Kaunas University of Technology
Faculty of Mathematics and Natural Sciences

Prediction of Recovery Time After a Global Stock Market Index Crash Using Neural Networks

Master's Final Degree Project

Mindaugas Švirinas
Project author

Assoc. Prof. Dr. Audrius Kabašinskas
Supervisor

Assoc. Prof. Dr. Rasa Norvaišienė
Supervisor

Kaunas, 2023

Kaunas University of Technology
Faculty of Mathematics and Natural Sciences

Prediction of Recovery Time After a Global Stock Market Index Crash Using Neural Networks

Master's Final Degree Project
Business Big Data Analytics (6213AX001)

Mindaugas Švirinas
Project author

Assoc. Prof. Dr. Audrius Kabašinskas
Supervisor

Assoc. Prof. Dr. Rasa Norvaišienė
Supervisor

Prof. Dr. Mayur Pal
Reviewer

Prof. Dr. Vytautas Snieška
Reviewer

Kaunas, 2023

Kaunas University of Technology
Faculty of Mathematics and Natural Sciences
Mindaugas Švirinas

Prediction of Recovery Time After a Global Stock Market Index Crash Using Neural Networks

Declaration of Academic Integrity

I confirm the following:

1. I have prepared the final degree project independently and honestly without any violations of the copyrights or other rights of others, following the provisions of the Law on Copyrights and Related Rights of the Republic of Lithuania, the Regulations on the Management and Transfer of Intellectual Property of Kaunas University of Technology (hereinafter – University) and the ethical requirements stipulated by the Code of Academic Ethics of the University;

2. All the data and research results provided in the final degree project are correct and obtained legally; none of the parts of this project are plagiarised from any printed or electronic sources; all the quotations and references provided in the text of the final degree project are indicated in the list of references;

3. I have not paid anyone any monetary funds for the final degree project or the parts thereof unless required by the law;

4. I understand that in the case of any discovery of the fact of dishonesty or violation of any rights of others, the academic penalties will be imposed on me under the procedure applied at the University; I will be expelled from the University and my final degree project can be submitted to the Office of the Ombudsperson for Academic Ethics and Procedures in the examination of a possible violation of academic ethics.

Mindaugas Švirinas

Confirmed electronically

Švirinas Mindaugas. Prediction of Recovery Time After a Global Stock Market Index Crash Using Neural Networks. Master's Final Degree Project / supervisors Assoc. Prof. Dr. Audrius Kabašinskas and Assoc. Prof. Dr. Rasa Norvaišienė; Faculty of Mathematics and Natural Sciences, Kaunas University of Technology.

Study field and area (study field group): Applied Mathematics (Mathematical Sciences).

Keywords: neural networks, financial markets, crisis, recovery, forecasting, prediction.

Vilnius, 2023. 76 pages.

Summary

The recovery of stock markets after a crash is a welcome development. The ability to estimate recovery length would help investors plan and achieve higher returns, protect against external risks and diversify better. Simultaneously, policy researchers and institutions could take more efficient fiscal and monetary policy actions during a crisis. However, predicting the length of the recovery process is rather challenging, as it can be impacted by various outside variables that change from crisis to crisis, such as the number of originators, the depth and duration of a crash, policy responses, an economic situation and psychological beliefs of individuals. This thesis used MSCI World Index monthly data between December 1969 and December 2022 and incorporated 19 popular features to forecast the index's recovery, defined as a return to pre-crash stock market peak price. For predictions, four neural networks-based models were chosen to be explored: multilayer perceptron, recurrent neural networks, long-short-term memory and gated recurrent unit. They were preferred over other machine learning models due to their ability to process sequential data. Statistical models were also not investigated due to prior research revealing their poorer performance attributable to stock market indices prices being chaotic, noisy and nonlinear. The final trained models, based on any historical 72 months information, are capable of providing price forecasts 72 months into the future. 72 months forecasting length was chosen due to research showing that crashes typically recover in less than six years. It was discovered that the multilayer perceptron model acquired the lowest MSE performance metric, equal to 0.09 on the entire test set, ranging from Great Recession to December 2022, with one hidden layer, eight nodes, a learning rate of 0.001, Adam optimiser and a ReLU activation function. The best MSE score measured only during the Great Recession period and equal to 0.03 was acquired by the multilayer perceptron with one hidden layer, 64 nodes, a learning rate of 0.001, Adam optimiser and a ReLU activation function. Regarding results from MAE and MAPE metrics perspective, multilayer perceptron models performed the best also for MAE and MAPE full test set data. The models did not beat the repeating time series baseline model only for MAPE and MAE values calculated during the Great Recession period. Yet, there was no drastic performance difference between the top baseline model and the runner-up neural network models in these cases. The performance of RNN, LSTM, and GRU models fell far short of MLP despite high hopes for them. Based on all MSE and MAE metrics and MAPE calculated on a full test dataset, the order of MLP being the best, followed by GRU, LSTM and then RNN was maintained. However, on MAPE calculated during the Great Recession period, GRUs scored a 13.35 % error, followed by RNNs at 15.06 %, MLPs at 15.99 % and LSTMs at 31.85 %. It was also noticed that the best models performed better during non-crash periods compared to crash periods. Interestingly, when forecasting recovery duration, the models usually commenced from a slightly larger or smaller price drop compared to the actual price but also typically displayed slower price-rising tendencies, eventually leading to an

almost perfect match between the actual price and anticipated price. If one used the best MLP model during the lowest price point of the Great Recession, then the recovery point would have been predicted on point. GRU would have led to prematurely announcing recovery by a few months and using the best RNN and LSTM models, recovery would have been announced two years prematurely. Finally, the US consumer price index and the price of the MSCI World Index were the two factors that the results showed to be most crucial for making forecasts, followed by information on treasury bills and notes, the number of months the current crisis period has been ongoing, Williams % R, MACD and Stochastic % K technical analysis indicators. The least important indicators were the percentage price decline from the pre-crisis peak, the length of months that the present non-crisis phase has been continuing and the change in the gold price.

Mindaugas Švirinas. Atsigavimo laiko po globalaus akcijų rinkos indekso nuosmukio prognozavimas pasitelkiant neuroninių tinklų modelius. Magistro baigiamasis projektas / vadovai doc. dr. Audrius Kabašinskas ir doc. dr. Rasa Norvaišienė; Kauno technologijos universitetas, Matematikos ir gamtos mokslų fakultetas fakultetas.

Studijų kryptis ir sritis (studijų krypčių grupė): Taikomoji matematika (Matematikos mokslai).

Reikšminiai žodžiai: neuroniniai tinklai, finansinės rinkos, krizė, atsigavimas, prognozavimas.

Vilnius, 2023. 76. p.

Santrauka

Akcijų rinkų atsigavimas po nuosmukio yra lauktinas įvykis. Galimybė įvertinti atsigavimo trukmę leistų investuotojams planuoti ir pasiekti didesnę grąžą, apsisaugoti nuo išorinės rizikos ir geriau diversifikuoti savo investicijas. Taip pat politikos tyrėjai ir institucijos galėtų imtis veiksmingesnių fiskalinės ir pinigų politikos veiksmų krizės metu. Tačiau prognozuoti atsigavimo trukmę yra gana sudėtinga, nes ją gali paveikti įvairios priežastys, kintančios priklausomai nuo krizės, pavyzdžiui, veiksnių, sukėlusių nuosmukį, skaičius, nuosmukio gylis ir trukmė, politiniai veiksmai, ekonominė padėtis ir individų įsitikinimai. Šioje disertacijoje buvo naudojami MSCI World Index mėnesiniai duomenys nuo 1969 m. gruodžio mėn. iki 2022 m. gruodžio mėn. ir įtraukta 19 populiarių ypatybių, leidžiančių prognozuoti indekso atsigavimą, apibrėžiamą kaip grįžimą į iki krizės buvusią aukščiausią akcijų rinkos kainą. Prognozavimui buvo pasirinkti keturi neuroniniais tinklais paremti modeliai: daugiasluoksnis perceptronas (MLP), rekurentiniai neuroniniai tinklai (RNN), ilgalaikės trumpalaikės-atminties neuroniniai tinklai (LSTM) ir sulaikomo pasikartojančio vieneto neuroniniai tinklai (GRU). Neuroninių tinklų algoritmai buvo pasirinkti vietoje kitų mašininio mokymosi modelių dėl jų gebėjimo apdoroti sekos duomenis. Statistiniai modeliai taip pat nebuvo tiriami dėl ankstesnių tyrimų, atskleidusių prastesnius jų rezultatus dėl akcijų rinkos indeksų kainų chaotiškumo, triukšmingumo ir netiesiškumo. Galutiniams modeliams pateikus istorinę 72 mėnesių informaciją, yra grąžinamos kainų prognozes 72 mėnesius į ateitį. 72 mėnesių prognozavimo trukmė pasirinkta pagrindžiant tyrimais, rodančiais, kad nuosmukiai paprastai atsigauna per mažiau nei šešerius metus. Atlikus tyrimą buvo nustatyta, kad daugiasluoksnis perceptrono modelis su vienu paslėptu sluoksniu, aštuoniais mazgais, mokymosi rodikliu lygiu 0.001, Adam optimizavimo funckija ir ReLU aktyvacijos funkcija įgijo žemiausią MSE rodiklio rezultatą, lygų 0.09 visame bandymo rinkinyje tarp 2007 metų krizės ir 2022 m. gruodžio mėn. Geriausias MSE rezultatas, išmatuotas tik 2007 metų krizės laikotarpiu, buvo lygus 0.03, kurį pasiekė kitas daugiasluoksnis perceptronas su vienu paslėptu sluoksniu, 64 mazgais, mokymosi rodikliu 0.001, Adam optimizavimo funckija ir ReLU aktyvacijos funkcija. Vertinant rezultatus iš MAE ir MAPE rodiklių perspektyvos, daugiasluoksniai perceptronų modeliai taip pat geriausiai pasirodė pagal MAE ir MAPE rodiklius, paskaičiuotus visam bandomajam rinkiniui. Modeliai neįveikė bazinio modelio, atkartojančio istorinę laiko eilutę, tik MAPE ir MAE vertėms, apskaičiuotoms 2007 metų finansinės krizės laikotarpiu. Tačiau šiais atvejais nebuvo ženklaus efektyvumo skirtumo tarp geriausio bazinio modelio ir antrąją vietą užimančio giliojo mokymosi modelio. Nepaisant didelių lūkesčių, RNN, LSTM ir GRU modelių rezultatai gerokai nusileido MLP. Remiantis abejais MSE ir MAE rodikliais ir MAPE rodikliu, apskaičiuotu panaudojant visą bandymo duomenų rinkinį, MLP modeliai buvo geriausi, po jų sekė GRU, LSTM ir RNN modeliai. Tačiau pagal MAPE, apskaičiuotą 2007 metų finansinės krizės laikotarpiu, GRU surinko 13.35 % paklaidą, po to RNN – 15.06 %, MLP – 15.99 % ir LSTM – 31.85 %. Taip pat

tyrimo metu buvo pastebėta, kad geriausi modeliai prognozuoja geriau ne nuosmukio laikotarpiais, palyginti su nuosmukio metu. Įdomu tai, kad prognozuojant atsigavimo trukmę, modeliai dažniausiai pradėdavo artimiausios kainos spėjimą nuo šiek tiek didesnio ar mažesnio kainų kritimo palyginti su faktine kaina, bet paprastai turėjo lėtesnes ateities kainų kilimo tendencijas, kurios galiausiai lėmė beveik tobulą tikrosios ir numatomos kainos atitiktį. Jei būtų naudojamas geriausias MLP modelis 2007 metų finansinės krizės žemiausios kainos metu, tuomet atsigavimo taškas būtų buvęs prognozuojamas tiksliai. GRU modelis būtų nustatęs atsigavimą keliais mėnesiais per anksti, o naudojant geriausius RNN ir LSTM modelius, atsigavimas būtų paskelbtas dvejais metais anksčiau. Galiausiai, JAV vartotojų kainų indeksas ir MSCI pasaulinio indekso kaina buvo svarbiausi du veiksniai atliekant prognozes. Taip pat modeliams buvo svarbi informacija ir apie iždo vertybinius popierius, kiek mėnesių jau tęsiasi krizė bei Williams % R, MACD ir Stochastinio % K techninės analizės rodikliai. Mažiausiai svarbūs rodikliai buvo procentinis kainos nuosmukis nuo prieš krizinės aukščiausios kainos, kiek mėnesių tęsiasi nekrizinis laikotarpis ir aukso kainos pokytis.

Table of Contents

List of Figures

List of Tables

List of Abbreviations and Terms

Abbreviations:

AI – Artificial intelligence;

CNN – Convolutional Neural Network;

DMLP – Deep Multilayer Perceptron;

Dr – Doctor;

ECB – European Central Bank;

FED – Federal Reserve Board;

GDP – Gross Domestic Product;

GRU – Gated Recurrent Unit;

LSTM – Long Short-term Memory;

MAPE – Mean Absolute Error;

MAPE – Mean Absolute Percentage Error;

MLP – Multilayer Perceptron;

MSCI – Morgan Stanley Capital International;

MSE – Mean Squared Error;

NBER – National Bureau of Economic Research;

RMSE – Root Mean Squared Error;

RNN – Recurrent Neural Network;

SHAP – Shapley Additive Explanations;

US – United States.

Introduction

Importance of topic. The recovery of stock markets after a crash is a welcome development as billions of dollars are traded daily and price swings significantly impact corporations and citizens' livelihoods. Moreover, the ability to estimate recovery length would allow investors to plan and earn higher returns, protect against external risks, share risks, diversify better and more. At the same time, policy researchers and institutions could take more efficient fiscal and monetary policy actions during a crisis. However, predicting the length of the recovery process is quite complex as it can be influenced by various external factors that vary from crisis to crisis and there is no set criteria for determining when the stock market has already recovered. Additionally, the research topic of stock market recovery following a crash has been previously assessed only to a minimal extent, with a majority of the research community focussing on either predicting crashes in advance or studying recoveries from a GDP perspective.

Questions analysed in the project. What is considered to be a crisis, a recession and an asset price bubble? What constitutes a recovery after a crisis? What are the drivers potentially influencing the length of recovery? Which neural networks-based method can predict stock market recovery after a crisis the best at any given moment? Do the models predict more accurately during a crisis or expansion after the recovery period? Which features contribute the most to the model's predictive power during a crash?

The object of the project – MSCI World Index, having monthly price recordings between December 1969 and December 2022.

Aim of the project – Identify crash periods and using neural networks-based methods create a model, capable of predicting recovery after a stock market crash at any given time.

Tasks of project.

1. Analyse current literature on crises, recessions and asset price bubbles to set definitions and understand the causes of crashes and drivers of recovery;
2. Review the literature on stock market price forecasting to find probable candidates for the best models and their optimisation techniques;
3. Select the most popular features, which potentially could improve models forecasting performance, create new derivative features using Markov Regime Switching and Performance Analytics package models to identify bull and bear episodes, pinpoint exact points of recovery;
4. Predict the stock market price multiple time steps in advance during the Great Recession period using different neural networks models and constitute the best model, which can help determine the remaining recovery length;
5. Understand which selected features contribute the most to the model's predictive performance during a crash;
6. Explore the best model prediction performance for other stock market regimes.

Structure of the project.

There are three sections to the project. The first part provides an overview of the definitions of crises, recessions, asset price bubbles and recoveries. It also analyses the causes and effects of crashes and possible factors that may affect the length of recovery. Furthermore, it overviews stock market forecasting difficulties, approaches and best models' optimisation practices. The second section provides descriptive statistics for the MSCI World Index, features used in forecasting and research methodology. The final section summarises the results by showing how the models performed throughout the crash and expansionary phases, analyses features' contribution to predictions and lists limitations and possible areas for further research.

1. Literature Analysis

As there is no widely accepted definition of a systemic crisis, recession, asset price bubble and recovery, various definitions will be overviewed in this literature analysis section. After that, the impact and behaviour of crashes will be analysed to understand the significance of research in predicting recovery length after a crash. Later, some factors that contribute to crashes and impact recovery time will be covered to reuse part of the features in modelling and be capable of better interpreting final research results.

Since most of the research community has been either focused on predicting crashes in advance or s tudying recoveries from a GDP perspective, the research topic of stock market recovery following a crash has only been evaluated to a minimal extent in the past. Due to the novelty of the subject, the literature study will draw inspiration from scholars who studied crises and recoveries from the persp ectives of both stock markets and economic indicators, where appropriate.

After overviewing the economic side, the literature analysis will shift focus to stock market forecasting aspects. In particular, challenges associated with forecasting the stock market, various features employed in predictions by other researchers, the typical length of time steps predicted and the statistical and machine learning models used will be discussed. Additionally, a review of how other researchers are optimising neural networks performance and turning the models from "black boxes" into explainable ones will be done.

1.1. Definitions of a Systemic Crisis, a Recession and an Asset Price Bubble

The occurrence of turbulences in a global economy is a recurrent phenomenon that can have significant negative economic and social consequences. Some of these longer-term turbulences can be categorised as systemic crises, some as recessions and some as asset price bubbles. However, it is essential not to use systemic crisis, recession and a bubble as interchangeable terms as they can represent different economic phenomena with potentially distinct impacts and implications depending on the authors' definitions. These definitions will be analysed in the following subsection.

1.1.1. Systemic Crisis

Caprio and Klingebiel [1] defined systemic crises as periods of mass bank liquidations or insolvencies, having a significant share of non-performing loans and large-scale government intervention to support banks. Homar and van Wijnbergen [2] used a similar systemic crisis definition where at least two conditions should be satisfied: the presence of a significant impact on the banking system, such as bank runs, large losses of bank capital or liquidations and significant policy interventions. Serious policy interventions condition, according to authors, can be met if at least three circumstances from below are in place: extensive liquidity support exists, gross bank restructuring costs reach at least 3 % of GDP, there are significant bank nationalisations, significant guarantees on bank liabilities are present, asset purchases amount to at least 5 % GDP and deposit freezes or bank holidays is in effect [2]. On the one hand, there were also some authors that gave less dimensional definitions of crises only through the prism of stock market pricing. For example, Mishkin and White [3] defined crises as just falls in the price of security or index below a certain threshold, in the authors' case below 20 % and over a certain period of time, in the authors' case over a day, five days, one month, three months and one year. On the other hand, Patel and Sarkar [4] calculated an indicator called CMAX that detects extreme price levels over a certain period. This indicator divides the current price by the maximum last 2-year price and defines a crisis where the indicator decreases below 35 % in emerging markets and 20 % in developed markets [4]. Hence, it is evident that the academic community can define crisis more globally, where an entire economic or financial system is at risk of collapse and several conditions are in place or more locally through the stock market price falling below some extreme value or ratio.

1.1.2. Recession

Conversely, a recession is often defined as a sustained period of economic decline. NBER definition characterises recession as a significant decrease in economic activity, which lasts more than a few months and affects an economy broadly, i.e. is not restricted to one sector [5]. NBER uses multiple monthly measures of economic activity, such as real personal income minus transfers, nonfarm payroll employment, employment, real personal consumption expenditures, wholesale-retail sales adjusted for price changes and industrial production [5]. Other definitions, such as by Matias Braun and Borja Larrain [6], defined as the period between peak to trough, where the trough is defined as cyclical GDP being more than one country-specific standard deviation below zero and the peak is described as a year in the past where cyclical GDP is larger than in the predate and post-date years. While both a systemic crisis and a recession can have drastic economic and social consequences, a systemic crisis is usually more severe and can lead to widespread instability.

1.1.3. Asset Price Bubble

Stock markets can experience additional phenomena, which can cause a sudden decrease in prices called asset price bubbles. Bubbles are "large, sustained mispricings of financial or real assets" [7]. Although not all mispricings are called bubbles, bubbles are usually associated with mispricings that are often explosive or price tremendously exceeds the fundamental price and assets are purchased due to resale opportunities, i.e. it is expected that the asset could be sold at an even higher price later [7]. The concept of bubbles is crucial because they can result in excessive investment in inflated assets. When bubbles burst and lead to a fall in stock market assets price, they could slow down real economic activity and negatively impact the balance sheets of businesses, individuals and other institutions [7]. This negative impact is especially possible if the assets were purchased using credit, which can have amplification and spillover effects [7].

The term "crash" will be preferred in this research thesis to describe periods of stock market decline because it incorporates the concepts of "systemic crises", "recessions" and "asset price bubbles". When appropriate, crises, recessions, and asset price definitions will also be used to describe directly relevant economic events that are taking place at the same time as the stock market decline period under study.

1.2. Definitions of Recovery After a Crash

There is also no universally accepted definition of recovery time after a systemic crisis, recession or asset price bubble as the recovery concept encompasses a complex set of multiple phenomena and would require describing what it means to return to normal, which could be subjective. Reinhart and Rogoff [8] have chosen the easiest-to-measure definition of recovery - the number of years it takes to reach the prior peak of the analysis object, which in their case was real per capita income. Fatás and Mihov discussed this definition's shortcomings. They argued that such a definition of returning to the peak is not ideal as it ignores the growth aspect of the trend, which can occur even during the recession and ignores the length of the recovery phase [9]. The growth aspect becomes vital during more prolonged financial crises where not including it could lead to the premature declaration of recovery [9]. However, other authors argue that as a trend is computed from a statistical calculation, it may not reflect an efficient output level and the efficient point may also decline during the crisis [10]. Other authors, such as Stiglitz [11], argued that the recovery should be defined through the prism of affected individuals, which covers multiple areas such as unemployment returning to normal levels and growth having resumed; other factors, such as an exchange rate stability, if a prolonged recession accompanies it, should not be included. However, these two variables have drawbacks of their own such as the existence of uncertainty around these measures and estimates of these variables may require revisions. These two variables also allow the economy to be below or above potential, posing a challenge of consistently dating recoveries. Thus, it is evident that there is no one widely accepted definition of what it means to return to normal: more straightforward methods, namely time to return to pre-crisis peak, are criticised for ignoring the growth aspect, while more complicated methods risk of bringing noise or inaccuracies due to difficulties of calculations.

The recovery of the stock market index to its pre-crash highest price will be the preferred definition of recovery in this research thesis. Nevertheless, the following literature analysis will still, for some authors, contain recovery definitions that are different from the preferred one. If so, this will be mentioned.

1.3. Causes of Systemic Crises, Recessions and Asset Price Bubbles

Understanding factors that might influence crises, recessions, and asset price bubbles are of high importance and can assist in establishing a more in-depth understanding of the dynamics of crisis and help predict the stock market’s recovery from a crash. Recessions and systemic crises may share some common origins, but the impact of mutual causes might vary in severity, thus, they will be studied together.

1.3.1. Causes of Systemic Crises and Recessions

Kannan et al. [12] analysed the most popular shocks that lead to a recession and found that the most popular shock causing a recession was oil shock, followed by monetary and fiscal policy shock and external demand shock. Additionally, some documentation shows that the beginning of international political crises typically increases the volatility of stock market returns [12]. The strength in opposing stock market volatility depends on the political crisis severity and the number of involved major powers [12]. Moreover, financial markets are affected by disasters, such as the Covid-19 pandemic, where the more extensive the severity of outbreak cases, the more substantial negative returns of financial markets were noticed [13]. Kannan et al. [12] also noticed that during the Great Moderation period, which lasted between the 1980s and the Financial crisis of 2007, recessions associated with financial elements became more common.

Regarding crises or recessions associated with financial components, ECB has distinguished five non-exclusive categories [14]. First, a systemic crisis or recession can be caused by the banking component, which is a type of risk where banks incur a particular type of hardship. Such distress can be driven by non-performance of loans or other assets, liquidity based, where withdrawals of money exceed available funds of the bank or interest rate based, forcing banks to pay larger amounts of money on its deposits and putting significant pressure [14]. Another type of crisis is a sovereign risk crisis, which could be defined by governments of countries having significant challenges in returning their sovereign debt interest or principal payments [14]. The third type of crisis is a currency risk crisis – a sharp decline in the foreign exchange value of the currency of a particular country, usually driven by market anticipation of domestic policies not being able to keep the exchange rate stable [14]. The fourth cause of the crisis is asset price correction, which happens when a rapid decrease in prices in a specific class of assets is happening [14]. Finally, the ECB had the fifth category of crisis defined as transitory [14]. A specific example of this kind of crisis is the collapse of the Soviet Union and its aftereffects.

Brunnermeier [7] also observed that the trigger that begins a crisis does not have to be a significant economic event. Due to amplification effects small triggers can create an enormous crisis, as happened during the 2007 Financial crisis when the subprime mortgage market made only 4 % of all mortgage market but managed to have a worldwide impact [7]. Such amplification effects can be direct, i.e. tied by direct contractual links such as bank runs or one bank's dependence on another bank to cover its obligations and indirect, caused by spillovers due to similar exposure such as liquidation of assets by one bank driving the decrease in balance sheet value of assets in another party [7].

