Hacettepe University Graduate School of Social Sciences

Department of Economics

ON THE YIELD CURVE FORECASTING: APPLICATIONS OF CONVENTIONAL AND NON-CONVENTIONAL TECHNIQUES

Hakan GENÇSOY

Ph.D Dissertation

Ankara, 2025

ON THE YIELD CURVE FORECASTING: APPLICATIONS OF CONVENTIONAL AND NON-CONVENTIONAL TECHNIQUES

Hakan GENÇSOY

Hacettepe University Graduate School of Social Sciences

Department of Economics

Ph.D Dissertation

Ankara, 2025

ACCEPTANCE AND APPROVAL

The jury finds that Hakan GENÇSOY has on the date of 06/12/2024 successfully passed the defense examination and approves his Ph.D. Dissertation titled "On The Yield Curve Forecasting: Applications of Conventional and Non-Conventional Techniques".

Prof. Dr. Lütfi ERDEN (Jury President)

Prof. Dr. Başak DALGIÇ (Main Adviser)

Prof. Dr. Özge KANDEMİR KOCAASLAN

Prof. Dr. İbrahim ÖZKAN

Prof. Dr. Burak GÜNALP

I agree that the signatures above belong to the faculty members listed.

Prof. Dr. Uğur ÖMÜRGÖNÜLŞEN

Graduate School Director

YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI BEYANI

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kağıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe Üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinlerin yazılı izin alınarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayınlanan ***"Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge"*** kapsamında tezim aşağıda belirtilen koşullar haricince YÖK Ulusal Tez Merkezi / H.Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

- o Enstitü / Fakülte yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren 2 yıl ertelenmiştir. (1)
- o Enstitü / Fakülte yönetim kurulunun gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren ... ay ertelenmiştir. (2)
- o Tezimle ilgili gizlilik kararı verilmiştir. (3)

06/12/2024

Hakan GENÇSOY

[1] *"Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge"*

(1) *Madde 6. 1. Lisansüstü tezle ilgili patent başvurusu yapılması veya patent alma sürecinin devam etmesi durumunda, tez* ***danışmanının*** *önerisi ve* ***enstitü anabilim dalının*** *uygun görüşü üzerine* ***enstitü*** *veya* ***fakülte yönetim kurulu*** *iki yıl süre ile tezin erişime açılmasının ertelenmesine karar verebilir.*

(2) *Madde 6. 2. Yeni teknik, materyal ve metotların kullanıldığı, henüz makaleye dönüşmemiş veya patent gibi yöntemlerle korunmamış ve internetten paylaşılması durumunda 3. şahıslara veya kurumlara haksız kazanç imkanı oluşturabilecek bilgi ve bulguları içeren tezler hakkında tez* ***danışmanının*** *önerisi ve* ***enstitü anabilim dalının*** *uygun görüşü üzerine* ***enstitü*** *veya* ***fakülte yönetim kurulunun*** *gerekçeli kararı ile altı ayı aşmamak üzere tezin erişime açılması engellenebilir.*

(3) *Madde 7. 1. Ulusal çıkarları veya güvenliği ilgilendiren, emniyet, istihbarat, savunma ve güvenlik, sağlık vb. konulara ilişkin lisansüstü tezlerle ilgili gizlilik kararı,* ***tezin yapıldığı kurum*** *tarafından verilir *. Kurum ve kuruluşlarla yapılan işbirliği protokolü çerçevesinde hazırlanan lisansüstü tezlere ilişkin gizlilik kararı ise,* ***ilgili kurum ve kuruluşun önerisi*** *ile* ***enstitü*** *veya* ***fakültenin*** *uygun görüşü üzerine* ***üniversite yönetim kurulu*** *tarafından verilir. Gizlilik kararı verilen tezler Yükseköğretim Kuruluna bildirilir.*
Madde 7.2. Gizlilik kararı verilen tezler gizlilik süresince enstitü veya fakülte tarafından gizlilik kuralları çerçevesinde muhafaza edilir, gizlilik kararının kaldırılması halinde Tez Otomasyon Sistemine yüklenir

******* *Tez* ***danışmanının*** *önerisi ve* ***enstitü anabilim dalının*** *uygun görüşü üzerine* ***enstitü*** *veya* ***fakülte yönetim kurulu tarafından karar verilir.***

ETİK BEYAN

Bu çalışmadaki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi, görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu, kullandığım verilerde herhangi bir tahrifat yapmadığımı, yararlandığım kaynaklara bilimsel normlara uygun olarak atıfta bulunduğumu, tezimin kaynak gösterilen durumlar dışında özgün olduğunu, **Prof. Dr. Başak DALGIÇ** danışmanlığında tarafımdan üretildiğini ve Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü Tez Yazım Yönergesine göre yazıldığını beyan ederim.

Hakan GENÇSOY

To my beloved family…

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisors, Prof. Dr. İbrahim Özkan and Prof. Dr. Başak Dalgıç for their invaluable help, guidance, and encouragement throughout every phase of this study. This dissertation would not have been possible without his support.

I would also like to thank my thesis committee, Prof. Dr. Özge Kandemir Kocaaslan, Prof. Dr. Lütfi Erden, and Prof. Dr. Burak Günalp for sharing their knowledge and insightful suggestions to improve this dissertation.

I want to thank my mother, Ümmühan Gençsoy, and my father, Fahrettin Gençsoy, for patiently supporting me in every aspect of my life.

Finally, I would like to express my special thanks to Sevil Kasar, Nurdan Alasun, Orhan Toptaş, Bilge Han Köse, and my valuable colleagues, who have been with me and supported me throughout this process, for their understanding, encouragement, and support.

I hope this study will be beneficial to everyone involved...

ABSTRACT

GENÇSOY, Hakan. *On The Yield Curve Forecasting: Applications of Conventional and Non-Conventional Techniques,* Ph. D. Dissertation, Ankara, 2025.

Forecasting the yield curves of government debt securities observed in the market is very crucial for both economic and financial units. Diebold and Li (2006) show that the AR(1) model makes the best forecasts with the Dynamic Nelson Siegel approach. Many researchers have later used this approach with different data and models. The differences and new approaches that emerge in the literature reveal the need for a more comprehensive study for yield curve forecasts. In this thesis, we forecast the yield curves of government bonds of G-7 countries, excluding Japan, for the period 2010-2022 using the Dynamic Nelson Siegel approach. We used conventional, non-conventional models, ensemble learning, and forecast combination models. Artificial Neural Networks (ANN), which have been frequently used in recent years and can investigate the data structure more effectively, have been used in the thesis. In the thesis, unlike the literature, we smooth the decay parameter (λ) of the Nelson-Siegel Model with the high-frequency Hodrick-Prescott Filter except for the United States. Thus, the λ parameter, like other factors, is included in the variables to be forecasted. In this thesis, we increase the flexibility of the yield curves, and the results of the forecasts are more meaningful. The results of the thesis are as follows. The yield curve fits the yields better with the float λ compared to the constant. In this thesis, we investigate models that make better forecasts than the random walk model, we show that although individual models mostly fail, successful forecasts can be easily made with forecast combination models we obtain from individual models. We also conclude that different ARIMA models can be used instead of AR(1), and that ensemble learning and forecast combination approaches can improve the forecasts of individual models. We conclude that ANN models are unstable in making successful forecasts.

Keywords

Yield Curve, Forecasting, Dynamic Nelson-Siegel, Artificial Neural Network, Forecast Combination

ÖZET

GENÇSOY, Hakan. *Getiri Eğrisi Tahmini Üzerine: Geleneksel ve Geleneksel Olmayan Tekniklerin Uygulamaları,* Doktora Tezi, Ankara, 2025.

Piyasada gözlemlenen devlet iç borçlanma senetlerinin getiri eğrisinin gelecekte ne olacağını tahmin etmek hem ekonomik hem de finansal birimleri için çok önemlidir. Diebold-Li (2006) Dinamik Nelson Siegel yaklaşımı ile yaptığı çalışmada AR(1) modelinin en iyi tahminleri yaptığını gösterir. Bu yaklaşımı daha sonra birçok araştırmacı farklı veri ve modellerde kullandılar. Literatürde ortaya çıkan farklılıklar ve yeni yaklaşımlar getiri eğrisi tahminleri için daha kapsamlı bir çalışma yapma ihtiyacı ortaya çıkartır. Bu çalışmada Japonya hariç G-7 ülkelerinin 2010-2022 dönemine ait devlet tahvillerinin getiri eğrilerini Dinamik Nelson Siegel yaklaşımı ile tahmin ederiz. Geleneksel, geleneksel olmayan modeller ile birlikte topluluk öğrenmesi ve tahmin birleştirmesi modelleri kullandık. Son yıllarda sıkça kullanılan ve veri yapısını daha etkin bir şekilde araştırabilen Yapay Sinir Ağları (YSA) tezde kullanılmıştır. Bu tezde literatürden farklı olarak, Nelson-Siegel Modelinin bozunma parametresini (λ) birleşik devletlerinki hariç yüksek frekanslı Hodrick-Prescott Filtresi ile düzleştiririz. Böylece diğer faktörler gibi λ parametresi de tahmin edilecek değişkenlere dahil edilir. Getiri eğrilerinin esnekliğini arttırdığımız bu çalışmada tahminlerin sonuçları daha anlamlıdır. Çalışmanın sonuçları şunlardır. Değişken λ ile sabite göre getiri eğrisi getirilere daha iyi uyum sağlar. Rassal yürüyüş modelinden daha iyi tahmin yapan modelleri araştırdığımız bu tezde bireysel modeller çoğunlukla başarısız olduğu halde bu modellerden elde ettiğimiz tahmin birleştirmesi modelleri ile kolaylıkla başarılı tahminler yapılabileceğini gösteririz. Ayrıca AR(1) yerine farklı ARIMA modellerinin kullanılabileceği, topluluk öğrenmesi ve tahmin birleştirmesi yaklaşımlarının bireysel modellerin tahminleri iyileştirebileceği sonucuna varırız. YSA modellerinin başarılı tahminler yapmada istikrarsız olduğu sonucuna varırız.

Anahtar Sözcükler

Getiri Eğrisi, Tahmin, Dinamik Nelson-Siegel, Yapay Sinir Ağı, Tahmin Kombinasyonu.

TABLE OF CONTENTS

ABBREVIATIONS

AI	Artificial Intelligence
ANN	Artificial Neural Network
ARIMA	Autoregressive Integrated Moving Average
BG	BG: Bates/Granger
CLS	Constrained Least Squares
DNS	Dynamic Nelson-Siegel
DT	Decision Tree
EL	Ensemble Learning
ELM	Extreme Learning Machines
FC	Forecast Combination
FFNN	Feed-forward Neural Networks
GRNN	General Regression Neural Network
HP	Hodrick-Prescott
kNN	k-Nearest Neighbors
LSTM	Long-Short Term Memory
MAE	Mean Absolute Error
ML	Machine Learning
MLP	MultiLayer Perceptron
NS	Nelson-Siegel
OLS	Ordinary Least Squares
RF	Random Forest
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Networks
RW	Random Walk
SA	Simple Avarage
SLFN	Single Hidden Layer Feed-forward Neural Network
TA	Trimmed Mean
TAR	Threshold Autoregression
XGB	Extreme Gradient Boosting
VAR	Vector Autoregression
WA	Winsorized Mean

TABLES INDEX

FIGURES INDEX

INTRODUCTION

The analysis of preventing economic uncertainties is crucial to manage potential risks towards economic crises. In this respect, understanding and adapting to changes in an economy over time, as well as well-designed and forward-looking policies are fundamental requirements to maintain a sustainable economy and balanced economic growth. One of the most important instruments of economic policy design is monetary policy and relatedly understanding the term structure of interest rates namely, the yield curve formation in an economy which is a well-defined predictor for future economic activity. Mishkin (1986) defines the yield curve, particularly as a key indicator of the bond market, where curves expresses interest rates of bonds with the same risk, liquidity, and tax conditions at different maturities. In other words, the yield curve show interest rates for debt contracts at various maturities, where these credit quality conditions determine the risk structure of the debt instruments. The risk structure, in turn, causes debt instruments with the same maturity to have different interest rates. Just as the fundamental production units of an economy are firms, they need to forecast these interest rates for planning future production and investments. Additionally, other economic actors such as the government, households, and foreigners need to understand the term structure of interest rates in order to make forecasts about their consumption and investments where debt instruments and the cost of borrowing matter.

There exist a significant number of studies on the formation of the yield curve particularly within finance, asset pricing theory, and macroeconomics literature. In this literature, the term structure of interest rates is usually analyzed to determine the relationship between maturity and zero-coupon bond yields where such analyses and forecasts allow for better controlled financial and economic practices. Within and in parallel with these studies, as the research on financial instruments, data structure, market theories, and models evolved, the yield curves have emerged as one of the most critical parameters in financial asset pricing, risk management, and portfolio management. For instance, the discount rate that determines the net value of cash flows is obtained from the yield curve while accurate yield curve forecasts reduce the risk of synthetic financial products. Yield curve forecasts further produce better fund and asset-liability management for economic actors in establishing their investment strategies.

Dynamic yield curve analysis is crucial for many purposes such as financial instrument pricing and debt restructuring. Government bonds as a fundamental financial instrument and relatedly derivative assets are priced and traded based on yield curves. To this end, forecasting the term structure of interest rates is crucial for monetary policy applications as well as for finance as the information provided by the economic indicators is reflected in the yield curve and term structure of interest rates. Obtaining the benchmark interest rates that the economy uses as a reference from the yield curve at a certain maturity level, rather than from a bond yield traded in the market, would allow for more accurate inferences for policy design. Accordingly, yield curve is one of the main instruments of central banks and monetary policy applications. On the other hand, short-term interest rates are related to monetary policy actions, whereas long-term interest rates are shaped by investors' expectations of future economic activities. Additionally, since long-term yield rates consist of future short-term interest rate expectations, the entire yield curve is related to the monetary policy rate in turn, where the monetary transmission mechanism directs the movements of short-term interest rates by influencing long-term interest rates over time.

Many studies investigate the connections between the term structure of the yield curve and macroeconomic indicators revealing the relationship between yield curve forecasts and business cycles. Level, slope, and curvature factors of the yield curves interact with many macroeconomic indicators, particularly inflation, output gap, and capacity utilization. Accordingly, market participants closely monitor these latent factors of the yield curve. Apart from monitoring, central banks try to change the shape of the yield curve from time to time in order to achieve their targets through influencing the financial and economic parameters. Therefore, forecasting the term structure of interest rates has always been critical for reflecting economic conditions and guiding monetary policy implementations.

Under free market conditions, the yield curve is a true reflection of economic agents' views on the state of the economy and their expectations. Since these expectations shape the future or positions taken by economic agents, near-term expectations are determined by the state of the yield curve where the yield curve forms a self-determining cycle. This cycle provides insights for yield curve analysis and forecasts. With this purpose, many models have been constructed to forecast the shape of the yield curve and its dynamics. Among them, conventional models such as Box-Jenkins, smoothing, threshold, and random walk have the advantages of easy application and interpretable outputs. Because of these features, they are usually preferred for the analysis and forecasting of the yield curves. However, much of the data in finance does not fit the structure of many conventional models. This leads to high forecasting and fitting errors, and many models yield worse results than those of the random-walk approach. Irregularities and variations in the yield data lead to violations of constraints such as stationarity and no missing data, which are the conditions that conventional models require for time series. Therefore, using conventional models to forecast yield curves may prevent modeling the actual patterns of them.

Issues with conventional financial data models in the yield curve forecasting have led researchers to use artificial intelligence (AI) applications. AI is applied in many financial, commercial, and economic fields, such as credit rating, forecasting bankruptcy, stock market, exchange rate determination, optimal capital structure, detecting financial crises and uncertainties as well as solving complex financial problems. For example, Machine Learning (ML) models and particularly Artificial Neural Networks (ANNs) can model high-frequency nonlinear financial data and have been of great interest for researchers and economic actors. Recently, Forecast Combinations (FC) of conventional and non-conventional models, and Ensemble Learning (EL) models have been used to complement one another. Regarding studies of yield curve forecasting using AI as well as the development of Graphics Processing Unit (GPU) technology in the 2000s has eliminated hardware problems and expanded the use of deep networks. As technological development accelerates, information can be processed faster using AI models making ML and ANN approaches more applicable with respect to conventional approaches.

Motivated by these, this study aims to forecast yield curves of G-7 countries, excluding Japan, over the period 2010-2022, mainly using the Dynamic Nelson Siegel (DNS) approach applied by Diebold and Li (2006), who forecast the Nelson Siegel (NS) factors as time series. We attempt to forecast yield curves using the conventional and non-conventional and particularly ANN models as well as using EL and FC approaches. We select Canada, Germany, France, Great Britain, Italy, and the United States (US) due to data availability for whom enough data exists to estimate latent factors efficiently. This thesis makes significant contributions to the existing literature. First of all, with a comprehensive approach, applying many models we investigate whether there is a conventional model that forecasts the structure of the data more efficiently i.e. better than the ANN models do. Further, keeping the model and parameter range wide, in k-Nearest Neighbors (kNN) and ANNs as non-conventional models, a more general optimization structure is investigated. Note that the large amount of data and hardware requirements for the increased computation volume are the most critical constraints in applying ANN models which is also valid for this study. In many studies within the related literature, ANN models provide excellent optimization which is supposed to be only coincidental. Accordingly, since optimization is problematic in ANN model applications, in this study, the structures and parameters are tested using different models. Secondly, we check the validity of comprehensive forecast combinations rather than exploring a single model. To our knowledge, there exists no study following such an investigation within the related literature. Only if the studies on Brazil are excluded, it is observed that the literature is limited in terms of the forecast combinations regarding model groups, data period, parameters used, etc.

Next, in contrast to many existing studies, by taking the parameter λ which denotes the decay parameter in Nelson-Siegel's yield curve model, as a float; and by this way including it among the factors to be forecasted instead of considering it as a constant, prevents us losing the flexibility of the yield curve. The yield curve flexibility relates with factors such as slope and curvature which include λ parameter. Accordingly, in this study, the yield curve, which becomes more flexible with a varying λ obtained by using the Hodrick Prescott Filter methodology, is expected to eliminate the issue of the non-coverage of the negative interest rates phenomena arised recently. Last but not least, unlike the existing studies, the sliding-window methodology is used to determine if the models would provide a random pattern causing them to be dysfunctional in some parts of data. The sliding window technique refers to a method of analyzing time series data by using a fixed-size subset of data points that moves through the entire dataset, allowing for the examination of changing relationships over time.

The remainder of this thesis is organized as follows. Chapter 1 labeled as "Yield Curve and Yield Curve Forecasting: A Brief Literature" introduces the key definitions and concept of the yield curve, term structure theories, and yield curve models as well as the forecasting literature. The second chapter labeled as "Considered Forecasting Methodologies" overviews conventional, non-conventional, ensemble learning and forecasting combination models employed. In Chapter 3 of “Data and Empirical Framework”, datasets and methodological frameworks are introduced. The last chapter is designed as "Empirical Analyses", where the evaluation methods and the empirical results of the models are presented.

CHAPTER 1:
YIELD CURVE AND YIELD CURVE FORECASTING: A BRIEF LITERATURE

The bond market is vital for economic activity financing the activities of companies and governments as well as determining the interest rates in an economy. Finance has many interest rates such as loans, markets, and benchmark interest rates. Due to the tendency of different interest rates to move together, the field of economics refers to a single interest rate which is a combination of these rates. Economists refer to this interest rate as the yield to maturity (Mishkin, 2021). The most accepted method of measuring the term structure of interest rates is yield curves identical to current rates or zero (interest) rates. Zero coupons are selected because there is no coupon effect. Zero coupon yields are not directly observed in the market; therefore, they must be estimated using yield curve models (De Pooter, 2007).

Yield curve models are designed based on their intended purpose. For instance, in macroeconomics, a smoother model is used to analyze the determinants of the curve, whereas a method that offers a better fit and considers volatility is preferred for pricing securities (Gürkaynak et al., 2007). The model for the yield curve should capture the general behavior of interest rates. The mathematical structure of the model determines its purpose and how well it fulfills these criteria. The evolution of models over time reflects the ongoing efforts to meet these needs.

This section begins by introducing the key definitions and equations used to develop yield curve models. It explains the theories behind the yield curve and the features of its shape. It covers models of the yield curve with different purposes and assumptions, their general characteristics, and the Dynamic Nelson-Siegel (DNS) model used in this thesis. Also, NS factors are mentioned. The section ends with a comprehensive literature review of the yield curve forecast methodologies.

1.1. INTEREST RATES

This title includes the definitions and equations that form the basis of yield curve models. The NS model used in this study is derived from these formulas. To begin with, a loan issued in the current period (at time t) has an interest rate r (t, T) and is repaid at time T. The yield rates at all times T are assembled to produce a yield curve (Davidson, 2014). Let the interest rate be constant within each interval but different at different intervals. Let r(t) be the instantaneous interest rate, over time interval [0, T], present value P and value (V) be

$$V(t) = P.\exp\left(\int_0^T r(t)dt\right) \quad (1.1)$$

(Campolieti & Makarov, 2014). The yield of a zero-coupon bond is the yield of a single payment at maturity and defines the implied yield. If the nominal value of the bond is 1:

$$P_t(\mathrm{m}) = e^{-\mathrm{m}r_t(m)} \quad (1.2)$$

$P_t(\mathrm{m})$ is price of the m-period discounted bond. $r_t(\mathrm{m})$ is continuously compounded nominal yield to maturity of a zero coupon bond (Diebold & Li, 2006).

Even if the interest rates between the maturities of the cash flows of a debt instrument are different, the rate obtained for equalizing the market value of the debt instrument is called yield to maturity (Davidson, 2014). Calculating the yield to maturity can sometimes be challenging. Instead, discount-based yields are used (Mishkin, 2021). The discount rate is the rate that equalizes the security value that will earn a return at maturity to its present value. The discount rate is calculated from the zero coupon yield curve (Akçay et al., 2012). A discount curve is obtained from the yield curve.

Forward interest rates are current period equivalents of future zero-coupon rates (Hull, 2018). Zero-coupon bond rates are forward rates issued at time t_0 (Davidson, 2014). The instantaneous (nominal) forward curve is obtained from the discount rate:

$$f_t(\mathrm{m}) = -P_t'(\mathrm{m})/P_t(\mathrm{m}) \quad (1.3)$$

(Diebold & Li, 2006). The spot interest rate (yield) is the average forward rate. The continuous compounded yield of a zero-coupon bond in period n is

$$y_t(n) = \frac{1}{n}\int_0^n f_t(m)dm = R(m) \quad (1.4)$$

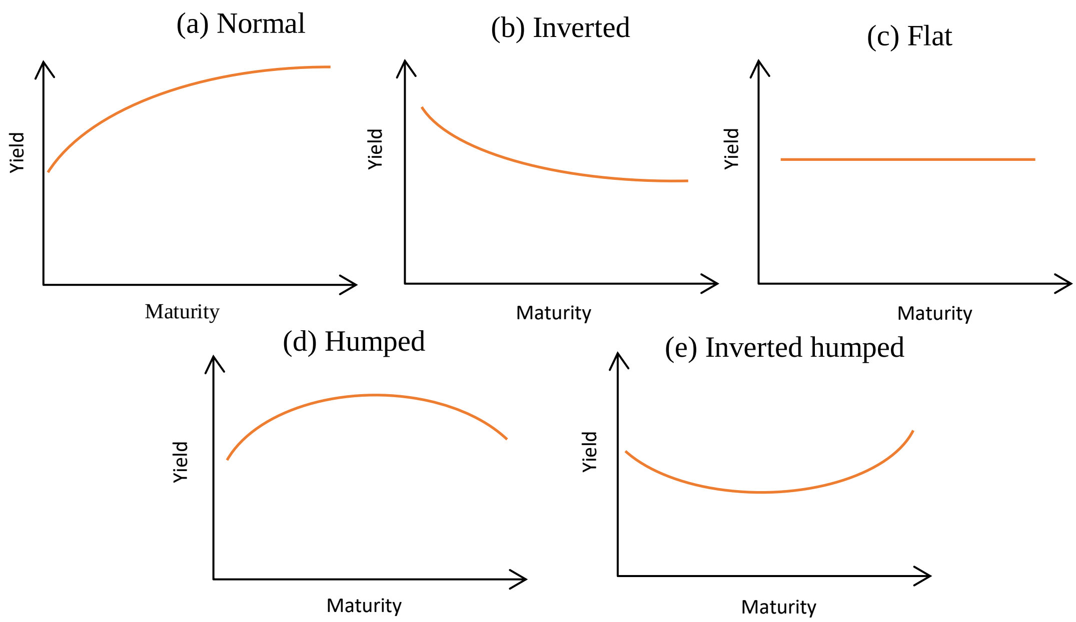
Where $f_t(m)$ defines forward rate and $R_t(m)$ is spot interest rate (Gürkaynak et al., 2007). Since these are data on long-term rates, forward interest rates contain the same information as the yield curve if the term premium is not considered (Svensson, 1995).

1.2. YIELD CURVE

Although there is no objection to the definition of the yield curve, the data to be used, their transformation, and the yield curve models vary. Using at least ten different securities data is recommended to reliably produce a yield curve (Akçay et al., 2012). Since zero coupon bond prices are not available in the market for all maturities, unobservable rates could be found by estimating the yield curve from the coupon prices of observed zero coupon bonds and coupons of coupon bonds (Gürkaynak et al., 2007). In generating zero-coupon yield curves for government bonds, zero-coupon bonds, as well as coupons of coupon bonds, can be used as inputs for separate security (Hull, 2018). However, in coupon bonds, the coupon rate changes the bond's actual maturity, which we call the duration. Additionally, if discount and coupon bonds are traded in two markets, their pricing dynamics may differ. Therefore, some researchers have stated that coupon rates are inappropriate to be used in the yield modeling (Svensson, 1995).

Yield curves are usually positively sloped but can be formed negatively sloped, flat, or very complex (Mishkin, 2021). While interest rates are expected to rise as maturity gets longer, long-term rates may be lower for various reasons, such as economic confidence (Kožíšek, 2018). The term structure cannot be described as a first-order Markov process because investors have information not included in the current yields (Duffee, 2012).

Figure 1: Types of Yield Curves



Explanations of the yield curve graphs, examples of which are shown in Figure 1, are as follows. (a) A normal yield curve has a positive slope, where yields increase as maturity increases. (b) In an inverted yield curve, long-term yields are lower than short-term yields. These negatively sloping yield curves are rare. (c) The yield spreads between maturities on a flat yield curve are very small. (d) In a humped yield curve, rising interest rates fall after reaching the maximum point of curvature and return to their initial levels. (e) In the inverted hump yield curve, interest rates slope downward in the short term and then increase again in the long term, reaching their initial level in the long term.

There is a relationship between the shape of the yield curve and the economy. The slope combines the term premium and interest rate expectations (and hence, inflation expectations). From a horizontal or inverted yield curve, it can be concluded that there is an expectation that short-term returns, economic activities, and inflation may decrease in the future. Fund holders who think that the economy will enter a recession may find that long-term demand for funds may decrease, future interest rates may decrease, and therefore higher yields will be obtained in the short term. Another possibility is that the contractionary monetary policy that causes the inverted yield curve could only reduce interest rates and inflation without causing a slowdown (Central Bank of Türkiye - Inflation Report, 2019-II). News of an economic upturn may increase short-term rates, leading to an inverted yield curve (Diebold & Rudebusch, 2013). A flat yield curve may signal the expectation of disinflation and slow growth (Central Bank of Türkiye - Inflation Report, 2019-II). On the other hand, a steeply sloping yield curve indicates that inflation will rise and a loose policy will follow (Mishkin, 2021). Changes in yield curve are signs of recession or economic recovery (Diebold & Rudebusch, 2013).

1.3. STYLIZED FACTS

The yield curve must have specific properties that must be met to be modeled. In 2006, Diebold and Li listed five of these critical properties known as stylized facts. First, yield curve is concave and increases in general. The normal yield curve shown in Figure 1(a) has these characteristics and possesses the term structure most commonly observed in the market. Second, the yield curve may assume various shapes over time, such as upward sloping, downward sloping, humped, and inverted-humped. These possible yield curve graphs are shown in Figure 1. Third, the yield dynamics are persistent (resilient, less volatile), whereas the spread dynamics are less persistent. Fourth, the short-term end of the yield curves is more volatile than the long-term end. Fifth, long-term rates are more persistent than short-term ones.

Many different observations have been made regarding the yield behavior. Some of these findings, which explain the shape and changes in the yield curve, can be summarized as follows. According to convexity in the compound-modified duration of the bond, capital loss from an increase in the interest rate is smaller than capital gain from a decrease in the rate. This causes the yield curve to become convex and hump-shaped (Gürkaynak et al., 2007). Yield curves are generally positively sloped, except when a sharp decrease in short-term rates is expected (Hull, 2018). In a developed bond market with a so-called "normal" positive yield curve, selling the bond before maturity is possible. In this way, despite the decreasing yield ratio as maturity decreases, a higher yield than the short-term yield can be obtained by selling the bond before its maturity (Campolieti & Makarov, 2014). It is rare for interest rates to change in the same way along the yield curve; that is, for the curve as a whole to shift by the same amount. The yield curve's slope usually changes because of changes in interest rates (Şişman, 2011).

Yield curves are not always expected to have all shapes because of arbitrage, which means that debt instruments with different yields can be traded for instant gain. For example, according to Davidson (2014), negative nominal interest rates are impossible. Of course, negative actual interest rates are also possible. Very low interest rates can be explained by low or negative inflation and lack of investment opportunities. However, the negative interest rates go beyond these reasons. The reason is that large investors and banks prefer to keep treasury bonds or central bank deposits to store their funds electronically despite the negative yields (Mishkin, 2021).

The interest rate risk, which refers to the risk of an asset arising from changes in interest rates, increases the risk of bonds with longer maturities. Prices and yields of bonds with longer maturities are more volatile because they are more sensitive to changes in interest rates (Mishkin, 2021). The risk of the bond is the difference between the actual behavior of the short-term rate and the behavior in the risk-free world and is priced positively. Interest rate risk is negatively priced because price and rate are negatively related. In the risk-free world, interest rates change more and have a larger expected future value (Hull, 2018).

Yield curve shows the bond market yield and reflects the market expectations. Interest rates are not expected to fluctuate as uncertainty decreases as the redemption date approaches. However, because there is more uncertainty in the long term, investors may demand a higher yield for this risk. This is why the yield curve typically slopes upwards. Sometimes, yields are higher in the short term and are expected to fall in the future (Campolieti & Makarov, 2014).

High yield spreads result in low future yield spreads (Hoogteijling 2020). According to Diebold and Rudebusch (2013), 12-month autocorrelations of yields exhibit persistent solid features. However, in contrast to these yield levels, spread autocorrelations, which are initially high, decline rapidly and remain smaller than level autocorrelations.

1.4. THEORIES FOR TERM STRUCTURE OF INTEREST RATES

The economic literature attempts to explain why the same debt instruments may have different yields. Theories have been developed to explain why the yields differ, instead of providing the same yields even if the maturity varies. Three different theories have come to the forefront to explain the term structure of yield curves. The expectations theory suggests that short-term current yields are the averages of future forward yields. Market segmentation theory suggests that short-, medium-, and long-term markets are entirely different. The liquidity premium theory suggests that investment in short-term instruments is preferred to preserve liquidity, whereas long maturity is preferred for borrowing (Hull, 2018).

The Expectations Theory, explains the empirical properties of the yield curve that (1) interest rates move together and (2) if short-term interest rates are low (high), the curve is most likely to be positively (negatively) sloped. The Market Segmentation Theory, explains the third empirical fact that (3) yield curves are generally positively sloped. Since both theories can explain the empirical features that the other cannot, the Liquidity Premium Theory was developed by combining the two theories to explain all three empirical features (Mishkin, 2021).

The average of the short-term interest rates that investors expect to realize during the maturity of the long-term bond is equal to the interest rate on the long-term bond. Since future short-term interest rates are expected to take different values, interest rates at different maturities are not equal. In the Expectations Theory, this claim assumes that bonds with different maturities are perfect substitutes. According to this theory, short-term interest rates increase when the yield curve has a positive slope. Moreover, according to this theory, since the long-term is expected to be the average of the short-term, interest rates in the long term are less volatile than interest rates in the short term. Because extreme deviations are expected in the short term, the yields eventually revert to the mean in the long term (Mishkin, 2021).

The Expectations Theory assumes that there is a relationship between the slope and future level of yields. Since this relationship is assumed to be linear in many studies, appropriate results could not be obtained from the studies (Modena, 2008). However, based on this relationship, indicated by the Expectations Theory, yield forecasting models can be established in which only term structure factors are considered inputs (parameters) and output. Litterman and Scheinkman (1991) and Jong (2000) claim that interest rate forecasts could only be made using term structure factors. Wood and Dasgupta (1995), Tappinen (1998), and Abid and Salah (2003) demonstrate this with the models they built with spreads and interest rate levels.

The Expectations Theory does not model the excess yields. This theory considers the risk premium constant and uses the average of historical data in its forecast. Although this theory does not relate risk premiums to any data, many studies have modeled them using macroeconomic data (Hoogteijling, 2020).

In finance applications, short-term interest rates are a linear function of several unobservable factors. In macroeconomics, they are the rates the central bank sets to stabilize the economy. Long-term interest rates are determined by the risk premium in finance. In contrast, in macroeconomics, they are determined by short-term interest rate expectations, and the Expectations Theory does not consider the risk premium mentioned in the finance literature. These differences illustrate the disconnection between the macroeconomics and finance literature (Diebold & Rudebusch, 2013). Besides, Diebold et al. (2006) show that the expectations hypothesis may not always be consistent with yield curves.

Contrary to Expectations Theory, Market Segmentation Theory assumes that the markets for bonds with different maturities are entirely separate and fragmented, are not substitutes for each other, and that price, supply-demand, and yield expectations do not affect each other (Mishkin, 2021). This theory, which argues that the formation of different yields in different maturities is due to different markets, is based on the positive slope of the curve in market uncertainty theory. According to the market uncertainty theory, maturity length and cash flow uncertainty are directly proportional (Akçay et al., 2012).

It proposes that long-term interest rates are the sum of the average of short-term interest rates suggested by expectations theory and the positive liquidity premium, which is the cost of choosing long-term bonds over short-term ones (Mishkin, 2021). If investors can obtain a premium that allows them to bear long-term risk, they may prefer long-term bonds to short-term ones (Kožíšek, 2018). Liquidity Premium Theory assumes bond markets are substitutes between terms, although not perfect substitutes (Mishkin, 2021). According to the liquidity premium theory, the risk premium of long-term bonds is high, as expected, and realized yields may differ. As the maturity gets longer, spreads are expected to decrease (Akçay et al., 2012).

According to the Preferred Habitat Theory, which is a version of the liquidity premium theory, if investors can earn higher yields, they prefer to buy bonds with maturities they do not prefer rather than bonds they prefer. Since investors prefer short-term bonds due to their low risk, they prefer long-term bonds only if they can earn higher yields. Therefore, the positively sloping liquidity premium increases as the time to maturity increases (Mishkin, 2021).

In the preferred habitat theory, the term premium can also take negative values instead of zero. The difference between the long- and short-term in this theory arises because these terms have different investors in different markets (Gibson et al., 2010). The no-arbitrage constraint in the theory explains shocks to the term structure and the central bank's ability to intervene in the yield curve (Kožíšek, 2018).

1.5. YIELD CURVE MODELS

Yield curve models are divided into categories, depending on the purpose or perspective. Diebold and Li (2006) divide yield curve models into two categories, Dolan (1999) into three, and Şişman (2011) into four categories. Diebold and Li (2006) divide yield curve models into no-arbitrage and equilibrium models. Accordingly, while no-arbitrage models aim at perfect matching in a way that does not allow arbitrage, equilibrium models aim to model the dynamics of the current rates. Equilibrium models use affine models after deriving yields at other maturities, subject to assumptions regarding the risk premium. Ultimately, these curves can be described as stochastic structures expressing a probabilistic description of the change in interest rates over time (Fabozzi, 2006).

Dolan (1999) divides yield curve models into stochastic, no-arbitrage, principal components, and fundamental models. Şişman (2011) divides yield curve models into factor models (affine), which are most frequently used because of their ease of calculation and flexibility; entire yield curve models, which aim to model the entire curve; market models that use market data as variables and can be used in option pricing; and other models such as consol models, price kernel models, positive rate models, non-linear models.

The majority of studies use the three categories as (i) affine-equilibrium, (ii) no-arbitrage, and (iii) statistical and parametric. Vasicek (1977) and Cox et al. (1985), Duffie and Kan (1996) create affine-equilibrium models. Hull and White (1990) and Heath et al. (1992) created no-arbitrage models. McCulloch (1971, 1975), Fisher et al. (1995), Vasicek and Fong (1982), Nelson and Siegel (1987), Svensson (1994), Bliss (1997), Rezende and Ferreira (2011) create statistical and parametric models.

Affine-equilibrium and no-arbitrage models are dynamic, while statistical and parametric models are static (Fabozzi, 2002). Spot interest rate forecasts are crucial parameters for portfolio management, whereas interest rate distribution is crucial for derivative pricing and risk management. No-arbitrage models prioritize spot fitting and do not contain much information about dynamics. On the contrary, equilibrium-affine models attempt to make predictions based on short-term dynamics. Unlike the in-sample fit, these models give poor results in out-of-sample forecasts (Diebold & Li, 2006). Static models are more suitable for out-of-sample forecasts. Models that attempt to construct yield curves with latent factors, such as NS, are the most efficient for forecasting (Vela, 2013). This study attempts to explain yield curve models according to affine-equilibrium, no-arbitrage and, statistical and parametric categories. The NS model is used in the study because it is suitable for out-of-sample forecast. We discuss the NS model, which is a parametric approach, and the Dynamic Nelson Siegel (DNS) model based on factor forecast under separate headings.

1.5.1. Affine -Equilibrium Models

Affine - Equilibrium models are used for derivative and current yields at different maturities, including a risk premium. These models first make assumptions for economic parameters. They then run a process (stochastic formulation) for short-term interest rates. Finally, they explain the interest derived from bond option prices (Hull, 2018). Affine-equilibrium models model instantaneous dynamics by considering the risk premium (He, 2013).

Affine-equilibrium models change short-term returns. The general equation is set up as

$$dr(t) = [\mu_0(t) + \mu_1(t)]dt + \sqrt[2]{\sigma_0^2(t) + \sigma_1^2(t)r(t)dW(t)} \tag{1.5}$$

Where r(t) defines instantaneous short rate at time t, μ is long term mean, dW(t) is the Wiener process and, σ is the volatility of the interest rates. These stochastic models consist of the sum of the drift and diffusion (Szenczi, 2016). These models assume that yield curves are linear functions of several factors (Kostyra & Rubaszek, 2020). These models have too many parameters to be optimized (Rodriguez, 2016).

In single-factor affine-equilibrium models, different shapes can be obtained by modeling all yields moving in the same direction but in different amounts (Hull, 2018). The yields in these models contain only one uncertainty term.

Forecasts are more efficient if Affine-equilibrium models include a no-arbitrage property and macroeconomic variables (Rezende & Ferreira, 2011). These models try to understand the economy based on economic indicators and short-term interest rates (Mineo et al., 2020).

High yields lead to a slowdown in the economy and lower demand for funds, which raises yields. Low yields lead to a demand for funds to fall, thereby increasing yields. This coercive system is a reversion to the mean in affine-equilibrium models (Hull, 2018).

1.5.2. No-arbitrage Models

No-arbitrage models have been developed because the estimated bond prices in affine-equilibrium models may differ from those in the market (Szenczi, 2016). No-arbitrage yield curve models find the points that best fit market data for government bonds. These models estimate the current yields by eliminating arbitrage opportunities at different maturities. These models are used primarily for derivative pricing. In the short-term rates in equilibrium models, the equilibrium model is converted into a no-arbitrage model by adding a function of time to the drift (Hull, 2018).

Affine-equilibrium models do not fit the term structure of current interest rates. This is because they found different term structures for different subjective parameters. The yield curve may change as the parameter chosen for optimization changes despite the same observed data. In equilibrium models, the term structures of today's interest rates are outputs. On the other hand, in no-arbitrage models, they are the inputs. In contrast to equilibrium models, drift is typically a function of time in no-arbitrage models. This ensures that yields in no-arbitrage models exactly fit the term structure (Hull, 2018). Unlike static models, these two dynamic model categories (affine-equilibrium and no-arbitrage) incorporate volatility into the model (He, 2013).

In macroeconomics, the interest-rate parameter refers to a single variable. It does not consider any phenomena such as risk and arbitrage. Conversely, the interest rate parameter is set for financial asset pricing, and macroeconomic effects are not considered. The no-arbitrage term structure model is an affine model that combines these two areas and incorporates the rational expectations theory of the New Keynesian Model. According to this model, the level is inflation, and the slope is monetary policy changes (Diebold and Rudebusch, 2013). Short-term rates of no-arbitrage models have high predictive power for GDP (Kožíšek, 2018).

1.5.3. Statistical and Parametric Models

Statistical and parametric models aim to obtain a curve that best fits observed yield data (Tüysüzoğlu, 2013). Statistical or parametric models are linear models, such as interpolation and bootstrapping, and nonlinear models, such as Nelson-Siegel and Svensson.

In cubic models, some bonds can distort the yield curve structure. Cubic interpolation may include extreme points. However, the Nelson-Siegel model can draw a curve that cancels these extreme points (Akıncı et al., 2006).

Nelson and Siegel (1987) assume that the τ values in the second and third terms (short- and medium-term) are equal because of the equation's overparameterization problem. Conversely, Bliss (1997) prefers a model with different values of this parameter because it provides more flexibility (De Pooter, 2007).

Models such as Svensson (1994) and Diament (1993) provide more valuable results by flattening the long-term portion of the yield curve (Akçay et al., 2012). In Svensson (1994), the second curvature improves the long-term fit, leading to more favorable results in terms of level and slope. This adjustment improves short-term fitting (Szenczi, 2016). In the Svensson (1994) model, decay parameters are assumed to have different values to avoid multicollinearity (De Pooter, 2007).

Björk and Christensen (1999) also propose a second 5-factor model. However, in this model, the yield curve does not flatten in the long term by approaching a certain level but instead becomes a model with a linear increase (De Pooter, 2007). Table 1 presents the yield curve models.

Table 1: Yield Curve Models

	Model	Formulas	Description
Equilibrium - Affine Models	Rendleman – Bartter	$dr = \mu r dt + \sigma r dw$	μ and σ are constants.
	Merton	$dr = a dt + \sigma dw,\ r_t = r_0 + at + \sigma W_t$	a is fixed.
	Dothan	$dr = \sigma r_t dw_t,\ r_t = r_0 e^{\sigma W_t - \frac{1}{2}\sigma^2 t}$	It is a model in which interest rates change only with the volatility of the market.
	GBM	$dr_t = \mu r_t dt + \sigma r_t dw_t$	μ is the coefficient of the mean.
	Mean - Reverting (Ornstein-Uhlenbeck) process or Vasicek	$dr_t = \theta(\mu - r_t)dt + \sigma dw_t$ $r_t = r_0 e^{\alpha t} + \alpha\mu t e^{\alpha t} + \int_0^t \sigma e^{\alpha(t-s)} dW_s$	θ gives the speed at which it will return to the mean. θ, μ, and σ are positive constants.
	Cox-Ingersoll-Ross (CIR)	$dr_t = \theta(\mu - r_t)dt + \sigma\sqrt{r_t} dw_t$ $r_t = \frac{\sigma^2}{4}\left(r_0^2 e^{-t} + \left(\int_0^t e^{(t-s)} dW_s \right)^2 + 2r_0 e^{-t/2} \int_0^t e^{(t-s)} dW_s \right)$	$\sqrt{r_t}$ term is added on the assumption that volatility will be high when interest rates are high and volatility will be low when interest rates are low. This model was developed to eliminate the possibility of finding a negative interest rate in the Vasicek model.
	Two-factor Markow	$dr=(u-\theta r)dt+\sigma_1 dw_1;\ du= -\mu u dt+\sigma_2 dw_2$	The level of reversion to the mean (u/θ) is not constant.
No-arbitrage Models	Ho-Lee	$dr = \theta(t)dt + \sigma dw$ $\theta(t) = F_t(0, t) + \sigma^2 t$	F_t are the instantaneous forward interest rates. θ(t) ensures the yield ratio matches the initial term structure model.
	Hull-White (One Factor)	$dr = [\theta(t) - ar]dt + \sigma dw$ $\theta(t) = F_t(0, t) + a[F(0, t) - r]$	It ensures full compliance with the initial maturity structure of the Vasicek model.
	Hull-White (Two Factor)	$df(r) = [\theta(t) + u - af(r)]dt + \sigma_1 dw_1$ $du = -budt + \sigma_2 dw_2$	u itself approaches zero. Negative yields can be found.
	Black-Derman-Toy	$d\ln r = [\theta(t) - a(t)\ln r]dt + \sigma(t)dw$	In practice, σ(t) is fixed and a(t) is zero.
	Black-Karansinski	$d\ln r = [\theta(t) - a(t)\ln r]dt + \sigma(t)dw$	There is no relationship between σ(t) and a(t).
	Interest rate tree	It contains stochastic processes. It is a discrete-time representation of a stochastic process for the short term. If the time transitions in the tree are Δt, the rate in the tree is the continuously compound Δt-period rate. The Δt-period rate (R) similarly follows the stochastic process of the instantaneous rate (r), using the continuous-time model.	It includes averages and reversion to the mean.

	Model	Formulas	Description
Statistical and Parametric Models	Interpolations	**Linear:** $R_c = R_a + (R_b - R_a) * \frac{(T_c - T_a)}{(T_b - T_a)}$ **Logarithmic:** $R_c = R_a + (R_b - R_a) * \frac{\ln(\frac{T_c}{T_a})}{\ln(\frac{T_b}{T_a})}$ **Cubic:** $R_c = R_a + (R_b - R_a) * \frac{(T_c^3 - T_a^3)}{(T_b^3 - T_a^3)}$ **Cubic Spline:** $R(T_i) = aR_i + bR_{i+1} + aR''_i + bR''_{i+1}$ $h = T_{i+1} - T_i$, $a = \frac{T_{i+1} - T}{h}$, $b = \frac{T - T_i}{h}$, $c = \frac{1}{6}(a^3 - a)h^2$, $d = \frac{1}{6}(b^3 - b)h^2$ **Quadratic:** $R_c = R_a + (R_b - R_a) * \frac{(T_c^2 - T_a^2)}{(T_b^2 - T_a^2)}$	It observed market returns are combined with various interpolation models to construct yield curves. R_a and R_b are the yield on two bonds traded in the market, T_a and T_b are maturities. For each unobserved T_c for each R_c interpolation models are applied to find the yield.
	Vasicek-Fong	Yields for all maturities from overnight to 30 years are calculated from the yields of the selected bond basket using the cubic spline method.	
	Bootstrapping	•Standard (Cluster) Bootstrapping: The nominal rates of standard maturities are found with the clustering technique. Cubic splines combine these standard maturities to form a nominal yield curve. Zero coupon rates are obtained with the bootstrapping method. •Iterative Bootstrapping: Zero coupon rates are calculated for each coupon period. Each time, zero-coupon rates are obtained until the bonds mature.	This method is based on arbitrage pricing, and prices reflect real prices. Despite the advantage of not requiring optimization, the resulting yield curve is a discrete function with poor estimating power for maturities outside the sample.
	Echols-Elliot	$\ln\big(1 + y(t, m_i)\big) =$ $a + b(m_i) + c\left(\frac{1}{m_i}\right) + dc_i + \varepsilon_i$	It can predict a negative interest rate. It is weak out of sample. Spot rate estimations may fail if there are no discounted bonds in the sample.
	Diament	$R_T = \frac{C_1(\frac{T}{C_3})^{C_4} + C_2}{(\frac{T}{C_3})^{C_4} + 1}$	If the curve has a hump, two more parameters are added to the formula.
	Mansi-Philip	$R_T = D_1 + D_2 e^{(D_4 T)} + D_3 e^{(2D_4 T)}$	It adapts to all curve shapes with this single formula.
	Nelson-Siegel	$R(m) = \beta_0 + \beta_1\left(\frac{1 - e^{-\lambda m}}{\lambda m}\right)$ $- \beta_2\left(\frac{1 - e^{-\lambda m}}{\lambda m} - e^{-\lambda m}\right)$	β_0, β_1, β_1 are level, slope, and curvature respectively, and λ is the maturity of the curvature at the point of rotation.
	Svensson (Extended Nelson - Siegel)	$R(m) = \beta_0 + \beta_1\left(\frac{1 - e^{-\lambda_1 m}}{\lambda_1 m}\right)$ $- \beta_2\left(\frac{1 - e^{-\lambda_1 m}}{\lambda_1 m} - e^{-\lambda_1 m}\right)$ $+ \beta_3\left(\frac{1 - e^{-\lambda_2 m}}{\lambda_2 m} - e^{-\lambda_2 m}\right)$	It adds a second curvature to the Nelson-Siegel model.

Source: Hull, 2018, pp.704-730; Akçay et al., 2012, pp.68-104.

1.6. NELSON-SIEGEL (NS)

Since Expectations Theory expresses current rates as a differential equation of forward rates, Nelson and Siegel (1987) derive the yield curve model from the solution of this equation. Accordingly, if the forward rate at maturity m is denoted by f(m), the root of the second-order differential equations is the solution.

$$\mathrm{f(m)} = \beta_0 + \beta_1 \exp(-m/\tau_1) + \beta_2 \exp(-m/\tau_2) \quad (1.6)$$

τ_1 and τ_2 are constants over time. β_0, β_1, and β_2 are coefficients determined by the initial conditions. Nelson and Siegel (1987) find that the model contained too many parameters to be solved. They assume $\tau_1=\tau_2=\tau$ in the model to be more parsimonious. The different values of τ do not affect the fit of βs. The authors also transform the model into a form with equal roots to obtain acceptable results. According to this model,

$$\mathrm{f(m)} = \beta_0 + \beta_1 \exp(-m/\tau) + \beta_2[(-m/\tau)\exp(-m/\tau)] \quad (1.7)$$

This equation produces monotonic, humped, and S-shaped forward rate curves. If we substitute Equation 1.7 into Equation 1.4, which converts forward rates to spot rates, we obtain a Nelson-Siegel yield curve model with equal roots.

$$\mathrm{R(m)} = \beta_0 + (\beta_1 + \beta_2)\,[1 - \exp(-m/\tau)]/(m/\tau) - \beta_2 \exp(-m/\tau) \quad (1.8)$$

This equation can produce monotonic, humped, and S-shaped curves just as the forward rate curve does. This model can be likened to the sum of a constant and a Laguerre function[1], which is the multiplication of a polynomial term, the convergence function, and an exponential decay term (Nelson & Siegel, 1987).

Nelson and Siegel (1987) create a set of values for τ to ensure the fitting of yield curves and find the values of β for each τ value using the linear least squares method. They choose whichever vector (τ, β_0, β_1, β_2) gave the best-fit value.

[1] Laguerre function: $xy'' + (1-x)y' + ny = 0$, n is not necessarily a nonnegative integer.

Nelson and Siegel (1987) interpret the coefficients of this model, which is a quadratic function, as β_0 long, β_1 short, and β_2 medium-term components. The long-term parameter is constant and has a non-zero limit, the medium-term parameter starts at an upward value and decreases towards zero, and the short-term terms start at zero, increase, and then decrease to zero. The NS model's interpretation of βs as maturity components is based on the l'Hôpital rule:

$$\lim_{m\to 0}\left(\frac{1-e^{-m/\tau}}{m/\tau}\right) = \lim_{m\to 0} e^{-m/\tau} = 1, \quad \lim_{m\to 0}\left(\frac{1-e^{-m/\tau}}{m/\tau} - e^{-m/\tau}\right) = 1-1=0 \tag{1.9}$$

The conclusion drawn from these two equations is that $R(0) = \beta_0 + \beta_1$. So, at time t=0, $\beta_0 + \beta_1$ is interpreted as a "short ratio."

$$\lim_{m\to \infty}\left(\frac{1-e^{-m/\tau}}{m/\tau}\right) = 0 = \lim_{m\to \infty}\left(\frac{1-e^{-m/\tau}}{m/\tau} - e^{-m/\tau}\right) \tag{1.10}$$

Based on these two equations, $R(\infty) = \beta_0$. So, at time t=∞, β_0 interpreted as a "long ratio." The difference between the long- and short-term

$$R(\infty)\text{-}R(0)=\beta_0 - (\beta_0 + \beta_1) = -\beta_1 \tag{1.11}$$

Moreover, it denotes yield spreads and is a proxy for curve steepness (Campolieti & Makarov, 2014).

Due to the computational difficulty of the interest rate term structure, it should be performed with the fewest parameters that provide the most explanatory power. Although the NS model is not no-arbitrage, it can model interest rates with high performance in in-sample adjustments and out-of-sample forecasts with three factors (Szenczi, 2016). Smoothing in the NS model is effective, avoids overfitting, and provides traceable and reliable results (Diebold & Rudebusch, 2013). NS is primarily effective in the discount bond market. However, they can adapt to the market conditions and produce more accurate forecasts with different yield curves. Despite these positive features, it is not easy to calculate the model parameters because of nonlinear optimizations or extreme values (Akçay et al., 2012).

This model graphs monetary policy expectations in the short term, business cycle fluctuation expectations in the medium term, and stable economic interest rates in the long term. β_0 gives the long-term interest rate expectation. According to the liquidity premium theory, this parameter is the sum of the interest rate and term premium. τ determines the point at which the curvature maximizes and the rate at which the slope and curvature approach zero as maturity lengthens. β_2 determines the magnitude and direction of the curvature.

NS is more prominent in analyzing and managing fiscal and monetary policies. However, NS is not an affine and no-arbitrage model (Aljinovic et al., 2012). Methods that best fit the yield data and model the entire yield curve, such as NS and Svensson, use more flexible constraints and can achieve better results than no-arbitrage models (Ishii, 2019).

1.7. DIEBOLD-LI (DYNAMIC NELSON-SIEGEL DNS)

Diebold and Li (2006) use the NS model, instead of the no-arbitrage or equilibrium approaches, to forecast the term structure of government bond yields. They add time as a third dimension to a two-dimensional graph of yield and maturity. Thus, they produce yield forecasts using NS model parameters. In this model, which they construct as a dynamic factor model, they reinterpret the parameters as the level, slope, and curvature. Their forecasts using autoregressive models give better results, especially for long horizons.

In the NS model, $(1 - e^{-m/\tau_t})/(m/\tau_t)$ and e^{-m/τ_t} similarly decreases monotonically. Hence, $\beta_1 + \beta_2$ and β_2 add similar terms into the model in the Equation 1.8. This not only complicates the intuitive interpretation of the factors of the NS model but also leads to multicollinearity; thus, forecasts cannot be accurately obtained (Diebold and Li, 2006). Diebold and Li (2006) replace $1/\tau$ with the term λ in the NS model and rearrange the yield curve model as

$$\hat{R}_{t+h/t}(m) = \hat{\beta}_{0,t+h/t} + \hat{\beta}_{1,t+h/t}\left(\frac{1-e^{-\lambda m}}{\lambda m}\right) + \hat{\beta}_{2,t+h/t}\left(\frac{1-e^{-\lambda m}}{\lambda m} - e^{-\lambda m}\right) \quad (1.12)$$

$$\hat{\beta}_{i,t+h/t} = \hat{c}_i + \hat{\Gamma}_i . \hat{\beta}_{it} \quad (1.13)$$

Like Nelson and Siegel (1987), Diebold and Li (2006) interpret β_0 as long-term, β_1 as short-term, and β_2 as medium-term, but they also call them dynamic latent factors as they add the time factor. They also associate these parameters with level, slope, and curvature in the curve, respectively. The "$(1-e^{-\lambda_t m})/\lambda_t m$" term decreases monotonically from 1 to 0 as maturity goes from zero to infinity. The "$(1-e^{-\lambda_t m})/\lambda_t m - e^{-\lambda_t m}$" term starts at 0, increases, and then decreases back to zero. The coefficient of β_0 is always 1. Since an increase in β_0 leads to the same increase in all maturities, they accept β_0 as the level. Some authors define the slope of the yield curve as $y_t(120)$-$y_t(3)$, whereas others define it as yt(∞)-yt(0), with a difference equals to β_1. Therefore, β_1 is associated with the slope of the yield curve. An increase in β_1 increases short-term yields more than long-term ones. β_2 is similar to $2y_t(24)$-$y_t(3)$-$y_t(120)$, so they associate it with curvature. An increase in β_2 affects short and long-term to a lesser extent and medium-terms to a greater extent (Diebold & Li, 2006).

Modeling a large dataset of observed factors as a function of a small dataset is called the factor model approach. DNS formulates yield datasets for various maturities with unobservable factors (Diebold et al., 2006). The dynamic factor models used in the yield curves contain several factors that summarize the price information for almost all the bonds. The factors of these models can be tracked statistically better than the yield data. These models prevent data mining and provide good out-of-sample forecasting results. Financial economic theory recommends the use of a factor structure.

Although many financial assets exist in the market, their expected returns are related to very few factors. For these reasons, dynamic factor models are preferred to summarize the price information of many nominal bonds at any given time (Diebold and Rudebusch, 2013). Early studies considered the level to be the only factor in yield curves. [e.g. Macaulay (1938) and Vasicek's no-arbitrage model (1977)]. The level dominates the movement of the yield curves. However, in reality, the yield curve contains more than one factor [Litterman & Scheinkman (1991) Willner, (1996) Bliss, (1997)]. Joslin et al. (2010) report that the first three principal components (3 factors: level, slope, and curvature) explain 95% of yields (Diebold and Rudebusch, 2013). Parsimonious factor models provide better results by combining factors in a linear model and forecasting each factor as a single time series, as in DNS and similar models. A shortcoming of such models is the lack of a theoretical foundation that incorporates the risk premium (Vela, 2013).

DNS models satisfy stylized facts mentioned in the Section 1.3. First, the yield curve produced by the DNS model is typically consistent with the averages of β_{0t}, β_{1t}, and β_{2t}. As a result, this model demonstrates that the yield curve is generally concave and tends to increase. Second, the DNS model allows for the yield curve to take on various shapes over time. Third, in DNS, yield dynamics are related to the dynamics of β_{0t}, spread dynamics are related to the dynamics of β_{1t}, and β_{0t} is more strongly persistent than β_{1t}. Therefore the property, which yield dynamics are persistent (resilient, less volatile), whereas the spread dynamics are less persistent, can also be satisfied. Fourth, in DNS, the short end is connected to β_0 and β_1, whereas the long end is connected to β_1. Because β_1 alone is expected to be less volatile, DNS is possible to satisfy the property which the short-term end of the yield curves is more volatile than the long-term end. Fifth, in DNS, long-term rates depend on β_0, and short-term rates depend on β_0 and β_1. As mentioned in fact three, β_0 is the most persistent factor that may satisfy the property which long-term rates are more persistent than short-term ones.

As seen in these properties, a good dynamic yield curve model is expected to fulfill its shape characteristics, ability to take different shapes at different times, strong persistence of yields (less volatile and stable), and weak persistence of spreads. It is possible that DNS can satisfy all these properties (Diebold & Li, 2006).

