after serving as editor for the Wall Street Journal for 20 years, Hamilton published *The Stock Market Barometer,* wherein he refers to his work as a “theory, a working hypothesis if nothing more”. This was the first time the ongoing work of Charles H. Dow and William Peter Hamilton was labelled and packaged as a theory, and Hamilton cited Dow explicitly throughout his 27 years of editor’s letters that pertained to “charting” (Bishop *et al.*, 2018). This body of work eventually caught the attention of later so-called “Dow Theorists”, Robert Rhea, Richard Shabacker and George Schaefer, among others. Toward the end of Hamilton’s life, his brother in law, Robert D. Edwards, joined him in active research and documentation of technical analysis, and Edwards eventually finished a piece of writing primarily written by Hamilton with the help of John Magee (Scott *et al.*, 2016). The work was published under the names of Edwards and Magee in 1948 as a book titled *Technical Analysis of Stock Trends,* and extensively and explicitly referenced Hamilton as chief contributor (Edwards *et al.*, 2007). According to Scott *et al.* (2016), the 1960s saw wide and popular use of charts and technical analysis as input for speculative decision making, and more mathematically sophisticated techniques began to enjoy attention as opposed to the process of visually examining charts in order to identify market phenomena.

Seeing as mathematical and statistical methods of analysing financial data is discussed extensively as a stand-alone concept in the following sections and subsections, this concludes the historical overview of technical analysis as an isolated concept. The next area of consideration is the field of time series analysis as an alternative or a complement to the groups of techniques discussed above.

2.4.2.2 Time series data analysis

A time series is a discrete time, continuous state process where time is a primary axis, and is structured as a set of certain discrete points spaced out over uniform intervals (Ullah, 2013). It is possible to perform analysis and forecasts on data in this form without paying any consideration to its underlying nature, for example whether it is financial data, medical data, weather data, etc. If the price of an asset is the variable under consideration, is presented in a time series, and no measurement representing its financial nature is considered, it becomes a pure time series prediction problem.

Time series forecasting techniques all have the same goal, which is to solve the problem of creating a statistically-methodised window into the future (Brockwell & Davis, 2002). However, these techniques approach the problem from some fundamentally different angles. Researchers have thus categorised these techniques into so-called “families”, depending on their fundamental starting point toward solving the forecasting problem. These families include, but are not necessarily limited to, Autoregressive (AR), Integrated (I), Moving Average (MA) models, Autoregressive Conditional Heteroskedasticity (ARCH) models, Vector Autoregressive (VAR) models, Spectral analysis models and State-space models. All of these families of models will be considered below, based on their ability to address the problem statement of this research.

2.4.2.2.1 Autoregressive Models

An autoregressive model is a statistical method to represent a stochastic process. It is used to describe time-varying processes across domains. The autoregressive model specifies that the output variable depends linearly on its own previous values (Wong & Li, 2000).

The simplest and most basic method of time series forecasting, often used as a benchmark for more complex models, is called the *Naïve* approach. The Naïve approach sets the value of the dependent variable equal to its previous value in the series, formally:

$$\hat{y}_{t+1} = y_t, \tag{2.2}$$

where $\mathrm{y_t}$ represents the value of variable y at the current point in time, and $\hat{\mathrm{y}}_{\mathrm{t+1}}$ represents the estimated value of variable y one time step ahead of the current point in time.

Setting a given value in a time series equal to its previous value may seem like a meaningless task, but equation (3.1) sets the standard upon which complexities are easier to add, as it can be seen as a regression equation with a β-value (regression coefficient) of 1 and no error term - no constant is included for simplicity (Asteriou & Hall, 2007).

Extending the Naïve approach to an AR model of order *p* (the number of lagged independent variables the model will have) involves the inclusion of coefficients and an error term as such:

$$Y_t = \sum_{i=1}^{p} \emptyset_i Y_{t-i} + u_t, \tag{2.3}$$

where p represents the number of lagged independent variables, $\emptyset$ represents the regression coefficients, and u the error term.

2.4.2.2.1.1 The AR(*p*) model and stationarity

A key concept underlying time series processes is stationarity. A time series is stationary when the following three conditions are met:

(a) $E(Y_t)$ = constant for all t;

(b) $Var(Y_t)$ = constant for all t; and

(c) $Cov(Y_t, Y_{t+k})$ = constant for all t and all $k \neq 0$;

or if its mean, variance and covariances remain constant over time (Wise, 1965).

According to Asteriou & Hall (2011), stationarity is so important to classical regression analysis, that results estimated based on non-stationary time series can be classified as "spurious" or meaningless. Stationarity is a requirement for AR(*p*) models, to ensure that temporary shocks are analysed in isolation, as any time series can be expressed as the summation of all of its previous shocks.

2.4.2.2.2 Moving average models

Moving average models can be divided into two main categories, namely those concerned with the past values of the dependent variable itself, and those concerned with the past values of the error terms. A discussion of the former follows below, after which the latter and finally the combination of AR and MA (ARMA) models.

2.4.2.2.2.1 MA models concerned with past values of the dependent variable

The moving average is a mathematical operation which creates a series of averages for a given point in time (Vandewalle *et al.*, 1999). The method creates a series of averages from data preceding the 0th time interval (t).

Below is the mathematical operation for what is known as the simple average, thereby equating the expected value of a given point in time to the average of all the values that come before it, formally:

$$\hat{y}_{t+1} = \frac{1}{x}\sum_{i=1}^{x} y_i. \qquad (2.4)$$

This method is applicable in cases where the time series in question fluctuates around a constant mean value, that is to say in a dataset where there is no seasonal or trend component present. However, if a time series exhibits a trend or seasonal component in its average, the simple average will need to be replaced by a moving average. Formally, a moving average model can be estimated as follows:

$$\hat{y}_i = \frac{1}{p}(y_{i-1} + y_{i-2} + y_{i-3} \dots + y_{i-p}). \tag{2.5}$$

The moving average model's accuracy depends greatly upon the chosen value for p – the number of previous observations from which the average is calculated.

In order to build more sensitivity toward different values of p into the model, the moving averages of previous time steps can be associated with a weight, by which the model becomes a weighted moving average model, formally:

$$\hat{y}_i = \frac{1}{m}(w_1 * y_{i-1} + w_2 * y_{i-2} + w_3 * y_{i-3} \dots + w_m * y_{i-m}), \tag{2.6}$$

where w represents the weight attached to the given time step value of y. Through the inclusion of a weight parameter to the moving average, we can develop a mathematical preference for values that are more recent with respect to y_i . However, the rate at which the weights are determined can also be refined through what is called simple exponential smoothing, formally defined as follows:

$$\hat{y}_{t+1|t} = \alpha * y_t + (1 - \alpha) * \hat{y}_{t|t-1}, \tag{2.7}$$

where $\propto$ represents a smoothing parameter. If the concept of exponential smoothing is incorporated into a model's specification, it assigns exponentially decreasing weights over time. Models that are specified with exponential smoothing tend to outperform those without – especially when seasonality is present in the data. However, data structures often exhibit level and trend fluctuations. The *level* is the average value in the series, where the *trend* is the general pattern of prices that can be observed within a given time period. Holt (1987) formally expressed a model incorporating both of these phenomena as follows:

$$\hat{y}_{t+h|t} = l_t + hb_t, \tag{2.8}$$

$$l_t = \alpha y_t + (1 - \alpha)(l_{t-1} + b_{t-1}), \text{ and} \tag{2.9}$$

$$b_t = \beta * (l_t - l_{t-1}) + (1 - \beta)b_{t-1}. \tag{2.10}$$

Holt further added to the effectiveness of the linear trend model by including a seasonal component. The model was called Holt's Winter Model and is formally expressed as follows:

$$L_t = \alpha(y_t - S_{t-s}) + (1 - \alpha)(L_{t-1} + b_{t-1}), \tag{2.11}$$

$$b_t = \beta(L_t - L_{t-1}) + (1 - \beta)b_{t-1}, \tag{2.12}$$

$$S_t = \gamma(y_t - L_t) + (1 - \gamma)S_{t-s}, \tag{2.13}$$

$$F_{t+k} = L_t + kb_t + S_{t+k-s}. \tag{2.14}$$

2.4.2.2.2.2 MA models concerned with past error terms

The moving average model can also be applied to error terms of a time series. The simplest moving average model concerned with past error terms is that of order one, also known as the MA(1) model, estimated as follows:

$$Y_t = u_t + \theta u_{t-1}. \tag{2.15}$$

The implication in (2.15) is thus that Y_t depends on the value of the immediate past error. The simple MA(1) model above can be extended to include regression coefficients for *q* amount of lagged error terms (and is then known as an MA(q) model) as follows:

$$Y_t = u_t + \sum_{j=1}^{q} \theta_j u_{t-j}, \tag{2.16}$$

where *j* is the space between the first relevant lagged dependent variable in the model and the current point in time.

2.4.2.2.2.3 The MA(*q*) process and invertibility

Invertibility is to the moving average what stationarity is to the autoregressive process. A time series is said to be invertible if it can be represented by a finite-order MA or a convergent autoregressive process (Asteriou & Hall, 2011).

In order to understand why this is of importance to the problem statement of this research, consider for a moment the representation of the AR(∞) process, where the most recent error can be written as a linear function of current and past observations:

$$u_t = \sum_{j=0}^{\infty} (-\theta)^j Y_{t-j}. \tag{2.17}$$

where it is posited that for an invertible process, $|\theta| < 1$, and so the most recent observations in the series have higher weights than observations from the more distant past. However, if we consider the case where $|\theta| > 1$, the weight coefficient for each term increases as the lags increase, which means that more distant observations are in time, the greater their influence become on the current error. Lastly, when $|\theta| = 1$, the weight coefficients are of constant size, which means that observations distant in time have the exact same effect on the current error as the most recent observation and everything in between.

2.4.2.2.3 ARIMA models

After the presentation of AR(*p*) models and MA(*q*) models, combinations of these two families should be considered. These are called ARMA(*p,q*) models, and are estimated in the following general form:

$$Y_t = \sum_{i=1}^{p} \emptyset_i Y_{t-i} + u_t + \sum_{j=1}^{q} \theta_j u_{t-j}. \tag{2.18}$$

In the ARMA(*p,q*) family of models, the conditions for stationarity pertain only to the AR(*p*) part of the model. Similarly, the condition for invertibility pertains only to the MA(*q*) part of the model. However, ARMA(*p,q*) models can only be specified for time series that are stationary, and since it is common for financial data to exhibit trend characteristics, ARMA(*p,q*) models need to be *integrated* in order to deliver meaningful results.

In order to induce stationarity to a time series, the data is *de-trended* through the application of a process called *differencing* as follows:

$$\Delta Y_t = Y_t - Y_{t-1}. \tag{2.19}$$

The equation (above) produces the first differences of a time series. According to (Asteriou & Hall, 2011), most economic and financial data show trends to some degree, and it is nearly always necessary to take first differences of a time series before continuing with estimation.

If a time series is stationary after performing operation (above, 2.19), the series is said to be integrated of order one and denoted by *I*(1). If after first differencing the series is not yet stationary, second differences must be taken as follows:

$$\Delta\Delta Y_t = \Delta^2 Y_t = \Delta Y_t - \Delta Y_{t-1}. \tag{2.20}$$

If the time series is stationary after second differencing (above, 2.20), it is integrated of order two and denoted by *I*(2). A time series is thus integrated of order *d* if it must be differenced *d* times in order to achieve stationarity, and is thus denoted by *I*(*d*).

We can therefore specify the model combining AR terms, MA terms and an integrated time series generally as: ARIMA(*p,d,q*) – where *p* is the number of lagged AR terms, *q* the number of lagged MA terms and *d* the number of times differencing had to be performed on the original series in order to achieve stationarity. Non-stationarity is often coupled with conditional properties of variance. This is referred to as heteroscedasticity, and is possible to model through the use of ARCH models.

2.4.2.2.4 Modelling variance with ARCH-GARCH models

Conventional econometric analysis operates under the assumption of homoscedasticity, which means that the variance of the disturbance terms is assumed to be constant over time (Asteriou & Hall, 2011). However, financial markets tend to exhibit periods of high and low volatility (Figlewski, 1997). This is precisely the type of problem that Engle (1982) addressed in the United Kingdom's inflation rate fluctuations of the 1980s and before, work which lead to the extensive development of advanced volatility modelling.

Engle's original procedure was called Autoregressive Conditional Heteroscedasticity (ARCH), and involved the assumption of autocorrelation among error terms over time. Since Engle's original breakthrough, many variations of ARCH models were developed, all with a slightly different approach to modelling variance (volatility in financial terms). The developments arising from Engle's original ARCH model include (but are not nearly limited to) Generalised ARCH, Exponential Generalised ARCH and Threshold Generalised ARCH. ARCH and GARCH are discussed here, after which all possible variations of ARCH models are considered holistically as per their ability to address the problem statement of this research.

Let us now consider Robert Engle's original idea of modelling volatility clustering through conditional heteroscedasticity, after which various complexities will be relatively easy to add into the equation. Engle's model suggests that the variance of the residuals at time t depends on the squared error terms from past periods. Engle suggested that modelling the variance and the mean of a series simultaneously will result in superior forecasting accuracy, the suspicion that conditional variance is not constant (Engle, 1982). If we consider a simple model:

$$Y_t = a + \beta' X_t + u_t \tag{2.21}$$

where X_t is a $\mathrm{k} \times 1$ of explanatory variables and β is a $\mathrm{k} \times 1$ vector of coefficients, it is possible to consider the conventional CLRM assumptions with specific attention paid to heteroscedasticity. Under CLRM assumptions, we would assume that u_t is independently distributed with a zero mean and constant variance σ^2.

Engle's idea starts off by letting the variance of the residuals (σ^2) depend on history, or similarly, to exhibit heteroscedasticity, because the variance will change over time. The simplest way to model this is to express the variance as dependent on *q* lagged time step of the squared error terms. The ARCH(*q*) model thus simultaneously models the mean and the variance of the series with the following specification (given equation above (2.22)):

$$u_t|\Omega_t \sim iid\ N(0, h_t), \tag{2.22}$$

$$h_t = \gamma_0 + \sum_{j=1}^{q} \gamma_j u_{t-j}^2, \tag{2.23}$$

where (2.22) is the mean equation, (2.23) is the variance equation, $\Omega_t$is the information set and $h_t = \sigma_t^2$.

One of the drawbacks of the ARCH specification, according to Engle (1995), was that it looked more like a moving average specification rather than an autoregressive one (which was his original intention). From this, Engle confirmed the research in *Generalised Autoregressive Conditional Heteroskedasticity* (GARCH). What made GARCH different from the original ARCH specification is that it includes lagged conditional variance terms as autoregressive terms. The general GARCH(*p,q*) model is specified as follows:

$$Y_t = a + \beta' X_t + u_t, \tag{2.24}$$

$$u_t|\Omega_t \sim iid\ N(0, h_t),$$

$$h_t = \gamma_0 + \sum_{i=1}^{p} \delta_i h_{t-i} + \sum_{j=1}^{q} \gamma_j u_{t-j}^2, \tag{2.25}$$

According to Asteriou & Hall (2011), this model specification usually performs very well and is easy to estimate because it has only three unknown parameters: $\gamma_0, \gamma_1$and δ_1. ARCH-GARCH models evidently possess the ability to model variance to a meaningful extent. However, they are not an appropriate family of models for the comparison performed in this study. This is because ARCH models do not lend themselves toward forecasts off of univariate time series sequences, but are rather expected to serve as an effective independent layer functioning around a baseline conditional classification problem. With that, the focus can now be shifted toward the VAR family of models.

2.4.2.2.5 VAR models

In the period leading up to Christopher Sims' publication of VAR models in 1980, statisticians and econometricians became increasingly concerned with what is now known as "the identification problem" or, more specifically, "non-fundamentalness" (Alessi *et al.*, 2008). Lutkepohl (2014) defines a shock to be non-fundamental if it cannot be retrieved as a forecast error from the observed variables in a given model. This definition seems to be well accommodated in the notion that the econometrician may not have all the information that individual agents may have (Hansen & Sargent, 1980, 1991; Lippi & Reichlin, 1993; Sargent &

Watson, 2007). It is therefore a reasonably popular argument that non-fundamentalness is an important factor to consider when forecasting accuracy is desired to be robust.

Even though the matter of non-fundamentalness is well-argued today, it was still an unnamed concept back in 1980. the year that Christopher Sims published his critique on the identification problem and simultaneously his work on VARs. Classic examples of this type of critique toward the forecasting model identification *status quo* pre-1980 is for example Boland's (1968) reconsideration of the identification problem and the validity of economic models, Sims' (1972) investigation of money, income and causality (a natural precursor to his publication of VAR's), Wallis' (1977) attempt to define the "final form" of econometric models and Hsiao's (1979) investigations of causality tests in econometrics. While this is by no means an exhaustive list of the pre-1980 discussions of the identification problem, it should indicate that there were – at the time – significant discussions in the literature about the opportunity of constructing modelling techniques that are non-fundamental.

These extensive attempts toward finding modelling techniques that are less dependent on prior estimation drove researchers toward the solution of the identification problem, which culminated in the eventual publication of Sims (1980). In his paper titled *Macroeconomics and Reality,* Sims heavily criticised contemporary identification techniques (Asteriou & Hall, 2011). These contemporary identification techniques involved the use of simultaneous equations to model econometric phenomena, which required clear prior identification of endogenous and exogenous variables – which was precisely Sims' point of concern. Furthermore, Sims argued that "the style in which identification is achieved is inappropriate, to the point at which claims for identification in these models cannot be taken seriously". To Sims, it was all too theoretical, and he argued therefore that the models of the time were not credible, due to their deeply theoretical nature. Sims' issue was with the strict *a priori,* which was a problem he set out to solve.

Sims (1980) formally proposed a Vector Autoregressive Model (VAR), specified as follows:

$$y_1(t) = a_1 + w_1 11 * y_1(t-1) + w_1 12 * y_2(t-1) + e_1(t-1), \tag{2.26}$$

$$y_2(t) = a_2 + w_1 21 * y_1(t-1) + w_1 22 * y_2(t-1) + e_2(t-1), \tag{2.27}$$

vectorised into a single equation:

$$\begin{bmatrix} y1(t) \\ y2(t) \end{bmatrix} = \begin{bmatrix} a1 \\ a2 \end{bmatrix} + \begin{bmatrix} w11 & w12 \\ w21 & w22 \end{bmatrix} \begin{bmatrix} y1(t-1) \\ y2(t-1) \end{bmatrix} + \begin{bmatrix} e1(t) \\ e2(t) \end{bmatrix}, \tag{2.28}$$

where: $\mathrm{y}_1$and y_2 are independent variables, a are constants, w are weights and e are error terms.

To any individual familiar with regression analysis under the classic linear regression model (CLRM) framework, it should be evident that - in the VAR model specification - each variable is simply expressed as a linear function of the past values of itself, and past values of all of the other variables in the dataset (Canova, 1999).

Toward the consideration of VAR models as a viable candidate in terms of solving the problem statement of this research, their benefits and detriments are summarised shortly below, with reference to (Asteriou & Hall, 2011):

Benefits	**Detriments**
Very simple identification	Atheoretic (Detached from Theory)
Very simple estimation (OLS)	Potential for loss of degrees of freedom
Relatively accurate forecasts compared to far more complex simultaneous equation models (Mahmoud, 1984; McNees, 1986)	Coefficients can be difficult to determine

It must be said at this stage that even though the atheoretic nature of VAR models is considered a detriment by (Asteriou & Hall, 2011), it will serve as an advantage to the robustness of a model, and therefore it rather drastically strengthens the case for their use here. This is because moving away from strict *a priori* is expected to contribute more toward a modelling solution that satisfies the requirements of the problem statement.

2.4.2.2.5.1 VAR models and causality

The atheoretic nature of the identification process pertaining to VAR models carries with it a slight complication, which is that it makes causality a necessary topic of concern when considering them to solve a forecasting problem (not taking away from the importance of causality measures in other specifications). Since initially there are no restrictions on any parameters in the identification procedure, in effect ‘everything causes everything’ (Asteriou & Hall, 2011). For this reason, causality tests become imperative the moment that VAR estimation attempts to become VAR inference (a problem not limited to this domain).

2.4.2.2.5.1.1 Granger causality

Upon consideration of causality there exist statistical tests, of which two are discussed here. Granger (1969) proposed an explicit and testable causality framework that is still in use today. A variable y_t is concluded to *Granger cause* x_t if a prediction of x_t including past values of y_t

delivers a superior result in accuracy compared to such a prediction excluding past values of y_t, *ceteris paribus.*

2.4.2.2.5.1.2 Sims Causality

Sims (1980) proposed an alternative causality test to Granger. Considering once again the possibility of causality between two variables y_t and x_t, and this time arguing from the premise that the future cannot cause the present, Sims proposes to estimate a VAR model with the following specification:

$$y_t = a_1 + \sum_{i=1}^{n} \beta_i x_{t-1} + \sum_{j=1}^{m} \gamma_j y_{t-j} + \sum_{\rho=1}^{k} \zeta_\rho x_{t+\rho} + e_{1t}, \quad (2.29)$$

$$x_t = a_2 + \sum_{i=1}^{n} \theta_i x_{t-i} + \sum_{j=1}^{m} \delta_j y_{t-j} + \sum_{\rho=1}^{k} \xi_\rho y_{t+\rho} + e_{2t}, \quad (2.30)$$

The approach with the above estimation, is that there are leading values of x included in the first (2.29) equation and leading values of y in the second (2.30) equation. If we consider the first (2.29) equation: if y_t causes x_t then it is logical to expect some form of relationship between y and the leading values of x. This means that we test for $\sum_{\rho=1}^{k} \zeta_\rho = 0$, and if it is rejected, causality is concluded to run from y_t to x_t, and not vice versa, since the future cannot cause the present.

In the process of exploring the various families of models that have been discussed in this section so far, it has become more evident that there lacks a certain versatility in the way that these families of models are able to approximate a function between two variables that correlate. The problem is that the models described so far rely on a linear equation being fit to the data. What if, however, it was possible to learn a function that is any combination of functions available to mathematicians, in terms of the way that two variables correlate with each other? This is, in fact, possible with non-linear techniques such as machine learning.

2.4.2.2.6 Machine learning

Forecasting techniques like machine learning have gained popularity in recent years because they are becoming an increasingly capable field, with abundant data and relatively abundant computational power compared to the era of its inception (Brownlee, 2013). According to Heaton *et al.,* (2016), methods relying on machine learning already outperform humans on tasks like language manipulation, image labelling and games of skill, and thus a deep learning neural network should also be able to find the relationship determining an asset's return no matter how complex or nonlinear.

A brief historical overview of the development timeline for Machine learning is outlined in this section, after which the available categories of forecasting techniques are discussed. Finally, these categories are considered in terms of their viability to solve the problem statement of this research.

2.4.2.2.6.1 A brief historical overview of machine learning

Machine learning (ML) find its roots long ago in history partly due to the fact that a significant portion of ML techniques have their underlying probabilistic theoretical premise in Bayes' theorem, which dates back to the work of Thomas Bayes himself (Bayes, 1763). Of course much of the ML process also involves data fitting, which traces back to the concept of an abstractly formulated state space in the analytical procedure which traces back to the Markov Chain (cited by Hayes, 1913).

However the first time that a tangible device unified the three concepts in the above paragraph was when Alan Turing proposed the so-called 'learning machine' which was originally designed to crack the Enigma code (encryption used by Germany in World War II). In Turing (1950), Alan Turing famously said: "*I propose to consider the question, Can machines think?"* This is then perhaps a suitable body of work to consider the origin of machines that have the ability to learn.