1.3.2. Causes of Bubbles

When it comes to asset bubbles or imbalances, many driving factors can strengthen the imbalance and, later, sudden decline in prices. Bubbles can be driven by belief distortions, such as investors thinking that “this time is different” due to a lack of sufficient data to notice the establishment of a bubble [7]. Investors may also ignore cautionary tales from history by believing that they will not repeat or not having the right expectations due to facing bubbles for the first time in their lifetime [7]. Insignificant news can also act as a trigger of significant volatility because they allow investors to synchronise selling strategies at the same time [15]. Moreover, the beginning of bubble formation usually happens in low volatility periods, when financing is more accessible and leverage is used to reduce the difference in returns between different risk-class assets [7]. Such situation can potentially disbalance financial markets when margin calls happen and some investors may be forced to sell the assets at a disadvantageous price, further depressing the prices and forcing other investors to sell their assets [7].

Moreover, there is evidence that sophisticated investors benefit from riding bubbles; for example, between 1998 and 2000, hedge funds invested heavily in overvalued technology stocks, capturing the returns and thus driving the bubble price even further and exciting just before downturns [16]. Investment through third parties portfolio managers can also expedite an asset bubble's formation. Such managers can have short-term thinking, refrain from long-term investment opportunities and buy bubble assets to avoid temporary losses that lead to fund outflows [15]. Furthermore, fund managers may be incentivised to buy overvalued assets even if they know they will start falling eventually to capture potential profit at the expense of investors who hire them and later “churn” on them [16]. Finally, when the bubble is near the explosion point and is volatile, risk shifting may lead to fund managers “double downing” on overpriced assets to recover the losses, prolonging the bubble and increasing its detrimental effects [7]. These challenges in identifying bubbles and some actors' incentives to invest in bubbles, even though negative externalities will be present, make overinvestment likely. For investors to protect against investments in bubbles, one would need to know the fundamental value of assets, which is a difficult task.

1.4. Effect of Crashes on Economic Variables and Stock Market Asset Prices

Even though the research community typically does not define and analyse crises through changes in stock market prices [17], the stock market can serve as a great reflection of the economic situation and, therefore, can be used as an object of research to analyse systemic crises and recessions impact and recovery. For example, Shuddhasattwa [18], by analysing stock prices responses in 17 countries over 145 years, shows that disasters and financial and non-financial recessions immediately pressurise stock prices and any monetary and fiscal policy intervention has an instantaneous effect on the stock market. Thus, the stock market analysed with the economic variables brings additional value and this subchapter will show similarities in the dynamics of economic variables and stock markets.

1.4.1. Effect of Crashes on Economic Variables

According to Reinhart and Rogoff [19], a prolonged decrease in real per capita GDP during the financial crisis is due to the extended recovery period. In a sample of 63 crises in advanced economies, it was noticed that, on average, it took 7.3 years to recover from the financial crash to before crash levels in real per capita GDP terms [19]. The median recovery time was six years and the average crash caused around a 9.6 % drop from the peak [19]. Moreover, on average, 42.9 % of crises experienced episodes of double dips, which is a renewed downturn before reaching prior levels [19]. Papell and Prodan have discovered similar findings where in advanced economies the return to the potential GDP after a financial crisis is longer than the return following other types of recessions and takes an average of 9 years to return [20]. Hence, there is evidence in the literature that systemic crises associated with financial components tend to be deep and long.

Wan and Jin [21] also analysed the crises impact on the economy, except the recovery metric chosen was not peak-to-peak but return to trend. The researchers noticed that the currency crisis cumulative output loss relative to the trend in developed countries was 14.83 % or 2.81 % annually, average recovery time was around 4.5 years. For the banking crisis, cumulative output loss relative to the trend was 17.20 % or 3.4 % annually and recovery time reached 6.08 years. It shows that the banking crisis is more profound and longer than the currency crisis in developed countries with respect to trend [21]. Laeven and Valencia investigated further into different effects of crises having financial components. They found that output loss during sovereign crises (median 40.38 %) is more considerable compared to banking crises (median 19.74 %), but banking crisis losses are more considerable than currency crises (median 3.04 %) [10]. Additionally, the researchers observed that currency crisis combined with sovereign crisis (median 61.19 %) on average generates a higher loss of output compared to banking crises or banking crises and currency crises combined (median 18.83 %) [10].

Regarding crisis dynamics, according to Bordo and Joseph [22], significant contractions in output tend to be followed by large business expansions and mild contractions by mild expansions making recoveries at least as rapid as the downward period. This pattern was noticed only during crisis recoveries. Noncrisis recoveries did not exhibit the following patterns.

1.4.2. Effect of Crashes on Stock Market Asset Prices

The recent capacity of the stock market to act as a near-real-time reflection of economic development can be potentially attributed to several factors, among which is the relative ease of participation facilitated by the low barriers to entry and the speed of transactions [18]. That, together with more active government stabilisation policies, can partly explain why stock market prices after a financial crisis are expected to rebound to pre-crisis levels within a relatively short span of 4-6 years in the post-World War II era [18]. Compared to the pre-World War II era, the expected recovery period was almost nine years [18]. Moreover, periods in the stock market with more significant price drops usually exhibit a positive association with a more robust recovery in stock market price levels [17]. Additionally, Goetzmann and Kim [23] found that during a decline in the markets of at least 50 %, significant positive returns are more probable compared to more modest crashes of 10-20 % decline, which sometimes exhibit different dynamics of more likely to being followed by a further decline, i.e. persistence is noticed instead of reversal. When analysing global financial market recovery after the 2008 Global Financial crisis, Foo [24] observed that market recovery is not homogenous, with increasing and decreasing trends among different markets, where more fluctuations were present in more emerging markets. Thus, similarly, as output recovery from a crisis period, stock market recovery exhibits similar patterns where the more extensive the crash, the quicker the recovery. However, the stock market recovery could also include increasing and decreasing trends.

1.5. Factors Contributing to Recovery Length

With regard to banking crises between 1970-2012 analysed through a GDP recovery perspective, countries experiencing later recovery and more prolonged periods can have one of the following domestic factors: having multiple types of crisis originators, having large banking sector, more significant budget deficits, overvalued currencies, and sizeable monetary expansion [25]. Wan and Jin also detected that control of private sector credit and financial openness, adjustments of the current account deficit and a favourable economic environment also contribute to changes in output recovery speed [21]. In addition, external variables such as low growth in world trade, increased uncertainty in financial markets, which can be described by volatility in gold or other stock market assets price and global interest rate shocks for middle-income countries that are reliant on external finance are also correlating with deeper recessions and more extended recovery periods [25]. Brunnermeier and Oehmke [7] also identified that currency devaluation could deepen banking crises if a debt is in foreign currency. Notwithstanding whether a recession has financial crisis elements, the depth and duration of the recession also have predictive power on the pace of recovery. Deeper recessions are associated with more robust growth in the first three years of recovery, while more prolonged recessions are associated with slower post-trough growth [26].

When it comes to assisting banks during systemic crises, Homar and van Wijnbergen analysed the effect of government and central bank intervention in 69 systematic banking crises between 1980-2014. They discovered that timely and sufficient bank recapitalisations reduce the duration of recessions on average by two years after the recapitalisation [27]. The authors argue that the reason behind this is a tendency for undercapitalised banks to be incentivised to pass loans to distressed borrowers instead of restructuring or liquidation [27]. That funds inefficient firms and limit credit to new borrowers. Laeven and Valencia [28] support this claim that recapitalisations increase the growth of financial firms. Chodorow-Reich [29] further provides evidence that decreased loans due to weakened banks affect employment negatively. However, the authors do not find evidence that other intervention mechanisms, including liquidity support, guarantees on bank liabilities or reserve money growth, contribute to decreased recession duration. In addition, even if policy responses, such as banking system recapitalisation, are present, recovery can still be sluggish as balance sheets of other parties may still be in bad shape [7].

Horii and Ono [30] found that the recovery process also depends on the realised frequency of the shock. If the shock occurs infrequently, it affects individuals' beliefs in a limited way and recovery is quicker [30]. Nevertheless, if people notice shocks multiple times, they start believing that the current situation is unstable and reduces consumption [30]. It takes a long time to reverse beliefs and increase consumption to previous levels [30]. In this case, the recovery speed is slow initially and gradually increasing and stabilising [30].

Regarding an asset bubble burst, some different dynamics could be in place. If the bubble formation is driven primarily by credit and high leverage, it leads to more robust amplification mechanisms and, thus, more profound crises [7]. For example, the technology bubble burst in 2000 led to immense wealth destruction, but the actual economic impact was minor compared to the 2007 financial crisis housing bubble burst, which was driven more by credit [7].

To summarise the economic part of the literature analysis, there is considerable evidence that crashes significantly affect the economy and stock market asset prices. Larger financial meltdowns can eradicate wealth drastically and take up to 6 years to recover. Moreover, the recovery may fluctuate and be influenced by factors that caused the crash and the reactions of institutions and investors. As a result, knowing when to expect stock market recovery is challenging and requires careful consideration of multidimensional factors for each case of a market downturn. The succeeding subchapters will analyse literature from a mathematical viewpoint to assess the applicability of various models for predicting the recovery duration of stock market asset prices.

1.6. Difficulties in Forecasting Stock Market Price

Fama [31] introduced the efficient market hypothesis, which states that future stock market prices cannot be predicted due to their behaviour depicting random walk. The efficient market hypothesis has three versions: weak, semi-strong and strong [31]. To begin with, a weak hypothesis states that past returns and return sequences are already included in stock market price, meaning technical analysis does not allow to outperform buy and hold strategy [31]. A semi-strong hypothesis states that stock prices reflect all publicly available information, i.e. economic, social, political, environmental and others, which further prohibits investors from predicting future developments and outperforming the market [31]. Finally, the strong hypothesis states that all information, including insider information, is included in the stock price [31]. A latter supposition provides a very extreme outlook on stock market prices and it is unlikely that such hypothesis can exactly reflect real life. Nevertheless, based on Fama [31], there were multiple examples where a weak and semi-strong form efficient market hypothesis was present at the time of his writing. If one believes in a weak hypothesis, it means that technical analysis will not assist in predicting future prices. On the other hand, if one trusts semi-strong, no external features will be helpful, only insider information.

Nonetheless, with time efficient market hypothesis gathered more criticism and stock market researchers managed to identify anomalies in the stock market pricing. As Self and Mathur [32] commented: "The true underlying market structure of asset prices is still unknown. <...> for a period of time, it behaves according to the classical definition of an efficient market; then, for a period, it behaves in such a way that researchers are able to systematically find anomalies to the behavior expected of an efficient market". Self and Mathur [32] provided anecdotal evidence of how elementary technical trading rules generated highly differentiating results between some trading days of a particular index, suggesting the potential existence of market information inefficiencies.

A newer version of the efficient market hypothesis can describe different efficiency periods, which includes evolutionary principles and allows for investors‘ behavioural biases such as loss aversion, overconfidence, and overreaction, called the adaptive market hypothesis [33]. According to the adaptive market hypothesis, participants of stock markets are affected by the following effects: “Individuals act in their own self-interest; individuals make mistakes; individuals learn and adapt; competition drives adaptation and innovation; natural selection shapes market ecology; evolution determines market dynamics” [33]. Adaptive market hypothesis argues that market dynamics are driven by “selfish individuals, competition, adaptation, natural selection, and environmental conditions” [33]. The competition factor means that the more species compete for scarce resources, the more the market is likely to be efficient. In contrast, the fewer individuals compete for rich resources, the more the market will be less efficient. The adaptive market hypothesis differs from the efficient market hypothesis in a way that Fama’s hypothesis prohibits investors from making mistakes and there are no learning or adaptation moments as markets are always in equilibrium [33]. Lo [33] provides a further example, in which rolling first-order autocorrelation for monthly returns of the Standard & Poor’s Composite Index from January 1871 to April 2003 was computed and revealed that autocorrelation cyclically varies through time, with periods in the 1950s being more efficient than in the early 1990s. According to Random Walk Hypothesis, returns should be serially uncorrelated and it is expected that serial autocorrelation would become progressively smaller with years as the market becomes more efficient. Lo [33] argues that the cyclical nature of returns can be reasoned by the changes in institutional market participants. Such adaptive market hypothesis ideas imply that risk returns are changing due to changes in market participants, institutions and conditions. In other words, investment products have superior and inferior moments of performance and new opportunities are continually appearing and disappearing after being exploited [34].

Goetzmann and Kim [23] also discussed difficulties in stock market forecasting, where smaller declines in stock market price led to persistence and more significant declines led to a rebound, which suggests more complex behaviour of the stock market compared to simple mean reversion.

Additionally, An and Loungani [35] described private and public sector forecasts in terms of GDP near the recession period and discovered that the ability to predict turning points is limited. Usually, forecasting starts close to the average. It begins to depart only later, showing that forecasters are aware of the changes. However, the magnitudes of the revisions are much smaller than they should be to forecast recessions accurately [35]. It could be argued that such a situation arises due to three potential factors: forecasters do not have enough information to reliably assess recessions because recession can occur due to political crises, for example, which are difficult to anticipate, forecasters may not have the incentive to predict a recession, i.e. reputation loss could be more significant for incorrectly calling a recession and third reason could be that those forecasters have their own biases and revise them to slow to respond to incoming information [35].

Thus, forecasting stock market performance is challenging, with some authors arguing that future predictions are impossible due to market efficiency. In contrast, others contend that errors made by investors provide opportunity windows for such forecasts to occur. Still, one must exercise caution in acknowledging these difficulties and recognise that forecasting is susceptible to making mistakes.

1.7. Features Used in Forecasting

As market anomalies during some periods are present in stock markets, it is noticeable that market participants are using historical market prices, company-specific information and other factors to include in their future predictions about the stock market [36]. The two core investment strategies used are fundamental and technical analysis [37]. Fundamental analysis "relies heavily on the analysis of current and past financial statement data to identify when underlying firm value differs from prevailing market prices" [38]. When the specific stock price is below its expected value, the stock should be purchased as it is believed to be priced at a fair value in the future. If a specific stock price is above its expected value, it should be sold or not purchased. In contrast, technical analysis is based on the premise that stock markets do not follow efficient market hypothesis and current price does not include past price information, thus past price and volume data can be used to predict future price movements [39].

Kumbure et al. [36] reviewed 138 articles published between 2000 and 2019 which used machine learning to predict stocks or indices prices and summarised the features used in stock market future price prediction. In total, 2173 unique variables were detected by Kumbure et al. [36], which were allocated to one of the four following categories: technical indicators, fundamental indicators, macro-economy and others. Altogether, 1348 or 62 % of all indicators were from the technical indicators group, making it the largest group, the macro-economy group had 279 variables (12.8 %), making it the third most popular group and lastly, fundamental indicators, having 157 variables (7.2 %). The remaining 18 % are other indicators, not falling in any of the three groups [36].

The table below summarises the indicators by category and subcategory and lists them in descending order of the number of articles that used them. In order to reduce the number of indicators to a comprehensible size, the following summary excludes indicators only used in one or two articles. In addition, other indicators were also excluded, as they include out-of-scope methods of this Master thesis. For example, textual analysis of Twitter tweets, variables derived from capital asset pricing model, GARCH or linear regression, or the indicators partly overlap with technical indicators group, except they are derived not from researched index or stock, but from related ones, such as major world indices or largest companies in a particular country or some derivate indices such as VIX [36].

Table 1. Indicators used to forecast stocks' or indices' future prices using machine learning techniques

Category	Subcategory	Indicator	Number of articles
Technical indicators	Basic technical indicators	Close price with lag 1	57
		High price with lag 1	33
		Low price with lag 1	31
		Volume of a single period	26
		Open price with lag 1	25
		% change in volume	7
		Volume of 5 periods	5
		Close price with lag 2	5
		Close price with lag 3	4
	Other technical indicators	Return	143
		Simple moving average	94

Category	Subcategory	Indicator	Number of articles
		Relative strength index	80
		Exponential moving average	54
		Stochastic %D	53
		Momentum	52
		Stochastic %K	49
		MACD	49
		Williams %R	35
		BIAS	28
		Bollinger bands	28
		Commodity channel index	24
		Price level	22
		Disparity	19
		Accumulation/distribution indicator	16
		Accumulation/distribution oscillator	12
		Price disagreement	12
		Price polarity	12
		Price oscillator	11
		Psychological line	11
		Directional movement	10
		Volume ratio	9
		On balance volume	8
		Triangular moving average	8
		Weighted moving average	7
		Accumulative ratio	5
		Volatility	4
		True Range	3
Macroeconomic variables	Exchange rates	USD / JPY	13
		USD / CNY	10
		USD / EUR	6
		USD / GBP	6
		USD / CAD	6
		USD / Multiple	4
	Commodities	% change in crude oil price	7
		% change in gold price	6
		Crude oil price with lag 1	3
	Economics	Consumer price index	7
		Industrial production	7
		Producer price index	4

Category	Subcategory	Indicator	Number of articles
		Unemployment rate	3
	Interest rates and money supply	3-month treasury bill	20
		10-year treasury note	15
		5-year treasury note	14
		6-month treasury bill	10
		1-month treasury bill	9
		1-year treasury bill	9
		30-year treasury bond	6
		Aaa corporate bond	6
		Baa corporate bond	6
		M2 money supply	6
		Term spread (10 years – 3 months)	6
Fundamental indicators	Stock information variables	Price/earnings ratio	10
		Earnings per share	5
		Market value of stock	5
		Price/sales ratio	5
		Book value of stock	4
		Dividend per share	4
	Balance sheet and profit & loss variables	Current ratio	7
		Debt to total assets ratio	7
		Return on assets	7
		Total assets turnover ratio	7
		Inventory turnover capital	6
		Fixed asset turnover ratio	5
		Interest coverage ratio	5
		Return on equity	5
		Accounts receivable turnover ratio	4
		Cash flow ratio	4
		Debt to equity ratio	4
		Current debt-to-equity ratio	3
		Current ratio	3
		Long-term debt-to-equity ratio	3
		Net income	3
		Percentage of gross profit to sales	3
		Percentage of net profit to sales	3

Category	Subcategory	Indicator	Number of articles
		Percentage of operating profit to sales	3

It is visible that five basic technical indicators, describing different previous period prices and volumes, were used more often than the rest of the technical indicators. From “other technical indicators”, the most diverse subgroup of all indicators, momentum indicators, measuring the speed with which price changes, were the most popular. It is followed by trend indicators describing the direction and strength of change, volatility indicators measuring fluctuations of specific variables and volume indicators focusing on the volume associated with changes [36]. From “macro indicators”, exchange rates are used, with the US dollar to other currencies being the most popular; commodities related to energy and precious metals; economic performance, primarily measuring national productivity; and interest rates and money supply, which includes various duration government debt obligations. From “fundamental indicators”, in the “stock information category”, the most popular group of variables was from the price ratio group, followed by earnings, market value and dividend indicators. From the “balance sheet and profit and loss statement” category, profitability ratio group, capitalisation ratio group and activity were the most popular. It is evident that there is no general agreement on which features should be used to predict the stock market price and dominating features tend to be the ones that are easier to extract or derive.

1.8. The Granularity of Data and Forecasting Horizon Overview

Usually, researchers can use 5-minute, daily, weekly or monthly level granularity of stock market data and typically include data from several months up to 10 years [36]. However, when it comes to forecasting horizon, the majority of literature between 2000 and 2019, which used machine learning techniques to forecast stock prices or movements, focused on daily level price or returns forecasting and 55 % of the researchers formulated research objective as regression type model, 44.3 % as classification type and the remaining 0.7 % as clustering [36, 40]. The research community's focus on daily forecasts demonstrates the necessity to investigate the potential of longer-term forecasts. Also, it presents the difficulty of comparing findings with those of other researchers due to their inexistence.

1.9. Models Used by Research Community

The methods utilised in stock market forecasting problems will be covered in this subsection.

1.9.1. Classical Econometric Methods

In the recent past, statistical time series models, including regression, exponential smoothing, ARIMA and Kalman filter were used in most stock market studies [41]. However, financial time series data generally is chaotic, noisy and nonlinear, influenced by various environmental factors thus, statistical techniques perform poorly while predicting stock market indices [41, 42]. It is especially noticeable when modelling with multiple explanatory variables because it introduces a multicollinearity problem causing correlations between explanatory variables and generating noise [43]. Furthermore, because econometric models often represent the average behaviour of the variables, whereas severe crises and steep recoveries are outlier occurrences, their performance is also known to deteriorate even more during crisis times and steep recoveries [43, 44]. Moreover, during times of crisis, the relationship between economic variables and output can be different, which brings a necessity to adjust model parameters. Some proposed time-varying parameter models still remain unsatisfactory as timing and estimating parameter changes require a large amount of information from similar occurrences [44]. Finally, sources of shocks or drives differ from crisis to crisis, which suggests a need for variables to change in the model over time, posing even further difficulties for forecasting [44]. Such challenges with econometric models made machine learning models more popular.

1.9.2. Machine Learning Methods

Machine learning methods are more capable of processing random, chaotic and non-linear data of stock market indices and are being used more as they can fit in-sample data very closely [42]. Machine learning, in particular, employs regularisation techniques that reduce uncertainty brought on by complex interactions between variables and minimises the influence of redundant information [45]. Some of the methods that are used by the research community are Random Forest, AdaBoost, SVM, KNN, Naive Bayes and other models [46]. However, with these machine learning models, when it comes to highly volatile financial time-series data, the probability of overfitting is great [47] and traditional machine learning models fail to capture the sequential dependencies of financial data [48]. One can pass historical information as new features, however, it is limiting and thus, neural networks-based methods are even more advanced and capable of solving limitations of other machine learning methods in stock market asset price forecasting [49].

1.9.3. Deep Learning Methods

Deep learning is a branch of machine learning methods based on neural networks, where multiple layers allow the modelling of complex patterns using simpler ones [50]. Due to their ability to capture sequential data, neural networks models gained favour in financial time series data modelling. Which models were employed by the research community in which areas are shown in the figure below from 2020 financial applications deep learning analysis [51]. It is evident from the figure below that dominating models are RNN, DMLP, LSTM and CNN. However, unlike LSTM, CNNs are more suited for classification problems and are used more for non-time varying or static data sets, and since financial data is time-varying, CNN may not be the best model for forecasting stock market price [51]. Nevertheless, the GRU model, which could be considered a similar simplified version of the LSTM model due to its capability of having short-term and long-term memory, may also be an interesting candidate for this research thesis despite its lower popularity.

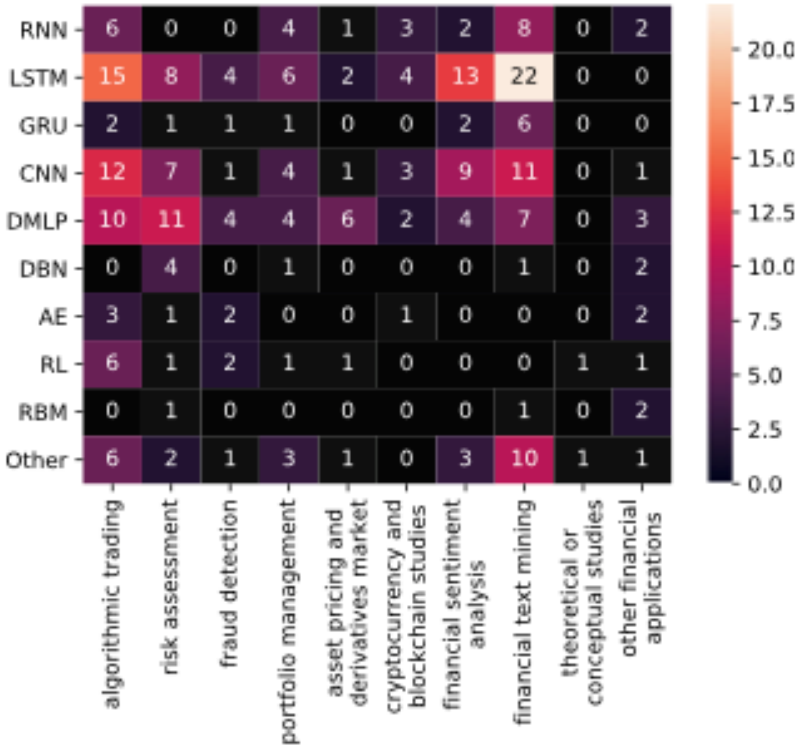

Fig. 1. Heatmap of used deep learning models in the financial area [51]

1.10. Hyperparameters Tuning of Neural Networks Models

Even though neural networks-based models are popular in financial time series forecasting problems, they can be difficult to tune. The following section will overview best practices from other scholars on how to achieve the most excellent stock market price prediction results.

