DOKUZ EYLÜL UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

FORECASTING FINANCIAL DATA UNDER STRUCTURAL BREAKS AND ARCH INNOVATIONS

by

Daud Ali ASER

January, 2023

IZMIR

FORECASTING FINANCIAL DATA UNDER STRUCTURAL BREAKS AND ARCH INNOVATIONS

A Thesis Submitted to the

Graduate School of Natural and Applied Sciences of Dokuz Eylül University

In Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Statistics

by

Daud Ali ASER

January, 2023

IZMIR

Ph.D. THESIS EXAMINATION RESULT FORM

We have read the thesis entitled **"FORECASTING FINANCIAL DATA UNDER STRUCTURAL BREAKS AND ARCH INNOVATIONS"** completed by **Daud Ali ASER** under supervision of **Prof. Dr. Esin FIRUZAN** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy.

Prof. Dr. Esin FIRUZAN
Supervisor

Prof.Dr. Ali Hakan BÜYÜKLÜ
Thesis Committee Member

Assoc.Prof. Idil YAVUZ
Thesis Committee Member

Prof.Dr. Erol EĞRIOĞLU
Examining Committee Member

Prof.Dr. Ceylan YOZGATLIGIL
Examining Committee Member

Prof.Dr. Okan FISTIKOĞLU
Director
Graduate School of Natural and Applied Sciences

ACKNOWLEDGEMENTS

First, praise be to Allah, the Almighty, for his countless blessings throughout my life and especially during my thesis to complete this research successfully.

I owe a great deal of appreciation to my distinguished supervisor, Prof. Dr. Esin FIRUZAN, for her indispensable guidance, unwavering encouragement, and tolerance while I was pursuing my Ph.D. Her vast knowledge and wealth of experience have inspired me in both my daily life and academic research. She was always there for me, even on holidays and outside of regular business hours.

I would like to express thanks Prof. Dr. Ali Hakan BÜYÜKLÜ, a member of the thesis committee, for his pertinent comments and contributions that assisted me develop this thesis. I would also like to thank Assoc. Prof. İdil YAVUZ, a member of the thesis committee, for her contributions and precious support, which was instrumental in shaping my modeling techniques, especially simulation designs.

I would desire to dedicate my work to my deceased father and my supportive mother. May Allah have mercy on them as they raised me when I was so little. I would likewise express my gratefulness to my wife, and my children. Without their assistance and inspiration over the past few years, I would have been unable to complete my thesis.

Finally, I want to acknowledge Yurtdışı Türkler ve Akraba Topluluklar Başkanlığı for their financial assistance. Thank you very much for providing me with a Ph.D. scholarship and the chance to complete one of my life's most significant goals.

Daud Ali ASER

FORECASTING FINANCIAL DATA UNDER STRUCTURAL BREAKS AND ARCH INNOVATIONS

ABSTRACT

Forecasting future values is fundamental for researchers, investors, practitioners and obtaining accurate predictions is critical in time series analysis. Accuracy is reliant on good modeling and good quality data. The latter is affected by unusual observations, changes over time, missing data, and structural breaks among others. Two major sources of complexity are structural breaks and ARCH effects, which affect the data quality and hence reduce forecast accuracy. Therefore, this thesis focuses on how structural breaks in conditional heteroscedastic financial data affect forecasting accuracy in short and long-term horizons. More specifically, we are interested in the impact of structural break location in the sample and break size on the predictive performance of linear and nonlinear forecasting models.

This study showed that the forecasting accuracy decreases when the structural break location approaches the end of the series. Moreover, it is shown that ARCH model is the best performing individual model in the absence of structural break while Multilayer Perceptron (MLP) Artificial Neural Network model outperforms all the individual models in the structural break presence, especially in large break sizes and long horizons.

In addition, we examined if the accuracy can be improved using combined forecasts in the cases of structural breaks and ARCH innovations, and which forecast combination schemes are optimal in those cases. It is found that regression-based forecast combination methods can enhance prediction accuracy in the cases of structural breaks and conditional heteroscedasticity. Dynamically weighted combinations show an improvement in accuracy than static version if the data contains a trend.

Keywords: Structural breaks, break location, forecasts Combination, ARCH effects, artificial neural networks

YAPISAL KIRILMALAR VE ARCH İNOVASYONLARI VARLIĞI DURUMUNDA FİNANSAL VERİLERİ ÖNGÖRÜLEME

ÖZ

Gelecekteki değerleri tahmin etmek araştırmacılar, yatırımcılar, uygulayıcılar için esastır ve doğru tahminler elde etmek zaman serisi analizinde kritik öneme sahiptir. Doğruluk, iyi modellemeye ve kaliteli verilere bağlıdır. Dahası, olağandışı gözlemlerden, zaman içindeki değişikliklerden, kayıp gözlemlerden ve hatta yapısal kırılmalardan etkilenir. İki temel çözümleme sorunu, verilerin kalitesini etkileyen ve dolayısıyla tahmin doğruluğunu azaltan, verilerdeki yapısal kırılmalar ve ARCH etkileridir. Bu nedenle, bu tez, koşullu değişen varyans ve yapısal kırıma varlığında finansal verilerin kısa ve uzun vadeli ufuklarda öngörü doğruluğunu nasıl etkilediğine odaklanmaktadır. Daha spesifik olarak, örneklemdeki yapısal kırılmanın konumunun ve kırılma boyutunun doğrusal tahmin modellerinin ve iki doğrusal olmayan modelin öngörü performansı üzerindeki etkisiyle ilgileniyoruz.

Bu çalışma, yapısal kırılma konumu serinin sonuna yaklaştığında tahmin doğruluğunun azaldığını göstermiştir. Ayrıca, ARCH modelinin yapısal kırılma olmadığında en iyi performans gösteren bireysel model olduğunu, Çok Katmanlı Algılayıcı Yapay Sinir Ağı modelinin yapısal kırılma varlığında, özellikle büyük kırılma boyutlarında ve uzun ufuklarda, uygulanan tüm bireysel modellerden daha iyi performans gösterdiği de gösterildi.

Ek olarak, yapısal kırılmalar ve ARCH inovasyonlarının varlığı durumlarında birleştirilmiş öngörüler kullanılarak doğruluğun iyileştirilip geliştirilemeyeceği ve bu durumlarda hangi öngörü kombinasyon şemalarının optimal olduğu incelendi. Regresyona dayalı öngörü kombinasyon yöntemlerinin, yapısal kırılmalar ve koşullu değişen varyans durumlarında öngörü doğruluğunu artırabileceği bulunmuştur. Veriler trend içeriyorsa, dinamik olarak ağırlıklı kombinasyonların, statik eşdeğerlerine göre doğrulukta daha iyi bir-performans gösterdikleri görülmüştür.

Anahtar Kelimeler: Yapısal kırılmalar, kırılma yeri, öngörü kombinasyonu, ARCH etkileri, yapay sinir ağları

CONTENTS

Page

PH.D. THESIS EXAMINATION RESULT FORM ii
ACKNOWLEDGEMENTS iii
ABSTRACT iv
ÖZ vi
LIST OF FIGURES xi
LIST OF TABLES xii

CHAPTER ONE - INTRODUCTION 1

CHAPTER TWO - LITERATURE REVIEW 5

2.1 Structural Breaks 5
2.2 ARCH Innovations 7
2.3 Implications of Structural Breaks and ARCH Innovations on Forecast Performance… 9
2.4 Literature of Forecasting under Structural Breaks and ARCH Innovations ... 10
2.4.1 Literature Related to Single Models 10
2.4.2 Literature Related to Combined forecasts 12

CHAPTER THREE - METHODOLOGY 15

3.1 Structural Break Tests 15
3.1.1 Chow Test 16
3.1.2 CUSUM Test 18
3.1.3 BAI-PERRON TEST of Multiple Structural Breaks 20
3.2 Autoregressive Conditional Heteroscedasticity (ARCH) Effect Test 22
3.3 Forecasting Models 23
3.3.1 Single Models 23
3.3.1.1 Exponential Smoothing (ETS) Model 23

3.3.1.2 Autoregressive Integrated Moving Average (ARIMA) Model 25
3.3.1.3 Combined ARIMA and ARCH (ARIMA - ARCH) Model..... 26
3.3.1.4 Artificial Neural Network (ANN) Model 27
3.3.2 Forecast Combinations 30
3.3.2.1 Simple Forecasts Combination Methods 31
3.3.2.1.1 Simple Average Based Forecast Combination..... 31
3.3.2.1.2 Median Based Forecast Combination 32
3.3.2.1.3 Bates/Granger Forecast Combination 32
3.3.2.2 Regression-based Forecast Combination Methods..... 33
3.3.2.2.1 Ordinary Least Squares (OLS) Regression-based Forecast Combination..... 33
3.3.2.2.2 Least Absolute Deviation (LAD) Regression-based Forecast Combination..... 34
3.4 Forecasting Accuracy Metrics..... 35
3.4.1 Root Mean Squared Forecast Error (RMSE)..... 35
3.4.2 Mean Absolute Percentage Error (MAPE)..... 35

CHAPTER FOUR - IMPACT OF STRUCTURAL BREAK LOCATION WITH ARCH INNOVATIONS ON FORECASTING ACCURACY: TRADITIONAL METHODS VERSUS ARTIFICIAL NEURAL NETWORK 37

4.1 Data Description and Simulation Designs..... 38
4.1.1 Data Generating Process in Simulation 38
4.2 Simulation Results..... 41
4.2.1 Case I: Forecasts of Stable Simulated Data..... 41
4.2.2 Case II: Forecasts when Break size (δ) = 3 42
4.2.3 Case III: Forecasts when Break size (δ) = 5 42
4.2.4 Case IV: Forecasts when Break size (δ) = 10..... 43
4.2.5 Summary of Simulation Results 44
4.3 Empirical Applications..... 46
4.3.1 Structural Break Testing..... 49
4.3.2 ARCH Effect Testing 50

4.4 Empirical Results 50
4.5 Chapter Conclusion 53

CHAPTER FIVE - IMPROVING FORECAST ACCURACY WITH COMBINATION OF FORECASTS UNDER STRUCTURAL BREAKS AND ARCH INNOVATIONS 55

5.1 Data Description and Simulation Designs 57
5.1.2 Simulation Designs 58
5.2 Simulation Results 62
5.2.1 Summary of Simulation Results 65
5.3 Empirical Applications 66
5.3.1 Testing ARCH Effects 68
5.3.2 Structural Break(s) Testing 68
5.4 Chapter Conclusion 74

CHAPTER SIX - CONCLUSION 75

REFERENCES 77

LIST OF FIGURES

Page

Figure 3.1 Artificial Neural Networks Architecture 28
Figure 3.2 Activation functions (tanh vs logistic)..... 29
Figure 4.1 Illustration of Simulated Data Prior Structural Break 39
Figure 4.2 Illustration of Break Locations and Break Sizes of Simulation 40
Figure 4.3 USD - TL Exchange Rate 47
Figure 4.4 Brent Oil Prices..... 47
Figure 4.5 Break Locations at Brent Oil Prices and TL – USD Exchange Rate..... 48
Figure 5.1 Graphical illustration of the simulated data..... 60
Figure 5.2 Simulated Data after Structural Break 61
Figure 5.3 BIST30 and BIST100 Data..... 67
Figure 5.4 Break Locations and Segments..... 69
Figure 5.5 BIST30 and BIST100 Means and Variations 73

LIST OF TABLES

Page

Table 3.1 R Functions for Forecasts Combination and their Description 34
Table 4.1 Summary of the Simulated Data Procedures and Train-Test Sets 40
Table 4.2 Hyperparameter Values 41
Table 4.3 RMSEs Values of Stable Simulated Data 41
Table 4.4 RMSE Values When Break Size (δ) = 3 42
Table 4.5 RMSE Values when Break Size (δ) = 5 43
Table 4.6 RMSE Values When Break Size (δ) = 10 44
Table 4.7 Break Dates and Segmentation Using Bai and Perron Test 49
Table 4.8 Significance of the Structural Break 49
Table 4.9 Lagrange-Multiplier test 50
Table 4.10 RMSE Values of Empirical Analysis 51
Table 4.11 MAPE Values of Empirical Analysis 51
Table 5.1 Summary of the Simulated Data Procedures and Train-Test Sets 61
Table 5.2 Design I Stable Case Results 62
Table 5.3 Design I with Structural Break 63
Table 5.4 Design II Stable Case Results 64
Table 5.5 Design II with Structural Break 65
Table 5.6 Lagrange-Multiplier Test Results 68
Table 5.7 BIST100 Break Locations and Segmentation 69
Table 5.8 BIST30 Breakpoints and Segmentation 70
Table 5.9 Train-Test Sets Plan of the Empirical Data 70
Table 5.10 BIST100 RMSEs Results 71
Table 5.11 BIST30 RMSEs Results 71
Table 5.12 BIST100 MAPE Results 72
Table 5.13 BIST30 MAPE Results 72

CHAPTER ONE
INTRODUCTION

Forecasting is about predicting the future as accurately as possible, given all of the information available, including historical data and knowledge of any potential future events that might impact the forecasts (Hyndman & Athanasopoulos, 2018). Time series analysis is used for many purposes, including forecasting future results, interpreting historical results, and formulating suggestions for policy.

Unexpected changes in model parameters called structural breaks are sometimes observed in financial time series data, possibly because of changes in policy, financial crises, and natural disasters, among others. The Great Moderation crisis of the mid-1980s, the enormous financial crisis known as the "Great Recession" of 2007–2008, during which the relationship between macroeconomic variables changed significantly, and the recent COVID-19 crises are notable examples of crises that brought about structural changes (Rossi, 2021). While failing to notice structural breaks might result in flawed conclusions and incorrect forecasts, doing so can help us better understand the fundamental mechanisms underlying data change.

Time-varying volatility, commonly referred to as conditional heteroscedasticity, is another detrimental attribute in financial time series. Actually structural breaks are often associated with heteroscedasticity in real-world datasets (Ahmed et al., 2017). Heteroscedasticity occurs when the variation of a variable is not constant when measured over time. Most types of volatility in financial time series result in serially correlated heteroscedasticity, which is thus conditional on times of increased variance. Consequently, we call such series conditional heteroscedastic. Assuming conditional homoscedasticity, especially in the case of structural breaks in forecasting time series may seem restrictive.

An accurate forecast about the future is vital in time series analysis, and it is always challenging for forecasters to accurately model complex structures in the data. That complexity is related to the identification of the underlying data-generating

mechanism, as it is not fully known in most cases. As the observed processes are too complicated to be accurately modeled with the exception of some natural sciences (Weiss et al., 2019), models are approximations of the real-world data-generating processes (Hansen, 2005).

Linearity assuming models such as ARIMA do not account for nonlinear patterns, which may result from structural breaks and changes in variance over time in financial time series and therefore suffer from generating inaccurate forecast values for the series featuring nonlinear patterns. Autoregressive conditional heteroscedastic (ARCH) models and their extension GARCH can be modeled if the change in variance can be correlated overtime. However, this is only a special case of nonlinear problems but may not perform well in general nonlinear patterns (Fischer et al., 2018) including the cases of structural breaks.

Many studies suggested machine learning models to make better prediction accuracy under structural breaks and ARCH effects. In one of the most recent studies on the subject, Kaushik et al. (2021) examined the performance of different deep neural networks (DNNs) when predicting time series with multiple structural breaks and high volatility. Short-term long memory (LSTM-RNN), multilayer perceptron (MLP), gated recurrent unit (GRU-RNN), and convolutional neural network (CNN) were the models they took into consideration. The DNN approaches produced convincing results for single-step forecasting, and their performance revealed that long prediction horizons have a negative impact on results.

Despite the extensive studies on the structural break and its impact on forecasting, little emphasis is paid to how the structural break location in time series affects predicting accuracy. Questions such as whether forecasting accuracy decreases as the structural break approaches the end of the sample series remain unaddressed and that is what inspired this thesis.

This thesis focuses on how structural breaks in conditional heteroscedastic financial data affect forecasting accuracy on short and long-term horizons. A significant portion

of the application chapters of this thesis is devoted to assessing the impact of the structural break location on forecast accuracy. This is done by placing structural breaks in different locations in the time series and measuring forecast error in each case. In doing this, we compared the predictive performance of four methods, namely Exponential Smoothing (ETS), Autoregressive Integrated Moving Average (ARIMA), Autoregressive Integrated Moving Average -Autoregressive Conditional Heteroskedastic (ARIMA – ARCH), and a Multilayer Perceptron (MLP) Artificial Neural Networks (ANN). This means we combined linear and non-linear forecasting methods on one side and classical and ANN on the other. These methods are implemented on simulated data and empirical data of daily close prices of brent oil and Turkish Lira – USD exchange rates after it is confirmed that data contain structural breaks and ARCH effects.

Another strategy to improve the forecasting accuracy is aggregating inputs from several forecasting techniques using forecasts. Combining the forecasts has been a helpful strategy to improve forecast accuracy for more than five decades since the original paper of Bates and Granger in 1969. Hence, this thesis also examines if the gains of combined forecasts are sustained in the cases of structural breaks and ARCH innovations. Several combination schemes including simple and regression-based combination procedures are explored by combining the forecasts of ETS, ARIMA and ANN models. The forecast combination methods are implemented in both simulated series and empirical data of two popular Turkish stocks namely BIST30 and BIST100. Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE) are the metrics used to compare the performances of the forecasting models. Chow, CUSUM, and Bai-Perron tests are applied to examine if structural breaks are present and to estimate the dates of the break.

This thesis makes several contributions to the forecasting field thus we think it benefits the scientific community in the following ways:

1. It examines not only the general impact of structural breaks in the forecasting accuracy, but also investigates how the location of the breaks affects the

accuracy of the forecasts. The break is positioned in three different test locations and namely first quartile (Q1), second quartile (Q2) and third quartile (Q3) of the time series sample and forecasting error is measured in each case. In addition, several magnitudes of breaks are examined to check if the size of the break affects the forecast accuracy.

2. Linear (ARIMA and Exponential Smoothing) and non-linear (ARIMA – ARCH and ANN) forecasting models are compared on one side and traditional and artificial neural networks on the other side.

3. It applies several forecast combination schemes to simulations and empirical financial data to improve accuracy of forecast.

The other sections of the thesis are organized as follows: Chapter 2 reviews the literature regards structural breaks and ARCH effects and the impacts that the presence of these features may have on forecasting. Chapter 3 provides information about the methods employed in forecasting, structural break tests, ARCH effect tests and metrics for forecasting accuracy. Chapter 4 focuses on the impact of structural break location on forecasting accuracy while Chapter 5 tries to improve the forecast accuracy based on combination of forecasts under structural breaks and ARCH innovations. Lastly, chapter 6 contains conclusion remarks based on the findings from the analysis.

CHAPTER TWO
LITERATURE REVIEW

Time series models are used for a wide range of reasons, including giving policy recommendations, interpreting historical data, and forecasting future results. The dynamics of a time series could abruptly change. That is, changes in policy, financial crises, natural disasters, and technological advancements may cause sporadic changes in the mean, slope, trend function, variance, or any other attribute. These shifts, known as structural breaks, usually cause parameter instability and diminish the model's ability to meet any of the above intentions, as they lead to inaccurate conclusions and inaccurate forecasts. Financial time series are also well known for time-varying volatility, also called conditional heteroscedasticity. In real data, structural breaks are often associated with heteroscedasticity.

This chapter focuses on the literature related to structural breaks in financial and economic indicators, implications of structural breaks on forecast performance, and the available literature on forecasting under of structural breaks and ARCH innovations.

2.1 Structural Breaks

The fundamental tenet of applied time series analysis and prediction is stationarity, which states that variables like mean, variance, and trend remain constant across time. However, it is a general concern in empirical economics and financial time series studies that unexpected changes (breaks) over time in model parameters are experienced. In the literature of econometrics and statistics, these unforeseen changes over time in model parameters are referred to as structural breaks. The structural change may impact any or all of the model parameters, such as intercept, which controls the mean of the series, the trend, the variance, or any other characteristic in the series and may have different implications (Hansen, 2001).

