ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

A NOVEL APPROACH FOR TIME SERIES FORECAST COMBINATIONS BASED ON MULTI-CRITERIA DECISION MAKING (MCDM) METHODS

M.Sc. THESIS

Tuğba Yasemin KARAGÖZ

Department of Industrial Engineering

Industrial Engineering Programme

JANUARY 2025

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

A NOVEL APPROACH FOR TIME SERIES FORECAST COMBINATIONS BASED ON MULTI-CRITERIA DECISION MAKING (MCDM) METHODS

M.Sc. THESIS

Tuğba Yasemin KARAGÖZ
(507221118)

Department of Industrial Engineering

Industrial Engineering Programme

Thesis Advisor: Asst. Prof. Erkan IŞIKLI

JANUARY 2025

ISTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

ZAMAN SERİSİ TAHMİN KOMBİNASYONLARI İÇİN ÇOK KRİTERLİ KARAR VERME YÖNTEMLERİNE DAYALI YENİ BİR YAKLAŞIM

YÜKSEK LİSANS TEZİ

Tuğba Yasemin KARAGÖZ
(507221118)

Endüstri Mühendisliği Anabilim Dalı

Endüstri Mühendisliği Bölümü

Tez Danışmanı: Dr. Öğr. Üyesi Erkan IŞIKLI

JANUARY 2025

Tuğba Yasemin KARAGÖZ, an M.Sc. student of ITU Graduate School with student ID 507221118, successfully defended the thesis entitled "A NOVEL APPROACH FOR TIME SERIES FORECAST COMBINATIONS BASED ON MULTI-CRITERIA DECISION MAKING (MCDM) METHODS", which she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor : **Asst. Prof. Erkan IŞIKLI** .....
Istanbul Technical University

Jury Members : **Assoc. Prof. Çiğdem KADAİFÇİ YANMAZ** .....
Istanbul Technical University

Assoc. Prof. Burak ALAKENT .....
Bogazici University

Date of Submission : 31 December 2024
Date of Defense : 20 January 2025

To the spirit of unconditional devotion that transcends reason,

FOREWORD

This thesis is the result of my in-depth research on time series forecasting combinations, an increasingly important topic in data science and forecasting. The completion of the thesis was made possible by the guidance and support of several people to whom I am grateful.

First and foremost, I would like to express my sincere and deepest gratitude to my advisor, Assist. Prof. Erkan Işıklı, for his invaluable guidance, insightful feedback, and constant encouragement throughout this process. I extend my thanks to the members of my thesis defense jury, Assoc. Prof. Çiğdem Kadaifçi Yanmaz and Assoc. Prof. Burak Alakent, for their thoughtful suggestions and constructive evaluations. I am also truly grateful to Prof. Dr. Burç Ülengin, whose excellent and instructive lecture provided me with critical insights and a strong theoretical foundation that shaped the direction of this study. Additionally, I am very thankful to Prof. Dr. Hür Bersam Sidal and Assoc. Prof. Nihan Yıldırım, whose lectures I have been gladly assisting, for their kind attitude, love, and support, which made it easier for me to adapt to the academic environment and gave me encouragement.

Finally, I would like to express my heartfelt thanks to my family. I am especially grateful to my dear mother, who kept me going throughout the writing of my thesis with her constant supply of coffee; to my little sisters, whose delicious desserts sweetened my days and sustained my energy; and to my brothers, whose support and presence provided moral strength.

It is my hope that this thesis contributes meaningfully to the study of time series forecasting and serves as a foundation for further research in this field.

January 2025

Tuğba Yasemin Karagöz
(Research Assistant)

TABLE OF CONTENTS

Page

ABBREVIATIONS

ACF : Autocorrelation Function
ADF : Augumented Dickey Fuller
AIC : Akaike Information Criteria
ANN : Artificial Neural Network
ARIMA : AutoRegressive Integrated Moving Average
CILOS : Criteria Impact LOSs
CLS : Constrained Least Squares
CRITIC : Criteria Importance Through Intercriteria Correlation
CSR : Complete Subset Regression
ÇKKV : Çok Kriterli Karar Verme
EIA : Energy Information Administration
ETS : Error, Trend, Seasonality
KPSS : Kwiatkowski-Phillips-Schmidt-Shin
LAD : Least Absolute Deviation
LM : Lagrange Multiplier
MCDM : Multi-criteria decision making
MAE : Mean Absolute Error
MAPE : Mean Absolute Percentage Error
MASE : Mean Absolute Scaled Error
ME : Mean Error
MED : Median Combination
ML : Machine Learning
MMcf : One million cubic feet
MPE : Mean Percentage Error
NNAR : Neural Network AutoRegressive
OLS : Ordinary Least Squares
PACF : Partial Autocorrelation Function
RF : Random Forest
RMSE : Root Mean Squared Error
SA : Simple Average

SMA **:** Simple Moving Average

sMAPE **:** Symmetric Mean Absolute Percent Age Error

STL **:** Seasonal and Trend decomposition using Loess

SVM **:** Support Vector Machine

TA **:** Trimmed Average

TBATS **:** Trigonometric, Box-Cox, ARMA errors, Trend, and Seasonal components

THETAM **:** Support Vector Machine

TOPSIS **:** The Technique for Order Preference by Similarity to an Ideal Solution

THETAM **:** Theta Model

WA **:** Winsorized Average

XGBoost **:** Extreme Gradient Boosting

SYMBOLS

$\boldsymbol{R}$	**:** The number of single prediction models combined
$\boldsymbol{f_{ji}}$	**:** The value of the forecast produced by the single prediction model *j* at time point *i*
$\boldsymbol{\lambda}$	**:** The correction factor
$\boldsymbol{n}$	**:** The number of observations
$\boldsymbol{a_i}$	**:** The actual value at time point *i*
$\boldsymbol{\hat{\omega}_j}$	**:** The weight assigned to the single prediction model of *j*
$\boldsymbol{\varepsilon_i}$	**:** The error value at time point *i*
$\boldsymbol{T}$	**:** The length of the time series
$\boldsymbol{k}$	**:** The number of lagged terms
$\boldsymbol{\Delta y_t}$	**:** The differenced time series
μ	**:** A constant term
$\boldsymbol{b}$	**:** The trend term
$\boldsymbol{\gamma}$	**:** The coefficient of the lagged level of the series
$\boldsymbol{S_t}$	**:** Cumulative sum of the residuals
$\boldsymbol{s^2}$	**:** Long-run variance of residuals
$\boldsymbol{e_i}$	**:** The remainder component at time *i*
$\boldsymbol{p}$	**:** The order of the autoregressive part
$\boldsymbol{q}$	**:** The order of moving average part

LIST OF TABLES

Page

LIST OF FIGURES

Page

A NOVEL APPROACH FOR TIME SERIES FORECAST COMBINATIONS BASED ON MULTI-CRITERIA DECISION-MAKING (MCDM) METHODS

SUMMARY

Forecast combinations in time series are quantitative forecasting methods that aim to obtain models with higher performance results in terms of accuracy by integrating the strengths of different time series forecasting models.

This study aims to develop a new multi-dimensional and objective approach using multi-criteria decision-making (MCDM) methods that outperform both single forecasting models and some of the most widely accepted forecasting combination models in the literature. In this context, a new combination model is developed in which the weights of the forecast models are assigned by using some hybrid MCDM approaches.

First, the model proposition of CRITIC-TOPSIS-based predictor assignment was applied to two different datasets. The reason to test the same model on two different datasets is to understand the efficacy of the proposed method for time series having different underlying patterns, characteristics, sizes, and frequencies. The initial dataset consists of daily data from the industry domain, sourced from the M4 competition file which is named D2035, and is publicly available as open data on GitHub. It includes irregular patterns, exhibits high-frequency noise, and volatility as well as is large in size. Then, the same proposed model was applied to a monthly dataset of deliveries of natural gas to electric power users (MMcf) for the United States of America (EIA's dataset), which includes more regular and various patterns, exhibits lower-frequency noise and volatility, as well as it is smaller in size.

After preliminary preparations and statistical tests, the datasets changed to conform to the specifications of the statistical application (if necessary) and were split into 3 subsets named train, validation, and test. All individual forecast models were trained on train sets and forecasted on validation and test set horizons. The selection of predictors (so, the components of combination forecast) is handled by visual inspection of the graphs over validation and test horizon. The decision matrix of predictors versus performance metrics was created as a next step. The performance metric results from the validation sets are used for each predictor to build this decision matrix. Using a hybrid MCDM approach, where the performance metric weights were assigned through the CRITIC method and the TOPSIS model first, was applied based on these weights, the closeness of each predictor to the ideal solution was determined. The normalized ratio corresponding to each predictor's closeness to the ideal solution will represent the weight within the new proposed combination model.

The relative test performance results derived from both datasets are compared with traditional single forecasting models, machine learning models, and several widely

used time series forecast combination models from the literature, as well as with each other.

The results indicate that the proposed CRITIC-TOPSIS hybrid MCDM-based forecast combination model outperformed all individual and existing combination models on natural gas delivery to USA electric power consumers (MMcf). On the other hand, the D2025 dataset did not yield performance results as outstanding as the MMcf dataset. However, it still provided promising outcomes with notably positive results based on certain performance metrics.

In this context, it can be stated that the proposed and implemented model performs better on datasets with more defined patterns, lower noise, and reduced volatility, rather than those with high randomness, high-frequency noise, irregularity, and ambiguous patterns. Additionally, the proposed model yielded highly effective results compared to its counterparts.

After superior outcomes of the proposed CRITIC-TOPSIS-based forecast combination model, especially on the natural gas delivery to USA electric power consumers dataset (EIA), it is decided to apply the same combination methodology on the same dataset with the CILOS-TOPSIS hybrid MCDM approach this time. Testing the proposed method with both the CRITIC and CILOS approaches enables a comparative analysis of how differing weight assignments for accuracy measurements based on their relative influence will affect the success of the combination forecast. The findings reveal that the proposed CILOS-TOPSIS hybrid MCDM-based forecast combination achieved small enhancements compared to the CRITIC-TOPSIS-based approach.

Additionally, another hybrid MCDM approach model built under the umbrella of the proposed model for forecast combination, the CILOS-ARAS-based model, was tested on EIA's dataset to evaluate potential improvements arising from modifications to the performance evaluation of the decision matrix. This approach facilitated the assessment of the utility function value, as opposed to relying on the closeness to the ideal solution using a TOPSIS-based evaluation when determining weights for the component forecast models.

All MCDM-based models created under the proposed framework are compared with each other and with other existing models in the literature.

ZAMAN SERİSİ TAHMİN KOMBİNASYONLARI İÇİN ÇOK KRİTERLİ KARAR VERME YÖNTEMLERİNE DAYALI YENİ BİR YAKLAŞIM

ÖZET

Nicel tahminleme teknikleri; geçmiş verilere ve bu verilerin altta yatan istatistiksel eğilimlerine dayalı olarak gelecekteki belirli bir periyotun sonuçlarını öngörmek için yıllardır çalışılmakta ve geliştirilmekte olan bir alan olmakla beraber kaynak tahsisi, üretim çizelgeleme, lojistik, risk yönetimi ve kısa/uzun vadeli stratejik planlama gibi karar verme, yönetim ve planlama faaliyetleri için önemli iç görüler sağlama potansiyelinde olmuştur. Belirli bir değişkenin geçmiş periyotta farklı veya düzenli zaman noktalarından toplanmış sıralı verilerinin analizine dayalı zaman serilerinin kullanımı da, nicel bir yaklaşım olarak literatürde yaygın bir şekilde kullanılmakta olup, istatistik tabanlı tahmin yöntemleri arasında öne çıkmaktadır.

ARIMA, mevsimsel ayrıştırma ve üstel düzleştirme gibi geleneksel zaman serisi modelleri, tahminleme literatüründe yıllardır yaygın olarak kullanılmaktadır. Son yıllarda, sinir ağları, rastgele orman ve destek vektör makinesi gibi doğrusal olmayan makine öğrenimi modellerini içeren modern tahmin teknikleri de zaman serisi analizi için kullanılmakta olup büyük ve karmaşık veri kümelerinde oldukça ümit verici test sonuçları vermektedir.

Tahminleme için ayrı ayrı modeller kullanmanın yanı sıra, birden fazla tekil yöntemi bütünleştirmek de mümkündür. Literatürde, özellikle son birkaç on yıldır zaman serisi tahmin kombinasyonları hakkında göz ardı edilemeyecek sayıda araştırma ve uygulama yapılmıştır. Çeşitli alanları kapsayan bu çalışmalar, en yalın tahmin tekniklerinin birleştirilmesiyle dahi öngörü performanslarının arttırılabileceğini ortaya koymuştur. Tekil tahmin modellerinin sonuçlarının basit aritmetik ortalamasını almaktan, optimizasyon/doğrusal temelli ağırlıklandırma yaklaşımlarına kadar birçok birleştirme yöntemi literatürde kendine yer bulmuşken, basit tahminleme tekniklerinin kendi içinde birleştirilmesinden makine öğrenimi algoritmalarının geleneksel yöntemlerle entegrasyonuna kadar birçok teknik de uygulamalarda test edilmiştir. Bu teknikler, farklı veri kümelerine uyarlanabilirlik, temel karmaşıklık ve hesaplama gereksinimleri açısından çeşitlilik gösterir.

Nihayetinde zaman serilerinde tahmin kombinasyonları, farklı tahmin modellerinin güçlü taraflarını bütünleştirerek doğruluk performansı daha yüksek modeller elde etmeyi amaçlayan nicel tahmin yöntemleri olarak literatürde var olagelmiştir. Tahminleme modellerinin performansının ölçümü ve karşılaştırılması adına da çeşitli metotlar bulunmaktadır. Grafik analizi, nicel ve nitel yaklaşımlar modellerin doğruluk ölçülerini ifade etmek için kullanılmaktadır. Bu çalışmada, sayısal ve objektif çıktı

vermeleri ile karar matrisine uyarlanabilir olmaları nedeniyle nicel performans ölçütleri üzerinden değerlendirme yapılacaktır.

Ayrıyeten literatürdeki çalışmaların çoğu, tahmin modellerinin performansını önceden belirlenmiş tek bir doğruluk ölçütüne göre veya farklı birkaç ölçütün "uzlaştığı" noktayı esas alarak bir arada değerlendirmektedir. Kurulan modellerin performanslarının farklı yönlerini yakalayabilen birbirinden farklı doğruluk ölçütlerinden uygun olan(lar)ının belirlenmesi de, bu yüzden önemli bir konu olarak karşımıza çıkmaktadır. Bu bağlamda, ortalama hata (ME), ortalama mutlak hata (MAE), kök ortalama kare hatası (RMSE), ortalama mutlak ölçekli hata (MASE), ortalama hata yüzdesi (MPE), ortalama mutlak yüzde hata (MAPE) ve ACF1 ölçümü olmak üzere farklı sınıflara ait 5 performans metriği, modellerin karşılaştırılması ve önerilen modelin içeriğinde kullanılması adına bu çalışmada yer almıştır.

Bu tez çalışması kapsamında, çok kriterli karar verme (ÇKKV) yöntemleri kullanılarak farklı performans metriklerinin bir arada göz önüne alındığı, hem tekli tahmin modellerinden hem de literatürde en yaygın kabul gören bazı tahmin kombinasyon modellerinden daha iyi performans gösteren çok boyutlu ve objektif yeni bir yaklaşım geliştirilmesi amaçlanmıştır. Bu bağlamda, bileşen tahmin modellerinin ağırlıklarının bazı hibrit ÇKKV yaklaşımları kullanılarak atandığı yeni bir kombinasyon modeli geliştirilmesi amaçlanmıştır.

Bu tez çalışması kapsamında, önerilen tahmin kombinasyonunu oluşturacak tahminleyici bileşenleri, ilk olarak geliştirilen CRITIC-TOPSIS tabanlı yaklaşım ile iki farklı veri seti üzerinde test edilmiştir. Aynı tekniğin iki farklı veri seti üzerinde test edilmesinin nedeni, önerilen yaklaşımın farklı özelliklere, boyutlara ve frekanslara sahip zaman serileri üzerindeki etkisini anlayabilmektir. İlk veri seti, GitHub'da açık veri şeklinde yer alan M4 yarışma dosyasından D2035 olarak adlandırılmış endüstri alanından günlük verilerden oluşmaktadır. İlgili veri seti, düzensiz desenler içeren, yüksek frekanslı gürültü ve oynaklık sergileyen büyük boyutlu bir veri setidir. Bu tez kapsamında önerilen aynı model daha sonra, daha düzenli desenler içeren, daha düşük frekanslı gürültü ve oynaklık sergileyen, daha küçük boyutlu başka bir veri seti olan "Amerika Birleşik Devletleri için elektrik enerjisi tüketicilerine (MMcf) aylık doğal gaz teslimatları" verisine (EIA'nın veri seti) de uygulanmıştır.

Ön hazırlıkların ve istatistiksel testlerin ardından, tahminleme modellerinin gerekliliklerine uygun hale getirilen veri setleri; eğitim, doğrulama ve test olarak üç ayrı alt kümeye ayrılmıştır. Tüm tekli tahmin modelleri eğitim setleri üzerinde eğitilmiş ve doğrulama ile test setleri üzerinden tahminlemesi yapılmıştır. Yeni tahmin kombinasyonu modelinin bileşenlerinin seçimi (yani, tekli tahmin modelleri), doğrulama ve test ufku üzerindeki tahmin sonuçlarının grafiklerinin görsel incelemesi ile seçilmiştir. Sonraki adımda ise belirlenen bu tahmin modelleriyle ve her modele karşılık gelen performans metrikleriyle beraber, tahmin modellerinin satırları, performans metriklerinin ise sütunları oluşturduğu bir karar matrisi meydana getirilmiştir. Bu karar matrisindeki performans metrikleri, doğrulama setleri üzerinden alınmıştır. Performans metriklerini temsil eden kriterlerin ağırlıklarının CRITIC yöntemiyle atandığı ve bu ağırlıklara dayalı olarak TOPSIS modelinin uygulandığı hibrit bir ÇKKV yaklaşımı kullanılarak, her bir tahmin modelinin ideal çözüme yakınlığı belirlenmiştir. Her bir tahmin modelinin ideal çözüme yakınlığına karşılık gelen normalleştirilmiş oranları, önerilen yeni kombinasyon modelindeki ağırlıkları temsil edecektir.

CRITIC-TOPSIS tabanlı önerilen bu yeni modelin her iki veri seti üzerinden elde edilen test performansı sonuçları, birbirleriyle, geleneksel tekli tahmin modelleriyle, makine öğrenimi modelleriyle ve literatürde yaygın olarak kullanılan çeşitli zaman serisi tahmin kombinasyon modelleri ile karşılaştırılmıştır.

Sonuçlar, önerilen CRITIC-TOPSIS hibrit ÇKKV tabanlı tahmin kombinasyon modelinin, ABD elektrik tüketicilerine (MMcf) doğal gaz teslimatı miktarını tahmin etmek adına, karşılaştırılan tüm tekli ve mevcut kombinasyon modellerinden daha iyi performans gösterdiğini göstermektedir. Öte yandan, D2025 veri kümesi üzerinden EIA'nın veri kümesi kadar üstün performans sonuçları vermese de, belirli performans metriklerine dayalı olarak önemli ölçüde olumlu sonuçlarla umut verici sonuçlar sağlanmıştır.

Bu bağlamda, önerilen ve uygulanan modelin, yüksek rastgelelik ve yüksek frekanslı gürültü içeren düzensiz ve belirsiz desenlere sahip veri kümelerinden ziyade, daha belirli örüntülere, daha düşük gürültüye ve daha az rastgeleliğe sahip veri kümelerinde daha iyi performans gösterdiği söylenebilir. Ayrıca önerilen modelin, kombinasyon modellerinin başarılı sonuç verdiği veri setleri için, kıyaslanan kombinasyon modellerine nazaran daha iyi sonuç verdiği görülmüştür.

Önerilen CRITIC-TOPSIS tabanlı tahmin kombinasyonu modelinin özellikle ABD elektrik tüketicilerine doğal gaz dağıtımı miktarı veri kümesi üzerindeki başarılı sonuçlarından sonra, aynı kombinasyon metodolojisinin CILOS-TOPSIS hibrit ÇKKV yaklaşımıyla da yine aynı veri kümesi üzerinde uygulanmasına karar verilmiştir. Önerilen yöntemde kriterlerin hem CRITIC hem de CILOS yaklaşımlarıyla ağırlıklandırılması ve karşılaştırmasının yapılması, kriterlerin arasındaki göreli ilişkinin farklı değerlendirilmesine dayalı olarak performans metriklerine farklı ağırlık ataması yapılmasının kombinasyon tahmininin başarısını nasıl etkileyeceğine dair bir analizin yapılabilmesini sağlamıştır. Bulgular, önerilen CILOS-TOPSIS hibrit ÇKKV tabanlı tahmin kombinasyonunun CRITIC-TOPSIS tabanlı yaklaşımına yakın sonuçlar vermesine karşın genel performansta iyileşme sergilediğini ortaya koymaktadır.

Ayrıca, bu tez kapsamında önerilen kombinasyon modelinin çatısı altında oluşturulan bir başka hibrit ÇKKV yaklaşımı modeli, CILOS-ARAS tabanlı model de, karar matrisinin performans değerlendirmesinde yapılan değişikliklerden kaynaklanan potansiyel iyileştirmeleri değerlendirmek için yine EIA'nın veri seti üzerinde test edilmiştir. Bu yaklaşımla, bileşen tahmin modellerine ağırlık atama sürecinde, TOPSIS tabanlı bir değerlendirme ile ideal çözüme yakınlık yerine fayda fonksiyonu değerine dayalı bir yöntem kullanmanın kombinasyon başarısı üzerindeki etkisi ortaya konulmuştur.