1.10.1. Activation Functions and Optimisers

By introducing nonlinearity, activation functions enable converting input data that is not linearly separable into more separable features [52]. Some of the more popular activation functions are Sigmoid, Tanh, ReLU, ELU and Swish, however, there are multiple other versions of activation functions where Dubey et al. [52] evaluated 18 of them. Nevertheless, the majority of practitioners explore the most popular ones. For example, Orimoloye et al. [53] analysed and detected that the ReLU activation function performs better than Tanh when employing deep neural networks across all time horizons during stock market index price forecasting. Nevertheless, they state that optimal parameter optimisation may produce satisfactory results for both activation functions. Rana et al. [54] also compared LSTM stock market predictions using four activation functions and seven optimisers, which are responsible for updating the weights and biases in neural networks. Two models producing the best outcome were retrieved by implementing the linear activation function with the Adamax optimizer and the Tanh activation function with the Adam optimizer. However, the performance of Adam and Adamax optimisers was almost identical in all cases and much better than other optimisers. Moreover, differences in results generated by linear, ReLU, Tanh and Sigmoid activations functions were negligible.

1.10.2. The Architecture of Neural Networks

Orimoloye et al. [53] showed that data size affects the performance of deep neural networks during index forecasting. One layer of neural networks performed better when fewer data, around 1500 observations, were present. Deeper layers were significantly better with large datasets, in their case around 800000 observations [53]. Such a situation happens because the more hidden layers the neural network has, the more complex representations it can learn. However, for datasets with fewer observations, the deeper model is more prone to overfitting the train data and, consequently, is incapable of generalising on unseen data. Thus, choosing the number of hidden layers is one of the most essential choices.

The number of neurons inside a hidden layer is another choice that has less influence but is nonetheless significant in the architecture of a neural network [55]. The more neurons, the more complicated patterns a model will be able to learn, but at the same time will also start overfitting [55].

1.10.3. Training Parameters of Neural Networks

Some of the main training parameters are learning rate and batch size. The learning rate is responsible for the pace at which a neural network learns or updates its values. Geron [55] generally describes an optimal learning rate as equal to about half of the maximum learning rate at which the training algorithm diverges. To begin tuning the learning rate, it is recommended to start with a very low one, around 0.00001 and increase to a very large around 10. When it comes to batch size, according to Masters and Luschi [56], small batch sizes produce the best generalisation and training stability for a variety of tests, usually ranging between 2 and 32. When substantial batch sizes are used, the range of usable learning rates significantly decreases, sometimes to such an extent that optimal learning rate could not be used potentially due to linear increase in the variance of the weight updates [56]. However, there are other opinions that large batch sizes are better because they do not add regularisation and allow for using more significant learning rates [57].

1.10.4. Overfitting Prevention

Overfitting is also a major problem in neural networks-based methods. To combat overfitting, the dropout technique is used, which randomly selects neurons and excludes them from data training, which leads to breaking co-adaptations of various neural networks that work for the training set but do not generalise on the test set [58]. It significantly reduces overfitting, gives major improvements compared to other regularisation methods and applies in multiple domains. One drawback of dropout is increased training time, typically longer by 2-3 times because parameter updates become noisy and gradients computed are not of the final architecture [58]. Another technique used is early stopping, which tracks validation set error levels. When the validation error does not improve for some steps, the training algorithm terminates and returns to the best validation error model, which prevents overfitting [59].

1.11. Neural Networks-based Models' Explainability

An additional problem with neural networks-based models is that they are essentially a "black box", making it impossible to evaluate the results. Consequently, such models can be ignored or not entirely trusted by some. Fortunately, SHAP values, developed by Lundberg and Lee [60], allow decomposing final prediction into the contributions of each input feature based on game theory. The SHAP method does that by taking a few input variables into different combinations and computing the average change in the prediction with or without the presence of certain variables [60]. The influence can be measured both as positive and negative and it was found that in stock market index predictions importance of specific variables are consistent regardless of any changes done to the model [45]. SHAP value returns high precision results because variables are calculated by considering all possible combinations. Additionally, it was observed that there was a more substantial agreement between human explanations and SHAP compared to other methods, also allowing to interpret input effects [60].

1.12. Final Project Topic and Tasks Validations

After reviewing the current literature, it is evident that analysing stock market recovery after crashes is a welcome research area, as billions of dollars are traded daily and significantly impacts countries', corporations' and citizens' livelihoods [61]. Moreover, with the ability to forecast recovery duration after a crash, investors could plan and earn higher returns than the market, protect against external risks, share-risks, diversify better, increase their liquidity and lower information and transaction costs. At the same time, policy researchers could complement their economic recovery forecast models by including stock market data that responds to the economic situation [62]. Additionally, strong motivation and the need to research recoveries are due to the lack of existing research in this area from the stock market perspective. The majority of study objects are related to investigating the recovery of GDP or production output, which only moderately correlates to changes in stock market prices, and therefore, the stock market has its unique dynamics potentially worth exploring [26, 63]. Thus, this research thesis will employ neural networks-based models due to their ability to process sequential data, in particular MLP, RNN, LSTM and GRU, and various features, the majority of them being crash-related, to create a model that will be able at any time to predict the future price of the stock market index 72 months in advance and as a consequence allow easier to identify when recovery is expected.

2. Methodology of Research and Data Overview

This chapter will overview the MSCI World Index, which will be used as the primary research object, and examine its composition, performance between the end of 1969 and 2022 and long-term memory to determine if it does not follow a random walk and can be forecastable. Furthermore, it will present an in-depth description of the models, forecasting features, crash and recovery periods calculations, features preprocessing, models' training scheme, performance metrics and tools used.

2.1. Research Structure

This Master's thesis will commence by collecting historical data on the MSCI World Index, followed by validating its forecastability using the Hurst exponent. Upon verifying the predictability, the Performance Analytics package will be used to identify crash and recovery dates. Furthermore, other explanatory features will be collected from various sources or derived from existing data using simple formulas and the Markov regime-switching model. Subsequently, the features will be divided into training, validation and test samples, then allocated into data windows of 144 months, which further will be assigned into batches. Later, features will be scaled between 0 and 1. Afterwards, MLP, RNN, LSTM and GRU models will be trained using different architectures and hyperparameters. Models' performance will be examined in two time periods: one solely during the Great Recession and the other lasting between Great Recession and December 2022. Finally, SHAP will be calculated for the best deep learning model. The schematic representation of the methodology is provided below.

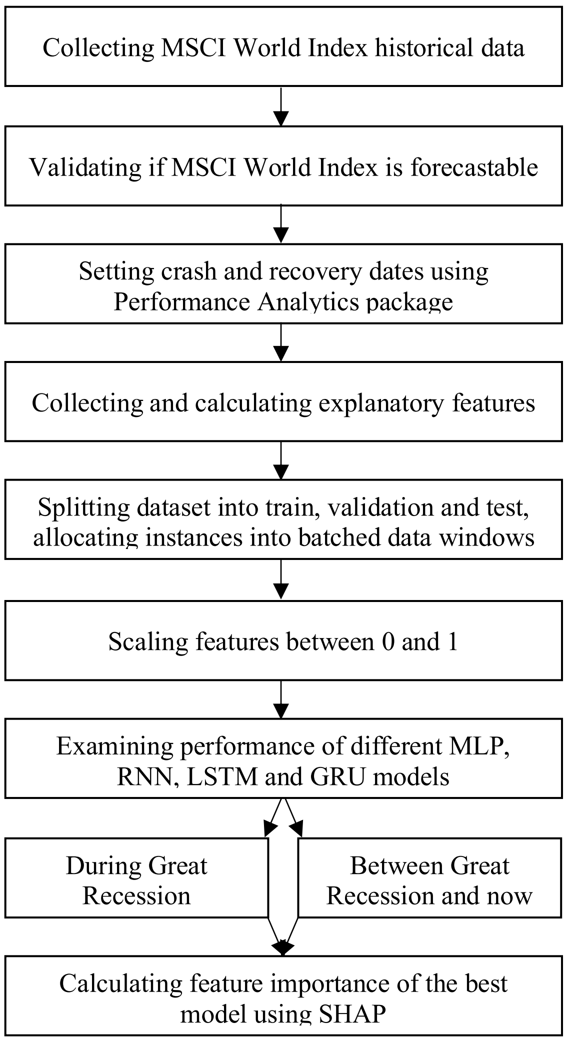

Fig. 2. Research Structure

2.2. MSCI World Index Overview

The main object of this research thesis will be the MSCI World Index, which captures large and mid-cap companies across 23 developed markets [64]. In total, 1509 entities are included in the index, which equals approximately 85 % of each country's free float-adjusted market capitalisation [64]. The stock index used in this thesis ranges between December 1969 and December 2022, has monthly values and is denominated in USD. However, the index itself was launched on 31 March 1986, but the availability of earlier values is possible due to back-testing done by MSCI [64].

An in-depth overview of the index structure shows that developed market countries include: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Hong Kong, Ireland, Israel, Italy, Japan, Netherlands, New Zealand, Norway, Portugal, Singapore, Spain, Sweden, Switzerland, the UK and the US [64]. As of 28 February 2023, US companies made 67.67 % of the total MSCI World Index market capitalisation, followed by Japan at 6.13 % and the United Kingdom at 4.43 % [64]. Additionally, the index is dominated by the information technology sector, having 21.25% of total weight, financials at 14.61 % of weight and health care at 13.31 % [64].

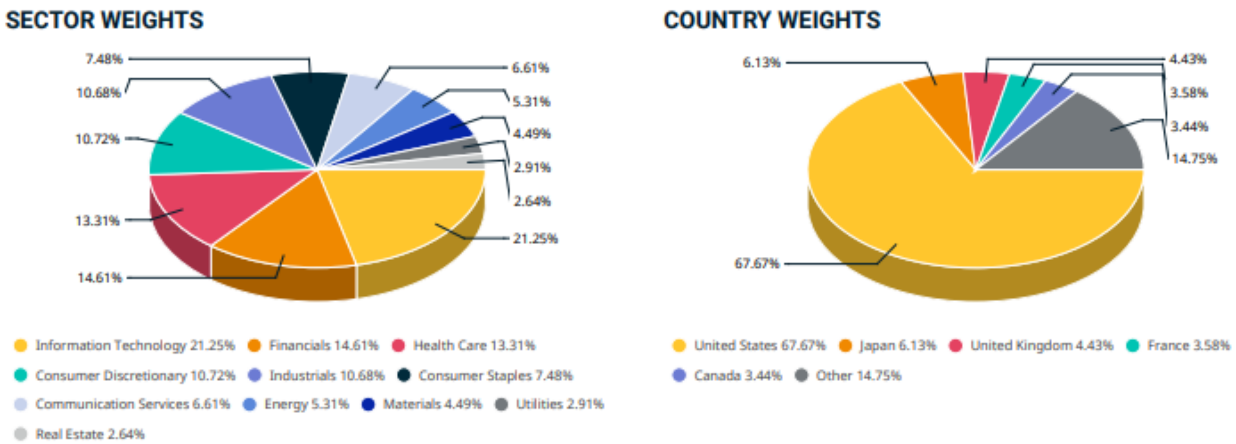

Fig. 3. Sector and country weights of MSCI World Index as of 28 February 2023 [64]

MSCI World Index is an attractive research object to investigate the dynamics of crashes and recovery periods because:

1. A respectable organisation developed the index and uses a reliable methodology;
2. The index covers global markets; thus, it exhibits international patterns in stock markets;
3. The index covers the majority of capitalisation in selected developed countries;
4. Index has a sufficient history of existence, allowing to witness a few stock market crashes and, consequently, enables to model more accurate long-term trends during turbulent times.

2.3. MSCI World Index Descriptive Statistics

There were 637 months in the MSCI World Index between 1969 December 1 and 2022 December 31. The prices observed ranged from 74.45 to 3231.73, with an average of 891.17.

Table 2. MSCI World Index descriptive statistics

Count	Mean	Std	Min	25%	50%	75%	Max
637	891.170	741.329	74.454	166.139	751.451	1334.930	3231.730

When observing the dynamics in monthly price change of the MSCI World index, several vital insights are noticeable (Fig. 4.). First of all, it is common to observe short-term fluctuations in the index price, which can be caused by a variety of factors as analysed in the first part of the thesis. However, these fluctuations are smoother compared to the dynamics of daily data and no multiple extreme values are present. Secondly, the trend of the MSCI World Index over a long period is positive, indicating that an increase in stock market index price during bullish periods has a more substantial effect than bearish periods. Thirdly, it is apparent that the index had a few periods of price crashes, which exhibited a longer-term persistent downward behaviour. Finally, when viewing through a perspective of logarithmic returns, there were more extreme negative returns falling below -0.1 compared to returns increasing above 0.1, in numerical terms, 15 to 6 months. It demonstrates that it was more common to experience more significant monthly decreases in returns compared to the same level increases (Fig. 5.).

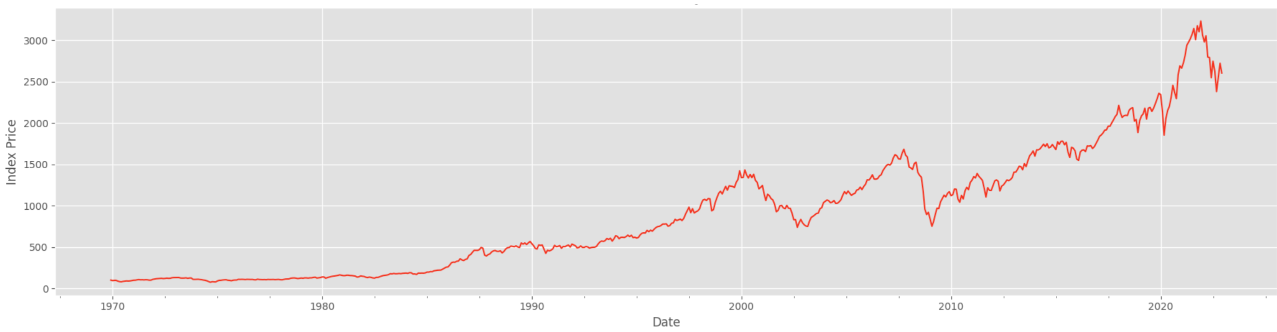

Fig. 4. MSCI World Index monthly price in USD

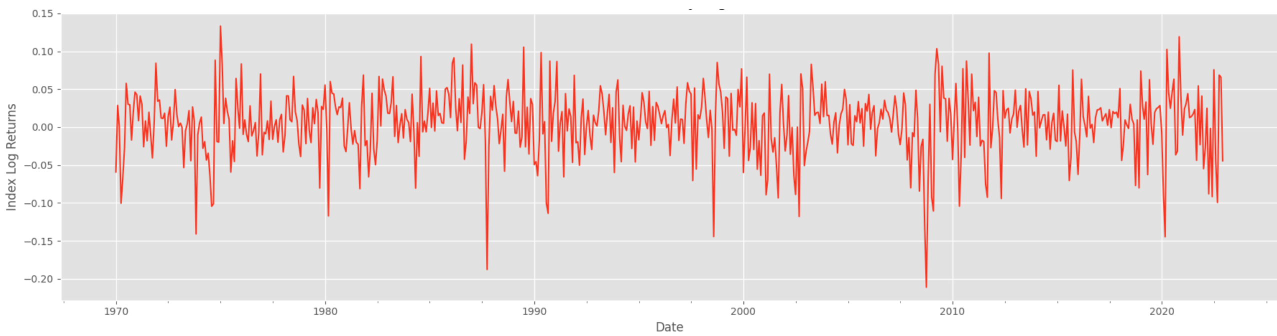

Fig. 5. MSCI World Index monthly logarithmic returns

Monthly and quarterly seasonality can typically reveal essential patterns of the index price dynamics that recur simultaneously each year. For that, a seasonal plot can be used, which displays monthly prices for all years on the same axis and a monthly plot, which plots separate lines for each month's prices and depicts the average price during each month with a red line. Given that price increases and decreases fluctuate from year to year without a pattern, the seasonal plot reveals no discernible monthly seasonality impacts (Fig. 6.). Fig. 7. verifies the findings from the previous plot and shows that the average monthly price differences are negligible. As there is no strong monthly seasonality, quarterly seasonality is also not visible, as shown in Fig. 8.

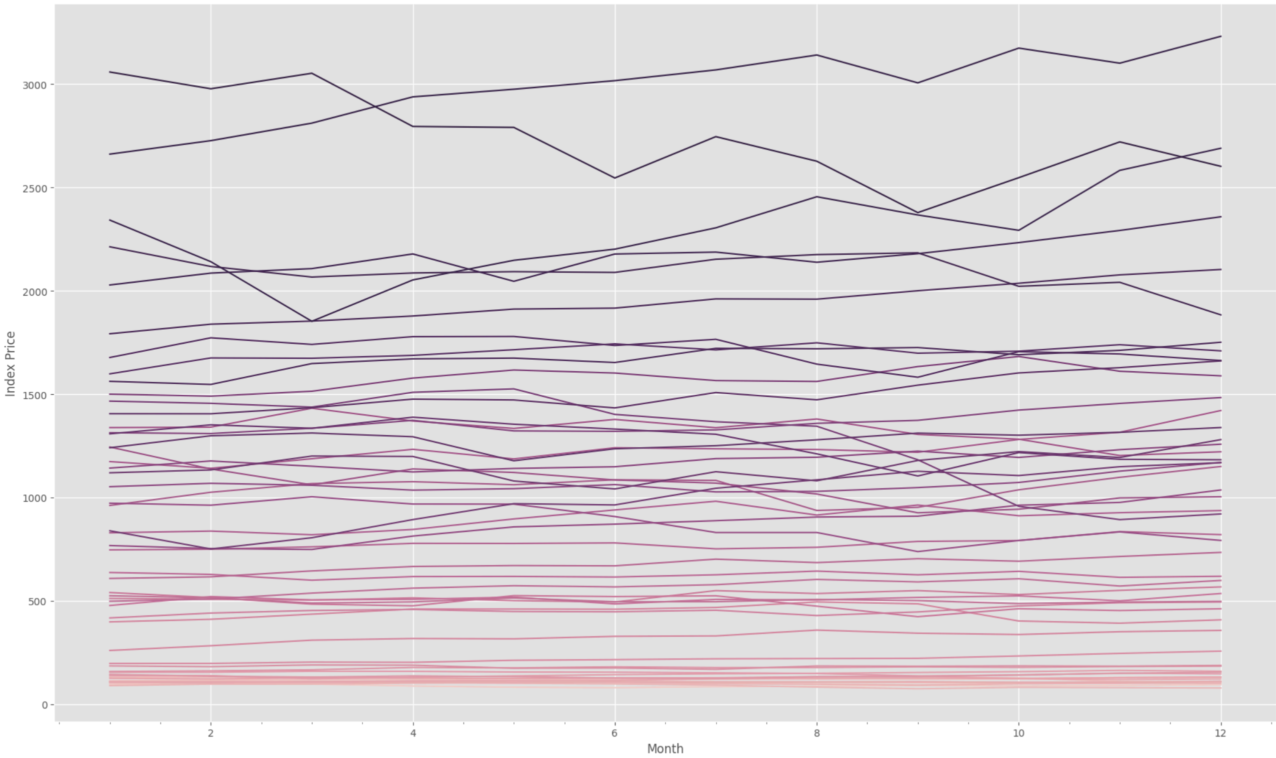

Fig. 6. MSCI World Index monthly price in USD – seasonal plot

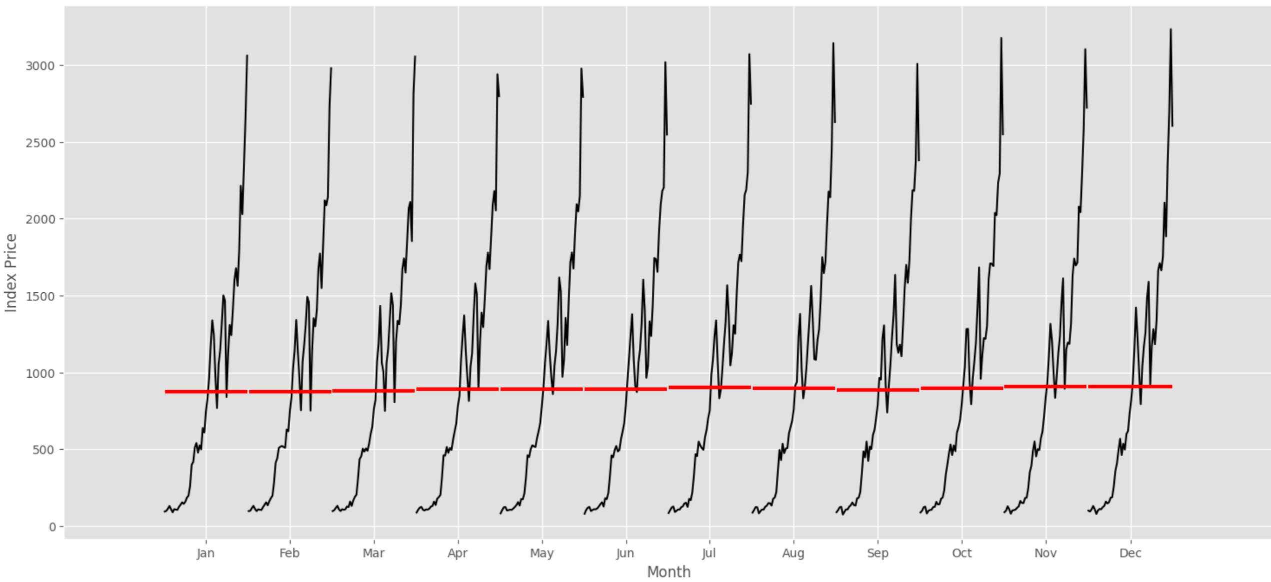

Fig. 7. MSCI World Index monthly price in USD – monthly plot

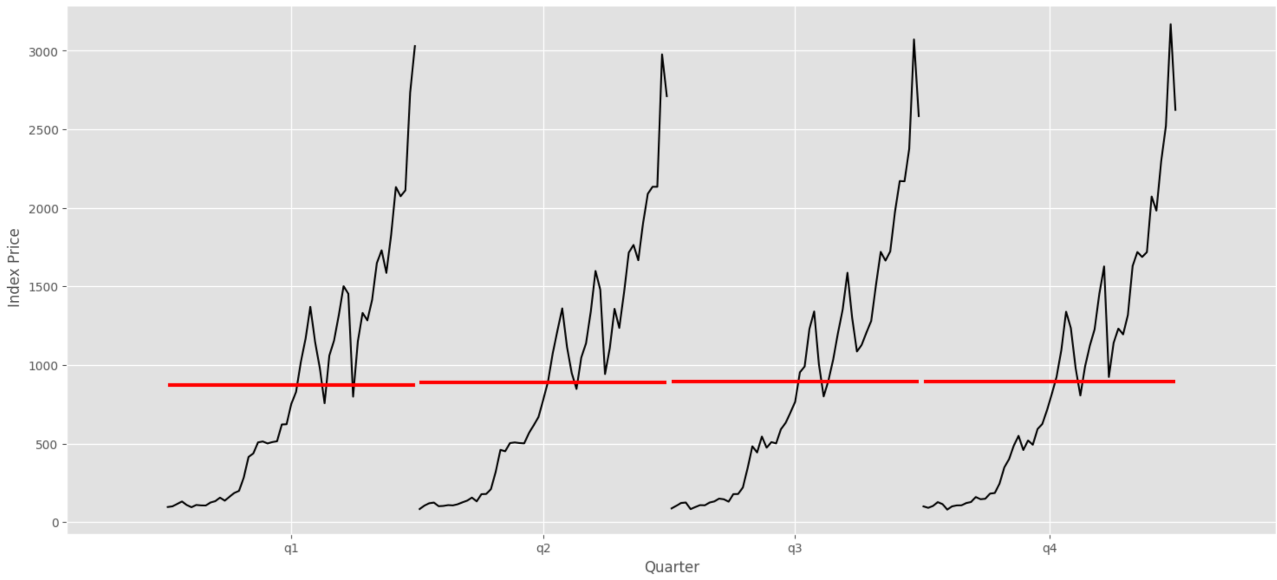

Fig. 8. MSCI World Index average monthly price in USD – Quarterly plot

2.4. MSCI World Index Long-term Memory Test

Before commencing predictions of recovery time, it is essential to understand if the analysed MSCI World index time series does not follow a random walk, which in case of random walk behaviour, time series would be announced as unpredictable. For that, the Hurst exponent will be used to measure the long-term memory of a time series, i.e. how drastically the time series deviates from random walk [65]. Hurst exponent strength has a simple interpretation where values typically range between 0 and 1. The equality of Hurst exponent to 0.5 means a geometric random walk. Values larger than 0.5 show a trending series. The closer to 1 Hurst exponent is, the stronger trend is present, meaning that high values will be followed by the same direction high values and low values by low values. Hurst exponent values below 0.5 exhibit mean reversion behaviour. The closer to 0, the more considerable strength of the mean-reversion effect. Mean reversion in practice means that high values will be followed by lower values and vice-versa [65].

There are multiple ways of measuring the Hurst exponent. However, this thesis will focus on the Hurst exponent established on estimating the rate of diffusive behaviour based on the variance of log prices. The calculations of the Hurst exponent will be calculated as follows:

1. For each lag in the considered range, a standard deviation of the differenced series will be calculated;
2. Then the slope of the log plot of lags versus the standard deviations will be calculated to retrieve the Hurst exponent.