Numerous studies have revealed that structural breaks are presented in many crucial and widely used financial indicators. For instance, Stock & Watson (1996) examined many U.S. economic relationships for model instability. The series examined key economic variables such as stock prices, industrial production, interest rates, and consumer expectations. Their findings showed that a significant proportion of the indicators have structural breaks and that these instabilities affect forecast performance. Koop and Potter (2001) studied this matter of structural break presence through an empirical analysis involving several macroeconomic indicators. Their findings showed that apparent threshold-type nonlinearities could be due to structural instability. Other study found structural breaks in the models of financial time series, including the monetary policy reaction function (Inoue & Rossi, 2010), global house prices (Knoll et al., 2014), labor productivity (Hansen, 2001), inflation and International interest rates (Rapach & Wohar, 2005), growth rates of industrial production (Siliverstovs and van Dijk 2002), the equity premium (C. J. Kim et al., 2005), CO2 emissions (Acaravci & Erdogan, 2016).

Policy adjustments, technological advancement, financial crises, and natural disasters are all possible causes of structural breaks in the literature. According to Antoch et al. (2019) and Andreou and Ghysels (2009), some of the causes of structural breaks include adjustments to factors related to the adoption of a single currency in Europe, price liberalization in developing countries, the global integration of equities markets, and adjustments to exchange rate regimes including the collapse of exchange rate mechanisms. Furthermore, Rossi (2021) identified three primary sources of structural breaks in economic time series, including the crises in the mid-1980s, characterized by a reduction in the volatility of business cycle fluctuations in developed nations compared to previous decades. This crisis is termed the "Great Moderation." Moreover, the "Great Recession", the enormous financial crisis in 2007-2009, during which the relationship among macroeconomic indicators changed dramatically, and the COVID-19 recent crises are notable causes of structural changes. Another study by Valadkhani et al. (2005) also stated that structural breaks might occur because of economic crises, changes in institutional arrangements, policy changes, and regime shifts. All the above-reviewed studies revealed the existence of structural

breaks, especially in most financial and macroeconomic indicators, and have a profound impact on the modeling and forecasting accuracy. The latter issue is discussed in the next section.

2.2 ARCH Innovations

Time-varying volatility, commonly referred to as conditional heteroscedasticity, is another detrimental attribute in financial time series. Heteroscedasticity occurs when the variation of a variable is not constant when measured over time. Most types of volatility in financial time series result in serially correlated heteroscedasticity, which is thus conditional on periods of increased variance. As a result, we call such series conditional heteroskedastic. Because the ARIMA models cannot account for conditional heteroscedasticity, a model that accounts for fluctuations in variance over time using past variance values is required, and this is what Autoregressive Conditional Heteroskedastic (ARCH) model does. The ARCH model developed by Engle (1982) was the first paper published to consider a parametric model for volatility. The ARCH model is a time series model that computes the variance of the present error term or innovation as a function of the sizes of the error terms in earlier time periods. Therefore, the variance is typically correlated with the squares of earlier innovations. When an autoregressive (AR) model is used to predict the error variance in a time series, the ARCH model is applicable.

To model a time series using an ARCH process, let ε_t represent the error terms, i.e., the series terms. This ε_t time series is separated into a stochastic part a_t and a time-dependent standard deviation σ_t that characterizes the typical size of the terms so that

$$\varepsilon_t = \sigma_t a_t \tag{2.1}$$

The random variable a_t in equation (2.1.) is a white noise process. The series σ_t^2 can be modeled as

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 \quad (2.2)$$

where $\omega > 0$ and $\alpha_i \geq 0$, $i > 0$.

Bollerslev (1986) extended the ARCH model by integrating the lag of conditional variance. Bollerslev model considered the conditional variance, which is represented by a moving average process rather than an autoregressive process and is known as generalized ARCH (GARCH) in equation (2.2).

In the GARCH (p, q) model, the variance is modeled as in equation (2.3).

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_p \sigma_{t-p}^2 \quad (2.3)$$

which can be represented as in equation (2.4).

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 \quad (2.4)$$

The ARCH-type models are extensively used in forecasting financial time series. As they can measure volatility, ARCH models are used by financial organizations in modeling risks and forecasting them into the future. Although ARCH models and their extensions are widely applied in financial forecasting, it could be rather challenging if the conditional heteroscedasticity is accompanied by instability of the other parameters of the model, such as the mean and/or the trend. Chapter three of this thesis discusses the methods to test if structural breaks and ARCH effects are presence.

2.3 Implications of Structural Breaks and ARCH Innovations on Forecast Performance

Structural breaks can cause massive forecasting errors, and overall model unreliability. Several studies have explicitly stated that the impending practical consequence of departures from constant parameter linear models is undeniable. Hansen (2001) emphasized this issue: "Structural change is pervasive in economic time series relationships, and it can be quite perilous to ignore. Inferences about economic relationships can go astray, forecasts can be inaccurate, and policy recommendations can be misleading or worse". Stock and Watson's (1996) study likewise investigated the impact structural breaks may have on accuracy of forecast especially when they are not appropriately accounted for in the model. In this study, the forecasting abilities of models with parameter adaptivity, such as rolling regressions, recursive least squares, and time-varying parameter models, were compared to fixed-parameter models. According to their findings, adaptive models outperform fixed-parameter models in more than half of the cases based on out-of-sample forecast. They concluded that poor forecast accuracy results from model misspecification caused by failing to take structural changes into account.

Clements and Hendry (1998, 1999) stated that macroeconomic models cannot forecast well because of the structural breaks and warned that ignoring the breaks leads to a significant divergence between theory and practice. A study conducted by Bauwens & Dufays (2011) on estimation and also forecasting structural breaks in financial time series and argued that ignoring structural breaks generates poor forecasts and frequently gives the false impression of the time series' integrated behavior. Rossi (2021) wrote a survey article touching on critical aspects of structural changes in financial time series. This paper investigated the importance of considering structural breaks when modeling, assessing the forecasting performance under parameters' instabilities, improving forecasts with structural breaks, and correctly measuring and assessing prediction uncertainty in unstable environments. Findings of this paper include that structural changes may cause instabilities in forecasting, and it is extremely crucial, both empirically and theoretically, to take them into account when

forecasting and evaluating forecasts and their uncertainties. It is also part of the findings that standard tests for forecast evaluation and methods to measure uncertainty around the forecast could be invalid under structural breaks. Andreou and Ghysels (2009) also argue that disregarding structural changes in financial processes can yield false persistence in conditional volatility. For example, neglected structural breaks in financial time series can generate integrated ARCH or long memory effects, which can also have consequences on forecasting.

2.4 Literature of Forecasting under Structural Breaks and ARCH Innovations

2.4.1 Literature Related to Single Models

Structural breaks and heteroskedasticity got significant attention in the empirical and theoretical literature. However, most of the studies on this end focused on each feature separately while they may interact in the real data because structural breaks are often associated with heteroskedasticity in real-world datasets (Ahmed et al., 2017). This section briefly discusses recent forecasting literature under structural breaks and/or ARCH effects.

Altansukh & Osborn (2022) conducted a study on forecasting time series using structural break inferences. This study recommended to average forecasts over sub-samples that are indicated by a confidence interval or established for the break date as opposed to relying on a point estimate of a coefficient break date in the event of structural breaks. Furthermore, they investigated whether utilizing a two-step process and explicitly taking into account a potential variance break improving forecast accuracy when compared to robust heteroskedasticity inference. Their simulation results and empirical implementation of U.S. productivity growth showed that averaging with the likelihood ratio-based confidence set typically outperforms other competing methods. Two-step inference is especially helpful when a variance break occurs simultaneously with or after any break in the coefficient.

Bouri et al. (2021) also examined how U.S. ethanol prices' price and volatility dynamics can be forecasted under structural breaks. The findings of this study revealed

that ARCH/ GARCH models integrating the breaks improves the forecast of U.S. ethanol market instability. Additionally, volatility perseverance tends to decrease when GARCH models incorporate structural breaks. Their findings suggested that ignoring such breaks could mislead the U.S. biofuel industry's risk assessment procedure.

Lee et al. (2022) argued that the Stein-like combined estimator they developed to forecast out-of-sample is optimal forecast under structural breaks. The interesting aspect of this study is how the new estimating method balances the consistency of the post-break estimators with the effectiveness of the complete sample estimator by utilizing both pre-break and post-break data. Real output growth forecasts across nine industrial economies were used to empirically test the proposed method's properties. The performance of the proposed procedure with a range of different techniques existing in the literature using simulation and the empirical showed the superiority of the new method for any break sizes and breakpoints.

Pettenuzzo and Timmermann (2011) also investigated how well stock returns and asset allocation might be predicted in the presence of structural breaks. They adopted a method that accounts for interruptions in the return prediction model's parameters both during the historical estimation period and at potential future points. The analysis of this study took into account a number of sources of uncertainty, including parameter uncertainty, model uncertainty, uncertainty regarding the number, location, and size of earlier breaks to model parameters, and uncertainty regarding upcoming breaks. The results of this research show that the parameters of conventional forecasting models are quite unstable and prone to many breaks, many of which coincide with important historical events. Furthermore, it has been found that an investor's ideal asset allocation is significantly impacted by the likelihood of past and future breaks. The main finding was that traditional return prediction model instability may indicate long-run risk estimations that are significantly higher than generally believed, which would make equities less desirable to long-term investors. Yin (2014) considered how to forecast a time series variable of interest in the case of parameter instability by emphasizing model selection criteria. This paper discusses several break sizes, different test sample

sizes, and various data-generating designs using Monte Carlo Simulation evidence and empirical analysis. All the experiments show that cross-validation weights outperform the competing methods under structural breaks and heteroscedasticity.

Bauwens et al (2011) also joined the other authors in revealing that neglecting change points in financial time series results in subpar projections and frequently creates the false impression that the time series is nearly integrated. They made a contribution in this area by developing an algorithm for Bayesian inference in AR-GARCH models subject to an unknown number of structural breaks at an unknown time using the Markov chain Monte Carlo (MCMC) approach. The marginal likelihood criteria are used to calculate the number of breaks using this method, which uses break dates as parameters. This study demonstrated how to forecast under structural breaks and took into account both pure-change point and recurrent regime specifications. Simulations and empirical analysis of eight financial time series of daily returns from 1987 to 2011 are used to demonstrate the algorithm's effectiveness.

Kaushik et al. (2021) examined the performance of different deep neural networks (DNNs) when forecasting time series with multiple structural breaks and high volatility. Short-term long memory (LSTM-RNN), multilayer perceptron (MLP), convolutional neural network (CNN), and gated recurrent unit (GRU-RNN) were the models they took into consideration. The DNN approaches produced convincing results for single-step forecasting, and their performance revealed that long prediction horizons have a negative impact on results.

2.4.2 Literature Related to Combined forecasts

As an alternative to forecasts based on individual models, it is possible to integrate forecasts produced from the same model but over various estimation windows, or from distinct single models. Although most studies focused on individual forecasting models under structural breaks and volatility, some studies touched on forecast combinations under structural breaks and ARCH effects. In this regard, Pesaran and Timmermann (2007) suggested forecast combinations that are constructed by

averaging forecasts made using all feasible window sizes, subject to a minimum length constraint. Based on the same idea, more complicated forecasting techniques have been suggested. For instance, Tian and Anderson (2014) suggested two novel weighting methods that take into account probable structural change by averaging forecasts made using different estimating windows. Forecasts based on inverted ordered CUSUM (extension of CUSUM) test statistics are given greater weight in the first approach, whilst forecasts based on more recent data are given more weight in the second. Their simulation and empirical evidence demonstrate that second weighting scheme-based combination forecasts outperform equally weighted combinations. De Gaetano (2018) also paid attention to forecasting schemes produced by the same model but using several estimating windows. They is analyzed heterogeneous autoregressive models of realized volatility along with many forecast combinations based on various weighting schemes and forecast combinations with equal weights. The results of this study demonstrated that forecast combination schemes beat the recursive forecasting approach, which is suitable in the absence of structural changes, and that weighting schemes that give greater weight to recent data are often preferable.

Despite the extensive literature on structural breaks and volatility analysis and their impact on forecasting, in reviewing that literature, two considerable gaps are realized, which this thesis will focus on. Firstly, the effect of structural break is generally considered looking to various aspects however, little emphasis is paid to how structural break location in time series affects predicting accuracy. Questions such as whether forecasting accuracy decreases whenever structural break approaches the ending of the series remain unaddressed. To address this aspect of structural break implication, we placed structural breaks in three different predetermined positions and measured the accuracy in each case using several forecasting models. Secondly, combining the forecasts has been a helpful strategy to improve forecast accuracy for more than five decades since the original paper of Bates and Granger in 1969, however, if the gains of combined forecasts are sustained in the cases of structural breaks and ARCH innovations is still understudied. Especially simple combination methods and regression-based forecast combination methods containing static and dynamic weighting strategies are considered in this thesis in cross-cutting with structural breaks

and ARCH effects. This kind of combination is not well taken into account in the reviewed literature.

CHAPTER THREE
METHODOLOGY

It is crucial to choose the appropriate methodology to generate accurate predictions in the case of structural breaks. In this chapter, information is given about the durable and adaptive methodologies, especially in cases where single or multiple structural breaks and ARCH effect are present in the series. Chow Test and CUSUM Test were used to detect the presence of a single structural break, Bai-Perron tests were introduced for multiple structural breaks, while Engle's Lagrange Multiplier tests were introduced to detect the presence of ARCH effect. A significant part of this chapter is also devoted to forecasting methods comprising individual models and forecast combination methods. The individual models considered here are ARIMA, Exponential Smoothing, ARIMA-ARCH, and Artificial Neural Networks models. At the same time, the forecast combination methods are subdivided into simple forecast combination methods and regression-based forecast combination methods. Simple average based forecast combination, median based forecast combination, and Bates-Granger based forecast combination are termed as simple forecast combination methods, whereas forecast combinations based on Ordinary Least Squares (OLS) Regression and Least Absolute Deviation (LAD) Regression are labeled as regression-based combination methods. Finally, methods for measuring forecasting accuracy are introduced and those employed in this thesis include Mean Absolute Percentage Error (MAPE) and Root Mean Squared Forecast Error (RMSE). The R functions used to implement these models are also briefly discussed.

3.1 Structural Break Tests

Identifying structural breaks in time series modeling is an essential topic in statistical and econometric literature. Numerous tests are available in the literature on structural break testing. The test that should be used in a specific circumstance depends on several possible factors, including whether the breakpoint is known or unknown and whether the data contains one break or more breaks. By considering the statistical

attributes of structural breaks and time series data, this study implemented three classic structural break tests and their extensions.

3.1.1 Chow Test

One of the initial tests that served as the basis for examining structural breaks was the Chow test, which was put forth by econometrician Gregory Chow in 1960. The Chow test is one of a class of F statistics-based structural change tests used in time series data analysis to determine the possibility for a single structural break to occur at a predetermined period (for example, a significant historical event such as a war or pandemic). Consider the classical linear regression model in equation (3.1).

$$y_t = x_t'\beta_t + \varepsilon_t \quad (t = 1, \dots, \mathrm{T}) \tag{3.1}$$

where y_t represents dependent variable's observation at time t, x_t = (1, $x_{t2}, \dots, x_{tk}$)′ denotes $k \times 1$ vector containing independent variables' observations and the first component is unity, u_t follows iid$(0, \sigma^2)$, and β_t is a $k \times 1$ vector of coefficients of regression. It compares the alternative of a known structural break existing at time T_b against the null hypothesis of no structural break. In order to perform this test, a linear model is built and then split into samples at a chosen break point.

$$y_t = x_t'\beta_1 + \varepsilon_{1t}, \qquad (1 \le t \le \mathrm{T}_b) \tag{3.2}$$

and

$$y_t = x_t'\beta_2 + \varepsilon_{2t}, \qquad (\mathrm{T}_b < t \le \mathrm{T}) \tag{3.3}$$

where T_b is some change point. Chow proposed fitting two separate regression models on the two subsamples as described in equation (3.2) and equation (3.3) and rejecting the null hypothesis whenever $\mathrm{F}_{\mathrm{T}_b}$ in equation (3.4)

$$F_{T_b} = \frac{\hat{\varepsilon}'\hat{\varepsilon} - \hat{e}'\hat{e}/k}{\hat{e}'\hat{e}/(T-2k)} \quad (3.4)$$

is too huge, where $\hat{e} = (\hat{\varepsilon}_1, \hat{\varepsilon}_2)'$ are the model's overall residuals, the subsamples' coefficients are estimated separately, and $\hat{\varepsilon}$ are the residuals from the restricted model, where only one fitting of the parameters is performed for all observations. The F_{T_b} test statistic has an asymptotically χ^2 distribution with k degrees of freedom whereas F_{T_b}/k has an exact F distribution with k and $T - 2k$ degrees of freedom. The fundamental drawback of Chow test is that the change point must be specified in advance; however, tests using Chow statistics exist that do not require explicitly stating a specific change point.

To estimate the unknown breakpoint, many authors extended the ideas of the Chow test. For instance, Quandt (1958, 1960) proposed the Supremum Test, a likelihood ratio test statistic for unknown structural break points, while Andrews (1993) provided analogous Wald and Lagrange Multiplier test statistics for it as in equation (3.5).

$$sup\mathrm{F} = \sup_{\underline{t} \le t \le \bar{t}} \mathrm{F}_t \quad (3.5)$$

These tests are computed by calculating Chow statistics for all possible change points in an interval $[\underline{t}, \bar{t}]$ and if any of those statistics exceeds a certain threshold, rejecting the null hypothesis. A formula interface specifies the model to be tested, while the limits $\underline{t}$ and $\bar{t}$ are respectively represented by from and to. To indications of observations, Sample fractions can also be used to specify these two parameters; for example, to take from=0.15 and implicitly to = 0.85. This can be easily performed using the R function *Fstats* from *strucchange* package developed by (Zeileis et al., 2002). While the Sup test is calculated to find the date that maximizes Chow Statistics, which is the most possible breakpoint, the Ave and Exp tests use all Chow Statistics values and are only enlightening about the presence of the break but not its date. The R function $sctest$ (structural change test) from the $strucchange$ package, which

produces an object of class "htest" with the test statistic and the appropriate p value, makes it simple to perform these tests.

3.1.2 CUSUM Test

Another widely used test for model parameters stability is the CUSUM test developed in Brown et al.'s seminal paper (1975), which contains cumulative sums of standardized residuals. This test comes under the class of generalized fluctuation test framework of structural change tests. This class of tests creates an empirical process that captures the variation in residuals or estimates by fitting a model to the provided data. It is possible to compute limits whose probability of crossing under the null hypothesis is α since the limiting processes for these empirical processes are known. If the fluctuation is unreasonably big if the empirical process route surpasses these thresholds, At the significance level α, the null hypothesis should be rejected. Brown et al. (1975) recommended taking into consideration cumulative sums of recursive residuals. Here the test statistic w_t is computed by equation (3.6):

$$w_t = \frac{1}{\hat{\sigma}\sqrt{T-k}} \sum_{i=k+1}^{t} \hat{\varepsilon}_i \tag{3.6}$$

where $\hat{\sigma}$ is defined as in equation (3.7):

$$\hat{\sigma} = \sqrt{\frac{\sum_{t=k+1}^{T} \hat{\varepsilon}_t}{T-k}} \tag{3.7}$$

and $\hat{\varepsilon}_t$ is provided by the series of recursive errors that are corrected for the distortion size as in equation (3.8):

$$\hat{\varepsilon}_t = \frac{y_t - x_t' \hat{\beta}_{t-1}}{\sqrt{1 + x_t'(X_{t-1}'X_{t-1}')^{-1}x_t'}} \tag{3.8}$$

where $\hat{\beta}_{t-1}$ is an estimate of β using data up to $t-1$ while X'_{t-1} is the corresponding matrix of exogenous variables. Under the alternative, the recursive residuals will only have zero mean up to t_0 when there is just one structural change point t_0. Therefore, the route of the process should be close to 0 up to t_0 and leave its mean afterwards. Ploberger & Krämer, (1992) proposed that the common OLS residuals' cumulative sums serve as the foundation for a structural break test. In that case $\hat{\varepsilon}_t$ is characterized as the typical OLS residual as in equation (3.9):

$$\hat{\varepsilon}_t = y_t - x'_t \hat{\beta}_t \tag{3.9}$$

The CUSUM test, according to Hansen (1991), is suitable for evaluating parameter instability in the intercept. Given a formula that describes the models to be tested the *efp* function in R generates an object of class "*efp*" which comprises a type-specific estimated empirical fluctuation process.