Turing was not far ahead of Marvin Minsky and Dean Edmonds, who built the first ever neural network, called the "Stochastic Neural Analog Reinforcement Calculator" (SNARC) in 1951 (Crevier, 1993), who were in turn not far ahead of Arthur Samuels, whose mechanical neural network machine was able to play draughts only a couple of years after (McCarthy & Feigenbaum, 2016). However, the next body of work that would have a truly significant effect on the underlying nature of machine learning as we know it today is that of the Perceptron (Rosenblatt, 1958). Rosenblatt's publication gathered widespread attention from academics and the more curious part of the public as well (Harding *et al.*, 1958). From the perspective of modern machine learning techniques, the perceptron can be defined as an algorithm that is cable of deciding whether a vector of numbers belongs to one class or the other. From a more modern perspective, therefore, the perceptron belongs to the supervised learning category of binary classifiers (Freund & Schapire, 1999). In 1969 however, Marvin Minsky and Seymour Papert published their book *Perceptrons* which set to lay out the limits and limitations of perceptrons and neural networks. This publication lead to what many would call the "AI winter" of the 1970s, where the future and general capability of methodology toward artificial intelligence was widely questioned and challenged, followed by pessimism in the AI community and the press, as well as cutback in funding (Crevier, 1993). As a result of this, very little progress was made in the machine learning field in the 1970s.

In 1980, exactly ten years after the onset of what Crevier (1993) describes as the AI winter, Fukushima (1980) published a discovery he called the *neocognitron*, which was a type of artificial neural network. The neocognitron was a mathematical model – a simulation – of a set of inputs algorithmically mapped to outputs. This contrasts the SNARC neural network described previously in that SNARC was a physically tangible hardware implementation powered by electric pulses which activate neurons, which is why it is called an *artificial* neural network. The work of Fukushima (1980) also contributed to what would later be known as *convolutional* neural networks, which are specifically known for their ability to parse data in convolutional phases, mimicking the animal brain's visual cortex (Le Cun *et al.*, 2015).

After the AI winter had passed, and Fukushima (1980) reignited the interest in the field of machine learning once again, the number of research publications grew, with De Jong (1981) introducing *explanation based learning.* Explanation based learning was proposed as a methodology for discarding data that is unimportant to a specific domain given a set of examples of data that can be categorised as relevant or not (Bernard, 2016). The year after saw the publication of recurrent neural networks by Hopfield (1982), a topic that was destined to receive much attention in more modern years. An example of what was possible at that time with machine learning is perhaps Sejnowski's (1985) *NetTalk,* a program that was able to learn how to pronounce words similar to the way a baby does (Bernard, 2016).

The year after the publication of NetTalk, Rumelhart *et al.,* (1986) were the first to introduce experiments of what is now known as *backpropagation* in a neural network environment. This was the idea of automatic differentiation (first applied theoretically to neural networks by Werbos (1974) in his PhD thesis), reversed. This means that automatic differentiation is applied with outputs from neurons nearer to the output layer being passed on as inputs to neurons nearer to the input layer. This method enables the neural network to learn representations within the hidden layers and pass them as inputs to earlier layers, which aids in the effectiveness of parsing and simplifying representations of information in the network (Rumelhart *et al.*, 1986).

The next major instalment in the chronology of the development of sophisticated learning models is the introduction of *Reinforcement* learning by Watson (1989), in which it was stated that the goal of the work was "to give a systematic analysis of possible computational methods of learning efficient behaviour". This is a significantly different methodology from what has been discussed in this overview so far, which was primarily the process of mapping inputs to outputs algorithmically. This significant difference in methodology creates an algorithm that finds a policy that is optimal in the sense that it maximises the expected value of the total reward over any and all successive steps, starting from a given state in any finite Markov decision process (FMDP). Reinforcement learning can, therefore, identify an optimal action-selection policy for any given FMDP, given infinite exploration time and a partly random policy to begin with (Melo, 2001). The specifications of reinforcement learning models were first denoted by Watson (1989) with the algorithm containing “Q”, which is representative of the reward used to provide the reinforcement and can be said to represent the “quality” of an action taken in a given state.

The period following the publication of Q-learning by Watson (1989) as described above saw the release of *Evolver*, which was a software package designed to implement genetic algorithms on personal computers (Markoff, 1990). While it is difficult to determine the direct outcome of the release, it marks a significant point in history when machine learning was being made more commercially available. A notable achievement of the era is also *TD-Gammon*, a computer program utilising machine learning to play the game Backgammon. The implementation consistently levelled human proficiency but failed to consistently surpass it (Tesauro, 1995). Three years after the release of TD-Gammon, Ho (1995) published a new specification of machine learning called the *Random Decision Forest Algorithm*, and in the same year, Cortes and Vapnik (1995) published the *Support Vector Machine* algorithm. Two years later, in 1997, IBM’s Deep Blue beat the world’s best Chess player: Gary Kasparov. It is noted, at this stage of the chronology, that this is the time during which Hochreiter and Schmidhuber (1997) published their work on the Long Short-Term Memory network, which is the network specification used in Chapter 4.

Another significant occurrence in the history of machine learning was the release of the software library called *Torch.* Its release marked the onset of notable popularisation of the entire machine learning field and vast accessibility improvements even to the general public (Collobert *et al.*, 2002). From this point onward, Platforms like *Kaggle* (a website that hosts machine learning competitions) created a widely engaged-in commercial market for machine learning developers. Many believe that developments such as *ImageNet* was a significant catalyst to developments in machine learning and whether that is considered true or not, by 2011 machine learning specifications were able to beat humans in games as complex and multi-faceted as *Jeopardy!*. Consequently, further progress in the field was commercialised continuously up until today, and instalments in the chronology become too frequent to review in this manner. These frequent developments can, however, be summarised in terms of an overarching categorisation of neural network architecture by domain, as is performed in the following section.

2.4.2.2.6.2 Categorisation of neural network effectiveness by domain

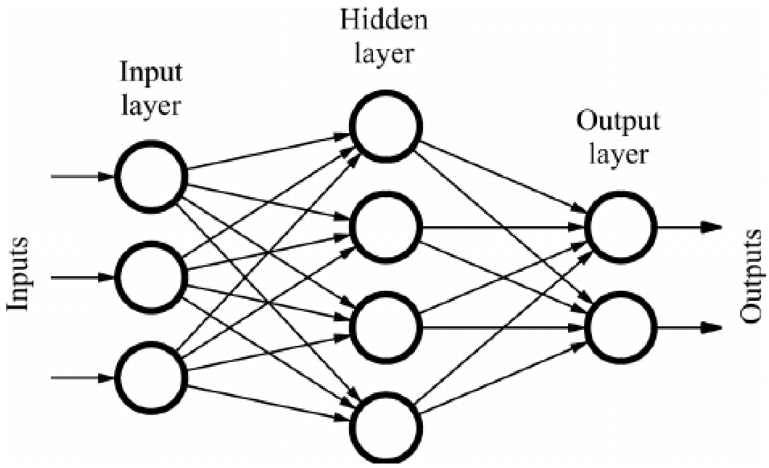
This section is designed to provide the reader with a categorisation of the general domains in which machine learning has been tested and accompanying measures of effectiveness. The purpose is to address the families of models commonly utilised within modern machine learning, and not the fine-tuned specifics within these families. It is also not an exhaustive summary of families of models in use today, but only those that may be relevant to this study. [1]

The perceptron (P) and the feed forward neural network (FFNN)

The perceptron was the genesis of what are now called *neurons*, the building blocks that neural networks consist of, and were originally designed as a stand-alone probabilistic model for information storage and organisation in the brain (Rosenblatt, 1958). The original perceptron was not proposed in the layered type of network that is standard today, but rather a theoretical proposal of how the operations of a single neuron (the perceptron) could be expressed mathematically and how it could operate similarly to organic neurons in that they generate a response from a stimuli with given characteristics. Rosenblatt (1958) referenced sources from neurology, behavioural science, automation studies, biology and statistics to propose the perceptron, which was at the time highly theoretical, as the computational power to position layers upon layers of these perceptrons in a network did not exist.

When neurons are stacked together in layers, they form a simple feed forward neural network as follows:

Figure 2-1: Simple Feed Forward Neural Network.



(Source: Quiza & Davim, 2011)

The FFNN in figure 2-1 is a relatively simple form of neural network configuration compared to more modern architectures. The simple FFNN has had its past successes in the contemporarily challenging environments object orientation detection, universal approximation problems

[1] The reader is encouraged to consider this section with Exhibit B (Van Veen, 2016) in the appendix close by.

(Hornik *et al., 1989)* and simple classification problems (Montana & Davis, 1989). The FFNN modelling technique offered researchers the opportunity to engage in solving problems that required nonlinear specification, because each neuron in such a network could have a nonlinear activation function determining whether its output will adjust the weight of the neuron that follows it (Huang & Ma, 1999). Among the first of FFNNs with interesting nonlinear activation functions was the Radial Basis Function (RBF) FFNN, which was created to aid in nonlinear interpolation whilst maintaining explicit learning rules (Broomhead *et al., 1988).* This proposed exciting solutions to, and drove research into themes such as, investigating effectiveness of FFNNs with a multitude of activation functions.

Before continuing with a discussion of various specifications of FFNNs, it must be addressed that there exists an ensemble of modelling techniques that could be considered a subfield of neural networks (if one is not too conservative with the inclusion of memory dependence). These models found their origin with the Hopfield Network (HN) and are discussed below, with their variations, before continuing with the discussion of NNs in the traditional sense.

Markov chains, Hopfield networks, Boltzmann machines and Restricted Boltzmann machines

We begin with the field of logic that precedes the Hopfield Network: the Markov Chain. Markov (1913) extended the theory of probability in a new direction with the work concerning state-dependence (cited by Hayes, 2013). This body of work has led to the consideration of state dependence in many fields of analytical tools over many years to come, and neural networks were no different. It was Hopfield (1982) who first proposed a system of interdependent and fully interconnected states which can correctly yield an entire information set's memory from any subpart of that system, given it is of sufficient size. This system promised favourable properties for generalisation, familiar pattern recognition and time series retention and was later called the Hopfield Network. This network configuration was utilised as a candidate for nearest neighbour problems (Montgomery & Kumar, 1986), image restoration (Paik & Katsaggelos, 1992) and missile guidance (Steck & Balakrishnan, 1994). The incorporation of a conceptual Markov into a neural network led to the creation of Boltzmann machine four years after the Hopfield network, and later the Restricted Boltzmann machine, all of which are relatively similar in how they differ from traditional FFNNs.

Boltzmann machines and Restricted Boltzmann machines are of a similar mechanical structure to Hopfield networks, apart from – in the former - how the structure approaches hidden nodes, and – in the latter – how the nodes are connected. Boltzmann machines hide some of the nodes in the network (thus they are not considered input nor output nodes) and those that are not hidden are considered inputs before training the network and outputs after the network is trained (Hinton *et al.,* 1986). Restricted Boltzmann machines behave similarly in this regard, with the added difference that not every node is connected to every other node. Rather, groups of nodes are algorithmically designed to share data among them (Smolensky, 1986). Boltzmann machines have been utilised in speech recognition (Bengio & De Mori, 1988), combinatorial optimisation (Aarts & Korst, 1989) and optics (Ticknor & Barrett, 1987).

This concludes the discussion of this ensemble of densely-connected, memory-addressable networks that do not feed forward from an input to an output layer. From here it is possible to consider the next chronological evolution in neural networks which emerged in 1988. The ensemble to be discussed in the following section is that of *Encoders*.

Autoencoders, Sparse Autoencoders, Vartiational Autoencoders and Denoising Autoencoders

Data compression (a process of encoding data to a smaller size) has always been a matter of interest to the technology industry as it leads to greater efficiency in the usage of storage space related resources and network speed. Encoder networks were first proposed by Bourlard & Kamp (1998) to serve as an efficient compression algorithm, mapping input to output through a confined space which represents the compressed data after training the network. It is therefore the same network architecture as FFNNs, only a different use case. The network consists of an input layer and an output layer of the same size, but (a) hidden layer(s) is restricted to a smaller size, and through this method, if the network maps input to output successfully, the same data can be represented within a smaller representation space i.e. a matrix of lower rank (Bourlard & Kamp, 1998).

Different methods were applied to the same end goal with the creation of Sparse Autoencoders. Sparse Autoencoders, however, have hidden layers consisting of more nodes than the input and output layer, and this is done in order to extract small differences in features from a given dataset. Autoencoders and Sparse Autoencoders share the commonality of proven effectiveness in information compression (Theis *et al.,* 2017). Sparse Autoencoders have been uniquely effective at complex classification (Xu *et al.,* 2016). They have also been effectively applied on multiple instances as pre-processing tools for feature engineering in convolutional networks (Vu *et al.,* 2017). A variation of the autoencoder structure applied to a completely different goal is the Variational Autoencoder.

Variational Autoencoder specifications exhibit the same architecture and primarily toward the same goal as standard autoencoders described above. However, the training process is not the same (Kingma & Welling, 2013). In Variational Autoencoders, training attempts to minimise data loss on an approximate probability distribution of the input samples. This is more closely related in methodology to Boltzmann – and Restricted Boltzmann machines described previously. Variational Autoencoder networks therefore combine the Bayesian measures of interdependence found in densely connected graphs such as the Boltzmann machine with the methodology of directed networks with informational flow mapped from input to output. This combination of methodologies is referred to as directed probabilistic models (Kingma & Welling, 2013). Variational Autoencoders have been found to be useful in detecting anomalies in data (An & Cho, 2015), image labelling (Pu *et al.,* 2016) and text modelling (Yang *et al.,* 2017). The effectiveness of Autoencoder models across domains would not have been possible without the contributions toward their ability to detect and isolate robust patterns, a task that Denoising Autoencoders are especially useful for.

The Denoising Autoencoder specification is the last model in the autoencoder ensemble to be brought under consideration here. This particular specification – with its more minor alterations included – was specifically designed with the purpose of exploiting autoencoders' aptitude for pattern recognition and representation whilst improving the robustness of results (Vincent *et al.,* 2008). Prior to the publication of Denoising Autoencoders, it was proven to be evident that stacked layers of neural networks provide improved results over single-layer architectures (Bengio, 2007). However, deeper networks brought with them more complex optimisation, and theoretical findings seemed difficult to scale to solutions that fit complex real-world problems. Although stacked layers of autoencoders producing simplified inputs to complex architectures were proven to improve results, the problem of robustness was addressed by Vincent *et al.* (2008). The idea was that introducing noise to autoencoders and mapping inputs with noise to noiseless outputs will improve robustness of pattern recognition, which was found to be successfully applied in the same study. This implicates that complex deep networks can be given inputs that are simpler, more robust and still more representative of data and more effective than before (by data pre-processing with the use of Denoising Autoencoders). Training networks to provide encoded/compressed inputs for other networks can, however, be done in many different ways – the proposal of Vincent *et al.* (2008) being just one of them. Another such a method is known colloquially as *greedy training.*

Greedy training is discussed on equal footing with Denoising Autoencoders above. The reason for this is that the greedy training methodology was proposed and publicised before Denoising Autoencoders, but greedy training relies on a much more complex neural network to be performed. The two theories are also very similar, and they each address largely the same domain of problem statement. In the original greedy training paper, Bengio *et al.* (2007) describes the same problem as Vincent *et al.* (2008). The problem in question states that sufficient evidence exists to suggest that deeper and more complex neural network architectures can be much more efficient than shallow architectures, in terms of the simplicity of the resulting function after training. Bengio *et al.* (2007) stated that this efficiency improvement is sometimes exponential. However, both groups of researchers also state that the challenge of training complex neural architectures requires more attention, and while Vincent *et al.* (2008) addressed this with the Denoising Autoencoder, Bengio *et al.* (2007) tested the Deep Belief Network that they called, more simply, greedy training. The Deep Belief Network stacks combinations of Restricted Boltzmann machine - and Variational Autoencoder layers in order to address the problem of simplifying complex non-linearities in inputs. What this means is that the data is pre-processed by the Deep Belief Network in order to be introduced to the learning algorithm as a probabilistic model. Deep Belief Networks have been utilised in various domains such as natural language processing (Sarikaya *et al.,* 2014), acoustic modelling (Mohamed *et al.,* 2012) and image classification (Krizhevsky *et al.,* 2012).

This concludes the overview of the conceptual variations of encoder networks. The evidence may suggest that encoder network methodologies are able to improve performance by mapping raw data to probabilistic representations, and this ensemble of architectures will be considered as a method toward solving the problem statement of this research. The encoder ensemble of network architectures has been used in the past to encode inputs to convolutional networks, which is the next network architecture under consideration.

Convolutional, Deconvolutional and Deep Convolutional Inverse Graphics Networks

Convolution is a mathematical operation performed on data from two functions, which outputs a third function, which in broad terms (so as to include Fourier analysis) is a technique fundamental to signal processing. Toward the addition of context: after the invention of the Perceptron, a significant amount of focus was set toward improving the results that they were able to achieve. This contributed toward various experiments in the way that Perceptrons were specified and how they are combined to form networks of Perceptrons. This is what led to the proposal of convolutional neural networks, which had their origin during the time that Multilayer Perceptrons were being perfected, when Fukushima (1980) proposed the specification of a so-called Neocognitron. The Neocognitron was essentially a FFNN consisting of multiple layers of Perceptrons, which categorises it more specifically under Multi-Layer Perceptrons. What made the Neocognitron different from contemporary MLPs was that it relied on convolution between layers (Fukushima, 1980). This implicates that the workload of finding causal and weight-bearing relationships among multitudes of neurons becomes limited to local connections, making the process more specialised and more efficient (Nebauer, 1998). Convolutional neural networks have been found comparatively superior in domains such as visual pattern recognition (LeCun *et al.,* 1998), handwritten digit recognition (Calderon *et al.,* 2003) and image classification (Chen *et al.,* 2016). This is perhaps unsurprising when one considers that the Neocognitron was originally designed, according to Fukushima (1980), to mimic broadly the operations performed by the visual cortex upon the reception of stimulus. There does, however, also exist potential for a model specification that operates in the same regard but in the other direction, which is to go from an informational abstraction of an arbitrary entity, to a graphical representation of it. The neural network architecture that is capable of performing such a task is the Deconvolutional Network.

Deconvolutional neural networks, even though their name and graphical representation may suggest otherwise, are very different from their convolutional counterparts. This is because the process of going from an image of a cat as input, to the string of letters “cat” as output, is so vastly different from the process of going from the string of letters “cat” as input to an image of a cat as output. The former is a typical example-capability of the above-described convolutional neural network. In such a specification, information is purposefully simplified from extremely specific details to robust features, edge primitives are taken from case-specific data and information such as intersections, parallelism and symmetry are destroyed (Zeiler *et al.,* 2010). In contrast, the latter is a typical example of the nature of problem that is addressed by deconvolutional networks, where a very simple abstraction of information is given as input to a network, and the output is expected as a robust set of features with as much instance-specific complexity as possible (Zeiler *et al.,* 2010). The result from such a methodology is a network capable of learning to synthesise images i.e. go from the string of characters “cat” as input to an image of a cat as output. Deconvolutional networks have been concluded to be useful in domains such as advanced feature learning (Zeiler *et al.,* 2011), image generation (Dosovitsky *et al.,* 2015) and semantic segmentation. Image synthesis, however, can be done more accurately and with greater precision given more sophisticated specifications combining convolution with other network features. One such an example is the Deep Convolutional Inverse Graphics Network.

The Deep Convolutional Inverse Graphics Network (DCIGN) is a specific network architecture with a specific task, which is to model interpretable representations of images, disentangled with respect to transformations such as out-of-plane rotations and lighting variations (Kulkarni *et al.,* 2015). The network relies on various other network specifications to get to this result. The input layer is connected first to various hidden convolutional layers, ending up in the “middle” of the network which represents the interpretable representation of the input information in the form of a variation autoencoder system. The representation is then developed from the variation autoencoder toward more hidden deconvolutional layers after which it is directed to the output (Kulkami *et al.,* 2015). This means that the features of an information set are encoded as probabilities in the VAE. The power that this lends to the network is that representations can be understood in a novel way, enabling the network to produce a picture of both a dog and a cat, having only ever been inputted pictures of dogs and cats separately (in keeping with the original example). These networks have been proven to be of utility in domains such as visual analogy-making (Reed *et al.,* 2015) and physics simulation (Wu *et al.,* 2017). The methodology of combining networks is also not nearly a new one, as neural networks have been used together in "teams" since very long ago. Just as they can be used in teams, they can also be pitted against one another as adversaries, resulting in an interesting breed of testing called Generative Adversarial Networks.

Generative Adversarial Networks

A good starting point for the underlying premise of the Generative Adversarial Network (GAN) is supervised learning, where in its simplest form, a labelled dataset is provided as input for an algorithm to map a function correlating to the relevant class (Kotsiantis *et al.,* 2007). The pitfall with this methodology is that the learning algorithm's accuracy depends on the amount of data available, which significantly narrows the domain in which the methodology can be tested, as well as its overall efficiency in the real world. A more ideal situation would be one where a specification's learning capability's most binding constraint is processing speed, which is a much less binding resource than labelled data available. Based broadly on this premise, GANs were invented to train one neural network against another neural network (Goodfellow *et al.,* 2014). The idea is that two networks are compiled: a generative network, and a discriminative network. The generative network has the purpose of capturing the distribution of the dataset. The discriminative network has the purpose of estimating the probability that a given sample originates from the training data rather than from the generative network. The training procedure involves maximising the probability that the discriminative network classifies incorrectly the origin of the given sample – outputting that the origin is from the real dataset rather than from the generative network. The discriminative network therefore can be seen as the judge of whether or not the sample is generative algorithm-generated or a true sample from the data (Goodfellow *et al.,* 2014). GANs have been found useful in domains such as texture synthesis, image manipulation (Zhang *et al.,* 2017) and compression (Agustsson *et al.,* 2018). The process of training adversarial networks, as with all of the networks discussed before, does not include an element of time sensitivity. This is the problem that recurrent networks attempt to address.

Recurrent Neural Networks, Long Short Term Memory Networks, Gated Recurrent Units, Neural Turing Machines and Bidirectional Variations

The network architectures discussed up until now all have one thing in common, which is that they will behave no differently at different points in time, unless an explicit time dimension is included in the input. Seeing as time is a concept underlying many interesting human behaviours, the sensitivity thereto is of considerable importance when modelling brain-mimicking connectionist models like neural networks. However, if space is not included as a spatial representation of a numerical-probabilistic dependency, how can a neural network become sensitive to it? How can, therefore, a network be configured to behave a certain way given a certain preceding set of anomalies in its input? One possible answer to this was first introduced by Jordan (1986) and later formally proposed to neural network architecture with evaluative tests by Elman (1990). The aim was to find structure in time by introducing the concept of recurrence.

Recurrent links are able to provide a neural network with dynamic memory (Elman, 1990). The result is that time is incorporated into the network through state-variability. Therefore the network is expected to behave differently based on its internal representation built on the recurrence of previous events. Elman (1990) reports that recurrent neural networks are indeed able to learn complex internal representations which incorporate tasks with definitive memory demands, and that they are able to behave highly dependent on context whilst still exhibiting generalisation across classes of items. The proposed application for recurrent neural networks was language. Until today, however, recurrent neural networks have been concluded useful in domains such as complex classification (Guler *et al.,* 2005), sentiment classification (Tang *et al.,* 2015) and widely in natural language processing (Graves *et al.,* 2013). The concept of introducing memory in the form of state dependence evolved further from the recurrent network, up to the point where the focus was specifically on how memory functions in the Long Short Term Memory.

Long Short Term Memory Neural Network

While the creation of state-sensitivity to time or to a sequence was good progress in the RNN architecture discussed previously, the solution did bring along its own set of problems, most notably the exploding/vanishing gradient problem. The problem of exploding gradients originates from the activation function in the neurons, which is mathematically discussed in Chapter 3, but a simplified high-level theoretical abstraction of this would be that the cells forget too quickly or barely at all. The performance of dynamic state networks depended too much on the researcher’s ability to diagnose gradient issues throughout the training process. This created the need for a recurrent neural network, with dynamic memory and state dependence, but with an improved solution to the gradient problem. To this end, Hochreiter and Schmidhuber (1997) propose a novel, efficient gradient based method toward the error flow during backpropogation called Long Short-Term Memory (LSTM). The LSTM proposes multiplicative gate units that can learn to open and close access to an error flow that is constant – a methodology which enables it to bridge minimal time lags in excess of 1000 discrete-time intervals whilst still preserving memory if the original anomaly recurs (Hochreither & Schmidhuber, 1997). Hochreiter and Schmidhuber (1997) compared the LSTM to contemporary algorithms and found that it boasted improved performance over all of them, and discredited further proposal through proofs that random weight guessing could solve complex problems faster than some of the proposed solutions of the day. The LSTM was found useful across many domains for many years to come, including natural language processing (Ghosh *et al.*, 2016), stock market predictions (Nelson *et al.,* 2017) and malware detection (Vinayakumar *et al.,* 2018). LSTMs also gave rise to variations of a similar methodology, under which Gated Recurrent Units.