1.7.1. β Factors (β_0, β_1, β_2)

The level is a highly persistent NS factor. The correlation between the level and inflation shows the consistency of the relationship between the yield curve level and inflationary expectations, referred to as the Fisher equation ($r=i-\pi$) in macroeconomics. The slope and curvature factors are not as persistent as the level. The slope has higher persistence and lower shock variance among these two factors, with equal unconditional variances. Slope and curvature are also related to the business cycle period. The correlation between slope and capacity utilization, an indicator of macroeconomic activity, suggests that slope is also linked to economic dynamics. However, no link was found between the curvature factor and economic indicators. Moreover, the strong correlations between $y_t(3)$-$y_t(120)$ and β_1 and between $2y_t(24)$-$y_t(3)$-$y_t(120)$ and β_2 suggest that the definition of NS parameters as slope and curvature is correct (Diebold et al., 2006).

The autocorrelation between lagged values is expected to be higher for the level factor than for other factors. The slope and curvature have more changeable characteristics (Szenczi, 2016). However, Duffee (2012) states in his literature survey that the level is problematic to forecast statistically, even though it can be forecasted in practice. The slope is forecastable, whereas the curvature is unrelated to the level and slope. He also notes that the level is unaffected by future excess yields, whereas a flatter curve and more curvature imply fewer excess yields. This level is unforecastable because it is unknown how much of the risk premium, which can move inversely to the change in yields, will reduce the movement in yields. For example, a recession in the economy increases the risk premium while reducing short-term interest rates. The term structure emerges according to the severity of the decrease and increase. However, the excess yield decreases. The slope has a positive relationship with the future excess yields.

It is unreliable in forecasting medium-term interest rate movements in between short and long term (Fama, 1984; Fama-Bliss, 1987; Campbell-Shiller; 1987, 1991). But, Campbell-Shiller-Schoenholts (1983), and Mankiw-Summers (1984) find contrary findings (Mishkin, 2021).

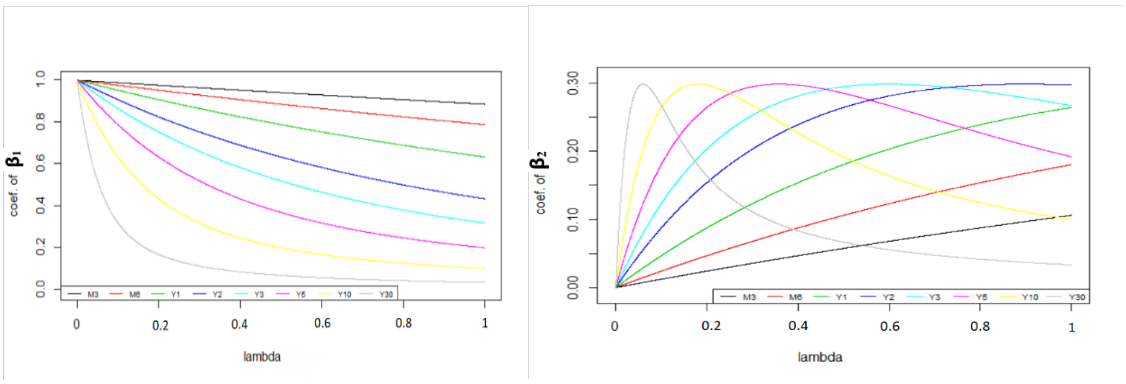
1.7.2. Lambda Factor (λ)

In the DNS model (Equation 1.12), it is expressed λ as $\lambda = 1/\tau$; for small λ, the decline is slow, and the curve fits better in the long term, whereas for large λ, the decline is fast, and the curve fits better in the short-term. According to Diebold and Li (2006), λ determines where the β_{2t} reaches its maximum. Nelson and Siegel (1987) impose the constraints $\beta_0>0$, $\beta_1+\beta_2>0$, $\tau,t>0$. In this model, β_0 is not affected by β_1 or β_2. β_1 and β_2 are affected by each other through λ. λ is the trade-off between the long and short terms. The optimal choice of λ is based on balancing long and short maturities (Szenczi, 2016).

No definitive study or theory determines the economic significance or relevance of λ. This deterioration parameter affects the yield curve fitting and significantly impacts the long-term fitting (Marek, 2015).[2]

Figure 2: Effect of λ on Slope (a) and Curvature (b).

(a) (b)



In Figure 2.a, the value of the term $(1 - e^{-\lambda m})/\lambda m$ at various maturities is observed as λ increases from 0 to 1. As λ approaches 1, the importance of β_1 and, hence, share in the errors of yields forecasts decreases, especially in longer terms. In other words, if λ is forecasted to be 1, the forecast improvement of β_1 will not be as effective for yields forecasts as at smaller λs.

In Figure 2.b, the value of the term $(1 - e^{-\lambda m})/\lambda m - e^{-\lambda m}$ at various maturities is observed as λ increases from 0 to 1. As λ approaches 1, the importance of β_2 and, hence, share in the error of yields forecasts first increases then decreases, especially for short maturities. In other words, if λ is forecasted to be 0, the forecast improvement in β_2 will not be as effective for yields forecasts as it is at larger λs. This effect increases faster as maturity increases and decreases after reaching different points for different maturities.

[2] Marek used NS: $R(m) = \beta_0 + \beta_1 \exp(-m\lambda) + \beta_2[(m\lambda)\exp(-m\lambda)$.

1.8. YIELD CURVE MODELS AND FORECASTING YIELD CURVE

Countries develop models to manage debt instruments and assess the economic impact on interest rates. Investors and researchers develop yield curve models suitable for risk management and investment decisions and publish their results as data or articles. The literature contains many academic studies and comments on digital platforms regarding yield curves in many countries.

Depending on the purpose of the study, some researchers aim for fitting, some for forecasting, and some for a balance between the two, and choose the model approach according to this goal. In the literature, conventional and non-conventional models (mostly ANN models) of forecasting yield curve, related especially with the DNS model of Diebold and Li (2006), which is the basis of this thesis, are discussed in two separate groups. Finally, studies of ensemble learning and forecast combinations are briefly discussed.

Nelson and Siegel (1987), following Milton Friedman's (1977) recommendation to produce a basic statistical model that can describe the entire term structure with a few parameters, develop a model that is simple, short, that is, sufficiently parsimonious meant to capture all yield curve shapes in a single model, in other words, to represent monotonic, humped and S-shaped yield curves in a single formula (Equation 1.8). Using this model, Nelson and Siegel obtain yield curves consistent with the United States bond yields for the 1981-1982 period with a correlation of 0.96. Moreover, the latent factors in the model are immediately adapted to monetary policy changes in 1982.

Diebold and Li (2006) propose the Dynamic Nelson Siegel (DNS) model to forecast monthly US yield curves (Equation 1.12 and 1.13). They propose an AR(1) model to forecast the factors in the second Equation 1.13. From the forecasted βs, they again obtain the yield curve forecasts. They compare them with VAR (1) model forecasts, using all βs as inputs. They forecast 10 different models for dynamic models, emphasizing AR and VAR methods. Although RW prevail in the short horizon, DNS using the AR(1) model yields better results, especially in the middle horizon and above. They claim that the reason forecasts in the short horizon fail to beat RW is mispricing due to insufficient liquid bonds. They also state that, although the model is not no-arbitrage, the smoothness of the forecasted yield curves may be an advantage over other models, as it does not lead to excessive positive positions. However, while extending the model by adding terms to the NS model, is suitable for in-sample fit, it can be said that extended models do not guarantee better performance since more conservative models are generally more successful in out-of-sample forecasting. Considering this, they, in line with NS, state that they chose to base their models on simplicity and parsimony, being based on a theory, not needing data mining, and being good at out-of-sample forecasting. This study utilizes the Diebold-Li (2006) model.

Diebold and Li (2006) used US Treasury bond yields between January 1985 and December 2000 as data and obtained good results from the DNS model. Although DNS does not exactly fit the data, it has become a preferred model for fitting and forecasting because its structure protects it from overfitting (Diebold & Rudebusch, 2013). Diebold and Rudebusch (2013) claim that most yield curve models have a good theory but give poor empirical results or vice versa. Despite successful empirical results, these models are not built on a robust theory. Therefore, they prefer the DNS model instead of conducting comprehensive research.

Diebold and Li's (2006) model has been widely studied, and many articles have been produced in parallel. However, some also argue that the study used the wrong approach and that the model produces inappropriate results. The success of DNS in out-of-sample forecasts has always been debated. Guidolin and Thornton (2008) show that DNS is no better than RW, even at long horizons.

Using US data, Rostan et al. (2017) show, that different optimal delay values can be selected then the p = 1 value in the DNS-AR(p) model with the Burg (1975)[3] approach, which is a signal processing model. The Burg approach is used to compare the shape of the yield curve (normal, humped, flat, or inverted) using various criteria.

[3] In the Burg approach, AR(p) aims to find the best-fitting model that minimizes the error of forward and backward forecasts in the time series in a way that satisfies the Levinson-Durbin recursion condition.

De Poorter (2007) analyzes US bond yields using the NS model, two-factor model[4], Björk and Christensen's (1999) four-factor model[5], Bliss' (1997) three-factor model[6], Svensson's (1994) four-factor model, and adjusted Svensson's (1994) four-factor model (β_3 is different[7]) with in-sample fit and out-of-sample forecasts with RW, AR(1), VAR(1) methods. De Poorter concludes that the higher the number of factors, the better the fit and forecasts. He finds no significant difference between taking λ as fixed for fit and estimating it. The iterative approach outperforms the direct forecast approach for the long horizons. The four-factor model and adjusted Svensson obtain the best forecast overall and in the subsamples. The four-factor VAR model with estimated λ provides a better forecast in the long run.

Using the same data as Diebold and Li (2006), Hays et al. (2012) obtain better forecasts than DNS-AR(1) with their proposed FDFM (functional dynamic factor model: a combination of DFM and FDA (functional data analysis)) model. Reschenhofer and Stark (2019) use DNS-AR(1) (factor and factor spreads) and AR(1) (yield and yield spreads) models for US yields and find that AR(1) (yield spreads) outperform RW.

Povala and Vasıl (2017) forecast German yields using the DNS-AR(1) model, and it could not perform better than RW in low interest periods. They state that changing λ or using time-varying λ does not affect forecasts.

Dauwe and Moura (2011) show that for euro swap yield curve, principal component analysis (PCA) gives good results in the short term and regression model and PCA in the long term in a group of models in which PCA models (AR, regression, regression with yield differences), DNS-AR(1) yield regression, yield difference regression, and RW models. Because the yields are not stationary, they recommend using PCA as the second model and the best forecast.

---

[4] $\beta_{0t} + \beta_{1t}\,[1 - \exp(-\lambda_t m)]/(\lambda_t m)$

[5] in addition to NS, $\beta_{3t}\,[1 - \exp(-2\lambda_t m)]/(2\lambda_t m)$

[6] β_1 and β_2 with different λ's

[7] In both Svensson and NS, β_3 is same but λ_t is different. Adjusted Svensson $\left[\exp\left(-\lambda_{2,t} m\right) + \left\{\left(2\lambda_{2,t} m\right) - 1\right\} \exp\left(-2\lambda_{2,t} m\right)\right]$

Frederick and Herzog (2012) include economic indicators in US yields, that the λ of the NS model and they find that the Kalman filter captures the yield curve dynamics well and that better forecasts could be made with respect to those of RW. Koopman et al. (2010) use a simple step function and a spline function for λ. This improves both the fit and forecast. He (2013) forecasts the yields of 10-year and 30-year US Treasury bonds using NS (λ estimation with Kalman filter), DNS-AR(1), RW, slope regression, AR(1) for yields, and VAR(1) for yields. The Kalman filter used to model the yield curve produces good results in fitting and forecasting in the dynamic model, particularly in the long term. He shows that λ can also be forecasted and is an important factor in the NS model. His study obtains better results with a non-constant λ and forecasts than with a fixed λ. He states that when λ is fixed in the DNS, βs and λ may fit at inappropriate values, and their interpretation and forecasts may not be meaningful. He concludes that estimating λ is more appropriate for interpreting economic forecasts and dynamic models.

Marek (2015), in his study on the estimation of yield curves of US, Eurozone, and UK government bonds with DNS-AR(1), obtains λ from Quasi-Newton class with BFGS (Broyden-Fletcher-Goldfarb-Shanno) optimization[8] and find β's in the second step. His proposed method provides more realistic results in in-sample fit, especially in the long term, and in out-of-sample forecasts, especially for data with high volatility.

Sambasivan and Das (2017) fit and forecast the yield curve with Dynamic Gaussian Process (GP)[9] better than DNS-AR models for medium and long-term US yields (2 years—30 years). Joslin et al. (2011) find that dynamic Gaussian models forecast US yields better than the VAR model. Reinicke (2019) uses AR(p), PCA, and Dynamic GP models to forecast US Treasury bond and German Bundesbank (Bundesbank) yields. Dynamic GP forecasts yield the best results. The AR(4) model forecasts better than the

---

[8] Gill and Leonard (2001)

[9] Yield curves can be modeled as functional forms consisting of factors of the term structure. In these models, which are formulated as $R = \phi\beta + \epsilon$, is expressed as:

Nelson-Siegel Base: $\phi = \left\{1, \frac{1-e^{-\lambda m}}{\lambda m}, \frac{1-e^{-\lambda m}}{\lambda m} - e^{-\lambda m}\right\}$

Exponential Base: $\phi = \left\{1, e^{\lambda_1 m}, e^{\lambda_2 m}, \dots\right\}$

Gaussian basis: $\phi = \left\{1, e^{-\lambda(m_1-c)^2}, e^{-\lambda(m_2-c)^2}, \dots\right\}$.

AR(1) model does. Reinicke takes the average of the calculated λ values as constant λ.[10] He uses a multistep (iterative) approach to examine the input data length.

Castellani and Santos (2006) find similar results in forecasting the 10-year yields of US Treasury bonds using conventional models (ARIMA, ECM) and AI models (FL with AI approach and manually generated FL, SOM, and MLP). They suggest that instead of single models, forecast combination or hybrid models combining ML and statistical models should be tried.

Kožíšek (2018) makes high-frequency US bond yield forecasts via NS parameters. LSTM- RV[11] give better results than AR(1), VAR(1), and VAR-RV models in daily and monthly forecast horizons in level and slope. Kožíšek concludes that using small samples (such as annual data) in LSTM reduces the forecasting power.

Hoogteijling (2020) forecasts US bond yields using linear models (PCR) and ML models (RF, NN).[12] His proposed a method prevailed in all the models. Putting constraints on the models initially increases predictive power. He obtains results contrary to the spanning hypothesis, which states that all information about future bond market movements is contained in the yield curve.

Zimmermann et al. (2002) forecast German yield curves better with their proposed ECNN (Error correction neural networks) model than the RNN and MLP models. Jacovides (2008) finds that MLP and SVM perform better than RW in forecasting daily changes in the levels and spreads[13] of UK interest rates and 6-month forwards. In the long run, SVM is more efficient than MLP in forecasting the direction of interest rates. Tappinen (1998) forecasts that the spreads between various maturities in US spot rates are better with MLP than with the OLS model. Gerhart et al. (2018) study the forecasts of EURIBOR interest

[10] This method was proposed by Arbia and Di Marcantonio (2015). Also Molenaars et al. (2015) found that the choice of λ does not affect the efficiency of prediction (Reinicke, 2019).
[11] RV: Realized Varyans. It adds the daily deviation $\left(RV_{\beta}^{D} = \sum_{t \in D} (\beta_t - \beta_{t-1})^2\right)$ to the model (Kožíšek, 2018).
[12] He calculated excess yields in models with $\left(-\left(\frac{m}{12} - 1\right)\left(R_{t+12}^{(m-12)} - R_t^{(m)}\right) + \left(R_t^{(m)} - R_t^{(12)}\right)\right)$; $m \geq 12,\ R: yields$ formula. Therefore, he estimated risk premiums.
[13] $\Delta R_{t+6}^{m} = f\left(S_t^{(n,m)}, R_t^{m}\right); S_t^{(n,m)} = (R_t^{n} - R_t^{m}) \quad n \in (6m, 1y, 3y, 5y, 10y, 20y)$

rates with RW, LSTM, and direction (increase or decrease) forecasts using SVM, RW, Linear Discriminant Analysis (LDA), and Quadratic Discriminant Analysis (QDA). They also use the bid-ask spreads. They use an SVM to forecast the upward or downward trends of level and slope. The ANN models provide good results, especially for long forecast horizons. In addition, adding spreads as data improves forecasting efficiency. The SVM yields the best results in terms of classification. Eklind (2020) shows that deep learning models (LSTM and TCN) outperform RW in forecasting the bond yields of the Bank of England (BOE) and the Bank of Sweden (Svenska Handelsbanken) over medium and long horizons. Wood and Dasgupta (1995) obtain an effective result in forecasting the UK interbank spot interest rate with MLP.

Rosadi et al. (2011) obtain Indonesia's yields by forecasting NS parameters (including λ). Regarding the forecast interval, MLP is good in the short term, whereas VAR was slightly better in the long term. Vela (2013) cannot demonstrate the superiority of the ANN approach in forecasting the NS and NSS factors of the yield curves of four Latin American countries (Colombia, Mexico, Peru, Chile) and the US using AR, VAR, and FFNN.

Kostyra and Rubaszek (2020) forecast Polish interbank market swap yields and the RF model does not statistically outperform the conventional models. No model could forecast a decline in yield levels. The Expectation Hypothesis model, based on the assumption that long-term returns are a predictor of short-term returns, yields better results, especially in the short term.

Researchers examine forecast combinations (FC), both theoretically and practically. Claeskens et al. (2014) examine what the weights of the forecast combination model should be, while Pesaran and Pick (2011) investigate the combination of forecasts from different data windows with the same data and model. Andrawis et al. (2011) study the combination of short and long-run forecasts. Genre et al. (2013) investigate whether other combination methods can make better forecasts than the simple average combination method for some economic data. Hyndman et al. (2007) experiment with hierarchy modeling in which they first made forecasts at each level and then combine them. Jore et al. (2008) try to forecast economic data (growth, inflation, short-term interest rates) using density combinations. Hsiao and Wan (2011) investigate geometric and average combinations for the optimal forecast combination. They compare forecast and information combinations. Steel (2017) examines model-averaging methods to reduce model uncertainty and studied Bayesian and frequency-averaging methods. The averaging method has been studied regarding growth, production, and finance. Adhikari and Agrawal (2013) examine forecast combinations to obtain better forecasts in time series and show theoretically and practically that choosing the arithmetic mean or median in combinations yields good results.

Rezende and Ferreira (2011) use the five-factor model (they added a second slope to the Svensson model with the same λ as the second bend), the NS, Bliss (1997), Svensson models for Brazilian yield curves. They forecast the parameters of the models with AR(1), VAR(1), and RW and their medians. The RW model provides the best forecasts for the short horizon, and the combination model (median) provides the best forecasts for the medium and long horizons. The forecasts worsen with an increase in the number of parameters. Caldeira et al. (2013) forecast Brazilian future yields using RW, AR(1), VAR(1), Bayesian VAR, DNS (with Kalman filter), and DNSS (with Kalman filter) models and their combinations (equally weighted, OLS, rank inverse weighted, RMSE inverse weighted, MSE inverse weighted). This combination yields better results with respect to those from the single models.

Araújo and Cajueiro (2014) show that forecast combinations for Brazilian yield curves of interest swaps are more consistent than forecasts using single models. They make forecasts using the AR, VAR, DNS, FSN-ECM (Functional Signal Plus Noise with an Equilibrium Correction Model) individual models, the simple mean, trimmed mean, WLS (Weighted Least Squares), inverted MSFE, and the worst individual model excluded, and the iteration of these methods. They make a comparison with the RW. For the individual models, AR gives the best and RW the worst results at the one-month horizon, whereas RW gives the best forecasts at other horizons. The models fail during the crisis period. Simple combinations yield better results for short horizons whereas WLS provides better results for long horizons. Araújo and Cajueiro remove the worst individual model from combination-improved forecasts and observe that the inverted MFSE model performs well for structural breaks.

CHAPTER 2:
CONSIDERED FORECASTING METHODOLOGIES

Forecasting involves making predictions about the future with high accuracy using current and historical information, patterns, and knowledge (Hyndman et al., 2021). The models used for forecasting can be classified into conventional approaches based on traditional data analysis and statistical methods, and non-conventional approaches beyond these.

Relying solely on a single model can be problematic. Models that work well in theory may not perform well when applied to actual data. Therefore, instead of sticking to a single model, exploring and working with various models to find suitable alternatives serves as a better approach (Lewis, 2017b). As such, this study utilizes multiple models with different theories and structures.

This chapter explains and compares conventional and non-conventional models. In addition, theoretical information about ensemble learning and forecast combination methods are also provided. ARIMA, ETS, VAR, RW, threshold models among others for conventional approaches, and kNN, some ANN models among others for non-conventional approaches are used for forecasting.

2.1. CONVENTIONAL FORECASTING MODELS

Conventional models usually make forecasts based on historical data using a specific structure or form. They are usually simple, understandable, and easy to implement. They expect data to follow a normal distribution and certain statistical assumptions.

Conventional models are used in almost every study to analyze and forecast financial data because they use time series. These models can form the basis of the study or be used as a criterion for evaluating other models. Regression, smoothing, autocorrelation, vector models, threshold models, and RW are the most used conventional models for time series forecasting. This section discusses the following models used in this study. They are ETS, TBATS, ARIMA models, ARFIMA, VAR, VECM, TAR, LLAR, LSTAR, SETAR and RW.

2.1.1. Exponential smoothing (ETS - TBATS) Models

Time series can be decomposed into their principal components which are trend, seasonality, and randomness. Trend is a continuous increase or decrease in the levels of a time series, seasonality is the fluctuations with values close to each other in the same seasons, and randomness is a part of the time series that represents the residual variations left after accounting for the trend and seasonality components of the series. In smoothing approaches, the error term can be used instead of randomness.

Exponential smoothing methods update the components of time series over time. It models the exponential decay of past observations as the time difference increases. The structure of these methods, which consist of simple exponential, trend, and Holt-Winters, and their variations, includes the components of the series and their smoothing formulas. These methods form the basis of models such as ETS and TBATS.

Automatic Exponential Smoothing Method (ETS) combines 18 different exponential smoothing methods, including error (E), trend (T), and seasonality (S) components. The error component can be multiplicative or additive; the trend component can be absent or additive or multiplicative or additive-damped; and the seasonality component can be absent or additive or multiplicative.

The TBATS model is a double seasonal exponential smoothing approach consisting of trigonometric seasonal, Box-Cox transformation, ARIMA error, trend, and seasonal components. Classical models with a seasonal structure are ineffective in forecasting time series where seasonality is not of a simple standard but more complex. This model includes time-varying Fourier series. It is an advantageous model for time series with multiple, high-frequency, non-integer seasonalities and other effects (Grmanová et al., 2016). For ETS and TBATS models, see Appendix 1.

2.1.2. Box-Jenkins (ARIMA) Models

A time series is considered autocorrelated if it exhibits a relationship with its past or future values at time t. Even the series has high autocorrelation, this current and previous data relationship does not necessarily imply causality. Box-Jenkins (1976) developed stochastic time series models, a combination of purely stochastic and autocorrelation parameters, for forecasting and analyzing time series. In the related literature, the method that best applies the ARMA model is the Box-Jenkins (1976) method. ARIMA(p,d,q):

$$x_t = \emptyset_0 + \emptyset_1 \nabla^d x_{t-1} + \emptyset_2 \nabla^d x_{t-2} + \ldots + \emptyset_p \nabla^d x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} + u_t \quad (2.1)$$

x: time series, $\emptyset$: autoregression coefficients, θ: moving average coefficients, ε: white noise process, p: number of lagged variables, q: number of lagged errors, u: residual, $\nabla^d x$: d^{th} order difference of x.

In this modeling, the series should be stationary. Certain conditions may reveal long memory in nonlinear time series. In series with long-term dependence, the d value in the model is in the range of [-0.5, 0.5]. In this model called Autoregressive Fractional Integrated Moving Averages (ARFIMA), the autoregressive function graph decreases very slowly. However, spurious long memory can be observed in stationary series. Partial non-stationarity is one reason for this problem in threshold models, while the entire series is stationary (Kuswanto & Sibbertsen, 2008).

2.1.3. Vector Autoregression (VAR) Model

VAR models all variables and their lagged values together. The model is constructed with the assumption that the variables affect each other. Impulse-response functions check this assumption. VAR(p) model:

$$x_t = \emptyset_0 + \emptyset_1 x_{t-1} + \cdots + \emptyset_p x_{t-p} + u_t \quad (2.2)$$

$$x_t = [x_t^1 \; x_t^2 \; \ldots \; x_t^m]' \quad (2.3)$$

$$u_t = [u_t^1 \; u_t^2 \; \ldots \; u_t^m]' \quad (2.4)$$

p: model degree, m: number of variables, $\emptyset$: coefficient matrix. In the model, the conditions of x_t^i $(i = 1 \dots m)$ being stationary and u_t^i $(i = 1 \dots m)$ being white noise and their correlation being zero are met, the VAR model is valid.

The interaction between variables in a VAR model can improve the accuracy of economic forecasts. This model also enables the examination of economic innovations and shocks (Zivot and Wang, 2007). The VAR model is an approach that analyzes the factor structure and the returns of financial assets that are suitable for this model. Since the NS model includes a factor structure, the VAR model is one of the most widely used models in yield curve forecasting studies, especially in the DNS approach (Diebold & Rudebasch, 2013). According to De Pooter (2007), VAR captures factor dynamics better than univariate AR(1) models. Building a dynamic model with VAR gives good results for short-term forecasts. Conversely, AR can provide good results in short- and long-term forecasts. Because the VAR model's forecasts may be poor despite the excellent fit. AR outperforms VAR in the DNS model because the correlation of NS factors that need to be estimated in VAR is insufficient (Vela, 2013).

2.1.4. Threshold (TAR-SETAR-LSTAR) and Locally Linear (LLAR) Models

Several threshold models have been suggested for piecewise autoregressive (AR) structures in time series. Additionally, locally AR models can be used for series that cannot provide linearity throughout the series.

The TAR (Threshold AR) (Tong, 1978) approach uses different AR models for series with different linear dynamics in different regimes. An m regime, p. degree TAR is defined as follows

$$x_t = \begin{cases} \emptyset_{0,1} + \sum_{i=1}^{p} \emptyset_{i,1} x_{t-i} + \sigma_1 \varepsilon_t, & x_{t-d} \leq r_1 \\ \emptyset_{0,2} + \sum_{i=1}^{p} \emptyset_{i,2} x_{t-i} + \sigma_2 \varepsilon_t, & r_1 < x_{t-d} \leq r_2 \\ \dots & \\ \dots & \\ \dots & \\ \emptyset_{0,m} + \sum_{i=1}^{p} \emptyset_{i,m} x_{t-i} + \sigma_m \varepsilon_t, & r_{m-1} < x_{t-d} \end{cases} \tag{2.7}$$

where m is determined previously; r is the threshold value, d is the delay parameter, and ε is random variable with 0 mean and unit variance ($-\infty < r_1 < \ldots < r_m < \infty$ and $d > 0$). Parameters are estimated by nonlinear least squares.

SETAR (Self-Exciting TAR) is a model that expresses a special case of the TAR approach. The series' past values determine the TAR model's threshold value. SETAR model:

$$x_t = F_1(x_{t-1}, \emptyset)(1 - I(x_{t-1} > r) + F_2(x_{t-1}, \emptyset)(I(x_{t-1} > r)) + \epsilon_t \quad (2.8)$$

F: autoregressive processes, I: indicator function, r: threshold value.

The LSTAR (Logistic Smooth Transition AR) model is a generalization of the SETAR model. This model includes different regimes and consists of the transition function between regimes in its structure (Skalin & Teräsvirta, 1999).

$$x_t = \left(\emptyset_1 + \emptyset_{10}x_t + \emptyset_{11}x_{t-d} + \cdots + +\emptyset_{1L}x_{t-(L-1)d}\right)\left(1 - G(r_t, \gamma, th)\right) +$$

$$\left(\emptyset_2 + \emptyset_{20}x_t + \emptyset_{21}x_{t-d} + \cdots + \emptyset_{2H}x_{t-(H-1)d}\right)G(r_t, \gamma, th) + \varepsilon_{t+s} \quad (2.9)$$

G: logistic function, r_t: threshold variable, L: low regime lag parameter, H: high regime lag parameter, G: transition function, γ: speed of the transition function.

LLAR (Locally Linear AR) is the AR model for locally linear time series. It is forecasted with d time-lagged time series models with m locally linear autoregressive series:

$$x_{t+s} = \emptyset_0 + \emptyset_1 x_t + \cdots + \emptyset_m x_{t-(m-1)d} \quad (2.10)$$

In the model, x_t is estimated for an interval of values ϵ at point m (Antonio; 2008).

2.1.5. Random Walk (RW)

RW refers to series that do not have a specific systematic structure and are stationary concerning the mean. A purely random series has a constant mean. RW is a stochastic model that takes the current data to forecast the data as the following

$$\hat{x}_{n+k}=x_n+\varepsilon_{n+k} \quad (2.11)$$

where ε_{n+k} is white noise, and ε_{n+k}~N(0,σ^2). According to Pearson (1906), forecasting models should be based on the idealized concept of RW. RW is based on the premise that data has an equal probability of going in all directions (Grinstead & Snell, 1997). Users of forecasting models want to produce models that make more efficient forecasts than RW forecasts (Szenczi, 2016).

2.2. NON-CONVENTIONAL FORECASTING MODELS

They generally require more data and more complex structure. They work with fewer assumptions and more flexibility. Non-conventional approach involves many ANN models that are suitable for forecasting. This section discusses only the models used in this study. They are kNN and ANN models which are MLP, NNAR, GRNN, ELM, GMDH, RNN, ENN, JNN, LSTM, and GRU.

2.2.1. k-Nearest Neighbour (kNN)

In the k-nearest neighbors (kNN), forecast is made with mean value of k nearest data points with numerical values. This algorithm is efficient, simple, and easy to understand, and it is nonparametric, namely, it makes no assumptions about the probability distribution of the data. There are no strict rules for choosing the algorithm's distance criterion and parameter k; trial and error is the best approach (Lewis, 2017b).

In kNN, the distances of the new input to the samples in the training data are calculated first. Note that, in this study, Euclidean distance approaches are used to find *d*. which is defined as follows:

$$d = \sqrt{\sum_{i=1}^{n}(x_z - x_i)^2} \quad (2.12)$$

where x_i is input data, y_i is output data, x_z is new input to predict.[14]

[14] For other distance measures see: Lewis, N. D. (2017b). *Machine Learning Made Easy With R*.

Next, the k smallest (closest) distances are selected in this study. Based on the calculated distances (d), the k neighbors (i.e. x and y values) with the smallest distance (nearest) are selected. The average of the target values (y) of the selected k neighbors is calculated.

$$\hat{y} = \frac{1}{k}\sum_{l=1}^{k} y_l \tag{2.13}$$

This mean is the regression value of the new point to be forecasted. Instead of the mean, the median or weighted mean can also be used in calculating the forecasted value. The value of k greatly affects the performance of the model. Small values of k (e.g. k=1) can introduce more noise into the model, while large values of k can help the model to be more general. Different distance measures can affect the performance of the model.

2.2.2. Artificial Neural Networks (ANN)

Artificial neural networks (ANN) are an important sub-branch of machine learning (ML) and a frequently used method in this field. ML, the building blocks of AI, forecasts the future by finding patterns from past data. Although many ML models do not give good results, they are preferable because they reach results faster and are less costly than employing human beings (Hoogteijling, 2020). On the other hand, well-implemented ML models are more successful than other approaches in revealing nonlinear relationships in financial data and simulating these data (Mitchell, 2006).

ANNs are mathematical algorithm-based computer systems that automatically develop the ability to generate, discover, and create new knowledge without assistance using learning methods like the human brain. Instead of ANNs, the terms "connectionist networks," "parallel distributed networks," and "neuromorphic systems" are also used. (Öztemel, 2016).

These algorithms can learn from examples and apply what they have learned. They are expressed graphically in network notation. Although the terminology differs, ANNs can be likened or adapted to statistical models.[15]

Cybenko (1989) introduces the Universal Approximation Theorem for ANNs and shows that a single hidden layer neural network with a sigmoid activation function can converge any continuous function. Hornik (1991) also proves this true for other arbitrary bounded activation functions (Hansson, 2017). According to the Universal Approximation Theorem, if an FFNN network has a single hidden layer and a random number of neurons, the network converges to any continuous function. The theorem is formulated as follows: Given any function f, $f \in C(I_D)$ and $\epsilon > 0$

$$F(x_1, \dots, x_D) = \sum_{j=1}^{M} \alpha_j \varphi \left(\sum_{i=1}^{D} w_{ij} x_i + b_j \right) \tag{2.14}$$

The sets of real constants α_j, b_j, w_{ij} define the integer M. $\varphi(\cdot)$: bounded, monotonically increasing continuous function. I_D: D-dimensional $[0,1]^D$ hypercube. $C(I_D)$: The space of the continuous function on I_D. j=1,...,M and i=1,...,D. Approximate realization of $f(\cdot)$ for all $x_1, x_2, \dots, x_D$ in the input space is

$$|F(x_1, \dots, x_D) - f(x_1, \dots, x_D)| < \epsilon \tag{2.15}$$

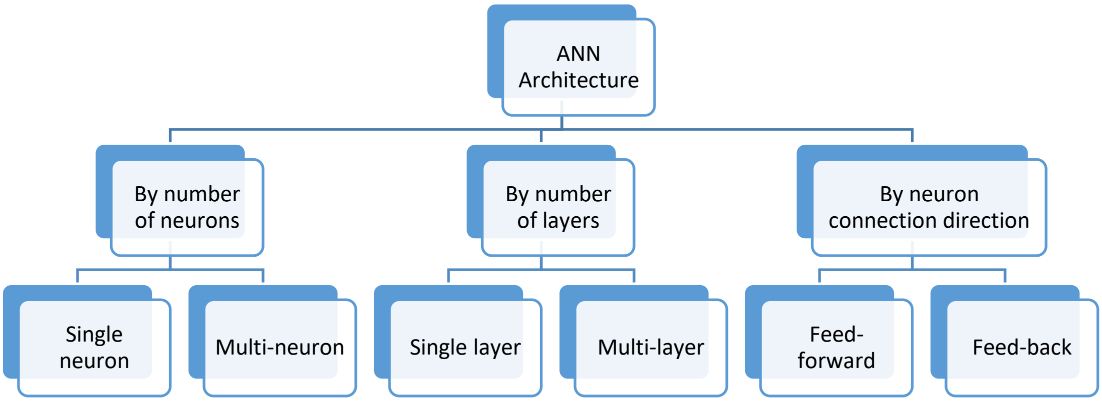
ANNs consist of architecture, activation function, and learning algorithm. ANN architecture consists of neurons, the basic unit of the network, and the network structure that connects them. One or more input values coming to the neuron are converted into output with the help of a function and transmitted to the neuron or neurons as input. The neurons of ANNs can be structured in layers. According to their structure, layers can be categorized into three main types: input, intermediate (hidden), and output layers. Data is sent to the input layer as inputs. The intermediate layers are multiplied by a weight, converted into output, and sent to the output layer. These layers may vary depending on the preferred architecture and type of ANN. For example, an ANN can be composed of a single layer or many different structures, such as a memory cell, Kohonen layer, Grosberg

---

[15] For statistical terminology and artificial neural terminology mappings made by Günay et al. (2007). See "Introduction To Single Variable Time Series Analysis".

layer, context layer, and content layer that can be added to the networks. Another type of architecture can be established by directly connecting the input layer neurons to the output layer (Günay et al., 2007). The architectural structure types of ANNs are shown in Figure 3.

Figure 3: ANN Architecture Structure Types



Source: Eğrioğlu et al. (2020).

Many ANNs and their derivatives have been modeled, and new models have been produced according to their purpose of use. In this study, we use feed-forward and feed-back ANNs with multi-neuron and multi-layer. ANNs with multi-neuron and multi-layer are expected to be efficient with nonlinear data.

Optimizing the adjustable parameters of an ANN is called "training the network" (Hansson, 2017). In network training, the connection weights of neurons are determined. The ability of the network to generalize as a result of training is called "network learning", and the rules to be followed in changing the weights are called "learning rules".

Advantages of ANNs could be stated as the following. First of all, ANNs give better results than econometric models with nonlinear data structures (Bajracharya, 2010). They can solve complex problems that are difficult or impossible to model mathematically. No prior knowledge is needed in modeling. It is not necessary to know the connections of the parameters with each other. There is also no need for assumptions about these connections (Öztemel, 2016). ANNs are easy to understand because they do not have a complex theory and can make better forecasts (Günay et al., 2007). Even if any neuron does not produce useful information, the neuron has fault tolerances that do not leave the whole network dysfunctional (Kuru, 2022). They adapt to the regime change in the data faster than conventional models (Gerhart et al., 2018). They can complete missing information. After training, they can work with incomplete information (Öztemel, 2016). ANNs can use different learning algorithms. Less affected by noise, chaotic components, and heavy queues (Masters, 1993). As a result, ANNs have emerged as an important option for forecasting nonlinear yield curves because they can handle any data structure flexibly (Vela, 2013).

On the other hand disadvantages of ANNs could be put forward as the following. First, there are too many initial parameters to be selected in ANN (Kaastra & Boyd, 1996). Constructing the network and training data requires experience and many decisions, such as training termination criterion, number of layers, and neurons, can be made by trial and error. Unlike traditional methods, it is not guaranteed to produce optimal solutions (Öztemel, 2016). Additionally, the optimized model may not be valid in every dataset (Kaastra & Boyd, 1996). Further, training the network takes time and ANNs tend to memorize the relationship between variables instead of extracting a general trend from the data, as the goal of learning algorithms is to make better predictions. This leads to the problem of overfitting (Szenczi, 2016). Optimizations are directly affected by sample data selection (Öztemel, 2016). "Black Box" refers to the fact that the network and the results cannot be represented or predicted by a theoretical method and that they cannot explain how they generate outputs while taking inputs and producing outputs as results (Öztemel, 2016; Kožíšek, 2018). The fact that the information in the network is hidden or distributed in neurons makes it challenging to interpret the behavior and results of ANNs (Atalay & Çelik, 2017). The inability to explain the behavior of the network leads to a trust problem. ANNs are skilled at identifying relationships in non-linear financial time series data. However, interpreting these connections is almost impossible because it is a black box. Since there is no certain rule, choosing parameters in optimization by trial and error is also referred as black art for ANNs (Castellani & Santo, 2006).

2.2.2.1. Fundamentals of Neural Networks

The structure of artificial neural networks (ANNs) involves the computation of neuron inputs (referred to as "net"), activation functions, computation of neuron outputs, learning rules dictating weight adjustments, and optimization methods. Many alternative approaches have been developed in ANN mathematics. However, since some of these approaches are prominent, many sources have treated only ANN algorithms that include these approaches as a general structure. For example, the delta learning rule, gradient descent optimization method, backpropagation, and calculation of the input of neurons are the topics most frequently covered and in this study they are explained in detail.

ANN aims to find the structural parameters that minimize the cost. For this purpose, a model calculates the revised weight ratios with fewer errors according to the selected learning rule. Cost function is defined as the following

$$\mathrm{Cost(w)} = \mathrm{RSS(w)} + \varrho l_p \qquad (2.16)$$

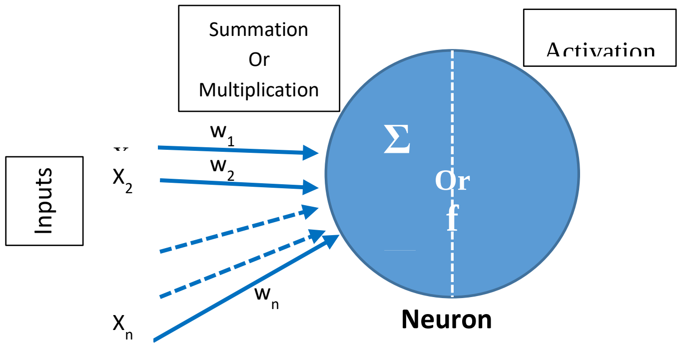
$$l_p\ norm: \|w\|_p = \left(\sum_i |w_i|^p\right)^{\frac{1}{p}} \qquad (2.17)$$

where w defines structural parameters, ϱ defines regularization parameter, p is the norm degree and RSS is the residuals sum of square.

In order to achieve this goal, the following mathematical steps are applied to find the best weight values (optimization) in ANN training: The network training begins with the selection of weights (usually random) and setting the maximum number of epochs/iterations. In the feedforward step, the "net" value arriving at each neuron in the input, hidden, and output layers is calculated. Then the activation formula in each neuron is computed. In the error evaluation step, the error is assessed by continuing the training until the error decreases to the predetermined level or until the predetermined number of iterations is reached. In the propagation step, the output layer propagates its errors backward and calculates error distribution gradients. In the tuning step, weights and biases are adjusted through the derivative of the activation function of neurons with error distribution gradients. This allows the network to learn (Lewis, 2017a).

Neurons are the basic processing unit of ANNs. Their functioning can be summarized as receiving inputs multiplied by weights that will be optimized in the process of running the model, producing an output with a selected activation function, and distributing it to the next neurons with weights that will be optimized (Jacovides, 2008).

Figure 4: ANN Neuron

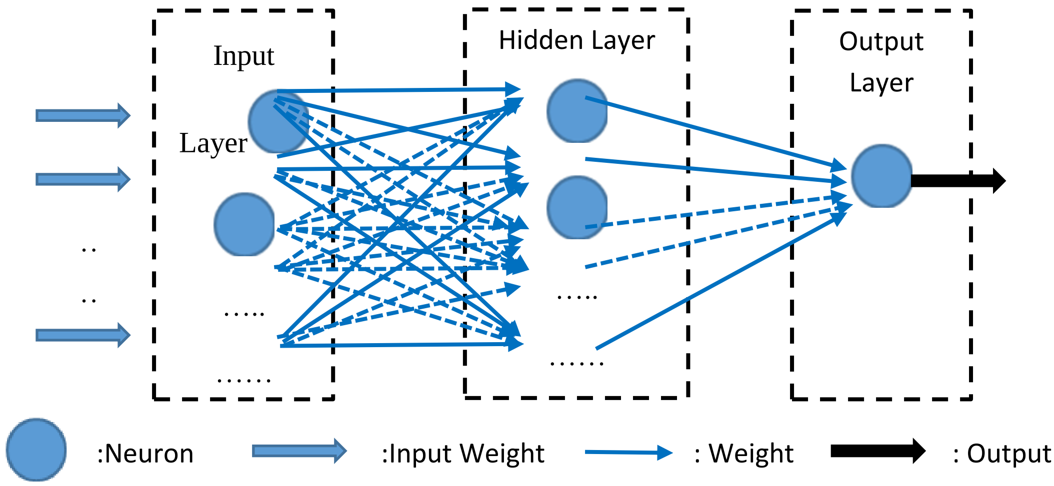


Source: Lewis (2017a).

Neuron, shown in Figure 4, has five basic elements: inputs, weights, sum function, activation function, and output. The sum function uses the net input sum, multiplication, maximum, minimum, median, pruned mean, majority, cumulative total net input, and their variations (Öztemel, 2016). As employed in this study, the weighted sum method is prominent in many.

Layer is a structural component of the network and contains groups of neurons that perform different tasks for processing data. Figure 5 shows the layers that make up an ANN, the neurons in these layers, and the connections between the neurons.

Figure 5: ANN Structure



Source: Lewis (2017a).

The information coming out of the neurons, also called nodes or units, activates the other neuron to which it is connected according to its weight. Input layer neurons distribute the inputs to hidden layer or output layer neurons. The hidden layer sends the information processed in its neurons to the output layer. Training ANNs can result in outputs with very large values, making it difficult to continue training. To avoid what Kaastra and Boyd (1996) refer to training paralysis, neurons use an activation function. Each neuron contains an activation function that ensures that the threshold value required for the input to activate the neuron and the output of the neuron is within a certain range (in ranges such as [0,1] or [-1,1]). In a layer, usually, all neurons have the same activation function. The neurons combine inputs (usually by taking a weighted sum), produce output with the activation function, and send it to the next neurons. Note that, in this study, unless stated otherwise, the input of a neuron is calculated (using summation) as

$$net_j = \sum_{i=1}^{n} w_{ij} x_j + b_j \tag{2.18}$$

w: weights, b: bias.

The hidden layer increases ANN's efficiency and flexibility. It is not preferred to use activation functions in the input and output layers. The activation function to be used in the hidden layers should be smooth, monotonically increasing, and differentiable (Hutchinson et al., 1994). The number of layers of the network refers to the depth of the network (Kožíšek, 2018). As the number of hidden layers increases, the risk of the network getting stuck in local minima increases. For this problem, it is recommended not to use more than two hidden layers.

The activation function matches the input and the output. This matching is done with linear or nonlinear functions (Eğrioğlu et al., 2020). Logistic (sigmoid) activation functions used mostly:

$$f(net) = \frac{1}{1+\exp(-d \times net)} \tag{2.19}$$

net: input values of neurons. d: slope parameter (constant and usually taken as 1). The output is in the range [0,1]. This differentiable function facilitates calculations (Lewis, 2017a). It is preferred in time series since it has a nonlinear differentiable structure. It learns the average behavior of the data well (Klimasauskas, 1993).

Another activation function used in this study is the Hyperbolic tangent:

$$f(net) = \frac{\exp(net) - \exp(-net)}{\exp(net) + \exp(-net)} \tag{2.20}$$

This function, which produces outputs in the range [-1,1], has the same properties as sigmoid but can be preferred for more complex nonlinear problems due to its wider output range. Although the function does not trivialize strongly negative inputs, it cannot always be claimed to give better results (Lewis, 2017a). This function is used in memory cells in the LSTM model.[16]

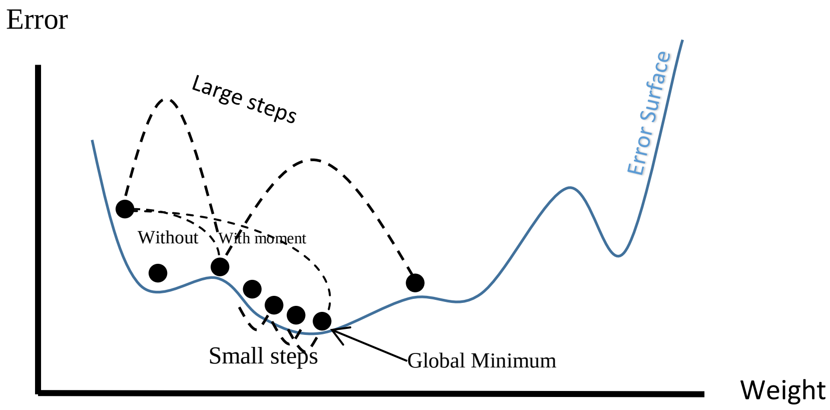
Various optimization algorithms have been developed using different approaches for reaching a solution of ANNs (Hansson, 2017). ANN aims to find appropriate weights (w) and thresholds (b) for inputs (x). In the error evaluation step, the difference between the value found by the neurons and the targeted value is found. For this purpose, the error

$$e(t)=y(t)-\hat{y}(t) \tag{2.21}$$

is calculated at any iteration. Figure 6 shows the change in the error surface, and the minimum error is searched on this surface with optimization. This search is stopped when the error decreases to a certain level or after a specified number of iterations. To reach global minimum levels, adjustments are made to the ANN weights. This adjustment method and additional structures (e.g. momentum) to be applied to this method provides optimization of the ANN.

[16] For some other activation functions used in the applications, see Lewis, 2016; Eğrioğlu et al., 2020; Öztemel, 2016; Gürsakal, 2017; Chen et al., 2019; Günay et al., 2007.

Figure 6: Error - Weight Graph



Source: Lewis (2017a).

Delta learning rule aims to minimize the difference between the output and target values. Gradient descent is used to minimize the cost (error) function. Backpropagation is a gradient descent method adapted to multilayer ANNs. These are used for optimization in ANNs. Epoch is the process of running the inputs through the network and optimizing the weights once. Batch size is the number of training samples used to train the input data. Momentum updates learning parameter. These terms relate to optimization that should be selected in the training of ANN models. The general mathematical structure of ANNs, which involves all these terms, has been discussed in Appendix 2. In the following sections, a brief explanation of feed forward and feed back models, which are ANN approaches used in the study are discussed

2.2.2.2. Feed-Forward Neural Networks (FFNN)

ANNs that receive output by running the input data forward in the network are called Feed Forward Neural Networks (Hansson, 2017). FFNNs are error-tolerant networks that can be applied to large data sizes where exact rules are basically not given as constraints. These networks, which are usually multilayer, perform nonlinear transformations in their hidden layers. In the layers, neurons perform mathematical operations to send data to the output neuron while the output neuron collects the incoming data in weighted form (Lewis, 2017a). MultiLayer Perceptron (MLP), Neural Network (NNET), Neural Network Time Series (NNETAR), Neural Network Nonlinear Autoregressive Model (NNETTS), General Regression Neural Network (GRNN), Extreme Learning Machines

(ELM), and General Method of Data Handling (GMDH), which are FFNN models with the general structure in Figure 5, are the forecast models/algorithms used in this study. Many other models can be constructed.

MultiLayer Perceptron (MLP) is a feed-forward model with hidden layer, and all neurons are connected (Hutchinson et al., 1994). It consists of an input layer, a hidden layer or layers, and an output layer (Lewis, 2015). In MLP, weights only connect neurons to neurons in different layers. Although there is usually one neuron in the output layer, there can be more than one depending on modeling or optional (Günay et al., 2007).

In MLP, also called propagation or backpropagation model, the generalized delta learning rule obtains the network output by forward calculation while revising the weight matrix by backward calculation (Öztemel, 2016).

Like other deep learning models, the MLP learns hierarchical representations, progressing from high-level features to low-level features (LeCun et al., 2015). Since MLP memorizes the dynamics in each sample, it can detect a general pattern that fits the input distribution well (Castellani & Santo, 2006).

For nonlinear data, a multilayer ANN is used because a single-layer network is not sufficient. However, a single hidden layer is sufficient for data whose input and output data take continuous values. ANN with two hidden layers can be said to have a structure that can be used for all types of data. The first layer defines the data i.e. determines the field. The second layer combines them with the "AND" function. The output reveals the desired result to be obtained from this field. The ANN in this structure converges to the constant in the Taylor expansion (Alpaydın, 2010).

As the number of hidden layers increases in MLP, the interpretation of neurons becomes more difficult. If there is only one hidden layer, it can be considered that high weights are the connections of positive outcomes, and low weights are the connections of negative outcomes.

Neural Network Auto-Regressive (NNAR): If the nonlinear autoregressive model is defined as

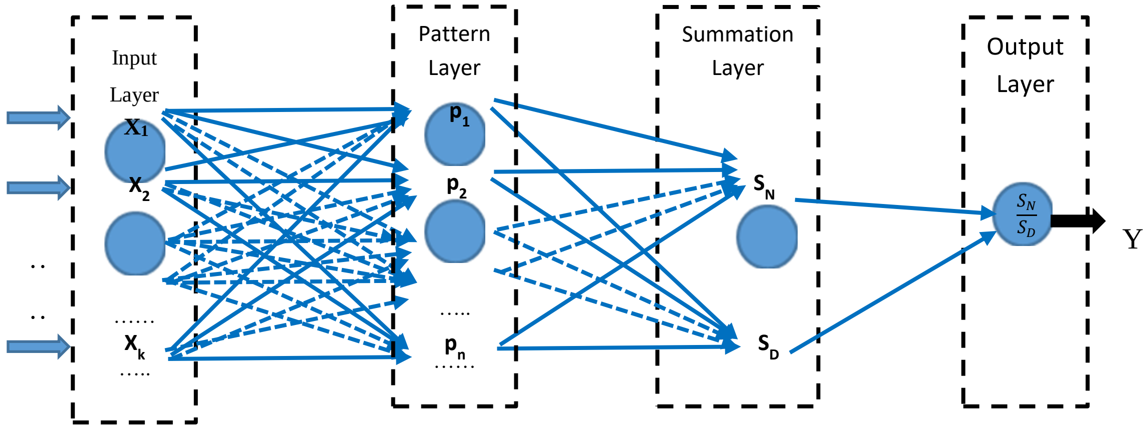
$$x_t = f(x_{t-1},\ldots, x_{t-p}) + \varepsilon_t \tag{2.22}$$

where f is a nonlinear function, supervised ANNs using ($x_{t-d}, \ldots , x_{t-d-m+1}$) values as input can be expressed as Neural Network Auto-Regressive (NNAR). The ANN representing the nonlinear function (f) is trained, and the results are used for forecasts (Stefani et al., 2017). Initial weights are chosen randomly in the model where missing values can be omitted. The average of the forecasts is considered as the result. The iterative forecast approach is used for multi-horizon forecasts.

For this approach, "NNET," "NNETAR," and "NNETTS" functions are executed Neural Network Time Series (NNETAR) is a Single hidden layer feedforward neural network (SLFN) model that models the forecasting of univariate time series. The neural network nonlinear autoregressive model (NNETTS) is a single hidden layer nonlinear neural network model with linear output and delayed values as input.

General Regression Neural Network (GRNN) consists of input, pattern, summation (two neurons), and output (or decision) layers. They are used in linear or nonlinear regression predictions with continuous dependent variables. These nonlinear mapping networks are fast, robust to extreme values, capable of converging solutions to each function structure with sufficient data, and have high performance even with a small number of samples (Lewis, 2015).

Figure 7: GRNN Model



Source: Lewis (2015).

Expected value of y:

$$E\,[y|x] = \frac{\int_{-\infty}^{+\infty} y f(x,y) dy}{\int_{-\infty}^{+\infty} f(x,y) dy} \tag{2.23}$$

x: attribute, f: compound probability density function.

GRNN estimates f from the samples. The input layer has k neurons for k features. The input layer distributes the inputs to the pattern layer. Activation function used in n number of neurons in the pattern layer for n number of observations is

$$p_i = \exp\left[-\frac{D^2}{2\sigma^2}\right] \tag{2.24}$$

$$D^2 = (X - X_i)^T (X - X_i) \tag{2.25}$$

i: neuron, D: Euclidean distance, X: input vector, X_i: i^{th} training input vector, σ: smoothing parameter. As σ gets larger, the distribution converges to a multivariate Gaussian as the importance of training data away from the predicted values increases and, moves away from a Gaussian distribution as σ gets smaller. The neurons in this layer are all connected to the two neurons of the summation layer. One neuron of the summation layer finds the weighted sum (S_N), and the other finds the unweighted sum (S_D) of the outputs of the pattern layer.

$$S_N = \sum_{i=1}^{n} y_i p_i \tag{2.26}$$

$$S_D = \sum_{i=1}^{n} p_i \tag{2.27}$$

Output layer finds the normalized prediction value by

$$\hat{Y}_i(\mathrm{X}) = \frac{S_N}{S_D} \tag{2.28}$$

In GRNN, the error function is MSE and the error minimization method is Conjugate GD. (Lewis, 2015).

It does not assign weights to input variables. It is a local approximator. The relationship to be found for predictions does not need to be valid in all data. It is sufficient to determine only the parameter σ at the beginning ($0<\sigma\leq1$) (Gheyas & Smith, 2009).

Time Series Forecasting with GRNN (GRNNTSF) is an autoregressive neural network that forecasts time series with GRNN regression using lagged values as input.

Extreme Learning Machines (ELM) network is a single hidden layer feed-forward neural network (SLFN) that can eliminate the negative features of FFNN, learn faster, and have better generalization ability most of the time. It does not have problems such as getting stuck in local minima, not being able to choose the optimal learning rate, and overfitting (Huang et al., 2006).

Unlike other ANNs, the training of the model, whose input weights and biases are chosen randomly at the beginning, is completed by first calculating the hidden layer output matrix and then the output weights (Huang et al., 2006). The calculation is completed by doing it all at once instead of gradual error reduction. Randomizing the weights between the input and hidden layers and linearizing the rest of the structure significantly increases the learning speed. The random feature matching (weight assignment) allows the nonlinear part to be passed quickly (Huang et al., 2004).

This setup of the model is based on the theory that; where N is the number of random samples, $\widetilde{N}$ is the number of hidden neurons, activation function is infinitely differentiable, and $\varepsilon>0$ is taken as a small number, it is necessary to make

$$||H_{N\times\widetilde{N}} W_{\widetilde{N}\times m} - T_{N\times m}|| < \varepsilon \tag{2.29}$$

where there will be a number of $\widetilde{N} \leq N$. W: weights connecting hidden layer neurons to the output neuron, b: biases of hidden layer neurons, T: target outputs of hidden layer neurons (Huang et al., 2006).[17]

ELM aims to minimize the size of the weight vector between the hidden layer and the output layer simultaneously with the training error (Huang et al., 2012). Using the Moore-Penrose generalized inverse[18] of these weights ensures linearity in the network. It does not require updates to the hidden layer parameters (Zhu et al., 2015).

If the activation functions of the hidden layer in the SLFN are infinitely differentiable, selecting the input layers and hidden layer bias will give similar results to gradient-based training. By choosing these parameters randomly, the rest of the network structure can be operated as a linear model. This essentially allows us to define the output weights as the generalized inverse of the output of the hidden layer. The ELM algorithm converges to the smallest weight set and error (Huang et al., 2006). ELM has too many neurons in its hidden layer for good updating ability (Zhu et al., 2015).

General Method of Data Handling (GMDH) Type Neural Networks are used to connect the MLP structure to a set of p attributes to the target variable. The connection between input and output variables is made by the infinite Volterra-Kolmogorov-Gabor polynomial.[19] A second-order polynomial to be used in a feedforward sensor converges to this polynomial (Ivakhnenko, 1971). This approach seems only practical for simple

[17] For detailed explanation see Appendix 3 and Appendix 4.