When the variability of the empirical process $efp(t)$ exceeds that of the limiting process by a huge margin, the CUSUM test rejects the null hypothesis of no structural break. The suitable boundary $b(t)$, which the limiting process simply passes with a specified probability α, serves as the basis for this comparison. Therefore, it must be decided that the fluctuation is exceptionally large, and at confidence level α, the null hypothesis will be rejected if $efp(t)$ crosses either $b(t)$or $-b(t)$ for any t (Zeileis et al., 2002). The standard boundaries for OLS-based CUSUM and the Recursive CUSUM are as in equation (3.10) and equation (3.1):

$$b(t) = \lambda.(1 + 2t) \tag{3.10}$$

$$b(t) = \lambda \tag{3.11}$$

respectively, where λ is a constant determining the confidence level and it depends on α. As Zeileis et al. (2002) suggest, a traditional significance test may still be necessary even though computing and visualizing the empirical fluctuation procedure with its limits gives and visualizes much of the data. The test statistic reflects what the border

crossing described. Hence the test statistic S_r from equation (3.12) for the residual-based processes is

$$S_r = \max_t \frac{efp(t)}{f(t)} \tag{3.12}$$

where $f(t)$ relies on the boundary's shape, i.e., $b(t) = \lambda \cdot f(t)$.

3.1.3 BAI-PERRON TEST of Multiple Structural Breaks

In many applications, notably macroeconomic time series, the coefficients shift from one stable regression relationship to another, and it is plausible to suppose that there are several structural breaks (m breakpoints) at these locations. As a result, we get m+1 segments with constant coefficients of regression. To that end, Bai & Perron (1998,2003) presented a thorough analysis of several issues in the context of multiple structural breaks, as well as some tests that rule out the existence of structural breaks based on least squares principles. The following model in equation (3.13) of multiple structural breaks with m breaks is taken into consideration by the Bai and Perron approach (m+1 regimes):

$$y_t = x_t'\beta + z_t'\delta_j + u_t, \quad t = T_{j-1} + 1, \dots, T \tag{3.13}$$

which can be simplified as in equations (3.14), (3.15) and (3.16):

$$y_t = x_t'\beta + z_t'\delta_1 + u_t, \quad t = 1, \dots, T_1 \tag{3.14}$$

$$y_t = x_t'\beta + z_t'\delta_2 + u_t, \quad t = T_1 + 1, \dots, T_2 \tag{3.15}$$

$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$

$$y_t = x_t'\beta + z_t'\delta_{m+1} + u_t, \quad t = T_m + 1, \dots, T \tag{3.16}$$

where y_t stands for the dependent variable at time t; x_t $(p \times 1)$ and $z_t(p \times 1)$ denote vectors of covariates, and β and δ_j $(j = 1, \dots, m+1)$ are the corresponding coefficients vectors; u_t is the error term at time t. The breakpoints $(T_1, \dots, T_m)$ are treated as unknown (it is convention that $T_0 = 0 \; and \; T_{m+1} = T$). The intention is to estimate the unknown coefficients of regression model together with the breakpoints that is if T observations on (y_t, x_t, z_t) are obtainable. The above multiple linear models of regression can be expressed in matrix form as in equation (3.17):

$$Y = X\beta + \bar{Z}\delta + U \tag{3.17}$$

where $Y = (y_1, \dots, y_T)'$, $X = (x_1, \dots, x_T)'$, $U = (u_1, \dots, u_T)'$, $\delta = (\delta_1', \delta_2' \dots, \delta_{m+1}')'$, whereas $\bar{Z}$ is the matrix which partitions Z diagonally at $(T_1, \dots, T_m)$, i.e. $\bar{Z} = diag(Z_1, \dots, Z_{m+1})$ with $Z_{\mathrm{i}} = (z_{T_{\mathrm{i}-1}+1}, \dots, z_{T_{\mathrm{i}}})'$. For any partition, the estimates of β and δ_j in equation (3.18) are those that minimize

$$(Y - X\beta - \bar{Z}\delta)'(Y - X\beta - \bar{Z}\delta) = \sum_{i=1}^{m+1} \sum_{t=T_{\mathrm{i}-1}+1}^{T} [\, y_t - x_t'\beta - z_t'\delta_i]^2 \tag{3.18}$$

Given m-partitions $(T_1, \dots, T_{\mathrm{m}})$, the estimates become $\hat{\beta}(\{T_{\mathrm{j}}\})$ and $\hat{\delta}(\{T_{\mathrm{j}}\})$. The coefficients and the m partitions are selected as the global minimizers of the total squared residuals through all partitions and the estimated breakpoints are in equation (3.19):

$$(\hat{T}_1, \dots, \hat{T}_m) = argmin_{T_1, \dots, T_{\mathrm{m}}} S_{\mathrm{T}}(T_1, \dots, T_{\mathrm{m}}) \tag{3.19}$$

where $S_{\mathrm{T}}(T_1, \dots, T_{\mathrm{m}})$ are the sum of squared residuals. The parameter estimates of the regression model are the estimates associated with the m-partition $\{\hat{T}_j\}$ i.e $\hat{\beta} = \hat{\beta}(\{T_{\mathrm{j}}\})$ and $\hat{\delta} = \hat{\delta}(\{T_{\mathrm{j}}\})$. The Bai and Perron test can be performed with the help of the *breakpoints* function R from the *strucchange* package of (Zeileis et al., 2002).

3.2 Autoregressive Conditional Heteroscedasticity (ARCH) Effect Test

The ARCH effect is well-known in econometrics and financial time series. The ARCH idea, created by economist Robert F. Engle in the 1980s, significantly enhanced financial modeling in terms of risk estimation by offering a volatility model that was more akin to actual markets. Autocorrelation in the squared series of residuals, a measure of conditional heteroscedasticity, is referred to as the autoregressive conditional heteroscedastic (ARCH) effect. If ARCH effects are not taken into account, traditional tests for autocorrelation in the conditional mean may be overly rejected (Taylor, 1984) and asymptotic efficiency problems may arise (Engle, 1982). The gold standard for identifying ARCH effects is Engle's ARCH-Lagrange multiplier (ARCH-LM) test. (Sjölander, 2011).

Engle's Lagrange multiplier test is calculated using an auxiliary test regression, and it follows the following procedure:

1. Determine the optimal fitting autoregressive model $AR(q)$ in equation (3.20)

$$y_t = \beta_0 + \beta_1 y_{t-1} + \cdots + \beta_q y_{t-q} + \varepsilon_t = \beta_0 + \sum_{i=1}^{q} \beta_i y_{t-i} + \varepsilon_t \qquad (3.20)$$

2. Obtain the squared residuals $\hat{\varepsilon}^2$ and regress them on a lagged and constant squared residuals up to order q:as in equation (3.21)

$$\hat{\varepsilon}_t^2 = \hat{\alpha}_0 + \sum_{i=1}^{q} \hat{\alpha}_i \hat{\varepsilon}_{t-i}^2 \qquad (3.21)$$

where q is the lags length of $ARCH$

3. The absence of ARCH effects is the null hypothesis, while the alternative hypothesis is that at least one of the estimated α_i coefficients is significant and hence the presence of the ARCH effect. The ARCH-LM test statistic is computed

by multiplying the number of observations in the sample of T residuals under the null hypothesis of no ARCH errors and the R^2 from the regression in equation (3.21). The LM test statistic is asymptotically distributed as χ^2distribution with q degrees of freedom; hence, if the calculated LM test statistic is greater than the Chi-square table value, we reject the null hypothesis and conclude there is an ARCH effect in the ARIMA model. The ARCH-LM test can be conducted using arch.test function in R from aTSA package.

3.3 Forecasting Models

In this section, implemented forecasting methods to generate forecasts in the cases of one or multiple structural breaks are discussed one by one. Individual forecasting models, namely ETS, ARIMA, ARIMA-ARCH and ANN, are first discussed, and then forecast combination models comprising simple forecast combination models and regression-based combination models are expressed.

3.3.1 Single Models

The individual methods we employed in this thesis are two linear (ARIMA and exponential smoothing) models and two non-linear models (ARIMA-ARCH and artificial neural network). These methods are described below.

3.3.1.1 Exponential Smoothing (ETS) Model

In the 1950s, the exponential smoothing class of models was introduced (Winters, 1960; Holt, 2004) and by using a weighted average of its previous values and giving more weight to recent observations, they can create time series forecasts. These models are made up of state equations that define how states (level, trend, and seasonal) vary over time and observation/ measurement equations that characterize the observed data, which is why they are called state-space models. Exponential smoothing models are labeled with the notation ETS for (Error, Trend, and Seasonal). The ETS models have got variations because of the trend and seasonal combinations. The trend component

may be none (N), may be additive damped (Ad) or additive (A), while seasonal component can be none (N), multiplicative (M), and additive (A), yielding nine variants of exponential smoothing models (Hyndman & Athanasopoulos, 2018).

The component form of the $ETS(A,N,N)$ model with additive errors in equation (3.22):

$$\text{Forecast equation} \qquad \hat{y}_{t+1|t} = \ell_t \tag{3.22}$$

$$\text{Smoothing equation} \qquad \ell_t = \alpha y_t + (1-\alpha)\ell_{t-1} \tag{3.23}$$

By re-arranging the smoothing equation (3.23), we get equation (3.24)

$$\ell_t = \ell_{t-1} + \alpha(y_t - \ell_{t-1}) = \ell_{t-1} + \alpha e_t \tag{3.24}$$

where $\ell_t$is the estimated level, $\hat{y}_{t+1|t}$is one step-ahead forecast for time $t+1$, which comes from the weighted average of all historical data, while $0 \leq \alpha \leq 1$ is the smoothing parameter and $e_t = y_t - \ell_{t-1} = y_t - \hat{y}_{t|t-1}$ is an error at time t.

Other ETS models can be written in a similar fashion for each of the exponential smoothing methods. An automated selection procedure is utilized to identify the exponential smoothing models by using ets R function from $forecast$ package developed by Hyndman and Khandakar (2008). This function automatically identifies which model best suits the given time series, estimates the model parameters, and returns information about the fitted model. It can use all information criteria, but the bias-corrected Akaike Information Criterion (AICc) is the default information criterion to select an appropriate model. The Akaike Information Criterion is calculated as in equation 3.25:

$$AIC = -2\log(L) + 2k \tag{3.25}$$

where k denotes number of parameters in the model and L represents estimated log-likelihood.

3.3.1.2 Autoregressive Integrated Moving Average (ARIMA) Model

One of the most extensively employed methods for time series forecasting is the Autoregressive Integrated Moving Average (ARIMA) model, and was first suggested by (Box and Jenkins, 1970). This model is a generalization of an ARMA model. ARIMA(p,d,q) is a non-seasonal ARIMA model with non-negative parameters p, d, and q, where p denotes number of time lags referred to as an order of the autoregressive part, d is the level of differencing, and q represents order of the moving-average part.

The ARIMA models transform the non-stationary series to a stationary series through a sequence of steps of differencing. A time series y_t denote the *integrated of order d*, if $\nabla^d y_t$ in equation (3.26) is stationary and

$$\nabla y_t = y_t - y_{t-1} \tag{3.26}$$

where y represents the time series and t is a time index.

After we transform the time series into stationary, the estimation is done like in equation (3.27)

$$\hat{y}_t = \mu + \sum_{i=1}^{p} \phi_i \, y_{t-i} + \sum_{i=1}^{q} \theta_i \, \varepsilon_{t-i} + \varepsilon_t \tag{3.27}$$

where ϕ_i denotes the coefficients of AR terms of order p and θ_i denote the coefficients of the MA terms of order q, and ε is a random term that is white noise, and μ is a constant term. The model parameters are calculated using maximum likelihood estimation (MLE).

For the best ARIMA model selection, we relied on $auto.arima\ R$ function from $forecast$ package developed by Hyndman and Khandakar (2008). This function identifies the best ARIMA model based on either the Bayesian information criterion (BIC), the Akaike information criterion (AIC), and the bias-corrected Akaike criterion (AICc) value. Within the given order limitations, this R function explores several models. The BIC is calculated as in equation (3.28)

$$BIC = -2\log(L) + klog(T) \tag{3.28}$$

where k stands for the total of parameters in the model, L is the estimated log-likelihood and T is the number of observations in the time series.

3.3.1.3 Combined ARIMA and ARCH (ARIMA - ARCH) Model

The error structures in ARIMA models are assumed to be generated from linear combinations of the innovations, with the innovations being, i.i.d. normal random variables, but that is not always the case. The variance fluctuates widely with time in many financial and economic time series, and ARIMA models cannot capture this behavior. To address these conditions, ARCH (autoregressive conditionally heteroscedastic) proposed by Engle(1982) model and GARCH (generalized ARCH) by Bollerslev (1986) have been constructed. For this reason, the ARIMA and GARCH models are linked. While the ARIMA part of the model estimates the mean, the conditional variance included in the residuals of the ARIMA is estimated by a GARCH part of the model. The estimation based on the combined model is as in equation (3.29):

$$\hat{y}_t = \mu + \sum_{i=1}^{p} \phi_i\, y_{t-i} + \sum_{i=1}^{q} \theta_i\, \varepsilon_{t-i} + \varepsilon_t \tag{3.29}$$

$$\varepsilon_t = \sigma_t a_t \tag{3.30}$$

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 \tag{3.31}$$

where a_t in equation (3.30) represents white noise ω and α_1 in equation (3.31) are the model parameters. Substituting for σ_t^2, we get as:

$$\varepsilon_t = a_t\sqrt{\omega + \alpha_1\varepsilon_{t-1}^2} \tag{3.32}$$

Then ε_t in equation (3.32) is ARCH (1) model, which stands for autoregressive conditionally heteroscedastic. An ARCH(q) process is provided in equation (3.33):

$$\varepsilon_t = a_t\sqrt{\omega + \sum_{i=1}^{q}\alpha_i\varepsilon_{t-i}^2} \tag{3.33}$$

GARCH(q,p) can be formulated if we define $\varepsilon_t = \sigma_t a_t$ where a_t white noise and σ_t^2 is given by:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q}\alpha_i\varepsilon_{t-i}^2 + \sum_{j=1}^{p}\beta_j\sigma_{t-j}^2 \tag{3.34}$$

where α_i and β_j in equation (3.34) are parameters of the model. The forecast is calculated by combination of output of the ARIMA part and the ARCH/GARCH part of the model. ARIMA - ARCH model can be implemented using the R functions of *ugarchspec*, *ugarchfit*, and *ugarchforecast* from *rugarch* package (Ghalanos, 2015).

3.3.1.4 Artificial Neural Network (ANN) Model

A multilayer perceptron (MLP) neural networks (also well-known as artificial neural networks or simply neural nets) are data driven machine learning algorithms commonly used for forecasting, particularly in financial forecasting (Shmueli & Jr., 2016). This machine learning technique first became popular in the 1990s, and Shmueli & Jr. (2016) argue that among machine learning algorithms, neural nets are the most popular for time series forecasting. The idea behind neural networks is to

mimic human brain functionality in capturing a complex relationship between output and inputs by creating layers of derived variables.

A neural network connects the predictors and the outcome through a series of layers. In each layer, some operation is performed on the input data to produce an output. The output of the previous layer becomes the input for the next layer. As shown in Figure 3.1, the structure of ANN is characterized by three layers:

1. An *input layer* that accepts the input values of the predictors and lagged terms in the time series situation.
2. *Hidden layers* that generate derived variables a hidden layer receives inputs from the preceding layer, which is the input layer, performs some calculations on those inputs based on some activation function, and creates an output.
3. An *output layer* receives input from the latest hidden layer and predicts future values.

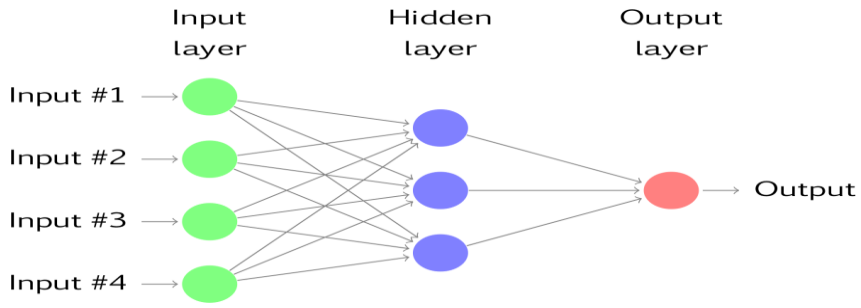


Figure 3.1 Artificial Neural Networks Architecture

Each layer contains nodes denoted by circles. While input layer nodes are the original predictors, hidden layer nodes are the weighted sum of the inputs to which some monotone function, called an activation function, is applied. The Common functions include linear, and s-shaped functions such as the logit and hyperbolic tangent or, in other words, sigmoid functions.

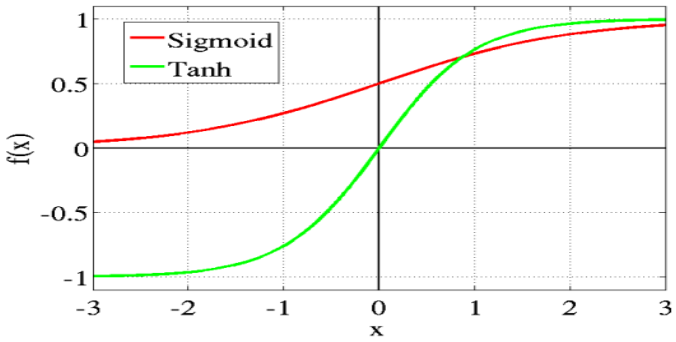


Figure 3.2 Activation functions (tanh vs logistic)

ANNs are versatile computational structures that may be utilized to represent a wide variety of nonlinear issues; hence they can be used to approximate nonlinearities in the series. The fact that ANN models are general approximators that can precisely estimate many functions gives them a substantial advantage over other kinds of nonlinear models. Their strength stems from the parallel processing of information from data. The process of building model does not require any earlier assumptions about the model format. Instead, the model is mainly determined by the attributes of the data under consideration (Adhikari et al., 2003).