After calculating the Hurst exponent for different lags, results indicate that for minimal lags, between 4 and 5, the time series exhibited slight mean reversion trends at 0.44 and 0.47. Between 6 and 7 lags, the time series revealed almost a random walk tendencies at 0.49 and 0.51. From 8 lag to 15 lag time, the series displayed slight trending tendencies and Hurst exponent values ranged between 0.52 and 0.55. From 20 lags onwards, the time series start to return to mean reversion trends, where the longer the period analysed, the more robust mean reversion trends are present. This test indicates that some periods of time series exhibit random walk patterns, but slight tendencies of trend and mean reversion in different periods can be seen, which proves that some parts of time series can be forecasted, especially when it involves long periods [66].

Cunado and Gracia did a similar long-term memory examination of the S&P500 index, which displays similar volatility to the MSCI World Index for 1929–2006 but tried to see the dynamics separately in bull and bear markets. They found four episodes of mean reversion during bull phases occurring within the latest years of the sample, between 1987 and 2006 [66]. Additionally, Kim et al. [67], while analysing the Dow Jones Industrial Average index between January 1900 and June 2009, detected strong evidence of returns predictability during economic or political crises with a moderate degree of uncertainty, while during bubbles, less predictability was noticeable.

Thus, acknowledging the limitations of the MSCI World Index that some shorter periods exhibit random walk patterns, this thesis will continue trying to forecast MSCI World Index's longer-term dynamics, as they display stronger long-term memory.

Table 3. Hurst exponent values for different lags of the MSCI World Index

Lag	**Hurst exponent value**	**Lag**	**Hurst exponent value**
4	0.4418	45	0.4432
5	0.4712	50	0.4325
6	0.4996	55	0.4255
7	0.5198	60	0.4208
8	0.5275	65	0.4159
9	0.5368	70	0.4085
10	0.5459	75	0.3983
15	0.5508	80	0.3863
20	0.5316	90	0.3629
25	0.5123	100	0.3508
30	0.4940	150	0.3431
35	0.4758	200	0.2905
40	0.4583	250	0.2557

2.5. Description of Models Used in Forecasting

As discussed in the literature review, neural networks are promising models capable of modelling complex sequential financial time series data. Therefore, this thesis, during modelling, will use neural networks-based models, which are described below.

2.5.1. MLP

A multilayer perceptron is a fully connected feedforward neural network consisting of one input layer, one or more hidden layers and one final layer called the output layer [68].

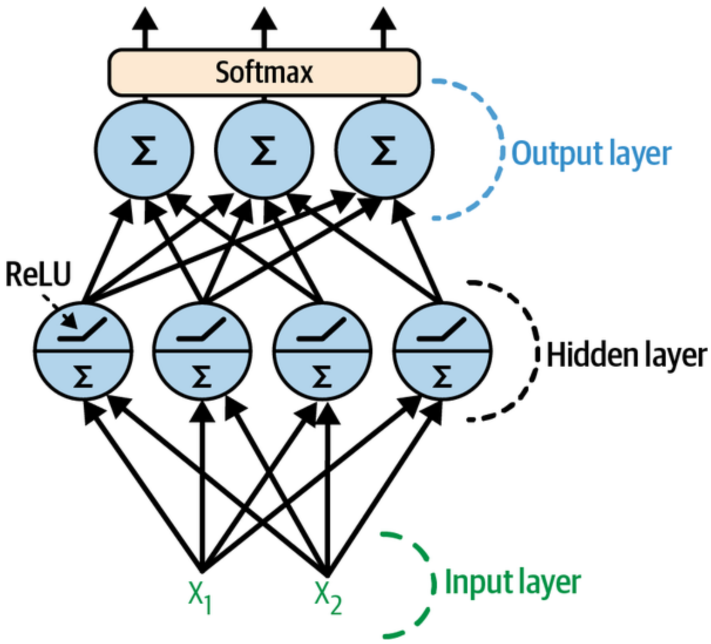

Fig. 9. A modern MLP structure [55]

The input is processed in one way in feed-forward neural networks and the output can be calculated using the formula below [68]:

$$O_t = f_1\big(W_y f_2(W_x x_t + b_h\big) + b_y) \quad (1)$$

Where O_t – output of multilayer perceptron;

f_1 – output activation function;

W_y – weight matrix of outputs;

f_2 – hidden layer activation function;

W_x – weight matrix inputs;

x_t – input;

b_h – bias vector;

b_y – bias vector.

2.5.2. RNN

RNNs are a type of neural network capable of processing sequential data, i.e. input at each time step depends on the previous inputs [69]. There are different types of RNN structures: one-to-many, many-to-one and many-to-many [51]. Unfolded traditional RNN cell is displayed below.

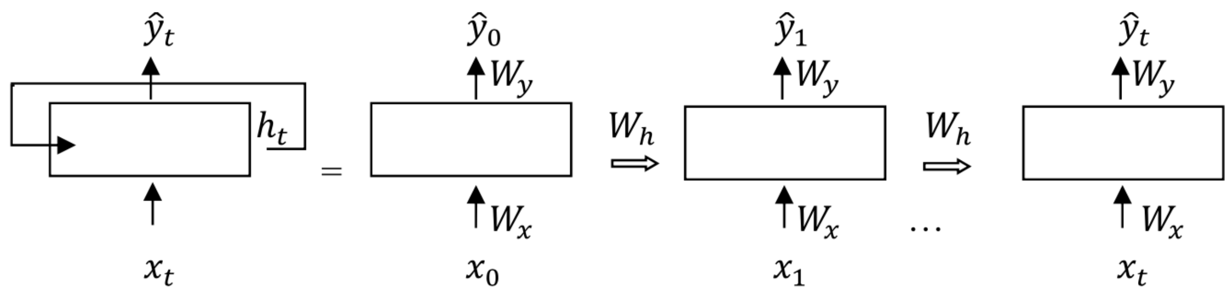

Fig. 10. RNN unfolded cell as displayed by Asrav and Erdal [59]

Recurrent neural networks repeatedly alter the cell states across time series using the following equation [68]:

$$h_t = f_2(W_x x_t + W_h h_{t-1} + b_h) \qquad (2)$$

Where h_t – output of recurrent neural network cell state transformation;

f_2 – hidden layer activation function;

W_x – weight matrix of inputs;

x_t – input;

W_h – hidden unit weight matrix;

h_{t-1} – previous timestep hidden unit output;

b_h – bias vector.

And therefore, the recurrent neural network can be then expressed as follows:

$$O_t = f_1\big(W_y f_2(W_x x_t + W_h h_{t-1} + b_h\big) + b_y) \qquad (3)$$

Where O_t – output of recurrent neural network;

f_1 – output activation function;

f_2 – hidden layer activation function;

W_x – weight matrix of inputs;

x_t – input;

W_h – hidden unit weight matrix;

h_{t-1} – previous timestep hidden unit output;

b_h – bias vector;

b_y – bias vector.

The vanishing gradient problem is an issue with RNNs that prevents them from learning long-term dependencies in the input data. However, they still proved superior performance in financial and economic time-series data analysis because, unlike multilayer perceptron, RNNs use their internal state to retrain the sequence of inputs and acknowledge previous steps during forecasting [45].

2.5.3. LSTM

LSTM is a recurrent gradient-based neural network that addresses the vanishing gradient problem present in standard RNN models and enables learning long-term dependencies in the input data [70]. It consists of an input layer, several hidden layers and an output layer, with memory cells also contained in the hidden layer [60]. The memory cell has three gates that maintain its state: forget gate, input gate and output gate, where forget gate is responsible for specific information removal, the input gate specifies information to add and the output gate specifies what information to output [60]. The structure of LSTM is illustrated in Fig 11. The main advantage of LSTMs is that they can handle long-term dependencies, noise, distributed representations, continuous values, do not require a priori choice of a finite number of states, generalise well, work great over a broad range of parameters and thus parameter fine-tuning importance decreases [70]. However, The LSTMs require each memory cell block to have two additional input and output gate units. Thus, it increases the number of weights [70].

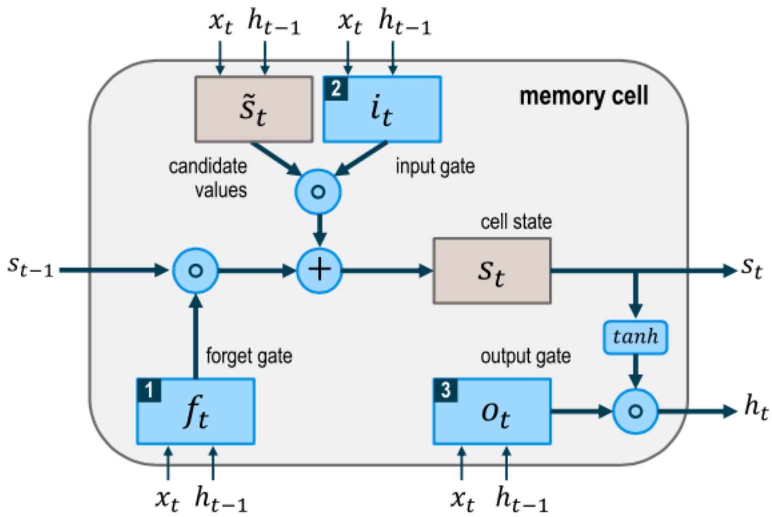

Fig. 11. Structure of an LSTM memory cell displayed by Fischer and Krauss [71]

The LSTM execution steps are the following [45]:

First of all, activation values f_t of the forget gate determines what information should be removed from the previous state by calculation:

$$f_t = sigmoid(W_{f,x}x_t + W_{f,h}h_{t-1} + b_f) \tag{4}$$

Where f_t – forget gate output;

$W_{f,x}$ – weight matrix of inputs;

x_t – input;

$W_{f,h}$ – hidden unit weight matrix;

h_{t-1} – previous timestep hidden unit output;

b_f – bias vector.

As one can see, f_t is computed by summing the current input x_t at time t, the output h_{t-1} of the hidden cell state at the previous time t-1 and the bias vector b_f. The sigmoid function scales the value into the range between 0 and 1, and the bias vector boosts the model's adaptability to suit the data. The possible values of f_t are 0, which is to entirely forget the information, and 1, which is to recall the information completely.

The second step determines how much the current time-series data in the new cell state must be updated. Two steps are required to do this: Firstly, the hyperbolic tangent function is used to calculate candidate values $\tilde{S}_t$ that might be present in the new cell state $\tilde{S}_t$. Secondly, the input gate's activation values i_t which candidate values should be added to the cell state S_t are calculated.

$$\tilde{S}_t = \tanh\left(W_{\tilde{S}_t,x}x_t + W_{\tilde{S}_t,h}h_{t-1} + b_{\tilde{S}_t}\right) \quad (5)$$

Where $\tilde{S}_t$ – candidate value;

$W_{\tilde{S}_t,x}$ – weight matrix of inputs;

x_t – input;

$W_{\tilde{S}_t,h}$ – hidden unit weight matrix;

h_{t-1} – previous timestep hidden unit output;

$b_{\tilde{S}_t}$ – bias vector.

$$i_t = sigmoid(W_{i,x}x_t + W_{i,h}h_{t-1} + b_i) \quad (6)$$

Where i_t – input gate value;

$W_{i,x}$ – weight matrix of inputs;

x_t – input;

$W_{i,h}$ – hidden unit weight matrix;

h_{t-1} – previous timestep hidden unit output;

b_i – bias vector.

In the input gate function, weight matrices W and bias vector b are the same as in the first equation f_t.

In the third step, the new cell states S_t are derived using the previous cell state and the current candidate value, where multiplication between f_t and S_{t-1} determines previous information amount to be forgotten and $i_t\tilde{S}_t$ indicates how much current information should be remembered with a formula:

$$S_t = f_t S_{t-1} + i_t\tilde{S}_t \tag{7}$$

Where S_t – new cell state;

f_t – forget gate output;

S_{t-1} – previous cell state;

i_t – input gate output;

$\tilde{S}_t$ – candidate cell state.

Lastly, the output h_t is controlled by the activation values O_t of the output gate using formulas:

$$O_t = sigmoid(W_{O,x}x_t + W_{O,h}h_{t-1} + b_O) \tag{8}$$

Where O_t – output gate;

$W_{o,x}$ – weight matrix of inputs;

x_t – input;

$W_{o,h}$ – hidden unit weight matrix;

h_{t-1} – previous timestep hidden unit output;

o – bias vector.

$$h_t = O_t \tanh(S_t) \tag{9}$$

Where h_t – current timestep hidden unit output;

O_t – output gate;

S_t – new cell state.

2.5.4. GRU

GRU is also a type of RNN that addresses the vanishing gradient problem [72]. GRU network uses a hidden unit consisting of two gates to control information flow through a network called reset and update gates [72]. The reset gate is responsible for choosing the amount of information from the previous state to be forgotten and the update gate determines how much new state should be retained [72]. It could be considered a simplified version of LSTM and trains faster by reducing the number of parameters [45]. However, the more sophisticated memory cell of LSTM can remember longer sequences [49].

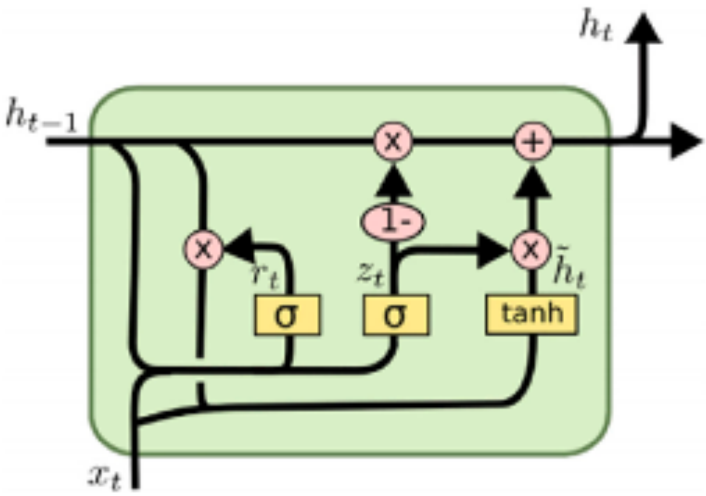

Fig. 12. Structure of a GRU unit displayed by Cho et al. [49, 72]

The GRU calculation logic is provided below [49]:

The weights W are learned during the training phase while h_t represents a memory unit that stores the information related to the reset gate.

The reset gate can be calculated with the following equation:

$$r_t = sigmoid(W_r x_t \ + W_{r,h} h_{t-1}) \tag{10}$$

Where r_t – reset gate output;

$W_{r,x}$ – weight matrix of inputs;

x_t – input;

$W_{r,h}$ – hidden unit weight matrix;

h_{t-1} – previous timestep hidden unit output.

The update gate can be calculated with the following equation:

$$z_t = sigmoid(W_z x_t \ + W_{z,h} h_{t-1}) \tag{11}$$

Where z_t – update gate output;

$W_{z,x}$ – weight matrix of inputs;

x_t – input;

$W_{z,h}$ – hidden unit weight matrix;

h_{t-1} – previous timestep hidden unit output.

The hidden layer output h_t is calculated as follows, where $\otimes$ is the product function of two vectors:

$$h'_t = tanh(W_c x_t \ + r_t(W_h h_{t-1})) \tag{12}$$

Where h'_t– intermediate hidden layer output before update gate application;

$W_{c,x}$ – weight matrix of inputs;

x_t – input;

r_t – reset gate output;

W_h – hidden unit weight matrix;

h_{t-1} – previous timestep hidden unit output.

$$h_t = (z_t) \otimes (h_{t-1} + (1 - z_t) \otimes h'_t) \quad (13)$$

Where h_t – hidden layer output;

z_t – update gate output;

h_{t-1} – previous timestep hidden unit output;

h'_t – intermediate hidden layer output before update gate application.

2.6. Identifying Low and High Regimes in MSCI World Index

It is accepted that the stock market experiences different regimes, called bullish and bearish periods. Bullish periods experience only mild price rises and fluctuations and bearish experience price decreases and volatility increases [73]. To identify statistically in timely manner periods where the fundamental environment of financial markets has changed, Markov regime-switching models are used. Markov regime-switching model is an autoregressive model where the mean of the process switches between a selected number of regimes and the probability of moving from one regime to another is governed by the Markov process [74]. Ang and Timmermann [75] identified that Markov regime-switching models are capable of capturing stylised facts of financial time series like fat tails, heteroskedasticity, skewness and autocorrelations. Haase et al. [73] used two regimes where regime zero corresponded to bull markets and regime one identified bear markets. The two regimes model was motivated by the possibility of differentiating between a volatile regime with a downward trend and a calmer regime with mostly positive returns. Also, more than two regimes can lead to unstable estimations in financial market time series [73]. Additionally, Haase's et al. [73] Markov regime-switching model was with a univariate setting to reduce identification problems and estimation uncertainty. Zens et al. [76] confirm that a small number of variables ensure a more stable estimation process with higher precision. The created model identified turning points from bear to bull markets by not exceeding four weeks delay [73].

As the stock market crisis phenomena consist of two stages: a recessionary period, described by a decrease in prices and a recovery period, characterised by the expansion of prices, this thesis will use the Markov regime-switching model to differentiate between volatile regimes with the downward trend and calmer regimes with mostly positive returns. Implementation will be done with the Statsmodels Python package [77]. Two regimes were chosen and no exogenous variables were added to the model, as more than two regimes and multiple variables can lead to unstable estimations. Furthermore, due to the stationarity of logarithmic returns, no trend will be included, the order will be selected as one, autoregressive coefficients will be allowed to be switched across regimes and the error term will be able to have a switching variance. Finally, the Markov regime-switching model results will be saved as a feature depicting the probability of low regime to assist neural networks in forecasting the MSCI World Index.

After applying the Markov regime-switching model to the MSCI World index, it is clear that it successfully distinguished between bullish and bearish patterns with minor deviations (Fig. 13). Some of these deviations can be observed during the Dot-com and Great Recession periods, where bullish regimes were announced slightly later than they actually happened or in some cases the model caught small changes in returns as regime change. Nevertheless, the probabilities may provide valuable insights into the potential index behaviour over time for deep learning models.

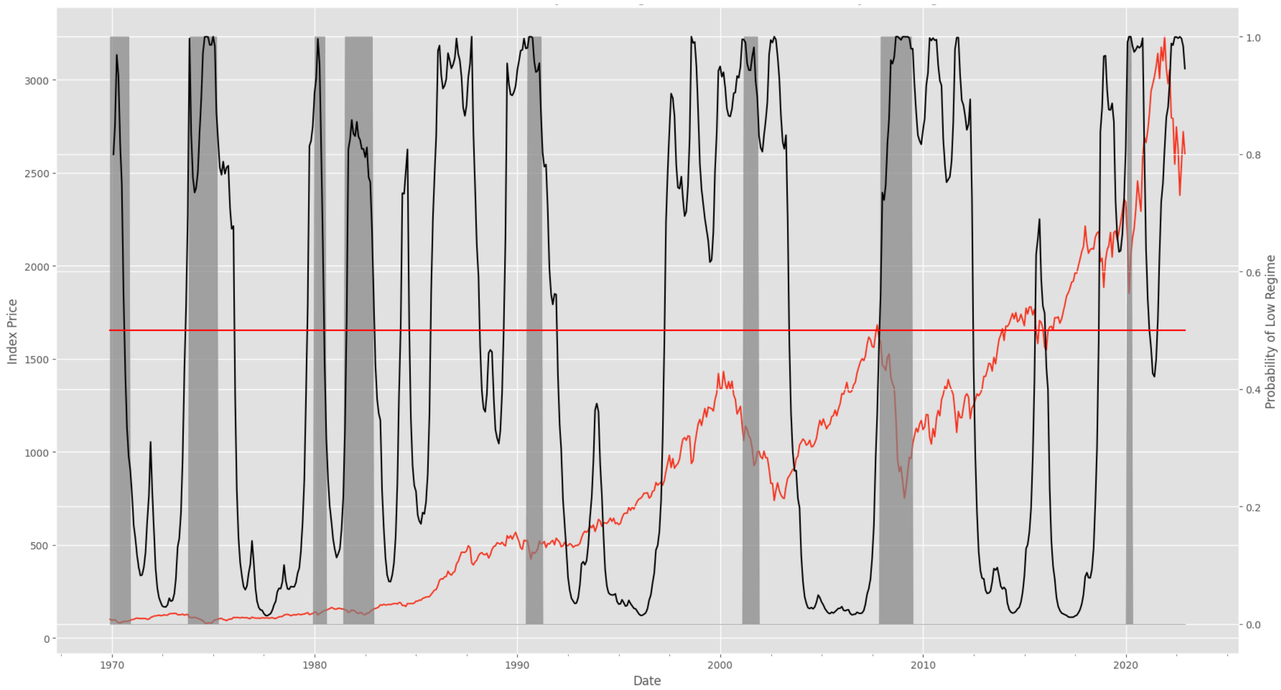

Fig. 13. MSCI World Index probability of low regime next to NBER business cycle dating in a shaded area

2.7. Identifying Recession and Recovery Periods in MSCI World Index

For this research thesis, as the study object will be the prices of a stock market index, the approach to defining the recovery through the perspective of the affected individuals, as suggested by Stiglitz, is difficult. Also, measuring recovery through price return to trend could be biased and include noise, as it is challenging to estimate whether the efficient level increased or decreased during a price crash. Thus, the stock market index will be declared as recovered, similarly to Reinhart's and Rogoff's definition: if it has returned to the peak pre-crash price levels. The potential growth aspect of the market will be ignored. However, partly it could be compensated by the initial higher peak value of the indices as the stocks making the indices can be overvalued before the crash [78].

As the objective is to identify the end of recovery periods after a stock market crash, the Performance Analytics package on R will be used to acquire an ordered list of crises by depth using simple/arithmetic chaining on logarithmic returns [79]. Each crisis period will have a starting date, trough date and recovery date. Performance Analytics identifies these periods by finding regimes with negative changes in returns, where the lowest point of drop in each regime interval is called trough and recovery is considered when price levels reach the same pre-crash peak level.

In general, when considering all crashes between December 1969 and December 2022, the average drawdown length was 11.7 months, average recovery took 7.7 months. It shows that, on average, recoveries take the more significant percentage of total drawdown time. When investigating the top 10 crashes, all except the one between December 1980 and March 1983 had longer recovery periods than decline ones.

In terms of recovery times, 24 of the 30 major accidents recovered within a year of the lowest crash point, one took less than two years, one took under three years and three crashes required more than four years (Fig. 14.). Unfortunately, one crash that started on January 2022 still have not recovered.

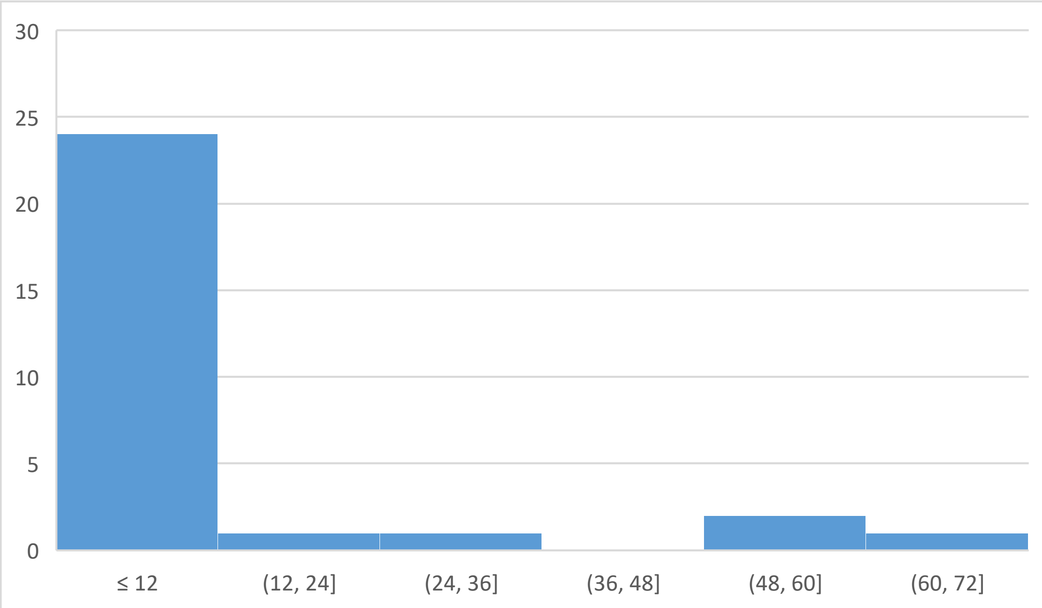

Fig. 14. MSCI World Index recovery duration from the lowest price point

After ranking the top 30 downturn periods, it was determined only to use the top 10 for the subsequent modelling because the other drawdown periods were either too short for this research scope, with most falling between 2 and 8 months, or not deep enough. These crashes will be transformed into a feature with a value of one if the relevant month has a crash and still has not recovered and a value of zero if there is no crash and the MSCI World Index is fully recovered.