Önerilen çerçeve altında oluşturulan tüm ÇKKV tabanlı modeller birbirleriyle ve literatürdeki diğer mevcut modellerle karşılaştırılmıştır.

1. INTRODUCTION

1.1 Background

Forecasting techniques have been developed for years to predict future outcomes subjecting to the uncertainty of current assets based on historical data and underlying statistical patterns and (or alternatively) knowledge and experience of any future occurrences that might affect the forecasts. These techniques have the potential to provide important insights for the decision-making activities of planning and management such as resource allocation, scheduling of production, transportation, risk management, and short-term/long-term strategic planning.

The selection of forecasting techniques depends on different factors. A forecaster should identify the forecasting problem first. Understanding the nature of the data will help to determine in which unit and how to collect the input. After that, the forecast horizon and the appropriate methods according to the dataset should be decided based on specific objectives. Organizational support is required for putting formalized forecasting methods into practice at the end.

The predictability of an event or quantity depends on several factors including (Hyndman and Athanasopoulos, 2018, p. 7):

1. The extent to which we perceive the contributing aspects

2. The quantity of data available

3. Whether the predictions have an impact on the object we're attempting to predict

Two major categories of forecasting methodologies are used in the literature. If sufficient quantitative information is available, then quantitative forecasting methods which involve using time series and regression models can be applied. When the situation involves adequate knowledge from qualitative sources but not sufficient quantitative information, qualitative approaches can be employed (Makridakis et al., 1998, p. 8).

Using time-series data as a quantitative approach is pretty common in the literature and the crux of the matter of statistical forecasting. Time-series analysis is an approach to statistics for examining data from several observations of a single unit or person at a regular pace over a substantial amount of observations, according to Velicer and Fava (2003). Regarding the modeling of time series data, two branches can be described. It is referred to as univariate time series forecasting if the data utilized to create the forecast solely includes the variable's historical data. However, it is referred to as multivariate time series forecasting if it incorporates values from the past of one or more variables that are not those for which forecasts are to be formed (Armstrong, 2007, p. 66). In this study, we work quantitatively with univariate time series data to test our novel approach.

Classical or traditional time series models such as ARIMA, seasonal decomposition, or exponential smoothing have been commonly used for years providing a framework to make a forecast by taking past observations as a reference. In recent years, modern forecasting models including machine learning methodologies like neural networks, random forest, or support vector machine have been used for time series analysis and the results are very promising. Recent empirical research demonstrates the tremendous utility of nonlinear machine learning models when paired with large and sophisticated datasets (Masini et al., 2023).

Besides using individual models to make a forecast, it is also possible to integrate multiple single methods. An unignorable amount of research and applications about time series forecast combinations have been seen in the literature, especially for the last couple of decades under the name of hybrid, ensemble, or combined forecasting time series models. Research in various fields has strongly suggested that the performance of prediction can be increased when sometimes even in a simple fashion forecasts are combined (Yang, 2004).

A variety of combining techniques has been generated ranging from taking simple averages of individual methods to hybrid models that combine ML algorithms with traditional methods. These techniques may diversify in terms of adaptability to different kinds of datasets, underlying complexity, and calculation requirements. A comprehensive overview of forecast combinations is provided in the literature section of this thesis.

Finally, both single and combined forecasting models are seeking for better accuracy. Several evaluation methods and performance metrics have been developed to measure the accuracy of forecasting models. A detailed explanation about error measures such us scale-dependents or percentage errors is given also in the literature part of this study.

1.2 Challenges and Limitations

Although time series forecast combinations have promising results, they also pose some limitations. The forecasting models to be combined should be selected appropriately, because classes (traditional or ML), diversity, and quality of single models can impact the effectiveness of the combination. Not only the quality of single models but also the weight assigned to each of them affects the accuracy gained. Then, one should decide on the aggregation rule for the models to reach optimal weights.

Additionally, various metrics for measuring model performance have been proposed. Comparing the performance metrics of different models individually may yield contradictory results. Using multiple error measures might lead to conflicting conclusions, which can be confounding and undesirable to decision-makers (Davydenko and Fildes, 2013). Every measure of performance is a single metric that represents only one facet of forecasting performance and a straight model performance metric by itself does not guarantee strategy profitability (Ülengin, 2024). To manage the uncertainty that arises from considering several contradicting performance metrics, it should be specified what kind of performance evaluation technique will be determined. If multiple performance metrics are considered, it is essential to assess their interdependence and adopt an objective approach to achieve consensus among them.

Another facet deserving attention is the generalization of how well forecasting models' error measurements will be made on unseen data. Franses (2011) stated that forecast models' individual performance histories do not ensure that they will successfully contribute to the combined forecasts; therefore, if the intention is to consideration of combined forecasts anyway, only working on the performance of a training set becomes obsolete. Because performance evaluations made only on training data are unreliable, assessing model accuracy requires separating datasets to validate and test. Splitting the available dataset into multiple subsets, and using validation and test sets separately can ensure an unbiased model evaluation and a reliable performance

estimation on unseen data. Inferring what generalized deduction about the model performances by using these subsets is an additional aspect worth noting.

1.3 Motivation and Scope of the Study

Most of the studies in the literature evaluate the performance of their forecasting models based on a single pre-determined accuracy measure or a couple of metrics together by taking the consensus of the majority as a basis. Obtaining the proper measurement metrics for accuracy is an additional challenge as they capture different facets of model performance (Mehdiyev et al., 2016). That is why, various performance measurements' concurrent consideration can enhance choosing the most effective forecasting model (Xu and Ouenniche, 2012). A limited number of studies addressed a multidimensional framework that takes several accuracy measurements into account simultaneously to evaluate the performance of forecasting models adopting an MCDM approach and select the best forecasting model based on this framework (Badulescu et al., 2021; Kumar and Kaur, 2021). Only one paper among these considered dependencies among accuracy measures within a multidimensional structure (Badulescu et al., 2021). To the best of our knowledge, there is a gap in the related literature for assigning combination weights to forecast models in a multidimensional framework while simultaneously considering different non-independent accuracy measures in an objective manner.

Besides considering several performance measures, several prediction models also need to be chosen and included in the combination. The selection of predictors in the ensemble is usually done by calculating the performance of each model toward the hold-out sample (Widodo and Budi, 2011). Moreover, there exist some studies that asserted that the combination of machine learning and traditional methods can give better accuracy results compared to single methods in their study (Mariñas-Collado, 2022).

So far, no studies have been encountered in the literature that propose a methodology for objectively determining forecast combination weights by competing multiple different-classed forecasting time series models and evaluating their performance by using various metrics simultaneously. The proposed approach also considers the intercorrelations between performance metrics.

Therefore, this thesis proposes a new *multidimensional and objective approach* based on multi-criteria decision-making (MCDM) framework to assign combination weights to time series forecast models. The goal is to obtain multi-dimensionally more accurate results compared to single models or other existing combination models existing in the literature.

1.4 The Proposed Approach

In this study, pre-determined forecasting models are evaluated as alternatives and accuracy metrics as criteria while adapting to a decision matrix. Beforehand we separate the datasets into three subsets as training, validation, and test to obtain performance results. The decision matrix mentioned is created over validation set results. In the subsequent phase, it will be used MCDM approaches for both weight estimation for accuracy metrics and performance estimation for alternatives (predictor forecasting models) considering dependence between attributes. The final rankings of each time series forecast model will be determined as a result of the chosen MCDM performance evaluation method. The normalized values of these final rankings will represent the weights of each individual model in the forecast combination. Considering the obtained weights for each MCDM approach we applied, a novel combination approach for time series forecasting will be designed. Figure 1.1 shows the first look of the proposed methodology.

Figure 1.1: A basic look at the the proposed methodology.

Within the context of this research, the absolute performances of each combination method obtained from hybrid MCDM methods will determined and we will compare their relative performances with each other, as well as with the most widely used combination methods that already exist in the literature. Our goal is to create a multi-dimensionally more accurate time series forecasting model on an objective structure compared to existing models in the literature.

The structure of this thesis is as follows: Chapter 2 provides an overview of the literature on time series forecast combinations. In Chapter 3, the methodology is presented, including the roadmap for implementation and a detailed explanation of all methods used. Chapter 4 discusses the implementation of the novel approach proposed in this study and presents the results. Finally, a conclusion and recommendation section is offered.

2. LITERATURE REVIEW ON TIME SERIES FORECAST COMBINATIONS

2.1 Introduction

The goal of time series forecast combinations is to achieve more accurate results compared to individual models in terms of performance metrics by integrating different models. In the literature, aggregating multiple individual forecasts is considered a valuable approach for obtaining more reliable results by leveraging the ability to capture different aspects that cannot be identified by a single model (Clemen, 1989; De Menezes et al., 2000; Hendry and Clements, 2004; Armstrong, 2007; Makridakis at al, 2020; Mariñas-Collado et al., 2022; Wang et al., 2023). According to Petropoulos et al. (2018), combining multiple forecasts is often better for reducing ambiguities around parameter specification and model forms such as data uncertainty and parameter uncertainty by incorporating multiple drivers of the data-generating process.

This is not a completely immature concept. More than five decades ago, John Bates and Clive Granger wrote a paper that can be described as a cornerstone of this field. They computed the combination weights for the models by utilizing the calculated mean squared prediction error matrix's diagonal components in their research and discovered that, when compared to separate models, the findings were more accurate in terms of MSE (Bates and Granger, 1969). Also, Clemen (1989) and Timmermann (2006) made major contributions to this field with their studies in the years to come.

Clemen (1989) provides a wide literature review about forecast combinations by focusing on early papers, benchmarks for forecast evaluation under both theoretical and empirical studies, and contributions from forecasting, psychology, statistics, and management science. Academic studies including simple averaging, regression-based models, Bayesian techniques, and mathematical programming approaches to assign weights for individual models were shared. Also, subjective combination procedures were discussed in his study.

According to our observations, Timmermann (2006) presents the most extensive paper on the theoretical foundations and scopes of forecast combinations up to the time the study was published. Combinations' dependence on asymmetric loss is examined, along with several potential issues with model misspecification, instability (non-stationarities), and estimate error. A combination of point, interval, and probability forecasts are discussed too. He lists arguments of the forecast combinations in the literature as;

1. Even when one forecasting method outperforms the other, combining both methods can often yield superior results compared to relying on either method individually.
2. Individual forecasts may be differently influenced by structural breaks in the data. By combining models with varying degrees of adaptability and stability to such breaks, the resulting forecast can often outperform those generated by single models.
3. Combining forecasts across diverse models enhances robustness, mitigating the effects of misspecification biases and measurement errors inherent in the datasets used for individual forecasts.
4. Forecasts based on differing loss functions may lead to improved outcomes when combined. However, in the example of individual forecasts y_1 and y_2 may produce identical expected losses, and it may not be optimal to combine them in certain cases - such as when the forecasts are perfectly correlated or due to estimation errors in determining combination weights.

In the following years, a research team led by Spyros Makridakis organized a series of forecasting competitions known as the Makridakis Competitions. They launched the first competition in 1982 and have since organized 6 competitions to evaluate and compare the performance of different forecasting methods. While not every Makridakis Competition specifically focused on forecast combination methods, their effectiveness has consistently been emphasized across competitions, particularly when diverse forecasting approaches are used or when hierarchical forecasting structures are involved. The detailed explanations and results of the competitions are summarized in Table 2.1 by using the references in the literature (Hyndman and Koehler, 2006; Jurman et al., 2012; Koning et al., 2005; Makridakis et al., 1982; Makridakis and

Hibon, 2000; Makridakis et al., 2020; 2022; Makridakis et al., 2022; 'Makridakis Competitions' 2023; Armstrong, 2007; Makridakis et al., 2023).

Wang et al. (2023, p.3) provided a graph that indicates the percentage of forecasting publications published in the Web of Science that deal with forecast combinations has been increasing over the previous 50 years, as shown in Figure 2.1. The proposal about combination forecasts is reaching 13.80% in 2021, as seen.

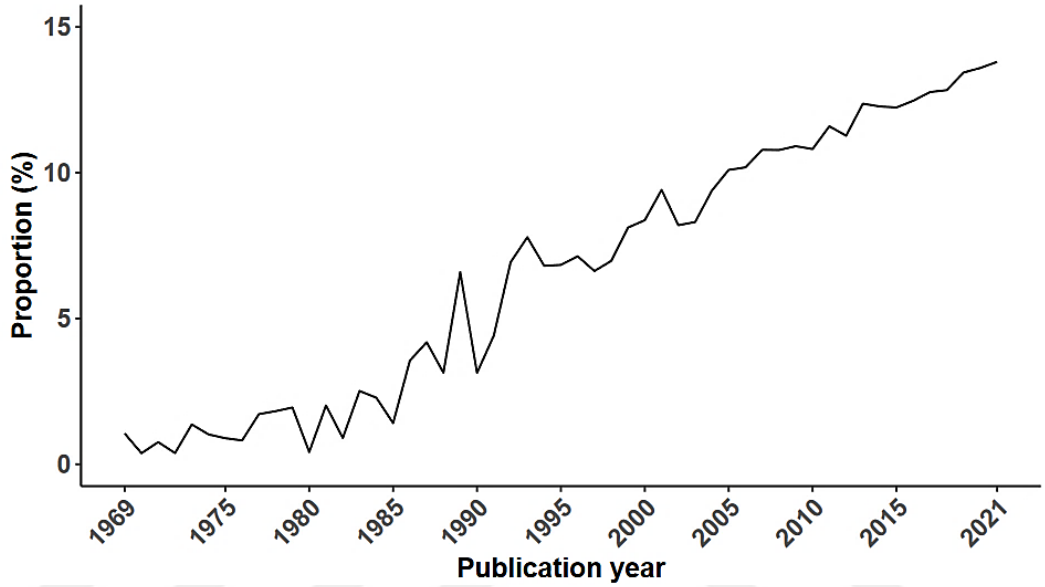

Figure 2.1: The percentage of forecast combination-related studies among all forecasting papers published between 1969 and 2021 and listed in Web of Science databases (Wang et al., 2023, p.3).

Finally, there is broad agreement following years of research that "combining forecasts reduces the final forecasting error" (Maté, 2021), however, there is no clear consensus on the best time series forecast combination method for a given situation (Blanc and Setzer, 2016). Which combination method to be used or the effectiveness of combination methods varies depending on the characteristics of the dataset being analyzed. That is why, in the literature, the methods related to time series forecast combinations vary specific to the data to be applied, and the studies in this field have been written accordingly. Especially in recent years, time series forecast combinations have been used in a variety of applications with encouraging outcomes, as can be noticed.

Table 2.1: Makridakis competitions.

Competition name	# of time series and forecast methods used	Hybrid methods	Focus	The results
M Competition (1982)	1001 time series with 15 (plus 9 variations) forecast methods	No	Short-term forecasting of univariate time series	Simpler techniques, such as exponential smoothing, frequently performed better than more intricate ones. The accuracy metric being utilized determines how the different methods' performances are ranked in relation to one another. The length of the forecasting horizon used determines how accurate the different approaches are.
M2 Competition (1993)	29-time series containing two combined forecasts, one overall average, and 16 forecast methods (including 5 human forecasters and 11 automatic trend-based approaches).	Not explicitly	Both univariate and multivariate time series forecasting	It was asserted that the outcomes were statistically identical to the M-Competition's. The most favorable outcomes were frequently obtained by combining forecasts from many methods (ensemble methods), but no single strategy consistently beat others across all series. (The MAPE of 11.7% was obtained by combining three exponential smoothing techniques, while the average component's MAPE was 12.2%.)
M3 Competition (2000)	3003 time series with 24 forecast methods	Not explicitly	Diverse types of time series data, including longer-term series and seasonal data	The authors claim that the M3 competition's results were comparable to those of the previous contests. It supported the M2 Competition's conclusions about how well forecasts from several approaches may be combined.
M4 Competition (2020)	100000 time series with all major ML and statistical methods	Yes	Forecasting monthly, quarterly, and yearly series, including univariate and multivariate data.	According to an analysis of more than 100 forecasting methodologies, hybrid and ensemble approaches—especially those that combine machine learning and statistical methods—tended to perform the best. An ML algorithm that was trained to minimize forecasting error through holdout testing determined the weights for the averaging, making the combination of seven statistical methods and one machine learning method the second most accurate.
M5 Competition (Initial results 2021, Final 2022)	All of the main forecasting techniques, including statistical, machine, and deep learning, were used to predict about 42,000 hierarchical time series.	Yes	Hierarchical time series forecasting, which involves forecasting at multiple levels of aggregation.	It brought new evaluation measures for assessing forecast accuracy and attracted an important participant number. The contest demonstrated how crucial it is to use sophisticated statistical and machine learning methods in addition to taking the data's hierarchical structure into account in order to get precise projections. The three top performers each used ensembles, or combinations, of independently trained and tweaked models, each with a unique training process and training dataset, in accordance with the M4 Competition.
M6 Competition: 2022 preliminary results, 2024 final results	50 S&P and 500 US equities and 50 overseas exchange-traded funds (ETFs) are used in this real-time financial forecasting competition, which uses all the main forecasting techniques, such as machine learning, deep learning, and statistics.	Yes	improving decision-making in financial markets through advanced forecasting methods	The M6 demonstrated that forecast combinations could improve decision-making in financial markets by leveraging diverse models to generate better predictions and investment strategies. It highlights the importance of designing combination strategies that adapt to market complexity and uncertainty. Filip Staněk, the global winner of the M6 Financial Forecasting Competition, used a special AI-based model to improve financial forecasting, referred to as MtMs (Meta-to-Mesa), which involves a hypernetwork that designs specific forecasting models tailored for different tasks.

Stock and Watson (2004a) applied time series forecast combinations with up to 73 predictors to estimate the output growth of seven countries by using an economic dataset covering 1959–1999 quarterly. Even though individual predictor-based forecasts are inconsistent over time and amongst countries, and they typically underperform an autoregressive benchmark, combination forecasts frequently outperform autoregressive forecasts. Kapetanios et al. (2008) described some features of the Suite[1] which focuses on combining a limited number of forecasts produced in an optimal way using various information sources and procedures. In their study, the time of bank independence (from 1997 Q2) is forecasted and evaluated over by MSE. Combinations of predictors produced accurate forecasts of the major macroeconomic indicators, which can be used as a benchmark forecast without bias to compare with the projections of policymakers.

Azevedo and Campos (2016) presented a forecast for the spot prices of Brent North Sea (Brent) and West Texas Intermediate (WTI) crude oil by three individual models (ARIMA, exponential smoothing, and dynamic regression) and the combination model via linear regression. The combination is made only with the ARIMA and exponential smoothing because the dynamic regression model was not valid. Models are compared on the validation set and for both oil prices, the combination model performed better than individual models including comparing models (neural network and naïve forecast) based on the MAPE performance metric.

In addition, new developments in computational intelligence have led researchers to evolve traditional time series methods or develop hybrid models that integrate them with machine learning algorithms in order to improve prediction accuracy on complex data sets (Isikli and Serdarasan, 2023). For example, Mariñas-Collado et al. (2022) compared and combined the forecast models generated by traditional methods and machine learning algorithms to predict urban bus passenger demand in Spain. They used a large dataset containing information on the hourly passenger arrivals at 272 bus stops. First, the bus stops were grouped into clusters. Then, several single forecast models (e.g., SVM, ARIMA, TBATS) were applied, as well as a combination of the

[1] As inputs to the forecasting process, the Bank of England has developed a "suite of statistical forecasting models" (the "Suite") that provide judgment-free statistical estimates of output growth and inflation as well as measures of pertinent news in the data.

best-performing models using methods like simple average, CLS regression, and Bates & Granger weighted mean. The results showed that the Bates and Granger method reduced the errors more effectively than the simple average method, although the difference was not very significant. Additionally, the SVM model performed as well as the combined models on its own.

Meng, Liu, and Feng (2022) sought to forecast monthly, quarterly, and annual numbers of occurrence fatalities and manufacturing safety incidents in Chinese housing municipal engineering by analyzing existing data from 2009 to 2019. In their research, they did not combine separate methods on the same time period; instead, they divided the time series according to the structure of the data, evaluated each segment with appropriate methods, and then combined the results. To divide the time series, they used the fractional-order accumulated Gray model, suitable for high dynamics, short sample sizes, and univariate models during January and February, which correspond to the Spring Festival period in China. For the rest of the remaining year, they used the ARIMA method, which is typically used for more regular and long-term samples. This approach resulted in satisfactory outcomes based on their methodology.

In the next section, theoretical modelings of major time series forecasting combination methods are presented with a schematic illustration.

2.2 Time Series Forecast Combination Methods and Schematic Outlines

From simple combination methods without estimate to complex strategies utilizing time-varying weights, nonlinear combinations, component correlations, and cross-learning, combination schemes have seen significant development (Wang et al., 2023).

Wei and Yang (2012) summarized the proposed methods for forecast combinations as those that rely on forecast optimization under variance-covariance estimation models similar to Bates and Granger (1969), Bayesian methods (Min and Zellner, 1993), and methods that consider structural breaks (Timmermann, 2006).

Similar to the work of Wei and Yang (2012), attempts have been made in a limited number of other studies to systematically outline combination methodologies (Wang et al., 2023; Weiss et al., 2019). Building upon all these studies, the following scheme in Figure 2.2 has been developed.