Gated Recurrent Units

Gated Recurrent Units (GRUs) are a slight variation of LSTMs (Chung *et al.*, 2015). GRUs were first proposed by Cho *et al.* (2014) in the form of a RNN Encoder-Decoder configuration with the purpose of application in meaningful representations of linguistic phrases. The LSTM and GRU have in common that both have the purpose of being capable of adaptively resetting or updating memory content, only the methods of doing so very slightly differ between the two (Chung *et al.,* 2015). The GRU has been referred to as a simplified variant of the LSTM by Greff *et al.,* (2017), which is suggested to make it a more efficient alternative. However in contrast, the RNN has also been found to be moderately less expressive than the LSTM, which might indicate that deeper architectures will be required to match the accuracy of the LSTM when using a GRU. Comparison finds very little to distinguish performance between the two, and thus parameter tuning of an LSTM may prove to be a favourable alternative to employing more-complex models. Other studies on complex problems also favour LSTMs over GRUs for their relative expressiveness (Athiwaratkun *et al.,* 2017). GRUs have been concluded to be useful in many of the same domains as LSTMs, including natural language processing (Yin *et al.,* 2017) and sequence modelling (Chung *et al.,* 2014). As the GRU is a variation of the LSTM in terms of the mechanism through which sensitivity to state (and therefore time) is achieved, variations in terms of other regards also exist. One such an example is the Neural Turing Machine, which addresses the concept of the memory state itself.

Neural Turing Machines

Neural Turing Machines are an extension of neural network capability through the investigation of memory (Zaremba & Sutskever*,* 2014). They are an interesting discussion for a partially different set of reasons than recurrent neural networks, but they are discussed here along with the recurrent network group, because they require memory-addressable state to perform the functions that they are designed to (Graves *et al.,* 2014). Neural Turing Machines have their underlying theory based on the three fundamental operations that computers are capable of performing: arithmetic & logic, flow control i.e. branching, and external memory (Von Neumann, 1945). Although neural networks have been widely successful as forecasting tools across many domains, as discussed in preceding sections, the investigations into the logical flow and external memory components of neural networks were largely neglected as a field of research until the release of the Neural Turing Machine publication by Graves *et al.* (2014). The study was an investigation of the memory states of recurrent neural networks analogous to working memory in humans, and explains how rapidly-created variables in memory can be analysed to illuminate some of the processes concurrent with training a memory-addressing network.

Even though Neural Turing Machines do not pose any efficiency or efficacy improvements as modelling or forecasting technique, they aim to improve the way we understand the operations that occur in the hidden layers. They can therefore be understood as regular recurrent neural networks with the added ability to read from and write to memory, and change state based on what is read. This makes them an investigation of the working of recurrent neural networks, rather than an attempt to improve their results or efficiency. However, there are more complex network configurations that attempt to improve modelling accuracy results based on recurrent networks. Among these is the group of models that arise when the back-propagation method in recurrent networks is altered to go in both directions toward and away from the output layer. These are called Bidirectional networks.

Bidirectional Networks

Bidirectional networks do not employ a methodology or network architecture that is separate from normal recurrent networks. They therefore look the same in a graph compared to their unidirectional counterparts. The difference between bidirectional and unidirectional networks is how they handle inputs with respect to time. Bidirectional networks are given previous and future time steps of their training data, all according to the original publishers Shuster and Paliwal (1997). This is achieved by training the network variations through positive and negative time direction. The bidirectional variations of various recurrent neural network architectures have been shown on multiple occasions to outperform their unidirectional counterparts (Graves & Schmidhuber, 2005). Bidirectional variations of recurrent neural networks have been successfully applied across various domains such as keyword spotting (Fernandez *et al.,* 2007), global stability analysis (Zhang & Yang, 2001) and speech recognition (Graves *et al.,* 2013).

This concludes the discussion of the group of networks that are sensitive to sequences and thus exhibit good performance on problems where time is a dimension in the data. It must be stated that many variations of these models exist, and that the above discussion is by no means exhaustive. However, as an overview with regard specifically to a neural architecture's sensitivity to time, the above is considered sufficient, seeing as most variations build complexity on the underlying premise of recurrence, which is thoroughly discussed.

The focus of the discussion can now be shifted toward a group of models that are categorised together for the way their neurons are similarly connected, although they have significant differences between them as well. These are somewhat special cases which can hardly be discussed with any other of the above categories. Also, the body of literature for these models is significantly less comprehensive than for most of the categories discussed so far.

Neural Networks with random connections

There exists a group of models with no set pattern for the connections between their neurons. They have not been formally categorised together with this characteristic, but seeing as they share it, and it results in their looking very similar, they are discussed together here. The first of these networks is the Extreme Learning Machine.

Extreme Learning Machines

The Extreme Learning Machine (ELM) is an architecture proposed originally by Huang *et al.* (2004) as an alternative to FFNNs, with the competitive edge of a much faster learning speed. Huang *et al.* (2004) criticise traditional FFNNs for their slow gradient-based learning algorithms as well as their iterative learning process, and propose (with the ELM) a similar node architecture, but two significant differences in the way it learns, the first of which is the way it attaches weights to neurons. The ELM randomly chooses input weights for all the neurons in the hidden layer, and then learns through an analytical process of adjusting the output weights from the hidden layer and minimising the error. The ELM specification was found to learn extremely fast and generalise well (Huang *et al.,* 2004). The ELM was usefully applied across various domains, including as an autoencoder for big data representation (Lekamalage *et al*., 2013), vehicle routing (Feng *et al.,* 2013) and data visualisation (Akusok *et al.,* 2013). ELMs therefore brought more efficiency to traditional techniques by minimising parameters that need to be learned in the training process, whilst maintaining a model specification that generalises well. In more specific terms, the idea of initialising a network with random weights and applying training only to the output has been utilised with other learning patterns as well, one of which is recurrence – found in Echo State Networks.

Echo State Networks

The Echo State Network (ESN) was originally proposed by Jaeger and Haas (2004) as a contribution toward nonlinear dynamical systems in engineering (specifically signal processing). At the time of its publication, the ESN improved performance on typical nonlinear dynamical systems modelling tasks and improved signal reception error rate significantly (Jaeger & Haas, 2004). The authors report that several learning algorithms exist, but criticise that they have not been widely employed in technical applications due to slow learning and suboptimal solutions.

The ESN is then proposed as an alternative, with improved learning rate. The ESN is also, like the ELM discussed above, initialised with weights that remain fixed throughout learning (however in the case of the ESN the input weights are not random). The output weights are also, like in the ELM, the only parameters that are learned during the training process. This contributes to the learning speed of the network. The ESN and the ELM are therefore very similar. What makes the ESN different from the ELM is that it also applies recurrence to the process which is otherwise the same. ESNs have been successfully utilised across various domains such as short term stock price prediction (Xin *et al.,* 2009) information processing (Boedecker *et al.,* 2012) and motor control. Initialising fixed weights for all neurons and random connections seems to give this group of networks some interesting abilities. ESNs and ELMs are, however, not the only networks that have the methodology of training only the synaptics between the hidden and output layers. More variations of the same network architecture also exist. One such an example is called the Liquid State Machine.

Liquid State Machines

The Liquid State Machine (LSM) looks a lot like the ELM and the ESN, because it also has randomly connected nodes between two structured layers of input and output. However, while the ELM and ESN share more commonalities than merely their graph, the LSM is truly quite different.

The Liquid State Machine is a spiking network. It was originally proposed by Wolfgang *et al.* (2002) as a solution to the problem of real-time, time-varying multimodal input processing. The LSM is different from any neural network based on a Turing machine methodology, i.e. transitions between well-defined states. This makes it a truly interesting model and comparatively unexplored territory. In the original paper, the LSM was proposed as a solution to problems in robotics, and has been utilised for this purpose (Antonelo *et al.,* 2007). LSMs have such unique application ability because of how they operate as spiking networks. Spiking networks do not use activation functions like most other neural networks do, rather they use thresholds that determine when a neuron’s weight is transferred. This means that each neuron is an accumulating memory cell rather than a matrix operation. Therefore, when updating a neuron, its value is not set to the outcome of its operation with inputs received from its neighbours, but rather its input value is added to its balance value. Once the preset threshold is reached, the neuron’s total balance is transferred to the next neuron, which adds this to its own balance. It means that a given neuron can be expressed as a certain condition being met and that any possible combination of events that could lead to the meeting of the current condition is described by the neuron value transfers that lead to the event of the given neuron spiking. This is what creates the spiking pattern, where an accumulation of events lead to a given neuron’s condition being true, and thus the given neuron will fire.

Outside of the originally intended domain (robotics), Liquid State Machines have been found particularly useful in domains such as speech recognition (Zhang *et al.,* 2015), molecular physics (Bartok *et al.,* 2013) and facial expression recognition (Grzyb *et al.,* 2009). Spiking models (under which LSM's) are an interesting group of models that can hardly be categorised based on commonality. More such networks are up for discussion in the next section, which begins with Deep Residual Networks.

Uncategorised architectures such as Deep Residual Networks, Support Vector Machines and Kohonen networks

This section is aimed toward addressing some neural network configurations that could be of value in terms of solving the problem statement, but do not fit easily into any of the above categorisations of networks. These configurations are therefore intrinsically unique enough to deserve not to be seen as a variation of the categories addressed. This is also not an attempt to subcategorise them together, as their underlying methodologies differ far from one another as well. This is therefore a discussion of miscellaneous architectures that have reason to be believed as possibly suitable toward solving the problem statement of this research. The first network model up for discussion is the relatively simple Deep Residual Network.

Deep Residual Networks

The Deep Residual Network (DRN) was first proposed by He *et al,* (2015) as a contribution toward an ongoing challenge faced in the neural network domain. The challenge in question is that, especially in the image classification realm, it has been established that deeper neural network configurations lead to more successful results (Szegedy *et al.,* 2015). However, deeper network specifications are also associated with accuracy saturation followed by accuracy drop, and higher training error. Given therefore that deeper networks yield improved results along with greater training error, the optimal solution would be to create a methodology that enables deeper networks with improved training error. This is the problem that He *et al.* (2015) address with the publication of the Deep Residual Network. The way the DRN approaches the problem of deeper layers with more accurate training, is by explicitly reformulating layers as learning residual functions with reference to their inputs, as opposed to other methodologies which attempt to learn unreferenced functions. He *et al.* (2015) also provide empirical evidence that this methodology leads to improved results, gained accuracy arising from greater depth and easier optimisation. DRNs were successfully applied in image classification (Yu *et al.,* 2017), image super resolution and language analysis (Tang *et al.,* 2018). DRNs are therefore concluded as one of the uncategorised network specifications, and the discussion can continue to Support Vector Machines.

Support Vector Machines

The Support Vector Machine (SVM) is a machine learning methodology first proposed by Cortes and Vapnik (1995). The SVM relates poorly to any of the concepts described in this section after the perceptron (which is the very first point of discussion). It is therefore given its own discussion of its own history from the true origins of its methodology seeing as it is such a powerful technique that cannot be ignored as a candidate toward solving the problem statement of this research.

The SVMs origins can be traced back to the first proposed algorithm for pattern recognition by Fisher (1936), where the optimal solution to a two-variable classification problem was first associated with a linear decision surface (Cortes & Vapnik, 1995). This technique can be seen in Rosenblatt's (1962) perceptron experiments as well, where each neuron within a perceptron implements a separating hyperplane, which can be seen as a piecewise linear separating surface. It must be mentioned at this stage that Rosenblatt's perceptron models did not allow for all the weights of the network to be adjusted in order to minimise local error for a given node, but rather only the weights associated with the output were adjustable at the time (Cortes & Vapnik, 1995).

The algorithmic solution toward adapting all weights in a neural network was only determined during the discovery of back-propagation (Rumelhart *et al.,* 1986). This involved a slight modification of the mathematics between neurons, but it did not change the fact that before SVMs, neural networks relied ultimately on piece-wise linear-type decision functions (Vapnik & Cortes, 1995). SVMs continued the use of linear separation with a different methodology, where input vectors are mapped into some high dimensional feature space through some non-linear mapping chosen *a priori*. The linear decision surface is then constructed with special properties that ensure high generalisation ability of the network. Cortes and Vapnic (1995) specify the technique in great detail, however for this research it is at this stage suitable to describe the SVM as a method that constructs an optimal hyperplane separating features in a non-linearly mapped feature space. The SVM has been found effective across various domains such as spam categorisation (Drucker & Vapnik, 1999), computational biology and phonetic classification (Clarkson & Moreno, 1999). SVMs are therefore quite unique in the machine learning domain, and bear comparatively little resemblance to previously discussed techniques. Another method that can be considered quite unique – but is still a neural network – is the Kohonen network, which is the next point of discussion.

Kohonen networks

The Kohonen network or, more accurately perhaps, the Self-Organised Topologically Correct Feature Map, was first analysed by Kohonen (1982) toward developing an unsupervised competitive learning strategy toward classification. It must be qualified that the Kohonen network is in some regards not a neural network, but it is still discussed here as a machine learning technique with neural network characteristics, as it may be a viable candidate toward solving the problem statement. The process of learning is unsupervised and as Kohonen (1982) describes, self-organising. Input is presented to a network of processing units constructed as an n-dimensional array. The network assesses with a spiking network-like threshold parameter which of the neurons most closely match the output, after which these specific neurons are adjusted to match the input even closer, with the neighbouring processes' movements measured and this measurement representing a way to classify which neurons suit the input data closest. Kohonen networks have been successfully utilised in acoustic emission classification (Huguet *et al.*, 2002), ultrasonic flaw detection (Margrave *et al.,* 1999) and drug design (Anzali *et al.,* 1998).

This concludes the discussion of our so-called miscellaneous models and ultimately the entire neural network section. Given the information presented in this section, it is the expectation that neural networks, especially those of which the specifications exhibit sensitivity to sequential data points, may possess the ability to perform well in a forecasting problem with the goal of maximising profit.

2.5 Valuation and trading of Bitcoin

This section aims to review the relevant literature as pertains to the valuation and trading aspects of Bitcoin. Having established, in preceding sections, a theoretical framework for the link between fintech and financial market complexity (of which Bitcoin is the relevant case study), and subsequently relative thereto having provided a background on asset valuation methodologies, the focus can now be shifted toward empirical evidence of strategies from the literature that employ these asset valuation methodologies in the form of forecasting techniques which to varying degrees of success harness these theoretically established market complexities in order to add value to an investment strategy.

The Bitcoin market system is investigated first, after which valuation practices in the literature are reviewed and, finally, the emphasis is placed on reported trading strategies and the extent to which they use market complexities to add value. This section serves therefore to summarise the empirical literature as pertains to valuation and trading of Bitcoin.

2.5.1 The market for Bitcoin

The dynamics in the Bitcoin market system can be differentiated in terms of its participants (users), and the phenomena which their collective actions form. Between the matters of Bitcoin users vs. the Bitcoin market, the literature is more concerned with the latter, to which the latter of the underlying subsections is dedicated to. We begin therefore with the market for Bitcoin as stratified by the demographics and even empirically suggested political views of its users.

Due to the anonymity of the Bitcoin network, it is difficult to determine who the users are. However, this did not stop Bohr & Bashir (2014) to ask the question: who uses Bitcoin? The results were obtained through the use of publicly available survey data, and the key aspects under investigation were wealth accumulation, optimism about the future of Bitcoin and what attracts the user base to the technology. The authors found that “age, time of initial use, geographic location, mining status, engaging on online discourse, and political orientation were all relevant factors that help explain various aspects of Bitcoin wealth, optimism and attraction. Baur et al., (2015) continues with the thread of applying survey data to the cause of further exploring the community of people behind Bitcoin, and finds that a generally positive attitude exists toward Bitcoin’s future across groups of interviewees. The underlying block chain technology is also seen as a potentially revolutionary way to create a more just society based on open platforms and open data. This supports a widely-held view that Bitcoin’s market participants may share a libertarian ideology for the future of money (Bashir et al., 2016).

Of course, a discussion of Bitcoin’s participants cannot go without reference to the aforementioned Silk Road. Athey *et al*., (2016) note that users who engage in illegal activity are more likely to try to protect their financial privacy and ultimately personal identity – a fact that drives this demographic of users toward Bitcoin given its comparatively superior privacy. The authors also found in the same work that investors and infrequent users (as opposed to users who use it as currency) of Bitcoin held the majority of the currency.

2.5.2 Phenomena

This section aims to review the research on market dynamics observed in the market for Bitcoin which does not pertain to participants or their demographics as such. It is therefore a review of the market mechanism and prevailing phenomena that can be measured quantitatively, not of the user base. The design here emphasises the consideration of the underlying nature of the Bitcoin market through the properties manifested in the form of change in available market metrics. The first area in which such consideration is performed is that of fundamental information.

The sources of fundamental information in the market for Bitcoin are not all the same as those in traditional capital markets. As a short summary for comparative purposes, traditional capital markets receive fundamental information pertaining to assets through established, formal and regulated channels (Abarbanell & Bushee, 1997). This is not the only source of fundamental information for traditional capital markets, but it is the source that these markets do not share with Bitcoin.

This is the main difference between fundamental informational supply in traditional capital markets and the market for cryptocurrency. Among other sources, e.g. social media signals, it is possible to draw comparisons between the market for cryptocurrency and other markets (Garcia & Schweitzer, 2015; Nguyen et al., 2015). This has given rise to multiple meaningful insights into the sentiment effects of social media content on price fluctuations in Bitcoin (Bukovina & Marticek, 2016). Connections between socio-economic signals and Bitcoin price were also found, proving quantitatively the impact of the social and economic fields described earlier on the functions of Bitcoin (Garcia et al., 2014). More narrowly scoped investigation of Bitcoin price formation as affected by the economic field (leaving out social) revealed clustering phenomena related weakly to sentiment in the form of an index called Bitcoin Attractiveness (inspired by the Barro (1979) model) (Ciaian et al., 2016). Opposing findings were generated by Empirical Mode Decomposition (EMD), suggesting that the true driver behind Bitcoin price formation is long-run economic variables (Bouoiyour et al., 2016).

Moving more toward the area of public perception once again, the empirical argument is strong that the use of other payment methods and customer knowledge seem to correlate strongly with Bitcoin price formation (Polasik et al., 2015). Finally, considering political drivers of Bitcoin price formation, contradicting evidence is found once again. Firstly, Yelowitz and Wilson (2015) argue that computer programming and illegal activity search terms correlate with Bitcoin search terms, and that political as well as investment related search terms do not, under the same modelling conditions. This while Bouoiyour and Selmi (2017) present - on equally empirically justified footing - that the political environments of the world are among the top three fundamental influencers of Bitcoin. Models to unify the two theories have not yet been found and therefore the only conclusion is that the underlying fundamental variables driving the price of Bitcoin vary significantly over time. Other interesting market phenomena in Bitcoin include that it has been found to follow the assumptions of the efficient market hypothesis (Jakub, 2015). Also, the correlation between Bitcoin price and global uncertainty metrics found Bitcoin to be a useful hedging instrument for the purpose of diversification (Bouri et al., 2017). Finally, Bitcoin’s international concurrency with respect to geographical dispersion of market participants yields opportunities for arbitrage (Makarov & Schoar, 2018).

The section concludes with the insight that the market phenomena discussed above will be considered in subsequent chapters that attempt to forecast the price of Bitcoin. In chapter 4, when models are fit to Bitcoin data, the research summarised in the above section will determine the inputs and to what extent modelling techniques will be allowed to vary. However, before it is possible to consider which models are suitable in the attempt to forecast Bitcoin's price, an overview of the available techniques must be provided, from which the best alternatives can be chosen in the light of the phenomena previously found to be present in the market in question. Having established an overview of relevant Bitcoin information above, it is now possible to consider the broader areas from which forecasting techniques can be drawn upon to perform forecasts with. This is the content of chapter 3 to follow, which provides the background to asset valuation.

2.5.3 Valuation methodologies as applied to Bitcoin

The matters concerning market dynamics of Bitcoin are to be discussed here. This section attempts to investigate Bitcoin's value proposition for rational investors. The contents are designed to review the literature in terms of the extent to which Bitcoin can be utilised toward improving the ratio of risk and return that an arbitrary portfolio composition exhibits. The first matter to investigate with regard to investment in Bitcoin is its utility toward portfolio diversification.

2.5.3.1 Bitcoin as a utility toward portfolio diversification

Diversification is described by O'Sullivan and Sheffrin (2003) as a process through which capital in a portfolio is exposed to a wider variety of assets - and their corresponding risks - rather than a narrower. Diversification is performed in order to reduce variance in the performance of a portfolio. A successfully diversified portfolio is one which possesses a variance in weighted average performance which is less than that of the weighted average performance of its least volatile asset.

The context under which a well-diversified portfolio is crucial was clarified during the recourse of the 2008 financial crisis (Fragkiskos, 2013). The notion that well-diversified portfolios contributes toward more robust performance is well supported in the literature that investigates systemic risk in financial markets (Verdier, 2013). Comparably, global markets that suffered more significant losses tend to be more poorly diversified (Bartram & Bodnar, 2009; Yang & Rea, 2017). It is for these reasons that investors should always be looking for opportunities to improve diversification in their portfolios, and Bitcoin may possess certain characteristics improving its inclination toward serving such a purpose.

The idea of using Bitcoin as a diversification instrument is not a new one. Kristoufek (2013) proposed the idea in a paper where the matter of investigation was not diversification, but rather a study of correlation between Google Trend searches and Bitcoin's value. In the paper, though, the author mentioned that Bitcoin's weak correlation with traditional markets may suggest it could be considered an instrument toward diversification. The matter did not enjoy much academic attention at the time, but interest was piqued in the topic around 2015, when multiple papers were published studying Bitcoin as a diversification tool. The inclusion of even a "small proportion" of Bitcoins into an already well-diversified investment portfolio may dramatically improve the risk-return characteristics thereof (Briere et al., 2015). Also, in the context of a Conditional Value-at-Risk (CVaR) methodology, Bitcoin should be included in optimal portfolios according to Eisl et al., (2015). Upon further investigation, the evidence becomes overwhelming in favour of Bitcoin as a diversification instrument (Bouri et al., 2017; Dyhrberg, 2016; Platanakis et al., 2018).

It is concluded then that Bitcoin has value as a portfolio diversification tool, but what about other functions? Is Bitcoin's entire value proposition limited to the fact that it bears little correlation to traditional assets, or can it offer value in the context of rational investors who engage in active speculative investment as well? If this is found to be true, to what extent is it so? The following sections aim to investigate this further.

CHAPTER 3 DATA AND METHODOLOGY

Bitcoin is considered a viable candidate to analyse using a data driven approach because it fulfils certain important criteria; firstly, that with regard to the availability of market data, Bitcoin is fully transparent, offering granular data freely available to everyone. Apart from the semantics of the forecasting process and the availability of market data, Bitcoin has been found to be weakly correlated with other assets (Briere *et al.,* 2015; Eisl *et al.,* 2015). This speaks to the practical utility of the asset as it offers utility toward improved diversification metrics of a portfolio. Simultaneously, a significant amount of findings were presented in recent years that suggested the application of forecasting- and trading strategies in the market for Bitcoin should lead to abnormal profits and a concurrent improvement in the risk-return ratio of the asset. If it is possible to gain a lower risk profile on an asset that simultaneously exhibits high levels of volatility and weak correlation with other assets, the investor can benefit from Bitcoin's diversification properties and still realise abnormal profits in that section of the portfolio. The question that this chapter aims to answer is whether it is possible to keep the diversification benefits of Bitcoin and still realise the abnormal profits that the literature suggests is possible in markets like Bitcoin. The answer to the question of how a variety of these neural networks perform comparatively in a controlled testing environment, however, must be tested and this is done in the following chapter. The methodology is easily expanded, however, to the problem of keeping diversification properties and adding short-term profit generation to a portfolio. This is done by means of a specific trading strategy. Determining the optimal way to test varying neural network architectures is difficult when nothing else is considered. However, from the world of trading strategies, it has been found that investing in two assets that bear little or no correlation diversifies a portfolio. Similarly, investing in assets that exhibit strong negative correlation leads to hedging. What the one asset gains, the other therefore loses. Consider then, that it is possible to fully hedge capital in a currency market (by investing in both sides of the same exchange rate), and then adjusting the weights of the assets according to the relevant forecast. It is through this mechanism that one can expect to gain short-term profit (reliant on forecasting accuracy), and still remain fully hedged whilst simultaneously enjoying the benefits of diversification properties of the separate instruments (if they exhibit them). Let us reserve this matter for a moment, to first consider the error rate of a forecast that could inform such a short-term trading decision.