Table 4. Top 30 drawdown periods in MSCI World Index

No.	Crash Start Date	Crash Lowest Price Date	Recovery Date	Depth	Total Length in months	Downturn Duration in months	Recovery Duration in months
1	1973-03-01	1974-09-01	1979-08-01	-0.4466	78	19	59
2	1970-01-01	1970-06-01	1971-03-01	-0.2252	15	6	9
3	2007-11-01	2009-02-01	2014-04-01	-0.2110	78	16	62
4	1980-12-01	1982-07-01	1983-03-01	-0.1827	28	20	8
5	2000-04-01	2002-09-01	2006-11-01	-0.1810	80	30	50
6	1990-01-01	1990-09-01	1993-05-01	-0.1072	41	9	32
7	1980-02-01	1980-03-01	1980-06-01	-0.0902	5	2	3
8	1987-09-01	1987-11-01	1989-01-01	-0.0898	17	3	14
9	1984-04-01	1984-07-01	1985-01-01	-0.0728	10	4	6

No.	Crash Start Date	Crash Lowest Price Date	Recovery Date	Depth	Total Length in months	Downturn Duration in months	Recovery Duration in months
10	2022-01-01	2022-09-01	Unknown	-0.0685	13	9	Unknown
11	1971-05-01	1971-10-01	1971-12-01	-0.0643	8	6	2
12	1979-10-01	1979-10-01	1980-01-01	-0.0618	4	1	3
13	2020-01-01	2020-03-01	2020-08-01	-0.0580	8	3	5
14	1998-07-01	1998-08-01	1998-11-01	-0.0435	5	2	3
15	2018-02-01	2018-12-01	2019-10-01	-0.0393	21	11	10
16	2015-06-01	2016-02-01	2017-01-01	-0.0360	20	9	11
17	1986-09-01	1986-10-01	1987-01-01	-0.0271	5	2	3
18	1997-08-01	1997-10-01	1998-02-01	-0.0226	7	3	4
19	1993-11-01	1993-11-01	1994-01-01	-0.0214	3	1	2
20	1994-02-01	1994-03-01	1994-08-01	-0.0211	7	2	5
21	1972-06-01	1972-06-01	1972-08-01	-0.0211	3	1	2
22	1994-09-01	1995-01-01	1995-03-01	-0.0198	7	5	2
23	2000-01-01	2000-01-01	2000-03-01	-0.0164	3	1	2
24	2020-09-01	2020-10-01	2020-11-01	-0.0163	3	2	1
25	1989-05-01	1989-06-01	1989-07-01	-0.0150	3	2	1
26	1972-09-01	1972-09-01	1972-11-01	-0.0140	3	1	2
27	1989-10-01	1989-10-01	1989-11-01	-0.0131	2	1	1
28	1983-07-01	1983-07-01	1983-09-01	-0.0129	3	1	2
29	1996-07-01	1996-07-01	1996-09-01	-0.0123	3	1	2
30	1984-02-01	1984-02-01	1984-03-01	-0.0118	2	1	1

The top 10 identified crises include four financial turbulence events and six more severe stock market crashes: The oil crisis bear market, Black Monday, the early 1990s recession, the Dot-com bubble, the Great Recession and the Covid-19 pandemic with the Russian invasion of Ukraine bear markets.

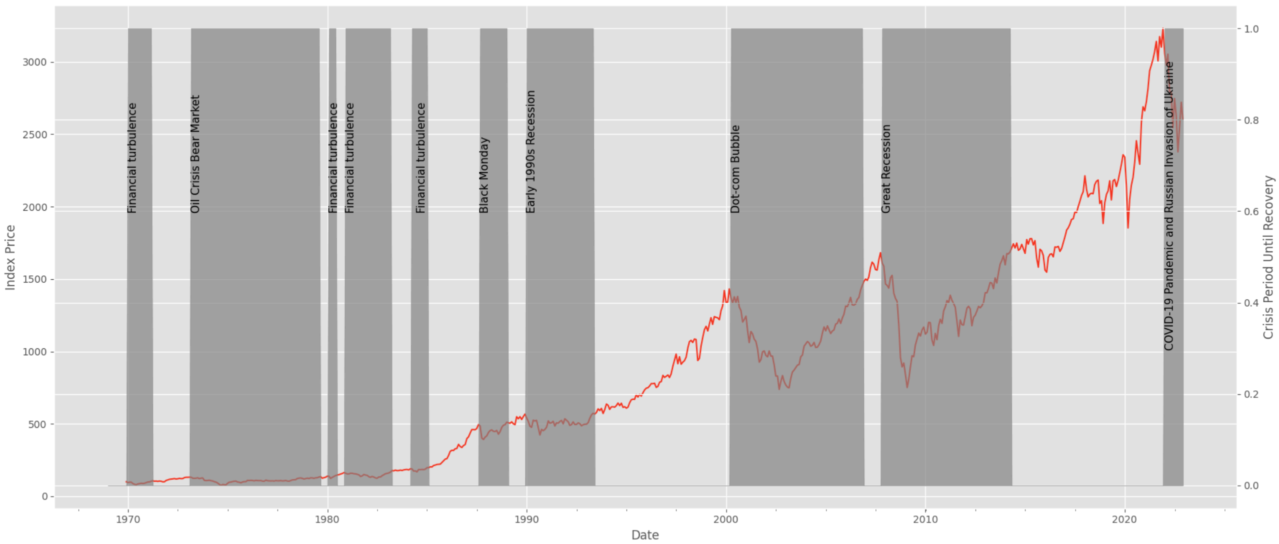

Fig. 15. MSCI World Index crash periods timing until the end of recovery

2.8. Overview of Features Used in Forecasting

In total, 19 features will be used in modelling. As machine learning algorithms minimise the influence of redundant information, more features were preferred to be explored [45]. Six features will be based on the most popular among research community technical indicators, adapted for longer-term forecasting. Two indicators will follow changes in the percentage prices of other assets: gold and oil. Two indicators will track the rate of 3-month US treasury bills and 10-year US treasury notes as some researchers found market linkage between different asset classes during the crisis period [80]. One indicator will measure the consumer price index of the US market. Due to the US being the principal country of the index, accounting for almost 68 % of the total index capitalisation, and the better accessibility of historical data, Treasury assets and the consumer price index of the US market were chosen as viable features. Two will depict historical index prices and logarithmic returns. One feature will identify the probability of a low regime based on the Markov regime-switching model. One feature will determine whether it is a crisis period, as identified by the Performance Analytics package. The remaining four features will derive from the Performance Analytics crisis flag feature: cumulative months in a crisis period, cumulative months in a non-crisis period, percentage of index price remaining from the pre-crisis peak and percentage of index price above the last crisis price.

Table 5. Features used in forecasting

No.	**Feature Name**	**Feature Description**
1	Index price	MSCI World Index monthly price. The feature will be forecasted into the future.
2	Crisis flag PA	Flag generated by Performance Analytics, which can acquire zero or one values and identifies if MSCI World Index is in a crisis period and has not recovered yet.
3	Cumulative months of in crisis period	The cumulative number of months current crisis period is present in the MSCI World Index. In the case of a non-crisis period, the value acquired is zero.
4	Cumulative months of in non-crisis period	The cumulative number of months current non-crisis period is present in the MSCI World Index. In case of a crisis period, the value acquired is zero.
5	Markov regime probability	The probability of the MSCI World Index being in a low regime, values range between zero and one.
6	Index log returns	Log returns of MSCI World Index.
7	Drop from peak	The percentage remaining of the MSCI World Index price is compared to the last pre-crisis price level. In the case of a non-crisis period, a zero value is given.
8	Rise from crisis	Percentage above MSCI World Index price compared to the last crisis period. In the case of a crisis period, a zero value is given.
9	RSI 14	The relative strength index of the last 14 months measures spending and change in price movements and oscillates between 0 and 100.
10	Stochastic %K	A stochastic oscillator is a momentum indicator that compares the price of an index to past prices over 14 months.
11	Stochastic %D	Stochastic %K indicator averages three months of stochastic oscillator results to show longer-term trends.
12	MACD	Moving average convergence/divergence is a trend-following momentum indicator that depicts the relationship between two exponential moving averages of an index price, which is calculated by subtracting the 26-period

No.	Feature Name	Feature Description
		exponential moving average from the 12-period exponential moving average.
13	Williams %R	Williams %R, often known as the Williams Percent Range, is a momentum indicator that monitors overbought and oversold conditions. It is calculated by subtracting the highest 14 months price from the last price and dividing by subtraction of the highest 14 months price from the lowest 14 months price. It ranges between 0 and -100.
14	Disparity index	The disparity index is a technical indicator that calculates the percentage difference between an index's monthly price and a 14-month moving average.
15	Crude oil price change [81]	% change in crude oil price compared to the previous month.
16	Gold price change [81]	% change in gold price compared to the previous month.
17	CPI [82]	Consumer Price Index of US market.
18	3-month treasury bill [83]	3-month US treasury bill rate.
19	10-year treasury note [84]	10-year US treasury note rate.

2.9. Features Preprocessing

All 19 selected features have no missing values, no extreme outliers and will not be used in statistical methods, which require strict compliance with assumptions, thus, the need for features preprocessing is minimal. In the scope of this thesis, only scaling of the features between zero and one will be implemented. It is suggested, especially for LSTM-based models, because it enables faster learning and better network convergence. After completing predictions, the data scale will be inverted to return to initial values for a more straightforward interpretation.

During feature scaling, it is critical to prevent data leakage from validation or test data sets into the training process, which could potentially inflate the performance metrics and degrade the model performance. Therefore, the scaler's minimum and maximum values will be calculated only on the training set to avoid data leakage. Then the computed minimum and maximum values will be applied to transform both training and test sets. It prevents data leakage, however, validation and test set values could fall outside the zero and one interval in case validation or test sets have larger or smaller values than the minimum or maximum value of the training set. Nevertheless, in the scope of this thesis, such situation does not affect the model performance.

2.10. Models' Training Strategy

In this thesis scope, different neural networks-based models: multilayer perceptron, recurrent neural networks, LSTMs and GRUs will be applied while varying combinations of hidden network layers, their nodes, learning and dropout rates and activation functions to achieve the best prediction score. Results acquired from neural networks-based models will be then compared to three baseline models – repeating the last value endlessly, repeating the last time series indefinitely and the simplest linear model.

Deep learning model training and forecasts will commence by, first of all, creating appropriate time windows: training, validation and test windows, as suggested by Jerez and Kristjanpoller [85]. The training phase will be responsible for searching the optimal parameters of the model [85]. The validation phase will allow evaluation of the performance of different combinations of hyperparameters, such as the width or depth of a network and learning or dropout rate, of the model on data that it has not seen during training but at the same time not overfit the test set [85]. Following good practices, validation of the models will commence from overfitting. Then, parameters will be modified gradually or the dropout rate will be introduced to reduce model complexity to reduce overfitting, i.e. increase training error.

Model performance in certain conditions strongly depends on the validation set [85], i.e. the model predicts price recovery during crisis periods better than during other periods if, during the validation phase, the model was validated during crisis periods. Therefore, the validation set will include a Dot-com bubble with 80 months of peak-to-peak length, where 50 months were dedicated to recovery.

One time window will consist of 72 months dedicated as training inputs and 72 months periods to either validate hyperparameters or test models' forecasts depending on the model stage. Because the past ten most significant crashes have not required more than 62 months of recovery, the 72-period forecasting window was chosen with some additional cushion. 72 months period is also in line with Rafiq's [18] findings that stock market prices after financial crises are expected to recover within 4-6 years.

Since MSCI World Index has over 600 months of available data and one entire data window consists of 144 months, deep learning models will receive multiple data windows, each one moved forward by one month, see Fig. 16.

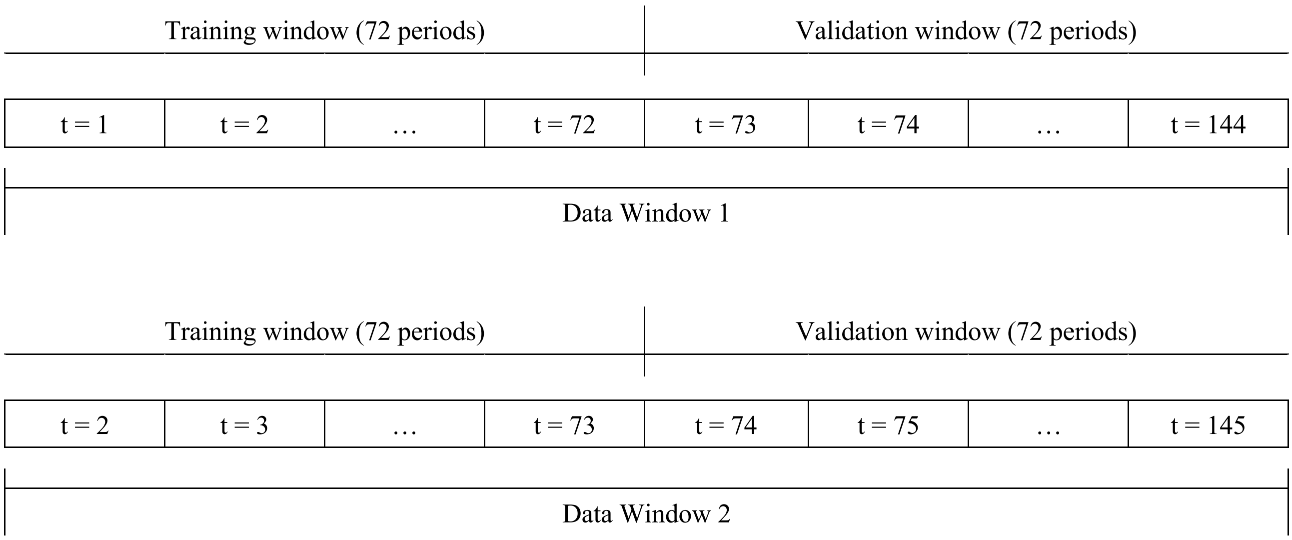

Fig. 16. Difference between data windows

Neural networks-based models are more computationally efficient when trained in batches instead of providing all data windows simultaneously [86]. Thus, in this thesis scope, the models will receive batches of data windows, where one batch will have 32 data windows. For further improvements in model performance, the data windows inside a batch will be shuffled as each data window is independent of all others [86]. It will not introduce data leakage as the time series order within each data window will be maintained, see Fig. 17.

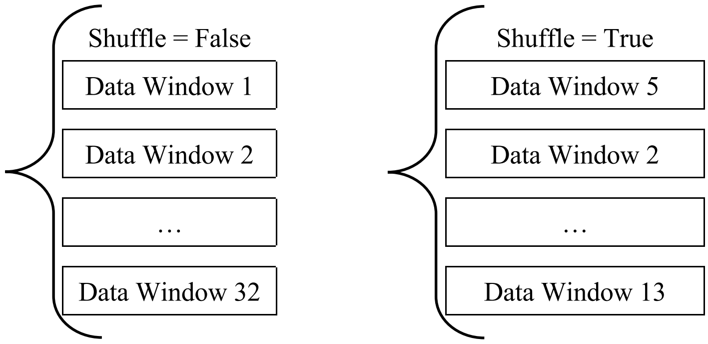

Fig. 17. Data windows shuffling effect on batch

35 % of the data was set aside for testing purposes and 65 % was dedicated to training and validating. The primary testing period of performance will be the Great Recession of 2007. When training the models, 1000 epochs will be allowed, where one epoch is defined as a model going through the training process on all batches once. However, all models converged before reaching the 1000 epochs limit. Additionally, each model will have a dense layer containing one neuron at the end to make the output more stable [54]. After acquiring the best performing model while using train and validation sets, the model will try forecasting a period of a test set, which until this time was not shown to the models. Consequently, after choosing the best method, as suggested by Yang [49], the maximum amount of available data will be used for training and validating and additional forecasts will be made for later periods starting from 2023 due to research interest.

2.11. Performance Metrics

Performance metrics are an important part of algorithms' success. They are a mathematical construct that measures how close actual outcomes are to what was expected or forecasted [87]. Botchkarev [87] analysed performance metrics and discovered that the three most popular metrics among the scientific community were MSE (or RMSE), MAE and MAPE. MAPE from the 1990s became the most popular and MAE took second place [87]. Additionally, Botchkarev [87] observed multiple instances where the most popular metrics have been strongly criticised, rejected or defended. Thus, there is no single metric that should be used universally. In this Master's thesis, MSE will be chosen as a main optimisation metric to penalise more significant deviations from the actual value because bigger deviations may mistakenly cross the recovery point, leading to the announcement of a recovery that is too early. However, MAPE and MAE will also be provided for reference. MAPE is useful because it shows the percentage difference between the predicted and actual values and is easy to interpret. At the same time, MAE penalises all errors equally, thus allows to see a more balanced view of errors. All three performance metrics show more satisfying predictions when closer to 0 and worse when going away from zero. However, MSE and MAE are scale-dependent metrics, meaning they can be compared only when used on the same data and are unsuitable for comparing data with different scales.

Table 6. Formulas of used performance metrics

Metric Abbreviation	Metric Name	Metric Formula [87]
MSE	Mean Squared Error	$MSE = \frac{1}{n}\sum_{i=1}^{n} e_i^2$
MAE	Mean Absolute Error	$MAE = \frac{1}{n}\sum_{i=1}^{n} \lvert e_i \rvert$
MAPE	Mean Absolute Percentage Error	$MAPE = \frac{100}{n}\sum_{i=1}^{n} \frac{\lvert e_i \rvert}{\lvert A_i \rvert}$

Where A_i – actual value;

e_i – error retrieved by subtracting from the actual value a predicted value;

n – the size of the dataset.

2.12. Used Technical Tools

In this Master's thesis, various tools were utilised to forecast the stock market's recovery after the crash. Python was the primary programming language, where libraries of Pandas and NumPy were used for data preparation purposes, Seaborn and Matplotlib were used for data visualisation. For feature preprocessing and prediction, Scikit-learn, Keras, and Tensorflow were utilised. Furthermore, the Markov Regime Switching model was implemented using Statsmodels and the SHAP package allowed to explain the models. Finally, R's Performance Analytics package was also used to identify crisis and recovery periods.

3. Results of Research

This section will examine the effects of altering the architecture and hyperparameters of the MLP, RNN, LSTM, and GRU models on the MSE performance metric calculated on the validation set. Then, after understanding the dynamics, multiple models' performance will be explored on validation and test sets using MSE, MAE and MAPE metrics. In order to assess model performance under various regimes, the test set will be further divided into two periods: one covering only the Great Recession and another ranging from the Great Recession to December 2022. Finally, after showing and analysing the model predictions, feature importance, limitations and suggestions for additional research will be provided.

3.1. Impact of Model Architecture and Hyperparameters on Performance

This subsection will analyse the impact of the model architecture and hyperparameters on the performance of the validation set through the MSE metric. Specifically, learning rates, number of hidden units, node sizes and activation functions of MLP, RNN, LSTM and GRU models will be explored. The goal is to understand better how these hyperparameters affect the models' capacity to generalise by investigating their effects on the validation set. The validation set metric was chosen to see the changes in models' performance on data not seen during training, but at the same time, not to overfit the test data.

3.1.1. Learning Rate

The learning rate is an essential hyperparameter that controls the step size of model updates during the training process in deep learning algorithms and its value can significantly affect the model's performance and training time. Having too small or too large of values will not allow the models to converge. The table below summarises the impact on the performance of the validation set MSE metric for each of the four models using different learning rates ranging from 0.00001 to 1. Each model investigated had two hidden layers with 32 nodes of the same type, activation function chosen was ReLU. The findings suggest that the performance of the models varied significantly across different learning rates. The best learning rates varied between 0.0001 and 0.01, where MLP and GRU achieved the best result at 0.001 rate, RNN at 0.01 and LSTM at 0.0001. Additionally, MLP and RNN performed very similarly at both 0.001 and 0.01 learning rate levels, whereas LSTM and GRU performance degraded more significantly outside their most optimal rates. Interestingly, with higher learning rates, such as 0.1 or 1, LSTM and RNN models did not converge, and low learning rates extended convergence duration multiple times. For example, a convergence of the model in the MLP case increased from 8.18 seconds for a 0.001 learning rate to 28.15 seconds for the lowest learning rate. These findings highlight the importance of selecting an appropriate learning rate for training deep learning models on financial time series data. During the primary models' optimisation stage, a 0.001 learning rate, which is also a default rate for Adam optimiser in the Keras package [88], will be preferred due to the best performance among a larger group of models and the trade-offs between model performance and computational efficiency.

Table 7. Learning rate impact on MSE validation set metric

Learning rate	0.00001		0.0001		0.001		0.01		0.1		1	
Model	**MSE**	**Time**	**MSE**	**Time**	**MSE**	**Time**	**MSE**	**Time**	**MSE**	**Time**	**MSE**	**Time**
MLP	0.58	28.15	0.18	15.89	0.03	8.18	0.03	4.59	0.20	3.03	0.37	4.19
RNN	0.35	76.26	0.29	12.56	0.09	28.19	0.07	19.42	0.34	7.94	NA	NA
LSTM	0.35	116.85	0.27	16.28	0.45	9.92	0.78	22.02	NA	NA	NA	NA
GRU	0.32	120.79	0.22	24.88	0.06	74.05	0.18	19.76	0.37	13.89	0.56	11.73

3.1.2. Number of Hidden Layers

Having analysed the impact of the learning rate on the MSE validation set metric and time to convergence, an investigation of the impact of hidden unit size will be conducted for each of the four types of neural networks-based models. The models investigated will have a fixed optimal learning rate of 0.001, ReLU activation function and 32 nodes in each hidden unit. The table depicting the results is provided below. The findings suggest that the best performance was achieved with three hidden layers for MLP, RNN and LSTM models, while GRU performed better with two hidden units. Nevertheless, the performance of MLP did not significantly differ when choosing between one and four hidden layers. Interestingly, the LSTM and RNN models achieved the highest MSE validation error at a hidden layer size of four. GRU performance was also the worst at three and four hidden layers compared to a smaller number of layers, indicating that increasing the hidden layer size beyond a certain point can lead to overfitting and decreased performance. The time to convergence also varied, for example, sometimes up to 10 times for LSTM, where the MLP model was the only one, where the duration until convergence for different numbers of layers did not differ that significantly. These results also underline the significance of choosing an appropriate hidden layer size for neural networks-based models. This Master's thesis will prefer models with one, two or three hidden layers during the main optimisation.

Table 8. Hidden layers size impact on MSE validation set metric

Hidden layer size	1		2		3		4	
Model	**MSE**	**Time**	**MSE**	**Time**	**MSE**	**Time**	**MSE**	**Time**
MLP	0.03	13.56	0.03	8.18	0.02	11.72	0.03	7.57
RNN	0.12	13.26	0.09	28.19	0.06	24.65	0.17	15.55
LSTM	0.36	7.76	0.45	9.92	0.15	80.72	0.60	22.23
GRU	0.08	23.47	0.06	74.05	0.11	51.38	0.10	59.19

3.1.3. Node Size in Hidden Layer

After analysing the impact of both learning rate and hidden layer size on the MSE validation set metric, the following exciting area for investigation is the effect of node size within each hidden layer. The table below outlines the influence on the MSE validation set when changing hidden layers' nodes size between 8 and 64. The models chosen have a learning rate of 0.001, ReLU activation function and three hidden units due to their effectiveness in the above investigations. The results show that MLP, RNN and LSTM performance is best at 32 nodes, for the GRU model at 64 nodes. Moreover, MLP performance is consistently good in the range of 8-32 nodes and starts to degrade at 64 nodes. On the other hand, RNN and LSTM models' performance is degrading at lower and higher node values with respect to 32 nodes. Regarding GRU, its performance persistently improves with more significant node sizes. Timewise RNN and LSTM models converged slower at smaller hidden layers node sizes. In contrast, GRU converged slower at the highest value and for MLP no significant difference in time was noticeable. These results indicate that node size also matters in financial time series analysis, despite some indicating it as a less impactful parameter.

Table 9. Hidden layers node size impact on MSE validation set metric

Node size	**8**		**16**		**32**		**64**	
Model	**MSE**	**Time**	**MSE**	**Time**	**MSE**	**Time**	**MSE**	**Time**
MLP	0.03	10.70	0.03	7.89	0.02	11.72	0.05	10.94
RNN	0.09	42.53	0.19	15.38	0.06	24.65	0.13	12.67
LSTM	0.36	61.04	1.91	21.87	0.15	80.72	0.48	16.76
GRU	0.24	42.68	0.11	55.14	0.11	51.38	0.04	83.54

3.1.4. Activation Function

Finally, the last important step is to explore how activation functions affect the model's performance. In this case learning rate is set to 0.001, all models will have three hidden units with 32 nodes except for GRU, which will have 64 nodes due to the effectiveness and activation functions will be varied between the ReLU, Tanh and linear. Results are provided in the table below. The ReLU activation function yields the lowest MSE values for MLP, LSTM and GRU models, while Tanh yields the lowest MSE value for RNN. MLP, RNN and GRU models' performance differs negligibly between different activation functions, while for LSTM activation functions have a more significant impact. These findings confirm previous research that satisfactory results can be achieved with multiple activation functions. Nevertheless, this research thesis will prefer the ReLU function except for RNNs, where Tanh will be explored more in-depth too.