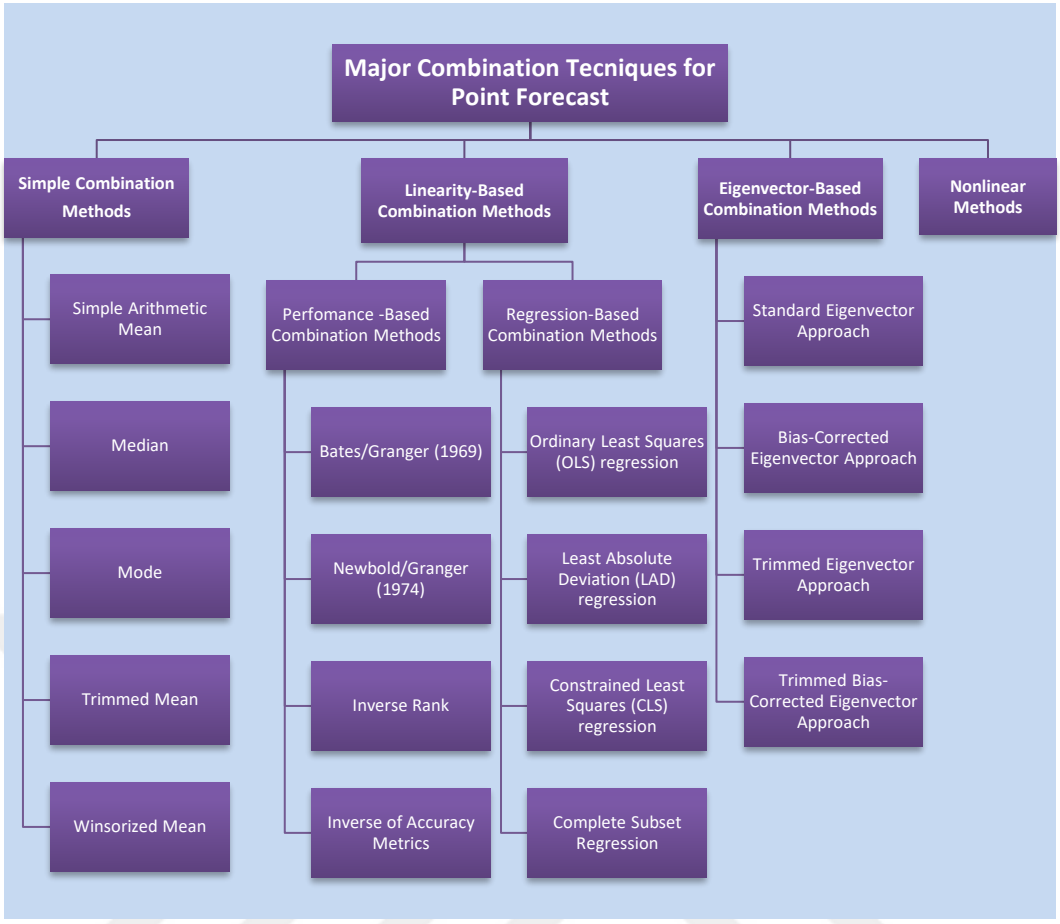

Figure 2.2: Major forecast combinations techniques.

In the rest of the study, explanations, formulations, and comparisons of the methods given in the context of this scheme are given. While presenting the formulations, consistent use of indices to represent the same notation has been ensured for whole methods. In this chapter, the following notations are used for the formulations: R represents the number of single prediction models combined; j donates the type of single prediction model. Additionally, f_{ji} represents the value of the forecast produced by the single prediction model j at time point i, while λ indicates the correction factor; n, is the number of observations predicted, a_i denotes the actual value at time point i; $\widehat{\omega}_j$ represents the weight assigned to the single prediction model of j and ε_i is the error value at time point i.

2.2.1 Simple combination methods

Compared to their counterparts, simple combinations can be implemented more effectively and with less computing load according to Wang et al. (2023, p. 1521). They distinguish themselves from other methods as they do not require parameter

(weight) estimation and, therefore, are not subject to estimation errors. Furthermore, these methods are computationally efficient and easy to apply. Simple combination techniques can be counted as simple arithmetic mean, geometric mean, harmonic mean, mode, median, trimmed mean, and winsorized mean. The formulations of each method can be seen in Table 2.2.

Table 2.2: Simple combination methods (Weiss et al., 2019).

Combination Method	The formulation	Combination Method	The formulation
Simple Arithmetic Mean	$f_i^c = \frac{1}{R}\sum_{j=1}^{R} f_{ji}$	The Median	For odd R: $f_i^c = f_{\left(\frac{R}{2}+0.5,i\right)}$ For even R: $f_i^c = \frac{1}{2}\left(f_{\left(\frac{R}{2},i\right)} + f_{\left(\frac{R}{2}+1,i\right)}\right)$
Geometric Mean	$f_i^c = \left(\prod_{j=1}^{R} f_{ji}\right)^{\frac{1}{R}}$	Winsorized Mean	$f_i^c = \frac{1}{R}\left[\lambda R f_{(\lambda R+1,i)} + \sum_{j=\lambda R+1}^{R-\lambda R} \lambda R f_{(R-\lambda R,i)}\right]$
Harmonic Mean	$f_i^c = \frac{R}{\sum_{j=1}^{R} \frac{1}{f_{ji}}}$	Mode	$f_i^c = Mode\{f_{1i}, f_{2i}, \ldots., f_{Ri}\}$
Trimmed Mean	$f_i^c = \frac{1}{R(1-2\lambda)} \sum_{j=\lambda R+1}^{(1-\lambda)R} f_{ji}$		

In the literature on time series forecast combinations, it is widely observed that the most commonly employed and accepted method for creating a combined forecast is the use of a simple average or equal weighting of the individual forecast models. According to Clemen (1989) averaging individual models is often the most effective strategy for generating a combined model. This concept has been referred to as the "forecast combination puzzle" in the literature, a term first introduced by (Stock and Watson, 2004). This argument has remained hard to beat for the past thirty years (Weiss et al., 2019; Wang et al. 2023). Key benefits of using the simple arithmetic average of time series forecast are also given by Palm and Zellner (1992):

1. Because combination weights are equal and no weight estimation is required,
2. In many situations, by averaging out the individual bias of predictors, it lowers variance and bias;

3. When how to estimate weights is uncertain, simple averaging can be the most effective way.

Supporting this claim, Blanc and Setzer (2016) stated that among the approaches studied, the simple arithmetic mean was not consistently outperformed by any other method in out-of-sample evaluations, as noted in their reviews.

In the literature, it is also discussed when the simple weighting outranks other alternatives. About this issue, Timmermann (2006, p. 146) mentions that when the individual forecast errors have identical variance and the same pair-wise correlations irrespective of the correlation between the two forecast errors, equal averaging of predictors is optimal. In this way, diversification gained through combination becomes highest.

Other simple combination methods, such as the median, and mode, besides trimmed and winsorized means also took attention in the literature. It is described that these methods are less vulnerable than a simple average to extreme forecasts and can be more useful for some scenarios (Lichtendahl and Winkler, 2020).

According to Jose and Winkler (2008), trimmed and winsorized means are promising due to their simplicity and good performance when there is a high level of variability among the predictors. Winsorized mean method does not completely remove outliers as in the trimmed mean method but it caps them at a certain level, that is why, it can be considered as a softer approach compared to the trimmed mean when dealing with outliers. Both methods are less sensitive to outliers than the simple average.

2.2.2 Linearity-based combination methods

Linearity-based forecast combinations cover assigning weights based on the accuracy metrics of component forecasts. These weights are determined through optimization procedures, such as minimizing forecast error (Bates and Granger, 1969; Granger and Newbold, 1974; Stock and Watson, 2004).

2.2.3 Performance-based combination methods

Under this title, some performance-based forecast combination methods are given in Table 2.3. The first one is Bates and Granger's method, which is considered the initiator of time series forecast combinations; the other one is the extended study of

them, which belongs to Granger and Newbold (1974). Rank-based weighting method is also given under the performance-based combination method.

Table 2.3: Performance-based combination methods.

Combination Method	The formulation	Combination Method	The formulation
Bates and Granger (1969)	$f_i^c = \sum_{j=1}^{R} f_{ji} \times \frac{\hat{\sigma}^{-2}(j)}{\sum_{j=1}^{R} \hat{\sigma}^{-2}(j)}$	Newbold and Granger (1974)	$f_i^c = \sum_{j=1}^{R} f_{ji} \times \frac{\Sigma_{jk}^{-1} e}{e^{\mathrm{T}} \Sigma_{jk}^{-1} e}$
Rank Based Weighting*	$\omega_j = \frac{\frac{1}{rank_j}}{\sum_{j=1}^{R} \frac{1}{rank_j}}$; $f_i^c = \sum_{i=1}^{R} \omega_i f_{ji}$	Weighting based on an inverse of performance metrics **	

Note: $\hat{\sigma}^{-2}(j)$ is the inverse of the error variance of predictor j, meaning forecasts with lower variance get higher weights; Σ is RxR error variance-covariance matrix of the predictors where σ_{jj} is the variance of forecast errors for model j, and σ_{jk} is the covariance between forecast errors of models j and k. In this case, Σ^{-1} is the inverse of Σ matrix. Finally, e is denoted as Rx1 vector of $(1,1,\ldots,1)^{\mathrm{T}}$.

*Aiolfi and Timmermann (2006)

**Some studies weighted the component forecasts by using the inverse of the performance metrics such as MSE, RMSE, or sMAPE (Nowotarski et al., 2014; Stock and Watson, 1998; Pawlikowski and Chorowska, 2020; Stock and Watson, 2004).

2.2.4 Regression-based combination methods

Crane and Crotty (1967) were the first to propose the idea of combining forecasts using regression. In subsequent years, Granger and Ramanathan (1984) also studied about this methodology for econometric models and regression methods became more popular for forecast combination literature.

The regression-based forecast combinations are linear functions of predictors where the weights are determined by utilizing individual forecasts as the explanatory variables and actual values as the response variable. It is possible to derive regression models with different objective functions and constraints. With these derivations, regression-based combination methods can be summarized as; ordinary least squares (OLS) regression, constrained least squares (CLS) regression, least absolute deviation (LAD) regression, and complete subset regression (CSR). Table 2.4 shows the formulation of all four techniques.

Table 2.4: Regression-based combination methods.

Combination Method	The formulation	Combination Method	The formulation
Ordinary Least Squares (OLS) Regression	$Z{=}min\ \sum_{i=1}^{n}(a_i - f_i^c)^2$ $f_i^c = \hat{\alpha} + \sum_{i=1}^{R} \widehat{\omega}_j f_{ji} + \varepsilon_i$	Constrained Least Squares (CLS) regression	$Z{=}min\ \sum_{i=1}^{n}(a_i - f_i^c)^2$ $f_i^c = \hat{\alpha} + \sum_{i=1}^{R} \widehat{\omega}_j f_{ji} + \varepsilon_i$ $\widehat{\omega}_j \geq 0\ \ \forall j\ , \sum_{j=1}^{R} \widehat{\omega}_j = 1$
Least Absolute Deviation (LAD) Regression	$Z{=}min\ \sum_{i=1}^{n}\lvert a_i - f_i^c\rvert$ $f_i^c = \hat{\alpha} + \sum_{i=1}^{R} \widehat{\omega}_j f_{ji} + \varepsilon_i$	Complete Subset Regression	$f_i^c = \frac{1}{K}\sum_{k=1}^{K}\left(a_k + \sum w_{j,k} f_{ji}\right)$ $w_j = \frac{exp(-1\ /\ 2\varphi_j)}{\sum_{j=1}^{n} exp(-1\ /\ 2\varphi_j)}$ where K is the number of subsets, S_k is the k-th subset of predictors, φ is the information criterion. For R predictors, there are 2^R - 1 subsets.

OLS regression minimizes squared residuals, whereas LAD regression minimizes absolute residuals. By adding a few more constraints to the same objective function of OLS, CLS regression can be modeled. Although CLS is theoretically suboptimal compared to Ordinary Least Squares (OLS) due to its additional constraints and lacks the asymptotic properties of OLS, it often performs better in practice, particularly when individual forecasts are highly correlated, and its weights are more interpretable (Weiss et al., 2019). Adding a constraint that the sum of the weights to one as in the CLS method guarantees that the combined forecast is also unbiased (Granger and Ramanathan, 1984). The two-step combination method CSR uses the full subset regression approach to generate *n* combined predictions using the original *R* number of forecasts as predictors first, then, these forecasts are combined using weights based on information criteria (Elliott et al., 2013).

2.2.5 Eigenvector-based combination models

The concept of minimizing the mean squared prediction error under a normalization condition is the foundation of the eigenvector-based forecast combination techniques presented by Hsiao and Wan (2014). The four eigenvector-based methods presented in Table 2.5 employ a normalization technique ensuring that the sum of the squares of the component weights equals one.

Combination weights are obtained from the estimated mean squared prediction error matrix using the conventional eigenvector method. This method takes the MSPE matrix's P positive eigenvalues are grouped in ascending order. At the formulations shown in Table 2.5, κ_l represents the eigenvector Θ_l. Additionally, $d_i = e \cdot \kappa_l$ where e is the transposed vector of $(1, 1, \ldots 1)_{P\times 1}$.

Table 2.5: Eigenvector-based combination methods.

Combination Method	The formulation	Combination Method	The formulation
Standard Eigenvector Approach	$\omega = \frac{1}{d_l}\kappa_l \, ; f_i^c = \sum_{j=1}^{R} \omega_i f_{ji}$	Bias-Corrected Eigenvector Approach	$\omega = \frac{1}{\tilde{d}_l}\tilde{\kappa}_l \, ; \, f_i^c = \alpha + \sum_{i=1}^{R} \omega_i f_{ji}$
Trimmed Eigenvector Approach	$\omega = \frac{1}{e'\kappa_l}\kappa_l;$ $f_i^c = \frac{1}{R(1-2\lambda)} \sum_{j=\lambda R+1}^{(1-\lambda)R} \omega_i f_{ji}$	Trimmed Bias-Corrected Eigenvector Approach	$\omega = \frac{1}{\tilde{d}_l}\tilde{\kappa}_l$ $f_i^c = \frac{1}{R(1-2\lambda)} \sum_{j=\lambda R+1}^{(1-\lambda)R} \omega_i f_{ji}$

NOTE: $\tilde{d}_l$ and $\tilde{\kappa}_l$ have the same definition as d_i and κ_l in the conventional eigenvector technique; the main distinction is that they relate to the spectral decomposition of the centered MSPE matrix instead of the original (uncentered) MSPE matrix. Both trimmed eigenvector approaches combines the advantageous changes of the first two approaches by pre-selecting component models that are used as input for the forecast combination; only a portion of the available forecast models are kept, and the models with the worst performance are discarded.

2.2.6 Nonlinear forecast combinations

Linear combination methods inherently assume a linear relationship between the constituent forecasts and the target variable. However, in such cases, it may be appropriate to relax this linearity assumption and explore nonlinear combination schemes of greater complexity, which have received limited research attention (Wang et al. 2023, p. 1525).

Two forms of nonlinearities in prediction combinations were distinguished by Timmermann (2006) one directly incorporates nonlinearities in the combination parameters, while the other combines nonlinear functions of individual forecasts with linear combination weights.

Because neural networks are capable of catching the underlying nonlinear link between individual forecasts and future events, they are frequently studied in time series forecast combination literature (Donaldson and Kamstra, 1996; Babikir and Mwambi, 2016; Krasnopolsky and Lin, 2012).

3. METHODOLOGY

3.1 Road of Implementation

As a road to the implementation of the whole methodology, the map in Figure 3.1 has been constituted and followed. First, the datasets to study are chosen. Then, an examination of the identified datasets is carried out by some statistical analysis. This analysis can be carried out with the help of graphs and correlograms of the datasets. Together with the results obtained, at this phase, the stationarity of the series should be examined by stationarity tests. If it is concluded that the dataset is stationary, whether it is white noise should be questioned. Datasets that are white noise are not suitable for building different forecasting models since the pure noise forecasts are given no weight in the optimum forecast combination (Timmermann, 2006, p. 162), and hence cannot serve the proposed model for this study. However, suppose it is concluded that the dataset is not stationary. In that case, it is necessary to determine whether the dataset needs a data imputation, outlier removal, or any kind of data transformation. After that, the presence of stochastic or deterministic trend should be questioned. Necessary actions are taken to achieve stationarity[2]. To apply forecasting models on a statistical basis, at least the estimated part of the dataset must be stationary (Timmermann 2006, 160). After reaching stationarity, the next step is to determine whether the transformed dataset is white noise. As mentioned earlier in this paragraph, it is only possible to apply the forecasting models on non-white noise datasets.

The prepared non-white noise datasets are now available to be separated into three subsets as training, validation, and test set. Individual forecasting models are trained on training set and then forecasted over validation and test set horizons. In this case, the length of the forecast horizon corresponds to the total length of the test and validation sets.

[2] Preparation tecniques and reaching stationarity steps are mention under the heading (3.2.1)

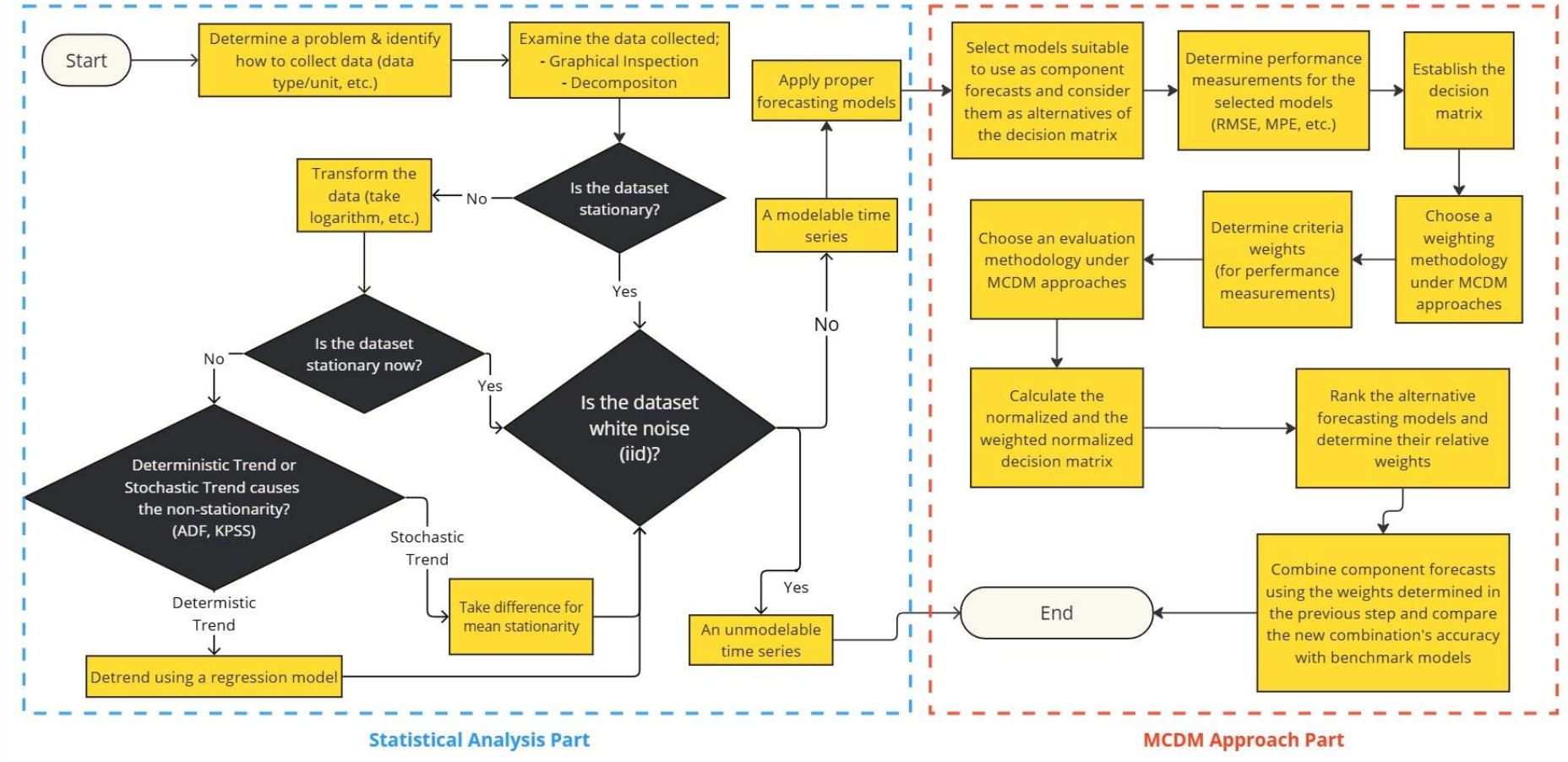

Figure 3.1: The implementation map.

Component forecasts[3] are selected based on their performance, diversity, and robustness considering visual analysis. In the decision matrix, which aligns predictors with performance measurements, each predictor's performance measurements are derived from the validation set. After that point, Multi-Criteria Decision Making (MCDM) techniques are applied.

First, the weights of the performance metrics, which represent the criteria weights, should be calculated. Then, performance evaluation for forecasting models should be carried out. For both assessments, the proper MCDM approaches are determined. In this study, the CRITIC and CILOS methods were utilized for criteria weight assessment, while TOPSIS and ARAS were employed for performance evaluation. According to which MCDM approach and weighting technique are applied, a weighted normalized decision matrix created can vary.

Finally, the stage is reached where normalization can be applied to the rankings obtained from the MCDM technique. The normalized ratio corresponding to each predictor will represent the weight within the combination model.

The following sections of this chapter provide a description of the methods and steps involved in the implementation path. The techniques used in the determination and preparation of the dataset and how the prediction methods to be included in the decision matrix were determined are explained in detail under separate headings. Accuracy performance measurement techniques and multi-criteria decision-making approaches used in the study are also explained in the following.