[18] Moore-Penrose solution: AGA=A, GAG=G, (AG)'=AG, (GA)'=GA, then G_{nxm} (G can be replaced by A^+), is Moore-Penrose generalized inverse is A_{mxn}. Least squares solution: In the linear equation system Ax=y.

$$\|Ax^* - y\| = \min_x \|Ax - y\|$$

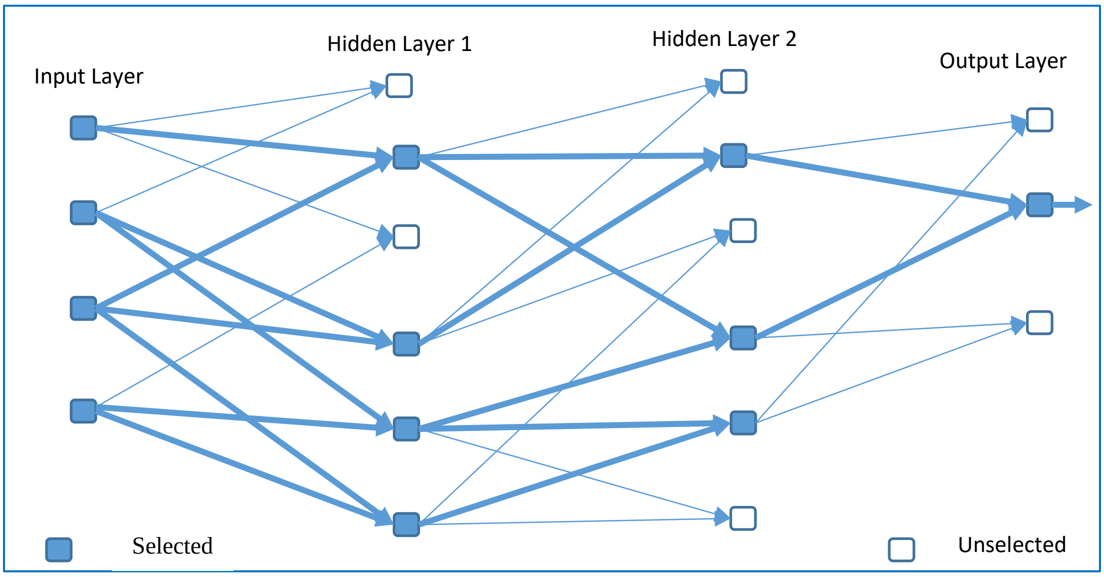
$A \epsilon R^m$: mxn matrix, $y \epsilon R^m$: vector, $x \epsilon R^n$: least squares solution. Let x be all least squares solutions: $\|x^*\| \leq \|x\|$
must be. If Ax=y and $G=A^+$ then $x^*=Gy$ is solution of the general system of linear equations (Rao and Mitra, 1971).

[19] Volterra-Kolmogorov-Gabor polinomial:

$$y = a_0 \sum_{i=1}^{p} a_i x_i + \sum_{i=1}^{p} \sum_{j=1}^{p} a_{ij} x_i x_j + \sum_{i=1}^{p} \sum_{j=1}^{p} \sum_{k=1}^{p} a_{ijk} x_i x_j x_k + \cdots$$

models, as finding a local optimum would interrupt network training and the addition of new units (Taušer & Buryan, 2011).

Figure 8: GMDH Network

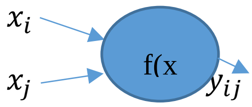


In the GRNN model, whose structure is shown in Figure 8, p input attributes are transmitted to the first hidden layer from p neurons in the network's input layer. Each neuron in the hidden layers and the output layer constructs the quadratic function of all binary combinations of the outputs of the neurons in the previous layers. The degree of polynomial in the layers is 2^l. l: layer order. The number of neurons in the first hidden layer is p(p-1)/2. In the first hidden layer, neurons use binary combinations of signals from the input layer as input. These signals are transmitted to the next layer using a second-order polynomial function. Additionally, the use of a second-order transfer function in the next layer causes its output to increase to the fourth order. This same process is applied to the next layers. As the layer progresses, the polynomial degree increases exponentially. However, at a certain level where the performance improvement does not increase sufficiently, the addition of new layers is stopped.

If the model uses binary combinations of input signals at each layer, it can become overextended. However, this issue can be addressed by eliminating neurons in these layers. The f functions obtained to reach the target estimate (y) are then compared with the Mean Squared Error (MSE) criterion. Neurons with poor estimations are removed, and the f functions that provide the best estimate of the desired quantity are used as input in the next layer. This process of comparing neurons and layers and selecting the best functions and neurons results in a self-organizing structure in which not all neurons are connected to each other (Lewis, 2017a).

This structure adds layers to the network, causing it to explore the entire state space. This structure is different from traditional BP. Figure 9 shows the output of a neuron in the first hidden layer.

Figure 9: GMDH Network Neuron Output



$$y_{ij} = f(x) = \alpha + \alpha_1 x_i + \alpha_2 x_j + \alpha_3 x_i^2 + \alpha_4 x_j^2 + \alpha_5 x_i x_j \qquad (2.30)$$

f: linear regression obtained with the training set (x_i, x_j): possible pairs of attributes p. α's are similar to MLP weights. With the test set, first, the coefficients of the f functions, called partial descriptors, are found.

2.2.2.3 Feed-Back Neural Networks

In FFNNs, neurons only receive signals from the previous layer and transmit them to the next layers. Recurrent neural networks (RNN), which are feed-back neural networks, on the other hand, are approaches that allow receiving signals from subsequent layers by adding a time-delayed structure to feed-forward networks (Szenczi, 2016).

In RNNs, the back connections of neurons loop their activation. This loop teaches the network the concept of time and provides it with a memory. The delay neuron in the context layer stores the previous time information and sends it back to the network with the next time data (Lewis, 2015). FFNN does not take continuity in time series into account. Since RNN models this in its algorithm, it is used in time series forecasting. It stores time-dependent information in its hidden layers, allowing it to learn time-dependent patterns in time series. Since the outputs are affected by lagged inputs, outputs, and hidden states in addition to current inputs, RNNs are dynamic (Öztemel, 2016).

In the RNN, the outputs of the delay unit are used as input to the hidden neurons along with the current input. This structure, called short-term memory, keeps the information obtained by the hidden layers in the network's memory since the activation value and/or output value of the hidden layer is given back to the network as input in the training with the previous time data (Lewis, 2017a).

RNNs can be constructed as full recurrent networks, in which there are forward and backward connections without specific rules. In these networks, all connections can be trained. Alternatively, partial recurrent networks can be created, where the recurrent connections are formed only between content units in the hidden layer neurons. There have been several models developed in addition to the simple RNN (SRNN) in the RNN approach. In this study, forecasts are created using ENN, JNN, LSTM, GRU, and SRNN algorithms, and are discussed in separate sections.

Simple Recurrent Neural Networks (SRNN): To learn patterns in sequential data, the SRNN uses an approach that "recurrent" patterns in predictions. The structure is suitable for remembering patterns in the short-term past from sequential observations, while the long-term past vanishes in training (Namın & Namın, 2018).

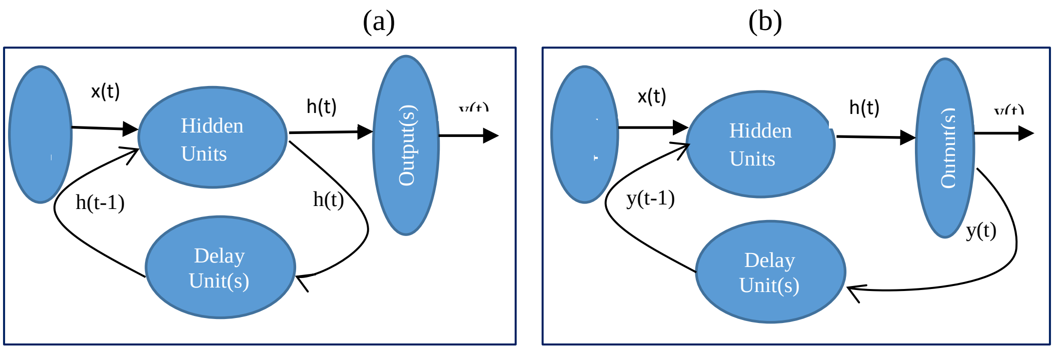
The SRNN algorithm relates the past features of the series as a whole to the outputs (Graves, 2012). In this algorithm, BP is applied to the entire past time. This leads to a reduction of weight revisions from the current time to the past. SRNN function:

$$h_t = \tanh(w_h h_{t-1} + w_x x_t + b) \tag{2.31}$$

h_t: new state, h_{t-1}: previous state, f(**.**): activation function, x_t: current input, w_h: previous state weight, w_x: current input weight.

Elman Neural Networks (ENN) and Jordan Neural Networks (JNN) include input, context (repetitive or delay), hidden, and output layers in their structure (Lewis, 2017a). In the context layer, ENN sends the hidden layer outputs to feedback, and JNN sends the output layer outputs to feedback (Eğrioğlu & Baş, 2020). These types neural networks learn short-term patterns. The structure of these networks is shown in Figure 10.

Figure 10: ENN (a) and JNN (b) Models



Source: Lewis (2017a).

In these networks, the activation function is not used in the input layer. In the output layer, there is a linear activation function where the inputs are collected. In the hidden layer, the activation function varies according to preference (Öztemel, 2016).

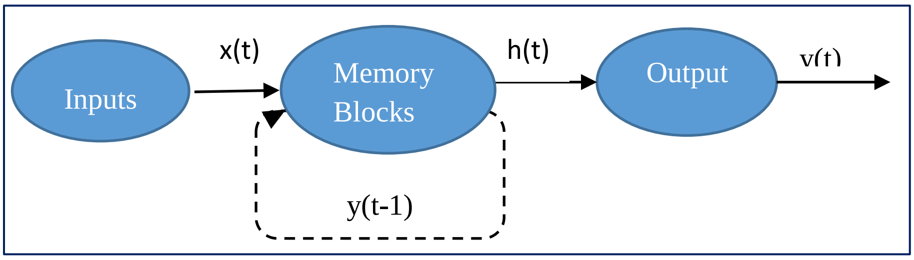
The number of neurons in the context layer equals that of the hidden layer, with connections established between all neurons of these two layers. Connection weights, which repeat the previous values of the hidden layer in Elman networks and the output layer in Jordan networks, with values equal to 1 in delay (context) units, serve as memory to provide input to the context layer. The information in this memory is given as additional input to the hidden layer neurons in the next time step. These networks aim to learn sequential and time-varying patterns.

Unlike MLP, in these networks where previous activation values from the context layer are also used, the learning rule is generalized delta, as in MLP. The algorithm in recurrent connection weights is the same as MLP. These networks are active in time series. Thanks to additional memory units, these networks generally make better predictions than FFNN (Adhikari & Agrawal, 2013a). ENN especially successfully models the first-order linear data structure.

Long-Short Term Memory (LSTM) network is a type of RNN with an AR structure where the data length can be chosen arbitrarily (Hansson, 2017). While RNN is used for data structures with a few lags, the LSTM model has been developed for very long lags. The hidden neurons in the RNN model are short-term, while the memory blocks used in the LSTM model are long-delay structures. The LSTM model bridges long-term previous information with memory blocks. With these additional features, LSTM is able to learn more successful, faster, and longer-term relationships (Lewis, 2017a). No matter which form the autoregressive structure in the data is in, the network has the ability to learn this structure. While in an ordinary RNN, the number of feedback loops needs to be determined by the implementer, in LSTM this is not necessary (Hansson, 2017). LSTM has been successfully used many times in forecasting stocks and indices (Abar, 2022).

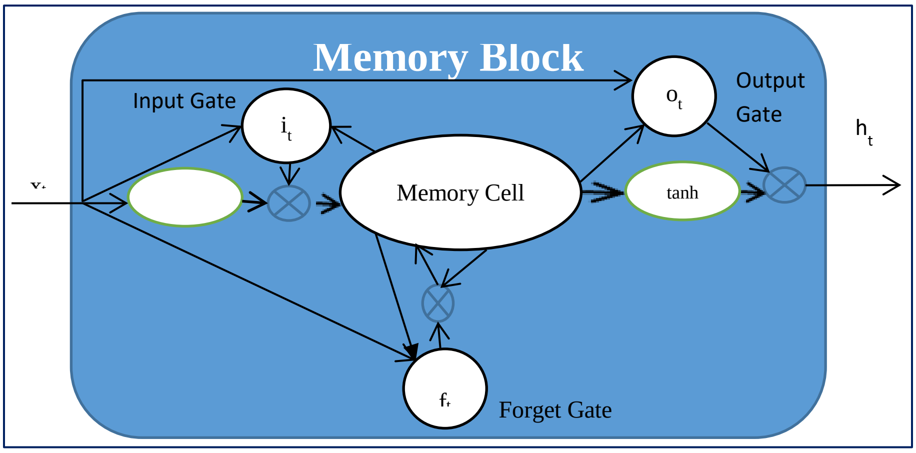
In traditional RNNs, the gradient equivalent to the weight matrix, which has a large exponential power due to being multiplied by itself many times, can cause error signals to shrink and disappear or grow exponentially fast. This vanishing or explosion causes the network to be slow or unable to learn the long-term dependency (Lewis, 2017a). If the errors grow to extremely large values, the weights oscillate on the surface of the error graph (see Figure 6) and unstable learning occurs (Hochreiter & Schmidhuber, 1997). LSTM solves the problem of vanishing or exploding gradients in RNNs with a memory cell. The memory cell prevents information from being lost or corrupted, which ensures that the cell's error remains constant or deletes the information of other units (Lewis, 2017a). The LSTM structure is shown in Figure 11. and the memory block in the structure is shown in Figure 12.

Figure 11: LSTM Network



Source: Lewis (2017a).

Figure 12: Memory Block Structure



Source: Lewis (2017a).

The memory block manages the updating of the block and the retention or exclusion of information (Kožíšek, 2018). Memory blocks have connections and gates that send signals to the block input and all gates (Eklind, 2020). Its structure also includes a constant error loop, output activation function, and surveillance links (Greff et al., 2015). The block has multiplicative input, output, forget, and other optional gates. The inputs of the block are multiplied by the activation function of the input gate, its outputs are multiplied by the output gate, and the previous block is multiplied by the forget gate. The general formula for input, forget, and output gates:

$$a_i^{(t)} = \sigma\left(b_i^a + \sum_j u_{ij}^a x_j^{(t)} + \sum_j w_{ij}^a h_j^{(t-1)}\right) \tag{2.32}$$

a: input (g), forget (f) and output (o) gates; σ: activation function, x: input, h: current hidden layer vector containing the output of all LSTM cells, b: constant, u: input weights, w: recurrent weights, i: cell, t: time. Input gate checks for new information coming into memory. At the forget gate, if $f_i^{(t)} = 1$ the door opens, if $f_i^{(t)} = 0$, it is closed. If the door is open, the cell state is first given as input to the cell, if not, it is not. Cell status revision:

$$s_i^{(t)} = f_i^{(t)} s_i^{(t-1)} + g_i^{(t)} \sigma\left(b_i^f + \sum_j u_{ij} x_j^{(t)} + \sum_j w_{ij} h_j^{(t-1)}\right) \tag{2.33}$$

In the addition process on the right side of this equation, the first term represents the cell state information controlled by the forgetting gate, and the second term represents the input information controlled by the input gate. Cell state information is also calculated at the output gate:

$$h_i^{(t)} = tanh\left(s_i^{(t)}\right)o_i^{(t)} \tag{2.34}$$

This gate behaves like the RNN hidden layer state (Abar, 2022).

The forget gate has been added later to block. In memory blocks, the weight of an input to the hidden layer is checked. If this input has no effect on the output, it is sent to the forget gate and this input in the block is reset to zero (Eklind, 2020). The peephole structure has been added to the network decides whether to reset the memory block completely by evaluating the inputs coming to the forgetting gate and the outputs coming from other cells using its own activation function, which we can call preselection. This application with the peephole allows the network to recognize rhythmic patterns in the series more effectively (Kožíšek, 2018).

LSTM can memorize the data sequence. There are algorithms for discarding, filtering, or adding data from each cell to subsequent cells. In the forget gate the outputs take a value between "0" ("forget everything") and "1" ("keep everything"). The memory gate decides which inputs to modify and which new data to store. The sigmoid function is used to change data and the tanh function is used to add new data. The output gate decides the efficiency of the cell (Namın & Namın, 2018).

LSTMs contain at least one loop. The number of neurons in the input layer is equal to the number of dependent variables (Fischer & Krauss, 2017). It involves many more parameters than traditional recurrent networks. LSTM does not require precise tuning of its parameters. It is a local approach in both data space and time, which is an advantage over other methods with complex training algorithms (Hochreiter & Schmidhuber, 1997).

In the *Gated Recurrent Unit (GRU)*, a hidden memory structure has been designed to record information, instead of using a separate memory cell like in LSTM. This structure includes the update gate (z), which combines the forget and input gates of the memory cell to predict how much the previous memory will be used, and the reset gate, which learns how to incorporate new inputs with the previous memory. If the previous hidden state is not relevant, the reset gate (r) is assigned a value of "0" to reset the GRU. The new input (x_t) is combined with the previous hidden state and sent to this gate.

The update gate allows transferring information that should have been kept in the previous hidden state to the current hidden unit. This gate avoids the gradient reset and explosion problem (Eğrioğlu & Baş, 2020). If the network has found long-term relation, it closes the update gate to use memory contents later. The last memory (hidden activation) is found by combining the current memory ($\tilde{h}_t$) and the weights revised by the previous memory update gate using the interpolation method. This network, which has a simpler structure than LSTM, can produce the same or better results as LSTM (Lewis, 2017a). GRU functions:

$$\text{Update gate: } u_t = \sigma(w_u x_t + r_u h_{t-1} + b_u) \tag{2.35}$$

$$\text{Reset gate: } r_t = \sigma(w_r x_t + r_r h_{t-1} + b_r) \tag{2.36}$$

$$\text{GRU memory: } c_t = tanh(w_c x_t + r_t \otimes u_c h_{t-1}) \tag{2.37}$$

$\otimes$: kronocker multiplication, $r_t \otimes u_c h_{t-1}$ destroys some of the knowledge from the past.

$$\text{GRU neuron output: } h_t = u_t \otimes h_{t-1} + (1 - u_t) \otimes c_t \tag{2.55}$$

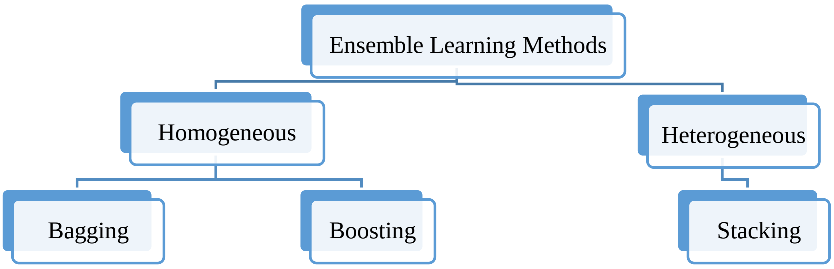
(Eğrioğlu & Baş, 2020).

2.3. ENSEMBLE LEARNING

Ensemble learning (EL) is combining the outputs of a set of learning algorithms (models) to forecast data (Pode & Mackworth, 2010). EL is also called learning together in harmony (Gürsakal, 2017).

EL aims to achieve high-performance results by combining the forecasts of individual models that may have poor predictive performance. Individual models are often termed "weak learners" because they produce forecasts that are only slightly better than random guessing, which is not considered satisfactory. In contrast, EL is considered to produce "strong learners". EL is generally more efficient than individual models of similar or different types (Lewis, 2017b). EL methods are shown in Figure 13.

Figure 13: EL Method



In EL, multiple forecasts are made and these forecasts are combined into a single result. In a homogeneous EL approach, models of the same type are used, while in a heterogeneous approach, models of different types are used. There are also EL approaches that combine these two groups. EL learns quickly how the distribution of target values changes and can be more accurate than single models (Grmanová et al., 2016).

Methods such as random forest, boosting, and bagging combine the strengths of simple basic models. However, these methods do not include probabilities, are not interpretable, and are not simple (Steel, 2017). In this study, the heterogeneous method (stacking) is not used and instead homogeneous models are used.

Tree-based algorithms are the most popular EL. Tree-based learning methods are based on decision trees. Decision trees can be defined as simple structures that divide observations into specific regions using features in the data set and make predictions from each region. However, a single decision tree can often result in high variance and therefore low generalization ability. Therefore, tree-based methods allow multiple decision trees to be filtered and combined using ensemble methods to obtain more complex and more accurate predictions (James et al., 2023).

Decision Trees (DTs) are simple models used for making decisions. They use tree-shaped decision rules to divide data based on specific criteria. DTs are non-parametric and do not make assumptions about the underlying data distribution. They model each step of decision-making using outcomes and conditions and are easy to interpret. They do not need data normalization, dummy variables, or missing value completion; their calculations are simple and fast, and they are effective models for big data. The disadvantages are that they are very sensitive to sample data; that is, they are unstable, and are prone to overfitting (Lewis, 2017b).

DTs are obtained by minimizing the generalization error from the data (Rokach & Maimon, 2005). These trees are built by selecting the most suitable function based on input sets and univariate or multivariate partitioning criteria such as information gain, gain ratio, Gini index, or others. The partitioning of the data into subsets is continued until the stopping criterion is met or the partition criterion is invalid (Balaban & Kartal, 2018). For an explanation of DTs, see Appendix 5.

Bagging (Bootstrap Aggregating) obtains predictions by generating training sets with the bootstrap method to reduce variance and then takes the average of these predictions. Bagging is a method that works by relying on a bootstrapped training set to create multiple trees. In this method, a random sample is taken from the training data. In this process, some observations are usually selected from the same data set in each training cycle and some are dropped. DTs are created on this sample. Each decision tree can have high variance because they are trained based on only a subset of the data set. However, when the predictions of these trees are averaged, the variance of the aggregate predictions is significantly reduced. In this way, the results obtained from the base trees provide a more reliable and stable prediction (James et al., 2023).

Bootstrap is an approach that obtains large observation data from a small set of observations by resampling. It is used when statistical methods are inadequate or the parametric assumption is invalid. It can be used in time series if the data structure is independent and identically distributed. The dependence of the series within itself should be transferred to the samples produced by bootstrap. If the classical bootstrap model is not suitable for this, non-overlapping block, which takes sequential observations of time series as a block, moving block, which takes sequential block observations and preserves the relationship of these observations, circular block, which turns the data into a sample in the form of a circle and applies the bootstrap approach equally weighted on any part of this circle, and stationary block bootstrap approaches using blocks with stationary properties of arbitrary lengths have been developed (İslamoğlu, 2020).

Bagging creates clusters/bag samples by randomly selecting from observations. One model is trained with each cluster. Some of the forecast models, such as ETS and ARIMA, are trained with clusters. The results are combined with averaging in regression problems (Alpaydın, 2010).

Random Forest (RF) is an EL approach that improves the performance of DTs by combining the predictions of many DTs obtained from different samples of the same data. Forecast combination is done by averaging in regression problems. RF makes trees that are weak learners become strong learners.

RF is a self-developed version of the bagging method. In this method, instead of the full set of predictors used in creating each tree, a randomly selected subset is used for each split. This randomness limits the repeated use of a strong predictor in the splitting processes in a tree, thus reducing the similarity of the trees to each other. As a result, RF provides more independent and reliable results by reducing the correlation of the predictions resulting from the similarity of the trees seen in bagging (James et al., 2023). RF is mostly efficient and requires almost no parameter or data tuning (Fischer & Krauss, 2017).

Random selection of subsamples allows the DTs to be differentiated. Unselected samples are used to calculate the forecast error and the weight in the forest. Trees are grown unless they need pruning. Finally, the predictions are combined (Lewis, 2017b).

Although unrelated decision trees are quite sensitive to the training data, when these trees are brought together in RF, the results obtained do not increase the bias and do not show this sensitivity, thus reducing the variance. They are easy to construct and compute and are trained quickly. Missing data can be filled with simple methods. They are effective with big data. They are resistant to outliers. Creating a structure that improves predictions without increasing bias reduces the problem of overfitting (Lewis, 2017b).

Boosting is a performance-enhancing method that combines weak learners with strong learners (Lewis, 2017b). In the Boosting approach, the error probability is tried to be reduced to the desired level by pushing them by training weak models in order (Alpaydın, 2010). Bagging is a homogeneous ensemble learning model that learns basic methods from different pieces of data while boosting is a homogeneous EL model that adapts after each sample or batch of data is generated (Grmanová et al., 2016). In this approach, which generally uses DT as the predictor model, unlike RF the growth of trees is not random. DT is grown according to the results of previous trees.

The application of the model is briefly as follows. First, a simple weak learner $\hat{f}^1(x)$ model is built. x: training data with certain characteristics. Then at each step, the greedy learning function that reduces the error the most is executed. If $\hat{f}^1(x)$ model is incorrect, it will be replaced by a duplicate model ($\hat{f}^2(x)$) is trained by giving more weight to incorrect predictions. Thus $\hat{f}^2(x)$, he focuses more on $\hat{f}^1(x)$'s mistakes and corrects the mistakes of the targets he is struggling to find. In this way, it improves performance by better predicting difficult predictions in each iteration. Training is stopped when a certain number of iterations or a certain performance level is reached. The last step is to merge the results

$$\hat{f}(x) = \sum_{b=1}^{B} \alpha^b \hat{f}^b(x) \qquad (2.38)$$

$\hat{f}^b(x)$: output of weak learner b, α: weight. It gives high weight to high performance (Lewis, 2017b).

Tree-based boosting is an approach based on the sequential growth of decision trees. In this method, each new tree is created with incorrect predictions from previous trees. Thus, the model constantly tries to overcome its errors. In this process, each tree is adapted to a modified version of the original dataset. In boosting, the complexity of each tree can usually be controlled, and these smaller trees slowly improve performance by focusing on areas where the model is weak. The shrinkage parameter controls the learning rate in the process and allows it to focus on the errors of more and different trees. Therefore, slow learning processes generally tend to produce stronger and more accurate results (James et al., 2023).

For the tree based approaches; model is initialized with $\hat{f}(x) = 0$ and r_i=y_i. r: residuals. Then at each b=1,…,B step; first a tree $\hat{f}^b(x)$ model is built. x: training data with certain characteristics. Then update $\hat{f}$ by $\hat{f}(x) \leftarrow \hat{f}(x) + \lambda \hat{f}^b(x)$. This adds a shrunken version of the new tree. Then update r_i by $r_i \leftarrow r_i + \lambda \hat{f}^b(x_i)$. After the end of this iteration (means b =B), the last step is to merge the results

$$\hat{f}(x) = \sum_{b=1}^{B} \lambda \hat{f}^b(x) \tag{2.39}$$

$\hat{f}^b(x)$: output of weak learner b, λ: shrinkage parameter (James et al., 2023).

Boosting is a simple, well-functioning approach that performs optimization without complex nonlinearity and careful tuning, is protected from overfitting, and is insensitive to redundant information. It has disadvantages such as interpretation difficulty, sensitivity to noise, and the possibility of outliers worsening overall performance. Although it reduces bias, it has overfitting and hyperparameter selection problems. Many types have been designed, such as AdaBoost, CatBoost, LPBoost, GBM, Light GBM, and XGBoost (Yakut & Kuru, 2022). The methodologies used in this study are briefly explained below.

First of all, in model-based boost (boost), the individual model (weak learner) and the optimization loss function can be selected as desired. Throughout this study, optimization is done with gradient boosting. Gradient Boosting with Component-wise Linear Models (glmboost) optimizes component-based linear models (weak learner) with the same approach. A generalized additive model by likelihood-based boosting (gamboost) makes predictions using the likelihood approach in cases where individual models (weak learners) are non-linear. This approach, which gives good results in model outputs with a Gaussian distribution, tries to adapt the model residuals to the data at each step with the B-spline method. Further, the data to be fitted are selected by deviation or other criterion. The adjustments made in this way are improved by adding to the previous ones. Gradient Boosting with Regression Trees (blackboost) uses regression trees as weak learners. It does the optimization with 'classic' gradient boosting. Note that, the regressions in the model are difficult to interpret because they are “black box” (Hothorn et al., 2023).

XGboost (Chen & Guestrin, 2016) is a supervised ML method that combines weak Gradient Boosting Machine (GBM) predictions with a boosting approach. They make it faster and more accurate against overfitting with parallel processing to strengthen models that are weak according to the errors between the prediction and the actual observation by iterations. This method, which has the features of tree pruning and working with missing data, is more effective than other methods despite overfitting and bias problems (Yakut & Kuru, 2022).

2.4. FORECAST COMBINATION

Forecast combinations are the process of re-forecasting the results of individual model's forecasts using various methods. Combination can be made directly or by using RF, bagging etc. The preference for combinations is based on the expectation that they forecast better than the model that forecasts best alone. Uncertainties in modeling the dynamics in the data and unmodeled and unpredictable situations reduce the effectiveness of forecasts. Combining models for forecasting can eliminate these problems. A parameter or dynamic not covered by one model might be covered by another model. The combination reduces model risks. Thus, models can cancel each other's negative effects. In a time-varying data structure, different models may work better at different times. Combinations can reduce forecast biases and errors (Raviv, 2016).

The forecast error obtained with these approaches can be, at most, as large as the best individual model. Moreover, the efficiency of the combination increases as the forecast horizon gets longer (Araújo & Cajueiro, 2014). As the efficiency of the models increases, the efficiency of the combination also increases (Andrawis et al., 2011).

It is more efficient to use different approaches for combining them by selecting the models that give the best results in groups containing the same approach. These combinations can be constructed in different models with different algorithms, different parameters, and different input representations. In the regarding literature, it is suggested that only one of the models, which has the same approach, should be used in the combination, the remaining models and failed models should not be included in the combination (Alpaydın, 2010).

Methods have been developed that combine averages, weights, regressions, shrinkage approaches, information criteria, performance, and many different forecast results. The models, which are used in this study, are listed below, where f^c defines forecast combination, N denotes the model number and f_i stands for the i^th^ model yet other methods can also be constructed for the purpose.

(i) Simple avarage

$$f^c = \frac{1}{N}\sum_{i=1}^{N} f_i \qquad (2.40)$$

(ii) Median

$$f^c = median(f_i) \qquad (2.41)$$

(iii) Trimmed average: K trim parameter,

$$f^c = \frac{1}{N-2K}\sum_{i=K+1}^{N-K} Kf_i \qquad (2.42)$$

(iv) Winsorized Mean: λ is trim factor; top/bottom 100 × λ% is winsorized measure; $K = \lambda N$ to be

$$f^c = \frac{1}{N}\left[Kf_{(K+1)} + \sum_{i=K+1}^{N-K} Kf_{(i-K)}\right] \qquad (2.43)$$

The simple average is a combination model that can be used as a benchmark. The trimmed mean takes the simple average of the remaining models, excluding the models with the largest and smallest forecasts. Winsorized Mean limits the effect of outliers on the mean by assigning them smaller absolute values according to a given criterion. Mean approaches such as median, trimmed mean, and Winsorized Mean are insensitive or less sensitive to outliers (Steel, 2017).

(v) Bates/Granger (1969) weight by variance:

$$w_i = \frac{\hat{\sigma}^{-2}(i)}{\sum_{j=1}^{N}\hat{\sigma}^{-2}(j)} \qquad (2.44)$$

Bates/Granger does not take into account the correlation between forecasts.

(vi) Constrained least squares (CLS) regression: The weights are positive and have the constraint of summing to 1. The closeness of the forecasts obtained from the models increases the efficiency of the forecast combination. β: coefficient matrix. CLS:

$$f^c = \beta f_i \tag{2.45}$$

(vii) Rank based weighting (Inverse RANK):

$$f^c = \sum_{i=1}^{N} f_i' \times \frac{Rank_i^{-1}}{\sum_{i=1}^{N} Rank_i^{-1}} \tag{2.46}$$

$Rank_i^{-1}$ is the order of the model in reverse sorting, done according to the MSE criterion.

(viii) Eigenvector approachs: The solution of the equation min(w'Σw), where the constraint w'w =1 is the model weights. w: weights, Σ: eigenvector. If Σ is unknown, replace

$$S = \frac{1}{T_1} \sum_{t=1}^{T_1} (ey_t - f_t)(y_t e' - f_t') \tag{2.47}$$

$$w^{VC} = (e'S^{-1}e)^{-1} S^{-1} e \tag{2.48}$$

S is diagonal. e is error. This is the standard method, and there are different variants of this approach, such as bias-corrected, trimmed, and trimmed and bias-corrected. Bias corrected eliminates bias (by subtracting the column means of the MSPE).

(ix) With Ensemble Learning models: In Bagging, Boosting and RF models, models can be trained with the fitted values of some individual models as input, and forecasts can be combined with the forecasts of these models.

$$F_{EL} = F(f \sim f_1 + f_2 + \cdots + f_k) \tag{2.49}$$

$$f^c = F_{EL}(\hat{f} \sim \hat{f}_1 + \hat{f}_2 + \cdots + \hat{f}_k) \tag{2.50}$$

First, the regression model is established with the fitted values of the individual models $(f_1, f_2, \ldots, f_k)$. Then, this regression (*f*) model is trained with one of the EL approaches (*F*). In the trained model (F_{EL}), the forecasts of the individual models $(\hat{f}_1, \hat{f}_2, \ldots, \hat{f}_k)$ are used as new inputs, and new forecasts (f^c) are obtained.

In forecast combining, the consistency of the approach should take precedence over its optimality. Breaks in data dynamics affect the performance of the combination. For this reason, Hsiao and Wan (2011) recommend a rolling window approach instead of the fixed window or updated window as input data in forecasting.

Many studies have concluded that the best forecast combination model is the simple average of forecasts. If the efficiency of forecasting models is poor, the simple average and the choice of the optimal model give very close results. In this case, it is more pragmatic to choose the mean (Huang & Lee, 2007). Since using a simple average approach does not require the estimation of weights or other parameters, the combination does not suffer from the loss or other inconveniences caused by these weights and similar forecasts (Adhikari & Agrawal, 2013a). Estimating the weights of the models to be used in forecast combining increases the variance. Even the optimal choice of weights may not ensure that the variance is lower than the variance of the individual models (Claeskens et al., 2014)

CHAPTER 3:
DATA AND EMPIRICAL FRAMEWORK

In this section, we study the yield curves of developed countries (G-7), excluding Japan. This section explains the data scope, and features, estimated factor data, operations performed and statistical information on these data, data processing, data-method-parameter selection, and optimization for forecasting. Unlike many other studies, we use daily data. Because obtaining this data is difficult, we directly use the yields at standard maturities, which are readily available in financial data terminals.

3.1. YIELD DATA

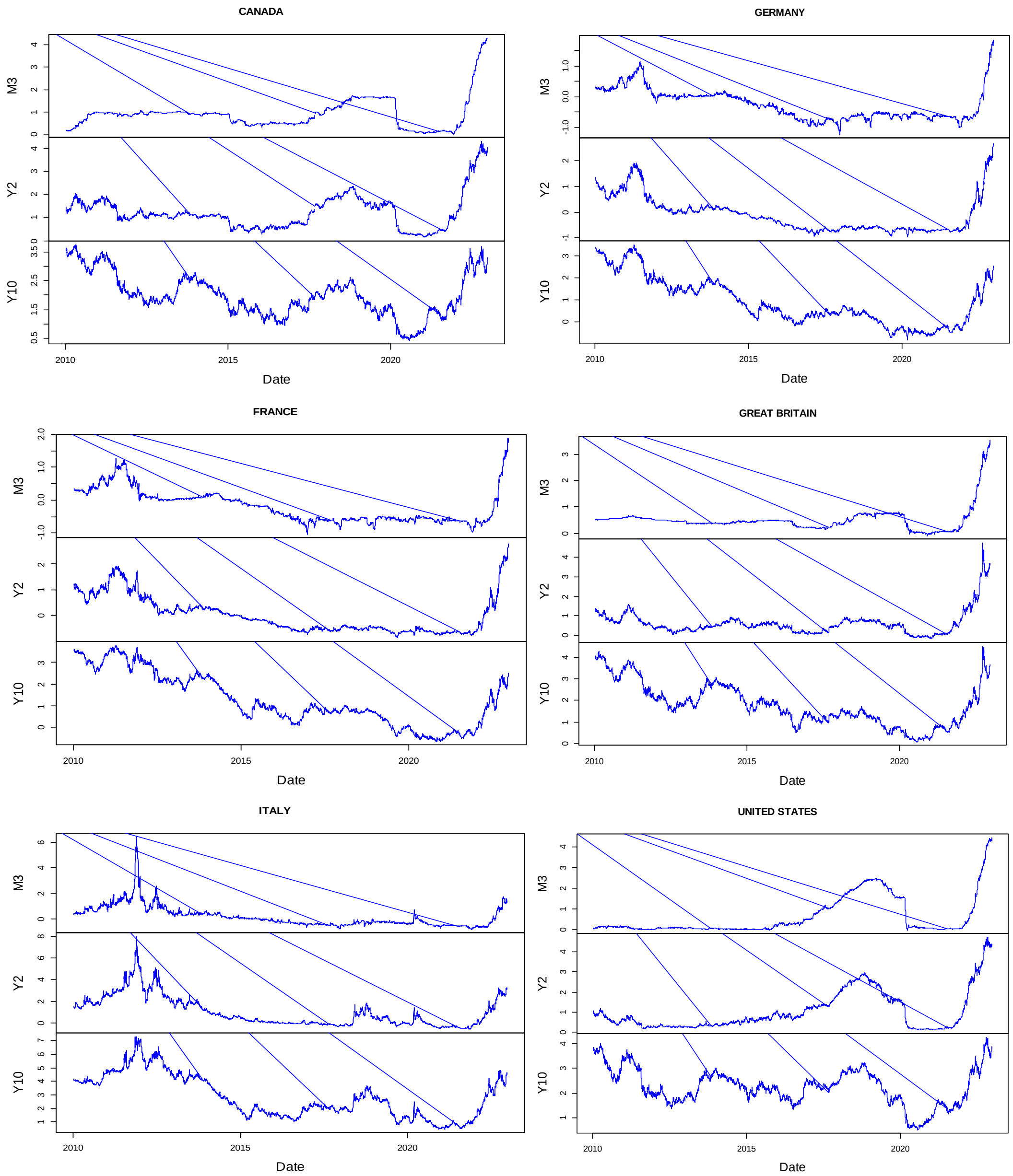
Using fewer maturities in yield curve and unequal spacing of maturities negatively affects the forecast performance (Reinicke, 2019). In the yield curve model, Gürkaynak et al. (2007) do not use government securities those with features that may imply options, those with maturities of less than three months for reasons such as lack of liquidity and different markets, and those with trading in a separate market as data.

This thesis aims to forecast the yield curves of G-7 countries, excluding Japan. Japan is excluded from the study due to insufficient data for the selected period. We collect data for Canada (CA), Germany (DE), France (FR), the United Kingdom (GB), Italy (IT), and the United States (US) from “the investing.com" website for the period over January 04.01.2010 -30.12.2022. We utilize daily yield data for 11, 17, 17, 17, 17, and 11 different standard maturities ranging from three months to 30 years for these countries, respectively. We calculate Nelson Siegel latent factors $\{\beta_{0t}, \beta_{1t}, \beta_{2t}, \lambda_t\}$ daily.

The term ‘missing data’ refers to data that significantly alters the distribution, thereby impacting the analysis and reducing the validity of the results (Suthar et al., 2012). Various methods such as Monte Carlo, interpolation, jumping, discarding, forward filling, and backward filling can address missing data (Akçay et al., 2012). For the empirical application part of this thesis, the data source does not provide regarding data for every day of maturity. To address this, we fill the missing data with the most recent data from the previous days (forward filling). The amount of missing data is less than 1% of the total data and is spread across the entire series period, thus not significantly affecting the series. Some standard maturities are not available during the some period and we only use them for calculating the NS parameters on days when data are available. Additionally, we exclude the data provided on weekends because of the lack of efficient market conditions.

Once the collected data is analyzed, fluctuations, which are not expected from the interest rate market, are observed in some maturities on days. This may be due to the inappropriate results of the model used to convert market interest rates into yields at standard maturities or the use of incorrect data. We adjust such data using a smoothing model. We identify yields that change more than three standard deviations, which lasted for at most four days, then return to their previous level and replace them with new yields obtained by interpolation. Thus, the changes in the series are more reasonable and modellable. For this reason, we take this data on changes as yields for the rest of the study. We calculated latent factors from these yield data. We use this final data version to calculate the forecast error (RMSE) and compare the yield forecasts. As seen in Figure 14 the yield graphs of the countries show that long-term yields follow a similar path.

Figure 14: Yield Graphs of Countries at Some Maturities



3.2. ESTIMATION OF NELSON SIEGEL PARAMETERS

Nelson-Siegel is a nonlinear model whose parameters have been discussed in previous studies. NS models have several solutions. The change in the structure of the yield curve over time and the maturity range analyzed may cause problems in the selection of λ in the NS model. Even if λ is not optimized well, the NS yield curves estimated by obtaining with βs can sometimes provide a better fit than those obtained from the data. (Szenczi, 2016). There may also be a multicollinearity problem between NS factors. In Nelson-Siegel's formula (see Equation 1.7), "$\exp(-m\lambda)$" and "$[(m\lambda)\exp(-m\lambda)]$" terms are similar to each other, so this problem arises. (Marek, 2015). These issues complicate the choice of an appropriate λ (Gilli et al., 2010). Following three approaches are used to estimate the parameters of the NS model.

(i) By finding the data set that gives the least error with OLS from a specified solution space,

(ii) By choosing a fixed λ and finding the βs that minimize the error from the NS converted to a linear model,

(iii) Nonlinear methods.

Note that the third method has problems such as getting stuck in local optima, excessive sensitivity to initial values, and high parameter instability, which make it difficult to interpret the results (León et al., 2018).

Determining λ and β using nonlinear models may not smooth the yield curve appropriately (Çepni et al., 2018). In NS models, using λ as a float significantly disrupts the stationarity of β (Rezende & Ferreira, 2011). If λ is considered a float parameter, the efficiency of the fit may increase, while the efficiency of the forecast may decrease, leading to significant errors in forecasts (Vela, 2013). Accordingly, some researchers have suggested that λ be chosen as a constant parameter. According to Szenczi (2016), to improve forecasts, λ should be chosen as a constant, obtained using an optimization method that minimizes the multicollinearity between NS factors. The fit of the yield curve to the data should be considered. λ is optimized over the medium-term factor. Because this optimization choice may affect the fit error in the long and short terms, a balance between mid-, long-, and short-term optimization should be considered.

Rosadi et al. (2011) state that partial or complete estimation can obtain the NS parameters. In partial estimation, either λ or β is taken as a constant, and β or λ is estimated. In the complete estimation, all parameters are estimated using sequential quadratic programming (SQP), the Nelder-Mead simplex method, and similar constrained optimization methods.

In the two-step NS method, which can be used to determine the factors, λ is first calibrated. Subsequently, other factors are found. The cost of choosing the two-step method is minimal. The one-step NS method estimates the full maximum likelihood by using a Kalman filter. Compared with the two-step method, a one-step estimation using the state-space structure is expected to yield better results. The two-step method with a calibrated λ is simple, convenient, and numerically stable. However, ignoring the estimation error in the first step may distort the result in the second step, resulting in worse optimization. However, the one-step method is mathematically and numerically challenging to follow, and its results are unreliable (Diebold & Rudebusch, 2013). The state-space approach leads to heteroskedasticity, missing data, heavy-tailed data, and measurement errors (Diebold et al., 2006).

Nelson and Siegel (1987) show that finding another τ (=1/ λ) for each dataset has little impact on the fit and that it is better to choose one constant τ for the whole dataset. Rezende and Ferreira (2011) perform the optimization to find an optimal and fixed value of λ with the formula (3.1).

$$\hat{\lambda} = arg \min_{\hat{\lambda}\epsilon\Omega} \left\{ \frac{1}{N} \sum_{n=1}^{N} \sqrt{\frac{1}{T} \sum_{t=1}^{T} \left(\mathrm{y}_t(\mathrm{m}_n) - \hat{\mathrm{y}}_t(\mathrm{m}_n, \lambda, \hat{\beta}_t) \right)^2} \right\} \tag{3.1}$$

They find βs for each λ in the number sequence Ω, which are limited to the maturity range of the yield data using OLS. From the results, they chose the λ as $\hat{\lambda}$ with the lowest RMSE.

Kožíšek (2018) finds λ with an optimization problem to find

$$\lambda = \max_{\lambda \in [2,20]} \left(\frac{1-e^{-\tau\lambda}}{\tau\lambda} - e^{-\tau\lambda} \right) \tag{3.2}$$

(λ =1, 1.5, 2, . . . 19.5, 20). Diebold and Li (2006) state that the curvature is maximum at λ = 0.0609, corresponding to 30 months. Molenaars et al. (2015) repeat Diebold-Li's (2006) λ optimization. However, in the same optimization, they find the exact λ for 30 months to be 0.0598.

Marek (2015) finds λ in a two-step approach. In this approach, NS is considered as a linear equation, assuming a definite constant λ value. He weights the errors in Equation 3.1, assuming that there is a difference between old and new observations.

Some researchers investigate the adverse effects of choosing a fixed λ and present their suggestions. According to Sambasivan and Das (2017), the fit in the NS model is not significantly affected by whether λ is a fixed or variable parameter. However, the yield curve shapes, providing an idea of the economy's future, can only be forecasted by estimating the exponential decay parameter (λ) that determines this shape. According to some researchers, Diebold-Li's (2006) decision to take λ as a constant and remove it as a dynamic factor constitutes a flaw in the model. The deterioration rate, curvature peak, and location are determined by using λ. Taking a fixed λ eliminates the dynamics of these points determined by the curvature in the yield curve (He, 2013).

Diebold and Rudebusch (2013) argue that the United States bond yields have recently been below zero in the short run, suggesting that it is inappropriate to set λ fixed at 0.0609 and that a time-varying λ can be used in the model so that yield curves can capture the curve to fit short-term negative yields.

Annaert et al. (2013) propose a ridge regression grid search method for finding NS parameters. According to this methodology, the λ with the lowest error from the solution space is found first. The conditions for the optimal λ are calculated. The coefficient is re-estimated only in cases in which this condition is satisfied. As long as this condition persists, λ remains constant. This method, which provides flexibility to the yield curve, also prevents multicollinearity. León et al. (2018) reach the same conclusions using the same approach. They perform fitting and forecasting using a roll apply procedure. They state that instead of βs estimated with a fixed λ, which is not economically meaningful, float λ selection and β values, which can express a connection in economic indicators, can be estimated.

This study calculates the level, slope, and curvature data from the yields and compares the findings with the Nelson–Siegel factors. In many studies, λ is taken as a fixed parameter, and the other factors are estimated. In this study, due to the aforementioned harmful properties of a fixed λ, λ is also one of the factors to be estimated where the following approaches are applied:

(i) Float λ: Determine the best-fit yield parameters within a specific range.

(ii) Mean λ: Taking the mean of the λ values obtained in the first approach and recalculating the β values

(iii) Median λ: Taking the median of the λ values obtained in the first approach and recalculating the β values

(iv) Mean-median λ: Taking the average of the λ obtained in the second and third approaches and recalculating the β's.

(v) HP Filter λ: Passing the λs obtained in the first approach through the Hodrick-Prescott filter and recalculating the β values.

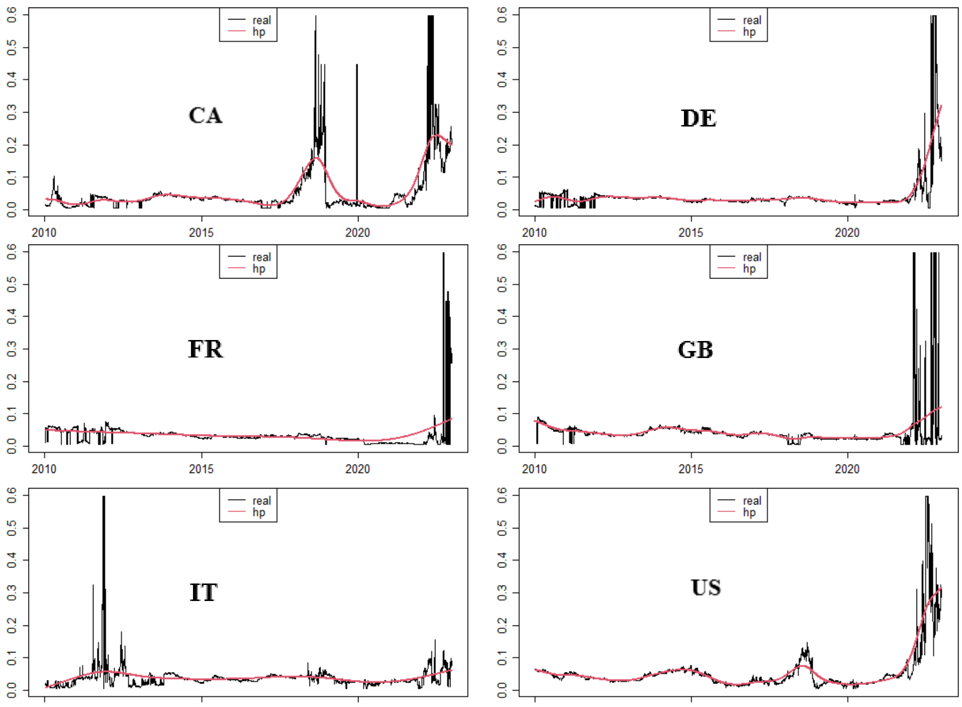
Accordingly, approach (i) creates a series with a standard increase between the shortest and most extended maturity, substitutes the values in this series for λ in the NS model, and finds βs using the linear regression method (Equation 3.1). It obtains more precise λ values by repeatedly optimizing two consecutive λ values with a slight error. The λ value is estimated between 0.001 and 1. For graphs of the obtained parameters, see Figure 15 and Figure 16 show that the NS parameters fluctuate, which is not expected. This approach optimizes the maturity date at the calculated time without considering the parameters of the previous or next day. It outputs the parameters with the smallest error in the linear regression. In this calculation, where no error tolerance is considered, the lowest error is found at different points on each yield curve. These calculations cause the parameters to fluctuate. This complicates the estimation of the series and renders results useless. Thus, we use alternative approaches to calculate the parameters addressing this problem. In these approaches, λ is assumed to be constant or to change slowly, and βs are determined by the linear regression method (OLS) over these λ values.

In approaches (ii), (iii), and (iv), we assume λ to be constant and unchanged throughout the series for each country. However, for the reasons mentioned earlier, λ is intended to be a time-varying parameter. Therefore, in approach (v), we apply a high-frequency HP filter to λ and take the "trend" data of the filter as λ. This approach is more in line with the economic and statistical foundations of the yield curve. The change in λ over time allows the yield curves to take different shapes and prevents excessive fluctuations in β that are inappropriate for the data structure.

Table 2: λ Values of The (ii), (iii), and (iv) Approaches.

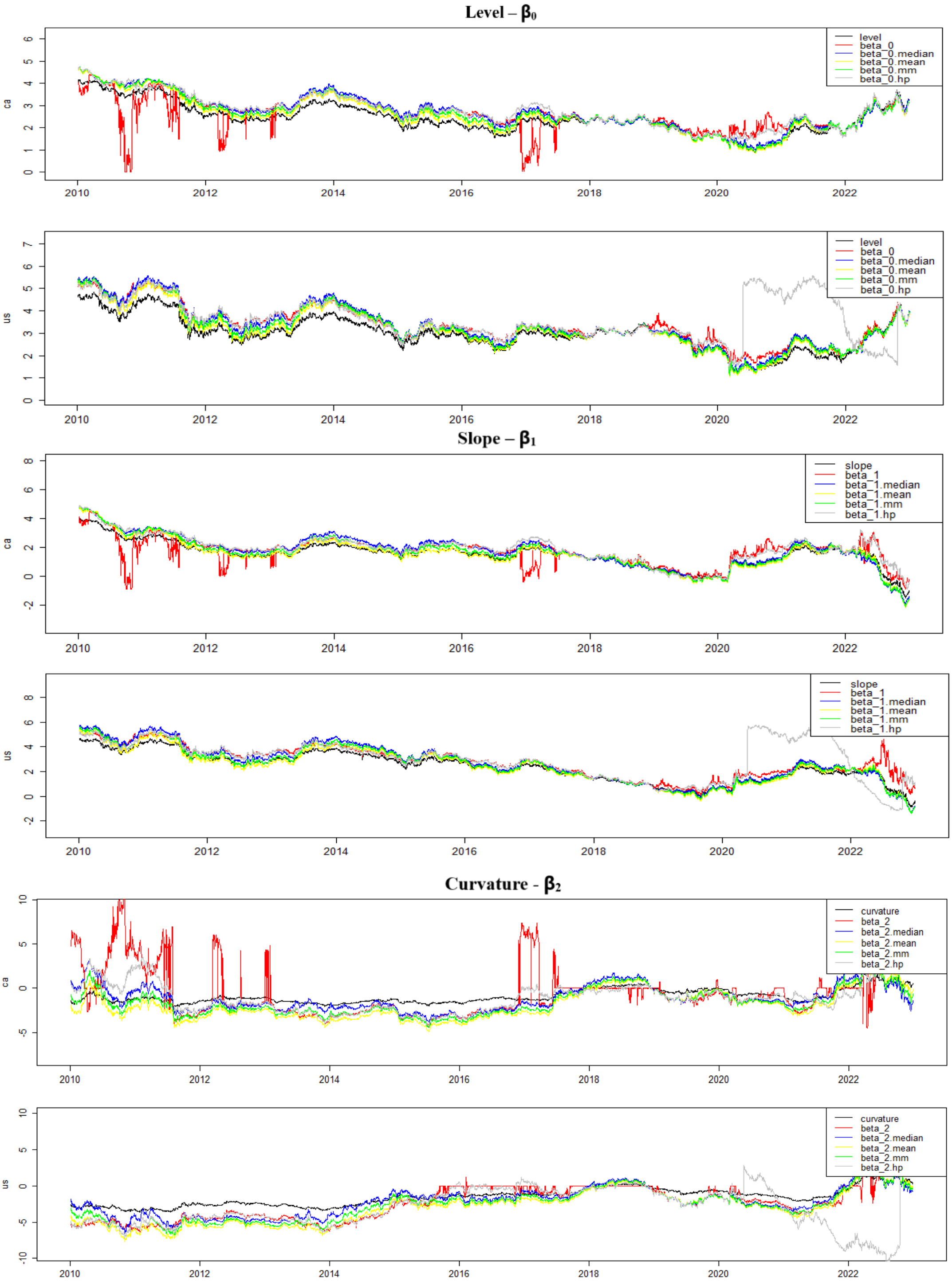
Country	Mean	Median	(Mean+Median)/2
CA	0.02975461	0.05250119	0.04112790
DE	0.03133113	0.04073675	0.03603394
FR	0.03000278	0.03403735	0.03202006
GB	0.03586918	0.04254933	0.03920925
IT	0.03335132	0.03752403	0.03543767
US	0.03795422	0.05471795	0.04633609

Figure 15: λ Plots of The (i) and (v) Approaches.



In Figure 15, the graphs show the factors obtained using the five approaches; the level, slope, and curvature series are expected to be consistent.

Figure 16: Level-β_0, Slope-β_1, Curvature-β_2 Plots of Canada and United States



In Figure 16, the series shown in red lines are the parameters obtained from the first approach and do not agree with the level, slope, and curvature. The averages and HP filter approaches significantly improve this agreement. If we compare below β_0 with the level (30-year rates), β_1 with the slope (the difference of {30 years–3 months}), and β_2 with the curvature ({2Y-(3 months + 30 years)}), the correlation serves as another proof of this in Table 3.

Table 3: Correlation of Factors with βs

LAMBDA	Level – β_0					
Country	CA	DE	FR	GB	IT	US
Float	0.6696592	0.8635520	0.7125214	0.9371354	0.9313586	0.9630079
Median	0.9571195	0.9919957	0.9876571	0.9937626	0.9886598	0.9702756
Mean	0.9875825	0.9961291	0.9900049	0.9956661	0.9913200	0.9856145
(Median+Mean)/2	0.9777326	0.9945138	0.9889389	0.9948824	0.9901322	0.9797084
HP Filter	0.9399208	0.9919038	0.9719311	0.9885996	0.9822358	0.2806470
LAMBDA	Slope – β_1					
Country	CA	DE	FR	GB	IT	US
Float	0.7952091	0.7587416	0.4470479	0.9192030	0.8346087	0.9675420
Median	0.9827336	0.9892613	0.9803132	0.9963625	0.9624464	0.9962506
Mean	0.9961735	0.9932542	0.9850866	0.9945922	0.9737494	0.9911767
(Median+Mean)/2	0.9955467	0.9924974	0.9829927	0.9957891	0.9689387	0.9951961
HP Filter	0.9567684	0.9709881	0.9404475	0.9850681	0.9562312	0.7163594
LAMBDA	Curvature - β_2					
Country	CA	DE	FR	GB	IT	US
Float	0.2921776	0.3441877	0.5018172	0.7935718	0.4332728	0.9253590
Median	0.7766510	0.8185985	0.5479502	0.7864305	0.8678648	0.9140500
Mean	0.9682878	0.9206214	0.6742652	0.8721786	0.8934716	0.9828278
(Median+Mean)/2	0.9135332	0.8805088	0.6153583	0.8353373	0.8813163	0.9610990
HP Filter	0.6590978	0.8056391	0.8401863	0.9524283	0.6870195	0.3127094

The following table (See Table 4) shows the error (RMSE) measures of how well the NS model fits some yields according to different λ approximations.

Table 4: RMSE of Some Forecasted Yields with βs According to All λ Approximations

	Maturity	Float	Median	Mean	(Median+ Mean)/2	HP Filter		Maturity	Float	Median	Mean	(Median+ Mean)/2	HP Filter
CA	M3	0,050	0,109	0,107	0,103	0,050	DE	M3	0,070	0,117	0,093	0,103	0,068
	Y1	0,065	0,128	0,107	0,113	0,066		Y1	0,066	0,098	0,082	0,088	0,064
	Y2	0,041	0,093	0,074	0,075	0,043		Y2	0,046	0,092	0,069	0,076	0,050
	Y5	0,046	0,075	0,053	0,060	0,049		Y5	0,040	0,041	0,040	0,039	0,041
	Y10	0,048	0,080	0,083	0,073	0,069		Y10	0,057	0,043	0,068	0,052	0,061
	Y30	0,051	0,071	0,069	0,058	0,070		Y30	0,084	0,110	0,124	0,114	0,110
	Mean	0,050	0,082	0,071	0,070	0,058		Mean	0,050	0,068	0,064	0,063	0,058
FR	M3	0,096	0,121	0,120	0,120	0,091	GB	M3	0,082	0,119	0,114	0,115	0,086
	Y1	0,074	0,090	0,084	0,087	0,087		Y1	0,080	0,093	0,088	0,089	0,086
	Y2	0,046	0,095	0,094	0,093	0,065		Y2	0,058	0,121	0,125	0,121	0,069
	Y5	0,064	0,064	0,069	0,066	0,057		Y5	0,060	0,059	0,059	0,059	0,066
	Y10	0,105	0,152	0,155	0,153	0,138		Y10	0,057	0,083	0,084	0,082	0,064
	Y30	0,090	0,160	0,152	0,156	0,146		Y30	0,066	0,088	0,094	0,090	0,076
	Mean	0,080	0,105	0,104	0,104	0,096		Mean	0,058	0,079	0,078	0,077	0,067
IT	M3	0,146	0,187	0,189	0,188	0,163	US	M3	0,039	0,109	0,104	0,102	0,421
	Y1	0,108	0,127	0,126	0,126	0,119		Y1	0,039	0,105	0,084	0,089	0,699
	Y2	0,106	0,140	0,142	0,140	0,123		Y2	0,029	0,100	0,088	0,088	0,610
	Y5	0,078	0,074	0,073	0,074	0,077		Y5	0,037	0,097	0,056	0,071	0,627
	Y10	0,085	0,101	0,106	0,103	0,097		Y10	0,059	0,098	0,089	0,090	0,782
	Y30	0,072	0,097	0,094	0,095	0,092		Y30	0,029	0,130	0,070	0,090	0,938
	Mean	0,089	0,101	0,101	0,101	0,097		Mean	0,044	0,093	0,080	0,082	0,732

As observed in the correlation and RMSE tables (See Table 3 and Table 4), taking λ as a constant (median, mean, or the average of these two) makes the βs in the NS model fit the yield curve with more error but transforms them into a better forecastable series with respect to the first approach. However, the HP Filter approach ensures that the fitting error is small, as in the case of the float λ, and produces better forecastable series, as in the case of constant parameters. On the other hand, these results cannot be obtained from the United States data. For the United States, the HP Filter causes β to be estimated with high deviations from the level, slope, and curvature after 2020. Prior to this period, it is observed that the parameters are stable without the HP filter in the first approach. For our forecasts, we use the HP filter approach for all countries except the United States and the original float λ approach for the United States.