Table 10. Activation function impact on MSE validation set metric

Activation function	**ReLU**		**Tanh**		**Linear**	
Model	**MSE**	**Time**	**MSE**	**Time**	**MSE**	**Time**
MLP	0.02	11.72	0.03	9.39	0.03	4.03
RNN	0.06	24.65	0.04	62.16	0.09	45.26
LSTM	0.15	80.72	0.19	48.81	0.25	40.06
GRU	0.04	83.54	0.06	79.59	0.04	88.32

3.2. Selection of the Best Models

The best performance measures were attained by experimenting with different architectures and hyperparameters combinations of MLP, RNN, GRU, and LSTM models. Three baseline models were also used for comparison. Baseline models included linear, endlessly repeating the last value and previous time series models. The MSE, MAE and MAPE performance on unseen test data was calculated in two ways: one considering unseen data between the Great Recession to the end of December 2022, called the test and another only taking the period when the Great Recession was present, called the crisis. Such decision to measure performance on different durations of test sets was made to understand:

1. How the model performs only during the crash periods;
2. How the model generalises both during crash periods and expansionary periods;
3. Are there any differences in performance, i.e. one period is forecasted notably better than another.

The multilayer perceptron model acquired the lowest MSE performance metric, equal to 0.09 on the entire test set, with one hidden layer, eight nodes, a learning rate of 0.001 and a ReLU activation function. The best MSE score measured only during the Great Recession period and equal to 0.03 was acquired by the multilayer perceptron with one hidden layer, 64 nodes, a learning rate of 0.001 and a ReLU activation function. Nonetheless, the best MLP model on the crisis set, according to the MSE metric, performed almost identically to the best MLP model on MSE full test set, which scored 0.04. Therefore, the difference between these top two MLP models is negligible.

Regarding the MAE score on the test set, the best model was a multilayer perceptron with two hidden layers, 32 nodes, a learning rate of 0.001 and a ReLU activation function and reached a score of 0.2. On the other hand, the baseline model, which repeated previous values, achieved the best MAE on the crisis set equal to 0.14. Nevertheless, the multilayer perceptron with three hidden layers, 32 nodes, 0.2 dropout and ReLU activation function was extremely close to the baseline model with MAE on the crisis set also equal to 0.14. As a consequence, the baseline model was the best but did not show superior performance compared to other models. Additionally, using the baseline model includes a sudden decrease in prices, performance strongly depends on the historical period taken and therefore lowers the trust in the model, as displayed below in the figure.

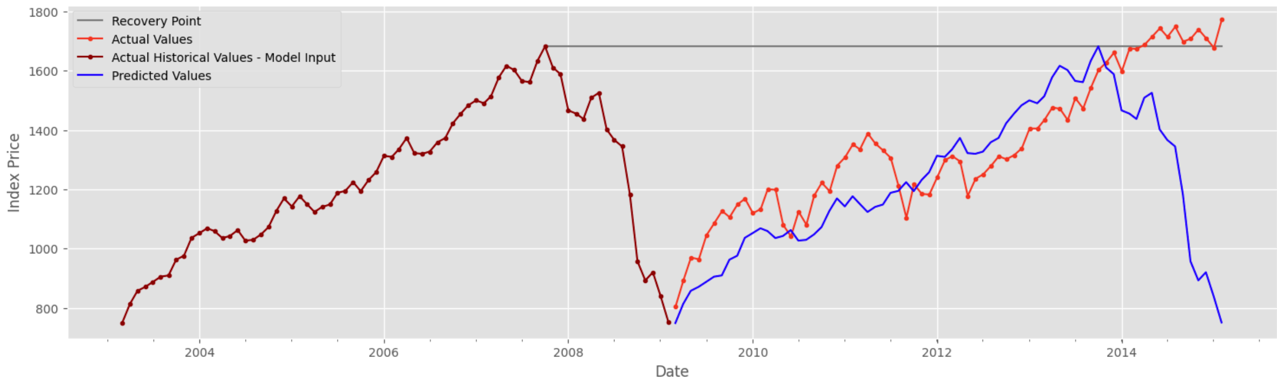

Fig. 18. MSCI World Index future price forecast from February 2009 using the best repeating time series baseline model

The best MAPE value of an entire dataset at 13.07 % was attained by the same multilayer perceptron model, which had the best MSE full test set performance. The MAPE crisis set best performance was equal to 12.55 %, which was also acquired by the baseline model, repeating the previous values. Still, the second-best model based on MAPE on crisis set was the GRU model with one hidden layer and 16 nodes, a ReLU activation function, a 0.001 learning rate and scored 13.35 %, which is less than one percentage point difference on average in prediction error. For that reason, once again, the baseline model did not show a much greater performance on the MAPE crisis set metric compared to neural networks-based models.

The findings above indicate that various metrics generate different best models. However, it is evident that in this financial time series price forecasting task, the clear winners from the four analysed models' groups are the multilayer perceptron models. They performed the best on the test and crisis set for the MSE performance metric and the best for MAE and MAPE test set data. Only for MAPE and MAE values on the Great Recession period the models did not beat the repeating time series baseline model. Yet, there was no drastic performance difference between the top baseline model and the runner-up neural networks in these cases.

However, being unable to display better neural networks-based model results compared to the baseline repeating model during the Great Recession period for MAE and MAPE metrics shows that the Great Recession period exhibited very similar recovery patterns to the pre-crisis expansion period and, thus, repeating pre-crisis period sequence allowed for great prediction results. However, if one looks at the big picture through a complete test set and compares the baseline repeating pattern model with the best multilayer perceptron model, the multilayer perceptron model showed two times better MSE performance of 0.2 compared to 0.42. The identical results were also generated on the MAPE metric, 13.07 % to 25.62 %. It displays that multilayer perceptron generalises better on multiple regimes of financial time series and potentially may demonstrate better results in future crises also. Especially if future crashes' recovery would resemble less the sequence of the previous growth.

Interestingly, all top 10 models exhibited more accurate predictions on the whole time series than the crisis dataset. For instance, the top 1 model's forecasts for the Great Recession were, on average, 17.77 % off from the actual values and only 13.07 % off from the entire dataset based on MAPE, indicating that crash periods future dynamics may be more difficult to forecast than non-crash periods.

When analysing the performance of other models, the baseline linear model performed well on the validation set and also showed impressive results on test sets. The performance of RNN, LSTM, and

GRU models fell far short of MLP despite high hopes for them based on MSE metrics. GRU was the best model if MLP-type models were excluded, with three hidden layers and 64 nodes in 10th place overall. It has a 0.13 MSE test result and 0.10 crisis dataset error, which is larger than the 0.09 and 0.04 scores generated by the best multilayer perceptron model. With MSE scores of 0.17 and 0.19 for the test and crisis datasets, respectively, the first LSTM model is ranked 11th in the final rankings. Finally, the first RNN model appears in 13th place with 0.19 and 0.25 MSE errors on full test and only crisis datasets.

When looking at MAE metrics, the rankings of model families are similar. For example, the best MAE values achieved on full and crisis datasets by the MLP models family were 0.2 and 0.14, where the best GRU had 0.27 and 0.15 scores, the best RNN had 0.37 and 0.17 values and the best LSTM had 0.31 and 0.34. Thus, on the basis of MAE metrics RNN family managed to outperform LSTM on the crisis dataset, but the order of the performances on the test set remained identical to the MSE metric, with MLP coming in first, followed by GRU, LSTM and then RNN.

When comparing results using MAPE metrics, there is greater variation in the performance of different types of neural networks-based models. The best scores acquired by MLPs on test and crisis datasets were 13.07 % and 15.99 %, GRUs scored 17.56 % and 13.35 %, RNNs scored 25.88 % and 15.06 %, LSTMs scored 22.36 % and 31.85 %. Therefore, the order of MLP being the first, followed by GRU, LSTM and then RNN was maintained for MAPE on the test set, however, for the crisis dataset GRU was the first, followed by RNN, MLP and LSTM. See the table below for more scores of performance metrics of selected neural networks-based models.

Table 11. Performance metric scores of selected neural networks-based models

No	Model	Description	MSE Val	MSE Test	MSE Crisis	MAE Val	MAE Test	MAE Crisis	MAPE Val	MAPE Test	MAPE Crisis
1	MLP	One hidden layer with eight nodes	0.04	0.09	0.04	0.16	0.20	0.16	18.65	13.07	17.77
2	MLP	Two hidden layers with 32 nodes both	0.03	0.09	0.07	0.13	0.20	0.23	14.32	13.91	24.13
3	Linear	Linear model	0.03	0.10	0.04	0.11	0.22	0.15	12.63	13.91	16.38
4	MLP	One hidden layer with 16 nodes	0.03	0.10	0.05	0.12	0.21	0.17	13.56	13.31	19.03
5	MLP	Three hidden layers with 32 nodes both	0.03	0.10	0.06	0.13	0.23	0.21	15.21	15.20	22.63
6	MLP	One hidden layer with 32 nodes	0.03	0.10	0.06	0.12	0.22	0.20	13.66	14.24	21.76
7	MLP	Two hidden layers, 32 nodes, 0.2 dropout	0.03	0.10	0.10	0.14	0.23	0.25	14.36	15.48	27.06
8	MLP	One hidden layer with 64 nodes	0.03	0.11	0.03	0.12	0.24	0.16	13.38	15.10	16.70

No	Model	Description	MSE Val	MSE Test	MSE Crisis	MAE Val	MAE Test	MAE Crisis	MAPE Val	MAPE Test	MAPE Crisis
9	MLP	Three hidden layers. 32 nodes. 0.2 dropout	0.05	0.12	0.04	0.17	0.24	0.14	17.63	14.91	15.99
10	GRU	Three hidden layers with 64 nodes	0.06	0.13	0.10	0.18	0.27	0.26	19.31	17.56	22.91
11	LSTM	One hidden layer, bidirectional, eight nodes	0.08	0.17	0.19	0.24	0.31	0.40	26.02	22.36	40.33
12	Baseline Last	Repeating last value	0.06	0.18	0.11	0.21	0.34	0.28	25.97	22.19	30.33
13	RNN	Three hidden layers with 32 nodes both	0.04	0.19	0.25	0.13	0.37	0.47	13.66	26.38	42.56
14	LSTM	One hidden layer with 32 nodes	0.14	0.23	0.18	0.32	0.38	0.34	34	26.45	31.85
15	RNN	Two hidden layers with 32 nodes both	0.07	0.24	0.09	0.20	0.40	0.22	21.73	25.88	20.94
16	Baseline Repeat	Repeating training set pattern	0.07	0.27	0.04	0.18	0.42	0.14	17.98	25.62	12.55
17	GRU	Two hidden layers with 32 nodes both	0.06	0.32	0.09	0.16	0.48	0.19	16.45	30.74	17.25
18	GRU	One hidden layer with 16 nodes	0.05	0.36	0.06	0.17	0.50	0.15	17.47	30.69	13.35
19	GRU	One hidden layer with 32 nodes	0.06	0.39	0.07	0.18	0.53	0.16	20.06	33.21	14.14
20	RNN	One hidden layer with 32 nodes	0.04	0.40	0.05	0.16	0.53	0.17	17.52	31.83	15.06
21	RNN	Two hidden layers, 32 nodes, 0.2 dropout	0.12	0.43	0.26	0.23	0.58	0.46	24.14	39.90	44.02
22	RNN	One hidden layer with 16 nodes	0.04	0.43	0.06	0.15	0.55	0.18	16.91	33.35	16.48
23	GRU	Two hidden layers, 32 nodes, 0.2 dropout	0.11	0.48	0.13	0.28	0.63	0.29	29.27	41.17	26.13
24	RNN	Three hidden layers with 32 tanh nodes both	0.04	0.55	0.09	0.16	0.64	0.25	16.57	39.70	21.89
25	GRU	Three hidden layers, 64 nodes, 0.2 dropout	0.12	0.63	0.20	0.28	0.73	0.39	28.78	48.03	35.72

No	Model	Description	MSE Val	MSE Test	MSE Crisis	MAE Val	MAE Test	MAE Crisis	MAPE Val	MAPE Test	MAPE Crisis
26	RNN	One hidden layer with eight nodes	0.11	0.71	0.16	0.29	0.76	0.36	29.74	47.75	31.69
27	RNN	Three hidden layers,32 tanh nodes, 0.2 dropout	0.07	0.71	0.15	0.20	0.75	0.34	20.63	47.41	29.62
28	LSTM	One hidden layer bidirectional with 16 nodes	0.19	0.72	0.33	0.37	0.77	0.55	40.27	51.15	51.32
29	LSTM	One hidden layer with eight nodes	0.23	0.82	0.41	0.41	0.86	0.62	45.37	57.98	58.18
30	LSTM	Two hidden layers, 32 nodes, 0.2 dropout	0.22	0.85	0.53	0.42	0.83	0.70	44.80	56.90	66.50
31	LSTM	One hidden layer with 16 nodes	0.20	1.05	0.55	0.35	0.98	0.73	38.51	65.78	68.76
32	LSTM	Three hidden layers, 32 nodes, 0.2 dropout	2.10	1.29	0.59	0.62	1.06	0.75	59.51	71.19	71.33
33	GRU	One hidden layer with eight nodes	0.30	1.65	0.70	0.52	1.23	0.82	54.85	81.72	77.34
34	LSTM	One hidden layer, bidirectional, 32 nodes	0.46	2.28	1.38	0.39	0.84	0.90	42.82	52.87	87.60
35	LSTM	Two hidden layers with 32 nodes	0.12	2.69	0.30	0.31	1.07	0.42	32.71	64.56	36.41

3.3. Forecasts of Neural Networks-based Methods During the Great Recession

The study's primary goal is to understand how quickly stock markets may rebound from a recession. As a result, beginning from various initial time points, this part will analyse the graphs when the best models declared the recovery during the Great Recession. As a reminder, the MSCI World Index is considered to be recovered when the price returns to its highest pre-crash price.

3.3.1. Best Model Overall

The best model based on the MSE test set metric, equal to 0.09, was acquired by the multilayer perceptron model with one hidden layer, eight nodes, learning rate of 0.001 and a ReLU activation function. To visualise how the best model is capable of predicting future time series values starting from different forecasting time intervals, four intervals were chosen: from December 2007, which marked one month into the Great Recession, December 2008, which marked 13 months into the crisis and began from almost the lowest point, February 2009, which was the lowest point of Great Recession and December 2009, which already showed strong recovery signals in index price.

When looking at the first case, starting predictions from December 2007, the model initially commenced from a slightly larger price drop compared to the actual price, but also displayed slower price-rising tendencies, which eventually led to an almost perfect match between the actual price and anticipated price during 2014. Short-term price swings were not captured by the model. However, since they can occur as a result of numerous unknowable and potentially unpredictable reasons, this thesis concentrates on the length of recovery.

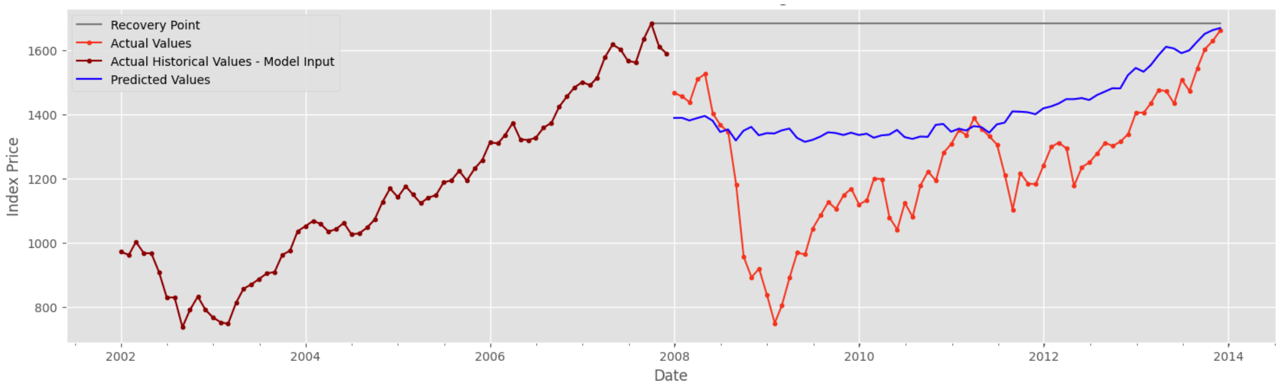

Fig. 19. MSCI World Index future price forecast from December 2007 using the best MLP model

In the second instance, starting with an index from a significant price drop on December 2008, the model initially began from a much higher index value but also showed slower price-rising tendencies, which eventually led to an exact match between the actual price and anticipated price close to the end of 2013.

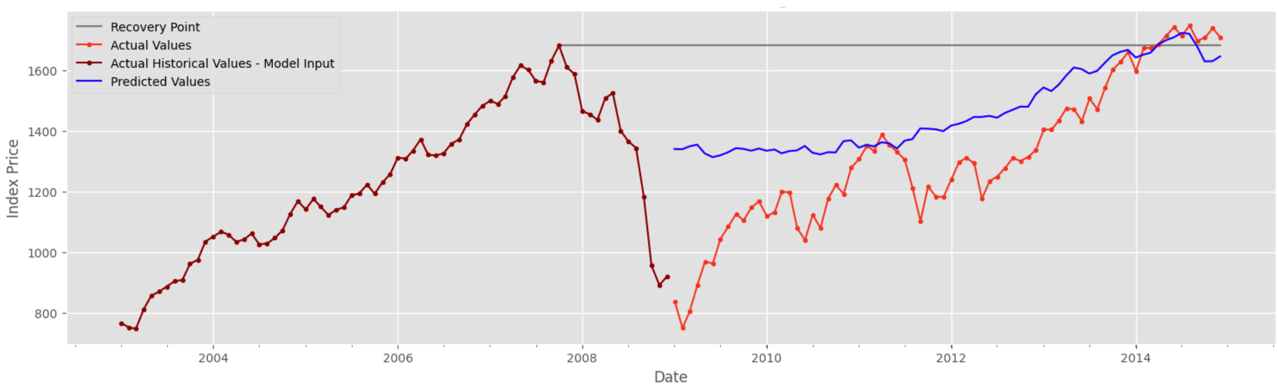

Fig. 20. MSCI World Index future price forecast from December 2008 using the best MLP model

In the third instance, starting with a price prediction from the lowest crisis point on February 2009, the model initially began from a much higher index value but also showed slower price-rising tendencies, which eventually led to an exact match between the actual price and anticipated price close to the end of 2013.

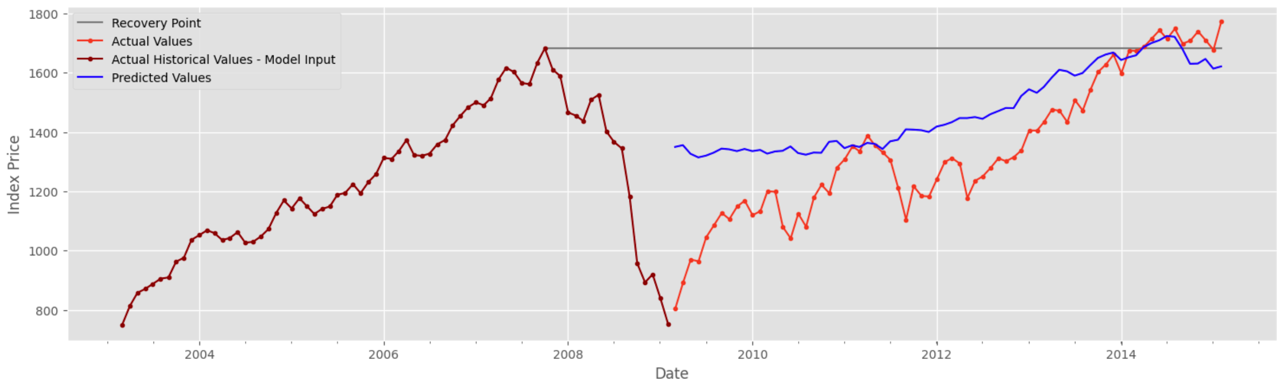

Fig. 21. MSCI World Index future price forecast from February 2009 using the best MLP model

In the fourth instance, starting with a price prediction from a moderately rebounded price on December 2009, the gap between the model's initial value and actual value was much smaller than in the previous two cases, showed slower price-rising tendencies and eventually led to an exact match between the actual price and anticipated price at the end of 2013.

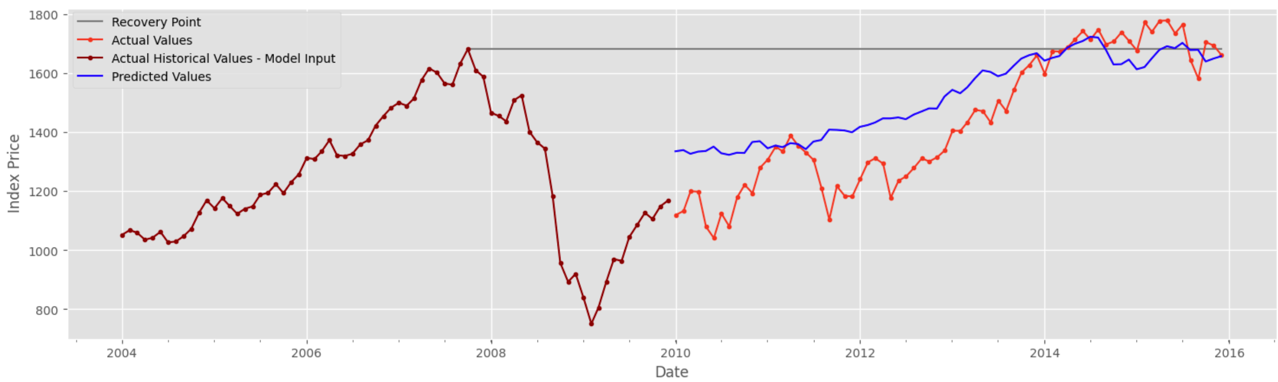

Fig. 22. MSCI World Index future price forecast from December 2009 using the best MLP model

It is clear from these projections that the strongest MLP model can predict recovery time reasonably accurately. However, when the starting value is extreme, it can occasionally fail to capture the initial point but makes up for it with a smaller trend. Furthermore, the model does not try to account for short-term fluctuations and tries to follow a long-term trend.

3.3.2. Best RNN model

The best RNN model, having three hidden layers with 32 nodes, when analysed from the lowest Great Recession price point on February 2009, had an initial price relatively close to the original one. However, the trend it caught was significantly larger than the real recovery trend. Therefore, the RNN model would have announced recovery prematurely by two years.

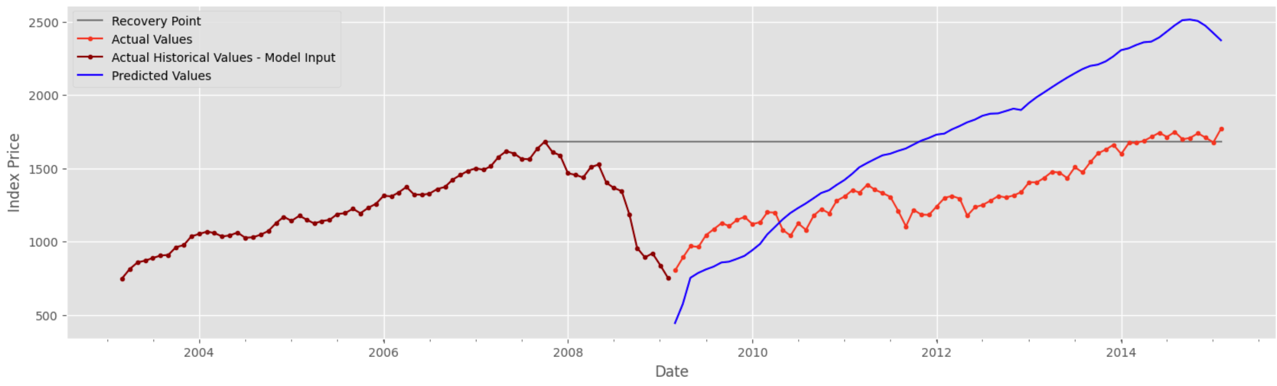

Fig. 23. MSCI World Index future price forecast from February 2009 using the best RNN model

3.3.3. Best LSTM model

The best LSTM model, having one hidden bidirectional layer with eight nodes, had a starting price that was extremely different from the original one and the price increase trend was excessively strong when analysed from the lowest Great Recession price point in February 2009. Thus, the model overshoots the recovery point by two years but then returns to the recovery point later. Nevertheless, it is visible that the LSTM model is struggling with the generalisation of performance and can be seen as unsuitable for these predictions.

