Analysis and Forecasting of Financial Time Series

Analysis and Forecasting of Financial Time Series:

Selected Cases

By

Jaydip Sen

Cambridge Scholars Publishing

Analysis and Forecasting of Financial Time Series: Selected Cases

By Jaydip Sen

This book first published 2022

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN (10): 1-5275-8884-X
ISBN (13): 978-1-5275-8884-4

Dedicated to my sister Nabanita who left us on September 27, 2021.

Jaydip

TABLE OF CONTENTS

List of Figures

LIST OF TABLES

About the Author

Jaydip Sen has experience in research, teaching, and industry over a span of 27 years. He had worked in reputed organizations like Oil and Natural Gas Corporation Ltd, India, Oracle India Pvt Ltd, Akamai Technology Pvt Ltd, Tata Consultancy Services Ltd and National Institute of Science and Technology, India, and Calcutta Business School, India. Currently, he is associated with Praxis Business School, Kolkata, INDIA, as a professor in the department of Data Science and Artificial Intelligence. His research areas include security in wired and wireless networks, intrusion detection systems, secure routing protocols in wireless ad hoc and sensor networks, trust, and reputation-based systems, privacy issues in ubiquitous and pervasive communication and the Internet of Things, machine learning, deep learning, and artificial intelligence in the financial domain. He has more than 200 publications in reputed international journals and referred conference proceedings and 20 book chapters in books published by internationally renowned publishing houses like Springer, CRC press, IGI-Global, etc. He has also authored two books which are published by two internationally reputed publishing houses. He has been listed among the top 2% of scientists in the world for the last three consecutive years 2019-2021, as per studies conducted by Stanford University, USA. Prof. Sen serves on the editorial board of three prestigious international journals and on the technical program committee in several international conferences of repute. Prof. Sen is a senior member of IEEE and ACM, USA.

PREFACE

The subject of financial time series analysis has attracted substantial attention in the last two decades, especially after Professors Robert Engle and Clive Granger won the Nobel awards. At the same time, this field has undergone rapid evolution and developments, especially in high-frequency finance, stochastic volatility, and the availability of sophisticated software and tools. While at a basic level, financial time series analysis is concerned with the theory and practice of asset valuation over time, at a broader level it includes diverse topics such as analysis of high-frequency price observations, arbitrage pricing theory, asset price dynamics, optimal asset allocation, cointegration, capital asset price models, and value at risk. It is a highly empirical discipline, but like other scientific fields theory forms the foundation for making inferences.

There is, however, a key feature that distinguishes financial time series analysis from other time series analyses. Both financial theory and its empirical time series contain an element of uncertainty. For example, there are various definitions of asset volatility, and for a stock return series, the volatility is not directly observable. As a result of the added uncertainty, statistical and econometric theories, and methods play an important role in financial time series analysis.

The chapters in the volume present several techniques of financial time series analysis and forecasting based on statistical and econometric approaches. The historical stock prices and sectoral index values for important stocks and sectors listed on the National Stock Exchange (NSE) of India and the Bombay Stock Exchange (BSE) are used in building financial models which are used to predict the future values of the index and stock prices. The statistical and econometric methods discussed in the chapters of this book include exponential smoothing, Holt and Winter trend and seasonality method, time series decomposition, autoregressive integrated moving average (ARIMA) method, ordinary least square regression (OLS), generalized autoregressive conditional heteroscedasticity (GARCH), and cointegration.

Chapter 1 titled *A Robust and Accurate Predictive Framework for the Indian Mid-Cap Sector Index* presents a time series decomposition-based approach for analyzing the salient characteristics of the time series of the mid-cap sector of the Indian economy from January 2010 to December

2021. The time series of the mid-cap sector index is decomposed into its three components - trend, seasonal, and random. Based on the decomposition results several interesting characteristics of the time series are identified. The month of March is found to exhibit the highest level of seasonality for the mid-cap time series, while July experiences the lowest seasonality effect. The time series is found to consist of a very moderate random component with the mean value of the percentage of the random component to the time series' aggregate value being as low as 4.73. The time series was found to be dominated by its trend component. Further, five predictive models are proposed for forecasting the future index values of the time series. Using the five models and based on the training data of the mid-cap sector's monthly index values from January 2010 to December 2021, the monthly index values for the year 2021 are forecasted. The models are compared based on their forecast accuracies.

