August 22, 2024

UNIVERSITY OF TWENTE.

EY

Thesis

One-step ahead forecasting IRSS using a hybrid approach: Combining time series models with machine learning models

University Supervisors	*Company Supervisors*
Prof. Dr. L. Spierdijk	J. van Lammerts Bueren
Dr. B. Roorda	T. Rikken

D. Broekman | s2101580 | Financial Engineering and Management

Colophon

Management:
Behavioural, Management and Social Sciences
Industrial Engineering and Management

Date:
August 22, 2024

Reference:
Master's Thesis IEM

Version:
1.0

Status:
Final

Project:
Master's Thesis IEM

Student Number:
s2012580

Author:
D. Broekman

Email:
davebroekman@gmail.com

Postal Adress:
P.O. Box 217
7500 AE Enschede

Website:
www.utwente.nl

Abstract

This study researches the forecasting accuracy of Interest Rate Swap Spreads (IRSS) using both econometric models and machine learning techniques, with a focus on integrating these approaches in a hybrid model to enhance predictive accuracy. Given the important role of IRSS for financial institutions, accurate forecasting is essential for effective risk management. Our research evaluates the performance of the Hull-White Two-Factor (HW2F) model, the Long Short-Term Memory (LSTM) networks, and introduces an integrative hybrid model combining the strengths of these two individual models.

Through an elaborate literature review, we have identified the important predictors influencing IRSS including financial predictors - such as the zero-coupon bond yield, the Treasury yield and the TED Spread - and macroeconomic predictors, such as the gross domestic product, the inflation rate and the unemployment rate. Our methodology involved data preprocessing strategies - including stationarity tests, normalisation and outlier adjustment - to ensure the quality of our dataset.

The HW2F model demonstrated high out-of-sample forecasting accuracy with a Root Mean Squared Error (RMSE) of 2.298, a Mean Absolute Error (MAE) of 1.550, and a Mean Absolute Percentage Error (MAPE) of 3.118%. It economically outperformed a naive model for forecasting IRSS. The LSTM model showed it is prone to overfitting risks and struggled in high volatility market conditions, with out-of-sample RMSE of 3.708, MAE of 2.571, and MAPE of 5.041%. The hybrid model, designed to leverage the HW2F's stability and the LSTM's pattern recognition capabilities using residual correction, showed promise in reducing forecast volatility. However, it did not consistently outperform the HW2F model in out-of-sample IRSS forecasting, with out-of-sample RMSE of 2.680, MAE of 2.131, and MAPE of 4.123%.

Our findings highlight the importance of balancing model complexity and stability in financial forecasting. Additionally, we recommend future research to research further model enhancements and improve the model interpretability, for example by Explainable Artificial Intelligence (XAI) techniques, ensuring that advanced models can provide transparent insights for financial institutions. This study highlights the continued relevance of traditional econometric models and explores the potential of integrating machine learning for improved financial forecasting.

Keywords: *Interest Rate Swap Spread, Financial Forecasting, Risk Management, Hull-White Two-Factor Model, Long Short-Term Memory Networks, Hybrid Model*

Acknowedgments

Dear Reader,

With this thesis, the journey of studying Industrial Engineering and Management with a specialisation in Financial Engineering and Management, will come to an end. This study and this research have been an incredible journey, in which I explored several topics in the financial industry. Not only will I leave the university with the gained knowledge from the courses, but also with a very great passion for financial engineering. This thesis was therefore an opportunity for me to apply all the knowledge and passion that I have gained over the past few years.

Firstly, I want to show my gratitude to my academic supervisors from the University of Twente, Laura Spierdijk and Berend Roorda. Their invaluable feedback over the past months has been of great help. Especially I appreciate the time and effort Laura has dedicated in assisting me with her guidance and feedback, which enhanced the quality and the structure of my thesis.

Furthermore, I am deeply thankful to my supervisors at EY, Jiri Lammerts van Bueren and Thom Rikken. All the practical insights from Jiri's professional expertise have not only guided me to become part of the great culture within EY, but will even help me in my future career. Additionally, the personal guidance of Thom has brought me a lot of motivation and enthusiasm throughout this period, as we met two or three times per week. Thank you for pushing my performance to a higher level, by giving your professional advice, and above all by sparkling my enthusiasm in the office.

Lastly, I would like to thank my family and friends for their continuous support throughout this whole project. Their moral support has helped me through a tough but inspiring phase of my study. Lastly, I would like to say a special thanks to my fellow interns at EY, who especially understood the ups and downs we all had gone through during this graduation, and who, after 6 months of working together, have now become friends.

Once again, I express my gratitude to all those who have supported me. I hope you enjoy reading this thesis! Sincerely,

Dave Broekman
Amsterdam, August 22, 2024

Abbreviations

ADF	Augmented Dickey-Fuller
ARIMA	Autoregressive Integrated Moving Average
CNN	Convolutional Neural Networks
CDS	Credit Default Swap
DOC	Direct Output Combination
EY	Ernst & Young
FNN	Feedforward Neural Networks
IQR	Interquartile Range
IRSS	Interest Rate Swap Spread
GARCH	Generalised Autoregressive Conditional Heteroskedasticity
GBM	Gradient Boosting Machines
GDP	Gross Domestic Product
HW1F	Hull-Whiter One-Factor
HW2F	Hull-White Two-factor
IRRBB	Interest Rate Risk in the Banking Book
LIBOR	London Interbank Offered Rate
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
PP	Parameter Prediction
PV	Present Value
NN	Neural Networks
OTC	Over The Counter
QQ	Quantile-Quantile
repo	Repurchase Agreement
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Networks
RC	Residual Correction
SDE	Stochastic Differential Equation
SS	Swap Spread
T	Treasury yield
tanh	Hyperbolic Tangent
TB	Treasury Bill rate
TEDS	TED-Spread
VAR	Vector Autoregression
XAI	Explainable Artificial Intelligence

Contents

List of Figures

List of Tables

1 Introduction

In this section, we will introduce the problem that will be covered throughout this thesis. First, we will go over some general background information on risk management in Section 1.1. After that, we will elaborate on the more specific context on the subject of Interest Rate Swap Spreads (IRSS) in Section 1.2, and we will identify the core problem of this research in Section 1.3. After that, we will define our research objective and research questions in Section 1.4. Lastly, we will cover the research design for this study in Section 1.5.

1.1 Background

Risk management in financial institutions consists out of identifying, assessing, monitoring and controlling risks to ensure the institution's stability, profitability and compliance with regulatory requirements. Risk management is a vital process for the banking sector due to the risks involved in banking activities, which range from lending, investing and trading large amounts of finances.

At the core of risk management is the trade-off between risks and capital requirements. Central banks require the banks to hold a certain amount of capital, known as the regulatory capital, to cover the risks they take on. Regulatory capital serves as a buffer to absorb sudden losses and protect depositors and the financial system in case of crises. However, the bigger the regulatory requirements for the bank, the less money can be used in financing processes, and hence, it is important for banks to keep their safety buffer as close to the regulatory requirements as possible. Hence, financial institutions are constantly focusing on forecasting risk types, to be able to cope with the ever-evolving landscape of the markets.

1.2 Problem Context

In the ever-evolving landscape of financial markets, swaps stand out as one of the most versatile and widely used derivative instruments. These complex financial contracts allow parties to exchange streams of cash flows over a specified period, catering for a broad spectrum of risk management, speculative and arbitrage needs. By transforming short-term deposits into longer-term liabilities, institutions can manage their exposure to fluctuations in rates effectively (McNulty 1990). The invention of swaps can be traced back to the early 1980s, and since then, they have grown exponentially in both volume and complexity, becoming integral to the global system (Smith Jr, Smithson, and Wakeman 1988). This expansion reflects the increasing need among financial institutions to hedge against the risks inherent in their operations. In Figure 1.1, it is depicted that interest rate swaps constitute a significant proportion of the interest rate derivative Over The Counter (OTC) market. In April 2022, the daily turnover in EUR swaps equalled $1.3 trillion (TBIS 2022). Because interest rate swaps are traded on the OTC market and not on the exchange market, the interest rate swap transactions are directly dealt with between two parties and can be customised to the needs of the participants.

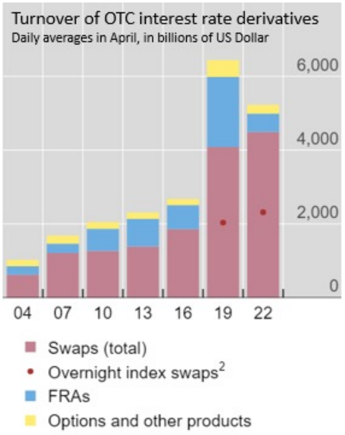


Figure 1.1: Interest Rate Derivatives Market Size in 2022 (TBIS 2022)

The mechanics of an interest rate swap transaction involve two parties agreeing to exchange cash flow streams over a set period (Hull 2012). These cash flows are often determined by different interest rates or currencies. For example, in an interest rate swap, one party might agree to pay a fixed interest rate on a notional principal amount, while receiving a floating rate from the party. In Figure 1.2, a case is visualised in which two companies have entered an interest rate swap agreement, in which Company A pays a fixed rate of interest of 3% and receives the floating London Interbank Offered Rate (LIBOR). The actual principal amounts typically do not change hands, instead the net difference in interest payments is exchanged. If Company A expects fluctuations in the interest rate, and it wants to hedge the risks that are a consequence of those fluctuations, the company would enter such a swap deal. To do so, the company must now pay a premium, which is typically above the initial LIBOR at the starting moment of the agreement, which is called the IRSS.

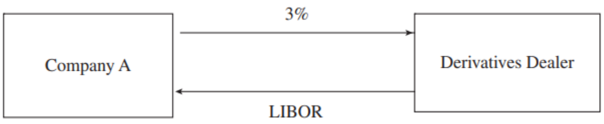


Figure 1.2: Example of an Interest Rate Swap (Hull 2012)

Swaps come in various forms, each designed to meet specific financial goals and manage different types of risks. The most common swap is a ”plain vanilla” interest rate swap where a fixed rate of interest is exchanged for LIBOR (McNulty 1990). Other swaps are currency swaps, where two parties help each other in hedging against currency risk; Commodity swaps where parties exchange floating and fixed rates to hedge against price volatilities of raw materials; and Credit Default Swaps (CDS), which provide insurance against risk of default by a third party (Hull 2012).

Swap spreads, particularly in the context of interest rate swaps, are a critical aspect of the swap market. They allow financial institutions to take care of interest rate risk management, and liquidity management and find arbitrage opportunities (Duarte, Longstaff, and Yu 2007). The swap rate is essentially the fixed rate that one party agrees to pay in exchange for receiving a floating rate, which is usually tied to a benchmark index such as LIBOR (Bicksler and A. H. Chen 1986). The determination of IRSS is influenced by several factors, including market conditions, market expectations, credit risks, and market supply and demand. If market participants are expecting economic growth or tighter monetary policy from central banks, there is expected to be a rise in the future interest rate (Tobin 1965). Rises in interest rates indicate rises in IRSS, ceteris paribus, and hence, the better the economy is expected to perform, the higher the IRSS are. On the contrary, when the economy is expected to fall, the central banks will step in and lower the interest rate, to increase market anticipation. Hence, IRSS will decrease in bad economic situations, ceteris paribus. However, when there is an increased risk of default in the market, this could imply a higher interest rate swap rate (Pereira 2015).

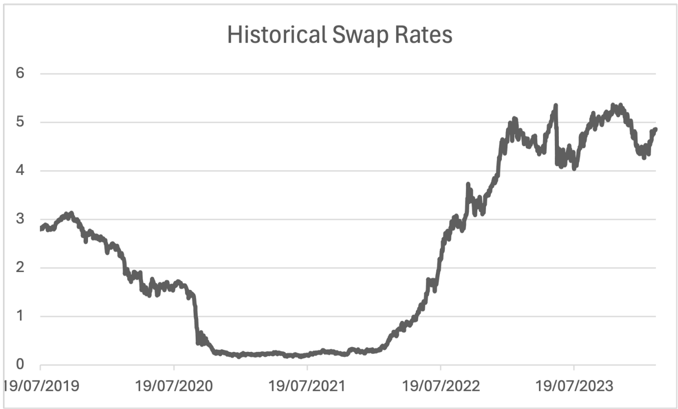


Figure 1.3: Historical Interest Rate Swap Rates (Investing.com 2024)

1.3 Core Problem

Swaps play a pivotal role in financial markets by providing a mechanism for managing risk and revealing information about market expectations on interest rates, currency movements, and credit risk (Bae, Karolyi, and Stulz 2003). Changes in IRSS are predictive of economic growth, inflation and monetary policy, ceteris paribus. The spread of a swap is calculated by the difference between the fixed rate and the floating rate, and since in most cases the floating rate equals the LIBOR, the uncertainty lies in the premium that is paid for the fixed rate, which is a prediction of the risk in the market (Hull 1993). Thus IRSS have a big impact on borrowing costs, investment decisions, and risk management strategies for market participants. Moreover, swaps contribute to market efficiency by facilitating the transfer of risk to those parties best equipped to bear them, thereby enhancing liquidity and supporting financial stability.

As mentioned, changes in the IRSS are predictive of macroeconomic shifts, and hence, the importance of accurate forecasting rises. Traditionally, the forecasting for swaps relied heavily on econometric models. Models that are now being used are Autoregressive

Integrated Moving Average (ARIMA) models and Generalised Autoregressive Conditional Heteroskedasticity (GARCH) models (Kontopoulou et al. 2023). These are complex mathematical models, as there are multiple factors influencing IRSS, which are often interrelated, which makes it very hard to forecast (Hull 2012). However, as the market for swaps has grown exponentially, so has the complexity of these instruments and the environments in which they are traded. This complexity necessitates advanced methodologies for accurately forecasting swaps.

Machine learning offers relatively new methods for the forecasting of IRSS, thanks to its ability to analyse larger datasets and recognise complex (non-linear) patterns in high dimensional spaces. Currently, financial institutions are exploring all possibilities with the help of machine learning in risk management (ING 2022; Rabobank 2022). Furthermore, new studies have found hybrid approaches between econometric and machine learning models, to improve the forecasting performance (Batool, Ahmed, and Ismail 2022).

The effects of the COVID-19 pandemic on the interest rate swap market were significant, and led to large fluctuations in the IRSS, as is visible in Figure 1.3, ceteris paribus. Market liquidity and functioning deteriorated, and there was an overall effect of investors unwinding their positions abruptly (Inoue, Miki, and Gemma 2021). Although a turmoil the size of COVID cannot be predicted accurately, there is an increased need for accurate forecasting within the interest rate swap market. The core problem that will be tackled during this research is:

”Literature has shown that there currently is a lack of forecasting accuracy using econometric models for IRSS forecasting.”

1.4 Research Objectives

The aim of this research is to investigate the IRSS forecasting performance of econometric and machine learning models. The goal of this is to be able to find the accurate characteristics of traditional forecasting and to combine them with the accurate characteristics of the machine learning model. In recent studies, automated machine learning methods to build time-series forecasting, Long-Short Term Memory (LSTM) deep learning were analysed together, and multiple time series and econometric models were tested, finding limitations to each model that was considered (Batool, Ahmed, and Ismail 2022; Do, Lakew, Sungrae Cho, et al. 2022; He et al. 2023; Jagero, Mageto, and Mwalili 2023; Kontopoulou et al. 2023).

The goal of this thesis is to address the gap in the financial literature regarding IRSS, a subject with limited comprehensive research. While existing studies primarily utilise econometric models, this work will introduce a machine learning approach and additionally, a hybrid approach will be introduced, merging the two individual models to analyse one-step ahead forecasting of IRSS more accurately. Additionally, the practical contribution of this thesis includes enhancing risk management for financial institutions, aiding investors in forming better strategies, providing policymakers with deeper market insights, advancing financial modelling techniques, and increasing market efficiency (Do, Lakew, Sungrae Cho, et al. 2022; Kontopoulou et al. 2023).

Ernst & Young (EY) is a global consultant in assurance, tax, transaction and advisory. At EY, the understanding of interest rates is critical for quantitative advisory services. It helps with enhancing credit risk assessment in mortgage portfolios, accurately valuing interest rate derivatives, and managing Interest Rate Risk in the Banking Book (IRRBB). Predicting future interest rates enables precise calculations of the expected credit loss metrics — such as Probability of Default (PD), Loss Given Default (LGD), and Exposure at Default (EAD) — improves derivative forecasting and informs the impacts on the Economic Value of Equity (EVE) and Net Interest Income (NII) for IRRBB reporting. Hence, the ability to forecast IRSS underlies the understanding of interest rate risks, which will be fundamental in the advisory towards financial institutions.

1.4.1 Main research question

The objective of this study is to address the following research question to be able to solve the core problem:

How do the predictive capacities of traditional financial models contrast with those of machine learning algorithms in one-step-ahead forecasting IRSS, and would an integrative approach using a hybrid model enhance the overall forecasting accuracy?

1.4.2 Sub-research question

To create a foundation for the main research question, six sub-research questions are produced to guide a step-by-step approach throughout this research.

I *What factors are the predictors that influence the dynamics of IRSS?*

In the first part of the research, the predictors that influence IRSS will be explored. To be able to build the best forecasting method, certain economical features should be selected, based on the predicting ability of the features. To find these values, a comprehensive exploration of current literature will be performed. Macroeconomic predictors, such as inflation rates, GDP growth, and unemployment rates, together with financial metrics such as the LIBOR are expected to build the right foundation for this research, upon which the models can be built.

II *Which econometric models are currently being used in the one-step forecasting of IRSS, and which are identified as the most accurate in literature?*

To answer this sub-research question, the research will focus on studying academic literature for various econometric models focusing on applications and accuracy. The goal is to find a model with the highest expected forecasting accuracy.

III *What are the assumptions and the limitations of these econometric models in the forecasting of IRSS?*

This sub-research question will research the theoretical assumptions of the selected econometric model. Furthermore, we will evaluate the forecasting accuracy using performance measures, that are selected to represent the error of the forecast. By doing this, the limitations of the econometric model can be analysed.

IV *Which machine learning models are expected to perform with the highest forecasting accuracy in the one-step forecasting of IRSS?*

Using academic literature and similar forecasting studies, we will evaluate multiple machine learning models, focusing on their potential forecasting accuracy for the forecasting of IRSS. One model will be selected based on our findings.

V *What are the assumptions and the limitations of these machine learning models in the forecasting of IRSS?*

We will perform a similar study, where we will combine the theoretical assumptions of the machine learning models, with the limitations found within the selected performance measures. Hence, after the model-building phase, we will create insights into the weak points of using machine learning forecasting for IRSS.

VI *How can a hybrid model that integrates econometric and machine learning approaches be optimally designed for one-step forecasting IRSS?*

Finally, after building both the econometric and the machine learning models for one-step-ahead forecasting IRSS, we will have gathered all the assumptions and the limitations of the individual models. We will perform a literature study on how to combine the models in a hybrid approach, to potentially overcome these limitations, and find and build the right hybrid model, to increase the forecasting accuracy.

1.5 Research Design

This research will be split up into multiple sections, to create a comprehensive approach to answer the sub-questions related to forecasting IRSS. The goal of this research is to find a hybrid approach of econometric and machine learning forecasting of the IRSS, that out-performs currently used methods. To do so, this research will first conduct a literature study on the key predictors that influence the IRSS, and how those key predictors are captured by the econometric and the machine learning models. This includes research on how the IRSS is currently set in the market, to create a real understanding of the IRSS. Additionally, within the literature study research will be conducted on current econometric models that are being used, and what types of machine learning models are promising for the forecasting process of IRSS. Based on this literature review, the models and their key features will be selected, that will be used during the model-building phase. After the models have been built, sensitivity analysis and analyses will be performed to confirm the forecasting accuracy of the model. If the models have been proven to forecast accurately, the predicting performance will be analysed, by comparing the results of the models with the historical values of the IRSS. This will be done based on the performance measures found in the research. The advantages and the limitations of both models will then be analysed, after which this information will be assessed in the hybrid-model building phase, which incorporates the advantages of both models and leverages these advantages to outperform the individual models in the forecasting performance of the IRSS. Finally, a conclusion will be synthesised regarding the suitability of the hybrid model for forecasting the IRSS in the financial markets.

1.6 Summary

Risk management in financial institutions involves identifying, assessing, monitoring and controlling risks to ensure stability, profitability and compliance with regulatory requirements. Among the financial instruments used in risk management, swaps stand out as one of the most versatile and widely used. They allow parties to exchange streams of cash flows over a specific period, mitigating a broad range of risks. Swaps have grown exponentially in volume and complexity since their inception in the early 1980s, becoming integral to the global financial system. IRSS play a critical role in this market as they are predictive of economic growth, inflation, and monetary policy. Accurate forecasting of IRSS is, therefore, crucial. While econometric models have been used in forecasting, the increasing complexity of swaps necessitates the exploration of more advanced methodologies, such as machine learning. This research aims to investigate the IRSS forecasting performance of econometric and machine learning models and to explore whether their integration could enhance overall forecasting accuracy.

2 Literature Framework

In this section, we will provide a comprehensive exploration of existing literature, to provide a fundamental understanding of the concepts used during this study, and to identify the gaps that currently exist. We will perform this literature study using a traditional literature review method, involving the identification, selection and critical analysis of previous research that directly relates to our research questions and objectives mentioned in Section 1.4.

For our study, it is important to study the concept of IRSS. Hence, in Section 2.1, we will find existing literature on the relevance of IRSS, with the aim of building an understanding of the concepts that are important in forecasting IRSS. Thereafter, if a good base of understanding is built, we will focus on finding the right financial and economic predictors for IRSS in Section 2.2. Furthermore, we will evaluate the relevant models for IRSS to find the highest predictive accuracy. First, we evaluate econometric models in Section 2.3, after which we will research machine learning models in Section 2.4. Lastly, we will find all relevant literature on building a hybrid model in Section 2.5.

2.1 Measure of IRSS

Interest rate swaps are agreements between two institutions in which each party commits to make periodic payments to the other based on a predetermined amount of notional principal for a predetermined life, called maturity. Furthermore, the swap spread is defined as the difference between the fixed rate on a swap and a Treasury yield with a comparable maturity, ceteris paribus (Hull 1993). Along with the volatility of the Treasury yield, the volatility in the swap spread is very important to determine the risk value of the swap positions to both of the counterparties.

In the business world, corporations face price risk, which stems from unexpected fluctuations in the fixed interest rates of swaps. If a company has entered into a swap, and the market rate for the swaps drops, ceteris paribus, the value of the swap decreases. In this situation, the fixed rate payer is now stuck on an ”above-market” fixed rate for LIBOR recipience. This is known as negative mark-to-market value. It presents a loss that, while not yet realised through actual transactions, may need to be reported within the company’s financial statements. For the other party, these fluctuations will be represented in an unexpected profit of the same magnitude. However, this gain comes with its own risk, as the fixed-rate receiver becomes vulnerable to the possibility that the counterparty might default on the agreement (K. C. Brown, Van Harlow, and Smith 1991). Hence, any volatility in the swap spread directly impacts the value of the swap position for both parties.

Market makers, who hedge unmatched positions in their swap books with treasury securities in both cash and futures markets, this volatility creates somewhat different problems. Because unexpected movements in the treasury components of the swap’s fixed rate can offset with any of several available instruments, the remaining source of risk is an unexpected change in the swap spread (K. C. Brown, Van Harlow, and Smith 1991). Given that these market makers typically operate with very tight profit margins — often as slim as 0.05% — even minor variations in the swap spread of a few basis points can significantly impact their profitability, ceteris paribus. Hence, if there were a stable and predictable relationship between the swap spread and other market rates, market makers could devise more effective hedging strategies, known as cross-hedges, to reduce their financial exposure. Of course, any residual risk resulting from the market maker's inability to create such a hedge will be priced into its bid/offer spread and ultimately affect the cost to corporations using swaps in their interest rate risk management program.

2.1.1 Financial dynamics

Within this theoretical framework, one of our goals is to answer Sub-Research Question I, in which our aim is to find the accurate predictors of IRSS. To do so, we will first find literature on the financial approach of IRSS. Decomposing the financial approach will create insights into the predictors of IRSS.

If an interest rate swap is agreed upon, Company A agrees to receive a fixed interest rate from Company B, which is determined by adding a specific Treasury yield (T) to a Swap Spread (SS). In return, Company A commits to pay a variable interest rate, typically indexed to the LIBOR. The LIBOR furthermore reflects the additional credit and liquidity risk premiums over the risk-free rate.

$$\text{Net Settlement} = (T_N + SS_N - \text{LIBOR}) \times NP,$$

where NP stands for the notional principal, which is essentially the reference amount upon which the interest payments are calculated, as the NP is not exchanged during a swap agreement. For a swap agreement that spans N periods, the fixed rate (represented as $T_N + SS_N$) is intended to match the expected average LIBOR throughout the duration of the contract, as we have discussed above.

$$NP \times \sum_{i=1}^{N} \left(\frac{T_N + SS_N}{(1+Z_i)^i} \right) = NP \times \sum_{i=1}^{N} \left(\frac{E[\text{LIBOR}_{i-1}]}{(1+Z_i)^i} \right),$$

where Z_i are the zero-coupon discount rates corresponding to each settlement date and $E[\cdot]$ is the expectation operator. In this equation, the pure expectations theory implies that the fixed rate is determined by the market's expectations of future short-term LIBOR. Using the zero-coupon bond discount rates provides an accurate, market-reflective method of calculating the present value of future swap payments compared to using a uniform treasury rate.

$$T_N + SS_N = \frac{\sum_{i=1}^{N} \left(\frac{E[\text{LIBOR}_{i-1}]}{(1+Z_i)^i} \right)}{\sum_{i=1}^{N} \left(\frac{1}{(1+Z_i)^i} \right)} = E_N[\text{LIBOR}], \tag{2.1}$$

where $E_N[\cdot]$ denotes the expectations cover N periods. We can analyse the components of

LIBOR at each date by breaking it down into two elements (K. C. Brown, Van Harlow, and Smith 1991). First, the Treasury Bill (TB) rate for that date, represents the yield on short-term government debt securities, which contains virtually no credit risk. Furthermore, the TED Spread (TEDS), is the difference between the interest rates on a three-month TB, and a three-month LIBOR. Hence, this spread represents the credit and liquidity risks. This decomposition is expected as

$$T_N + SS_N = E_N[\text{TB} + \text{TEDS}] = E_N[\text{TB}] + E_N[\text{TEDS}] \tag{2.2}$$

The use of the TB in this model aligns with the short-term nature, matching the short-term benchmarks used for the floating LIBOR. This in return can be rearranged as follows.

$$SS_N = (E_N[\text{TB}] - T_N) + E_N[\text{TEDS}], \tag{2.3}$$

which implies that in the financial market, the swap spread should equal the difference between the expected future T-bill rates and the Treasury bond yield, plus the expected average of future TED spreads. Furthermore, according to the pure expectations theory, $E_N[TB]$ is expected to be the same as the yield for an N-period zero-coupon Treasury bond, represented as Z_N (Hejazi, Lai, and Yang 2000). Therefore, continuously reinvesting in one-period Treasury bills over N periods, each at its expected yield as initially forecasted, should yield the same compound rate of return as investing in an N-period zero-coupon Treasury bond with a yield of Z_N. Substituting this obtains

$$SS_N = (Z_N - T_N) + E_N[\text{TEDS}] \tag{2.4}$$

The equation indicates that the IRSS consists of two elements, the first element, $Z_N - T_N$, reflects the coupon bias found within the treasury yield curve (Caks 1977), this bias suggests that, depending on the yield curve's direction, a ten-year zero-coupon can yield higher or lower than a ten-year coupon-bearing bond of similar face value. Specifically an upward-sloping yield curve will typically result in a higher yield for the zero-coupon bond compared to the par-value bond, thereby increasing the IRSS, assuming other conditions remain constant. The second element affecting the IRSS is the anticipated future levels of the TED spread, summarised by its average value. An increase in this expected average would also lead to an increase in the IRSS (K. C. Brown, Van Harlow, and Smith 1991). However, it is important to note, that the future trajectories of LIBOR and TB rates may not move parallel to each other, leading to variable outcomes for the TED spread. Here as the swap matures, the longer-term Z_N become more relevant, reflecting broader economic expectations and providing more stability.