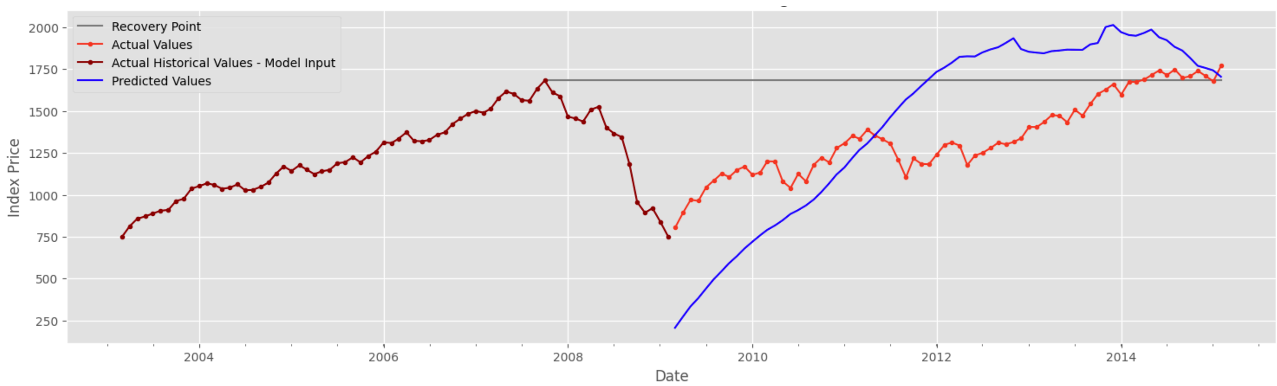

Fig. 24. MSCI World Index future price forecast from February 2009 using the best LSTM model

3.3.4. Best GRU model

The best GRU model captured the trend exceptionally close to the original price when analysed from the lowest Great Recession price point in February 2009. One, who trusted the model during the Great Recession, would have declared the recovery just a few months earlier than the actual recovery happened.

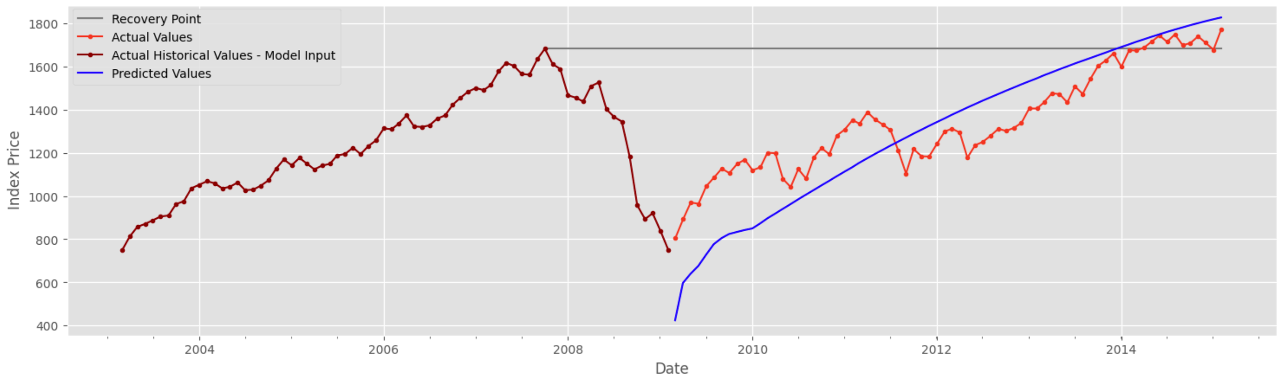

Fig. 25. MSCI World Index future price forecast from February 2009 using the best GRU model

3.4. Forecasts of Neural Networks-based Methods After the Great Recession

To understand how the best MLP model generalises to different stock market regimes, future long-term price projections will be visualised in this chapter during non-crash periods.

3.4.1. Best Model Overall

Starting with predictions from May 2014, it is visible that initial and predicted values are incredibly close to each other, showing that it is easier for the model to predict the closest prices when lower volatility was present in the time series before. However, the model did not catch a more extreme spurt in stock market prices between 2016 and 2018. Nevertheless, the predictions and actual values converged due to the Covid-19 crash in 2020.

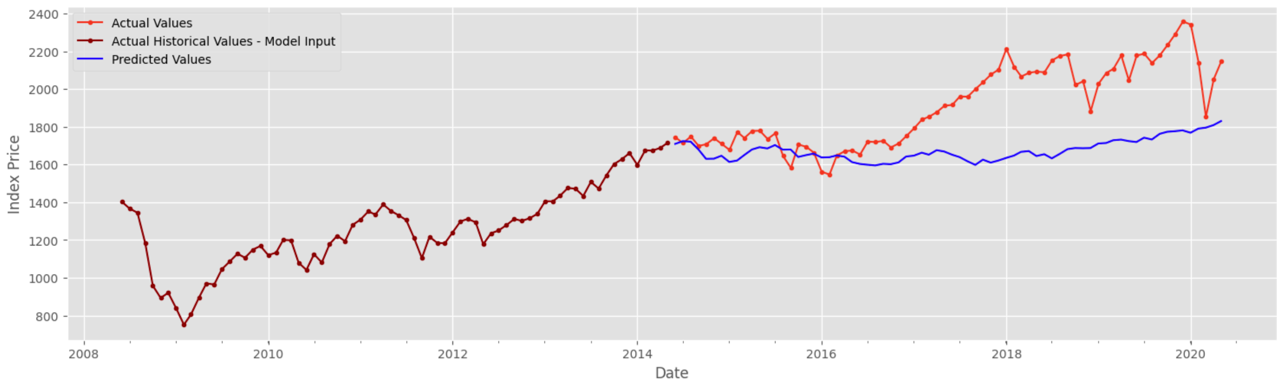

Fig. 26. MSCI World Index future price forecast from May 2014 using the best MLP model

Analysing estimates from December 2016 reveals a very similar pattern: initial anticipated values are extremely near at first, then the gap widens for three years until the Covid-19 crash, which equalised the prices, but then the gap once again started growing from 2020 onward. It is evident that the price increase between 2020 and 2022 was unusually fast, therefore, that could explain why the model struggled to catch such spurt.

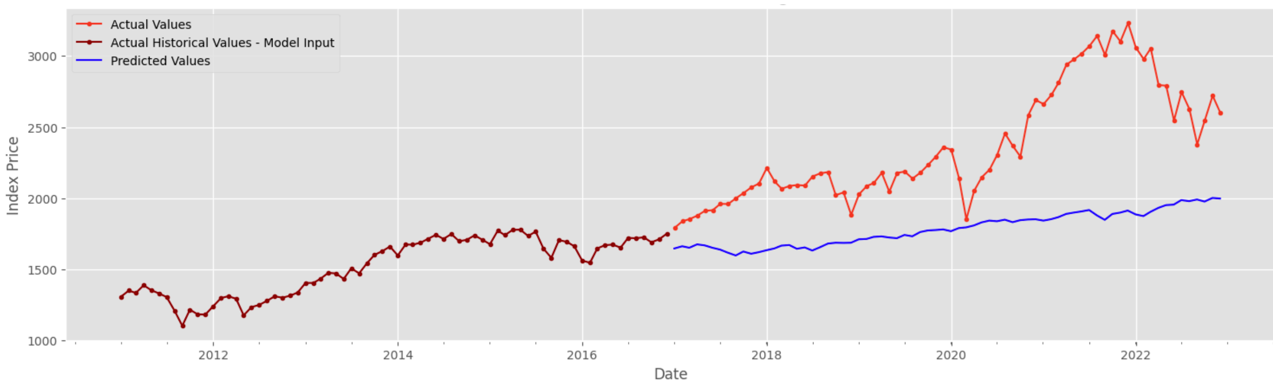

Fig. 27. MSCI World Index future price forecast from December 2016 using the best MLP model

3.5. Forecasts of Neural Networks-based Methods from 2023

3.5.1. Best Model Overall

Out of academic curiosity, the best MLP model was used to forecast prices from January 2023. It coincides with the bear market from which the MSCI World Index has not recovered yet. As in the Great Recession period, the initial gap between the original and predicted price is noticeable and the predicted time series shows a gradual price increase. The Great Recession period showed that even though the beginning values differ moderately when looking at index values from a longer-term perspective, the prices start to match better and better. When comparing to the baseline model of historical values repetition, it is visible that the general predictions trend is similar, where price levels are very close to each other between 2024 and 2026, then they start to diverge, but once again converge in 2029. Nevertheless, the original best model shows that recovery is not expected within 72 months. Thus, in this case, a larger quantity of crash data is needed to capture more dynamics. Potentially, model's retraining or adding new features specific to this crash may also help.

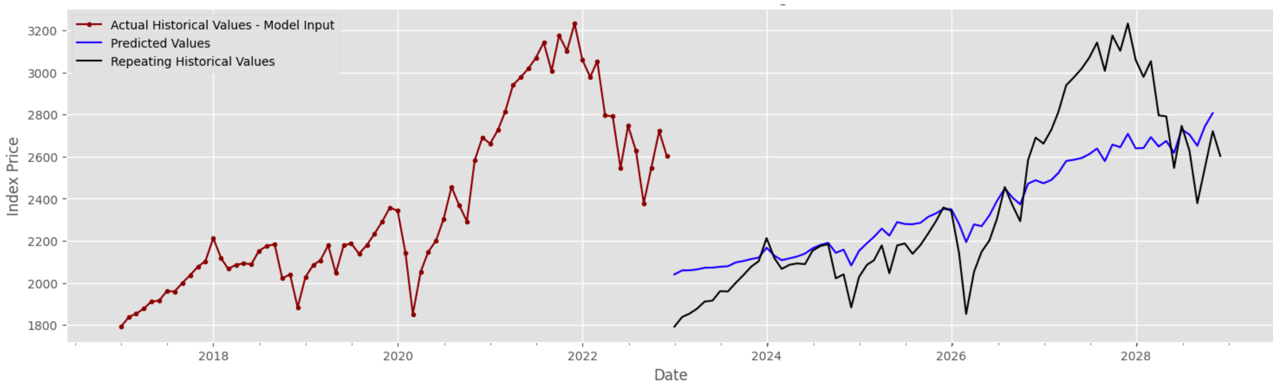

Fig. 28. MSCI World Index future price forecast from January 2023 using best MLP model next to repeating historical values model

3.5.2. Retrained Model on all Available Data

For this part, the best multilayer perceptron model with one hidden layer, eight nodes, a learning rate of 0.001 and a ReLU activation function was retrained on all available data and used for forecasting the same period. The forecast for January 2023 closely resembles the repeating pattern of previous values. However, when considering the Great Recession patterns, repeating historical patterns was the best scoring model for the MAE and MAPE metrics. Nevertheless, the new model results should be viewed with caution as, typically, in case of a significant number of new data entries, it is better to retrain the models with fresh parameters because the top-performing model may change. Moreover, this crash was caused by different dynamics compared to the ones recorded before, among them the weakened economy caused by the Covid-19 pandemic and Russia's invasion of Ukraine, thus, different factors could be in place, which may affect the recovery of MSCI World Index in a way not seen before and the model may not generalise well.

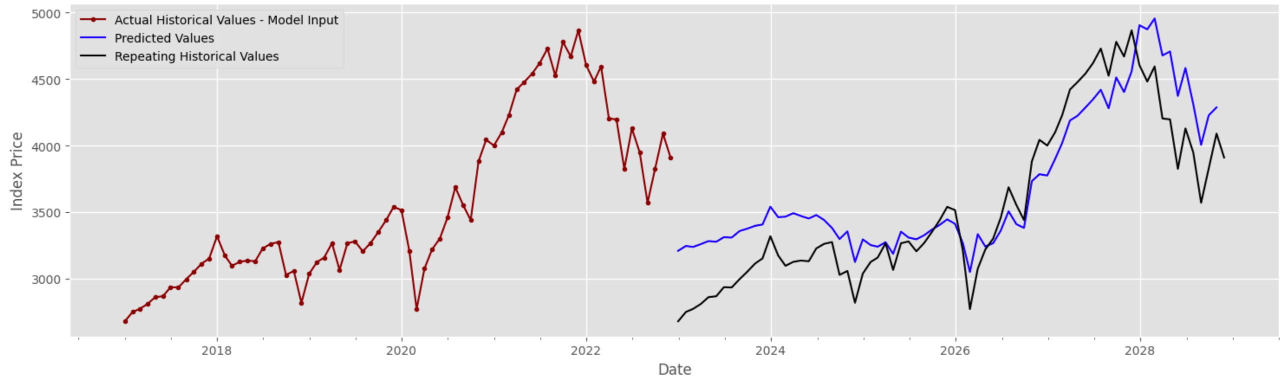

Fig. 29. MSCI World Index future price forecast from January 2023 using best retrained MLP model next to repeating historical values model

3.6. Features Contribution to Predictions

Neural networks-based models are known to be a "black box", which prevents from understanding each feature contribution to every prediction. Nevertheless, with SHAP values, one can measure the impact of each feature on predictions. This research thesis calculated SHAP values for the best multilayer perceptron model during two critical periods: the first month of the Great Recession and the lowest point of the Great Recession market during the forecasted recovery period. The results indicate that the two most essential features for predictions were the consumer price index and the MSCI World Index price, where price effects were more significant during the lowest Great Recession point compared to the first month. Moreover, technical analysis indicators of Williams %R, MACD and Stochastic %K, information on treasury bills and notes and how many months the current crisis period is ongoing were also consistently important. The least important indicators were the percentage drop in price from the pre-crisis peak and the number of months current non-crisis period is ongoing. The gold price change was the only feature with no predictive power. These findings suggest that the selected features have similar contributions to forecasts with minor deviations between the beginning of the crisis and the lowest period of the crisis. Calculated SHAP values are provided below.

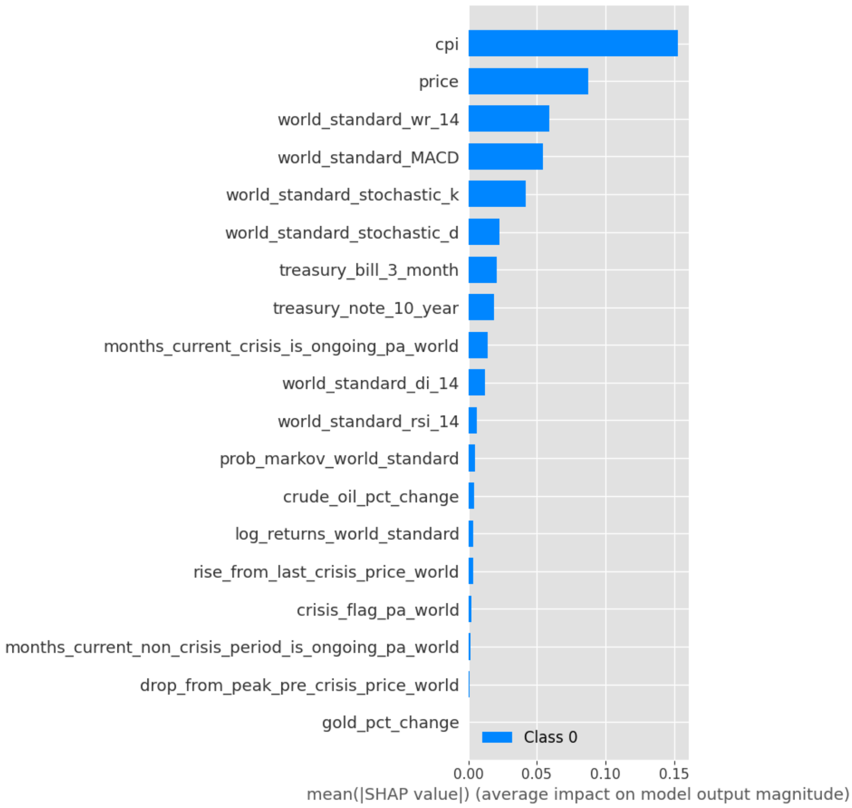

Fig. 30. Feature importance of the best MLP model for forecasts beginning from the first month of the Great Recession

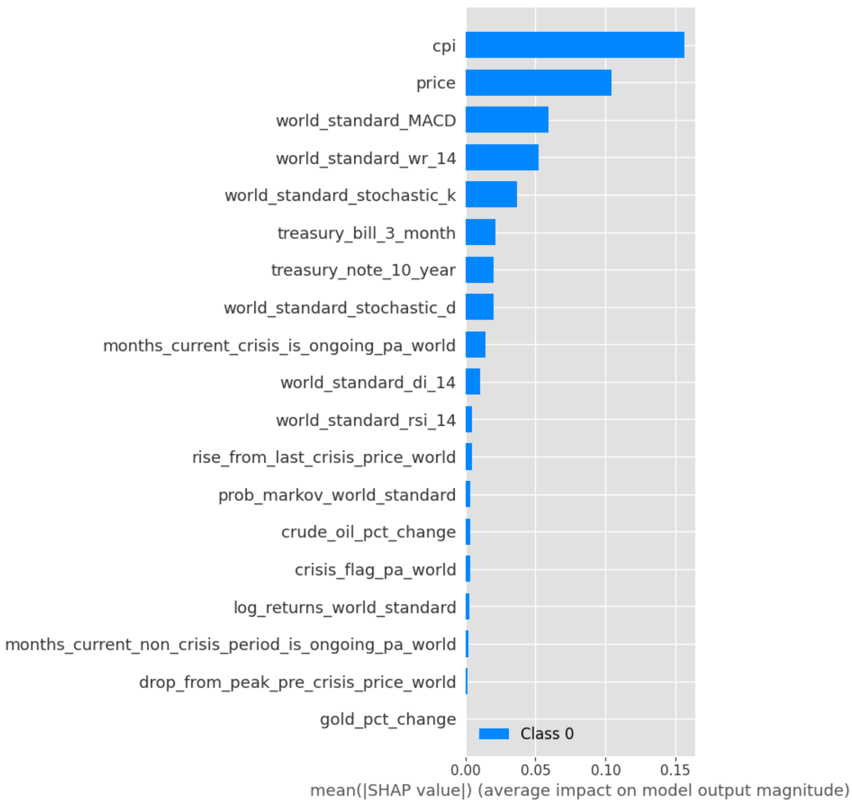

Fig. 31. Feature importance of the best MLP model for forecasts beginning from the lowest point of the Great Recession

3.7. Limitations and Further Research Opportunities

As with any research project, there are some limitations and further potential research opportunities what can be explored more. First of all, only a single index covering global developed markets and medium to large enterprises was investigated. It would be intriguing to develop models that generalise well in other areas, such as in emerging markets, local countries or would try to predict the dynamics of only smaller enterprises. Second of all, there were limited occurrences of crashes due to data availability from December 1969. It would be advantageous to explore more extensive periods before 1969 or repeat the modelling process after new crashes emerge. Third of all, although several neural networks-based models were investigated in this research thesis, it might be interesting to experiment with combining various layers to see how they perform, for instance, by combining MLP with GRU. Finally, new features used in the modelling process could be added as different crises can have different originators. It is clear from the research that other characteristics, like those related to pandemics, politics, or bank recapitalisations, may also have a predictive impact.

Conclusions

1. The literature review revealed that there is no single, widely accepted definition of what it means to recover after a crash. More straightforward approaches, such as the time it takes to reach pre-crash peak prices, are criticised for ignoring the growth component, while more complicated approaches, such as time to return to trend, risk introducing noise or inaccuracies because of the challenging calculations. This study chose a simple approach of return to peak pre-crash price as recovery due to easier interpretability and less risk of introducing inaccuracies.
2. Analysis of literature showed that crashes could be brought on by a multitude of factors, including oil and political shocks, natural disasters such as pandemics or financial related such as distressed banking system, inability to cover sovereign debt payments, substantial currency devaluation or asset price bubbles formation due to biases of investors and institutions.
3. The length of recovery after a crash depends on the number of originators, the depth and duration of a crash, policy responses such as banking recapitalisation and monetary expansion, an economic situation such as budget deficits, private sector credit and financial openness, external factors like growth in world trade, interest rate or volatility in gold or other asset prices and psychological beliefs of individuals.
4. The main research object MSCI World Index between 1969 and 2022 incurred one crash that took less than two years to recover from the lowest crash point, one that took under three years, three crashes that required more than four years and over 20 smaller price fallings below one year. Additionally, one crash did not yet finalise. Such findings are in line with other researchers' results that detected that recoveries take 4-6 years for larger crashes due to the stock market acting as a nearly real-time reflection of economic development, driven by low barriers to entry and speed of transactions.
5. MSCI World Index exhibited unforecastable time horizons of six and seven months based on the Hurst exponent. Nonetheless, longer periods, for example, between the eighth and fifteenth months, demonstrated weak trending tendencies and above 20 months has shown strengthening mean reversion trends indicating prediction potential. These findings are consistent with those of other researchers, who detected that the stock market occasionally exhibits anomalous behaviour, deviating from efficient market hypothesis and hence being forecastable. Therefore, this research thesis attempted to forecast recovery time following a crash in stock markets.
6. For forecasting, four neural networks-based models were chosen to be explored: multilayer perceptron, recurrent, long-short-term memory and gated recurrent unit neural networks. They were preferred over other machine learning models due to their ability to process sequential data. Statistical models were not investigated due to prior research revealing their poorer performance attributable to stock market indices prices being chaotic, noisy, nonlinear and influenced by various environmental factors. Moreover, statistical methods' capabilities especially deteriorate during crisis times and steep recoveries as they do not exhibit average behaviour and can be driven by different elements each time.
7. The study found that the multilayer perceptron model with one hidden layer, eight nodes, a learning rate of 0.001, Adam optimiser and a ReLU activation function acquired the lowest MSE performance metric, equal to 0.09 on the entire test set, ranging from Great Recession to December 2022. During the Great Recession, it scored 0.04. The best MSE score measured only during the Great Recession period and equal to 0.03 was acquired by the multilayer perceptron with identical parameters as the model above, except with 64 nodes. Nonetheless, the performance difference between these top two models on the crisis set is negligible. Additionally, shallow

neural networks performing better than deeper ones on MSCI World Index small data size is consistent with other researchers' findings.

8. Regarding results from MAE and MAPE metrics perspective, multilayer perceptron models performed the best also for MAE and MAPE full test set data. Only for MAPE and MAE values on the Great Recession period the models did not beat the repeating time series baseline model. Yet, there was no drastic performance difference between the top baseline model and the runner-up deep learning models in these cases.
9. The performance of RNN, LSTM, and GRU models fell far short of MLP despite high hopes for them due to their ability to correct themselves and have a "memory". Based on MSE metrics, GRUs were the second best type of models, with three hidden layers, 64 nodes, Adam optimiser and ReLU activation function, in 10th place overall. It had a 0.13 MSE full test result and 0.10 crisis dataset error. With MSE scores of 0.17 and 0.19 for the full test and crisis datasets, respectively, the first LSTM model is ranked 11th in the final rankings. Finally, the first RNN model appears in 13th place with 0.19 and 0.25 MSE errors on full test and crisis datasets. Based on all MAE metrics and MAPE complete test set metric, the order of MLP being the first, followed by GRU, LSTM and then RNN was maintained. However, on MAPE calculated during the Great Recession period, GRUs scored a 13.35 % error, followed by RNNs at 15.06 %, MLPs at 15.99 % and LSTMs at 31.85 %.
10. Interestingly, a significant majority of the best models exhibited more accurate predictions on the whole time series than the Great Recession dataset while looking at the MAPE metric. It may indicate that crash periods' future prices may be more difficult to forecast than non-crash periods due to their more volatile nature, potentially caused by risk-aversion, panic selling and uncertainty about the severity and duration of crashes, the effectiveness of policy responses and more.
11. When forecasting the recovery duration from a crash, the best models initially commenced from a slightly larger or smaller price drop compared to the actual price, did not capture short-term swings, but also typically displayed slower and more even price-rising tendencies, eventually leading to almost perfect match between the actual price and anticipated price. If one had trusted the best RNN and LSTM model, then recovery during the Great Recession at a moment of lowest price would have been announced two years prematurely, while for GRU, only a few months prematurely and MLP would have predicted on point.
12. The research showed that the US consumer price index and the price of the MSCI World Index were the two most important factors for predictions. Moreover, information on Treasury bills and notes, the number of months the current crisis period has been ongoing, Williams % R, MACD, and Stochastic % K technical analysis indicators were all consistently significant. The percentage price decline from the pre-crisis peak and the change in gold price were the least significant.
13. The general recommendation is to employ simpler neural networks-based models to forecast price recovery. Additionally, a comprehensive examination of the crash's causes and remedies in comparison to earlier ones is also necessary. In the case of different downturn drivers and treatment actions, forecasting performance may suffer, necessitating the development of new models with more relevant features to the specific crash.
14. Investors for new positions may use the model predictions to schedule their market entry better and have expectations of price movements to anticipate in the future. Investors with existing positions could determine whether to wait for the rebound. Lastly, governmental institutions could use the model to plan the economic impact of the crash and determine if more drastic measures compared to the historical ones should be taken to mitigate the downturn impacts quicker.