Chapter 2 titled *A Deep Learning Model for Sectoral Profitability Study of the Indian Stock Market* presents an LSTM model for predicting future stock prices. The model is optimized with suitably designed layers and regularized using the dropout regularization method. The historical stock prices for 140 stocks from fourteen sectors listed in NSE, India are extracted from the web from January 1, 2010, to July 13, 2022. The model is used for predicting future stock prices with a forecast horizon of one day and based on the predicted output of the model, buy/sell decisions of the stocks are taken. The total profit earned from the buy/sell transactions for a stock is normalized by its mean price over the entire period to arrive at the profitability measure of the stock. The profitability figures of all stocks of a sector are summed up to derive the overall profitability index of the sector. It is observed that while the *fast-moving consumer goods* (FMCG) sector is the most profitable one, the *realty* sector has the lowest profitability index. The accuracy of the LSTM model is measured using three metrics, Huber loss, *mean absolute error* (MAE), and the *accuracy score*. It is observed that the LSTM model is highly accurate.

Chapter 3 titled *Sectoral Volatility Analysis of the Indian Stock Market using GARCH*, illustrates the design and analysis of several volatility models based on different variants of GARCH have been presented. The models are built on the historical stock prices from five important sectors listed on the NSE of India from January 1, 2010, to April 30, 2022. From each of the five sectors, the top ten stocks are selected based on the report published by the NSE on December 31, 2021. The five sectors are auto, banking, consumer durables, information technology, and pharma. The GARCH models are fine-tuned and then backtested on the out-of-sample data to estimate their accuracies in the prediction of future volatility of the stocks. While it is observed that asymmetric GARCH models outperform their symmetric counterparts, EGARCH is found to have yielded the most accurate results for all the stocks analyzed in this work.

Chapter 4 titled *Portfolio Design using Mean-Variance Optimization and Hierarchical Risk Parity Approach* presents portfolio design approaches for stocks chosen from thirteen sectors and NIFTY 50 stocks on the National Stock Exchange of India. Based on the historical prices of the ten stocks with the largest free-float market capitalization from each sector and the 50 stocks in the NIFTY group, the *mean-variance portfolio* (MVP) and the *hierarchical risk parity* (HRP) portfolios are designed. The portfolios are backtested on both training and the test data to identify the portfolio with the higher cumulative return and higher Sharpe Ratio for each sector. It is found that while on the training data, the MVP portfolio yielded the higher cumulative returns for seven among the fourteen sectors, the return yielded by the HRP portfolio is found to be higher for eleven sectors on the test data. Since for a portfolio, its performance on the test data is what matters to the investors, the results of the study indicate that the HRP portfolio is a better choice over the MVP for the investors of the Indian stock market.

Chapter 5 titled *A Comparative Study of Performance Metrics in Mean-Variance Portfolio Optimization* illustrates three different approaches to portfolio design based on the maximization of three ratios – the *Sharpe ratio*, the *Sortino ratio*, and the *Calmar ratio*. Twelve important sectors listed on the National Stock Exchange of India are first selected, and portfolios are designed based on the three approaches for ten stocks from each sector on the historical stock prices from January 1, 2017, to December 31, 2020. The performances of the portfolios are evaluated based on their cumulative returns for the period January 1, 2021, to December 31, 2021. It is observed that, for the test period, the cumulative returns of portfolios built on the maximization of the Sharpe ratio are the highest for eight sectors of the twelve sectors studied in the work. The results demonstrate the superiority of the *Sharpe ratio* as the metric in portfolio optimization.

Chapter 6 titled *A Cointegration-Based Approach to Pair Trading of Stocks from Selected Sectors of the Indian Stock Market* presents a cointegration-based pair-trading approach that identifies stock pairs that exhibit robust cointegration in their prices over three years from January 1, 2018, to December 31, 2020. The stocks are chosen from twelve sectors listed on the National Stock Exchange of India. Once the cointegrated pairs are identified, the pair-trading portfolios are formed and the performance of the portfolios is observed over a test period of one year from January 1, 2021, to December 31, 2021. Suitable trigger points are identified so that the short and the long positions for both stocks are identified accurately. The realty and the PSU bank sectors are found to have produced the best results as all the cointegrated pairs from these two sectors yielded positive returns. However, four pairs from the auto sector out of a total of ten pairs yielded negative returns.

While the chapters in this volume do not cover the basic theories of the topics involved, all the relevant principles and fundamentals are discussed in brief in the chapters for the sake of completeness. Hence, even if some background knowledge of statistics and econometrics may be useful, the readers are not expected to have advanced knowledge in those fields. I am sure that the volume will be a valuable resource to anybody interested in gaining knowledge in financial time series analysis. However, the primary target reader groups for the book are the advanced postgraduate and doctoral students of finance, econometrics, management, data science, computer science, and information technology. Further, faculty members of graduate schools and universities, and practitioners in the industry working in the areas of financial analytics, risk management, security analysis, and portfolio management, are also expected to find the book quite useful.