The prediction of mature financial markets such as the markets for stocks and bonds has been researched at length in past research endeavours (Kaastra & Boyd, 1996). However, it is expected that past results obtained for these asset classes may not be directly extrapolated to the market for Bitcoin, as Bitcoin seems to have fundamental characteristics that differ from those of the more established markets.[2] One example of this is the prevalence of high levels of volatility and heteroscedasticity in the Bitcoin market (Briere *et al.,* 2013). These patterns do not exclude volatility-clustering and may be among the chief causes of the weak dependency between Bitcoin and other traditional assets. This brings about an opportunity for investors (or researchers) looking to attempt predictive modelling on Bitcoin data.

Given the discussion of time series modelling techniques in the preceding chapter, it is logical to narrow down the models which should undergo testing in the context of the problem statement. The first and most obvious to exclude (due to the univariate nature of the analysis conducted here) is the VAR family of models which are suitable for multivariate time series analysis (Chatfield, 2016). A similarly simple exclusion is the ARCH family of models along with its variants that consider conditional heteroscedasticity. While Bitcoin price time series may be likely to exhibit ARCH properties, their modelling is not the purpose of this research. It is therefore stated that a conditional heteroscedasticity perspective is most welcome as an addition to any of the models that are presented here, but simultaneously that the implementation of ARCH models as a stand-alone predictive process lacks the ability to signal a profit maximising agent appropriately as to the forecasted direction of the market (Awartani & Corradi, 2005). In future research, the forecasting solutions found to be superior here will almost definitely benefit from the inclusion of ARCH-modelling to ensure appropriate adjustments in investment behaviour pertaining to positive correlations perhaps between volatility and model austerity. However, the scope for such fine-tuning may remove from the rigour and robustness of this comparison.

Yet another simple exclusion would be fundamental and technical analyses based on their high levels of subjectivity (Lo *et al.,* 2000). These are the only *a priori* exclusions of the techniques described in chapter three, and therefore all but these exclusions mentioned above are carried over to the testing environment. From the domain of classical linear statistical models, some specifications are estimated here, with no specific argument based upon their results, but rather they are serving as a benchmark with which to emphasise the forecasting ability of neural networks. The model configurations that will be included are: AR and two time series techniques incorporating exponential smoothing, namely Simple Exponential Smoothing (SES), and Holt-Winters Exponential Smoothing (HWES). This considers the domain of classical linear statistical forecasting techniques well-represented in terms of a benchmark with which to provide context to the forecasting prowess of neural networks. The discussion can thus continue to the domain of machine learning techniques.

---

[2] See chapter 2 for an in-depth discussion on the differences between Bitcoin and other asset classes.

However, even when the domain of forecasting techniques is narrowed down to machine learning, the question of which of the many architectures (discussed in chapter 3) and their variations to bring over to a testing environment, still remains. The answer to this problem is found in the literature for natural language processing. Given the complexity of price-related phenomena in the Bitcoin market, deep learning techniques make for an interesting technologically driven solution based on their performance in areas that can be argued to be quite similar to the Bitcoin market in certain ways. Let us consider, for example, the domain of natural language processing (NLP), and the deep learning techniques that have been proven highly effective therein. NLP is, theoretically, in certain ways comparable to the problem of forecasting the price for Bitcoin, in the sense that it also deals with sequences of data points and techniques to find and predict patterns in such sequences.

Such a scientific inquiry into the literature of NLP narrows down the dauntingly vast (and ever-developing) domain of neural networks from which to choose to those which include temporal sensitivity, i.e. the RNN and CNN groups. Further investigation into these groups, their suitable subtypes and why they may offer effective solutions to the problem statement is done in the comparison section of this chapter. Before this is done, however, the attention is temporarily diverted first to the dataset in question. This is done in order to ensure that all techniques discussed in this chapter are employed with the utmost sensitivity to the underlying data upon which they are fit. Before any estimations of prediction are done, the data will first be described and visualised in the following section.

3.1 Investigating the lag structures and autoregressive properties in Bitcoin's price

The dataset in question is restricted to a univariate analysis of the daily price of Bitcoin, because Bitcoin was indeed (in retrospect) a speculative bubble. In keeping with the literature, it is therefore expected that autoregressive properties are to be found in Bitcoin's price fluctuations (Hencic *et al.,* 2015). However, expanding the dataset to a multivariate analysis could possibly hide autoregressive effects and simultaneously make results between models less comparable. Seeing as the focus here is on Bitcoin as a speculative asset, and to compare models in that context, the price values at t-x timesteps are expected to yield meaningful results. The dataset originates from the online database CoinMarketCap and contains a single field of daily prices for Bitcoin, measured in dollars. The first ten rows are printed below:

Table 3-1: The first ten rows of the univariate Bitcoin price dataset

Price
135.30
134.44
144.00
139.00
116.38
106.25
98.10
112.90
115.98
112.25

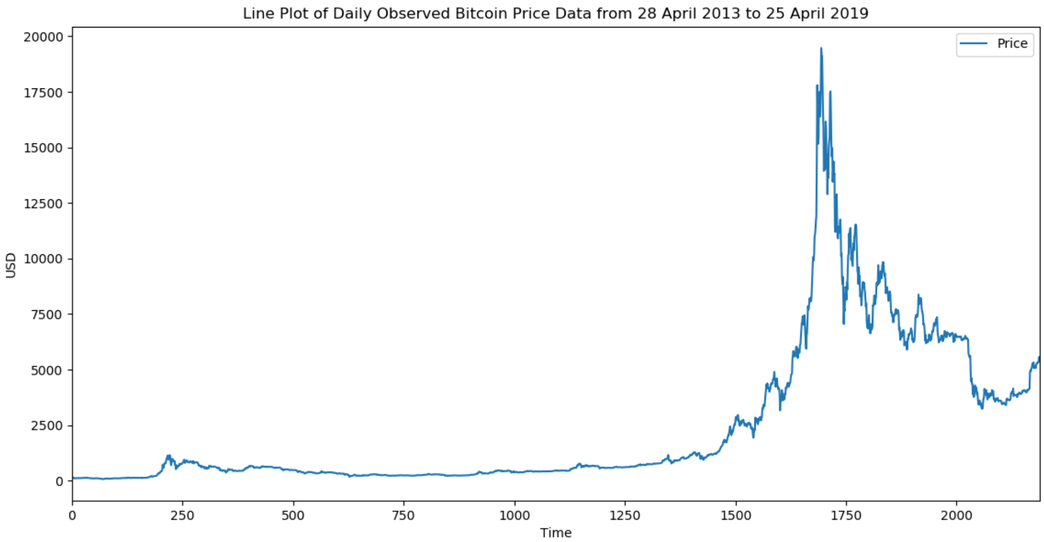
The total number of observations in the dataset is equal to 2189, separated by 24 hour increments. The data, therefore, is observed from 28 April 2013 to 25 April 2019. Some preliminary descriptive statistics of the series can be described as follows:

Table 3-2: A summary of the descriptive statistics

Price	
Mean	2401.116
Standard Error	71.64731
Median	625.97
Mode	418.42
Standard Deviation	3352.145
Sample Variance	11236876
Kurtosis	3.871929
Skewness	1.941181
Range	19407.3
Minimum	68.5
Maximum	19475.8
Count	2189

It is possible to see the dramatic fluctuations in price values over a relatively short period of time. In order to gain insight into the range of dates during which such large fluctuations took place, a line graph of the data is presented in figure 4-1 below.

Figure 3-1: Line graph of the daily observed price data from 28 April 2013 to 25 April 2019



Through visual inspection of the above line plot of the time series in question, it is already possible to conclude with relative certainty that the time series is not stationary. As this is an investigation of the lag structures of Bitcoin price, it is important to draw all conclusions from operations after the data has been transformed to a stationary time series. An augmented Dickey-Fuller (ADF) test of the format:

$$\Delta y_t = \alpha + \gamma y_{t-1} + v_t \tag{3.1}$$

is conducted first and foremost, in order to determine measurably whether the visual observations of non-stationarity for the price variable are correct. The results are summarised in table 3-3 below.

Table 3-3: The ADF (constant) results before differencing

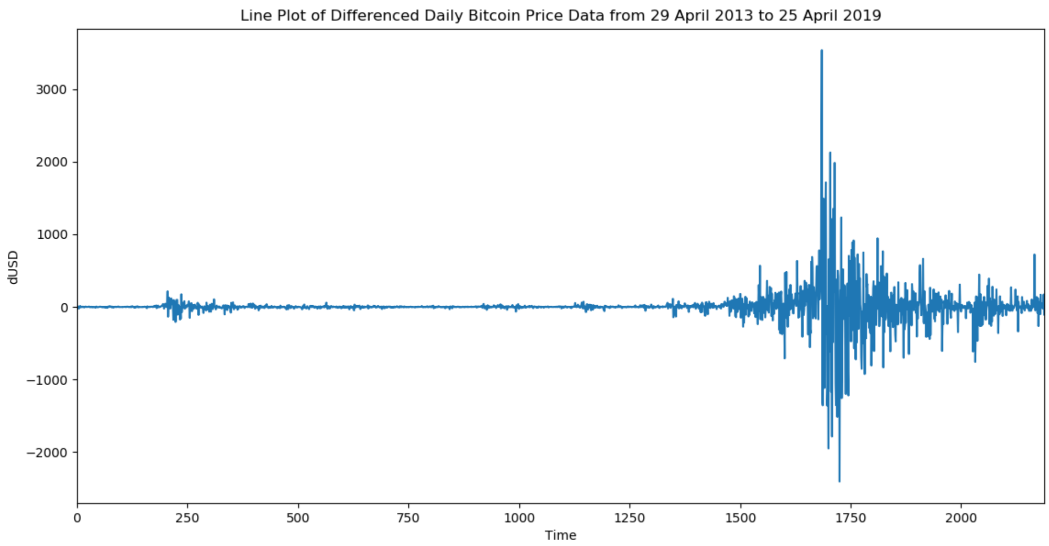
ADF Statistic:	-2.029692
p-value:	0.273708
Critical Values:	
1%:	-3.433
5%:	-2.863
10%:	-2.567

Table 3-4: The ADF (constant and trend) results before differencing

ADF Statistic:	-2.923007
p-value:	0.154893
Critical Values:	
1%:	-3.413
5%:	-3.128
10%:	-3.963

The p-value = 0.273708 (0.154893), therefore the null hypothesis that Price follows a unit root process is not rejected even at a 10% confidence interval. The series is concluded to be non-stationary in non-differenced form as it has been described thus far in both intercept and trend and intercept cases. After differencing the data once, a line plot is generated and presented in figure 4-2 below.

Figure 3-2: Line graph of differenced daily Bitcoin price data from 29 April 2013 to 25 April 2019.



Through visual observation of the line plot of differenced daily prices, it seems more likely that the differenced series is stationary. To determine mathematically whether this is true, the ADF test is carried out once more, with the results summarised in table 4-4 below.

Table 3-5: The ADF (constant) results after first differencing

ADF Statistic:	-8.568892
p-value:	0.000000
Critical Values:	
1%:	-3.433
5%:	-2.863
10%:	-2.567

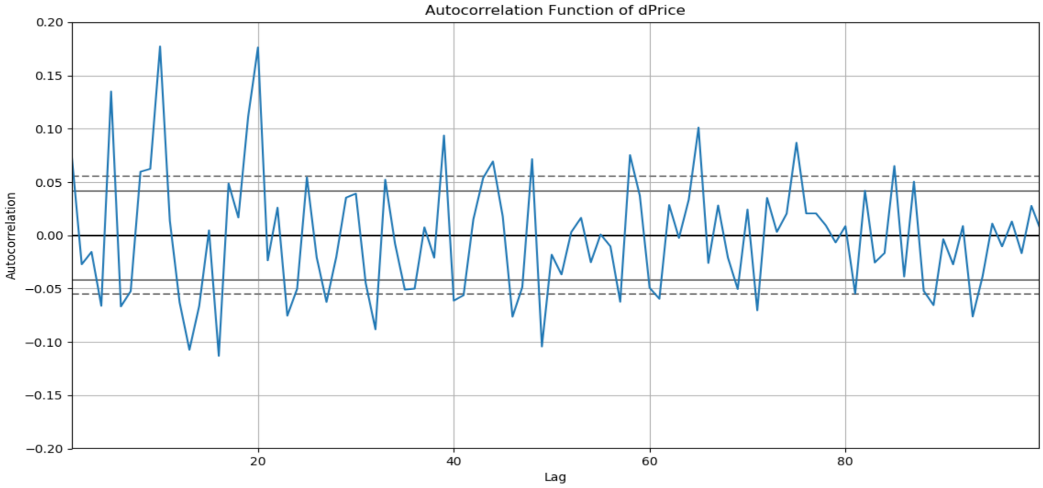
Table 3-6: The ADF (constant and trend) results after first differencing

ADF Statistic:	-8.567257
p-value:	0.000000
Critical Values:	
1%:	-3.128
5%:	-3.413
10%:	-3.963

The p-value = 0.000000, therefore the null hypothesis that Price follows a unit root process is rejected at a 1% confidence interval. The series is concluded to be stationary after first-differencing in both constant and constant and trend specifications.

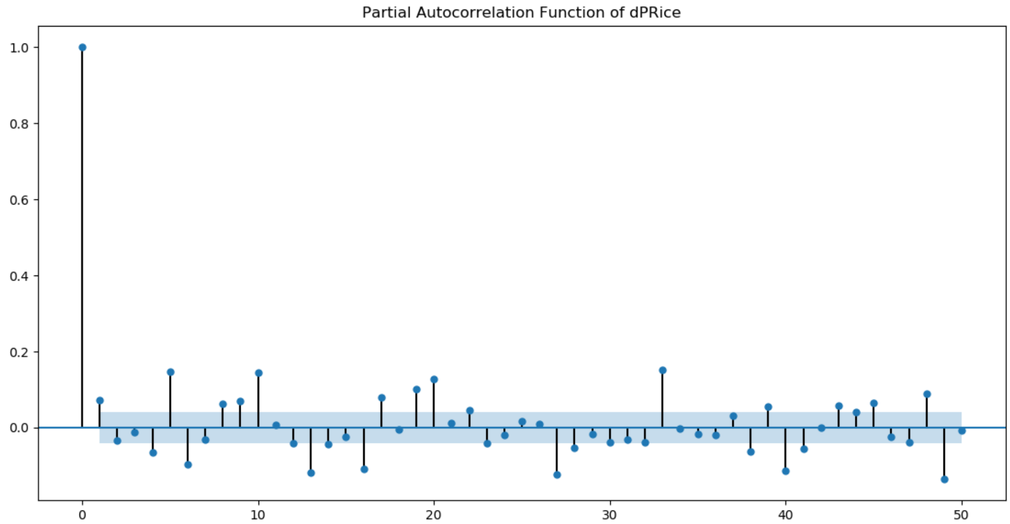
The focus can now be shifted toward finding significant lags that may exhibit forecasting ability. In order to gain further insight into the lag structures of the data, an autocorrelation function (ACF) is estimated on the differenced series and is presented in figure 4-3 below.

Figure 3-3: Autocorrelation function of differenced price values



The ACF plot suggests that the most significant lags can be found at timesteps t-5, t-10, t-15 and t-20. A partial autocorrelation function is estimated in order to gain more insight into the structure of the autoregressive properties of the time series and is presented in figure 4-4 below.

Figure 3-4: Partial autocorrelation function of differenced price



Significant lags with respect to t are once again suggested at t-5, t-10, t-15 and t-20. An autocorrelation matrix for 25 lagged observations of the dependent variable is estimated and presented in table 4-5 below (Only the first column of the autocorrelation matrix is printed here, due to size constraints).

Table 3-7: Autocorrelation matrix (first column only) of the differenced data

	dPrice
dPrice	1
dPrice t-1	0.072874
dPrice t-2	-0.02725
dPrice t-3	-0.01558
dPrice t-4	-0.06618
dPrice t-5	0.135039
dPrice t-6	-0.0668
dPrice t-7	-0.05243
dPrice t-8	0.05967
dPrice t-9	0.062421
dPrice t-10	0.177368
dPrice t-11	0.013619
dPrice t-12	-0.06299
dPrice t-13	-0.10752
dPrice t-14	-0.06625
dPRice t-15	0.004747
dPRice t-16	-0.11322
dPrice t-17	0.048667
dPrice t-18	0.016641
dPrice t-19	0.111495
dPrice t-20	0.176554

The autocorrelation matrix confirms lags that may add to the informational content of a forecast at t-5, t-10, t-20. Plotting these lags on a system of axes produces the lag plots visualised in figures 4-5 to 4-7 below

Figure 3-5: Lag plot showing visually the correlation between price at t and t-5

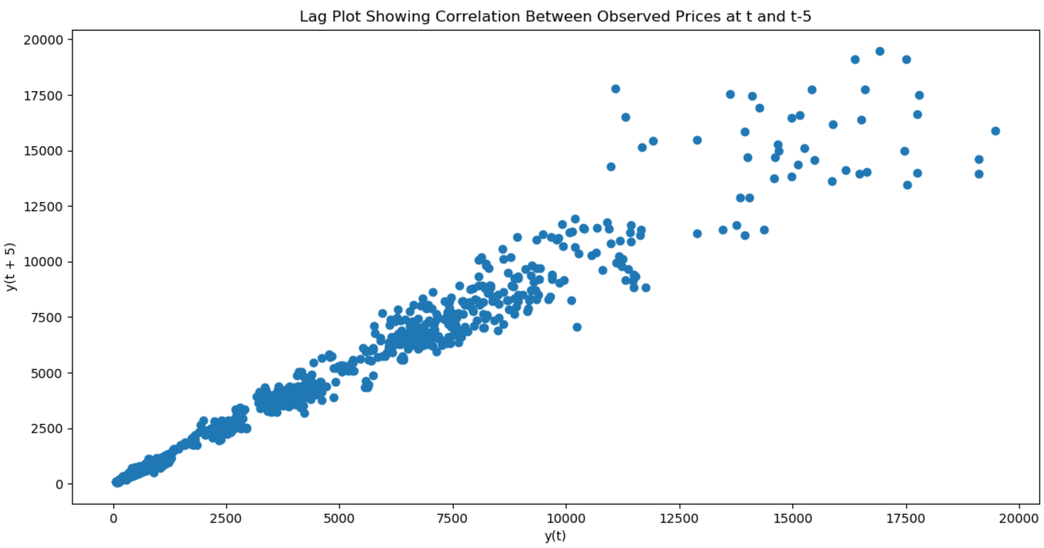


Figure 3-6: Lag plot showing visually the correlation between observations at t and t-10

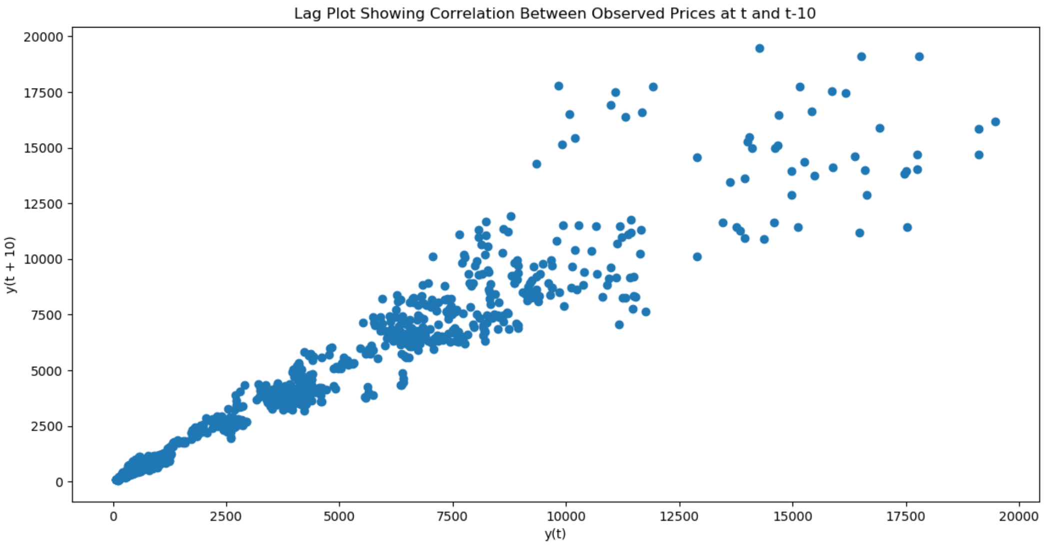
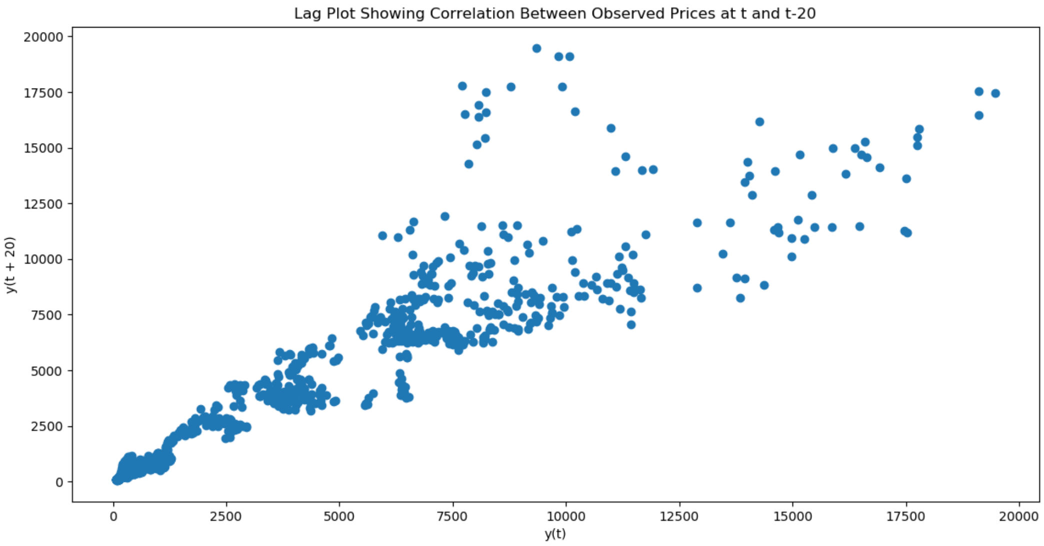


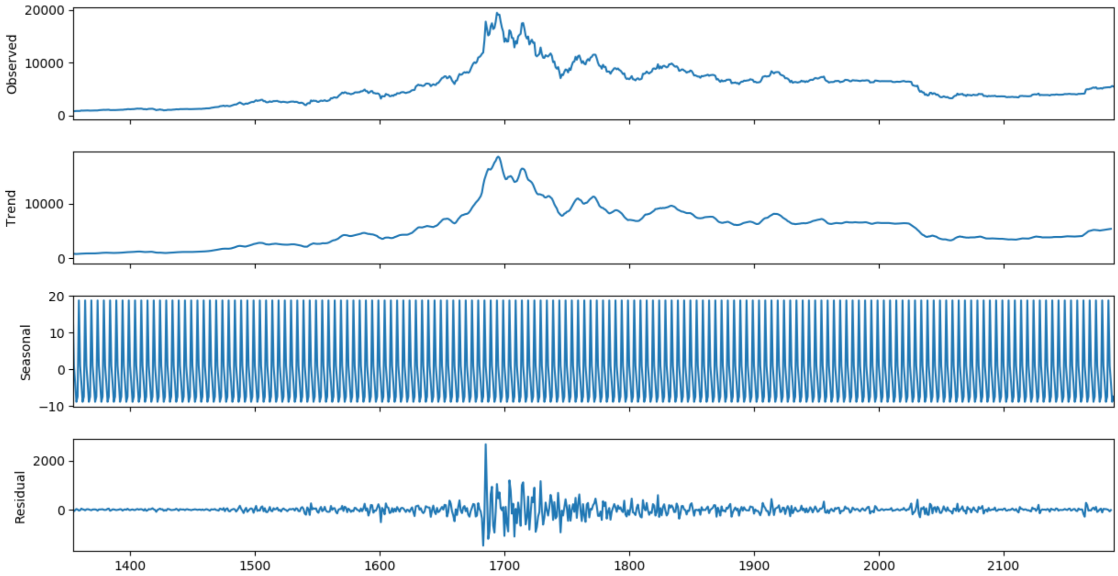
Figure 3-7: Lag plot showing visually the correlation between observations at t and t-20



The lag plots presented in figures 4-5 to 4-7 suggest that prices act more erratically when they are at extreme highs therefore that variance in price expressed as a percentage of price is greater when price exceeds USD 10 000. This leaves only one component of a univariate time series to consider, which is its decomposition into the relevant subparts i.e. trend and seasonal fluctuations.