List of References

1. Gerard Caprio and Daniela Klingebiel. *Bank Insolvencies: Cross-Country Experience.* The World Bank, Nov 30, 1999 Available from: http://elibrary.worldbank.org/doi/book/10.1596/1813-9450-1620 DOI 10.1596/1813-9450-1620.
2. HOMAR, T. and VAN WIJNBERGEN, S.J.G. Bank Recapitalization and Economic Recovery After Financial Crises. *Journal of Financial Intermediation*, Oct 01, 2017, vol. 32. pp. 16-28. Available from: https://dx.doi.org/10.1016/j.jfi.2016.11.004 CrossRef. ISSN 1042-9573. DOI 10.1016/j.jfi.2016.11.004.
3. MISHKIN, F.S. and WHITE, E.N. *U.S. Stock Market Crashes and their Aftermath: Implications for Monetary Policy.* Cambridge, Mass: National Bureau of Economic Research. Jun 01, 2002 Available from: http://www.nber.org/papers/w8992 Business Premium Collection. ISBN 0898-2937. DOI 10.3386/w8992.
4. PATEL, S.A. and SARKAR, A. Crises in Developed and Emerging Stock Markets. *Financial Analysts Journal*, 1998, vol. 54, no. 6. pp. 50-59. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-0002111017&doi=10.2469%2ffaj.v54.n6.2225&partnerID=40&md5=ee2725fffb3f1bcbe3033b581d21d8ef Scopus. ISSN 0015-198X(ISSN). DOI 10.2469/faj.v54.n6.2225.
5. National Bureau of Economic Research. *Business Cycle Dating.* Available from:https://www.nber.org/research/business-cycle-dating.
6. BRAUN, M. and LARRAIN, B. Finance and the Business Cycle: International, Inter-Industry Evidence. *The Journal of Finance (New York)*, Jun, 2005, vol. 60, no. 3. pp. 1097-1128. Available from: https://api.istex.fr/ark:/67375/WNG-SG1GC0KH-M/fulltext.pdf CrossRef. ISSN 0022-1082. DOI 10.1111/j.1540-6261.2005.00757.x.
7. BRUNNERMEIER, M.K. and OEHMKE, M. Bubbles, Financial Crises, and Systemic Risk. *National Bureau of Economic Research*, September, 2012 [viewed March 22, 2023]. Available from: https://www.nber.org/system/files/working_papers/w18398/w18398.pdf DOI 10.3386/w18398.
8. REINHART, C.M. and ROGOFF, K.S. *This Time is Different: Eight Centuries of Financial Folly.* Princeton: Princeton University Press, 2009 ISBN 9781400831722.
9. FATÁS, A. and MIHOV, I. *Recoveries.* CEPR Discussion Paper No. DP9551, 17 July 2013.
10. LAEVEN, L. and VALENCIA, F. Systemic Banking Crises Database. *MF Econ Rev 61*, 26 June, 2013. pp. 225-270. Available from: https://link.springer.com/article/10.1057/imfer.2013.12#citeas DOI https://doi.org/10.1057/imfer.2013.12.
11. STIGLITZ, J.E. Reforming the Global Economic Architecture: Lessons from Recent Crises. *The Journal of Finance (New York)*, Aug, 1999, vol. 54, no. 4. pp. 1508-1521. Available from: https://api.istex.fr/ark:/67375/WNG-1P33K1MR-Q/fulltext.pdf CrossRef. ISSN 0022-1082. DOI 10.1111/0022-1082.00154.
12. KANNAN, P., SCOTT, A. and TERRONES, M. *From Recession to Recovery: How Soon and how Strong?* , Jan 01, 2014 Available from: https://search.proquest.com/docview/1863199187 EconLit. ISBN 9781475543407.
13. ZHANG, D., HU, M. and JI, Q. Financial Markets Under the Global Pandemic of COVID-19. *Finance Research Letters*, 2020, vol. 36. pp. 101528. Available from: https://www.sciencedirect.com/science/article/pii/S1544612320304050 ISSN 1544-6123. DOI 10.1016/j.frl.2020.101528.
14. LO DUCA, M. *A New Database for Financial Crises in European Countries: ECB/ESRB EU Crises Database.* SSRN, 2017 Available from:

http://www.econis.eu/PPNSET?PPN=1791275982 CrossRef. ISBN 1556-5068. DOI 10.2139/ssrn.3011859.
15. FAIR, R.C. Events that Shook the Market. *The Journal of Business*, 2002, vol. 75, no. 4. pp. 713-731. Available from: http://www.jstor.org/stable/10.1086/341640 JSTOR. ISSN 0021-9398,15375374. DOI 10.1086/341640.
16. ALLEN, F. and GORTON, G. Churning Bubbles. *The Review of Economic Studies*, 1993, vol. 60, no. 4. pp. 813-836. Available from: http://www.jstor.org/stable/2298101 JSTOR. ISSN 0034-6527,1467937X. DOI 10.2307/2298101.
17. MUIR, T. Financial Crises and Risk Premia. *The Quarterly Journal of Economics*, May 01, 2017, vol. 132, no. 2. pp. 765-809. Available from: https://www.jstor.org/stable/26495149 CrossRef. ISSN 0033-5533. DOI 10.1093/qje/qjw045.
18. RAFIQ, S. How did House and Stock Prices Respond to Different Crisis Episodes since the 1870s?. *Economic Modelling*, Sep, 2022, vol. 114. pp. 105913. Available from: https://dx.doi.org/10.1016/j.econmod.2022.105913 CrossRef. ISSN 0264-9993. DOI 10.1016/j.econmod.2022.105913.
19. REINHART, C.M. and ROGOFF, K.S. Recovery from Financial Crises: Evidence from 100 Episodes. *American Economic Review*, May 01, 2014, vol. 104, no. 5. pp. 50-55. Available from: https://www.jstor.org/stable/42920909 CrossRef. ISSN 0002-8282. DOI 10.1257/aer.104.5.50.
20. PAPELL, D.H. and PRODAN, R. The Statistical Behavior of GDP After Financial Crises and Severe Recessions. *The BE Journal of Macroeconomics*, 2012, vol. 12, no. 3.
21. WAN, C. and JIN, Y. Output Recovery After Financial Crises: An Empirical Study. *Emerging Markets Finance and Trade*, 2014, vol. 50, no. 6. pp. 209-228. Available from: https://doi.org/10.1080/1540496X.2014.1011936 ISSN 1540-496X. DOI 10.1080/1540496X.2014.1011936.
22. BORDO, M.D. and HAUBRICH, J.G. Deep Recessions, Fast Recoveries, and Financial Crises: Evidence from the American Record. *Economic Inquiry*, 2017, vol. 55, no. 1. pp. 527-541.
23. GOETZMANN, W.N. and KIM, D. Negative Bubbles: What Happens After a Crash. *European Financial Management : The Journal of the European Financial Management Association*, Mar, 2018, vol. 24, no. 2. pp. 171-191. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/eufm.12164 CrossRef. ISSN 1354-7798. DOI 10.1111/eufm.12164.
24. FOO, J. and WITKOWSKA, D. A Comparison of Global Financial Market Recovery After the 2008 Global Financial Crisis. *Folia Oeconomica Stetinensia*, 2017, vol. 17, no. 1. pp. 109-128. Available from: https://doi.org/10.1515/foli-2017-0009 DOI 10.1515/foli-2017-0009.
25. AMBROSIUS, C. *What Explains the Speed of Recovery from Banking Crises?* Kiel und Hamburg: ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, 2016 [viewed March 21, 2023]. Available from: https://www.econstor.eu/handle/10419/145606.
26. HOWARD, G., MARTIN, R. and WILSON, B.A. Are Recoveries from Banking and Financial Crises really so Different?, 2011.
27. HOMAR, T. and VAN WIJNBERGEN, S.J.G. Bank Recapitalization and Economic Recovery After Financial Crises. *Journal of Financial Intermediation*, 2017, vol. 32. pp. 16-28. Available from: https://www.sciencedirect.com/science/article/pii/S1042957316300638 ISSN 1042-9573. DOI 10.1016/j.jfi.2016.11.004.

28. LAEVEN, L. and VALENCIA, F. The Real Effects of Financial Sector Interventions during Crises. *Journal of Money, Credit and Banking*, 2013, vol. 45, no. 1. pp. 147-177. Available from: http://www.jstor.org/stable/23320080 JSTOR. ISSN 0022-2879,15384616.
29. CHODOROW-REICH, G. The Employment Effects of Credit Market Disruptions: Firm-Level Evidence from the 2008–9 Financial Crisis. *The Quarterly Journal of Economics*, 2014, vol. 129, no. 1. pp. 1-59.
30. HORII, R. and ONO, Y. Financial Crisis and Slow Recovery with Bayesian Learning Agents. *International Journal of Economic Theory*, 2022, vol. 18, no. 4. pp. 578-606. Available from: https://doi.org/10.1111/ijet.12322 ISSN 1742-7355. DOI 10.1111/ijet.12322.
31. FAMA, E.F. Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance (New York)*, May, 1970, vol. 25, no. 2. pp. 383 CrossRef. ISSN 0022-1082. DOI 10.2307/2325486.
32. SELF, J. and MATHUR, I. Asymmetric Stationarity in National Stock Market Indices: An MTAR Analysis. *Journal of Business*, Nov 01, 2006, vol. 79, no. 6. pp. 3153-3174. Available from: https://www.jstor.org/stable/10.1086/505252 CrossRef. ISSN 0021-9398. DOI 10.1086/505252.
33. LO, A.W. Reconciling Efficient Markets with Behavioral Finance: The Adaptive Markets Hypothesis. *Journal of Portfolio Management*, 2005, vol. 7, no. 2. pp. 21-44. Available from: https://ssrn.com/abstract=1702447.
34. LIM, K. and BROOKS, R. The Evolution of Stock Market Efficiency Over Time: A Survey of the Empirical Literature. *Journal of Economic Surveys*, Feb, 2011, vol. 25, no. 1. pp. 69-108. Available from: https://api.istex.fr/ark:/67375/WNG-TFC4NLV9-V/fulltext.pdf International Bibliography of the Social Sciences (IBSS). ISSN 0950-0804. DOI 10.1111/j.1467-6419.2009.00611.x.
35. AN, Z., JALLES, J.T. and LOUNGANI, P. How Well do Economists Forecast Recessions?. *International Finance*, 2018, vol. 21, no. 2. pp. 100-121. Available from: https://doi.org/10.1111/infi.12130 ISSN 1367-0271. DOI 10.1111/infi.12130.
36. KUMBURE, M.M., LOHRMANN, C., LUUKKA, P. and PORRAS, J. Machine Learning Techniques and Data for Stock Market Forecasting: A Literature Review. *Expert Systems with Applications*, 2022, vol. 197. pp. 116659. Available from: https://www.sciencedirect.com/science/article/pii/S0957417422001452 ISSN 0957-4174. DOI 10.1016/j.eswa.2022.116659.
37. LOHRMANN, C. and LUUKKA, P. Classification of Intraday S&P500 Returns with a Random Forest. *International Journal of Forecasting*, 2019, vol. 35, no. 1. pp. 390-407. Available from: https://www.sciencedirect.com/science/article/pii/S0169207018301481 ISSN 0169-2070. DOI 10.1016/j.ijforecast.2018.08.004.
38. ABARBANELL, J.S. and BUSHEE, B.J. Abnormal Returns to a Fundamental Analysis Strategy. *The Accounting Review*, 1998, vol. 73, no. 1. pp. 19-45. Available from: http://www.jstor.org/stable/248340 JSTOR. ISSN 0001-4826.
39. BLUME, L., EASLEY, D. and O'HARA, M. Market Statistics and Technical Analysis: The Role of Volume. *The Journal of Finance*, 1994, vol. 49, no. 1. pp. 153-181.
40. NAZARETH, N. and RAMANA REDDY, Y.V. Financial Applications of Machine Learning: A Literature Review. *Expert Systems with Applications*, 2023, vol. 219. pp. 119640. Available from: https://www.sciencedirect.com/science/article/pii/S0957417423001410 ISSN 0957-4174. DOI 10.1016/j.eswa.2023.119640.
41. EFENDI, R., ARBAIY, N. and DERIS, M.M. A New Procedure in Stock Market Forecasting Based on Fuzzy Random Auto-Regression Time Series Model. *Information Sciences*, 2018, vol. 441.

pp. 113-132. Available from: https://www.sciencedirect.com/science/article/pii/S0020025518300975 ISSN 0020-0255. DOI 10.1016/j.ins.2018.02.016.
42. CHEN, Y. and HAO, Y. A Feature Weighted Support Vector Machine and K-Nearest Neighbor Algorithm for Stock Market Indices Prediction. *Expert Systems with Applications*, 2017, vol. 80. pp. 340-355. Available from: https://www.sciencedirect.com/science/article/pii/S0957417417301367 ISSN 0957-4174. DOI 10.1016/j.eswa.2017.02.044.
43. CEPNI, O., GÜNEY, I.E. and SWANSON, N.R. Nowcasting and Forecasting GDP in Emerging Markets using Global Financial and Macroeconomic Diffusion Indexes. *International Journal of Forecasting*, 2019, vol. 35, no. 2. pp. 555-572. Available from: https://www.sciencedirect.com/science/article/pii/S0169207018301985 ISSN 0169-2070. DOI 10.1016/j.ijforecast.2018.10.008.
44. FORONI, C., MARCELLINO, M. and STEVANOVIC, D. Forecasting the Covid-19 Recession and Recovery: Lessons from the Financial Crisis. *International Journal of Forecasting*, 2022, vol. 38, no. 2. pp. 596-612.
45. PARK, S. and YANG, J. Interpretable Deep Learning LSTM Model for Intelligent Economic Decision-Making. *Knowledge-Based Systems*, 2022, vol. 248. pp. 108907. Available from: https://www.sciencedirect.com/science/article/pii/S0950705122004348 ISSN 0950-7051. DOI 10.1016/j.knosys.2022.108907.
46. MINTARYA, L.N., et al. Machine Learning Approaches in Stock Market Prediction: A Systematic Literature Review. *Procedia Computer Science*, 2023, vol. 216. pp. 96-102. Available from: https://www.sciencedirect.com/science/article/pii/S1877050922021937 ISSN 1877-0509. DOI 10.1016/j.procs.2022.12.115.
47. MULLAINATHAN, S. and SPIESS, J. Machine Learning: An Applied Econometric Approach. *The Journal of Economic Perspectives*, 2017, vol. 31, no. 2. pp. 87-106. Available from: http://www.jstor.org/stable/44235000 JSTOR. ISSN 0895-3309.
48. WANG, W., LI, W., ZHANG, N. and LIU, K. Portfolio Formation with Preselection using Deep Learning from Long-Term Financial Data. *Expert Systems with Applications*, 2020, vol. 143. pp. 113042. Available from: https://www.sciencedirect.com/science/article/pii/S0957417419307596 ISSN 0957-4174. DOI 10.1016/j.eswa.2019.113042.
49. YANG, S. A Novel Study on Deep Learning Framework to Predict and Analyze the Financial Time Series Information. *Future Generation Computer Systems*, 2021, vol. 125. pp. 812-819. Available from: https://www.sciencedirect.com/science/article/pii/S0167739X2100279X ISSN 0167-739X. DOI 10.1016/j.future.2021.07.017.
50. GOODFELLOW, I., BENGIO, Y. and COURVILLE, A. *Deep Learning*. The MIT Press, 2016 ISBN 9780262337373.
51. OZBAYOGLU, A.M., GUDELEK, M.U. and SEZER, O.B. Deep Learning for Financial Applications: A Survey. *Applied Soft Computing*, 2020, vol. 93. pp. 106384.
52. DUBEY, S.R., SINGH, S.K. and CHAUDHURI, B.B. Activation Functions in Deep Learning: A Comprehensive Survey and Benchmark. *Neurocomputing*, 2022, vol. 503. pp. 92-108. Available from: https://www.sciencedirect.com/science/article/pii/S0925231222008426 ISSN 0925-2312. DOI 10.1016/j.neucom.2022.06.111.
53. ORIMOLOYE, L.O., SUNG, M., MA, T. and JOHNSON, J.E.V. Comparing the Effectiveness of Deep Feedforward Neural Networks and Shallow Architectures for Predicting Stock Price Indices. *Expert Systems with Applications*, 2020, vol. 139. pp. 112828. Available from:

https://www.sciencedirect.com/science/article/pii/S0957417419305305 ISSN 0957-4174. DOI 10.1016/j.eswa.2019.112828.
54. RANA, M., UDDIN, M.M. and HOQUE, M.M. *Effects of Activation Functions and Optimizers on Stock Price Prediction using LSTM Recurrent Networks.* , 2019.
55. GÉRON, A. *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow.* " O'Reilly Media, Inc.", 2022.
56. MASTERS, D. and LUSCHI, C. Revisiting Small Batch Training for Deep Neural Networks. *arXiv Preprint arXiv:1804.07612*, 2018.
57. SMITH, L.N. A Disciplined Approach to Neural Network Hyper-Parameters: Part 1--Learning Rate, Batch Size, Momentum, and Weight Decay. *arXiv Preprint arXiv:1803.09820*, 2018.
58. SRIVASTAVA, N., et al. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *The Journal of Machine Learning Research*, 2014, vol. 15, no. 1. pp. 1929-1958.
59. LECUN, Y., BENGIO, Y. and HINTON, G. Deep Learning. *Nature*, 2015, vol. 521, no. 7553. pp. 436-444. Available from: https://doi.org/10.1038/nature14539 ISSN 1476-4687. DOI 10.1038/nature14539.
60. LUNDBERG, S.M. and LEE, S. A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*, 2017, vol. 30.
61. HOSEINZADE, E. and HARATIZADEH, S. CNNpred: CNN-Based Stock Market Prediction using a Diverse Set of Variables. *Expert Systems with Applications*, Sep 01, 2019, vol. 129. pp. 273-285. Available from: https://dx.doi.org/10.1016/j.eswa.2019.03.029 CrossRef. ISSN 0957-4174. DOI 10.1016/j.eswa.2019.03.029.
62. SHAH, J., VAIDYA, D. and SHAH, M. A Comprehensive Review on Multiple Hybrid Deep Learning Approaches for Stock Prediction. *Intelligent Systems with Applications*, 2022. pp. 200111.
63. BALL, C. and FRENCH, J. Exploring what Stock Markets Tell Us about GDP in Theory and Practice. *Research in Economics*, 2021, vol. 75, no. 4. pp. 330-344. Available from: https://www.sciencedirect.com/science/article/pii/S1090944321000399 ISSN 1090-9443. DOI 10.1016/j.rie.2021.09.002.
64. MSCI. *MSCI World Index.* MSCI [viewed Mar 27, 2023]. Available from: https://www.msci.com/documents/10199/178e6643-6ae6-47b9-82be-e1fc565ededb.
65. Lewinson Eryk. *Python for Finance Cookbook: Over 80 Powerful Recipes for Effective Financial Data Analysis.* 2nd ed. Packt Publishing, December 30, 2022 ISBN 1803243198.
66. CUNADO, J., GIL-ALANA, L.A. and GRACIA, F.P.d. Mean Reversion in Stock Market Prices: New Evidence Based on Bull and Bear Markets. *Research in International Business and Finance*, 2010, vol. 24, no. 2. pp. 113-122. Available from: https://www.sciencedirect.com/science/article/pii/S0275531909000397 ISSN 0275-5319. DOI 10.1016/j.ribaf.2009.10.001.
67. KIM, J.H., SHAMSUDDIN, A. and LIM, K. Stock Return Predictability and the Adaptive Markets Hypothesis: Evidence from Century-Long U.S. Data. *Journal of Empirical Finance*, 2011, vol. 18, no. 5. pp. 868-879. Available from: https://www.sciencedirect.com/science/article/pii/S0927539811000612 ISSN 0927-5398. DOI 10.1016/j.jempfin.2011.08.002.
68. ASRAV, T. and AYDIN, E. Physics-Informed Recurrent Neural Networks and Hyper-Parameter Optimization for Dynamic Process Systems. *Computers & Chemical Engineering*, 2023, vol. 173. pp. 108195. Available from: https://www.sciencedirect.com/science/article/pii/S0098135423000649 ISSN 0098-1354. DOI 10.1016/j.compchemeng.2023.108195.

69. HOPFIELD, J.J. Neural Networks and Physical Systems with Emergent Collective Computational Abilities. *Proceedings of the National Academy of Sciences*, 1982, vol. 79, no. 8. pp. 2554-2558. Available from: https://doi.org/10.1073/pnas.79.8.2554 DOI 10.1073/pnas.79.8.2554.
70. HOCHREITER, S. and SCHMIDHUBER, J. Long Short-Term Memory. *Neural Computation*, Nov 15, 1997, vol. 9, no. 8 [viewed 3/25/2023]. pp. 1735-1780. Available from: https://direct.mit.edu/neco/article/doi/10.1162/neco.1997.9.8.1735 MEDLINE. ISSN 0899-7667. DOI 10.1162/neco.1997.9.8.1735.
71. FISCHER, T. and KRAUSS, C. Deep Learning with Long Short-Term Memory Networks for Financial Market Predictions. *European Journal of Operational Research*, 2018, vol. 270, no. 2. pp. 654-669. Available from: https://www.sciencedirect.com/science/article/pii/S0377221717310652 ISSN 0377-2217. DOI 10.1016/j.ejor.2017.11.054.
72. CHO, K., et al. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. *arXiv Preprint arXiv:1406.1078*, 2014.
73. HAASE, F. and NEUENKIRCH, M. Predictability of Bull and Bear Markets: A New Look at Forecasting Stock Market Regimes (and Returns) in the US. *International Journal of Forecasting*, 2023, vol. 39, no. 2. pp. 587-605. Available from: https://www.sciencedirect.com/science/article/pii/S0169207022000048 ISSN 0169-2070. DOI 10.1016/j.ijforecast.2022.01.004.
74. HAMILTON, J.D. A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, 1989, vol. 57, no. 2. pp. 357-384. Available from: http://www.jstor.org/stable/1912559 JSTOR. ISSN 0012-9682,14680262. DOI 10.2307/1912559.
75. ANG, A. and TIMMERMANN, A. Regime Changes and Financial Markets. *Annual Review of Financial Economics*, 2012, vol. 4, no. 1. pp. 313-337. Available from: https://doi.org/10.1146/annurev-financial-110311-101808 ISSN 1941-1367. DOI 10.1146/annurev-financial-110311-101808.
76. ZENS, G. and BÖCK, M. A Factor-Augmented Markov Switching (FAMS) Model. *arXiv Preprint arXiv:1904.13194*, 2019.
77. *Markov Switching Autoregression Models — Statsmodels.* [viewed Apr 1, 2023]. Available from: https://www.statsmodels.org/dev/examples/notebooks/generated/markov_autoregression.html.
78. BEN-DAVID, I., FRANZONI, F., KIM, B. and MOUSSAWI, R. Competition for Attention in the ETF Space. *The Review of Financial Studies*, Aug 04, 2022, vol. 36, no. 3. pp. 987 CrossRef. ISSN 0893-9454. DOI 10.1093/rfs/hhac048.
79. PETERSON, B.G., et al. *PerformanceAnalytics: Econometric Tools for Performance and Risk Analysis in R.* , 2020.
80. DING, L. and PU, X. Market Linkage and Information Spillover: Evidence from Pre-Crisis, Crisis, and Recovery Periods. *Journal of Economics and Business*, 2012, vol. 64, no. 2. pp. 145-159. Available from: https://www.sciencedirect.com/science/article/pii/S0148619511000567 ISSN 0148-6195. DOI 10.1016/j.jeconbus.2011.11.001.
81. *Commodity Markets, The World Bank.* Available from:https://www.worldbank.org/en/research/commodity-markets.
82. Robert Shiller. *Online Data Robert Shiller.* Available from:http://www.econ.yale.edu/~shiller/data.htm.
83. *Board of Governors of the Federal Reserve System (US), 3-Month Treasury Bill Secondary Market Rate, Discount Basis [TB3MS].* Available from:https://fred.stlouisfed.org/series/TB3MS.

84. *Board of Governors of the Federal Reserve System (US), Market Yield on U.S. Treasury Securities at 10-Year Constant Maturity, Quoted on an Investment Basis [GS10].* Available from:https://fred.stlouisfed.org/series/GS10.
85. JEREZ, T. and KRISTJANPOLLER, W. Effects of the Validation Set on Stock Returns Forecasting. *Expert Systems with Applications*, 2020, vol. 150. pp. 113271. Available from: https://www.sciencedirect.com/science/article/pii/S0957417420300968 ISSN 0957-4174. DOI 10.1016/j.eswa.2020.113271.
86. PEIXEIRO, M. *Time Series Forecasting in Python.* 1st ed. Manning, August, 2022 ISBN 9781617299889.
87. BOTCHKAREV, A. A New Typology Design of Performance Metrics to Measure Errors in Machine Learning Regression Algorithms. *Interdisciplinary Journal of Information, Knowledge, and Management*, Jan 01, 2019, vol. 14. pp. 45-76. Available from: https://search.proquest.com/docview/2616630522 CrossRef. ISSN 1555-1229. DOI 10.28945/4184.
88. Keras. *Keras Documentation: Adam.* Available from:https://keras.io/api/optimizers/adam/.