3.2 Determination and Preparation of Datasets

3.2.1 Selection of datasets

While determining the datasets, the time frequency and sample size of the observations and the presence of patterns that are suitable for statistical estimation are taken into consideration. For time series analysis to be empirically successful and robust to outliers, a sufficiently large sample size must be used to distinguish between noise and

[3] Selection of the component forecasts are mentioned under the heading (3.3.1)

the main trend, ensuring the data size is adequate relative to the number of models (N) (Timmermann, 2006).

The required sample size can change based on several things like complexity and frequency of the dataset, however, generally having at least 50-100 data points is recommended to capture underlying patterns like seasonality or trend. For this thesis study, since combination models are considered and it is necessary to examine relatively complex datasets using many prediction models, we aimed to select datasets with at least 200 observations.

Since models rely on historical data, there is a risk that the fitting process may primarily capture transient features. Therefore, it is essential to validate the models using a dataset distinct from the one employed for estimation (Ülengin, 2024). With this in mind, in order to observe how the proposed model works on datasets with different frequency levels, 2 datasets with monthly and daily frequencies were found and tested.

Moreover, it was studied on datasets with different patterns, such as trend or seasonal effects, so that different forecasting models can capture different effects during the combination and allow the success of combination models to increase.

3.2.2 Data preparation

Data preparation is the first step that should be taken for time series forecasting and should be done before forecasting models are estimated. This process may involve transformation, outlier removal loading in data, identifying missing values, decomposition, and other pre-processing tasks (Hyndman and Athanasopoulos, 2021). Before applying these tasks, first, we should understand the features of the data.

3.2.2.1 Graphical inspection and correlograms

Graphical inspection is an essential step for understanding the features of the dataset. Visualization allows us to see common patterns, extreme observations, and changes over time. For this study, the Eviews program is used to obtain plots of the datasets.

As a more quantitative approach, the linear correlation (autocorrelation measures) between a time series' lagged values can give an idea about the behavior of this dataset (Hyndman and Athanasopoulos, 2021). If autocorrelation between successive values of the time series is not equal to zero, this indicates an association between values of

the identical variable across time periods which implies that the series is predictable and can be modelled. However, for a white noise time series, all autocorrelation measures give results as zero, as there is no dependency between values over time. As r_1 represents the measure of the relationship between y_t and and y_{t-1}, r_2 represents the measure of the relationship between y_{t-1} and y_{t-2} and, so on. In this case, r_k can be formulated as in Equation (3.1) (Hyndman and Athanasopoulos, 2021);

$$r_k = \frac{\sum_{t=k+1}^{T}(y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{T}(y_t - \bar{y})^2} \tag{3.1}$$

where T is the length of the time series.

If the correlation between values at different time lags, rather than just successive values, needs to be measured, *partial autocorrelations* can be calculated. Partial autocorrelations assess the relationship between y_t and y_{t-k} after excluding the effects of lags 1,2,3,..., *k-1* (Hyndman and Athanasopoulos, 2021). To understand the significance of both partial and ordinary autocorrelations, confidence intervals are constructed. Whereas in literature generally $\pm 1.96/\sqrt{T}$ is used as a critical value, in this study the Bartlett formula[4] based confidence intervals were used as in Equation (3.2):

$$\pm 1.96 \cdot \sqrt{\frac{1}{T}\left(1 + 2 \cdot \sum_{i=1}^{k-1} r_i^2\right)} \tag{3.2}$$

If r_k exceeds outside this confidence interval above, it is considered that the autocorrelation exists and is significantly different from zero at a 5% significance level.

The plot displaying autocorrelations as bars is called the ACF plot, while the one showing partial autocorrelations is called the PACF plot. These are sometimes been called the correlogram (Box et al. 2016, 30). Ziegel (1995) mentions that the ACF (the autocorrelation function) and PACF (the partial autocorrelation function) act as helpful measures in the method of recognizing and estimating time series proposed by Box and Jenkins (1976). Correlograms of ACF and PACF are important plots that should

[4] A Bartlett interval refers to a confidence interval for the variance of a normal distribution, derived using Bartlett's chi-squared test for variance.

be inspected before choosing and applying forecast models to understand the behavior of datasets.

These plots are not used as a stationarity test for the dataset neither help to understand if the data contains stochastic or deterministic trend, but help to understand trend/seasonal patterns. For a time series with trend, the PACF will drop to zero relatively quickly, while the ACF bars decrease slowly (Hyndman and Athanasopoulos, 2021). For this study, ACF/PACF plots with Bartlett formula-based confidence intervals are used by the Gretl Program.

3.2.2.2 Investigating stationarity

Kwiatkowski et al. (1992) described that a stationary time series is one whose statistical characteristics are independent of the series' observation time. Stationarity is essential for time series forecasting and should be guaranteed before using forecasting models, as many of them assume that the statistical characteristics of the data (mean, variance, and autocorrelation) remain constant across time. By eliminating (or reducing) trend and seasonality, transformations like logarithms can help stabilize the variance, moreover, differencing or detrending can assist in stabilizing the mean of a time series (Hyndman and Athanasopoulos, 2021).

If the dataset shows a deterministic trend, *detrending* should be handled to get a stationary series. On the other hand, the non-stationary time series which includes a unit root/ stochastic trend should be *differenced*. If a series containing stochastic trends is detrended, then it will fail to achieve stationarity (Ülengin, 2024; Ryan et al., 2023).

The stationarity test helps us to understand if the dataset is stationary or non-stationary with a stochastic trend or non-stationary with a deterministic trend.

For this study, we will use ADF and KPSS stationarity tests simultaneously on Eviews software.

Augmented Dickey-Fuller (ADF) test: The test, which was created by Dickey and Fuller (1979), examines a unit root in a time series, indicating non-stationarity with a stochastic trend. The null hypothesis is that a unit root exists. The test can be tested on three different equations numbered with 3.3, 3.4 and 3.5 given below.

- Test equation without intercept or trend notations:

$$\Delta y_t = \gamma y_{t-1} + \sum_{i=1}^{k} \beta_i \Delta y_{t-i} + \varepsilon_t \quad (3.3)$$

- Test equation with intercept notation only:

$$\Delta y_t = \mu + \gamma y_{t-1} + \sum_{i=1}^{k} \beta_i \Delta y_{t-i} + \varepsilon_t \quad (3.4)$$

- Test equation with trend and intercept notations:

$$\Delta y_t = \mu + bt + \gamma y_{t-1} + \sum_{i=1}^{k} \beta_i \Delta y_{t-i} + \varepsilon_t \quad (3.5)$$

where:

Δy_t is the differenced time series,

μ is a constant term,

b represents the trend,

γ is the coefficient of the lagged level of the series,

k is the number of lagged terms.

The ADF test checks if γ is different from zero significantly. If it is not, then the series has a stochastic trend and we reject the null hypothesis (stationary). In this study, the order of application of the tests will be carried out by going from general (test equation with trend and intercept notations) to specific (test equation without intercept or trend notations).

Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test: Kwiatkowski et al. (1992) proposed the KPSS test of the null hypothesis, which assesses the degree to which a time series deviates from a stationary mean or trend and assumes that a measurable series is stationary around a deterministic trend. The test of the hypothesis that the random walk has zero variance is the LM (Lagrange multiplier) test, and the series is written as the sum of the deterministic trend, random walk, and stationary error. The equation used for KPSS statistics can be seen in Equation (3.6).

$$\text{KPSS statistics} = \frac{1}{T^2} \sum_{t=1}^{T} S_t^2 \,/\, s^2 \quad (3.6)$$

where:

S_t is cumulative sum of the residuals,

s^2 is long-run variance of residuals,

T is the length of the time series.

3.2.3 Splitting the dataset

For this study, the dataset will be split into three subsets: *train*, *validation*, and *test.* The forecast model parameters are determined on the train set. Forecast model weights within the combination will be determined by validation set and finally, the test set will help to understand the accuracy of the model created via train and validation set on unseen data. This splitting strategy ensures that the generalization of the model is made well for unknown future data, avoiding overfitting and evaluating a variety of potential models' performance reliably (Géron, 2019).

Furthermore, the size of the selected subsets is important for the appropriate determination of the parameters of the forecast models and their weights in the combination. It is crucial to take into account whether models estimated with daily intervals can accurately capture the process dynamics across longer timeframes, like weeks or months, when deciding how far to extend the forecast horizon (Timmermann 2006, 160). The validation set should be large enough to detect significant weighting for component forecasts and even a small weight in the forecast combination can potentially improve forecast performance. It is customary to reserve 20% of the data for testing and utilize the remaining 80% for training. In this thesis study, we have also tried to stick to these ratios based on this generalization and separate nearly 75% of the series for training, 15% for validation, and 10% for testing.

3.3 Forecasting Model Selection as Rows of Decision Matrix

3.2.1 How to select predictors for combination?

To create a combined forecast model, first developing multiple individual models based on the time series data is held independently. Then, the forecasts from each model can be combined to extract more information from the original dataset, which can enhance the accuracy of the overall forecast (Wang et al., 2023).

However, selecting the component forecast models to be combined, which will form the alternatives in the decision matrix, is another step that needs to be addressed. Based on Timmermann (2006), the combination’s success depends largely on selecting the

right forecasts to combine. Ideally, the forecasts should differ, with some being higher and others lower than the actual outcome, in order for the forecast errors to counterbalance one another. This rarely happens in practice, though, because predictors are often trained using alike data and techniques. It is doubtful that the combination will increase accuracy if all of the individual forecasts chosen are identical or contain a lot of overlapping data (Wang et al., 2023).

Lichtendahl and Winkler (2020) also emphasized a high degree of diversity among single forecast models is essential for the success of the combination. They also noted that the accuracy of the selected models impacts the success of the combined forecast. For Franses (2011), the individual forecast models that contribute significantly to the overall accuracy of the forecast are the ones that need to be combined.

To utilize a variety of models, researchers have attempted to incorporate numerous forecasts into a regression framework (Wang et al. 2023), Armstrong (1989) suggested that, instead of combining all forecasts, it is often beneficial to discard the poorest-performing models - a process known as 'trimming' - and focus only on the best models under the principle of ‘using sensible models.’ Supporting this claim, Timmermann (2006) mentioned that once the best prediction models have been identified, the final model combines the different forecasts from each model to provide comprehensive information too.

Hendry and Clements (2002) highlighted that forecast combinations should result in even greater MSE reductions when two or more positively linked individual models are susceptible to shifts in opposite directions.

Swanson and Zeng (2001) investigated selection techniques based on regression to decide which models to include in forecast combinations.

Aligning combining rules to forecasting scenarios will be more successful the more we know which sets of underlying assumptions are connected to which combining rules. In the model proposed in this thesis study, component forecasts are selected by excluding the poorest-performing models, ensuring a high degree of diversity among individual forecasting models, and accounting for potential shifts in opposing directions.

3.2.2 Forecast models for time series analysis

3.2.2.1 STL (Seasonal and trend decomposition using Loess)

The STL methodology, which was originated by Cleveland (1990), breaks down a time series into three parts*: "trend," "seasonality,"* and *"residual"* .These parts are then estimated independently using the locally weighted regression method (Loess). It is considered particularly useful for understanding the underlying patterns in datasets that contain significant seasonal effects.

The general form of the STL model is in Equation (3.7):

$$f_i = b_i + S_i + e_i \tag{3.7}$$

where:

f_i is the forecasted values of the time series at time point i.

b_i is the trend component at time i and can be modelled as $b_i = \alpha + \beta i$ where α is intercept and β is the slope of this trend regression model.

S_i is the seasonal component at time point *i* and allows to capture the seasonality by assigning seasonal indices as $S_i = s_{season}(i)$.

e_i is the remainder component at time i.

Although STL can manage any kind of seasonality, allowing users to determine the trend's smoothness and enable the seasonal component to alter over time, it only offers additive decomposition capabilities and does not automatically handle calendar or trading day variations (Hyndman and Athanasopoulos, 2021).

3.2.2.2 ETS (Error, Trend, Seasonality)

A class of exponential smoothing methods called ETS models is used to capture time series *error, trend,* and s*easonality*; structured in various ways based on different combinations of these three components (Hyndman et al., 2008). So, the extention of traditional exponential smoothing methods becomes possible via ETS framework by different combinations of these components. ETS model notation is donated by ETS (E, T, S) where each parameter represents the component of error, trend, and seasonality as can be seen in Figure 3.2.

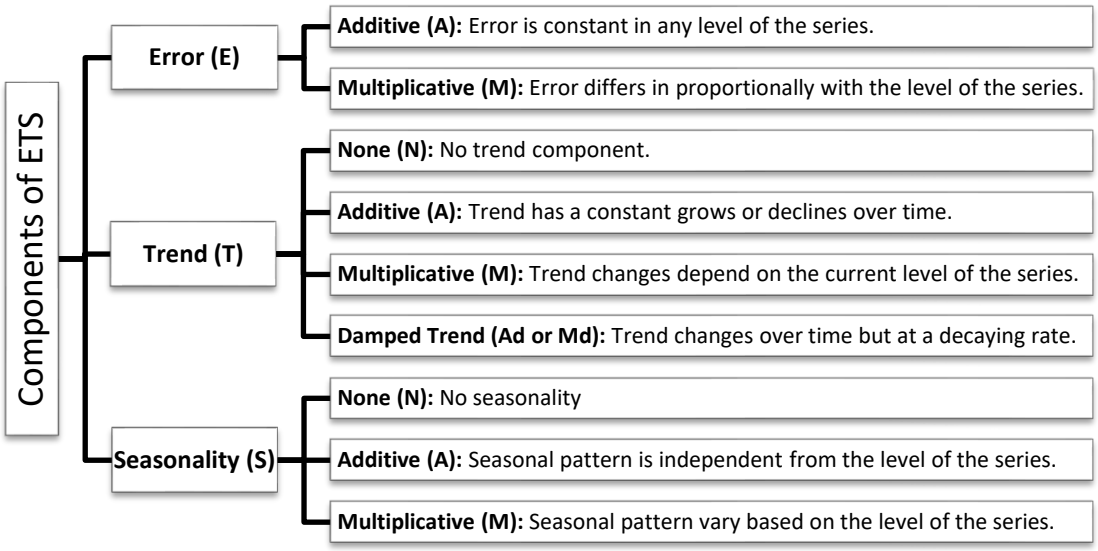

Figure 3.2: ETS framework.

Figure 3.3 also shows the state space equations for each scenario (ℓ_t represents the level of series at time *t*, b_t the slope at time *t*, s_t the seasonal component of the series at time *t*, and *m* the number of seasons in a year; the smoothing parameters are α, β^*, γ and ϕ; $\phi_h = \phi + \phi^2 + \cdots + \phi^h$; and the integer fraction of $(h-1)/m$ is *k*.

The selection of the ETS model varies on the characteristics of the time series and often, the estimation of model parameters is held by the Maximum Likelihood Estimation (MLE) technique. This approach maximizes the "likelihood", and so minimizes the sum of squared errors while optimizing the parameters.

Table 3.1 presents some of the most commonly used ETS methods, along with their notations and references.

Table 3.1: Commonly used ETS methods.

ETS method	Notation	Referance
Simple Exponential Smoothing (SES)	ETS(A, N, N)	(Brown, 1959)
Holt's Linear Trend method	ETS(A, A, N)	(Holt, 2004)
Holt-Winters' Additive Method	ETS(A, A, A)	(Winters, 1960)
Holt-Winters' Multiplicative Method	ETS(M, A, M)	(Winters, 1960)
Holt-Winters' Damped Method	ETS (A, A_d, A)	(Gardner, 1985)

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	$y_t = (\ell_{t-1} + \phi b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$

Figure 3.3: Equations in state space for every model in the ETS framework (Hyndman and Athanasopoulos, 2021).

3.2.2.3 ARIMA (AutoRegressive Integrated Moving Average)

By integrating autoregressive (AR), integrated (I), and moving average (MA) processes, the Box-Jenkins ARIMA family of linear statistical models, which are based on the normal distribution, can be used to simulate the behavior of a wide variety of real-time series (Siegel, 2016). Based on the primary study of Box et al. (2016), the model parameters (p, d, q) are chosen according to the data characteristics.

AR (autoregressive) part: It captures the relationship between each data point and its past values, effectively creating a regression of the variable on its own previous values. So, the AR part uses lagged values as predictors, with the number of lag terms specified by the order *p*. A common way to express an autoregressive model of order *p* is shown as in Equation (3.8):

$$y_t = \phi_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t \qquad (3.8)$$

where ϕ_0 is constant, ε_t is white noise, y_t is the variable we investigate its forecast for and p is the order of autoregressive part. For an AR(1) model:

- when $\phi_1 = 0$ and $\phi_0 = 0$, y_t is corresponding to *white noise*;
- when $\phi_1 = 1$ and $\phi_0 = 0$, y_t is *a random walk*;
- when $\phi_1 = 1$ and $\phi_0 \neq 0$, y_t is equated to *a random walk with drift*;
- when $\phi_1 < 0$, y_t tends to fluctuate around the mean.

Because stationarity is required, autoregressive model parameters are restricted by some additional constraints.

MA (moving average) part: While the AR part uses the past values of the time series, the MA part uses past forecast errors with the number of lag terms specified by the order q within the same regression structure as shown in Equation (3.9).

$$y_t = \phi_0 + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \tag{3.9}$$

where ϕ_0 is constant, ε_t is white noise and q is the order of moving average part.

Because we do not observe the values of ε, the regression model has a different sense compared to the AR part. For the moving average part, it tries to hit the target by correcting the next prediction error based on the current error with the number of error terms specified by the order q (Ülengin, 2024).

I (integration) part: The differencing operation making the dataset stationary is represented by this part. The order of differencing which is represented by the parameter d, determines how many times the series has been differentiated.

A belief that ARIMA models are broader than exponential smoothing techniques is one that is frequently spread. Although linear exponential smoothing models are special cases of ARIMA models, there are no analogous equivalents for non-linear exponential smoothing techniques in the ARIMA framework. Conversely, numerous ARIMA models exist without corresponding representations in the domain of exponential smoothing. Notably, all Exponential Smoothing State Space (ETS) models are inherently non-stationary, whereas ARIMA models encompass both stationary and non-stationary processes, highlighting fundamental differences in their scope and application Hyndman and Athanasopoulos (2021). Table 3.2 shows the equivalency interactions of ETS and ARIMA models.

Table 3.2: ETS and ARIMA model equivalency interactions (Hyndman and Athanasopoulos, 2021).

ETS forecast model	ARIMA forecast model
$ETS(A,N,N)$	$ARIMA(0,1,1)$
$ETS(A,A,N)$	$ARIMA(0,2,2)$
$ETS(A,A_d,N)$	$ARIMA(1,1,2)$
$ETS(A,N,A)$	$ARIMA(0,1,m)(0,1,0)_{\mathrm{m}}$
$ETS(A,A,A)$	$ARIMA(0,1,m+1)\ (0,1,0)_{\mathrm{m}}$
$ETS(A,A_d,A)$	$ARIMA(1,0,m+1)\ (0,1,0)_{\mathrm{m}}$

3.2.2.4 NNAR (Neural Network AutoRegressive)

As machine learning techniques continue to advance, neural networks—also referred to as artificial neural networks, or ANNs—have emerged as one of the most popular approaches for time series data in recent years. These networks serve as the foundation for deep learning algorithms used in predictive models (Su et al., 2024, p. 3530).

Similar to how lagged values are utilized in a linear autoregression model, they may also be used as inputs to a neural network when dealing with time series data. Neural network autoregression (NNAR) is the name of this time series-adapted forecasting model, which makes use of artificial neural networks to identify complicated and non-linear trends (Davidescu et al., 2021). The linear combination function can be represented as in Equation (3.10):

$$net_n = \sum_x w_{xz} y_{xz} \tag{3.10}$$

where w_{xz} is the weight parameter connecting input *x* to the hidden node *z* and y_{xz} represents the input value from node *x* for the *z*-th node in the hidden layer. These weights are learned by the model during training to capture relationships in the time series data.

After calculating the weighted sum of net_n, the NNAR model applies a nonlinear sigmoid function which squashes the output of the linear combination to be between 0 and 1 for processing the input before passing it to the next layer with the Equation (3.11):

$$s(z) = \frac{1}{1+e^{-z}} \tag{3.11}$$

If the seasonal component is also included, the notation for NNAR can be written as $NNAR(p,P,k)_{\mathrm{m}}$ where p is the number of lagged inputs utilized as predictors (which

is equivalent to the AR order in an ARIMA model), *P* is the number of seasonal lags, *k* represents the number of neurons in the hidden layer and *m* is the seasonal period.

A schematic illustration of the NNAR $(p, P, k)_{\mathrm{m}}$ structure can be seen in Figure 3.4.