In order to gain a greater understanding of the seasonal and trend fluctuations in the data, a decomposition plot is estimated and presented in figure 3-8 below.

Figure 3-8: Decomposition plot showing the observed, trend, seasonal and residual values for the data.

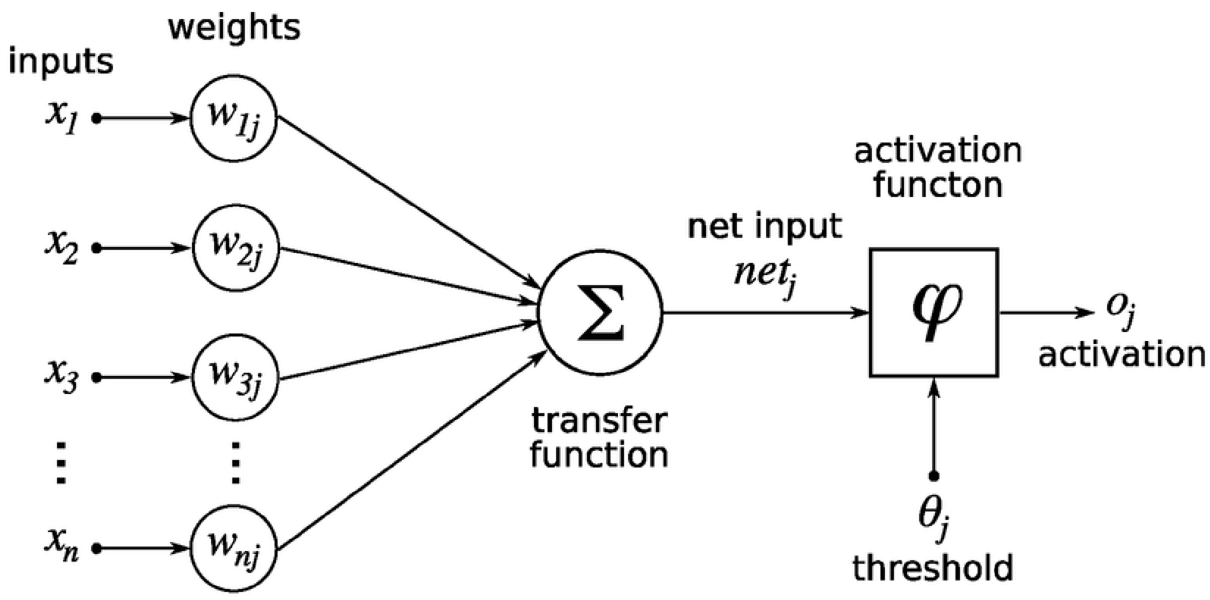


The trend line of the time series visualised in the second subplot of figure 3-8 suggests what has already been formally confirmed by the ADF tests and differencing procedures completed previously in this section. Therefore, it is at best a confirmation that the data is not stationary in non-differenced form and the trend that the data does follow can at least be visually seen in the figure. The seasonal component (subplot 3) of figure 4-8, however, suggests that the data is expressed meaningfully with a seasonal component equal to 5 days (freq=5). This suggests that the use of the observation at t-5 may deliver a meaningful forecast to the observation at time t. It is reported that the frequency set to 5 rendered the decomposition plot in the desired way, and that setting frequency of seasonality to 10 and 20 delivered undesirable results.

Furthermore, with regard to the residuals, it is once more confirmed that the data exhibits more erratic behaviour during times of higher price values. This is known as heteroscedasticity and can be modelled using the ARCH family of models. This refers back to the explanation that can be found in the beginning of this section, pertaining to the fact that while the inclusion of ARCH-modelling may improve arbitrary base-forecasts, it is not suitable to model both the direction and variance of price with ARCH modelling. Therefore, the use of ARCH models is preserved for future research in a conditional specification environment where the forecasting problem may be framed initially as a binary classification problem, with ARCH properties included after classification in order to determine probable magnitude. Seeing, however, as this research aims to compare models, the use of such combinations is not expected to yield meaningful results, and the suggestions from the decomposition plot are therefore limited to influence the models that are suitable for comparison. A short summarising statement of the results is in order as such: firstly, in section 3.1, the autoregressive properties of Bitcoin price were investigated, through which three concrete findings can be presented. Firstly, Bitcoin data is non-stationary in non-differenced form, and stationarity is obtained after first differencing. Secondly, the data exhibits significant (p<1%) autoregressive behaviour in differenced form in terms of the observations at t-5, t-10 and t-20. Thirdly, the data is confirmed to exhibit a seasonal component with a frequency of 5 (days). Therefore, the data was deemed suitable for forecasting with models that relate input to past variables.

Building on the conclusions arrived at by the end of chapter 2 it is now possible to consider the CNN, RNN and LSTM neural networks in greater detail. For supporting arguments as to why these are networks of choice, refer to chapter 3. The description here is meant to be a more detailed, mathematical discussion of the workings of the CNN, RNN and LSTM specifically. The starting point is a visualisation of the workings of the individual neuron:

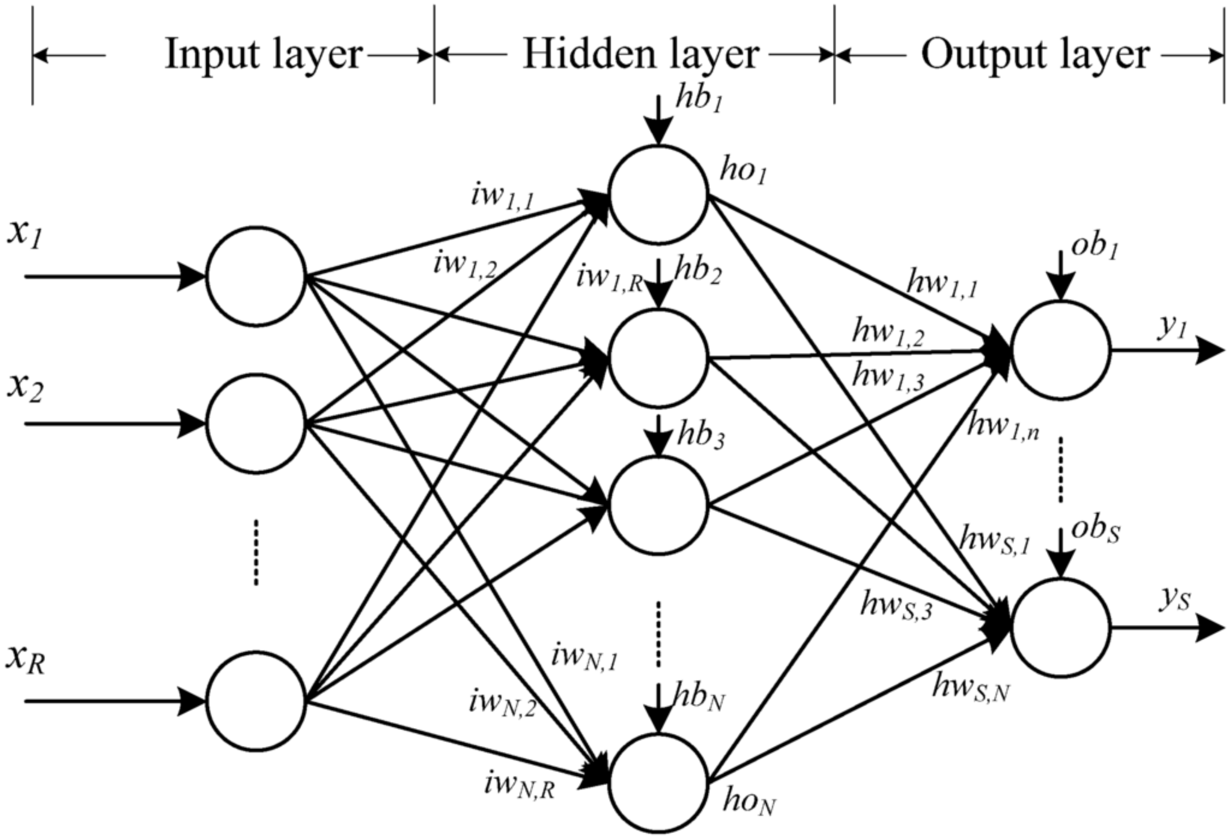
Figure 3-9: Graphical representation of the functioning of a single neuron's operation in a neural network



(source: Keilbar, 2018)

where $\{x_1, x_2, x_3, \ldots, x_n\}$ represent the input signals, $\{w_{1j}, w_{2j}, w_{3j}, \ldots, w_{nj}\}$ represent the synaptic weights, $\sum$ represents the transfer function, φ represents the activation function, θ_j represents the threshold (also referred to as the bias), and o_j the value passed on to the next entity in the network. Coupling an arbitrary number of such neurons together forms a neural network. Organising these groups of neurons into groups – each with its own purpose – forms what is sometimes referred to as the vanilla neural network. Presumably vanilla refers to the simplicity of such a network. The network can be visually represented by breaking down neuronal groups by function and identifying them visually as layers wherein information flows from left to right (feedforward neural network). The vanilla neural network can be visualised as:

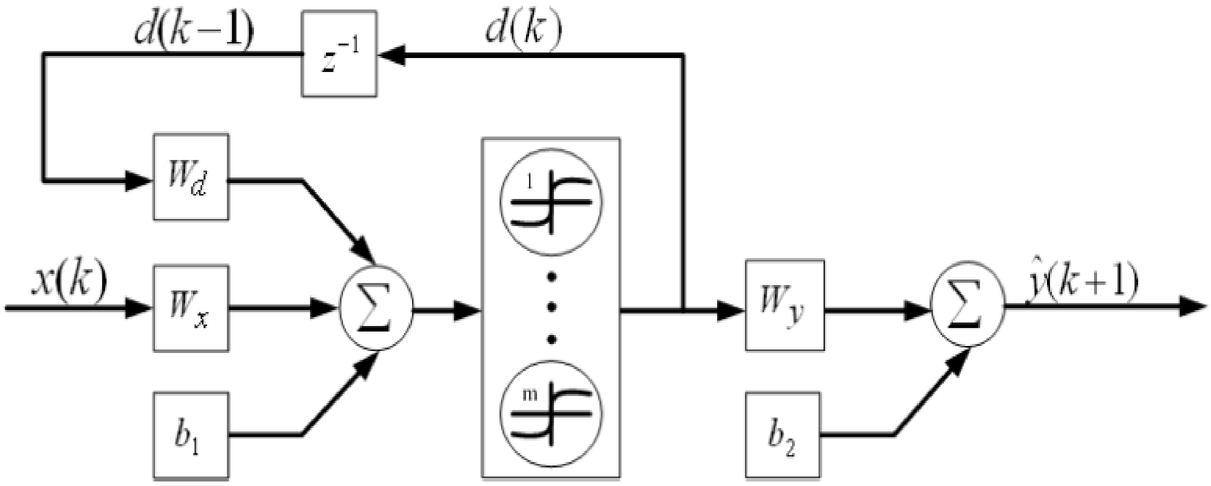
Figure 3-10: Graphical representation of a simple feedforward neural network depicting operations between neurons



(source: Chen *et al.,* 2015)

where $ho_i = f(\sum_{j=1}^{R} iw_{i,j}.x_j \ + \ hb_i)$, for $i = 1,..,N$, and $y_i = f\left(\sum_{k=1}^{N} hw_{i,k}.ho_k + ob_i\right)$, for $i = 1,..,S$. The calculations in each neuron are the same as is explained in figure 4-10 above, which emphasises the fact that a neural network can be expressed as a matrix algebra problem, and the visual representation as in figure 4-10 is merely toward a logistical understanding of informational flows (to liken the functioning of a neural network to that of the human brain, in keeping with the original purpose of the field as is described in the section on the history of neural networks). The description of the vanilla neural network thus far makes it possible to add into the equation the first element that makes the LSTM different from other neural networks, which is its temporal element. The temporal element that is present in the LSTM architecture must be explained from where it originates, which is from the concept of recurrence. The LSTM is a specialised subtype of recurrent neural networks in that it introduces recurrence along with long-term memory neurons. The LSTM is therefore different from the vanilla feedforward network in the way it passes information between neurons (including recurrence) and the way neurons operate individually. Let recurrence be considered first, after which convolution, and finally the focus will be shifted toward the functioning of the individual LSTM neuron. Recurrence can be visualised as follows:

Figure 3-11: Graphical representation of a single neuron capable of recurrence



(source: Molina *et al.,* 2011)

where k represents the time-step, W_x the input matrix, b the bias, d the delay, W_d the delayed input matrix (input from the previous time-step i.e. the recurrence) and $\hat{y}$ the output of the recurrent operation. The same operation can be expressed algebraically as:

$$\hat{y}(k+1) = f(W_x \times x(k) + W_d \times d(k-1) + b_1) \times W_y + b_2 \quad (3.2)$$

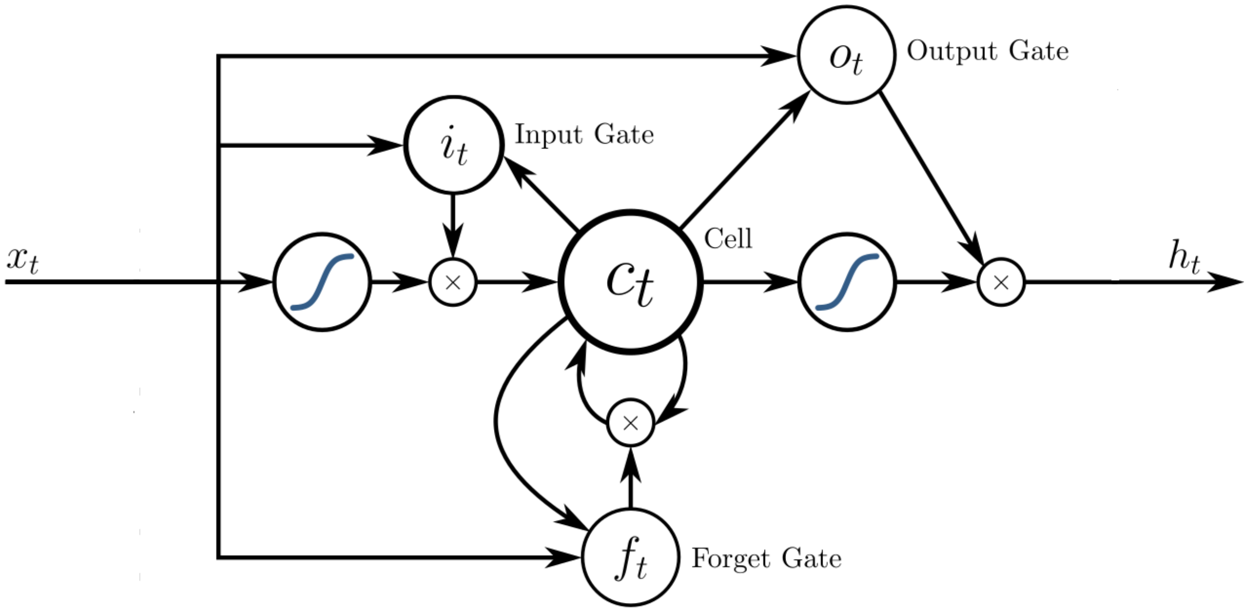
with the variables assigned to the same entities as for the visualisation above. This emphasises once again that even recurrent neural networks are essentially complex matrix algebra problems which can be illustrated by graphic representations to aid in their purpose of imitating the logistical functioning of the human brain. Recurrence, however, does not come without a trade-off, and the negative side of this trade-off is the computational complexity of the operations involved. The process of training RNNs involves calculating the gradient of the error with respect to the parameters in the network as with other neural networks, but the recurrence adds the implication that RNN parameters are altered for all time steps after their impact is measurable. This implies exponential complexity in the matrix algebra that has to be performed when training a neural network through computationally expensive methods such as Real-Time

Recurrent Learning (RTRL), Back Propagation Through Time (BPTT) or truncated BPTT (Prokhorov *et al.,* 2001).

The matter of convolution, even though often mentioned together with recurrence and the LSTM, functions quite differently than these counterparts. What distinguishes the CNN from other networks is not so much the functioning of individual neurons, but rather the way in which outputs of neurons relatively later in the network refer their informational content back to earlier neurons. The CNN's properties have been found particularly useful in image processing and, to a lesser extent, language modelling (for the more detailed discussion of the CNN, refer to chapter 2).

Building on this process, the focus can now be shifted toward the LSTM, which categorises under recurrent neural networks, but adds more complexity in the way patterns in data can alter the parameters of the network to imitate semantic memory even more accurately than has been described up until this point. The LSTM neural network is finally introduced as a whole:

Figure 3-12: Graphical representation of the functioning of the LSTM neuron



(source: Graves *et al.,* 2013)

where x_t represents the input (usually in the form of a vector), f_t represents the forget gate's activation vector, i_t the input gate's activation vector, o_t the output gate's activation vector, c_t the cell state vector (the operation performed on the input), and h_t the output vector of the LSTM unit. The LSTM network is Turing complete, which means it is able to solve any problem a Turing machine can (Siegelmann & Sontag, 1992). While this is already impressive in comparison to other neural networks, the LSTM network is also known for its ability to find patterns in sequences of data. It has been labelled the most commercial achievement in AI, as it aids in a wide range of tasks varying from disease prediction to the composition of music (Vance, 2018). The network is very different from the standard feedforward (vanilla) Multilayer Perceptron as has been described above. However, it even differs quite significantly from its super category, which is the recurrent neural network class of models. What makes the LSTM network different from regular recurrent neural networks is the inclusion of gates that behave like neurons themselves, independently of the cell operation (Hochreiter & Schmidhuber, 1997). This network configuration and variants thereof implicate the network's ability to "remember" given phenomena in sequences of data over arbitrary time intervals. One such of these gates in the LSTM network's architecture independently learns to activate its mathematical operation, given a set of circumstances are found to be true in a given sequence of data that flows through the neuron. These memory gates imitate the functioning of the human brain quite differently from traditional feed-forward neural networks, as patterns within stimuli can now activate the neuron's influence on its neighbours. This is true even if such reaction may be thought to have been dormant in the brain for any given amount of time. The LSTM network was originally created to solve the problem of vanishing/exploding gradients found in normal recurrent neural networks (Hochreiter & Schmidhuber, 1997). Solving this problem allowed the LSTM network configuration to excel at numerous tasks with unprecedented success in a non-supervised environment, to such a vast extent that upon occasion it has been considered a key element in some of the greatest breakthroughs of artificial intelligence (Rodriguez, 2018).

It is due to such findings in the literature that it seems reasonable to include a comparison of an LSTM network's performance with that of CNN's and RNN's (on the same problem). Therefore, these models are tested under the exact same conditions. Each network is iteratively shown an input vector of 50 preceding values and asked to predict one time step into the future. The results are printed to an output file of the following nature (from which the error metrics as presented in chapter 4) are calculated:

Table 3-8: A summarising excerpt from the DAILY MODEL PREDICTIONS file

t	Predicted t+1	Real t+1
112.25	104.6159	109.6
109.6	105.9734	113.2
113.2	114.465	112.8
112.8	113.0575	117.7
117.7	119.2114	115.64
115.64	115.2516	114.82
114.82	116.1174	117.98
117.98	118.1215	111.4
111.4	117.6439	114.22

It is reported that such a file will be produced for every model in the tests in chapter 4, after which error metrics on forecasting will be calculated and, in the section thereafter, a profit maximising agent (a simple computer program) will use this same file as input to the investment decision. In each row of table 3-6 therefore, the results for the optimal model specification are reported. This is achieved with a consistent development pipeline specified as follows: each model begins by considering the first 50 entries of the dataset. The model is specified, estimated and made to predict the 51st day's price value using the learned function (implemented in Python). Upon the next iteration, the exact same model specification is fit to observations 2 to 51 of the dataset and made to predict observation number 52. The third iteration estimates the model using observations 3 to 52 and predicts observation number 53 and so forth until all of the remaining data is predicted. This methodology is performed in order to reach the fourth research objective, and the trading strategy is specified in section 4.2.

This section is concluded then, with a reiteration of the summary of the insights gained from the description, visualisation and preparation of the data in the beginning of this chapter. Firstly, the data is concluded to be stationary after first differencing. Secondly, the data exhibits significant ($p<1\%$) autoregressive behaviour in differenced form in terms of the observations at t-5, t-10 and t-20. Thirdly, the data is confirmed to exhibit a seasonal component with a frequency of 5 periods (days).

CHAPTER 4 RESULTS AND DISCUSSION

This chapter aims firstly to determine the extent to which various time-sensitive neural networks possess the ability to minimise error in a forecasting environment and, secondly to determine whether fund managers can draw benefit from Bitcoin and assets with similar market complexities through a trading strategy, whilst still benefiting from the diversification properties. In the previous chapter, the data and descriptive statistics on the involved Bitcoin time series have been reported. Given, however, the complex nature of financial technology, it is expected that time-sensitive time series modelling techniques such as the recently popularised field of neural networks may exhibit superior predictive capabilities compared to more archaic methods. Firstly, the forecasting error of the three best-performing neural networks is reported for comparison, after which the findings will be placed back in the context of an investment portfolio as per the fourth objective.

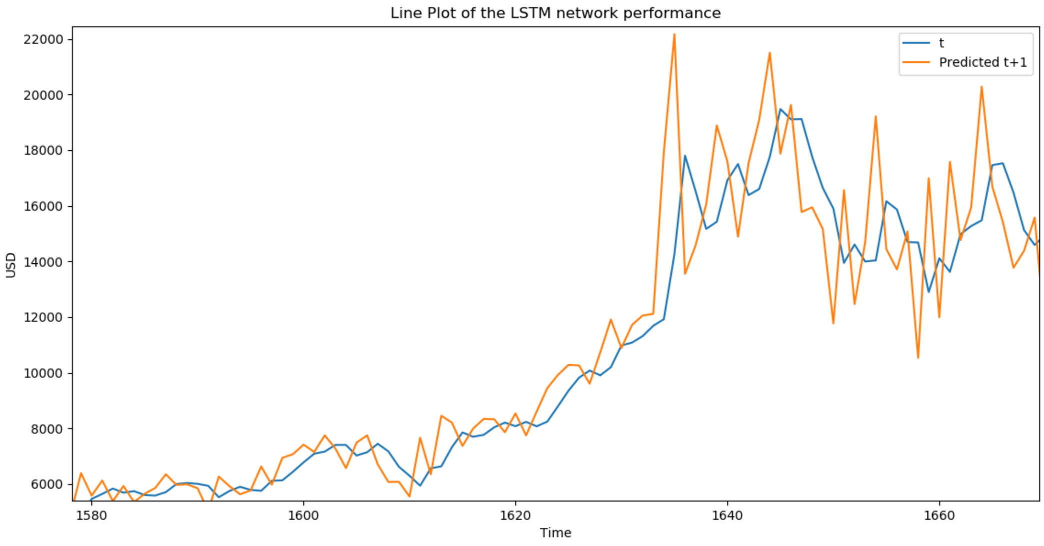
Before the results are discussed, however, one small issue that is often brought up in the literature regarding neural networks must first be addressed. This issue is that of training time. Some problems become extremely mathematically complex when they are modelled with neural networks. So much so, in fact, that it time becomes a consideration to trade off with accuracy. In the current specification, however, the problem was still comparatively simple. The input vector to this problem at no point exceeded 50 entries, and the output vector was a single number throughout all of the experiments. The training times were therefore negligible. The longest any of the neural networks took to train was a couple of seconds. Iteration through the entire dataset (rolling forecast equal to 50 input observations, one output observation and increments equal to one) took under 7 minutes on an Nvidia GTX 550 Ti GPU. If more input data is considered this will surely increase, but at this scale it is almost irrelevant. Let the results then be reported in the following section.