2.1.2 Macroeconomic dynamics

In the previous section, we have decomposed the financial approach of IRSS. However, this can be a significant simplification, as the IRSS is influenced by macroeconomic dynamics. One of the primary considerations is the costs incurred by market makers, especially in relation to hedging open positions. The repurchase agreement (repo) rate stands out as a critical factor. In a repo, a financial institution that owns securities agrees to sell the securities for a certain price and to buy them back at a later time for a slightly higher price. As the repo rate increases, ceteris paribus, so does the cost of hedging, leading market makers to adjust their bid/offer spread accordingly. This adjustment impacts the fixed rate paid on the swap, highlighting a significant role hedging costs play in the forecasting of longer-term IRSS.

Another vital element is the credit risk associated with swap dealers. Dealers with higher credit ratings tend to offer narrower bid/offer spreads, ceteris paribus, as they are perceived to be lower risk (Sun, Sundaresan, and C. Wang 1993). This difference in credit reputation among dealers directly affects IRSS forecasting. Furthermore, external market conditions, including economic uncertainties and shifts in central bank rates, also play a crucial role. These factors, alongside liquidity and credit spreads between different types of bonds, influence market expectations and, consequently, the forecasting of swaps.

Additionally, transaction costs are another factor influencing IRSS. Although swap markets generally feature narrower bid-offer spreads than interbank markets, higher transaction costs can lead to increased quoted yields, reflecting the elevated costs associated with the trade execution (Amihud and Mendelson 1986). Under simplifying assumptions — no transaction costs or default — swap forecasting theory implies arbitrage-free markets, which is not true in real-world (Sun, Sundaresan, and C. Wang 1993). Under simplifying assumptions (no transaction costs or default), swap forecasting theory implies that the arbitrage-free rate for a generic interest rate swap should equal the yield on a par bond with the same maturity, which is not the case in real-world swaps. Lastly, the relationship between IRSS and Treasury yield varies with maturity, generally widening as maturity increases.

2.1.3 Summary

In this section, we have discussed the importance of the forecasting of swaps and the swap spreads. In financial markets, corporations face price risk from fluctuations in fixed IRSS. If market rates decline, the value of existing swaps falls, leading to a negative mark-to-market value for the fixed-rate payer, who is stuck paying an above-market rate. Conversely, the fixed-rate receiver might see an unrealised gain but faces default risk from the counterparty. Market makers, hedging unmatched positions with treasury securities, face risks from volatility, particularly from changes in the IRSS. These changes can significantly affect their profitability, especially since they operate with tight profit margins.

Furthermore, we have decomposed the financial approach of IRSS, and by doing so, we have found that the IRSS stems from the coupon bias found within the Treasury yield curve ($Z_N - T_N$) and the anticipated future levels of the TED spread ($E_N[TEDS]$). However, real-world forecasting of swaps is more complex. It includes factors like hedging costs, credit risks of swap dealers, market conditions, economic uncertainties, central bank rate shifts, liquidity, and credit spreads. Furthermore, the relationship between IRSS and Treasury yields, which typically widens with maturity, adds another layer of complexity.

With this information, we will focus on the main predictors of IRSS in the following section. We will study the financial predictors that we have found, and will furthermore focus on macroeconomic predictors and the forecasting of swaps.

2.2 Predictors of IRSS

During this research, we will use multiple models to forecast IRSS. To be able to forecast, our models need predictors. Predictors for machine learning models are input features or variables that are used to train and make predictions (S. Wang et al. 2015). Predictors are selected based on their relevance to the target outcome, which in the case of our research is the IRSS. The quality of the predictors directly affects the performance of the models, making the process of feature selection a critical step in the model development process.

In general, the optimal number of predictors in machine learning depends on the complexity of the problem, the available data, and the specific architecture and training methods used (He et al. 2023). As in Section 2.1, we have found that IRSS is influenced by both financial and macroeconomic dynamics, and hence, this section will focus on finding the right predictors in these fields.

2.2.1 Financial predictors

In Section 2.1, we have concluded that within the financial decomposition of IRSS the coupon bias within the Treasury yield curve ($Z_N - T_N$), and the anticipated future levels of the TED spread ($E_N[TEDS]$) are important elements for forecasting. Hence the first predictors that we select are as follows:

- **Treasury yield curve**: Kurpiel (2003) has identified the slope of the yield curve as one of the variables potentially driving IRSS. As in a steep curve environment, the demand for the fixed-rate receiver in a swap increases, ceteris paribus. Hence, not only the position of the Treasury yield but additionally the slope will be considered. R. Brown, In, and Fang (2002) have found that the curvature of the yield curve does not have a significant contribution.

- **N-period zero-coupon bond**: Lekkos and Milas (2001) examine the ability of factors such as the level, volatility and slope of the zero-coupon government bond yield curve. They have found that the slope of the term structure has a significant counter-cyclical effect across maturities.

- **TED spread**: The TED spread, which represents the default risk premiums, however, has been found to play a small role and the significance varies across maturities (R. Brown, In, and Fang 2002), ceteris paribus, with a higher significance for longer maturities (Lekkos and Milas 2001).

2.2.2 Macroeconomic predictors

Furthermore, in Section 2.1, we have found that real-world forecasting of swaps is more complex. It includes factors like hedging costs, credit risks of swap dealers, market conditions, economic uncertainties, central bank rate shifts, liquidity, and credit spreads.

- **Gross domestic product**: The Gross Domestic Product (GDP) and IRSS are closely interconnected, with GDP having a significant impact on IRSS (Gargano and Timmermann 2014). GDP influences IRSS through its impact on monetary policy and market expectations, ceteris paribus. A growing GDP can lead to higher interest rates to control inflation, making IRSS more attractive.

- **Unemployment rate**: The unemployment rate impacts IRSS by influencing central bank policies and market sentiment (Gargano and Timmermann 2014). High unemployment typically leads to lower interest rates to encourage borrowing and investment, ceteris paribus, making IRSS more attractive.

- **Inflation rate**: The inflation rate directly influences IRSS by affecting central bank interest rate decisions and investor expectations (Beenstock and Chan 1988; S.-S. Chen 2009). High inflation often prompts central banks to raise interest rates to stabilise prices, ceteris paribus, making interest rate swaps more desirable to lick in than current rates.

2.2.3 Summary

In this section, we have focused on identifying predictors for forecasting IRSS using machine learning models. The selection of relevant predictors is crucial as it directly impacts model performance. Financial predictors identified include the Treasury yield curve's position and slope, the N-period zero-coupon bond's slope and the TED spread.

Macroeconomic predictors cover broader economic indicator swaps, such as GDP, which influences interest rates and swap attractiveness through monetary policy and market expectations. The unemployment rate impacts IRSS by guiding central bank policies and market sentiment, while inflation rate adjustment by central banks affects investor expectations and IRSS.

Concluding, within Section 2.2 and with the help of Section 2.1, we have answered Sub-Research Question I of our research. In the next section, we will study the literature surrounding Sub-Research Question II.

2.3 Econometric Models

In the following section, we will study traditional financial time-series models, in order to answer Sub-Research Question II. Traditional financial time-series models apply statistical and mathematical methods to the analysis of economic data. It aims to to give empirical content to economic theories to test hypotheses and forecast future trends. As mentioned in Section 1.2, many models have been used and evaluated in their forecasting accuracy of financial time-series (Duffie and Singleton 1997; Lekkos and Milas 2004; Sun, Sundaresan, and C. Wang 1993). In this section, we will discuss the relevant econometric models for our study, after which a decision can be made on which model will be used. The models that we will discuss are found in current literature, where they predominantly have been used in interest rate forecasting. As interest rate swaps are an interest rate-driven derivative, multiple studies have found that these models are therefore applicable to IRSS forecasting (Cortes 2003; Y. Huang, C. R. Chen, and Camacho 2008; K. Zhang and Liang 2008). This close relationship is underpinned by the fundamental mechanics of interest rate movements and their direct impact on the valuation of interest rate derivatives, such as IRSS, as we have found in Section 2.1.

2.3.1 Econometric models

- **Vector Autoregression:** Vector Autoregression (VAR) is a statistical model, used to capture the linear interdependencies among multiple time series. The VAR model generalises the autoregressive model by allowing for more than one evolving variable, all of which are interdependent (S. Kim and Roubini 2000). This method is widely used in econometrics and financial analysis for forecasting economic and financial variables.

  In VAR model, each variable is expressed as a linear combination of past values of itself and past values of all other variables in the system. This setup allows for the mutual influence of variables to be captured. A typical VAR model for a system of N variables looks like as follows:

  $$y_{i,t} = c_i + \sum_{j=1}^{n} \sum_{k=1}^{p} \beta_{ij,k} y_{j,t-k} + \varepsilon_{i,t}, \tag{2.5}$$

  where $y_{i,t}$ represents the value of the i-th variable at time t, c_i represents a constant term for the i-th equation, $\beta_{ij,k}$ are the coefficients measuring the value of the j-th variable's lagged value on the i-th variable, and $\varepsilon_{i,t}$ is the white noise error term. A big limitation of VAR is that there's a lack of theoretical guidance, as it treats all variables endogenously.

  When modelling interest rates, we will adapt the first formula as follows.

  $$r_t = c_1 + \sum_{j=1}^{n} \sum_{k=1}^{p} \beta_{1j,k} y_{j,t-k} + \varepsilon_{1,t},$$

  where $i = 1$, and the other individual predictors within the formula will be a version of Formula 2.5, where each predictor is represented as $y_{i,t}$, for each $i \geq 2$.

- **ARIMA:** The ARIMA model is a widely used statistical approach for forecasting and analysing time series data. ARIMA models are particularly popular in economics and finance for their applicability to a wide range of stationary and non-stationary series, such as stocks and economic indicators (Shumway et al. 2017).

The model consists of an integrated part, which involves differentiating the time series to achieve stationarity, an autoregressive part, which represents the current value of the time series as a linear combination of all its parameters, and a moving average part, where the current value of the series is modelled as a linear combination of its past forecast values. This makes the model very flexible, however, it does not fully incorporate external factors (Khan, Urooj, and Muhammadullah 2021).

$$(1 - \phi_1 B - ... - \phi_p B^p)(1 - B^d)y_t = (1 + \theta_1 B + ... + \theta_q B^q)\epsilon_t, \tag{2.6}$$

where B denotes the backshift operator, y_t represents the original time series, ϕ_i denotes the parameters in the autoregression, θ_i are the parameters in the moving average, and ϵ_t are the error terms. In the case of this study, within Formula 2.6 the interest rate would be represented as a substitute for y_t

$$(1 - \phi_1 B - ... - \phi_p B^p)(1 - B^d)r_t = (1 + \theta_1 B + ... + \theta_q B^q)\epsilon_t$$

- **GARCH:** The GARCH model is a statistical approach used extensively in financial econometrics to forecast financial time series, such as IRSS (Hull 2012). The general formula is as follows.

$$\sigma_n^2 = \gamma V_L + \sum_{i=1}^{p} \alpha_i u_{n-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{n-j}^2, \tag{2.7}$$

where V_L represents the long-run variance rate, σ represents the volatility and u represents the return rate. γ, α, and β are the weights given to each of the parameters, which together sum up to 1. Furthermore, p and q are the orders of the model specifying the number of lag terms. The most commonly used setup of the model is called GARCH(1,1) and replaces p and q with those values respectively.

The GARCH model makes use of mean reversion, in which the model tends to revert to the long-term average. Furthermore, the model also used volatility clustering, in which it captures the moment where high-volatility events tend to cluster together, which is a close representation of the financial market. However, this element affects the model as it assumes symmetric shocks to volatility.

- **Vasicek:** The Vasicek model is instrumental in the field of interest rate modelling. Its ability to predict interest rate movements is built upon predicting the "short-rate", which refers to the instantaneous interest rate at which an entity can borrow money for an infinitesimally short period. This model is central in the forecasting processes of interest rate derivatives, such as IRSS (Orlando, Mininni, and Bufalo 2020).

In the Vasicek model, the short-rate is assumed to follow a mean-reverting stochastic process, which means that the interest rate tends to revert towards a long-term average level over time. The mathematical representation of this behaviour is a Stochastic Differential Equation (SDE), which in the case of the Vasicek model is expressed as follows:

$$dr = a(b - r)dt + \sigma dz, \tag{2.8}$$

where a, b and σ are constants (Vasicek 1977). Here dr represents the change in the short-rate, a represents the speed of the mean reversion, b is the long-term mean level of the interest rate, σ is the volatility of the rate, and dz represents the increment of a Wiener Process, representing the random market risk factor. The disadvantage of this model is that the parameters are constant over time, which does not fully capture the actual market behaviour. Furthermore, the model can produce negative interest rates, which are not realistic in many financial scenarios.

- **Hull-White:** The Hull-White One Factor (HW1F) model is an extension to the Vasicek and the CIR models, to better accommodate the empirical characters of interest rates and fit the observed market data more accurately (Hull and White 2001). Hence, the formula is an evolution to Formula 2.8,

$$dr = [\theta - art]dt + \sigma dz, \tag{2.9}$$

where θ represents the time-dependent parameter designed to ensure the model fits the current yield curve. The HW1F model allows for zero and negative rates. It has proven to be very suitable for recent interest rate forecasting practices.

Furthermore, the Hull-White Two-Factor (HW2F) formula has evolved as the following:

$$df(r) = [\theta(t) + u - af(r)]dt + \sigma_1 dz_1, \tag{2.10}$$

where $f(r)$ is a function of interest rate r, and u has an initial value of zero, after which it follows the following process

$$du = -budt + \sigma_2 dz_2$$

This model provides a richer pattern of the yield curve movements and a richer pattern of volatilities than one-factor models of r (Hull 2012).

2.3.2 Comparison

In the literature, the HW2F has been found very effective for IRSS forecasting (Cortes 2003; Rodríguez, López, and Benedicto 2024; Seppälä, Poon, and Schröder 2013). Furthermore, based on the comparative analysis in Table 2.1, we find that HW2F emerges as the most accurate model for forecasting term structures and interest rate derivatives, such as IRSS. As IRSS are very complex interest rate derivatives, this model we will use this model during this study.

Table 2.1: Comparison of Econometric Models

Model	**Advantages**	**Disadvantages**	**Applications**
VAR	Captures relationships between multiple variables	Can lead to parameter overload; assumes linearity; lack of theoretical guidance	Multivariate time series analysis; Forecasting network of interrelated financial quantities
ARIMA	Flexible; effective for short-term forecasting	Assumes linearity; stationary requirement; does not incorporate external factors	Stock prices; Economic indicators
GARCH	Captures time-varying volatility and volatility clustering; suitable for conditional variance modelling	Can be complex to calibrate	Volatility forecasting; Risk management
Vasicek	Analytically tractable; mean-reverting interest rate model; foundational model	Can produce negative interest rates; simplistic assumptions	Interest rate derivative pricing; Risk management; Bond pricing
Hull-White	Flexible, allows for fitting to current term structures; mean reversion; captures interest rate movements	Potential complexity in parameter estimation	Interest rate derivative pricing; Asset-liability management

2.3.3 HW2F

In this Section, we will consider all relevant information about the HW2F model, that is needed in the model-building phase of this research. For the model-building phase, we will use the discretised equations of the continuous model mentioned in Formula 2.10. Hence the formulas that we will use are

$$f(r_{t+\Delta t}) = f(r_t) + [\theta(t) + u_t - af(r_t)]\Delta t + \sigma_1\sqrt{\Delta t}\varepsilon_1 \tag{2.11}$$

$$u_{t+\Delta t} = u_t + (-bu_t)\Delta t + \sigma_2\sqrt{\Delta t}\varepsilon_2 \tag{2.12}$$

The HW2F model allows for the correlation between the two stochastic processes, an important element when dealing with real-world interest rates that often correlate between different maturities. The correlation between dz_1 and dz_2 is defined as ρ, where $-1 \leq \rho \leq 1$ (Blanchard 2014). This ρ plays an important role in accurately capturing the interaction between short-term fluctuations and long-term economic trends. The correlation affects the shape and dynamics of the yield curve, realistically influencing how it reacts to changes in economic conditions.

Furthermore, as mentioned the θ_t represents the time-dependent mean reversion level of the short-rate. It ensures that the model fits the initial term structure of interest rates. The process θ_t typically considers elements such as current market conditions, the mean reversion speed and the volatility of the interest rates. To determine the θ_t we will use a bootstrapping method where we set the condition that the model prices of zero-coupon bonds match the observed prices at time 0 (Hull 1996). Given the zero-coupon bond price $P(0,t)$, which represents the future payments of the interest rate swap, at time 0 maturing at time t, the forward rate $f(0,t)$ is

$$f(0,t) = -\frac{\delta \ln P(0,t)}{\delta t} \tag{2.13}$$

Using this, the time-dependent drift term $\theta(t)$ can be derived as follows

$$\theta(t) = \frac{\delta f(0,t)}{\delta t} + af(0,t) + \frac{\sigma_1^2}{2a}(1-e^{-2at}) + \frac{\sigma_1\sigma_2\rho}{b-a}(e^{-at}-e^{-bt}) \tag{2.14}$$

2.3.4 Summary

During this section, we have performed research on various econometric models that focus on an application in financial time-series forecasting, such as the VAR, which is useful for finding interdependencies among multiple time series but lacks theoretical guidance, ARIMA, known for its flexibility and applicability to time-series data, however without capturing external factors, and GARCH, which functions well in capturing volatility but assumes symmetric volatility shocks.

Furthermore, interest rate models like the Vasicek model have been analysed, which is foundational yet sometimes too simplistic, together with the Hull-White models. The HW2F model can accurately account for complex market dynamics for interest rate derivatives. Consequently, this model has been selected for use during this study. With this information, we have found the answer to Sub-Research Question II. In the next section, we will elaborate on the existing literature surrounding Sub-Research Question IV. Furthermore, we have now built sufficient basis to research Sub-Research Question III in Section 4.1.

2.4 Machine Learning Models

In the following section, we will study machine learning models, in order to answer Sub-Research Question IV.

2.4.1 Machine learning

Machine learning can be performed in three categories of models (Bishop and Nasrabadi 2006). Supervised learning, in which training data comprises examples of input vectors along with their corresponding target vectors, unsupervised learning, in which the training data consists of a set of input vectors without corresponding target values, and reinforcement learning, which is concerned with the problem of finding suitable actions to take in a given situation in order to maximise the final reward. In our case, to forecast the behaviour of IRSS using machine learning, supervised learning is the most applicable. This is the case because supervised training uses historical IRSS and economic indicators to predict the future behaviour of IRSS.

Furthermore, within supervised machine learning, there are two discrete categories: Classification and regression. Classification is about aiming to assign each input vector to one of a finite number of discrete categories. Regression output consists of one or more continuous variables. During this study in forecasting IRSS, we are going to use models that have the function of creating regressions over the historical data.

There are many types of regressive supervised machine learning techniques. However, due to the special nature of IRSS, and their complex and non-linear relationships and interactions between variables, three models are suitable for our research.

- **Random forest:** A random forest algorithm operates by constructing multiple decision trees during training time and outputting the mean of the prediction from the individual trees. Each tree in the forest is built from a random sample of data, drawn with replacement, known as a bootstrap sample. This process introduces variability among the trees, which leads to a more robust and accurate model (Kumar and Thenmozhi 2006).

  The advantages of random forests are that they are very accurate in handling both linear and non-linear data and that there is great control of overfitting. However, there is a large usage of memory, as combining all individual trees can lead to slow performance and large model sizes.

- **Gradient boosting machines:** Gradient Boosting Machines (GBM) are powerful techniques used for regression tasks. GBM focuses on converting weak learners, such as decision trees, into strong collective models through an iterative process. The main principle is to build new models that progressively correct errors made by previous models. At each step, a new decision tree is added that predicts the residuals or errors of the entire sample. Instead of minimising the residuals in the space of the data, such as random forests, GBM minimises them in the space of the previous model's predictions, performing a gradient descent in the model's prediction space (Derbentsev et al. 2020).

  The advantage of GBM is that it has superior predictive accuracy compared to other machine learning models on structured data. Furthermore, it provides insights into the significance of each feature in the prediction process. However, it is very complex, as there are large numbers of hyperparameters, and the sequential nature of boosting can lead to very long training times compared to models that can.

- **Neural networks:** Neural Networks (NN) represent a framework in the field of machine learning and artificial intelligence. A typical NN consists of an input layer, one or more hidden layers, and an output layer. They process information by passing inputs through these interconnected layers, while each neuron applies a weighted sum of its inputs and passes the result through a non-linear activation function (Bishop and

Nasrabadi 2006).

The big advantage of NN is that they have got great understanding of forecasting complex non-linear relationships, as it can identify complex patterns in high-dimensional data. Furthermore, Furthermore, NN can handle very large datasets, making them suitable for big data applications and deep learning. Neural networks are less sensitive to error term assumptions and they can tolerate noise, chaotic components, and heavy tails better than most other methods (Kaastra and Boyd 1996). Other advantages include greater fault tolerance, robustness, and adaptability compared to expert systems due to the large number of interconnected processing elements that can be 'trained' to learn new patterns. However, the model is very prone to overfitting. Furthermore, they are criticised due to the black-box nature of their solutions.

Given these characteristics, some research shows that NN is considered optimal for forecasting IRSS (Roy et al. 2020). Recent advances in neural network technologies showed very promising results in comparison to other machine learning techniques, and hence, there is a lot of interest in the financial sector in exploring the possibilities in risk management (J. Huang, Chai, and Stella Cho 2020; Trippi and Turban 1992). Furthermore, in Section 2.2, we have identified ten financial and macroeconomic predictors for IRSS, and hence, our input data will be high-dimensional and very complex, which the NN will be able to cope with.

2.4.2 Neural networks

There are many types of NN and many studies have been performed on the subject of financial forecasting (Bishop and Nasrabadi 2006; Cui, W. Chen, and Y. Chen 2016; Jagero, Mageto, and Mwalili 2023; Kaastra and Boyd 1996; Trippi and Turban 1992; S. Wang et al. 2015). For our study, we will consider the most relevant neural network types, for optimal IRSS forecasting.

- **Feedforward Neural Network:** Feedforward Neural Networks (FNN) is a foundational architecture of NN and is seen as the most basic version (Fine 2006). FNN is mainly designed for recognition and classification tasks and is characterised by layers of neurons that process inputs and pass them forward, with each neuron receiving input from multiple predecessors, as visualised in Figure 2.1. FNN does not produce any feedback, as there are no loops within the network. Data thus flows from the input layer through one or more hidden layers to the output layer.

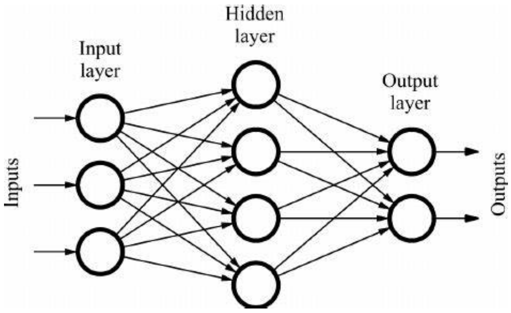


Figure 2.1: Feed Forward Neural Network (DeepAI 2019)

In comparison to other NNs, FNN is a very straightforward architecture, which makes it relatively easy to interpret. Furthermore, this has the effect that the speed of the model is faster, as there are no feedback loops, and hence, the model is able to converge faster than other models. However, FNNs do not have a memory mechanism, unlike other NNs. They treat each input independently without considering its relationship to previous or future inputs. This makes the FNN model less effective for time series analysis.

- **Convolutional Neural Network:** Convolutional Neural Networks (CNN) are a type of artificial neural network that has shown significant promise in financial forecasting. CNNs are particularly well-suited to this task due to their ability to learn and extract features from time series data, which can be used to predict future trends and data (Watson 2003).

  One key advantage of CNN is their ability to handle multivariate time series data. This means that they can be trained on multiple financial time series, such as our specified predictors in Section 2.2, to learn relationships between these variables. Furthermore, CNNs also have the ability to learn long-term dependencies in time series data. This is achieved through the use of dilated convolutions, which allow the model to capture both short-term and long-term trends (Cui, W. Chen, and Y. Chen 2016; Nouri 2014). This tends to be important in finance, as long-term trends can be influenced by a variety of factors such as economic indicators or geopolitical events.

- **Recurrent Neural Network:** Recurrent Neural Networks (RNN) are an improvement of FNN and are distinguished by their ability to send information over sequential steps. This means that unlike traditional NN, which assumes that all inputs are independent of each other, RNNs are built on the premise that the order and context of elements in a sequence matter to each other. Hence, in time series data in financial markets, this characteristic provides high accuracy. The most important characteristic that is the cause of this is the fact that RNNs work with a memory, which allows them to not only individual data points but entire sequences of data (Watson 2003).

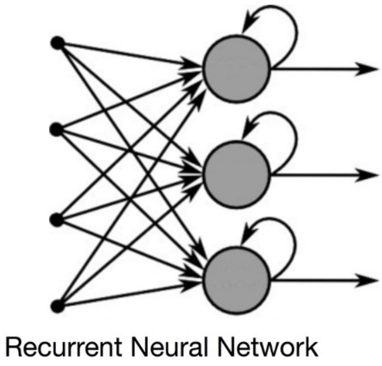


Figure 2.2: Recurrent Neural Network (Jeans 2019)

However, traditional RNNs are particularly prone to problems of vanishing gradients and exploding gradients during the training process. This means that if the contribution of information decays geometrically over time, it makes it difficult for the network to learn and retain long-range dependencies within the sequence. It thus prioritises short-term memory over the long-term information of the time series data. This issue is significant because it can impede the ability of RNN to learn from data effectively.

An addition to the RNN, these challenges have led to the development of an RNN variant, the Long Short-Term Memory (LSTM) network. Although an LSTM network is also designed as an RNN, Hochreiter and Schmidhuber (1997) have introduced a complex mechanism called a cell state, which runs through the chain of LSTM units. A cell state acts like a transfer system, transporting information straight down the entire chain of the network with only minimal linear interactions. This design allows information to flow relatively unchanged through many steps, making it easier for an RNN to transport and preserve long-term dependencies.

An LSTM cell exists out of three main components, as visible in Figure 2.3. A forget gate determines what information should be discarded from the cell state, an input gate decides what new information is going to be stored in the cell gate, and an output gate decides what the next hidden state should be. With this system, both long-term and short-term dependencies are captured within the network (Pawar, Jalem, and Tiwari 2019).