In time series, the relationship between the inputs $(y_{t-1}, y_{t-2}, \dots, y_{t-p})$ and the output (y_t) can be represented by the mathematical representation in equation (3.35):

$$y_t = \alpha_0 + \sum_{j=1}^{q} \alpha_{jg} \left(\beta_{0j} + \sum_{i=1}^{p} \beta_{ij} y_{t-i} \right) + \varepsilon_t, \tag{3.35}$$

where α_j $(j = 0,1,2, \dots, q)$ and β_{ij} $(i = 0,1,2, \dots, p; j = 1,2, \dots, q)$ represent the model parameters which are often called the connecting weights; p denotes the input nodes while q stands for number of hidden nodes. The logistic function is often used as the hidden layer transfer function, that is as in equation (3.36),

$$g(x) = \frac{1}{1 + e^{-x}} \tag{3.36}$$

Hence, from previous observations, the ANN model conducts mapping of a non-linear functional $(y_{t-1}, y_{t-2}, \dots, y_{t-p})$ to the future y_t, as in equation (3.37).

$$y_t = f\left(y_{t-1}, y_{t-2}, \dots, y_{t-p}, w\right) + \varepsilon_t \tag{3.37}$$

Since we are using the lagged values from the time series as inputs to a neural network in the same way that lagged values are used in a linear autoregression model, it is called a neural network autoregression or NNAR model. In this paper, we employed feed-forward networks based on $NNAR(p, k)$ where p indicate lags as inputs and k nodes in the hidden layer (Hyndman & Athanasopoulos, 2018).

Neural network autoregression models were identified using "$nnetar$" function from "$forecast$" package in R produced by Hyndman (2012). This function automatically fits ANN model to the given time series data using lagged values of the series as inputs, so it is a nonlinear autoregressive model.

3.3.2 Forecast Combinations

Forecast combination has been a well-established strategy for improving forecasting accuracy since the seminal paper of Bates and Granger (1969). In the theoretical and empirical studies, numerous combination procedures have been proposed, but so far there are no theoretical foundations for the ideal technique to combine diverse forecasts; instead, a lot relies on the features of the available data. Regularly used approaches for combining forecasts include simple combinations that ignore forecast error correlations and more sophisticated combination schemes to estimate optimal combination weights such as regression - based and eigenvector weighting approaches. Even Andrawis et al. (2011) proposed combining the combined forecasts and called that strategy hierarchical forecast combination.

To illustrate notations, we denote $F_{T\times P}$ as the forecast's matrix with $T \times P$ dimension where T stands for the rows and P represents for the number of columns.

We also denote f_i as the forecast attained from model i, $i \epsilon \{1 \dots P\}$. The associated weight with that forecast in the global combined forecast is represented as w_i, while the combined forecast is denoted as as f^c.

3.3.2.1 Simple Forecasts Combination Methods

We start with some simple ways to combine forecasts. They are simple in that they ignore correlations between forecast errors and that we do not need to estimate the weight every forecast to be assigned for in the total contribution.

3.3.2.1.1 Simple Average Based Forecast Combination

This is the most intuitive approach for forecast combination perhaps because its simplicity (Weiss et al., 2019). This is just the arithmetic average of the single forecasts, and it is given in equation (3.38),

$$f^c = \frac{1}{P}\sum_{i=1}^{P} f_i \tag{3.38}$$

where f^c is the combined forecast, P denotes number of to be combined single models and f_i as is the forecast obtained from model i.

This equal weighting strategy which is also called “forecast combination puzzle” as coined by Stock and Watson (2004) may in some situations emerge as the better than the sophisticated forecast combination approaches as Clemen (1989) argues. A rigorous empirical and theoretical explanation of why this innocent approach outperforms more sophisticated techniques is provided by Claeskens et al. (2016) and Smith and Wallis (2009).

3.3.2.1.2 Median Based Forecast Combination

This is another simple and location measure combination method that robust to outliers. This technique uses the median of the forecasts component and can be appropriate for some applications. Simple average is not an appropriate combination strategy if some of the forecasts component are biased as Palm and Zellner (1992) suggested, and such a case median combination becomes handy. The combined forecast using median method is given in equations (3.39) and (3.40):

- For odd P:

$$f^c = f_{(\frac{p}{2}+0.5)} \tag{3.39}$$

- For even P:

$$f^c = \frac{1}{2}(f_{(\frac{p}{2})} + f_{(\frac{p}{2}+1)}) \tag{3.40}$$

3.3.2.1.3 Bates/Granger Forecast Combination

The influential paper of Bates and Granger (1969) proposed the idea of combining forecasts. The estimated mean squared prediction error matrix is used in this method to determine the combination weights, and therefore the combined forecast is computed as in equation (3.41),

$$f^c = \sum_{i=1}^{P} f_i' \times \frac{\hat{\sigma}_i^{-2}}{\sum_{j=1}^{P} \hat{\sigma}_j^{-2}} \tag{3.41}$$

where $\hat{\sigma}_i^{-2}$ is the estimated mean squared prediction error of model i. Even though this was derived under the assumption of uncorrelated forecasts, this method can work well in practice.

3.3.2.2 Regression-based Forecast Combination Methods

In this thesis, we implemented two regression-based combination models namely Ordinary Least Squares (OLS) regression-based combination method, and Least Absolute Deviation (LAD) regression-based. We consider both time-invariant combination weighting (static version) and time-varying combination weighting (dynamic version) strategies in both regression-based combinations. That means we considered the regression-based methods a total of four.

The dynamic version of regression based forecast combination allows for time-varying weights. It builds on the idea of time series-cross validation by taking the training set as starting point. The computation is done in an expanding window fashion rather than a single in- and out-of-sample split.

3.3.2.2.1 Ordinary Least Squares (OLS) Regression-based Forecast Combination

One of the sophisticated rules of combining forecasts is OLS. The idea of regression-based forecast combination was established by Crane and Crotty (1967) and later Granger and Ramanathan (1984) expanded its usage more successfully. This approach assumes that the forecasts combined are linear function of the single forecasts, and we obtain weights by regressing actual values on individual forecasts as in equation (3.42).

$$y = \alpha + \sum_{i=1}^{P} w_i f_i + \varepsilon, \tag{3.42}$$

The combined forecast can be computed as in equation (3.43):

$$f^c = \hat{\alpha} + \sum_{i=1}^{P} \hat{w}_i f_i, \tag{3.43}$$

The excellent feature of the OLS combinations is that, because of the intercept in the equation (3.43), the combined forecast is unbiased even if one of the individual forecasts is biased. A drawback could be that the method does not restrict the combination weights; hence, they do not add up to 1 and there is possibility to be negative), which complicates interpretation.

3.3.2.2.2 Least Absolute Deviation (LAD) Regression-based Forecast Combination

Instead, the sum of squared errors is minimized when estimating the coefficient in equation (3.43), we may prefer those coefficients to be estimated by minimizing the absolute sum of squares. The LAD mode is less sensitive to outliers and can retain its stability once the component forecasts are highly correlated in contrast to OLS. The LAD strategy of combination is be preferred over OLS in the above situations (Nowotarski et al., 2014).

The combination schemes used in this study can be implemented with the help of R functions from ForecastComb package developed by (Weiss et al., 2019) as shown in Table 3.1.

Table 3.1 R Functions for Forecasts Combination and their Description

Function	**Description**
Simple Forecast Combination Functions	
comb_SA	Simple Average forecast combination
comb_MED	Median Forecast Combination
comb_BG	Bates/Granger (1969) forecast combination
Regression-Based Forecast Combination Functions	
comb_OLS	Ordinary Least Squares (OLS) forecast combination
comb_LAD	Least Absolute Deviation (LAD) forecast combination
rolling_combine	Computes the dynamic version of the combined forecast (time-varying combination weights). Inside this function we need to specify the combination method that should be used (OLS or LAD in this paper).

3.4 Forecasting Accuracy Metrics

Several methods are available in the literature for measuring the accuracy of the forecasts and comparing the model's forecasting performance. Among those, two metrics are used in this study to compare the forecast methods and assess the accuracy of out-of-sample projections against the reserved observed values in the evaluation sample termed the test set.

3.4.1 Root Mean Squared Forecast Error (RMSE)

Root Mean Squared Error (RMSE) is one of the commonly employed accuracy measures based on squared errors and is computed as in equation (3.44):

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{T}(\hat{y}_{t+h|t} - y_{t+h})^2}{T}} \tag{3.44}$$

where $\hat{y}_{t+h|t} - y_{t+h}$ states forecast error. It is also given training data by $\{y_1, \ldots, y_t\}$ and the test data is given by $\{y_{t+1}, y_{t+2}, \ldots\}$.

Root mean squared error (RMSE) is widely used to compare forecast methods applied to an individual time series or several time series having the same units (Hyndman & Athanasopoulos, 2018). Since our time series data simulations are replicated based on the same data generation process, RMSE is an appropriate measure for comparison of the predictive performance of the considered forecast methods.

3.4.2 Mean Absolute Percentage Error (MAPE)

The mean absolute percentage error (MAPE) can measure the forecast accuracy based on percentage errors. Unlike RMSE, it is easy to interpret, scale-independent, and calculated as follows. Let y_t and $\hat{y}_t$ represent the actual and forecasted values at the point t, respectively. Then, MAPE is calculated as in equation (3.45):

$$\text{MAPE} = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{y_t - \hat{y}_{t+h}}{y_t}\right| \times 100 \tag{3.45}$$

where t stands number of data points, y_t is the actual value, and $\hat{y}_t$ is the forecast value.

Because of its scale independence and interpretability, MAPE become one of the most widely used metrics for forecast accuracy (S. Kim & Kim, 2016). Besides the popularity and interpretability of MAPE, it has the drawback of producing infinite or undefined values in the case when the actual values are zero or close to zero. This is a problem we met in the simulation part of this study and hence we utilized MAPE only in the empirical analysis which we did not face that problem in it. Both RMSE and MAPE can be calculated using the accuracy R function in forecast package.

CHAPTER FOUR
IMPACT OF STRUCTURAL BREAK LOCATION WITH ARCH INNOVATIONS ON FORECASTING ACCURACY: TRADITIONAL METHODS VERSUS ARTIFICIAL NEURAL NETWORK

Since forecasting future values is fundamental for researchers, investors, practitioners, etc., obtaining accurate predictions is critical in time series analysis. The accuracy is reliant on good modelling and good-quality data. The latter is affected by unusual observations, changes over time, missing data, and structural breaks among others. Economic crises such as the 2007–2008 financial crisis and COVID-19 are the major causes of data instability.

As we discussed in Chapter two of this thesis, despite the extensive literature on the structural break and its impact on forecasting, little emphasis is paid to how the structural break location in time series affects predicting accuracy. Questions such as whether forecasting accuracy decreases as the structural break approaches the end of the sample series remain unaddressed. Another concern usually experienced in time series is the problem of conditional heteroscedasticity. Assuming conditional homoscedasticity, especially in the case of structural breaks in forecasting macroeconomic conditions or financial time series may seem restrictive. Therefore, this chapter focuses on how structural breaks in conditional heteroscedastic financial and macroeconomic data affect forecasting accuracy on short and long-term horizons. More specifically, the impact of the structural break location on forecast accuracy is assessed by placing the structural break in different locations in the time series and measuring forecast error in each case. In addition, several magnitudes of breaks are examined to check if the size of the break affects the forecast accuracy.

In this chapter, we compared the predictive performance of four models, namely Exponential Smoothing (ETS), Autoregressive Integrated Moving Average (ARIMA), Autoregressive Integrated Moving Average -Autoregressive Conditional Heteroskedastic (ARIMA – ARCH), and Artificial Neural Networks (ANN). This means we combined linear and non-linear forecasting methods on one side and

classical and artificial neural networks on the other. These methods are implemented on simulated data and empirical data of daily close prices of brent oil and Turkish Lira – USD exchange rates. The metrics used to compare the performances of the forecasting models are the Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE). The Chow, CUSUM, and Bai and Perron tests are applied to examine the presence of structural break (s) and estimate the date (s) of the break.

4.1 Data Description and Simulation Designs

In this chapter, both simulated and empirical data are used. The empirical data used here are daily Turkish Lira against US Dollar (USD – TL) exchange rate and Brent oil prices, as discussed in Section 4.3. The simulations are carried out in a way that generated data contain conditionally – heteroscedastic errors and structural breaks in predefined test locations, i.e., first quartile of the sample series (Q1), second quartile (Q2), or third quartile sample series (Q3). Furthermore, different sizes of structural breaks are used. The procedure of simulation is detailed in Section 4.1.1.

4.1.1 Data Generating Process in Simulation

In generating data, it is considered the design of the AR(2) process in equation (4.1) with autoregressive conditionally heteroscedastic errors ARCH(2),

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \tag{4.1}$$

$$\varepsilon_t = \sigma_t a_t \tag{4.2}$$

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 \tag{4.3}$$

where a_t in equation (4.2) is white noise and ω, α_1 and α_2 in equation (4.3) are the parameters of the variance model. Substituting for σ_t^2 we receive:

$$\varepsilon_t = a_t\sqrt{\omega + \alpha_1\varepsilon_{t-1}^2 + \alpha_2\varepsilon_{t-2}^2} \tag{4.4}$$

with the help of equation (4.4), $AR(2) + ARCH(2)$ process can be generated.

The parameters of both the AR part and ARCH part of the data generating process mentioned above were specified as: $\mu = 5, \phi_1 = 0.65, \phi_2 = 0.25, \omega = 0.2, \alpha_1 = 0.45, \alpha_2 = 0.3$. These values were chosen arbitrarily. The graphical visualization of the simulated data is shown in Figure 4.1 for illustration purposes.

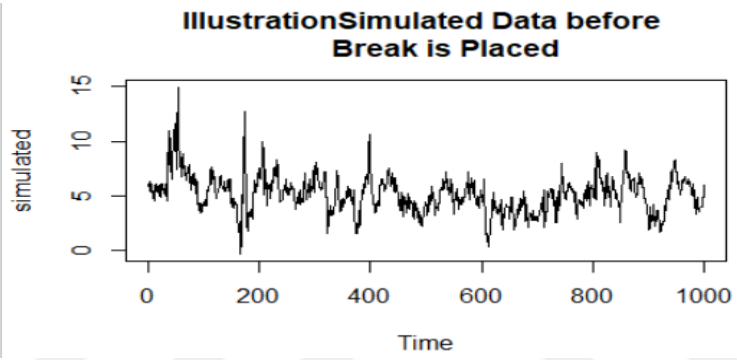


Figure 4.1 Illustration of Simulated Data Prior Structural Break

To examine the effect of structural break location on forecasting accuracy, we first consider the case (case I) of forecasting with stable data, i.e., before the structural break is placed. We then construct three unstable cases by allowing a one-time permanent break in the mean of the series in each case. The breaks were positioned at 25%th observation termed as a break at the first quartile(Q1), 50%th observation as the second quartile (Q2) and 75%th observation as the third quartile (Q3) of the training sample (n). We let the structural break take the multiplicative form, which means that if the pre-break mean is μ, then the post-break value is $\delta\mu$, where δ is a break size parameter. In each situation we considered three different break sizes namely break of small magnitude ($\delta = 3$), medium break size ($\delta = 5$) and large break size ($\delta = 10$). For an illustration purpose of the location of the structural break and break sizes see Figure 4.2.

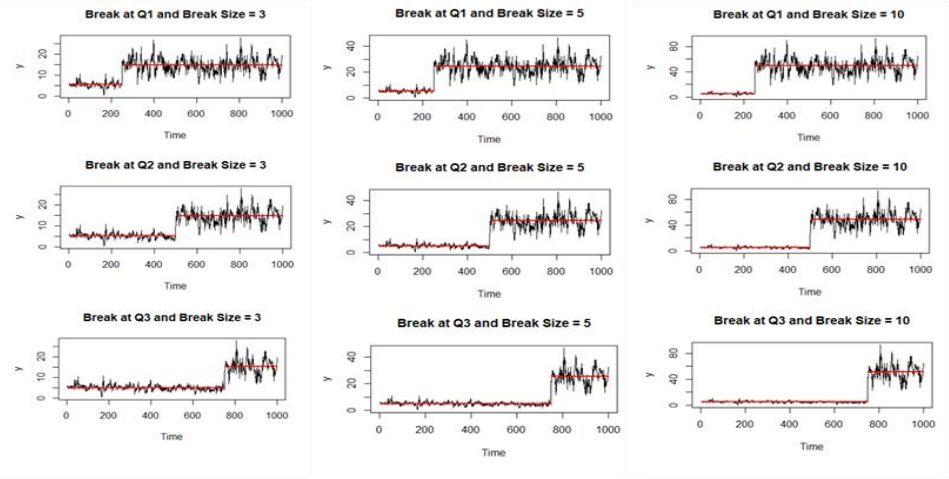


Figure 4.2 Illustration of Break Locations and Break Sizes of Simulation

As shown in Table 4.1, the total sample size, $T = n + h$ is 1000. To investigate if the choice of the horizon size has an impact on forecasting results, we reserve the first 990, 995 and 997 (n = 990, n = 994 and n = 997) observations as the training sample and the rest as the prediction sample (h = 10, h = 6 and h = 3) in three separate experiments in each case. The simulation process is conducted with 5000 times replications.

Table 4.1 Summary of the Simulated Data Procedures and Train-Test Sets

Case	**Break Size**	**Break Location**	**Sample Size**	**Training Set(size)**	**Test set (size)**
Case I	-	-	1000	990, 994, 997	10, 6, 3
Case II	3	Q1 (25%), Q2 (50%), Q3 (75%)	1000	990, 994, 997	10, 6, 3
Case III	5	Q1 (25%), Q2 (50%), Q3 (75%)	1000	990, 994, 997	10, 6, 3
Case IV	10	Q1 (25%), Q2 (50%), Q3 (75%)	1000	990, 994, 997	10, 6, 3

In each case, to evaluate and compare predictive performance, we generate forecasts using four methods: ARIMA, ETS, ARIMA − ARCH, and ANN. Root mean squared forecast error (RMSE) is used to assess their predictive performance.

4.2 Simulation Results

Simulation results are presented in four cases. The first case is the stable data where there is no structural break while the other three cases are based on the break size and in each of the three break locations are considered. The hyperparameter values of the ANN based on the RMSE metric are given in Table 4.2.

Table 4.2 Hyperparameter Values

Break Size	**No. of optimal lags (inputs)**	**No. of hidden layers**	**No. neurons in the hidden layer**	**Model structure notation**
No break	4	1	3	NNAR (4,3)
(δ) = 3	8	1	5	NNAR (8,5)
(δ) = 5	9	1	5	NNAR (9,5)
(δ) = 10	6	1	4	NNAR (6,4)

4.2.1 Case I: Forecasts of Stable Simulated Data

Simulation results for the stable case are shown in Table 4.3. We can see from Table 3 that all the models are performing well when there is no structural break as the RMSE values are small. It can also be seen that the combined (ARIMA-ARCH) model outperforms all the classical ARIMA, Exponential Smoothing (ETS), and even the Artificial Neural Network (ANN) models across all considered test horizons. At the same time, ANN is the next best model.

Table 4.3 RMSEs Values of Stable Simulated Data

Model	**h = 3**	**h = 6**	**h = 10**
ARCH	0.773	0.941	1.083
ARIMA	0.791	0.975	1.147
Exp	0.818	1.017	1.216
ANN	0.785	0.961	1.108

The ARIMA and exponential smoothing methods are less performing models in this case. This was expected because they failed to capture the nonlinearity of the process under consideration. RMSEs increase monotonically as the horizon size increases in all models.

4.2.2 Case II: Forecasts when Break size (δ) = 3

Simulation results for case II where the structural break size is relatively small (δ = 3) are presented in Table 4.4. It can be noticed from the table that the RMSEs of all the models for all test horizons increased in comparison with the RMSEs in Table 3 of stable data. It is also clear that ANN outperforms all the other models across all considered test horizons if the break is in Q1 or in Q2. ARIMA takes over the ANN when the structural break is in the third quartile (Q3). Moreover, RMSEs increase monotonically as the horizon size increases in all models. RMSEs also slightly increase as the break moves from Q1 through Q2 to Q3.