4.1 Determine the extent to which neural networks are able to forecast Bitcoin price

This section aims to test whether this is true, as pertains to the third research objective (delineated in chapter one). The objective of this section is therefore to determine the extent to which three promising neural network architectures are able to minimise the forecasting error in a time series prediction problem using the Bitcoin dataset. The problem of specifying a suitable neural network configuration is described at length in chapter 2 and will be summarised here along with more detail being provided about the specific architectures in question: the LSTM, CNN and RNN. The RNN, CNN and LSTM included the input layer and 50 hidden nodes in one layer. The CNN included 64 filters, a kernel size = 2 and max pooling size = 2. All networks used the “relu” activation function, “adam” optimiser and minimised the “mse” loss function.

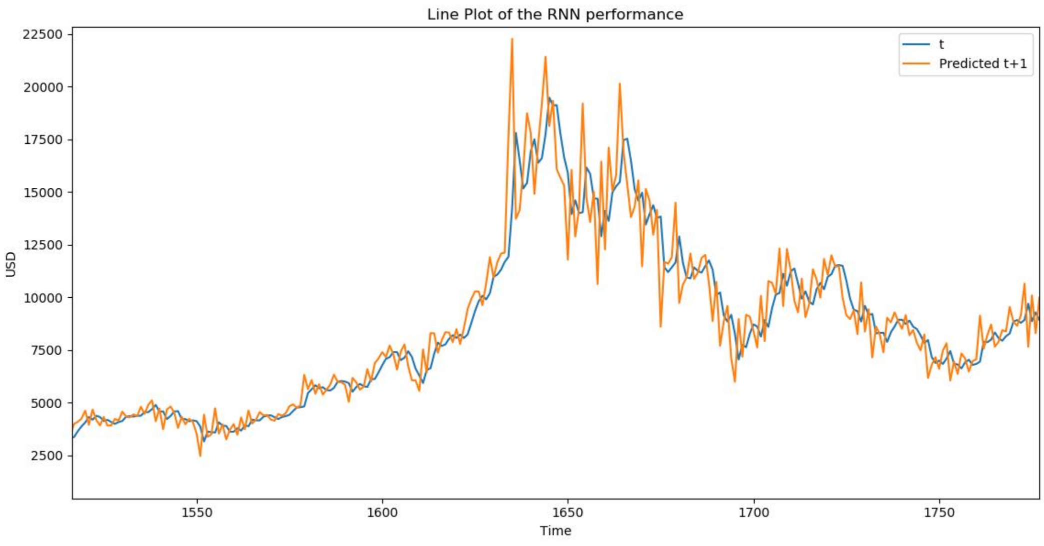
Before the results are reported, a simple line graph of the LSTM's performance can be visualised as follows:

Figure 4-1: Line graph (excerpt) of the LSTM network's forecasting performance



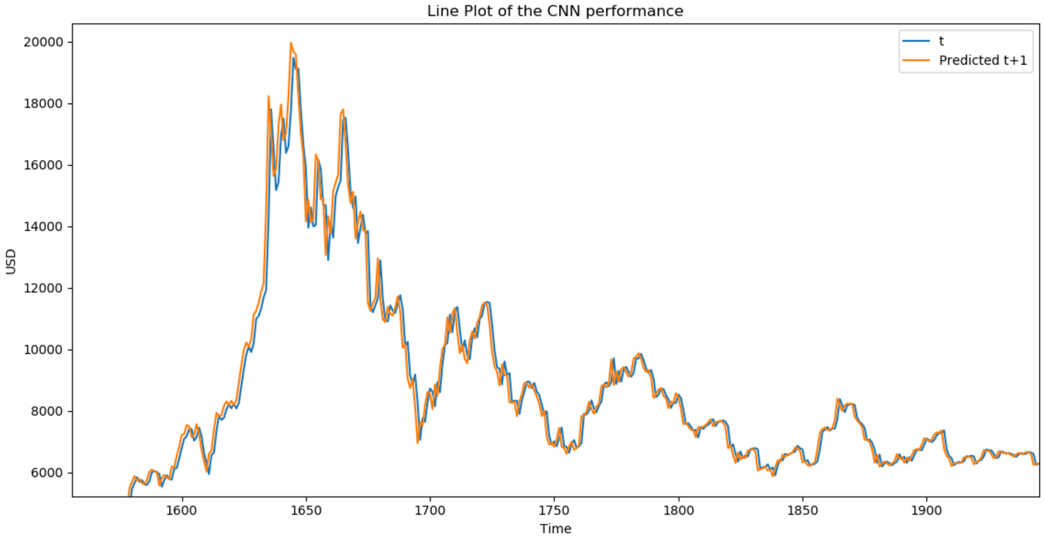
Similarly, that of the RNN:

Figure 4-2: Line graph (excerpt) of the RNN's forecasting performance



And finally, the same graph as pertains to the CNN:

Figure 4-3: Line plot (excerpt) of the CNN's forecasting performance



Closer visual inspection reveals an example of the LSTM, CNN and RNN's ability to anticipate a drop in price 24 hours in advance, even after consecutive days of upward pressure – whilst considering as input only the previous 50 days' price observations. In order to draw the comparison as meaningfully as possible, the results achieved by these networks are summarised in the table below, as measured by various error metrics.

Table 4-1: A print-out of model performance

Model	MAPE	RMSE	RMSLE
AR(5)	3.65	309.16	0.06
SES	2.87	246.34	0.05
HWES	2.86	238.60	0.05
LSTM	5.36	333.02	0.21
CNN	0.56	50.15	0.01
RNN	3.177	304.46	0.05

From table 4-1, it is possible to see that the model which minimised the error metrics is the CNN, of which the MAPE=0.56, RMSE=50.15 and RMSLE=0.01. However, these results (even though they may seem promising), will add value to the empirical literature and risk management practitioners only if it is contextualised in terms of the effects that it could have on the risk-return ratio of an investment portfolio. The discussion of these performances is therefore reserved for the next section.

4.2 Placing the forecasting results back into the context of an investment portfolio

This section aims to emphasise the contribution of these findings to the fund manager-portion of the intended audience of this research. This is done by determining whether fund managers are able to benefit from Bitcoin simultaneously as an active-trading instrument and a portfolio diversifier. The *a priori* null hypothesis is therefore that fund managers are able to simultaneously benefit from Bitcoin as a trading instrument and as a portfolio diversifier. Bitcoin’s value proposition as a portfolio diversifier is well-established. However, the mechanism through which speculative profit is expected to be added to the value of diversification relies on the trading strategy. Apart from the semantics of the forecasting process and the availability of market data, Bitcoin has been found to be weakly correlated with other. This speaks to the practical utility of the asset as it offers utility toward improved diversification metrics of a portfolio. Simultaneously, a significant amount of findings were presented in recent years that suggested the application of forecasting- and trading strategies in the market for Bitcoin should lead to abnormal profits and a concurrent improvement in the risk-return ratio of the asset. If it is possible to gain a lower risk profile on an asset that simultaneously exhibits high levels of volatility and weak correlation with other assets, the investor can benefit from Bitcoin’s diversification properties and still realise abnormal profits in that section of the portfolio. The question that this chapter aims to answer is whether it is possible to keep the diversification benefits of Bitcoin and still realise the abnormal profits that the literature suggests is possible in markets like Bitcoin. The answer to the question of how a variety of these neural networks perform comparatively in a controlled testing environment, however, must be tested and this is done in the following chapter. The methodology is easily expanded, however, to the problem of keeping diversification properties and adding short-term profit generation to a portfolio. This is done by means of a specific trading strategy. Determining the optimal way to test varying neural network architectures is difficult when nothing else is considered. However, from the world of trading strategies, it has been found that investing in two assets that bear little or no correlation diversifies a portfolio. Similarly, investing in assets that exhibit strong negative correlation leads to hedging. What the one asset gains, the other therefore loses. Consider then, that it is possible to fully hedge capital in a currency market (by investing in both sides of the same exchange rate), and then adjusting the weights of the assets according to the relevant forecast. It is through this mechanism that one can expect to gain short-term profit (reliant on forecasting accuracy), and still remain fully hedged whilst simultaneously enjoying the benefits of diversification properties of the separate instruments (if they exhibit them). The key would be to hold Bitcoin and whatever currency is chosen to trade in (in this case USD). This will result in equal gains/losses if the exchange rate fluctuates. However, it should be possible to change only the weights of these two assets in a perpetual rebalancing trading strategy. Thereby holding slightly more of the expected-to-appreciate asset every time that the forecasting procedure suggests a price fluctuation, and conversely slightly less of the expected-to-depreciate currency. The margin with which these weights change is variable, and 10% of the expected-to-depreciate asset is proposed (based on Bitcoin's high volatility). Such a trading strategy can be expressed in pseudocode as below.

Trade strategy expressed in pseudocode

```
Start
Initialize currency variables:
 USD = 100
 BTC = 1.0
for every ROW in DAILY MODEL PREDICTIONS:
    if Predicted Price at t+1 > Price at t:
        Buy BTC equal to 10% of USD holdings @ current price
    else:
        Buy USD equal to 10% of BTC holdings @ current price
print(USD)
print(BTC)
End
```

Let this program be referred to as the investment decision agent, whose task is to perpetually rebalance the USD:BTC holdings in the slight (currently 10% of the holdings in the expected-to-depreciate asset) favour of the expected-to-appreciate asset. It is expected that this strategy will enable speculative profit as informed by the machine learning models, whilst still preserving sufficient exposure to both assets so as to not trade off the diversification properties of Bitcoin. The forecast methodology conjoined with the trading strategy should also be able to counteract any structural breaks in the data, as the traditional train/test split fails to do. By relying only on the previous 50 days' price data in order to forecast the fluctuation on the next day, the methodology is free from influence of market forces that manifested during a different timeframe. If therefore, there exists any structural breaks, the model will automatically (through the application of a rolling window forecast conjoined with the trading strategy) only consider the market structure as it is at any given time step. This drastically reduces the risk of a structural break causing misleading results, as the model does not rely on data that stretches far from the current time step. Let this trading strategy be implemented in a back-testing scenario and the results reported therefrom.

Table 4-2: A reprint of model performance, including results obtained during back-testing

Model	True Classifications	MAPE	RMSE	RMSLE	Profit/Loss
AR(5)	1045	3.65	309.16	0.06	2 171.46
SES	1029	2.87	246.34	0.05	380.63
HWES	1029	2.86	238.60	0.05	479.51
LSTM	1888	5.36	333.02	0.21	84 830.43
CNN	1980	0.56	50.15	0.01	120 284.28
RNN	2008	3.177	304.46	0.05	102 341.81

From table 4-2, it is possible to conclude that there exists significant profit-generating opportunities in short term rebalancing of opposing currencies like BTC and USD. Let this section (and therewith this chapter) be concluded then with a summary of the results obtained in the respective sections. The model that minimised the MAPE, RMSE and RMSLE was the CNN. The classical linear statistical model that maximised profit during back-testing was the AR(5) model. The machine learning model that maximised profit during back-testing is the CNN. Therefore, the CNN exhibited the greatest ability to minimise error and maximise profit when conjoined with this particular trading strategy. The implications of these results for the theoretical and empirical contribution of this study, as well as for the risk management practitioner are discussed in the final chapter.

CHAPTER 5 CONCLUSION AND FUTURE WORK

5.1 Conclusion

The evolution of financial technology correlates positively with the complexity of financial markets. This occurs in two forms, the first of which is the process of technological integration in traditional markets (in which case the effects can be explained relatively effectively as exogenous variables to the existing system). The second form is the rapid demand-driven development of new financial markets for non-traditional, inherently technological assets (in which case the effects manifest significantly more chaotically as they are endogenous to the system). The latter was the responsible mechanism for the research undertaken here. This is so, because even though fintech-associated markets exhibit significantly more adverse risk profiles than traditional asset markets, the empirical literature still suggests a value proposition in these markets both for traders and portfolio managers. This value proposition is firstly in the form of diversification and secondly in the form of a trading strategy-based profit generating technique. Where the literature suggested several such trading strategies to be effective, this research aimed to determine how the underlying forecasting methodologies of these trading strategies compare in a controlled testing environment and how their application could impact portfolio risk. The gap is that this research aimed to address, was therefore that, even though the literature suggests the significance of AI-based trading results in fintech markets, these findings are fragmented, and it is unclear how they may generalise in the greater context of portfolio management amidst rapid global fintech advancement. A significant amount of focus in the empirical literature was toward testing the field of neural networks in terms of this problem. Even though this has been done on several occasions, the question remained how the previous findings offer a value proposition to an investment portfolio. More specifically, it needed to be determined if the conjunction of a trading strategy could enable risk managers to benefit from the well-established diversification properties in Bitcoin, whilst still generating abnormal trading profits in the short-run.

The general objective of this paper was therefore to leverage the established relationship between financial technology and financial market complexity in order to propose a theoretical explanation for recent empirical findings that suggest significant speculative wealth creation implications in the AI-driven Bitcoin investment domain of the literature. The relevant theory was materialised through the application of empirical tests regarding the extent to which market characteristics associated with higher levels of financial technology may be utilised in contributing toward an improved risk-return profile for a given investment portfolio. The theoretical value added was in the conceptualisation of fintech as a playing field upon which the forces of risk and return interact. The empirical contribution was to provide a comparison between the various existing “AI-in-Bitcoin trading” methodologies in terms of the extents to which they are able to mitigate the adverse risk profiles in complex, fintech-endogenous markets. The contribution was emphasised by placing the findings back into the context of an investment portfolio.

This general objective was broken down into specific objectives so as to provide an answer to the research question, and thereby solve the problem statement. These specific objectives were to provide a theoretical framework through which the link(s) between financial technology and different forms of financial market complexities may be understood, to provide an empirical overview of the real-world market effects of the link between financial technology and financial market complexities, and how these complexities have been utilised toward speculative wealth creation in the past, to compare the performance of various neural network architectures in the context of a speculative wealth creation problem and to place the findings back into the context of the risk-return profile of an investment portfolio.

The process of meeting the first objective illustrated fintech as a playing field upon which the forces of risk and return interact. This showed that financial technology and finance are inherently inseparable phenomena and, that even though there exists clear interdependency with increasingly short feedback loops, the relationship of cause and effect between fintech and market complexity is not clear enough.

In meeting the second objective, various market complexities (which were explore during the first objective) were found in fintech-endogenous markets for which Bitcoin served as a case study. Also, the extent to – and the ways in which these market complexities had been leveraged in the past were explored in terms of how they could contribute to the findings here. It was found that neural network-based price forecasting is a phenomenon that is well-represented in the literature, specifically as pertains to fintech-endogenous markets like Bitcoin. Whether this is because of inherent aspects of the two concepts, or rather due to their concurrent popularity is still unclear.

In the pursuit of reaching the third specific objective, various neural networks were estimated and compared in terms of their ability to minimise the error in a forecasting problem. Statistical models were provided as a benchmark. It was found that neural networks possess a significant ability to minimise error in Bitcoin price data. This concurs with empirical findings from the literature. The specific neural network architecture that performed best in terms of error metrics was the CNN.

In the process of reaching the fourth and final research objective, the forecasting abilities of the neural networks were placed in the context of a back-tested trading strategy with very specific properties. The aim was to determine whether it is possible for the modern risk manager to realise an abnormal profit in the short-run through active speculation informed by AI, whilst remaining well-hedged (as exposure to both sides of the exchange rate is maintained) and simultaneously benefitting from the empirically-established diversification properties of Bitcoin. It was found that this is indeed possible, which is emphasised as the key finding of this research.

The implication for the risk management practitioner is that the specific trading strategy used here, joined with an appropriate forecasting technique (such as the NN's provided) can provide well-hedged, abnormal short-term profit without trading off the diversification properties of the underlying assets.

The conjoined achievement of the four specific research objectives provided an answer to the research question, which solves the problem statement of this research.

5.2 Future work

The most pertinent point in terms of future work is that machine learning results are expected to be much more effective if the amount of training examples is expanded, as this is expected to result in more accurate function approximation. Therefore, a logical test would be to determine the informational content added by more information when synthesised into artificial intelligence by neural networks, and then compare the cost of data-housing, in order to determine the exact trade-off between expenditure on storage and improvement of forecasting results.

The second point of future work to be done is to expand the problem to a multivariate analysis. The nature of this study placed a logical limit on the amount of variables that could be included. The comparison would not have been as clear between the models involved if more variables had been included. Now, however, that it has been established that neural networks are worth testing in more challenging environments and how their underlying methodologies compare in a controlled environment, it is considered very suitable to expand the variables that the neural network could take into account to produce forecasts.

The third point of recommended future work is added granularity of data. Given the highly effective success rates of profit maximisation strategies presented here (on daily data), it is expected that adding greater granularity to the data may produce improved results. While more granular data may not be as readily available, it is very possible to obtain through API integration with some of the major exchanges.

Miscellaneous recommendations for future work include improved neural network hyper-parameters and even overarching specifications, i.e. deeper stacked layers of neural networks. The inclusion of ARCH-GARCH modelling in the context of a binary classification problem is also recommended. Finally, improved pre-processing methods are recommended (especially pertaining to methodologies arising from the field of spectral analysis). The future research opportunities combining Artificial Intelligence and Finance are truly abundant.

REFERENCE LIST

Aarts, E.H. & Korst, J.H. 1989. Computations in massively parallel networks based on the Boltzmann machine: a review. *Parallel computing*, 9(2):129-145.

Abarbanell, J.S. & Bushee, B.J. 1997. Fundamental analysis, future earnings, and stock prices. *Journal of accounting research*, 35(1):1-24.

Agarwal, S., Amromin, G., Ben-David, I., Chomsisengphet, S. & Evanoff, D.D. 2014. Predatory lending and the subprime crisis. *Journal of financial economics*, 113(1):29-52.

Agustsson, E., Tschannen, M., Mentzer, F., Timofte, R. & Van Gool, L. 2018. Generative adversarial networks for extreme learned image compression. *arXiv preprint* arXiv:1804.02958.

Akusok, A., Lendasse, A., Corona, F., Nian, R. & Miche, Y. 2013. ELMVIS: a nonlinear visualization technique using random permutations and ELMs. *IEEE intelligent systems*, 28(6):41-46.

Alessi, L., Barigozzi, M. & Capasso, M. 2008. A review of nonfundamentalness and identification in structural VAR models.

Alsudairi, M.A. & Vasista, T. 2012. Successful Implementation Of Electronic Bill Payment and Presentment System–A Sadad Case Study. *IMT case journal*, 2(2):1-12.

An, J. & Cho, S. 2015. Variational autoencoder based anomaly detection using reconstruction probability. *Special Lecture on IE*, 2:1-18.

Anabtawi, I. & Schwarcz, S.L. 2013. Regulating ex post: how law can address the inevitability of financial failure. *Texas law review*, 92:75.

Antonelo, E.A., Schrauwen, B. & Van Campenhout, J. 2007. Generative modeling of autonomous robots and their environments using reservoir computing. *Neural processing letters*, 26(3):233-249.

Anzali, S., Gasteiger, J., Holzgrabe, U., Polanski, J., Sadowski, J., Teckentrup, A. & Wagener, M. 1998. The use of self-organizing neural networks in drug design. *Perspectives in drug discovery and design*, 9:273-299.

Arinaminpathy, N., Kapadia, S. & May, R.M. 2012. Size and complexity in model financial systems. *Proceedings of the National Academy of Sciences*, 109(45):18338-18343.

Arner, D.W. 2016. FinTech: Evolution and Regulation. Режим доступу: http://law. unimelb. edu. au/__data/assets/pdf_file/0011/1978256/D-Arner-FinTech-Evolution-Melbourne-June-2016. pdf.

Asteriou, D. & Hall, S. 2011. Applied Econometrics. 2nd ed. Hampshire, UK: Palgrave Macmillan.

Athey, S., Parashkevov, I., Sarukkai, V. & Xia, J. 2016. Bitcoin pricing, adoption, and usage: Theory and evidence.

Athiwaratkun, B. & Stokes, J.W. 2017. Malware classification with LSTM and GRU language models and a character-level CNN. (In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing* (ICASSP) p. 2482-2486).

Atsalakis, G.S., Atsalaki, I.G., Pasiouras, F. & Zopounidis, C. 2019. Bitcoin price forecasting with neuro-fuzzy techniques. *European Journal of Operational Research*, 276(2):770-780.

Banner, S. 1997. What causes new securities regulation--300 years of evidence. Wash. *ULQ,* 75:849

Barberis, J. 2012. The 2007 Meltdown – A Legal Phenomenon. University of Birmingham, p. 60.

Barratt, M.J. 2012. Silk Road: E-bay for drugs. *Addiction*, 107(3):683-683.

Bartók, A.P., Gillan, M.J., Manby, F.R. & Csányi, G. 2013. Machine-learning approach for one- and two-body corrections to density functional theory: applications to molecular and condensed water*.* *Physical Review B*, 88(5):054104.

Bartram, S.M. & Bodnar, G.M. 2009. No place to hide: The global crisis in equity markets in 2008/2009. *Journal of international Money and Finance*, 28(8):1246-1292.

Bashir, M., Strickland, B. & Bohr, J. 2016. What motivates people to use bitcoin?. (*In* International Conference on Social Informatics. Springer, Cham. p. 347-367).

Baur, A.W., Bühler, J., Bick, M. & Bonorden, C.S. 2015. Cryptocurrencies as a disruption? empirical findings on user adoption and future potential of bitcoin and co. (*In* Conference on e-Business, e-Services and e-Society. Springer, Cham. p. 63-80).

Baur, D.G., Hong, K. & Lee, A.D. 2018. Bitcoin: medium of exchange or speculative assets?. *Journal of international financial markets, institutions and money*, 54:177-189.

Bayes, T. 1763. An Essay towards solving a Problem in the Doctrine of Chance. *Philosophical Transactions*, 53: 370–418.

Benedetti, H. & Kostovetsky, L. 2018. Digital tulips? Returns to investors in initial coin offerings.

Bengio, Y. 2009. Learning deep architectures for AI. *Foundations and trends in machine learning,* 2(1):1-127.

Bengio, Y. & De Mori, R. 1988. Use of neural networks for the recognition of place of articulation. (*In* International Conference on Acoustics, Speech, and Signal Processing*. p.103-106). IEEE.*

Bengio, Y., Lamblin, P., Popovici, D. & Larochelle, H. 2007. Greedy layer-wise training of deep networks. (*In* Advances in neural information processing systems, p. 153-160).

Bishop Jr, G.W. 1961. Evolution of the Dow Theory. *Financial Analysts Journal*, 17(5):23-26.

Boedecker, J., Obst, O., Lizier, J.T., Mayer, N.M. & Asada, M. 2012. Information processing in echo state networks at the edge of chaos. *Theory in Biosciences*, 131(3):205-213.

Bohr, J. & Bashir, M. 2014. Who uses bitcoin? An exploration of the bitcoin community. (*In* 2014 Twelfth Annual Conference on Privacy, Security and Trust (PST), p. 94-101). IEEE.

Boland, L.A. 1968. The identification problem and the validity of economic models. *South African Journal of Economics*, 36(3):236-240.

Bouoiyour, J. & Selmi, R. 2017. The Bitcoin price formation: Beyond the fundamental sources. *arXiv preprint* arXiv:1707.01284.

Bouoiyour, J., Selmi, R., Tiwari, A.K. & Olayeni, O.R. 2016. What drives Bitcoin price. *Economics bulletin*, 36(2):843-850.

Bouri, E., Gupta, R., Tiwari, A.K. & Roubaud, D. 2017. Does Bitcoin hedge global uncertainty? Evidence from wavelet-based quantile-in-quantile regressions. *Finance research letters,* 23:87-95.

Bouri, E., Molnár, P., Azzi, G., Roubaud, D. & Hagfors, L.I. 2017. On the hedge and safe haven properties of Bitcoin: Is it really more than a diversifier? *Finance research letters*, 20:192-198.

Bourlard, H. & Kamp, Y. 1988. Auto-association by multilayer perceptrons and singular value decomposition. *Biological cybernetics,* 59(88): 291-294.