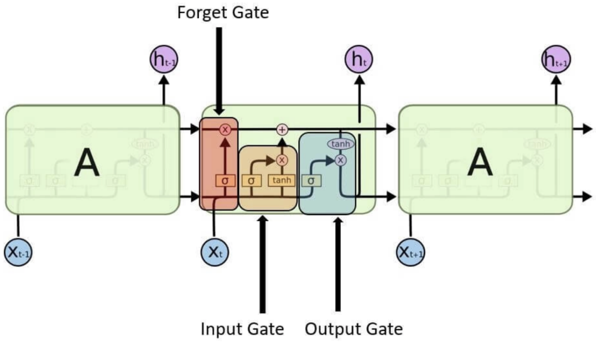


Figure 2.3: Schematic Representation of LSTM Cell (Tashmit 2023)

However, LSTM networks are computationally intensive and often require substantial training times. Their complex internal mechanics are challenging to design, understand and interpret compared to the simpler models.

2.4.3 Comparison

With the scope of this research, we are aiming to create valuable new insights for the current literature. Currently, there have not been any studies on forecasting IRSS using neural networks, and hence, we will be investigating its potential. In similar studies where the credit default swap spreads have been analysed, and term structures have been forecasted, LSTM networks emerge with high potential (Mao et al. 2023; Vukovic et al. 2022; Xiong et al. 2019). Although interest rate swaps are not the same product as credit default swaps, we can expect similar dynamics of the products given the fact that are sensitive to changes in credit conditions and market perceptions of risk (Giglio 2016). Furthermore, term structures and IRSS are highly related, as both are influenced by expectations of future interest rates and economic conditions (Cortes 2003). Term structures provide a snapshot of the current interest rate environment across different maturities, which directly influences the pricing and valuation of interest rate swaps. Hence, we will base this comparison on the existing literature, and the potential for applying it to IRSS forecasting.

Based on the comparative analysis provided in Table 2.2, LSTM networks emerge as the most suitable model for forecasting IRSS due to their specific advantages. Their ability to manage different sequence lengths and mitigate the vanishing gradient problem from an RNN. Furthermore, currently, some research indicates that LSTM dominates machine learning algorithms in financial forecasting (Firat et al. 2017; H. Y. Kim and Won 2018; Zahn et al. 2021). Despite the higher computational costs and their complexity, LSTM has the highest forecasting accuracy.

Table 2.2: Comparison of Neural Network Models in Financial Forecasting

Model	**Advantages**	**Disadvantages**	**Applications within Financial Forecasting**
FNN	Simple and fast to train; Good for static pattern recognition; Easier to understand and interpret	Lack of contextual memory; Prone to overfitting; Fixed input and output size	Stock market prediction; Credit risk assessment; Fraud detection; Portfolio management
CNN	Efficient in processing spatial data; Automatic feature extraction; Robust to spatial variations in input	High computational cost; Poor generalisation to non-visual data; Requires large datasets for training	Market sentiment analysis from visual content; Fraud detection in financial documents; Feature extraction from financial charts for algorithmic trading
RNN	Can process sequential data effectively; Contextual information processing across time steps; Flexible input and output lengths	Prone to vanishing and exploding gradient problems; Computationally intensive for long sequences; Difficult to parallelise	Time series forecasting; Anomaly detection in financial transactions; Predictive analysis for market trends
LSTM	Excellent at handling long-term dependencies; Versatile in managing various sequence lengths; Mitigates vanishing gradient problem	Resource-intensive; Complex architecture leading to longer training times; Can be difficult to tune	High-frequency trading analysis; Long-term stock market forecasting; Risk management; Sequential data analysis for economic indicators

2.4.4 LSTM

In Section 2.4.2, we have already discussed the background of LSTM models. In short, we have learned that RNNs enhance traditional networks by preprocessing sequences of data in order, to capture contextual relationships. In contrast to other NN models, RNN contains a memory component that handles entire time-series data. However, RNNs can struggle with exploding gradients, limiting their ability to learn long-term dependencies. To solve this problem, LSTM networks were developed, which incorporate mechanisms to capture long-term dependencies, together while handling short-term dependencies.

In this section, we will elaborate on the mechanics of LSTM networks, in order to be able to prepare the model architecture in Section 4.2. Therefore, we will elaborate more on the structure of the LSTM cells, as depicted in Figure 2.3. The LSTM memory itself is called a ”gated” cell, where the word gate reflects its ability to make the decision of preserving or ignoring certain memory information (Siami-Namini, Tavakoli, and Namin 2019). An LSTM model can capture important features from inputs and preserve this information based on the weight values assigned to the information during the training process. Therefore, an LSTM model can learn what types of information are worth preserving and what types are not. As visible in the figure, an LSTM cell exists out of the three following gates:

- **Forget gate:** The forget gate makes the decision on what information needs to be removed. This is done with a sigmoid function,

  $$\sigma(x) = \frac{1}{1+e^{-x}}, \tag{2.15}$$

  which creates a value which is often either close to 0 or close to 1 for each number in the cell state, as the function converges to these values (Song et al. 2020). This decision is based on the current input x_t and the previous hidden state h_{t-1}. The output of this gate is computed as follows:

  $$f_t = \sigma(W_{f_h}[h_{t-1}], W_{f_x}[x_t], b_f), \tag{2.16}$$

  where b_f operates as a constant which is called the bias value, and W_f are the weights. An output close to 0 means to forget the information, and an output close to 1 means to keep the information (Siami-Namini, Tavakoli, and Namin 2019).

- **Input gate:** The input gate makes the decision whether or not the new information will be added to the LSTM memory. This gate consists of two layers, the sigmoid layer as in Formula 2.15, and a hyperbolic tangent (tanh) layer (Song et al. 2020):

  $$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{2.17}$$

  Similarly to the forget gate, the sigmoid function generates a value often close to 0 or close to 1, acting as a filter that decides how much of the new values are let through. Furthermore, the tanh creates a value between -1 and 1, to develop a vector of new candidate values that will be added to the LSTM memory.

  $$i_t = \sigma(W_{i_h}[h_{t-1}], W_{i_x}[x_t], b_i) \tag{2.18}$$

  $$\tilde{C}_t = \tanh(W_{c_h}[h_{t-1}], W_{c_x}[x_t], b_c), \tag{2.19}$$

  where W_{i_h}, W_{c_h}, W_{i_x} and W_{c_x} are the weights for the input gate and the candidate cell state respectively, and b_i and b_c are the biases. i_t represents whether the value will be updated or not, and c_t indicates the vector of the new candidate values (Siami-Namini, Tavakoli, and Namin 2019).

After the forget gate and the output gate, a new cell state C_t will be calculated with the outputs from gates. This will be done with the following formula

$$C_t = f_t \times c_{t-1} + i_t \times \tilde{C}_t, \tag{2.20}$$

where C_{t-1} refers to the previous cell state. Combining these two layers from the input gate with the forget gate will therefore yield a new cell state C_t, for the output gate.

- **Output gate:** The output gate uses two layers to make a decision. The first layer is a sigmoid layer, as in Formula 2.15, on what part of the LSTM memory contributes to the output. After that, it performs a non-linear tanh function, as in Formula 2.17.

$$o_t = \sigma(W_{o_h}[h_{t-1}], W_{o_x}[x_t], b_o) \tag{2.21}$$

$$h_t = o_t \times tanh(\tilde{C}_t), \tag{2.22}$$

where o_t is the output value, and h_t is the next hidden state, which contains as a value between -1 and 1.

2.4.5 Summary

There are three categories of machine learning, supervised learning, unsupervised learning and reinforcement learning. Forecasting IRSS fall into the category of regression-supervised learning. In this section, the most significant models have been considered, such as RF, GBM and NN. After analysing and comparing these models, this study has found that NN has the highest forecasting accuracy. Although the black-box nature of NN and the risks of overfitting, the biggest advantages are that NN can find non-linear relationships and that it can work with large multi-dimensional datasets.

NN can vary widely but for financial forecasting FNN. CNN and RNN are relevant. From these models, we have found that FNN and CNN have too many drawbacks, in comparison to RNN, as RNN works with a memory, which more accurately captures the current trend of financial time series. However, traditional RNN models have big risks of vanishing gradients, and thus, LSTM models are considered, as these are designed to overcome these limitations. Although LSTM models have higher computational demands, research supports that the forecasting accuracy is superior. Based on these findings, we have now sufficient evidence to answer Sub-Research Question IV. Additionally, based on this information we can focus on studying Sub-Research Question V in Section 4.2 Similarly, we will provide the right foundation to answer Sub-Research Question VI in the next section.

2.5 Hybrid Model

Now that we have found the models that will be used for the IRSS forecasting with the highest expectation for forecasting accuracy, we will study the option to create an econometric-machine learning hybrid, in order to answer Sub-Research Question VI.

2.5.1 Hybrid model

In recent years, a lot of studies have been performed to evaluate the performance of hybrid forecasting, and the strengths offer a lot of potential. G. P. Zhang and Qi (2005) have introduced an ARIMA-ANN model, demonstrating enhanced forecasting accuracy than the individual models. In an economic context, hybrid models have proven effective as well, as Khan, Urooj, and Muhammadullah (2021) applied hybrid models to predict gold prices, and Mucaj and Sinaj (2017) focused on currency exchange rates time-series. However Devi et al. (2021)'s research on wheat production forecasting and Musa and Joshua (2020)'s study on Nigerian stock market returns, still faced limitations due to their specific contexts.

Although the hybrid models have been found to show potential in these studies, there have not been any studies yet on IRSS forecasting using hybrid models, or any similar interest rate derivative. Hence, we will use the methodology of the existing hybrid model studies mentioned above, and find the best approach for forecasting IRSS.

- **Parameter prediction:** As discussed in Section 2.3 and Section 2.4, both econometric models and machine learning models are set up by a lot of (hyper-)parameters. The setup of these parameters can be very difficult, as the setup of these parameters defines the whole topology of the model (Gorgolis et al. 2019). In the Parameter Prediction (PP) approach, the machine learning model will be used to accurately predict the most accurate parameters of the econometric model.
- **Residual correction:** In the Residual Correction (RC) approach, the prior knowledge of the econometric model is used to forecast. Using a similar setup to an individual econometric model, an initial forecast is made. However, as discussed in Section 2.3, econometric models could limited consider macroeconomic factors or other input parameters. Hence, with the use of machine learning models, the real value of the error of the econometric model will be forecasted, whilst considering the macroeconomic state. In this research, we will refer to the real value of the error as the residual error. Due to the nature of ML models, this creates the opportunity to enhance our forecast with multidimensional states.
- **Direct output combination:** In the Direct Output Combination (DOC), multiple models are independently used to predict a future value, after which these models will be combined. This can be done by assigning weights to each of the outputs, for example, based on the general accuracy based on the historical data.

2.5.2 Comparison

Based on the comparative analysis provided in Table 2.3, we find that RC stands out as the best hybrid model approach for our study. Whilst leveraging the strengths of econometric models, additionally, we will decrease the impact of the black box effect of the ML model. If we were to use the parameter prediction approach, we would discard a lot of our findings in the LSTM model, as we would have to build a new model to find the specific parameters. Furthermore, the direct output combination method is a relatively simplistic approach, in which the combining process of the models is very limited.

Table 2.3: Comparison of Hybrid Model Types for Forecasting Time Series Data

Model	**Advantages**	**Disadvantages**	**Applications**
PP	Enhances parameter setting of econometric models using ML techniques, potentially increasing model accuracy.	Relies heavily on the quality and appropriateness of the ML model for parameter estimation, which can be complex.	Useful in complex financial modelling where parameter setting is critical, like in risk assessment and pricing models.
RC	Leverages the strengths of econometric models and uses ML to correct for their errors, improving overall accuracy.	Dependence on the initial model's outputs; improvements limited to the correction capacity of the ML model.	Effective in enhancing the forecast accuracy of established models in economics and finance, especially where residuals are systematic.
DOC	Combines predictions from multiple models to potentially reduce bias and variance through ensemble techniques.	Requires careful calibration of weights and integration methods, which can be non-trivial.	Suitable for diverse scenarios in financial markets forecasting, where multiple models provide varied insights and forecasts.

2.5.3 Summary

In this section, we have discussed multiple types of hybrid models between econometric models and machine learning models. In the literature three main strategies were explored: PP, which used ML to optimise the parameters of econometric models, RC, which applies econometric models for initial forecasts and then uses ML models to correct any errors, and DOC, which merges outputs of multiple models by a given weight to each model. The analysis concludes that, given that we have found in Section 1.3 that the forecasting accuracy of econometric models lacks for IRSS, the RC approach is most suitable for our study due to its ability to leverage established strengths of econometric models whilst addressing their limitations with the accuracy enhancements provided by ML.

2.6 Summary

In this section, a comprehensive literature study was performed to create a fundamental understanding of the concepts that will be used during this study. We have studied the dynamics of IRSS, both by addressing the theoretical and the practical applications. From understanding the dynamics, and the risks associated with forecasting the IRSS including hedging costs, credit risks and economic conditions, we were then able to find the most relevant predictors for the IRSS. Both include financial predictors, to capture the risk-free movements of the IRSS, such as the Treasury yield curve, and the real-world macroeconomic predictors, such as GDP and inflation rates.

After selecting the most important predictors, we then focused on the models that could be used for our research objective. We have evaluated various econometric models, such as the VAR, ARIMA, and GARCH, and financial models, Vasicek and Hull-White. From this analysis, we found that the HW2F model will perform with the highest forecasting accuracy within our study. Furthermore, we have also evaluated machine learning approaches, particularly NN models. Among the different NN models, the LSTM model is highlighted for its superior forecasting accuracy, due to its ability to distinguish long-term and short-term relationships within financial time-series data.

Furthermore, we finalise this theoretical study by analysing hybrid models used for forecasting financial time-series data. Among multiple types of hybrid types, we find that the RC method emerges as the most promising, leveraging the theoretical strengths of both the econometric model and the ML model. In the RC approach, we will apply econometric models for initial forecasts and then use ML models to correct any errors.

3 Data

In Section 2.2, we have performed research on the financial and macroeconomic predictors IRSS. In this section, we will continue our research on our predictors, and transform them into inputs for our model. In Section 3.1, we will elaborate on how we collect the data, and in Section 3.2, we will perform an analysis on our data, in order to find improvements for our data in order to improve forecasting accuracy. Lastly, we will prepare the date in Section 3.3.

3.1 Data Collection

In order to solve our main research question, one of our objectives is to monitor the validity throughout our research. When doing quantitative research, the data used in the study has a high impact on this validity (Kaastra and Boyd 1996). Hence, there are a few requirements that are set for the data before it can be used in our models. Firstly, as described in Section 1.5, the data should be publicly available, to ensure transparency and replicability. When data is publicly available, other researchers can verify our results by replicating the analysis. This enhances the credibility of our findings. Secondly, the vendor of our data should have a high reputation, in order to increase the accuracy and the reliability of our outcomes. Errors in data can lead to incorrect conclusions, and hence, a study should be performed on the reputation of the vendor. Lastly, the data should be retrieved from similar sources. A similar source means that the number of distinctions between the types of data are minimised, by for instance choosing a specific scope for the financial landscape the data will be retrieved from (Kaastra and Boyd 1996). Therefore, and for data availability reasons we will only investigate European data for this research. Furthermore, daily data will provide less overfitting of the LSTM model and a better performance of our HW2F model, and it enhances feature learning, as there is more data to learn from. Lastly, as mentioned in Section 2.1, we found that the relationship between IRSS and Treasury yield varies with maturity, generally widening as maturity increases. Hence, for more significant predictions, we will use products with 10-year maturity.

As mentioned in the requirements above, to monitor the validity of our research, we should thus ensure that we retrieve our data from highly reputable sources. Furthermore, we need to ensure a similar format for our data. Hence, in Table 3.1, we enlist the sources and the format of the input data to our models.

Table 3.1: Sources of datasets

Data	Function	Source	n
Interest rate swap rate	Output	Refinitiv Eikon	5216
10-year zero-coupon bond	Predictor	Investing.com	5372
Treasury yield curve	Predictor	FRED Economic Data	5235
TED spread	Predictor	Refinitiv Eikon	5092
GDP	Predictor	Refinitiv Eikon	240
Unemployment rate	Predictor	Refinitiv Eikon	240
Inflation rate	Predictor	Refinitiv Eikon	240

3.2 Data Analysis

Although the vendors have been checked on their reputation of providing high-quality data, all data should still be checked for errors, by examining the basic statistics of the data (Kaastra and Boyd 1996). In this section, we will analyse the descriptive statistics of each of the datasets that we use. By doing so, we can look for inconsistencies and big outliers. Additionally, we will analyse the trends of our data, which are visualised in Appendix A. Furthermore, we will perform the Augmented Dickey-Fuller (ADF) test. The ADF test is a statistical test used in time series analyses checking for stationarity. If the data is non-stationary, the statistical properties, such as the mean and variance, change over time (Lopez 1997). The model underlying the ADF test follows the following process:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{i=1}^{p} \delta_i \Delta y_{t-i} + \varepsilon_t, \tag{3.1}$$

where Δy_t is the first difference of the series, α is a constant term, βt represents the deterministic trend, γ is the coefficient on the lagged level of the series y_{t-1}, and δ_i are the coefficients for the lagged differences Δy_{t-i}. the error term ε_t is assumed to be white noise. The null hypothesis H_0 of the ADF test is that $\gamma = 0$, which means that the data is non-stationary. The alternative hypothesis H_1 is that $\gamma < 0$. If the test statistic is less than the critical value of 5%, which is the standard level of significance used in many statistical tests to minimise the risk of Type I errors, the null hypothesis is rejected, indicating that the series is stationary.

In Figure 3.1, we visualise the correlation on a scale from 1 to -1, which represents perfect positive correlation and perfect negative correlation respectively. Some key takeaways from this figure are that IRSS shows a moderate positive correlation with inflation, which suggests that higher correlation to increase IRSS, ceteris paribus. Furthermore, we see that a higher unemployment rate is linked to lower IRSS observations, due to the moderate negative correlation. Furthermore, we see a strong positive correlation between the zero-coupon bond yields and the Treasury yields, as we expect, since both measure long-term borrowing costs. Lastly, we see a negative correlation between the unemployment rate and the inflation rate.

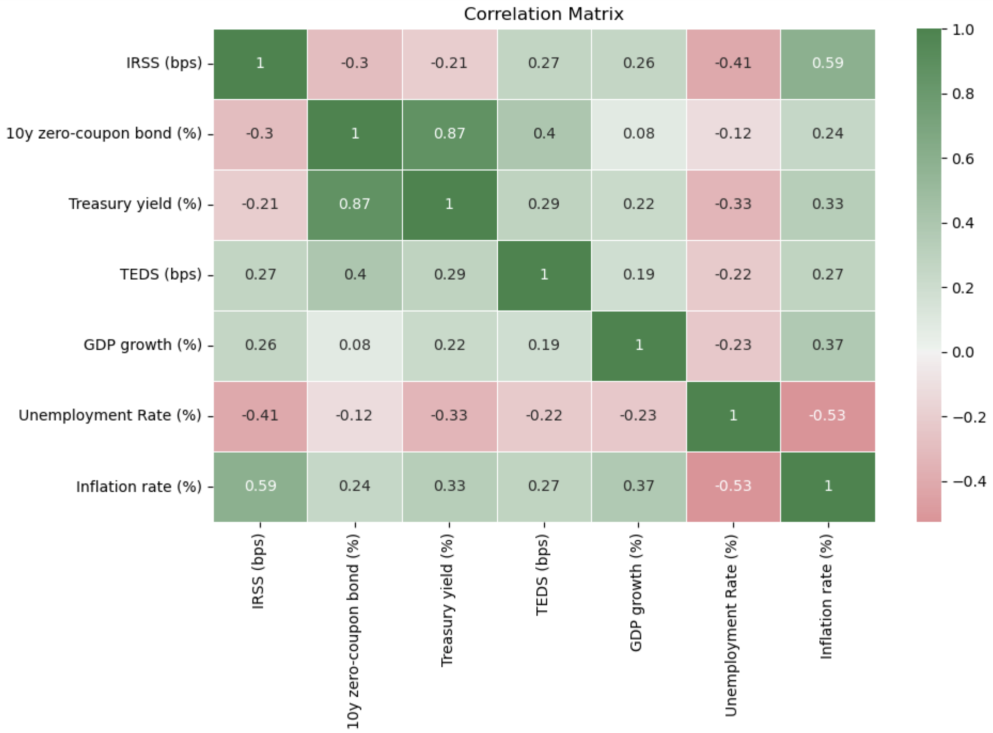


Figure 3.1: Correlation Matrix of Data

3.2.1 Interest rate swap rate

First, we will examine the interest rate swap rate data obtained from Refinitiv Eikon. This data will be used as the benchmark parameter for the output of our models. However, we first have to translate this interest rate swap rate into the IRSS. this is done by the following formula,

$$\text{IRSS} = \left(\frac{\text{Bid}_{\text{Swap Rate}} + \text{Offer}_{\text{Swap Rate}}}{2} - \text{10-Year Zero-Coupon Bond Yield} \right) * 100 \quad (3.2)$$

The trends within the IRSS are depicted in Figure 3.2. The graph shows both short-term fluctuations and long-term trends, indicating that IRSS is influenced by a mix of shocks and underlying economic factors. The IRSS fluctuates significantly over time, with sharp peaks in 2008, 2012 and 2021, in times of economic crises. Table 3.2 shows that the standard deviation of 16.81 bps is relatively high, and together with the very large range of the values from the minimum value to the maximum value, we can conclude that the series is very volatile. However, the ADF test shows stationarity, suggesting that there is no unit root within the time series, which is ideal for our analysis.

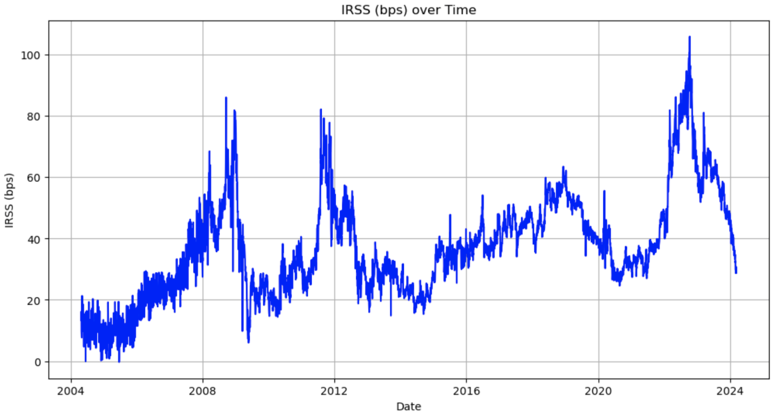


Figure 3.2: Trend of the IRSS

Table 3.2: Descriptive Statistics for Interest Rate Swap Rate

Count	**Mean**	**std**	**Min**	**25%**	**50%**	**75%**	**max**	**ADF**
5117	36.98 bps	16.78 bps	-0.3 bps	25.3 bps	35.7 bps	46.8 bps	106 bps	Reject H_0

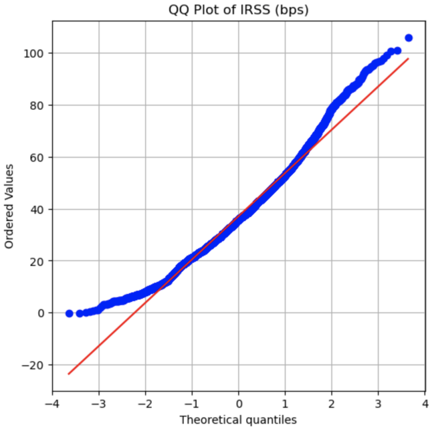


Figure 3.3: QQ Plot for IRSS Data

In Figure 3.3, we see a deviation from the normal distribution in the Quantille-Quantile (QQ) Plot, especially in the upper tail, indicating a positive skewness and potential outliers, which in accordance with Figure 3.2 and the histogram in Figure 3.4.

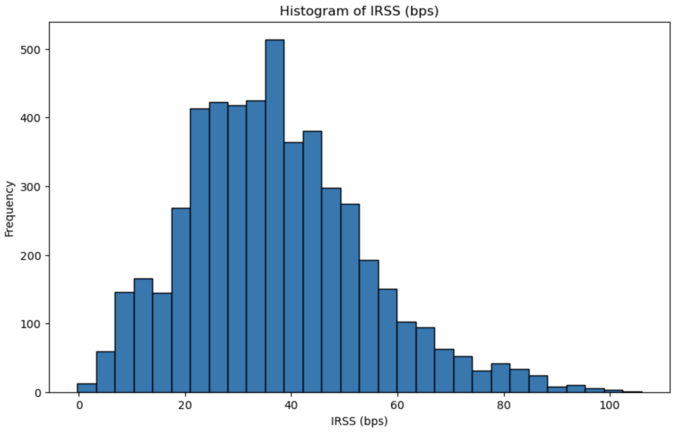


Figure 3.4: Histogram of IRSS Data

3.2.2 10-year zero-coupon bond

The 10-year zero-coupon yield shown in Appendix A.1 contains significant variance, which is confirmed by the descriptive statistics in Table 3.3. Furthermore, it is noticeable that the minimum value is negative, which means that despite the negative yield, it was still perceived as relatively safe in the conditions of the market. Furthermore, we find with the ADF test that the historical bond yield is non-stationary, as the H_0 hypothesis is accepted. However, the structural break between the training data and the testing data could cause difficulties, as COVID-19 introduced a rising bond yield.

Table 3.3: Descriptive Statistics for 10-Year Zero-Coupon Bond

Count	**Mean**	**std**	**Min**	**25%**	**50%**	**75%**	**max**	**ADF**
5117	1.80%	1.58%	-0.854%	-0.355%	1.70%	3.26%	4.69%	Not Reject H_0

Furthermore, the QQ plot in Appendix C.1 indicates deviation from normality, particularly in the tails, as we can see by the many peaks in the histogram in Appendix D.1.

3.2.3 Treasury yield curve

The Treasury yield curve, as visualised in Appendix A.2, fluctuates a lot, as it is influenced by economic data, central bank policies and market sentiment. The graphs suggest that there are a lot of short-term interest rate changes. The descriptive statistics in Table 3.4 show a relatively moderate standard deviation. Furthermore, the interval from the minimum value to the maximum is according to the policy decisions of the past twenty years. Although the moderate standard deviation, the time series does not reject the ADF test null hypothesis and is therefore non-stationary. Lastly, the structural break of the Treasury yield curve could not be ideal, as, after the break, the index rises quickly, unlike before.

Table 3.4: Descriptive Statistics for Treasury Yield Curve

Count	**Mean**	**std**	**Min**	**25%**	**50%**	**75%**	**max**	**ADF**
5117	2.89%	1.13%	0.00%	2.00%	2.74%	3.84%	5.26%	Not reject H_0

The QQ plot in Appendix C.2 shows a slight deviation from normality, with thick tails to both ends. The histogram in Appendix D.2 shows a slightly left-skewed distribution, but mainly confirms the thick tails.

3.2.4 TED Spread

The TEDS in Appendix A.3 shows variability with spikes, indicating periods of financial stress, similar to the other predictors of IRSS. The standard deviation, given in Table 3.5, is relatively high, mainly caused by the financial instability before 2010. In times of financial stability, the TEDS has proven to be very stable. Therefore, the ADF test shows the series to be stationary.

Table 3.5: Descriptive Statistics for TED Spread

Count	**Mean**	**std**	**Min**	**25%**	**50%**	**75%**	**max**	**ADF**
5092	41.1 bps	35.0 bps	-19.0 bps	20.9 bps	31.4 bps	49.8 bps	324 bps	Reject H_0

The histogram and the QQ plot in Appendix D.3 and Appendix C.3 relatively, indicate a right-skewed distribution, with some upside outliers, and a significant deviation from normality.