3.3. DATA FRAMEWORK

In artificial neural networks (ANNs), enhancement is achieved using big data, making high-frequency data to be the preferred choice. However, the feedback effect of volatility on the relationship between volatility and yields diminishes its effectiveness when high-frequency data is used (Dufour et al. 2012). High-frequency yield data may introduce statistical problems such as time-varying volatility, conditional variance, fat tails, or leptokurtic unconditional distributions (Diebold & Rudebusch, 2013). In this thesis, daily data for the countries is considered as high-frequency data where the statistics and graphs demonstrate the presence of these issues, and approaches such as ANN are anticipated to help address these problems.

3.3.1. Data Description and Statistics

Descriptive statistics on data provide preliminary information on the characteristics of the series being studied and the selection of appropriate models. Statistical tests and regression analyses are considered valid when the stationarity condition is met. In time series, stationarity is tested using unit root tests. One of the most widely used methods for testing stationarity is the Augmented Dickey-Fuller (ADF) test. Table 5 and Table A.2 present statistical information on the NS factors, level-slope-curvature components, some maturities of the countries, and the ACF and PACF results of their lagged values. Finally, Table 5 presents these series' ADF results and stationarity tests of Canada. The other countries’ statistics are in Table A.2 in Appendix 9. The fact that the ACF declines over time at lagged values, whereas the PACF decays rapidly, indicates that the series in the statistics contain an autoregressive structure. Germany (all maturities), France (maturities up to fifteen years), the United Kingdom (maturities up to eight years), Italy (maturities up to six years), and the United States (maturities up to three months only) have negative yields on the same days. According to the NS parameters, yield curves with positive and negative slopes are observed for each country. Moreover, the positive and negative curvatures indicate that the yield curves can take any shape. According to the ADF tests, some parameters are not stationary.

Table 5: Descriptive Statistics of Canada Data

CA	MEAN	SD	MİN	MAX	ACF(1)	ACF(10)	ACF(21)	ACF(252)	PACF(1)	PACF(10)	ADF
M3	0,910	0,701	0,028	4,300	0,996	0,956	0,903	-0,064	0,996	-0,006	0,99
Y1	1,131	0,782	0,120	4,600	0,997	0,963	0,920	-0,022	0,997	-0,004	0,99
Y2	1,225	0,748	0,151	4,291	0,997	0,969	0,934	0,043	0,997	0,005	0,99
Y5	1,548	0,723	0,308	3,856	0,997	0,969	0,938	0,179	0,997	0,009	0,96
Y10	1,979	0,737	0,435	3,741	0,997	0,971	0,943	0,304	0,997	0,007	0,92
Y30	2,428	0,695	0,888	4,152	0,998	0,975	0,952	0,458	0,998	0,012	0,93
β_0	2,752	0,720	1,251	4,736	0,997	0,973	0,946	0,551	0,997	-0,006	0,57
β_1	-1,908	0,961	-4,90	0,876	0,996	0,959	0,914	0,392	0,996	-0,014	0,32
β_2	-1,008	1,784	-4,14	5,333	0,996	0,961	0,916	0,258	0,996	0,000	0,90
λ	0,053	0,054	0,010	0,229	0,999	0,987	0,969	0,047	0,999	-0,016	0,01
Level	2,428	0,695	0,888	4,152	0,998	0,975	0,952	0,458	0,998	0,012	0,93
Slope	1,518	0,864	-1,49	3,972	0,997	0,963	0,919	0,327	0,997	-0,012	0,63
Curvature	-0,888	0,778	-2,09	1,689	0,997	0,975	0,949	0,328	0,997	0,002	0,51

3.3.2. Data Selection

Extending the time series to be used as input in the past provides more information from the series and statistical improvement of the models. (Ganguli & Dunnmon, 2017). In a time series, keeping the starting point constant and repeatedly making the forecast with each new data entry effectively processes the information provided by new data. However, repeating the forecasts by shifting the series starting point at each new data entry is preferable when comparing results. The optimal amount of data to be used in the models is also related to the location and size of the data break. If this break coincides with the end of the data period, the search for the optimal amount of data does not yield reliable results (Pesaran & Pick, 2011). In this study, we determine the inputs and the number of inputs used in the models according to the optimization results at specific intervals.

3.3.3. Data Pre-Processing

Data preprocessing reduces noise, learns patterns, and identifies relationships by analyzing and transforming data (Kaastra & Boyd, 1996). Transformations stabilize the variance and adapt the data to a normal distribution. Logarithmic transformation, normalization, and scaling are applied to enhance the model's effectiveness (Lewis, 2017b). In artificial neural networks, preprocessing of input data is essential

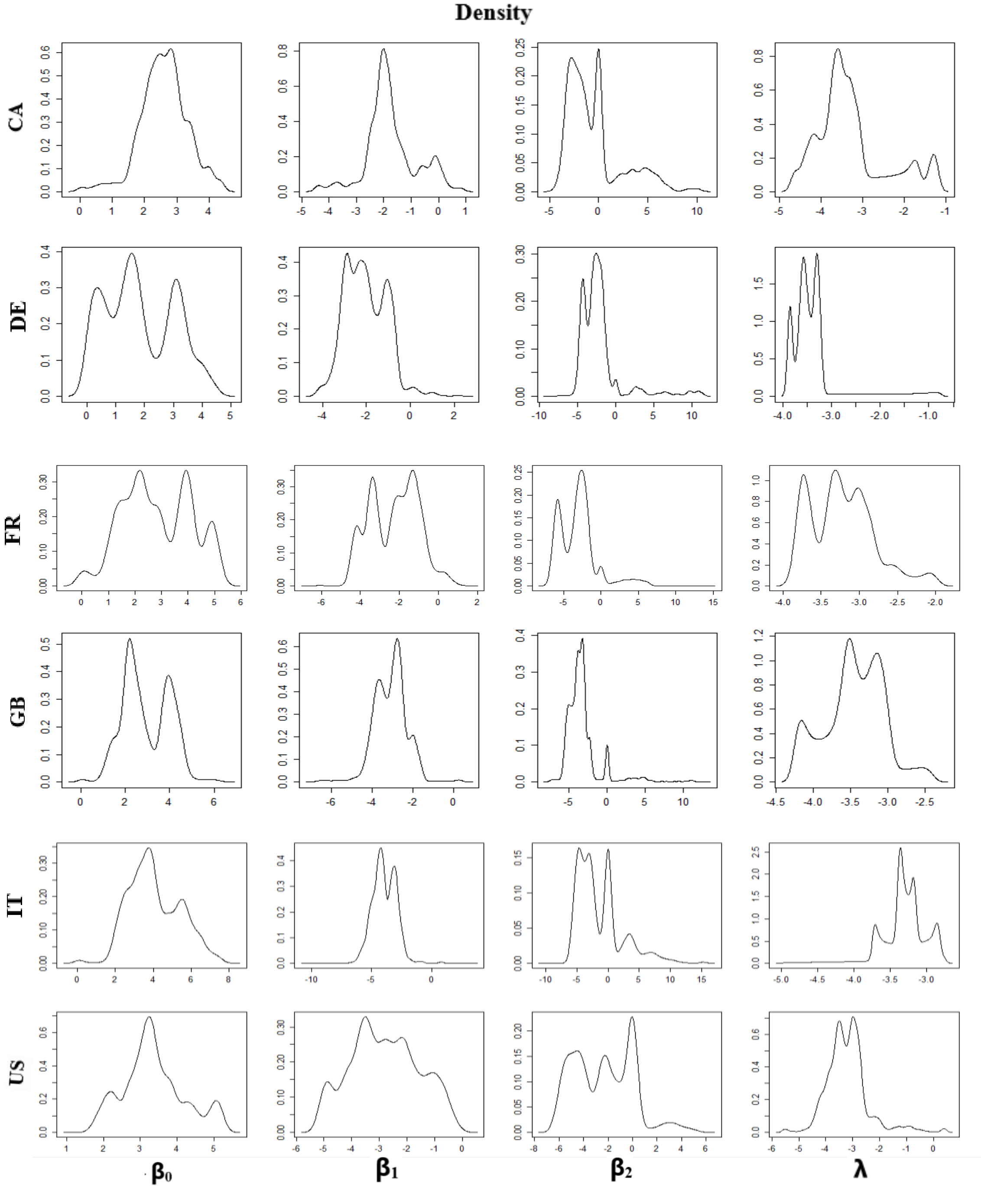
(Gajowniczek & Zabkowski, 2017). Scaling protects network training by preventing neurons from receiving excessively large or small inputs (Öztemel, 2016). Logarithmic transformation is an effective normalization method for positive time series (Lewis, 2015), and is preferred in cases where variables exhibit right-skewed distributions. Transforming variables into proportional data reduces the number of inputs without affecting degrees of freedom (Kaastra & Boyd, 1996).

Normalization standardizes the distribution. The most commonly used normal distribution is the Gaussian distribution, in which data standardized as $\mathcal{N}(0,1)$ performs better. This transformation is done with $z_i = \frac{x_i - \bar{x}}{\sigma_x}$ (Lewis, 2017a). For faster and more efficient processing of ANNs, a minimum-maximum normalization transformation can be preferred, where the relationship between variables is preserved, but the size of the variables is reduced (Li et al., 2001). For scaling bipolar data processing ($z_i = \frac{2x_i - x_{max} - x_{min}}{x_{max} - x_{min}}$) or unipolar data processing ($z_i = \frac{x_i - x_{min}}{x_{max} - x_{min}}$) methods can be used. Data are converted to the [-1,1] scale with bipolar data processing and the [0,1] scale with unipolar data processing. Klimasauskus (1991) proposes bipolar data processing for prediction.

Observations visibly separated from other observations are called outliers. Unwanted discrepancies in observations are referred to as noise. Noise can be caused by attributes that are not considered or by a recording error called teacher noise (Alpaydın, 2010). In this study, some of this type of noise is eliminated by smoothing the yield data. In addition, the choice of λ (HP Filter approach) eliminates most of the noise and outliers in the latent factors. Modeling the anomalies in the data reveals the effectiveness of data preprocessing and forecasting methods.

For the peak curvature to be in the maturity range of the data in the yield curve, λ should be estimated in the range [0.001:1]. Therefore, in this study, we perform, in individual models, $log\left(\frac{\lambda - 0{,}001}{1-\lambda}\right)$ transformation. We convert the forecasts of the new values back to values in the range [0.001:1] by applying reverse transformation. Figure 17 below shows the distribution of factors in the countries' yield curves. The λ graphs show the logarithmic transformation values.

Figure 17: Density Graphs of NS Factors



According to the graphs presented in Figure 17, the data generally does not conform to a normal distribution and cannot be identified to contain outlier values. We apply the Shapiro-Wilk normality test to the data. In all the test results regarding the NS factors of the countries, the result of "p-value < 2.2e-16" is obtained implying that none of them fit the normal distribution.

Based on the optimization results, we select normal, scaled, or logarithmically transformed data for most ANN models. We does not apply standard distribution transformation, as it do not improve the forecasts. We perform scaling using unipolar data processing and apply only this transformation to ANN models that are not optimized using these three methods.

3.3.4. Data Splits

In the triple separation approach, observations are divided into training, validation, and testing, and model selection is performed according to the performance of the validation training. The performance of the model is tested where training and testing data partitioning are performed against the possibility of finding spurious correlations or patterns (Lewis, 2017b).

The training set is used to optimize the weights of the ANN models. This process is called the goodness of fit. The larger the training set, the better the network performance (Zhang et al., 1998; Nam and Schafer, 1995). Using training data with large sizes and percentages and a delayed input can increase the predictive ability of models (Tealab et al., 2017). As data exploration and the number of tests increases, the model increases the probability of finding an optimal solution, thereby increasing the danger that the optimal solution will exhibit random performance.

In separation approaches, such as k-partition cross-validation or leave-one-out methods, each partition (the part taken for training) obtain different sets of weights and bias values, so it can be a problem to determine which set is more appropriate to use.

In this study, using the sliding window approach, we conduct a training study for a certain period, and select the parameters that yield the best results among the model groups and use in these models. We take the NS factor series in a specific window, forecast repeatedly on the same series as the sliding window, and compare errors with the test data to determine which model gave better results. This method also are used in forecasting studies using ensemble-learning models. In addition, we employ forecast combination (FC) using the models that provide the best results.

In this study, each series to be forecasted contains 3389 data points, the first 1000 of which we reserve as training data. We reserve the data between 1001 and 1070, the last 40 of which are the forecast horizons, for daily forecasts for use in optimization. We repeat this process every 200 data points. We make forecasts for the following 200 data points using we obtain the parameters in the optimization.

3.4. MODELLING FRAMEWORK

There are three main uncertainties in economic models. These are the theory uncertainty that forms the basis of the economy, which arise with debates about which factors will explain an economic structure or indicator; specification uncertainty about how to model economic theories; and heterogeneity uncertainty; which refers to whether a model fits for different data (Brock et al., 2003).

Different methods can be followed for forecasting according to the aim of the study. That is one-step forecasting without re-forecasting, multi-step forecasting without re-forecasting, and multi-step forecasting with re-forecasting. Using same input, in the first one, every data is forecasted separately; in the second one, a data series is forecasted at the same time; and in the last one, the data forecast is performed step-by-step using the forecasted data as input again in the forecast (Namın & Namın, 2018). In iterative multistep forecasting, the forecast error at short horizons increases when feeding on itself at longer horizons. Although direct forecasting does not accumulate errors, it is deprived of information from observations lagged by the previous forecast horizon. This increases the forecast error (Taieb et al., 2012). Past data and future values are directly related to the multi-input, multi-output approach, and forecasts are made simultaneously, which is the preferred single-output approach (Sagheer et al., 2021). Short forecast horizons are more accurate because they use more recent information (Lewis, 2017a).

It is supposed that yield curves are indicators that include economic, financial, political, and other types of information. In this case, it is not expected to be adequate to use information other than the information revealed by the curve as input in the forecasting models. For example, according to the expectations hypothesis, yield curve forecasting is based on expectations of short-term returns, and the risk premium is not used as a variable in forecasting (Povala & Vasıl, 2017).

As long-term economic contracts can be made based on future medium-term contract prices, cointegration emerges between economic medium and long maturities. Because yield curves are also an indicator of economic dynamics, cointegration may occur between maturities (Mineo et al., 2020). This encourages the forecasting of yield curves with fewer parameters instead of yields.

Once the data structure is examined, it is observed that regularity did not occur in almost any data; therefore, optimization in parameter selection is performed at regular intervals, not once. Our aim is to improve the forecasts by recalibrating a suitable model in one period if the goodness of the forecast deteriorates over time despite the change in the data structure. We determine the intervals at which the optimizations should be performed using the AR model. First, we perform tests on the AR(1) model using different input numbers. We repeat these tests for 100, 200, and 300 data points, and we determine the number of data points used in each data interval. With this amount of data, we forecast 40 days for each piece of data. According to the criterion we use in the forecasts, find the optimization frequency with the smallest error and apply to all models.

We use the average of the RMSE values of the validation set of all models (same model with different parameter sets) and their forecasts errors at a given forecast horizon to select parameters for a short period at fixed intervals. We find the values of the last forecasts of the few models that give the best RMSE values and their yields at standard maturities, and we calculate the area between the forecasted yield curve and the original yield curve. We use the parameters with the smallest area in all latent factors over a specific time interval, assuming that they make the best forecast.

3.5. PARAMETERS SELECTIONS

Sample selection, how inputs and outputs are used in the network, the numerical representation method of inputs and outputs, the choice of initial values, learning and momentum coefficients, the period of revising the weights, the scale of inputs and outputs, the stopping criterion, and the topology of the network are practical in the performance of the network (Öztemel, 2016).

The ML algorithm is susceptible to parameter choices. This sensitivity makes it difficult to make practical adjustments in the parameter selection (Hoogteijling, 2020). ANN researchers and practitioners try to set theoretical or experience-based rules for selecting parameters, such as several neurons, layers, and initial weights. However, these rules are subjective, and their effectiveness is debated by comparing them with approaches such as random selection, grid search, and stepwise increase or decrease. However, these rules are subjective, and their effectiveness is debated by comparing them with approaches such as random selection, grid search, and stepwise increase or decrease.

In deep learning, if the number of neurons in the layer increases, the risk of overfitting increases, as the modeling can increase the learning of patterns and noise, not statistical connections. Therefore, out-of-sample forecast performance is decreased. The number of neurons should be increased as the number of patterns increases without reducing the generalization capability. The number of layers should be chosen to learn all the data patterns that will benefit the model's performance. As the data features become more complex, more hidden layers are used to utilize each level of data features in training. An ANN assumes that it has efficient capacity utilization, where each neuron learns a different feature of the inputs. The number of neurons is chosen accordingly (Lewis, 2016). Networks with fewer neurons in more than one hidden layer may have an advantage over multi-neuron networks with a single hidden layer. This is because modeling with fewer parameters can generalize better.

The hidden layer units determine the analyzed data's characteristics, capture the data's systematics, and provide a curvilinear mapping between the input and output. Cybenko (1989) and Hornik et al. (1989) suggest a single hidden layer for complex nonlinear data.

In general, one to four hidden layers are recommended (Kaastra & Boyd, 1996). Lippman (1987), Cybenko (1988), and Lapades and Faber (1988) show that using three or more hidden layers does not improve the forecast results. Trial and error are the preferred methods for determining the number of units (neurons) of hidden layers. It is recommended that this number be at least 10. Some determine this number using specific methods.[20] Tang and Fishwick (1993) claim that the number of hidden units does not affect the forecast performance (Günay et al., 2007). Any proposed number standard is based on the researchers' experience. The best method is trial and error. However, the training set must be considered when selecting the number of neurons. The number of neurons determines the number of weights of the network, and if there is insufficient training set data, the problem of overfitting arises. Klimasauskas (1993) states that the number of hidden layer neurons should be enough to have at least five times as much training data as the number of weights.

Tang and Fishwick (1993) relate the number of input units to the number of autoregressive (AR) terms. In contrast, Zhang et al. (1998) disagree because the MA model does not include AR terms. While Sharda and Potil (1992) and Tang and Fishwick (1993) find 12 input units for monthly data and 4 for quarterly data to be intuitively appropriate, Eğrioğlu et al. (2008) show that setting the number of input units of seasonal data equal to its period does not give good results. Lachtermacher and Fuller (1995) state that taking too many inputs is not good for a forecast horizon, which is 1, but it is suitable for a forecast horizon, which is more than one. Zhang et al. (1998) claim that parameters should be determined with curvilinear regression studies.

To increase the performance, the network topology can be experimented step-by-step by narrowing from large to small or expanding from small to large (Öztemel, 2016). Goodfellow et al. (2016) state that the grid search method is used in the hyperparameter selection of an ANN if the number of parameters to be determined is three or fewer.

[20] If number of inputs is n, for number of hidden units Lippman (1987) and Hecth (1990) chose 2n+1; Wong (1991) chose 2n; Tang and Fishwick (1993), Kong (1991) chose n/2 (Günay et al., 2007). Baily and Thompson (1990) suggested n.3/4; Katz (1992) suggested from n.3/2 to 3n; Ersoy (1990) suggested up to 2n. Another criterion for choosing the number of neurons can be (number of inputs + outputs)/2 or its square root (Lewis, 2015). In selecting the number of neurons in single hidden layer neural networks, $\sqrt{(n \times m)}$ (n: input, m: output neurons) is one of the approaches used (named pyramid rule). Moreover, this number can vary between half and 2 times this formula.

Eklind (2020) states that although random selection facilitates the operation of the ANN, he could not comment on the difference with grid search. Reviewing previous studies, Bergstra and Bengio (2012) state that searching hyperparameters randomly yields the same results as a grid search, but is more efficient regarding computational cost. If a parameter selected in a random search fails, the failure can be eliminated immediately by replacing it. A grid search is reliable for parameters of low dimensionality. In a grid search, there is the problem of testing parameters that change is less critical than those whose change is necessary. Larochelle et al. (2007) train ANNs with random search in less time than grid search and obtain better results on most data. The fact that an ANN depends on hardware capacity causes a significant disadvantage in the trial-and-error approach.

Another strategy to avoid getting stuck in local optima in ANNs is to choose random weights for the initialization. Because choosing different initial weights changes the results, it is preferable to run the ANNs many times with these different weights and average the results. However, this method does not guarantee optimal results (Szenczi, 2016). It is recommended to randomly assign initial weights within a specific range (usually [-1,0.1] is preferred). It is observed that if this range is wide, the network continuously navigates at the local optima. In contrast, if it is narrow, it causes an increase in the number of iterations required to find the solution. If the learning rate (η), which determines the step of change of the weights, is large, the network oscillates around the local optima on the error surface. If it is small, the learning time increases. This ratio is mainly preferred in the range [0.2,0.4] (Öztemel, 2016).

Overfitting can be prevented by using more data, stopping training after reaching the minimum test error or dropout, or adding a penalty term for an increased fit (Gürsakal, 2017). Masters (1993) states that it is more appropriate to try to expand the training data first instead of stopping the training in case of a performance decrease in training because the approach of stopping the training when the error starts to increase during the training of ANNs focuses on the result instead of the cause, and the alternative cost is much higher. A new and expanded test set should also be created.

The training is stopped if no further improvement can be made in error owing to training with random weights in the ANN training. This is known as the convergence approach. Another approach that can be used to stop training is training-test interruption, where the network is tested by stopping training at specific iterations. If there is not enough improvement, training is resumed. However, with this approach, it is possible that the result may not be better if training is continued. However, the convergence approach does not accept the overtraining criterion. The aim is to achieve a global minimum. The training-test interruption approach, on the other hand, avoids overfitting by controlling the training by interrupting it. However, this approach has difficulty in choosing the number of iterations and finding the optimal training test interruption (Kaastra & Boyd, 1996).

According to the literature, there is no difference in the efficiency between grid search and random selection in parameter selection. In this study, we periodically perform parameter selections in many models. Random selection requires considerable attention and effort because it must be performed manually. In this study, it is necessary to apply the random selection approach many times. For this reason, we use a systematic grid search method to calculate the errors of the predictions based on a certain number of selected parameters. We use the parameter group that provides the best results according to the error criterion in the forecasting models until the following optimization. We do not use the early stopping method because it does not sufficiently benefit ANNs. The experiments we conduct with the data show that early stopping does not significantly affect the model training. Although the selected parameters cannot find a good error value in the training, we do not apply this approach because we try other alternatives, as recommended by Masters (1993).

CHAPTER 4:
EMPIRICAL ANALYSES

This section presents a comparison of the forecasting model results. The section begins by elaborating on the comparative framework and detailing the methodological approach, statistical parameters, and analytical procedures for the evaluation. We utilize a total of 62 distinct models. A few of them are identical and differ only in their epoch or iteration counts. Owing to the large number of models, there are many results when comparing NS factors and yield curves, as well as many forecast graphs. Given the constraints of this study, a comprehensive display of all the models is not feasible. Instead, the analysis proceeds in two phases. The first phase involves individually examining approaches, such as evaluating all FFNNs. The second phase entails a comparative assessment of the most influential models identified using the different approaches. The RW model serves as a benchmark in various assessment procedures.

4.1. EVALUATION METHODS

RMSE is considered suitable for evaluating the fit of the NS model (Szenczi, 2016). This metric, commonly employed in yield curve forecasting models, exhibits sensitivity to data levels (Reinicke, 2019). As an absolute measure of errors and a second-order loss function, the RMSE requires model errors to conform to the Gaussian distribution to effectively compare the model performance. Furthermore, when comparing models, it is crucial to assess how the distribution changes based on the forecast horizon (Szenczi, 2016). The RMSE is not affected by the number of forecasts, making it suitable for model comparison (Hyndman and Koehler, 2006). In this study, we compute along with RMSE, other error measures such as MAE, MASE, and SMAPE for the forecast outcomes.

$$RMSE \text{ (Root Mean Squared Error)} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(x_t - \hat{x}_t)^2} \quad (4.1)$$

$$MAE \text{ (Mean Absolute Error)} = \frac{1}{ntest}\sum_{t=1}^{ntest}|x_t - \hat{x}_t| \quad (4.2)$$

$$MAPE \text{ (Mean absolute percentage error)} = \frac{100}{ntest}\sum_{t=1}^{ntest}\left|\frac{x_t - \hat{x}_t}{x_t}\right| \quad (4.3)$$

$$SMAPE \text{ (Symmetric MAPE)} = 200 \times \frac{1}{ntest} \sum_{t=1}^{ntest} \frac{|x_t - \hat{x}_t|}{|x_t| + |\hat{x}_t|} \tag{4.4}$$

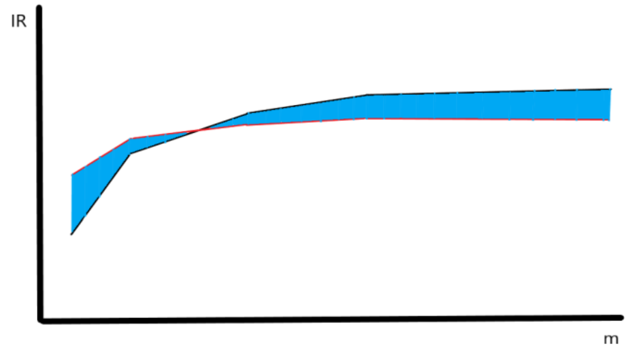
$$MASE \text{ (Mean Absolute Scaled Error)} = \frac{1}{ntest} \sum_{t=1}^{ntest} \left| \frac{x_t - \hat{x}_t}{\frac{1}{n-1} \sum_{i=m+1}^{n} |x_t - x_{t-m}|} \right| \tag{4.5}$$

MASE measures performance relative to the naive forecast. M is period value in seasonal series; in non-seasonal series, it is “1”. The MAPE measure is skewed to the right because of extreme outliers, which can yield misleading comparison results (Gajowniczek and Zabkowski, 2017). Consequently, we do not use the MAPE measure. We opt for SMAPE, a metric that assesses relative errors, to examine percentage discrepancies among time series with different absolute values (Grmanová et al., 2016).

Evaluating model effectiveness across all models relies on total error measurements. If we express $y_{t+h}(m_i)$ as data of the entire yield curve at horizon h and a yield curve as $x_t = y_{t+h}(m_i)$, using the RMSE criterion, we convert the entire model's error into a single value. Merging all maturities into a single series to compute an error is not exclusive to RMSE. We also use this consolidation approach for other assessment metrics.

We further evaluate the magnitude of the area between the forecasted and actual yield curves as an additional error metric. A smaller area indicates a closer alignment between the forecasted yield curves derived from the NS factor forecasts and actual curves. This measurement can be interpreted as an indicator of a model's forecasting effectiveness and overall predictive capability. The area calculation depicted in Figure 18 serves multiple purposes. It is used to identify the optimal optimization frequency, establish the parameter sets for model implementation, and evaluate the effectiveness of the forecasts.

Figure 18: Area Between The Actual and Forecasted Yield Curves



The computation of the area involves summing the basic geometric shapes found between the actual and forecasted yields at the standard maturities depicted in the graph. Appendix 8 contains the code used for area calculation.

Hypothesis tests are utilized to examine the data structure and establish statistical differences in forecast errors. Hypothesis rejection occurs when the probability associated with the calculated statistical value in the test is exceptionally low. In the case of VAR(1) models, a causality test is implemented; in the Granger causality test, H_0 defines that there is no causality and H_0 is rejected when $p<p_{critical}$. However, in the opposite case, it is not possible to definitively state "There is causality". These tests are inadequate for examining nonlinear associations. As demonstrated in Appendix 9, the Granger Causality Test results reveal that causality is not universally applicable among the factors utilized as inputs in VAR(1) models. This may account for the inefficiency of the forecasts generated by the VAR(1) model.

When implementing these statistical analyses, we apply the t-test, Wilcoxon, and Diebold-Mariano (1995) tests to examine the forecast errors associated with NS factors in yield curves. We undergo each model's individual comparisons with the RW using these analytical methods. The t-test tests whether the valid sample and population means are different. The test explores the impact of random variation by explicitly focusing on sampling errors. When the test statistic exceeds the critical value ($t>t_{critical}$), it indicates a statistically significant difference between the means that cannot be attributed to chance alone (Ghatak, 2017).

The Wilcoxon signed-rank test is a nonparametric test that checks whether the error of the best model is statistically lower than that of other models. This test is more appropriate when forecast errors are not normally distributed (Grmanová et al., 2016). The Wilcoxon test statistically tests the difference between the performances of the two models. Let A and B be two models.

$$\text{Wilcoxon test: } W^{+} = \sum_{i=1}^{M} I(u_i > 0) Rank(|u_i|) \quad (4.6)$$

$$u_i = SMAPE_B(i) - SMAPE_A(i) \quad (4.7)$$

M is the number of observations or forecasts, and I is the indicator (sign) function (Andrawis et al., 2011).

The Diebold-Mariano (DM) test statistic tests the hypothesis that there is a statistical difference between the forecast errors of the two models (Castellani and Santo, 2006). Diebold-Mariano test:

$$DM = \frac{\bar{d}}{\sqrt{\frac{2\pi \hat{f}_d(0)}{n}}} \quad (4.8)$$

$$d(t) = g(e_{1t}) - g(e_{2t}) \quad (4.9)$$

g(t) is a loss function (exponential, logarithmic, quadratic, etc.). e_{1t} and e_{2t} are forecast errors of the models. $\hat{f}_d(0) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} \gamma_d(k)$ is spectral density estimate at the zero point. γ is autocovariance and k is a lag step. The test statistic (DM) has a standard normal distribution; $DM \sim \mathcal{N}(0,1)$. If the p-value is large, no difference is concluded (Eğrioğlu & Baş, 2020). Forecast errors can be subject to asymmetric penalties. The DM test indicates that the covariance of the differences between the losses of the models must be stationary for the two models to be considered the same (Szenczi, 2016). The loss function does not necessarily need to be symmetric or quadratic. Correlations may exist among forecast errors, and their means may deviate from zero. Additionally, the error distribution may not adhere to a Gaussian distribution.

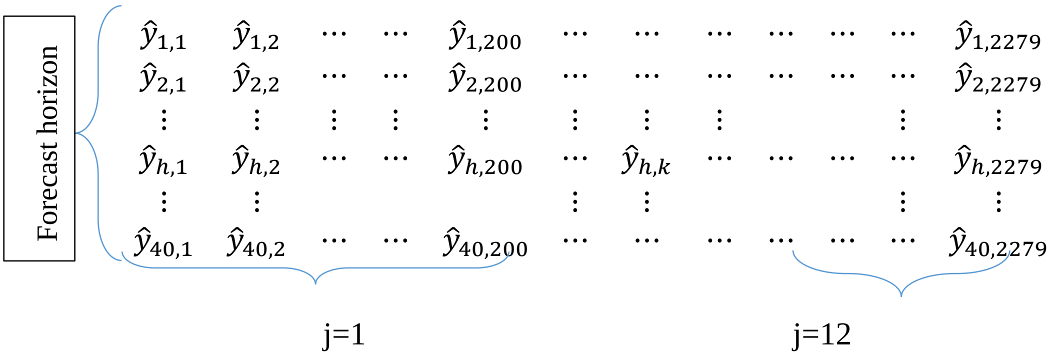
Algorithms that model time series should be able to predict patterns outside the data time interval and find a statistical confidence interval in predictions. Many ANN methods can only be effective in the time interval of the training data, and even if a confidence interval is desired, they cannot withstand a statistical distribution and will not be reliable.

In this study, we forecast NS factors as a univariate time series, except for the VAR(1) model (all used as input). We forecast using the sliding window method at a daily frequency of 2279 (2239 in FC versions 2 and 3). The forecast horizon is 40 days and there are 12 forecasting periods. Forecasts have three dimensions: day, forecast horizon and yields of standard maturities in forecasted yield curves. We calculate three types of errors in this study. First, we obtain individual RMSE error values using the NS factor forecasts. In the calculation made with data in the forecast horizon, which is the interval from 1 to i (i=1,…,40), for each day, we create RMSE matrices of size 40 × 2279 (40 × 2239 in FC versions 2 and 3).

Second, we calculate the area error of the yield curves obtained from the factor forecasts for each day and forecast horizon. For the calculation, we create an area matrix of size 40 × 2279 (40 × 2239 in FC versions 2 and 3). Third, we calculate the absolute errors and RMSE values of the standard maturity yields obtained from the factor forecasts for the entire forecast horizon, day, or all for each forecast horizon, and forecasting period. We also calculate the MAE, SMAPE, and MASE for all these yields for each forecast horizon. In the calculation, we create an error matrix of size 1x2279 (1x2239 in FC versions 2 and 3) for the daily dimension, 40x1 for all of them for each forecast horizon, and 40x12 for each forecasting period and each forecast horizon.

Figure 19 represents any model's factor forecasts or yield curve ($\hat{y}_{i,h}$ is series of yield curve forecasts of the ith day in the hth horizon).

Figure 19: Forecasted Yields Matrix



For factors, we calculate the RMSE value for the forecast horizon and forecasting period (of 40x12 size) and daily (of 2279 or 2239 size). For yield curves, we calculate the absolute error for the forecast horizon and forecasting period (of 40x12 size) and only daily (2279 or 2239 size). Besides, we calculate the error measures only according to the forecast horizon (of 40x1 size). We compare and test these matrices and series obtained from the models with those of the RW model. Thus, we investigate the effectiveness of the models in terms of the forecast horizon, forecasting period, etc. In comparisons made with the RW model, we take those whose error is lower than RW as "1" and those whose error is higher as "0". Thus, we find that the number of days, yield curves, forecast horizons, and forecasting periods result in less error than RW. We also test the hypotheses regarding whether the models provide better results than RW with 95% confidence intervals. We design the following hypotheses:

H_0: $RW^{AREA} \leq MODEL^{AREA}$.

H_1: $RW^{AREA} > MODEL^{AREA}$.

The hypothesis is tested using t-test, Wilcoxon, and DM tests.

H_0: $RW^{RMSE} \leq MODEL^{RMSE}$.

H_1: $RW^{RMSE} > MODEL^{RMSE}$.

The hypothesis is tested using t-tests and Wilcoxon tests.

H_0: $RW^{MAE} \leq MODEL^{MAE}$.

H_1: $RW^{MAE} > MODEL^{MAE}$.

We test this hypothesis using t-test, Wilcoxon, and DM tests. In these tests, if the error is lower than that of RW, we write "1" (i.e., acceptance of H_0) and "0" vice versa, where we calculate the number of better forecasts.

4.2. PARAMETER OPTIMIZATION

Structural changes in time-series data can increase the model fit and forecast errors if the model parameters remain static. ARIMA techniques model time series by analyzing temporal variations. Thus, utilizing the AR(1) model is the most basic method for identifying intervals in which parameter adjustments should be implemented.

Forecasts are generally anticipated to be more accurate for shorter horizons. However, it is undesirable for models to exhibit substantial increases in error as the forecast horizon increases. Specific parameter configurations may sometimes yield favorable outcomes for near-term forecasts but show significant error growth over longer horizons. Conversely, other parameter sets may produce consistent error levels across all the forecast periods. The equal-weighted error criterion is inappropriate for eliminating some parameter sets with reasonable early forecasts if they can have worse late forecasts. In contrast, the other parameter sets that start the forecast horizon by worse forecasting have reasonably forecasts. Therefore, we select parameters with relatively good initial forecasts that do not introduce an overestimation error as the forecast horizon lengthens. Rather than employing uniform weights for parameter selection, we implement a weighted sum that diminishes as the forecast horizon increases. The weight is set at 1 for a one-day forecast horizon and systematically decreases arithmetically to 0.25 for a 40-day forecast horizon. We apply this weighting scheme exclusively during parameter selection. When comparing model forecasts, we revert to using equal weights.

We followed the same practice in the test set to determine the models' optimization frequency and parameter sets. Each country has 3389 data. We reserve the first 1000 data for the first optimization. To determine the number of inputs, we create parameter sets with 200, 400, 600, 800, and 1000 alternative numbers of input data and other parameters, with the last data set being the 1000th data. We use a multistep method to estimate each NS factor with these parameter sets at the 40-day forecast horizons (between 1001 and 1040). We repeat these forecasts five more times for every six data points in the test data. We determine the RMSE values of the 40-day forecasts of each parameter set on each forecast day and average them. Depending on the number of parameter sets, we find the yield curves and their weighted-area error measures from the last day's forecasts only, using the first few parameter sets that provide the lowest average RMSE for each forecast. We use the parameter sets of the NS factors with the lowest weighted area error as the input to forecast the following data until optimization. We repeat these practices in subsequent optimizations. Some models are inherently inaccurate regarding parameter sets, and we exclude these sets from the alternatives. If the chosen optimal parameter set fails when we use it for forecasting the series, we use the next-best alternative parameter set that does not fail.

We find the AR(1) model forecasts of the NS factors for 100, 200, and 300 data points with inputs between 200 and 1000 to determine the optimization frequency. We obtain the yield curves from these univariate series forecasts using the NS formula. We find the weighted sum of the areas between the forecasted and actual yield curves over the forecast horizon. We forecast the model over the forecasting period using the number of inputs with the smallest total error. We then compare the area between the actual and forecasted curves to find the optimization frequency that gives the best results and obtain averaging tables (See Appendix 10). For each country, we calculate the average area for each forecast horizon. Although the results are very close, optimization at a frequency of 200 data points yields the least error at almost all horizons in Great Britain, short and medium horizons in Canada, medium and long horizons in Germany, and short horizons in France and the United States. Therefore, we optimize every 200 data points in this study. We repeat the optimization process for every 200 data points, starting from the 1000th data point. Using the parameters obtained in the optimization, we forecast the following 200 data points. The training, testing, and forecast phases of each series involved 12 iterations. Table 6 presents the designated models, inputs, and parameter sets to be implemented in this study.

Table 6: Considered Models, Inputs, and Parameter Sets

(a) Conventional Models

	MODEL	INPUT	INPUT PARAMETERS	OPTIMIZATION	OPTIMIZATION PARAMETERS (Alternatives)
CONVENTIONAL MODELS	Auto ARIMA (AA), AR(1), ARFIMA	Univariate TS	NS Parameters (Individually)	Number of inputs	200,400,600,800,1000
	ARIMA (p,d,q)	Univariate TS	NS Parameters (Individually)	Number of inputs	200,400,600,800,1000
				p,q	0,1,2
				D	0,1
	ETS-TBATS	Univariate TS	NS Parameters (Individually)	Number of inputs	200,400,600,800,1000
				Model	ETS, TBATS
				Lambda	"auto", "NULL"
	LLAR-LSTAR-SETAR	Univariate TS	NS Parameters (Individually)	Number of inputs	200,400,600,800,1000
				M (embedding dimension)	2,4,10
				eps (neighborhood size)	1,2,4,10,20
				delay (Time Delay)	0,1
	TAR	Univariate TS	NS Parameters (Individually)	Number of inputs	200,400,600,800,1000
				p1, p2 (AR order of the lower-upper regime)	1,2
	RW	Univariate TS	NS Parameters (Individually)		
	VAR(1)	Multivariate TS	NS Parameters of country	Number of inputs	200,400,600,800,1000

(b) Non-Conventional Models

	MODEL	INPUT	INPUT PARAMETERS	OPTIMIZATION	OPTIMIZATION PARAMETERS (Alternatives)
NON-CONVENTIONAL MODELS	kNN	Univariate TS	NS Parameters (Individually)	Number of inputs	200,400,600,800,1000
				Lag number of inputs	2,3,4,5,10,20,40
				a string	"median," "weighted," "mean."
	ELM-ELM-A	Univariate TS	NS Parameters (Individually)	Number of inputs	200,400,600,800,1000
			ELM:5, ELM-A: 40 (Number of networks to train)	Number of hidden nodes	10,20,50,80, ”NULL”
				scaled data	original, [0,1] scaled
	ENN-ENN-A /JNN-JNN-A / MLP-MLP-A	Lagged variables of univariate TS	NS Parameters (Individually)	Number of inputs	200,400,600,800,1000
				Lag number of inputs	2,4,10,20
		Maximum of iterations to learn	ENN-JNN-MLP:1000; ENN-A-JNN-A-MLP-A:5000	Number of hidden nodes and layers	first layer: 4,8,20; second layer: 0,4,8,20 (not for JNN, JNN-A)
				parameters for the learning function	.001,.005,.01,.05,.1
				scaled data	original, [0,1] scaled, normalization
	NNET-NNET-A	Lagged variables of univariate TS	NS Parameters (Individually)	Number of inputs	200,400,600,800,1000
				Lag number of inputs	2,4,10,20
		Maximum of iterations to learn	NNET:5000; NNET-A:20000	Number of hidden nodes and layers	4,8,20
				parameter for weight decay	0, 0.000001
	RNN-RNN-A /LSTM / GRU	Lagged variables of univariate TS	NS Parameters (Individually)	Number of inputs	200,400,600,800,1000
				Lag number of inputs	4,10,20
		Number of iterations	RNN:80; RNN-A:300; LSTM-GRU:50	Number of hidden nodes and layers	first layer: 4,8,20; second layer: 4,8,20 (not for LSTM, GRU)
				learning rate to be applied for weight iteration	.001,.005,.01,.025,.05

	NNETAR-NNETAR-A	Lagged variables of univariate TS	NS Parameters (Individually)	Number of inputs	200,400,600,800,1000
				Lag number of inputs	2,4,10,20
				Number of hidden nodes	4,8,20
		number of networks to fit with different random starting weights	NNETAR:20; NNETAR-A:50	parameter for weight decay	.001,.005,.01,.05,.1
				scaled data	original, [0,1] scaled, normalization
				Lambda	"auto", "NULL"
	GRNN-GRNNTSF	Lagged variables of univariate TS	NS Parameters (Individually)	Number of inputs	200,400,600,800,1000
				Lag number of inputs	2,4,10,20
				scaled data	original, [0,1] scaled, normalization
				sigma (smooth function scalar)	.01,.05,.1,.5,.1,2,3,4
	GMDH	Lagged variables of univariate TS	NS Parameters (Individually)	Number of inputs	200,400,600,800,1000
				Lag number of inputs	2,4,10,20
				scaled data	original, [0,1] scaled, normalization
				Number of hidden layers	1,2,3
	NNETTS	Univariate TS	NS Parameters (Individually)	Number of inputs	200,400,600,800,1000
				scaled data	original, [0,1] scaled, normalization
				Number of hidden nodes	4,8,20
				embedding dimension	2,4,10,20
				thDelay (Time Delay)	0,1,5
				forecasting steps	0,1,5

(c) Ensemble Learning Models

	MODEL	INPUT	INPUT PARAMETERS	OPTIMIZATION	OPTIMIZATION PARAMETERS (Alternatives)
ENSEMBLE LEARNING	BAGGING (forecast package).	Univariate TS	NS Parameters (Individually)	Number of inputs	200,400,600,800,1000
				Function	ETS- AUTO ARIMA
	RANDOM FOREST MODELS (randomForest packages)	Lagged variables of univariate TS	NS Parameters (Individually)	Number of inputs	200,400,600,800,1000
		Output and residuals of AR(1) model	Forecast trend by AR(p)	Number of trees	500,5000
		Trend and cycle of HP filter		P	1,2,3
	XGB (forecastxgb package)	Univariate TS	NS Parameters (Individually)	Number of inputs	200,400,600,800,1000

(d) Forecast Combination Models

	MODEL	INPUT	INPUT PARAMETERS	OPTIMIZATION	OPTIMIZATION PARAMETERS (Alternatives)	OUTPUT
FORECAST COMBINATION	FORECAST COMBINATION (forecastComb package) – Version 1.	Univariate TS	NS Parameters (Individually)	Optimization: Parameters are selected from individual models results (13 models' fitted and forecasts).	All models are recalculated using 1000-day data, but parameters are taken from individual model optimizations. The best-performing individual model alternatives in parentheses are used in forecast combining.	1) SA, 2) CLS, 3) EIG1, 4) EIG2, 5) EIG3, 6) InvW, 7) MED, 8) TA, 9) WA *
		Fitted values of individual models	Individual Models: 1) ARIMA (ARIMA, AUTO ARIMA, ARFIMA, AR(1)), 2) ETS (ETS, TBATS), 3) ELM (ELM, ELM-A), 4) ENN (ENN, ENN-A), 5) JNN (JNN, JNN-A), 6) GRNN, 7) MLP (MLP, MLP-A), 8) NNET, 9) NNETAR (NNETAR, NNETAR-A), 10) RNN (RNN, RNNA), 11) RW, 12) TAR, 13) VAR.			
	FORECAST COMBINATION (forecastComb package) – Version 2.	Univariate TS	NS Parameters (Individually)	Models are selected from individual models results	The best-performing individual model alternatives in parentheses are used in forecast combining.	1) BG**, 2) CLS, 3) InvW, 4) MED, 5) TA,

	Forecasts of individual models	1) ARIMA (ARIMA, AUTO ARIMA, ARFIMA, AR(1)), 2) ETS (ETS, TBATS), 3) ELM (ELM, ELM-A), 4) ENN (ENN, ENN-A), 5) JNN (JNN, JNN-A), 6) GRNN, 7) MLP (MLP, MLP-A), 8) NNET, 9) NNETAR (NNETAR, NNETAR-A), 10) RNN (RNN, RNNA), 11) RW, 12) TAR, 13) VAR.			6) WA, 7) SA
FORECAST COMBINATION (forecastComb package) – Version 3.	Univariate TS	NS Parameters (Individually)	Models are selected from individual models results	The best performing of the individual model alternatives in parentheses are used in forecast combining	1) BG, 2) CLS, 3) InvW, 4) MED, 5) SA, 6) TA, 7) WA
	Forecasts of individual models	Individual Models: 1) ARIMA (ARIMA, AUTO ARIMA, ARFIMA, AR(1)), 2) ETS (ETS, TBATS), 3) ELM (ELM, ELM-A), 4) ENN (ENN, ENN-A), 5) RW, 6) NNET OR NNETAR (NNETAR, NNETAR-A),			
FORECAST COMBINATION (forecast package) – Version 4.	Univariate TS	NS Parameters (Individually)	Number of inputs	200,400,600,800,1000	
	Models	AUTO.ARIMA, ETS, THETAM (Theta method model), NNETAR, STLM (Seasonal Decomposition of Time Series by Loess), TBATS.	Weights	"equal", "insample.errors", "cv.errors"	
BOOSTING (mboost package).	Univariate TS	NS Parameters (Individually)	Number of inputs	200,400,600,800,1000	
	Fitted values of individual models	AR(1), ETS, TBATS, RW, ELM, NNETAR	Model	glmboots, gamboost, mboost, blackboost	
RANDOM FOREST MODELS (randomForest packages)	Univariate TS	NS Parameters (Individually)	Number of inputs	200,400,600,800,1000	Two models: randomForest, cforest
	Fitted values of individual models for «cforest» and «ramdomForest»	AR(1), ETS, TBATS, RW, ELM, NNETAR	number of trees	500,5000 (for randomForest)	

*SA: Simple Avarage, CLS: Constrained Least Squares, EIG1: Standard Eigenvector, EIG2: Bias-Corrected Eigenvector, EIG3: Trimmed Eigenvector, InvW: Inverse Rank, MED: Median, OLS: Ordinary Least Squares, TA: Trimmed Mean, WA: Winsorized Mean

* BG: Bates/Granger (1969)

Note: "…-A" means alternative number of networks to train/epoch/iteration/… etc. in Model-A

4.3. RESULTS

In this study, we first employ *Conventional Models* because of the abovementioned advantages. Although Diebold-Li (2006) takes "p" as 1 in AR(p) model, we also consider other alternatives of Box-Jenkis approaches to forecasting models, taking the possibility that changing this parameter may improve the forecasts in some cases into account. With the "auto.Arima" function, the parameters "p,d,q" are determined freely each time, whereas with the ARIMA(p,d,q) model, we determine the parameters by optimizing. We also apply the ARFIMA model if the factors have a long-run structure.

Unlike most studies in the regarding literature, in this thesis, instead of the yields themselves, we accept the NS factors, which are shown to be related to macroeconomic parameters in the economics literature, as inputs in RW, and we calculate yields from the forecasts of these factors with the NS model.

Once we examine the error graphs and values[21], we observe that the Auto ARIMA model generally forecasts better than the other Box-Jenkins approaches. On the other hand, the ARIMA and ARFIMA models have high forecasting errors. Auto ARIMA produces much higher error forecasts on some days especially as the forecast horizon increases. We understand that these highly erroneous forecasts are due to the inability to make effective forecasts for one of the β factors in some optimizations. For example, the forecast errors of β_2 in Canada, β_0 in France, and β_1 in Great Britain are high on certain days. AR(1) has less error in βs; however, we generally forecast the yield curve more effectively with Auto ARIMA in terms of area and tests. VAR(1) has more errors and extreme results than the Box–Jenkins approach. The model exhibits extreme errors, particularly as the forecast horizon increases.

In general, the TAR model is more effective[22]. This is because of the extreme errors that other threshold models occasionally made in the β_0 forecast. The TAR model is more

[21] For the sake of brevity we do not include these results here/in this thesis. Data includes four latent factors for each six country. 24 series must be forecasted by every model. All series have error matrices have dimension 40 x 2279. In addition, yield curve have matris too. Also, graphs of these matrices and test result are produced. Showing these result in this thesis would be problematic.

[22] For the sake of brevity we do not include these results here/in this thesis.

stable and has fewer errors. In contrast, the other threshold models show an exponential increase in error as the forecast horizon lengthens, significantly deteriorating their efficiency.

Using the ETS-TBATS exponential smoothing method, we produce the best forecasts with no extreme errors compared with RW[23]. In conventional models, ETS-TBATS methods, which provide good whole yield curve forecasts, and RW can generally obtain better results than the Box-Jenkins approach, but vice versa for βs.

The non-conventional models employed are kNN, feed-forward (FFNN), and feed-back (RNN) models. The kNN model yields results with increasing errors as the forecast horizon increases[24]. In ANN models, because of the time-consuming nature of optimization, in some models, the parameters to be used, which are found through optimization, remain constant. We make alternative forecasts using the same parameters and more epochs/iterations. Thus, we observe whether forecasts improve as the number of cycles increases.

FFNN models, including EML, ELMA, GMDH, GRNN, MLP, MLP-A, NNET, and NNET-A, are compared[25]. ELM models produce stable and good (low error) forecasts over the entire series. The fact that the GMDH has a considerable number of extreme forecasting errors, significantly as the forecast horizon increases, and that the MLP model errors increase regionally (in some forecasting periods) across all forecast horizons, significantly reduces the overall forecasting efficiency of these models. NNET models and GRNN (except Italy) locally have error increases for some factors and yield curve. When the test results are compared with RW, the ELM and GRNN models are more efficient.

We compare FFNN models with time-series algorithms in terms of their functions[26]. NNETAR models are generally the models with the least error and have the least error for most days. They have better results in the tests, except for some parts of some factors

[23] For the sake of brevity we do not include these results here/in this thesis.
[24] For the sake of brevity we do not include these results here/in this thesis.
[25] For the sake of brevity we do not include these results here/in this thesis.
[26] For the sake of brevity we do not include these all results here/in this thesis.

in terms of the factor basis and area criterion. The NNETTS model occasionally produces extreme errors. In addition, this model has a relatively high error rate in all forecast horizons. The GRNNTSF model has relatively high forecast errors for all the forecast horizons. These errors increase as the forecast horizon increases. NNETAR models stand out in the tests against the RW model's forecast results and overall error measure.

When we analyze recurrent models[27], RNN, RNN-A, GRU, and LSTM models have more errors than ENN, ENN-A, JNN, and JNN-A regarding factors and yield curves in general and all forecast horizons. The JNN and JNN-A models exhibit high errors in some regions. We believe this is because the input parameters obtained due to optimization in that historical interval worsen the forecasts.

In Ensemble Learning (EL) models, we use NS factors as univariate inputs for Bagging and Extreme Gradient Boosting (XGB) models, whereas their lagged values are inputs in the RF model. We examine EL approaches that use individual models as inputs in the forecast combination approach. In the RF model with the AR(1) model, we formulate the fitted values and residual values found by using the AR(1) model from the NS factors as input and output and use the AR(1) forecasts as new inputs. In the RF model with the Hodrick Prescott (HP) filter[28], we formulate trend and cycle values found with the HP filter as input and output and use AR(1) forecasts of trend values as new inputs. The models in this group generally produce stable forecasts without extreme errors. Therefore, we can say that it generally yields better results than the individual models. Canada, Germany, France, and the United States's bagging models agree with this approach. Regarding the area error measure, Great Britain and Italy's bagging models outperform the RW model in the mid- and long-forecast horizons. Individually, βs have a weaker forecast performance.

Forecast combinations are created using five different approaches[29]. By "Version"s we mean forecast combinations made with different inputs. We give all versions we employed and their inputs, parameters, and models, in Table 6. In Version 1, the SA

---

[27] For the sake of brevity we do not include these results here/in this thesis.
[28] We use HP filter for factors in this model.
[29] For the sake of brevity we do not include these all results here/in this thesis.

model has high errors looking at the overall forecasts for all forecast horizons. The EIG3 and EIG4 models made exact forecasts as RW. The EIG1 and EIG2 models generally produce extreme local errors. Thus, the CLS model is superior. The errors in the other models are slightly higher. When we compare the RMSE measure calculated for each forecasting period with that of the RW model in β factors and area measure, we obtain the best results by the CLS model and the WA and MED models in λ. In Appendix 11, Table A.4.(a) shows the models with the best test results for area criteria. In the DM test, the CLS model outperforms the daily forecast horizon and forecasting period. The INVW model outperforms the forecasting period. In the t-test and Wilcoxon test, the INVW, SA, TA, and WA models yield satisfactory results. In the FC derived from the fit values of the models, the problem of model error trade-off in the forecast error can be an essential factor. The model error should not be small and the forecast error should be significant. However, a significant model error is undesirable, even if the forecast error is small. This suggested that these models are not dominant in the data, and their forecasts are unreliable (Szenczi, 2016).

In Version 2, the first striking result of the approach is that the forecasts of the factors outperform the RW model at short forecast horizons[30]. Even if SA makes extreme error forecasts, unlike the first approach, SA is not always the worst model for forecasting β factors. However, in contrast to the first approach, the forecast errors of CLS are higher. Regarding the factor and yield curve forecasts, we can say that the MED model generally yields better results at short and medium forecast horizons and the BG model at long horizons. The INVW model is better for Great Britain and Italy in the medium- and long-run. Despite these promising results, the INVW and BG models also produce extreme forecast errors. The tests for the area criteria are presented in Table A.4.(b) (See Appendix 11). Although the MED model is superior in the DM tests, SA is prominent in the other tests.

In version 3, we remove models that produce extreme errors, such as RNN. Although this choice makes forecasts more stable in terms of errors, the efficiency of the forecasts is reduced[31]. Because there are no model which has extreme errors in this approach, the

[30] For the sake of brevity we do not include these results here/in this thesis.
[31] For the sake of brevity we do not include these all results here/in this thesis.

results of the models are closer. This approach is superior to RW, particularly for short and medium forecast horizons. Again, SA is not always the model with the worst results in forecasting β factors. The CLS model produces more errors than the other models do. Regarding the factor and yield curve forecasts, we can say that the MED model generally yields better results at short forecast horizons. The MED and BG models provide better results for medium and long forecast horizons. The INVW model is occasionally better for medium and long forecast horizons. The results of the tests are presented in Table A.4.(c) (See Appendix 11). The superiority of the BG and SA models comes to the fore in the tests.

FC Version 4 provides promising results in tests against the forecast results of the RW model. The Canada, Germany, and France's models outperform the RW model in forecasting periods and daily tests (ttest, wtest, and DM) (See Table 7).

FC Version 5 uses the EL method for forecasting[32]. In the Boosting, Conditional RF, and RF models, we re-forecast with individual models: AR (1), ETS, TBATS, RW, ELM, and NNETAR, and calculate the fitted values. We use fitted values as input and forecast by using the forecast value of the individual models as a new input. In the boosting model, we use gradient boosting with regression trees, gradient boosting for additive models, and gradient boosting with component-wise linear model approaches in the optimizations and choose according to the results. We can see that the boost model errors are higher than those of the others. We also observe that approaches that use the fitted values of individual models as input generally yield the worst results.

FC versions 3, 4, and 5 generally produce stable forecasts without extreme errors.[33] To see how ANN models affect, we repeat FC versions 1,2 and 3 by excluding these models. In the DM tests, the ANN models mainly improved their forecast accuracy in short horizons in Version 2 and all horizons in Version 3. In both versions, we see these improvements in the forecasts of βs for short forecast horizons and λ for all horizons (the United States only in short horizons). It also increases the efficiency of the CLS of version

[32] For the sake of brevity we do not include these all results here/in this thesis.
[33] For all versions, see Table 7.

1 in the forecasting factors. Other models of Version 1 sometimes yield much better results.

4.4. CONCLUDING REMARKS

In this study, we use 62 different models or alternatives. We forecast all four factors of six countries by all models separately. We must produce at least 1488 graphs, error matrices… etc. It is impossible to show all the results and graphs of all models. In addition, some models make worse forecasts than others for each forecast horizon and period. Eliminating these models will simplify the presentation of the results and make the comparison between good models more understandable. For this purpose, we use the RW model as a benchmark. We divide the factor series into 12 periods and calculate the errors of all yields for each model for each forecast horizon and period using the methods described at the beginning of this section. For example, the calculation of the total yield error is based on TRMSFE, TMAFE, TMASFE, and TSMAPFE. We calculate the error values of all models at the forecast horizons according to these formulas, as shown in Appendix 12. Table A.5 in Appendix 12 is ranked according to the TRMSFE values. In Table A.5, this ranking does not change sufficiently for other error measures to affect grouping and evaluation. In some models, contrary to the general situation, extremely erroneous forecasts can be made at a few points, which leads to an overestimation of the forecast error of the entire series. However, the opposite situation may occur. We attempt to differentiate the models according to whether they always, occasionally, or never make good forecasts. We base this differentiation on the RW forecast errors. Thus, we eliminate some of the models as mentioned above.

Accordingly, we find kNN, MLP (and MLP-A), NNET (and NNET-A), NNETTS, GMDH, RNN (and RNN-A), LSTM, and GRU to be inefficient models with high error forecasts in every dimension (daily, forecast horizon and yields of standard maturities) and we eliminate these non-conventional models.[34] We further consider the RMSE of the NS factors and the area error measures of the yield curves and hypothesis tests in this elimination process. They have high forecast errors (See Table A.5). According to the

---

[34] For the sake of brevity we do not include these all results here/in this thesis.

analysis, the kNN model does not make extremely inaccurate forecasts; however, it does not generally make better forecasts. The GMDH model results in extreme errors for some forecast horizons.