Brandvold, M., Molnár, P., Vagstad, K. & Valstad, O.C.A. 2015. Price discovery on Bitcoin exchanges. *Journal of international financial markets, institutions and money*, 36:18-35.

Briere, M., Oosterlinck, K. & Szafarz, A. 2015. Virtual currency, tangible return: portfolio diversification with bitcoin. *Journal of asset management*, 16(6):365-373.

Brito, J., Castillo, A. 2013. Bitcoin: a primer for policymakers. Mercatus Center. George Mason University. Retrieved 22 October 2013.

Brockwell, P.J., Davis, R.A. & Calder, M.V. 2002. Introduction to time series and forecasting (Vol. 2), NY: springer.

Broomhead, D.S. & Lowe, D. 1988. Radial basis functions, multi-variable functional interpolation and adaptive networks. *Royal Signals and Radar Establishment Malvern (United Kingdom).*

Brownlee, J. 2013. Machine learning. https://machinelearningmastery.com/machine-learning-is-popular Date of access: 3/28/2018.

Bruton, G., Khavul, S., Siegel, D. & Wright, M. 2015. New financial alternatives in seeding entrepreneurship: Microfinance, crowdfunding, and peer-to-peer innovations. *Entrepreneurship theory and practice,* 39(1):9-26.

Buckley, R.P. 2016. Reconceptualizing Global Financial Regulation. *Oxford journal of legal studies 2016.*

Buffett, W. 1988. Chairman's letter to the shareholders of Berkshire Hathaway. Inc. 2005 Annual Report.

Bukovina, J. & Marticek, M. 2016. Sentiment and bitcoin volatility (No. 2016-58). Mendel University in Brno, Faculty of Business and Economics.

Butler, T. & O'Brien, L. 2019. Understanding RegTech for Digital Regulatory Compliance. (*In* Disrupting Finance, Palgrave Pivot: Cham. p. 85-102).

Calderón, A., Roa, S. & Victorino, J. 2003. Handwritten digit recognition using convolutional neural networks and gabor filters. (*In* Proceedings of the international conference on computational intelligence. Canberra, ACT: Unalmed.

Canova, F. 1999. Vector autoregressive models: specification, estimation, inference, and forecasting. Handbook of Applied Econometrics Volume 1: Macroeconomics. Hoboken, NJ: Wiley-Blackwell.

Carse, D. 1999. Keynote: Regulatory Framework of e-banking. http://www.bis.org/review/r991012c.pdf Date of access: 26 March 2019.

Cerny, P.G. 1994. The dynamics of financial globalization: Technology, market structure, and policy response. *Policy sciences*, 27(4):319-342.

Chancellor, E. 2018. The Bitcoin Premonition. *Financial history*, (125):13-16.

Chatfield, C. 2016. The analysis of time series: an introduction. London, LDN: Chapman and Hall.

Chatfield, C. & Yar, M. 1988. Holt-winters forecasting: some practical issues. *The Statistician,* :129–140.

Chaum, D. 1983. Blind signatures for untraceable payments. *Advances in cryptology: proceedings of crypto*, 82(3):199–203.

Cheah, E.T. & Fry, J. 2015. Speculative bubbles in Bitcoin markets? An empirical investigation into the fundamental value of Bitcoin. *Economics letters*, 130:32-36.

Chen, C.Y.H., Härdle, W.K., Hou, A.J. & Wang, W. 2018. Pricing Cryptocurrency options: the case of CRIX and Bitcoin. *SSRN* 3159130.

Chen, J.F., Do, Q. & Hsieh, H.N. 2015. Training artificial neural networks by a hybrid PSO-CS algorithm. *Algorithms*, 8(2):292-308.

Chen, L. 2016. From fintech to finlife: the case of fintech development in China. *China economic journal,* 9(3):225-239.

Chen, S., Wang, H., Xu, F. & Jin, Y.Q. 2016. Target classification using the deep convolutional networks for SAR images. *IEEE transactions on geoscience and remote sensing,* 54(8):4806-4817.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H. & Bengio, Y. 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint* arXiv:1406.1078.

Cohan, U.W. 2017. A History of Bitcoin. SSRN 3047875.

Choron, H. & Choron, S. 2011. Money: Everything You Never Knew About Your Favorite Thing to Find, Save, Spend & Covet. *Chronicle books*, page 22.

Chung, J., Gulcehre, C., Cho, K. & Bengio, Y. 2014. Empirical evaluation of gated recurrent neural networks on sequence modelling. *arXiv preprint* arXiv:1412.3555.

Chung, J., Gulcehre, C., Cho, K. & Bengio, Y. 2015. Gated feedback recurrent neural networks. (*In* International Conference on Machine Learning, p. 2067-2075).

Ciaian, P., Rajcaniova, M. & Kancs, D.A. 2016. The economics of BitCoin price formation. *Applied economics*, 48(19):1799-1815.

Clarkson, P. & Moreno, P.J. 1999. On the use of support vector machines for phonetic classification. (*In* 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing, Proceedings. p. 585-588). IEEE.

Coccia, M. 2018. A theory of the general causes of long waves: War, general purpose technologies, and economic change. *Technological forecasting and social change*, 128:287-295.

Coinmarketcap. 2019. https://coinmarketcap.com. Date accessed 6 May 2019.

Collier, P. 1999. On the economic consequences of civil war. *Oxford economic papers*, 51(1):168-183.

Collobert, R., Bengio, S. and Mariéthoz, J. 2002. Torch: a modular machine learning software library. Idiap.

Cortes, C., Vapnik, V. 1995. Support-vector networks. *Machine learning*, 20(3): 273–297.

Corzo, T., Prat, M. and Vaquero, E. 2014. Behavioral Finance in Joseph de la Vega's Confusion de Confusiones. *Journal of behavioral finance*, 15(4)341-350.

Dai, W. 1998. B-money. *Consulted*, 1:2012.

Damodaran, A. 2012. Investment valuation: tools and techniques for determining the value of any asset. Vol. 666. New York, NY: John Wiley & Sons.

Daniel, E. 1999. Provision of electronic banking in the UK and the Republic of Ireland. *International journal of bank marketing,* 17(2):72-83.

Davies, G. 2010. History of money. Cardiff, CWL: University of Wales Press.

Demertzis, M., Merler, S. & Wolff, G.B. 2018. Capital Markets Union and the fintech opportunity. *Journal of financial regulation*, 4(1):157-165.

Drucker, H., Wu, D. & Vapnik, V.N. 1999. Support vector machines for spam categorization. *IEEE transactions on neural networks*, 10(5):1048-1054.

Dyhrberg, A.H. 2016. Bitcoin, gold and the dollar–A GARCH volatility analysis. *Finance research letters*, 16:85-92.

Dyhrberg, A.H. 2016. Hedging capabilities of bitcoin. Is it the virtual gold? *Finance Research Letters*, 16:139-144.

Edwards, R.D., Magee, J. & Bassetti, W.H.C. 2018. Technical analysis of stock trends. Boca Raton, FL: CRC press.

Elman, J. 1990. Finding structure in time. *Cognitive science* 14(2):179-211.

Engelen, E., Erturk, I., Froud, J., Leaver, A. & Williams, K. 2010. Reconceptualizing financial innovation: frame, conjuncture and bricolage. *Economy and society*, 39(1):33-63.

EY. 2018. Fintech Ecosystem Playbook. https://www.ey.com/Publication/vwLUAssets/EY-fintech-ecosystem-playbook/$FILE/EY-fintech-ecosystem-playbook.pdf. Date of access: 28 Feb. 2019.

Fan, P.S. 2018. Singapore approach to develop and regulate FinTech. (*In* Deng, R.H., *ed.* Handbook of Blockchain, Digital Finance, and Inclusion, Volume 1. Cambridge: Academic Press. p. 347-357)

FARRAR, E. 1962. The investment decision under uncertainty. Englewood Cliffs, NJ: Prentice-Hall.

Feng, L., Ong, Y.S. & Lim, M.H. 2013. ELM-guided memetic computation for vehicle routing. *IEEE intelligent systems*, 28(6):38-41.

Fernández, S., Graves, A. & Schmidhuber, J. 2007. An application of recurrent neural networks to discriminative keyword spotting. (*In* International Conference on Artificial Neural Networks, Heidelberg:Berlin. p. 220-229.)

Figlewski, S. 1997. Forecasting volatility. *Financial markets, institutions & instruments*, 6(1):1-88.

Fisher, R.A. 1936. The use of multiple measurements in taxonomic problems. *Annals of eugenics,* 7(2):179-188.

Fragkiskos, A. 2014. What is portfolio diversification? *SSRN* 2331475

Freund, Y. & Schapire, R. E. 1999. Large margin classification using the perceptron algorithm. *Machine learning*, 37(3):277–296.

Fukushima, K. 1980. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, 36(4):193-202.

Gai, P., Haldane, A. & Kapadia, S. 2011. Complexity, concentration and contagion. *Journal of monetary economics,* 58(5):453-470.

Garcia, D. & Schweitzer, F. 2015. Social signals and algorithmic trading of Bitcoin. *Royal Society open science*, 2(9):150288.

Garcia, D., Tessone, C.J., Mavrodiev, P. & Perony, N. 2014. The digital traces of bubbles: feedback cycles between socio-economic signals in the Bitcoin economy. *Journal of the Royal Society interface,* 11(99):20140623.

Ghosh, S., Vinyals, O., Strope, B., Roy, S., Dean, T. & Heck, L. 2016. Contextual lstm (clstm) models for large scale nlp tasks. *arXiv preprint* arXiv:1602.06291.

Gilbert, A.M., Hunt, D. & Winch, K.C. 1997. Creating an integrated payment system: The evolution of Fedwire. Federal Reserve Bank of New York. *Economic policy review*, 3(2):1-7.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. & Bengio, Y. 2014. Generative adversarial nets. *Advances in neural information processing systems,* :2672-2680.

Graves, A. & Schmidhuber, J. 2005. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural networks,* 18(5-6):602-610.

Graves, A., Mohamed, A.R. & Hinton, G. 2013. Speech recognition with deep recurrent neural networks. (*In* 2013 IEEE international conference on acoustics, speech and signal processing, p.6645-6649).

Graves, A., Greg, W., & Ivo Danihelka. 2014. Neural turing machines. *arXiv preprint* arXiv:1410.5401.

Greenberg, A. 2013. Fbi says it’s seized $28.5 million in bitcoins from ross ulbricht, alleged owner of silk road. https://www.forbes.com/sites/andygreenberg/2013/10/25/fbi-says-its-seized-20-million-in-bitcoins-from-ross-ulbricht-alleged-owner-of-silk-road/#69082cf72765 Date of access: 27 March 2019.

Greff, K., Srivastava, R.K., Koutník, J., Steunebrink, B.R. & Schmidhuber, J. 2017. LSTM: A search space odyssey*. IEEE transactions on neural networks and learning systems*, 28(10):2222-2232.

Gruber, S. 2013. Trust, identity and disclosure: are Bitcoin exchanges the next virtual havens for money laundering and tax evasion*. Quinnipiac*, 32(1):135.

Grzyb, B.J., Chinellato, E., Wojcik, G.M. & Kaminski, W.A. 2009. Facial expression recognition based on liquid state machines built of alternative neuron models. (*In* 2009 International Joint Conference on Neural Networks, p. 1011-1017).

Güler, N.F., Übeyli, E.D. & Güler, I. 2005. Recurrent neural networks employing Lyapunov exponents for EEG signals classification. *Expert systems with applications*, 29(3):506-514.

Hamilton, J.L. 1978. Marketplace organization and marketability: NASDAQ, the stock exchange, and the national market system. *The journal of finance*, 33(2):487-503.

Hayes, B. 2013. First links in the Markov chain*. American scientist,* 101(2):252.

He, K., Zhang, X., Ren, S. & Sun, J. 2016. Deep residual learning for image recognition. (*In* Proceedings of the IEEE conference on computer vision and pattern recognition, p. 770-778).

Heaton, J.B., Polson, N.G. & Witte, J.H. 2017. Deep learning for finance: deep portfolios. *Applied stochastic models in business and industry*, 33(1):3-12.

Hencic, A. & Gouriéroux, C. 2015. Noncausal autoregressive model in application to bitcoin/USD exchange rates. (*In* Econometrics of risk, Cham: Springer. p. 17-40).

Hengeveld, B. & Rooijakkers, J. 2019. Adding Friction to Frictionless Payments: A Haptic Interface for Digital Transactions. (*In* Proceedings of the Thirteenth International Conference on Tangible, Embedded, and Embodied Interaction, Tempe, Arizona. New York, NY: ACM. p. 243-250).

Herrera, N.S. 2017. What's All This Talk About Fintech? An Analysis of Media Conversations Around Fintech in 2016. Southern California: USC. (Thesis – PhD)

Hinton, G. & Sejnowski, T. 1986. Learning and releaming in Boltzmann machines. *Parallel distributed processing: explorations in the microstructure of cognition,* 1: 282-317.

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

Hooker, R., King, L.W. 1996. Mesopotamia: the code of Hammurabi. *Washington State University.*

Hopfield, J. 1982. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, 79:2554–2558.

Hornik, K., Stinchcombe, M. and White, H. 1989. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359-366.

Hsiao, C. 1979. Causality tests in econometrics. *Journal of economic dynamics and control*, 1(4):321-346.

Huang, D.S. & Ma, S.D. 1999. Linear and nonlinear feedforward neural network classifiers: a comprehensive understanding. *Journal of intelligent systems*, 9(1):1-38.

Huguet, S., Godin, N., Gaertner, R., Salmon, L. & Villard, Y.D. 2002. Use of acoustic emission to identify damage modes in glass fibre reinforced polyester. *Composites science and technology*, 62(10-11):1433-1444.

Hurley, R., Gong, X. & Waqar, A. 2014. Understanding the loss of trust in large banks. *International journal of bank marketing*, 32(5):348-366.

Huwart, J.Y. & Verdier, L. 2013. The 2008 financial crisis–a crisis of globalization. *Economic globalisation: origins and consequences*, 1:126-143.

Jaeger, H. & Harald, H. 2004. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science* 304(5667):78-80.

Jakub, B. 2015. Does Bitcoin follow the hypothesis of efficient market?. *International journal of economic sciences,* 4(2):10-23.

Jarrell, G.A. 1984. Change at the exchange: The causes and effects of deregulation. *The journal of law and economics*, 27(2):273-312.

Jordan, M.I. 1986. An introduction to linear algebra in parallel distributed processing. *Parallel distributed processing*, 1:365-422.

Kaastra, I & Boyd, M. 1996. Designing a neural network for forecasting financial and economic time series. *Neurocomputing*, vol. 10(3): 215–236.

Kamo, T. & Dagli, C. 2009. Hybrid approach to the Japanese candlestick method for financial forecasting. *Expert systems with applications*, 36(3):5023-5030.

Kaza, G. 2000. In defense of fundamental analysis: securities arbitrage and the efficient market theory*. The quarterly journal of Austrian economics*, 3(1):59-61.

Kingma, D.P. & Welling, M. 2013. Auto-encoding variational bayes. *arXiv preprint* arXiv:1312.6114 (2013).

Kirkpatrick II, C.D. & Dahlquist, J.A. 2010. Technical analysis: the complete resource for financial market technicians. Upper Saddle River, NJ: FT press.

Kosse, A. & Hielkema, P. 2019. Innovative yet robust supervision in a rapidly changing payments landscape. *Journal of digital banking*, 3(3):232-241.

Kotsiantis, S.B., Zaharakis, I. & Pintelas, P. 2007. Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering,* 160:3-24.

Krizhevsky, A., Sutskever, I. & G. Hinton. 2012. Imagenet classification with deep convolutional neural networks.(*In* Neural Information Processing Systems, Lake Tahoe, Nevada. La Jolla, LJ: NIPS Foundation p. 113-311).

Kshetri, N. 2015. Success of crowd-based online technology in fundraising: An institutional perspective. *Journal of international management*, 21(2):100-116.

Kulkarni, T.D., Whitney, W.F., Kohli, P. & Tenenbaum, J. 2015. Deep convolutional inverse graphics network. *Advances in neural information processing systems,* pp.2539-2547.

Larsen, K. & Gilani, S. 2017. RegTech is the New Black-The Growth of RegTech Demand and Investment. *Journal of financial transformation*, 45:22-29.

Laurent, P., Chollet, T. & Herzberg, E. 2015. Intelligent automation entering the business world. https://www2.deloitte.com/content/dam/Deloitte/lu/Documents/operations/lu-intelligent-automationbusiness-world.pdf. Date of access: 16 March 2019.

LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278-2324.

Lekamalage, L., Kasun, C., Zhou, H., Huang, G.B. & Cambria, E. 2013. Representational learning with ELMs for big data. *IEEE intelligent* systems, 11:31-34.

Lev, B. & Thiagarajan, S.R. 1993. Fundamental information analysis. *Journal of Accounting research*, p.190-215.

Lo, A.W., Mamaysky, H. & Wang, J. 2000. Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. *The journal of finance*, 55(4):1705-1765.

LTP Team. 2015. Survey shows Americans trust technology firms more than banks and retailers. Let's talk payment. http://letstalkpayments.com/survey-shows-americans-trust-technology-firms-more-than-banks-and-retailers. Date of access: 24 March 2019.

Lütkepohl, H. 2014. Fundamental problems with nonfundamental shocks. *Essays in nonlinear time series econometrics*, p. 198-214.

Makarov, I. & Schoar, A. 2018. Trading and arbitrage in cryptocurrency markets. Journal of financial economics (2019).

Malkiel, B.G. & Fama, E.F. 1970. Efficient capital markets: a review of theory and empirical work. *The journal of Finance*, 25(2):383-417.

Margrave, F.W., Rigas, K., Bradley, D.A. & Barrowcliffe, P. 1999. The use of neural networks in ultrasonic flaw detection. *Measurement,* 25(2):143-154.

Marino, J. 2012. Goldman Sachs is a Tech Company. http://www.businessinsider.com/goldman-sachs-has-more-engineers-than-facebook-2015-4. Date of access: 23 March 2019.

Markoff, J. 1990. BUSINESS TECHNOLOGY; What's the Best Answer? It's Survival of the Fittest. http://www. nytimes. com/1990/08/29/business/business-technology-what-sthe-best-answer-it-s-survival-of-the-fittest. html. Date of access: 21 March 2019.

Markov, A.A. 1913. An example of statistical study on the text of 'Eugene Onegin' illustrating the linking of events to a chain. *Izvestiya Akademii Nauk*, 6:153-162.

McCarthy, J., Feigenbaum, E. 1958. Arthur Samuel: Pioneer in Machine Learning. *Association for the Advancement of Artificial Intelligence*, (3):10.

McNeil, I. ed. 2002. (*In* Encyclopedia of the history of technology).

Melo, F.S. 2001. Convergence of Q-learning: A simple proof. Institute Of Systems and Robotics*. Tech. Rep*, :1-4.

Millon-Cornett, M.H. & Tehranian, H. 1989. Stock market reactions to the depository institutions deregulation and monetary control act of 1980. *Journal of banking & finance*, 13(1):81-100.

Mohamed, A.R., Dahl, G.E. & Hinton, G. 2012. Acoustic modeling using deep belief networks. *IEEE Transactions on audio, speech, and language processing*, 20(1):14-22.

Molina, D., Liang, J., Harley, R. & Venayagamoorthy, G.K. 2011. Comparison of TDNN and RNN performances for neuro-identification on small to medium-sized power systems. *Computational intelligence applications in smart grid,* p. 1-8.

Mondres, T. 2019. Five Global Fintech Trends in 2018. *ABA banking journal*, 111(1):52.

Montana, D.J. & Davis, L. 1989. Training Feedforward Neural Networks Using Genetic Algorithms. *International journal of computational artifical intelligence,* 89:762-767.

Montgomery, B.L. & Kumar, B.V. 1986. Evaluation of the use of the Hopfield neural network model as a nearest-neighbor algorithm. *Applied optics*, 25(20):3759-3766.

Moody's. 2015. Global Credit Research - Moody's: Threat of Cyber Attack on US Utilities cushioned by likelihood of government support. https://www.moodys.com/research/Moodys-Threat-of-cyber-attack-on-US-utilities-cushioned-by--PR_336640 Date of access: 21 March 2019.

Moore, G. 1965. Moore's law. *Electronics magazine,* 38(8):114.

Moore, T., Christin, N. & Szurdi, J. 2018. Revisiting the risks of bitcoin currency exchange closure. *ACM transactions on internet technology*, 18(4):50.

Mourlon-Druol, E. 2015. Trust is good, control is better': The 1974 Herstatt Bank Crisis and its Implications for International Regulatory Reform. *Business history*, 57(2)3:11-334.

Mpofu, R.T., De Beer, J.S. & Marx, J. 2016. Investment Management. 4th ed. Pretoria, GT: Van Schaik.

Myers, S.C. 1984. Finance theory and financial strategy. *Interfaces*, 14(1):126-137.

Nakamoto, S., 2008. Bitcoin: A peer-to-peer electronic cash system.

Nebauer, C. 1998. Evaluation of convolutional neural networks for visual recognition. *IEEE transactions on neural networks*, 9(4):685-696.

Nelson, D.M., Pereira, A.C. & de Oliveira, R.A. 2017. Stock market's price movement prediction with LSTM neural networks. (*In* 2017 International Joint Conference on Neural Networks, Anchorage, AL: IEEE. p.1419-1426).

Nguyen, T.H., Shirai, K. & Velcin, J. 2015. Sentiment analysis on social media for stock movement prediction. *Expert systems with applications*, 42(24):9603-9611.

O'Hagan, A. 2016. The Satoshi affair. *London review of books*, 38(13):7-28.

Online Etymology Dictionary. 2019. Finance. https://www.etymonline.com/word/finance Date of Access: 11 March 2019.

Online Etymology Dictionary. 2019. Technology. https://www.etymonline.com/word/technology Date of Access: 11 March 2019.

O'sullivan, A. & Sheffrin, S.M. 2003. Economics: principles in action. Upper Saddle River. NJ: Pearson Prentice Hall.

Patrick, L. and Julia, S.O. eds., 2017. OECD Insights Debate the Issues: Complexity and Policy making. OECD Publishing.

Paik, J.K.& Katsaggelos, A.K. 1992. Image restoration using a modified Hopfield network. *IEEE transactions on image processing*, 1(1):49-63.

Palle, S. 2019. Artificial Intelligence Using DBS-QOS In Banking Organizations. *Journal of scientific research & engineering trends,* 5(1):2395-566X.

Palmer, M., Boyd-Barrett, O. & Rantanen, T. 1998. Global financial news. *The globalization of news*, p.61-78.

Penman, S.H. 2010. Financial forecasting, risk and valuation: Accounting for the future. *Abacus*, 46(2):211-228.

Platanakis, E., Sutcliffe, C. & Urquhart, A. 2018. Optimal vs naïve diversification in cryptocurrencies. *Economics letters*, 171:93-96.

Polasik, M., Piotrowska, A.I., Wisniewski, T.P., Kotkowski, R. & Lightfoot, G. 2015. Price fluctuations and the use of Bitcoin: An empirical inquiry. *International journal of electronic commerce,* 20(1):9-49.

Proença, J. 2019. Challenges and opportunities of mobile payments innovations. Porto: UCP. (Thesis - PhD).

Prokhorov, D., G. Puskorius, & L. Feldkamp. 2001. Dynamical neural networks for control. (*In* J. Kolen and S. Kremer., *eds*. A Field Guide to Dynamical Recurrent Networks. Hoboken, NJ: IEEE Press).

Pu, Y., Gan, Z., Henao, R., Yuan, X., Li, C., Stevens, A. & Carin, L. 2016. Variational autoencoder for deep learning of images, labels and captions. *Advances in neural information processing systems* p.2352-2360.

Qian, Y. 2019. Central Bank Digital Currency: optimization of the currency system and its issuance design. *China economic journal*, :1-15.

Quiza, R. & Davim, J.P. 2011. Computational methods and optimization. *Machining of hard materials*. London, LDN: Springer.

Reed, S.E., Zhang, Y., Zhang, Y. & Lee, H. 2015. Deep visual analogy-making. *Advances in neural information processing systems,* :1252-1260.

