Multi-Horizon Forecasting and Inter-Asset Causal Dynamics on the NZX: A Deep Learning Study of LSTM-Based Models with Technical Indicators

Khin Le Le

A project report submitted to the Auckland University of Technology
in partial fulfillment of the requirements for the degree of
Master of Computer and Information Sciences (MCIS)

2025

School of Engineering, Computer & Mathematical Sciences

Abstract

Financial forecasting plays a vital role in investment and policy planning, yet traditional models such as ARIMA often fail to capture nonlinear dependencies and multi-horizon patterns in financial markets. This study develops Long Short-Term Memory (LSTM)–based deep-learning frameworks enhanced with technical indicators and attention mechanisms to forecast New Zealand Exchange (NZX) returns across multiple horizons. Daily data from 2015–2024, covering the NZX 50 and eleven sectoral indices together with macroeconomic variables, were used to train and evaluate models over horizons of one day (h = 1), one month (h = 21), one quarter (h = 63), and one year (h = 252).

Our results show that ARIMA achieved the lowest short-term error (MAE = 0.0049 at h = 1), while attention-enhanced LSTM models produced the highest directional accuracy (DA = 49.9%) and maintained consistent performance across longer horizons ($MAE \approx$ 0.058–0.060, $Sharpe \approx$ 0.23). Although cumulative returns were similar to ARIMA ($\approx$ 3.43 at h = 252), attention-based architectures delivered smoother and more stable risk-adjusted outcomes, showing that LSTM-based models achieved more stable and interpretable performance across multiple horizons. Granger-causality and attention analyses identified short-term causal influences from Energy, Real Estate, Consumer Discretionary, Industrials, and Information Technology sectors, along with macro-financial effects from NZD/USD and the Official Cash Rate (OCR), confirming that inter-asset lagged effects are statistically and economically meaningful.

These findings demonstrate that LSTM-based multi-horizon frameworks effectively improve multi-horizon forecasting performance. The study provides interpretable evidence of cross-sector and macro-financial spillovers in a small open market, offering practical insights for portfolio diversification and risk-management strategies in the New Zealand context.

Keywords: Multi-horizon forecasting, LSTM, Technical indicators, Temporal attention, Granger causality, New Zealand Exchange (NZX)

Table of Contents

List of Figures

List of Tables

Attestation of Authorship

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (except where explicitly defined in the acknowledgments), nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a university or other institution of higher learning.

Signature: Khin Le Date: 2 November 2025

Acknowledgment

I would like to express my deepest gratitude to my supervisor, Professor Minh Nguyen, for his invaluable guidance, encouragement, and patience throughout this research journey. His expertise and insightful feedback have been instrumental in shaping the direction and quality of this study.

I would also like to express my gratitude to the Auckland University of Technology for providing me with the opportunity and resources to pursue my Master's degree. My appreciation extends to all the lecturers and staff members of the School of Engineering, Computer and Mathematical Sciences for their continuous support and dedication.

Khin Le Le

Auckland, New Zealand

October 2025

Chapter 1
Introduction

This chapter is composed of five parts. First, it introduces the background and motivation for this research, followed by the research questions that guide the study. Next, the chapter outlines the contributions and objectives of this work. Finally, it provides an overview of the report structure.

1.1 Background and Motivation

Forecasting financial markets is a complex and high-impact task, influencing investment decisions, portfolio management, and policy planning. Accurate forecasts enable investors to anticipate market movements and reduce risk exposure, while policymakers rely on such insights to design effective regulatory measures. However, predicting financial returns remains challenging due to the inherent volatility, non-linearity, and dynamic nature of market behavior (Sunki et al., 2024).

Traditional statistical models such as ARIMA and SARIMA have been widely applied in time-series forecasting. These models assume linear relationships and are generally effective for short-term horizons. However, their performance deteriorates in the presence of long-term dependencies and structural breaks, making them less suitable for multi-horizon forecasting. Additionally, they fail to capture complex, non-linear interactions between sectors, which are critical for understanding market dynamics (Arumugam et al., 2023).

With advancements in machine learning, deep learning techniques have emerged as powerful tools for financial forecasting. Among them, Long Short-Term Memory (LSTM) networks—a type of recurrent neural network—are particularly well-suited for sequential data modeling due to their ability to capture long-range dependencies. Recent studies have demonstrated that LSTM models outperform traditional models in stock price forecasting tasks, primarily in short-term scenarios (Fischer et al., 2018) (Nelson et al., 2017). However, most research focuses on large markets such as the S&P 500, leaving smaller markets like the New Zealand Exchange (NZX) largely unexplored.

Another critical area of interest is multi-horizon forecasting, which extends beyond the traditional one-day-ahead paradigm. Investors and policymakers often require multi-step predictions—from short-term (1 day) up to long-term (1 year)—to enable both tactical and strategic decision-making. Research shows that incorporating technical indicators along with macroeconomic variables into deep learning models improves forecast accuracy (Latif et al., 2025).

Finally, understanding inter-asset causal dynamics is essential for risk management and portfolio optimization. Financial markets exhibit complex interconnections where shocks in one sector can propagate to others over time. As Diebold et al. (2011) demonstrate, measuring connectedness among financial assets provides a framework to capture how shocks are transmitted across markets. Su et al. (2021) emphasize that correctly identifying sectoral volatility spillovers is crucial for risk management, since unanticipated interactions between industries can significantly alter portfolio outcomes. More recent evidence by Polat (2024) further demonstrates that sectoral spillovers are both time-varying and asymmetric, highlighting the importance of monitoring inter-sector linkages in dynamic market conditions. Identifying these relationships helps investors design diversification strategies and improve resilience against systemic risks.

This study addresses two major gaps: (1) the lack of research on multi-horizon forecasting in smaller markets like the NZX using advanced deep learning techniques, and (2) the limited exploration of inter-asset causal dynamics and their interpretability across multiple time horizons.

1.2 Research Questions

In this report, the main objective of this research is to investigate whether deep learning models, particularly Long Short-Term Memory (LSTM) networks enhanced with technical indicators, can outperform traditional ARIMA models in forecasting NZX50 returns across multiple time horizons:

- next day,
- next month,
- next 3 months, and
- next year.

Additionally, the study aims to explore whether lagged relationships across different sectors and assets can be identified and modeled. For example:

- A sharp drop in the Technology sector (Asset A) may lead to a rise in Consumer Staples (Asset B) in the following month.
- which in turn may affect Industrial (C) and Finance sectors (D) after a longer lag

Therefore, the main research questions of this report are:

- How can deep learning models, specifically LSTM networks enhanced with technical indicators, improve multi-horizon return forecasts compared to classical ARIMA models?

To refine this research question, we can split it again:

- Can deep learning improve multi-horizon return forecasts compared to classical methods?
- How can technical indicators be effectively integrated into deep learning models for enhanced predictive power?
- Can temporal dependencies between asset classes be identified and modelled?
- Which lagged inter-asset relationships are statistically significant and economically interpretable?
- What implications do these cross-sector signals hold for portfolio strategy and economic forecasting?

The core focus of this research is on building robust multi-horizon forecasting models and analyzing cross-sector causal dynamics in the NZX market using deep learning techniques. Additionally, attention will be given to interpretability and the practical value of the identified relationships for real-world applications.

1.3 Contributions

The focus of this research is to improve financial forecasting in the New Zealand Exchange (NZX) by developing LSTM-based deep learning models capable of multi-horizon predictions and detecting causal dynamics between asset returns on the NZX. The proposed framework introduces advanced methods for forecasting financial time series and analyzing sectoral dependencies. Specifically, this study aims to:

- Develop LSTM-based models enriched with technical indicators to predict NZX returns across multiple horizons (1 day, 1 month, 3 months, and 1 year).
- Build a comprehensive dataset combining NZX50 sectoral indices and relevant macroeconomic indicators for robust modeling.
- Incorporate technical indicators into the input feature set to enhance predictive performance and capture market dynamics.
- Design and train a suite of LSTM architecture with attention mechanisms to forecast return trajectories across time
- Evaluate model performance using multiple metrics, including Mean Absolute Error (MAE), directional accuracy, and economic utility measures.
- Apply Granger causality tests and interpretable attention layers to identify and analyze lagged relationships across different sectors (e.g., Technology → Consumer Staples → Industrial → Financial).

By the end of this research, the following key contributions will be realized:

- A multi-horizon forecasting framework in predicting NZX returns.
- A systematic approach for integrating technical indicators into LSTM models for improved accuracy in financial forecasting.
- A methodology for uncovering inter-sector causal dynamics using deep learning interpretability techniques and statistical causality analysis
- Practical insights into cross-sector signals that can inform portfolio optimization and risk management strategies in small markets like New Zealand.

In addition, the study will conduct comparative analysis between LSTM and ARIMA models to highlight the advantages of deep learning over classical time-series methods for long-term and multi-horizon forecasting.

1.4 Objectives of This Report

In this report, we present a comprehensive research framework designed to evaluate the effectiveness of LSTM-based deep learning models for multi-horizon financial forecasting on the New Zealand Exchange (NZX). Our approach integrates technical indicators with LSTM architectures to enhance predictive performance and interpretability, addressing the limitations of traditional time-series models like ARIMA. Additionally, this study explores inter-asset causal dynamics across NZX sectors to identify lagged relationships that influence long-term market behavior. This report includes the following major components:

- A literature review of classical time-series forecasting models, deep learning methods for financial prediction, and multi-horizon forecasting techniques to establish the research context.
- A methodology for building an enriched dataset of NZX50 sector indices and relevant macroeconomic indicators, engineering technical indicators, and training LSTM-based models with attention mechanisms for multi-horizon predictions.
- An analytical framework for detecting inter-asset causal relationships through Granger causality tests and interpretability techniques applied to attention weights in deep learning models

The specific objectives of this report are twofold:

- To develop and evaluate LSTM-based models for forecasting NZX returns across multiple time horizons (1 day, 1 month, 3 months, and 1 year) and compare their performance with ARIMA models using metrics such as MAE, directional accuracy, and economic utility
- To identify and interpret cross-sector causal relationships within the NZX using statistical methods and deep learning interpretability, providing insights for portfolio optimization and economic forecasting

The research will conclude with a comparative analysis highlighting the advantages and limitations of LSTM-based models over traditional approaches, offering practical implications for financial practitioners and contributions to academic literature in time-series forecasting and causal modeling.

1.5 Structure of this Report

The structure of this report is organized as follows:

- In Chapter 2, we conduct a literature review focusing on financial forecasting techniques. We begin with traditional statistical models such as ARIMA and SARIMA, then explore the progress of deep learning approaches, particularly LSTM models and the use of attention mechanisms. Finally, we review studies related to multi-horizon forecasting and inter-asset causal dynamics.
- In Chapter 3, we present the research methodology, including data collection from NZX, feature engineering with technical indicators, and the design of LSTM-based models. We also describe the baseline ARIMA models, the implementation of attention mechanisms, and the evaluation framework using statistical and economic metrics.
- In Chapter 4, we describe the experimental design and implementation. This includes training the models, performing multi-horizon forecasts, and analyzing results. We also visualize forecast outputs and assess inter-asset causal relationships using both attention-based methods and Granger causality tests.
- In Chapter 5, we provide a detailed discussion and analysis of the results. We compare the performance of LSTM models with ARIMA, examine the interpretability of causal patterns, and discuss the implications for portfolio optimization and financial decision-making.
- In Chapter 6, we conclude the report by summarizing key findings, reflecting on the research objectives, and suggesting future directions such as extending the framework to real-time forecasting and exploring other deep learning architectures.

Chapter 2 Literature Review

This chapter reviews prior research on financial forecasting models, covering both traditional statistical approaches and recent deep learning methods

2.1 Introduction

Financial time-series forecasting has been a central topic in quantitative finance, driven by the need for accurate predictions to inform investment strategies, portfolio management, and economic decision-making. The volatility and dynamic behavior of financial markets make forecasting complex, requiring models that can adapt to changing conditions and capture hidden patterns in data (Zhao et al., 2024). Over the years, researchers have proposed various models, ranging from classical statistical approaches to advanced machine learning and deep learning techniques, to improve prediction accuracy and robustness.

This chapter provides a comprehensive review of existing research relevant to multi-horizon forecasting and inter-asset causal dynamics. It begins by discussing traditional forecasting models, such as Autoregressive Integrated Moving Average (ARIMA) and Seasonal ARIMA (SARIMA), which have served as foundational tools in time-series analysis. It then explores the evolution of deep learning approaches, particularly Long Short-Term Memory (LSTM) networks, which have demonstrated superior capabilities in modeling sequential data. Additionally, this review examines recent developments in multi-horizon forecasting strategies and methods for analyzing causal relationships across financial assets. Finally, the chapter concludes by identifying research gaps that this study aims to address, emphasizing the need for advanced models capable of both accurate long-term forecasting and interpretability in smaller markets such as the New Zealand Exchange (NZX).

2.2 Financial Time-Series Forecasting and Traditional Models

Traditional time-series forecasting methods have long been the foundation of financial prediction models. Among these, the Autoregressive Integrated Moving Average (ARIMA) model introduced by Box and Jenkins remains one of the most widely used. ARIMA models are designed to capture autocorrelations in stationary time series and are expressed as ARIMA(p, d, q), where p denotes the order of the autoregressive component, d the degree of differencing, and q the order of the moving average component (Siami-Namini, 2018).

Seasonal ARIMA (SARIMA) extends the ARIMA framework by including seasonal components, making it suitable for datasets exhibiting periodic. SARIMA models are particularly useful in contexts where financial data exhibit recurring seasonal effects, such as quarterly earnings announcements or holiday-driven consumption patterns. However, these models operate under the assumption of linearity and stationarity, which, while practical for many applications, limits their ability to capture the non-linear and dynamic behaviors observed in financial markets (Khashei et al., 2012).

ARIMA models demonstrate consistent effectiveness for short-term stock index forecasting across multiple markets. Dassanayake et al. (2021) found that ARIMA (1,1,0) with intercept was the optimal model for predicting New Zealand's NZX50 index, evaluated using MAE, MAPE, and RMSE metrics. Similarly, Hou (2023) successfully applied ARIMA to forecast China's CSI 300 index volatility trends over five years (2017-2022), demonstrating the model's capability in handling complex market situations that linear models cannot address. YiChen et al. (2017) confirmed ARIMA's strong potential for short-run forecasting using Apple Inc. stock data, noting its competitiveness with other prediction methods for investment guidance. Rotela Junior et al. (2014) validated ARIMA's performance on Brazil's Ibovespa index using the Box-Jenkins method, achieving lower MAPE values compared to other smoothing models. These studies collectively establish ARIMA as a reliable tool for stock index prediction, particularly excelling in short-term forecasting scenarios across diverse financial markets.

The SARIMA model demonstrates mixed effectiveness in forecasting financial markets across different applications. In currency exchange forecasting, SARIMA shows strong performance, with studies finding it provides better forecasting ability for INR/USD exchange rates by incorporating seasonal data patterns (D et al., 2024). However, SARIMA's effectiveness varies in stock market applications. While one study successfully applied SARIMA to forecast the NIFTY 50 index using 729 model parameter combinations selected via AIC criteria (Tewari, 2020), another found that SARIMA underperformed compared to ARIMA for stock price prediction, despite its capability in detecting seasonal variations

(Kavitha et al., 2025). The research suggests SARIMA's effectiveness depends on the specific financial instrument and market conditions being analyzed.

ARIMA and SARIMA models demonstrate several key limitations when applied to economic forecasting. A primary constraint is their linear assumption, which can lead to significant errors during nonlinear market events like crashes, as these models assume future values have linear relationships with current and past values (Khashei et al., 2009; Liu, 2024). Both models require substantial historical data to generate accurate results, limiting their effectiveness with small datasets (Asadi et al., 2012; Khashei et al., 2009). ARIMA models particularly struggle with long-term predictions, showing increased errors over extended forecasting horizons while performing better for short-term .Additionally, these models' reliance on historical data means they may inadequately account for external real-world factors that influence economic systems (Zhao et al., 2024).

Traditional ARIMA/SARIMA models are fundamentally univariate and lack native support for incorporating external explanatory variables such as macroeconomic indicators or sectoral interactions. These models rely solely on historical patterns of the variable being modeled, abandoning classical econometric approaches that incorporate explanatory variables suggested by economic theory (Petrică et al., 2016). Extensions like ARIMAX and SARIMAX are possible but require significant modification. Vector Autoregression enables modeling multiple interrelated series, but it becomes computationally prohibitive as dimensionality increases and remains confined to linear relationships (Nicholson et al., 2017). These limitations underscore the need for non-linear and adaptive forecasting methods. Hybrid models such as the ARIMA combined with neural networks have demonstrated improved performance in capturing complex temporal dependencies in financial time series (Zhang, 2003).

Machine learning methods such as Support Vector Machines (SVM) and Random Forests have been introduced as alternatives. Comparative studies show mixed performance between these algorithms. While SVMs demonstrate superior classification accuracy in predicting stock market direction (Kumar et al., 2006) and credit operations (Teles et al., 2021), Random Forests offer advantages in operational simplicity and processing speed (Teles et al., 2021). However, both approaches face fundamental challenges in capturing temporal dependencies inherent in financial data.

2.3 Deep Learning for Financial Forecasting

The limitations of traditional time-series models, such as ARIMA and SARIMA, in handling non-linear patterns and long-term dependencies have motivated researchers to explore advanced computational methods. Early approaches, including machine learning techniques like Support Vector Machines (SVM), Random Forests, and Gradient Boosting, showed improvements in capturing non-linear relationships but were limited by their inability to model sequential dependencies effectively (Patel et al., 2015; Zhang et al., 1998). These models rely on engineered features rather than learning temporal structures directly, making them less suitable for time-series prediction.

Deep learning emerged as a powerful alternative due to its capability to automatically extract features from raw data and capture complex dependencies (LeCun et al., 2015). Recurrent Neural Networks (RNNs) were among the first architectures used for sequence modeling, as they maintain hidden states that propagate through time. However, standard RNNs suffer from vanishing and exploding gradient problems, which restrict their ability to learn long-term dependencies (Bengio et al., 1994). This limitation paved the way for the development of Long Short-Term Memory (LSTM) networks. LSTM networks were introduced to address gradient instability in RNNs (Hochreiter et al., 1997)**.** LSTM achieves this by introducing memory cells and gating mechanisms:

- Input gate: Controls which new information enters the memory cell.
- Forget gate: Decides which information to discard.
- Output gate: Determines what information is output at each time step.

This gating system enables LSTM models to capture both short- and long-term dependencies effectively, making them particularly well-suited for financial time-series forecasting where historical patterns influence future market movements. Unlike traditional models that assume linearity, LSTMs learn non-linear dependencies and adapt to changing market conditions dynamically.

To enhance the ability of recurrent models to capture both local and global temporal patterns, hybrid architectures have been proposed. For instance,Lai et al. (2018) introduced the Long- and Short-Term Time-series Network (LSTNet), which combines convolutional and recurrent layers to model short-term fluctuations and long-term dependencies simultaneously. Such CNN–RNN hybrids further extend the capacity of LSTMs for complex financial time-series forecasting.

Beyond recurrent architectures, Borovykh et al. (2017) demonstrated that convolutional neural networks can model temporal dependencies directly through causal convolutions, offering a simpler yet effective alternative for financial time-series prediction.

Studies demonstrate the effectiveness of LSTM networks in forecasting stock prices, volatility, and macroeconomic indicators. For example, Fischer et al. (2018) applied LSTM models to predict stock returns for the S&P 500 and reported significant accuracy improvements over logistic regression and random forests. Similarly, Nelson et al. (2017) confirmed LSTM's superiority in predicting Brazilian stock market prices compared to traditional and widely used neural networks such as MLP, Random Forest and a pseudo-random model.

Although LSTMs address long-term dependency issues, their performance can still degrade when sequences become very long or when interpretability is essential. To overcome these challenges in time-series forecasting, Qin et al. (2017) proposed a dual-stage attention-based RNN model that dynamically focuses on informative time steps and relevant input features. Attention mechanisms, initially developed for natural language processing tasks, have been adapted for financial forecasting to focus on the most relevant time steps in a sequence (Bahdanau et al., 2014; Vaswani et al., 2017). Attention assigns different weights to input elements, allowing models to "attend" to crucial information and ignore less important parts.

Temporal Attention LSTM models and Transformer-based architectures (such as Temporal Fusion Transformer) have gained popularity in multi-horizon forecasting due to their superior ability to handle long-range dependencies and their inherent interpretability (Lim et al., 2021). These models can reveal which historical periods most influence predictions, providing insights for traders and analysts.

Deep learning models, particularly LSTM and its variants, have been widely applied to various financial forecasting problems:

- Stock Price Prediction: Fischer et al. (2018) used LSTMs to predict daily returns of the S&P 500, outperforming traditional classifiers and simple neural networks.
- Volatility Forecasting: Bao et al. (2017) integrated wavelet transforms with LSTM networks to predict stock price volatility, improving accuracy in noisy financial environments.
- Exchange Rate Forecasting: Galeshchuk et al. (2017) demonstrated the advantage of LSTM over ARIMA in forecasting currency exchange rates.
- Portfolio Optimization: Sezer et al. (2020) explored Deep learning approaches, including LSTM and attention-based models, have also been explored for portfolio optimization and asset allocation tasks.

Despite their advantages, deep learning models are not without challenges. They require large datasets for effective training, are computationally intensive, and risk overfitting when applied to small markets like NZX. This necessitates techniques such as dropout regularization, hyperparameter tuning, and attention-based interpretability to ensure reliable performance,(Liu et al., 2025).