The GRU model, which is claimed to be better than LSTM, generally leads to worse forecasts. The RNN-A model with increased epochs obtains more inaccurate results than the RNN model, whereas MLP-A vs. MLP and NNET-A vs. NNET obtain almost identical results.

We eliminate the EIG1, EIG2, and EIG3 models of the FC(Version 1) approach[35]. Because EIG3 obtains the same results as RW, the EIG2 model is mainly identical to RW, and the forecasts are highly inaccurate in the remaining parts. EIG1 produces exact forecasts as RW, but some countries make poor forecasts in one period or good forecasts for only a few days, which is insignificant in the comparisons. We conclude this result by comparing the tests with RW. We see that similiarities by results of total forecast errors in Table A.5. We compare each model in the FC(Version 2) models with Version 3 models[36], and because the errors of Version 2 are always higher, we also eliminate Version 2 models. We conclude that the models that produce extreme errors must be excluded. The SA model in version 1 always has higher errors than the other versions because the forecasts of some of the individual models in version 1 are used as highly inaccurate inputs, and these errors are directly incorporated into the SA model. Therefore, we eliminate this model.

We compare the results of the TAR model with those of the LLAR-LSTAR-SETAR approachs. The TAR model almost consistently outperformed the LLAR-LSTAR-SETAR approachs in terms of forecast efficiency. Therefore, we also eliminate the LLAR-LSTAR-SETAR models. We compare the GRNN and GRNNTSF models and eliminate GRNNTSF because RNN is always superior. We also eliminate the RF-C (cforest) model, as it almost always obtained worse results than the other RF models.

[35] For the sake of brevity we do not include these all results here/in this thesis.
[36] For the sake of brevity we do not include these all results here/in this thesis.

We compare the forecasting models designed with their alternatives in each country separately and eliminate less-efficient alternatives. We select ELM in Germany, Great Britain, and the United States; ELM-A in other countries; ENN in Germany, Great Britain, Italy, and the United States; ENN-A in other countries; JNN in Great Britain, Italy, and the United States; JNN-A in other countries; and NNETAR in Canada and France, NNETAR-A in other countries. We eliminate these alternatives.

Using elimination methods, we exclude 29 different models from the comparison. Thus, the number of models to be analyzed is reduced from 62 to 33. Among these models, there are still models constructed using the same approach, but they are not eliminated. Therefore, we select models representing this approach from the models using the same approach. The selection is based on the forecast error measures and hypothesis testing. To compare the efficiency of the models, we select those that give the best results.

In FC(Version 3) models, for all series, according to the forecasting periods and forecast horizon, the results of the hypothesis tests and values, such as the forecast error area measure of yields, we compare TRMSFE and the results of the hypothesis tests with each other and with the RW model. The MED and BG models generally provide the best results. Although other FC models are sometimes ahead, we can say that this is due to a slight difference and does not affect the ranking of the approach. As the MED and BG models can represent FC(Version 3), we exclude the other five models. We perform the same study in Version 1 and select the CLS model.

Because the XGB model is a special boosting approach and appears to provide more accurate forecasts than the BOOST model, we select the XGB model and exclude the BOOST model from the comparisons.

We also exclude the ARFIMA and ARIMA models because they generally rank lower than the other autoregression approaches, their error values are high, and their test results are insufficient compared with those of the Auto Arima and AR(1) models. Therefore, we analyze Auto ARIMA and AR(1) in the results.

Among the RF models, we select the RF and RF-AR (RF-HP for the United States) models. We observe that RF-AR provides better results at short and medium forecast horizons and RF long horizons (all horizons for the United States) compared to other RF models.

As a result of these selections, we reduce the number of models from 33 to 19. We summarize the forecast results of these models in Table 7. We compare the values with RW in Table 7. We obtain NS factors table from the RMSE values of the factor forecasts. We show the number of $RMSE_{RW}$>$RMSE_{MODEL}$ at "1-10", "11-20", "21-30", and "31-40" forecast horizons and at all forecast horizons (calculated daily) in Table 7. In the AREA columns of Table 7, we directly and individually compare each area with RW.

In the hypotheses test columns of Table 7, we show that models provide better results than RW in terms of the entire full forecast horizon (as in the daily RMSE calculations of the factors) and the forecasting periods in each forecast horizon. We show this in Table 7 as the daily sum (D.SUM). In Table 7, the models and periods with more numbers indicate fewer errors than RW. We have 120 forecast horizons for every ten periods, 480 forecast horizons for all periods, and 2279 forecast days (2239 in FC version 3). We also have a 22790 (22390 in FC version 3) area for every ten periods and a 91160 (89560 in FC version 3) area for all periods. If the numbers we find are more than half of the total numbers, we consider the model to be more efficient than RW. These are shown in bold font in Table 7. In the remainder of this section, we discuss only the models listed in Table 7.

Table 7: The Forecasting Period and Day Numbers Which Have Better RMSE, Tests and Area Errors of Models than RW

Canada	β_0					β_1					β_2					λ					AREA				
MODEL	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*
AA	35	34	37	37	143	42	29	15	15	101	31	26	30	31	118	**120**	**120**	**120**	**120**	**480**	10339	10632	10960	**11533**	43464
AR(1)	36	44	50	50	180	39	50	50	58	197	**70**	**76**	**80**	**80**	**306**	30	30	30	30	120	11258	9718	9357	9310	39643
ELM	10	12	7	21	50	28	21	19	30	98	24	30	39	50	143	**120**	**113**	**104**	**99**	**436**	8273	7545	7910	8486	32214
ENN	17	24	30	28	99	7	12	13	12	44	19	36	30	20	105	**103**	**101**	**87**	**85**	**376**	7663	7706	8195	8542	32106
ETS/TBATS	30	24	33	20	107	32	44	40	35	151	54	53	51	**63**	221	**120**	**120**	**120**	**120**	**480**	10921	11348	**11673**	**12167**	**46109**
GRNN	8	9	1	10	28	0	0	16	26	42	4	16	26	13	59	17	21	30	37	105	5392	5767	6276	6756	24191
JNN	10	12	16	13	51	6	0	0	3	9	9	6	10	33	58	**108**	**110**	**110**	**100**	**428**	7680	7898	9055	9637	34270
NNETAR	5	12	21	38	76	6	10	17	34	67	7	32	29	24	92	**79**	**60**	40	40	219	8262	7677	8541	9118	33598
TAR	37	37	30	24	128	23	28	40	40	131	28	50	57	50	185	**120**	**118**	**110**	**110**	**458**	10382	9906	10210	10870	41368
VAR(1)	3	10	7	0	20	0	0	0	4	4	6	1	0	0	7	73	69	52	50	244	4785	5808	6299	6717	23609
BAG	2	3	8	14	27	16	30	27	33	106	0	0	0	0	0	**120**	**120**	**120**	**120**	**480**	10202	10957	11240	11650	44049
FC(V4)	32	34	43	53	162	39	34	43	50	166	**67**	**64**	**72**	57	**260**	**120**	**120**	**120**	**120**	**480**	10894	11091	11269	12044	45298
RF	6	24	29	22	81	1	15	27	22	65	14	31	26	20	91	6	20	20	20	66	8991	9068	9292	9539	36890
F-AR	18	12	16	20	66	2	20	31	35	88	21	46	53	57	177	19	20	20	20	79	9461	8874	8404	8240	34979
XGB	4	7	30	32	73	1	20	24	35	80	7	11	19	26	63	1	0	0	0	1	8631	9373	9902	10043	37949
FC(V1).CLS	**61**	**63**	**61**	**68**	**253**	**64**	**71**	**74**	**82**	**291**	**75**	**79**	**82**	**84**	**320**	**120**	**120**	**120**	**120**	**480**	**11565**	**12486**	**12664**	**13133**	**49848**
FC(V3).BG	**110**	**93**	18	10	231	**116**	**100**	12	10	238	**120**	**99**	23	5	**247**	**120**	**120**	**109**	58	**407**	9662	9528	10009	10644	39843
FC(V3).MED	**113**	**95**	16	10	234	**110**	**91**	3	0	204	**118**	**100**	25	0	**243**	**120**	**120**	44	10	**294**	10106	9396	9710	9693	38905

	T_TEST_err						W_TEST_err						DM_TEST_err						DMTEST_area					
	1-10	*11-20*	*21-30*	*31-40*	*SUM*	*D.SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*D.SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*D.SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*D.SUM*
AA	8	20	21	24	73	561	0	11	20	23	54	433	7	25	27	25	84	959	6	10	15	31	62	1013
AR(1)	9	10	10	10	39	432	0	0	0	1	1	416	8	10	10	10	38	687	26	20	20	20	86	788
ELM	0	0	0	0	0	453	0	0	0	0	0	406	1	0	0	0	1	585	9	8	0	0	17	647
ENN	4	8	10	10	32	421	3	10	10	10	33	413	5	10	10	10	35	559	3	10	10	10	33	591
ETS/TBATS	11	43	60	**69**	183	614	0	3	11	28	42	449	7	32	38	43	120	1129	7	27	22	29	85	1070
GRNN	0	0	0	0	0	255	0	0	0	0	0	237	0	0	0	0	0	337	0	0	0	0	0	384
JNN	0	3	29	30	62	502	0	6	12	20	38	466	0	3	19	30	52	594	1	2	18	27	48	632
NNETAR	0	13	14	10	37	456	0	9	10	10	29	406	0	8	10	10	28	603	1	14	15	10	40	613
TAR	1	25	30	30	86	527	2	10	13	35	60	479	4	17	30	30	81	773	7	16	21	34	78	899
VAR(1)	0	0	0	0	0	236	0	0	0	0	0	216	0	0	0	0	0	342	0	0	0	0	0	416
BAG	10	38	38	44	130	629	2	19	30	24	75	521	9	24	37	34	104	1022	0	0	2	20	22	995
HYBRİD	20	45	53	**65**	183	575	0	12	17	20	49	488	10	41	59	**68**	178	1112	11	35	38	48	132	1034
RF	0	5	10	12	27	452	0	0	5	10	15	442	0	2	10	10	22	717	0	0	16	22	38	704
RF-AR	0	3	2	0	5	284	0	0	0	0	0	270	0	0	0	0	0	448	0	12	15	10	37	577
XGB	0	0	0	0	0	389	0	0	0	0	0	360	0	0	0	0	0	723	0	0	0	0	0	693
FC(V1).CLS	39	**65**	**75**	**85**	**264**	31	0	0	0	0	0	30	32	55	60	**74**	221	**1274**	22	36	38	41	137	**1176**
FC(V3).BG	0	8	19	49	76	551	0	0	17	20	37	506	0	5	17	34	56	815	2	0	27	46	75	815
FC(V3).MED	1	10	19	43	73	391	0	0	0	0	0	373	0	3	14	36	53	750	5	0	18	34	57	776

Germany	β_0					β_1					β_2					λ					AREA				
MODEL	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*
AA	25	28	16	10	79	33	37	25	33	128	**70**	58	46	43	217	**120**	**120**	**120**	**120**	**480**	10757	**11546**	**12212**	**12873**	**47388**
AR(1)	12	10	13	20	55	50	50	51	60	211	58	50	40	40	188	20	20	20	20	80	10586	10224	9541	9064	39415
ELM	19	20	20	25	84	20	20	22	27	89	34	50	50	47	181	**119**	**101**	**78**	59	**357**	8774	8011	7697	7567	32049
ENN	10	10	10	18	48	3	28	30	38	99	11	35	53	40	139	**101**	**92**	**86**	**90**	**369**	7549	7856	8355	8595	32355
ETS/TBATS	27	24	21	20	92	38	31	30	30	129	56	59	57	57	229	**120**	**120**	**120**	**120**	**480**	11131	**11948**	**12809**	**13344**	**49232**
GRNN	4	11	11	11	37	0	18	20	16	54	8	39	50	42	139	1	0	0	0	1	5976	7100	7521	7513	28110
JNN	9	10	14	30	63	16	30	40	40	126	12	29	30	28	99	**87**	**87**	**72**	**70**	**316**	7835	8255	8821	9384	34295
NNETAR	0	9	21	27	57	9	20	20	20	69	14	30	30	29	103	**91**	**78**	**65**	53	**287**	8510	8904	9703	9788	36905
TAR	36	47	56	42	181	33	37	39	40	149	48	**65**	52	50	215	**110**	**110**	**100**	**98**	**418**	10273	10280	10164	10183	40900
VAR(1)	4	20	20	26	70	2	18	23	30	73	8	10	35	36	89	**92**	**90**	**90**	**80**	**352**	6994	7632	7561	7247	29434
BAG	3	0	6	23	32	1	10	20	20	51	12	14	27	27	80	**120**	**120**	**120**	**120**	**480**	10453	11178	**11660**	**12221**	45512
HYBRİD	44	49	50	48	191	44	30	44	54	172	60	60	**66**	60	**246**	**119**	**120**	**120**	**120**	**479**	11147	**12188**	**12725**	**12973**	**49033**
RF	0	20	35	30	85	5	28	37	33	103	5	26	47	60	138	0	0	0	0	0	8375	8639	8858	9185	35057
RF-AR	5	22	34	40	101	12	34	40	40	126	19	51	43	38	151	23	20	20	20	83	8416	8327	8054	7713	32510
XGB	0	0	10	10	20	2	23	34	36	95	16	43	51	54	164	0	0	0	0	0	7954	8861	8824	8855	34494
FC(V1).CLS	34	27	30	30	121	**76**	**65**	56	47	**244**	**95**	**96**	**80**	**88**	**359**	**116**	**110**	**110**	**116**	**452**	**11727**	**12610**	**12798**	**13221**	**50356**
FC(V3).BG	**110**	**84**	13	7	214	**118**	**93**	18	10	239	**110**	**90**	12	0	212	**120**	**120**	**105**	44	**389**	10152	10925	11694	12093	44864
FC(V3).MED	**110**	**86**	7	0	203	**120**	**101**	19	10	**250**	**110**	**85**	8	0	203	120	120	45	10	**295**	10429	10640	10843	10986	42898

	T_TEST_err						W_TEST_err						DM_TEST_err						DMTEST_area					
MODEL	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*D.SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*D.SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*D.SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*D.SUM*
AA	24	36	35	40	135	662	4	19	34	40	97	616	26	31	37	52	146	**1198**	18	19	14	24	75	1138
AR(1)	18	20	20	13	71	547	0	6	10	1	17	537	9	17	20	14	60	828	16	20	20	15	71	825
ELM	4	10	10	10	34	523	0	1	17	20	38	502	0	10	10	10	30	597	4	10	10	10	34	627
ENN	0	6	10	13	29	376	0	6	10	12	28	360	0	6	10	14	30	492	7	10	10	10	37	560
ETS/TBATS	41	**61**	52	60	214	639	3	23	36	40	102	580	41	60	54	60	215	**1256**	28	54	50	54	186	**1188**
GRNN	0	5	16	20	41	342	0	0	10	8	18	335	0	3	10	19	32	483	0	9	10	17	36	494
JNN	2	10	11	17	40	445	0	8	11	22	41	427	0	10	17	23	50	543	4	10	12	26	52	590
NNETAR	0	0	0	7	7	470	0	0	0	0	0	435	0	0	0	0	0	594	0	0	0	0	0	704
TAR	5	20	22	23	70	717	1	2	0	0	3	647	5	15	19	10	49	907	0	3	10	10	23	848
VAR(1)	1	18	17	18	54	681	0	16	25	27	68	617	0	15	16	18	49	819	0	0	0	0	0	522
BAG	13	24	39	40	116	699	8	23	30	36	97	644	17	30	30	38	115	1093	7	12	16	20	55	1054
HYBRİD	35	**62**	**65**	60	222	770	2	31	42	40	115	684	38	**64**	**62**	60	224	1215	22	45	60	60	187	1138
RF	0	0	0	0	0	388	0	0	0	0	0	373	0	0	0	0	0	621	0	0	1	2	3	630
RF-AR	0	1	4	4	9	440	0	3	9	0	12	380	0	2	9	0	11	551	2	8	14	7	31	586
XGB	0	0	0	0	0	379	0	0	0	0	0	363	0	0	0	0	0	630	0	0	0	0	0	615
FC(V1).CLS	60	60	**68**	**80**	**268**	5	0	0	0	0	0	3	57	**62**	**61**	**74**	**254**	**1288**	13	40	43	50	146	**1183**
FC(V3).BG	6	20	26	41	93	733	0	3	18	30	51	680	4	20	30	46	100	1049	9	20	27	41	97	991
FC(V3).MED	11	10	21	23	65	640	0	0	4	20	24	603	5	10	16	25	56	959	14	19	21	37	91	917

France	β_0					β_1					β_2					λ					AREA				
MODEL	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*
AA	13	8	16	20	57	22	26	30	33	111	50	39	58	30	177	**120**	**120**	**120**	**120**	**480**	10580	10868	11082	11297	43827
AR(1)	21	21	30	30	102	49	55	60	60	224	**65**	52	60	51	228	40	40	40	40	160	11014	10350	10029	9821	41214
ELM	20	22	31	30	103	20	31	40	40	131	8	16	30	39	93	**120**	**120**	**120**	**120**	**480**	7797	7084	7534	7712	30127
ENN	0	2	10	10	22	20	30	26	26	102	22	24	30	30	106	**109**	**114**	**120**	**120**	**463**	7378	6395	6825	7022	27620
ETS/TBATS	37	40	42	40	159	35	33	23	17	108	53	42	45	35	175	**120**	**120**	**120**	**120**	**480**	**11659**	**12468**	**12690**	**12905**	**49722**
GRNN	3	14	18	13	48	0	17	20	20	57	10	15	19	39	83	6	20	20	20	66	5224	5895	6534	7275	24928
JNN	0	21	23	36	80	11	31	48	50	140	25	33	35	40	133	**110**	**110**	**110**	**110**	**440**	7097	7060	7444	7439	29040
NNETAR	0	4	14	29	47	18	14	28	39	99	15	29	15	16	75	**114**	**120**	**113**	**110**	**457**	7914	7398	7816	8275	31403
TAR	21	25	20	20	86	26	30	24	17	97	20	22	20	28	90	**120**	**120**	**120**	**120**	**480**	9700	9102	9265	9534	37601
VAR(1)	0	0	0	0	5	0	0	0	9	9	0	0	9	10	19	**118**	**110**	**110**	**103**	**441**	2622	3253	3761	4592	13912
BAG	2	1	3	18	24	0	7	15	23	45	5	8	23	25	61	**120**	**120**	**120**	**120**	**480**	9971	10575	10790	11016	42352
HYBRİD	36	23	40	40	139	42	40	40	40	162	55	52	59	53	219	**120**	**120**	**120**	**120**	**480**	**11226**	**11631**	**12121**	**12485**	**47463**
RF	0	11	33	40	84	0	16	27	26	69	9	12	34	39	94	0	0	0	0	0	8814	9519	10114	10211	38658
RF-AR	0	15	20	27	62	14	50	50	50	164	27	55	52	49	183	12	0	0	0	12	9245	8966	8487	8563	35261
XGB	0	0	14	24	38	0	5	17	26	48	8	17	26	44	95	0	0	0	0	0	8622	9345	9849	10210	38026
FC(V1).CLS	**72**	**67**	**62**	**76**	**277**	59	56	**68**	**62**	**245**	**93**	**101**	**93**	**95**	**382**	**88**	**100**	**110**	**113**	**411**	**12419**	**13041**	**13339**	**13383**	**52182**
FC(V3).BG	**110**	**85**	17	10	222	**114**	**85**	13	10	222	**113**	**94**	11	0	218	**120**	**120**	**117**	**85**	**442**	8836	8642	9234	9694	36406
FC(V3).MED	**110**	**89**	20	10	229	**118**	**87**	13	10	228	**119**	**94**	11	0	224	**120**	**120**	**102**	26	**368**	9813	9552	9812	9831	39008

	T_TEST_err						W_TEST_err						DM_TEST_err						DMTEST_area					
MODEL	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*D.SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*D.SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*D.SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*D.SUM*
AA	17	22	25	31	95	349	0	9	19	14	42	288	16	30	30	29	105	993	17	30	30	25	102	993
AR(1)	4	10	10	10	34	466	2	2	4	14	22	382	19	10	15	20	64	780	18	10	20	24	72	876
ELM	0	7	10	10	27	417	13	7	10	10	40	389	7	6	10	10	33	511	3	10	10	10	33	606
ENN	0	0	6	10	16	267	6	0	3	10	19	266	3	0	6	10	19	377	2	0	3	10	15	448
ETS/TBATS	12	35	42	50	139	376	0	2	14	34	50	321	18	36	42	50	146	1126	25	32	42	50	149	1152
GRNN	0	0	9	11	20	286	3	0	0	2	5	267	2	0	5	11	18	434	2	6	19	20	47	467
JNN	0	0	4	4	8	338	0	0	10	3	13	310	0	0	10	3	13	504	0	8	10	3	21	571
NNETAR	0	0	0	6	6	356	0	0	0	12	12	339	0	0	0	16	16	507	0	0	0	20	20	552
TAR	0	17	20	20	57	460	8	0	0	0	8	447	4	8	15	11	38	676	4	13	12	10	39	772
VAR(1)	0	0	0	0	0	196	0	0	0	0	0	204	0	0	0	0	0	225	0	0	0	0	0	220
BAG	3	17	20	29	69	563	0	4	20	20	44	515	4	20	20	20	64	978	1	12	20	20	53	938
HYBRİD	19	34	44	56	153	530	0	7	20	23	50	437	19	28	50	60	157	1067	14	15	35	40	104	1089
RF	0	7	10	10	27	395	0	0	0	0	0	340	0	9	10	9	28	797	0	0	0	0	0	763
RF-AR	1	9	10	10	30	392	5	9	0	5	19	362	6	10	10	10	36	558	8	10	10	10	38	633
XGB	0	0	0	0	0	358	0	0	0	0	0	309	0	0	0	0	0	753	0	0	0	0	0	738
FC(V1).CLS	30	46	52	41	169	58	0	0	10	10	20	50	33	50	50	48	181	**1223**	29	43	49	35	156	**1203**
FC(V3).BG	3	10	20	20	53	532	3	0	3	8	14	541	3	1	16	20	40	749	4	0	0	3	7	751
FC(V3).MED	5	10	10	20	45	560	8	10	10	10	38	573	11	10	16	21	58	824	5	0	0	8	13	807

Great Britain	β_0					β_1					β_2					λ					AREA				
MODEL	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*
AA	23	28	27	26	104	51	45	30	30	156	29	27	48	48	152	**120**	**120**	**120**	**120**	**480**	11297	**12024**	**12158**	**12435**	**47914**
AR(1)	21	34	40	48	143	20	20	42	42	124	51	50	54	60	215	30	30	30	30	120	10670	9934	9981	9651	40236
ELM	4	0	0	8	12	23	28	40	40	131	26	32	50	44	152	**120**	**102**	**83**	**72**	**377**	7679	6259	6380	6475	26793
ENN	0	3	11	20	34	0	18	30	30	78	3	17	28	30	78	**96**	**90**	**90**	**71**	**347**	7100	6807	7116	7481	28504
ETS/TBATS	41	44	46	44	175	40	35	49	47	171	60	**67**	56	**64**	**247**	**120**	**120**	**120**	**120**	**480**	**11500**	**11984**	**12297**	**12650**	**48431**
GRNN	3	10	14	16	43	0	6	28	32	66	4	18	32	35	89	7	11	14	20	52	5570	6536	6964	6936	26006
JNN	7	10	9	16	42	0	16	40	40	96	17	26	34	26	103	**61**	50	50	59	220	7063	6123	5879	6191	25256
NNETAR	3	2	17	22	44	1	17	25	37	80	13	20	24	41	98	**90**	**82**	**63**	45	**280**	7979	7617	8108	8116	31820
TAR	20	20	18	17	75	17	27	40	40	124	34	40	40	34	148	**120**	**116**	**110**	**100**	**446**	9848	9127	8663	8281	35919
VAR(1)	8	0	8	10	26	0	0	11	20	31	5	10	10	10	35	**70**	**62**	60	60	**252**	6803	7905	7988	7333	30029
BAG	0	4	17	24	45	3	10	13	16	42	5	10	10	14	39	**120**	**120**	**120**	**120**	**480**	4637	8991	12500	16264	42392
HYBRİD	35	48	34	30	147	31	23	21	37	112	41	40	40	40	161	**120**	**120**	**120**	**120**	**480**	4919	9664	13506	17318	45407
RF	1	10	20	26	57	1	13	37	41	92	2	20	32	37	91	0	4	10	10	24	5095	9713	13831	17700	46339
RF-AR	10	17	20	20	67	9	27	30	36	102	27	50	50	49	176	15	13	20	20	68	5174	10738	13371	12963	42246
XGB	0	5	11	19	35	8	19	44	52	123	2	21	43	46	112	0	0	0	0	0	4856	9498	13542	17458	45354
FC(V1).CLS	59	56	**63**	55	233	57	**62**	**71**	**64**	**254**	**76**	**82**	**81**	**86**	**325**	**119**	**112**	**110**	**110**	**451**	**12153**	**12993**	**13409**	**13814**	**52369**
FC(V3).BG	**110**	**77**	5	2	194	**119**	**91**	15	10	235	**117**	**82**	18	10	227	**120**	**118**	**106**	**45**	**389**	9161	8900	9214	9304	36579
FC(V3).MED	**110**	**80**	8	0	198	**117**	**90**	17	10	234	**117**	**89**	18	10	234	**120**	**120**	58	12	**310**	9453	8851	8862	8610	35776

	T_TEST_err						W_TEST_err						DM_TEST_err						DMTEST_area					
MODEL	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*D.SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*D.SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*D.SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*D.SUM*
AA	35	51	60	54	200	525	0	0	7	20	27	447	33	39	47	50	169	1168	5	25	35	46	111	1066
AR(1)	10	10	2	0	22	488	0	0	0	0	0	477	10	1	0	0	11	729	17	15	10	0	42	812
ELM	0	0	8	10	18	405	0	9	10	6	25	390	0	9	10	10	29	465	0	5	9	7	21	498
ENN	0	0	0	0	0	326	0	0	0	0	0	307	0	0	0	0	0	434	0	0	0	0	0	435
ETS/TBATS	29	36	49	73	187	408	0	8	0	8	16	350	39	47	38	55	179	1165	28	30	39	45	142	1091
GRNN	0	0	0	0	0	363	0	0	0	0	0	366	0	0	0	0	0	452	0	0	0	0	0	477
JNN	0	0	0	0	0	332	0	0	0	0	0	311	0	0	0	0	0	400	0	0	0	0	0	380
NNETAR	0	0	3	1	4	456	0	0	0	0	0	435	0	0	2	0	2	573	0	0	7	10	17	567
TAR	1	0	0	0	1	444	0	9	6	0	15	435	0	0	0	0	0	642	3	0	0	0	3	683
VAR(1)	5	10	13	30	58	479	6	24	21	5	56	470	5	18	20	27	70	540	6	10	10	10	36	575
BAG	4	18	24	32	78	610	0	4	5	20	29	565	0	14	17	39	70	1027	0	0	55	**113**	168	800
HYBRİD	15	30	33	43	121	514	5	19	15	4	43	480	22	40	22	23	107	951	0	3	**95**	**120**	218	872
RF	0	0	1	6	7	414	0	0	0	0	0	394	0	0	0	0	0	674	0	10	**103**	**120**	233	880
RF-AR	0	0	0	0	0	474	0	0	0	0	0	454	0	0	0	0	0	555	0	23	**88**	**88**	199	761
XGB	0	0	0	0	0	462	0	0	0	0	0	431	0	0	0	0	0	764	0	7	**93**	**120**	220	850
FC(V1).CLS	59	**61**	**69**	**75**	**264**	13	0	0	0	0	0	12	56	55	59	**63**	233	1332	23	33	40	24	120	**1223**
FC(V3).BG	0	2	5	10	17	473	0	0	0	0	0	450	0	0	0	3	3	731	0	0	0	3	3	703
FC(V3).MED	12	7	7	10	36	484	0	0	0	0	0	469	8	1	0	0	9	726	11	7	0	0	18	668

Italy	β_0					β_1					β_2					λ					AREA				
MODEL	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*
AA	25	16	28	25	94	21	8	4	4	37	38	42	53	56	189	**120**	**120**	**120**	**120**	**480**	**11675**	**11654**	**11872**	**11881**	**47082**
AR(1)	38	34	30	30	132	40	40	41	49	170	53	**66**	60	52	231	10	10	10	10	40	9558	9033	8877	8478	35946
ELM	9	10	10	20	49	25	22	27	25	99	16	9	18	20	63	**114**	**108**	**90**	**90**	**402**	8433	8050	8403	8388	33274
ENN	8	10	22	30	70	14	20	25	27	86	13	18	20	25	76	**108**	**110**	**110**	**110**	**438**	7034	7181	7530	7819	29564
ETS/TBATS	34	18	11	14	77	36	50	33	25	144	16	18	8	17	59	**120**	**120**	**120**	**120**	**480**	11368	11010	11335	**11809**	45522
GRNN	3	8	5	2	18	8	16	10	10	44	3	14	27	33	77	0	3	0	0	3	6527	7308	7670	8033	29538
JNN	0	22	40	46	108	10	1	12	20	43	4	15	21	44	84	**118**	**120**	**115**	**110**	**463**	6810	7613	8151	8750	31324
NNETAR	17	21	20	20	78	14	17	10	10	51	17	9	22	20	68	**99**	**110**	**98**	**80**	**387**	8552	8277	8107	8402	33338
TAR	34	22	20	20	96	20	20	20	20	80	26	20	20	29	95	**120**	**120**	**120**	**118**	**478**	10049	9427	9404	9100	37980
VAR(1)	1	9	10	10	30	1	37	31	32	101	0	0	3	10	13	**110**	**95**	**81**	**86**	**372**	5841	6195	5764	5575	23575
BAG	7	22	30	34	93	3	22	27	22	74	2	11	19	21	53	**120**	**120**	**120**	**120**	**480**	3925	8648	**12760**	**16518**	41851
HYBRİD	39	35	39	35	148	46	50	50	50	196	45	33	30	30	138	**120**	**120**	**120**	**120**	**480**	4138	9418	**13946**	**18255**	**45757**
RF	0	0	0	0	0	0	12	22	35	69	2	32	37	42	113	0	0	0	0	0	4520	10362	**14756**	**18380**	**48018**
RF-AR	7	21	25	34	87	18	40	40	40	138	13	38	50	50	151	8	0	0	0	8	4418	10591	**13663**	**14478**	43150
XGB	4	9	10	18	41	7	17	35	50	109	10	25	33	49	117	0	0	0	0	0	4294	9819	**14460**	**18423**	**46996**
FC(V1).CLS	47	54	54	52	207	58	**67**	**78**	**70**	**273**	**84**	**86**	**81**	**67**	**318**	**105**	**110**	**108**	**110**	**433**	**12296**	**12543**	**12788**	**13262**	**50889**
FC(V3).BG	**110**	**87**	19	10	226	**110**	**85**	18	10	223	**119**	**87**	26	10	242	**120**	**120**	**118**	**69**	**427**	9482	9582	9798	10152	39014
FC(V3).MED	**110**	**92**	14	10	226	**110**	**79**	10	10	209	**120**	**86**	23	10	239	**120**	**116**	27	10	**273**	10055	9927	9552	9597	39131

	T_TEST_err						W_TEST_err						DM_TEST_err						DMTEST_area					
MODEL	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*D.SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*D.SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*D.SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*D.SUM*
AA	24	23	27	40	114	639	0	0	9	25	34	565	22	21	21	30	94	1089	18	18	30	22	88	1048
AR(1)	14	20	20	20	74	562	0	1	11	14	26	529	6	13	20	20	59	734	10	17	27	30	84	721
ELM	0	0	0	0	0	507	0	1	0	0	1	461	1	0	0	1	2	638	1	0	1	10	12	660
ENN	0	0	7	20	27	381	0	0	9	19	28	354	0	0	9	20	29	464	0	0	13	20	33	522
ETS/TBATS	18	11	19	16	64	487	0	0	4	18	22	449	16	14	31	30	91	1068	11	8	16	34	69	995
GRNN	0	0	9	15	24	314	0	3	9	11	23	294	0	0	9	10	19	450	0	8	18	20	46	499
JNN	0	0	0	0	0	383	0	0	0	0	0	334	0	0	0	0	0	521	0	0	3	10	13	555
NNETAR	0	0	7	10	17	459	0	0	10	14	24	455	0	0	9	10	19	580	0	0	9	10	19	634
TAR	3	0	1	10	14	561	0	0	0	0	0	521	0	0	0	4	4	698	10	12	0	5	27	775
VAR/VECM	0	0	0	0	0	291	0	0	0	0	0	267	0	0	0	0	0	370	0	0	0	0	0	220
BAG	9	24	20	20	73	637	1	5	22	30	58	573	9	28	30	30	97	986	0	0	63	**110**	173	792
HYBRİD	35	33	43	40	151	548	0	7	16	20	43	492	30	32	40	41	143	1015	0	3	**94**	**120**	217	854
RF	0	0	0	0	0	384	0	0	0	0	0	373	0	0	0	0	0	611	0	9	**104**	**120**	233	956
RF-AR	2	5	18	16	41	488	0	0	11	10	21	436	0	4	13	10	27	590	0	26	**87**	**106**	219	746
XGB	6	6	16	16	44	482	0	0	6	20	26	438	0	6	17	20	43	758	0	7	**101**	**120**	228	926
FC(V1).CLS	57	**80**	**79**	**80**	**296**	19	0	0	0	0	0	14	56	**80**	**79**	**80**	**295**	**1231**	32	57	49	**66**	204	**1166**
FC(V3).BG	4	10	11	30	55	588	3	7	12	20	42	558	4	15	12	24	55	815	4	18	15	31	68	828
FC(V3).MED	2	20	10	10	42	515	3	6	10	10	29	480	6	10	10	10	36	792	1	16	18	16	51	790

United States	β0					β1					β2					λ					AREA				
MODEL	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*
AA	60	34	26	22	142	58	**65**	50	49	222	59	**74**	**64**	46	**243**	56	**66**	**64**	52	238	10143	10115	10272	10127	40657
AR(1)	20	20	20	27	87	14	0	10	12	36	40	**27**	11	10	88	41	40	40	40	161	8957	7555	7188	7073	30773
ELM	8	10	10	15	43	5	18	20	20	63	16	10	10	17	53	22	29	30	30	111	5737	5125	5064	5230	21156
ENN	4	10	10	23	47	7	2	0	0	9	26	39	34	30	129	3	9	14	22	48	6418	6545	6466	6723	26152
ETS/TBATS	**65**	56	50	54	225	74	**65**	**64**	42	**245**	**78**	**89**	**88**	**90**	**345**	**71**	**93**	**66**	**63**	**293**	10758	10625	10541	10005	41929
GRNN	10	0	0	16	26	3	2	19	15	39	19	27	35	42	123	11	22	37	33	103	4532	5373	5813	6050	21768
JNN	6	0	8	20	34	6	6	0	2	14	18	23	23	34	98	20	31	23	28	102	7315	7372	7633	7960	30280
NNETAR	8	8	10	19	45	3	7	6	10	26	8	4	13	18	43	7	3	10	16	36	6946	6180	6251	6314	25691
TAR	19	32	40	38	129	23	36	37	12	108	43	20	17	11	91	52	45	40	31	168	8718	7866	7656	7347	31587
VAR(1)	36	28	26	30	120	50	57	56	49	212	11	10	10	10	41	33	54	35	25	147	8478	8057	8062	8029	32626
BAG	19	14	3	12	48	23	29	36	36	124	45	57	47	31	180	18	30	34	26	108	10211	9967	9711	9565	39454
HYBRİD	47	59	50	42	198	**69**	**69**	56	50	**244**	48	43	40	40	171	**68**	**72**	59	49	**248**	10035	9927	10375	10740	41077
RF	13	9	20	20	62	21	16	15	22	74	25	30	41	46	142	14	20	34	41	109	7683	7880	8684	8998	33245
RF-HP	5	13	41	49	108	9	21	32	45	107	21	16	19	22	78	53	50	47	42	192	7881	8642	8003	7316	31842
XGB	5	19	24	31	79	1	16	24	30	71	19	25	35	40	119	6	23	28	33	90	7942	8866	9588	9941	36337
FC(V1).CLS	**77**	**70**	**75**	**65**	**287**	55	**62**	**64**	**62**	**243**	**96**	**72**	**72**	**65**	**305**	**67**	56	59	**64**	**246**	11104	**11421**	**11586**	**11837**	**45948**
FC(V3).BG	**110**	**79**	5	8	202	**115**	**84**	2	0	201	**107**	**69**	21	5	202	**113**	**84**	19	10	226	8916	8784	8885	8690	35275
FC(V3).MED	**110**	**78**	5	8	201	**118**	**88**	2	0	208	**98**	**71**	21	5	195	**120**	**83**	21	10	234	9072	8911	9395	9214	36592

	T_TEST_err						W_TEST_err						DM_TEST_err						DMTEST_area					
MODEL	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*D.SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*D.SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*D.SUM*	*1-10*	*11-20*	*21-30*	*31-40*	*SUM*	*D.SUM*
AA	10	30	33	40	113	489	0	20	30	31	81	446	7	28	40	40	115	938	0	10	33	36	79	874
AR(1)	0	10	10	10	30	263	0	0	0	0	0	197	0	1	10	2	13	397	2	4	10	13	29	544
ELM	0	0	0	0	0	178	0	0	0	0	0	167	0	0	0	0	0	222	0	0	0	0	0	342
ENN	0	0	0	0	0	207	0	0	0	0	0	194	0	0	0	0	0	292	0	0	0	0	0	370
ETS/TBATS	9	10	9	10	38	261	0	0	3	7	10	242	9	6	9	10	34	860	5	5	1	7	18	862
GRNN	0	0	0	0	0	195	0	0	0	0	0	159	0	0	0	0	0	298	0	0	0	0	0	339
JNN	0	0	7	10	17	319	0	0	9	10	19	301	0	0	8	10	18	437	0	5	14	10	29	508
NNETAR	0	0	0	0	0	216	0	0	0	0	0	177	0	0	0	0	0	289	0	0	0	0	0	384
TAR	0	0	0	0	0	293	0	0	0	0	0	265	0	0	0	0	0	414	0	0	0	0	0	580
VAR(1)	0	12	31	30	73	526	0	12	29	30	71	510	0	12	30	30	72	646	1	15	25	15	56	655
BAG	10	23	30	30	93	444	7	21	30	23	81	393	12	23	30	24	89	940	0	22	13	10	45	799
HYBRİD	0	3	23	25	51	382	0	0	10	20	30	329	0	4	17	20	41	891	0	0	9	18	27	840
RF	0	0	0	0	0	307	0	0	0	0	0	265	0	0	0	0	0	533	0	0	0	1	1	564
RF-A	0	2	10	12	24	221	0	0	0	0	0	186	0	0	0	0	0	320	0	9	10	11	30	462
XGB	0	0	0	0	0	330	0	0	0	0	0	289	0	0	0	0	0	610	0	0	0	0	0	602
FC(V1).CLS	20	17	19	25	81	0	0	0	0	0	0	0	20	22	22	21	85	1107	1	0	19	27	47	993
FC(V3).BG	0	0	2	19	21	316	0	0	0	11	11	284	0	0	0	17	17	536	0	0	14	12	26	621
FC(V3).MED	0	0	7	14	21	285	0	0	0	6	6	260	0	0	2	6	8	594	0	0	10	11	21	658

In Table 7, we compare the Auto ARIMA and AR(1). Accordingly, Auto ARIMA is superior in every comparison, except for the βs of Canada, France, Great Britain, and Italy. The TAR model is statistically worse than autoregressive models and according to hypothesis tests in Table 7, except for the long horizon of Canada. The TAR model also has a better βs forecast than AR(1).

In Table 7. ENN, GRNN, VAR(1), and ELM (except for some βs) do not provide good results and are not worse than autoregressive models. VAR(1) is slightly more efficient in the United States, but not as efficient as in Auto ARIMA. ELM is usually an ANN model that produces the best factor results.

According to Table 7, bagging model appears to have an advantage over RW for medium and long forecast horizons. This advantage disappears with NS factors and forecasting period tests. Including the forecasting factors, the FC(Version 4) model is one of the best models for RW. RF seems to provide better results than RW in Great Britain and Italy in medium and long horizons. One can say that the RF model yields better results with AR(1) as the input. However, this is not always reflected in the tests.

We see that ETS/TBATS is less efficient than FC(Version 4) in factors other than the United States in Table 7. We can say that the ETS/TBATS and FC(Version 4) models perform well compared to the other models except for the FC(Version 1)-CLS, FC(Version 3)-MED, and FC(Version 3)-BG models.

The FC models outperform RW in terms of factors and yield curves at short and medium forecast horizons in Table 7. FC(Version 1)-CLS, on the other hand, achieves good results at almost all forecast horizons, and in the tests, this superiority is observed at medium and long horizons. The fact that models such as FC(Version 1)-CLS, FC(Version 4), and ETS/TBATS, which give good results, are sometimes worse in the daily comparison (D.SUM columns in Table 7) in the tests is thought to be due to the fact that the daily forecast series are very good in certain parts, not in all horizons.

Diebold and Li (2006) state that they use monthly frequency data because the skewness and kurtosis of the daily frequency data do not comply with the normal distribution, which negatively affects the forecasts. Table 8 shows the skewness and kurtosis values of the data used in this study according to the daily and monthly frequencies. We observe that converting the frequency to monthly can disrupts these momentums even more. Therefore, reducing the frequency would not have a positive effect on some forecasts. Unexpectedly, RW is effective for non-normally distributed data.

Table 8: Kurtosis and Skewness Value for Daily and Monthly Factor Data

	DAILY								MONTHLY							
	Kurtosis				Skewness				Kurtosis				Skewness			
	β0	**β1**	**β2**	**λ**	**β0**	**β1**	**β2**	**λ**	**β0**	**β1**	**β2**	**λ**	**β0**	**β1**	**β2**	**λ**
CA	2,60	3,85	3,98	5,97	0,33	0,10	0,98	2,02	4,52	3,68	5,09	13,64	-0,52	0,38	1,51	3,12
DE	2,01	2,51	5,38	23,94	0,34	-0,28	0,91	4,53	1,94	3,43	11,11	42,37	0,25	0,57	2,69	6,12
FR	1,77	2,18	3,79	4,90	0,19	0,08	0,52	1,05	2,58	6,63	10,43	68,68	0,16	-0,22	2,43	7,51
GB	1,92	2,05	2,35	6,85	0,24	-0,20	-0,08	1,78	2,10	2,15	7,84	127,00	0,02	0,01	1,75	10,68
IT	2,04	2,01	7,35	3,30	0,31	0,02	1,40	0,10	2,19	2,38	4,00	8,68	0,39	-0,09	1,13	1,88
US	2,73	2,16	2,65	32,15	0,38	0,08	0,40	4,94	2,69	2,13	2,78	34,03	0,35	0,12	0,43	5,21

According to the NS model, the long term should correspond to β_0, and the short term should correspond to the sum of β_0 and β_1 (See Equation 1.9 and Equation 1.10). The FC(Version 3)-MED model makes the best forecasts for β_0 for short and medium forecast horizons, and the FC(Version 1)-CLS and RW models make the best forecasts for long forecast horizons. FC(Version 3)-MED forecasts the best sum of β_0 and β_1, which corresponds to the short term for short forecast horizons, and FC(Version 1)-CLS model forecasts for medium and long forecast horizons[37].

[37] For the sake of brevity we do not include these all results here/in this thesis.

CONCLUSION

Interest rates are fundamental to the economy and finance, where the yield curves derived from interest rates provide crucial information for various fields. The development of financial markets and the rapid growth of money market volumes further highlight the significance of these curves. This study mainly focuses on forecasting yield curves, which is essential for constructing and managing the future and finances on an economy. In order to achieve accurate forecasts, the Nelson-Siegel factors are forecasted effectively where we recognize that solely relying on Diebold-Li's AR(1) model may not suffice for optimal forecasting of yield curves. To this aim, in addition to conventional models, we employ non-conventional approaches such as ANN and k-nearest neighbors (kNN). Further, we explore forecast combinations that integrate multiple models to enhance the accuracy instead of limiting ourselves to a single model.

This study analyzes daily yield data of G-7 countries, excluding Japan, over the period 2010-2022 focusing on the maturities from three months to 30 years. We examine the implications of treating the Nelson-Siegel Model's λ parameter as a constant and criticisms of the constant value proposed by Diebold and Li (2006). Varying parameter λ disrupts the stationarity of the β parameters, which diminishes the efficiency of forecasting. On the contrary, maintaining a constant λ inhibits the curvature and slope components' dynamic nature, thereby compromising the yield curves' flexibility and, potentially preventing negative yields where multicollinearity issues in the slope and curvature terms might arise. To address this concern, we apply the Hodrick-Prescott (HP) filter to λ (excluding the United States data), resulting in slowly varying λ values that reflect economic realities more accurately and align with Nelson-Siegel's theory, ultimately enhancing the reliability of our forecasts.

In this study, we categorize the data into training, testing, and forecasting. We determine the parameters through optimization in a specific sequence. With a multistep approach, we allow errors tend to accumulate as the forecast horizon increases. However, because of the absence of the correlation in the lagged values of the data, we cannot anticipate effective results from direct forecasting methods, which either by forecasting one step at each time, or by simultaneously forecasting the entire horizon. In this study, using a multistep approach, we use the sliding window methodology to forecast a 40-day horizon. Note that since random search techniques could be time-consuming, and in ANNs many parameters should be estimated, these networks are essentially black boxes where identifying the parameters to focus on might be challenging. Thus, in this study, we favor grid search for parameter optimization where it is a more systematic approach.

In this thesis we generate forecasts of yield curves using 62 models across conventional, non-conventional, ensemble learning (EL), and forecast combination (FC) groups. We discard some models because they perform inadequately or yield more significant inaccuracies compared to other models using similar approaches. Accordingly, we are left with and assess 19 models based on total yield and yield curve area errors. We compare those remaining models based on the NS factors and forecast errors such as TRMSFE (total root mean square forecast error) calculated over the entire yield curve. We utilize error criteria and area errors to assess whether the models outperform random walk (RW). Once compared with RW, models including MLP (and MLP-A), GMDH, RNN (and RNN-A), LSTM, GRU, NNET (and NNET-A), KNN, NNETTS, and FC (Version 1)-EIG2 consistently perform poorly across all forecast horizons and periods, ranking at the bottom according to total error criteria (such as TRMSFE). Besides, the lackluster performance of some recurrent ANNs might be attributed to the large number of parameters that require an estimation. In these models, increasing the number of epochs does not improve forecasts.

According to our further results, models such as ARFIMA, ARIMA, LLAR-LSTAR-SETAR, GRNNTSF, BOOSTING, RF-C, RF-M, and RF-HP (RF-AR in the United States), along with all models of FC (Version 2) and FC (Version 1)-EIG1, FC (Version 1)-SA, FC (Version 1)-TA, FC (Version 1)-WA, FC (Version 1)-INVW, FC (Version 1)-MED, FC (Version 3)-CLS, FC (Version 3)-SA, FC (Version 3)-TA, FC (Version 3)-WA, and FC (Version 3)-INVW display greater inaccuracies with respect to the other models within the same approaches. Additionally, the FC (Version 1)-EIG3 produces exactly the same forecasts as RW. Thus, we exclude these models from further comparisons. Accordingly, we select the alternative epoch/iteration providing the best forecasts among the ELM, ENN, JNN, and NNETAR models while excluding the rest. We conclude that the most successful model compared to the RW in all countries is FC

(Version 1)-CLS. This model, which makes fewer forecast errors compared to RW in most day data, also surpasses RW in the Diebold-Mariano tests (except for the United States sample). The ETS/TBATS and FC (Version 4) models follow the FC (Version 1)-CLS model. Furthermore, the EL-based models in Great Britain and Italy are effective for longer forecast horizons where they outperform RW in all tests.

One should note that while the AR(1) model used by Diebold and Li (2006) is effective for short-term forecasting; using Auto ARIMA might yield even better results. Accordingly, one notable finding of this study is that among all the models under investigation, those incorporating autoregressive algorithms, such as NNETAR, and RF-AR, whether selected or eliminated, tend to perform better compared to the other models in their respective categories. In terms of the forecast horizons, our results show that the FC (Version 3)-MED produces the best forecasts for Nelson-Siegel Model's β_0 factor, which refers to the long term, in both the short- and medium-forecast horizons. In contrast, the FC (Version 1)-CLS and the RW model perform better for long-term forecasts. For the sum of β_0 and β_1, which refers to the short term, the FC (Version 3)-MED gives the most accurate forecasts in short forecast horizons. Lastly, the FC (Version 1)-CLS model is more effective for medium and long forecast horizons as it emphasizes fewer extreme errors.

Looking at the business cycles, in times of economic or financial crisis, any data with RW approach produces a stable forecast which means it is of low variance and close to the average of sample, helping prevent misforecasts from impacting subsequent forecast periods. According to the Expectations Hypothesis, long-term yields also follow RW as long as the short-term yields follow an RW path (Mishkin, 2007). One research on yield curve forecasting by Guidolin and Thornton (2008) demonstrates that the RW method yields better results for long- and short-term horizons. However, in this study, we identify a forecast combination approach that outperforms RW.

In this thesis, we also achieve a forecastable λ factor with minimal error. Accordingly, we do not isolate this factor from its economic and theoretical context, which ensures a strong correlation between the β parameters and latent factors of the yield curve. As a result, our forecasts typically demonstrate only slight deviations from actual data, apart from those caused by the inherent structures of the models. We note that models forecasted for the United States, where the λ factor is not smoothed using the HP filter, tend to perform less effectively with respect to those for other countries. This finding could be attributed to the extreme fluctuations in United States data in more recent periods, which contribute to higher errors in the λ factor across all the models.

This study presents several key findings to the related literature. First, we develop a variable λ approach that offers a better fit for yield data with respect to that with the constant λ method, resulting in a more flexible yield curve. Second, we identify autoregressive models that perform better than the AR(1) model. Third, our ANN models do not consistently outperform the other individual forecasting models. Fourth, we enhance the forecasts from the autoregressive model using ensemble learning-based methods. Finally, we conclude that some FC approaches outperform RW models. Additionally, we observe that using ANN models as inputs improves the forecasts of the FC models.

Results of this study provide several significant conclusions and insights. First of all, suppose that an individual model encounters high forecast errors while the training data does not exhibit extreme errors. In such case, the overall error of the forecast combination model might increase. Yet, if the model consistently produces extreme forecast errors, its weight in forecast combination decreases, leading to improved forecasts. Thus, our results suggest selecting individual models with stable forecasting performance and utilizing them as inputs in forecast combinations in order to benefit from the diversity of models. Once we examine the results of the forecast combinations, models with high and variable errors, such as RNN in FC (Version 1), reduce the efficiency of the forecasts. On the other hand, the FC (Version 1)-CLS model, which excludes these models makes better forecasts. Moreover, the ANN models show significant improvements in forecast accuracy, particularly in shorter horizons for Version 2 of FC and across all horizons for Version 3 of FC.

Next, one should theoretically expect that the Vector Autoregression (VAR) model yield the best results because of the relationship between the Nelson-Siegel (NS) factors. However, in this thesis, we find that the weak relationship between these NS factors hinders the effectiveness of the VAR model. Furthermore, the VAR model requires estimating more parameters than autoregressive models where it seems that the relationship between the NS factors is not sufficiently robust for optimal parameter estimation. Additionally, we assert that increasing the number of epochs or iterations in ANN models has an inconsistent impact on forecasting. In this study, the models developed with varying epochs/iterations demonstrate that increasing these numbers does not necessarily improve the forecasting performance. Relatedly, the variability in outcomes across different countries, along with the lack of a theoretical foundation for this choice, could be defined as a limitation of this study.

One should further note that models of the yield curve constructed from simple data may struggle to fully incorporate interest rate dynamics. That is although the term structure of interest rates is expected to accurately reflect the rational expectations of economic agents, numerous economic and international factors influence the yield curve dynamics. Thus, in a further research, we can explore models such as the Nonlinear Autoregressive Exogenous (NARX) model which uses exogenous data. Besides, within the context of ANNs, we can also enhance network complexity, utilize more lagged values, and improve the training data as future work. Note that, ANNs are designed as forecast-oriented structures, and many studies have shown that they can identify patterns. They forecast bending over long horizons more effectively compared to models such as Box-Jenkins, which tend to converge to a constant over time in shorter horizons. In this study, ANNs generally underperform other models achieving this capability even if we can see this effectiveness occasionally. This study further shows that some models which forecast factors poorly can produce good yield curve forecasts. This is because of the fact that these factors can cancel each other's forecasting errors due to their connections within a model. Such connections specifically arise between β_0 and β_1 of Nelson-Siegel, where the errors from the two datasets can eliminate each other's total error within a model.

Last but not least, in this study, the Extreme Learning Machine (ELM) model performs better than other ANNs. This result suggests that it might be useful to create algorithms that blend different methods instead of adhering to traditional ANN structures, such as specific neurons and linear structure found in the ELM model. Besides, in a future framework we may develop hybrid models that can simultaneously estimate the parameters from multiple models within one model. Moreover, in future research, we may also check ANN structures that combine different types of neurons and functions within their design.

Overall, the results of this study show that although the theoretical framework of the models utilized to forecast yield curves is robust, they rarely demonstrate statistical success across our entire sample. Most models tend to achieve partial success in certain forecast periods or horizons. Despite the failures of most individual models, it can be argued that some of them may still be useful thanks to their statistical information, which can enhance the efficiency of other models when combined. We can harness this utility by combining forecasts or incorporating various components, such as the forecast horizon, period, and the impact of different factors on the overall success of the forecasts. To sum up, this thesis briefly demonstrates that better forecasts of yield curves could be obtained by employing forecast combinations, rather than using conventional and non-conventional individual models. Furthermore, including the λ parameter of the NS factors among the factors to be forecasted rather than taking it as constant, improves the forecast performance.

BIBLIOGRAPHY

Abar, H. (2022). BIST 100 Endeksi İçin Fiyat Öngörüsü: ARIMA ve LSTM [Price Forecast for BIST 100 Index: ARIMA and LSTM]. Ö. F. Rençber (Ed.), *Veri Madenciliğinde Kullanılan Regresyon Modelleri ve R ile Uygulamalı Örnekler* [Regression Models Used in Data Mining and Applied Examples with R]. 173-194. İstanbul: Nobel Akademik Yayıncılık.

Abid, F., and Salah, M. B. (2003). Estimating The Term Structure of Interest Rates: Neural Network vs One Factor Parametric Models. *Transactions on Information and Communications Technologie, 29,* 421-428.

Adhikari, R. and Agrawal, R. K. (2013b). Hybridization of Artificial Neural Network and Particle Swarm Optimization Methods For Time Series Forecasting - - *International Journal of Applied Evolutionary Computation, 4(3)*, 75-90. doi: 10.4018/jaec.2013070107.

Adhikari, R.and Agrawal, R. K. (2013a). A Linear Hybrid Methodology For Improving Accuracy of Time Series Forecasting. *Neural Computing and Applications, 25(2)*. doi:10.1007/s00521-013-1480-1.

Akçay, M. B., Kasap, M., Doğuç, T., and Kasap, G. (2012). *Türev Piyasalar Ve Yapılandırılmış Ürünler Finans Mühendisliği Ve Risk Yönetimi Perspektifiyle* [Derivative Markets and Structured Products from a Financial Engineering and Risk Management Perspective]. İstanbul: Scala Yayıncılık. ISBN:9786054650002.

Aljinovic, Z., Poklepovic, T., and Katalinic, K. (2012). Best Fit Model for Yield Curve Estimation. *Croatian Operational Research Review, 3,* 28–40.

Alpaydın, E. (2010). *Introduction to Machine Learning (2nd ed.)*. London: The MIT Press.

Andrawis, R. R., Atiya, A. F., and El-Shishiny, H. (2011). Combination of Long Term and Short Term Forecasts, with Applicatıon to Tourism Demand Forecasting. *International Journal of Forecasting, 27,* 870–886.

Annaert, J., Claes, A., De Ceuster, M. and Zhang, H. (2013). Estimating The Spot Rate Curve Using The Nelson-Siegel Model: A Ridge Regression Approach. *International Review of Economics and Finance, 27*, 482-496.

Araújo, R. C. D. and Cajueiro, D. O. (2014). Combining Term Structure of Interest Rate Forecasts: The Brazilian Case, *EconomiA, 14(2),* 102–121. http://dx.doi.org/10.1016/j.econ.2013.08.007.

Arbia, G. and Di Marcantonio, M. (2015). Forecasting Interest Rates Using Geostatistical Techniques. *Econometrics, 3(4)*, 733 - 760.

Atalay, M. and Çelik, E. (2017). Büyük Veri Analizinde Yapay Zekâ Ve Makine Öğrenmesi Uygulamaları (Applications of Artificial Intelligence and Machine Learning in Big Data Analysis). *Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 9(22),* 155-172. doi: 10.20875/makusobed.309727.

Baily, D. and Thompson, D.M. (1990). Developing Neural Network Applications. *Al Expert, 5(6),* 38-47.

Bajracharya, Dinesh (2010). *Econometric Modeling vs Artificial Neural Networks – A Sales Forecasting Comparison*. Master Thesis, University of Borås, Borås.

Balaban, M. E. and Kartal, E. (2018). *Veri Madenciliği ve Makine Öğrenmesi Temel Algoritmaları ve R Dili ile Uygulamaları (2nd ed.)* [Data Mining and Machine Learning Basic Algorithms and Applications with R Language]. İstanbul: Çağlayan Kitabevi. ISBN:9789754361353.

Bates, J. M. and Granger, C. W. J. (1969). The Combination of Forecasts. *Operations Research Quarterly, 20*, 451-468.

Bergstra, J. and Bengio, Y. (2012). Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research 13,* 281-305.

Björk, T. and Christensen, B.J. (1999). Interest Rate Dynamics and Consistent Forward Rate Curves. *Mathematical Finance, 9,* 323-348. https://doi.org/10.1111/1467-9965.00072.

Bliss, R. R. (1997). Testing Term Structure Estimation Methods. *Advances in Futures and Options Research, 9,* 197–231.

Box, G. E. P., and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control* (2nd ed.). San Francisco: Holden-Day.

Brock,W. A. and Durlauf, S. N. (2001). Growth Empirics and Reality. *World Bank Economic Review 15(2),* 229–272.

Burg, J. P. (1975). *Maximum entropy spectral analysis.* Ph.D Thesis, Stanford University, California.

Caldeira, J. F., Moura, G. V., Santos, A. A. P., and Tourrucôo, F. (2016). Nelson–Siegel Model: Brazilian Evidence. *EconomiA, 17,* 221–237.

Campbell, J. Y., and Shiller, R.J. (1991). Yield Spreads and Interest Rate Movements: A Bird's Eye View. *Review of Economic Studies, 58(3),* 495–514. https://doi.org/10.2307/2298008.

Campbell, J., and Shiller, R. (1987). Cointegration and Tests of Present Value Models. *Journal of Political Economy, 95(5),* 1062–1088. https://doi.org/10.1086/261502.