I express our sincere thanks to all without whose help and support this project would not have been a success. Special thanks are due to Adam Rummens, Commissioning Editor, Amanda Miller, Typesetting Manager, and Courtney Dixon, Designer of Cambridge Scholars Publishing for their support and cooperation. The members of my family have always been the sources of motivation and inspiration for all my scholastic and academic works of mine. I dedicate my work to my beloved sister Ms. Nabanita Sen, who, unfortunately, left us on 27th September 2021, due to the deadly disease of cancer. My sister was always the pillar of strength for me, and she was the primary source of support and motivation for my effort toward this volume. Last but not the least, I gratefully acknowledge the immense support and motivation I received from my wife Ms. Nalanda Sen, my daughter Ms. Ritabrata Sen, and my mother Ms. Krishna Sen. Without their sacrifice and support, the publication of this volume would not have been possible. Many thanks to all of them!

Jaydip Sen
Praxis Business School
Kolkata, INDIA

Program Codes: The program codes associated with the chapters of this book are available at the following GitHub link:
https://github.com/JaydipSen/Analysis-and-Forecasting-of-Financial-Time-Series-Selected-Cases

CHAPTER 1

A ROBUST AND ACCURATE PREDICTIVE FRAMEWORK FOR THE INDIAN MID-CAP SECTOR INDEX

Introduction

Developing an accurate and efficient forecasting framework for the robust prediction of stock prices has been one of the most exciting yet complex challenges researchers face in machine learning and analytics. Researchers have proposed numerous technical, fundamental, and statistical indicators for accurately predicting the prices of stocks. Sen and Datta Chaudhuri proposed a novel approach based on time series decomposition and analysis for efficient portfolio diversification and prediction of stock prices (Sen, 2022; Sen, 2018b; Sen, 2018c; Sen, 2017a; Sen, 2017b; Sen & Datta Chaudhuri, 2018; Sen & Datta Chaudhuri, 2017a; Sen & Datta Chaudhuri, 2017b; Sen & Datta Chaudhuri, 2017c; Sen & Datta Chaudhuri, 2016a; Sen & Datta Chaudhuri, 2016b; Sen & Datta Chaudhuri, 2016c; Sen & Datta Chaudhuri, 2016d; Sen & Datta Chaudhuri, 2016e). The authors hypothesized that all sectors of an economy do not exhibit a similar pattern of variations in their stock prices. It is more usual to find that various sectors exhibit different patterns in their trend, different characteristics in their seasonality behavior, and varying degrees of randomness in their time series values. While on the one side, the efficient market hypothesis has argued for the randomness aspect of stock price movements. On the other side, there are propositions to counter the hypothesis by delving into various fundamental characteristics of different sectors and different stocks in those sectors. We argue that in addition to the differences in the fundamental attributes among stocks belonging to multiple companies, the performances of different stocks are also very much dependent on and coupled to the sectors to which the stocks belong. Since the behavior of each sector of the economy is influenced by its unique set of factors, the pattern of the price movement of stocks belonging to various sectors is also determined and influenced by these factors.

In this chapter, our goal is to study the behavioral pattern exhibited by the time series of the *mid-cap* sector of India so that the salient properties of that sector can be better understood. By its definition, a *mid-cap* company has a market capitalization between Indian Rupees (INR) 50 billion to INR 200 billion. For our study, the monthly average index values of the *mid-cap* sector are used for the period January 2010 - December 2016 as per the Bombay Stock Exchange (BSE). The monthly time series data is decomposed into its three components using functions defined in the R programming language. Based on the decomposition results, we demonstrate how several exciting characteristics of the time series can be extracted to gain valuable insights into its behavioral pattern. We particularly illustrate how a more in-depth analysis of the trend, seasonal and random components, provides us with helpful information about the growth pattern, seasonal properties, and randomness exhibited by the time series index values. For predicting future behavioral patterns, we also propose an extensive framework for time series forecasting consisting of five methods of prediction of time series index values. The five forecasting methods are critically analyzed in terms of forecasting accuracy.

The organization of the chapter is as follows. The section titled *Related Work* presents a brief literature survey on some of the current work on time series analysis and forecasting. In the section titled *Methodology*, we present a detailed description of the methodology we followed in this work. We discuss in detail the method of decomposition of the *mid-cap* sector time series into its various components. The section titled *Time Series Decomposition Results* presents extensive decomposition results of the time series values into its trend, seasonal and random components. The decomposition results are analyzed in-depth to understand several essential characteristics and behavior revealed by the time series. In the section titled *Proposed Forecasting Methods*, we propose a set of five forecasting methods for predicting the future values of the time series. In the section titled *Forecasting Results*, we provide results on the performance of five forecasting methods on the *mid-cap* sector time series data. Each of the proposed algorithms is evaluated and compared based on two performance metrics. The first metric is based on the computation of the *root mean square error* (RMSE). However, to express the RMSE as the percentage of the mean value of the target variable, we design a derived metric that is computed as the ratio of the RMSE to the mean values of the target variable. The second metric used for comparing the performance of the predictive model is *mean absolute percentage error* (MAPE). Finally, in the section titled *Conclusion*, we conclude the chapter and highlight some future directions of work.