Table 4.4 RMSE Values When Break Size (δ) = 3

Break size (δ) = 3	**Structural Break at Q1**			
	Model/Horizon	**h = 3**	**h = 6**	**h = 10**
	ARCH	2.411	2.990	3.559
	ARIMA	2.407	2.963	3.501
	Exp	2.460	3.058	3.654
	ANN	2.380	2.905	3.392
	Structural Break at Q2			
	ARCH	2.419	3.010	3.597
	ARIMA	2.414	2.982	3.520
	Exp	2.465	3.060	3.661
	ANN	2.405	2.946	3.454
	Structural Break at Q3			
	ARCH	2.461	3.106	3.786
	ARIMA	2.432	3.033	3.591
	Exp	2.459	3.104	3.693
	ANN	2.464	3.071	3.646

4.2.3 Case III: Forecasts when Break size (δ) = 5

Simulation results for case III where the structural break size is relatively medium (δ=5) are reported in Table 4.5.

Table 4.5 RMSE Values when Break Size (δ) = 5

Break size (δ) = 5	Structural Break at Q1			
	Model/Horizon	**h = 3**	**h = 6**	**h = 10**
	ARCH	4.041	5.024	5.996
	ARIMA	4.008	4.941	5.841
	Exp	4.099	5.094	6.089
	ANN	3.966	4.830	5.661
	Structural Break at Q2			
	ARCH	4.056	5.086	6.074
	ARIMA	4.036	5.032	5.944
	Exp	4.087	5.164	6.142
	ANN	4.042	5.016	5.859
	Structural Break at Q3			
	ARCH	4.150	5.237	6.401
	ARIMA	4.080	5.095	6.035
	Exp	4.104	5.173	6.158
	ANN	4.090	5.086	5.982

RMSEs increase monotonically with the increase of break magnitude and horizon size in all models. RMSEs also marginally increase as the break moves from Q1 through the Q2 to the Q3. For horizons 6 and 10, ANN is shown to be the best choice wherever the break is located. But for horizon 3, ANN is best if the break is at Q1, while if the break is at Q2 and Q3, ARIMA gives better results than ANN.

4.2.4 Case IV: Forecasts when Break size (δ) = 10

Simulation results for case IV where the structural break size is relatively large (δ = 10) are reported in Table 4.6. Same to Case II and III, RMSEs increase with the break size, horizon size, and location of the structural break.

Table 4.6 RMSE Values When Break Size (δ) = 10

Break size (δ) = 10	Structural Break at Q1			
	Model/Horizon	**h = 3**	**h = 6**	**h = 10**
	ARCH	8.101	10.086	12.047
	ARIMA	8.020	9.890	11.719
	Exp	8.177	10.181	12.171
	ANN	7.936	9.626	11.284
	Structural Break at Q2			
	ARCH	8.179	10.272	12.290
	ARIMA	8.103	10.141	11.951
	Exp	8.211	10.358	12.318
	ANN	8.067	9.927	11.583
	Structural Break at Q3			
	ARCH	8.504	11.137	13.465
	ARIMA	8.193	10.445	12.148
	Exp	8.228	10.640	12.365
	ANN	8.170	10.378	12.012

On the other hand, regardless of the horizon size and the location of the break, ANN is the best performing model among the applied models when the break size is large (δ=10).

4.2.5 Summary of Simulation Results

In the simulation part of this paper, we examined how the structural break affects forecasting accuracy. In doing so we investigated three different test locations (Q1, Q2, Q3) for the structural break, three break sizes (δ = 3, δ = 5, δ = 10), and three forecast horizons (h = 3, h = 6, h = 10). RMSE is used for the comparison tool. It is found that:

1. The results of the stable case give good guidance for researchers showing how is difficult to analyze financial and economic time series despite the non-existence of structural break in the series. Table 4.3 shows the performance of forecasting with stable data. According to the results in Table 4.3, the ARIMA and Exponential models are not performing well. It can be explained by the fact that the data under study is generated from a non-linear process. Since the asymmetry

and nonlinearity properties exist naturally in financial and economic time series data regardless of the presence of structural break, it is proposed that ARIMA and Exponential Smoothing models should be used cautiously in non-linear processes.

2. For the case of relatively small break magnitude ($\delta = 3$), ANN is superior to all the other models across all considered test horizons if the break is in Q1 or Q2. ANN loses its dominance and ARIMA overtakes when the structural break is in Q3. The failure of ANN when the break size is too small and the structural break is in the third quartile can be explained as ANN models associated with the testing part of the series (Soy Temür & Yıldız, 2021), the effect of structural break which is in the testing part may have a further negative impact on the performance of ANN. This could be the first reason for this result. The shifting of the series because of the structural break in the last quartile, the non-linear behavior of economic time series could turn over to linear behavior. Hence, ARIMA generates better performance than ANN, especially when the researchers deal with small size structural breaks. This could be the second reason for result 2.

3. In the case of comparatively medium break size ($\delta = 5$) in Table 4.5, ANN is the best for horizons ($h = 6$) and ($h = 10$) wherever the break is located. For horizon ($h = 3$), ANN is still the best if the break is at Q1, while ARIMA gives better results than ANN if the break is at Q2 and Q3. Although ANN models are associated with the testing part of the series, it is not affected by the location of the structural break, especially for Q3 which its size is 5. On the contrary the case of breaks size 3, ANN recovers itself in the long-term horizon. On the other hand, ARIMA is a univariate model and is also known as a method that gives very successful results in short-term forecasts (Soy Temür & Yıldız, 2021). ARIMA models need the stability of the data as it requires basic statistical properties such as mean, variance, covariance, or autocorrelation to be constant over time periods (Babu & Reddy, 2014). So, the behavior of the financial time series may change around the statistical properties (mean, variance, etc.) because of the larger break size of the series. Therefore, the performance of ARIMA for the long-term horizons is not good as the short-term horizon in the presence of structural breaks.

This could mean that ANN gives better results for the longer horizons, but ARIMA's performance is getting worse as break size and horizon increase. From the point of economical view, making forecasts for the long term is more valuable but riskier. This is because the economic time series data has dynamic properties and may have been affected by sudden changes, so the researcher must keep in mind those bad events. As a result, researchers should prefer ANN for the larger size structural break in the series.

4. In the case of a relatively large break size ($\delta = 10$), regardless of the horizon size and the location of the break, ANN is the best performing model among the applied models. The results for the largest break size ($\delta = 10$) supported the interpretations of a relatively smaller break size ($\delta = 5$).

5. RMSEs increase with horizon length, break size, and the break location moving from the first quartile, second quartile to the third quartile of the data. All the models have the smallest RMSEs when the horizon is short ($h = 3$), the break size is small ($\delta = 3$) and the break is located at the beginning of the training set (Q1). The largest RMSEs are found when the horizon is long ($h = 10$), the break size is large ($\delta = 10$) and the break locates close to the end of the training set (Q3).

For guidance:

When the break size is 3 ($\delta=3$, around the mean) and in the last quartile, ARIMA which is a mean-based method should be preferred in comparison with ANN. When the break size is medium as ($\delta=5$) and the break locates in Q2 and Q3, ARIMA should be used for the short-term horizons. Except for conditions in which ARIMA is performing well, we propose using the ANN in any other case.

4.3 Empirical Applications

In addition to the simulation evidence, we applied considered methods to forecast the daily Turkish Lira against US Dollar (USD – TL) exchange rate and Brent oil

prices. We selected these two financial data because they exhibit typical structural breaks we are going to investigate in this study. The Figures 4.3 and 4.4 show the time series data of the USD – TL exchange rate and Brent oil prices and the confidence interval of potential structural breaks are shown. The graphs show extracted prices before it is divided into sub-windows based on the interested test locations of structural breaks.

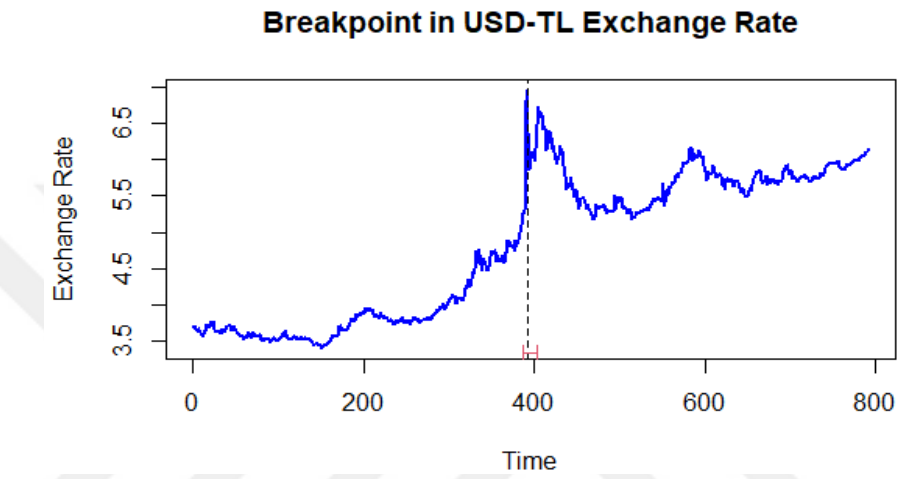


Figure 4.3 USD - TL Exchange Rate

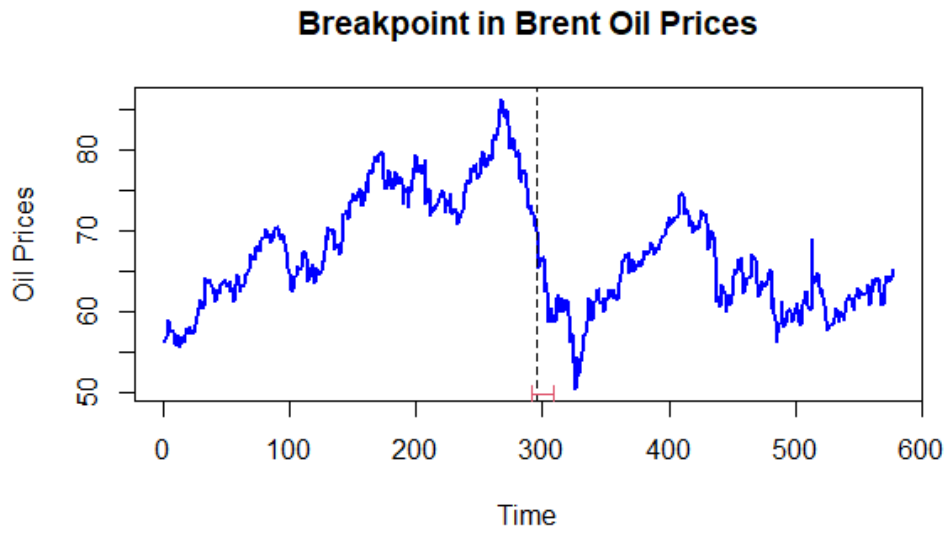


Figure 4.4 Brent Oil Prices

We extracted data by windowing according to the time periods where structural breaks were found. It is extracted in such a way that the break location fits the interested test locations (Q1, Q2, Q3) in line with the simulations section.

Both Brent Oil Prices and the Turkish Lira - US Dollar exchange rate experienced sudden shocks and fluctuations in recent years, making it challenging to forecast their future values. The Close prices of Brent Oil ranging from (20.09.2017 – 13.12.2019) are accessed through the quantmod package in R. There is a sharp jump changes in the last quarter of 2018, which can be considered as change points that suddenly reverses from the rapid rise, or as a reflection of economic practice. The close prices of the USD - TL exchange rate in the range (9.02.2017 – 26.02.2020) are extracted from Yahoo Finance. Likewise, in the USD - TL exchange rate data, there are abrupt changes around July and August 2018, due to the crisis and depreciation of the Turkish Lira against the US Dollar. To be consistent with what we have done in the simulations section, we drew three windows of size 400 each for Brent Oil and three windows of size 500 each for TL – USD to represent the three structural break test locations (Q1, Q2, Q3) as illustrated in Figure 4.5.

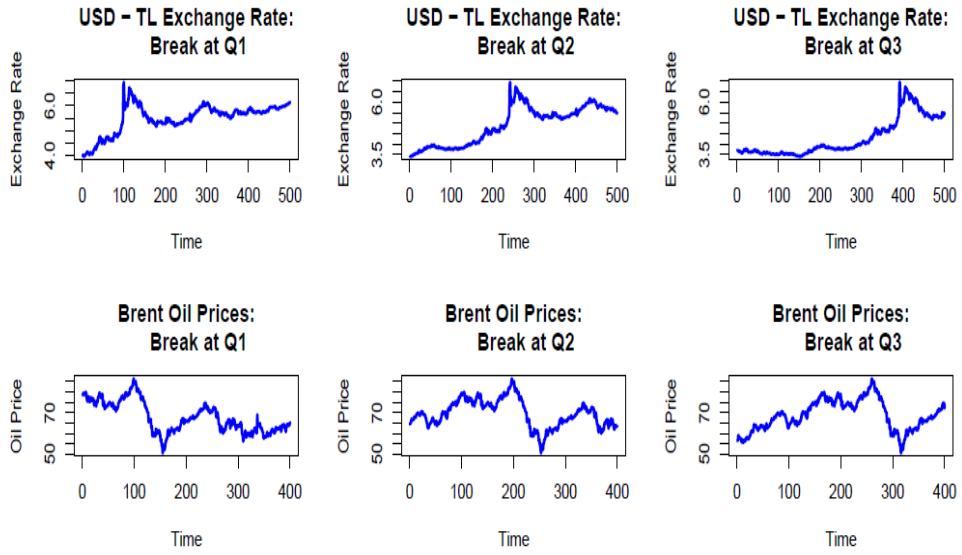


Figure 4.5 Break Locations at Brent Oil Prices and TL – USD Exchange Rate

The upper part of Figure 4.5 illustrates the USD – TL exchange rates while the lower part shows Brent Oil Prices. While both data sets exhibit at least one structural break in each case, crude oil prices show much volatility greater than the exchange rate.

4.3.1 Structural Break Testing

Before fitting and forecasting with the models under consideration, we first check the presence of structural breaks in our data and estimate the break dates. Bai and Perron, (2003) made possible to estimate the break date and the intercept of each segment and that is done with the help of the R function *breakpoints* from the *strucchange* package in R developed by (Zeileis et al., 2002). Table 4.7 shows the estimated break date, segments, and intercepts in both brent oil and USD – TL exchange rate data. The presence of structural break is also tested using Chow and CUSUM tests and the results are presented in Table 4.7.

Table 4.7 Break Dates and Segmentation Using Bai and Perron Test

Data	**Break Date**	**Segment**	**Intercept**
USD - TL Exchange Rate	10/7/2018	9/2/2017 - 10/7/2018	3.817
		11/7/2018 - 26/2/2020	5.696
Brent Oil Prices	9/11/2018	20/9/2017 - 9/11/2018	70.519
		12/11/2018 - 13/12/2019	63.525

Table 4.8 Significance of the Structural Break

Data	**Test Statistic**		**P-value**
USD - TL Exchange Rate	Chow (supF) test	39.661	8.774e-08
	CUSUM Test	13.185	2.2e-16
Brent Oil Prices	Chow (supF) test	22.098	0.0004
	CUSUM Test	5.4359	2.2e-16

Test statistics and the accompanied p-values results of Chow test and CUSUM test in Table 4.8 show that there is significant structural break. As can be seen in Table 4.7, USD - TL Exchange Rate contain structural break in 10/7/2018 and estimated intercept in the segment of 9/2/2017 - 10/7/2018 is 3.817 while the segment 11/7/2018 -

26/2/2020 should have 5.696 as intercept. Brent Oil Prices also contain structural break at 9/11/2018. The estimated intercepts for 20/9/2017 - 9/11/2018 and 12/11/2018 - 13/12/2019 segments are 70.519 and 63.525 respectively.

4.3.2 ARCH Effect Testing

ARCH Engle's Test for Residual Heteroscedasticity is applied to the empirical data to examine if the data contain ARCH effects. The squared residuals are a sequence of white noise, according to the null hypothesis in the Lagrange Multiplier test, which assumes that the residuals are homoscedastic. The test statistic is calculated starting from Lag 1 and then after every 4 steps until 24th Lag.

Table 4.9 Lagrange-Multiplier test

USD - TL Exchange Rate				**Brent Oil Prices**		
Lag	**LM Test Statistic**	**P.value**		**Lag**	**LM Test Statistic**	**P.value**
4	822	0.000		4	918	0.000
8	400	0.000		8	447	0.000
12	250	0.000		12	276	0.000
16	185	0.000		16	204	0.000
20	145	0.000		20	159	0.000
24	115	0.000		24	127	0.000

The LM Test is conducted using arch.test function in R from aTSA package. The results in Table 4.9 reveal that both empirical time series data in hand contain ARCH effect.

4.4 Empirical Results

Like simulations, we employed the same four forecasting methods, then forecasted out–of–samples and calculated Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE). We also chose evaluation sample sizes same as those in simulations (h = 3, h = 6, h = 10). The smaller the value of RMSE and MAPE for a given method compared to the other methods the better that method is performing. The empirical results of both considered data are presented in Tables 4.10 and 4.11 side by side.

Table 4.10 RMSE Values of Empirical Analysis

Brent Crude Oil Price				USD - TL Exchange Rate			
Structural Break at Q1				**Structural Break at Q1**			
Model/Horizon	**h = 3**	**h = 6**	**h = 10**	**Model/Horizon**	**h = 3**	**h = 6**	**h = 10**
ARCH	0.2573	0.8358	1.1164	ARCH	0.0256	0.0635	0.0725
ARIMA	0.3074	1.1437	1.5038	ARIMA	0.0314	0.0653	0.079
EXP	0.3072	1.1440	1.5034	EXP	0.0272	0.0625	0.0733
ANN	0.2342	0.816	1.017	ANN	0.0259	0.0511	0.0571
Structural Break at Q2				**Structural Break at Q2**			
ARCH	0.4911	1.3326	1.9648	ARCH	0.0370	0.0919	0.1165
ARIMA	0.5308	1.4246	1.9287	ARIMA	0.0335	0.0892	0.1255
EXP	0.4434	1.2377	1.9873	EXP	0.0388	0.0910	0.1176
ANN	0.347	1.1628	1.9339	ANN	0.0282	0.0708	0.0739
Structural Break at Q3				**Structural Break at Q3**			
ARCH	1.3475	2.2675	2.4315	ARCH	0.0341	0.1509	0.1693
ARIMA	1.6042	2.1624	2.2808	ARIMA	0.0379	0.1008	0.132
EXP	1.3213	2.2603	2.2804	EXP	0.0813	0.1045	0.1333
ANN	1.3109	2.1177	2.1115	ANN	0.0295	0.0994	0.0887

Table 4.11 MAPE Values of Empirical Analysis

Brent Crude Oil Price				USD - TL Exchange Rate			
Structural Break at Q1				**Structural Break at Q1**			
Model/Horizon	**h = 3**	**h = 6**	**h = 10**	**Model/Horizon**	**h = 3**	**h = 6**	**h = 10**
ARCH	0.3198	1.4840	1.1238	ARCH	0.3155	0.9478	1.0430
ARIMA	0.3272	2.0954	1.6325	ARIMA	0.0301	0.9838	1.1616
EXP	0.3274	2.0948	1.6330	EXP	0.3503	0.9290	1.0590
ANN	0.3368	1.2937	1.0751	ANN	0.3309	0.7557	0.8175
Structural Break at Q2				**Structural Break at Q2**			
ARCH	0.7548	2.0108	2.3451	ARCH	0.6715	1.5175	1.9302
ARIMA	0.8143	2.1529	2.2915	ARIMA	0.5875	1.4859	2.0943
EXP	0.6747	1.8413	2.3819	EXP	0.7037	1.4992	1.9523
ANN	0.4439	1.6855	2.3547	ANN	0.5113	1.1475	1.2131
Structural Break at Q3				**Structural Break at Q3**			
ARCH	1.1743	2.6640	2.7130	ARCH	0.5832	2.4816	2.5424
ARIMA	1.4595	2.5129	2.5080	ARIMA	0.6121	1.4692	1.9731
EXP	1.1532	2.6594	2.5073	EXP	1.4087	1.6269	1.9690
ANN	1.1412	2.4429	2.2787	ANN	0.5056	1.3908	1.5016

The findings and discussions related to the empirical analysis are summarized here:

i. Brent oil prices continuously fluctuate causing instability in the economy. The sudden ups and downs of oil prices have direct or indirect impacts on economics since they produce specific issues in global economics. Due to this rapid variability, predicting oil prices has global importance however the behavior of the brent oil prices is unstable over time and it is difficult to predict it. Among the considered models for Brent Oil Prices, ANN outperforms the others including the ARCH model in almost all horizons no matter where the break locates (Q1, Q2, Q3). Even though Oil prices are much more volatile, ANN is more resilient to variance heterogeneity and the structural break than the other forecasting techniques in consideration. Imbalanced data, which has non-linear trend, changeable cyclical fluctuations, and high volatility, is a big issue in time series analysis. It should also be noted that the existence of high volatility may mask the effect of structural break on the performance as well. For complex economical or financial data, ANN should be the first preferable methodology.

ii. After 2017, there is a rapid fluctuation in Turkish Lira value against Dollar. The Turkish US Dollar – Turkish Lira exchange rate series has more statistically stable properties in comparison with brent oil price series. The series has trend component and less volatility. In this sense, the structural break location is important for the data. If the structural break is in Q1, the series seems to be more volatile in comparison with when the structural break in Q2 and Q3. For USD - TL exchange rate, ANN still gives the smallest RMSEs and MAPEs in all the cases except the short horizon (h = 3) when the break is in Q1. In this case, the ARCH model takes the lead for RMSE metric and ARIMA for MAPE metric. Hence, the ARCH and ARIMA models may be preferred for short–term forecasts of less volatile data when the break is at the very beginning of the training set (Q1). The reason could be that the variance spectrum of the data is broadened because of location and size of structural break, and the volatility becomes more effective. Therefore, the ARCH and ARIMS models in Q1 for the

short-term horizon generate less RMSE and MAPE respectively in comparison to ANN.

iii. All models taken into account here are affected by the location of the break as RMSEs and MAPEs are smallest when the break locates at Q1 and largest when the break is in Q3. This result is true for both exchange rate and oil price data in most horizons applied.

iv. For Crude oil prices data, which is relatively more volatile, RMSEs increase faster as the break location progresses from Q2 to Q3 than its progress from Q1 to Q2 for short horizons (h = 3, h = 6), and it is the opposite for long horizons (h = 10). This is true for all models used here.

v. For the exchange rate data, which is relatively less volatile, the only agreement between the models is that RMSEs increase as the break location moves from Q1 to Q2 and that is not the case in Q3. There is no considerable agreement between the used models regarding the speed of RMSEs increase.