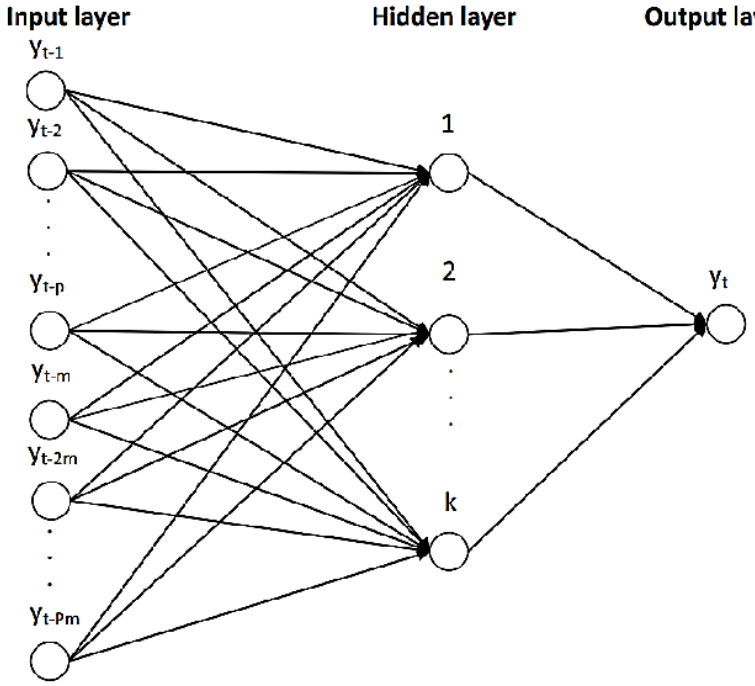

Figure 3.4: A schematic illustration of the NNAR $(\boldsymbol{p}, \boldsymbol{P}, \boldsymbol{k})_{\mathrm{m}}$ structure (Alonso et al., 2002).

A $NNAR\ (p, P, 0)_{\mathrm{m}}$ model is equivalent to an $ARIMA(p, 0{,}0)(P, 0{,}0)_{\mathrm{m}}$ model but without the restrictions on the parameters that ensure stationarity (Hyndman and Athanasopoulos, 2021).

3.2.2.5 TBATS (Trigonometric, Box-Cox, ARMA errors, Trend, and Seasonal components)

De Livera et al. (2011) suggested the TBATS technique, which uses the letters B, A, T, and S to stand for the Box-Cox transformation, the ARMA model, trend pattern, and seasonality in the object time series, respectively. The exponential smoothing state space model and a novel trigonometric representation of seasonal behaviors based on the Fourier series and the Box-Cox transform are combined in this automated approach to handle complicated, numerous, and overlapping seasonal patterns. In contrast to dynamic harmonic regression, it permits seasonality to gradually shift over time. (De Livera et al., 2011).

TBATS differentiates from many traditional forecasting techniques in terms of the ability to handle complex seasonal patterns, non-linear trends, and variable variance in time series data.

3.2.2.6 THETAM (Theta model)

Proposed by Fotios Petropoulos and Konstantinos Nikolopoulos (2000) as an improved version of the original theta model introduced by (Assimakopoulos and Nikolopoulos 2000) which forecasts based on θ lines constructed to capture different aspects of the data through simple exponential smoothing. The ThetaM model is a more sophisticated version of theta that takes into account seasonal effects.

3.2.2.7 Naïve

The Naïve model for forecasting time series assumes the prediction of the future value as the last directly observed data point(s) and can be defined as one of the simplest time series forecasting methods.

3.2.2.8 Simple moving average (SMA)

It is a simple estimation method that is widely used in the literature to identify trends in time series data and to make short-term forecasts, based on the calculation of the average value depending on the number of periods to be determined.

3.2.2.9 Random Forest

The Random Forest is an ensemble machine-learning method that was first put forth by Breiman (2001) and is applied to both regression and classification issues where a large number of decision trees are combined. The data must first be converted into a supervised learning format so that temporal dependencies (lags, seasonal effects, etc.) may be included before it can be used for time series forecasting. Once the transformed data has been trained, forecasting can be performed.

3.2.2.10 Support Vector Machine

It is a vector space-based machine learning method that tries to find the best hyperplane separating the two most distant classes in the training dataset and maximizes the margin between classes (Schölkopf and Smola, 2002). Finding a hyperplane with a big margin that offers higher generalization for the prediction of additional data points is

the goal of the margin, which is the distance between the hyperplane and the support vectors. Unlike traditional methods such as ARIMA, which assume linear relationships, SVR is successful in dealing with complex, nonlinear patterns often found in time series data[5].

3.2.2.11 XGBoost (Extreme Gradient Boosting)

The XGBoost model, an improved and optimized version of the gradient boosting algorithm, is a decision tree-based ensemble algorithm that incrementally builds a model from weak learners, typically decision trees (Chen and Guestrin, 2016). Similar to other machine learning algorithms for time series, it can make predictions by transforming time series data into a form of supervised learning with lagged features, rolling statistics, and date-based features.

3.4 Accuracy Performance Measurement Techniques

According to earlier research, performance evaluations can be carried out quantitatively using a variety of evaluation metrics that capture distinct facets of model performance, qualitatively by weighing the advantages and disadvantages of alternative strategies, or statistically through statistical tests (Jurman et al, 2012). In this study, our focus will be on the quantitative analysis of performance metrics as they are numerical and adaptable to the decision matrix. In Table 3.3, the definitions and formulas of the performance metrics used in this study are given.

Each error measure represents a single metric that highlights one aspect of forecasting performance, with each having its own advantages and limitations. Table 3.4 summarizes the strengths and constraints of the performance metrics used in the proposed model, with an extended version provided by Badulescu et al. (2021, p.7). This thesis aims to adopt a multi-dimensional and objective approach by incorporating all these metrics when determining the component forecast weights for forecast combinations.

[5] Retrieved from: https://www.geeksforgeeks.org/time-series-forecasting-with-support-vector-regression/

Table 3.3: List of performance metrics used for performance evaluation of the proposed forecasting method (Theodosiou, 2011, p.1186; Diebold, 2007; Hyndman and Athanasopoulos, 2018).

n is the number of forecasts, f_i is the forecasted value, a_i is the actual value;	
Forecast Errors	
Mean Error (ME): Represents the average of the differences between model predicted values and actual observed values.	$ME = \frac{1}{n}\sum_{i=1}^{\Pi}(f_i - a_i)$
Scale-Dependent Errors	
Mean Absolute Error (MAE): It provides an estimate of the size of the prediction errors but not the direction of them. It is the mean of the absolute differences between the model's forecasted values and the actual observed values.	$MAE = \frac{1}{n}\sum_{i=1}^{n}\lvert f_i - a_i\rvert$
Root Mean Squared Error (RMSE): It is calculated as the square root of the mean square of the squares of the variations between the observed and model-predicted values.	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(f_i - a_i)^2}$
Scaled Errors	
Mean Absolute Scaled Error (MASE): It correlates the prediction model's mean absolute error (MAE) to that of a basic naive model in order to scale it.	$MASE = \frac{\frac{1}{n}\sum_{i=1}^{n}\lvert f_i - a_i\rvert}{\frac{1}{n-1}\sum_{i=2}^{n}\lvert a_i - a_{i-1}\rvert}$
Percentage Errors	
Mean Percentage Error (MPE): It is the average of the percentage errors of the differences between the model predicted values and the observed actual values compared to the actual values. It gives an idea about the bias that may occur in forecasting models.	$MPE = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{f_i - a_i}{a_i}\right) \times 100$
Mean Absolute Percentage Error (MAPE): It is the mean of the absolute percentage errors between the actual values observed and the values predicted by the model. It gives a clear picture of the magnitude of errors are in comparison to actual values.	$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left\lvert\frac{f_i - a_i}{a_i}\right\rvert \times 100$
Correlation Metrics	
ACF1 Measurement: In the context of time series forecasting models, the ACF1 of the residuals is used to assess whether there is any autocorrelation remaining after the model is fitted. Ideally, a well-fitted model should produce approximately uncorrelated residuals. In this case, the ACF1 of the residuals should be close to zero, indicating that the model adequately captures all temporal patterns in the data.	$ACF1 = \frac{Cov(f_i - a_i, f_{i-1} - a_{i-1})}{Var(f_i - a_i)}$

Table 3.4: Pros and cons of the error measures.

Error measure	Pros	Cons
Mean Error (ME)	-Exhibits variance, range, and bias.	-Positive and negative errors can cancel out each other. -Scale-dependent. -Not robust to outliers.
Mean Absolute Error (MAE)	-Exhibits variance, range, and bias. -Positive and negative errors do not cancel out each other.	-A skewed mean is produced by outliers. -Scale-dependent. -Challenging to compare datasets with different units.
Mean Absolute Scaled Error (MASE)	-Scale-independent. -Enabling cross-dataset comparison. -Provides context by comparison against a naive model.	-Does not provide information about the direction of errors. -Needs a baseline forecast to calculate, which makes it challenging to comprehend without knowing the benchmark.
Mean Percentage Error (MPE)	-Percentage-based accuracy, independent of scale. -Permit cross-dataset comparisons -Relative size and direction of bias.	-Sensitive to large outliers. -Positive and negative errors can cancel out each other.
Mean Absolute Percentage Error (MAPE)	-Percentage-based accuracy, independent of scale. -Permit cross-dataset comparisons.	-Greater penalty for positive errors, making it asymmetrical.
ACF1	-Gives information on the forecast errors' independence. -Facilitates the identification of residual correlation, a sign of model misspecification.	-Does not indicate the magnitude of errors. -Without further context, it might not be possible to compare datasets directly.

3.5 Multi-Criteria Decision Making Approach

3.2.1 CRITIC as a weighting method

The Criteria Importance Through Intercriteria Correlation (CRITIC), which was proposed by Diakoulaki et al. (1995), takes into account the level of contrast and the conflict between criteria via the variation of each criterion, simultaneously. It determines the relative importance weights of the criteria based on the correlation between them and models the information obtained from decision-makers by taking into account the contrast density (Diakoulaki et al., 1995). So, the method assigns weights based on the amount of information a criterion conveys, which is measured by its standard deviation and the correlation between other criteria. One of the distinctive features of CRITIC compared to other weighting methods is that it can be

applicable even for dependent attributes and is an objective weighting method (Alinezhad and Khalili, 2019). With these features, it is possible to assess potentially dependent performance metrics that objectively measure forecasting models' accuracy.

The steps for applying the method can be summarized as follows (Diakoulaki et al., 1995; Alinezhad and Khalili, 2019);

Step 1: Building of the normalized decision matrix is handled. Because all of our criteria are in terms of error terms, they can be considered as negative attributes and their normalization can be done by the Equation (3.12);

$$x_{ij} = \frac{r_{ij} - r_i^+}{r_i^- - r_i^+} \quad i = 1, \dots, m \quad j = 1 \dots, n \tag{3.12}$$

where r_{ij} is the *j*th performance value of alternative *i* of the original matrix, x_{ij} is the element of the normalized decision matrix for ith alternative in jth attribute, $r_i^+ = max(r_1, r_2, \dots, r_m)$ and $r_i^- = min(r_1, r_2, \dots, r_m)$.

Step 2: Standard deviation of each attribute is measured as in Equation (3.13);

$$\sigma_j = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}\left(x_{ij} - \bar{x}_j\right)^2}; \quad i = 1, \dots m \tag{3.13}$$

where $\bar{x}_j$ is the mean of the *j*th attribute and *n* is the number of alternatives.

Step 3: The correlation coefficient between attributes is calculated by the Equation (3.14);

$$\rho_{jk} = \sum_{i=1}^{m}\left(x_{ij} - \bar{x}_j\right)\left(x_{ik} - \bar{x}_k\right) / \sqrt{\sum_{i=1}\left(x_{ij} - \bar{x}_j\right)^2 \sum_{i=1}^{m}\left(x_{ik} - \bar{x}_k\right)^2} \tag{3.14}$$

where $\bar{x}_k$ is the mean of the kth attribute

Step 4: The index (C) is estimated as in Equation (3.15);

$$C_j = \sigma_j \sum_{k=1}^{n}\left(1 - \rho_{jk}\right); \; j = 1 \dots, n \tag{3.15}$$

Step 5: Finally, criteria weights can be determined as shown in Equation (3.16);

$$w_j = \frac{C_j}{\sum_{j=1}^{n} C_j};\ j = 1 \dots, n \tag{3.16}$$

3.2.2 CILOS as a weigthing method

Building on a concept originally proposed by Mirkin et al. (1974), the CILOS (Criteria Importance Through Intercriteria Correlation) method was further developed and refined by Zavadskas and Podvezko (2016). It can be used as an objective weighting method similar to CRITIC and takes into account the loss of importance of the remaining criteria when a criterion reaches the ideal value. In the CRITIC method, a criterion's weight increases with its relative loss of effect, whereas in CILOS, a smaller relative loss corresponds to a greater weight. Therefore, testing the proposed method with both approaches and comparing the results allows us to observe the impact of the way evaluating accuracy measurements from different perspectives within the forecast combination. The calculation steps are given as follows (Zavadskas and Podvezko, 2016):

Step 1: The following equation is processed to convert the negative attributes of the decision matrix. Since all our criteria are expressed as error terms and thus represent negative attributes, it is necessary to apply the following formulation in Equation (3.17) to all criteria;

$$\hat{r}_{ij} = \frac{\min_{\mathrm{i}} r_{ij}}{r_{ij}};\ \mathrm{i}, j \in \{1, \dots, n\} \tag{3.17}$$

Step 2: Afterwards, the normalized values of the decision matrix are calculated by the Equation (3.18);

$$d_{ij} = \frac{r_{ij}}{\sum_{i=1}^{n} r_{ij}};\ j = 1, \dots, n \tag{3.18}$$

where $\bar{r}_{ij}$ represents the normalized values of the decision matrix of *i* th forecasting model in the *j* th accuracy measurement.

Step 3: At the next step, the square matrix is created according to the Equation (3.19);

$$c_{\mathrm{i}} = \max_{i} \bar{r}_{ij} = c_{k_i j};\ i,j \in \{1,\dots,n\} \tag{3.19}$$

where $c_{k_i j}$ is the maximum value of *j* th criteria, derived from the decision matrix using k_i rows to form a square matrix $c_{ij} = c_{k_i j}$ and $c_{jj} = c_j$. The *i* th row of square matrix contains the elements corresponding to the k_i th row of decision matrix.

Step 4: As the next step, based on the values obtained in the previous step, the elements of the relative impact loss matrix can be calculated using the Equation (3.20);

$$p_{ij} = \frac{c_{jj} - c_{ij}}{c_{jj}};\qquad i,j \in \{1,\dots,n\},\qquad p_{jj} = 0 \tag{3.20}$$

p_{ij} denotes the relative impact loss associated with the *j*-th attribute when this attribute considered the best value.

Step 5: By using p_{ij} values, the weight system matrix is built as seen in Equation (3.21);

$$\mathrm{F} = \begin{pmatrix} -\sum_{i=1}^{n} P_{i1} & P_{12} & \cdots & P_{1\mathrm{n}} \\ P_{21} & -\sum_{i=1}^{n} P_{i2} & \cdots & P_{2\mathrm{n}} \\ \vdots & \vdots & \ddots & \vdots \\ P_{\mathrm{n}1} & P_{\mathrm{n}2} & \cdots & -\sum_{i=1}^{n} P_{i\mathrm{n}} \end{pmatrix}_{n\times n} \tag{3.21}$$

Step 6: The weight of attributes [q_1, q_2, ..., q_n] can be assigned through the solution of the following Equation (3.22);

$$Fq^T = 0 \tag{3.22}$$

3.2.3 TOPSIS method

The Technique for Order Preference by Similarity to an Ideal Solution (TOPSIS) method chooses the option that is most similar to the positive ideal solution and most dissimilar from the negative ideal solution is the foundation of the strategy created by Hwang and Yoon (1981). The method's application phases are as follows (Hwang and Yoon, 1981; Gaur et al., 2023):

Step 1: The normalized decision matrix is constructed. The *dij* element of the normalized decision matrix for TOPSIS can be found as follows in Equation (3.23);

$$d_{ij} = \frac{r_{ij}}{\sum_{i=1}^{m} r_{ij}^2} ; i = 1, \dots, m \quad j = 1 \dots, n \tag{3.23}$$

Step 2: The normalized decision matrix is weighted by using the Equation (3.24);

$$b_{ij} = d_{ij} \cdot w_j ; i = 1, \dots, m \quad j = 1 \dots, n \tag{3.24}$$

Step 3: Identification of Positive Ideal and Negative Ideal solutions is evaluated in terms of cost, as the normalized performance values of forecasting models represent error measures. However, for error measures that measure bias and can take negative values, the "best value" is not the lowest value on the numerical scale, but the value closest to zero. Therefore, when selecting the best value of the mean error (ME), mean percentage error (MPE) and ACF1 criteria, the correction was made by taking the value closest to zero, not the smallest numerical value. Likewise, the "worst value" was taken as the value furthest from zero.

Let v_j^* be the positive ideal solution donating the "best value" in the column of *j*th criterion, and v_j^- be the negative ideal solution showing the "worst value" in the column of *j*th criterion. Positive Ideal (S_i*) and Negative Ideal (S_i^-) solutions can be formulated in Equation (3.25);

$$S_i* = \sqrt{\sum_j \left(v_{ij} - v_j^*\right)^2} \; and \; S_i^- = \sqrt{\sum_j \left(v_{ij} - v_j^-\right)^2} \tag{3.25}$$

Step 4: Relative closeness to the ideal solution is calculated by Equation (3.26);

$$C_{i^*} = \frac{S_i^-}{(S_i^* + S_i^-)} \tag{3.26}$$

Based on the coefficient C_{i^*}, which represents the ratio of the distance to the negative ideal solution in the total distance, it is inferred that the prediction model with the largest C_{i^*} value should have the most weight in the combination.

The TOPSIS method does not take into account the relationships between the criteria, nor does it provide an easy method for determining the criteria (Tao et al., 2012), however, CRITIC or CILOS does. So, the hybrid CRITIC-TOPSIS or CILOS-TOPSIS

technique will help to evaluate component forecasts, also considering interdependencies between performance metrics in different ways.

3.2.4 ARAS method

In the context of Multi-Criteria Decision-Making (MCDM), the Additive Ratio Assessment (ARAS) method, first presented by Zavadskas and Turskis (2010) has become an effective and simple-to-use tool. Based on the proportional assessment principle, ARAS evaluates each alternative's performance concerning an ideal alternative that is created by combining the best values of all criteria. Unlike TOPSIS, which balances closeness to ideal and worst solutions, ARAS solely considers the proportionate performance in comparison to the optimal solution. Because of its focus on proportionality, ARAS is both computationally efficient and simple to understand. The following are the phases of application for the method:

Step 1: Each criterion's optimal values are identified, and the components at the row of ideal values can be displayed as r_{oj}. Since all our criteria are expressed as error terms and thus represent negative attributes, the optimal values can be expressed as the following Equation (3.27);

$$r_{oj} = \min r_{ij};\ i = o, 1, \dots., m, j = 1, \dots., n \tag{3.27}$$

However, this formulation is not accepted for the criteria that can take negative values, such as ME or MPE. For these criteria, the “optimal value” is considered as the value closest to zero under the proposed model in this thesis.

Step 2: A normalized decision matrix is created by Equation (3.28);

$$d_{ij} = \frac{r_{ij}}{\sum_{i=0}^{m} r_{ij}};\ j = 1, \dots, n \tag{3.28}$$

Step 3: The weighted normalized decision matrix is derived from the weight of the criteria as can be formulated in Equation (3.29);

$$b_{ij} = d_{ij} \cdot w_j\ ; i = o, 1, \dots, m\ \ j = 1 \dots, n \tag{3.29}$$

Step 4: The optimality function (S_i) can be formulated as shown in Equation (3.30) and defined as the value that is thought to be better the larger it is;

$$S_i = \sum_{j=1}^{n} b_{ij}\ ;\ i = o, 1 \ldots, m \tag{3.30}$$

Step 5: The utility degree (K_i) for every prediction model is determined in order to assess the final rankings for each forecast model and formulated with Equation (3.31);

$$K_i = \frac{S_i}{V_o}\ ;\ i = o, 1 \ldots, m \tag{3.31}$$

where V_o is the optimality value of S_i.

4. MODEL IMPLEMENTATION AND RESULTS

The success of the proposed approach is tested on two different datasets. While these datasets are selected, it is considered for both to have different patterns, sizes, and frequencies. While the first dataset includes more irregular patterns, exhibits high-frequency noise and volatility as well as being much larger in size, the second one is more regular, has clearer seasonal patterns and linear trends, and is much smaller in size. The proposed hybrid-MCDM method-based time series forecast combination models are applied. The results are presented and compared with individual and existing combination models in the literature. R programming was utilized to execute all models, including both individual and combined approaches, and to generate results across all subsets of the datasets. The most frequently used R libraries during this implementation phase were `forecast`, `ForecastComb`, `ggplot2`, `zoo` and `Metrics`. Preliminary statistical inspections were also conducted in R, while additional analyses were performed using EViews and Gretl. The results from the validation set, obtained through R, were used to construct a decision matrix. This decision matrix was then employed to apply the steps of the multi-criteria decision-making (MCDM) approach using Excel.

4.1 The Subset of D2035 Dataset from Open Access File of M4 Competition

4.2.1 Data pre-processing

The first dataset is a daily dataset from the industry domain, taken from the M4 competition file, which is shared as open data on GitHub[6]. The properties of the dataset used are in Table 4.1 as follows:

Table 4.1: D2035 dataset.

M4id	Category	Frequency	Horizon	SP	Starting Date
D2035	Industry	1	14	Daily	30-03-01 12:00

[6] https://github.com/Mcompetitions/M4-methods/blob/master/Dataset/M4-info.csv

The selected subset of the D2035 dataset covers 4197 observations between 30/03/2001 and 24/09/2012 on a daily basis. Figure 4.1 shows the graph of the mentioned dataset. It can be seen that this dataset includes irregular patterns, exhibits high-frequency noise and volatility as well as is large in size. Also, the seasonal patterns or linear trends are obscure.

Figure 4.1: The graph of the D2035 dataset from the open access file of M4 competition between 30/03/2001 and 24/09/2012.