Campolieti, G., & Makarov, R.N. (2014). *Financial Mathematics: A Comprehensive Treatment*. New York: Chapman and Hall/CRC. https://doi.org/10.1201/9781315373768.

Castellani, M., and Santo, E. A. D. (2006). Forecasting Long-Term Government Bond Yields: an Application of Statistical and AI Models. *Economics and Management Institute Working Papers*.

Chen, K., Chen, K., Wang, Q., He, Z., Hu J., and He, J.(2019). Short-Term Load Forecasting With Deep Residual Networks. *IEEE Transactions on Smart Grid, 10(4)*, 3943-3952. doi: 10.1109/TSG.2018.2844307.

Chen, T. and Guestrin, C. (2016). Xgboost: A Scalable Tree Boosting System. *In Proceedings of the 22nd ACM SIGKDD International Conferance on Knowledge Discovery and Data Mining*. 785-794. California.

Claeskens, G., Magnus, J., Vasnev, A., and Wang, W. (2014). The Forecast Combination Puzzle: A Simple Theoretical Explanation - *Tinbergen Institute Discussion Paper*, *127(3).*

Cox, J., Ingersoll, J., and Ross, S. (1985). A Theory of The Term Structure of Interest Rates. *Econometrica, 53*, 385-407.

Cybenko, G. (1988). Continuous Valuet Neural Networks with Two Hidden Layers Are Sufficient, *Tecnical Report*, *Tuft University*.

Cybenko, G. (1989). Approximation by Superpositions of a Sigmoidal Function. *Mathematics of Control Signals Systems*, *2*. 303-314.

Çepni, O., Güney, İ. E., Küçüksaraç, D. and Yılmaz, M. H. (2018). The Interaction between Yield Curve and Macroeconomic Factors. *Research Notes in Economics, 2018-02*. Ankara: The Central Bank of The Republic of Türkiye.

Dauwe, A. and Moura, M. L. (2011). Forecasting The Term Structure of The Euro Market Using Principal Component Analysis. *Insper Working Paper* (WPE: 233/2011), Kortrijk: Institute of Education and Research.

Davidson, M. (2014). *Quantıtatıve Fınance A Sımulatıon-Based Introductıon Usıng Excel*. Chapman and Hall/CRC Press. https://doi.org/10.1201/b16039.

De Livera, A. M., Hyndman, R. J., and Snyder, R. D. (2011). Forecasting Time Series with Complex Seasonal Patterns Using Exponential Smoothing, *Journal of the American Statistical Association, 106(496)*, 1513-1527.

De Pooter, M. (2007). Examining The Nelson-Siegel Class Of Term Structure Models: In-Sample Fit Versus Out-Of-Sample Forecasting Performance. *SSRN Electronic Journal*.

Diament, P. (1993). Semi-Empirical Smooth Fit to The Treasury Yield Curve, *Journal of Fixed Income, 2*, 55-70.

Diebold, F. X., and Li, C. (2006). Forecasting the Term Structure of Government Bond Yields. *Journal of Econometrics, 130*, 337-364.

Diebold, F. X., and Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics, 13*, 253-263.

Diebold, F. X., and Rudebusch, G. D. (2013). *Yıeld Curve Modelıng And Forecastıng: The Dynamıc Nelson-Sıegel Approach* (1th ed.). United Kingdom: Princeton University Press Princeton and Oxford. ISBN: 9780691146805.

Diebold, F. X., Rudebusch, G. D., Aruoba., S. B. (2006). The Macroeconomy and The Yield Curve: A Dynamic Latent Factor Approach. *Journal of Econometrics, 131*, 309–338.

Dolan, C. P. (1999). Forecasting the Yield Curve Shape: Evidence in Global Markets. *The Journal of Fixed Income*, 92-99.

Duffee, G. R. (2012). Forecasting interest rates, *Economics Working Paper Archive,599*. The Johns Hopkins University, Department of Economics.

Duffie, D., and Kan, R. (1996). A Yield Factor Model of Interest Rates. *Mathematical Finance, 6*, 379-406.

Dufour, J. M., Garcia, R., and Taamouti, A. (2012). Measuring High Frequency Causality between Returns, Realized Volatility, and Implied Volatility. *Journal of Financial Econometrics, 10(1)*, 124-163.

Eğrioğlu, E., Aladağ, Ç. H., and Günay, S. (2008). A New Architecture Selecting Strategy Solving Seasonal Autoregressive Time Series by Artificial Neural Networks. *Hacettepe Journal of Mathematics and Statistics, 37(2)*, 185-200.

Eğrioğlu, E., and Baş, E. (2020). *Zaman Serileri Ve Öngörü Yöntemleri (R Uygulamalı)* [Time Series and Forecasting Methods (R Applied)]. Ankara: Nobel Akademik Yayıncılık. ISBN: 9786254390128.

Eğrioğlu, E., Yolcu, U., and Baş, E. (2020). *Yapay Sinir Ağları - Öngörü Ve Tahmin Uygulamaları* (2nd ed.) [Artificial Neural Networks - Forecasting and Forecasting Applications]. Ankara: Nobel Akademik Yayıncılık. ISBN: 9786050331530.

Eklind, A. (2020). *Deep Neural Networks to Ensure The Quality of Calculated Yield Curves In Banking*. Master Thesis, KTH Royal Instıtute Of Technology, Stockholm.

Ersoy, O. (January 1990). *Tutorial at Hawaii International Conference on Systems Sciences,* Hawaii.

Fabozzi, F. J. (2002). *Interest Rate, Term Strucuture, and Valuation Modeling* . New Jersey: John Wiley & Sons, Inc.

Fabozzi, F. J. (2006). *Bond Markets, Analysis and Strategies (6th ed.)*. New Jersey: Pearson-Prentice Hall.

Fama, E. F. (1984). Forward and spot exchange rates. *Journal of Monetary Economics, 14(3),* 319-338. https://doi.org/10.1016/0304-3932(84)90046-1.

Fama, E. F., and Bliss, R. R. (1987). The Information in Long-Maturity Forward Rates. *American Economic Review, 77*, 680–692.

Fischer, T. and Krauss, C. (2017). Deep Learning with Long Short-Term Memory Networks for Financial Market Predictions. *Friedrich-Alexander-Universität Discussion Papers in Economics, No. 11*.

Fisher, M., Nychka, D., Zervos, D. (1995). Fitting the Term Structure of Interest Rates with Smoothing Splines. *Board of Governors of the Federal Reserve System, Finance and Economics Discussion Series, 1*.

Frederick, J. and Herzog, J. O. (2012). Forecasting Performance of Three-Factor Nelson-Siegel Models - Methods in Modeling and Forecasting The Term Structure. *BBVA Research, Economic Analysis.*

Friedman, M. (1977). Time Perspective in Demand for Money. *Unpublished Paper*, University of Chicago.

Gajowniczek, K. and Zabkowski, T. (2017). Two-Stage Electricity Demand Modeling Using Machine Learning Algorithms. *Energies, 10.* doi:10.3390/en10101547.

Genre, V., Kenny, G., Meyler, A., and Timmermann, A. (2013). Combining Expert Forecasts: Can Anything Beat The Simple Average?. *International Journal of Forecasting, 29(1),* 108-121. https://doi.org/10.1016/j.ijforecast.2012.06.004.

Gerhart, C., Lütkebohmert, E., and Weber, M. (2018). Robust Forecasting of Multiple Yield Curves. O. Valenzuela, F. Rojas, H. Pomares, and I. Rojas (Ed.). *Theory and Applications of Time Series Analysis Selected Contributions from ITISE 2018* (p.187-202). Granada: Springer Nature Switzerland AG.

Ghatak, A. (2017). *Machine Learning With R.* Kolkata: Springer Nature Singapore Pte Ltd.. doi:10.1007/978-981-10-6808-9.

Gheyas, I. A. and Smith L. S. (July 2009). A Neural Network Approach to Time Series Forecasting. London: *Proceedings of the World Congress on Engineering 2009, Vol II.*

Gibson, R., Lhabitant, F., and Talay, D. (2010). Modeling The Term Structure of Interest Rates: A review of The Literature. *Foundations and Trends in Finance, 5*, 1-156. http://dx.doi.org/10.1561/0500000032.

Gill, P. E. and Leonard, M. W. (2001). Reduced-Hessian Quasi-Newton Methods For Unconstrained Optimization. *SIAM Journal of Optimization, 12(1),* 209-237.

Gilli, M., Große, S., and Schumann, E. (2010). Calibrating the Nelson-Siegel-Svensson Model. *COMISEF Working Papers Series*.

Goodfellow, I., Bengio, Y., and Courville A. (2016). *Deep Learning*, MIT Press.

Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., and Schmidhuber, J. (2015). LSTM: A Search Space Odyssey. *IEEE Transactions on Neural Networks and Learning Systems, 28(10).* doi:10.1109/TNNLS.2016.2582924.

Grinstead, C. M. and Snell, J. L. (1997). Introduction to Probability. *American Mathematical Society*, p. 471.

Grmanová, G., Laurinec, P., Rozinajová, V., Ezzeddine, A. B., Lucká, M., Lacko, P. et. al. (2016). Incremental Ensemble Learning For Electricity Load Forecasting. *Acta Polytechnica Hungarica, 13(2)*, 97-117.

Guidolin, M., and Thornton, D. L. (2008). Predictions of Short-Term Rates and The Expectations Hypothesis of The Term Structure of Interest Rates. *European Central Bank Working Paper Series, 977*.

Günay, S., Eğrioğlu, E., and Aladağ, Ç. H. (2007). *Tek Değişkenli Zaman Serileri Analizine Giriş* [Introduction to Univariate Time Series Analysis]. Ankara: Hacettepe Üniversitesi Yayınları. ISBN: 9789754912425.

Gürkanyak, R. S., Sack, B., and Wright, J. H. (2007). The U. S. Treasury Yield Curve: 1961 to the Present. *Journal of Monetary Economics, 54*, 2291-2304. https://doi.org/10.1016/j.jmoneco.2007.06.029.

Gürsakal, N. (2017). *Makine Öğrenmesi ve Derin Öğrenme* [Machine Learning and Deep Learning]. Bursa: Dora Yayınları. ISBN: 9789752447486

Hansson, M. (June 2017). *On Stock Return Prediction with LSTM Networks*. Master Thesis, Lund University, Lund.

Hays, S. E., Shen, H., and Huang, J. (2012). Functional Dynamic Factor Models with Application to Yield Curve Forecasting. *Annals of Applied Statistics, 6(3),* 870-894. doi:10.1214/12-AOAS551.

He, C. (2013). *Forecastıng The Yıeld Curve Of Government Bonds: A Comparatıve Study*. Master Thesis, Brock University, Ontario.

Heath, D., Jarrow, R., and Morton, A. (1992). Bond Pricing and The Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation. *Econometrica, 60,* 77-105.

Hecth-Nielsen, R. (1990). *Neurocomputing*, Menlo Park, CA: Addison-Wiley.

Hochreiter, S. and Schmidhuber, J. (1997) Long Short-Term Memory. *Neural Computation*, *9(8),* 1735-1780. doi: 10.1162/neco.1997.9.8.1735.

Hoogteijling, T. (2020). *Forecasting Bond Risk Premia With Machine Learning.* Master Thesis, Erasmus University, Rotterdam.

Hornik K. (1991). Approximation Capabilities of Multilayer Feedforward Networks. *Neural Networks, 4*, 251-257.

Hornik, K., Stinchcombe, M., and White, M. (1989). Multilayer Feedforward Networks are Universal Approximators. *Neural Networks*, *2*, 359-366.

Hothorn, T., Bühlmann, P., Kneib, T. & Schmid, M., and Hofner, B. (2010). mboost: Model-Based Boosting. R package. https://cran.r-project.org/web/packages/mboost/mboost.pdf.

Hsiao, C. and Wan, S. K. (2014). Is There an Optimal Forecast Combination?. *Journal of Econometrics Volume 178(2),* 294-309. https://doi.org/10.1016/j.jeconom.2013.11.003.

Huang, G. B., Zhou, H., Ding, X., and Zhang, R. (2012). Extreme Learning Machine for Regression and Multiclass Classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 42(2)*, 513-529.

Huang, G. B., Zhu, Q. Y., and Siew, C. K. (2006). Extreme Learning Machine: Theory and Applications. *Neurocomputing, 70,* 489–501. doi:10.1016/j.neucom.2005.12.126.

Huang, G.-B., Zhu, Q. Y., and Siew, C. K. (25-29 Jul., 2004). Extreme Learning Machine: A New Learning Scheme of Feedforward Neural Networks. Budapest: *Proceedings of International Joint Conference on Neural Networks (IJCNN2004), 2*, 985–990.

Huang, H. and Lee, T. H. (2007). To Combine Forecasts or to Combine Information?. *Econometric Reviews, 29*, 534-570. 10.1080/07474938.2010.481553.

Hull, J. C. (2018). *Optıons, Futures, And Other Derıvatıves* (10th ed.). United States – New York: Pearson Education Inc.. ISBN: 9780134472089.

Hull, J., and White, A. (1990). Pricing Interest Rate Derivative Securities. *Review of Financial Studies, 3*, 573-592.

Hutchinson, J. M., Lo, A. W., and Poggio., T. (1994). A nonparametric approach to pricing and hedging derivative securities via learning networks. *Journal of Finance, 49(3)*, 851-889.

Hyndman, R. J. and Khandakar, Y. (2008). Automatic Time Series Forecasting: The Forecast Package For R. *Journal of Statistical Software, 27(3).* doi:10.18637/jss.v000.i00.

Hyndman, R. J. and Koehler, A. B. (2006). Another Look at Measures of Forecast Accuracy. *International Journal of Forecasting, 22(4)*, 679-688.

Hyndman, R. J., Ahmed, R. A., and Athanasopoulos, G. (2007). Optimal Combination Forecasts For Hierarchical Time Series. *Working Paper 09/07, Department of Econometrics and Business Statistics, Monash University*.

Hyndman, R.J., and Athanasopoulos, G. (2021). *Forecasting: Principles And Practice* (3rd ed). Australia: OTexts.com/fpp3. Accessed on 10.07.2024.

Ishii, H. (2019). Forecasting Term Structure of Interest Rates In Japan. *International Journal of Financial Studies, 7(3)*. doi:10.3390/ijfs7030039.

Ivakhnenko, A. G. (1971). Polynomial Theory of Complex Systems. *Systems, Man and Cybernetics, IEEE Transactions, 4,* 364-378.

İslamoğlu, E. (2020). *Modern Zaman Serileri ve Yöntemleri* [Modern Time Series and Methods]. Ankara: Nobel Akademik Yayıncılık -ISBN: 9786254062698.

Jacovides, A. (2008). *Forecasting Interest Rates from The Term Structure: Support Vector Machines vs Neural Network.* Master Thesis, University of Nottingham, Nottingham.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2023). *An Introduction to Statistical Learning (2nd ed.)* [PDF]. Springer.

Jong, F. D. (2000). Time Series and Cross-Section Information in Affine Term-Structure Models. *Journal of Business and Economic Statistics*, *18(3),* 300-314. https://doi.org/10.2307/1392263.

Jore, A. S., Mitchell, J., and Vahey, S. P. (2008). Combining Forecast Densities From VARs with Uncertain Instabilities. *Journal of Applied Econometrics*. doi:10.2307/40607049

Joslin, S., Priebsch, M., and Singleton, K. J. (2010). Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks. *MIT and Stanford University, Manuscript*.

Joslin, S., Singleton, K. J., and Zhu, H. (2011). A New Perspective on Gaussian Dynamic Term Structure Models. *Review of Financial Studies, 24,* 926–970.

Kaastra, I., and Boyd, M. (1996). Designing A Neural Network For Forecasting Financial and Economic Time Series. *Neurocomputing*, *10,* 215—236.

Katz, J.O. (1992). Developing Neural Network Forecasters for Trading. *Technical Analysis of Stocks and Commodities 10(4),* 58-70.

Klimasauskas, C. C. (1993). Applying Neural Networks. R.R. Trippi and E. Turban (ed.). *Neural Networks in Finance and Investing: Using Artificial Intelligence to Improve Real World Performance,* 64-65.

Klimasauskus, C. C. (1991). Applying Neural Networks, Part 3: Training A Neural Network, *PC-AI, May-June,* 20-24.

Kong, S. (1991). *An Investigation of The Use of Feedforward Neural Networks for Forecasting*. Ph. D. Thesis, Kent State University, Ohio.

Koopman, S. J., Mallee, M. I. P., and van der Wel, M. (2010). Analyzing The Term Structure of Interest Rates Using The Dynamic Nelson-Siegel Model with Time-Varying Parameters. *Journal of Business & Economic Statistics, 28(3),* 329-343.

Kostyra, T. P., and Rubaszek, M. (2020). Forecasting The Yield Curve For Poland. *Econometric Research in Finance, 5(2),* p.103-117. doi: 10.2478/erfin-2020-0006.

Kožíšek, J. (2018). *Forecasting Term Structure Of Government Bonds Using High Frequency Data.* Master Thesis, Charles University, Prague.

Kuru, Ö. (2022). Yapay Sinir Ağları [Artificial Neural Networks]. Ö. F. Rençber (Ed.), *Veri Madenciliğinde Kullanılan Regresyon Modelleri ve R ile Uygulamalı Örnekler* [Regression Models Used in Data Mining and Applied Examples with R]. 153-171. İstanbul: Nobel Akademik Yayıncılık.

Kuswanto, H. and Sibbertsen, P. (2008). A Study on "Spurious Long Memory in Nonlinear Time Series Models". *Applied Mathematical Sciences, 2(55),* 2713-2734.

Lachtermacher, G. and Fuller, J. D. (1995). Backpropagation in Time Series Forecasting. *Journal of Forecasting, 14,* 381-393.

Lapades, A. and Faber, R. (1988). How Neural Nets Work. D. Z. Andersan (Ed.). *Neural Information Processing Systems,* 442-456.

Larochelle, H., Erhan, D., Courville, A., Bergstra, J., and Bengio, Y. (2007). An Empirical Evaluation of Deep Architectures on Problems with Many Factors of

Variation. Z. Ghahramani (Ed.), *Proceedings of the Twenty-fourth International Conference on Machine Learning (ICML'07)*, 473–480.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep Learning. *Nature, 521*, 436–444.

León, A., Rubia, A., and Sanchis-Marco, L. (2018). On Multicollinearity and The Value of The Shape Parameter in The Term Structure Nelson-Siegel Model - *The IEB International Journal of Finance, 16*, 8-29. doi: 10.5605/IEB.16.1

Lewis, N. D. C. (2015). *Build Your Own Neural Network Today!*. CreateSpace Independent Publishing Platform. ISBN: 9781519101235.

Lewis, N. D. C. (2016). *Deep Learning Made Easy With R: A Gentle Introduction For Data Science*. CreateSpace Independent Publishing Platform. ISBN: 9781519514219.

Lewis, N. D. C. (2017a). *Neural Networks for Time Series Forecasting with R: An Intuitive Step By Step Blueprint for Beginners*. CreateSpace Independent Publishing Platform. ISBN: 9781544752952.

Lewis, N. D. C. (2017b). *Machine Learning Made Easy with R: An Intuitive Step by Step Blueprint for Beginners*. CreateSpace Independent Publishing Platform. ISBN13: 9781546483755.

Li, H., Chen, P. C. L., and Huang, H. P. (2001). *Fuzzy Neural Intelligent Systems: Mathematical Foundation and the Applications in Engineering*. Boca Raton: CRC Press.

Lippman, R. P. (1987). An Introduction to Computing with Neural Nets, *IEEE ASSP Magazine, 4*, 4-22.

Litterman, R., and Scheinkman, J. (1991). Common Factors Affecting Bond Returns. *Journal of Fixed Income, 1*, 54-61. http://dx.doi.org/10.3905/jfi.1991.692347.

Macaulay, F. (1938). Some Theoretical Problems Suggested by the Movements of Interest Rates, Bond Yields, and Stock Prices in the United States Since 1865. *National Bureau of Economic Research.*

Mankiw, N. G. and Summers, L. H. (1984). Do Long-Term Interest Rates Overreact to Short-Term Interest Rates?. *Brookings Papers on Economic Activity, 15(1),* 223-248.

Marek, J. (2015). *The Nelson-Siegel Model: Present Application and Alternative Lambda Determination.* Bachelor Thesis, Charles University, Prague.

Masters, T. (1993). *Practical Neural Network Recipes in C++*. New York: Academic Press.

McCulloch, J. H. (1971). Measuring the Term Structure of Interest Rates. *Journal of Business, 44*, 19–31.

McCulloch, J. H. (1975). The Tax Adjusted Yield Curve. *Journal of Finance, 30*, 811–830.

Mineo, E., Alencar, A.P., Moura, M., and Fabris, A. E. (2020). Forecasting The Term Structure of Interest Rates with Dynamic Constrained Smoothing B-Splines. *Journal of Risk and Financial Management, 13.* doi:10.3390/jrfm13040065.

Mishkin, F. S. (1986). *The Economics of Money, Banking, and Financial Markets*. Boston: Little Brown and Co..

Mishkin, F. S. (2007). *The Economics of Money, Banking, and Financial Markets* (8th ed.). Boston: Addison Wesley Co..

Mishkin, F. S. (2021). *The Economics of Money, Banking, and Financial Markets Global Edition* (13th ed.). Essex: Pearson Education. ISBN: 9780136894353.

Mitchell, T. M. (2006). The Discıpline of Machine Learning. *Machine Learning Department, Pittsburgh: Carnegie Mellon University.*

Modena, M. (2008). The Term Structure and The Expectations Hypothesis: A Threshold Model. *MPRA Paper*, *9611*.

Molenaars, T. K., Reinerink, N. H., and Hemminga, M. A. (2015). Forecasting The Yield Curve: Art or Science?. *Magazine De Actuaris, 22(4)*, 38 - 40.

Nam, K. and Schafer, T. (1995). Forecasting International Airline Passenger Traffic Using Neural Network Logistics and Transportation. *The Logistics and Transportation Review, 31 (3)*, 239-251.

Namın, S. S. and Namın, A. S. (2018). Forecasting Economic and Financial Time Series: ARIMA vs. LSTM. *ArXiv,* abs/1803.06386.

Nelson, C. R., and Siegel, A. F. (1987). Parsimonious Modeling of Yield Curves. *Journal of Business*, *60*, 473-489.

Öztemel, E. (2016). *Yapay Sinir Ağları (4th ed.)* [Artificial Neural Networks]. İstanbul: Papatya Yayıncılık Eğitim. ISBN: 9789756797396.

Pearson, K. (1906). *A Mathematical Theory of Random Migration*. London: Dulau & Co.

Pesaran, M. H. and Pick, A. (2011). Forecast Combination Across Estimation Windows. *Journal of Business & Economic Statistics*, *29*, 307 – 318.

Pode, D. L. and Mackworth, A. K. (2010). *Artificial Intelligence Foundations of Computational Agents*. Newyork: Cambiridge University Press. ISBN: 9780521519007.

Poole, D. L., and Mackworth, A. K. (2023). *Artificial Intelligence: Foundations of Computational Agents* (3rd ed.). Cambridge: Cambridge University Press. https://doi.org/10.1017/9781009258227.

Povala, P., and Vasıl, R. (2017). Forecasting The Term Structure of Interest Rates In Slovakia. *Economic analysis*, *42.*

Rao, C.R. and Mitra, S.K. (1971). *Generalized Inverse of Matrices and Its Applications*. New York: Wiley.

Raviv, E. (2016). *Forecast Combinations in R Using The ForecastCombinations Package a Manual*. eranraviv.com. Accessed on 10.01.2024.

Reinicke, S. (2019) *Modeling and Forecasting Yield Curves A Comparison of Published Methods.* Master Thesis, Ludwig-Maximilians-Universität, München

Reschenhofer, E., & Stark, T. (2019). Forecasting the Yield Curve with Dynamic Factors. *Romanian Journal of Economic Forecasting*, *22(1)*, 101-113.

Rezende, R. B. D., and Ferreira, M. S. (2011). Modeling and Forecasting The Yield Curve by an Extended Nelson-Siegel Class of Models: A Quantile Autoregression Approach. *Journal of Forecasting.* http://dx.doi.org/10.2139/ssrn.1290741.

Rodriguez, C. (2016). *Estimating and Forecasting The Term Structure of Interest Rates: US and Colombia Analysis*. Master Thesis, Universidad del Rosario, Bogota.

Rokach, L. and Maimon, O. (2005). Decision Trees. L. Rokach, and O. Maimon (Ed.), *Data Mining and Knowledge Discovery Handbook*, 165-192. Boston: Springer.

Rosadi, D., Nugraha, Y. A. and Dewi, R. K. (2011). Forecasting The Indonesian Government Securities Yield Curve Using Neural Networks and Vector Autoregressive Model. *Bank for International Settlments Events*. Dublin.

Rostan, P., Belhachemi, R., and Racicot, F. E. (2017). Yield Curve Forecasting with The Burg Model. *Journal of Forecasting, 36*, 91–99. doi: 10.1002/for.2416.

Sagheer, A., Hamdoun, H., and Youness, H. (2021). Deep LSTM-Based Transfer Learning Approach for Coherent Forecasts in Hierarchical Time Series. *Sensors 2021, 21*. https://doi.org/10.3390/s21134379.

Sambasivan, R. and Das, S. (2017). A Statistical Machine Learning Approach to Yield Curve Forecasting. *arXiv preprint arXiv:1703.01536*.

Sharda, R. and Potil, R. B. (1992). Connectionist Approach to Time Series Prediction: An Emprical Test. *Journal of Intelligent Manifacturing, 3*, 317-323.

Shiller, R. J., Campbell, J. Y., and Schoenholtz, K. L. (1983). Forward Rates and Future Policy: Interpreting the Term Structure of Interest Rates. *Brookings Papers on Economic Activity, 14(1),* 173-224.

Skalin, J. and Teräsvirta, T. (1999). Another Look at Swedish Business Cycles, 1861-1988. *Journal of Applied Econometrics, 16(4),* 359–378.

Steel, M. F. J. (2019). Model Averaging and Its Use In Economics. *MPRA Paper, 91970*.

Stefani, J. D., Caelen, O., Hattab, D., and Bontempi, G. (September 2017). Machine Learning for Multi-Step Ahead Forecasting of Volatility Proxies. *MIDAS 2017 - 2nd Workshop on Mining Data for Financial Applications.*

Suthar, B., Patel, H., and Goswami, A. (2012). A Survey: Classification of Imputation Methods in Data Mining, *International Journal of Emerging Technology and Advanced Engineering, 2(1)*, 309-312.

Svensson, L. E. O. (1994). Estimating and Interpreting Forward Interest Rates: Sweden 1992-1994. *NBER Working Paper Series, 4871*.

Svensson, L. E. O. (1995). Estimating Forward Interest Rates with the Extended Nelson & Siegel Method, *Quarterly Review, 3*, 13-26.

Szenczı, R. (2016). *Forecasting The Hungarian Term Structure Of Interest Rates With Econometrics And Neural Network Methods*. Master Thesis, Corvınus University, Budapest.

Şişman, M. Ö. (2011). *Faiz Oranları Getiri Eğrisi Simülasyonu Yöntemleri Ve Bankacılıkta Aktif Pasif Yönetimi Üzerine Etkileri: Türkiye'de Ticari Bankalar Üzerine Bir Uygulama* [Interest Rates Yield Curve Simulation Methods and Their Effects on Asset Liability Management in Banking: An Application on Commercial Banks in Turkey]. Ankara: BDDK Kitapları. ISBN:9789759004118.

Taieb, S. B., Bontempi, G., Atiya, A., and Sorjamaa, A. (2012). A Review and Comparison of Strategies for Multi-Step Ahead Time Series Forecasting Based on

The NN5 Forecasting Competition. *Expert Systems with Applications*, *39(8)*, 7067–7083. https://doi.org/10.1016/j.eswa.2012.01.039.

Tang, Z. and Fishwick, P. A. (1993). Feedforward Neutral Nets as Models for Time Series Forecasting, *Operation Research Society of America*, *5(4)*, 374-385.

Täppinen, J. (1998). Interest Rate Forecasting with Neural Networks. *Government Institute for Economic Research*, *Working paper*, *170*.

Taušer, J., and Buryan, P. (2011). Exchange Rate Predictions in International Financial Management by Enhanced GMDH Algorithm. *Prague Economic Papers*, *20.3*, 232-249.

Tealab, A., Hefny, H., and Badr, A. (2017). Forecasting of Nonlinear Time Series Using ANN. *Future Computing and Informatics Journal*, *2*, 39-47.

The Central Bank of The Republic of Türkiye. (2019). *Inflation Report* (2019-II). Ankara: The Central Bank of The Republic of Türkiye.

Tong, H. (1975). Autoregressive Model Fitting with Noisy Data by Akaike's Information Criterion. *IEEE Transactions on Information Theory*, *21(4)*, 476-480. doi:10.1109/TIT.1975.1055402.

Vasicek, O. (1977). An Equilibrium Characterization of The Term Structure. *Journal of Financial Economics*, *5*, 177-188.

Vasicek, O.A., Fong, H.G. (1982). Term Structure Modeling Using Exponential Splines. *Journal of Finance*, 37, 339–348.

Vela, D. (2013). Forecasting Latin-American Yield Curves: An Artificial Neural Network Approach. *Borradores de Economia*, *761*.

Willner, R. (1996). A New Tool for Portfolio Managers: Level, Slope and Curvature Durations. *Journal of Fixed Income*, *June*, 48–59.

Wong, F. S. (1991). Time Series Forecasting Using Backpropagation Neural Networks. *Neurocomputing*, *2*, 147-159.

Wood, D., and Dasgupta, B. (1995). Modeling The Term Structure Of Sterling Interbank Interest Rates. *In the Proceedings of the Third International Conference on Neural Network in the Capital Mark*ets. Eds Refenes & Apostolos, John Wiley & Sons, 6-49.

Yakut, E., and Kuru, Ö. (2022). XGBoost Yöntemi [XGBoost Method]. Ö. F. Rençber (Ed.), *Veri Madenciliğinde Kullanılan Regresyon Modelleri ve R ile Uygulamalı Örnekler* [Regression Models Used in Data Mining and Applied Examples with R]. 103-126. İstanbul: Nobel Akademik Yayıncılık.

Zhang, G., Potuwa, B. E., Hu, Y. M. (1998). Forecasting with Artificial Neural Networks, The State of Art. *International Journal of Forecasting, 14,* 35-62.

Zhu, W., Miao, J., and Qing, L. (2015). Constrained Extreme Learning Machines: A Study on Classification Cases. *arXiv*. doi: arxiv-1501.06115.

Zimmermann, H. G., Tietz, C., and Grothmann, R. (2002) Yield Curve Forecasting by Error Correction Neural Networks and Partial Learning. *ESANN 2002 10th European Symposium on Artificial Neural Networks*, Bruges (Belgium), 407-412.

Zivot, E., and Wang, J (2007). *Modeling Financial Time Series with S-PLUS*. New York: Springer.

APPENDIX 1: SMOOTHING METHODS

A.1.a. SMOOTHING METHODS

x: time series, L: level component, b: growth component, s: seasonal component.

*Simple Exponential Smoothing: 1≥ α≥0.

$$\hat{y}_{t+1} = \alpha y_t + (1-\alpha)\hat{y}_t \qquad \text{(A.1.1)}$$

*Holt Linear Trend Exponential Smoothing:

$$\hat{y}_{t+h} = \hat{\ell}_t + h\hat{b}_t \qquad \text{(A.1.2)}$$

$$\hat{\ell}_t = \alpha y_t + (1-\alpha)\hat{y}_t \qquad \text{(A.1.3)}$$

$$\hat{b}_t = \beta\left(\hat{\ell}_t - \hat{\ell}_{t-1}\right) + (1-\beta)\hat{b}_{t-1} \qquad \text{(A.1.4)}$$

It is used in time series with trend. It performs level and trend update. 1> α, β>0. Initial values are found by linear trend regression. α: smoothing coefficient of the mean, β smoothing coefficient of the slope. Holt's two-parameter exponential smoothing technique does not use a second smoothing formula and only performs trend smoothing.

*Damped Trend Exponential Smoothing:

$$\hat{y}_{t+h} = \hat{\ell}_t + (\emptyset + \emptyset^2 + \emptyset^h)\hat{b}_t \qquad \text{(A.1.5)}$$

$$\hat{\ell}_t = \alpha y_t + (1-\alpha)(\hat{\ell}_{t-1} - \emptyset\hat{b}_{t-1}) \qquad \text{(A.1.6)}$$

$$\hat{b}_t = \beta(\hat{\ell}_t - \hat{\ell}_{t-1}) + (1-\beta)\emptyset\hat{b}_{t-1} \qquad \text{(A.1.7)}$$

It adds the damping parameter to the Holt smoothing.

* Holt-Winters' Exponential Smoothing:

1) Additive:

$$\hat{y}_{t+h} = \hat{\ell}_t + h\hat{b}_t + s_{t+h-m(k+1)} \qquad \text{(A.1.8)}$$

$$\hat{\ell}_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\hat{\ell}_{t-1} + \hat{b}_{t-1}) \qquad \text{(A.1.9)}$$

$$\hat{b}_t = \beta\left(\hat{\ell}_t - \hat{\ell}_{t-1}\right) + (1-\beta)\hat{b}_{t-1} \tag{A.1.10}$$

$$s_t = \gamma(y_t - \hat{\ell}_{t-1} - \hat{b}_{t-1}) + (1-\gamma)s_{t-m} \tag{A.1.11}$$

It is used for time series with trend and seasonality. It adds a seasonal component update to the Holt smoothing and a seasonal term to the forecast equation. m: period, k: integer part of (h-l)/m. The initial value is found by the additive decomposition method. This system of equations is used for point forecasts in all methods.

2) Multiplicative:

$$\hat{y}_{t+h} = (\hat{\ell}_t + h\hat{b}_t)s_{t+h-m(k+1)} \tag{A.1.12}$$

$$\hat{\ell}_t = \alpha(y_t/s_{t-m}) + (1-\alpha)(\hat{\ell}_{t-1} + \hat{b}_{t-1}) \tag{A.1.13}$$

$$\hat{b}_t = \beta\left(\hat{\ell}_t - \hat{\ell}_{t-1}\right) + (1-\beta)\hat{b}_{t-1} \tag{A.1.14}$$

$$s_t = \gamma(y_t/(\hat{\ell}_{t-1} - \hat{b}_{t-1})) + (1-\gamma)s_{t-m} \tag{A.1.15}$$

In two methods Λs are found by MSE minimization.

*HW damped exponential smoothing update equations:

1) Additive:

$$\hat{y}_{t+h} = \hat{\ell}_t + (\emptyset + \emptyset^2 + \emptyset^h)\hat{b}_t + s_{t+h-m(k+1)} \tag{A.1.16}$$

$$\hat{\ell}_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\hat{\ell}_{t-1} + \emptyset\hat{b}_{t-1}) \tag{A.1.17}$$

$$\hat{b}_t = \beta\left(\hat{\ell}_t - \hat{\ell}_{t-1}\right) + (1-\beta)\emptyset\hat{b}_{t-1} \tag{A.1.18}$$

$$s_t = \gamma(y_t - \hat{\ell}_{t-1} - \emptyset\hat{b}_{t-1}) + (1-\gamma)s_{t-m} \tag{A.1.19}$$

2) Multiplicative:

$$\hat{y}_{t+h} = (\hat{\ell}_t + (\emptyset + \emptyset^2 + \emptyset^h)\hat{b}_t)s_{t+h-m(k+1)} \tag{A.1.20}$$

$$\hat{\ell}_t = \alpha(y_t/s_{t-m}) + (1-\alpha)(\hat{\ell}_{t-1} + \emptyset\hat{b}_{t-1}) \tag{A.1.21}$$

$$\hat{b}_t = \beta\left(\hat{\ell}_t - \hat{\ell}_{t-1}\right) + (1-\beta)\emptyset\hat{b}_{t-1} \tag{A.1.22}$$

$$s_t = \gamma(y_t/(\hat{\ell}_{t-1} - \emptyset\hat{b}_{t-1})) + (1-\gamma)s_{t-m} \quad (A.1.23)$$

Simple Exponential Smoothing can be expressed as: $\varepsilon_t = x_t - \hat{\ell}_{t-1}$ or

$$y_t = \hat{\ell}_{t-1} + \varepsilon_1; \qquad \hat{\ell}_t = \hat{\ell}_{t-1} + \alpha\varepsilon_1 \quad (A.1.24)$$

In multiplicative:

$$y_t = \hat{\ell}_{t-1}(1+\varepsilon_1); \qquad \hat{\ell}_t = \hat{\ell}_{t-1}(1+\alpha\varepsilon_1). \quad (A.1.25)$$

Error here calculated by

$$\varepsilon_t = \frac{x_t - \hat{x}_{t-1}}{\hat{x}_{t-1}} \quad (A.1.26)$$

(Eğrioğlu & Baş, 2020; İslamoğlu, 2020; Hyndman & Khandakar, 2008)

A.1.b. ETS (N:None; A: Additive, A_d: Additive Damped)

ADDITIVE ERROR MODELS

Trend	Seasonal N	Seasonal A	Seasonal M
N	$y_t = \ell_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$	$y_t = \ell_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = \ell_{t-1}s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/\ell_{t-1}$
A	$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$	$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1})s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1} + \beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + b_{t-1})$
A_d	$y_t = \ell_{t-1} + \phi b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$	$y_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = \phi b_{t-1} + \beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + \phi b_{t-1})$

MULTIPLICATIVE ERROR MODELS

Trend	Seasonal N	Seasonal A	Seasonal M
N	$y_t = \ell_{t-1}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$	$y_t = (\ell_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \alpha(\ell_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + s_{t-m})\varepsilon_t$	$y_t = \ell_{t-1}s_{t-m}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
A	$y_t = (\ell_{t-1} + b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1})s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
A_d	$y_t = (\ell_{t-1} + \phi b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$

Source: Hyndman and Athanasopoulos (2021).

A.1.c. TBATS

Reduced forms: $\phi_p(L)\eta(L)y^{(\omega)}{}_t = \theta_q(L)\delta(L)\varepsilon_t$, L: lag operatör.

$$\eta(L) = (1-L)(1-\phi L)\prod_{i=1}^{T}\prod_{j=1}^{k_i}(1 - 2\cos\lambda_j^{(i)}L + L^2)$$

$$\delta(L) = [L^2\phi(1-\alpha) + L(\alpha + \phi\beta - \phi - 1) + 1]\prod_{i=1}^{T}\prod_{j=1}^{k_i}(1 - 2\cos\lambda_j^{(i)}L + L^2) +$$

$$(1-L)(1-\phi L)\sum_{i=1}^{T}\sum_{j=1}^{k_i}\prod_{\tilde{i}=1,\tilde{i}\neq i}^{T}\prod_{\tilde{j}=1,\tilde{j}\neq j}^{k_{\tilde{i}}}(1 - 2\cos\lambda_{\tilde{j}}^{(\tilde{i})}L + L^2)\left[(\cos\lambda_j^{(i)}\gamma_{1i} + \sin\lambda_j^{(i)}\gamma_{2i})L^2 - \gamma_{1i}L^3\right]$$

$$+ (1-L)(1-\phi L)L\prod_{i=1}^{T}\prod_{j=1}^{k_i}(1 - 2\cos\lambda_j^{(i)}L + L^2)\sum_{i=1}^{T}k_i\gamma_{1i}.$$

Source: De Livera et al. (2011).

APPENDIX 2: MATHEMATICAL STRUCTURE OF ARTIFICIAL NEURAL NETWORK

A.2.a. EPOCH

The number of epochs is the number of times the entire training sample is used in the model (Kožíšek, 2018). The selection of epochs is a critical and problematic issue in training models. Researchers often study the impact of the number of epochs on training performance. One key question is whether increasing the number of epochs always leads to improved performance. For instance, Namın and Namın (2018) demonstrated that increasing the number of epochs in LSTM had a stochastic effect on the results. To address the challenge of selecting the optimal number of epochs, researchers explore alternative parameters, adjust the step size, modify error and activation functions, and consider different algorithms. (Lewis, 2015).

At the beginning of network training, the error usually decreases rapidly. (See Figure 6). When the error of the network starts to increase again, the model suffers from overfitting if it is not stopped early. The trained model obtained as a result of early stopping is used in the test data. This is one of the effective methods against overfitting (Lewis, 2016). However, there is a risk that the this point may be the local optimum instead of the global optimum. Another solution to this problem is to limit the number of iterations (Öztemel, 2016).

A.2.b. BATCH SIZE

In machine learning training, the data is divided into batches, with each batch representing a subset of the training data (Namın & Namın, 2018). In traditional Backpropagation used in artificial neural networks (ANNs), the gradient is calculated for each neuron and epoch. However, in the case of large networks or datasets, this process requires considerable computation. Batching involves computing the gradient for several training data or groups of training data together (simultaneously) rather than for each individual training data. When using batching in training, multiple samples are passed in forward/backward signal transmission within a single epoch (Lewis, 2017a).

For recurrent neural networks (RNNs), the batch size refers to the number of training samples used in the back-and-forth transmission of RNN signals in order to perform weight updates. In ML, a large batch of input data requires more computation and memory, whereas a small batch can lead to highly variable parameters and local minima problems (Hoogteijling, 2020). Choosing a larger size reduces the generalization and performance ability as a result of less use in the training of certain input group characteristics that need to be explored in the network (Kožíšek, 2018).

A.2.c. DELTA AND DELTA BAR DELTA LEARNING

Some ANN models used in the study were trained with delta and delta bar delta learning rules. Delta learning rule aims to minimize the difference between the output and target values. For this, it revises the weight by taking the derivative of error (E) with respect to its weight. Weight revision formulas:

$$\Delta w_{ij} = \eta \frac{\partial E}{\partial w_{ij}} = \eta(o_j - d_j)\frac{\partial f(net_j)}{\partial net_j} x_i \tag{A.2.1}$$

$$o_j = f\left(net_j\right) \tag{A.2.2}$$

$$E = \frac{1}{2}(o_j - d_j)^2 \tag{A.2.3}$$

η: learning parameter, x: neuron input, o: neuron output, d: target value $(\ )_{ij}$: from neuron i to neuron j, *f*: aktivation function. The weights are changed at each calculation by the delta rule expressed as the system of equations

$$w_{i+1}=w_i+\Delta w_i \tag{A.2.4}$$

$$\Delta w_i=\eta\times x_i\times e \tag{A.2.5}$$

(Jacovides, 2008)

In the delta bar delta learning rule, the learning parameter is found separately for each weight. Weight revision formulas:

$$\Delta w_{ij}^{(k)} = \eta_{ij}^{(k)}(k)(o_j^{(k)} - d_j^{(k)})\frac{\partial f(net_j^{(k)})}{\partial net_j^{(k)}}x_i \quad (A.2.6)$$

$$\eta_{ij}^{(k)} = \eta_{ij}^{(k)} - \Delta\eta_{ij}^{(k)} \quad (A.2.7)$$

$$\Delta\eta_{ij}^{(k)} = \begin{cases} A & ; D_{ij}^{(k-1)}\frac{\partial E}{\partial w_{ij}^{(k)}} > 0 \\ -\varphi\eta_{ij}^{(k)} & ; D_{ij}^{(k-1)}\frac{\partial E}{\partial w_{ij}^{(k)}} < 0 \\ 0 & ; otherwise \end{cases} \quad (A.2.8)$$

$$D_{ij}^{(k-1)} = (1-\theta)\frac{\partial E}{\partial w_{ij}^{(k-1)}} + \theta\frac{\partial E}{\partial w_{ij}^{(k-2)}} \quad (A.2.9)$$

k: iteration number. φ, θ and A: constant. (Eğrioğlu et al., 2020).

A hidden layer is used in ANNs to model data with a piecewise continuous function relationship between input and output. Hornik et al. stated that an ANN with hidden layers that converge to this continuous function can be found in a data space of any dimension (Lewis, 2017a).

Optimization can be challenging when the error function is neither convex nor concave. If the second-order derivative (Hessian matrix) of the error function is not positive or negative semi-definite, the training can become stuck at local minima. The minimum error found by the network is influenced by the error surface, meaning that it depends on where the network starts training on this surface.(Lewis, 2017a).

A.2.d. GRADIENT DESCENT (GD)

Since slope is the derivative of curves in mathematics, derivatives of the loss function are used in this approach (Gürsakal, 2017). The GD attempts to minimize the network error by an iterative update of the parameters. In GD, weights increase (decrease) if the partial derivative is negative (positive). The learning rate determines the size of the steps of these updates (Lewis, 2017a). In GD, the learning rate can be made adjustable instead of fixed. (Alpaydın, 2010).

The search direction, step length (a) and new output are calculated at each iteration until the stopping condition is met or up to a certain number.

$$d^{(k)} = \nabla f(x^{(k)} \quad \text{(A.2.10)}$$

$$x^{(k+1)} = x^{(k)} + a^{(k)} d^{(k)} \quad \text{(A.2.11)}$$

$d^{(k)}$: search direction, $x^{(k)}$: point, ∇: gradient (Eğrioğlu et al., 2020).

Iterative optimization is performed to find the minimum value of RSS. For this purpose, the weight function in the cost gradient (objective function) is tried to be minimized with a quadratic solution using Taylor's Theorem (Ghatak, 2017).

If the derivative of the error function can be taken in ANN,

$$E^t(w|x,y) = \frac{1}{2}(y-o)^2 = \frac{1}{2}[y-(wo)]^2 \quad \text{(A.2.13)}$$

equality represents the error, while in the gradient descent,

$$\Delta w^{(k)} = \eta(y-o)x^{(k)} \quad \text{(A.2.14)}$$

$$w_{k+1} = w_k + \Delta w_k \quad \text{(A.2.15)}$$

indicates revision (x,y): data sample, k=0,…,m is weight revision (Alpaydın, 2010). If $w^{(k)}$ takes a value that underestimates the effect of x in the regression, then the effect of feature j (i.e. network connectivity) on $(y-o)$ becomes positive. The $w^{(k+1)}$ value will be increased (Ghatak, 2017).

In GD, each iteration is done with the entire dataset. Since recalculating the gradients with samples with similar data structures in these updates will lead to the same results, stochastic gradient descent (SGD) can be applied by randomly selecting single samples at each iteration to avoid unnecessary computation. This method usually provides much faster training. SGD with decreasing learning rate converges in the same way as GD. It finds general or global minima in non-convex loss functions (Lewis, 2017a).

Newton Method uses a quadratic approximation for the objective function. Search direction, which is the difference with the Gradient Descent approach, is expressed with

$$d^{(k)} = -\left[\nabla^2 f\left(x^{(k)}\right)\right]^{-1} \nabla f\left(x^{(k)}\right) \qquad \text{(A.2.16)}$$

$\nabla^2 f\left(x^{(k)}\right)$: is the Hessian matrix of the objective function at $x^{(k)}$. If this matrix is positive definite, $d^{(k)}$ indicates a decrease. Weight regression:

$$w^{(k+1)} = w^{(k)} - \left[\nabla^2 f\left(w^{(k)}\right)\right]^{-1} \nabla f\left(w^{(k)}\right) \qquad \text{(A.2.17)}$$

The algorithm is the same as the GD except for the search direction (Eğrioğlu et al., 2020).

Levenberg-Marquardt, which revises the weights only once in an epoch, is an algorithm that combines the steepest descent method, which finds the global optimum most shortly and aims for the smallest steps, and the Gauss-Newton method, which finds the global optimum by the second order derivative of the total error function (Szenczi, 2016). Gauss-Newton algorithm is wanted to avoid the disadvantages of this large dimension. Weight revision:

$$w^{(k+1)} = w^{(k)} - [J'J + \mu]^{-1} J'e \qquad \text{(A.2.18)}$$

Jacobian matrix (J), μ (Marquardt constant) (Eğrioğlu at al., 2020). The Levenberg-Marquardt training algorithm can be fast with good data scaling, as in steepest descent, and can be advantageous with large data, as in Gauss-Newton (Szenczi, 2016).

The LM algorithm does not calculate the Hessian matrix. It is a second-order derivative, so it calculates fast. The Jacobian matrix is calculated. It is the first-order derivative of the network errors. It makes Newton-like adjustments (Ruiz et al., 2016).

A.2.e. BACKPROPAGATION (BP)

The gradual adjustment of the weights backward from the output to minimize the error sum of the network is called error propagation or backpropagation (Lewis, 2017a). BP (generalized delta rule) is the search for a gradual reduction with the steepest descent approach on the surface of the error space (See Figure 6) to minimize the sum of the error squares (Pade & Mackworth, 2010). It optimizes the preferred error function according to the characteristics of the data set without constraints (Günay et al., 2007).

BP makes repeated updates with the derivative of the function used in the network to converge to the minimum error in ANN (Eklind, 2020). BP calculates the distance between the output of the ANN and the target output. It then uses this error measure to determine the associated error of the previous neurons. This error propagation is carried back to the input layer. The backward error propagation is then utilized to adjust the weights and intercepts (Lewis, 2017a).

Weight revision of k neurons of the hidden layer:

$$\Delta w_{j,k}(t)=\eta \times y_h(t) \times \delta_h(t) \tag{A.2.19}$$

The error reduction term $\delta_k(t)$, which replaces e(t) in the perceptron approach, is obtained by multiplying the derivative of the activation function by the error (ϵ).

$$\delta_h(t) = \frac{\partial y_k(t)}{\partial x_h(t)} \times \epsilon_h(t) \tag{A.2.20}$$

(Jacovides, 2008). The derivatives of the error function only require multiplying the deltas by the inputs (Hansson, 2017). Changes are made in the bias value with the same formula and procedures as in the weight. Error calculation of hidden layer neuron:

$$\delta_j(t) = \frac{\partial y_h(t)}{\partial x_h(t)} \times \sum_{k=1}^{l} \delta_h(t)\, w_{j,h}(t) \tag{A.2.21}$$

l: number of output neurons (Jacovides, 2008).

A.2.f. BACKPROPAGATION REVISION FORMULAS

Hidden-Output layers weights (*v* instead of w to avoid confusion with input-hidden layer):

$$v_j^{(t+1)} = v_j^{(t)} + \Delta v_j^{(t+1)} = v_j^{(t)} + \eta\left(-\frac{\partial E}{\partial v_j}\right) \tag{A.2.22}$$

$$\frac{\partial E}{\partial v_j} = \frac{\partial E}{\partial o_l}\frac{\partial o_l}{\partial net_o}\frac{\partial net_o}{\partial v_j} \tag{A.2.23}$$

$$\frac{\partial E}{\partial v_j} = \underbrace{-(d_l - o_l)}_{\frac{\partial E}{\partial o_l}}\ \underbrace{o_l}_{\frac{\partial o_l}{\partial net_o}}\ \underbrace{(1 - o_l)od_j}_{\frac{\partial net_o}{\partial v_j}} \tag{A.2.24}$$

$$v_j^{(t+1)} = v_j^{(t)} + \eta(d_l - o_l)o_l(1 - o_l)od_j \tag{A.2.26}$$

Output layer bias weights:

$$b_o^{(t+1)} = b_o^{(t)} + \eta\left(-\frac{\partial E}{\partial b_o}\right) \tag{A.2.27}$$

$$\frac{\partial E}{\partial b_o} = \frac{\partial E}{\partial o_l}\frac{\partial o_l}{\partial net_o}\frac{\partial net_o}{\partial b_o} \tag{A.2.28}$$

$$\frac{\partial E}{\partial b_o} = -(d_l - o_l)o_l(1 - o_l) \tag{A.2.29}$$

$$b_o^{(t+1)} = b_o^{(t)} + \eta(d_l - o_l)o_l(1 - o_l) \tag{A.2.30}$$

Input-hidden layers weights:

$$w_{ij}^{(t+1)} = w_{ij}^{(t)} + \Delta w_{ij}^{(t+1)} = w_{ij}^{(t)} + \eta\left(-\frac{\partial E}{\partial w_{ij}}\right) \tag{A.2.31}$$

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial o_l}\frac{\partial o_l}{\partial net_o}\frac{\partial net_o}{\partial od_j}\frac{\partial od_j}{\partial net_j}\frac{\partial net_j}{\partial w_{ij}} \tag{A.2.32}$$

$$\frac{\partial E}{\partial w_{ij}} = -(d_l - o_l)o_l(1 - o_l)v_j od_j\left(1 - od_j\right)x_{li} \tag{A.2.33}$$

$$w_{ij}^{(t+1)} = w_{ij}^{(t)} + \eta(d_l - o_l)o_l(1 - o_l)v_j^{(t+1)} od_j\left(1 - od_j\right)x_{li} \tag{A.2.34}$$

Hidden layer bias weights:

$$b_j^{(t+1)} = b_{ij}^{(t)} + \eta\left(-\frac{\partial E}{\partial b_{ij}}\right) \tag{A.2.35}$$

$$\frac{\partial E}{\partial b_j} = \frac{\partial E}{\partial o_l}\frac{\partial o_l}{\partial net_o}\frac{\partial net_o}{\partial od_j}\frac{\partial od_j}{\partial net_j}\frac{\partial net_j}{\partial b_j} \tag{A.2.36}$$

$$\frac{\partial E}{\partial b_j} = -(d_l - o_l)o_l(1 - o_l)v_j od_j\left(1 - od_j\right) \quad \text{(A.2.37)}$$

$$bd_j^{(t+1)} = bd_j^{(t)} + \eta(d_l - o_l)o_l(1 - o_l)v_j^{(t+1)} od_j\left(1 - od_j\right) \quad \text{(A.2.38)}$$

(Eğrioğlu et al., 2020).

A.2.g. STANDARD BACK PROPAGATION (BP) ALGORITHM

1) Initial parameters are determined (η, ε, maxit).
2) t=0 (iteration)
3) k=0 (cycle)
4) k=k+1
5) If k<maxit, go to "6". If not, the process is finished.
6) In iterations except, k=0 If the condition $|TE^k - TE^{k-1}| < \varepsilon$ is met, the process is finished. TE: Total Error. If not, it is continued. $TE^k = \sum_{l=1}^{n} \frac{1}{2} e_l^2 = \sum_{l=1}^{n} \frac{1}{2}(o_l - h_l)^2$.
7) l=0 learning counter.
8) l=l+1
9) t=t+1
10) If l<n, go to "11", otherwise go to "4".
11) For the current training sample, the output of the network is calculated with steps "3"-"6".
12) Weights and biases between the hidden and output layer are revised.
13) Weights and biases between input and hidden layer are updated (Eğrioğlu et al., 2020).

While BP is simplified as the derivative of the error function with respect to the weights in feed-forward ANN, this derivative is multiplied in recurrent ANN in parallel with its repetitive structure (Lewis, 2017a). Global minimum avoidance and convergence speed efficiency in BP are limited (Adhikari & Agrawal, 2013b).

A.2.h. MOMENTUM

The learning rate determines how quickly the parameters change in each cycle, while the momentum determines how quickly the previously calculated parameter values are updated. If parameter changes in the same direction happen one after the other, these changes accumulate and can potentially skip local minima (Hoogteijling, 2020). A large learning rate leads to fast learning but may cause training to miss the global minimum, resulting in poor performance or no learning at all. Conversely, choosing a small learning rate may lead to a longer time to find the minimum (Lewis, 2017a).

Momentum adds the previous revision to the weight revision at a certain rate.

$$w_{ij}^{(k+1)} = w_{ij}^{(k)} + \Delta w_{ij}^{(k+1)} + \Delta\gamma w_{ij}^{(k)} \quad \text{(A.2.39)}$$

γ: momentum constant. When the training reaches the optimal point, the derivative sign changes, due to the significant oscillation of the error surface. This change has an inverse effect on the weight adjustment, reducing the step length or increasing it.(Eğrioğlu et al., 2020).

With a high rate of momentum, there is a risk of skipping global minima along with local minima. If chosen at a low rate, it may lose the ability to skip local minima. Finding the appropriate optimum value by trial and error may be preferable. The learning rate should be reduced if the momentum is chosen high (Lewis, 2017a). Momentum memorizes the values of the previous weight update. It flattens the error surface (Alpaydın, 2010).

APPENDIX 3: SINGLE HIDDEN LAYER FEED-FORWARD NEURAL NETWORK (SLFN)

N: number of random samples, (x_i, t_i): random samples.With random hidden neurons

$$\sum_{i=1}^{\tilde{N}} W_i f_i(x_j) = \sum_{i=1}^{\tilde{N}} W_i f(w_i \cdot x_j + b_i) = o_j, \quad j = 1, \ldots, n \tag{A.3.1}$$

$$x_i = [x_{i1}, x_{i2}, \ldots, x_{in}]^T \in R^n, t_i = [t_{i1}, t_{i2}, \ldots, t_{im}]^T \in R^m \tag{A.3.2}$$

$\tilde{N}$: number of hidden neurons, $f(\cdot)$: activation function. $w_i = [w_{i1}, w_{i2}, \ldots, w_{in}]^T$: connection weight of input neurons with hidden neuron i. $W_i = [W_{i1}, W_{i2}, \ldots, W_{im}]^T$: weights connecting hidden neuron i to the output neuron, b_i: bias of hidden neuron i. SLFN error: $\sum_{j=1}^{N} \|o_j - t_j\| = 0$. If N neurons are randomly selected in this model $\sum_{i=1}^{\tilde{N}} W_i f(w_i \cdot x_j + b_i) = t_j$ equality is obtained. HW=T matrix.

$$H(w_1, \ldots, w_{\tilde{N}}, b_1, \ldots, b_{\tilde{N}}, x_1, \ldots, x_N) = \begin{bmatrix} f(w_1 \cdot x_1 + b_1) & \ldots & \ldots f(w_{\tilde{N}} \cdot x_1 + b_{\tilde{N}}) \\ \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots \\ f(w_1 \cdot x_N + b_1) & \ldots & \ldots f(w_{\tilde{N}} \cdot x_N + b_{\tilde{N}}) \end{bmatrix}_{N \times \tilde{N}}$$

i^{th} hidden layer output (A.3.3)

H: hidden layer output matrix.

$$W = \begin{bmatrix} W_1^T \\ . \\ . \\ . \\ W_{\tilde{N}}^T \end{bmatrix}_{\tilde{N} \times m} \text{ and } T = \begin{bmatrix} t_1^T \\ . \\ . \\ . \\ t_N^T \end{bmatrix}_{N \times m} \tag{A.3.4}$$

Let a standard SLFN network have N hidden neurons and an infinitely differentiable activation function. If this network is trained with N random ($x_i \in R^n$, $t_i \in R^m$) samples, according to the continuous probability distribution $w_i \in R^n$, $b_i \in R$ are randomly selected. From the output of the hidden layer $\|HW - T\| = 0$ can be formulated. If $\varepsilon > 0$ is taken as a small number, there are the required number of $\tilde{N} \leq N$ making $\|H_{N \times \tilde{N}} W_{\tilde{N} \times m} - T_{N \times m}\| < \varepsilon$ in the same way.