For guidance:

For the imbalanced and complex economic and financial data, ANN should be preferred in any case regardless of the break size and the break location. The dynamical behavior of the time series before or after structural break does not affect the success of ANN. In fact, just in the Q1 case, which is covered by the training set, structural break has caused high volatility and that affected ANN negatively in this case. If structural break causes volatility in economic data or the structural break and volatility overlapped in time and the break is in the first quartile, ARCH model should be used.

4.5 Chapter Conclusion

The aim of this chapter was to answer the question of how a structural break affects forecasting performance. More specifically, the focus was on the impact of break

location and break size on the predictive performance of competing methods (ARIMA, Exponential, ARIMA - ARCH and Artificial Neural Network) in various simulation cases. The above techniques are applied to forecasting daily close prices of brent oil and TL – USD exchange rates out – of – sample and compare the performance of the applied models.

The findings show that apart from the impact of structural break on the forecasting accuracy, break location and break size have implications on prediction. This result is found both in simulations and empirical analysis. Forecasting error increases with the increase of break magnitude and the location of the break as it is getting closer to the end of the series. The error is minimum when the break is in the first quartile, it is more when the break is in the second quartile, and it is the most when break locates in the third quartile.

Comparing the competing models, ARCH model is the best performing model in the absence of structural break while ANN is the next in performance. For the cases of the various structural break sizes and break locations, ANN outperforms all the other models with the exception of few places. ANN also shown superiority in the empirical analysis in most of the cases under investigation. Hence, while structural break affects the performance of all the examined models, ANN is the most resilient for the breaks and instability of the parameters.

As the results suggest that the effect of structural break location is worst when it is in the third quartile (Q3), it is recommended that this issue be scrutinized intensely to improve forecast accuracy in this case. One of the strategies to improve forecast accuracy is to combine the forecast components of different models based on sophisticated combination schemes and that is what next chapter (Chapter 5) is discussing.

CHAPTER FIVE
IMPROVING FORECAST ACCURACY WITH COMBINATION OF FORECASTS UNDER STRUCTURAL BREAKS AND ARCH INNOVATIONS

An accurate forecast about the future is vital in time series analysis, and it is always challenging for forecasters to accurately model complex structures in the data. That complexity is related to the determination of the underlying data-generating mechanism as it is not fully known in most cases. As the observed processes are too complicated to be accurately modeled with the exception of some natural sciences (Weiss et al., 2019), models are approximations of the real-world data-generating processes (Hansen, 2005). The sources of the complexity of the data structures include the unexpected changes (breaks) we experience in model parameters in financial time series, perhaps because of policy shifts, technological advancement, financial crises, and natural disasters. Many researchers testified that structural breaks are present in economic and financial time series. At the forefront of those researchers are Stock and Watson (1996) who used several standard statistical tests to examine many U.S. economic time series relationships for model instability, Koop and Potter (2001), Siliverstovs and van Dijk (2003). Structural breaks reduce forecasting accuracy as the studies of Clements and Hendry (1998, 1999) and Hansen (2001) show and consequently mislead policy recommendations and other prediction purposes.

The traditional approach of forecasting is based on the existence of best method among the available methods and can be identified but the selection of that best method depends upon the features of that time series. Aggregating inputs from several forecasting techniques using forecasts is an alternative to the conventional strategy. This solves the issue of having to choose just one method and rely solely on its forecasts (Winkler & Makridakis, 1983). Numerous studies have demonstrated that using several different models on the same time series and averaging the forecasts components is a simple way to improve forecast accuracy. For more than five decades John Bates and Clive Granger wrote their seminal paper (Bates & Granger, 1969) and showed that often, combination of forecasts improves forecast accuracy. Two decades

later Clemen (1989) wrote a review paper and argued that the accuracy of the forecast is improved by combining multiple forecasts. Moreover, Clemen's findings showed that in many cases by merely averaging the forecasts, one can make impressive performance improvements. Although the idea of forecast combination dates back to half a century, and even in recent days, papers describing novel combination techniques are published in prestigious journals, which encourage additional study (Weiss et al., 2019).

Several combining rules are proposed in looking for an optimal combination of forecasts. The proposed techniques include a combination based on variance-covariances of the forecast components by Bates and Granger (1969) and other equal weighting methods including simple average and median forecast combinations. Moreover, Ordinary Least Squares (OLS) regression-based combination is proposed by Granger and Ramanathan (1984), while Hsiao and Wan (2014) developed eigenvector-based combination approaches.

Given the prominence of the forecast combination topic in both theoretical and empirical studies, it is applied more generically and needs further studies if the gains of combined forecasts are sustained in specific cases such as the cases of structural breaks and ARCH effects. The studies concerned in this area are very limited in literature. In this chapter, we examine how the forecast accuracy can be improved in the presence of structural break(s) and ARCH/GARCH innovations and which combination schemes are optimal in the considered cases. This is done by combining the component forecasts from single models in either simple weighting techniques or regression-based weighting strategies. In the later, we considered both time - varying combination weights (dynamic version) and time – non-varying combination weights (static version). Before the combination is implemented, the component forecasts are generated from three individual models called Autoregressive Integrated Moving Average (ARIMA), Exponential Smoothing (ETS), and Artificial Neural Network (ANN). These models are implemented in simulated and empirical data while the Mean Absolute Percentage Error (MAPE) and root mean squared forecast error

(RMSE) are used to assess the accuracy and predictive performance of the models under study.

5.1 Data Description and Simulation Designs

Data used in this chapter contains simulations and real data. As shown in Section 5.2, the simulated data is generated based on autoregressive (AR) processes with autoregressive conditionally heteroscedastic errors and it contains structural break in the third quartile of the time series sample. An empirical data of two popular Turkish financial stocks namely BIST30 and BIST100 are employed in this here. We consider BIST data because it has the features of structural breaks we are interested in our investigation.

Several studies in the literature used various methods to forecast BIST100 or BIST30. For instance, Aygören et al. (2012) studied BIST100 forecasting using traditional time series models such as ARIMA, numerical search models, and artificial neural network models. They argued that ANN outperformed Newton Numerical Search Models and conventional time series models. Telli and COŞKUN (2016) also forecasted BIST100 Index using Artificial Neural Network models using daily data between July and November 2015. Among several models, structured multilayer perceptron (MLP) is shown as the best model. ÜNVAN and ERGENÇ (2022) additionally compared the predictive ability of ANN and Regression models in application to BIST 100 closing prices between 2010 and 2020. This research found that ANN performs better than the regression models considered.

Raşo and Demirci (2019) used deep learning methods to forecast the Turkish Stock Market of BIST30 from January 2016 to April 2018. The study's findings revealed that the deep learning model outperforms other techniques, including Support Vector Regression (SVR). Furthermore, ALP et al. (2020) conducted a comparative study on predicting BIST price indices. They compared the performance of ARIMA against two deep learning methods, namely Gated-Recurrent Unit (GRU) and Long Short-Term Memory (LSTM), in predicting the BIST 100, BIST 50, and BIST 30 price indices.

This study found that ARIMA models outperform deep learning models in forecasting the price indices under consideration.

Aker (2022) likewise examined price volatility of BIST100 by comparing LSTM and Facebook Prophet (Fbprophet) methods. This study suggested that LSTM model outperformed Fbprophet model based on evaluation metrics of RMSE, MAE and MSE. PAKEL and ÖZEN (2021) also investigated volatility analysis of BIST100 using GARCH models. This study identified two significant shocks in the BIST100 namely currency shock and COVID-19 pandemic shock in 2018 and 2020 respectively. This research revealed that stock market volatility increased significantly during 2018-2020 periods, and that the increase was more persistent during the COVID-19 pandemic period. Yılmaz and Kale (2022) also analyzed the short and long-term asymmetrical impacts of company's financial ratios on their risks. These study's findings demonstrate that risk and financial ratios do not have an asymmetric relationship in the short run.

All the aforementioned recent studies indicate that it is still important to produce forecasts based on the BIST30 and BIST100 data, which are influenced by the undesirable conditions experienced by the Turkish economy in the short and long term. Some of these articles compared classical and modern forecasting methods, while others tried to model variance. Unlike the articles cited, classical time series models and ANN models were compared under structural breaks and ARCH innovations conditions in the BIST30 and BIST100 data, where it is difficult to produce a high-frequency, volatile, and low-error prediction. Although the results obtained in our study are partially parallel to the results found in these studies, comparison of combining forecast performances under different conditions of structural breaks and conditional heteroscedasticity could be a contribution to the literature.

5.1.2 Simulation Designs

In the simulation procedure, we considered two different designs of data-generating processes:

1. Design I is based on AR (3) process with autoregressive conditionally heteroscedastic errors ARCH (2) as shown in equation (5.1):

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \varepsilon_t \tag{5.1}$$

$$\varepsilon_t = \sigma_t a_t \tag{5.2}$$

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 \tag{5.3}$$

where a_t in equation (5.2) is white noise ω, α_1 and α_2 in equation (5.3) are the parameters of the variance model. Substituting for σ_t^2 we receive equation (5.4):

$$\varepsilon_t = a_t \sqrt{\omega + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2} \tag{5.4}$$

2. Design II is based on ARIMA (3,1,0) process with autoregressive conditionally heteroscedastic errors ARCH (2). The formulation of design II is same as the first design except that the second design process generates time series data with trend in addition to ARCH effects.

Simulations in both designs are conducted by specifying fixed stationary and invertible parameters. For design I, parameters of the AR part specified as $\mu = 5, \phi_1 = 0.65, \phi_2 = -0.55$ and $\phi_3 = 0.3$ while parameters of the ARCH part are $\omega = 0.2, \alpha_1 = 0.45, \alpha_2 = 0.3$. similarly for design II, stationary and invertible parameters of ARIMA part are specified as $\mu = 5, \phi_1 = 0.45, \phi_2 = 0.3$ and $\phi_3 = -0.75$ and ARCH part parameters are same as those of design I ARCH model. It is worth noting that these values were chosen arbitrarily. Graphical visualization of the simulated data is in Figure 5.1.

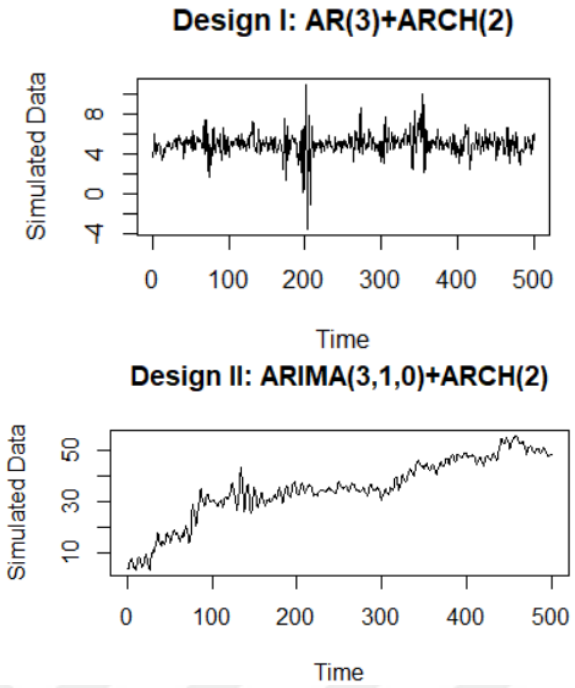


Figure 5.1 Graphical illustration of the simulated data

The forecasting models under study are applied to the simulated data before the structural break is placed to forecast out-of-sample observations. A single permanent structural break is allowed in the generated time series' mean in both designs. The structural break is positioned in the 75%th observation of the total sample. The break took multiplicative form and so the pre-structural break mean of μ becomes $\delta\mu$ after the break, where δ is a break size which is 10 in this case. Figure 5.2 below illustrates data after structural break.

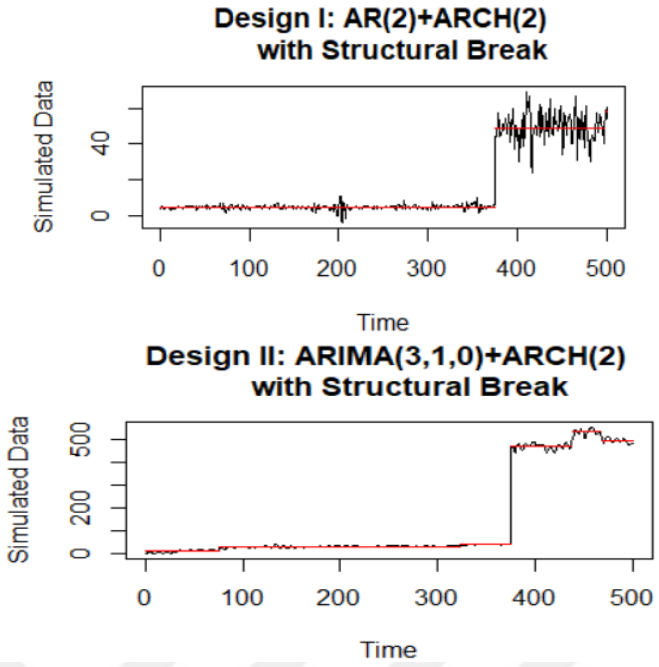


Figure 5.2 Simulated Data after Structural Break

As shown in Table 5.1, the total sample size in each case ($T = n + h$) is 500 where T stands for the total sample size, n denotes training sample size, and h presents the prediction sample size.

Table 5.1 Summary of the Simulated Data Procedures and Train-Test Sets

Design	**Case**	**Break Size**	**Break Location**	**Sample Size**	**Training Set(size)**	**Test set (size)**
Design I	Stable	-	-	500	480, 490, 495	20, 10, 5
	Break	10	75th% Observation	500	480, 490, 495	20, 10, 5
Design II	Stable	-	-	500	480, 490, 495	20, 10, 5
	Break	10	75th% Observation	500	480, 490, 495	20, 10, 5

To examine if the horizon size has an impact on forecasting accuracy, we reserve the first 480, 490, and 495 (n = 480, n = 490 and n = 495) observations as the training sample and the rest as the prediction sample (h = 20, h = 10 and h = 5) in three separate experiments in each case. The simulation process is repeated 3000 times in each case.

In each case, we generate forecasts using three individual methods: ARIMA, ETS, and ANN. We sought to increase forecasting accuracy by combining component forecasts from single models into simple forecast combination and regression-based

combination models. Simple forecast combination models considered in this chapter are simple average forecast combination, Bates–Granger forecast combination and median forecast combination. The regression-based combination models used OLS forecast combination (both statistic and dynamic versions) and LAD forecast combination (both static and dynamic versions). Mean Absolute Percentage Error (MAPE) (only in empirical analysis) and Root mean squared forecast error (RMSE) are used to assess predictive ability of the various methods in consideration.

5.2 Simulation Results

In the following tables, we present RMSE values used to measure forecasting accuracy. If the RMSE value is less than the others, it indicates that the method under consideration outperforms the other competing methods. The smaller the RMSE is, the better the performance of the forecasting is for that given sample.

Table 5.2 Design I Stable Case Results

Model	**h = 20**	**h = 10**	**h = 5**
Individual Models			
ETS	0.03464	0.02664	0.02516
ARIMA	0.03150	0.02310	0.01503
ANN	0.03038	0.02159	0.01013
Simple Forecast Combination			
Comb_SA	0.03200	0.02320	0.01586
Comb_MED	0.03125	0.02311	0.01551
Comb_BG	0.03145	0.02278	0.01437
Regression-Based Forecast Combination			
Comb_OLS_static	0.02912	0.02106	0.01031
Comb_OLS_dynamic	0.02907	0.02109	0.01033
Comb_LAD_static	0.02909	0.02117	0.00966
Comb_LAD_dynamic	0.02906	0.02123	0.00963

Simulation results for the AR (3) with ARCH error design before the structural break is placed are presented in Table 5.2. The meanings of the symbols in the table are listed below;

- Comb_SA: Simple Average forecast combination,

- MED: Median Forecast Combination,
- Comb_BG: Bates/Granger forecast combination
- Comb_OLS_static: Ordinary Least Squares forecast combination (statistic version)
- Comb_OLS_dynamic: Ordinary Least Squares forecast combination (dynamic version)
- Comb_LAD_static: Least Absolute Deviation forecast combination (statistic version)
- Comb_LAD_dynamic: Least Absolute Deviation forecast combination (dynamic version)

The accuracy evaluation delivers that ANN outperforms the other stand-alone models used here in all considered horizons. Regression-based combination models (in their static and dynamic versions) exhibit improved accuracy compared to single models and simple combinations. In comparison to the static versions of the models, letting combination weights to vary with time (dynamic version) does not appear to change much accuracy.