To be able to apply a statistically significant forecasting model with the relevant dataset, a number of preliminary preparations and tests were carried out to prepare the data for statistical application. First, the Box-Cox transformation is tested to see whether we need any transformation for the response variable (Box and Cox, 1964) and was shown in Figure 4.2.

```
> lambda=BoxCox.lambda(m4d);lambda
[1] 0.1094047
```

Figure 4.2: Box-Cox transformation result from R.

Because λ is close to zero, it can be concluded that the log-transform of the dataset can be applied. So, the stationarity tests are handled on a log-transformed D2035 dataset in order to question the existence of a stochastic trend or deterministic trend. For the ADF test, the Akaike-info criterion was taken into consideration when considering sample size. The optimum maximum lag length is determined as 30 based on Schwert's Rule at R by the urca library (see Appendix A). ADF test settings from Eviews for the log-transformed D2035 dataset is in Figıre 4.3.

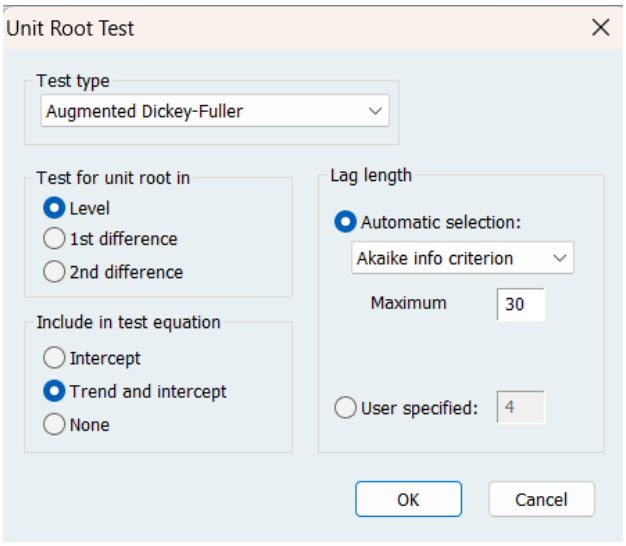

Figure 4.3: ADF test settings for the log-transformed D2035 dataset.

The results are as follows in Table 4.2. See Appendix A to see full table outputs taken from E-Views for these tests.

Table 4.2: Stationarity test results for the log-transformed data.

Unit-Root Test	**Test Equation**	**Critical value percentages – *t*-statistics**	**Result**
ADF test	With the equation with trend and intercept notation, both notations are significant with a *p*-value of 0.033 and 0.049.	at the level of 1% (-3.96) at the level of 5% (-3.41) at the level of 10% (-3.12)	*t*-statistics: -2.08 *p*-value: 0.5571
KPSS test	With the equation with trend and intercept notation, both notations are significant with a *p*-value of 0.000.	at the level of %1 (0.22) at the level of %5 (0.15) at the level of %10 (0.12)	LM-statistics: 0.69

When we start the ADF test analysis of our dataset with the equation with trend and intercept notation, we can say that the notations are significant, so we are on the right equation. When we test for the presence of a unit root through this equation, we can say that we can't reject the presence of a stochastic trend according to the result of the test. The results for the KPSS test are also consistent with the ADF test. In light of these results, we infer that there is a stochastic trend in our dataset and it should be taken the difference of the log-transformed series to reach the mean stationarity. See the graph of the log-differenced series in Figure 4.4.

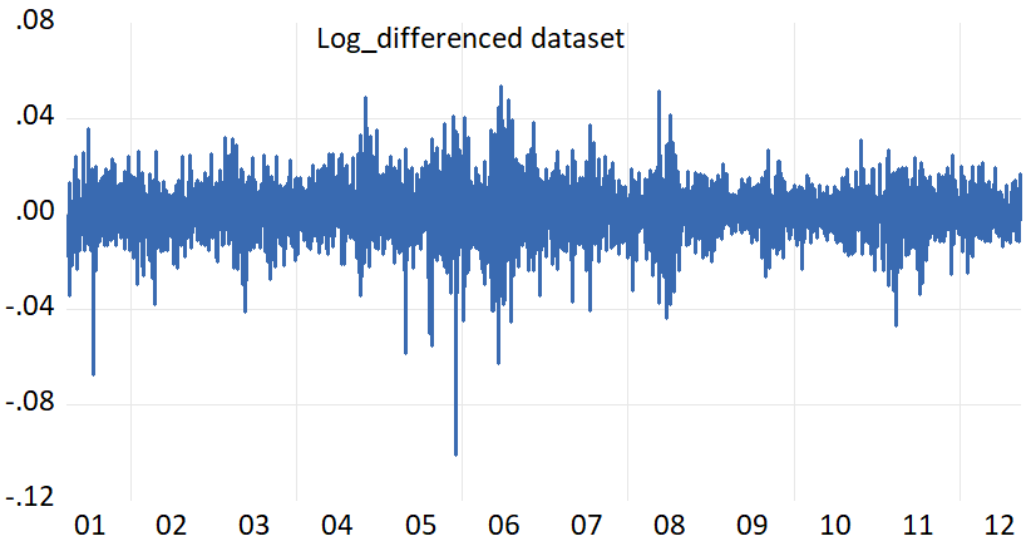

Figure 4.4: Graph of log-differenced series.

We apply stationarity tests again but this time on log-differenced datasets to be sure that finally, the series is stationary after all the transformations (see Table 4.3).

Table 4.3: Stationarity test results for the log-differenced dataset.

Unit-Root Test	**Test Equation**	**Critical value percentages – *t*-statistics**	**Result**
ADF test	The equation with no trend and no intercept notation.	at the level of 1% (-2.57) at the level of 5% (-1.94) at the level of 10% (-1.61)	*t*-statistics: -30.83 *p*-value: 0.0000
KPSS test	The equation with no trend and no intercept notation.	at the level of %1 (0.74) at the level of %5 (0.46) at the level of %10 (0.35)	LM-statistics: 0.08

We reject the null hypothesis that the log-differenced dataset has a unit root by the ADF test. The KPSS test also yields results consistent with the ADF test. See Appendix A to see full table outputs taken from E-Views for these tests.

For the next step, it is required to check whether the stationary series is white noise. For this, see the correlogram and autocorrelation function results of the log-differenced D2035 dataset taken from Gretl in Figure 4.5 and Figure 4.6.

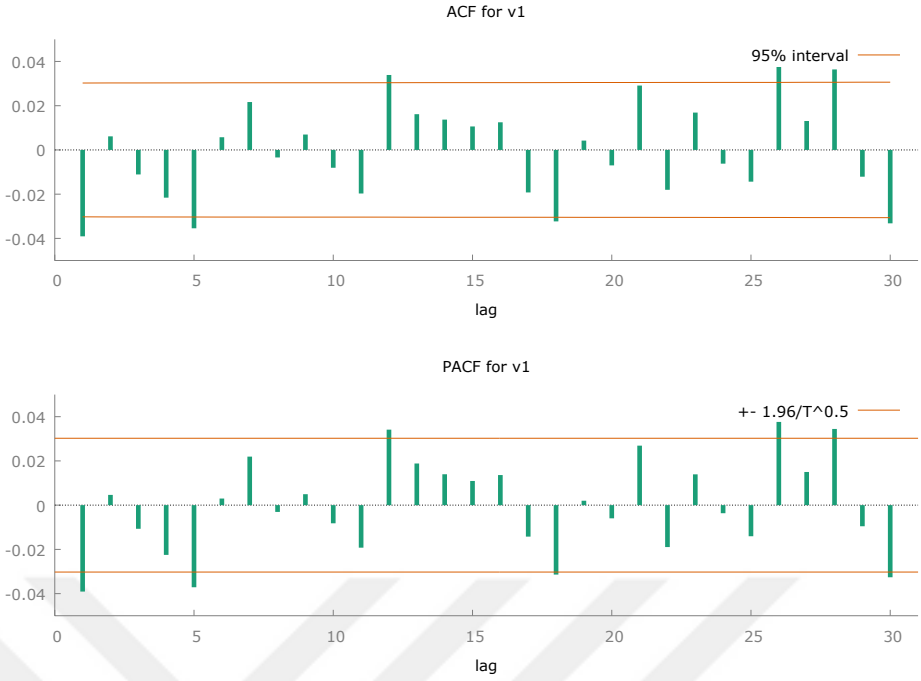

Figure 4.5: Correlogram of log-differenced dataset.

gretl: correlogram

Autocorrelation function for v1
***, **, * indicate significance at the 1%, 5%, 10% levels
using Bartlett standard errors for ACF

LAG	ACF		PACF		Q-stat.	[p-value]
1	-0.0391	**	-0.0391	**	6.4111	[0.011]
2	0.0061		0.0046		6.5697	[0.037]
3	-0.0111		-0.0107		7.0864	[0.069]
4	-0.0216		-0.0225		9.0435	[0.060]
5	-0.0354	**	-0.0371	**	14.3231	[0.014]
6	0.0057		0.0030		14.4615	[0.025]
7	0.0217		0.0219		16.4337	[0.021]
8	-0.0034		-0.0031		16.4825	[0.036]
9	0.0070		0.0049		16.6861	[0.054]
10	-0.0081		-0.0082		16.9594	[0.075]
11	-0.0197		-0.0192		18.5924	[0.069]
12	0.0339	**	0.0341	**	23.4303	[0.024]
13	0.0162		0.0188		24.5311	[0.027]
14	0.0137		0.0140		25.3220	[0.032]
15	0.0106		0.0109		25.7961	[0.040]
16	0.0125		0.0136		26.4547	[0.048]
17	-0.0192		-0.0142		28.0133	[0.045]
18	-0.0323	**	-0.0314	**	32.4148	[0.020]
19	0.0042		0.0020		32.4898	[0.028]
20	-0.0070		-0.0060		32.6955	[0.036]
21	0.0291	*	0.0269	*	36.2712	[0.020]
22	-0.0180		-0.0189		37.6427	[0.020]
23	0.0169		0.0139		38.8483	[0.021]
24	-0.0062		-0.0036		39.0107	[0.027]
25	-0.0143		-0.0141		39.8792	[0.030]
26	0.0375	**	0.0376	**	45.8214	[0.010]
27	0.0131		0.0150		46.5454	[0.011]
28	0.0364	**	0.0344	**	52.1468	[0.004]
29	-0.0121		-0.0096		52.7663	[0.004]
30	-0.0332	**	-0.0326	**	57.4208	[0.002]

Figure 4.6: Autocorrelation function results of log-differenced dataset.

Significant bars observed in the correlogram indicate that the data exhibits non-white noise characteristics, suggesting the presence of serial correlation rather than pure randomness. This finding implies that the data is stationary but retains autocorrelated which is suitable for the implementation of the proposed combination model.

For the next step, training-validation and test sets are determined:

- Training set: 30/03/2001-31/12/2009 (3199 observations)
- Validation set: 01/01/2010-31/12/2011 (730 observations)
- Test set: 01/01/2012-24/09/2012 (268 observations)

In the next stage, the dataset, which is divided into three parts as training, validation, and test set, is estimated with the traditional and machine learning forecasting algorithms mentioned in the methodology section. Individual forecast models were trained on a train set and forecasted on validation and test set horizons. All the results are noted, and the graphs for each predictor over the test horizon are inspected visually. Auto.arima, TBATS, THETAM, ETS, and Naïve yielded a forecast equivalent to the mean of the test horizon, suggesting that the model relies on an average-based approach for its predictions rather than capturing nuanced patterns within the data. Because these predictors are out of our scope, the decision matrix is created by the rest of the forecast models (NNAR, SMA, SVM, STL, XGBoost, RandomForest, Box-Jenkins).

A graph presenting the observed values for the validation and test sets, alongside the forecasted values generated by each individual forecasting model to be used in the decision matrix, is below (Figure 4.7). The graph for Box-Jenkins is given separately in Figure 4.8, as this model was developed using EViews rather than RStudio.

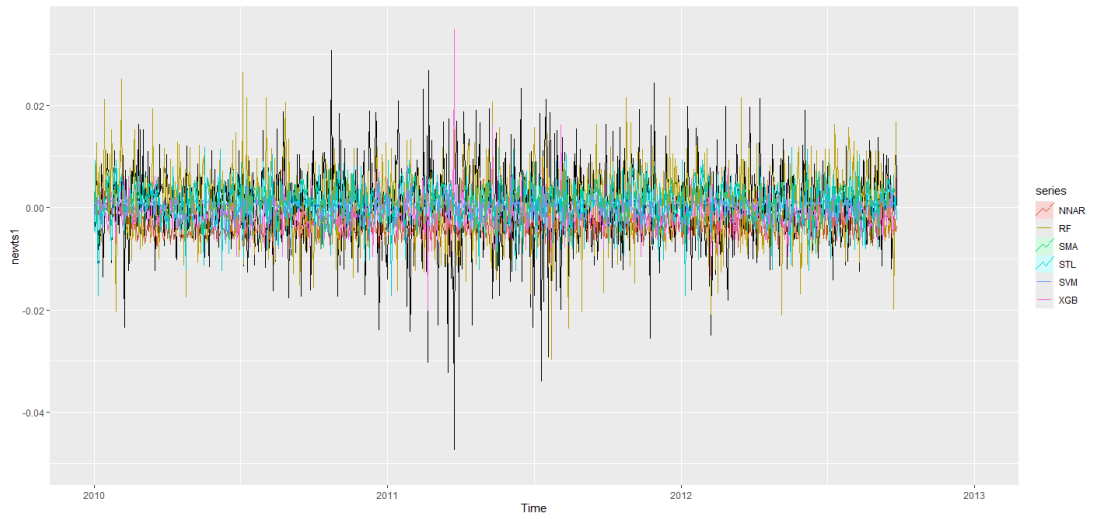

Figure 4.7: Actual and forecasted values on validation and test sets.

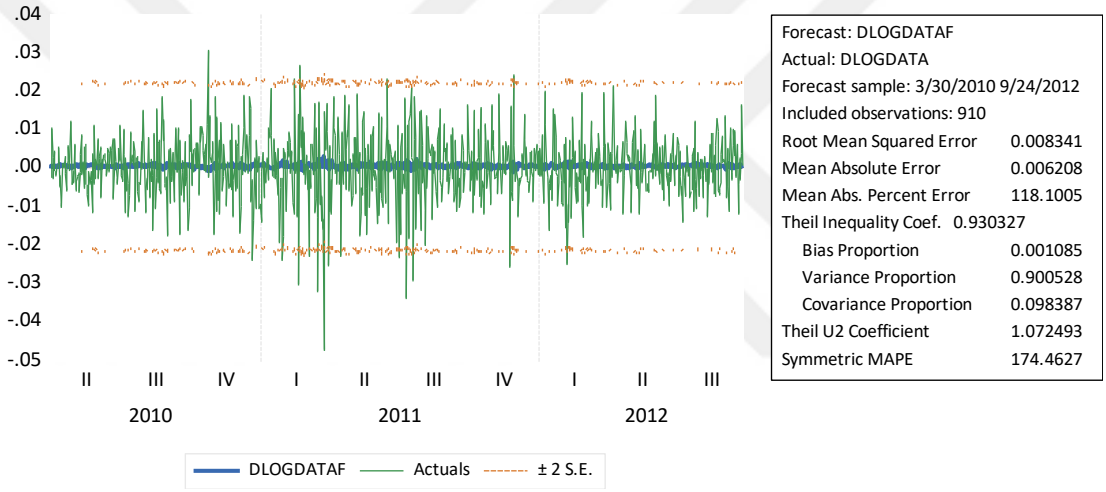

Figure 4.8: Box-Jenkins (ar(1) ar(5) model forecasts) model from Eviews.

The decision matrix is constructed by using the performance results of each predictor on the validation set (see Table 4.4).

Table 4.4: Decision matrix created over the validation set.

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
NNAR	0.0038	0.0097	0.0075	68.624	243.9467	3.8670	-0.0109
SMA	-0.0018	0.0092	0.0069	220.288	311.4079	4.1838	0.0221
SVM	-0.0004	0.0090	0.0069	133.007	152.5781	8.2188	-0.0280
STL	-0.0002	0.0096	0.0073	176.265	345.9326	1.4423	-0.0318
XGB	0.0014	0.0103	0.0078	175.606	222.9316	2.4097	0.0153
RF	-0.0005	0.0113	0.0087	128.276	401.9167	1.1597	0.0569
BOX	-0.0002	0.0087	0.0065	111.757	118.4925	5226.32	-0.0334

Figure 4.9 shows the correlation plot of the performance metrics on these decision matrix created over the validation set by using the library of *corrplot* from RStudio.

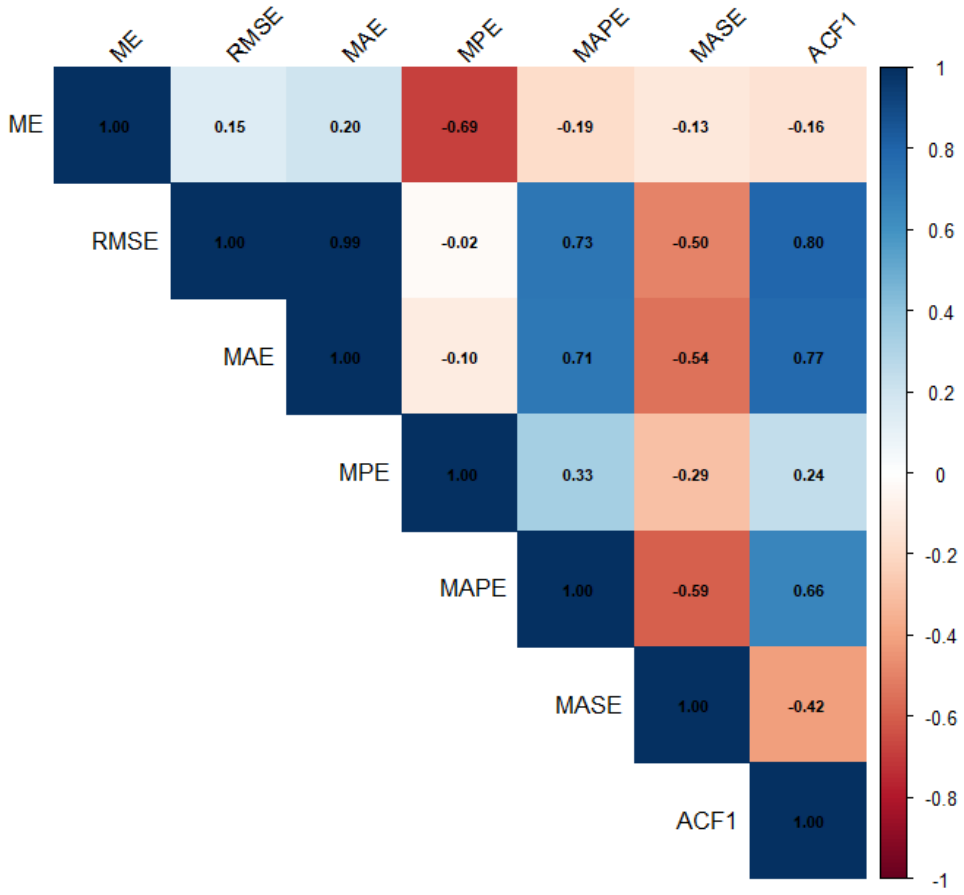

Figure 4.9: The correlation plot between performance metrics for D2035 dataset.

The correlation plot showS that ME and MPE seem to be less correlated with other performance metrics compared to evaluating of other metrics individually. RMSE and MAE are resulted as highly positively correlated. The remaining performance metrics have, to varying degrees, approached of both positive and negative correlations with each other and with other performance metrics. In the measurement where values close to 1 indicate a positive correlation and those close to -1 indicate a negative correlation, all results are shown in Figure 4.9. Considering these findings, it can be seen that the performance metrics do not give completely independent results from one another. When assigning combination weights, it is crucial to take into account the correlations between these metrics in order to give greater weight to metrics that have informational value. This in turn suggests the potential need to use objective weighting methods that take into account interdependencies.

4.2.2 Implementation of CRITIC-TOPSIS hybrid MCDM technique

The CRITIC technique was used to establish the weights of the performance metrics. All tables for each step at below. First, the normalized decision matrix for CRITIC and standard deviations for each column are found in Table 4.5.

Table 4.5: Normalized decision matrix for CRITIC with σ_j values.

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
NNAR	0.000	0.619	0.540	1.000	0.557	0.999	0.750
SMA	1.000	0.807	0.806	0.000	0.319	0.999	0.385
SVM	0.758	0.886	0.800	0.575	0.880	0.999	0.939
STL	0.729	0.635	0.628	0.290	0.198	1.000	0.981
XGB	0.441	0.397	0.428	0.295	0.632	1.000	0.460
RF	0.769	0.000	0.000	0.607	0.000	1.000	0.000
BOX	0.725	1.000	1.000	0.716	1.000	0.000	1.000
σ_j	***0.322***	***0.338***	***0.326***	***0.329***	***0.362***	***0.378***	***0.377***

Intercriteria correlations matrix (Table 4.6) and the matrix constructed with 1-ρ_{jk} values, and index C_j (Table 4.7) are determined.

Table 4.6: Intercriteria correlations matrix.

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
ME	1.000	0.148	0.204	-0.686	-0.188	-0.127	-0.155
RMSE	0.148	1.000	0.992	-0.023	0.726	-0.496	0.799
MAE	0.204	0.992	1.000	-0.103	0.712	-0.542	0.772
MPE	-0.686	-0.023	-0.103	1.000	0.334	-0.292	0.242
MAPE	-0.188	0.726	0.712	0.334	1.000	-0.594	0.657
MASE	-0.127	-0.496	-0.542	-0.292	-0.594	1.000	-0.415
ACF1	-0.155	0.799	0.772	0.242	0.657	-0.415	1.000

Table 4.7: Matrix of 1-ρ_{jk} values and index C_j.