2.4 Multi-Horizon Forecasting

Financial forecasting traditionally focuses on one-step-ahead predictions, where models estimate the next value in a series. However, decision-makers often require forecasts across multiple time horizons ranging from short-term (days) to long-term (months or years) to guide portfolio allocation, hedging strategies, and macroeconomic planning (Lim et al., 2021). Multi-horizon forecasting involves predicting a sequence of future values, not just the immediate next observation, making it essential for strategic financial decisions.

Multi-horizon forecasting builds upon multi-step forecasting strategies. Two primary approaches dominate this area:

- Recursive (Iterative) Strategy: A model trained for one-step forecasting is applied repeatedly, feeding its previous output as input for the next prediction (Taieb et al., 2016). While computationally efficient, this method suffers from error accumulation as horizon length increases.
- Direct Strategy: A separate model is trained for each forecast horizon (e.g., 1 day, 1 month, 3 months). Although more accurate for longer horizons, this approach increases complexity and requires substantial data (Taieb et al., 2016).

Advanced approaches combine these strategies or adopt Sequence-to-Sequence (Seq2Seq) architectures inspired by machine translation in NLP. These models predict entire sequences in a single pass, reducing error propagation (Lim et al., 2021). Temporal Fusion Transformers (TFT) are a recent innovation offering interpretability and scalability for multi-horizon financial predictions.

Multi-horizon forecasting models have been applied to equity returns, commodity prices, and macroeconomic indicators. For example,Kroujiline et al. (2016) developed an agent-based model for forecasting stock market returns over multiple horizons, from intraday to monthly, demonstrating that heterogeneous investor behavior can generate distinct predictive patterns across timescales. Despite progress, key challenges remain:

- Data scarcity: Longer horizons require extensive training data.
- Concept drift: Market dynamics change over time, affecting accuracy.
- Interpretability: Multi-horizon models are often treated as “black boxes,” limiting trust in predictions.

2.5 Inter-Asset Causal Modeling and Cross-Sector Relationships

While multi-horizon forecasting focuses on predicting the future path of a single asset or index, an equally important challenge lies in understanding how shocks propagate between assets and sectors. This perspective shifts the emphasis from forecasting in isolation to modeling inter-asset causal dynamics, which are essential for systemic risk assessment, contagion detection, and effective portfolio diversification.

Early studies highlighted that financial markets function as interconnected systems, where disturbances in one sector can be transmitted across others and amplify risks. Diebold et al. (2011) formalized this concept with a connectedness index derived from forecast-error variance decompositions, offering a rigorous and quantitative approach to mapping spillovers between financial firms. Building on this foundation, Diebold et al. (2014) extended the connectedness framework, introducing a generalized approach that enabled the measurement of total, directional, and net spillovers across financial markets. Subsequent studies, such as Baruník et al. (2018), further refined the framework by decomposing connectedness into short- and long-term frequency components, capturing how transmission dynamics vary across investment horizons. More recently, this methodology has been applied directly to stock markets. For example, Choi et al. (2023) used the same connectedness framework to analyze upside and downside risk spillovers among international equity markets during periods of financial turmoil, revealing that downside risk connects more intensely across markets and that developed markets act as net transmitters of risk to their emerging counterparts.

A widely used statistical tool in this domain is Granger causality which tests whether past values of one series improve the forecast of another (Granger, 1969). Network-based models extend Granger causality to visualize interconnectedness across markets, helping identify systemically important sectors (Billio et al., 2012). However, these approaches assume linear relationships and may fail to capture complex dependencies present in financial data.

Recent studies have sought to address this limitation by integrating Granger-based connectedness with nonlinear models. For instance, Boonpong et al. (2023) investigated sector return connectedness in Thailand's stock market and demonstrated that incorporating inter-sector linkages into CNN-LSTM models improved forecasting performance compared to plain LSTM.

Beyond Granger-based extensions, deep learning has opened new directions for modeling inter-asset dependencies. Transformer architectures such as the Temporal Fusion Transformer illustrate how attention mechanisms can improve interpretability in time-series forecasting (Lim et al., 2021). Recent studies have employed graph neural networks (GNNs) with attention mechanisms to explicitly model markets as networks of interconnected assets, uncovering spillovers and systemic linkages across sectors (Xiang et al., 2022). These approaches are particularly suited to high-dimensional settings, where traditional linear methods struggle.

Understanding causal chains between sectors enables investors to anticipate spillover effects and optimize portfolios. For instance, a shock in the energy sector may influence industrials and transportation due to input cost dependencies. Incorporating these insights into forecasting models enhances resilience against systemic risks and supports proactive risk management in volatile markets.

2.6 Research Gaps and Summary

This literature review reveals a clear evolution in financial forecasting, from traditional linear models to advanced deep learning architectures. While substantial progress has been made, several critical research gaps remain:

- **Limited Multi-Horizon Forecasting in Small Markets:** Most deep learning studies focus on large financial markets such as the U.S. or global indices. Research on smaller markets like the New Zealand Exchange (NZX) remains rare.
- **Insufficient Integration of Causal Dynamics:** Although Granger causality and network models are well established for analysing inter-asset relationships, few studies integrate these insights into predictive deep learning frameworks. Combining interpretability with predictive accuracy is still an open challenge.

- **Interpretability and Trustworthiness:** Advanced neural networks often function as "black boxes," limiting their adoption in financial decision-making. More work is needed on explainable AI and attention-based visualization to make models transparent and actionable for investors.

Addressing these gaps is essential for improving forecasting accuracy and reliability in smaller markets. This research therefore aims to bridge these challenges by developing LSTM-based models with technical indicators for multi-horizon predictions and by incorporating inter-asset causal analysis to enhance interpretability and support decision-making.

Chapter 3 Methodology

This chapter outlines the methodological framework used to evaluate the forecasting performance of ARIMA, LSTM, and attention-based LSTM models on NZX50 index. It describes the data sources, preprocessing and feature engineering steps, and model design for multi-horizon prediction. The chapter also details the evaluation metrics and causality analyses employed to assess predictive accuracy and inter-asset relationships.

3.1 ARIMA

ARIMA models have long been regarded as a benchmark for time-series forecasting because of their ability to capture short-term linear dependencies (Box et al., 2015). The central principle is that the current value of a series can be explained by its own past values (autoregression) and by past forecast errors (moving averages). The autoregressive (AR) process of order *p* assumes that today's value depends linearly on its past *p* observations:

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t \tag{3.1}$$

The moving average (MA) process of order *q* models the dependence of the current observation on *q* past shocks, as shown below:

$$y_t = \mu + \epsilon_t + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} \tag{3.2}$$

Combining both yields the ARMA (p, q) model:

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} \tag{3.3}$$

While ARMA is suitable for stationary series, many financial time series are not stationary. To address this, differencing can be applied, producing the ARIMA (p, d, q) model, where *d* is the order of differencing needed to remove non-stationarity:

$$\nabla^d y_t = c + \sum_{i=1}^{p} \phi_i \nabla^d y_{t-i} + \epsilon_t + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} \tag{3.4}$$

For return series, such as those used in this study, stationarity is typically achieved with no differencing (*d = 0*), since returns fluctuate around a relatively stable mean (Tsay, 2010).

The fitting of ARIMA models generally follows the Box–Jenkins methodology: (i) testing for stationarity using the Augmented Dickey–Fuller (ADF) and KPSS tests, (ii) examining the autocorrelation (ACF) and partial autocorrelation (PACF) plots to guide the selection of AR and MA orders, (iii) estimating candidate models and comparing them using information criteria such as AIC and BIC, and (Nelson et al.)conducting residual diagnostics, including the Ljung–Box test, to confirm adequacy of fit. To maintain parsimony, the search for orders is restricted to small values (e.g., $p, q \leq 2$).

3.2 Long Short-Term Memory (LSTM) Networks

While ARIMA provides a transparent baseline for time-series forecasting, its linear structure limits its ability to capture nonlinear dependencies and long-range temporal effects that are often present in financial markets. To address these limitations, researchers have turned to neural network-based models, which can learn complex nonlinear relationships directly from data without strong statistical assumptions.

Among neural networks, Artificial Neural Networks (ANNs) were the first to be widely applied in forecasting. ANNs consist of layers of interconnected nodes that compute weighted sums of inputs and apply nonlinear activation functions. Although effective for static pattern recognition, feedforward networks process each input independently and cannot capture sequential dependencies. This limitation makes them unsuitable for modelling financial time series, where temporal structure is central. The figure is used from the articles An Introduction to Neural Networks for Beginners (Thomas, 2017).

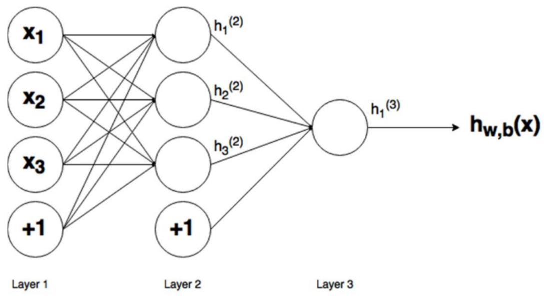


Figure 3.2.1 Three Layer Neural Network

To overcome the limitation of feedforward networks in modeling temporal dependencies, Recurrent Neural Networks (RNNs) introduce recurrence, allowing information to persist over time. The hidden state at each time step depends on both the current input and the previous hidden state, as represented by the general form

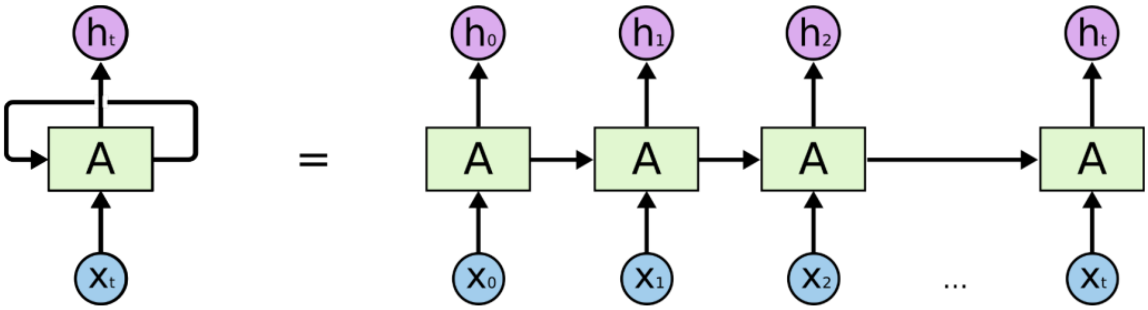
$$h^{(t)} = f\left(h^{(t-1)}, x^{(t)}; \theta\right) \tag{3.5}$$

where $f(\cdot)$denotes a nonlinear activation function and θrepresents the learnable parameters (Goodfellow et al., 2016). This formulation can be expressed more concretely as

$$h_t = f(W_{xh}x_t + W_{hh}h_{t-1} + b_h) \tag{3.6}$$

where $W_{xh}$and $W_{hh}$are the weight matrices for the input and recurrent connections, and $b_h$is the bias term. This recurrence provides a mechanism for memory across time, enabling RNNs to model sequential patterns in financial data. However, when trained through backpropagation through time (BPTT), standard RNNs often suffer from vanishing or exploding gradients, which hinder their ability to learn long-term dependencies (Bengio et al., 1994). The figure is used from Colah's Blog,(Olah, 2015).

Figure 3.2.2 A Basic RNN-Unrolled Recurrent Neural Network



To address these challenges, Hochreiter et al. (1997) introduced the Long Short-Term Memory (LSTM) network, a specialized RNN architecture. Instead of a simple recurrent unit, the LSTM cell contains a cell state C_t acting as a memory highway, and three gates (input, forget, output) that regulate information flow:

- **Forget gate:** decides what information to discard.
- **Input gate:** selects which new information to store.
- **Cell update**: combines retained past memory and candidate values.
- **Output gate:** determines what information contributes to the hidden state h_t.

The internal operations of an LSTM cell can be expressed through the following set of equations (Olah, 2015):

$$\begin{aligned} f_t &= \sigma(W_f[h_{t-1}, x_t] + b_f) \\ i_t &= \sigma(W_i[h_{t-1}, x_t] + b_i) \\ \tilde{C}_t &= \tanh(W_C[h_{t-1}, x_t] + b_C) \\ C_t &= f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \\ o_t &= \sigma(W_o[h_{t-1}, x_t] + b_o) \\ h_t &= o_t \odot \tanh(C_t) \end{aligned} \tag{3.7}$$

where sigmoid (σ) functions act as filters, while tanh bounds memory contributions within [−1, 1]. Together, they allow the LSTM to retain long-term dependencies while flexibly adapting to new inputs. This design is especially valuable in finance: sigmoid gates mimic selective market responses, while tanh ensures stability under volatility. Figure 3.2.3 which is adapted from Colah's Blog (Olah, 2015) illustrating the input, output, and cell states along with the gating mechanisms.

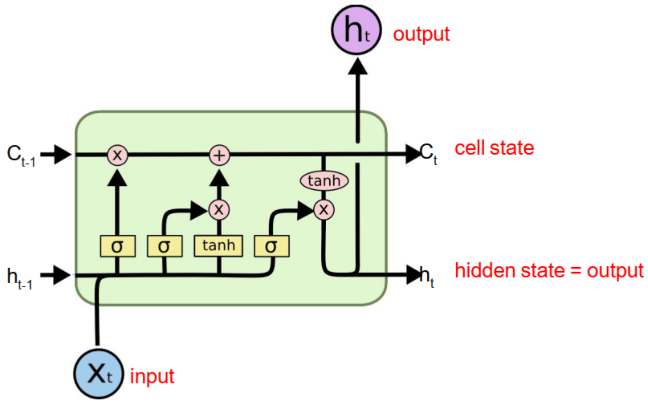


Figure 3.2.3 LSTM cell structure

While LSTMs effectively capture temporal dependencies, they treat all inputs with equal importance. However, in real-world financial data, certain time steps or features may carry more significance than others. To overcome this, attention mechanisms have been developed, which will be discussed in the following section.

3.3 Attention Mechanism

While LSTMs can capture sequential dependencies, they treat each input in the sequence with equal weight. In practice, certain time steps or features may be more relevant than others. To address this limitation, attention mechanisms (Bahdanau et al., 2014) dynamically assign weights to hidden states, allowing the model to focus on the most informative signals.

Formally, for a hidden state h_t, an alignment score is computed as:

$$e_t = v^\top \tanh(W_h h_t + b) \tag{3.8}$$

These scores are normalized through a softmax function to produce attention weights α_t :

$$\alpha_t = \frac{\exp(e_t)}{\sum_k \exp(e_k)} \tag{3.9}$$

The final context vector is a weighted sum of hidden states:

$$c = \sum_t \alpha_t h_t \tag{3.10}$$

where α_t reflects the relative importance of each time step. The context vector c can then be combined with the current hidden state to improve prediction. The overall structure of this additive attention mechanism is illustrated in Figure 3.3.1 which is adapted from Bahdanau et al. (2014), illustrating how attention weights ($\alpha_{t,i}$) are computed over encoder hidden states.

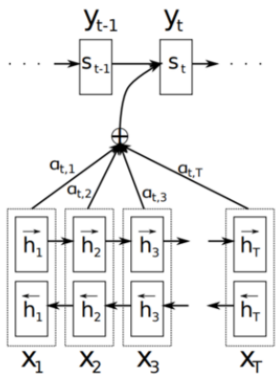


Figure 3.3.1 Bahdanau attention mechanism

In this report, attention is integrated with LSTM networks to achieve two objectives:

(1) Enhanced performance by enabling the model to prioritize relevant past signals and reduce noise.

(2) Interpretability by revealing which time steps and features contribute most to forecasts, complementing statistical tests such as Granger causality.

3.4 Data

(1) Dataset collection

Daily index values for the S&P/NZX 50 and eleven sectoral indices were obtained from S&P Dow Jones Indices LLC. The dataset spans from 1 January 2015 to 31 December 2024, totaling 2,505 daily records per index. The indices include Energy, Materials, Industrials, Consumer Discretionary, Consumer Staples, Health Care, Financials, Information Technology, Communication Services, Utilities, and Real Estate. The Index Value field was selected as the primary measure, representing the official daily closing level of each index (S&P, 2024). Other columns (e.g. Market Capitalization, Divisor, and Count) were excluded as they are relevant only for index construction and not for predictive modelling.

Each sector index in the dataset follows the official naming convention published by S&P Dow Jones Indices. Table 3.4.1 lists the sector indices, their corresponding codes, and the variable names used in this study. Daily log returns were computed for each index and used as model input features (variables named with the suffix _ret, denoting daily returns).

Table 3.4.1 Description of S&P/NZX Sector Indices

Variable Name	**Index Name**	**Index Code**	**Description**
SPNZXALL60_ret	S&P/NZX All Real Estate (Sector)	SPNZXALL60	Property investment and real estate development companies
SPNZXALL10_ret	S&P/NZX All Energy (Sector)	SPNZXALL10	Energy generation and distribution companies
SPNZXALL15_ret	S&P/NZX All Materials (Sector)	SPNZXALL15	Mining, raw-material extraction, and processing industries
SPNZXALL20_ret	S&P/NZX All Industrials (Sector)	SPNZXALL20	Manufacturing, transport, and logistics firms
SPNZXALL25_ret	S&P/NZX All Consumer Discretionary (Sector)	SPNZXALL25	Retail, leisure, and consumer-service companies
SPNZXALL30_ret	S&P/NZX All Consumer Staples (Sector)	SPNZXALL30	Food, beverage, and essential-goods producers
SPNZXALL35_ret	S&P/NZX All Health Care (Sector)	SPNZXALL35	Medical, pharmaceutical, and biotechnology firms
SPNZXALL40_ret	S&P/NZX All Financials (Sector)	SPNZXALL40	Banks, insurance, and other financial institutions
SPNZXALL45_ret	S&P/NZX All Information	SPNZXALL45	Technology, software, and IT-service firms

	Technology (Sector)		
SPNZXALL50_ret	S&P/NZX All Communication Services (Sector)	SPNZXALL50	Telecommunications and digital media providers
SPNZXALL55_ret	S&P/NZX All Utilities (Sector)	SPNZXALL55	Electricity, water, and infrastructure services
SPNZXNZ50_ret	S&P/NZX 50 Index	SPNZXNZ50	Market-wide benchmark representing the 50 largest NZX-listed companies

Macroeconomic series were collected from the Reserve Bank of New Zealand (RBNZ, 2024), namely:

- NZD/USD exchange rate,
- Official Cash Rate (OCR), and
- Consumer Price Index (CPI)

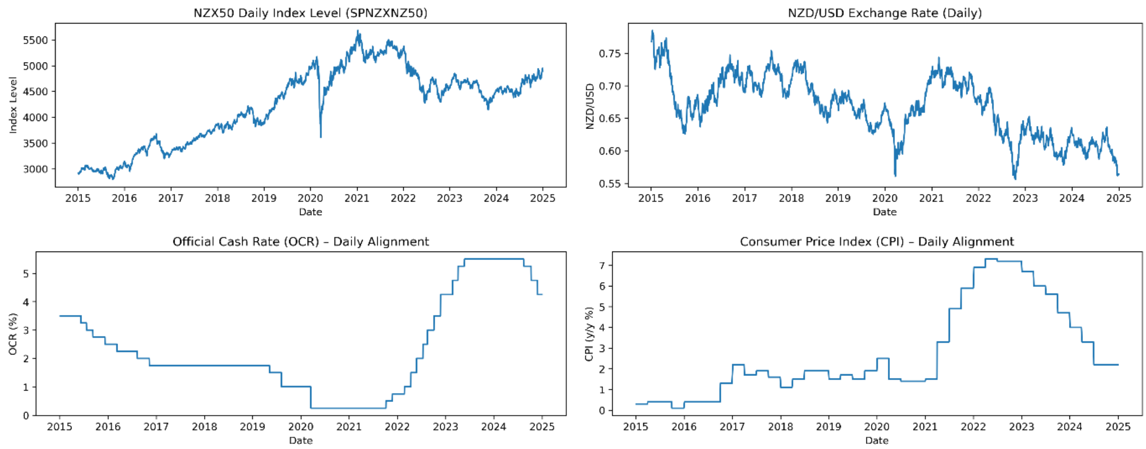
Indicators were selected based on established evidence: inflation (CPI) and short-term interest rates are classic predictors of equity returns (Fama et al., 1977; Kaul, 1987); exchange rates influence stock performance via trade and currency channels (Phylaktis et al., 2005); and policy-rate announcements (OCR) impact New Zealand equity markets (Wang et al., 2012). These macroeconomic variables play a crucial role in shaping movements in the NZX 50 Index, which reflects New Zealand's overall market performance.

The Official Cash Rate (OCR) directly influences borrowing costs and corporate investment decisions because changes in monetary policy often trigger revaluations of equities as investors adjust return expectations. The Consumer Price Index (CPI) captures inflationary trends that affect consumer purchasing power, cost structures, and profit margins, which in turn influence earnings forecasts and stock valuations. The NZD/USD exchange rate impacts export-oriented companies within the NZX 50, as a stronger local currency reduces international competitiveness while a weaker NZD can boost foreign revenue translation. Together, these variables represent the macro-financial channels through which domestic and global economic conditions transmit to the New Zealand equity market.