APPENDIX 4: ELM NETWORK

Minimization of the cost function in SLFN,

$$\left\|H(\hat{w}_1, \dots, \hat{w}_{\tilde{N}}, \hat{b}_1, \dots, \hat{b}_{\tilde{N}})\hat{W} - T\right\| = \min_{w_i, b_i, \beta} \|H(w_1, \dots, w_{\tilde{N}}, b_1, \dots, b_{\tilde{N}})W - T\| \quad \text{(A.4.1)}$$

According to this,

$$E = \sum_{j=1}^{N}\left(\sum_{i=1}^{\tilde{N}} W_i f(w_i \cdot x_j + b_i) - t_j\right)^2 \quad \text{(A.4.2)}$$

If we train SLFN according to this cost; $\hat{w}_i, \hat{b}_i, \hat{W}(i = 1, \dots, \tilde{N})$ parameters can be found.

**Gradient-based solution*: If H is unknown, in gradient-based learning weights and biases is revised with

$$\omega_K = \omega_{K-1} - \eta \frac{\partial E(\omega)}{\partial \omega} \quad \text{(A.4.3)}$$

to minimize

$$\|HW - T\| = 0 \quad \text{(A.4.4)}$$

$\omega = (w_i, W_i, b_i)$: parameter set, η: learning rate.

**Minimum norm least squares solution*: In training, input weights and hidden layer bias can be left constant without revision. The H matrix can remain constant after initially assigning random parameters. The cost function minimization becomes linear.

$$\left\|H(w_1, \dots, w_{\tilde{N}}, b_1, \dots, b_{\tilde{N}})\hat{W} - T\right\| = \min_{\beta} \|H(w_1, \dots, w_{\tilde{N}}, b_1, \dots, b_{\tilde{N}})W - T\| \quad \text{(A.4.5)}$$

With the cost function, $\hat{W}$ is found. Solution of this equation is

$$\hat{W} = H^{\dagger}T \quad \text{(A.4.6)}$$

$H^{\dagger}$ is the generalized inverse of H. If $\tilde{N} = N$, when w_i, b_i chosen at random H is an inverted square matrix and the SLFN error can approach zero with training. But most of the time $\tilde{N} \ll N$ becomes. H is a non-square matrix. Therefore $w_i, b_i, W_i (i = 1, \ldots, \tilde{N})$ parameters and HW=T solution may not exist. In SLFN with N sigmoid hidden neurons, if b_i is set optimally, it can be trained with N observations by assigning random weights. The number of hidden neurons required here must be

$$\sum_{i=1}^{L} s_i \gg max_i(s_i) \tag{A.4.7}$$

and common hidden neurons should not be shared. In ELM, SVD (singular value decomposition) can be used in any case to find the Moore-Penrose generalized inverse of matrix H. The iterative method contradicts the search and non-iteration property of ELM. The orthogonal project method can be used if H^TH is not singular and

$$H^{\dagger}=(H^TH)^{-1}H^T \tag{A.4.8}$$

(Huang et al., 2006). Regularized ELM is solved with

$$W = H^T(\frac{I}{\lambda} + HH^T)^{-1}T \tag{A.4.9}$$

I: unit matrix λ: regularization factor found by cross-validation in training (Zhu et al., 2015).

APPENDIX 5: DECISION TREES

In decision trees, similar data are partitioned in the same branch. In tree-based methods, the results of individual decision trees are combined. Individual decision trees have bias and variance problems (Hoogteijling, 2020). In traditional decision trees, a local optimum is found at each node. However, depending on the decision made at the next node, the result of the previous decision may be sub-optimal. Universal algorithms can only approximate The global optimum tree (Lewis, 2017b).

The Classification and Regression Trees (CART) (Breiman et al., 1984) approach builds a model from the training sample for a fixed classifier or regression that makes piecewise predictions. It selects the good tree using a penalization criterion.

The decision tree (DT) regression approach divides the feature space into subgroups or leaves. Simple regressions model the data samples grouped in each leaf. In particular, a network node is taken to process all the data and then partitioned according to its optimal function.

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \frac{N_L}{N_a} H\big(B_L(\theta)\big) + \frac{N_R}{N_a} H\big(B_R(\theta)\big) \tag{A.5.1}$$

$N_L, N_R \in [N_{min}, N_a]$. L: left, R: right, N: number of data. For each leaf a, regression model is constructed with the univariate observation mean.

$$\hat{y} = \frac{1}{N_a} \sum_{i=1}^{N_a} y_{a,i} \tag{A.5.2}$$

$$\bar{y}_{a,j} = \frac{1}{N_a} \sum_{i=1}^{N_a} y_{a,i,j}, \forall j = 1, \dots, m \tag{A.5.3}$$

$$H(x_a) = \frac{1}{mN_a} \sum_{j=1}^{m} \sum_{i=1}^{N_a} (y_{a,i,j} - \bar{y}_{a,j})^2 \tag{A.5.4}$$

(Burger & Moura, 2015).

Entropy, one of the methods used in data partitioning in decision trees, measures randomness.

$$Entropy(X) = \sum_{i=1}^{k} -p_i log_2(p_i) \quad \text{(A.5.5)}$$

p_k: the probability that an sample belongs to class k.

$$p_i = \frac{|C_{i,X}|}{|X|} \quad \text{(A.5.6)}$$

Entropy can be considered as a measure to find homogeneity (Lewis, 2017b).

The ID3 Decision Tree approach is the simplest decision tree approach using categorical attributes. The criterion for tree building is information gain.

$$Information\ Gain(A) = Entropy(X) - Entropy_A(X) \quad \text{(A.5.7)}$$

$$Entropy_A(X) = \sum_{j=1}^{v} \frac{|X_j|}{|X|} Entropy(X_j) \quad \text{(A.5.8)}$$

C_i: number of classes, i=1,…,k; $C_{i,X}$: set of observations of class C_i in X, |·|: number of observations. This criterion is the information gain that a partition A in the tree can provide (Balaban & Kartal, 2018).

C4.5 Decision Tree approach; uses the partition criterion as a criterion.

$$Partition\ Information_A(X) = -\sum_{j=1}^{v} \frac{|X_j|}{|X|} * log_2\left(\frac{|X_j|}{|X|}\right) \quad \text{(A.5.9)}$$

A: training data attribute. This measure indicates the information that can be obtained by dividing A into v parts. The attribute with the highest gain according to

$$\text{Gain Rate(A)} = \frac{\text{Information Gain(A)}}{\text{Partition Information (A)}} \quad \text{(A.5.10)}$$

ratio is selected as a node. The numerical equivalent of the attributes of the observations can be dichotomized (≤θ and >θ) according to a threshold value (θ). Further partitioning can be done using other methods. Threshold value can be taken as the midpoint in the forms of

$$\theta_i = \frac{v_i + v_{i+1}}{2} \quad \text{(A.5.11)}$$

Another approach is the largest value that does not exceed the average value can be selected in the form of

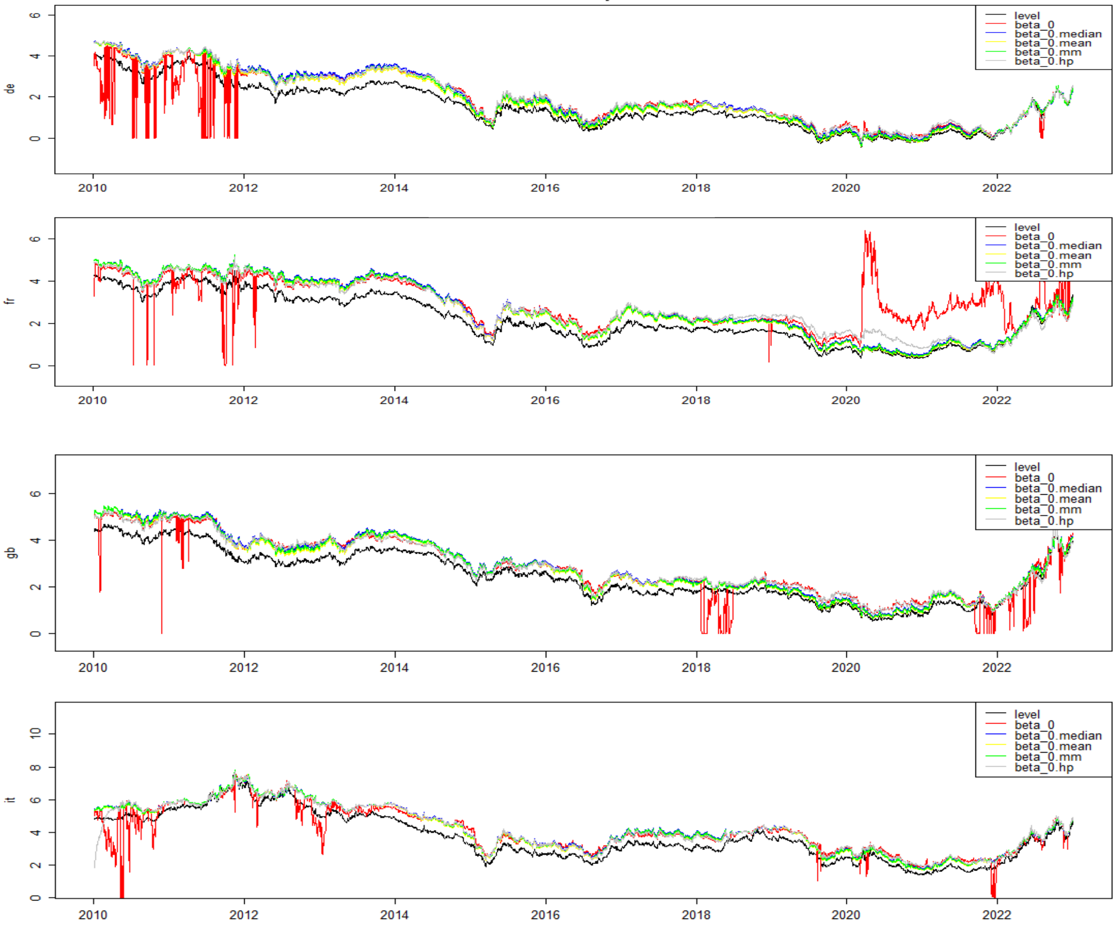
$$\theta_i = max\left\{v|v \leq \frac{v_i+v_{i+1}}{2}\right\} \tag{A.5.12}$$

(Berzal at al., 2004).

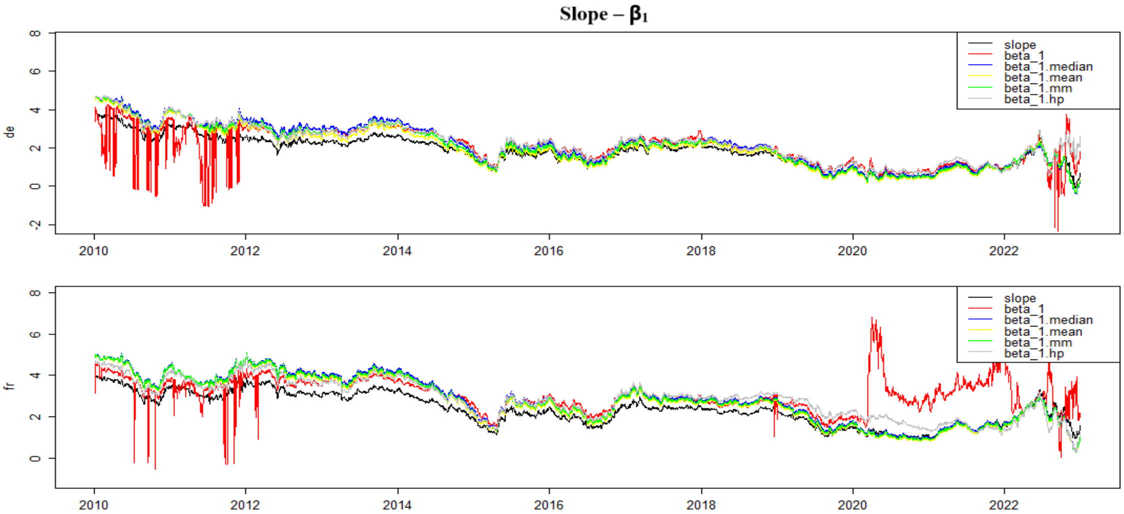
APPENDIX 6: PLOTS OF LEVEL, SLOPE, AND CURVATURE VS NS FACTORS

Figure A.1: Level-β_0, Slope- β_1, Curvature- β_2 Plots Of Germany, France, Great Britain And Italy

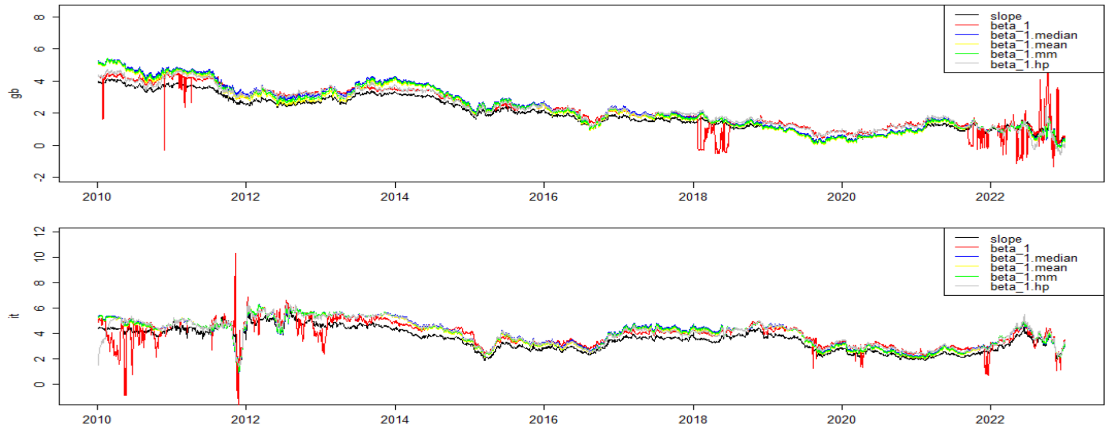
Level – β_0



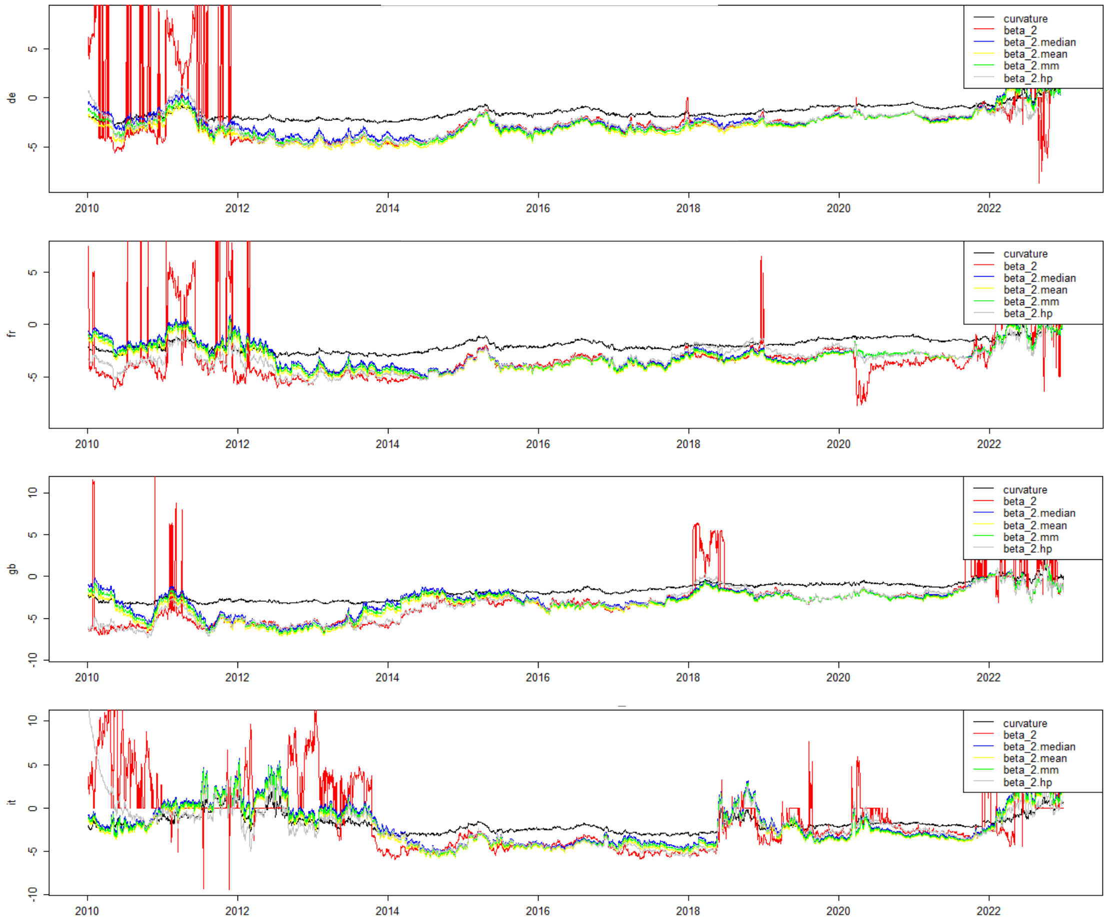
Slope – β_1



Slope – β_1



Curvature - β_2



APPENDIX 7: DESCRIPTIVE STATISTICS

Table A. 1: Descriptive Statistics of Some Yields and Yield Curve Factors Data

Germany	MEAN	SD	MİN	MAX	ACF(1)	ACF(10)	ACF(21)	ACF(252)	PACF(1)	PACF(10)	ADF
M3	-0,266	0,502	-1,226	1,871	0,996	0,960	0,915	0,476	0,996	0,026	0,99
Y1	-0,152	0,627	-0,941	2,633	0,996	0,960	0,919	0,366	0,996	0,010	0,99
Y5	0,263	0,897	-0,953	2,819	0,997	0,972	0,946	0,522	0,997	0,017	0,99
Y10	0,889	1,073	-0,801	3,497	0,998	0,981	0,963	0,640	0,998	0,015	0,99
Y30	1,516	1,130	-0,437	4,111	0,998	0,983	0,967	0,686	0,998	0,018	0,99
β0	1,983	1,280	-0,199	4,748	0,999	0,985	0,969	0,713	0,999	0,022	0,96
β1	-2,225	0,950	-4,759	0,515	0,997	0,965	0,934	0,628	0,997	0,029	0,05
β2	-2,521	1,378	-4,991	4,026	0,994	0,930	0,863	0,327	0,994	0,038	0,40
λ	0,041	0,043	0,021	0,323	0,994	0,937	0,868	0,003	0,994	0,003	0,99
Level	1,516	1,130	-0,437	4,111	0,998	0,983	0,967	0,686	0,998	0,018	0,99
Slope	1,782	0,790	-0,095	3,828	0,997	0,966	0,928	0,622	0,997	0,003	0,04
Curvature	-1,443	0,719	-2,854	1,340	0,995	0,952	0,913	0,403	0,995	0,009	0,74

France	MEAN	SD	MİN	MAX	ACF(1)	ACF(10)	ACF(21)	ACF(252)	PACF(1)	PACF(10)	ADF
M3	-0,191	0,503	-1,031	1,889	0,996	0,960	0,924	0,487	0,996	0,012	0,99
Y1	-0,069	0,645	-0,835	2,790	0,997	0,960	0,916	0,373	0,997	-0,005	0,99
Y5	0,546	0,988	-0,775	3,045	0,998	0,977	0,956	0,603	0,998	0,010	0,99
Y10	1,255	1,229	-0,638	3,788	0,999	0,987	0,973	0,723	0,999	0,004	0,99
Y30	2,165	1,134	0,325	4,469	0,999	0,986	0,971	0,717	0,999	0,022	0,99
β0	2,713	1,196	0,561	5,133	0,999	0,986	0,970	0,723	0,999	0,007	0,82
β1	-2,851	0,924	-4,956	0,276	0,997	0,967	0,927	0,601	0,997	-0,010	0,06
β2	-3,253	1,197	-5,839	1,288	0,993	0,920	0,858	0,322	0,993	-0,004	0,26
λ	0,034	0,013	0,015	0,084	0,998	0,975	0,948	0,409	0,998	-0,001	0,99
Level	2,165	1,134	0,325	4,469	0,999	0,986	0,971	0,717	0,999	0,022	0,99
Slope	2,356	0,778	0,908	3,981	0,997	0,971	0,939	0,626	0,997	0,017	0,16
Curvature	-1,939	0,645	-3,220	0,649	0,993	0,935	0,883	0,388	0,993	0,022	0,61

Great Britain	MEAN	SD	MİN	MAX	ACF(1)	ACF(10)	ACF(21)	ACF(252)	PACF(1)	PACF(10)	ADF
M3	0,508	0,480	-0,091	3,545	0,994	0,937	0,866	-0,103	0,994	0,003	0,99
Y1	0,544	0,567	-0,175	4,240	0,992	0,943	0,887	-0,038	0,992	-0,030	0,99
Y5	1,138	0,777	-0,137	4,688	0,996	0,957	0,911	0,251	0,996	-0,017	0,98
Y10	1,786	0,982	0,102	4,501	0,997	0,974	0,946	0,499	0,997	-0,015	0,99
Y30	2,479	1,099	0,549	4,892	0,999	0,984	0,967	0,657	0,999	0,001	0,99
β0	2,917	1,189	0,831	5,324	0,999	0,986	0,970	0,702	0,999	-0,010	0,98
β1	-2,311	1,169	-4,748	0,598	0,998	0,977	0,949	0,697	0,998	-0,022	0,01
β2	-3,304	1,867	-7,367	4,245	0,996	0,961	0,935	0,585	0,996	-0,010	0,02
λ	0,043	0,019	0,022	0,120	0,997	0,969	0,935	0,198	0,997	-0,004	0,93
Level	2,479	1,099	0,549	4,892	0,999	0,984	0,967	0,657	0,999	0,001	0,99
Slope	1,970	1,049	0,101	4,158	0,998	0,983	0,965	0,724	0,998	-0,013	0,17
Curvature	-1,743	1,029	-3,419	2,572	0,998	0,977	0,962	0,672	0,998	-0,017	0,01

Italy	MEAN	SD	MİN	MAX	ACF(1)	ACF(10)	ACF(21)	ACF(252)	PACF(1)	PACF(10)	ADF
M3	0,168	0,809	-0,804	6,407	0,996	0,944	0,865	0,495	0,996	-0,037	0,01
Y1	0,616	1,158	-0,583	6,508	0,998	0,971	0,934	0,531	0,998	-0,040	0,73
Y2	1,006	1,395	-0,571	7,979	0,998	0,974	0,945	0,573	0,998	-0,020	0,82
Y5	1,875	1,573	-0,139	7,825	0,998	0,981	0,962	0,608	0,998	-0,008	0,96
Y10	2,839	1,580	0,483	7,311	0,999	0,986	0,972	0,649	0,999	0,005	0,98
Y30	3,694	1,425	1,393	7,677	0,999	0,990	0,979	0,717	0,999	-0,008	0,96
β0	4,183	1,388	1,834	7,684	0,998	0,987	0,973	0,708	0,998	0,018	0,52
β1	-4,039	0,982	-6,502	-1,402	0,995	0,942	0,866	0,474	0,995	-0,020	0,01
β2	-2,188	2,294	-5,739	12,31	0,988	0,881	0,793	0,259	0,988	0,008	0,01
λ	0,038	0,010	0,008	0,062	0,998	0,976	0,948	0,161	0,998	-0,004	0,91
Level	3,694	1,425	1,393	7,677	0,999	0,990	0,979	0,717	0,999	-0,008	0,96
Slope	3,526	0,892	0,901	5,897	0,996	0,946	0,879	0,540	0,996	-0,046	0,01
Curvature	-1,849	0,979	-3,289	2,243	0,989	0,904	0,839	0,240	0,989	0,027	0,29

US	MEAN	SD	MİN	MAX	ACF(1)	ACF(10)	ACF(21)	ACF(252)	PACF(1)	PACF(10)	ADF
M3	0,631	0,933	-0,046	4,457	0,997	0,970	0,934	0,288	0,997	0,009	0,99
Y1	0,815	1,017	0,041	4,788	0,997	0,973	0,941	0,243	0,997	-0,012	0,99
Y2	1,017	0,963	0,105	4,728	0,997	0,976	0,947	0,220	0,997	-0,012	0,99
Y5	1,603	0,791	0,192	4,448	0,996	0,968	0,933	0,137	0,996	0,003	0,93
Y10	2,249	0,731	0,512	4,247	0,996	0,965	0,928	0,184	0,996	0,002	0,87
Y30	2,970	0,765	1,066	4,834	0,997	0,973	0,946	0,401	0,997	-0,004	0,85
β0	3,363	0,833	1,315	5,307	0,997	0,975	0,951	0,518	0,997	0,001	0,64
β1	-2,792	1,231	-5,163	-0,113	0,996	0,973	0,949	0,689	0,996	-0,048	0,33
β2	-2,348	2,410	-6,347	5,405	0,994	0,962	0,929	0,456	0,994	-0,054	0,61
λ	0,055	0,070	0,005	0,598	0,975	0,836	0,762	0,020	0,975	-0,039	0,09
Level	2,970	0,765	1,066	4,834	0,997	0,973	0,946	0,401	0,997	-0,004	0,85
Slope	2,339	1,183	-0,868	4,695	0,998	0,978	0,951	0,627	0,998	-0,020	0,50
Curvature	-1,568	1,185	-3,591	1,962	0,998	0,985	0,967	0,462	0,998	-0,026	0,53

APPENDIX 8: AREA FUNCTION FOR R SOFTWARE

```
function(t, x1, x2){
  a.f <- matrix(NA, nrow = (length(t)-1), ncol = 1)
  for (i in 1:(length(t) - 1)) {
    if ((x1[i] - x2[i]) * (x1[i+1] - x2[i+1]) <0) {
      a.f[i,] <- (t[i+1] - t[i]) * ((abs(x1[i] - x2[i]) + abs(x1[i+1] - x2[i+1])) / 2-
                             abs(x1[i]-x2[i]) * abs(x1[i+1] - x2[i+1]) / (abs(x1[i]- x2[i]) +
                         abs(x1[i+1] - x2[i+1])))
    }else{
      a.f[i,] <- (t[i+1] - t[i]) * (abs(x1[i] - x2[i]) + abs(x1[i+1] - x2[i+1])) / 2
    }
  }
  return(sum(a.f))
}
```

APPENDIX 9: GRANGER CAUSALITY TEST

Table A.2: Granger Causality Test

P 1	P 2	p value	P 1	P 2	p value	P 1	P 2	p value	P 1	P 2	p value	P 1	P 2	p value	P 1	P 2	p value	P 1	P 2	p value
CA.b0	CA.b1	0,863956	DE.b0	DE.b1	0,000000	FR.b0	FR.b1	0,004413	GB.b0	GB.b1	0,978902	IT.b0	IT.b1	0,000000	TR.b0	TR.b1	0,679174	US.b0	US.b1	0,774750
CA.b0	CA.b2	0,000001	DE.b0	DE.b2	0,000000	FR.b0	FR.b2	0,000000	GB.b0	GB.b2	0,000000	IT.b0	IT.b2	0,000000	TR.b0	TR.b2	0,065771	US.b0	US.b2	0,681870
CA.b0	CA.l	0,294846	DE.b0	DE.l	0,013690	FR.b0	FR.l	0,000000	GB.b0	GB.l	0,000022	IT.b0	IT.l	0,001175	TR.b0	TR.l	0,421256	US.b0	US.l	0,019801
CA.b1	CA.b0	0,000009	DE.b1	DE.b0	0,000000	FR.b1	FR.b0	0,017760	GB.b1	GB.b0	0,000000	IT.b1	IT.b0	0,000000	TR.b1	TR.b0	0,000009	US.b1	US.b0	0,815532
CA.b1	CA.b2	0,000494	DE.b1	DE.b2	0,000000	FR.b1	FR.b2	0,000000	GB.b1	GB.b2	0,630640	IT.b1	IT.b2	0,015100	TR.b1	TR.b2	0,000757	US.b1	US.b2	0,000016
CA.b1	CA.l	0,036548	DE.b1	DE.l	0,361638	FR.b1	FR.l	0,001024	GB.b1	GB.l	0,004765	IT.b1	IT.l	0,000009	TR.b1	TR.l	0,000733	US.b1	US.l	0,216129
CA.b2	CA.b0	0,000015	DE.b2	DE.b0	0,000000	FR.b2	FR.b0	0,001864	GB.b2	GB.b0	0,000000	IT.b2	IT.b0	0,000002	TR.b2	TR.b0	0,730891	US.b2	US.b0	0,045325
CA.b2	CA.b1	0,055121	DE.b2	DE.b1	0,000000	FR.b2	FR.b1	0,031936	GB.b2	GB.b1	0,000000	IT.b2	IT.b1	0,000213	TR.b2	TR.b1	0,000046	US.b2	US.b1	0,004692
CA.b2	CA.l	0,545208	DE.b2	DE.l	0,035063	FR.b2	FR.l	0,000031	GB.b2	GB.l	0,253004	IT.b2	IT.l	0,414755	TR.b2	TR.l	0,123689	US.b2	US.l	0,000001
CA.l	CA.b0	0,164854	DE.l	DE.b0	0,000000	FR.l	FR.b0	0,000000	GB.l	GB.b0	0,000000	IT.l	IT.b0	0,000000	TR.l	TR.b0	0,002477	US.l	US.b0	0,154431
CA.l	CA.b1	0,000000	DE.l	DE.b1	0,000000	FR.l	FR.b1	0,000049	GB.l	GB.b1	0,000000	IT.l	IT.b1	0,000063	TR.l	TR.b1	0,000000	US.l	US.b1	0,519174
CA.l	CA.b2	0,200106	DE.l	DE.b2	0,000000	FR.l	FR.b2	0,000000	GB.l	GB.b2	0,000000	IT.l	IT.b2	0,000000	TR.l	TR.b2	0,000075	US.l	US.b2	0,008966
CA.b0	US.b0	0,000000	DE.b0	US.b0	0,000000	FR.b0	US.b0	0,000000	GB.b0	US.b0	0,000000	IT.b0	US.b0	0,000001	TR.b0	US.b0	0,067153	US.b0	DE.b0	0,102881
CA.b0	US.b1	0,000000	DE.b0	US.b1	0,000000	FR.b0	US.b1	0,000000	GB.b0	US.b1	0,000000	IT.b0	US.b1	0,002097	TR.b0	US.b1	0,033923	US.b0	DE.b1	0,947937
CA.b0	US.b2	0,001123	DE.b0	US.b2	0,000000	FR.b0	US.b2	0,000000	GB.b0	US.b2	0,000194	IT.b0	US.b2	0,028513	TR.b0	US.b2	0,332514	US.b0	DE.b2	0,810501
CA.b0	US.l	0,009051	DE.b0	US.l	0,027947	FR.b0	US.l	0,053331	GB.b0	US.l	0,139362	IT.b0	US.l	0,521802	TR.b0	US.l	0,528017	US.b0	DE.l	0,019433
CA.b1	US.b0	0,029184	DE.b1	US.b0	0,000000	FR.b1	US.b0	0,001256	GB.b1	US.b0	0,000000	IT.b1	US.b0	0,000000	TR.b1	US.b0	0,117660	US.b1	DE.b0	0,118625
CA.b1	US.b1	0,000000	DE.b1	US.b1	0,000000	FR.b1	US.b1	0,000000	GB.b1	US.b1	0,000000	IT.b1	US.b1	0,000186	TR.b1	US.b1	0,154162	US.b1	DE.b1	0,268598
CA.b1	US.b2	0,000022	DE.b1	US.b2	0,000000	FR.b1	US.b2	0,000000	GB.b1	US.b2	0,000000	IT.b1	US.b2	0,000191	TR.b1	US.b2	0,001657	US.b1	DE.b2	0,614019
CA.b1	US.l	0,565722	DE.b1	US.l	0,520406	FR.b1	US.l	0,803250	GB.b1	US.l	0,477768	IT.b1	US.l	0,624111	TR.b1	US.l	0,001397	US.b1	DE.l	0,216109
CA.b2	US.b0	0,022949	DE.b2	US.b0	0,000648	FR.b2	US.b0	0,450431	GB.b2	US.b0	0,000000	IT.b2	US.b0	0,002500	TR.b2	US.b0	0,949751	US.b2	DE.b0	0,158342
CA.b2	US.b1	0,593166	DE.b2	US.b1	0,083912	FR.b2	US.b1	0,092498	GB.b2	US.b1	0,000000	IT.b2	US.b1	0,004807	TR.b2	US.b1	0,207242	US.b2	DE.b1	0,333523
CA.b2	US.b2	0,246758	DE.b2	US.b2	0,334381	FR.b2	US.b2	0,000007	GB.b2	US.b2	0,000000	IT.b2	US.b2	0,012654	TR.b2	US.b2	0,103512	US.b2	DE.b2	0,981300
CA.b2	US.l	0,017879	DE.b2	US.l	0,000031	FR.b2	US.l	0,000000	GB.b2	US.l	0,000766	IT.b2	US.l	0,053946	TR.b2	US.l	0,102530	US.b2	DE.l	0,028781
CA.l	US.b0	0,000000	DE.l	US.b0	0,000000	FR.l	US.b0	0,000000	GB.l	US.b0	0,000000	IT.l	US.b0	0,000000	TR.l	US.b0	0,000000	US.l	DE.b0	0,366687
CA.l	US.b1	0,000000	DE.l	US.b1	0,066407	FR.l	US.b1	0,000000	GB.l	US.b1	0,000000	IT.l	US.b1	0,000000	TR.l	US.b1	0,000021	US.l	DE.b1	0,439886
CA.l	US.b2	0,000185	DE.l	US.b2	0,000000	FR.l	US.b2	0,000000	GB.l	US.b2	0,000000	IT.l	US.b2	0,000000	TR.l	US.b2	0,013254	US.l	DE.b2	0,722056
CA.l	US.l	0,000000	DE.l	US.l	0,000000	FR.l	US.l	0,000000	GB.l	US.l	0,000000	IT.l	US.l	0,000000	TR.l	US.l	0,000000	US.l	DE.l	0,000000

APPENDIX 10: AREA ERRORS OF AR(1) MODEL

Table A.3: Area Errors of AR(1) Model for 100, 200 and 300 Forecasting Periods

CA	100	200	300	DE	100	200	300	FR	100	200	300	GB	100	200	300	IT	100	200	300	US	100	200	300
1	19,61	18,01	18,03	1	15,23	15,22	15,22	1	28,45	28,40	28,45	1	17,93	17,93	18,00	1	21,68	23,08	20,60	1	15,74	15,58	15,76
2	22,63	21,09	21,12	2	18,38	18,39	18,37	2	30,50	30,47	30,47	2	21,66	21,65	21,77	2	26,84	28,02	25,84	2	20,59	20,38	20,60
3	25,46	23,98	24,02	3	21,18	21,24	21,22	3	32,44	32,48	32,39	3	25,06	25,04	25,25	3	31,30	32,49	30,40	3	24,92	24,72	24,90
4	28,02	26,67	26,72	4	23,94	24,00	23,98	4	34,23	34,29	34,15	4	28,15	28,11	28,37	4	35,36	36,50	34,53	4	28,88	28,78	28,88
5	30,34	29,08	29,14	5	26,34	26,39	26,38	5	35,91	36,04	35,81	5	30,96	30,86	31,18	5	38,93	39,96	38,18	5	32,54	32,57	32,51
6	32,40	31,23	31,32	6	28,61	28,59	28,70	6	37,53	37,70	37,40	6	33,50	33,40	33,81	6	42,39	43,37	41,70	6	35,99	36,15	35,95
7	34,40	33,32	33,45	7	30,58	30,60	30,74	7	39,14	39,41	38,99	7	35,96	35,83	36,25	7	45,65	46,64	45,10	7	39,37	39,70	39,42
8	36,29	35,29	35,46	8	32,46	32,49	32,74	8	40,72	41,09	40,60	8	38,20	38,06	38,47	8	49,00	50,01	48,47	8	42,39	42,89	42,46
9	38,01	37,11	37,31	9	34,39	34,45	34,71	9	42,33	42,75	42,21	9	40,42	40,26	40,65	9	52,18	53,24	51,61	9	45,50	46,16	45,49
10	39,74	38,95	39,22	10	36,24	36,33	36,57	10	43,90	44,46	43,79	10	42,61	42,41	42,78	10	55,31	56,42	54,72	10	48,42	49,23	48,51
11	41,56	40,88	41,17	11	38,13	38,27	38,54	11	45,57	46,23	45,44	11	44,85	44,60	45,05	11	58,18	59,31	57,50	11	51,41	52,32	51,53
12	43,35	42,77	43,10	12	39,97	40,15	40,41	12	47,09	47,88	46,96	12	47,05	46,72	47,31	12	61,02	62,11	60,24	12	54,20	55,32	54,47
13	45,12	44,67	45,04	13	41,82	42,05	42,29	13	48,63	49,55	48,52	13	49,19	48,81	49,62	13	63,76	64,94	62,91	13	56,99	58,34	57,45
14	46,89	46,55	46,98	14	43,64	43,81	44,12	14	50,11	51,16	49,99	14	51,38	50,91	51,83	14	66,50	67,68	65,48	14	59,63	61,18	60,23
15	48,51	48,30	48,75	15	45,32	45,50	45,81	15	51,66	52,90	51,53	15	53,52	52,98	54,05	15	69,08	70,36	68,10	15	62,17	63,88	62,94
16	50,19	50,03	50,54	16	46,94	47,16	47,44	16	53,19	54,58	53,06	16	55,63	54,96	56,12	16	71,56	72,89	70,36	16	64,58	66,47	65,56
17	51,78	51,72	52,21	17	48,56	48,73	49,10	17	54,65	56,15	54,59	17	57,74	56,91	58,26	17	73,97	75,35	72,67	17	66,84	68,91	67,99
18	53,48	53,48	53,99	18	50,12	50,22	50,74	18	56,07	57,70	56,10	18	59,80	58,80	60,32	18	76,28	77,79	74,97	18	69,05	71,31	70,49
19	55,21	55,28	55,75	19	51,70	51,73	52,40	19	57,62	59,34	57,66	19	61,72	60,56	62,37	19	78,74	80,34	77,41	19	71,25	73,73	73,03
20	56,86	56,98	57,46	20	53,28	53,28	54,00	20	59,21	61,06	59,28	20	63,66	62,38	64,34	20	81,14	82,85	79,76	20	73,47	76,16	75,50
21	58,54	58,67	59,19	21	54,84	54,77	55,57	21	60,78	62,70	60,85	21	65,54	64,12	66,30	21	83,55	85,36	82,06	21	75,65	78,53	77,90
22	60,19	60,36	60,85	22	56,34	56,23	57,15	22	62,30	64,29	62,37	22	67,42	65,84	68,24	22	85,91	87,70	84,43	22	77,85	80,92	80,37
23	61,95	62,11	62,62	23	57,66	57,57	58,52	23	63,66	65,74	63,87	23	69,08	67,35	69,95	23	88,21	90,07	86,74	23	80,02	83,28	82,78
24	63,65	63,82	64,28	24	58,92	58,81	59,86	24	65,05	67,23	65,35	24	70,80	68,93	71,77	24	90,47	92,46	88,90	24	82,11	85,48	85,06
25	65,34	65,53	65,94	25	60,29	60,11	61,24	25	66,45	68,71	66,76	25	72,47	70,49	73,56	25	92,64	94,77	91,01	25	84,07	87,68	87,23
26	66,96	67,20	67,58	26	61,66	61,44	62,59	26	67,80	70,19	68,23	26	74,14	72,02	75,27	26	95,03	97,22	93,32	26	86,09	89,82	89,39
27	68,48	68,77	69,11	27	62,92	62,66	63,89	27	69,07	71,65	69,67	27	75,82	73,55	77,03	27	97,38	99,62	95,66	27	88,04	91,92	91,49
28	70,02	70,26	70,55	28	64,25	63,92	65,23	28	70,41	73,15	71,10	28	77,47	75,03	78,77	28	99,63	102,02	97,94	28	89,95	94,01	93,56
29	71,46	71,76	72,02	29	65,59	65,21	66,57	29	71,73	74,65	72,59	29	79,24	76,63	80,59	29	101,88	104,46	100,20	29	91,83	96,07	95,50
30	72,75	73,08	73,32	30	66,92	66,47	67,86	30	73,11	76,17	74,06	30	80,95	78,22	82,40	30	104,15	106,75	102,52	30	93,68	98,05	97,56
31	73,97	74,32	74,54	31	68,23	67,67	69,17	31	74,51	77,73	75,59	31	82,67	79,77	84,19	31	106,44	109,04	104,82	31	95,43	99,94	99,46
32	75,25	75,63	75,81	32	69,53	68,86	70,52	32	75,83	79,22	77,05	32	84,28	81,24	85,90	32	108,71	111,39	107,09	32	97,17	101,86	101,41
33	76,56	76,90	77,10	33	70,88	70,01	71,85	33	77,13	80,67	78,47	33	85,94	82,78	87,69	33	110,75	113,42	109,18	33	99,01	103,89	103,41
34	77,83	78,16	78,34	34	72,13	71,15	73,13	34	78,40	82,12	79,83	34	87,52	84,26	89,34	34	112,95	115,52	111,50	34	100,76	105,80	105,31
35	79,03	79,35	79,50	35	73,38	72,31	74,40	35	79,70	83,56	81,18	35	89,00	85,62	90,92	35	115,22	117,68	113,83	35	102,44	107,55	107,12
36	80,21	80,57	80,74	36	74,63	73,47	75,72	36	80,96	84,95	82,53	36	90,39	86,91	92,40	36	117,41	119,85	116,06	36	104,10	109,23	108,89
37	81,36	81,83	82,00	37	75,97	74,64	77,12	37	82,10	86,22	83,82	37	91,80	88,26	94,01	37	119,55	121,94	118,22	37	105,71	110,91	110,69
38	82,52	83,11	83,28	38	77,27	75,77	78,52	38	83,34	87,53	85,08	38	93,33	89,69	95,67	38	121,52	124,06	120,23	38	107,41	112,66	112,41
39	83,61	84,31	84,53	39	78,52	76,88	79,90	39	84,55	88,85	86,31	39	94,87	91,13	97,39	39	123,52	126,14	122,29	39	109,06	114,32	114,19
40	84,75	85,53	85,77	40	79,64	77,84	81,11	40	85,72	90,11	87,50	40	96,36	92,53	99,01	40	125,52	128,16	124,29	40	110,79	116,13	116,03

APPENDIX 11: THE BEST FORECAST COMBINATION MODELS

Table A.4: The Best Forecast Combination Models of FC Version 1

	t-testi		Wilcoxon testi		DM testi	
	Forecast horizon (days)	Forecasting period (Every 200 data period)	Forecast horizon (days)	Forecasting period (Every 200 data period)	Forecast horizon (days)	Forecasting period (Every 200 data period)
CA	TA	TA	WA	WA	CLS	CLS
DE	INVW	SA	INVW	SA	CLS	CLS
FR	INVW	WA	INVW	WA	CLS	CLS
GB	SA	SA	SA	SA	CLS	CLS
IT	SA	WA	INVW	WA	CLS	CLS
US	WA	NONE	WA	NONE	CLS	CLS

Table A.5: The Best Forecast Combination models of FC Version (2)

	t-testi		Wilcoxon testi		DM testi	
	Forecast horizon (days)	Forecasting period (Every 200 data period)	Forecast horizon (days)	Forecasting period (Every 200 data period)	Forecast horizon (days)	Forecasting period (Every 200 data period)
CA	MED	BG	BG	BG	MED	MED
DE	BG	INVW	BG	MED	BG	MED
FR	INVW	SA	INVW	MED	MED	SA
GB	SA	SA	SA	BG	MED	SA
IT	SA	WA	SA	SA	MED	MED, TA
US	WA	MED	WA	MED, BG	BG	BG

Table A.6:The Best Forecast Combination models of FC Version (3)

	t-testi		Wilcoxon testi		DM testi	
	Forecast horizon (days)	Forecasting period (Every 200 data period)	Forecast horizon (days)	Forecasting period (Every 200 data period)	Forecast horizon (days)	Forecasting period (Every 200 data period)
CA	CLS, BG	CLS, BG	CLS	CLS	CLS, BG	BG
DE	BG	BG	BG	BG	BG	WA, TA
FR	SA	SA	SA	SA	MED	SA
GB	SA	NONE	SA	NONE	INVW	SA
IT	BG	BG	BG	BG	SA	BG
US	SA	SA	CLS	SA	MED	INVW

APPENDIX 12 THE ERRORS OF YIELD CURVE FORECASTS

Table A.7: Canada's The Errors of Yield Curve Forecasts for All Yields at Some Forecast Horizons

CA - 1 day	RMSE	MAE	MASE	SMAPE	CA - 20 day	RMSE	MAE	MASE	SMAPE	CA - 40 day	RMSE	MAE	MASE	SMAPE
RW	0,067	0,051	0,215	0,054	RW	0,193	0,134	0,569	0,122	HYBRID	0,291	0,203	0,859	0,204
FC(V1).EIG3	0,067	0,051	0,215	0,054	FC(V1).EIG3	0,193	0,134	0,569	0,122	RW	0,292	0,200	0,849	0,175
FC(V1).CLS	0,067	0,051	0,215	0,054	BAGGING	0,194	0,139	0,589	0,143	FC(V1).EIG3	0,292	0,200	0,849	0,175
FC(V1).EIG1	0,067	0,051	0,215	0,054	HYBRID	0,194	0,137	0,580	0,135	BAGGING	0,294	0,207	0,881	0,206
ETS/TBATS	0,067	0,051	0,215	0,054	FC(V1).CLS	0,195	0,135	0,573	0,123	ETS/TBATS	0,295	0,198	0,841	0,177
HYBRID	0,067	0,051	0,216	0,054	ETS/TBATS	0,196	0,136	0,576	0,129	FC(V1).CLS	0,295	0,201	0,854	0,177
AR(1)	0,067	0,051	0,216	0,054	Auto.ARIMA	0,209	0,146	0,614	0,163	RF	0,306	0,216	0,916	0,203
Auto.ARIMA	0,067	0,051	0,216	0,055	RF	0,211	0,152	0,646	0,152	XGB	0,309	0,214	0,916	0,197
TAR	0,068	0,051	0,218	0,054	FC(V1).MED	0,212	0,150	0,637	0,136	FC(V1).MED	0,315	0,217	0,921	0,190
FC(V3).MED	0,068	0,051	0,221	0,055	XGB	0,215	0,151	0,646	0,148	FC(V3).BG	0,316	0,215	0,927	0,203
FC(V1).MED	0,068	0,051	0,218	0,054	FC(V3).BG	0,215	0,151	0,649	0,148	FC(V3).INVW	0,317	0,216	0,929	0,204
FC(V3).WA	0,069	0,052	0,222	0,055	FC(V1).TA	0,216	0,152	0,645	0,137	FC(V3).MED	0,318	0,218	0,937	0,195
FC(V3).TA	0,069	0,052	0,222	0,055	FC(V3).INVW	0,216	0,151	0,649	0,147	FC(V1).TA	0,318	0,219	0,929	0,192
FC(V1).TA	0,069	0,052	0,219	0,054	FC(V3).MED	0,217	0,151	0,649	0,142	FC(V3).TA	0,319	0,217	0,930	0,195
FC(V3).BG	0,069	0,052	0,223	0,056	FC(V3).WA	0,217	0,151	0,650	0,141	FC(V3).WA	0,319	0,217	0,930	0,195
FC(V1).WA	0,069	0,052	0,220	0,055	FC(V3).TA	0,217	0,151	0,650	0,142	FC(V1).WA	0,321	0,221	0,935	0,194
FC(V3).INVW	0,069	0,052	0,224	0,056	AR(1)	0,218	0,154	0,654	0,143	FC(V3).CLS	0,326	0,225	0,962	0,229
FC(V3).SA	0,069	0,052	0,224	0,056	FC(V1).WA	0,219	0,154	0,652	0,139	FC(V2).MED	0,327	0,227	0,981	0,199
LL/LST/SET-AR	0,069	0,052	0,222	0,056	FC(V1).INVW	0,219	0,156	0,667	0,145	FC(V3).SA	0,328	0,222	0,951	0,199
BAGGING	0,069	0,052	0,222	0,056	FC(V3).SA	0,221	0,153	0,658	0,145	Auto.ARIMA	0,331	0,223	0,931	0,247
GMDH	0,070	0,052	0,222	0,057	FC(V2).MED	0,224	0,158	0,684	0,147	AR(1)	0,336	0,234	0,994	0,209
FC(V2).MED	0,070	0,052	0,225	0,057	FC(V2).CLS	0,227	0,159	0,690	0,148	FC(V1).INVW	0,336	0,222	0,947	0,194
FC(V3).CLS	0,071	0,053	0,230	0,058	FC(V3).CLS	0,229	0,162	0,692	0,171	FC(V2).WA	0,340	0,231	1,001	0,203
ELM-A	0,072	0,055	0,232	0,060	FC(V2).TA	0,232	0,163	0,704	0,151	FC(V2).TA	0,343	0,232	1,001	0,202
ELM	0,072	0,055	0,233	0,060	FC(V2).WA	0,233	0,164	0,707	0,152	RF-C	0,343	0,231	0,974	0,214
ARFIMA	0,073	0,055	0,232	0,059	TAR	0,233	0,160	0,682	0,154	ENN	0,344	0,255	1,099	0,263
NNET-AR-A	0,073	0,055	0,234	0,060	RF-C	0,233	0,161	0,677	0,153	ENN-A	0,352	0,257	1,100	0,266
NNET-AR	0,073	0,055	0,235	0,060	RF-HP	0,236	0,168	0,716	0,163	BOOSTING	0,359	0,249	1,019	0,239
RF-A	0,074	0,057	0,243	0,064	FC(V2).INVW	0,237	0,166	0,720	0,156	RF-HP	0,366	0,254	1,078	0,232
RF	0,077	0,058	0,247	0,062	ENN	0,252	0,185	0,790	0,199	RF-M	0,370	0,253	1,016	0,227
FC(V1).INVW	0,080	0,061	0,260	0,068	RF-A	0,254	0,179	0,763	0,180	FC(V1).SA	0,373	0,260	1,118	0,238
XGB	0,080	0,060	0,256	0,067	RF-M	0,259	0,183	0,739	0,177	NNET-AR	0,375	0,256	1,068	0,242
ENN-A	0,083	0,060	0,254	0,070	ENN-A	0,259	0,186	0,788	0,209	NNET-AR-A	0,377	0,256	1,069	0,240
FC(V2).TA	0,084	0,062	0,266	0,070	FC(V1).SA	0,263	0,196	0,848	0,195	ARIMA	0,406	0,273	1,188	0,250
FC(V2).WA	0,085	0,062	0,267	0,071	NNET-AR	0,264	0,187	0,785	0,192	RF-A	0,411	0,285	1,204	0,267
ENN	0,085	0,063	0,269	0,079	BOOSTING	0,266	0,187	0,765	0,184	ARFIMA	0,443	0,312	1,294	0,281
FC(V2).CLS	0,088	0,064	0,275	0,071	NNET-AR-A	0,267	0,187	0,786	0,191	ELM-A	0,466	0,309	1,271	0,325
RF-HP	0,088	0,067	0,284	0,075	FC(V2).SA	0,272	0,197	0,850	0,190	GRNN	0,467	0,328	1,435	0,315
RF-C	0,090	0,065	0,274	0,069	FC(V2).CLS	0,288	0,191	0,819	0,188	ELM	0,470	0,313	1,295	0,332
FC(V2).INVW	0,099	0,070	0,302	0,078	FC(V1).EIG1	0,293	0,143	0,597	0,125	JNN-A	0,471	0,279	1,179	0,261
JNN	0,114	0,074	0,313	0,090	ARFIMA	0,305	0,214	0,897	0,203	NNET-A	0,489	0,338	1,394	0,348
RF-M	0,124	0,095	0,387	0,103	ARIMA	0,320	0,210	0,904	0,195	NNET	0,492	0,340	1,404	0,347
FC(V2).SA	0,128	0,094	0,406	0,111	ELM-A	0,321	0,215	0,889	0,234	GRNN-TSF	0,496	0,356	1,532	0,342
FC(V1).SA	0,130	0,100	0,429	0,120	ELM	0,324	0,217	0,900	0,240	JNN	0,498	0,298	1,244	0,284
BOOSTING	0,142	0,092	0,376	0,094	JNN-A	0,380	0,209	0,878	0,202	FC(V1).EIG1	0,509	0,216	0,894	0,180
JNN-A	0,152	0,082	0,349	0,110	GRNN	0,385	0,266	1,150	0,279	MLP-A	0,687	0,560	2,494	0,438
FC(V2).CLS	0,156	0,091	0,390	0,104	NNET	0,397	0,266	1,097	0,280	KNN	0,701	0,511	1,986	0,506
GRNN	0,188	0,127	0,542	0,170	NNET-A	0,406	0,269	1,110	0,283	MLP	0,712	0,581	2,588	0,451
VAR(1)	0,199	0,127	0,537	0,125	KNN	0,414	0,309	1,272	0,344	TAR	0,741	0,271	1,109	0,227
ARIMA	0,208	0,102	0,433	0,091	GRNN-TSF	0,416	0,295	1,270	0,300	LSTM	0,761	0,561	2,261	0,514
VAR/VECM	0,221	0,133	0,564	0,130	JNN	0,431	0,239	0,995	0,224	FC(V2).CLS	0,807	0,276	1,162	0,248
NNETTS	0,225	0,141	0,587	0,201	MLP-A	0,625	0,510	2,273	0,409	RNN	0,832	0,555	2,413	0,460
GRNN-TSF	0,226	0,158	0,679	0,196	MLP	0,649	0,530	2,360	0,421	GRU	0,932	0,618	2,638	0,525
NNET-A	0,236	0,127	0,524	0,141	VAR/VECM	0,653	0,305	1,263	0,284	RNN-A	0,965	0,708	3,362	0,502
NNET	0,236	0,127	0,524	0,141	LL/LST/SET-AR	0,668	0,195	0,815	0,165	NNETTS	1,118	0,422	1,672	0,377
KNN	0,264	0,178	0,750	0,217	LSTM	0,694	0,513	2,067	0,486	FC(V2).INVW	1,490	0,264	1,115	0,205
MLP-A	0,554	0,453	2,023	0,378	RNN	0,767	0,505	2,191	0,437	FC(V2).SA	1,509	0,300	1,253	0,238
MLP	0,580	0,474	2,118	0,390	NNETTS	0,775	0,305	1,241	0,314	FC(V2).CLS	1,866	0,268	1,124	0,200
RNN-A	0,585	0,389	1,685	0,348	GRU	0,876	0,585	2,497	0,511	VAR/VECM	19	0,967	2,676	0,384
LSTM	0,629	0,463	1,858	0,456	RNN-A	0,903	0,656	3,071	0,478	LL/LST/SET-AR	89	3,120	5,017	0,226
RNN	0,642	0,396	1,684	0,388	GMDH	67	1,681	3,284	0,208	GMDH	43662	544	1,215	0,303
GRU	0,818	0,549	2,344	0,495	VAR(1)	330	7,102	4,229	0,277	FC(V1).EIG2	9E+05	5E+04	4,230	0,201
FC(V1).EIG2	8,524	0,461	1,409	0,062	FC(V1).EIG2	4E+05	2E+04	4,231	0,146	VAR(1)	2E+06	4E+04	4,832	0,376

Table A.8: Germany's The Errors of Yield Curve Forecasts for All Yields at Some Forecast Horizons