3.2.5 GDP

The GDP growth in Appendix A.4 generally shows a growth in the past twenty years. The GDP series contains, in contrast to the other series, not daily but monthly observations. Although the series visibly represents the economic turmoils in the past twenty years, according to the descriptive statistics in Table 3.6 and to the ADF test, the series is stationary. Lastly, we see in the graph that the structural break is not ideal, as the GDP experiences big shocks due to COVID-19.

Table 3.6: Descriptive Statistics for GDP

Count	**Mean**	**std**	**Min**	**25%**	**50%**	**75%**	**max**	**ADF**
5113	0.956%	3.06%	-14.7%	0.500%	1.40%	2.10%	14.3%	Reject H_0

The QQ plot for GDP in Appendix C.4 shows significant deviations from normality, with clusters of extreme values on both ends. The histogram in Appendix D.4 confirms a left-skewed distribution, with some extremes on the positive side.

3.2.6 Unemployment rate

Appendix A.5, shows periods of stability and periods of fluctuations of the unemployment rate in Europe. The descriptive statistics in Table 3.7 show a very moderate variance in the series. However, the figure shows that over the time period of twenty years, there has been a very large trend, suggesting a unit root within the series. This is confirmed by the ADF test, and hence, the unemployment rate is non-stationary.

Table 3.7: Descriptive Statistics for Unemployment Rate

Count	Mean	std	Min	25%	50%	75%	max	ADF
5117	8.92%	1.68%	6.40%	7.50%	8.70%	10.1%	12.2%	Not reject H_0

In Appendix C.5, we see that there is no normality in the unemployment rate. Similarly to the Treasury yield, the histogram of the unemployment rate in Appendix D.5 shows several peaks, with no clear distribution.

3.2.7 Inflation rate

Throughout the past twenty years, the inflation rate in Appendix A.6 shows a relatively stable trend, with minor fluctuations. However, since the COVID-19 pandemic, the series has shown a very large peak, which has an effect on the interval between the minimum and maximum value, and on the standard deviation of the series. Due to this large peak, the inflation rate of the past twenty years is non-stationary according to the ADF test. Lastly, due to COVID-19, we see a big shock in the inflation rate, which is not ideal for the structural break.

Table 3.8: Descriptive Statistics for Inflation Rate

Count	Mean	std	Min	25%	50%	75%	max	ADF
5117	2.12%	2.06%	-0.600%	0.900%	1.80%	2.50%	10.6%	Not reject H_0

Appendix D.6 and Appendix C.6 show the histogram and the QQ plot of the inflation rate. It shows a left-skewed distribution with very noticeable outliers on the positive side.

3.3 Data Preparation

Once the data is collected, it must be prepared for analysis in our models. This involves cleaning the data and handling missing or inaccurate observations. Missing observation which often exists, can be handled in various ways. All missing observations can be dropped, or the observations can be created by interpolating nearby values in the data (Kaastra and Boyd 1996). In our research, we will investigate the missing observations closely. A single missing observation will be filled with the average value of the two surrounding values. If there are many missing observations, we will look for other data sources. Lastly, the data should be transformed into a suitable format for the analysis. For example, this means that all data should be set to similar timelines, and to normalise all the data, in order to prevent a single feature from dominating the whole analysis (Prince 2023). We can achieve this by scaling the data in a range using a min-max standardisation, using the following formula

$$x_{i,norm} = 2 * \frac{x_{i,obs} - x_{i,min}}{x_{i,max} - x_{i,min}} - 1, \tag{3.3}$$

where x_{max} and x_{min} are the maximum and the minimum values of predictor i respectively. This equation will result in a range of values from [-1,1], which is optimal for machine learning, as all values revolve around 0.

Furthermore, in Appendix B, we analyse the box plots visualising a part of the descriptive statistics given in Section 3.2. Here, we aim to find the outliers of each of the predictors, as outliers can significantly impact the forecasting accuracy of our models. To do so, we use the Interquartile Range (IQR), defined as the interval between the first quartile (25%) and the third quartile (75%), with $IQR = Q3 - Q1$. Outliers are identified for observations x of predictor i for $x_i < Q1_i - (c * IQR_i)$ or $x_i > Q3_i + (c * IQR_i)$. For our analysis, we take $c = 3$, as our dataset is very large, and we are looking for $\alpha = 0.05$ (Zafeirelli and Kavroudakis 2024), similarly to our ADF analysis in Section 3.2. With this analysis, we find that the predictors contain a lot of outliers, as visualised in Table 3.9. In Figure 3.5, we visualise the box plot of the IRSS observations. We find that although there are outliers, none of them fall outside of the $3 * IQR$ range. This is confirmed by the table. After identifying the outliers, we now transform these outliers and set them to the minimum or maximum value within the $3 * IQR$ range.

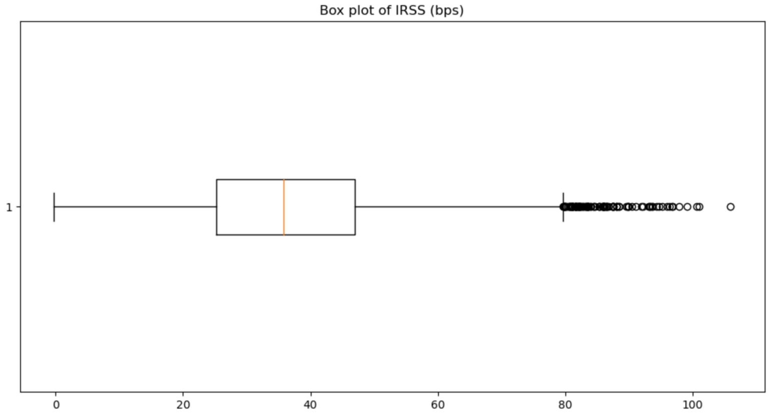


Figure 3.5: Box-plot of IRSS Observations

Table 3.9: Percentage of Outliers Per Parameter

Parameters	**Percentage of Outliers**
Interest rate swap rate	0%
10-year zero-coupon Bond	0%
Treasury Yield	0%
TED Spread	0.0293%
GDP	0.0642%
Unemployment rate	0%
Inflation	0.0506%

Lastly, in Section 3.2 we have identified that the 10-year zero-coupon bond, the treasury yield curve, the unemployment rate and the inflation rate are non-stationary time series. Econometric models assume stationarity to produce reliable forecasts, hence, non-stationary data can lead to inaccurate model parameters. A common technique to use for non-stationary time series is to apply the first difference. In this case, we transform the data with the following formula:

$$x_t^\star = x_t - x_{t-1}, \tag{3.4}$$

for every IRSS observation x_t. This transformation helps stabilise the mean of the time series by removing trends and reducing the impact of seasonality. After applying this first-difference method, we find that our predictors are stationary according to the ADF test.

After preparing our data, we show the renewed descriptive statistics in Table 3.10. We mainly see changes in the standard deviation and maximum values for the TEDS, GDP and Inflation rates, which can be explained by the removal of the outliers of those predictors. Furthermore, we see that all data is stationary, as each H_0 hypothesis of the ADF test is rejected. Additionally, no outliers are found within the data.

Table 3.10: Descriptive Statistics of Prepared Data

Parameter	**Count**	**Mean**	**std**	**Min**	**25%**	**50%**	**75%**	**Max**	**ADF**	**Outliers**
Interest rate swap rate	5116	37.03 bps	16.81 bps	-0.3000 bps	25.20 bps	35.80 bps	46.97 bps	106.0 bps	Reject H_0	0
10-year zero-coupon bond	5116	1.797%	1.577%	-0.8540%	0.3550%	1.701%	3.256%	4.687%	Reject H_0	0
Treasury yield curve	5116	2.901%	1.126%	0.5200%	2.010%	2.750%	3.850%	5.260%	Reject H_0	0
TED spread	5116	39.74 bps	29.66 bps	-18.97 bps	20.95 bps	31.18 bps	49.56 bps	135.4 bps	Reject H_0	0
GDP	5116	1.014%	2.126%	-4.300%	0.5000%	1.400%	2.100%	6.900%	Reject H_0	0
Unemployment rate	5116	8,924%	1.675%	6.400%	7.500%	8.700%	10.10%	12.20%	Reject H_0	0
Inflation rate	5116	2.038%	1.792%	0.6000%	0.9000%	1.800%	2.500%	7.300%	Reject H_0	0

3.4 Summary

To ensure the validity of our research, we have set criteria for the quality of our data sources. Firstly, the sources that we use should be publicly available for transparency and replicability, the sources should be reputable vendors to ensure accuracy and reliability, and we should minimise the distinctions between the data that we have retrieved, such as using the same data interval, to improve the comparison between data types.

Within our data analysis, we have analysed all the predictors of IRSS for errors using visual analysis of the time series and descriptive statistics. Furthermore, we have checked the stationarity of the predictors using the ADF test.

Lastly, in the data preparation phase, we handled missing observations by interpolating single gaps and seeking alternative sources for larger gaps. Furthermore, we have normalised the data using min-max standardisation and we have aligned the data to similar timelines. Outliers were then identified and adjusted using the IQR method to prevent them from having too much impact on our forecasting accuracy. Non-stationary time series, where the mean and variance change over time, are made stationary using the first-difference method, which subtracts each data point from the previous one to eliminate the unit root. This transformation is important for reliable forecasting, as the models used in our study require stable data to produce accurate predictions.

4 Empirical Implementation

After analysing the data in the previous section, we will now focus on the empirical implementation of our models. In this section, we will elaborate on each individual model, and describe the model architecture that we will use, in order to increase our forecasting accuracy. First, we will In section 4.1, we will elaborate on the HW2F model, after which we will discuss the LSTM model in Section 4.2. The knowledge and the architectures of these models will then be combined in our hybrid model in Section 4.3. We will finally discuss the performance estimators in Section 4.4.

4.1 HW2F Model

In Section 2.3, we have studied the theoretical information needed for our HW2F model. In this section, we will learn how the architecture of the model is structured, and how it operates. To do this, we will specify the parameters of the HW2F model and its calibration to historical data, ensuring that it accurately reflects the dynamics of IRSS.

In Figure 4.1, we show the train-test split of the HW2F data that will be used in our model. The structural break between the training and testing datasets reflects that the model will be tested against realistic and volatile scenarios. The data before the break contains realistic stress moments, similar to the data after the break. Hence, this will provide a good evaluation of its performance.

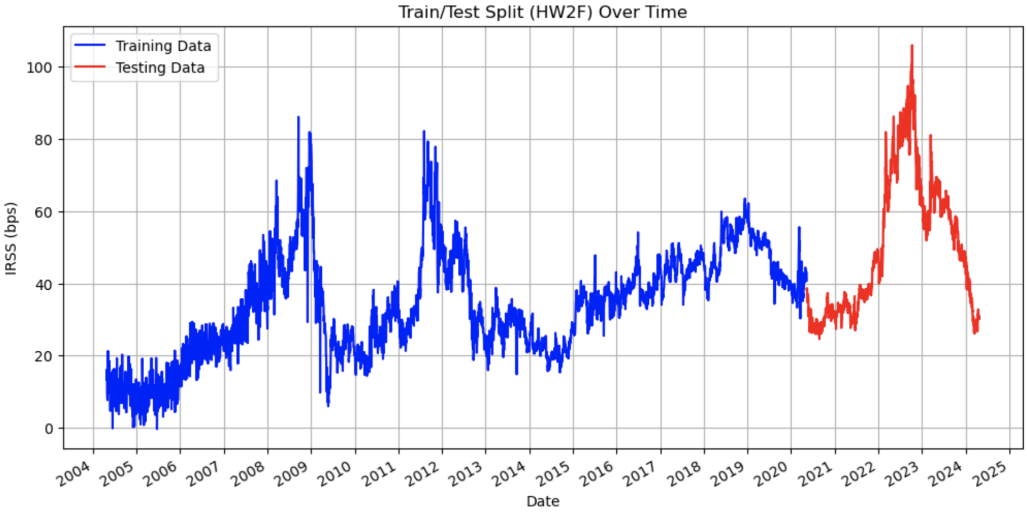


Figure 4.1: Train-Test Split of the HW2F Data

A significant part of structuring the architecture of the model is by setting the model parameters. The first step is to calibrate the parameters of the HW2F. A loss function will be used, namely the RMSE, which will be introduced in Formula 4.3. During our study, we will use the programming environment Python to optimise the complex parameters. During this optimisation process, we will use a grid search with a large number of simulations to find the ideal setup of the model, minimising the RMSE. Each simulation will perform a full training cycle, and after a number of training cycles, we take the average RMSE. By doing this, we ensure to eliminate the randomness of the Wiener Process, which is mentioned in Formula 2.10. To find the right number of simulations, we use a stopping rule that stops the simulation after the convergence of the RMSE. Lastly, know that the range for a and b are both non-negative and have a practical upper bound, as very high mean parameter values would suggest too quick mean reversion. There is no information in the literature on the boundary conditions of the parameters, so for this study, we set the constraints to $0.01 \leq a \leq 1.00$ and $0.01 \leq b \leq 1.00$. The results of this grid search will be considered in Section 5.1.1. For the long-term and the short-term volatility of the model, we use dynamic volatilities which change over time, instead of static parameters. By using changing volatility parameters, we aim to create a more flexible and adaptive model with improved matching of the volatility behaviour of the IRSS data.

With all the information that we gathered on the HW2F model, we are able to build the model. In Figure 4.2, a schematic overview of the model is shown. In step 1, we read and preprocess the data. The input to our model is the historical IRSS observations, which will first be normalised, using the min-max standardisation, after which the data will be split into the train and test set, as mentioned in Section 3.3. After this, we build the environment of the HW2F model. This step mainly consists of building the right functions for the calculation of the HW2F model. Firstly, we will calculate the daily return rates, which are then directly used to find the correlation between the short-term and the long-term changes. This correlation is used in the Cholesky decomposition of the correlated Wiener processes in Formula 2.10. Lastly, we set up the function for calculating the process of θ_t according to Formula 2.14 and the general HW2F function according to Formula 2.10. After setting up the environment, we can now train our model, and find the ideal setup for our a and b parameters, by running through all possible settings. We evaluate the average loss function per setup, and after all settings have been run through, we find the setup that forecasts the most accurately. This setup is then used in our testing step, where we apply our model to completely new data. The output of our model is the IRSS prediction of day $t+1$, which we can compare to the real observation, and evaluate the RMSE.

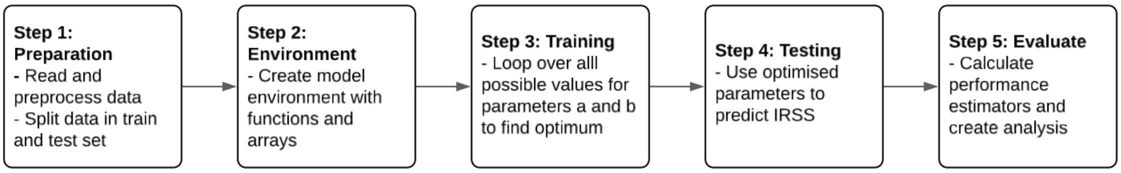

Step 1: Preparation
- Read and preprocess data
- Split data in train and test set
Step 2: Environment
- Create model environment with functions and arrays
Step 3: Training
- Loop over alll possible values for parameters a and b to find optimum
Step 4: Testing
- Use optimised parameters to predict IRSS
Step 5: Evaluate
- Calculate performance estimators and create analysis

Figure 4.2: Schematic Overview HW2F Model

4.2 LSTM Model

In Section 2.4, we have considered the structure of LSTM cells. In order to make these LSTM cells operate, the whole model will be configured appropriately. In this section, we will consider the model architecture and the training dynamics of our LSTM model. The architecture of the model is crucial to how the model learns and makes conclusions from the data. The main goal of our LSTM model is that learn from in-sample data, and that the model can apply this knowledge to out-of-sample data, in order to forecast the fluctuations in IRSS. We will discuss the method of handling information within the network, and the methods to perform best out-of-sample, without overfitting on the in-sample data.

During the testing phase, input data will first be passed forward through the system. The input data that will be used are the predictors of IRSS, as found in Section 2.2. This means that during a simulation, the predictor data goes through the LSTM cells, as described in Section 2.4. The state of each cell is influenced by the input at that time step and the previous cell state. After a complete pass through the system, the loss function is calculated. Appropriate financial time series is to use the RMSE for this loss function, which is given in Formula 4.3. This function measures the difference between the predicted output of the LSTM and the actual observation. This information is then backpropagated through the network, starting from the output layer, towards the input layer (Rumelhart and Zipser 1985). This process involves the calculation of the gradient of the loss function with respect to each weight in the network by applying the chain rule. These gradients give information on how changes in each parameter affect the overall loss function. As our LSTM model will handle data sequentially and maintain a memory of previous inputs through their hidden states, the gradients of the loss function that were calculated are not only important to the current sequences but also for the dependencies (Bishop and Nasrabadi 2006). After the updates of the gradients, the network will look for a local minimum, which is called the gradient descent (Prince 2023). With this updated gradient, the weights and the biases of the LSTM network are updated, with the use of the following formula

$$w_{new} = w_{old} - \eta * \nabla_w J(w), \tag{4.1}$$

where η represents the learning rate, and $\nabla_w J(w)$ the gradient of the loss function J with respect to weight w. By doing so, the model aims to reduce the loss in subsequent iterations.

During these simulations, the model optimises its knowledge in each state at each time step. After many runs, the goal is that the model has had enough experience with the data, to be able to predict the next step. However, as for our HW2F model, it is important that our model is not prone to overfitting. Hence, within the LSTM model, we will apply three methods to prevent overfitting. Firstly, we split the training data set from the testing data set. The out-of-sample testing set is crucial to provide an unbiased evaluation of the prediction accuracy of new data, ensuring that the LSTM model generalises well beyond the training data. As training the model is very important for the learning process of the model, we have to create a large training data set. This will be done by taking a 70/10/20 train-validation-test split, as visualised in Figure 4.3. This split ensures that the test-IRSS data is similar to the HW2F test-IRSS data. This ensures a reliable comparison of the results. The structural breaks between the training, validation and testing datasets are chosen to segment different phases of market conditions. The training phase and the testing phase both contain a stress period, to make sure that the model will be tested against realistic scenarios. The validation dataset contains relatively less volatility, to make sure that the validation process is smoother. Furthermore, as mentioned in Section 3.1, we will use daily data as our input. Secondly, we will use a dropout mechanism, which randomly sets a fraction of the input data equal to zero, to prevent neurons from co-adapting too much. By doing this, the dropout forces the network to learn more robust features that are useful in predicting IRSS. Thirdly, we will create a weight decay, that adds a penalty to the loss function. This encourages the model to keep the weights small. These three strategies will make sure that the model is simplified, and will help prevent it from overfitting. The final output of the model is a single value that estimates the IRSS value on the next day.

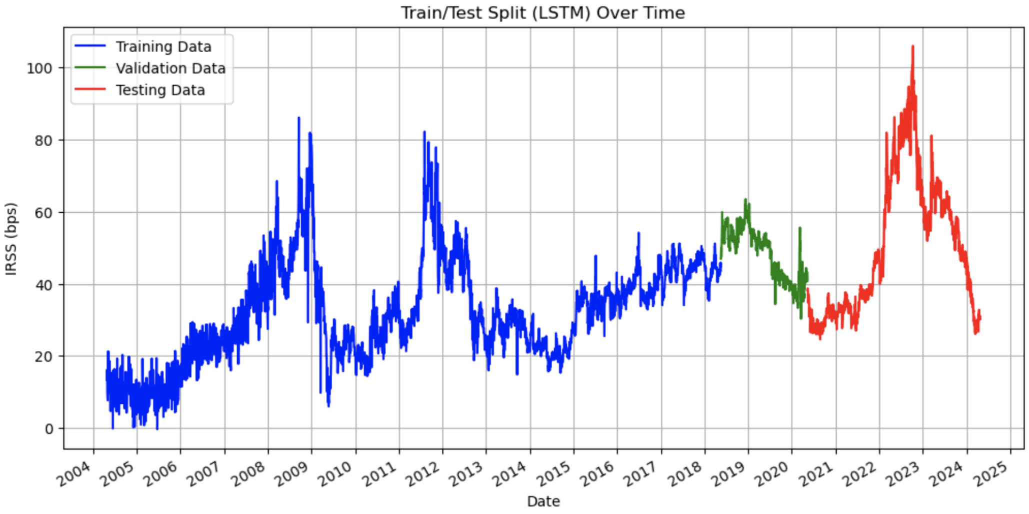


Figure 4.3: Train-Test split of the LSTM Data

Furthermore, despite the useful functionalities of NN, they are typically characterised by a large set of hyperparameters. These hyperparameters define the network's topology, computational power and a lot more (Gorgolis et al. 2019). Therefore, hyperparameters are essential for machine learning algorithms since they directly control the training algorithm's behaviours and have a significant effect on the performance of machine learning models (Bakhashwain and Sagheer 2021). Hence, these hyperparameters have to be configured properly in order to increase the predictive accuracy of the network. Hyperparameter optimisation aims to find the global optimum for $\mathbf{x}^*$ of an unknown black-box function f where $f(\mathbf{x})$ can be evaluated for any arbitrary $\mathbf{x} \in \chi$. That

$$\mathbf{x}^* = argmax_{\mathbf{x}\in\chi} f(\mathbf{x}),$$

where χ is a hyperparameter space that can contain categorical discrete and continuous variables, which are set for our research in Table 4.1 (H. Cho et al. 2020).

Optimising these hyperparameters can be very difficult, and using a random value process can take a lot of iterations before finding the right setup, as multiple hyperparameters correlate with each other through a lot of dimensions (Goodfellow et al. 2013; Larochelle et al. 2012). Furthermore, having to run the entire NN requires a lot of computational effort and takes a lot of time. Hence, apart from randomly selecting the hyperparameters, the implementation automation of hyperparameter tuning is required in our model. During our study, we make use of Bayesian approximations, which use the probabilistic model to make informed decisions on where to evaluate the function next within parameter space χ (Nguyen 2019).

There are many hyperparameters that should be set for an LSTM model. The most important ones are discussed here (Reimers and Gurevych 2017). Firstly, the number of layers of the model determines the depth of the LSTM. A deeper NN can capture more complex relationships but may be harder to train. The batch size represents the number of samples that are processed before the LSTM model is updated. Smaller batch sizes can lead to faster convergence, but it might be more noisy. Furthermore, the number of epochs represents the number of complete passes through the entire training set. More epochs allow more learning opportunities but lead to more overfitting. The type and the ranges of each hyperparameter are enlisted in Table 4.1 (Gal and Ghahramani 2016).

Table 4.1: Possible Hyperparameters Setting Ranges

Hyperparameter	**Type**	**Range**
Number of layers	Integer	[1,4]
Batch size	Integer	32,64,128,256,512
Units in each layer	Integer	16,32,64,128,256,512
Epochs	Integer	[100,1000]
Activation Function	Categorical	relu, elu, selu, tanh, sigm
Dropout rate	Float	0.05,0.1,0.025,0.5
Learning rate	Float	[0.001,0.1]

To conclude, we now have gathered all the information needed to start building the model. In Figure 4.4, we find a schematic overview of the LSTM model that is built. The model consists of 5 steps, in which the first step is to prepare the data. The input data that is used in the LSTM model, are the predictors of IRSS that have been identified in Section 2.2, which have been analysed in Section 3.2. The data is then split into the training, validation and testing sets as discussed above. In the next step, we will set up the LSTM model itself. This step consists of building the cells, after which the Bayesian approximations hyperparameter tuner will find the ideal setup for the model, given the input data. The ideal setup is then saved, so that it can be called upon when needed in the training and the testing step. The next step is to start training the model, from which the improvement of the weight distribution of the model is the most important element, as discussed above.

The model starts using all the input parameters to find the output, which is the prediction of the IRSS value on day $t+1$. With the use of gradient descent, an optimal setup is found for the testing step. In this step, the model is applied to completely new data and predicts the IRSS. The prediction can then be compared with the actual observation, after which we can analyse the error.

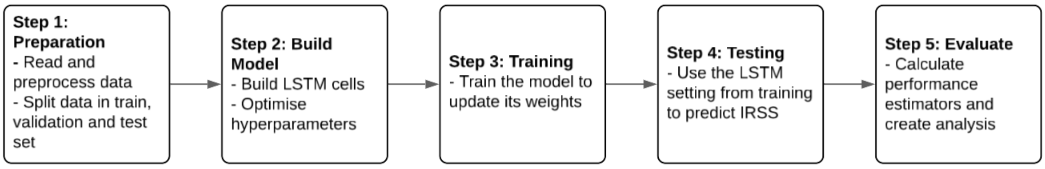

Step 1: Preparation
- Read and preprocess data
- Split data in train, validation and test set
Step 2: Build Model
- Build LSTM cells
- Optimise hyperparameters
Step 3: Training
- Train the model to update its weights
Step 4: Testing
- Use the LSTM setting from training to predict IRSS
Step 5: Evaluate
- Calculate performance estimators and create analysis

Figure 4.4: Schematic Overview LSTM Model

4.3 Hybrid Model

In Section 2.5, we have performed a study on approaches to create an integrative hybrid approach to enhance the forecasting accuracy of individual traditional models and ML models. In that study, we have found that RC contains the highest potential to enhance our forecasting accuracy. Within RC, the prior knowledge of the econometric model is used to forecast, after which the residual error of the traditional model will be forecasted. To apply this to our models, we will use HW2F to do the main prediction. As discussed in Section 1.3, we know that this model's predictive accuracy is limited, and hence we will leverage our LSTM model to predict the residual error of our HW2F model. This hybrid model approach is thus aimed as an enhancement of current forecasting techniques.

As discussed, with the aim of enhancing the individual models we set up a hybrid model. The hybrid model consists of five steps, as visualised in Figure 4.5. Firstly, we will create an environment that is very similar to the HW2F environment in Section 4.1. We set up the parameters and the functions as described in Figure 4.2. The mean-reversing parameters a and b will be trained during the training of the HW2F model phase, after which it can directly be implemented in the hybrid model, as that setting produces the best HW2F forecasting accuracy.

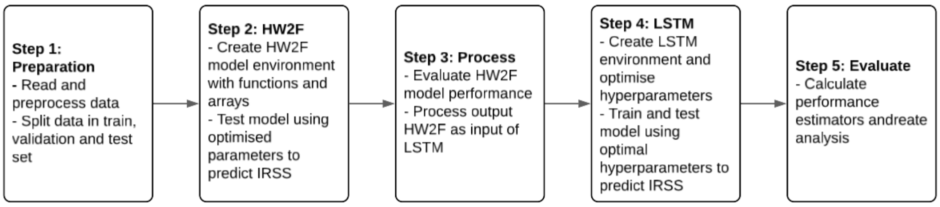

Step 1: Preparation
- Read and preprocess data
- Split data in train, validation and test set
Step 2: HW2F
- Create HW2F model environment with functions and arrays
- Test model using optimised parameters to predict IRSS
Step 3: Process
- Evaluate HW2F model performance
- Process output HW2F as input of LSTM
Step 4: LSTM
- Create LSTM environment and optimise hyperparameters
- Train and test model using optimal hyperparameters to predict IRSS
Step 5: Evaluate
- Calculate performance estimators andreate analysis

Figure 4.5: Schematic Overview Hybrid Model

The next step is to process the error of the HW2F. For each time step, we calculate the error of the HW2F model. The goal is to predict this error with our LSTM model, using the RC approach as discussed in Section 2.5. Differently from our individual LSTM model, where the IRSS rates are the input and the output of the model, our hybrid model now uses the residual error as input and output. Due to this significant change in input and output variables, we will find the ideal hyperparameter setting using the Bayesian approximations parameter optimiser, as discussed in Section 4.2. With these hyperparameters, we start training the model, using the same train/validation/test split. After that, our model will be tested on our new testing data, during which it will use its built-up knowledge to predict the error of the HW2F model. This output will then be evaluated with Formula 4.2, which calculates the error of the hybrid model.

$$\text{Error Hybrid Model} = (\text{Prediction HW2F} - \text{Observed IRSS}) - \text{Prediction LSTM} \tag{4.2}$$

Within our hybrid model, it is important to prevent overfitting. To make sure that the model will not analyse the in-sample data too closely, we will use the same three procedures for the HW2F part of the hybrid model, and the three procedures for the LSTM part of the model. For the HW2F part, we make sure to split the data into training data and testing data, we will use daily data to create sufficient historical data, and we will empirically check whether the optimised parameters are economically plausible. Furthermore, for the LSTM part, we will use a training/validation/testing split within our data, that aligns with the split from the HW2F. Furthermore, we will use the same dropout mechanism as mentioned in Section 4.2, and the same weight decay.