Reuter, P. 1985. The organization of illegal markets: An economic analysis. Vol. 84, No. 9. Washington, DC: US Department of Justice, National Institute of Justice.

Rodriguez, J. 2018. The Science Behind OpenAI Five that just Produced One of the Greatest Breakthrough in the History of AI. https://towardsdatascience.com/the-science-behind-openai-five-that-just-produced-one-of-the-greatest-breakthrough-in-the-history-b045bcdc2b69Retrieved 2019-01-15 Date of access: 4 April 2019.

Roland, A. 1995. Science, technology, and war. *Technology and C\culture,* 36(2):83-100.

Romānova, I. & Kudinska, M. 2016. Banking and Fintech: a challenge or opportunity?. *Contemporary issues in finance: current challenges from across Europe* :21-35.

Rosenblatt, F. 1958. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386–408.

Rucker, B. & Leach, M.W., Ntt Data Inc. 2019. Providing quantitative evaluations of friction within a customer experience to reduce abandonment and improve conversion of transactions. (Patent Application: U.S. 15/663,016).

Rumelhart, D., Hinton, G. & Williams, R. 1986. Learning representations by back-propagating errors. *Nature,* 323:533–536.

Sabine, R. 1869. The History and Progress of the Electric Telegraph. *Virtue & co,* p.206.

Sarikaya, R., Hinton, G.E. & Deoras, A. 2014. Application of deep belief networks for natural language understanding. *IEEE/ACM transactions on audio, speech and language processing*, 22(4):778-784.

Sassen, S. 2005. The embeddedness of electronic markets: The case of global capital markets. *The sociology of financial markets*, :17-37.

Sathye, M. 2005. The impact of internet banking on performance and risk profile: Evidence from Australian credit unions. *Journal of banking regulation*, 6(2):163-174.

Saunders, A., Cornett, M.M. & McGraw, P.A. 2006. Financial institutions management: A risk management approach. New York, NY, USA: McGraw-Hill.

Schenk, C.R. 2014. Summer in the city: banking failures of 1974 and the development of international banking supervision. *The English historical review*, 129(540):1129-1156.

Schmidhuber, J. 2015. Deep learning in neural networks: An overview. *Neural networks*, 61:85-117.

Schooner, H.M. & Taylor, M.W. 2009. Global bank regulation: principles and policies. Cambridge, MA: Academic Press.

Schoonover, T.D. 2013. Uncle Sam's War of 1898 and the Origins of Globalization. Lexington, KY: University Press of Kentucky.

Scott, G., Carr, M. & Cremonie, M. 2016. Technical Analysis: Modern Perspectives. *Research foundation reviews*, 11(1):1-36.

Selmi, R., Tiwari, A.K. and Hammoudeh, S., 2018. Efficiency or speculation? A dynamic analysis of the Bitcoin market. *Economics bulletin*, 38(4):2037-2046.

Sethi, A. 2019. Digital China. *Chinese consumers,* :33-54.

Shareef, M.A., Baabdullah, A., Dutta, S., Kumar, V. & Dwivedi, Y.K. 2018. Consumer adoption of mobile banking services: An empirical examination of factors according to adoption stages. *Journal of retailing and consumer services*, 43:54-67.

Siegelmann, T & Sontag, E. 1992. On the Computational Power of Neural Nets. (*In* Fifth annual workshop on computational learning theory, New York, NY: ACM publications. p. 440–449)

Sims, C.A. 1972. Money, income, and causality. *The American economic review*, :540-552.

Smith, S. 2018. Contactless payments to represent 1 in 3 in-store transactions globally by 2020. https://www.juniperresearch.com/press/press-releases/contactless-payments-to-represent-1-in-3-in-store Date of access: 25 February 2019.

Smith, T. 2019. pmdarima. https://www.alkaline-ml.com/pmdarima. Date of access: 15 Feb. 2018.

Smolensky, P. 1986. Information processing in dynamical systems: foundations of harmony theory. *Parallel distributed processing: explorations in the microstructure of cognition*, 1194-281.

Spooner, M.C. 1984. Origin of fundamental analysis. *Financial analysts journal*, 40(4):79-80.

Steck, J.E. & Balakrishnan, S.N. 1994. Use of Hopfield neural networks in optimal guidance. *IEEE transactions on aerospace and electronic systems*, 30(1):287-293.

Stemler, A.R. 2013. The JOBS Act and crowdfunding: Harnessing the power—and money—of the masses. *Business horizons*, 56(3):271-275.

Szabo, N. 2008. Bit Gold. http://unenumerated.blogspot.com/2005/12/bit-gold.html Date of access: 21 December 2019.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A. 2015. Going deeper with convolutions. (*In* Proceedings of the IEEE conference on computer vision and pattern recognition, Boston, MA: IEEE Press, p. 1-9).

Tang, D., Qin, B. & Liu, T. 2015. Document modeling with gated recurrent neural network for sentiment classification. In *Proceedings of the 2015 conference on empirical methods in natural language processing* (pp. 1422-1432).

Tang, R. & Lin, J. 2018. Deep residual learning for small-footprint keyword spotting. (*In* 2018 IEEE International Conference on Acoustics, Speech and Signal Processing, Calgary, AB: IEEE Pess. p. 5484-5488).

Tesauro, G. 1995. Temporal Difference Learning and TD-Gammon. *Communications of the ACM*. 38(3).

Trautman, L.J. 2014. Virtual currencies; Bitcoin & what now after Liberty Reserve, Silk Road, and Mt. Gox?. *Richmond journal of law and technology*, 20(4).

Theis, L., Shi, W., Cunningham, A. & Huszár, F. 2017. Lossy image compression with compressive autoencoders. *arXiv preprint.* arXiv:1703.00395.

Thomas, M. & Morse, J. 2017. FinTech—Origins and Prognosis*. Nacva quickreads.* https://quickreadbuzz.com/2017/07/26/fintech-origins-prognosis/ Date of access: 15 January 2019.

Ticknor, A.J. & Barrett, H.H. 1987. Optical implementations in Boltzmann machines. *Optical engineering,* 26(1):260116.

Tirole, J. 2017. There are many reasons to be cautious about bitcoin*. Financial Times*, 30:2017.

Töpfer, L.M. 2018. China’s integration into the global financial system: toward a state-led conception of global financial networks. *Dialogues in human geography*, 8(3):251-271.

Trębacz, T. 2019. FinTech as an innovative banking sector. *World scientific news*, 122:83-95.

Turing, A. 1950. Computing Machinery and Intelligence. *Mind*. 59(236):433–460.

Ullah, M. 2013. Time Series Analysis and Forecasting. http://itfeature.com/time-series-analysis-and-forecasting/time-series-analysis-forecasting. Date of Access: 8 February 2019.

Van Hout, M.C. & Bingham, T. 2014. Responsible vendors, intelligent consumers: Silk Road, the online revolution in drug trading. *International journal of drug policy*, 25(2):183-189.

Van Loo, R. 2018. Making Innovation More Competitive: The Case of Fintech. *UCLA law review*, 65(1).

Vance, A. 2018. This Man Is the Godfather the AI Community Wants to Forget
Jürgen Schmidhuber says he'll make machines smarter than us. His peers wish he'd just shut up . https://www.bloomberg.com/news/features/2018-05-15/google-amazon-and-facebook-owe-j-rgen-schmidhuber-a-fortune Date of access: 16 Jan 2019.

Vandewalle, N., Ausloos, M. & Boveroux, P. 1999. The moving averages demystified. *Physica A: statistical mechanics and its applications*, 269(1):170-176.

Vincent, P., Larochelle, H., Bengio, Y. & Manzagol, P.A. 2008. Extracting and composing robust features with denoising autoencoders. (*In* Proceedings of the 25th international conference on Machine learning, New York, NY: ACM. p. 1096-1103).

Von Neumann, J. 1945. First draft of a report on the edvac, contract no. w-670-ord-402 moore school of electrical engineering, univ. of penn., philadelphia.. *Origins of Digital Computers: Selected Papers*. Reprinted (in part) in Randell, Brian. 1982.

Vu, T.D., Yang, H.J., Nguyen, V.Q., Oh, A.R. & Kim, M.S. 2017. Multimodal learning using convolution neural network and Sparse Autoencoder. (*In* 2017 IEEE International Conference on Big Data and Smart Computing, Jeju, KO: IEEE p. 309-312).

Waldrop, M.M. 1987. Computers Amplify Black Monday: The sudden stock market decline raised questions about the role of computers; they may not have actually caused the crash, but may well have amplified it. *Scienc.*, 238(4827):602-604.

Wallace, B. 2011. The rise and fall of Bitcoin. https://www.wired.com/2011/11/mf-bitcoin/ Date of access: 16 March 2018.

Wallis, K.F. 1977. Multiple time series analysis and the final form of econometric models. *Econometrica: journal of the Econometric Society*, :1481-1497.

Welch, B. 1999. Electronic Banking and Treasury Security*. Elsevier,* :48.

Werbos, P. 1974. Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences. Cambridge: Harvard University. (Thesis – PHD)

Wilmoth, J. 2018. Happy Bitcoin Pizza Day!. https://www.ccn.com/happy-bitcoin-pizza-day. Date of access: 10 Jan 2019.

Wise, J. 1956. Stationarity conditions for stochastic processes of the autoregressive and moving-average type. *Biometrika*, 43(1/2):215-219.

Wong, C.S. & Li, W.K. 2000. On a mixture autoregressive model. *Journal of the Royal Statistical Society: series b* (statistical methodology), 62(1):95-115.

Wright, M. 2018. Banking on AI:How artificial intelligence will impact the financial world. https://www.accenture.com/t00010101t000000z__w__/gb-en/_acnmedia/pdf-68/accenture-redefine-banking.pdf Date of access: 17 March 2019

Wu, J., Lu, E., Kohli, P., Freeman, B. & Tenenbaum, J. 2017. Learning to see physics via visual de-animation. *Advances in neural information processing systems,* :153-164.

Wulan, V.R. 2017. Financial Technology (Fintech) A New Transaction In Future. Journal of electrical engineering and computer science, 2(1).

Xu, J., Xiang, L., Liu, Q., Gilmore, H., Wu, J., Tang, J. & Madabhushi, A. 2016. Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images. *IEEE transactions on medical imaging,* 35(1):119-130.

Yang, Z., Hu, Z., Salakhutdinov, R. & Berg-Kirkpatrick, T. 2017. Improved variational autoencoders for text modeling using dilated convolutions. (*In* Proceedings of the 34th International Conference on Machine Learning-Volume 70, Sydney, AU: JMLR. p. 3881-3890).

Yao, Q. 2018. A systematic framework to understand central bank digital currency. *Science China information sciences*, 61(3):033101.

Yelowitz, A. & Wilson, M. 2015. Characteristics of Bitcoin users: an analysis of Google search data. *Applied economics letters*, 22(13):1030-1036.

Yu, L., Chen, H., Dou, Q., Qin, J. & Heng, P.A. 2017. Automated melanoma recognition in dermoscopy images via very deep residual networks. *IEEE transactions on medical imaging,* 36(4):994-1004.

Zeiler, M.D., Krishnan, D., Taylor, G.W. and Fergus, R. 2010. Deconvolutional networks. (*In* Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA: IEEE. p. 7).

Zeiler, M.D., Taylor, G.W. & Fergus, R. 2011. Adaptive deconvolutional networks for mid and high level feature learning. *ICCV* , 1(2):6.

Zetzsche, D.A., Arner, D.W., Buckley, R.P. & Weber, R.H. 2019. The Future of Data-Driven Finance and RegTech: Lessons from EU Big Bang II. *Available at SSRN 3359399*.

Zhang, H., Sindagi, V. & Patel, V.M. 2017. Image de-raining using a conditional generative adversarial network. *arXiv preprint* arXiv:1701.05957.

Zhang, J. & Yang, Y. 2001. Global stability analysis of bidirectional associative memory neural networks with time delay. *International Journal of Circuit Theory and Applications*, 29(2):185-196.

Zhang, Y., Li, P., Jin, Y. & Choe, Y. 2015. A digital liquid state machine with biologically inspired learning and its application to speech recognition. *IEEE transactions on neural networks and learning systems,* 26(11):2635-2649.

ANNEXURES

Exhibit A

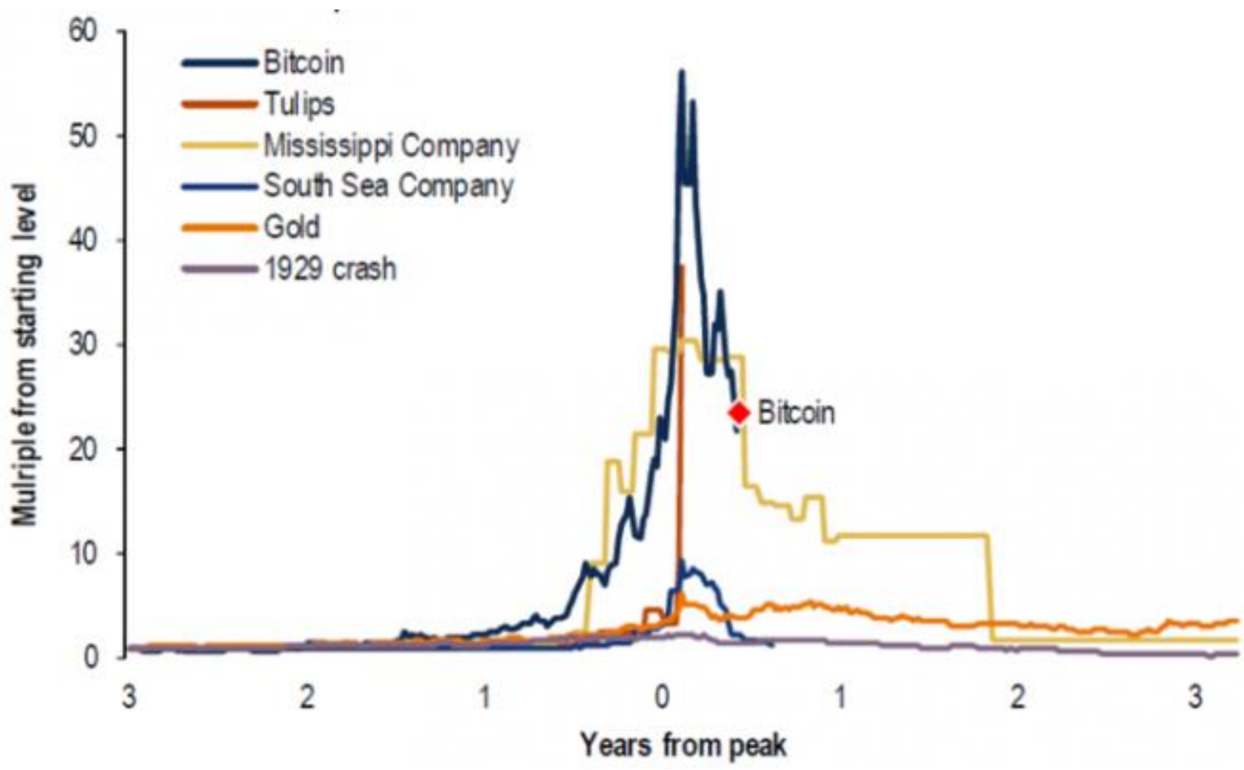


Exhibit B

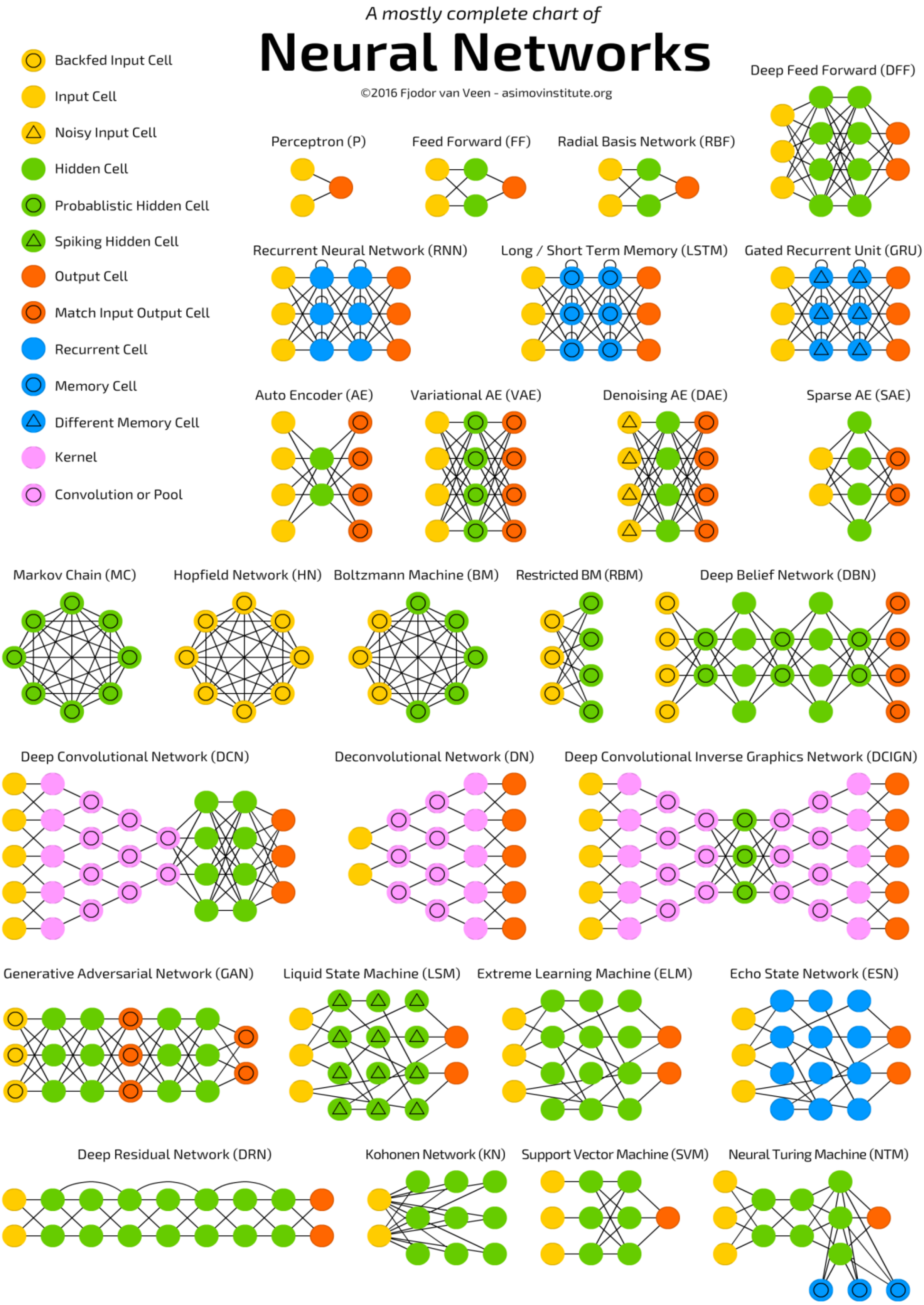

A mostly complete chart of
Neural Networks
©2016 Fjodor van Veen - asimovinstitute.org
Backfed Input Cell
Input Cell
Noisy Input Cell
Hidden Cell
Probablistic Hidden Cell
Spiking Hidden Cell
Output Cell
Match Input Output Cell
Recurrent Cell
Memory Cell
Different Memory Cell
Kernel
Convolution or Pool
Perceptron (P)
Feed Forward (FF)
Radial Basis Network (RBF)
Deep Feed Forward (DFF)
Recurrent Neural Network (RNN)
Long / Short Term Memory (LSTM)
Gated Recurrent Unit (GRU)
Auto Encoder (AE)
Variational AE (VAE)
Denoising AE (DAE)
Sparse AE (SAE)
Markov Chain (MC)
Hopfield Network (HN)
Boltzmann Machine (BM)
Restricted BM (RBM)
Deep Belief Network (DBN)
Deep Convolutional Network (DCN)
Deconvolutional Network (DN)
Deep Convolutional Inverse Graphics Network (DCIGN)
Generative Adversarial Network (GAN)
Liquid State Machine (LSM)
Extreme Learning Machine (ELM)
Echo State Network (ESN)
Deep Residual Network (DRN)
Kohonen Network (KN)
Support Vector Machine (SVM)
Neural Turing Machine (NTM)

Exhibit C

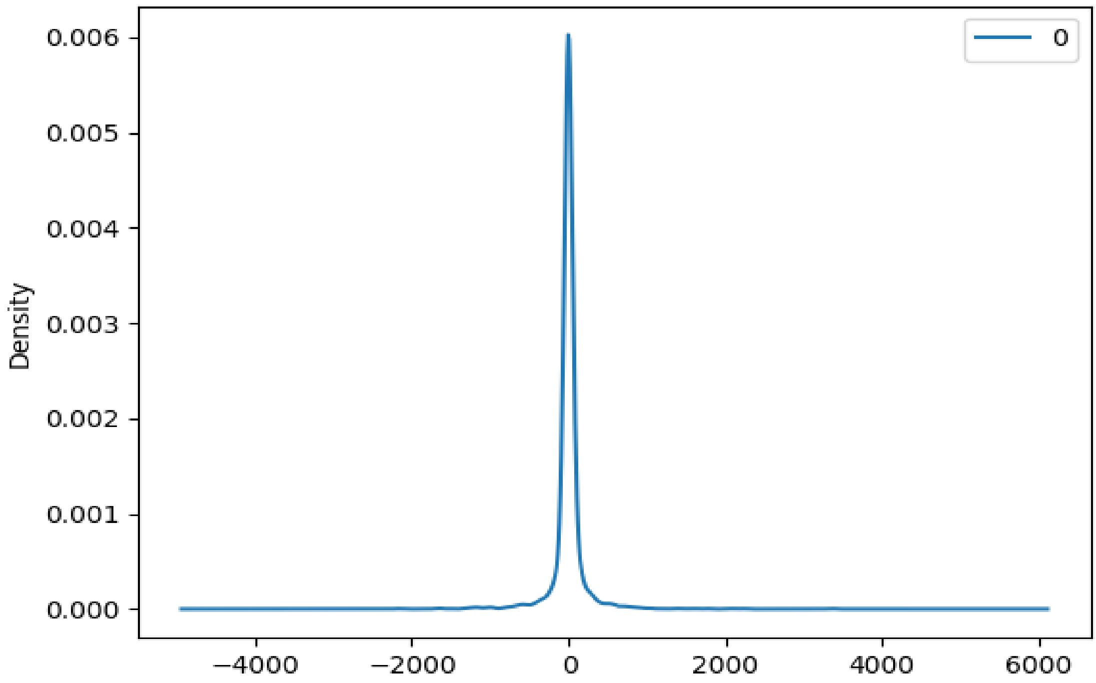
Here follows the regression results for all of the classical linear statistical models used in chapter 3:

AR model summary:

```
                              AR Model Results
==============================================================================
Dep. Variable:                  D.Price   No. Observations:               2188
Model:                            AR(5)   Log Likelihood            -14959.482
Method:                         css-mle   S.D. of innovations          225.448
Date:                              2019   AIC                        29932.963
Time:                          06:00:44   BIC                        29972.799
Sample:                               1   HQIC                       29947.524

=================================================================================
                    coef    std err          z      P>|z|      [0.025      0.975]
---------------------------------------------------------------------------------
const             2.4366      5.466      0.446      0.656      -8.276      13.149
ar.L1.D.Price     0.0837      0.021      3.956      0.000       0.042       0.125
ar.L2.D.Price    -0.0330      0.021     -1.561      0.119      -0.074       0.008
ar.L3.D.Price    -0.0014      0.021     -0.065      0.948      -0.043       0.040
ar.L4.D.Price    -0.0762      0.021     -3.602      0.000      -0.118      -0.035
ar.L5.D.Price     0.1453      0.021      6.876      0.000       0.104       0.187
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1           -1.1000          -0.9064j            1.4253           -0.3903
AR.2           -1.1000          +0.9064j            1.4253            0.3903
AR.3            0.5722          -1.3480j            1.4644           -0.1861
AR.4            0.5722          +1.3480j            1.4644            0.1861
AR.5            1.5799          -0.0000j            1.5799           -0.0000
-----------------------------------------------------------------------------
```

AR residual plot:


0.006
0.005
0.004
0.003
0.002
0.001
0.000
Density
−4000
−2000
0
2000
4000
6000
0