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
ME	0.000	0.852	0.796	1.686	1.188	1.127	1.155
RMSE	0.852	0.000	0.008	1.023	0.274	1.496	0.201
MAE	0.796	0.008	0.000	1.103	0.288	1.542	0.228
MPE	1.686	1.023	1.103	0.000	0.666	1.292	0.758
MAPE	1.188	0.274	0.288	0.666	0.000	1.594	0.343
MASE	1.127	1.496	1.542	1.292	1.594	0.000	1.415
ACF1	1.155	0.201	0.228	0.758	0.343	1.415	0.000
C_j	***2.194***	***1.302***	***1.291***	***2.148***	***1.578***	***3.199***	***1.548***

Finally, weights of performance criteria according to CRITIC method is obtained in Table 4.8.

Table 4.8: Weights of performance criteria according to CRITIC method.

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
16.5%	9.8%	9.7%	16.2%	11.9%	24.1%	11.7%

After determining the criteria weights, the decision matrix is ready for the application of the TOPSIS method.

For starting to implementation of the TOPSIS method, vector normalization is applied to the matrix in Table 4.4 as a first step, and Table 4.9 is constructed.

Table 4.9: Vector normalized decision matrix.

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
NNAR	0.8529	0.3768	0.3831	0.1706	0.3368	0.0007	-0.1306
SMA	-0.3943	0.3577	0.3534	0.5477	0.4299	0.0008	0.2641
SVM	-0.0924	0.3497	0.3542	0.3307	0.2106	0.0016	-0.3347
STL	-0.0557	0.3752	0.3734	0.4382	0.4776	0.0003	-0.3800
XGB	0.3026	0.3993	0.3957	0.4366	0.3078	0.0005	0.1830
RF	-0.1066	0.4396	0.4436	0.3189	0.5549	0.0002	0.6805
BOX	-0.0509	0.3381	0.3318	0.2779	0.1636	1.0000	-0.4003

By multiplying criteria weights obtained from the CRITIC method with elements of vector normalized decision matrix, a weighted normalized decision matrix can be built (Table 4.10).

Table 4.10: Weighted normalized decision matrix.

Weights:	**16.5%**	**9.8%**	**9.7%**	**16.2%**	**11.9%**	**24.1%**	**11.7%**
	ME	**RMSE**	**MAE**	**MPE**	**MAPE**	**MASE**	**ACF1**
NNAR	0.14111	0.03699	0.03731	0.02764	0.04007	0.00018	-0.01524
SMA	-0.06524	0.03512	0.03442	0.08874	0.05115	0.00019	0.03083
SVM	-0.01528	0.03433	0.03449	0.05358	0.02506	0.00038	-0.03907
STL	-0.00922	0.03684	0.03636	0.07101	0.05682	0.00007	-0.04436
XGB	0.05006	0.03921	0.03853	0.07074	0.03662	0.00011	0.02136
RF	-0.01763	0.04316	0.04319	0.05167	0.06601	0.00005	0.07944
BOX	-0.00842	0.03320	0.03231	0.04502	0.01946	0.24125	-0.04672

Positive and negative ideal solutions can be obtained as shown in Table 4.11. For ME and ACF1, the value closest to zero is taken as the positive ideal solution and the value furthest from zero is taken as the negative ideal solution. For the other criteria, the minimum value represents the positive ideal and the maximum value represents the negative ideal.

Table 4.11: Positive and negative ideal solutions for each criterion.

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
P/S	-0.00842	0.03320	0.03231	0.027644	0.01946	0.00005	-0.01524
N/S	0.14111	0.04316	0.04319	0.088740	0.06601	0.24125	0.07944

Table 4.12 shows the deviation rates (S_i^*), of each model from the positive ideal solution values, deviation rates (S_i^-) from the negative ideal solution values, relative

closeness values ($\mathbf{C}_{i^*}$) to the ideal solution, and the weights of the prediction models obtained by normalizing ($\mathbf{C}_{i^*}$) values within the combination for this series.

Table 4.12: TOPSIS method results under CRITIC-weighted criteria weights (with normalized individual forecasting model weights).

	S_i*	S_i^-	$\mathbf{C}_{i^*}$	**The weights of the component models in the combination**
NNAR	0.1511	0.2675	0.639	**12.6%**
SMA	0.1005	0.3216	0.762	**15.0%**
SVM	0.0364	0.3156	0.897	**17.7%**
STL	0.0645	0.3108	0.828	**16.3%**
XGB	0.0836	0.2665	0.761	**15.0%**
RF	0.1096	0.2911	0.726	**14.3%**
BOX	0.2439	0.2063	0.458	**9.0%**
		∑	5.072	100.0%

Figure 4.10 shows the observed values for the validation and test sets, alongside the forecasted values generated by the proposed combination forecast.

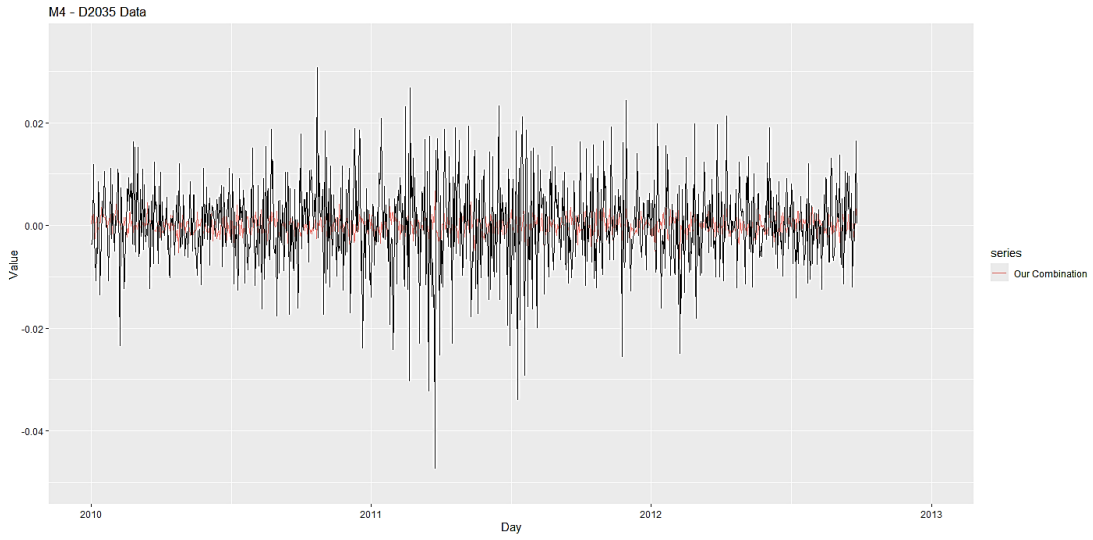

Figure 4.10: The forecasted values generated by the proposed combination forecast, alongside the observed values for the validation and test sets.

4.2.3 Interpretation of findings

The new combination model with the weights obtained after the hybrid MCDM model was applied to the validation set was compared with the performance of the single models and other combination models on the test set and the following results were obtained (Table 4.13).

The results show that the proposed combination model performs better in terms of mean error compared to all individual models and the other combination models in the literature. In terms of root mean square error and mean absolute error, the results are close to each other in general. However, it can be claimed that the proposed model did a great job compared to most of the combination models given in the table and individual models. Although certain models achieve notably lower mean percentage error, mean absolute percentage error, and mean absolute scale error compared to the proposed model, the proposed model still outperforms several others, particularly existing combination models, according to these accuracy metrics. Regarding the ACF1 metric, the proposed model yielded higher values compared to its counterparts, indicating a higher level of residual correlation, which is generally undesirable.

Table 4.13: Results of individual forecasting models, combination models, and the proposed CRITIC-TOPSIS combination method on the test set for D2035 dataset.

	Performance Result / Forecasting Model	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Single Models	**STL**	-0.00042	0.00833	0.00677	-233.062	674.129	1.37127	0.03495
	ETS	-0.00041	0.00703	0.00539	108.759	142.725	Inf	0.02918
	Auto.ARIMA	-0.00041	0.00703	0.00539	108.738	142.603	Inf	0.02918
	NNAR	0.00332	0.00801	0.00622	87.727	550.626	2.71329	0.00434
	TBATS	-0.00028	0.00702	0.00537	106.370	129.162	Inf	0.02918
	THETAM	-0.00064	0.00704	0.00542	113.485	170.035	82822.65	0.02911
	NAÏVE	0.00388	0.00801	0.00621	26.129	574.060	Inf	0.02918
	SMA(n=3)	-0.00194	0.00779	0.00610	16.981	513.241	3.64383	0.13201
	RF	-0.00027	0.00953	0.00771	413.325	1102.966	1.00477	0.04127
	SVM	-0.00074	0.00707	0.00544	119.668	213.491	7.72092	0.08122
	XGBoost	0.00127	0.00758	0.00576	-17.304	284.493	1.99789	0.08841
	Box-Jenkins	-0.00041	0.00703	0.00539	108.767	142.774	Inf	0.02918
Proposed Model		**-0.00013**	**0.00715**	**0.00551**	**55.59425**	**287.5708**	**3.417**	**0.07140**
Other Comb. Models	**SA**	-0.00834	0.01094	0.00936	251.2177	1346.090	6.03513	0.05047
	OLS	-0.00077	0.00976	0.00789	332.8623	1118.659	0.98305	0.03545
	MED	-0.00031	0.00708	0.00547	60.17042	202.5419	5.22389	0.05478
	TA	-0.00834	0.01094	0.00936	251.2177	1346.090	6.03513	0.05047
	WA	-0.00834	0.01094	0.00936	251.2177	1346.090	6.03513	0.05047
	CLS	0.00022	0.00841	0.00672	278.2672	778.8293	1.18969	0.05384
	LAD	0.00057	0.00976	0.00786	301.1401	1033.509	0.97633	0.03186

Note: Points that perform worse than our combination model are painted with pink filler.

The median-based approach for time series forecast combination yields the best performance among the combination methods for this dataset; however, our proposed combination model achieves comparably close and promising results.

4.2 USA Deliveries of Natural Gas To Electric Power Users (Mmcf) Dataset

4.2.1 Data pre-processing

The effectiveness of the proposed approach is further evaluated using a dataset with distinct characteristics and a different underlying pattern than the initial dataset. The deliveries of natural gas to electric power users (MMcf) dataset from the U.S. Energy Information Administration (EIA)[7] is used for this time. The series covers 281 data points between January 2001 and May 2024 and appears as a monthly time series. Figure 4.11 shows the graph of the relevant dataset. It includes more regular patterns, exhibits lower-frequency noise, and volatility, as well as it is smaller in size compared to the first dataset. The reflections of trend and seasonal effects can be observed from the graph apparently.

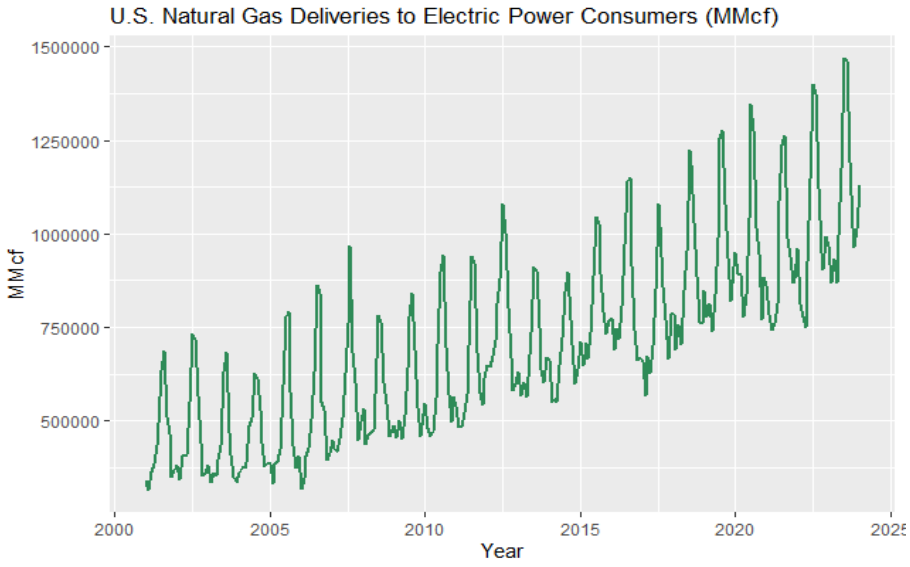

Figure 4.11: Graph of the natural gas delivery to USA electric power consumers (MMcf) dataset.

As previous dataset, a number of preliminary preparations and tests were carried out to prepare the data for statistical application. To question the existence of a stochastic trend or deterministic trend, unit-root tests are performed. For the ADF test, this time Hannan-Quinn criterion is taken into consideration when considering sample size (see Figure 4.12). The optimum maximum lag length is determined as 15 based on Schwert's Rule at R by the urca library (see Appendix B).

[7] Retrieved from: https://www.eia.gov/

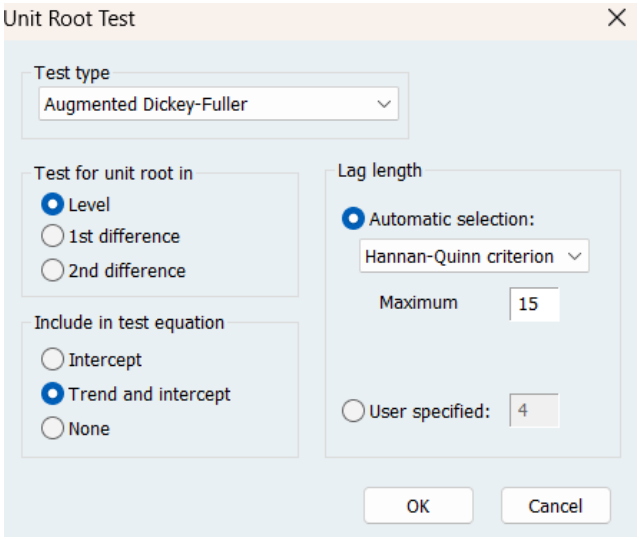

Figure 4.12: ADF test settings for the raw U.S. deliveries of natural gas to electric power users (MMcf) dataset.

The summary of the stationarity test results is given in Table 4.14. For the whole output see Appendix B.

Table 4.14: Stationarity test results for the raw data.

Unit-Root Test	**Test Equation**	**Critical value percentages – *t*-statistics**	**Result**
ADF test	With the equation with trend and intercept notation, both notations are significant with a *p*-value of 0.000 and 0.000.	at the level of 1% (-3.99) at the level of 5% (-3.42) at the level of 10% (-3.13)	*t*-statistics: -4.32 *p*-value: 0.0033
KPSS test	With the equation with trend and intercept notation, both notations are significant with a *p*-value of 0.000.	at the level of %1 (0.22) at the level of %5 (0.15) at the level of %10 (0.12)	LM-statistics: 0.10

Initiating the ADF test analysis on our dataset with the equation incorporating both trend and intercept terms, we find that these terms are statistically significant. Under this equation, the test results allow us to reject the presence of a stochastic trend, indicating no unit root in the series. Similar to the ADF test, the result of the KPSS test, which is used to test whether a univariate time series has a unit root, is consistent with the ADF test. Although the results of the tests suggest that there is no stochastic trend, the graphical analysis clearly reveals the presence of a trend. In light of these results, we infer that there is a deterministic trend in our dataset and we can detrendize it. The graph of detrended series was shown in Figure 4.13.

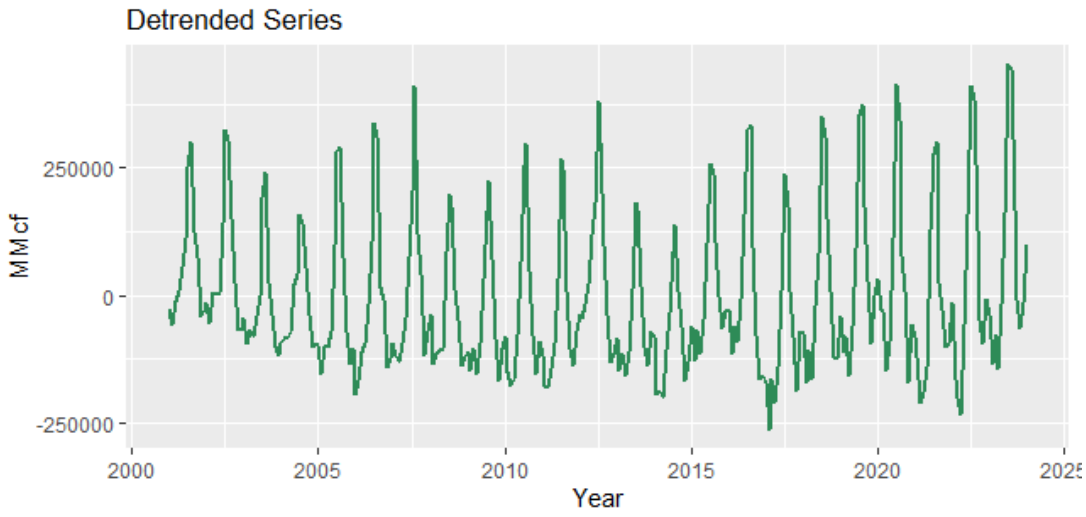

Figure 4.13: Detrended U.S. deliveries of natural gas to electric power users (MMcf) dataset.

When the same stationarity tests are applied to the detrended series the results point that the transformed series are strictly stationary (see Table 4.15). See Appendix B to see full table outputs taken from E-Views for these tests.

Table 4.15: The stationarity test results for the detrended dataset.

Unit-Root Test	**Test Equation**	**Critical value percentages – *t*-statistics**	**Result**
ADF test	The equation with no trend and no intercept notation.	at the level of 1% (-2.57) at the level of 5% (-1.94) at the level of 10% (-1.62)	*t*-statistics: -4.28 *p*-value: 0.0000
KPSS test	The equation with no trend notation.	at the level of %1 (0.74) at the level of %5 (0.46) at the level of %10 (0.35)	LM-statistics: 0.10

The correlogram of the detrended dataset with Bartlett confidence intervals can be seen in Figure 4.14.

Significant bars observed in the correlogram indicate that the data exhibits non-white noise characteristics, suggesting the presence of serial correlation rather than pure randomness. This finding implies that the data is stationary but remains autocorrelated. Because it is not white noise, we can go on to apply the proposed model.

Training validation and test sets were determined for the next steps:

- Training set: 01/2001-12/2017 (204 observations)
- Validation set: 01/2018-06/2022 (54 observations)

- Test set: 07/2022-05/2024 (23 observations)

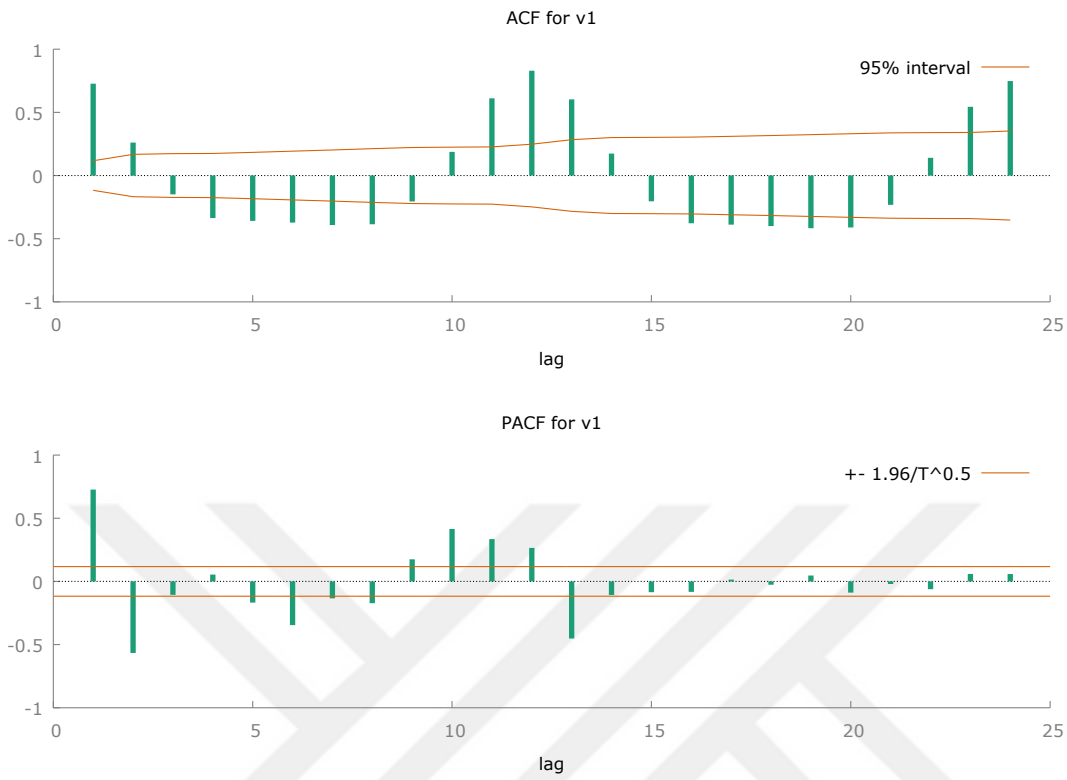

Figure 4.14: Correlogram of detrended series under Bartlett confidence intervals.