4.4 Performance Estimators

To evaluate the prediction accuracy of our HW2F, LSTM and hybrid models, we conduct several experiments on the results of the models. To do this, we find the right performance estimators in the current literature on forecasting financial time-series data. Although the RMSE functions as the loss function of the LSTM model, we will evaluate the performance of the models using the following three performance estimators, to improve our result analysis:

- **Root Mean Squared Error:** The RMSE is commonly used to assess the accuracy of time-series data prediction using regression analyses (Bakhashwain and Sagheer 2021). The RMSE captures the standard deviation of the prediction error, and is calculated as follows:

  $$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_t^{obs} - y_t^{pred})^2} \tag{4.3}$$

  The RMSE is sensitive to large errors, due to the squaring of prediction errors within the formula. Hence, large outliers are penalised.

- **Mean Absolute Error:** The Mean Absolute Error (MAE) is a straightforward metric in regression models, reflecting the average error in a set of predictions, without considering their direction (Bakhashwain and Sagheer 2021). The MAE is calculated as follows:

$$MAE = \frac{\sum_{t=1}^{T} |y_t^{obs} - y_t^{pred}|}{T} \tag{4.4}$$

A key advantage of MAE is its interpretability. Furthermore, it is robust to outliers since it does not square the errors.

- **Mean Average Percentage Error:** The Mean Average Percentage Error (MAPE) provides a useful measure of prediction accuracy in a forecasting method which usually expresses accuracy as a percentage (Song et al. 2020). The MAPE is calculated as follows:

$$MAPE = \left(\frac{1}{T} \sum_{t=1}^{T} \frac{y_t^{obs} - y_t^{pred}}{y_t} \right) \times 100 \tag{4.5}$$

A big advantage to adding MAPE as a performance estimator is that the MAPE calculates the relative error instead of an absolute error. This allows for the comparison of forecasts across different scales, which increases the cross-model comparison.

During our study, we will use both these performance estimators to evaluate the performance of each model. The strengths of these performance estimators will provide a lot of information on the nature of the errors. If the MAE is higher than the RMSE, we will know that there are multiple errors, but those errors are not very large. On the other hand, if the RMSE is higher than the MAE, we know that although the model could perform well, there are a few very large errors. Furthermore, the MAPE will provide us with a relative error, which increases the accuracy of our comparison.

4.5 Summary

In this chapter, we have elaborated on the architecture of the three models of this research in forecasting IRSS. The HW2F model's structure is developed by analysing historical IRSS data, which provided initial estimates for the model parameters. The model calibration involves optimising the mean-reverting parameters a and b, using sensitivity testing with the parameter constraint of $0.01 \leq a$, $b \leq 1.0$. To overcome overfitting, the model uses cross-validation, a train/test data split and daily data.

The architecture of the LSTM model and its hyperparameters is optimised using a Bayesian approximations hyperparameter tuner. During the training phase of the model, a lot of effort is put into preventing the model to start overfitting. This is done by using a training/validation/testing data split, a dropout mechanism and a weight decay.

Lastly, the hybrid model combines the architecture of the HW2F and the LSTM models to enhance forecasting accuracy. The HW2F model provides the initial predictions, while the LSTM model predicts the residual error of the HW2F model. This approach leverages the strengths of both models. The hybrid model follows a structured process from parameter setup to error prediction, whilst incorporating all the aforementioned methods to prevent overfitting.

5 Results

In this section, we will elaborate on the results of our model. After that, we will go over the individual results of each of the models. First, the HW2F model results will be discussed in Section 5.1, after which the LSTM model results will be discussed in Section 5.2, and lastly, our hybrid model results are shown in Section 5.3. After analysing the individual models, we will do a comparative study of the results in Section 5.4.

5.1 Results HW2F

In this section, we will discuss the results of the econometric HW2F model. We will first go over the results of our parameter analysis in Section 5.1.1, and continue in Section 5.1.2 on the HW2F model results.

5.1.1 Parameters

In Section 4.1, we have set up the environment for the HW2F model. We have discussed As concluded in that section, the architecture of the model determines the behaviour of the model. Hence, the settings of mean-reversion parameters a and b are very important to our HW2F model. During the training phase of our HW2F model, we will perform a scenario analysis for all the model setups where $0.01 \leq a, b \leq 1.0$. For each setup, we simulate the HW2F model many times, after which we will compare the best RMSE results for each of the setups. In Appendix E, we visualise the outcomes of the simulations. From this analysis, we have found that the model is best trained for values $a = 0.97$ and $b = 0.73$. This setting will therefore be used in the testing phase of the model, in order to find the most accurate out-of-sample prediction results. Furthermore, in Appendix F we have performed sensitivity analyses on both variables and find that these values minimise the IRSS prediction performance estimators. The values for a and b indicate a very strong mean reversion, to capture the nature of the IRSS.

5.1.2 Model

After setting up the model as discussed in previous sections, the results of the model are presented in Table 5.1.

Table 5.1: Performance Estimators HW2F

	RMSE	**MAE**	**MAPE**
Train	3.371	2.287	15.50%
Test	2.298	1.550	3.118%

Furthermore, in Figure 5.2, we see the prediction error per time step of the HW2F model. Most errors per time step are within a range of -5 bps and +5 bps, which indicates a generally stable forecasting accuracy. This can be caused due to the high mean-reversion, smoothing out the short-term fluctuations. However, when comparing the error per time step with the return rate per time step in Figure 5.1, we see that this has a negative correlation. The extremes in the return rate result in spike rates in the error, and hence, we expect a naive model, as the model is sensitive to non-linear conditions.

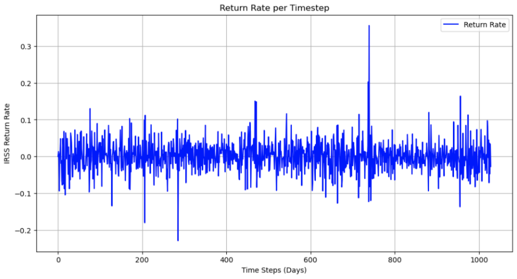


Figure 5.1: IRSS Return Rate Per Time Step

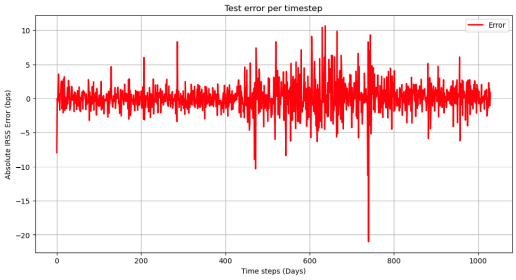


Figure 5.2: Residual Prediction Error per Time Step

When we then compare the results of the prediction with the observations of the IRSS data in Appendix G.1 and Figure 5.3, we see that the model closely fits the prediction in the next time step to the observation before. This could suggest that the model is using lazy naive forecasting, where $IRSS_{t+1} \approx IRSS_t$. However, if we look at Formula 2.10, this does seem normal, as we only calculate the small change in r, dr, from each day of the observations, as we are using a one-step ahead predictor.

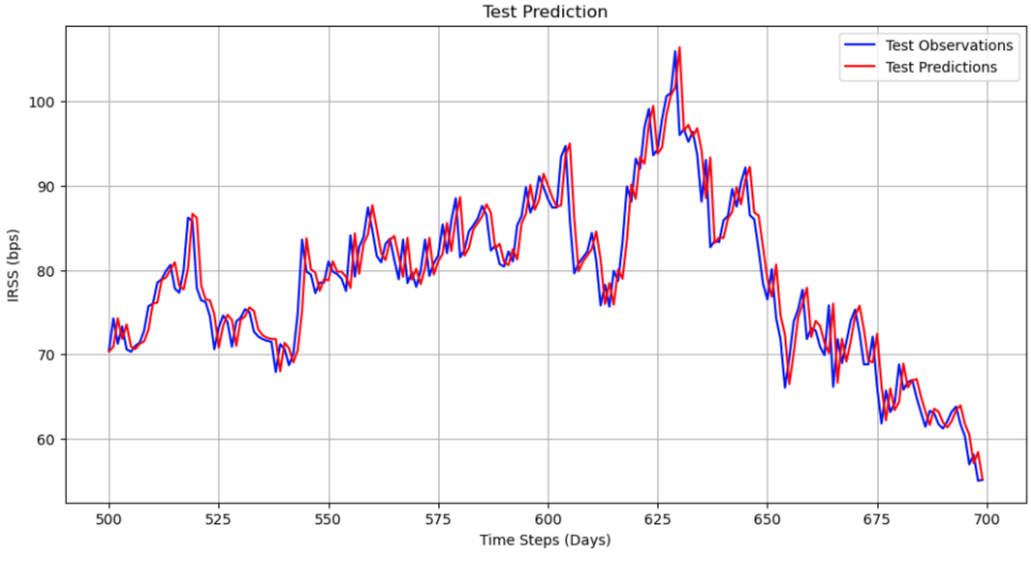


Figure 5.3: Snippet of Test Predictions (HW2F)

To make sure that the model is not a naive forecasting model, we test a naive predictor, which calculates the error for $IRSS_{t+1} = IRSS_t$. The results of the naive predictor on the test data, given in Table 5.2, we see that our HW2F model outperforms the naive predictor.

Table 5.2: Comparison Test Performance HW2F Model and Naive Predictor

	RMSE	MAE	MAPE
HW2F Model	2.298	1.550	3.118%
Naive Predictor	2.306	1.555	3.119%

Lastly, in Figure 5.5, we visualise the distribution of our prediction in a box plot. We can see that the box plot of the test error is narrower than the box plot of the train error, however, that can be explained by the higher volatility of the IRSS train data, as visualised in Figure 5.4. The comparable distribution of the underlying distributions indicates that the model is showing a similar performance to the naive model.

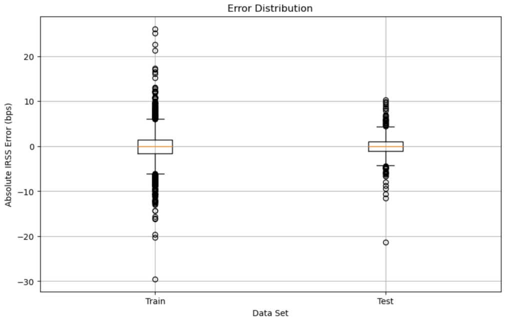


Figure 5.4: Naive Error Box-Plot

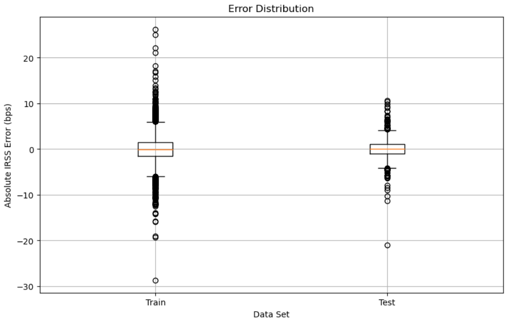


Figure 5.5: HW2F Error Box-Plot

5.2 Results LSTM

In this section, we will elaborate on the results of the LSTM model. First, we will present our findings of the hyperparameter analysis in Section 5.2.1, after which we will discuss the LSTM model results in Section 5.2.2.

5.2.1 Parameters

In section 4.2, we have found that the hyperparameters of our LSTM model have a significant effect on the output of our model. Hence, we have decided to use the Bayesian approximations hyperparameter tuner. Before the training phase of the model, this tuner will create the best environment for the model, during many iterations of the hyperparameter settings. The optimal hyperparameter settings are noted in Table 5.3.

Table 5.3: Optimal Hyperparameters Setting

Hyperparameter	**Optimal Value**
Number of layers	3
Batch size	32
Units in each layer	150, 200, 100
Epochs	200
Activation Function	tanh
Dropout rate	0.1, 0.3
Learning rate	0.0030

The hyperparameters suggest a moderate complexity of the model, with 3 LSTM layers having 150, 200 and 100 units respectively. The number of layers and units indicates a large capacity to learn complex patterns. As mentioned in Section 2.4, the tanh activation function allows the model to handle both positive and negative inputs, which is effective for our data as we use a standardisation that ranges from -1 to 1.

5.2.2 Model

During the first phase of the LSTM, we train the model using the hyperparameters from Table 4.1 and the inputs as mentioned in Section 4.2. During the epochs, we find that the model converges quickly as shown in Figure 5.6, after which the early stopping algorithm stops the training, to prevent overfitting. From the figure, we see a spike in loss at epoch 50, however, we can ignore this, as the loss stabilises again right after the peak.

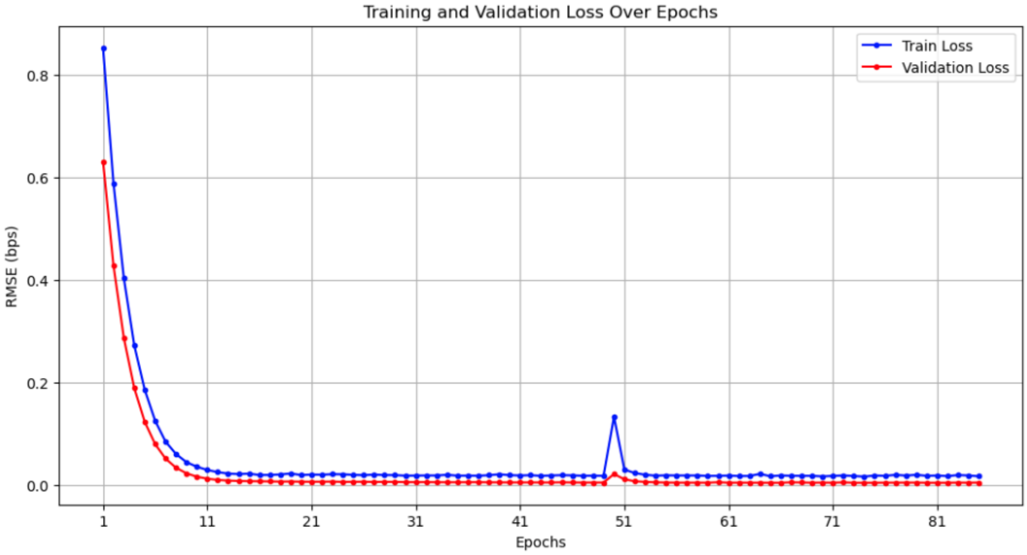


Figure 5.6: LSTM Loss per Epoch

The results of the LSTM model are presented in Table 5.4, which makes a split between the training, validation and testing data. It is evident that the value of the performance estimators is influenced by the number of observations in the set. Due to the fact that the training data set is bigger than the validation data set and the test data set, this has the effect that the performance estimators are significantly higher as well. It is therefore not possible to compare the results of one model, but as the other models will be set up in the comparable data splits, this creates an opportunity for cross-validating our model.

Table 5.4: Performance Estimators LSTM

	RMSE	MAE	MAPE
Train	3.195	2.249	12.50%
Validate	2.154	1.537	3.332%
Test	3.708	2.571	5.041%

In Figure 5.7, we see that the average error of the residual error per time step fluctuates within a range from -5 bps and +10 bps. We see that the residual error per time step has a trend, where the first 400 days of the forecast are accurately predicted, after which the error then increases. This is most likely caused by the extreme volatility of the IRSS in combination with the out-of-sample performance is too unpredictable for the model to understand. Furthermore, the spikes in the observations, visualised in Figure 5.1, are not fully picked up by the forecast of the LSTM model.

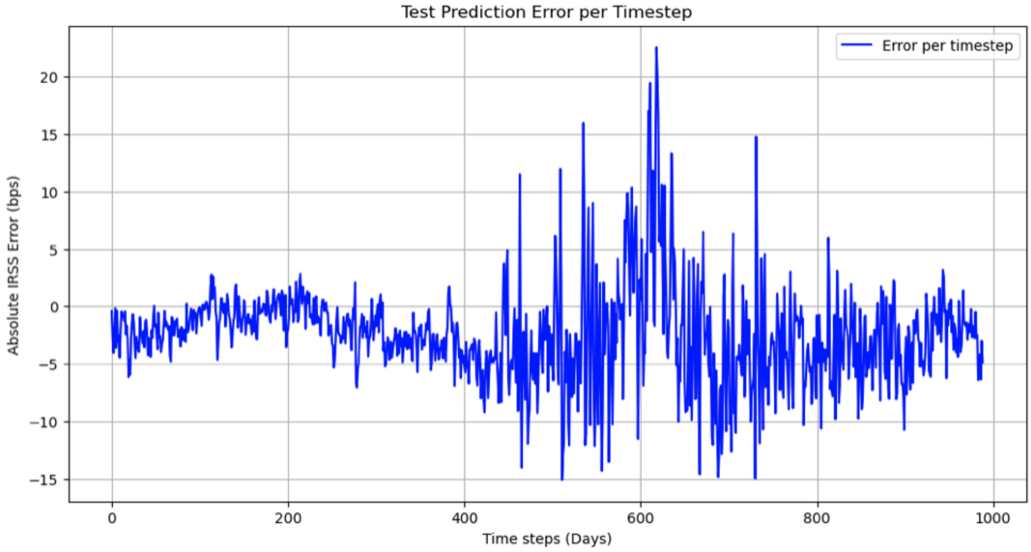


Figure 5.7: Residual Prediction Error per Time Step (LSTM)

Furthermore, Figure 5.8 and Figure 5.9, which show a small snippet of the output, show that the model closely follows the actual observations of the IRSS data, suggesting a good fit. It is visible in Appendix G.2, where the full test prediction is attached, that the model captures most fluctuations in the data, indicating effective learning and generalisation to the validation and testing set.

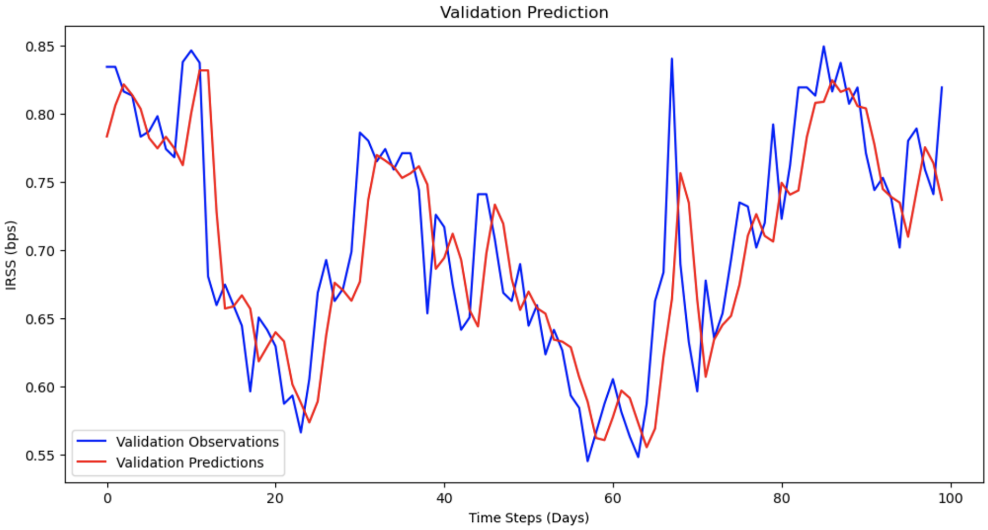


Figure 5.8: Snippet of Validation Predictions (LSTM)

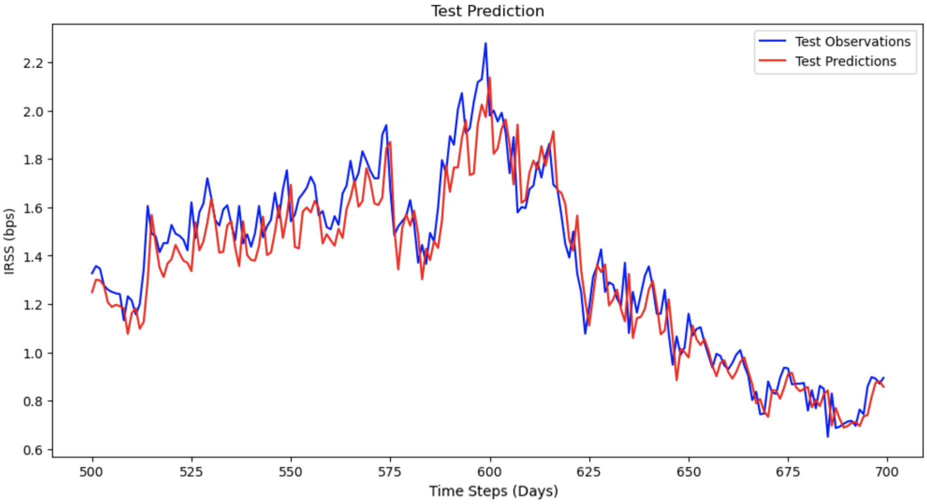


Figure 5.9: Snippet of Test Predictions (LSTM)

The analysis of these graphs indicates a model that performs well in capturing the general movements and volatility of IRSS in both the validation and the testing phases. The model's ability to fit close to the actual data in most of the observed range demonstrates its prediction accuracy.

Lastly, in Figure 5.11, we find the box plot of the residual errors generated by our LSTM model. It is noticeable that the box plot of the test error is narrower relative to the box plot of the volatility of the data splits as visible in Section 5.16. This implies that the model is functioning well in capturing the volatility of the IRSS data. However, we see a negative median, which implies a negative bias of the LSTM model. Lastly, the in-sample error has decreased significantly, which suggests risks for overfitting.

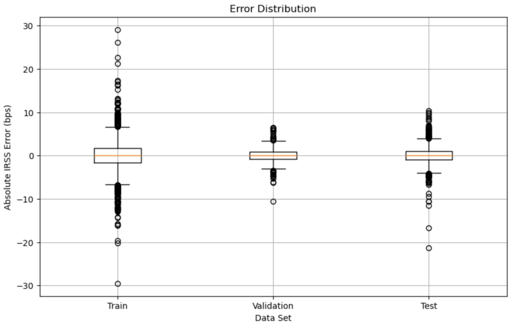


Figure 5.10: Naive Error Box-Plot

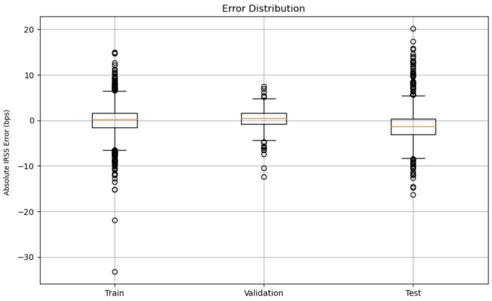


Figure 5.11: LSTM Error Box-Plot

5.3 Results Hybrid Model

In this section, we will go over the results of our hybrid model. We will start by presenting our findings on the hyperparameter settings in Section 5.3.1, after which we will continue discussing the results of the hybrid model in Section 5.3.2.

5.3.1 Parameters

As mentioned in Section 4.3, within the RC hybrid approach, we will first apply our knowledge from the HW2F model that was built to produce the most accurate IRSS forecast. As the model setting with the highest forecast accuracy has already been established in Section 5.1.1, we will use the same setup, where our mean reverting parameters a and b are 0.97 and 0.73, respectively.

After processing the error of the HW2F model, our LSTM model's inputs will now change significantly. Hence, we will again use the Bayesian hyperparameter tuner as explained in Section 4.3. The optimal hyperparameter settings for this model are shown in Table 5.5.

Table 5.5: Optimal hyperparameters setting

Hyperparameter	**Optimal Value**
Number of layers	2
Batch size	32
Units in each layer	100,50
Epochs	200
Activation Function	tanh
Dropout rate	0.1
Learning rate	0.0026

The model has two layers LSTM layers, with 100 and 50 units respectively. This architecture results in a simple model reducing computational complexity. The tanh activation function maps inputs to a range between -1 and 1, which fits the input and the output of our model as mentioned in Section 3.3. The learning rate is slightly conservative, ensuring a stable and precise convergence of the model during training.

5.3.2 Model

In this section, we will discuss the results of our proposed hybrid model, according to the specifications discussed above. As visible in Figure 5.12, the model is converging to its optimal performance, finding the most accurate forecasting for the IRSS rate. However, as mentioned, due to the setting of the hyperparameters, the convergence is relatively slow and stable.

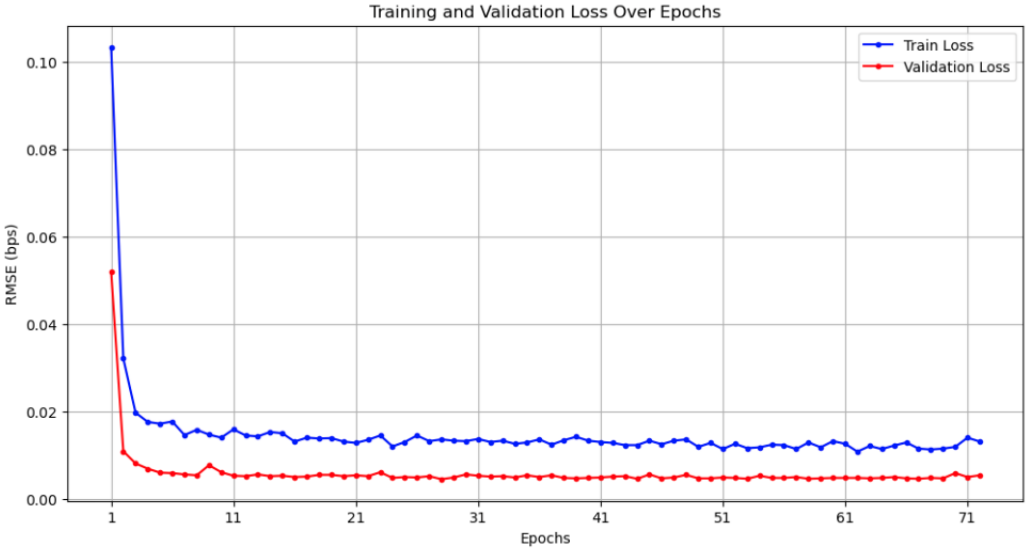


Figure 5.12: Hybrid Loss per Epoch

The results of the hybrid are visualised in Table 5.6. We see that the values among the train, validation and test data sets differ significantly, however, this can be traced back to the size of each of the data sets.