Table 5.3 Design I with Structural Break

Model	**h = 20**	**h = 10**	**h = 5**
Individual Models			
ETS	0.6977	0.4308	0.2059
ARIMA	0.6676	0.2698	0.1788
ANN	0.4962	0.2299	0.1405
Simple Forecast Combination			
Comb_SA	0.6057	0.2696	0.1549
Comb_MED	0.6256	0.2529	0.1405
Comb_BG	0.5928	0.2646	0.1531
Regression-Based Forecast Combination			
Comb_OLS_static	0.4320	0.1676	0.1395
Comb_OLS_dynamic	0.4339	0.1683	0.1399
Comb_LAD_static	0.4428	0.1971	0.1345
Comb_LAD_dynamic	0.4451	0.1988	0.1354

In Table 5.3 we reported the simulation results of AR (3) with ARCH error design with the structural break. Again, ANN performs better than its single-model counterparts of ARIMA and ETS in all considered horizons. The results also suggest that all the simple combination models are better than single models other than ANN. Moreover, regression-based forecast combination models are performing better than any other model considered here while the OLS method in its two versions performs best. As in the previous case, dynamic combination weights do not change, and even the static version is better suggesting that combination weights estimates are not changing much over time.

Table 5.4 Design II Stable Case Results

Model	**h = 20**	**h = 10**	**h = 5**
Individual Models			
ETS	1.8587	1.2474	0.9491
ARIMA	1.7695	1.2573	0.4621
ANN	1.7107	1.1923	0.4426
Simple Forecast Combination			
Comb_SA	1.7429	1.0797	0.5097
Comb_MED	1.8063	1.1909	0.4440
Comb_BG	1.7504	1.0989	0.4904
Regression-Based Forecast Combination			
Comb_OLS_static	1.6734	1.2092	0.4141
Comb_OLS_dynamic	1.6162	1.2031	0.4150
Comb_LAD_static	1.7027	1.1716	0.4295
Comb_LAD_dynamic	1.6825	1.1682	0.4300

Simulation results of the stable case of design II where data are generated from AR (3) with trend and ARCH (2) errors are stated in Table 5.4. Results show that ANN exhibits the best accuracy in comparison with the non-combined models. Moreover, all simple forecast combination models perform well in comparison with non-combined methods when the forecasting horizon is moderate (h = 10) even better than ANN but lose that dominance in long and short horizons. Although all regression-based models show improvement in the accuracy of forecasting, the OLS model is the best-performing one. In contrast to design I, dynamic combination weights show improvement in the results for both OLS and LAD forecast combination suggesting that combination weights estimates are changing much over time.

It is reported in Table 5.5 the simulation results of design II where data are generated from AR (3) with trend and ARCH (2) errors after structural break is positioned in the data. As in the previous cases, ANN is the best single model in all horizons considered. Simple combinations show some improvement in accordance with ARIMA and ETS and it is worth mentioning that the simple average based combination model is the most superior in the short horizon (h = 5).

Table 5.5 Design II with Structural Break

Model	**h = 20**	**h = 10**	**h = 5**
Individual Models			
ETS	11.1834	8.6215	5.4916
ARIMA	11.1829	8.6216	5.4857
ANN	11.0805	8.5229	5.3504
Simple Forecast Combination			
Comb_SA	11.1034	7.2160	2.3155
Comb_MED	11.1831	8.6215	5.4857
Comb_BG	11.1058	7.2384	5.4886
Regression-Based Forecast Combination			
Comb_OLS_static	10.9607	7.5337	4.5216
Comb_OLS_dynamic	10.7620	7.4997	4.4685
Comb_LAD_static	10.1257	7.5863	2.8762
Comb_LAD_dynamic	9.9617	7.5088	2.8344

Regression-based forecast combination methods again improve the forecasting accuracy, especially dynamic combination weighting show significant improvement in comparison with their static counterparts.

5.2.1 Summary of Simulation Results

1. ANN performs better than its non-combined counterparts across all considered horizons, in all examined cases and data generating designs.

2. Simple forecast combinations (simple average, median and Bates-Granger) perform better than ETS and ARIMA in most cases.

3. Regression-based forecast combination methods outperform all other competing models in all cases no matter which design data is from or if the structural break is present or not.

4. Allowing time-varying combination weights (dynamic version) shows a significant improvement in accuracy comparing with the static counterparts in the case of trend, and that suggests that the estimated weights of combination fluctuate much over time. In contrast, dynamic combination weights do not change much in the case of trend absence.

5. Other studies including Hsiao & Wan (2014) showed that if one of the component forecasts outperforms the rest which is also the case here, regression-based approaches are favorable and our results agreed with that conclusion.

6. If the series contains both trend and structural break, simple average and Bates/Granger combinations perform better than ANN for h = 10. Especially, in regression based forecast combinations for both static and dynamic also perform better than ANN for all horizons. This shows that existence of trend and structural break at the same time adversely affects the performance of ANN.

7. The presence of trend in the series, and additionally break in trend increases in the same way the RMSEs for all designs, across all horizons and methods. Even removing trend by taking the difference could not prevent the increase in RMSEs.

8. The RMSEs are smaller in the short horizon (h = 5) compared to longer horizons (h=10 and 20) under all conditions. This also shows that it is always safer to work under short term period of forecasting.

5.3 Empirical Applications

In line with the simulation evidence, all considered models in the simulation are applied to forecast daily close prices of two popular Turkish financial stocks data namely BIST30 and BIST100. We consider BIST data because it has the features we

are interested in our study. As it is realized in Tables 5.6, 5.7 and 5.8 of this section, both considered stock data contain multiple structural breaks and ARCH effects.

We plotted BIST30 and BIST100 data in Figure 5.3. For example, both considered stock data contain multiple structural breaks and ARCH effects as shown in Figure 5.4 and Tables 5.6, 5.7 and 5.8. Stocks data in the period between Jan 4, 2010, to Jun 20, 2022, were extracted from this website (investing.com).

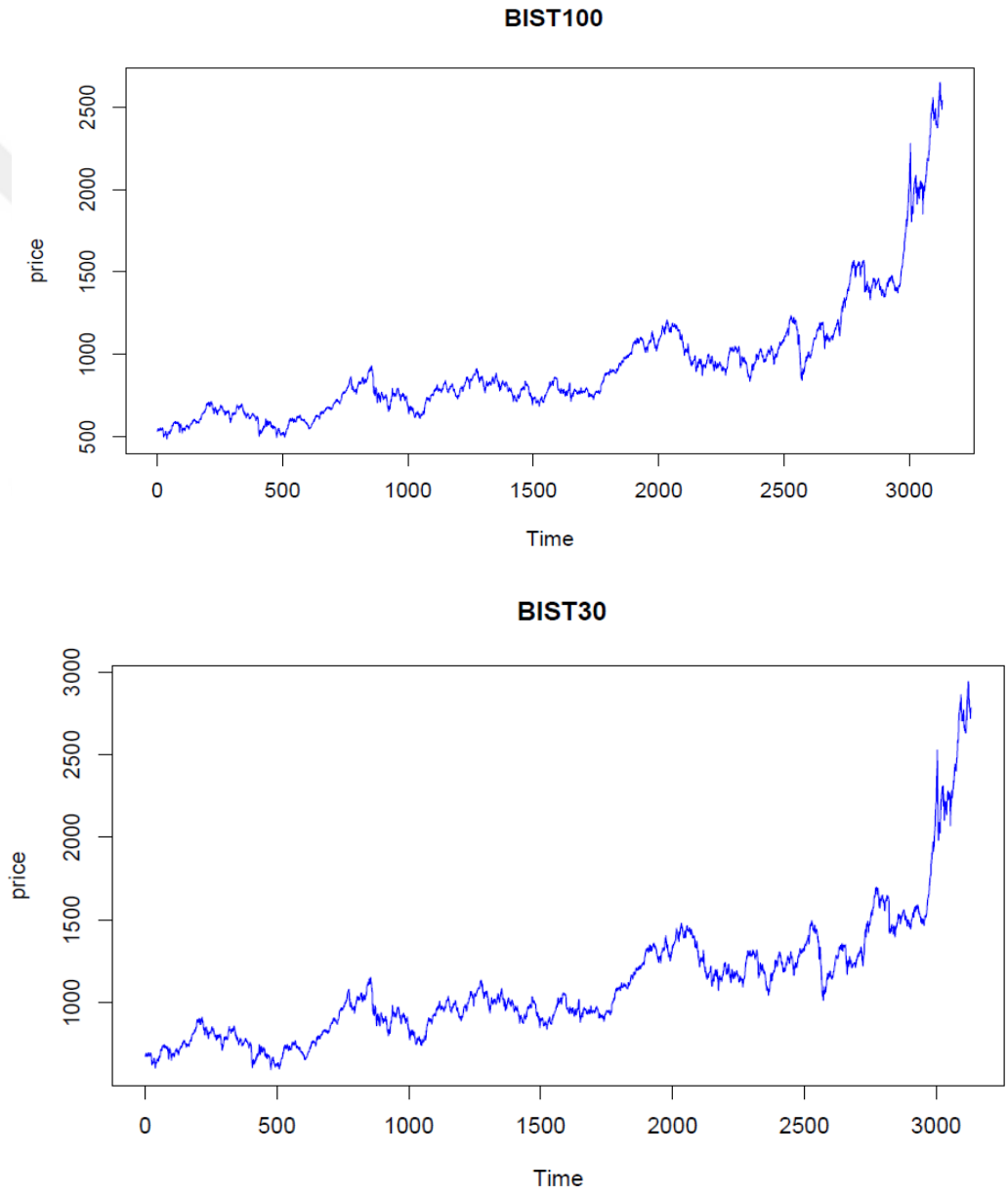


Figure 5.3 BIST30 and BIST100 Data

5.3.1 Testing ARCH Effects

To check the existence of ARCH effect in the series, Engle's Lagrange Multiplier test is used. This test is conducted using arch.test function in R from aTSA package.

Table 5.6 Lagrange-Multiplier Test Results

BIST100				BIST30		
Lag	**LM Test Statistic**	**P.value**		**Lag**	**LM Test Statistic**	**P.value**
4	4970	0.000		4	3293	0.000
8	1927	0.000		8	1611	0.000
12	1250	0.000		12	1028	0.000
16	886	0.000		16	753	0.000
20	691	0.000		20	594	0.000
24	513	0.000		24	437	0.000

Since p values in Table 5.6 are zero or close to zero, we reject the null hypothesis of no ARCH effects present in the residuals. Therefore, we may conclude the presence of ARCH effects in both two series under study.

5.3.2 Structural Break(s) Testing

Before we forecast with the models under consideration, we first check the presence of structural breaks in the data under consideration. This is performed with Bai and Perron (2003) test for multiple structural breaks with the help of the R function breakpoints from the strucchange package.

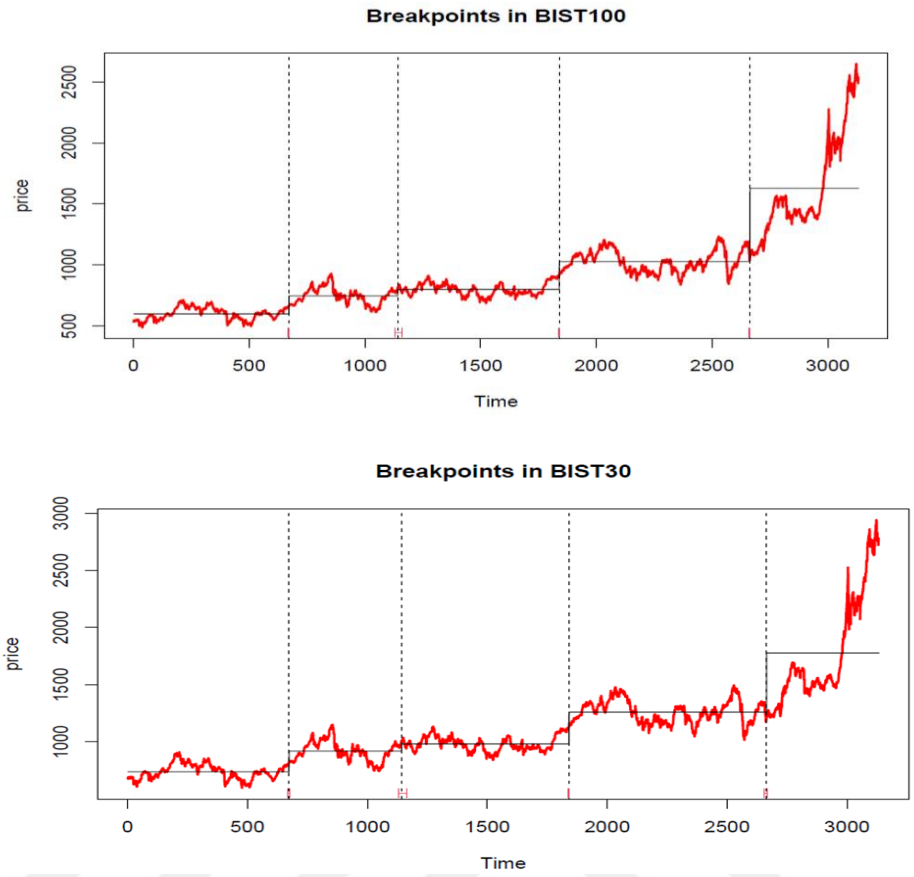


Figure 5.4 Break Locations and Segments

Figure 5.4 suggests that both stocks data contain at least four significant structural breaks. Table 5.7 and 5.8 show the breakpoints, break dates estimated and optimal segmentation accompanied by the estimated intercepts for each segment. Both BIST100 and BIST30 have got the same number of structural breaks and similar locations for break dates.

Table 5.7 BIST100 Break Locations and Segmentation

Breakpoint	Breakdate	Segment	Intercept
672	Aug29,2012	Jan4,2010 – Aug29,2012	600.233
1141	Jul14,2014	Aug30,2012 – Jul14,2014	748.157
1839	Apr19,2017	Jul15,2014 – Apr19,2017	800.134
2663	Aug7,2020	Apr20,2017 – Aug7,2020	1028.809
		Aug8,2020 – Jun20,2022	1628.825

Table 5.8 BIST30 Breakpoints and Segmentation

Breakpoint	Breakdate	Segment	Intercept
672	Aug29,2012	Jan4,2010 – Aug29,2012	738.596
1141	Jul14,2014	Aug30,2012 – Jul14,2014	919.286
1839	Apr19,2017	Jul15,2014 – Apr19,2017	981.905
2663	Aug7,2020	Apr20,2017 – Aug7,2020	1260.563
		Aug8,2020 – Jun20,2022	1777.168

It is now clear that the empirical data we have contain structural breaks and ARCH effects and it is time to continue the forecasting process and measure its accuracy.

Table 5.9 Train-Test Sets Plan of the Empirical Data

Stock Index	Sample Size	Training Set(size)	Test set (size)
BIST100	3132	3112, 3122, 3127	20, 10, 5
BIST30	3132	3112, 3122, 3127	20, 10, 5

As Table 5.9 shows, the total sample size in each case ($T = n + h$) is 3132 where T stands for total sample size, n stands for training sample size, and h represents prediction sample size. Although the total sample size of empirical data is much more than that of simulated data, we examined the same prediction sample sizes (horizon size) as the simulation part. In each case, we reserve the first 3112, 3122 and 3127 (n = 3112, n = 3122 and n = 3127) observations as the training sample and the rest as the prediction sample (h = 20, h = 10 and h = 5) in three separate experiments in each case.

Tables 5.10, 5.11, 5.12 and 5.13 present the empirical results of the two Turkish financial stocks data namely BIST100 and BIST30 in RMSEs and MAPE respectively. The results of the empirical data are similar to that of design II of simulated data. The possible explanation to that similarity is that their components are similar in the sense that both data contain ARCH effects, trend, and structural break(s).

Table 5.10 BIST100 RMSEs Results

Model	h = 20	h = 10	h = 5
Individual Models			
ETS	164.6264	117.8719	22.70763
ARIMA	165.9428	120.8585	22.85353
ANN	154.2198	89.08681	22.38252
Simple Forecast Combination			
Comb_SA	161.5748	109.0967	22.62900
Comb_MED	164.6226	117.7284	22.67512
Comb_BG	161.4729	108.7855	22.62865
Regression-Based Forecast Combination			
Comb_OLS_static	150.7041	69.4551	22.0219
Comb_OLS_dynamic	138.3404	55.8600	22.0339
Comb_LAD_static	153.8752	99.2405	20.8992
Comb_LAD_dynamic	150.4317	88.8441	20.8992

Table 5.11 BIST30 RMSEs Results

Model	h = 20	h = 10	h = 5
Individual Models			
ETS	167.9629	160.405	23.49172
ARIMA	172.8163	156.102	23.48209
ANN	158.327	127.077	23.13754
Simple Forecast Combination			
Comb_SA	166.3248	147.7268	23.28021
Comb_MED	167.9673	156.0968	23.54386
Comb_BG	166.3106	147.2373	23.27600
Regression-Based Forecast Combination			
Comb_OLS_static	156.9727	101.259	23.05612
Comb_OLS_dynamic	135.9082	86.07532	23.15337
Comb_LAD_static	158.0176	121.7372	23.08587
Comb_LAD_dynamic	150.6015	114.8549	23.10114

Table 5.12 BIST100 MAPE Results

Model	h = 20	h = 10	h = 5
Individual Models			
ETS	5.8526	4.3566	0.8166
ARIMA	5.9051	4.4600	0.8130
ANN	5.4400	3.3045	0.8091
Simple Forecast Combination			
Comb_SA	5.7326	4.0389	0.8129
Comb_MED	5.8520	4.3504	0.8155
Comb_BG	5.7285	4.0277	0.8129
Regression-Based Forecast Combination			
Comb_OLS_static	5.3019	2.5470	0.8085
Comb_OLS_dynamic	4.7503	1.8772	0.8092
Comb_LAD_static	5.4289	3.6710	0.7763
Comb_LAD_dynamic	5.2776	3.2411	0.7763

Table 5.13 BIST30 MAPE Results

Model	h = 20	h = 10	h = 5
Individual Models			
ETS	5.2615	5.4096	0.7478
ARIMA	5.4534	5.2652	0.7063
ANN	4.9033	4.2930	0.7607
Simple Forecast Combination			
Comb_SA	5.2060	4.9720	0.7383
Comb_MED	5.2649	5.2608	0.7596
Comb_BG	5.2055	4.9549	0.7389
Regression-Based Forecast Combination			
Comb_OLS_static	4.8389	3.4216	0.7734
Comb_OLS_dynamic	3.8009	2.7548	0.7770
Comb_LAD_static	4.8933	4.1093	0.7367
Comb_LAD_dynamic	4.5240	3.8523	0.7375

1. Among the non-combined models ANN is much better than its counterparts in all considered horizons. Unlike simulated data, the difference between ANN and its counterparts is big in the medium and long horizons (h = 10 and h = 20) while the difference is smaller in the short horizon (h = 5).

2. For simple combination models, Comb_SA and Comb_BG are performing better than non-combined models other than ANN in all horizons.

3. Regression based forecast combination models are significantly improving forecasting accuracy. Dynamic versions of combination weighting give best results especially OLS combined weighting is the best model in both BIST30 and BIST100 data in all horizons.

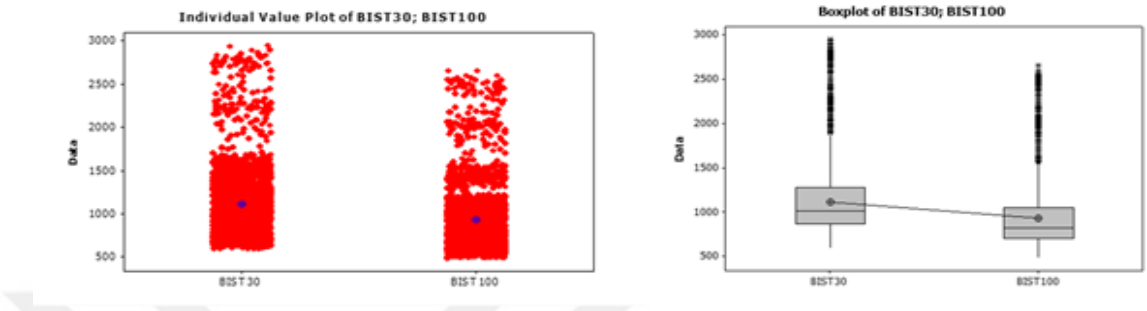


Figure 5.5 BIST30 and BIST100 Means and Variations

4. Both BIST30 and BIST 100 time series graphs show similar behavior and structural break breaks, but BIST100 RMSEs are smaller in all cases. This could be explained that BIST100 includes more companies, so its average and standard deviation are lower than BIST30 as Figure 5.5 shows. The statistical significance of the difference between the variances of the two series (BIST30 and BIST100) was tested with the F test and Levene test and showed that there is a statistically significant difference between the two variances. The significance of difference between the means of the two series was tested with the Mann-Whitney test and found it significance. Therefore, our models perform better in BIST100 data because of the less variation compared with BIST30.