Since macroeconomic indicators are published at varying frequencies, data alignment was necessary. CPI (quarterly) was mapped to all trading days of its reference quarter (e.g., ‘Mar 2024’ applied to Jan–Mar 2024), yielding a complete daily series. The Official Cash Rate (OCR), although reviewed approximately every six weeks, is published by RBNZ as a daily series where values remain constant between review dates; therefore, it was merged directly without further conversion. The NZD/USD exchange rate series was aligned on its daily frequency. The raw dataset was first aligned on trading days to ensure consistency across the NZX50 index, sectoral indices, exchange rates, and macroeconomic variables. Missing values were checked, though no major gaps were identified in the core return series.

Figure 3.4.1 presents a compact visual overview of the core series to see data continuity and frequency alignment of NZX50, NZD/USD, OCR and CPI. These plots are descriptive only and do not inform model estimation.

Figure 3.4.1 The Trend charts of NZX50, NZD/USD, OCR and CPI.



(2) Data Preprocessing

Return construction

Raw index levels were transformed into log returns to achieve stationarity and comparability across indices:

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \tag{3.11}$$

where P_t is the index value at time t (Tsay, 2010).

Missing Value Handling

Missing values can undermine analysis if left unaddressed, so we first checked the raw source series (NZX indices and NZD/USD, OCR, CPI). This initial inspection found no missing value across the 2015–2024 period. Missing entries appeared only after feature engineering, and the pattern depended on the model's input set:

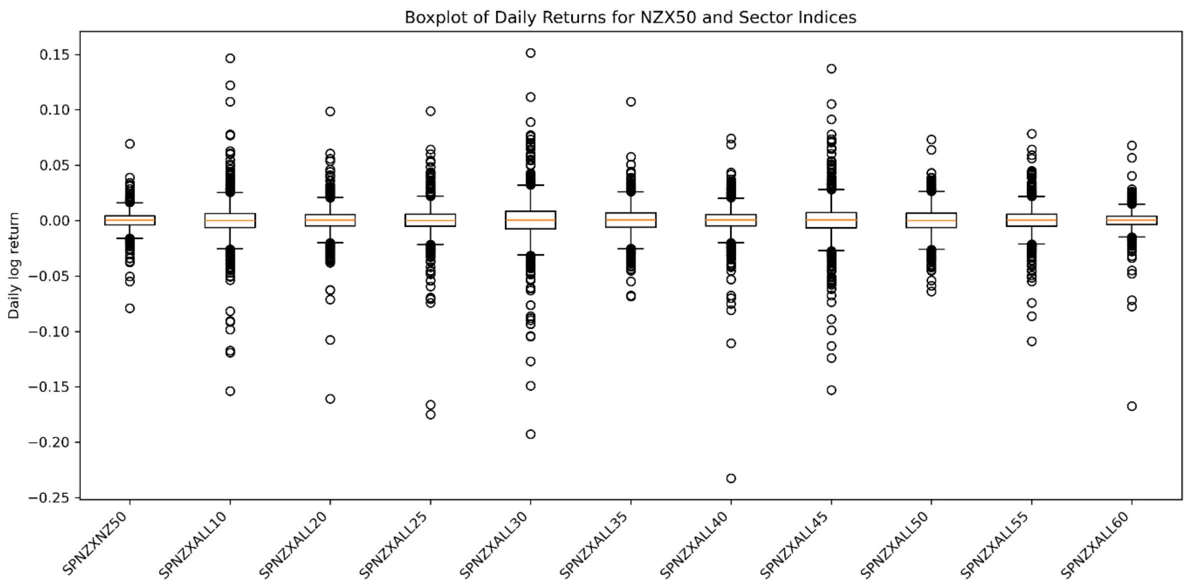
- **ARIMA model:** Used only NZX50 returns, so no additional missing values were introduced.
- **Baseline LSTM model:** Incorporated basic technical indicators such as moving averages and RSI. Rolling calculations created undefined values at the beginning of the series. To address this, the first 19 rows were removed, resulting in 2,486 usable observations.
- **Attention-LSTM models:** Expanded to include sector indices, broader technical indicators, and macroeconomic variables. Longer rolling windows introduced more undefined values. The first 252 rows were dropped and reduced the dataset from 2,505 to 2,253 observations.

The missing values in this report occur only at the beginning of the dataset and are a direct result of rolling-window calculations, such as moving averages, volatility, and momentum. These indicators require a certain number of past observations before they can be computed, so the earliest rows cannot produce valid outputs. Importantly, these missing entries are not caused by errors in the data collection process or actual breaks in the market data but arise from the way features are constructed. Instead of trying to impute, such as by backward filling or mean substitution, the more appropriate solution is to remove the affected rows. Backward filling would create look-ahead bias, because it would use future information to estimate earlier values, which is not realistic in a forecasting context. Other imputation methods would also change the statistical behaviour of returns, for example by smoothing volatility or altering momentum patterns. By removing the initial rows, the dataset is kept in its original form without introducing misleading adjustments. This approach is widely accepted in financial time-series analysis, as the loss of early observations does not meaningfully reduce the sample size for modelling.

Handling Outliers

Daily returns occasionally contain extreme values caused by sudden shocks, policy changes, or market news. These are not data errors but real features of financial markets. Removing them would alter the distribution of returns and weaken the realism of the study. Figure 3.4.2 shows boxplots for daily returns for the NZX50 and sector indices. While most values are clustered around zero, a few points extend beyond the whiskers, highlighting the heavy-tailed nature of financial returns. Some sectors display wider spreads and more outliers. However, NZX50 index appears relatively stable due to diversification. These outliers represent real market conditions. Therefore, retained them to ensure that models are trained on data that reflects the full dynamics of the market.

Figure 3.4.2 Boxplot of daily returns for the NZX50 and sector indices



(3) Feature Engineering

Feature engineering was tailored to the requirements of each model class. While ARIMA relies purely on past returns, the LSTM-based models incorporate technical, sectoral, and macroeconomic features to capture richer market dynamics.

- **ARIMA model:**

No additional features were engineered beyond the NZX50 daily returns.

- **Baseline LSTM model:**

  The baseline LSTM included the NZX50 returns along with four standard technical indicators designed to capture short- and medium-term market patterns:

  - **SMA20** (20-day Simple Moving Average) – trend indicator.
  - **EMA20** (20-day Exponential Moving Average) – smoother trend indicator with more weight on recent prices.
  - **RSI14** (14-day Relative Strength Index) – momentum indicator reflecting overbought/oversold conditions.
  - **MACD** (12–26, 9) – momentum and trend reversal indicator based on the difference of short- and long-term EMAs with a 9-day signal line.

- **Attention-LSTM models (Temporal and Feature Attention):**

  The attention-based models expanded the feature space to capture both market structure and macroeconomic influences. In addition to the NZX50 returns, sectoral indices (SPNZXALL10–SPNZXALL60) were incorporated to model cross-sector dynamics. A broader set of technical indicators was engineered, including:

  - **SMA20, SMA60** and **EMA20, EMA60** for short- and medium-term trend detection.
  - **RSI14** and **MACD** (12–26, 9) for momentum and reversal signals.
  - **Volatility** (Vol21) measured as 21-day rolling standard deviation.
  - **Momentum** (Mom63) based on 63-day return differences to capture quarterly shifts.

  Macroeconomic variables included the Official Cash Rate (OCR) and Consumer Price Index (CPI), represented both as levels and as temporal differences (1-day, 3-month, and 12-month changes). The NZD/USD exchange rate was incorporated as raw returns and corresponding technical indicators (SMA, EMA, RSI, MACD, volatility, momentum), reflecting international capital flows in New Zealand's open economy.

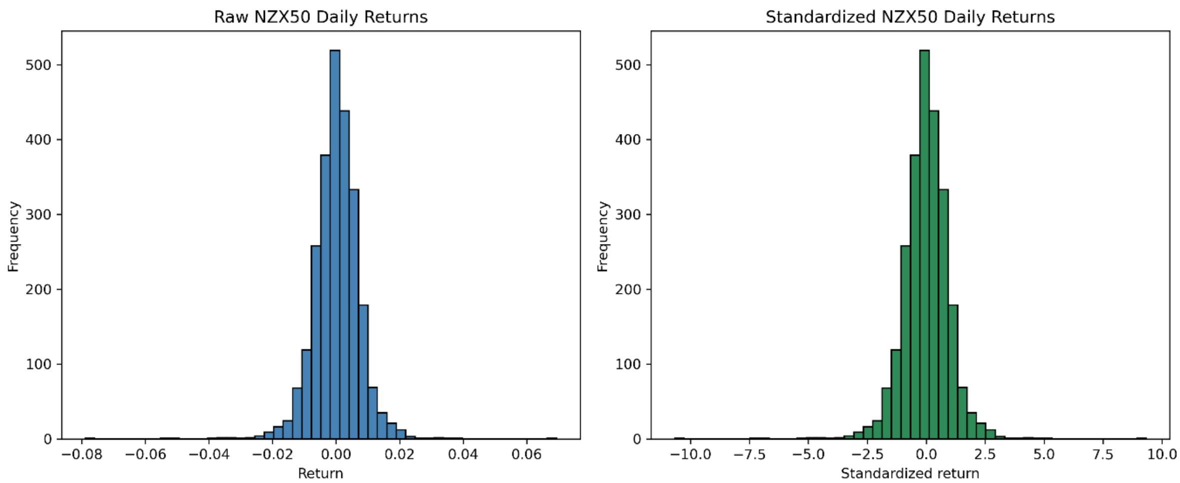
(4) Normalization & Scaling

The predictors used in this study include raw returns, technical indicators, and macroeconomic variables. These variables differ in their numerical ranges; for example, exchange rates and index levels are expressed in different units, while technical indicators such as RSI or MACD have bounded or scaled values. Feeding such heterogeneous inputs directly into a neural network can bias the learning process, since variables with larger magnitudes may dominate the optimization. To address this, all predictors for the LSTM-based models were standardized using z-score normalization:

$$x' = \frac{x-\mu}{\sigma} \tag{3.12}$$

where μ and σdenote the mean and standard deviation of each variable, estimated from the training set. The same transformation was applied to the validation and test sets to avoid information leakage. Figure 3.4.3 illustrates the effect of standardization on NZX50 returns: the raw distribution is rescaled to mean zero and unit variance without altering its heavy-tailed nature. This ensured that all inputs contributed on a comparable scale and helped stabilize the training process. Z-score normalization was chosen over Min–Max scaling because it preserves distributional properties and is less sensitive to outliers, which is preferable for financial return series.

Figure 3.4.3 Distribution of NZX50 daily returns before and after z-score normalization



In contrast, the ARIMA model was estimated directly on the NZX50 return series. Because returns are already stationary and centred around zero, no additional normalization or scaling was required.

(5) Train–Validation–Test Split

The dataset was partitioned chronologically into training (2015–2020), validation (2021–2022), and test (2023–2024) sets. The training set was used to fit models, while the validation set was primarily employed for hyperparameter tuning and model selection in the LSTM-based approaches. For ARIMA, the validation set is used only for stability checks. The test set was held out for final evaluation, simulating real-world forecasting on unseen data. This scheme ensured consistent and fair comparability across models.

(6) Hyperparameter Selection

A grid search was conducted to identify optimal hyperparameters for the LSTM and Attention-LSTM models. Because LSTM performance is highly sensitive to architectural and training parameters, each configuration in the predefined grid was trained on the training set and evaluated on the validation set, with the model achieving the lowest validation MAE selected as optimal. The grid was deliberately kept compact to maintain reproducibility and computational efficiency under the rolling-origin evaluation setup.

In this study, the following key hyperparameters were considered:

- Number of layers and hidden units:

We used two stacked LSTM layers, with the first containing 64–128 units and the second 32–64 units. While a single-layer LSTM can capture basic temporal dependencies, stacking layers enhances the model’s ability to learn more complex sequential patterns. However, deeper architectures may increase the risk of overfitting, especially with limited financial data such as in the NZX market. The chosen hidden unit ranges (64–128 / 32–64) represent a moderate network capacity that balances learning flexibility and generalization.

- Sequence length (look-back window):

Look-back windows of 30 and 60 trading days were tested to capture both short-term and medium-term dependencies. This choice balances capturing memory without excessive computational overhead.

- Dropout rate:

Dropout values of 0.2 and 0.3 were tested to prevent overfitting by randomly deactivating a fraction of neurons during training. Moderate dropout is effective in regularizing neural networks without significantly hindering learning capacity. Lower rates (e.g., <0.1) may provide insufficient regularization, while higher rates (>0.5) can lead to underfitting. The chosen range balances model generalization and convergence stability, which is particularly important for financial data prone to noise and limited sample size.

- Batch size:

Mini-batch sizes of 32 and 64 were considered. Batch size influences both training stability and speed, with smaller batches providing more frequent updates at the cost of noisier gradients.

- Learning rate:

Two candidate values, 0.001 and 0.0005, were tested to control how quickly the model updates its weight during training. The value 0.001 is the standard default for the Adam optimizer (Kingma & Ba, 2015) and is widely used in LSTM applications, while 0.0005 was included as a more conservative alternative to improve stability on noisy financial data. A very high learning rate can make training unstable, whereas a very small one can slow learning. These values provide a practical balance between convergence speed and stability for the NZX dataset.

- Epochs:

The maximum number of training epochs was set to 50. This limit provides enough iterations for the model to learn meaningful patterns while avoiding unnecessary computation and overfitting. Early Stopping was applied to monitor validation loss and automatically stop training once improvement plateaued, ensuring efficient convergence without manual tuning

- Early stopping patience: The patience parameter controls how many epochs the training continues after the validation loss stops improving. A patience value of 10 was used for the baseline LSTM and 8 for the attention-based models. The slightly shorter patience in the attention models reflects their faster convergence, as attention layers accelerate learning by focusing on relevant temporal or feature patterns. In both cases, these settings provided a good balance between allowing sufficient training time and preventing overfitting or unnecessary computation.

3.5 Evaluation

A forecasting model should not be judged solely by its in-sample statistical fit, but by its ability to deliver accurate and useful predictions in an out-of-sample setting. To this end, three complementary evaluation dimensions are employed: forecast accuracy, directional accuracy, and economic utility. Together, these measures provide a holistic assessment of whether deep learning models such as LSTM with attention offer meaningful improvements over the classical ARIMA benchmark.

(1) Mean Absolute Error (MAE)

The Mean Absolute Error is a widely used measure of forecast accuracy. It quantifies the average magnitude of prediction errors, regardless of direction:

$$MAE = \frac{1}{N}\sum_{t=1}^{N} |y_t - \hat{y}_t|, \tag{3.13}$$

where y_t is the observed return at time t, $\hat{y}_t$ is the forecasted return, and N is the number of forecasts.

MAE is more robust to outliers than the Mean Squared Error (MSE), making it particularly suitable in financial markets where large shocks may distort squared-error measures. In this study, MAE is reported separately for each horizon (1 day, 1 month, 3 months, and 1 year) to evaluate model performance across different forecast lengths.

(2) Directional Accuracy (DA)

While MAE measures the magnitude of forecast errors, investors are often more concerned with whether the model correctly predicts the direction of price movement (up or down). Directional Accuracy (DA) is calculated as:

$$DA = \frac{1}{N}\sum_{t=1}^{N} I\big(\text{sign}(y_t) = \text{sign}(\hat{y}_t)\big), \tag{3.14}$$

where $I(\cdot)$ is an indicator function equal to 1 if the forecast and actual returns share the same sign, and 0 otherwise.

A DA of 0.5 corresponds to random guessing in a binary up/down market. Values above 0.5

indicate predictive value, while values close to 1 suggest near-perfect directional forecasting. This measure is particularly relevant to trading strategies, where even small improvements in directional accuracy can translate into significant economic gains.

(3) Economic Utility

Ultimately, the value of a forecasting model lies in its ability to improve investment outcomes. To measure this, a simple trading strategy is simulated:

- If the forecasted return is positive $(\hat{y}_t > 0)$, take a long position.
- If the forecasted return is negative $(\hat{y}_t < 0)$, take a short position.

The realized strategy return at time t is then:

$$r_t^{\text{strategy}} = \text{sign}(\hat{y}_t) \cdot y_t. \tag{3.15}$$

From these returns, cumulative profits and the Sharpe ratio are computed:

$$Sharpe = \frac{E[r_t^{\text{strategy}}]}{\sigma(r_t^{\text{strategy}})}, \tag{3.16}$$

where the numerator is the mean strategy return, and the denominator is its standard deviation. This evaluation provides a bridge between statistical accuracy and real-world profitability. A model with slightly better MAE or DA may nonetheless deliver substantially higher cumulative returns if it consistently positions on the correct side of the market.

(4) Out-of-Sample Evaluation Protocol

To avoid look-ahead bias, all models are assessed using a rolling-origin evaluation. At each origin, the model is trained on data up to time t and produces forecasts for horizons h{1,21,63,252}. The window is then advanced, and the process is repeated. This protocol mirrors realistic trading conditions, where forecasts must be generated sequentially without access to future information. Performance metrics (MAE, DA, Sharpe ratio, cumulative return) are computed at each horizon, ensuring a fair and robust comparison of model effectiveness.

3.6 Causality Analysis

Forecast accuracy alone is not sufficient for understanding the dynamics of financial markets.

Investors and policymakers are often equally interested in the drivers of market movements and the interdependencies between sectors. To address this, the present study incorporates two complementary approaches to causality analysis: Granger causality tests, which provide a statistical measure of temporal influence, and attention interpretability, which provides model-based insight into feature relevance. Together, these methods enrich our understanding of cross-sector dynamics in the New Zealand equity market.

(1) Granger Causality

Granger causality is a statistical framework used to test whether one time series contains predictive information about another (Granger, 1969). A series x_t is said to Granger-cause y_t if past values of x_t improve the forecast of y_t beyond what is possible using past values of y_t alone. Formally, consider the following bivariate autoregressive model:

$$y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{j=1}^{q} \beta_j x_{t-j} + \varepsilon_t. \tag{3.17}$$

The null hypothesis of no Granger causality is:

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_q = 0. \tag{3.18}$$

If this hypothesis is rejected, then past values of x_t contain useful information for predicting y_t . In this report, daily returns of NZX sector indices are tested pairwise to uncover directional linkages. For example, the test may reveal whether shocks in the Financial sector lead returns in the Real Estate sector, reflecting inter-industry dependencies. Such findings help identify causal chains (e.g., Energy → Industrials → Consumer Discretionary) that shape overall market behavior.

(2) Attention-Based Interpretability

While Granger causality provides a classical statistical perspective, the LSTM with attention offers a model-based view of causality. The attention mechanism assigns weights α_t to hidden states or features, effectively quantifying their relative importance in producing the forecast:

$$c = \sum_{t=1}^{T} \alpha_t h_t. \tag{3.19}$$

where a higher weight α_t indicates that the model has focused more strongly on that particular lag or feature when generating the prediction. This study leverages attention distributions in two ways:

a. **Cross-feature insight**: By examining attention weights on sectoral returns and macroeconomic variables, we can identify which inputs exert the greatest influence at different forecast horizons.
b. **Temporal dynamics**: Attention distributions reveal whether short-term shocks (e.g., last 5 days) or longer-term signals (e.g., 2–3 months ago) are driving forecasts.
c. **Interpretability**: This approach helps to demystify the “black box” nature of deep learning, providing an interpretable link between inputs and predictions.

(3) Integration of Approaches

The two causality frameworks serve complementary purposes. Granger causality offers a formal hypothesis test grounded in linear statistical models, providing clear null-hypothesis conclusions. Attention interpretability captures nonlinear dependencies and cross-feature interactions that Granger causality may miss.

By combining these approaches, this study not only evaluates predictive performance but also provides actionable insights into the causal structure of the NZX market. This dual perspective strengthens both the academic contribution and the practical relevance of the research.

Chapter 4 Results

The main content of this chapter presents the empirical results obtained from forecasting the NZX50 index across multiple horizons (h=1, 21, 63, 252). Results are reported for baseline models (ARIMA), benchmark deep learning models (LSTM), and proposed attention-based extensions.

4.1 Data Description

The dataset covers daily trading data for the NZX50 index, eleven sector indices, and three macroeconomic variables (NZD/USD exchange rate, Official Cash Rate (OCR), and Consumer Price Index (CPI). Returns were calculated as daily log-changes for all financial variables, while first-differences were applied to OCR and CPI to capture discrete adjustments.

Descriptive statistics show that mean daily returns across indices are close to zero, consistent with efficient market behaviour. The NZX50 return has a mean of –0.00002, with a standard deviation of 0.0074, reflecting moderate volatility. Sector indices exhibit higher dispersion. The most volatile sectors include SPNZXALL15(Material) and SPNZXALL30(Consumer Staples). More defensive sectors, such as SPNZXALL20(Industrial) and SPNZXALL60(Real Estate), show lower variability. Skewness and kurtosis values indicate that returns are non-normal, with fat tails and occasional extreme movements. This feature is typical of financial markets and supports the use of non-linear models in later chapters. For macroeconomic series, NZD/USD returns are less volatile. OCR and CPI appear more volatile, but this reflects infrequent step-like changes rather than daily variation. Table 4.1.1 presents the descriptive statistics for daily returns and macroeconomic changes. Figure 4.1.1 presents a histogram of NZX50 returns, showing the high concentration near zero with heavy tails.