Table 5.6: Performance Estimators LSTM

	RMSE	MAE	MAPE
Train	3.599	2.478	17.55%
Validate	1.650	1.172	2.475%
Test	2.680	2.131	4.123%

When we analyse the error per time step, as visualised in Figure 5.13, we see that the prediction of the model is very stable up to time step 400. After this, as the IRSS time series gets less stationary, as seen in Figure 1.3, the prediction accuracy of the hybrid model decreases.

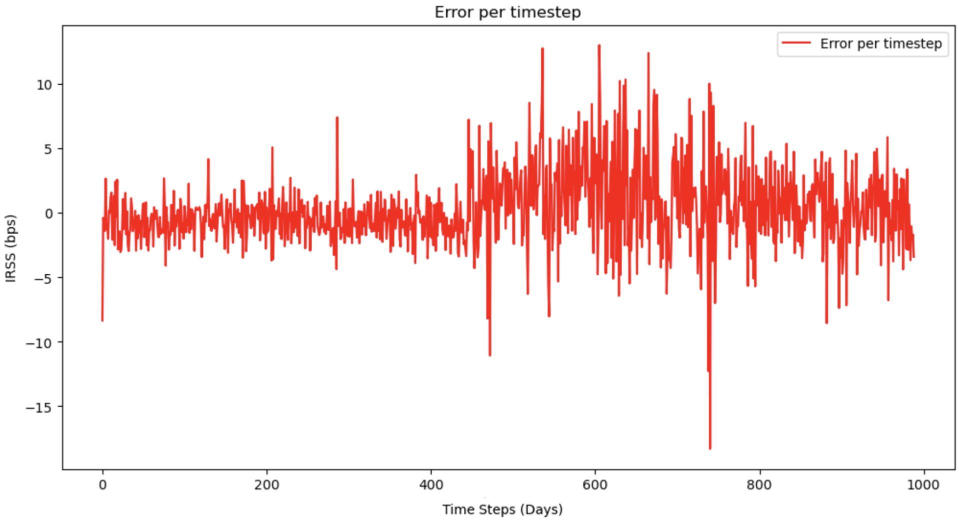


Figure 5.13: Hybrid Residual Error per Time Step

In Figure 5.14, we see that the validation loss is very small. The margins between the validation observations and the validation predictions are very small, as the model captures the movement of the IRSS data. However, when analysing Figure 5.15 we find that out-of-sample the model has more trouble with accurately fitting a prediction to the observation. The trends of the IRSS observations are generally followed well, however, the predictions are a lot more volatile. In Appendix G.3, we attach a full overview of the out-of-sample prediction.

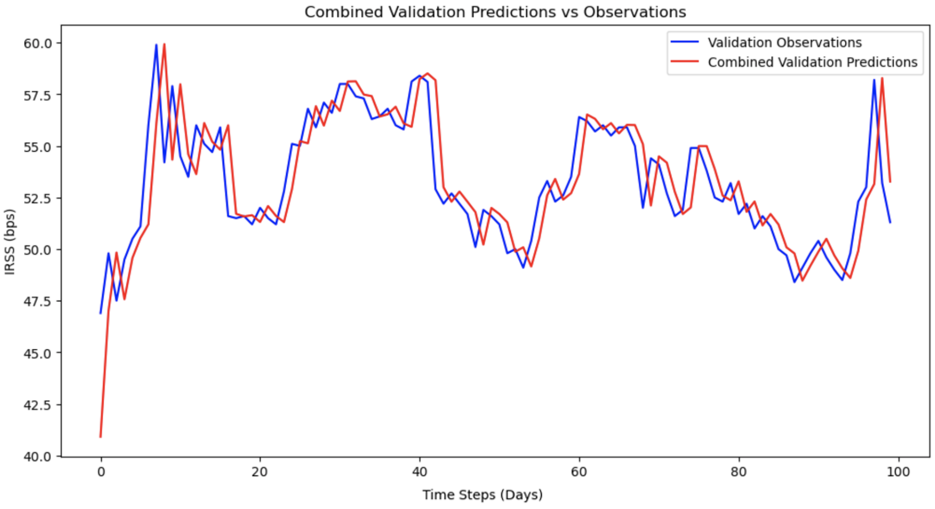


Figure 5.14: Hybrid Validation Predictions Snippet

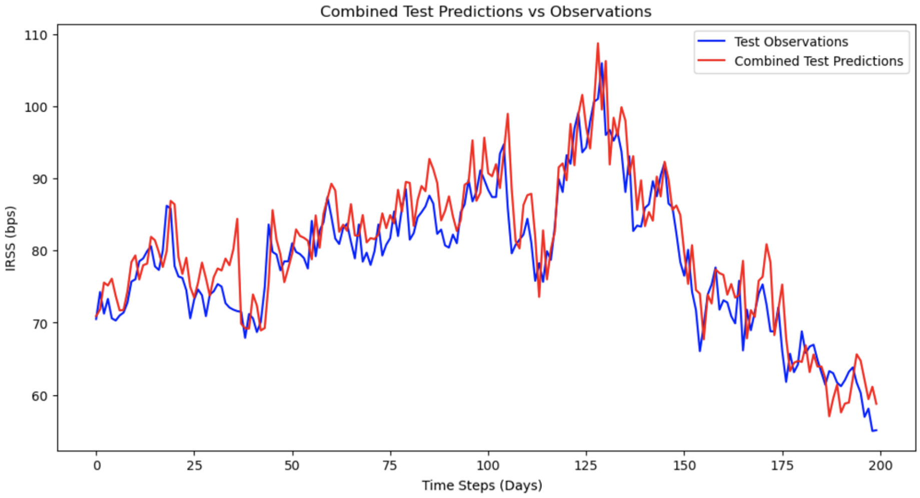


Figure 5.15: Hybrid Test Predictions Snippet

Lastly, we analyse the error distribution of the hybrid model in Figure 5.17, in comparison to the return rates of the data in Figure 5.16. Here we find that the in-sample performance of the model is very well, as the box plots of the training data set and the validation data set have been narrowed significantly. For the out-of-sample performance, we see a lot fewer extreme outliers, which implies a good predictive behaviour of the model.

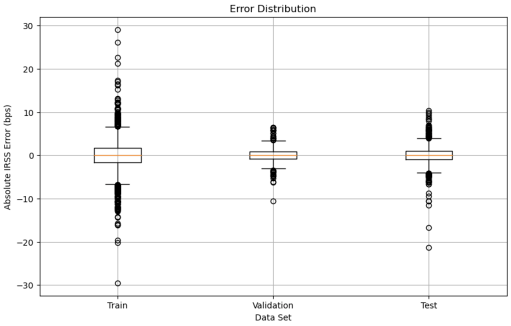


Figure 5.16: Naive Error Box-Plot

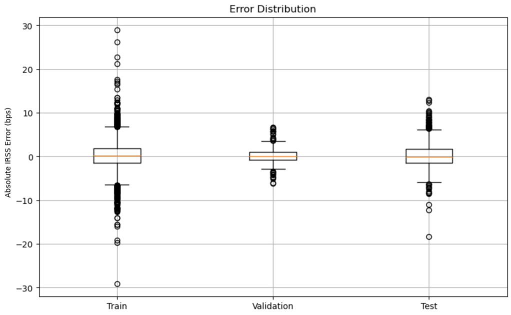


Figure 5.17: Hybrid Error Box-Plot

5.4 Comparative Analysis

In this section, we will perform a comparative analysis, based on all our findings. We will start by analysing the model complexity in Section 5.4.1, followed by the visual analysis in Section 5.4.2. After that, we will focus on the performance estimators as discussed in Section 5.4.3, and the error statistics in Section 5.4.4. Finally, we will analyse the relative performance per time step on out-of-sample data in Section 5.4.5.

5.4.1 Model complexity

We start analysing the complexity of the model by comparing the parameters of the models. Firstly, the HW2F model is relatively simple to set up, as the financial literature surrounding it already exists, requiring no further knowledge. After calibration of the parameters, the model's computational demand is minimal, making it an efficient choice for practical applications.

Furthermore, the LSTM model is more complex and challenging to set up. Its deeper architecture and higher number of units per layer lead to a larger number of parameters, allowing it to capture more intricate patterns in the data. This is shown in Table 5.7. This increased complexity, however, needs more computational resources and poses a higher risk of overfitting. Our LSTM model's increased complexity is highlighted by its deeper architecture and higher number of units per layer. This can also be found in the relatively quick convergence of the model, as shown in Figure 5.6, in comparison to the hybrid model's convergence in Figure 5.12.

Lastly, the hybrid model combines the setup of both HW2F and LSTM models, requiring further insights and understanding of both models. This dual setup increases its complexity and demands a comprehensive understanding of the underlying mechanisms of each component model.

Table 5.7: Optimal Hyperparameters Setting Comparison

Hyperparameter	**LSTM**	**Hybrid**
Number of layers	3	2
Batch size	32	32
Units in each layer	150,200,100	100,50
Epochs	200	200
Activation Function	tanh	tanh
Dropout rate	0.1, 0,3	0.1
Learning rate	0.0030	0.0026

5.4.2 Graphs

When analysing the model outputs, as visualised in the sections above, and Appendix G, we find that in terms of accuracy, the HW2F model shows the most accurate predictions out-of-sample. The LSTM model produces very volatile outputs in periods of stress, which are partly neutralised by our hybrid model. Furthermore, all the models capture the trend of the IRSS well, but the LSTM model cannot cope accurately with the volatility of the IRSS.

Based on the graphs, we find that the HW2F model is the most robust model in terms of out-of-sample performance, maintaining a close prediction with the IRSS observations across different periods. The Hybrid model performs reasonably well but shows more volatile outputs in stress periods. Lastly, the LSTM model is the least consistent, showing very significant deviations in the IRSS predictions.

5.4.3 Performance estimators

As discussed in Section 4.4, we will use three performance estimators to analyse the forecasting accuracy of the models. By analysing the in-sample performance of the models in Table 5.8, we find that the LSTM is the best-performing model. The LSTM model produces the lowest RMSE, MAE and MAPE, which indicates that it provides on average the most accurate predictions. Furthermore, the HW2F shows good performance, similar to the naive model, with slightly higher performance estimators than the LSTM model. The hybrid model has the lowest in-sample forecasting accuracy according to the performance estimators.

Table 5.8: In-Sample Estimators Comparison

	Naive	**HW2F**	**LSTM**	**Hybrid**
RMSE	3.403	3.371	3.195	3.599
MAE	2.306	2.287	2.249	2.478
MAPE	15.48%	15.50%	12.50%	17.55%

After analysing the out-of-sample performance in Table 5.9, we find that the HW2F produces the most accurate predictions for the IRSS. The HW2F model has the lowest RMSE, MAE and MAPE, which indicates that it performs the best on unseen data. Although the hybrid model performs worse than the HW2F model, the hybrid model seems to perform better than the LSTM model.

Table 5.9: Out-of-Sample Performance Estimators Comparison

	Naive	**HW2F**	**LSTM**	**Hybrid**
RMSE	2.306	2.298	3.708	2.680
MAE	1.555	1.550	2.571	2.131
MAPE	3.119%	3.118%	5.041%	4.123%

Altogether, the HW2F demonstrates the most consistent and reliable performance, for both in-sample and out-of-sample data. The LSTM model performs very well on in-sample data but is not able to produce accurate predictions on out-of-sample data. This suggests that the model is very prone to overfitting. Lastly, the hybrid model shows potential but needs further improvement to match the reliability of the HW2F model.

5.4.4 Error statistics

We continue our comparison by comparing the box plots provided in the sections above, which provide the distribution of the residual errors of the models. In Table 5.10, we analyse the distribution of the in-sample performance. The naive model contains the lowest mean IRSS error. The HW2F model shows a slight overestimation of the IRSS but has low variability in its predictions. The LSTM has the lowest mean error, and the lowest standard deviation, which indicates that the model provides the most accurate in-sample forecast. Additionally, the spread of the LSTM model, the distance between the maximum and minimum value, is also the lowest, implying the most consistent predictions. Lastly, the hybrid model performs less accurately than the other models in terms of the distribution of errors.

Table 5.10: In-Sample Residual IRSS Error Distribution (bps)

	Naive	**HW2F**	**LSTM**	**Hybrid**
Mean	0.0008	0.1058	-0.02596	0.1563
Std	3.403	3.379	3.019	3.596
Min	-29.49	-29.11	-33.23	-29.08
Max	26.10	26.04	15.02	29.00
Median	0	0.07416	0.1999	0.1425
Spread	55.59	55.15	48.25	58.08

Furthermore, in Table 5.11, we visualise the error distribution of the models on the out-of-sample date. Here, we find that the HW2F appears to be the most stable model with the lowest standard deviation and the most consistent prediction errors. The LSTM model lacks performance, as the variability of the error is significant in comparison to the other models, as can be seen from the spread of the model errors. The hybrid model performs slightly less accurately than the HW2F model, but it does have the lowest spread.

Table 5.11: Out-of-Sample Residual IRSS Error Distribution (bps)

	Naive	**HW2F**	**LSTM**	**Hybrid**
Mean	0	0.09475	-1.103	0.2548
Std	2.306	2.299	3.890	2.969
Min	-21.29	-21.01	-16.26	-18.31
Max	10.35	10.72	20.15	13.30
Median	0.0	0.03587	-1.246	-0.08422
Spread	31.64	31.73	36.41	31.61

In terms of consistency and stability, the hybrid model and the HW2F stand out as the most accurate models, both during in-sample and out-of-sample forecasting. These models show the lowest standard deviation and the lowest estimation bias. Furthermore, the LSTM produces the lowest average in-sample error but significantly lacks the ability to forecast unseen data. Although the hybrid shows do not match the accuracy of the HW2F in-sample performance, it does have a similar performance on out-of-sample data.

5.4.5 Relative performance

In Table 5.12, we find the individual comparison between two models, where we analyse what percentage of time steps the "row"-model outperforms the "column"-model, by having a smaller error than the other. Here we see that overall, the HW2F is the strongest model, by consistently outperforming the LSTM and the hybrid model, in the majority of the time steps. Furthermore, the results show that the hybrid model outperforms only the LSTM model, which is the least-performing model on out-of-sample data. The behaviours of the naive model and the HW2F model are very similar

Table 5.12: Percentage of Time Steps Each Individual Model Outperforms

	Naive	**HW2F**	**LSTM**	**Hybrid**
Naive	-	52%	72%	65%
HW2F	48%	-	73%	66%
LSTM	28%	27%	-	32%
Hybrid	25%	34%	68%	-

Furthermore, when we analyse the three models together, which outperform the out-of-sample data in Table 5.13, we see that the HW2F model still shows the highest percentage of lowest errors, followed by the hybrid model. Altogether, based on outperforming the other models per time step, we find that the HW2F model is the most robust model, and the weakest performance is shown by the LSTM model.

Table 5.13: Percentage of Time Steps All Models Outperform

	Outperforms
Naive	29%
HW2F	22%
LSTM	23%
Hybrid	27%

Lastly, in Table 5.14 we perform the Wilcoxon signed-rank test. The Wilcoxon signed-rank test is a statistical test used to compare two related samples, which do not follow normal distributions (Oyeka, Ebuh, et al. 2012). If the p-value $< \alpha$, the null hypothesis gets rejected, meaning that there is a significant difference between the samples. From the results, we see that there is no significant difference between the naive errors and the HW2F errors. The other pairs do have significant differences, meaning one model outperforms the other significantly. Based on this, we find that the naive model and the HW2F model are best performing, followed by the hybrid model, and lastly the LSTM model.

Table 5.14: Results Wilcoxon Signed-Rank Test with $\alpha = 0.05$

Model 1	**Model 2**	**P-Value**	α	**Reject H_0?**
Naive	HW2F	2.547e-01	0.05	No
Naive	LSTM	7.837e-63	0.05	Yes
Naive	Hybrid	8.561e-63	0.05	Yes
HW2F	LSTM	5.082e-64	0.05	Yes
HW2F	Hybrid	6.748e-32	0.05	Yes
LSTM	Hybrid	1.723e-23	0.05	Yes

5.4.6 Economical outperformance

Lastly, from the result comparison, we see that the naive model and the HW2F model have similar performance. In Table 5.8 and Table 5.9, we have found that the in-sample and out-of-sample performance estimators are similar, as well as the in-sample and out-of-sample IRSS error distributions in Table 5.10 and Table 5.11. Furthermore, we found in Table 5.13 that the two models do not significantly outperform each other. Hence, in this section, apart from the small statistical outperformance, we will analyse whether there is an economical outperformance of the model.

In the results for the hybrid model, we have used the IRSS prediction error of the HW2F model as an input, as visualised in Figure 4.5. Now, we will create a similar hybrid model, which uses the IRSS prediction error of the HW2F model as an input. By doing so, we get the following results of the performance estimators, as visualised in Table 5.15. We find that when using the HW2F model prediction error, we consistently outperform the hybrid model that uses the naive model prediction error as input, both in-sample and out-of-sample.

Table 5.15: Performance Estimators Comparison Hybrid Model with Different Input Data

	Naive (In-Sample)	**HW2F (In-Sample)**	**Naive (Out-of-Sample)**	**HW2F (Out-of-Sample)**
RMSE	3.634	3.599	3.019	2.680
MAE	2.541	2.478	2.168	2.131
MAPE	17.82%	17.55%	4.172%	4.123%

In Table 5.16, we find the distribution of the residual IRSS prediction error of the hybrid model, when using different input samples. We can see that despite the higher in-sample mean error, the hybrid model with the HW2F input has a better performance out-of-sample.

Table 5.16: Residual IRSS Error Distribution (bps) of Hybrid Model with Different Input Data

	Naive (In-Sample)	**HW2F (In-Sample)**	**Naive (Out-of-Sample)**	**HW2F (Out-of-Sample)**
Mean	0.06200	0.1563	0.3192	0.2548
Std	3.616	3.596	3.002	2.969
Min	-29.42	-29.08	-18.77	-18.31
Max	19.18	29.00	13.03	13.30
Median	0.05160	0.1425	0.08580	-0.08422
Spread	48.60	58.08	31.80	31.61

When applying the Wilcoxon signed-rank test to the data, we find that the P-value is less than $a = 0.05$, hence, we reject H_0. Hence, with this study, we have found that the HW2F model economically outperforms the naive model, within our hybrid approach study.

Table 5.17: Results Wilcoxon Signed-Rank Test with $\alpha = 0.05$ of Hybrid Model with Different Input Data

Input Model 1	**Input Model 2**	**P-Value**	α	**Reject H_0?**
Naive	HW2F	3.458e-02	0.05	Yes

5.5 Summary

In this section, we have discussed the results of our three models. We will base these results on three performance estimators: The RMSE, which measures the standard deviation of prediction errors and is sensitive to large outliers; The MAE, which reflects the absolute error and is more robust to outliers; and the MAPE, which expresses the accuracy of the models as a percentage, as is useful for comparing across scales.

The HW2F model is simple and efficient with minimal computational demand after parameter calibration. The LSTM model is more complex requiring deeper architecture and more computational demand. Lastly, the hybrid model combines the HW2F and the LSTM models, which increases the complexity of the model and requires an understanding of both models.

Furthermore, the HW2F model provides the most accurate out-of-sample predictions, based on the visual outputs of the models. The LSTM model shows very high volatility during stress periods, while the hybrid model neutralises a lot of this volatility, but is still less consistent than the HW2F model.

In terms of the performance estimators, on in-sample data, the LSTM model outperforms the other models. Although the HW2F model performs well on in-sample data, it excels the other models on out-of-sample data. Here, the hybrid model, however, shows improved performance in comparison to the in-sample performance of the model.

By analysing the box plots of the individual models, which give an understanding of the distribution of the residual error of the models per time step, we find that on in-sample prediction, the LSTM performs better than the other models. However, on out-of-sample data, we find that the HW2F model is the most stable with the lowest standard deviation. The hybrid model performs slightly less accurately than the HW2F.

Lastly, in terms of relative performance, we analysed what percentage of the time steps each model has had the lowest error in comparison to the other models. Here we have found that the HW2F model consistently outperforms the LSTM model and the hybrid model, from which the LSTM has the lowest overall relative performance.

Throughout the model comparisons, we have found very similar forecasting performance among the naive model and our HW2F model. We continued our research and used these two separate models as inputs to our hybrid model, and found that within our study, the results however do have an economic difference.

Based on these results, we have now found sufficient insights for answering Sub-Research Question III, Sub-Research Question V and Sub-Research Question VI, which will be discussed in the next section.

6 Conclusion, Discussion, and Future Research

In this paper, we studied the forecasting performance in predicting IRSS. Interest rate swaps are critical financial instruments for risk management, and accurate IRSS forecasting is crucial due to the predictive nature of IRSS regarding economic growth, inflation, and monetary policies. Given the growing complexity of swaps, it is essential to explore advanced methodologies beyond traditional econometric models. Hence, this research explores the forecasting accuracy of IRSS of machine learning models and, additionally, researches the potential benefits of using an integrative hybrid approach. Therefore, in this section, we will focus on concluding our findings in our research.

6.1 Conclusion

Our study was structured to individually address each of the sub-research questions before formulating a comprehensive answer to the main research question. Therefore, to conclude our thesis, we will first elaborate on the sub-research questions.

I *What factors are the predictors that influence the dynamics of IRSS?*

In a comprehensive literature study, we found that the predictors influencing the dynamics of IRSS include financial predictors, such as the Treasury yield curve, a zero-coupon bond yield, and the TEDS. Furthermore, macroeconomic predictors of IRSS are the GDP, the unemployment rate, and the inflation rate.

II *Which econometric models are currently being used in the forecasting of IRSS, and which are identified as the most accurate in literature?*

The econometric models currently used in forecasting IRSS found in our literature study include the VAR, ARIMA, GARCH, Vasicek, and Hull-White models. Among these, the HW2F model is identified as the most accurate in our study due to its ability to capture the dynamics of IRSS effectively. Furthermore, we have found that the HW2F model has an economical advantage in forecasting accuracy over the use of a naive model.

III *What are the assumptions and the limitations of these econometric models in the forecasting of IRSS?*

The econometric models assume that the predictors are stationary, linear and normally distributed. More specifically, the HW2F model relies on the assumption of constant mean-reversion parameters to estimate the interest rate movement, which may not hold true under all market conditions. Furthermore, after the model-building phase, we find that the model performs well under stable conditions, but shows spikes in prediction error during periods of high IRSS return rate volatility. Lastly, the model prediction closely aligns with the previous value, suggesting a form of naive forecasting.

IV *Which machine learning models are expected to perform with the highest forecasting accuracy in the forecasting of IRSS?*

During our literature study, by studying random forest algorithms, GBM techniques and NN models, we found that NN will fit our IRSS prediction the best, as it can identify complex non-linear relationships from high-dimensional input data. From additional research, we found that LSTM models show the highest forecasting potential, as they can handle different sequence lengths and mitigate the vanishing gradient problem from RNN.

V *What are the assumptions and the limitations of these machine learning models in the forecasting of IRSS?*

In our study, we found that the LSTM model requires significant computational resources and very careful hyperparameter tuning. We have found that although incorporating overfitting prevention techniques, predicting IRSS is too difficult for LSTM, as it is still prone to overfitting, causing the model to lack forecasting accuracy on out-of-sample predictions. This results in very high fluctuating error rates in our analysis.

VI *How can a hybrid model that integrates econometric and machine learning approaches be optimally designed for forecasting IRSS?*

For forecasting IRSS, we have found that an optimal hybrid can be designed by combining the initial forecasts from the HW2F econometric model with error corrections predicted by the LSTM model, using the RC methodology found in our literature study. This approach leverages the strengths of both models, with the HW2F providing stability and financial knowledge, and the LSTM capturing complex patterns. During evaluation, the hybrid model demonstrated improved forecasting accuracy compared to the LSTM model alone, reducing volatility in out-of-sample data. Specifically, the hybrid model achieved an RMSE of 2.680, MAE of 2.131, and MAPE of 4.123%, which were all improvements over the LSTM model's RMSE of 3.708, MAE of 2.571, and MAPE of 5.041%. However, despite these improvements, the hybrid model did not consistently outperform the HW2F model, which had an RMSE of 2.298, MAE of 1.550, and MAPE of 3.118%. Notably, the hybrid model still exhibited more volatility during periods of market stress than the HW2F model. Furthermore, the increased model complexity and the higher computational demand are both limiting the performance of the model.

After addressing the sub-research questions individually, and finding the answer to them in our research, we can now conclude our research by considering the main research question, stated below:

How do the predictive capacities of traditional financial models contrast with those of machine learning algorithms in one-step-ahead forecasting IRSS, and would an integrative approach using a hybrid model enhance the overall forecasting accuracy?

Our analysis has found that the HW2F model consistently provides an accurate out-of-sample forecast. We have found that the HW2F maintains a very stable prediction, due to the high mean-reversion parameters that match the behaviour of the IRSS fluctuations. Furthermore, due to the characteristics of the HW2F being a relatively simple model with very low computational demand, we find that the HW2F contrasts with our machine learning model. The LSTM model, with high computational demand and a high risk of overfitting, showed a very high volatility in prediction error in stress periods. Although it performs very well in in-sample data, the forecast on out-of-sample data shows very significant volatility, indicating an overfitting model.

The objective of this paper was to propose a hybrid model aimed at leveraging the strengths of the HW2F model and the LSTM model. However, using the setup discussed in this paper, the hybrid model did not surpass the performance of the HW2F model in terms of out-of-sample stability and accuracy. In our research, we have found that in terms of the model complexity, the visual outputs, the performance estimators and the relative performance of the models, the hybrid model neutralises the out-of-sample error volatility produced by the individual LSTM model.

To conclude, this study highlights the importance of balancing complexity and stability in forecasting models and suggests that traditional econometric models remain highly valuable tools, despite the rapid growth of machine learning tools in financial forecasting.

6.2 Discussion

During our study, we have made assumptions, that can have a significant effect on the outcome of our study. Hence, in order to ensure the replicability and validity of our study, in this section we will discuss these assumptions. The first assumption of our study is that the market functions efficiently. As discussed in Section 2.1, changes in IRSS are built up from two elements: The financial element and the macroeconomic element. This assumption is built upon the fact that the market prices fully reflect all the available information. However, markets can exhibit inefficiencies due to factors like regulatory changes or behavioural biases. During our study, inefficiencies are taken out of the equation when forecasting the IRSS.

Throughout our research, we have found that we were limited due to the very limited research that has currently been performed on forecasting IRSS. Due to the limited research, we had to make assumptions in our methodology for all three models. The HW2F model has been used in a limited number of studies to forecast IRSS. Despite this, it has been used to forecast interest rates in multiple studies. Hence, we have used the assumption that as IRSS is an interest rate-driven derivative, this model would function well for our IRSS forecast. NN has not been used for forecasting IRSS, and hence, similarly to the HW2F model, we assume that its forecasting results on similar forecast such as interest rate and CDS, imply that there is potential in forecasting IRSS using NN. Lastly, the subject of hybrid models using both econometric and machine learning models has not been discussed in financial forecasting.

Furthermore, as discussed in Section 3, the structural breaks made between the training/validation/testing set of the predictors might not have been optimal. The data limitation due to the COVID-19 pandemic made sure that the data on which the models have been trained did not closely represent the stress period of the data after the breaks, affecting the ability of the model to capture the financial and macroeconomic movements of the world.