5. Although the results obtained in our study under individual models are partially parallel to the results found in reviewed studies, our application forecast combination models and comparison of combining forecast performances under different conditions is a contribution to the literature.

5.4 Chapter Conclusion

This study aimed to examine if the generally accepted concept regarding prominence of forecasts combination in forecasting accuracy is sustained in specific cases of structural breaks and conditional heteroscedasticity. These are well known phenomena in financial time series. This paper also explored which combination schemes are optimal in those cases by combining ETS, ARIMA and ANN models. Simple and regression-based combination techniques are used to combine the individual components of forecasts. These methods are implemented using simulation and empirical based data.

The study found that regression-based forecast combination methods are the best forecast combination schemes in the cases of structural breaks and conditional heteroscedasticity. Allowing time-varying combination weights (dynamic version) shows a significant improvement in accuracy compared with the static counterparts if the time series under consideration contains trend. In contrast, dynamic combination weights do not change much in the case of trend absence in comparison with its static version. Although simple combination schemes including simple average model exhibit wonderful performance in the literature of forecasting, they just perform better than ETS and ARIMA but hardly beat ANN. This is an indication that structural break(s) affect the performance of simple combination strategies. Furthermore, ANN is the best individual model in all cases of simulated and empirical data compared to its single counterparts and even outperforms simple forecast combination methods in most of the cases.

CHAPTER SIX
CONCLUSION

In this thesis we focused on forecasting accuracy of several forecasting models under structural breaks and ARCH effects. Firstly, we examined the impact of structural break location and break size on predictive performance of four competing methods (ARIMA, Exponential, ARIMA - ARCH and Artificial Neural Network) in various cases of simulations and empirical data. The first quartile (Q1), second quartile (Q2), and third quartile (Q3) of the sample time series data are the three test sites where the impact of the break positions is examined. We also considered three different break magnitudes ($\delta = 3$, $\delta = 5$ and $\delta = 10$) and three lengths of horizons ($h = 3$, $h = 6$, $h = 10$).

In this application, we found that apart from the impact of structural break on the forecasting accuracy as a general case, break location and break size have implications on prediction. This result is found both in simulations and empirical analysis. Forecasting error increases with the increase of break magnitude and the location of the structural break. The error is minimum when the break is in the first quartile and break size is small ($\delta = 3$), it is more when the break is in the second quartile and break size is moderate ($\delta = 5$), and it is the most when break locates in the third quartile and break size is comparatively large ($\delta = 10$).

Comparing the competing models, ARCH model is the best performing model in the absence of structural break while ANN is the next in performance. For the cases of the various structural break sizes and break locations, ANN outperforms all the other models except for a few places. ANN also shown superiority in the empirical analysis in most of the cases under investigation. Hence, while structural break affects the performance of all the examined models, ANN is the most resilient for the breaks and instability of the parameters.

To deeply study the impact of the structural break on forecasting accuracy, we conducted another investigation to scrutinize this matter more especially when the break is in the third quartile of the sample. We examined if the generally accepted

concept regarding prominence of forecasts combination in forecasting accuracy is sustained in specific cases of structural breaks and conditional heteroscedasticity. We also explored which combination schemes are optimal in those cases by combining ETS, ARIMA and ANN models. Simple and regression-based combination techniques are used to combine the individual components of forecasts. These methods are implemented using simulation and empirical based data.

The findings showed that regression-based forecast combination methods are the best forecast combination schemes in the cases of structural breaks and conditional heteroscedasticity. Allowing for time-varying combination weights (dynamic version) shows a significant improvement in accuracy comparing with the static counterparts if the time series under consideration contains trend. In contrast, dynamic combination weights do not change much in the case of trend absence in comparison with its static version. Although simple combination schemes including simple average model exhibit wonderful performance in the literature of forecasting, they just perform better than ETS and ARIMA but hardly beat ANN. This is an indication that structural break(s) affect the performance of simple combination strategies. Furthermore, ANN is the best individual model in all cases of simulated and empirical data compared to its single counterparts and even outperforms simple forecast combination methods in most of the cases.

This study recommends that the impact of structural break location on forecasting accuracy to be examined more using deep learning methods such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory Networks (LSTMs). These methods may improve the forecast accuracy since they offer a lot of promise for time series forecasting. Moreover, separate effects of either structural breaks or ARCH innovations need to be investigated as this thesis focused on the combined effect of the two features.

REFERENCES

Acaravci, A., & Erdogan, S. (2016). The convergence behavior of CO2 emissions in seven regions under multiple structural breaks. *International Journal of Energy Economics and Policy*, *6*(3), 575–580.

Adhikari, R., Bijari, M., & Zhang, G. P. (2003). Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model. *Neurocomputing*, *50*, 159–175. www.elsevier.com/locate/neucom

Ahmed, M., Haider, G., & Zaman, A. (2017). Detecting structural change with heteroskedasticity. *Communications in Statistics - Theory and Methods*, *46*(21), 10446–10455. https://doi.org/10.1080/03610926.2016.1235200

Aker, Y. (2022). Analysis of Price Volatility in BIST 100 Index With Time Series: Comparison of Fbprophet and LSTM Model. *European Journal of Science and Technology*, *35*, 89–93. https://doi.org/10.31590/ejosat.1066722

Alp, S., Yiğit, Ö. E., & Öz, E. (2020). Prediction Of BIST Price Indices: A Comparative Study Between Traditional and Deep Learning Methods. Sigma Journal of Engineering and Natural Sciences, 38(4), 1693 – 1704.

Altansukh, G., & Osborn, D. R. (2022). Using structural break inference for forecasting time series. *Empirical Economics*, *63*(1), 1–41. https://doi.org/10.1007/s00181-021-02137-w

Andrawis, R. R., Atiya, A. F., & El-Shishiny, H. (2011). Combination of long term and short term forecasts, with application to tourism demand forecasting. *International Journal of Forecasting*, *27*(3), 870–886. https://doi.org/10.1016/j.ijforecast.2010.05.019

Andreou, E., Ghysels, E. (2009). Structural Breaks in Financial Time Series. In: Mikosch, T., Kreiß, JP., Davis, R., Andersen, T. (eds) Handbook of Financial Time Series. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-71297-8_37

Andrews, D. W. K. (1993). Tests for Parameter Instability and Structural Change With Unknown Change Point. *Econometrica*, *61*(4), 821. https://doi.org/10.2307/2951764

Antoch, J., Hanousek, J., Horváth, L., Hušková, M., & Wang, S. (2019). Structural breaks in panel data: Large number of panels and short length time series. *Econometric Reviews*, *38*(7), 828–855. https://doi.org/10.1080/07474938.2018.1454378

Aygören, H., Sarıtaş, H., & Moralı, T. (2012). İMKB 100 Endeksinin Yapay Sinir Ağları ve Newton Nümerik Arama Modelleri ile Tahmini. *Uluslararası Alanya İşletme Fakültesi Dergisi*, *4*(1), 73–88.

Babu, C. N., & Reddy, B. E. (2014). A moving-average filter based hybrid ARIMA-ANN model for forecasting time series data. *Applied Soft Computing Journal*, *23*, 27–38. https://doi.org/10.1016/j.asoc.2014.05.028

Bai, B. Y. J., & Perron, P. (1998). *Estimating and Testing Linear Models with Multiple Structural Changes Author (s): Jushan Bai and Pierre Perron Published by : The Econometric Society Stable URL : https://www.jstor.org/stable/2998540*. *66*(1), 47–78.

Bai, J., & Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, *18*(1), 1–22. https://doi.org/10.1002/jae.659

Bates, A. J. M., & Granger, C. W. J. (1969). *The Combination of Forecasts Stable URL : http://www.jstor.org/stable/3008764*. *20*(4), 451–468.

Bauwens, L., Dufays, A., & DE BACKER, B. (2011). Estimating and forecasting structural breaks in financial time series. Center for Operations Research and Econometrics.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. Journal of Econometrics , 31,307–327.

Bouri, E., Dutta, A., & Saeed, T. (2021). Forecasting ethanol price volatility under structural breaks. *Biofuels, Bioproducts and Biorefining*, *15*(1), 250–256. https://doi.org/10.1002/bbb.2158

Box, G. and Jenkins, G. (1970). Time series analysis: forecasting and control. Holden-Day.

Brown, R. L., Durbin, J., & Evans, J. M. (1975). Techniques for Testing the Constancy of Regression Relationships Over Time. *Journal of the Royal Statistical Society:*, *37*(2), 149–163. https://doi.org/10.1111/j.2517-6161.1975.tb01532.x

Chow, G. C. (1960). Tests of Equality Between Sets of Coefficients in Two Linear Regressions. Econometrica, 28(3), 591-605. https://www.jstor.org/stable/1910133

Claeskens, G., Magnus, J. R., Vasnev, A. L., & Wang, W. (2016). The forecast combination puzzle: A simple theoretical explanation. *International Journal of Forecasting*, *32*(3), 754–762. https://doi.org/10.1016/j.ijforecast.2015.12.005

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, *5*(4), 559–583. https://doi.org/10.1016/0169-2070(89)90012-5

Clements, M.P. and Hendry, D.F. (1998). Forecasting Economic Time Series, Cambridge University Press.

Clements, M.P. and Hendry D.F (1999). Forecasting Non-stationary Economic Time Series, The MIT Press.

Crane, D. B. and Crotty, J. R. (1967). A two-stage forecasting model: Exponential smoothing and multiple regression. Management Science, 13(8):501–507. https://doi.org/10.1287/mnsc.13.8.B501.

De Gaetano, D. (2018). Forecast combinations in the presence of structural breaks: Evidence from U.S. equity markets. *Mathematics*, *6*(3), 1–19. https://doi.org/10.3390/math6030034

Engle, R.F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. Econometrica, 50 (4), 987-1007.

Fischer, T. ;, Krauss, C. ;, & Treichel, A. (2018). Machine learning for time series forecasting - a simulation study. *FAU Discussion Papers in Economics*, *02*, No. 02/2018. http://hdl.handle.net/10419/173659www.econstor.eu

Ghalanos, A. (2015). Introduction to the rugarch package. (GARCH models). *October*, 1–36.

Granger, C. W. and Ramanathan, R. (1984). Improved methods of combining forecasts. Journal of Forecasting, 3 (2):197–204,. URL https://doi.org/10.1002/for.3980030207.

Hansen, B.E. (1991). a comparison of tests for parameter instability: an examination of asymptotic local power. University of Rochester.

Hansen, B. E. (2001). The New Econometrics of Structural Change. *Journal of Economic Perspectives*, *15*(4), 117–128.

Hansen, B. E. (2005). Challenges for econometric model selection. *Econometric Theory*, *21*(1), 60–68. https://doi.org/10.1017/S0266466605050048

Holt, C. C., (2004). *Forecasting seasonals and trends by exponentially weighted moving averages*. *20*, 5–10. https://doi.org/10.1016/j.ijforecast.2003.09.015

Hsiao, C., & Wan, S. K. (2014). Is there an optimal forecast combination? *Journal of Econometrics*, *178*(PART 2), 294–309. https://doi.org/10.1016/j.jeconom.2013.11.003

Hyndman, R. J., & Khandakar, Y. (2008). *Automatic Time Series Forecasting : the forecast Package for R*. *27*(3). https://doi.org/10.18637/jss.v000.i00

Hyndman, R.J.,(2012). New in forecast 4.0. December 2012. https://robjhyndman.com/hyndsight/forecast4/

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting : Principles and Practice*.

OTexts, Melbourne. OTexts.com/fpp2.

Inoue, A., & Rossi, B. (2010). *Identifying the Sources of Instabilities in Macroeconomic Fluctuations*. 1–48.

Kaushik, R., Jain, S., Jain, S., & Dash, T. (2021). Performance evaluation of deep neural networks for forecasting time-series with multiple structural breaks and high volatility. *CAAI Transactions on Intelligence Technology*, *6*(3), 265–280. https://doi.org/10.1049/cit2.12002

Kim, C. J., Morley, J. C., & Nelson, C. R. (2005). The structural break in the equity premium. *Journal of Business and Economic Statistics*, *23*(2), 181–191. https://doi.org/10.1198/073500104000000352

Kim, S., & Kim, H. (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, *32*(3), 669–679. https://doi.org/10.1016/j.ijforecast.2015.12.003

Knoll, K., Schularick, M., & Steger, T. (2014). No Price Like Home: Global House Prices, 1870-2012. Center for Economic Studies and ifo Institute (CESifo), Munich. http://hdl.handle.net/10419/103123

Koop, G., & Potter, S. (2001). Are Apparent Findings of Nonlinearity Due to Structural Instability in Economic Time Series? The Econometrics Journal, 4(1), pp. 37–55. https://doi.org/10.2139/ssrn.163151

Lee, T. H., Parsaeian, S., & Ullah, A. (2022). Optimal forecast under structural breaks. *Journal of Applied Econometrics*, *37*(5), 965–987. https://doi.org/10.1002/jae.2908

Nowotarski, J., Raviv, E., Trück, S., & Weron, R. (2014). An empirical comparison of alternative schemes for combining electricity spot price forecasts. *Energy Economics*, *46*, 395–412. https://doi.org/10.1016/j.eneco.2014.07.014

Pakel, C., & Özen, K. (2021). Daily Volatility Analysis of Bist 100 Constituents Between 2018-2020. *M U Iktisadi ve Idari Bilimler Dergisi*, *42*(2), 340–360.

https://doi.org/10.14780/muiibd.854509

Palm, F. C. and Zellner, A. (1992). To combine or not to combine? issues of combining forecasts. Journal of Forecasting, 11(8):687–701. https://doi.org/10.1002/for.3980110806.

Pesaran, M. H., & Timmermann, A. (2007). Selection of estimation window in the presence of breaks. *Journal of Econometrics*, *137*(1), 134–161. https://doi.org/10.1016/j.jeconom.2006.03.010

Pettenuzzo, D., & Timmermann, A. (2011). Predictability of stock returns and asset allocation under structural breaks. *Journal of Econometrics*, *164*(1), 60–78. https://doi.org/10.1016/j.jeconom.2011.02.019

Ploberger, W., & Krämer, W. (1992). The Cusum Test with Ols Residuals Author (s): Werner Ploberger and Walter Krämer Published by : The Econometric Society Stable URL : http://www.jstor.org/stable/2951597 REFERENCES Linked references are available on JSTOR for this article : You may need. *Econometrica*, *60*(2), 271–285.

Rapach, D. E., & Wohar, M. E. (2005). Regime Changes in International Real Interest Rates: Are They a Monetary Phenomenon? *Journal of Money, Credit, and Banking*, *37*(5), 887–906. https://doi.org/10.1353/mcb.2005.0057

Raşo, H., & Demirci, M. (2019). Predicting the Turkish Stock Market BIST 30 Index using Deep Learning. International Journal of Engineering Research and Development, 11(1), 253 - 265. https://doi.org/10.29137/umagd.425560.

Rossi, B. (2021). Forecasting in the Presence of Instabilities: How We Know Whether Models Predict Well and How to Improve Them. *Journal of Economic Literature*, *59*(4), 1135–1190. https://doi.org/10.1257/JEL.20201479

Quandt, R.E. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes. Journal of the American Statistical Association, 53(284), 873-880. https://www.jstor.org/stable/2281957

Quandt, R.E. (1960). Tests of the hypothesis that a linear regression system obeys two separate regimes. Journal of the American Statistical Association 55(290), 324-330

Shmueli, G., & Jr., K. C. L. (2016). *Practical Time Series Forecasting: A Hands-On Guide*. 202. http://galitshmueli.com/practical-time-series-forecasting-book

Siliverstovs, B., & van Dijk, D. (2002). Forecasting Industrial Production with Linear, Nonlinear, and Structural Change Models. *Econometric Institute Report EI 2003-16*.

Sjölander, P. (2011). A stationary unbiased finite sample ARCH-LM test procedure. *Applied Economics*, *43*(8), 1019–1033. https://doi.org/10.1080/00036840802600046

Smith, J., & Wallis, K. F. (2009). A simple explanation of the forecast combination puzzle. *Oxford Bulletin of Economics and Statistics*, *71*(3), 331–355. https://doi.org/10.1111/j.1468-0084.2008.00541.x

Soy Temür, A., & Yıldız, Ş. (2021). Comparison of Forecasting Performance of ARIMA LSTM and HYBRID Models for The Sales Volume Budget of a Manufacturing Enterprise. In *Istanbul Business Research* (Vol. 50, Issue 1). https://doi.org/10.26650/ibr.2021.51.0117

Stock, J. H., & Watson, M. W. (1996). Evidence on Structural Instability in Macroeconomic Time Series Relations. Journal of Business & Economic Statistics,14:3,11–30

Stock, J. H., & Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, *23*(6), 405–430. https://doi.org/10.1002/for.928

Taylor, S. (1984). Modelling Financial Time Series, Wiley,New York.

Tellİ, Ş., &COŞKUN, M. (2016). Forecasting the BIST 100 Index Using Artificial Neural Networks with Consideration of the Economic Calendar Forecasting the

BIST 100 Index Using Artificial Neural Networks. International Review of Economics and Management, 4(3), 26–46. https://doi.org/10.18825/irem.67309.

Tian, J., & Anderson, H. M. (2014). Forecast combinations under structural break uncertainty. *International Journal of Forecasting*, *30*(1), 161–175. https://doi.org/10.1016/j.ijforecast.2013.06.003

Ünvan, Y. A., & Ergenç, C. (2022). Forecasting BIST 100 Index With Artificial Neural Networks and Regression Analysis. *BİLTÜRK Journal of Economics and Related Studies*, *4*(1), 20–32. https://doi.org/10.47103/bilturk.1039669

Valadkhani, A., Allan P. Layton, & Mosayeb Pahlavani. (2005). Multiples Structural Breaks in Australia's Macroeconomic data: An Application of the Lumsdaine and Papell Test. *International Journal of Applied Econometrics and Quantitative Studies*, *2*.

Weiss, C. E., Raviv, E., & Roetzer, G. (2019). Forecast combinations in R using the ForecastComb package. *R Journal*, *10*(2), 262–281. https://doi.org/10.32614/RJ-2018-052

Winkler, R. L., & Makridakis, S. (1983). The Combination of Forecasts. *Journal of the Royal Statistical Society. Series A (General)*, *146*(2), 150. https://doi.org/10.2307/2982011

Winters, P.R., (1960). Forecasting Sales by Exponentially Weighted Moving Averages. Management Science, 6(3), 324-342. https://www.jstor.org/stable/2627346.

Yin, A. (2014). *Out-of-Sample Forecast Model Averaging with Parameter Instability*. *515*, 1–40.

Yılmaz, K., & Kale, S. (2022). Risk ve Finansal Göstergeler Arasindaki Asimetrik İlişki: BIST İmalat Sektöründe Bir Uygulama. *Ekoist: Journal of Econometrics and Statistics*, *0*(0), 0–0. https://doi.org/10.26650/ekoist.2022.36.1035097

Zeileis, A., Leisch, F., Homik, K., & Kleiber, C. (2002). strucchange: An R Package

for Testing for Structural Change. *Journal of Statistical Software*, *7*(2), 1–38.