Figure 4.1.1 Histogram of NZX50 returns

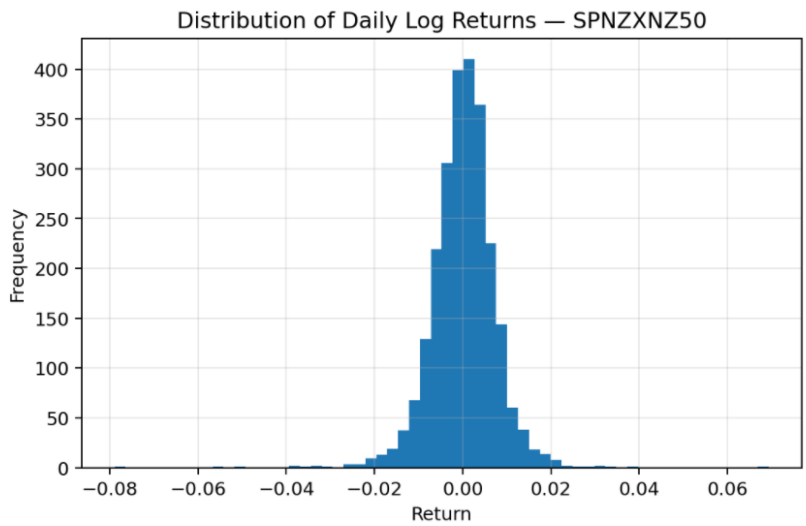


Table 4.1.1 Descriptive Statistics of Daily Log Returns and Changes

Variable	count	mean	std	min	25%	50%	75%	max	skew	kurtosis
SPNZXALL10_ret	2504	0.000102	0.014307	-0.15383	-0.00636	0	0.006433	0.146495	-0.31392	20.15916
SPNZXALL15_ret	2504	-0.00013	0.016474	-0.11861	-0.00903	-1.9E-05	0.009139	0.137312	-0.30664	6.387111
SPNZXALL20_ret	2504	0.000252	0.010343	-0.16067	-0.00476	0.000302	0.005435	0.098457	-1.6608	34.40382
SPNZXALL25_ret	2504	-0.00025	0.012058	-0.17487	-0.00525	0.00006	0.005659	0.09884	-2.30593	37.42185
SPNZXALL30_ret	2504	0.000322	0.017179	-0.19263	-0.00751	0.000271	0.008343	0.151475	-0.79927	16.51766
SPNZXALL35_ret	2504	0.000422	0.011334	-0.06836	-0.00591	0.000702	0.006956	0.107207	0.159938	5.417016
SPNZXALL40_ret	2504	0.000021	0.011323	-0.23245	-0.00481	0.00057	0.005327	0.07416	-4.25594	78.82452
SPNZXALL45_ret	2504	0.000469	0.015166	-0.15281	-0.00654	0.000516	0.007285	0.137347	-0.32335	14.20477
SPNZXALL50_ret	2504	0.000107	0.011464	-0.06386	-0.0064	0.0001	0.006708	0.073367	-0.06476	2.462094
SPNZXALL55_ret	2504	0.000301	0.010545	-0.10881	-0.00503	0.000413	0.005736	0.07827	-0.5744	11.06197
SPNZXALL60_ret	2504	0.000016	0.008363	-0.16723	-0.0037	0.000186	0.003773	0.067753	-3.50003	72.06205
SPNZXNZ50_ret	2504	0.000206	0.007418	-0.0789	-0.0038	0.000417	0.004346	0.069366	-0.57739	11.85909
NZDUSD_ret	2504	-0.00012	0.00683	-0.053	-0.00412	-0.00014	0.00416	0.029258	-0.2131	2.237532
OCR_chg	2504	0.0003	0.042104	-0.75	0	0	0	0.75	2.238528	163.6854
CPI_chg	2504	0.000759	0.080966	-1.1	0	0	0	1.8	6.937689	226.9485

4.1.2 Correlation Analysis

To further understand the relationships among the NZX indices and macroeconomic variables, pairwise correlations of daily returns were computed. The results are shown in Table 4.1.2 and visualised in Figure 4.1.2.

The analysis shows that sector indices are positively correlated overall, but the strength of the relationships varies. The strongest correlations are observed between sector indices and the NZX50 benchmark. For example, SPNZXALL20(Industrials) and NZX50 have a correlation of 0.73, while SPNZXALL35(Healthcare) and NZX50 have a correlation of 0.72. These results indicate that movements in sectors such as Industrials and Healthcare strongly influence the overall market index.

In contrast, the sectors show very weak correlations with each other. For instance, SPNZXALL15(Material) and SPNZXALL30(Consumer Staples) are only weakly associated (0.10), and SPNZXALL10(Energy) with SPNZXALL30(Consumer Staples) also shows a low correlation (0.10). This suggests that certain sector pairs behave independently and do not move in tandem.

Overall, correlations among the sector indices range from as low as 0.10 to as high as 0.73, highlighting both tightly linked and loosely connected relationships within the NZX market. This variation supports the later use of feature attention in the forecasting models, as the degree of interdependence differs across sectors. NZD/USD returns display weak positive correlations with most equity indices, typically below 0.20. OCR and CPI changes show negligible correlations with daily returns, reflecting their lower frequency and stepwise adjustments. While these macro variables do not exhibit strong daily co-movement with equities, they remain important explanatory factors that may influence returns over longer horizons.

Figure 4.1.2 Correlation Heatmap

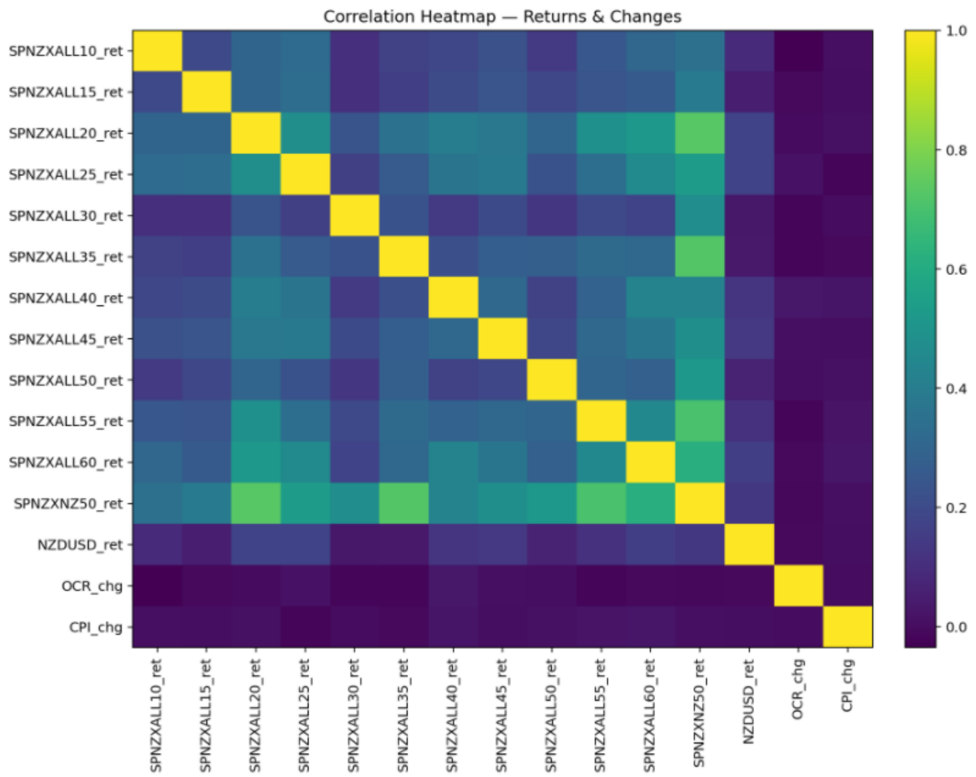


Table 4.1.2 Correlation Matrix Table

	SPNZX ALL10_ ret	SPNZX ALL15_ ret	SPNZX ALL20_ ret	SPNZX ALL25_ ret	SPNZX ALL30_ ret	SPNZX ALL35_ ret	SPNZX ALL40 _ret	SPNZX ALL45_ ret	SPNZX ALL50_ ret	SPNZX ALL55_ ret	SPNZX ALL60 _ret	SPNZX NZ50_re t	NZDUSD _ret	OCR_chg	CPI_chg
SPNZXALL10_ret	1	0.192	0.2999	0.3273	0.1076	0.1711	0.1872	0.2238	0.1397	0.2474	0.3075	0.3487	0.0944	-0.0348	0.0086
SPNZXALL15_ret	0.192	1	0.298	0.3339	0.1035	0.1585	0.206	0.2387	0.184	0.2399	0.2582	0.3889	0.0522	-0.0066	0.0035
SPNZXALL20_ret	0.2999	0.298	1	0.4757	0.2348	0.351	0.3985	0.3782	0.3041	0.4855	0.5159	0.7298	0.1717	-0.0038	0.0127
SPNZXALL25_ret	0.3273	0.3339	0.4757	1	0.1631	0.2641	0.3634	0.3837	0.2249	0.339	0.4563	0.5365	0.1732	0.017	-0.0192
SPNZXALL30_ret	0.1076	0.1035	0.2348	0.1631	1	0.2252	0.1402	0.1987	0.1309	0.194	0.1714	0.4727	0.0317	-0.016	0.0014
SPNZXALL35_ret	0.1711	0.1585	0.351	0.2641	0.2252	1	0.2185	0.2749	0.2777	0.3245	0.3108	0.7215	0.0378	-0.0209	-0.0102
SPNZXALL40_ret	0.1872	0.206	0.3985	0.3634	0.1402	0.2185	1	0.315	0.1655	0.2886	0.4284	0.4272	0.1248	0.0328	0.0232
SPNZXALL45_ret	0.2238	0.2387	0.3782	0.3837	0.1987	0.2749	0.315	1	0.1872	0.3105	0.3691	0.4796	0.1358	0.0096	0.0024
SPNZXALL50_ret	0.1397	0.184	0.3041	0.2249	0.1309	0.2777	0.1655	0.1872	1	0.3039	0.284	0.5178	0.0681	0.0029	0.0122
SPNZXALL55_ret	0.2474	0.2399	0.4855	0.339	0.194	0.3245	0.2886	0.3105	0.3039	1	0.4499	0.7009	0.1143	-0.0194	0.0197
SPNZXALL60_ret	0.3075	0.2582	0.5159	0.4563	0.1714	0.3108	0.4284	0.3691	0.284	0.4499	1	0.6142	0.158	-0.0076	0.0261
SPNZXNZ50_ret	0.3487	0.3889	0.7298	0.5365	0.4727	0.7215	0.4272	0.4796	0.5178	0.7009	0.6142	1	0.1312	-0.0135	0.0073
NZDUSD_ret	0.0944	0.0522	0.1717	0.1732	0.0317	0.0378	0.1248	0.1358	0.0681	0.1143	0.158	0.1312	1	-0.0068	0.0032
OCR_chg	-0.035	-0.0066	-0.0038	0.017	-0.016	-0.021	0.0328	0.0096	0.0029	-0.019	-0.008	-0.0135	-0.0068	1	-0.0001
CPI_chg	0.0086	0.0035	0.0127	-0.0192	0.0014	-0.01	0.0232	0.0024	0.0122	0.0197	0.0261	0.0073	0.0032	-0.0001	1

4.2 ARIMA

The ARIMA model is implemented as the statistical benchmark for forecasting NZX50 returns. Following the Box–Jenkins methodology, the training data is first tested for stationarity using the Augmented Dickey–Fuller (ADF) and KPSS tests. Both tests confirmed that the return series is stationary without differencing, allowing the differencing order to be set at d=0. Autocorrelation (ACF) and partial autocorrelation (PACF) plots were examined to identify potential autoregressive (Sunki et al.) and moving average (MA) components.

To avoid overfitting, the search for candidate models was restricted to small orders, with $p, q \leq 2$. AIC and BIC values were computed for each specification, and the best-performing order was selected. While BIC favoured a parsimonious constant-mean model (ARIMA(0,0,0)), residual diagnostics indicated significant autocorrelation, suggesting the model was inadequate. AIC, on the other hand, selected ARIMA (2,0,0), which reduced residual dependence and provided a better overall fit. Based on these diagnostics, ARIMA (2,0,0) was retained as the baseline statistical model.

Rolling forecasts were then generated for each horizon ($h = 1, 21, 63, 252$) using a recursive approach. At each test origin, the model was re-estimated on all available data to ensure that parameter estimates reflected the most recent information. Forecast accuracy was measured using statistical metrics (MAE, DA) and economic metrics (Sharpe ratio, cumulative return , and maximum drawdown. Table 4.3 presents the out-of-sample results of ARIMA forecasts on the NZX50 index across four horizons, evaluated on the 2023–2024 test set.

Table 4.2.1 ARIMA Forecasting Results on NZX50 (Test Set: 2023–2024)

Horizon	MAE	DA (%)	Mean Ret.	Sharpe	Cum. Ret.	Max DD
h1	0.0049	49.10%	-0.0004	-0.0616	-0.1958	7.4613
h21	0.0196	50.30%	0.0011	0.048	0.5456	6.2441
h63	0.0331	54.20%	0.0029	0.0729	1.276	62.7615
h252	0.0701	45.20%	0.0138	0.2257	3.4327	0

Table 4.2.1 reports that the model performs close to random at the daily horizon (*h*=1), with Directional Accuracy of 49.1% and a negative Sharpe ratio (–0.06). At the monthly and quarterly horizons (*h*=21, *h*=63), performance improves slightly, with Sharpe ratios of 0.05 and 0.07, respectively. The strongest results occur at the annual horizon (*h*=252), where the model achieves a Sharpe ratio of 0.23 and a cumulative return of 3.43, indicating some ability to capture longer-term market dynamics.

Figure 4.2.1 ARIMA Actual vs Predicted Cumulative Returns

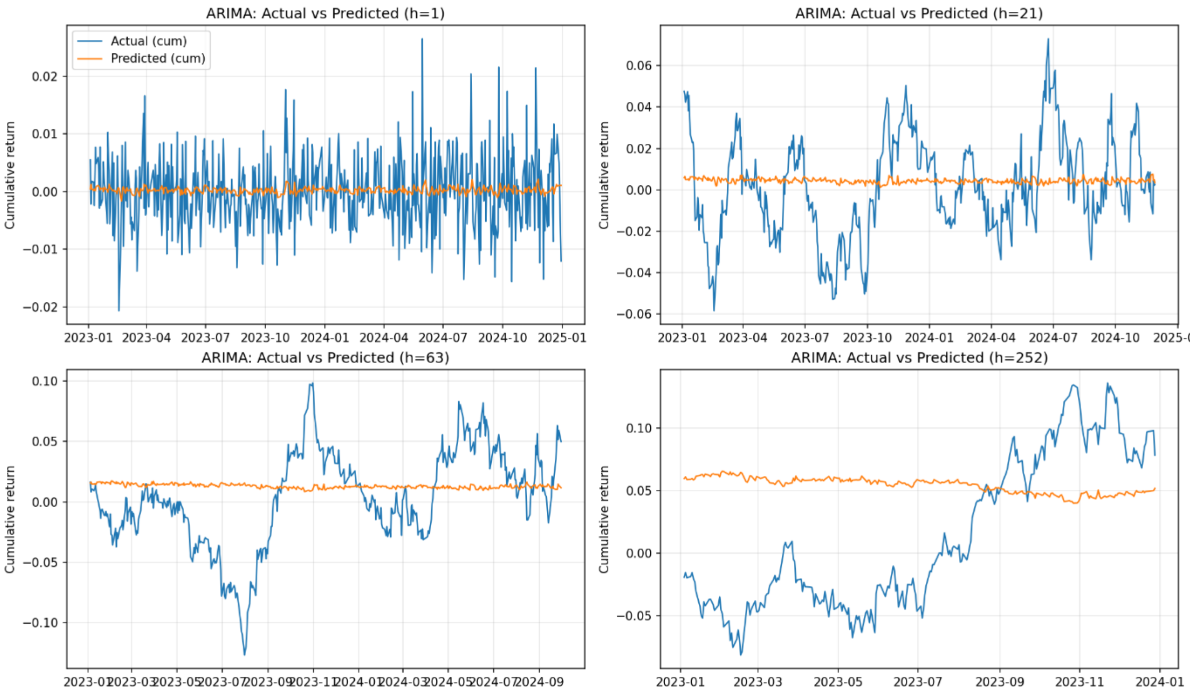


Figure 4.2.1 shows actual (blue) and predicted (orange) cumulative returns across horizons *h*=1,21,63,252. ARIMA forecasts remain almost flat at all horizons, failing to capture the volatility of actual returns. Only at the annual horizon (*h*=252) does the model weakly reflect the upward trend in the data.

Figure 4.2.2 ARIMA Scatter Plots of Realized vs Predicted Returns

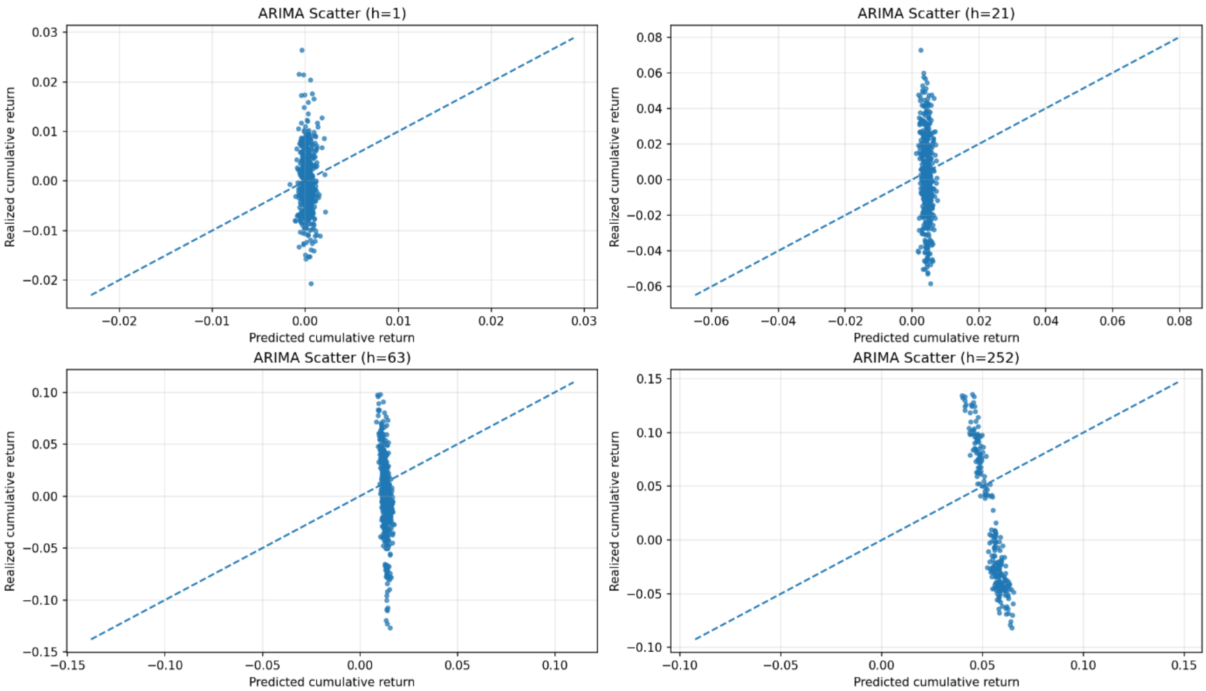


Figure 4.2.2 plots show against predicted returns across horizons. Predictions cluster tightly near zero, while realized returns vary more widely, confirming ARIMA's limited ability to capture daily and medium-term fluctuations.

Table 4.2.2 ARIMA vs Buy & Hold Comparison

Horizon	ARIMA ($1)	Buy & Hold ($1)
h1	$0.81	$1.05
h21	$1.51	$1.51
h63	$2.51	$2.51
h252	$19.38	$19.39

Table 4.2.2 compares the final value of $1 invested using ARIMA forecasts against a passive Buy & Hold strategy. The results show that at the daily horizon (*h*=1), ARIMA underperforms Buy & Hold, reducing final wealth from $1.05 to $0.81. At monthly, quarterly,

and annual horizons, the outcomes of ARIMA are nearly identical to Buy & Hold, suggesting that the model provides little improvement over a passive investment strategy.

The ARIMA results highlight three main points. First, predictive accuracy is weak at short horizons, with daily forecasts performing no better than random guessing. Second, modest improvements appear on medium horizons, though the Sharpe ratios remain low. Finally, at the annual horizon, ARIMA captures some longer-term trends, producing positive risk-adjusted returns. However, when benchmarked against a passive Buy & Hold strategy, ARIMA does not generate additional economic value. At the daily horizon, it even underperforms, while on longer horizons its outcomes are nearly identical to Buy & Hold. Overall, ARIMA has limited ability to forecast short-term NZX50 movements and offers little practical advantage over passive investment.

4.3 LSTM

The Long Short-Term Memory (LSTM) neural network was implemented as a deep learning benchmark for comparison with ARIMA. In this report, a two-layer LSTM architecture is employed, with hyperparameters tuned using a grid search over the training and validation sets. The hyperparameter search space included the number of hidden units in the first and second layers ([32,64] and [16,32], respectively), dropout rates [0.2,0.3], learning rates [0.001,0.0005], batch size [32], epochs [50], and lookback windows [30,60]. Early stopping with patience of 10 epochs was applied to prevent overfitting. Based on validation performance at horizon h=1, the best hyperparameters were: 64 units in the first LSTM layer, 32 units in the second layer, *dropout* = 0.2, *learning rate* = 0.005, *batch size* = 32, *epochs* = 50, and a 30-day lookback window. These parameters were then applied consistently across all horizons to ensure comparability. Rolling forecasts were generated for the test set (2023~2024) across four horizons (h=1,21,63,252) using a recursive prediction approach. Performance was evaluated using the same set of metrics as ARIMA: statistical (MAE, Directional Accuracy) and economic (Sharpe ratio, cumulative return, and maximum drawdown).