DE - 1 day	RMSE	MAE	MASE	SMAPE	DE - 20 day	RMSE	MAE	MASE	SMAPE	DE - 40 day	RMSE	MAE	MASE	SMAPE
RW	0,058	0,041	0,235	0,188	FC(V4)	0,171	0,114	0,645	0,377	FC(V4)	0,245	0,158	0,901	0,470
FC(V1).EIG3	0,058	0,041	0,235	0,188	ETS/TBATS	0,171	0,114	0,647	0,384	RW	0,253	0,161	0,911	0,473
ETS/TBATS	0,058	0,042	0,235	0,188	RW	0,172	0,114	0,646	0,380	FC(V1).EIG3	0,253	0,161	0,911	0,473
FC(V4)	0,058	0,042	0,235	0,187	FC(V1).EIG3	0,172	0,114	0,646	0,380	FC(V1).CLS	0,255	0,164	0,926	0,477
FC(V1).CLS	0,058	0,042	0,236	0,188	FC(V1).CLS	0,173	0,115	0,652	0,382	FC(V3).BG	0,255	0,166	0,944	0,488
AR(1)	0,058	0,042	0,235	0,188	FC(V1).EIG1	0,174	0,116	0,655	0,383	ETS/TBATS	0,257	0,161	0,911	0,465
FC(V1).EIG1	0,059	0,042	0,236	0,188	FC(V1).MED	0,183	0,121	0,678	0,397	FC(V1).EIG1	0,257	0,165	0,935	0,479
FC(V3).MED	0,059	0,042	0,239	0,190	FC(V1).TA	0,184	0,122	0,684	0,403	FC(V3).INVW	0,264	0,172	0,978	0,502
Auto.ARIMA	0,059	0,042	0,238	0,189	FC(V3).BG	0,184	0,125	0,708	0,403	FC(V3).WA	0,264	0,171	0,969	0,492
TAR	0,059	0,042	0,239	0,189	FC(V3).MED	0,184	0,126	0,715	0,406	FC(V3).TA	0,264	0,171	0,970	0,492
FC(V3).WA	0,059	0,042	0,241	0,191	FC(V1).WA	0,185	0,123	0,688	0,407	FC(V1).TA	0,264	0,168	0,936	0,483
FC(V3).TA	0,059	0,042	0,241	0,191	FC(V3).WA	0,186	0,125	0,710	0,401	FC(V1).WA	0,264	0,168	0,935	0,483
FC(V1).MED	0,059	0,042	0,238	0,190	FC(V3).TA	0,186	0,125	0,711	0,401	FC(V1).MED	0,265	0,169	0,942	0,480
FC(V1).TA	0,060	0,042	0,239	0,192	FC(V1).INVW	0,187	0,128	0,718	0,417	FC(V1).INVW	0,265	0,174	0,970	0,502
FC(V2).MED	0,060	0,042	0,243	0,192	FC(V2).MED	0,187	0,127	0,726	0,412	FC(V3).MED	0,266	0,174	0,989	0,499
FC(V1).WA	0,060	0,042	0,239	0,192	AR(1)	0,187	0,126	0,703	0,401	FC(V2).WA	0,269	0,180	1,031	0,523
FC(V3).SA	0,060	0,043	0,244	0,193	FC(V3).INVW	0,189	0,128	0,726	0,410	FC(V3).SA	0,270	0,176	0,993	0,507
FC(V3).BG	0,060	0,043	0,245	0,194	FC(V3).SA	0,191	0,129	0,729	0,412	FC(V2).MED	0,270	0,178	1,018	0,509
FC(V3).INVW	0,060	0,043	0,245	0,194	FC(V2).WA	0,192	0,132	0,757	0,430	FC(V3).CLS	0,280	0,186	1,060	0,533
ELM-A	0,060	0,043	0,242	0,192	BAGGING	0,198	0,128	0,722	0,405	RF	0,283	0,187	1,065	0,527
ELM	0,060	0,043	0,243	0,192	FC(V2).TA	0,198	0,134	0,764	0,431	AR(1)	0,285	0,187	1,022	0,509
BAGGING	0,060	0,043	0,244	0,192	FC(V2).CLS	0,203	0,130	0,742	0,415	XGB	0,285	0,191	1,083	0,540
GMDH	0,060	0,043	0,243	0,194	RF-C	0,205	0,143	0,783	0,456	ARIMA	0,288	0,178	0,995	0,503
LL/LST/SET-AR	0,062	0,044	0,247	0,194	FC(V3).CLS	0,211	0,143	0,811	0,444	RF-C	0,292	0,197	1,064	0,544
ARFIMA	0,063	0,045	0,253	0,201	Auto.ARIMA	0,212	0,134	0,756	0,416	BOOSTING	0,303	0,215	1,180	0,598
FC(V3).CLS	0,063	0,045	0,257	0,199	XGB	0,212	0,145	0,823	0,447	RF-M	0,308	0,208	1,099	0,579
NNET-AR-A	0,067	0,048	0,268	0,207	RF	0,213	0,141	0,802	0,434	FC(V1).SA	0,314	0,227	1,253	0,647
NNET-AR	0,067	0,048	0,269	0,208	RF-A	0,216	0,145	0,798	0,453	ENN	0,315	0,217	1,218	0,602
RF-A	0,068	0,050	0,283	0,217	RF-HP	0,217	0,145	0,810	0,453	BAGGING	0,320	0,189	1,062	0,507
RF	0,070	0,049	0,279	0,210	FC(V2).INVW	0,221	0,137	0,781	0,432	ELM	0,325	0,229	1,252	0,620
ENN-A	0,072	0,050	0,283	0,219	ARIMA	0,222	0,134	0,756	0,424	ELM-A	0,326	0,232	1,262	0,631
FC(V1).INVW	0,073	0,055	0,308	0,239	ELM-A	0,223	0,157	0,864	0,489	NNET-AR	0,327	0,217	1,205	0,583
XGB	0,074	0,053	0,298	0,225	ELM	0,225	0,157	0,865	0,483	RF-HP	0,327	0,204	1,127	0,558
FC(V2).WA	0,074	0,053	0,304	0,234	BOOSTING	0,230	0,163	0,903	0,498	NNET-AR-A	0,328	0,217	1,203	0,582
FC(V2).TA	0,075	0,054	0,311	0,238	RF-M	0,233	0,164	0,876	0,512	RF-A	0,330	0,216	1,164	0,577
ARIMA	0,076	0,050	0,284	0,207	VAR(1)	0,238	0,143	0,806	0,428	Auto.ARIMA	0,347	0,199	1,117	0,516
ENN	0,081	0,056	0,316	0,237	FC(V1).SA	0,245	0,186	1,034	0,580	GRNN	0,366	0,251	1,452	0,651
FC(V2).CLS	0,081	0,053	0,308	0,232	FC(V2).SA	0,246	0,152	0,865	0,486	ARFIMA	0,377	0,229	1,244	0,583
RF-HP	0,084	0,061	0,343	0,244	NNET-AR	0,251	0,172	0,951	0,495	VAR(1)	0,403	0,216	1,205	0,534
FC(V2).INVW	0,089	0,058	0,333	0,248	NNET-AR-A	0,251	0,173	0,953	0,497	ENN-A	0,404	0,254	1,435	0,621
RF-C	0,092	0,063	0,353	0,265	ENN	0,256	0,178	0,995	0,521	GRNN-TSF	0,409	0,280	1,635	0,700
JNN	0,104	0,067	0,373	0,266	ARFIMA	0,268	0,165	0,904	0,478	JNN	0,434	0,253	1,397	0,622
FC(V2).SA	0,106	0,079	0,455	0,324	GRNN	0,273	0,192	1,105	0,545	FC(V2).CLS	0,536	0,246	1,377	0,589
JNN-A	0,113	0,064	0,357	0,249	FC(V2).CLS	0,278	0,173	0,983	0,490	NNETTS	0,566	0,319	1,775	0,721
VAR(1)	0,120	0,070	0,396	0,274	JNN	0,303	0,204	1,132	0,566	JNN-A	0,579	0,265	1,488	0,605
BOOSTING	0,127	0,074	0,419	0,286	GRNN-TSF	0,318	0,221	1,280	0,590	RNN	0,674	0,383	2,236	0,765
FC(V1).SA	0,140	0,114	0,638	0,421	ENN-A	0,333	0,202	1,134	0,533	NNET	0,692	0,375	2,090	0,746
GRNN	0,141	0,094	0,537	0,326	JNN-A	0,381	0,195	1,093	0,523	LSTM	0,693	0,418	2,140	0,835
RF-M	0,153	0,110	0,598	0,416	KNN	0,409	0,313	1,727	0,761	NNET-A	0,695	0,378	2,099	0,749
FC(V2).CLS	0,167	0,084	0,483	0,293	NNETTS	0,488	0,265	1,468	0,629	GRU	0,708	0,457	2,271	0,909
GRNN-TSF	0,180	0,119	0,673	0,367	RNN	0,612	0,334	1,948	0,721	KNN	0,742	0,559	2,978	1,014
KNN	0,244	0,175	0,965	0,513	NNET	0,614	0,318	1,768	0,655	MLP-A	0,745	0,625	3,546	1,338
NNETTS	0,293	0,148	0,837	0,423	NNET-A	0,615	0,319	1,772	0,656	MLP	0,775	0,652	3,679	1,341
RNN	0,509	0,248	1,432	0,596	LSTM	0,641	0,387	1,983	0,798	RNN-A	0,815	0,577	3,647	1,040
RNN-A	0,525	0,320	1,945	0,718	GRU	0,651	0,427	2,122	0,882	FC(V2).TA	6,615	0,301	1,466	0,526
NNET	0,534	0,226	1,266	0,467	MLP-A	0,710	0,593	3,366	1,304	FC(V2).CLS	26	0,807	3,053	0,505
NNET-A	0,534	0,226	1,267	0,467	RNN-A	0,735	0,518	3,283	0,991	FC(V2).INVW	37	1,054	3,577	0,526
GRU	0,589	0,395	1,967	0,855	MLP	0,737	0,618	3,490	1,308	FC(V2).SA	48	1,325	3,999	0,582
LSTM	0,593	0,362	1,852	0,769	TAR	1,568	0,183	0,994	0,429	TAR	570	13	6,698	0,543
MLP-A	0,672	0,558	3,176	1,267	GMDH	17	0,543	2,460	0,505	GMDH	2605	92	7,914	0,659
MLP	0,703	0,588	3,334	1,280	FC(V1).EIG2	3E+08	1E+07	8,947	0,397	FC(V1).EIG2	7E+08	2E+07	8,936	0,491
FC(V1).EIG2	5912	140	8,295	0,204	LL/LST/SET-AR	1E+19	2E+17	5,263	0,458	LL/LST/SET-AR	2E+43	5E+41	5,152	0,572

Table A.9: France's The Errors of Yield Curve Forecasts for All Yields at Some Forecast Horizons

FR - 1 day	RMSE	MAE	MASE	SMAPE	FR - 20 day	RMSE	MAE	MASE	SMAPE	FR - 40 day	RMSE	MAE	MASE	SMAPE
RW	0,107	0,071	0,295	0,253	RW	0,197	0,128	0,534	0,371	ETS/TBATS	0,269	0,178	0,741	0,468
FC(V1).EIG3	0,107	0,071	0,295	0,253	FC(V1).EIG3	0,197	0,128	0,534	0,371	FC(V4)	0,276	0,179	0,746	0,468
FC(V1).EIG1	0,107	0,071	0,295	0,253	FC(V1).EIG1	0,197	0,128	0,534	0,371	FC(V1).EIG1	0,278	0,179	0,748	0,460
AR(1)	0,107	0,071	0,295	0,254	ETS/TBATS	0,197	0,129	0,538	0,382	RW	0,278	0,179	0,747	0,460
ETS/TBATS	0,107	0,071	0,295	0,254	FC(V4)	0,197	0,129	0,537	0,381	FC(V1).EIG3	0,278	0,179	0,747	0,460
FC(V4)	0,107	0,071	0,295	0,254	FC(V1).CLS	0,198	0,128	0,533	0,372	FC(V1).CLS	0,280	0,179	0,746	0,461
FC(V1).CLS	0,107	0,071	0,295	0,253	FC(V1).MED	0,210	0,139	0,575	0,394	XGB	0,287	0,192	0,801	0,488
Auto.ARIMA	0,108	0,071	0,296	0,255	FC(V1).WA	0,211	0,139	0,578	0,396	FC(V3).BG	0,290	0,193	0,813	0,504
TAR	0,108	0,071	0,297	0,256	FC(V1).TA	0,211	0,139	0,577	0,395	FC(V3).TA	0,291	0,193	0,813	0,497
ARFIMA	0,108	0,073	0,302	0,258	AR(1)	0,212	0,142	0,587	0,400	FC(V3).WA	0,292	0,193	0,812	0,496
FC(V1).MED	0,108	0,071	0,295	0,250	FC(V3).MED	0,212	0,141	0,591	0,406	FC(V1).WA	0,292	0,191	0,794	0,487
FC(V1).WA	0,108	0,071	0,296	0,251	FC(V3).TA	0,213	0,142	0,598	0,409	FC(V3).MED	0,292	0,192	0,810	0,494
FC(V3).MED	0,108	0,071	0,301	0,255	FC(V3).WA	0,213	0,142	0,598	0,408	FC(V1).TA	0,293	0,191	0,794	0,485
FC(V1).TA	0,108	0,071	0,296	0,251	FC(V2).MED	0,215	0,146	0,614	0,426	FC(V3).INVW	0,294	0,196	0,823	0,506
BAGGING	0,109	0,072	0,299	0,256	BAGGING	0,215	0,137	0,571	0,393	RF	0,295	0,195	0,818	0,503
FC(V2).MED	0,109	0,072	0,302	0,255	FC(V3).BG	0,215	0,144	0,603	0,419	FC(V1).MED	0,295	0,193	0,801	0,488
FC(V3).WA	0,109	0,072	0,303	0,257	XGB	0,218	0,146	0,610	0,411	FC(V3).SA	0,298	0,197	0,827	0,500
FC(V3).TA	0,109	0,072	0,303	0,257	FC(V3).INVW	0,218	0,145	0,610	0,421	FC(V2).MED	0,300	0,204	0,860	0,525
ELM-A	0,109	0,072	0,300	0,254	FC(V3).SA	0,218	0,146	0,613	0,418	FC(V2).TA	0,309	0,211	0,890	0,548
GMDH	0,109	0,072	0,299	0,258	RF	0,219	0,145	0,608	0,413	FC(V2).WA	0,309	0,212	0,892	0,549
ELM	0,109	0,072	0,300	0,254	FC(V1).INVW	0,222	0,150	0,620	0,430	AR(1)	0,310	0,208	0,856	0,514
LL/LST/SET-AR	0,109	0,072	0,302	0,257	FC(V2).TA	0,223	0,153	0,642	0,448	FC(V3).CLS	0,314	0,216	0,908	0,560
FC(V3).BG	0,109	0,073	0,306	0,261	ARIMA	0,223	0,150	0,622	0,429	RF-C	0,319	0,219	0,896	0,557
FC(V3).SA	0,109	0,073	0,307	0,261	FC(V2).WA	0,223	0,153	0,644	0,450	ARIMA	0,322	0,210	0,869	0,525
FC(V3).INVW	0,109	0,073	0,307	0,260	RF-C	0,226	0,157	0,647	0,448	BAGGING	0,322	0,200	0,829	0,496
NNET-AR-A	0,110	0,074	0,307	0,256	RF-HP	0,232	0,157	0,648	0,439	NNET-AR	0,325	0,226	0,929	0,560
NNET-AR	0,110	0,074	0,307	0,256	FC(V3).CLS	0,238	0,161	0,678	0,462	NNET-AR-A	0,326	0,228	0,935	0,563
FC(V3).CLS	0,112	0,075	0,315	0,267	RF-A	0,244	0,166	0,676	0,453	RF-HP	0,343	0,227	0,924	0,552
RF-A	0,113	0,077	0,323	0,274	NNET-AR	0,251	0,174	0,715	0,474	ENN	0,346	0,251	1,056	0,633
RF	0,113	0,077	0,322	0,272	NNET-AR-A	0,252	0,175	0,717	0,474	RF-M	0,349	0,240	0,980	0,586
XGB	0,117	0,080	0,334	0,277	Auto.ARIMA	0,252	0,155	0,646	0,421	BOOSTING	0,363	0,262	1,078	0,608
ARIMA	0,118	0,079	0,329	0,280	RF-M	0,261	0,187	0,767	0,516	RF-A	0,367	0,245	0,982	0,576
FC(V1).INVW	0,119	0,080	0,333	0,293	BOOSTING	0,266	0,190	0,784	0,493	ENN-A	0,368	0,264	1,108	0,646
ENN-A	0,120	0,081	0,339	0,287	ARFIMA	0,273	0,174	0,714	0,447	GRNN	0,381	0,264	1,081	0,598
FC(V2).WA	0,120	0,082	0,345	0,294	ELM-A	0,273	0,190	0,780	0,500	ARFIMA	0,382	0,240	0,974	0,546
FC(V2).TA	0,121	0,082	0,346	0,294	ELM	0,273	0,191	0,782	0,503	ELM	0,385	0,268	1,094	0,619
RF-HP	0,123	0,086	0,357	0,295	ENN	0,280	0,200	0,835	0,544	ELM-A	0,387	0,269	1,099	0,618
RF-C	0,124	0,087	0,361	0,310	ENN-A	0,288	0,205	0,852	0,546	GRNN-TSF	0,422	0,297	1,231	0,658
FC(V2).CLS	0,133	0,086	0,362	0,302	GRNN	0,289	0,207	0,848	0,529	FC(V1).INVW	0,423	0,207	0,856	0,516
ENN	0,133	0,093	0,386	0,321	FC(V1).SA	0,292	0,214	0,868	0,592	Auto.ARIMA	0,423	0,241	1,001	0,544
FC(V2).INVW	0,143	0,091	0,381	0,321	TAR	0,292	0,168	0,705	0,456	JNN-A	0,451	0,289	1,135	0,629
BOOSTING	0,143	0,088	0,369	0,294	GRNN-TSF	0,353	0,250	1,030	0,591	JNN	0,498	0,329	1,290	0,699
FC(V2).SA	0,156	0,114	0,472	0,396	JNN-A	0,377	0,227	0,887	0,529	NNET	0,524	0,360	1,426	0,749
RF-M	0,169	0,131	0,541	0,435	JNN	0,413	0,258	1,008	0,598	NNET-A	0,547	0,370	1,465	0,758
GRNN	0,175	0,129	0,534	0,403	NNET-A	0,449	0,302	1,188	0,661	GRU	0,675	0,485	1,766	0,878
FC(V1).SA	0,177	0,137	0,558	0,449	KNN	0,454	0,338	1,385	0,761	MLP-A	0,737	0,613	2,579	1,131
GRNN-TSF	0,228	0,153	0,631	0,424	NNET	0,455	0,303	1,193	0,662	MLP	0,758	0,625	2,629	1,135
KNN	0,249	0,166	0,692	0,473	NNETTS	0,497	0,268	1,067	0,592	KNN	0,830	0,619	2,446	1,023
NNETTS	0,253	0,171	0,682	0,456	GRU	0,634	0,451	1,642	0,841	FC(V1).SA	0,863	0,276	1,114	0,655
FC(V2).CLS	0,265	0,143	0,592	0,397	MLP-A	0,704	0,580	2,441	1,096	LSTM	0,952	0,550	2,274	0,793
JNN-A	0,269	0,120	0,471	0,304	MLP	0,725	0,592	2,492	0,000	RNN	1,007	0,665	2,796	1,013
JNN	0,275	0,129	0,506	0,336	FC(V2).CLS	0,735	0,240	0,981	0,555	NNETTS	1,138	0,344	1,354	0,671
NNET	0,334	0,201	0,792	0,456	FC(V2).INVW	0,823	0,177	0,734	0,466	RNN-A	1,150	0,919	4,018	1,303
NNET-A	0,335	0,202	0,798	0,458	LSTM	0,924	0,518	2,142	0,754	FC(V2).CLS	4,810	0,403	1,553	0,671
VAR(1)	0,551	0,288	1,039	0,571	RNN	0,955	0,608	2,555	0,953	TAR	10	0,761	2,746	0,583
GRU	0,597	0,423	1,539	0,808	RNN-A	1,098	0,865	3,783	1,264	GMDH	182	4,819	6,010	0,662
MLP-A	0,671	0,551	2,344	1,077	FC(V2).CLS	1,128	0,180	0,744	0,454	FC(V2).INVW	247	5,466	6,101	0,565
MLP	0,694	0,568	2,407	1,088	LL/LST/SET-AR	1,431	0,232	0,928	0,489	LL/LST/SET-AR	345	7,909	4,535	0,606
RNN-A	0,735	0,492	2,115	0,940	FC(V2).SA	1,949	0,223	0,904	0,527	FC(V2).CLS	354	7,748	6,562	0,561
RNN	0,771	0,422	1,761	0,765	GMDH	6,791	0,435	1,546	0,530	FC(V2).SA	486	11	6,902	0,615
LSTM	0,901	0,488	2,020	0,717	VAR(1)	202	4,707	5,738	0,756	VAR(1)	5E+05	1E+04	7,826	0,845
FC(V1).EIG2	1,571	0,207	0,911	0,357	FC(V1).EIG2	3E+07	1E+06	4,627	0,465	FC(V1).EIG2	6E+07	3E+06	4,635	0,548

Table A.10: Great Britain's The Errors of Yield Curve Forecasts for All Yields at Some Forecast Horizons

GB - 1 day	RMSE	MAE	MASE	SMAPE	GB - 20 day	RMSE	MAE	MASE	SMAPE	GB - 40 day	RMSE	MAE	MASE	SMAPE
ETS/TBATS	0,077	0,054	0,289	0,117	FC(V1).CLS	0,230	0,146	0,785	0,229	FC(V1).CLS	0,330	0,202	1,086	0,292
RW	0,077	0,054	0,289	0,116	RW	0,230	0,146	0,786	0,229	RW	0,330	0,202	1,087	0,292
FC(V1).EIG1	0,077	0,054	0,289	0,116	FC(V1).EIG1	0,230	0,146	0,786	0,229	FC(V1).EIG1	0,330	0,202	1,087	0,292
FC(V1).EIG3	0,077	0,054	0,289	0,116	FC(V1).EIG3	0,230	0,146	0,786	0,229	FC(V1).EIG3	0,330	0,202	1,087	0,292
FC(V1).CLS	0,077	0,054	0,289	0,116	ETS/TBATS	0,236	0,149	0,798	0,232	ETS/TBATS	0,336	0,205	1,102	0,295
Auto.ARIMA	0,078	0,054	0,289	0,117	FC(V4)	0,240	0,151	0,811	0,238	XGB	0,338	0,214	1,151	0,319
AR(1)	0,078	0,054	0,289	0,117	AR(1)	0,246	0,159	0,837	0,274	RF	0,340	0,219	1,175	0,327
FC(V4)	0,078	0,054	0,289	0,117	RF	0,248	0,163	0,877	0,265	FC(V4)	0,346	0,212	1,140	0,309
TAR	0,078	0,055	0,293	0,120	XGB	0,251	0,161	0,866	0,259	ARIMA	0,354	0,229	1,187	0,359
FC(V1).MED	0,079	0,054	0,291	0,117	ARIMA	0,257	0,169	0,883	0,285	FC(V3).TA	0,357	0,226	1,220	0,342
FC(V3).MED	0,079	0,054	0,297	0,119	BAGGING	0,258	0,156	0,840	0,250	FC(V3).WA	0,358	0,226	1,221	0,342
FC(V1).TA	0,079	0,054	0,292	0,117	Auto.ARIMA	0,259	0,156	0,838	0,244	FC(V3).BG	0,358	0,226	1,226	0,337
FC(V1).WA	0,079	0,055	0,292	0,117	FC(V1).INVW	0,260	0,163	0,866	0,257	FC(V3).INVW	0,358	0,225	1,217	0,340
ARFIMA	0,080	0,055	0,295	0,118	FC(V3).MED	0,265	0,163	0,882	0,265	FC(V3).MED	0,359	0,227	1,224	0,342
FC(V2).MED	0,080	0,055	0,300	0,120	FC(V3).TA	0,268	0,166	0,903	0,268	FC(V3).SA	0,360	0,228	1,226	0,348
FC(V3).WA	0,080	0,056	0,305	0,121	FC(V2).MED	0,269	0,168	0,912	0,268	AR(1)	0,364	0,233	1,205	0,368
FC(V3).TA	0,080	0,056	0,305	0,121	FC(V3).WA	0,269	0,167	0,905	0,269	FC(V1).MED	0,365	0,226	1,189	0,317
ELM-A	0,081	0,056	0,300	0,124	FC(V1).WA	0,269	0,164	0,869	0,252	FC(V1).WA	0,365	0,229	1,207	0,326
BAGGING	0,081	0,056	0,301	0,121	FC(V1).TA	0,269	0,164	0,865	0,250	FC(V1).TA	0,366	0,229	1,203	0,323
ELM	0,081	0,056	0,300	0,124	FC(V3).BG	0,270	0,168	0,911	0,265	FC(V2).MED	0,369	0,235	1,272	0,349
LL/LST/SET-AR	0,082	0,056	0,300	0,121	FC(V1).MED	0,270	0,162	0,859	0,246	FC(V1).INVW	0,373	0,227	1,203	0,323
FC(V3).SA	0,084	0,059	0,321	0,126	RF-HP	0,271	0,175	0,918	0,292	RF-C	0,382	0,250	1,306	0,368
FC(V3).BG	0,085	0,058	0,319	0,126	FC(V3).SA	0,272	0,170	0,918	0,275	RF-M	0,390	0,256	1,296	0,364
FC(V3).INVW	0,086	0,058	0,320	0,126	FC(V3).INVW	0,272	0,168	0,910	0,267	FC(V3).CLS	0,391	0,247	1,331	0,365
RF-A	0,086	0,061	0,327	0,136	FC(V2).TA	0,272	0,173	0,943	0,285	BAGGING	0,392	0,222	1,189	0,320
GMDH	0,087	0,056	0,301	0,120	FC(V2).WA	0,272	0,173	0,945	0,284	Auto.ARIMA	0,394	0,224	1,209	0,323
RF	0,087	0,060	0,322	0,130	RF-A	0,274	0,178	0,932	0,285	RF-HP	0,395	0,248	1,289	0,359
NNET-AR-A	0,087	0,059	0,316	0,126	RF-C	0,281	0,184	0,968	0,290	GRNN	0,401	0,264	1,369	0,409
NNET-AR	0,088	0,059	0,317	0,126	RF-M	0,285	0,193	0,989	0,300	GRNN-TSF	0,404	0,265	1,375	0,390
XGB	0,089	0,062	0,333	0,137	FC(V2).SA	0,299	0,195	1,068	0,315	BOOSTING	0,411	0,272	1,397	0,433
FC(V1).INVW	0,093	0,065	0,346	0,152	GRNN	0,300	0,204	1,059	0,341	FC(V1).SA	0,412	0,269	1,402	0,366
FC(V2).WA	0,094	0,065	0,354	0,149	FC(V3).CLS	0,306	0,188	1,016	0,292	NNET-AR	0,419	0,275	1,482	0,423
FC(V2).TA	0,095	0,066	0,359	0,150	VAR(1)	0,313	0,209	1,163	0,318	RF-A	0,421	0,276	1,406	0,402
ARIMA	0,097	0,066	0,349	0,137	BOOSTING	0,315	0,207	1,075	0,342	NNET-AR-A	0,421	0,274	1,480	0,422
RF-HP	0,102	0,071	0,381	0,153	FC(V1).SA	0,316	0,206	1,077	0,314	JNN-A	0,421	0,294	1,577	0,407
FC(V3).CLS	0,110	0,064	0,349	0,136	ARFIMA	0,316	0,192	0,988	0,294	JNN	0,434	0,306	1,622	0,449
FC(V2).CLS	0,115	0,070	0,384	0,140	GRNN-TSF	0,321	0,213	1,109	0,339	ARFIMA	0,437	0,271	1,368	0,378
JNN-A	0,120	0,069	0,371	0,136	JNN	0,330	0,235	1,254	0,366	ENN	0,444	0,269	1,433	0,399
RF-C	0,125	0,081	0,431	0,158	ELM	0,332	0,224	1,145	0,364	VAR(1)	0,454	0,300	1,647	0,410
JNN	0,125	0,074	0,397	0,144	FC(V2).INVW	0,334	0,183	0,995	0,279	ELM	0,457	0,314	1,570	0,460
FC(V2).SA	0,138	0,096	0,524	0,204	ELM-A	0,336	0,227	1,157	0,363	ELM-A	0,463	0,318	1,584	0,460
FC(V2).INVW	0,140	0,079	0,431	0,160	NNET-AR	0,343	0,210	1,127	0,341	ENN-A	0,472	0,284	1,498	0,412
GRNN	0,142	0,100	0,520	0,224	NNET-AR-A	0,344	0,209	1,124	0,339	MLP-A	0,690	0,500	2,614	0,529
RF-M	0,145	0,110	0,569	0,215	JNN-A	0,344	0,230	1,229	0,338	MLP	0,691	0,499	2,566	0,536
BOOSTING	0,157	0,097	0,510	0,200	FC(V2).CLS	0,361	0,179	0,969	0,269	NNET	0,728	0,400	2,111	0,541
FC(V1).SA	0,159	0,113	0,598	0,226	ENN	0,368	0,224	1,190	0,361	NNET-A	0,729	0,401	2,113	0,542
GRNN-TSF	0,170	0,113	0,589	0,239	ENN-A	0,422	0,239	1,250	0,369	LSTM	0,776	0,499	2,385	0,578
ENN	0,179	0,086	0,459	0,168	KNN	0,533	0,384	2,009	0,552	GRU	0,861	0,578	2,773	0,657
KNN	0,188	0,123	0,656	0,238	MLP-A	0,636	0,458	2,395	0,503	RNN-A	0,919	0,730	4,163	0,657
ENN-A	0,204	0,083	0,442	0,154	MLP	0,639	0,458	2,351	0,509	RNN	0,947	0,673	3,533	0,695
NNETTS	0,208	0,116	0,610	0,227	FC(V2).CLS	0,648	0,231	1,228	0,321	KNN	0,977	0,701	3,398	0,796
VAR(1)	0,213	0,119	0,642	0,182	NNET	0,654	0,334	1,762	0,464	NNETTS	1,472	0,548	2,699	0,542
FC(V2).CLS	0,295	0,118	0,640	0,201	NNET-A	0,656	0,335	1,765	0,466	FC(V2).SA	16	0,682	2,424	0,381
MLP	0,573	0,404	2,081	0,467	TAR	0,672	0,210	1,093	0,298	FC(V2).WA	16	0,655	2,341	0,363
MLP-A	0,574	0,404	2,102	0,466	LSTM	0,724	0,456	2,179	0,546	FC(V2).TA	16	0,655	2,343	0,364
NNET	0,579	0,220	1,169	0,295	GRU	0,821	0,542	2,601	0,630	FC(V2).INVW	60	1,828	3,277	0,359
NNET-A	0,579	0,220	1,170	0,295	RNN-A	0,891	0,688	3,936	0,624	FC(V2).CLS	78	2,316	3,396	0,351
RNN-A	0,617	0,427	2,369	0,492	RNN	0,913	0,623	3,275	0,655	FC(V2).CLS	178	4,906	3,763	0,396
LSTM	0,688	0,417	2,001	0,504	NNETTS	1,257	0,403	2,029	0,445	TAR	182	5,241	3,836	0,427
RNN	0,761	0,471	2,456	0,559	LL/LST/SET-AR	23	0,835	2,604	0,327	GMDH	194	4,669	3,644	0,502
GRU	0,774	0,503	2,414	0,602	GMDH	23	0,866	2,977	0,357	LL/LST/SET-AR	5E+04	906	3,735	0,438
FC(V1).EIG2	2,260	0,104	0,525	0,119	FC(V1).EIG2	1E+07	4E+05	4,974	0,236	FC(V1).EIG2	3E+07	7E+05	4,969	0,298

Table A.11: Italy's The Errors of Yield Curve Forecasts for All Yields at Some Forecast Horizons

IT - 1 day	RMSE	MAE	MASE	SMAPE	IT - 20 day	RMSE	MAE	MASE	SMAPE	IT - 40 day	RMSE	MAE	MASE	SMAPE
RW	0,087	0,064	0,167	0,180	RW	0,276	0,188	0,493	0,337	RW	0,394	0,276	0,724	0,435
FC(V1).EIG3	0,087	0,064	0,167	0,180	FC(V1).EIG3	0,276	0,188	0,493	0,337	FC(V1).EIG3	0,394	0,276	0,724	0,435
FC(V4)	0,087	0,064	0,167	0,181	FC(V1).CLS	0,276	0,189	0,495	0,339	FC(V1).CLS	0,394	0,278	0,727	0,438
FC(V1).CLS	0,087	0,064	0,167	0,180	FC(V1).EIG1	0,278	0,189	0,496	0,339	FC(V1).EIG1	0,397	0,279	0,732	0,438
FC(V1).EIG1	0,087	0,064	0,167	0,180	FC(V3).MED	0,287	0,201	0,530	0,368	FC(V3).BG	0,405	0,294	0,777	0,487
Auto.ARIMA	0,088	0,064	0,168	0,185	FC(V3).BG	0,289	0,204	0,539	0,386	FC(V3).MED	0,407	0,293	0,774	0,473
FC(V3).MED	0,088	0,064	0,169	0,184	FC(V3).WA	0,289	0,202	0,532	0,375	FC(V3).TA	0,409	0,294	0,775	0,479
TAR	0,088	0,065	0,169	0,183	FC(V3).TA	0,289	0,202	0,533	0,376	FC(V3).WA	0,410	0,294	0,776	0,479
FC(V1).MED	0,088	0,065	0,170	0,184	ETS/TBATS	0,291	0,196	0,512	0,349	FC(V3).INVW	0,416	0,303	0,798	0,497
FC(V1).TA	0,089	0,065	0,171	0,183	FC(V4)	0,295	0,197	0,515	0,352	XGB	0,418	0,298	0,783	0,470
FC(V1).WA	0,089	0,065	0,171	0,183	FC(V3).INVW	0,295	0,210	0,553	0,396	FC(V2).MED	0,420	0,297	0,786	0,469
ETS/TBATS	0,090	0,065	0,169	0,181	FC(V1).MED	0,298	0,211	0,553	0,383	ETS/TBATS	0,421	0,287	0,750	0,447
FC(V2).MED	0,090	0,065	0,172	0,186	FC(V2).MED	0,300	0,209	0,552	0,373	FC(V1).MED	0,427	0,313	0,816	0,502
ELM-A	0,091	0,067	0,175	0,199	AR(1)	0,301	0,208	0,543	0,363	RF	0,427	0,314	0,829	0,498
ELM	0,092	0,068	0,177	0,201	XGB	0,301	0,211	0,556	0,378	FC(V3).SA	0,430	0,311	0,818	0,501
ARFIMA	0,092	0,068	0,177	0,190	FC(V1).TA	0,303	0,217	0,566	0,389	FC(V2).WA	0,432	0,305	0,807	0,484
BAGGING	0,092	0,067	0,176	0,187	FC(V3).SA	0,305	0,217	0,571	0,401	FC(V1).TA	0,433	0,320	0,835	0,503
FC(V3).WA	0,093	0,067	0,177	0,192	FC(V1).WA	0,305	0,218	0,569	0,391	FC(V4)	0,435	0,285	0,745	0,454
FC(V3).TA	0,094	0,067	0,178	0,193	FC(V2).TA	0,306	0,214	0,564	0,385	FC(V1).WA	0,436	0,323	0,843	0,506
NNET-AR-A	0,095	0,068	0,179	0,192	FC(V2).WA	0,307	0,215	0,568	0,389	AR(1)	0,436	0,314	0,818	0,482
NNET-AR	0,095	0,069	0,180	0,192	BAGGING	0,309	0,208	0,542	0,374	FC(V1).INVW	0,440	0,327	0,852	0,510
FC(V3).BG	0,099	0,070	0,185	0,198	RF	0,311	0,223	0,588	0,402	RF-HP	0,453	0,331	0,861	0,499
FC(V3).INVW	0,099	0,071	0,188	0,201	RF-HP	0,313	0,223	0,583	0,394	FC(V3).CLS	0,456	0,332	0,877	0,539
RF-A	0,101	0,075	0,195	0,208	FC(V3).CLS	0,322	0,231	0,609	0,434	BAGGING	0,459	0,314	0,816	0,490
XGB	0,103	0,076	0,199	0,208	RF-A	0,322	0,224	0,582	0,383	FC(V2).TA	0,463	0,307	0,813	0,483
RF	0,104	0,075	0,197	0,204	FC(V1).INVW	0,327	0,236	0,616	0,412	ARIMA	0,465	0,336	0,887	0,502
FC(V3).SA	0,105	0,075	0,198	0,211	RF-M	0,343	0,258	0,686	0,475	RF-M	0,472	0,356	0,941	0,573
FC(V3).CLS	0,106	0,074	0,197	0,208	RF-C	0,346	0,248	0,649	0,427	RF-C	0,481	0,345	0,896	0,521
AR(1)	0,113	0,073	0,191	0,191	BOOSTING	0,351	0,257	0,669	0,452	BOOSTING	0,487	0,363	0,939	0,558
LL/LST/SET-AR	0,119	0,071	0,186	0,191	Auto.ARIMA	0,354	0,210	0,548	0,365	RF-A	0,490	0,352	0,905	0,511
FC(V2).TA	0,119	0,082	0,216	0,218	ARIMA	0,355	0,250	0,665	0,421	NNET-AR-A	0,507	0,375	1,003	0,588
FC(V2).WA	0,120	0,082	0,217	0,219	NNET-AR-A	0,381	0,272	0,722	0,481	NNET-AR	0,507	0,375	1,001	0,587
FC(V2).CLS	0,121	0,084	0,222	0,213	NNET-AR	0,385	0,273	0,724	0,479	GRNN	0,514	0,376	0,983	0,560
FC(V1).INVW	0,123	0,086	0,224	0,213	ARFIMA	0,387	0,265	0,680	0,425	ENN	0,515	0,381	0,985	0,589
RF-HP	0,124	0,089	0,233	0,229	ELM-A	0,389	0,281	0,721	0,472	GRNN-TSF	0,524	0,383	1,037	0,548
RF-C	0,124	0,087	0,230	0,228	ELM	0,390	0,284	0,729	0,480	FC(V1).SA	0,540	0,412	1,067	0,616
GMDH	0,127	0,067	0,175	0,185	GRNN	0,397	0,282	0,738	0,480	ENN-A	0,545	0,404	1,044	0,602
FC(V2).INVW	0,130	0,090	0,239	0,225	ENN	0,411	0,293	0,751	0,509	ARFIMA	0,545	0,375	0,955	0,524
ARIMA	0,137	0,094	0,250	0,240	GRNN-TSF	0,425	0,303	0,820	0,489	ELM-A	0,568	0,405	1,030	0,588
BOOSTING	0,141	0,089	0,232	0,224	ENN-A	0,430	0,304	0,783	0,512	ELM	0,568	0,410	1,043	0,600
GRNN	0,163	0,113	0,297	0,267	FC(V2).SA	0,434	0,263	0,693	0,451	Auto.ARIMA	0,583	0,317	0,821	0,473
FC(V2).SA	0,166	0,124	0,327	0,287	FC(V1).SA	0,435	0,324	0,842	0,535	JNN-A	0,606	0,406	1,059	0,568
RF-M	0,172	0,135	0,364	0,351	JNN-A	0,453	0,314	0,822	0,506	JNN	0,679	0,431	1,133	0,567
ENN-A	0,190	0,102	0,264	0,254	FC(V2).CLS	0,457	0,229	0,603	0,395	MLP-A	1,041	0,838	2,191	0,871
FC(V2).CLS	0,190	0,117	0,309	0,276	JNN	0,507	0,334	0,875	0,512	RNN	1,059	0,764	2,000	0,794
ENN	0,200	0,119	0,305	0,293	FC(V2).INVW	0,638	0,241	0,634	0,409	NNET	1,081	0,663	1,723	0,759
GRNN-TSF	0,215	0,142	0,379	0,322	KNN	0,709	0,492	1,270	0,704	NNET-A	1,094	0,671	1,744	0,762
JNN	0,231	0,118	0,308	0,256	MLP-A	0,987	0,791	2,065	0,846	MLP	1,132	0,896	2,322	0,891
FC(V1).SA	0,247	0,169	0,442	0,364	NNET	0,988	0,557	1,441	0,664	GRU	1,241	0,868	2,215	0,874
VAR(1)	0,259	0,165	0,435	0,302	NNET-A	0,991	0,561	1,451	0,667	KNN	1,357	0,911	2,250	0,920
KNN	0,310	0,212	0,554	0,411	RNN	0,993	0,692	1,813	0,751	LSTM	1,375	0,914	2,337	0,838
JNN-A	0,316	0,148	0,384	0,303	MLP	1,081	0,851	2,203	0,873	RNN-A	1,375	1,009	2,714	0,913
NNETTS	0,532	0,276	0,710	0,440	NNETTS	1,176	0,550	1,389	0,699	NNETTS	2,631	0,855	2,073	0,835
NNET	0,718	0,309	0,806	0,436	GRU	1,197	0,819	2,090	0,850	FC(V2).SA	41	2,068	3,224	0,562
NNET-A	0,722	0,313	0,815	0,439	RNN-A	1,321	0,937	2,522	0,869	FC(V2).CLS	53	2,550	3,574	0,506
RNN	0,738	0,497	1,299	0,645	LSTM	1,337	0,860	2,198	0,803	FC(V2).INVW	82	3,269	4,103	0,520
RNN-A	0,873	0,577	1,535	0,692	VAR(1)	1,387	0,374	0,956	0,510	GMDH	90	2,750	3,416	0,626
MLP-A	0,928	0,742	1,958	0,815	FC(V2).CLS	1,679	0,315	0,825	0,478	VAR(1)	215	5,302	5,241	0,655
MLP	1,025	0,804	2,108	0,844	GMDH	2,254	0,373	0,942	0,456	FC(V2).CLS	219	5,511	5,582	0,584
GRU	1,150	0,775	1,977	0,826	TAR	3,023	0,408	1,027	0,452	TAR	416	16	5,647	0,598
LSTM	1,305	0,812	2,076	0,779	LL/LST/SET-AR	236	6,626	4,503	0,446	LL/LST/SET-AR	3E+06	8E+04	5,522	0,576
FC(V1).EIG2	40,19	2,128	3,036	0,212	FC(V1).EIG2	5E+07	3E+06	5,801	0,371	FC(V1).EIG2	1E+08	5E+06	5,806	0,470

Table A.12: United States' The Errors of Yield Curve Forecasts for All Yields at Some Forecast Horizons

US - 1 day	RMSE	MAE	MASE	SMAPE	US - 20 day	RMSE	MAE	MASE	SMAPE	US - 40 day	RMSE	MAE	MASE	SMAPE
RW	0,054	0,039	0,096	0,078	RW	0,204	0,134	0,330	0,156	RW	0,321	0,207	0,510	0,212
FC(V1).CLS	0,054	0,039	0,096	0,078	FC(V1).EIG3	0,204	0,134	0,330	0,156	FC(V1).EIG3	0,321	0,207	0,510	0,212
FC(V1).EIG3	0,054	0,039	0,096	0,078	FC(V1).CLS	0,204	0,134	0,330	0,156	FC(V1).CLS	0,321	0,207	0,510	0,212
ETS/TBATS	0,057	0,041	0,101	0,085	ETS/TBATS	0,213	0,139	0,344	0,173	ETS/TBATS	0,333	0,218	0,539	0,239
AR(1)	0,058	0,042	0,103	0,083	BAGGING	0,213	0,143	0,356	0,196	BAGGING	0,333	0,226	0,564	0,265
VAR(1)	0,058	0,042	0,104	0,084	Auto.ARIMA	0,221	0,146	0,363	0,202	FC(V4)	0,348	0,231	0,571	0,279
BAGGING	0,058	0,042	0,104	0,087	FC(V4)	0,227	0,149	0,367	0,197	Auto.ARIMA	0,353	0,233	0,583	0,279
Auto.ARIMA	0,058	0,042	0,102	0,088	XGB	0,257	0,172	0,426	0,239	XGB	0,364	0,239	0,589	0,286
FC(V4)	0,061	0,042	0,104	0,086	FC(V1).INVW	0,260	0,176	0,430	0,227	RF	0,370	0,242	0,597	0,290
ARFIMA	0,062	0,045	0,110	0,092	RF	0,262	0,176	0,433	0,242	FC(V1).INVW	0,382	0,256	0,624	0,287
FC(V1).MED	0,062	0,043	0,107	0,084	VAR(1)	0,269	0,186	0,460	0,289	FC(V3).MED	0,389	0,250	0,621	0,271
TAR	0,063	0,045	0,111	0,099	FC(V3).BG	0,278	0,178	0,443	0,217	FC(V3).WA	0,399	0,257	0,638	0,288
FC(V3).MED	0,066	0,045	0,114	0,089	FC(V3).MED	0,285	0,178	0,443	0,214	FC(V3).TA	0,399	0,257	0,638	0,289
FC(V1).TA	0,066	0,047	0,115	0,092	FC(V3).INVW	0,288	0,184	0,458	0,229	FC(V3).BG	0,400	0,258	0,642	0,277
FC(V2).MED	0,066	0,045	0,113	0,086	FC(V3).TA	0,290	0,183	0,454	0,230	FC(V3).INVW	0,413	0,266	0,660	0,293
FC(V3).WA	0,068	0,047	0,117	0,091	FC(V3).WA	0,290	0,183	0,454	0,230	VAR(1)	0,416	0,287	0,705	0,375
FC(V1).WA	0,068	0,048	0,117	0,092	FC(V1).MED	0,290	0,189	0,460	0,235	FC(V3).SA	0,426	0,274	0,675	0,313
FC(V3).TA	0,068	0,047	0,118	0,091	FC(V1).TA	0,296	0,197	0,477	0,245	FC(V1).TA	0,428	0,287	0,694	0,313
FC(V3).BG	0,068	0,048	0,120	0,096	FC(V1).WA	0,299	0,199	0,482	0,248	FC(V1).MED	0,429	0,282	0,681	0,308
FC(V3).INVW	0,069	0,048	0,121	0,097	FC(V2).MED	0,300	0,193	0,479	0,219	FC(V1).WA	0,431	0,289	0,699	0,316
FC(V3).SA	0,071	0,050	0,125	0,097	FC(V3).SA	0,302	0,194	0,479	0,250	FC(V2).MED	0,432	0,282	0,699	0,285
GMDH	0,075	0,048	0,119	0,105	ARFIMA	0,311	0,215	0,511	0,253	FC(V1).SA	0,453	0,315	0,757	0,349
RF	0,078	0,056	0,137	0,119	RF-HP	0,320	0,201	0,484	0,257	RF-C	0,460	0,304	0,719	0,354
FC(V3).CLS	0,081	0,053	0,133	0,107	RF-C	0,321	0,217	0,515	0,281	ENN	0,461	0,312	0,772	0,369
FC(V1).INVW	0,082	0,060	0,147	0,130	FC(V1).SA	0,336	0,238	0,574	0,305	FC(V2).CLS	0,472	0,307	0,764	0,352
RF-A	0,087	0,061	0,150	0,136	RF-M	0,337	0,225	0,535	0,250	FC(V3).CLS	0,479	0,300	0,747	0,332
FC(V2).WA	0,087	0,061	0,154	0,121	FC(V3).CLS	0,343	0,212	0,528	0,265	RF-M	0,482	0,318	0,756	0,318
FC(V2).TA	0,089	0,063	0,157	0,123	FC(V2).CLS	0,349	0,219	0,544	0,286	ARFIMA	0,484	0,331	0,776	0,354
NNET-AR-A	0,091	0,058	0,142	0,112	AR(1)	0,361	0,235	0,568	0,276	NNET-AR-A	0,496	0,351	0,852	0,454
RF-HP	0,091	0,064	0,157	0,144	NNET-AR-A	0,368	0,251	0,609	0,353	NNET-AR	0,496	0,350	0,851	0,456
XGB	0,091	0,063	0,156	0,140	NNET-AR	0,373	0,252	0,612	0,355	JNN-A	0,502	0,318	0,780	0,362
NNET-AR	0,092	0,058	0,143	0,113	ENN	0,376	0,251	0,621	0,327	RF-HP	0,515	0,330	0,780	0,349
ELM-A	0,097	0,066	0,162	0,124	JNN-A	0,380	0,237	0,590	0,296	ENN-A	0,517	0,337	0,827	0,420
ELM	0,097	0,066	0,162	0,126	ENN-A	0,387	0,255	0,625	0,355	ARIMA	0,521	0,347	0,860	0,351
FC(V2).CLS	0,098	0,065	0,163	0,112	BOOSTING	0,392	0,272	0,640	0,401	JNN	0,530	0,336	0,824	0,414
JNN	0,109	0,070	0,171	0,144	ARIMA	0,394	0,252	0,622	0,273	BOOSTING	0,530	0,379	0,881	0,492
FC(V2).INVW	0,109	0,073	0,183	0,123	JNN	0,410	0,245	0,603	0,324	AR(1)	0,562	0,371	0,887	0,386
RF-C	0,121	0,084	0,203	0,162	RF-A	0,424	0,279	0,672	0,322	GRNN-TSF	0,585	0,401	0,943	0,467
ARIMA	0,128	0,081	0,200	0,122	GRNN-TSF	0,476	0,321	0,755	0,420	GRNN	0,586	0,374	0,877	0,397
RF-M	0,136	0,099	0,239	0,170	GRNN	0,480	0,299	0,701	0,355	ELM	0,644	0,439	1,027	0,458
ENN-A	0,140	0,070	0,173	0,144	ELM	0,485	0,327	0,774	0,380	RF-A	0,644	0,433	1,021	0,430
FC(V2).SA	0,145	0,109	0,271	0,179	ELM-A	0,503	0,343	0,819	0,380	NNET-A	0,664	0,442	1,042	0,516
FC(V1).SA	0,157	0,116	0,283	0,225	KNN	0,507	0,368	0,903	0,479	ELM-A	0,673	0,465	1,100	0,463
LL/LST/SET-AR	0,166	0,064	0,156	0,123	NNET	0,577	0,357	0,842	0,444	NNET	0,674	0,444	1,048	0,509
ENN	0,167	0,088	0,216	0,162	NNET-A	0,578	0,360	0,847	0,448	MLP-A	0,810	0,610	1,512	0,519
FC(V2).CLS	0,185	0,089	0,224	0,148	NNETTS	0,640	0,386	0,913	0,470	MLP	0,837	0,635	1,578	0,537
BOOSTING	0,197	0,116	0,282	0,203	MLP-A	0,740	0,557	1,383	0,495	LSTM	0,890	0,630	1,407	0,606
GRNN	0,253	0,153	0,365	0,200	MLP	0,765	0,580	1,444	0,513	KNN	0,926	0,659	1,590	0,662
JNN-A	0,254	0,117	0,287	0,190	LSTM	0,810	0,571	1,274	0,577	RNN	0,997	0,694	1,800	0,556
KNN	0,265	0,185	0,450	0,297	RNN	0,939	0,624	1,614	0,525	NNETTS	1,152	0,573	1,303	0,541
NNETTS	0,315	0,203	0,491	0,329	GRU	1,099	0,782	1,663	0,756	GRU	1,184	0,847	1,799	0,783
GRNN-TSF	0,320	0,181	0,428	0,281	RNN-A	1,156	0,794	2,170	0,561	RNN-A	1,200	0,851	2,361	0,577
NNET	0,376	0,190	0,454	0,259	FC(V2).WA	1,817	0,241	0,579	0,241	FC(V1).EIG1	112	3,448	2,256	0,250
NNET-A	0,376	0,190	0,455	0,260	FC(V2).TA	1,817	0,241	0,579	0,242	GMDH	205	5,574	2,196	0,438
RNN	0,643	0,446	1,122	0,501	FC(V2).CLS	1,877	0,230	0,551	0,220	TAR	3262	68	2,207	0,384
MLP-A	0,674	0,516	1,290	0,486	FC(V2).SA	1,921	0,287	0,683	0,284	FC(V2).TA	6798	132	2,693	0,306
MLP	0,694	0,533	1,336	0,496	GMDH	3,247	0,331	0,769	0,336	FC(V2).WA	6798	132	2,693	0,307
RNN-A	0,715	0,511	0,000	0,497	FC(V2).INVW	3,570	0,276	0,643	0,235	FC(V2).CLS	7069	141	2,816	0,282
LSTM	0,725	0,514	1,150	0,551	TAR	4,515	0,303	0,680	0,284	FC(V2).SA	7092	147	2,860	0,340
GRU	1,018	0,725	1,539	0,733	FC(V1).EIG1	79	2,408	2,021	0,192	FC(V2).INVW	1E+04	268	2,806	0,300
FC(V1).EIG2	2,396	0,184	0,411	0,106	FC(V1).EIG2	3E+06	8E+04	3,015	0,184	FC(V1).EIG2	7E+06	2E+05	3,017	0,240
FC(V1).EIG1	18,36	0,505	0,910	0,107	LL/LST/SET-AR	2E+21	4E+19	4,500	0,310	LL/LST/SET-AR	8E+64	2E+63	4,500	0,418

APPENDIX 13: ORIGINALITY REPORT

	HACETTEPE ÜNİVERSİTESİ SOSYAL BİLİMLER ENSTİTÜSÜ	Doküman Kodu *Form No.*	FRM-DR-21
		Yayım Tarihi *Date of Pub.*	04.01.2023
	FRM-DR-21 **Doktora Tezi Orijinallik Raporu** *PhD Thesis Dissertation Originality Report*	Revizyon No *Rev. No.*	02
		Revizyon Tarihi *Rev.Date*	25.01.2024

HACETTEPE ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
İKTİSAT ANABİLİM DALI BAŞKANLIĞINA

Tarih: 06/12/2024

Tez Başlığı: Getiri Eğrisi Tahmini Üzerine: Geleneksel ve Geleneksel Olmayan Tekniklerin Uygulamaları.
Tez Başlığı (Almanca/Fransızca)*::...

Yukarıda başlığı verilen tezimin a) Kapak sayfası, b) Giriş, c) Ana bölümler ve d) Sonuç kısımlarından oluşan toplam 173 sayfalık kısmına ilişkin, 06/12/2024 tarihinde şahsım/tez danışmanım tarafından Turnitin adlı intihal tespit programından aşağıda işaretlenmiş filtrelemeler uygulanarak alınmış olan orijinallik raporuna göre, tezimin benzerlik oranı % 6'dır.

Uygulanan filtrelemeler**:

1. ☒ Kabul/Onay ve Bildirim sayfaları hariç
2. ☒ Kaynakça hariç
3. ☐ Alıntılar hariç
4. ☒ Alıntılar dâhil
5. ☒ 5 kelimeden daha az örtüşme içeren metin kısımları hariç

Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Uygulama Esasları'nı inceledim ve bu Uygulama Esasları'nda belirtilen azami benzerlik oranlarına göre tezimin herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumlarda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.

Hakan Gençsoy/İmza

Öğrenci Bilgileri					
	Ad-Soyad	Hakan GENÇSOY			
	Öğrenci No	N11249442			
	Enstitü Anabilim Dalı	İKTİSAT			
	Programı	İKTİSAT			
	Statüsü	**Doktora**	☐	**Lisans Derecesi ile (Bütünleşik) Dr**	☒

DANIŞMAN ONAYI

UYGUNDUR.
(Prof. Dr. Başak Dalgıç, İmza)

*Tez **Almanca** veya **Fransızca** yazılıyor ise bu kısımda tez başlığı **Tez Yazım Dilinde** yazılmalıdır.
**Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü Tez Çalışması Orjinallik Raporu Alınması ve Kullanılması Uygulama Esasları İkinci bölüm madde (4)/3'te de belirtildiği üzere: Kaynakça hariç, Alıntılar hariç/dahil, 5 kelimeden daha az örtüşme içeren metin kısımları hariç (Limit match size to 5 words) filtreleme yapılmalıdır.

	HACETTEPE ÜNİVERSİTESİ SOSYAL BİLİMLER ENSTİTÜSÜ	Doküman Kodu *Form No.*	FRM-DR-21
		Yayım Tarihi *Date of Pub.*	04.01.2023
	FRM-DR-21 **Doktora Tezi Orijinallik Raporu** *PhD Thesis Dissertation Originality Report*	Revizyon No *Rev. No.*	02
		Revizyon Tarihi *Rev.Date*	25.01.2024

TO HACETTEPE UNIVERSITY
GRADUATE SCHOOL OF SOCIAL SCIENCES
DEPARTMENT OF ECONOMICS

Date: 06/12/2024

Thesis Title (In English): On The Yield Curve Forecasting: Applications of Conventional and Non-Conventional Techniques.

According to the originality report obtained by myself/my thesis advisor by using the Turnitin plagiarism detection software and by applying the filtering options checked below on 06/12/2024 for the total of 173 pages including the a) Title Page, b) Introduction, c) Main Chapters, and d) Conclusion sections of my thesis entitled above, the similarity index of my thesis is 6%.

Filtering options applied**:

1. ☒ Approval and Decleration sections excluded
2. ☒ References cited excluded
3. ☐ Quotes excluded
4. ☒ Quotes included
5. ☒ Match size up to 5 words excluded

I hereby declare that I have carefully read Hacettepe University Graduate School of Social Sciences Guidelines for Obtaining and Using Thesis Originality Reports that according to the maximum similarity index values specified in the Guidelines, my thesis does not include any form of plagiarism; that in any future detection of possible infringement of the regulations I accept all legal responsibility; and that all the information I have provided is correct to the best of my knowledge.

I respectfully submit this for approval.

Hakan Gençsoy /

Student Information				
	Name-Surname	Hakan GENÇSOY		
	Student Number	N11249442		
	Department	ECONOMICS		
	Programme	ECONOMICS		
	Status	**PhD** ☐	**Combined MA/MSc-PhD** ☒	

SUPERVISOR'S APPROVAL

APPROVED
(Prof. Dr. Başak Dalgıç, Signature)

**As mentioned in the second part [article (4)/3]of the Thesis Dissertation Originality Report's Codes of Practice of Hacettepe University Graduate School of Social Sciences, filtering should be done as following: excluding refence, quotation excluded/included, Match size up to 5 words excluded.

APPENDIX 14: ETHICS COMMISSION FORM

	HACETTEPE ÜNİVERSİTESİ SOSYAL BİLİMLER ENSTİTÜSÜ	Doküman Kodu *Form No.*	FRM-DR-12
	FRM-DR-12 **Doktora Tezi Etik Kurul Muafiyeti Formu** *Ethics Board Form for PhD Thesis*	Yayım Tarihi *Date of Pub.*	22.11.2023
		Revizyon No *Rev. No.*	02
		Revizyon Tarihi *Rev.Date*	25.01.2024

HACETTEPE ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
İKTİSAT ANABİLİM DALI BAŞKANLIĞINA

Tarih: 06/12/2024

Tez Başlığı (Türkçe): Getiri Eğrisi Tahmini Üzerine: Geleneksel ve Geleneksel Olmayan Tekniklerin Uygulamaları.
Tez Başlığı (Almanca/Fransızca)*::..
....

Yukarıda başlığı verilen tez çalışmam:

1. İnsan ve hayvan üzerinde deney niteliği taşımamaktadır.
2. Biyolojik materyal (kan, idrar vb. biyolojik sıvılar ve numuneler) kullanılmasını gerektirmemektedir.
3. Beden bütünlüğüne veya ruh sağlığına müdahale içermemektedir.
4. Anket, ölçek (test), mülakat, odak grup çalışması, gözlem, deney, görüşme gibi teknikler kullanılarak katılımcılardan veri toplanmasını gerektiren nitel ya da nicel yaklaşımlarla yürütülen araştırma niteliğinde değildir.
5. Diğer kişi ve kurumlardan temin edilen veri kullanımını (kitap, belge vs.) gerektirmektedir. Ancak bu kullanım, diğer kişi ve kurumların izin verdiği ölçüde Kişisel Bilgilerin Korunması Kanuna riayet edilerek gerçekleştirilecektir.

Hacettepe Üniversitesi Etik Kurullarının Yönergelerini inceledim ve bunlara göre çalışmamın yürütülebilmesi için herhangi bir Etik Kuruldan izin alınmasına gerek olmadığını; aksi durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.

Hakan GENÇSOY/

Öğrenci Bilgileri				
	Ad-Soyad	Hakan GENÇSOY		
	Öğrenci No	N11249442		
	Enstitü Anabilim Dalı	İKTİSAT		
	Programı	İKTİSAT		
	Statüsü	**Doktora** ☐	**Lisans Derecesi ile (Bütünleşik) Dr** ☒	

DANIŞMAN ONAYI

UYGUNDUR.
(Prof. Dr. Başak Dalgıç, İmza)

* Tez **Almanca** veya **Fransızca** yazılıyor ise bu kısımda tez başlığı **Tez Yazım Dilinde** yazılmalıdır.

	HACETTEPE ÜNİVERSİTESİ SOSYAL BİLİMLER ENSTİTÜSÜ	Doküman Kodu *Form No.*	FRM-DR-12
		Yayım Tarihi *Date of Pub.*	22.11.2023
	FRM-DR-12 **Doktora Tezi Etik Kurul Muafiyeti Formu** *Ethics Board Form for PhD Thesis*	Revizyon No *Rev. No.*	02
		Revizyon Tarihi *Rev.Date*	25.01.2024

HACETTEPE UNIVERSITY
GRADUATE SCHOOL OF SOCIAL SCIENCES
DEPARTMENT OF ECONOMICS

Date: 06/12./2024

ThesisTitle (In English): On The Yield Curve Forecasting: Applications of Conventional and Non-Conventional Techniques.

My thesis work with the title given above:

1. Does not perform experimentation on people or animals.
2. Does not necessitate the use of biological material (blood, urine, biological fluids and samples, etc.).
3. Does not involve any interference of the body's integrity.
4. Is not a research conducted with qualitative or quantitative approaches that require data collection from the participants by using techniques such as survey, scale (test), interview, focus group work, observation, experiment, interview.
5. Requires the use of data (books, documents, etc.) obtained from other people and institutions. However, this use will be carried out in accordance with the Personal Information Protection Law to the extent permitted by other persons and institutions.

I hereby declare that I reviewed the Directives of Ethics Boards of Hacettepe University and in regard to these directives it is not necessary to obtain permission from any Ethics Board in order to carry out my thesis study; I accept all legal responsibilities that may arise in any infrigement of the directives and that the information I have given above is correct.

I respectfully submit this for approval.

Hakan GENÇSOY /

Student Information				
	Name-Surname	Hakan GENÇSOY		
	Student Number	N11249442		
	Department	ECONOMICS		
	Programme	ECONOMICS		
	Status	**PhD** ☐	**Combined MA/MSc-PhD** ☒	

SUPERVISOR'S APPROVAL

APPROVED
(Prof. Dr. Başak Dalgıç, Signature)