Lastly, another limitation of this study has to do with the characteristics of a NN. The combination of the black box effect and the computational demand of the model limited this research in the ability to fine-tune the model. After iterating the, and finding non-optimal performance, tuning the model and computing new results would often take up a lot of time. Furthermore, due to the black box effect, we weren't always sure whether the adaption of the model would have a positive effect. After training our NN, the model created weights that are used to make a forecast and validate the forecasting accuracy of the NN. However, these weights are not transparent, nor can they be changed manually. This makes it complex to determine the significance of the different inputs that were used in our NN. Additionally, due to the very large size of our multidimensional input data, the trial and error part of the hybrid model took a lot of computational effort. Despite using techniques like cross-validation and dropout ratio, our LSTM model and our hybrid model constantly showed the behaviour of overfit on the training data. Due to this high computational effort, the study failed to overcome this problem, resulting in low forecasting accuracy of both the LSTM model and the hybrid model.

6.3 Contribution to Literature

As discussed in Section 1.4, this research was aimed to address the gap in the financial knowledge regarding IRSS. The currently existing literature focuses on utilising econometric models for IRSS forecasting. Our contribution aimed at introducing a machine learning model to this literature, combined with an integrative hybrid model.

We studied an area without extensive research in financial literature, hybrid models. Multiple models were examined, and an elaborate framework was built to integrate the strengths of individual models. Furthermore, we applied a comparative study providing insights into the current performance of IRSS forecasting across various models, establishing a framework for comparing multiple models. Additionally, we provided an extensive methodology for data preparation and model calibration, enhancing the replicability and robustness of future forecasting studies.

Our findings in this study confirm the stability and reliability of traditional econometric models like HW2F over newer machine learning techniques, particularly in out-of-sample forecasting. The study showed that the machine learning techniques underperform traditional econometric models, as they highlight the overfitting and volatility risks of LSTM models.

6.4 Recommendations

Based on the conclusions drawn from the study, and the limitations that were drawn within the discussion, we now focus on formulating recommendations for the company and for future research.

Firstly, we recommend continuing the study for further model enhancements. Within our methodology, we made assumptions that scope our research. For future research, we suggest exploiting multiple extensions to the research, such as using other strategies to optimise the hyperparameters. As we have used the Bayesian approximations hyperparameter tuner, we use a technique with a high computational demand. As discussed in Section 6.2, this computational demand has limited us in our research, so by using other hyperparameter tuners, we create the opportunity to further enhance our model. Furthermore, in further research, we recommend further analysis of feature engineering, by introducing lags, moving averages or Sharpe ratios. By performing feature engineering, we aim for our model to better understand the dynamics of the real world. Additionally, in order to find improved forecasting accuracy, we suggest exploiting more NN models, en more hybrid models. Lastly, we have currently investigated one-step ahead forecasting. After increasing the forecasting accuracy of the one-step ahead hybrid model, multi-step forecasting will become an opportunity.

Our second recommendation is to work on the interpretability of the hybrid model's prediction. Given the complexity of combining econometric and machine learning techniques, for the risk management of financial institutions, it could be critical to ensure that the results are understandable to both technical and non-technical stakeholders. Therefore, we recommend the exploitation of explainable Artificial Intelligence (XAI) techniques, which aim to make the model's decision-making process more transparent. This will increase the trust and adoption of hybrid models for financial institutions.

Lastly, the sequential nature of the hybrid model, where the LSTM model corrects the HW2F error, assumes that error patterns are consistent and predictable, which turned out to not be the case. Additionally, due to the sequential nature of the hybrid model, we assume that the optimal setup of the individual HW2F model found in the grid search will imply the same optimal setup of the HW2F model for the hybrid model. Hence, further research should include scenario analyses on multiple model settings, in order to improve the hybrid forecasting accuracy.

References

Amihud, Yakov and Haim Mendelson (1986). “Liquidity and stock returns”. In: *Financial Analysts Journal* 42.3, pp. 43–48.

Bae, Kee-Hong, G Andrew Karolyi, and Reneé M Stulz (2003). “A new approach to measuring financial contagion”. In: *The Review of Financial Studies* 16.3, pp. 717–763.

Bakhashwain, Norah and Alaa Sagheer (2021). “Online Tuning of Hyperparameters in Deep LSTM for Time Series Applications.” In: *International Journal of Intelligent Engineering & Systems* 14.1.

Batool, Komal, Mirza Faizan Ahmed, and Muhammad Ali Ismail (2022). “A Hybrid Model of Machine Learning Model and Econometrics’ Model to Predict Volatility of KSE-100 Index”. In: *Reviews of Management Sciences* 4.1, pp. 225–239.

Beenstock, Michael and Kam-Fai Chan (1988). “Economic forces in the London stock market”. In: *Oxford Bulletin of Economics and Statistics* 50.1, pp. 27–39.

Bicksler, James and Andrew H Chen (1986). “An economic analysis of interest rate swaps”. In: *The Journal of Finance* 41.3, pp. 645–655.

Bishop, Christopher M and Nasser M Nasrabadi (2006). *Pattern recognition and machine learning*. Vol. 4. 4. Springer.

Blanchard, Arnaud (2014). “The two-factor Hull-White model: pricing and calibration of interest rates derivatives”. In: *KTH Royal Insitute of Technology*.

Brown, Keith Cates, William Van Harlow, and Donald J Smith (1991). *An empirical analysis of interest rate swap spread*. Boston University, School of Management.

Brown, Rob, Francis In, and Victor Fang (2002). “Modeling the determinants of swap spreads”. In.

Caks, John (1977). “The coupon effect on yield to maturity”. In: *The Journal of Finance* 32.1, pp. 103–115.

Chen, Shiu-Sheng (2009). “Predicting the bear stock market: Macroeconomic variables as leading indicators”. In: *Journal of Banking & Finance* 33.2, pp. 211–223.

Cho, Hyunghun et al. (2020). “Basic enhancement strategies when using Bayesian optimization for hyperparameter tuning of deep neural networks”. In: *IEEE access* 8, pp. 52588–52608.

Cortes, Fabio (2003). “Understanding and modelling swap spreads”. In: *Bank of England Quarterly Bulletin, Winter*.

Cui, Zhicheng, Wenlin Chen, and Yixin Chen (2016). “Multi-scale convolutional neural networks for time series classification”. In: *arXiv preprint arXiv:1603.06995*.

DeepAI (May 2019). *Feed Forward Neural Network*. URL: https://deepai.org/machine-learning-glossary-and-terms/feed-forward-neural-network.

Derbentsev, Vasily et al. (2020). “Machine learning approaches for financial time series forecasting”. In: CEUR Workshop Proceedings.

Devi, Monika et al. (2021). “Forecasting of wheat production in Haryana using hybrid time series model”. In: *Journal of Agriculture and Food Research* 5, p. 100175.

Do, Trong-Hop, Demeke Shumeye Lakew, Sungrae Cho, et al. (2022). “Building a Time-Series Forecast Model with Automated Machine Learning for Heart Rate Forecasting

Problem". In: *2022 13th International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, pp. 1097–1100.

Duarte, Jefferson, Francis A Longstaff, and Fan Yu (2007). "Risk and return in fixed-income arbitrage: Nickels in front of a steamroller?" In: *The Review of Financial Studies* 20.3, pp. 769–811.

Duffie, Darrell and Kenneth J Singleton (1997). "An econometric model of the term structure of interest-rate swap yields". In: *The Journal of Finance* 52.4, pp. 1287–1321.

Fine, Terrence L (2006). *Feedforward neural network methodology*. Springer Science & Business Media.

Firat, Orhan et al. (2017). "Multi-way, multilingual neural machine translation". In: *Computer Speech & Language* 45, pp. 236–252.

Gal, Yarin and Zoubin Ghahramani (2016). "A theoretically grounded application of dropout in recurrent neural networks". In: *Advances in neural information processing systems* 29.

Gargano, Antonio and Allan Timmermann (2014). "Forecasting commodity price indexes using macroeconomic and financial predictors". In: *International Journal of Forecasting* 30.3, pp. 825–843.

Giglio, Stefano (2016). *Credit default swap spreads and systemic financial risk*. Tech. rep. ESRB working paper series.

Goodfellow, Ian et al. (2013). "Multi-prediction deep Boltzmann machines". In: *Advances in Neural Information Processing Systems* 26.

Gorgolis, Nikolaos et al. (2019). "Hyperparameter optimization of LSTM network models through genetic algorithm". In: *2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA)*. IEEE, pp. 1–4.

He, Yong et al. (2023). "Application of LSTM model optimized by individual-ordering-based adaptive genetic algorithm in stock forecasting". In: *International Journal of Intelligent Computing and Cybernetics* 16.2, pp. 277–294.

Hejazi, Walid, Huiwen Lai, and Xian Yang (2000). "The expectations hypothesis, term premia, and the Canadian term structure of interest rates". In: *Canadian Journal of Economics/Revue canadienne d'économique* 33.1, pp. 133–148.

Hochreiter, Sepp and Jürgen Schmidhuber (1997). "Long short-term memory". In: *Neural computation* 9.8, pp. 1735–1780.

Huang, Jian, Junyi Chai, and Stella Cho (2020). "Deep learning in finance and banking: A literature review and classification". In: *Frontiers of Business Research in China* 14.1, p. 13.

Huang, Ying, Carl R Chen, and Maximo Camacho (2008). "Determinants of Japanese Yen interest rate swap spreads: Evidence from a smooth transition vector autoregressive model". In: *Journal of Futures Markets: Futures, Options, and Other Derivative Products* 28.1, pp. 82–107.

Hull, John (1993). *Options, futures, and other derivative securities*. Vol. 7. Prentice Hall Englewood Cliffs, NJ.

— (1996). "Using Hull-White interest rate trees". In: *Journal of Derivatives* 3.3, pp. 26–36.

— (2012). *Risk management and financial institutions*. Vol. 733. John Wiley & Sons.

Hull, John and Alan White (2001). "The general Hull–White model and supercalibration". In: *Financial Analysts Journal* 57.6, pp. 34–43.

ING (2022). *Annual Report.*

Inoue, Shiori, Shota Miki, and Yasufumi Gemma (2021). "The Japanese Yen Interest Rate Swap Market in the Time of COVID-191". In: *Proceedings 63rd ISI World Statistics Congress*. Vol. 11, p. 16.

Investing.com (Feb. 2024). *EUR 10 years IRS interest rate swap Bond Historical Data*. URL: https://www.investing.com/rates-bonds/eur-10-years-irs-interest-rate-swap-historical-data.

Jagero, Barry Agingu, Thomas Mageto, and Samuel Mwalili (2023). "Modelling and Forecasting Inflation Rates in Kenya Using ARIMA-ANN Hybrid Model". In: *American Journal of Neural Networks and Applications* 9.1, pp. 8–17.

Jeans, Nathalie (Jan. 2019). *How I classified images with Recurrent Neural Networks*. URL: https://medium.com/@nathaliejeans/how-i-classified-images-with-recurrent-neural-networks-28eb4b57fc79.

Kaastra, Iebeling and Milton Boyd (1996). "Designing a neural network for forecasting financial and economic time series". In: *Neurocomputing* 10.3, pp. 215–236.

Khan, Faridoon, Amena Urooj, and Sara Muhammadullah (2021). "An ARIMA-ANN hybrid model for monthly gold price forecasting: empirical evidence from Pakistan". In: *Pakistan Econ Rev* 4.1, pp. 61–75.

Kim, Ha Young and Chang Hyun Won (2018). "Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models". In: *Expert Systems with Applications* 103, pp. 25–37.

Kim, Soyoung and Nouriel Roubini (2000). "Exchange rate anomalies in the industrial countries: A solution with a structural VAR approach". In: *Journal of Monetary economics* 45.3, pp. 561–586.

Kontopoulou, Vaia I. et al. (2023). "A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks". In: *Future Internet* 15.8. ISSN: 1999-5903. URL: https://www.mdpi.com/1999-5903/15/8/255.

Kumar, Manish and M Thenmozhi (2006). "Forecasting stock index movement: A comparison of support vector machines and random forest". In: *Indian institute of capital markets 9th capital markets conference paper*.

Kurpiel, A (2003). "Forecasting Swap Spread Dynamics". In: *Societe Generale*.

Larochelle, Hugo et al. (2012). "Learning algorithms for the classification restricted Boltzmann machine". In: *The Journal of Machine Learning Research* 13.1, pp. 643–669.

Lekkos, Ilias and Costas Milas (2001). "Identifying the factors that affect interest-rate swap spreads: Some evidence from the United States and the United Kingdom". In: *Journal of Futures Markets: Futures, Options, and Other Derivative Products* 21.8, pp. 737–768.

— (2004). "Common risk factors in the US and UK interest rate swap markets: Evidence from a nonlinear vector autoregression approach". In: *Journal of Futures Markets: Futures, Options, and Other Derivative Products* 24.3, pp. 221–250.

Lopez, J Humberto (1997). "The power of the ADF test". In: *Economics Letters* 57.1, pp. 5–10.

Mao, Weifang et al. (2023). “Forecasting and trading credit default swap indices using a deep learning model integrating Merton and LSTMs”. In: *Expert Systems with Applications* 213, p. 119012.

McNulty, James E (1990). “The pricing of interest rate swaps”. In: *Journal of Financial Services Research* 4, pp. 53–63.

Mucaj, Roneda and Valentina Sinaj (2017). “Exchange Rate Forecasting using ARIMA, NAR and ARIMA-ANN Hybrid Model”. In: *Exchange* 4.10, pp. 8581–8586.

Musa, Yakubu and Stephen Joshua (2020). “Analysis of ARIMA-artificial neural network hybrid model in forecasting of stock market returns”. In: *Asian Journal of Probability and Statistics* 6.2, pp. 42–53.

Nguyen, Vu (2019). “Bayesian optimization for accelerating hyper-parameter tuning”. In: *2019 IEEE second international conference on artificial intelligence and knowledge engineering (AIKE)*. IEEE, pp. 302–305.

Nouri, Daniel (2014). “Using deep learning to listen for whales”. In: *URl: http://danielnouri.org/notes/2014/01/10/using-deep-learning-to-listen-for-whales.*

Orlando, Giuseppe, Rosa Maria Mininni, and Michele Bufalo (2020). “Forecasting interest rates through Vasicek and CIR models: A partitioning approach”. In: *Journal of Forecasting* 39.4, pp. 569–579.

Oyeka, Ikewelugo Cyprian Anaene, Godday Uwawunkonye Ebuh, et al. (2012). “Modified Wilcoxon signed-rank test”. In: *Open Journal of Statistics* 2.2, pp. 172–176.

Pawar, Kriti, Raj Srujan Jalem, and Vivek Tiwari (2019). “Stock market price prediction using LSTM RNN”. In: *Emerging Trends in Expert Applications and Security: Proceedings of ICETEAS 2018*. Springer, pp. 493–503.

Pereira, John (2015). “An empirical investigation of corporate credit default swap spreads and returns”. PhD thesis. Kingston University.

Prince, Simon J.D. (2023). *Understanding Deep Learning*. The MIT Press. URL: http://udlbook.com.

Rabobank (2022). *Annual Report*.

Reimers, Nils and Iryna Gurevych (2017). “Optimal hyperparameters for deep lstm-networks for sequence labeling tasks”. In: *arXiv preprint arXiv:1707.06799*.

Rodríguez, Merche Galisteo, Isabel Morillo López, and Teresa Preixens Benedicto (2024). “CVA with wrong-way risk and correlation between defaults: An application to an interest rate swap”. In: *Revista de Economía y Finanzas*. URL: https://api.semanticscholar.org/CorpusID:267958859.

Roy, Sanjiban Sekhar et al. (2020). “Random forest, gradient boosted machines and deep neural network for stock price forecasting: a comparative analysis on South Korean companies”. In: *International Journal of Ad Hoc and Ubiquitous Computing* 33.1, pp. 62–71.

Rumelhart, David E and David Zipser (1985). “Feature discovery by competitive learning”. In: *Cognitive science* 9.1, pp. 75–112.

Seppälä, Heikki, Ser-Huang Poon, and Thomas Schröder (2013). “Closed Form Approximation of Swap Exposures”. In: *Econometrics: Mathematical Methods & Programming eJournal*. URL: https://api.semanticscholar.org/CorpusID:150836581.

Shumway, Robert H et al. (2017). “ARIMA models”. In: *Time series analysis and its applications: with R examples*, pp. 75–163.

Siami-Namini, Sima, Neda Tavakoli, and Akbar Siami Namin (2019). “The performance of LSTM and BiLSTM in forecasting time series”. In: *2019 IEEE International conference on big data (Big Data)*. IEEE, pp. 3285–3292.

Smith Jr, Clifford W, Charles W Smithson, and Lee Macdonald Wakeman (1988). “The market for interest rate swaps”. In: *Financial Management*, pp. 34–44.

Song, Xuanyi et al. (2020). “Time-series well performance prediction based on Long Short-Term Memory (LSTM) neural network model”. In: *Journal of Petroleum Science and Engineering* 186, p. 106682.

Sun, Tong-Sheng, Suresh Sundaresan, and Ching Wang (1993). “Interest rate swaps: An empirical investigation”. In: *Journal of Financial Economics* 34.1, pp. 77–99.

Tashmit (June 2023). *Understanding an RNN cell.* URL: https://www.codingninjas.com/studio/library/understanding-an-rnn-cell.

TBIS (Oct. 2022). *OTC interest rate derivatives turnover in April 2022.* URL: https://www.bis.org/statistics/rpfx22_ir.htm.

Tobin, James (1965). “Money and economic growth”. In: *Econometrica: Journal of the Econometric Society*, pp. 671–684.

Trippi, Robert R and Efraim Turban (1992). *Neural networks in finance and investing: Using artificial intelligence to improve real world performance.* McGraw-Hill, Inc.

Vasicek, Oldrich (1977). “An equilibrium characterization of the term structure”. In: *Journal of financial economics* 5.2, pp. 177–188.

Vukovic, Darko B et al. (2022). “Are CDS spreads predictable during the Covid-19 pandemic? Forecasting based on SVM, GMDH, LSTM and Markov switching autoregression”. In: *Expert systems with applications* 194, p. 116553.

Wang, Sheng et al. (2015). *Protein secondary structure predictions using deep convolutional neural fields.* eprint: 1512.00843.

Watson, Mark W (2003). “Macroeconomic forecasting using many predictors”. In: *Econometric Society Monographs* 37, pp. 87–114.

Xiong, Rick Yuankang et al. (2019). “Forecasting credit spreads: A machine learning approach”. In: *Semantic Scholar*.

Zafeirelli, Sofia and Dimitris Kavroudakis (2024). “Comparison of outlier detection approaches in a Smart Cities sensor data context”. In: *International Journal on Smart Sensing and Intelligent Systems.* URL: https://api.semanticscholar.org/CorpusID:267687278.

Zahn, Rebecca et al. (2021). “Application of a long short-term memory neural network for modeling transonic buffet aerodynamics”. In: *Aerospace Science and Technology* 113, p. 106652.

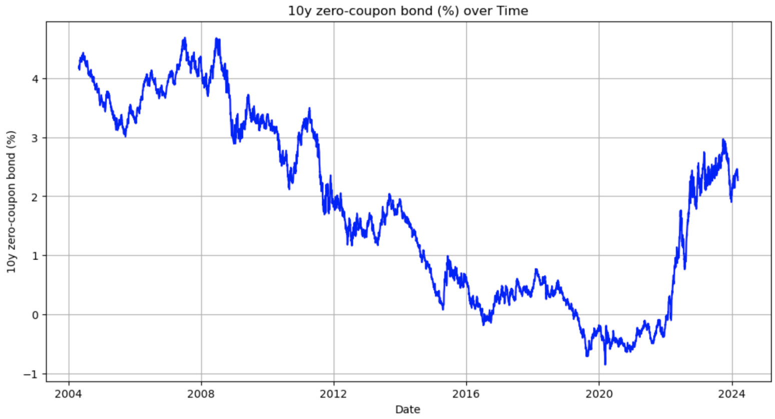
Zhang, G Peter and Min Qi (2005). “Neural network forecasting for seasonal and trend time series”. In: *European journal of operational research* 160.2, pp. 501–514.

Zhang, Ke and Bing Liang (2008). “An Empirical Analysis for Determinants of Interest Rate Swap Spread”. In: URL: https://api.semanticscholar.org/CorpusID:154334225.