Table 4.3.1 LSTM Forecasting Results on NZX50 (Test Set: 2023~2024)

Horizon	MAE	DA(%)	Mean Ret.	Sharpe	Cum. Ret.	Max DD
h1	0.0051	46.30%	0.0001	0.0193	0.061	3.0107
h21	0.0199	50.30%	0.0011	0.048	0.5456	6.2441
h63	0.033	54.20%	0.0029	0.0729	1.276	62.7615
h252	0.0601	45.20%	0.0138	0.2257	3.4327	0

Table 4.3.1 reports the out-of-sample forecasting results for the LSTM model. At the daily horizon (*h*=1), performance is close to random, with Directional Accuracy of 46.3% and a near-zero Sharpe ratio. At monthly and quarterly horizons (*h*=21, *h*=63), results improve slightly, with Sharpe ratios of 0.05–0.07 and cumulative returns between 0.55 and 1.28. At the annual horizon (*h*=252), the model produces a Sharpe ratio of 0.23 and a cumulative return of 3.43, indicating some ability to capture long-term market dynamics.

Figure 4.3.1 LSTM Actual vs Predicted Cumulative Returns

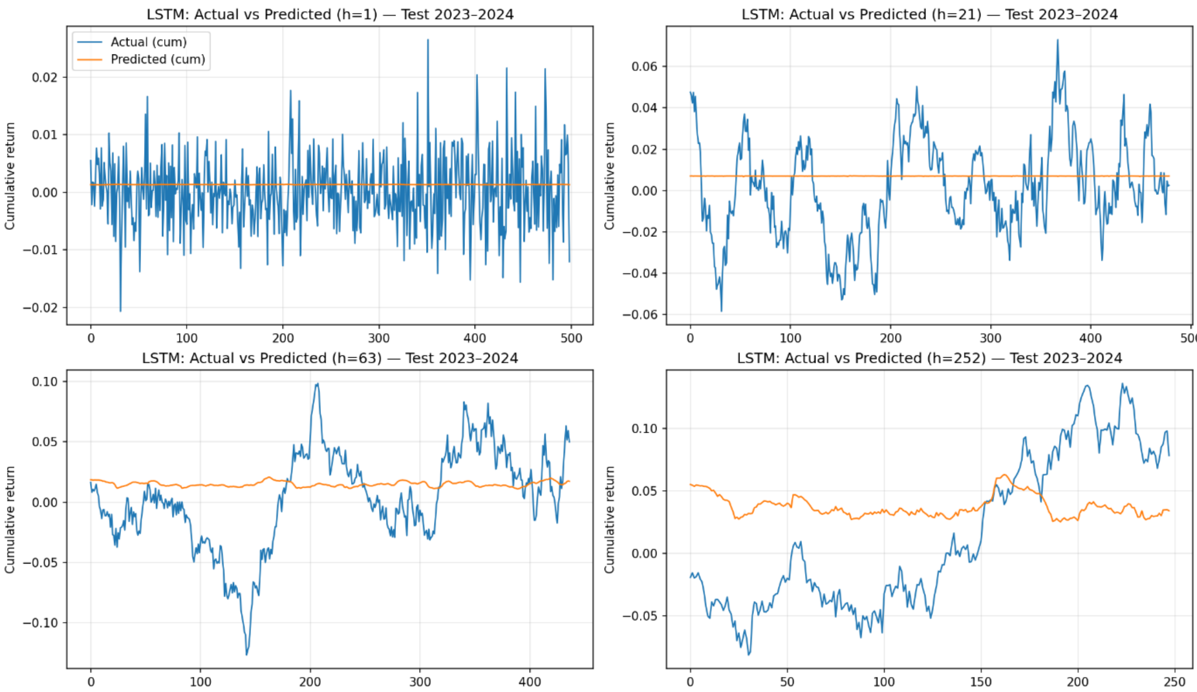


Figure 4.3.1 compares actual (blue) and predicted (orange) cumulative returns for the LSTM model. As with ARIMA, predictions remain flat at short horizons (*h*=1,21), while actual returns fluctuate substantially. At longer horizons (*h*=63,252), predictions align more closely with the upward trend, though they underestimate the magnitude of movements.

Figure 4.3.2 LSTM Scatter Plots of Realized vs Predicted Returns

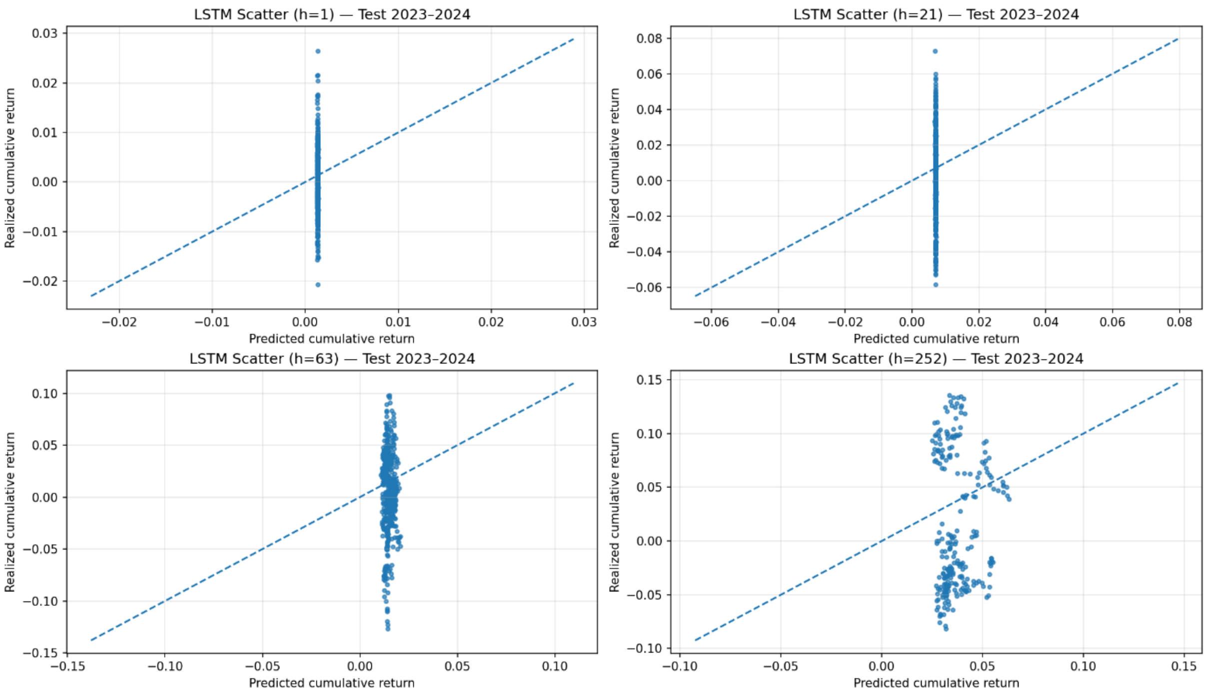


Figure 4.3.2 plots realized against predicted returns. Predictions cluster near zero at short horizons, but dispersion increases at h=252, suggesting the LSTM captures some long-run variation in market returns.

Table 4.3.2 LSTM vs Buy & Hold Comparison

Horizon	LSTM ($1)	Buy & Hold ($1)
h1	$1.05	$1.05
h21	$1.51	$1.51
h63	$2.51	$2.51
h252	$19.38	$19.39

Table 4.3.2 reports the cumulative wealth trajectories of the LSTM model compared with Buy & Hold. Like ARIMA, the LSTM forecasts yield nearly identical outcomes to the passive benchmark, indicating that the model does not generate significant excess value. Overall, the LSTM model demonstrates limited predictive power at short horizons, with results close to random chance. Performance improves modestly at monthly and quarterly horizons but remains weak relative to the variability of actual returns. At the annual horizon, the model achieves positive risk-adjusted returns and tracks broad market direction, though predictions remain conservative compared to realized values. However, when compared with a passive Buy & Hold strategy, the LSTM forecasts add no economic value. Across all horizons,

cumulative wealth outcomes are nearly identical to Buy & Hold, indicating that the model does not deliver meaningful financial gains despite modest improvements in directional accuracy.

4.4 Temporal Attention LSTM

This section extends the baseline analysis by incorporating attention mechanisms into LSTM models, with the aim of improving the modelling of temporal dependencies in NZX50 returns. The attention-based LSTM was tuned using grid search. The parameter grid spanned variations in lookback window (30, 60 days), hidden units (64~128 in the first layer, 32–64 in the second), attention dimension (16, 32), dropout rate (0.2, 0.3), and learning rate (1e-3, 5e-4). Early stopping with patience of 8 epochs was employed to prevent overfitting. The best configuration, identified from h=1 tuning, specified a 60-day lookback window, 64 units in both LSTM layers, attention dimension of 16, dropout 0.3, and learning rate of 0.001, with training conducted for up to 50 epochs and batch size of 32.

Attention weights are produced by a two-layer scoring network over the LSTM hidden sequence, followed by a softmax across the lookback dimension. The resulting weights form a convex combination of hidden states (context vector), allowing the model to emphasize the most informative lags for each forecast. Forecasting was performed for four horizons (h=1, 21, 63, 252 trading days), with returns aggregated cumulatively for horizons beyond one day.

Table 4.4.1 Temporal Attention-LSTM Forecasting Results on NZX50 (2023–2024 Test Set)

Horizon	MAE	DA(%)	Mean Ret.	Sharpe	Cum. Ret.	Max DD
h1	0.0063	49.90%	-0.0002	-0.0346	-0.1091	3.7388
h21	0.0491	50.10%	0.0012	0.049	0.5564	6.2441
h63	0.0709	54.20%	0.0029	0.0729	1.276	62.7616
h252	0.0588	45.20%	0.012	0.1938	2.9658	86.5336

Table 4.4.1 shows that predictive accuracy remains weak at the daily horizon, with DA 50% and negative Sharpe ratios. At medium horizons (h=63), the model achieves its best performance, with DA = 54.2% and modestly positive Sharpe ratios. At the annual horizon (h=252), the model generates a higher cumulative return (2.97), but with high volatility ($MaxDD$ > 80%).

Figure 4.4.1 Temporal Attention-LSTM Actual vs Predicted Returns

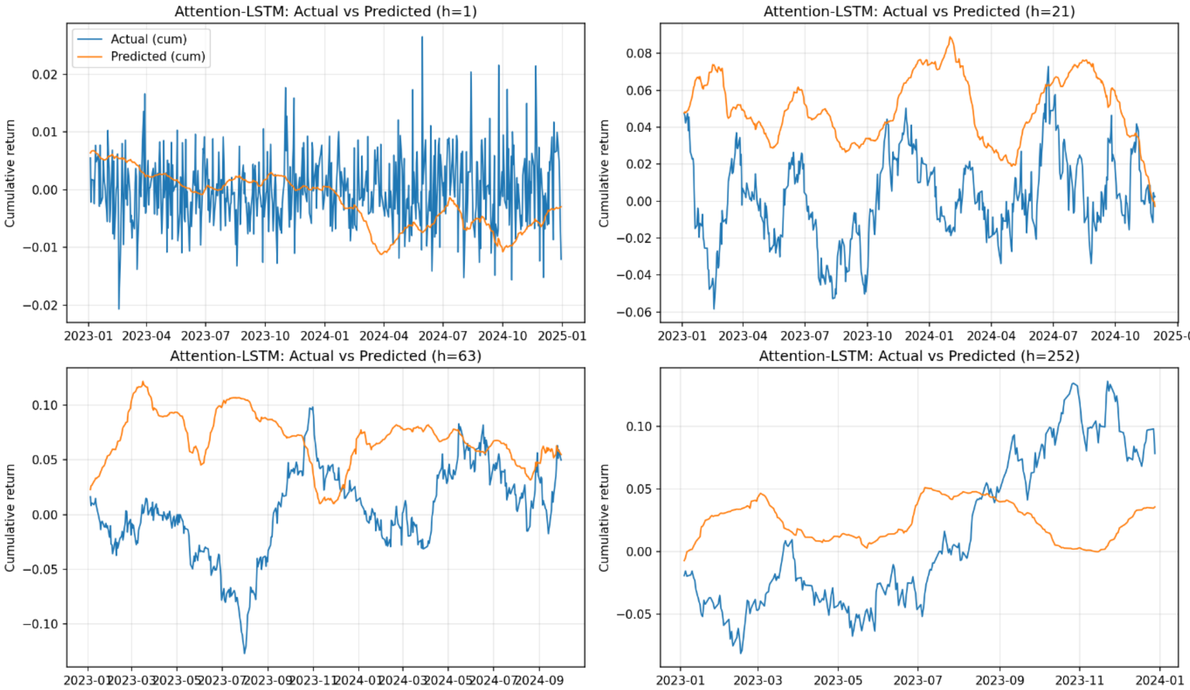


Figure 4.4.1 shows actual (blue) and predicted (orange) cumulative returns across horizons. At the daily horizon (h=1), the model produces smoother predictions than the actual returns. While this reduces noise, it also means the model fails to capture sharp day-to-day fluctuations in the NZX50. At the medium horizons (h=21, 63), the predictions move up and down in a way that is closer to the actual patterns. At h=252, the model predicts the overall upward trend but fails to capture the sharp rises and drops of the market.

Figure 4.4.2 Scatter of Predicted vs Realized Returns

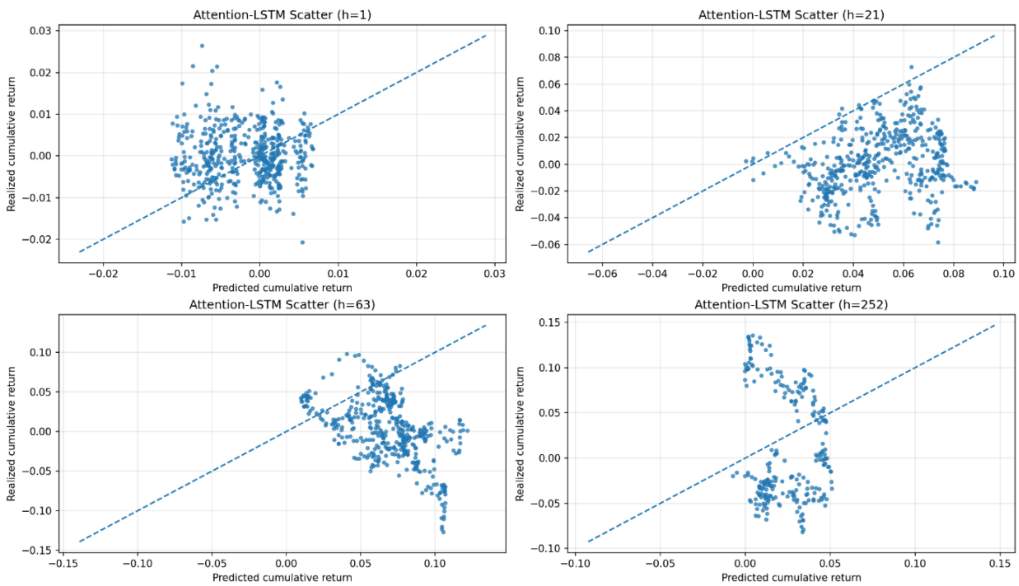


Figure 4.4.2 shows predicted versus actual cumulative returns across horizons. At h=1, points

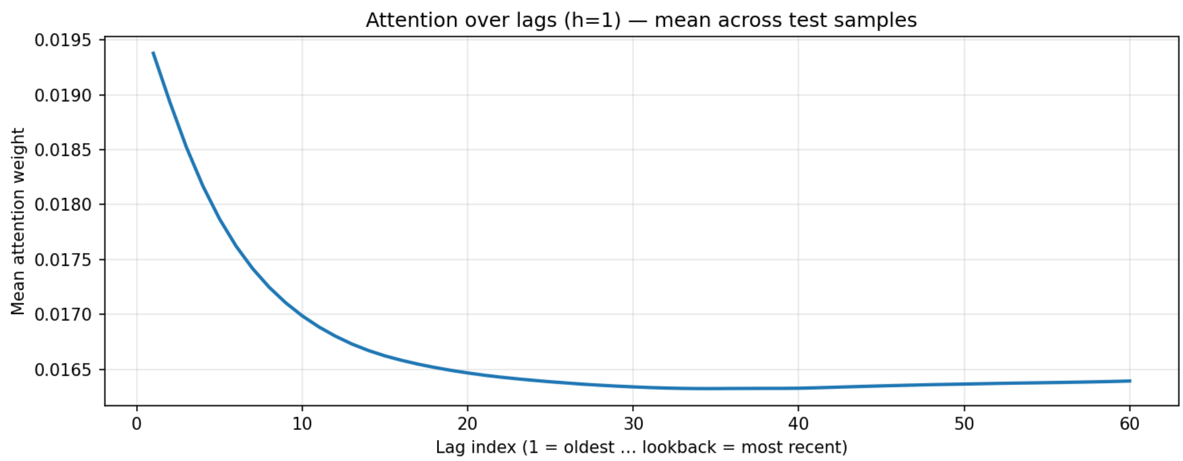
cluster near zero, indicating little day-to-day predictive power. At h=21 and h=63, scatter spreads more widely with some alignment to the 45° line, reflecting modest improvement in tracking medium-term cycles. At h=252, points still follow the upward slope but deviate from the diagonal, meaning the model predicts the overall trend but underestimates the magnitude of large swings.

Table 4.4.2 Temporal Attention-LSTM vs Buy & Hold Comparison

Horizon	Attention-LSTM ($1)	Buy & Hold ($1)
h1	$0.89	$1.06
h21	$1.52	$1.06
h63	$2.51	$1.06
h252	$12.13	$1.06

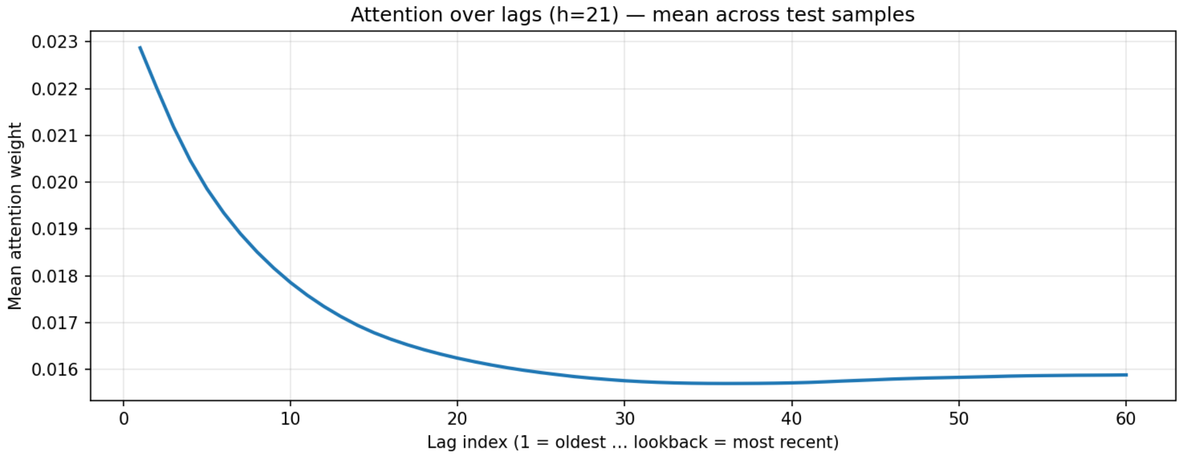
Table 4.4.2 shows that the Attention-LSTM underperforms Buy & Hold at the daily horizon (h=1) but significantly outperforms at h=21 and h=63. At h=252, the model produces much higher cumulative wealth ($12.13 vs $1.06), though with very high drawdowns, making long-horizon results unstable. Figure 4.4.3 to Figure 4.4.6 compares actual vs. predicted cumulative returns.

Figure 4.4.3 Mean Attention Weights Over Lags (h=1)



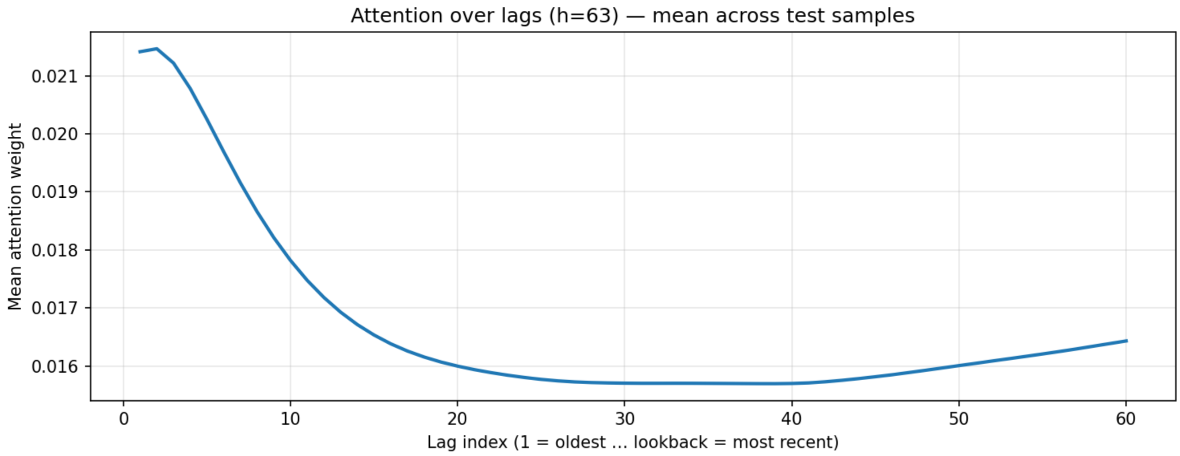
At daily horizon (h=1), the model gives slightly more weight to older lags than recent ones, meaning daily forecasts rely more on long-term memory than just yesterday's data.