Related Work

In the literature, researchers have proposed several approaches and techniques for forecasting the daily prices of stocks and index values of various sectors of the economy. Neural network-based approaches are the most popular among these propositions.

Mostafa presented a method of forecasting the movement of stock prices in Kuwait that utilized the concepts of *artificial neural networks* (ANNs) (Mostafa, 2010). The author used historical stock price records from the Kuwait Stock Exchange (KSE) from 2001 to 3003 and builds two models, (i) a multi-layer perceptron (MLP) and (ii) a generalized regression neural network, for predicting the close price of the stocks listed in KSE. The results show that the generalized regression neural network is more accurate in forecasting future stock prices.

Kimoto et al. proposed neural network models based on historical accounting data and various macroeconomic factors. The neural network models were found to be accurate in predicting the movement patterns in several stock returns (Kimoto et al., 1990).

Leigh et al. illustrated various approaches to predicting stock prices and stock market index movements in the New York Stock Exchange (NYSE) from January 1981 to December 1999 (Leigh et al., 2005). The authors used linear regression and simple neural network models in developing their proposed approaches.

Hammad et al. demonstrated how the output of an ANN model can be forced to converge after executing a finite number of iterations and then producing highly accurate predicted values of stock prices (Hammad et al., 2007).

Dutta et al. designed predictive models using ANNs for forecasting the closing index values of the BSE from January 2002 to December 2003 (Dutta et al., 2006).

Tsai and Wang carried out a study investigating how and why the forecasting accuracy produced by a BN-based approach usually is higher than that yielded by a traditional regression and neural network-based method (Tsai & Wang, 2009).

Tseng et al. showed how techniques like traditional time series decomposition, HoltWinters models, Box-Jenkins method, and artificial neural networks can be applied in predicting the prices of a randomly selected set of 50 stocks from September 1998 to December 2010 (Tseng et al., 2012). The authors observed that the forecasting errors were minor for the Box-Jenkins method, HoltWinters model, and normalized neural network model. However, higher error values were observed in the time series decomposition method and non-normalized neural network model.

Moshiri and Cameron proposed a *backpropagation network* (BPN) with econometric models for predicting the inflation level of the economy using the following techniques: (i) Box-Jenkins Autoregressive Integrated Moving Average (BJARIMA) model, (ii) Vector Autoregressive (VAR) model, and (ii) Bayesian Vector Autoregressive (BVAR) model (Moshiri & Cameron, 2010).

Thenmozhi demonstrated how the chaos theory can be applied to identifying the pattern of changes in stock prices in the Bombay Stock Exchange (BSE) from August 1980 to September 1997. The author discovered that the daily and weekly returns of the BSE index exhibited non-linear trends, while the BSE index showed a weakly chaotic movement pattern (Thenmozhi, 2006).

Hutchinson proposed a novel approach based on the principle of learning networks for accurately predicting the price of a derivative (Hutchinson et al., 1994).

Mehtab and Sen presented a series of work on the design of predictive models for the future stock price and index values and movements based on innovative machine learning and deep learning architectures (Sen, 2021; Sen, 2018a; Sen & Mehtab, 2022; Mehtab & Sen, 2022; Mehtab & Sen, 2021; Mehtab & Sen, 2020a; Mehtab & Sen, 2020b; Mehtab & Sen, 2020c; Mehtab & Sen, 2019; Mehtab et al., 2021; Mehtab et al., 2020). The models proposed by the authors used historical stock prices and stock index values at daily intervals or 5 minutes intervals. The power of *convolutional neural networks* (CNNs) and *long- and short-term memory* (LSTM) networks was exploited for achieving a very high level of accuracy in the out-of-sample of data. The authors proposed four CNN models and six LSTM models that differ in architecture and input data (i.e., univariate or multivariate time series data and the size of the input data). The models were compared on their *root mean square error* (RMSE) values. The results elicited two interesting observations: (i) the performances of the CNN models are superior to that of the LSTM models, and (ii) the models based on the univariate data yield more accurate forecasting. In another set of papers, Sen et al. proposed further variants of CNN and LSTM-based models for predicting future stock price values and stock price movements for designing optimum portfolios of stocks (Sen & Datta Chaudhuri, 2018; Sen et al., 2021a; Sen et al., 2021b; Sen et al., 2021c; Sen & Mehtab, 2021d; Sen et al., 2021e; Sen et al., 2021f; Sen & Mehtab, 2021g; Sen et al., 2020). The authors reported extensive results for comparing the performance of the models.