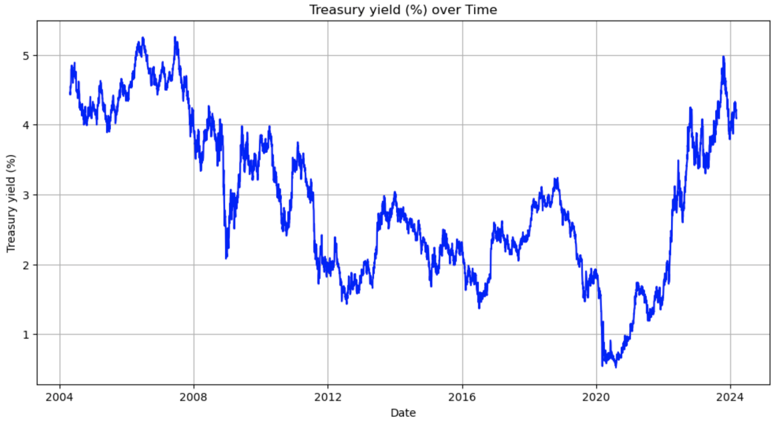
Appendices

A Data Trends

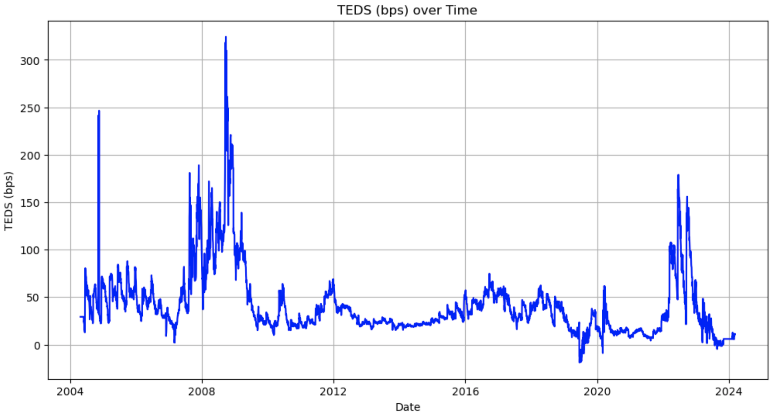
A.1 10-Year Zero-Coupon Bond



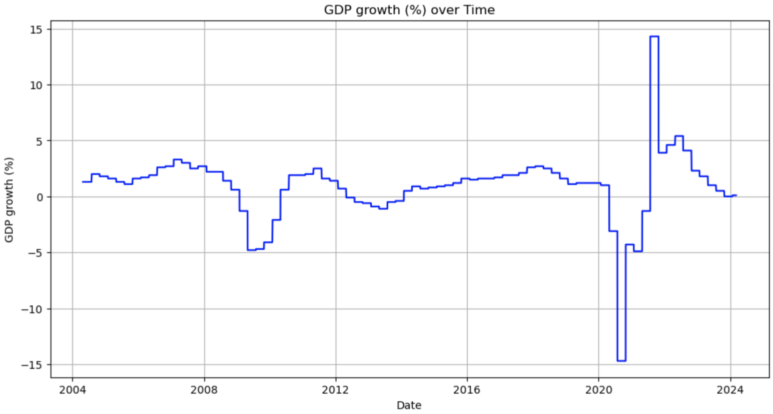
A.2 Treasury Yield



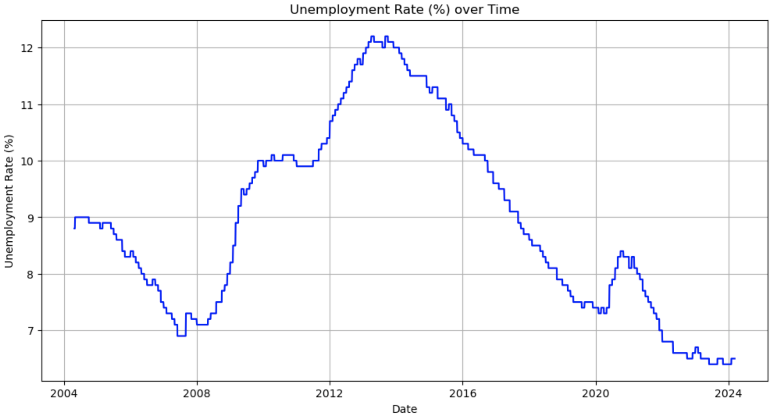
A.3 TED Spread



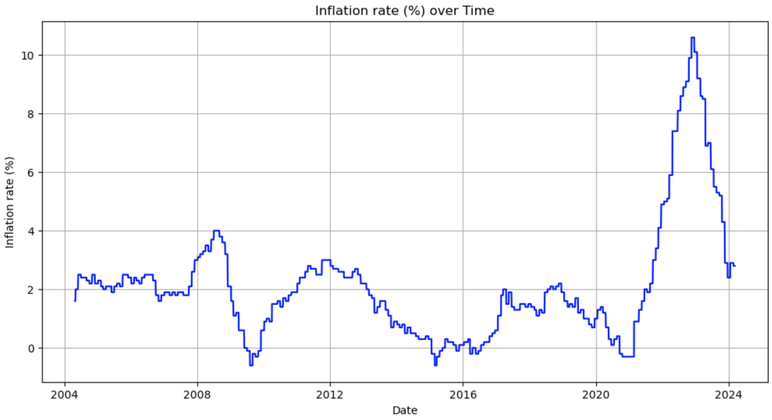
A.4 GDP



A.5 Unemployment Rate

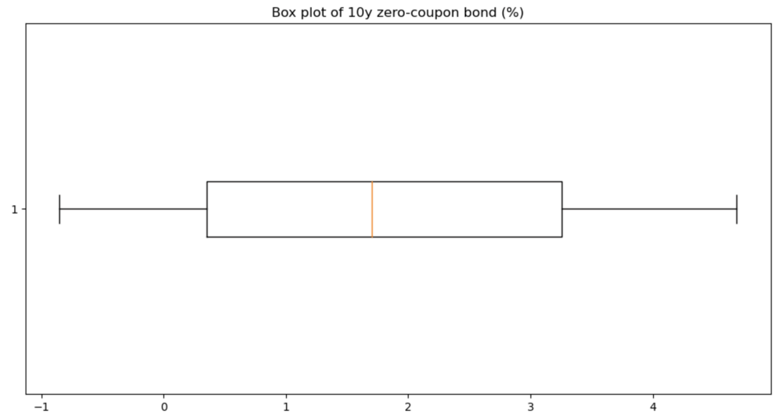


A.6 Inflation Rate

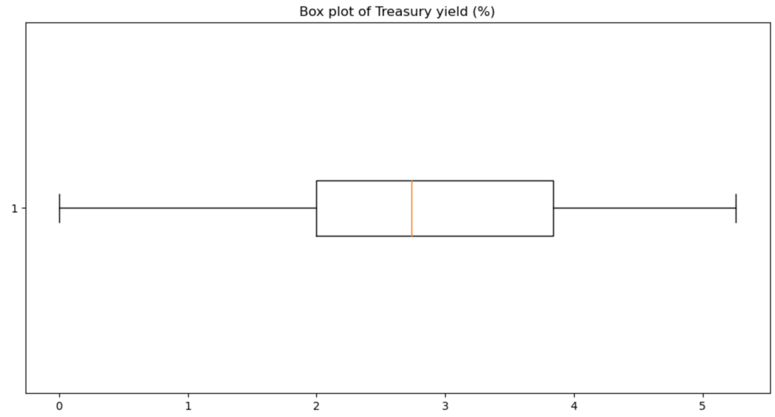


B Data Outliers

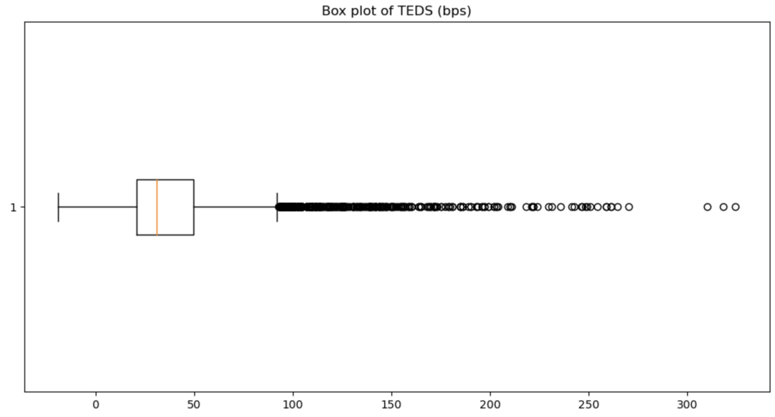
B.1 10-Year Zero-Coupon Bond



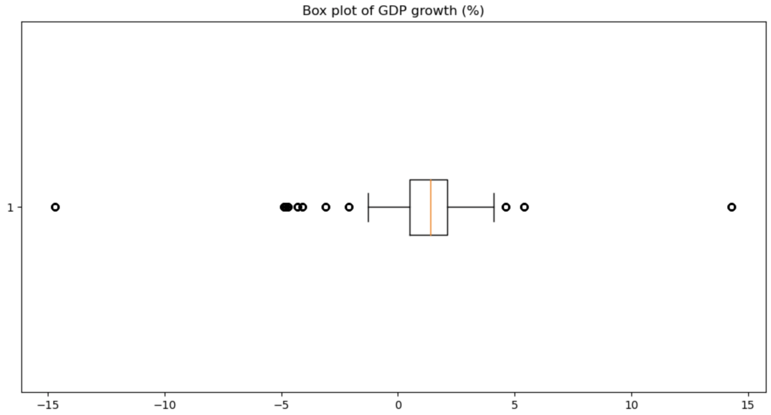
B.2 Treasury Yield



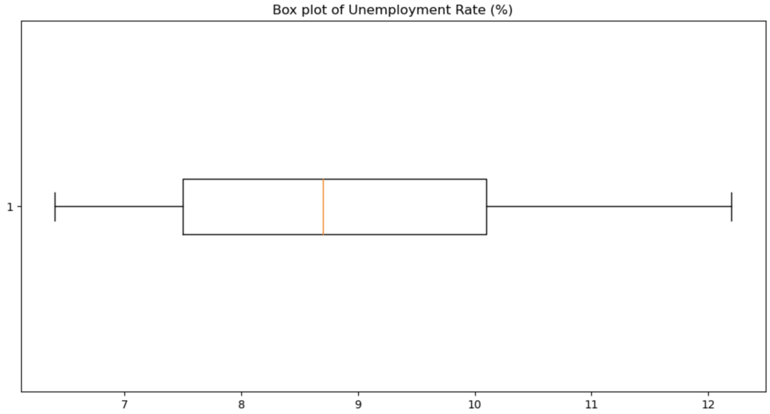
B.3 TED Spread



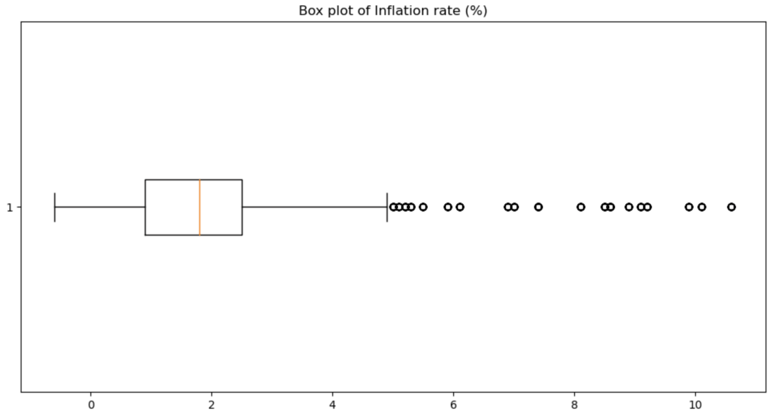
B.4 GDP



B.5 Unemployment Rate

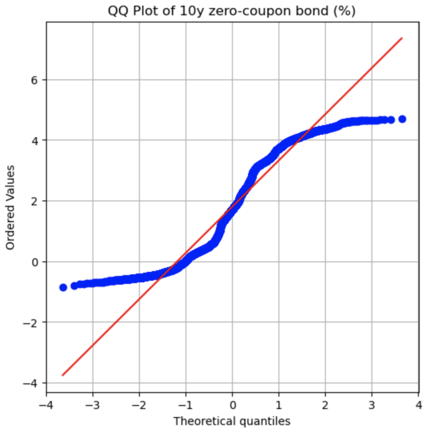


B.6 Inflation Rate

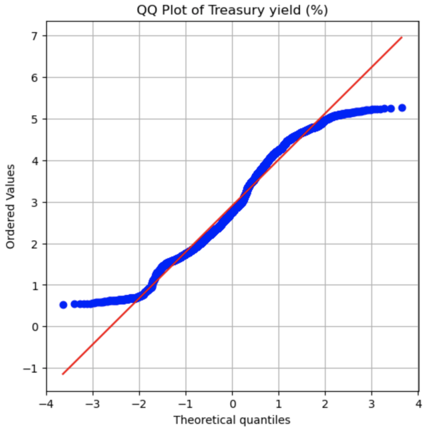


C Data QQ-Plots

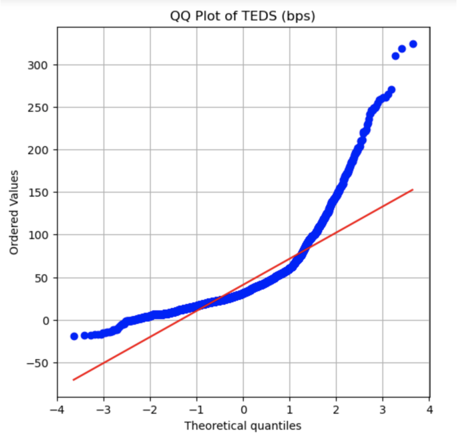
C.1 10-Year Zero-Coupon Bond



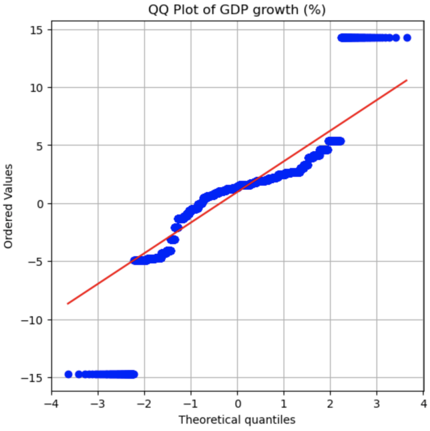
C.2 Treasury Yield



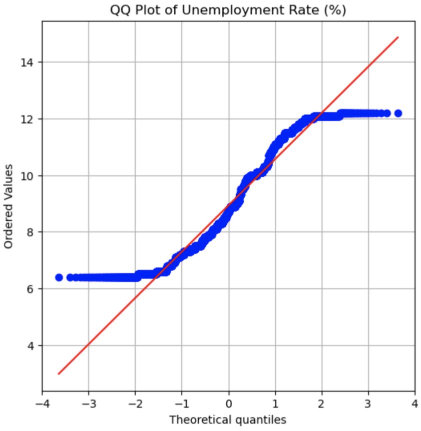
C.3 TED Spread



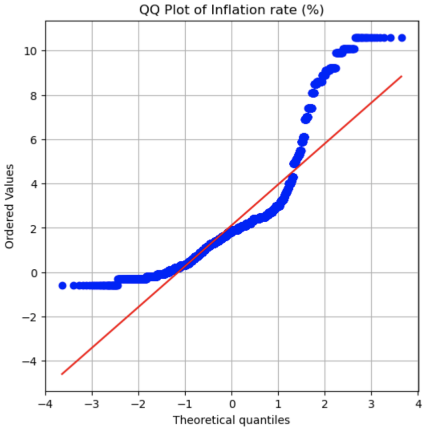
C.4 GDP



C.5 Unemployment Rate

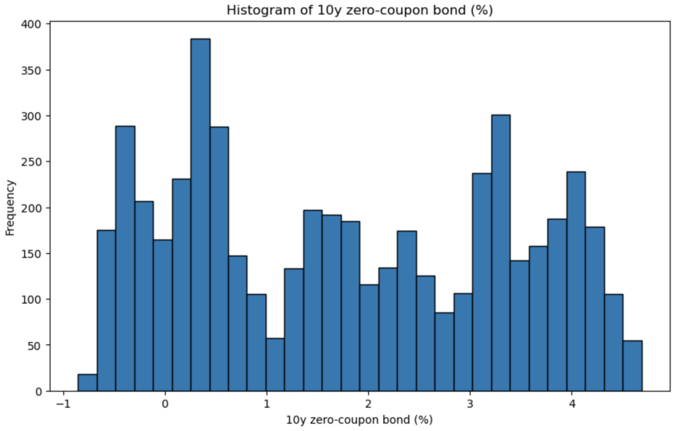


C.6 Inflation Rate

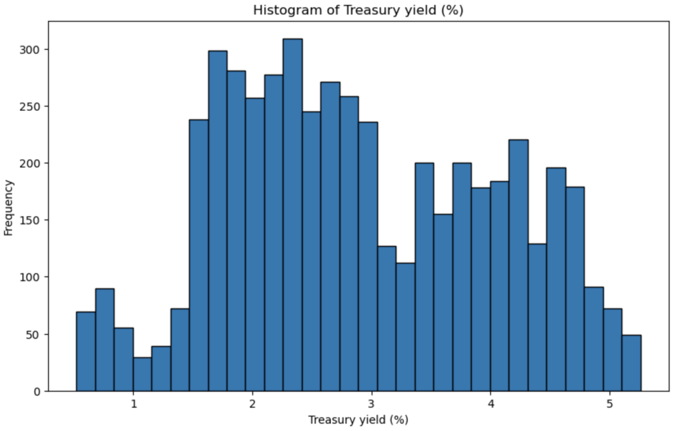


D Data Histograms

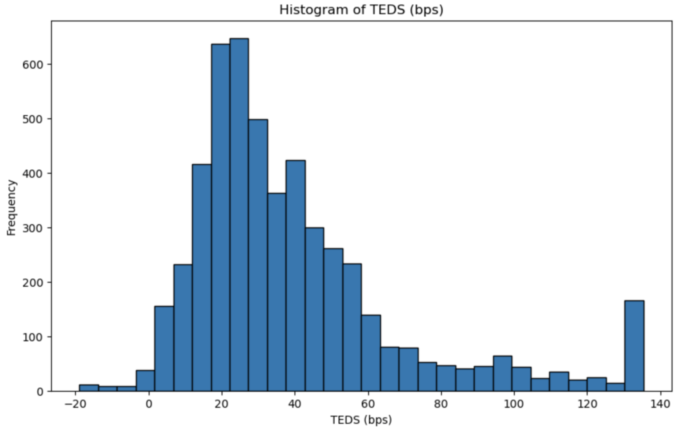
D.1 10-Year Zero-Coupon Bond



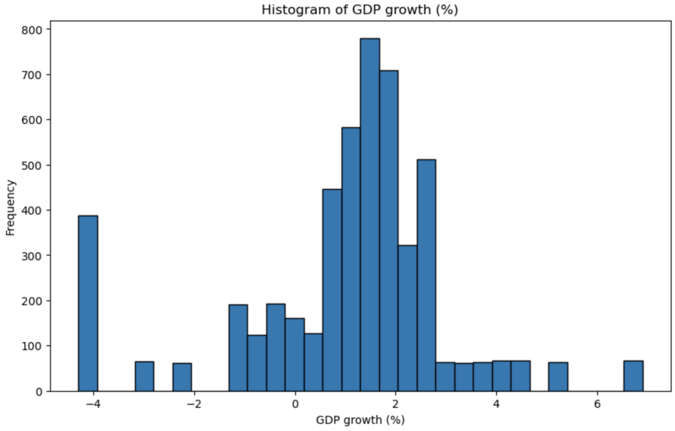
D.2 Treasury Yield



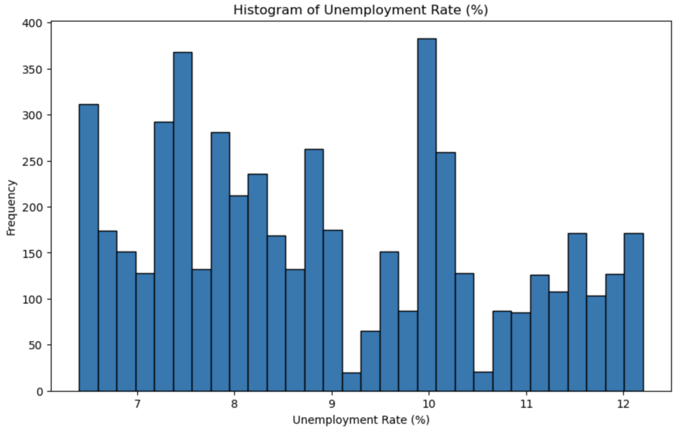
D.3 TED Spread



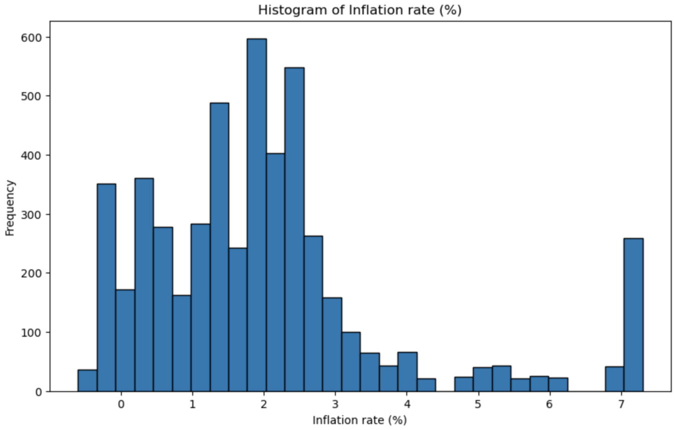
D.4 GDP



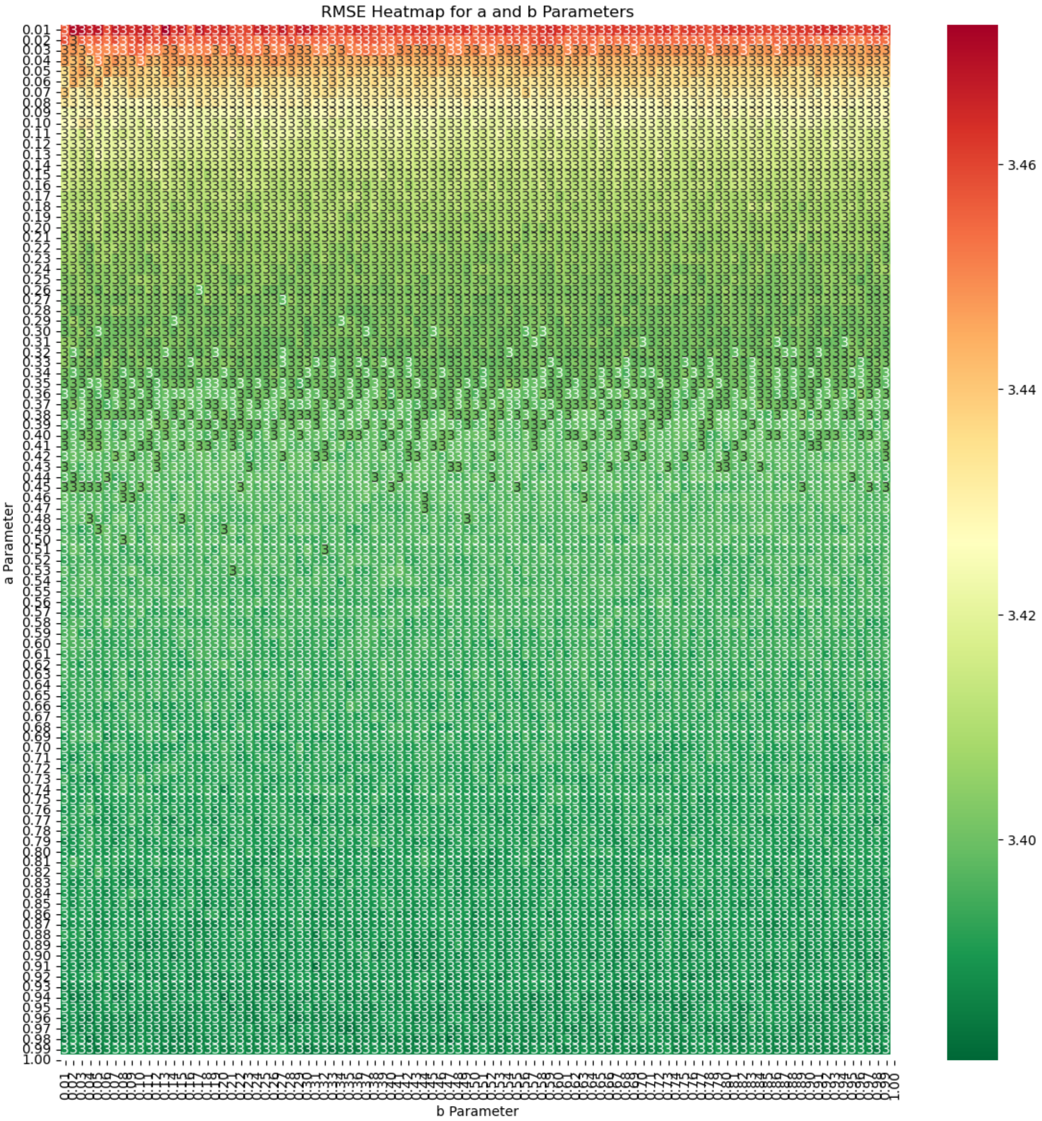
D.5 Unemployment Rate



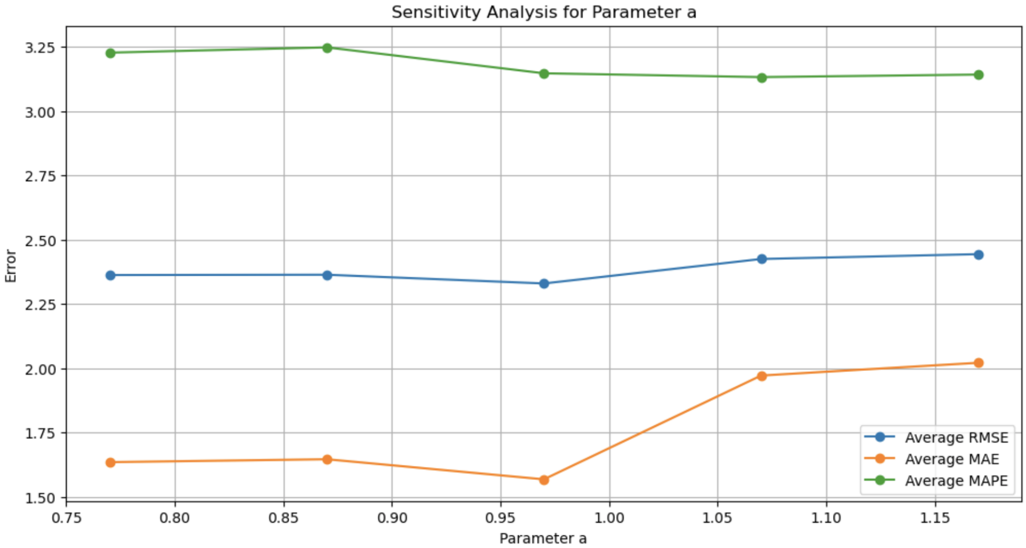
D.6 Inflation Rate

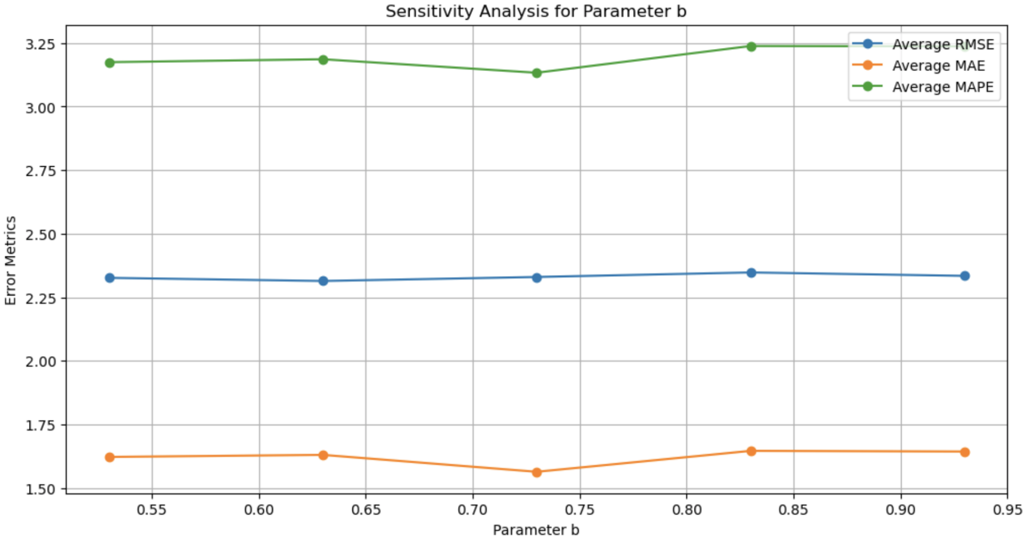


E RMSE Heatmap



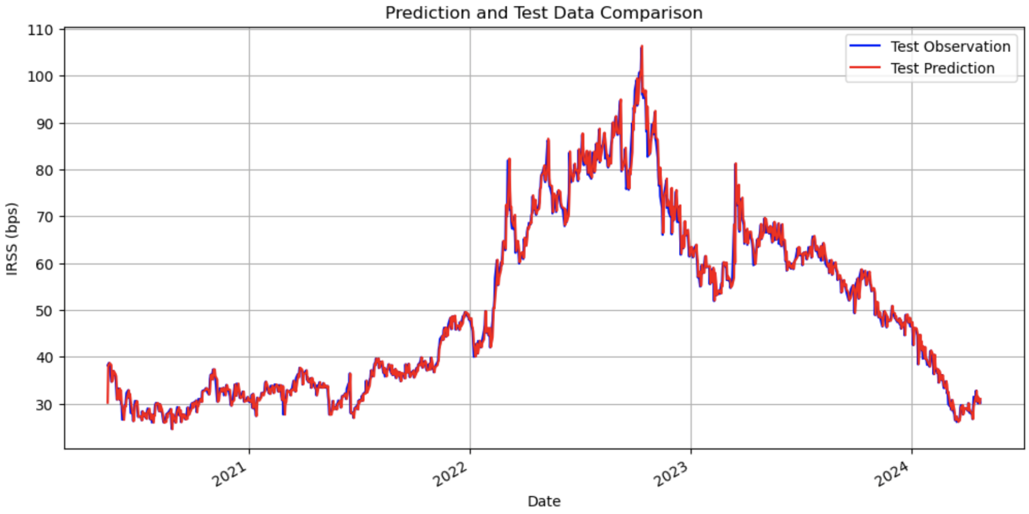
F Sensitivity Analysis HW2F Parameters



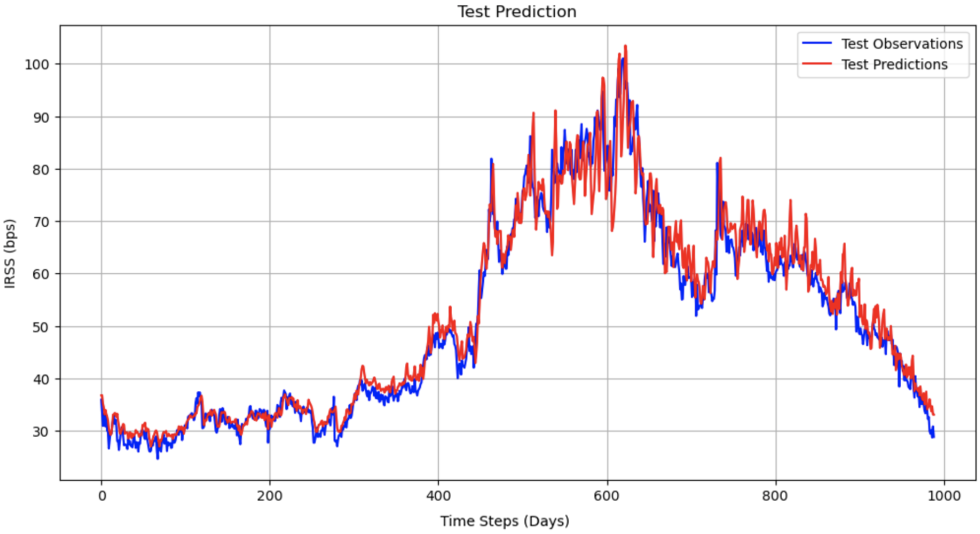


G Full Test Predictions

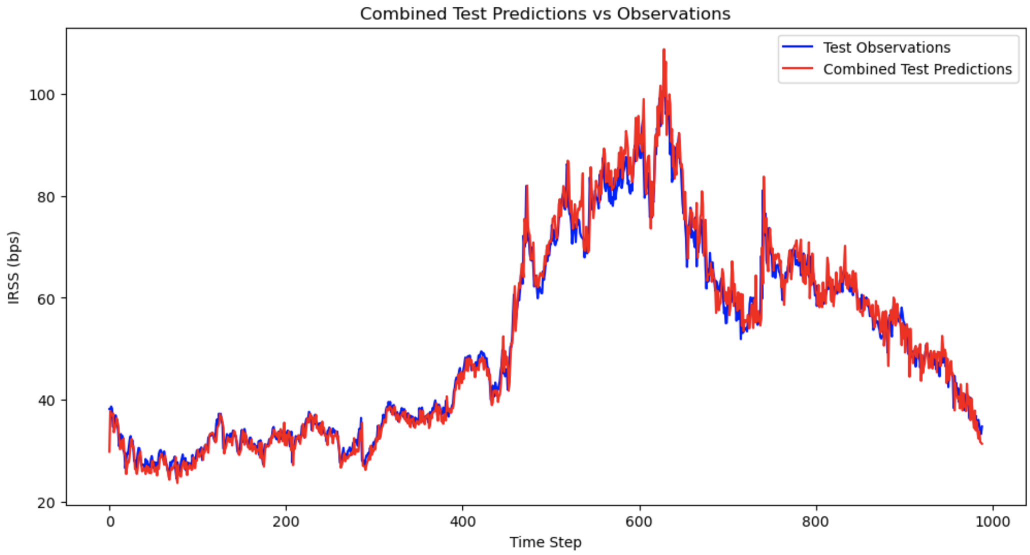
G.1 HW2F



G.2 LSTM



G.3 Hybrid


Combined Test Predictions vs Observations
Test Observations
Combined Test Predictions
IRSS (bps)
20
40
60
80
100
0
200
400
600
800
1000
Time Step