Figure 4.4.4 Mean Attention Weights Over Lags (h=21)



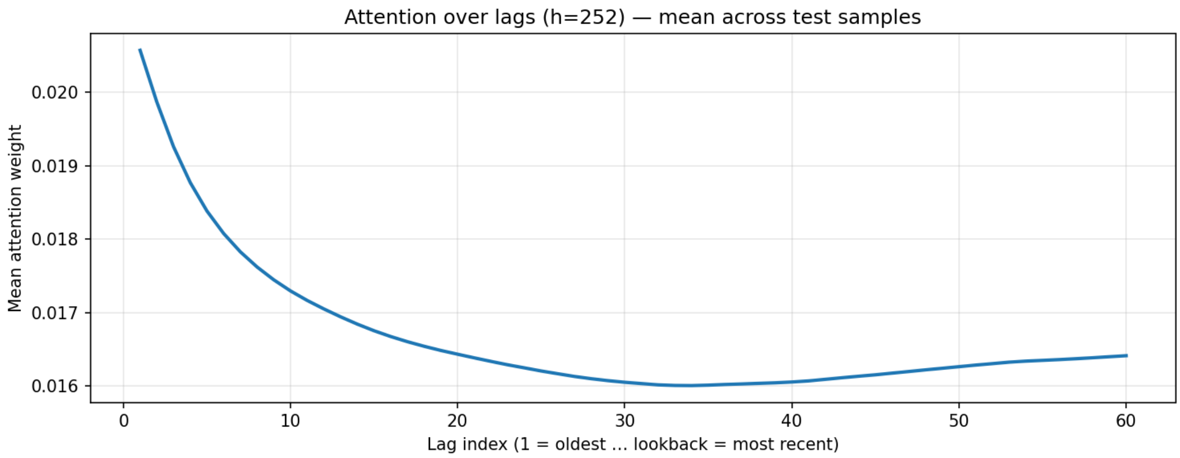
At the monthly horizon (h=21), older lags still matter more, but the weighting is more balanced, showing the model uses a wider history.

Figure 4.4.5 Mean Attention Weights Over Lags (h=63)



At quarterly horizon (h=63), the model shows a U-shape pattern, paying attention to both older and recent lags while downplaying the middle ones.

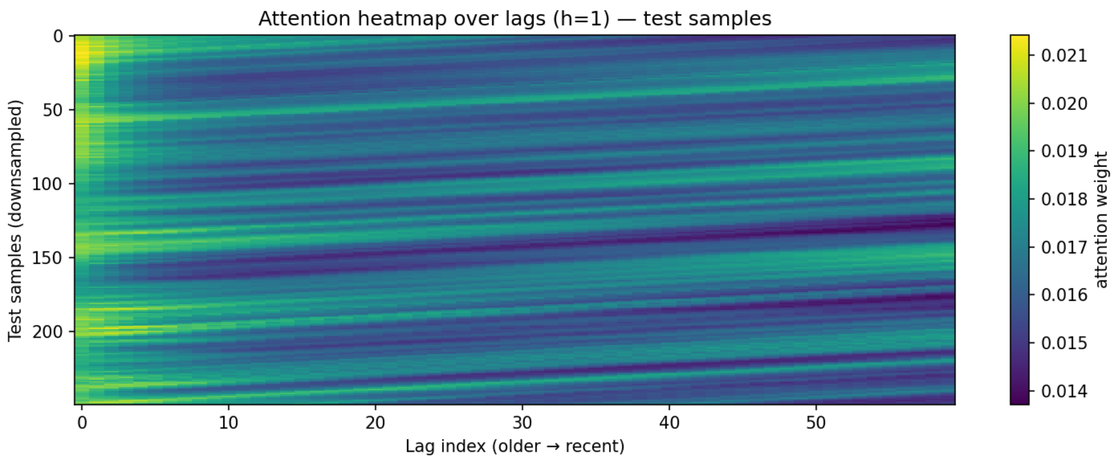
Figure 4.4.6 Mean Attention Weights Over Lags (h=252)



At the yearly horizon (h=252), weights are spread more evenly, with both older and recent lags contributing to predictions. Across all horizons, the Attention-LSTM does not exhibit a pure recency bias. Instead, it tends to assign higher weight to older lags. This indicates that the model values information from both distant and recent history, depending on the forecasting horizon, which aligns with the intuition that short-term and long-term market drivers operate differently.

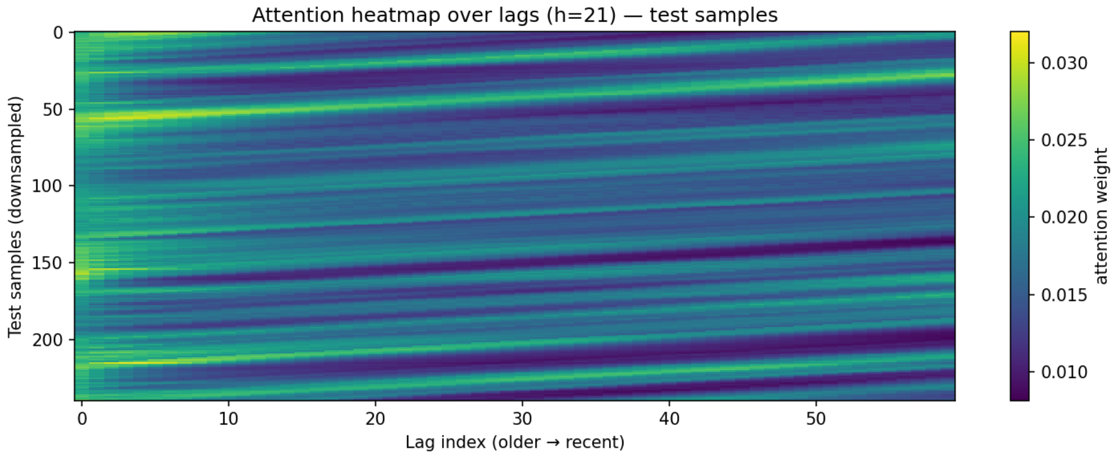
A key advantage of the Attention-LSTM is its interpretability. Heatmaps of attention weights (Figures 4.10a~d) visualizes attention weights across test samples (rows) and lags (columns).

Figure 4.4.7 Attention weight heatmap over lags (h=1)



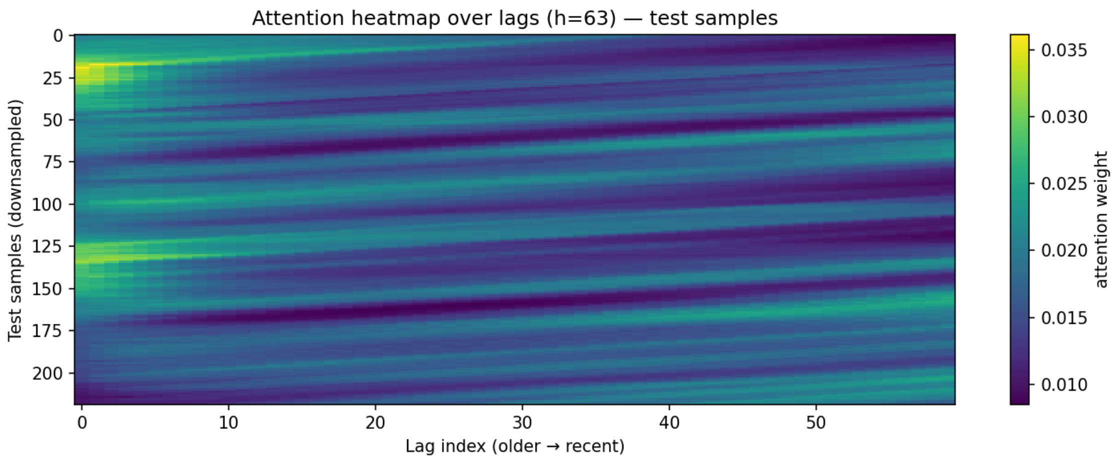
The heatmap for daily forecasts shows bright weights on older lags, while recent lags are darker. This means the model relies more on earlier parts of the lookback window than the most recent days.

Figure 4.4.8 Attention weight heatmap over lags (h=21)



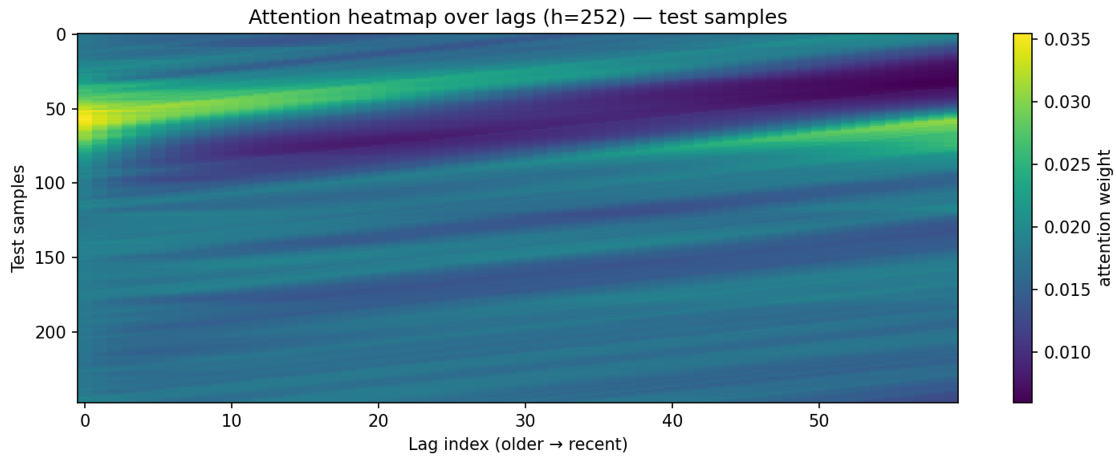
At the monthly horizon, the heatmap displays stronger diagonal streaks extending across the panel. Compared to *h*=1, the attention is less concentrated on just the oldest lags and instead spreads more evenly across history.

Figure 4.4.9 Attention weight heatmap over lags (*h*=63)



The quarterly horizon shows a darker overall map, with most weights concentrated in the blue-green range. This means attention is weaker and more diffuse across lags.

Figure 4.4.10 Attention weight heatmap over lags (h=252)



At the annual horizon, the heatmap reveals broader and more balanced streaks across the full lag range. Across horizons, the attention heatmaps show how the model's focus shifts with prediction length. Daily forecasts concentrate on older lags, while monthly horizons introduce clear cyclic patterns around one trading month. Quarterly forecasts are more diffuse, with relatively low attention intensity, and annual horizons spread attention broadly across the entire history. These patterns indicate that the model adapts its temporal focus to the forecast horizon, from narrow short-term to broad long-term context.

The Temporal Attention-LSTM demonstrates limited predictive value at short horizons. Performance improves at medium horizons (h=21, h=63), the model outperforms the Buy & Hold benchmark, both in terms of risk-adjusted measures and cumulative wealth growth. At the annual horizon, the model delivers very high cumulative returns, but also extreme volatility and drawdowns. Importantly, the attention mechanism highlights different patterns at different horizons. This makes the model more interpretable than a standard LSTM.

4.5 Cross-Sector Forecasting with Granger Causality and Attention-LSTM

This section investigates whether lagged signals from NZX sectoral indices and macroeconomic variables can improve forecasts of the NZX50 index. Two complementary approaches were employed: (i) Granger causality tests, and (ii) an extended Attention-LSTM model with temporal and feature attention.

Granger Causality Tests

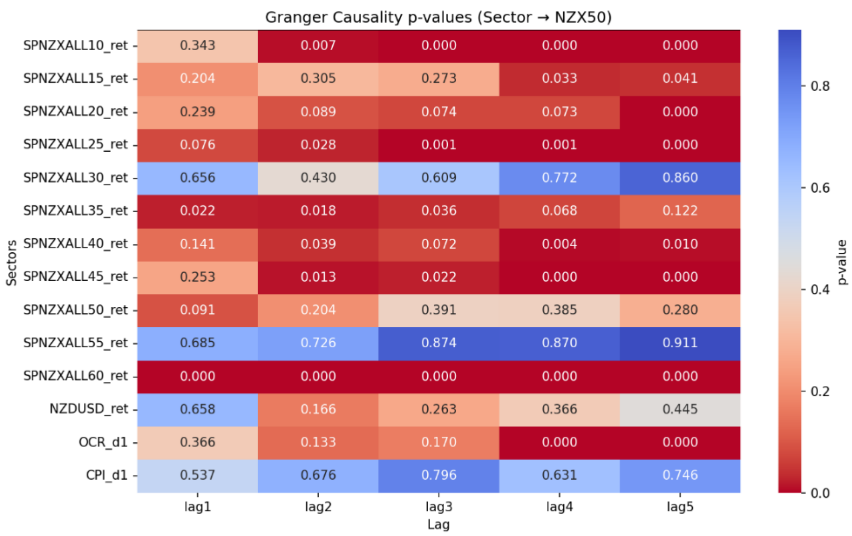
We first applied Granger causality tests to daily log returns of all NZX sector, together with macroeconomic variables. Tests were conducted for lags 1~5, capturing short- to medium-term spillovers. The null hypothesis of each test is that past values of a sector do not help forecast NZX50 returns. Rejection at the 5% significance level indicates that the sector provides incremental predictive power. Table 4.5.1 shows Granger causality test p-values for NZX sector returns, NZD/USD, OCR, and CPI (lags 1–5)

Table 4.5.1 Granger causality test p-values for NZX sector returns, NZD/USD, OCR, and CPI (lags 1–5)

	10_ret	15_ret	20_ret	25_ret	30_ret	35_ret	40_ret	45_ret	50_ret	55_ret	60_ret	NZDUSD_ret	OCR_d1	CPI_d1
lag1	0.3434	0.204	0.2392	0.0758	0.6556	**0.0218**	0.1407	0.2531	0.0915	0.6852	0	0.6582	0.3662	0.5369
lag2	**0.0073**	0.3055	0.0892	0.0276	0.4298	**0.0182**	**0.0392**	**0.013**	0.2038	0.7258	0	0.1657	0.1329	0.6763
lag3	0.0004	0.2731	0.0739	0.0012	0.6092	**0.0364**	0.072	0.0218	0.3908	0.8735	0	0.2631	0.1695	0.7957
lag4	0	0.0332	0.073	0.0009	0.7723	0.068	0.0036	0.0001	0.3849	0.87	0	0.3657	0	0.6307
lag5	0	0.0412	0	0.0003	0.8604	0.1217	0.0096	0.0002	0.2802	0.9109	0	0.4454	0	0.7462

The heatmap shown in Figure 4.5.1 highlights Energy, IT, Real Estate, and OCR showing strong predictive roles across multiple lags.

Figure 4.5.1 Heatmap of Granger causality p-values (lags 1–5)



The Granger Causality tests reveal that several sectors provide statistically significant predictive signals for NZX50 returns. Energy (10) emerges as the strongest driver, showing significance across lags 2~5 (p < 0.01), which indicates that shocks in the energy sector propagate quickly and persist over several trading days. Consumer Discretionary (25) becomes predictive at medium horizons (lags 3~5, p < 0.001), consistent with demand-driven cycles that influence market movements beyond the short term. Industrials (20) is significant only at lag 5, reflecting slower, capital-intensive dynamics that transmit with delay. Information Technology (45) contributes at lag 2~3 (p = 0.013; 0.0218) and strongly at lag 4~5 (p = 0.0001; 0.0002), suggesting innovation-related shocks diffuse into the broader market after short-to-medium delays. Real Estate (60) is highly significant across all lags, consistent with the interest-rate sensitivity and capital intensity of the sector. In contrast, Health Care (35) shows only a weak but still statistically significant effect at lag 2 (p = 0.0182), suggesting a modest short-horizon influence. Among macro variables, the Official Cash Rate (OCR) is significant at lags 4–5 (p = 0.0000), implying a delayed monetary-policy transmission into equities, while CPI and NZD/USD do not show predictive power in this window. Overall, these results reveal cross-sector causal chains—especially from Energy, Discretionary, Industrials, IT, and Real Estate, as well as delayed effects of monetary.

Table 4.5.2 highlights the strongest lag and minimum p-value for each driver, along with a short economic interpretation.

Table 4.5.2 Key Granger-Causal Sectors and macro variables for NZX50 (p-values < 0.05)

Sector / Variable (GICS)	**Strongest lag(s)**	**Minimum p-value**	**Interpretation (summary)**
Energy (10)	2–5	0.0000	Fast and persistent propagation from energy shocks.
Industrials (20)	5	0.0000	Delayed, cyclical effects from capital-intensive activity.
Consumer Discretionary (25)	3–5	0.0003	Medium-horizon demand-cycle signals.
Health Care (35)	2	0.0182	Short-horizon influence only.
Financials (40)	2, 4–5	0.0036	Recurring, delayed effects consistent with balance-sheet/credit channels.
Information Technology (45)	2–5	0.0001	Innovation shocks diffuse with short-to-medium delays.
Real Estate (60)	1–5	0.0000	Persistent, rate-sensitive leadership.
OCR (macro)	4–5	0.0000	Monetary policy transmits with delay.

Overall, these results confirm that NZX50 dynamics are shaped by lagged spillovers from key sectors and delayed monetary policy effects. This motivates the use of an Attention-LSTM model to capture and interpret these cross-sectoral dependencies more flexibly.

Attention-LSTM with temporal and feature attention

To complement the Granger analysis, the forecasting framework was extended with an Attention-LSTM model that incorporates both temporal and feature attention. Hyperparameters were tuned at the one-day horizon (*h*=1) using grid search, yielding the following optimal configuration: *lookback* = 60, *units1* = 128, *units2* = 64, *attention_dim* = 16, *dropout* = 0.2, learning *rate* = 0.001, *batch* = 32, *epochs* = 50, and *patience* = 8. This tuned model was then applied consistently across all horizons (*h*=1, *h*=21, *h*=63, *h*=252). Temporal attention layers assigned dynamic weights to historical lags within the lookback window, while feature attention layers determined the relative importance of each sector and macroeconomic feature group.

Table 4.5.3 Cross-sector attention LSTM Forecasting Results on NZX50 (2023–2024 Test Set)

Horizon	MAE	DA(%)	Mean Ret.	Sharpe	Cum. Ret.	Max DD
h1	0.0158	46.30%	0.0001	0.0193	0.061	3.0107
h21	0.0443	50.10%	0.0011	0.0481	0.5472	6.2441
h63	0.047	54.20%	0.0029	0.0729	1.276	62.7616
h252	0.0584	45.20%	0.0138	0.2257	3.4327	0

Table 4.5.3 reports model performance across horizons. Notably, horizon *h*=63 achieved the highest directional accuracy (54.2%) and strong cumulative performance, while horizon *h*=252 produced the largest raw cumulative return with increased volatility and unstable Sharpe ratios. Short-term horizons (*h*=1, *h*=21) showed weaker predictive performance, suggesting that cross-sector dependencies may take time to detect in aggregate.

Figure 4.5.2 Cross-sector attention LSTM Actual vs Predicted Returns

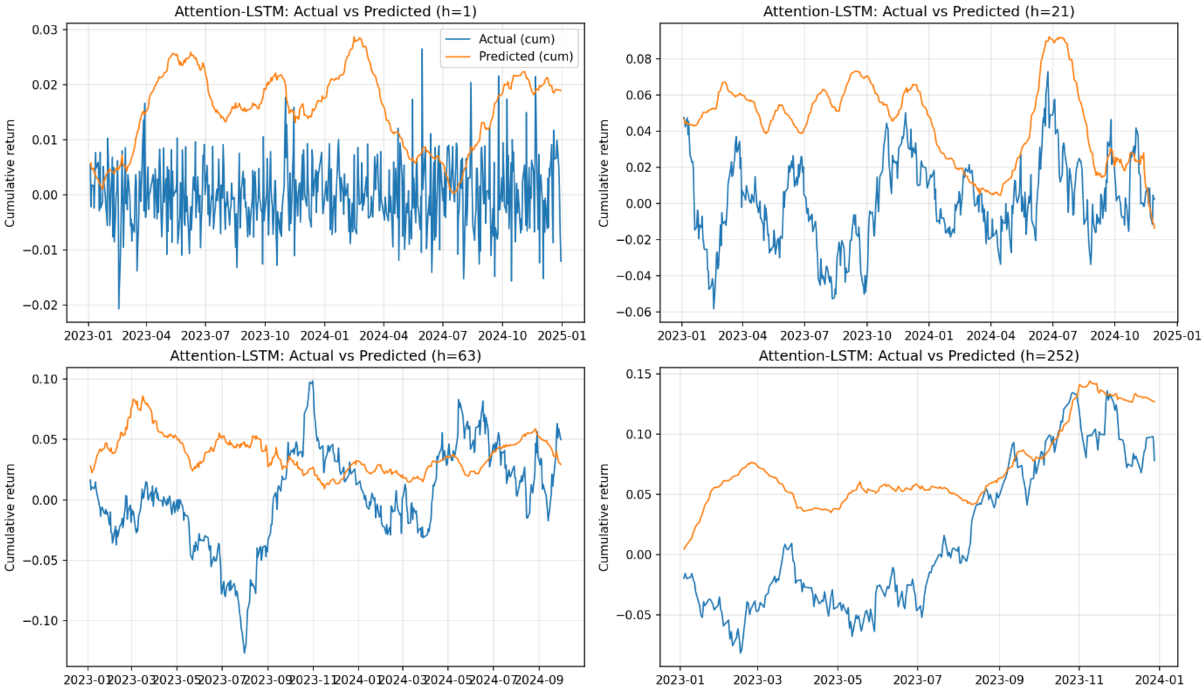


Figure 4.5.2 shows actual (blue) and predicted (orange) cumulative returns across horizons. For *h*=1, the model makes smooth predictions but misses the fast daily ups and downs. At *h*=21, it follows the medium-term trends better and matches the rises and falls more closely. For *h*=63, the model still tracks the general direction but misses some big drops and recoveries. At *h*=252, it captures the long-term upward trend but struggles to quickly detect around turning points.

Figure 4.5.3 Cross-sector attention LSTM Scatter of Predicted vs Realized Returns

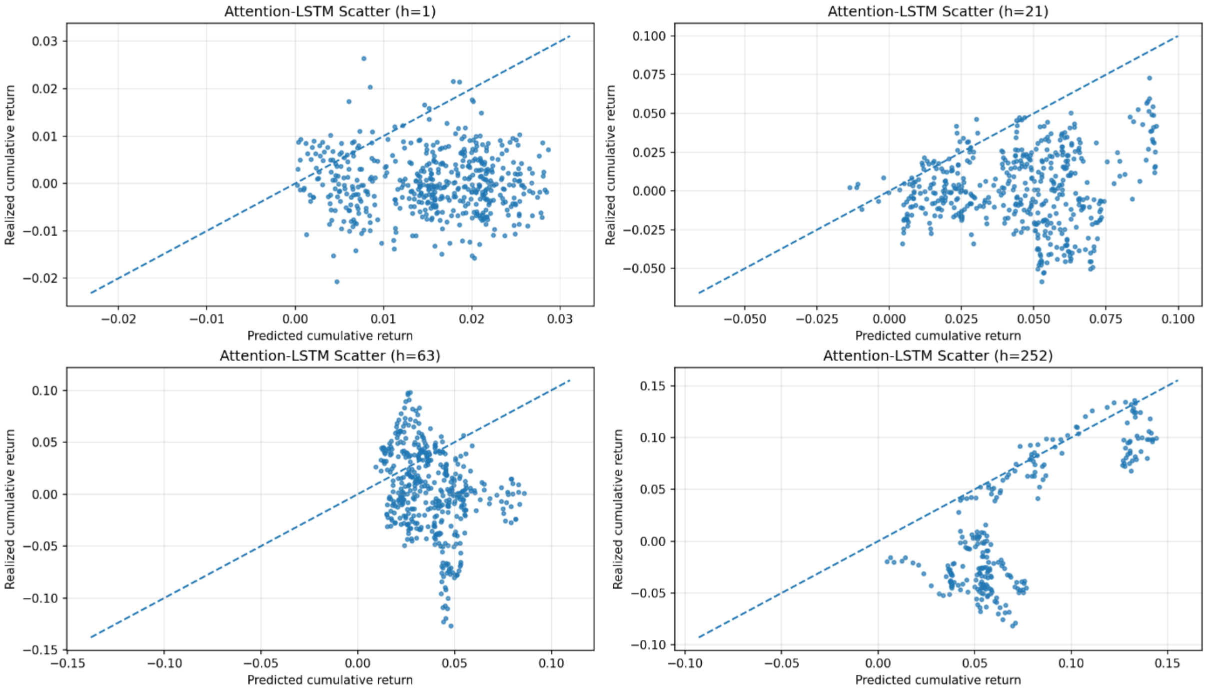


Figure 4.5.3 shows predicted versus actual cumulative returns across horizons. At *h*=1, predictions cluster between 0~0.03, while actual values are more spread (−0.02, +0.03). At *h*=21, the model over-predicts positive returns as many points lie below the line (*x*>0.05). At *h*=63, the model captured the overall direction while underestimating the scale of longer swings. At *h*=252, predictions followed realized returns more closely, although downturns were still under-predicted. Overall, the scatter plots indicate that the model performs weakly at short horizons but improves over longer horizons, where it captures broad market trends more reliably.

Table 4.5.4 Cross-sector attention LSTM vs Buy & Hold Comparison

Horizon	LSTM ($1)	Buy & Hold ($1)
h1	$1.05	$1.06
h21	$1.51	$1.06
h63	$2.51	$1.06
h252	$19.38	$1.06

Table 4.5.4 shows that $1 invested with the Cross-Sector Attention-LSTM could grow much faster than buy-and-hold. Buy-and-hold stayed around $1.06, but the model reached $1.51 after a month, $2.51 after three months, and as high as $19.38 after a year. This means the model picked up useful sector signals that allowed it to compound returns. However, the

numbers are likely too high for real trading, since the test does not include real-world issues like trading costs and market liquidity.

Figure 4.5.4 Feature attention (h=1)

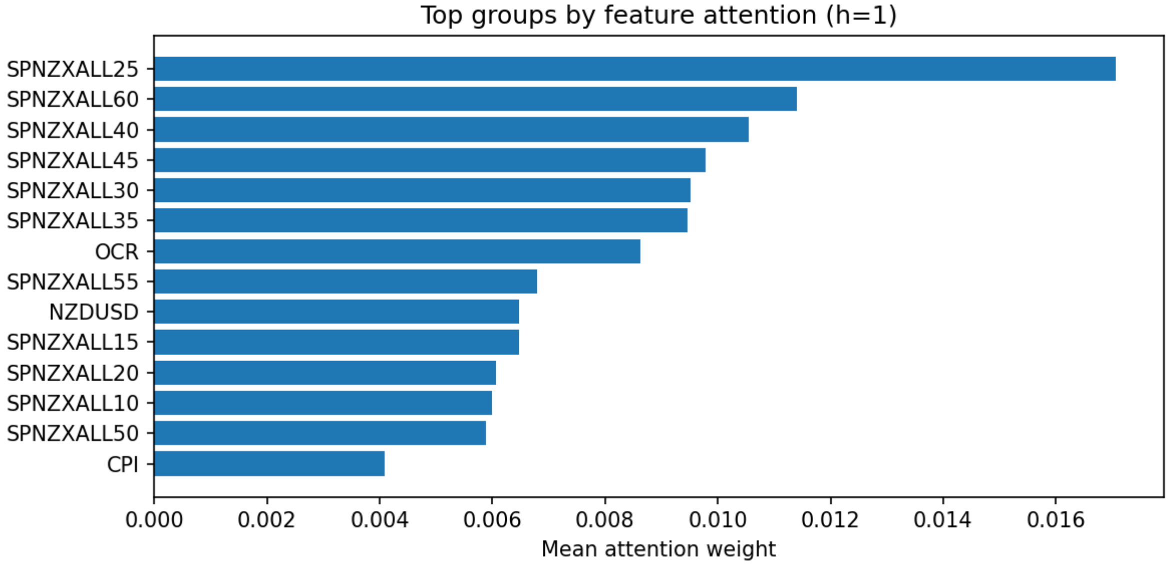


Figure 4.5.5 Feature attention (h=21)

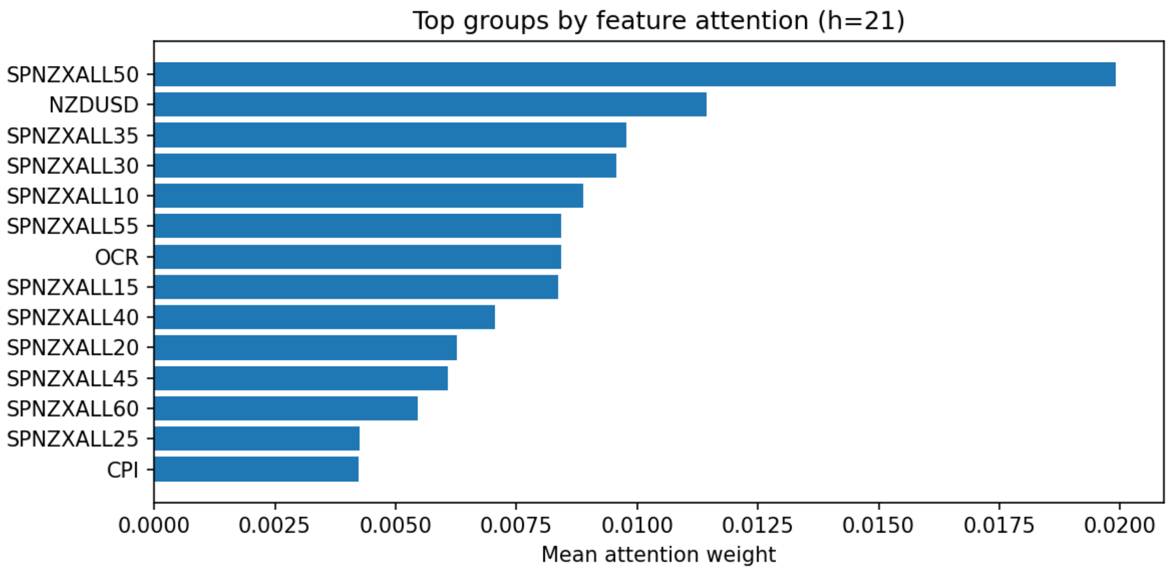


Figure 4.5.6 Feature attention (h=63)

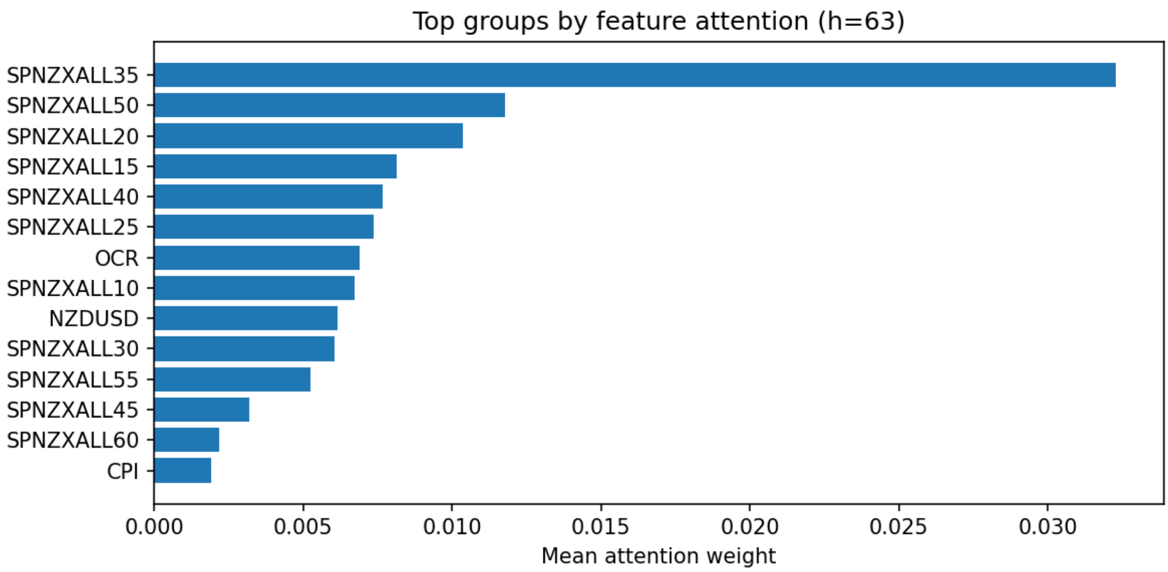
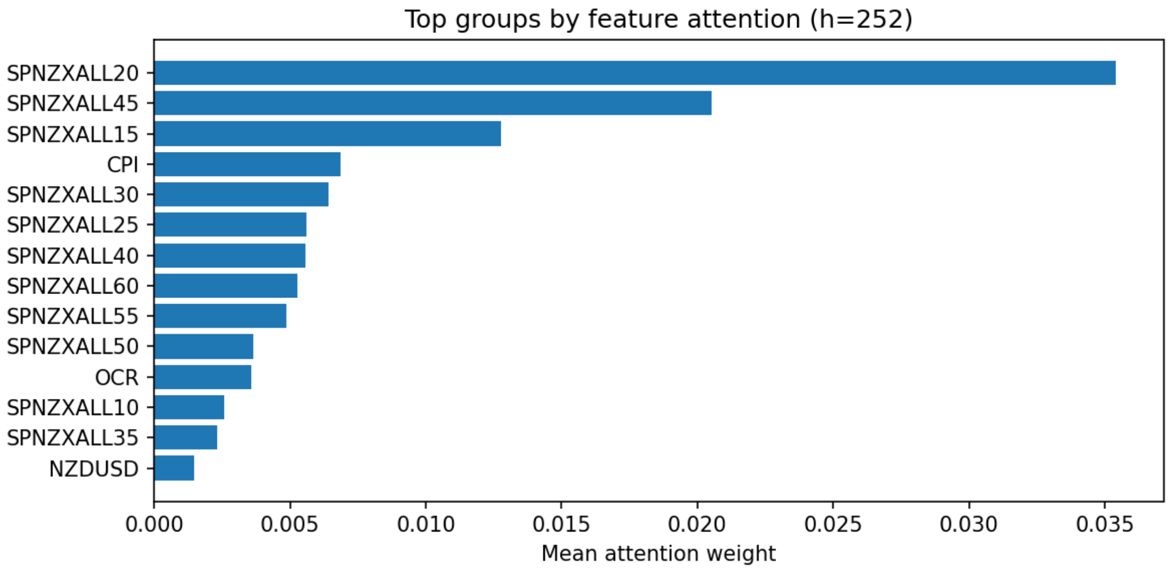
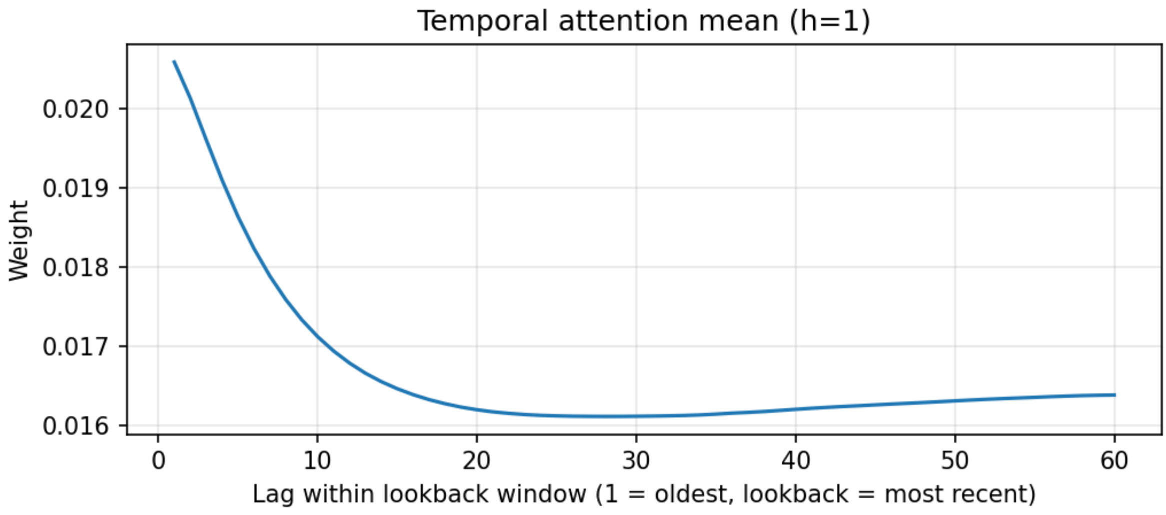


Figure 4.5.7 Feature attention (h=252)



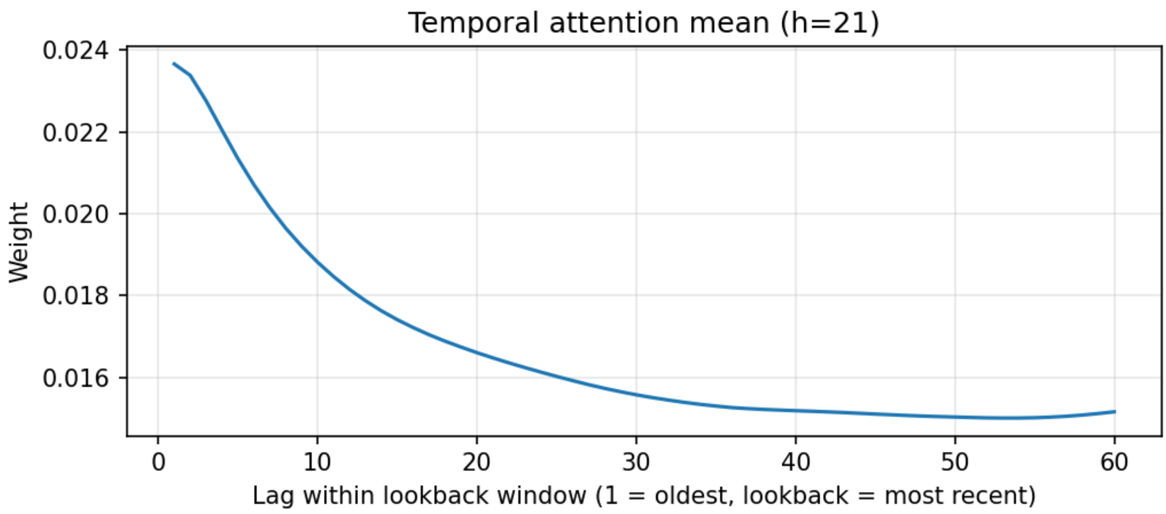
Feature attention plots (Figures 4.5.4 to Figure 4.5.7) highlight which sectors and macro variables the model prioritizes at each horizon. At h=1, the model paid most attention to sector indices like SPNZXALL25 (Consumer Discretionary), SPNZXALL60 (Real Estate), and SPNZXALL40 (Financials), with OCR also showing relevance, while CPI had the least weight. By h=21, the focus shifted, with SPNZXALL50 (Telecommunication Services), NZD/USD and SPNZXALL35 (Health Care) becoming dominant, suggesting short-term exchange rate effects influenced monthly returns. At h=63, SPNZXALL35 (Health Care) clearly stood out as the most important sector, followed by SPNZXALL50 (Telecommunication Services) and SPNZXALL20 (Industrials), showing that mid-term signals concentrated in specific sectors. By h=252, attention narrowed further, with SPNZXALL20 (Industrials) and SPNZXALL45 (Information Technology) strongly dominating, while macro factors like CPI gained more weight, and NZD/USD lost influence. Overall, the feature attention shows a progression that many sectors matter for short horizons. But as the horizon lengthens, the model concentrates on fewer sectors. While macro factors like CPI become more visible for long-term predictions. Figure 4.14 (a~d) compares actual vs. predicted cumulative returns.

Figure 4.5.8 Temporal Mean Attention Weights Over Lags (h=1)



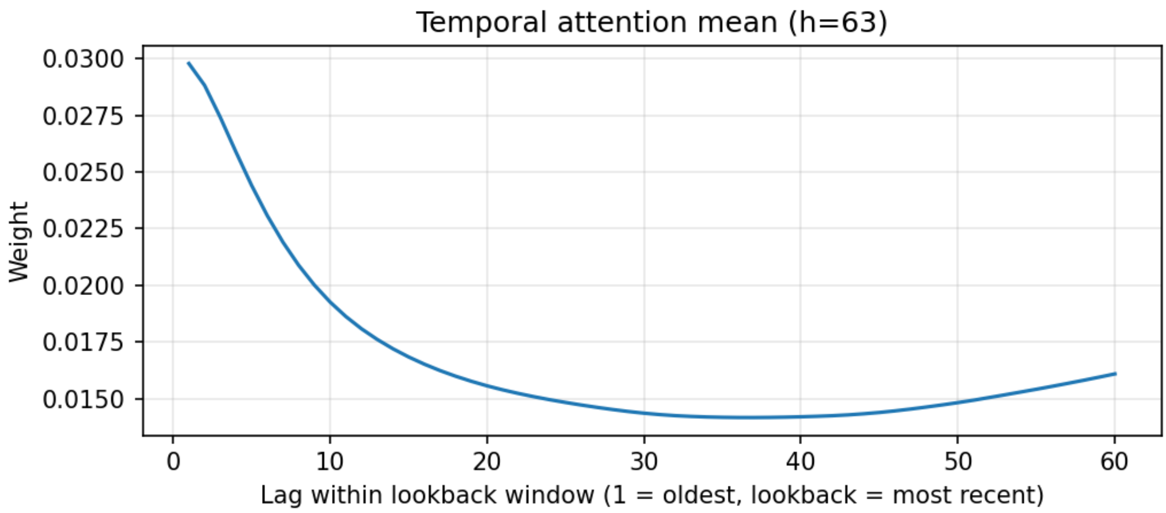
At daily horizon, the model places its highest weight on the farthest lags, with sharp declining and seeing small rises on the recent days.

Figure 4.5.9 Temporal Mean Attention Weights Over Lags (*h*=21)



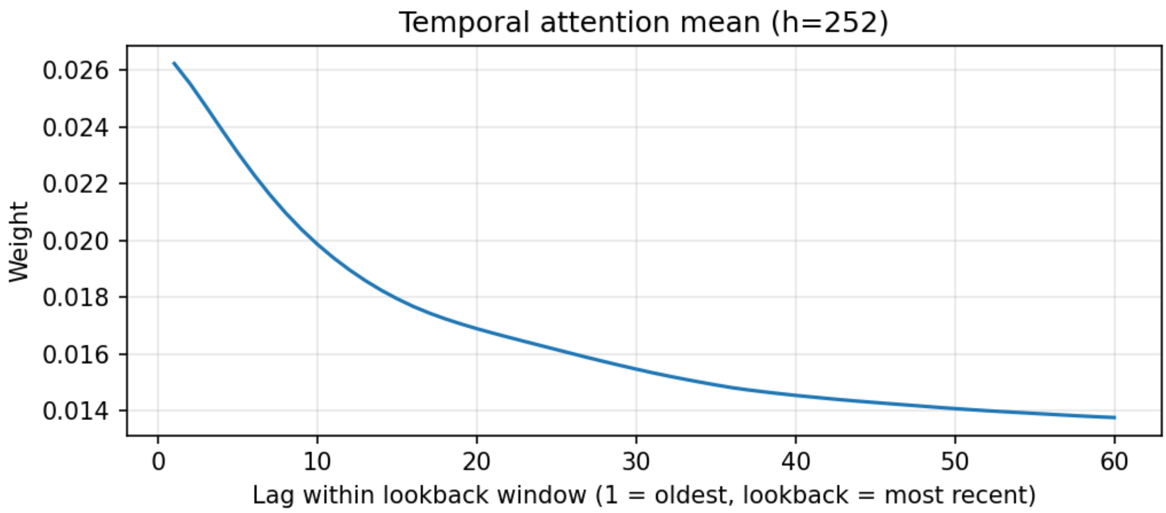
At the monthly horizon, a similar decay is observed, with strong emphasis on the first few lags (oldest), while weights flatten for intermediate days.

Figure 4.5.10 Temporal Mean Attention Weights Over Lags (h=63)



At quarterly horizon, the pattern is similar, but we see a rise in weights again near the most recent days, showing that both long memory and some recent signals play a role.

Figure 4.5.11 Temporal Mean Attention Weights Over Lags (*h*=252)



At the yearly horizon, the pattern is strongest here: attention concentrates almost entirely on the earliest lags, while weights for more recent days flatten close to zero. Overall, the results indicate that the model interprets historical context from months ago as more predictive than immediate past returns. This suggests that in the NZX, useful signals take longer to affect the index, so the model learns to rely more on older patterns rather than only the most recent movements.

Cross-Asset Causal Chains

The results from Granger Causality and Attention-LSTM point to a three-step chain of influence on the NZX50. Granger tests show that Energy (10) and Real Estate (60) act as early drivers, with strong predictive signals at short lags. Consumer Discretionary (25) becomes important at medium horizons, linking demand-driven effects to the index. At longer horizons, Industrials (20) and Information Technology (45) gain influence, reflecting slower structural dynamics. Attention weights support this pattern, with Real Estate and Discretionary dominant in the short run, and Industrials and IT taking over in the long run. Overall, shocks appear to start in Energy and Real Estate, flow into Discretionary spending, and then spread through Industrials and IT before reaching the NZX50.

Chapter 5
Analysis and Discussions

In this chapter, experimental results are analyzed and compared. Comparisons of the results under various conditions will be explored. In the end, in this chapter, we also discuss the limitations of this project.

5.1 Comparative Analysis of Models

The results across all horizons highlight clear differences in performance between traditional statistical methods and deep learning approaches. At short horizons (h=1), none of the models produced consistent predictive power, confirming the difficulty of exploiting very short-term signals in an efficient and noisy market. ARIMA was slightly better at minimizing errors, but LSTM-based models aligned marginally better with profitability.

At the 21-day and 63-day horizons, all models achieved nearly identical directional accuracy, mean return, Sharpe ratio, and cumulative return. The only difference was in MAE: ARIMA and the baseline LSTM achieved lower errors while the attention-based models showed higher values. This suggests that attention mechanisms did not yield an advantage at the monthly and quarterly horizon, as the models captured the same directional patterns and profitability, but with increased prediction error.

At the 252-day horizon, all models achieved substantial long-term gains compared to the buy-and-hold benchmark. Cumulative returns were clustered around 3.4 for ARIMA, the baseline LSTM, and the cross-sector attention LSTM, while the temporal-only attention model underperformed slightly at ~3.0. Directional accuracy remained modest at 45.2%, yet profitability was driven by the persistence of positive yearly returns rather than short-term prediction accuracy. These results suggest that all variants delivered similar long-term outcomes, with only minor variation in errors and cumulative returns. To illustrate these trade-offs clearly, Table 5.1.1 presents a consolidated comparison of all models across horizons.

Table 5.1.0.1 Consolidated Performance of Models Across Horizons

Horizon	**Model**	**MAE**	**DA (%)**	**Mean Return**	**Sharpe**	**Cumulative Return**
h=1	ARIMA	0.0049	49.1%	–0.0004	–0.0616	–0.1958
	LSTM (Tech Indicators)	0.0051	46.3%	0.0001	0.0193	0.0610
	Temporal Attention-LSTM	0.0063	49.9%	–0.0002	–0.0346	–0.1091
	Cross-sector attention LSTM	0.0158	46.3%	0.0001	0.0193	0.0610
h=21	ARIMA	0.0196	50.3%	0.0011	0.0480	0.5456
	LSTM (Tech Indicators)	0.0199	50.3%	0.0011	0.0480	0.5456
	Temporal Attention-LSTM	0.0491	50.1%	0.0012	0.0490	0.5564

	Cross-sector attention LSTM	0.0443	50.1%	0.0011	0.0481	0.5472
h=63	ARIMA	0.0331	54.2%	0.0029	0.0729	1.2760
	LSTM (Tech Indicators)	0.0330	54.2%	0.0029	0.0729	1.2760
	Temporal Attention-LSTM	0.0709	54.2%	0.0029	0.0729	1.2760
	Cross-sector attention LSTM	0.0470	54.2%	0.0029	0.0729	1.2760
h=252	ARIMA	0.0701	45.2%	0.0138	0.2257	3.4327
	LSTM (Tech Indicators)	0.0601	45.2%	0.0138	0.2258	3.4332
	Temporal Attention-LSTM	0.0588	45.2%	0.0120	0.1938	2.9658
	Cross-sector attention LSTM	0.0584	45.2%	0.0138	0.2257	3.4327

5.2 Interpretation of Attention and Granger Results

The Granger causality analysis revealed that several sectors serve as leading indicators of NZX50 performance. Energy, Real Estate, Consumer Discretionary, Industrials, and Information Technology all showed statistically significant predictive effects across different lag structures. For instance, Energy shocks transmitted rapidly, while Industrials influenced returns only with longer delays. Real Estate stood out as consistently predictive across all lags, consistent with its strong sensitivity to interest rates and capital cycles.

The feature attention results from the LSTM models closely aligned with these findings. At short horizons, the model concentrated on Real Estate, Consumer Discretionary, and Financials, showing that immediate signals often come from interest rate–sensitive and demand-driven sectors. At medium horizons, Health Care and Telecommunications gained prominence, while macro variables such as NZD/USD also became relevant, reflecting cyclical influences on monthly returns. At long horizons, Industrials and Information Technology became the most influential, while CPI gained weight, suggesting that slow-moving macroeconomic trends shape annual performance.

Temporal attention offered additional insight by emphasizing older lags rather than the most recent days. This indicates that information diffusion in the NZX is slower compared to larger and more liquid markets. Combining these strands, the evidence points to cross-sector causal chains where shocks in leading sectors—such as Energy and Real Estate—propagate into discretionary spending and eventually shape Industrials and IT performance before feeding into the NZX50.

5.3 Practical and Theoretical Implications

The analysis of these results carries important implications for forecasting in the NZX. First, short-term horizons remain highly unpredictable due to efficiency and noise. Even advanced models cannot consistently produce meaningful accuracy at daily scales. Second, medium horizons show modest value, with LSTM-based methods capturing some cyclicality but still converging closely with ARIMA. Third, the strongest results appear at long horizons, where temporal attention sustains predictive value by focusing on relevant lag structures. This suggests that investment strategies in smaller markets like NZX may benefit more from medium- and long-term perspectives rather than attempting to exploit short-term fluctuations. Finally, the collapse of the cross-sector model highlights that interdependencies must be modeled selectively. While sectoral relationships are important, naive pooling can overwhelm the signal with noise, especially in smaller markets with limited data depth.

5.4 Limitations

The findings must be interpreted in light of several limitations. The NZX dataset is relatively small compared to global markets, which constrains generalizability and limits the robustness of deep learning models. While attention mechanisms improve interpretability by highlighting influential lags and features, the models remain partly opaque, which may challenge adoption in risk-sensitive contexts. Furthermore, the analysis incorporated key domestic macroeconomic indicators such as the NZD/USD exchange rate, CPI, and OCR. However, broader global drivers — including commodity prices and international interest rate movements — were not included, even though they are also likely to influence NZX sector dynamics. Finally, model performance was sensitive to hyperparameter tuning and required considerable computational resources, raising practical barriers to real-world implementation.

Chapter 6 Conclusion and Future Work

In this chapter, we will provide a brief conclusion of the study and outline potential directions for future research.

6.1 Conclusion

In summary, this study has compared traditional and deep learning models, interpreted sectoral and temporal attention patterns, and discussed both practical implications and methodological limitations. The evidence shows that while short-horizon predictions remain difficult, attention-based LSTMs significantly improve long-horizon forecasting and provide interpretable insights into sectoral linkages. Granger causality and feature attention both highlight Energy, Real Estate, Industrials, and IT as key drivers, while temporal attention suggests slower information transmission in the NZX. Despite limitations, the hybrid approach of combining statistical tests with interpretable deep learning offers a powerful framework for understanding cross-sector dynamics and forecasting small equity markets more effectively.

6.2 Future Work

Future research could expand the dataset by incorporating global drivers such as commodity prices and international interest rates, which may strongly influence NZX movements. Additional experiments with Transformer-based architecture could improve long-horizon predictions. Further work is also needed to enhance interpretability, making models more practical for real-world financial forecasting.

References

Arumugam, V., & Natarajan, V. (2023). Time series modeling and forecasting using autoregressive integrated moving average and seasonal autoregressive integrated moving average models. *Instrumentation Mesure Métrologie*, *22*(4), 161-168. https://doi.org/10.18280/i2m.220404

Asadi, S., Tavakoli, A., & Hejazi, S. R. (2012). A new hybrid for improvement of auto-regressive integrated moving average models applying particle swarm optimization. *Expert Systems with Applications*, *39*(5), 5332-5337. https://doi.org/https://doi.org/10.1016/j.eswa.2011.11.002

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Bao, W., Yue, J., & Rao, Y. (2017). A deep learning framework for financial time series using stacked autoencoders and long-short term memory. *PLOS ONE*, *12*(7), e0180944.

Baruník, J., & Křehlík, T. (2018). Measuring the frequency dynamics of financial connectedness and systemic risk. *Journal of Financial Econometrics*, *16*(2), 271-296. https://doi.org/10.1093/jjfinec/nby001

Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, *5*(2), 157-166.

Billio, M., Getmansky, M., Lo, A. W., & Pelizzon, L. (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics*, *104*(3), 535-559. https://doi.org/https://doi.org/10.1016/j.jfineco.2011.12.010

Boonpong, N., & Phuensane, P. (2023). Predicting returns in emerging markets: A CNN-LSTM approach to sectoral network analysis. *Fuzzy Economic Review*, *28*(1).

Borovykh, A., Bohte, S., & Oosterlee, C. W. (2017). Conditional time series forecasting with convolutional neural networks. *arXiv preprint arXiv:1703.04691*.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control*. John Wiley & Sons.

Cheng, X., Zhu, W., Yan, W. (2025) Centrality-aware collaborative network embedding for overlapping community detection. IEEE Transactions on Network Science Methodologies.

Choi, K.-H., & Yoon, S.-M. (2023). Risk connectedness among international stock markets: Fresh findings from a network approach. *Systems*, *11*(4), 207.

D, D. T., Bung, D. P., & M, D. J. (2024). Predicting exchange rate between the us dollar (USD) and Indian rupee (INR): An empirical analysis using Sarima model. *International Journal of Research in Finance and Management*, *7*(1), 56-63. https://doi.org/10.33545/26175754.2024.v7.i1a.283

Dassanayake, W., Ardekani, I., Gamage, N., Jayawardena, C., & Sharifzadeh, H. (2021). Effectiveness of stock index forecasting using ARIMA model: Evidence from New Zealand. International Conference on Advancements in Computing (ICAC),

Diebold, F., & Yilmaz, K. (2011). On the Network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics*, *182*.

https://doi.org/10.2139/ssrn.1937613

Diebold, F. X., & Yılmaz, K. (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics*, *182*(1), 119-134. https://doi.org/https://doi.org/10.1016/j.jeconom.2014.04.012

Fama, E. F., & Schwert, G. W. (1977). Asset returns and inflation. *Journal of Financial Economics*, *5*(2), 115-146. https://doi.org/https://doi.org/10.1016/0304-405X(77)90014-9

Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, *270*(2), 654-669. https://doi.org/https://doi.org/10.1016/j.ejor.2017.11.054

Galeshchuk, S., & Mukherjee, S. (2017). Deep networks for predicting direction of change in foreign exchange rates. *Intelligent Systems in Accounting, Finance and Management*, *24*. https://doi.org/10.1002/isaf.1404

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, *37*(3), 424-438. https://doi.org/10.2307/1912791

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

Hou, J. (2023). Stock index prediction based on ARIMA model. *Advances in Economics, Management and Political Sciences*, *14*(1), 319-328. https://doi.org/10.54254/2754-1169/14/20230846

Huan, Y., Yan, W. (2025) Semaphore recognition using deep learning. Electronics 14 (2), 286

Kaul, G. (1987). Stock returns and inflation: The role of the monetary sector. *Journal of Financial Economics*, *18*(2), 253-276. https://doi.org/https://doi.org/10.1016/0304-405X(87)90041-9

Kavitha, K., Sivapuram, S., Tirumalasetty, S. V., Kummari, V., Varada, S. N., & Ravuri, V. (2025). Stock prices using arima and sarima with macroeconomic indicators predicting. International Conference on Inventive Research in Computing Applications (ICIRCA),

Khashei, M., Bijari, M., & Hejazi, S. R. (2012). Combining seasonal ARIMA models with computational intelligence techniques for time series forecasting. *Soft Computing*, *16*(6), 1091-1105. https://doi.org/10.1007/s00500-012-0805-9

Khashei, M., Bijari, M., & Raissi Ardali, G. A. (2009). Improvement of auto-regressive integrated moving average models using fuzzy logic and artificial neural networks (ANNs). *Neurocomputing*, *72*(4), 956-967.

Kroujiline, D., Gusev, M., Ushanov, D., Sharov, S. V., & Govorkov, B. (2016). Forecasting stock market returns over multiple time horizons. *Quantitative Finance*, *16*(11), 1695-1712. https://doi.org/10.1080/14697688.2016.1176241

Kumar, M., & Thenmozhi, M. (2006). Forecasting stock index movement: A comparison of support vector machines and random forest. Indian Institute of Capital Markets 9th Capital Markets Conference

Lai, G., Chang, W.-C., Yang, Y., & Liu, H. (2018). Modeling long-and short-term temporal patterns with deep neural networks. The ACM SIGIR Conference on Research & Development in Information Retrieval,

Latif, S., Aslam, F., Ferreira, P., & Iqbal, S. (2025). Integrating macroeconomic and technical indicators into forecasting the stock market: A data-driven approach. *Economies*, *13*(1).

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436-444.

Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, *37*(4), 1748-1764.

Liu, B., & Lai, M. (2025). Advanced machine learning for financial markets: A PCA-GRU-LSTM approach. *Journal of the Knowledge Economy*, *16*(1), 3140-3174.

Liu, J. (2024). Navigating the financial landscape: The power and limitations of the ARIMA model. *Highlights in Science, Engineering and Technology*, *88*, 747-752.

Ma, X., Yan, W. (2021) Banknote serial number recognition using deep learning. Springer Multimedia Tools and Applications.

Nelson, D. M. Q., Pereira, A. C. M., & Oliveira, R. A. d. (2017). Stock market's price movement prediction with LSTM neural networks. International Joint Conference on Neural Networks (IJCNN),

Nicholson, W. B., Matteson, D. S., & Bien, J. (2017). VARX-L: Structured regularization for large vector autoregressions with exogenous variables. *International Journal of Forecasting*, *33*(3), 627-651.

Olah, C. (2015). *Understanding LSTM networks*. https://colah.github.io/posts/2015-08-Understanding-LSTMs/

Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications*, *42*(1), 259-268.

Peng, D., Yan, W. (2025) Test-time training with adaptive memory for traffic accident severity prediction. Computers, MDPI.

Petrică, A.-C., Stancu, S., & Tindeche, A. (2016). Limitation of ARIMA models in financial and monetary economics. *Theoretical & Applied Economics*, *23*(4).

Phylaktis, K., & Ravazzolo, F. (2005). Stock prices and exchange rate dynamics. *Journal of International Money and Finance*, *24*(7), 1031-1053.

Polat, O. (2024). Interlinkages across US sectoral returns: Time-varying interconnectedness and hedging effectiveness. *Financial Innovation*, *10*(1), 51.

Qin, Y., Song, D., Chen, H., Cheng, W., Jiang, G., & Cottrell, G. (2017). A dual-stage attention-based recurrent neural network for time series prediction. *arXiv preprint arXiv:1704.02971*.

RBNZ. (2024). *Macroeconomic indicators: Official cash rate, consumer price index, and exchange rate data*. https://www.rbnz.govt.nz/statistics

Rotela Junior, P., Salomon, F. L. R., & de Oliveira Pamplona, E. (2014). ARIMA: An applied time series forecasting model for the Bovespa stock index. *Applied Mathematics*, *05*(21), 3383-3391. https://doi.org/10.4236/am.2014.521315

S&P. (2024). *S&P/NZX sector indices data.* S&P Dow Jones Indices LLC. https://www.spglobal.com/spdji

Sezer, O. B., Gudelek, M. U., & Ozbayoglu, A. M. (2020). Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. *Applied Soft Computing*, *90*, 106181.

Siami-Namini. (2018). Forecasting economics and financial time series: ARIMA vs. LSTM. *ArXiv*. https://doi.org/abs/1803.06386

Su, X., & Liu, Z. (2021). Sector volatility spillover and economic policy uncertainty: Evidence from China's stock market. *Mathematics*, *9*(12).

Sunki, A., SatyaKumar, C., Surya Narayana, G., Koppera, V., & Hakeem, M. (2024). Time series forecasting of stock market using ARIMA, LSTM and FB prophet [10.1051/matecconf/202439201163]. *MATEC Web Conf.*, *392*.

Taieb, S. B., & Atiya, A. F. (2016). A bias and variance analysis for multistep-ahead time series forecasting. *IEEE Transactions on Neural Networks and Learning Systems*, *27*(1), 62-76.

Teles, G., Rodrigues, J. J., Rabêlo, R. A., & Kozlov, S. A. (2021). Comparative study of support vector machines and random forests machine learning algorithms on credit operation. *Software: Practice and Experience*, *51*(12), 2492-2500.

Tewari, A. (2020). *Forecasting NIFTY 50 benchmark index using seasonal ARIMA time series models*. https://doi.org/10.13140/RG.2.2.10332.95364

Thomas, A. (2017). An introduction to neural networks for beginners. *adventuresinmachinelearning. com*.

Tsay, R. (2010). Financial time series and their characteristics. *Analysis of Financial Time Series*, *44*, 47-71.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *30*.

Wang, S., & Mayes, D. G. (2012). Monetary policy announcements and stock reactions: An international comparison. *The North American Journal of Economics and Finance*, *23*(2), 145-164. https://doi.org/https://doi.org/10.1016/j.najef.2012.02.002

Yan, W. (2019) Introduction to Intelligent Surveillance: Surveillance Data Capture, Transmission, and Analytics. Springer.

Yan, W. (2023) Computational Methods for Deep Learning Theory, Algorithms, and Implementations. Springer.

Yan, W. (2026) Robotic Vision: From Deep Learning to Autonomous Systems. Springer.

Yang, G. (2025) ChatPPG: Multi-Modal Alignment of Large Language Models for Time-Series Forecasting in Table Tennis. Master's Thesis, Auckland University of Technology, New Zealand.

Xiang, S., Cheng, D., Shang, C., Zhang, Y., & Liang, Y. (2022). Temporal and heterogeneous graph neural network for financial time series prediction. ACM International Conference on Information & Knowledge Management,

YiChen, D., Siyi, L., & Xueqin, G. (2017, 2017/04). Time series analysis: An application of ARIMA model in stock price forecasting. International Conference on Innovations in Economic Management and Social Science (IEMSS 2017),

Zhang, G., Eddy Patuwo, B., & Y. Hu, M. (1998). Forecasting with artificial neural networks:: The state of the art. *International Journal of Forecasting*, *14*(1), 35-62.

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, *50*, 159-175.

Zhao, L., Yan, W. (2024) Prediction of currency exchange rate based on transformers. Journal of Risk and Financial Management, 17(8): 332, MDPI.

Zhao, Y., Zhang, W., & Liu, X. (2024). Grid search with a weighted error function: Hyperparameter optimization for financial time series forecasting. *Applied Soft Computing*, *154*, 111362.