Individual forecast models were trained on the train set and forecasted on validation and test set horizons. ETS resulted in the ETS(A,N,A) model. Also, auto.arima resulted as ARIMA(1,0,0)(0,1,1)[12] (For the output details of both models see the Appendix B). They both gave so similar and compatible results with each other. Moreover, TBATS and STL gave nearly the same results as each other too. So, including one of them in the decision matrix will be wiser for differentiation of predictors, so the quality of the combination. For the Box-Jenkins model, suitable results were not found to model this dataset and naïve gave underestimated results based on its visual inspection. That is why they are also not included in the decision matrix. All the results are noted, and the graphs over the test horizon for each predictor used in the decision matrix are inspected visually. Figure 4.15 presents a graph of the observed values for the validation and test sets, alongside the forecasted values generated by each individual forecasting model to be used in the decision matrix.

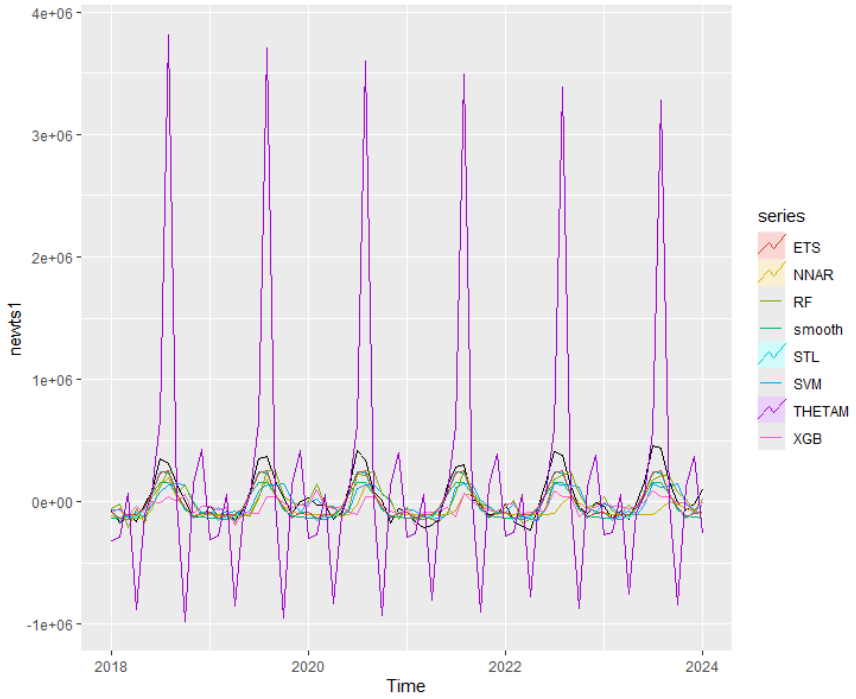

Figure 4.15: Actual and forecasted values over validation and test sets.

Based on the prediction methods described earlier and the performance results of each method on the validation set, the following decision matrix is constructed (Table 4.16).

Table 4.16: Decision matrix created over the validation set.

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
STL	30564.5334	68454.7	55121.34	-35.5175	155.6079	0.704979	0.516842
ETS	23342.6479	69064.08	56298.13	-45.2766	152.8521	0.809448	0.505613
NNAR	65322.45	137444.7	99896.25	-31.8329	186.327	2.494934	0.541119
THETAM	-157886.1	983127.3	523706.8	601.5116	1140.552	0.537545	0.001429
SMA(n=3)	50676.27	121633.3	93221.74	-120.263	214.8115	1.947799	0.49149
RF	17520.61	118274.3	93115.1	-46.8808	157.5301	1.013367	0.283048
SVM	22898.52	125190	100549.3	-95.307	200.4804	1.748163	0.462011
XGBoost	52743.96	165314.1	114099.7	27.28779	123.7993	2.129218	0.562711

Figure 4.16 shows the correlation plot of the performance metrics on this decision matrix generated over the validation set using RStudio's corrplot library. The results show that ME & ACF1, MPE & MAPE, MAE & MPE, MAE & MAPE, RMSE & MAE, RMSE & MAPE, RMSE & MPE duos are almost perfectly positively correlated. ME and ACF1 individually are mostly negatively correlated with other performance measures.

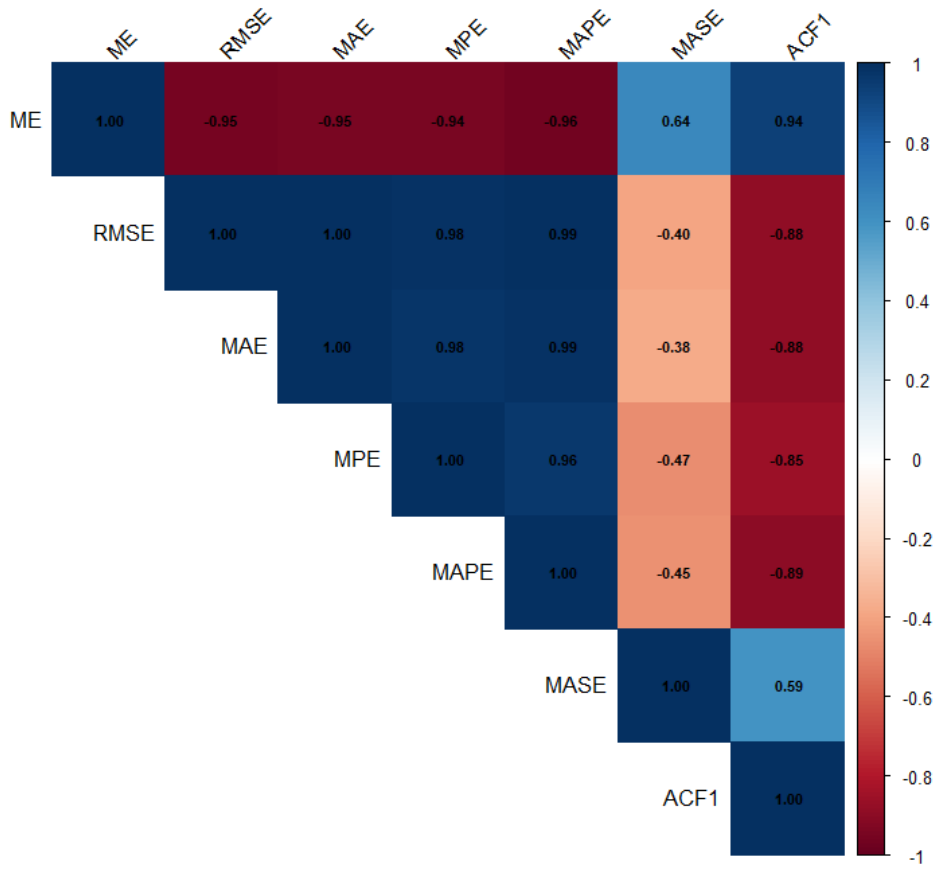

Figure 4.16: The correlation plot between performance metrics for EIA's dataset.

With varying degrees of correlation, all results are shown in Figure 4.16. Considering these results, it was observed that the performance metrics gave highly dependent results, especially compared to the D2035 dataset. Therefore, the potential need for the use of objective weighting methods that take into account interdependencies seems to be necessary, as was the case with the D2035 dataset.

4.2.2 Application of CRITIC-TOPSIS hybrid MCDM tecnique

All the steps for CRITIC tecnique presented at methodology section is applied on matrix in Table 4.16. The final weights of the performance metrics were determined as shown in Table 4.17.

Table 4.17: Weights of performance criteria according to the CRITIC method.

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
17.7%	12.0%	11.7%	11.7%	12.2%	16.5%	18.2%

After determining the criteria weights, the decision matrix is ready for the application of the TOPSIS method. Table 4.18 shows the deviation rates (S_i*), of each model from

the positive ideal solution values, deviation rates (S_i^-) from the negative ideal solution values, relative closeness values (C_{i^*}) to the ideal solution, and the weights of the prediction models obtained by normalizing (C_{i^*}) values within the combination.

Table 4.18: TOPSIS method results under CRITIC-weighted criteria weights (with normalized individual forecasting model weights).

	S_i^*	S_i^-	C_{i^*}	**The weights of the component models in the combination**
STL	0.073906	0.28026767	0.791	**13.9%**
ETS	0.072016	0.27612413	0.793	**13.9%**
NNAR	0.11455	0.28675393	0.715	**12.6%**
THETAM	0.265686	0.10710601	0.287	**5.1%**
SMA(n=3)	0.094117	0.28565463	0.752	**13.2%**
RF	0.044657	0.26930329	0.858	**15.1%**
SVM	0.080977	0.2675508	0.768	**13.5%**
XGBoost	0.10557	0.27460595	0.722	**12.7%**
		∑	5.686	**100.0%**

Figure 4.17 shows the observed values for the validation and test sets, alongside the forecasted values generated by the proposed CRITIC-TOPSIS hybrid MCDM-based forecast combination.

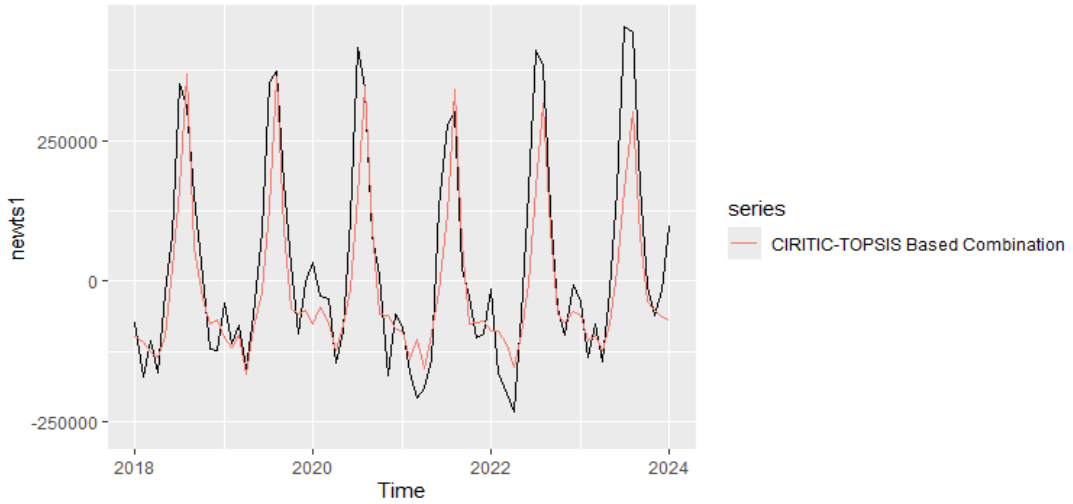

Figure 4.17: The forecasted values generated by the proposed CRITIC-TOPSIS combination forecast, alongside the observed values for the validation and test sets.

4.2.2.1 Interpretation of findings

The new combination model weighted with the weights obtained after the hybrid CRITIC-TOPSIS MCDM model applied to the validation set is compared with the performance of the single models and other combination models on the test set and the following results are obtained (Table 4.19).

Table 4.19: Results of individual forecasting models, combination models, and the proposed CRITIC-TOPSIS combination method on the test set for EIA's dataset.

	Performance Result / Forecasting Model	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Single Models	STL	74540,3	99709,6	78176,6	-214,246	263,446	1,07420	0,51081
	ETS	67309,5	98850,8	76377,7	-186,928	240,237	1,17603	0,46837
	NNAR	1.23495,7	223858,8	144827,6	-260,786	367,994	6,44352	0,59011
	Auto.arima	115974.11	134657.2	115974.1	-396.357	458.2368	1.68304	0.494684
	THETAM	-148631.5	929510,9	515490,8	-448,630	1227,260	0,52000	-0,03472
	NAÏVE	126833.401	142819.6	126833.4	-429.083	494.9744	1.22219	0.255235
	SMA(*n*=3)	87749,0	158351,0	110997,5	-285,622	442,194	2,34143	0,45119
	RF	13825,3	134342,0	110412,8	-255,887	475,126	1,27798	0,25941
	SVM	41110,8	152384,0	120226,9	-188,174	470,140	1,96327	0,40108
	XGBoost	72068,1	167241,7	121791,2	-202,141	313,970	2,78121	0,59907
Proposed Model		**26318.21**	**84862.2**	**64488.09**	**-10.2217**	**130.4874**	**0.77646**	**0.135978**
Other Comb. Models	SA	42351.02	104748	79085.48	-265.05	320.7602	0.57996	-0.25954
	OLS	50137.12	113073.4	87264.95	-262.577	418.4049	1.33108	0.283218
	MED	72317.86	121157.7	88724.85	-270.565	337.9998	2.08044	0.473203
	TA	71016.62	119188.1	86665.65	-266.487	331.6697	2.02347	0.503087
	WA	71517.6	119900.9	87317.18	-268.057	333.8027	2.05576	0.492067
	CLS	constraints are inconsistent, no solution!						
	LAD	34178.6	115754.4	93065.17	-258.691	494.5812	1.26067	0.24991

Note: Points that perform worse than our combination model are painted with pink filler.

The results indicate that the proposed CRITIC-TOPSIS hybrid MCDM-based forecast combination outperformed all individual and existing combination models in forecasting natural gas delivery to US electric power consumers (MMcf).

4.2.3 Application of CILOS-TOPSIS hybrid MCDM technique

CRITIC-TOPSIS-based predictor weight assessment for the combination gave robust results. How might the results differ, if instead of considering contrast intensity and inter-criteria correlation for performance metrics at the decision matrix, we assess their relative importance using cumulative impact and interactions with other criteria? As a methodology, CILOS assigns greater weights to criteria that have less correlation with one another in order to completely eliminate redundancy. The main emphasis is on unique information; variability (spread) is not explicitly taken into consideration. CRITIC, on the other hand, considers both variability and redundancy. It assigns weight to criteria with high variability (a measure of how much the criterion changes or spreads out) in addition to criteria with low correlation (like CILOS). The CILOS method followed by a re-application of the TOPSIS approach is also tested to investigate how neglecting the spreads within the criteria would affect the criteria weights and therefore the overall combination success. By using the decision matrix

given in Table 4.16, the criteria are changed to the maximized ones, making the highest value of a criterion the optimal (best) one. (see Table 4.20).

Table 4.20: Transforming the minimized criteria for CILOS.

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
STL	0.573	1.000	1.000	-0.768	0.796	0.762	0.003
ETS	0.751	0.991	0.979	-0.603	0.810	0.664	0.003
NNAR	0.268	0.498	0.552	-0.857	0.664	0.215	0.003
THETAM	-0.111	0.070	0.105	0.045	0.109	1.000	1.000
SMA(n=3)	0.346	0.563	0.591	-0.227	0.576	0.276	0.003
RF	1.000	0.579	0.592	-0.582	0.786	0.530	0.005
SVM	0.765	0.547	0.548	-0.286	0.618	0.307	0.003
XGBoost	0.332	0.414	0.483	1.000	1.000	0.252	0.003
SUM	3.924	4.661	4.851	-2.278	5.358	4.008	1.022

The normalized values for the elements of Table 4.20 is given in Table 4.21. The bold elements are the highest values for each column.

Table 4.21: Normalized values of decision matrix for CILOS.

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
STL	0.146	**0.215**	**0.206**	0.337	0.148	0.190	0.003
ETS	0.191	0.213	0.202	0.265	**0.151**	0.166	0.003
NNAR	0.068	0.107	0.114	**0.376**	0.124	0.054	0.003
THETAM	-0.028	0.015	0.022	-0.020	0.020	**0.249**	**0.979**
SMA(n=3)	0.088	0.121	0.122	0.100	0.108	0.069	0.003
RF	**0.255**	0.124	0.122	0.256	0.147	0.132	0.005
SVM	0.195	0.117	0.113	0.126	0.115	0.077	0.003
XGBoost	0.085	0.089	0.100	-0.439	0.187	0.063	0.002

The square matrix is created in Table 4.22.

Table 4.22: The square matrix.

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
ME	**0.255**	0.124	0.122	0.256	0.147	0.132	0.005
RMSE	0.146	**0.215**	0.206	0.337	0.148	0.190	0.003
MAE	0.146	0.215	**0.206**	0.337	0.148	0.190	0.003
MPE	0.068	0.107	0.114	**0.376**	0.124	0.054	0.003
MAPE	0.191	0.213	0.202	0.265	**0.151**	0.166	0.003
MASE	-0.028	0.015	0.022	-0.020	0.020	**0.249**	**0.979**
ACF1	-0.028	0.015	0.022	-0.020	0.020	**0.249**	**0.979**

The relative impact loss matrix is given below (Table 4.23).

Table 4.23: The relative impact loss matrix.

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
ME	0.000	0.421	0.408	0.321	0.030	0.470	0.995
RMSE	0.427	0.000	0.000	0.104	0.018	0.238	0.997
MAE	0.427	0.000	0.000	0.104	0.018	0.238	0.997
MPE	0.732	0.502	0.448	0.000	0.180	0.785	0.997
MAPE	0.249	0.009	0.021	0.297	0.000	0.336	0.997
MASE	1.111	0.930	0.895	1.053	0.866	0.000	0.000
ACF1	1.111	0.930	0.895	1.053	0.866	0.000	0.000

The F matrix is created in Table 4.24.

Table 4.24: The F-matrix.

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
ME	-4.057	0.421	0.408	0.321	0.030	0.470	0.995
RMSE	0.427	-2.793	0.000	0.104	0.018	0.238	0.997
MAE	0.427	0.000	-2.667	0.104	0.018	0.238	0.997
MPE	0.732	0.502	0.448	-2.931	0.180	0.785	0.997
MAPE	0.249	0.009	0.021	0.297	-1.977	0.336	0.997
MASE	1.111	0.930	0.895	1.053	0.866	-2.065	0.000
ACF1	1.111	0.930	0.895	1.053	0.866	0.000	-4.984

For the weight assignment for each criterion, The Excel Solver is used. Finally, the weights in Table 4.25 are found for each performance metric as a result of the CILOS technique. After determining the criteria weights, the decision matrix is ready for the application of the steps of the TOPSIS method.

Table 4.25: Criteria weights determined by the CILOS technique.

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
9.5%	8.8%	9.2%	17.8%	14.7%	28.3%	11.7%

As a result, it is possible to find weights of component forecasts by the proposed model in this thesis study under the CILOS-TOPSIS hybrid approach (see Table 4.26).

Table 4.26: TOPSIS method results under CILOS-weighted criteria weights (with normalized individual forecasting model weights).

	S_i^*	S_i^-	C_{i^*}	**The weights of the component models in the combination**
STL	0.0517	0.2823	0.845	**14.9%**
ETS	0.0532	0.2804	0.840	**14.8%**
NNAR	0.1366	0.2594	0.655	**11.5%**
THETAM	0.2461	0.1337	0.352	**6.2%**
SMA(n=3)	0.1100	0.2759	0.715	**12.6%**
RF	0.0455	0.2734	0.857	**15.1%**
SVM	0.0947	0.2681	0.739	**13.0%**
XGBoost	0.1147	0.2487	0.684	**12.0%**
		$\sum$	5.688	100.0%

Figure 4.18 shows the observed values for the validation and test sets, alongside the forecasted values generated by the proposed CILOS-TOPSIS hybrid MCDM-based forecast combination.

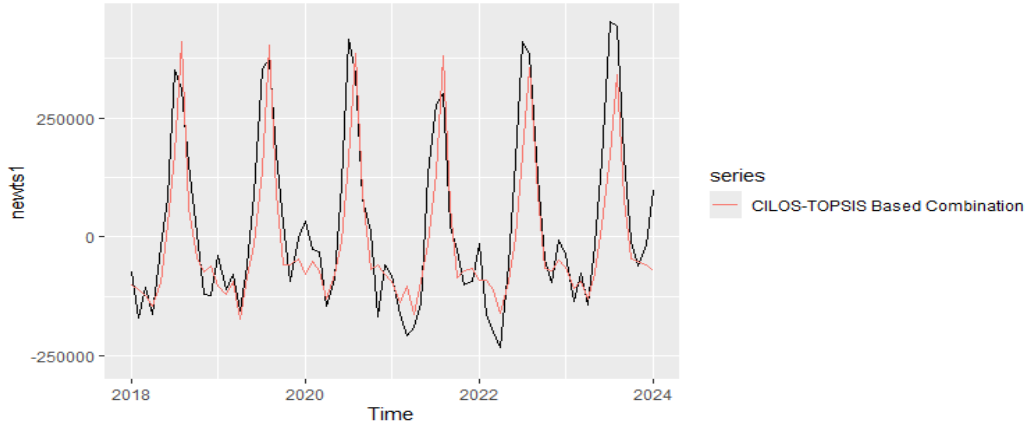

Figure 4.18: The forecasted values generated by the proposed CILOS- TOPSIS combination forecast, alongside the observed values for the validation and test sets.

4.2.3.1 Interpretation of findings

The new combination model weighted with the weights obtained after the hybrid CILOS-TOPSIS MCDM model applied to the validation set was compared with the performance of the single models and other combination models on the test set and the following results were obtained (see Table 4.27).

Table 4.27: Results of individual forecasting models, combination models, and the proposed combination method on the test set.

	Performance Result / Forecasting Model	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Single Models	STL	74540,3	99709,6	78176,6	-214,246	263,446	1,07420	0,51081
	ETS	67309,5	98850,8	76377,7	-186,928	240,237	1,17603	0,46837
	NNAR	1.23495,7	223858,8	144827,6	-260,786	367,994	6,44352	0,59011
	Auto.arima	115974.11	134657.2	115974.1	-396.357	458.2368	1.68304	0.494684
	THETAM	-148631.5	929510,9	515490,8	-448,630	1227,260	0,52000	-0,03472
	NAÏVE	126833.401	142819.6	126833.4	-429.083	494.9744	1.22219	0.255235
	SMA(*n*=3)	87749,0	158351,0	110997,5	-285,622	442,194	2,34143	0,45119
	RF	13825,3	134342,0	110412,8	-255,887	475,126	1,27798	0,25941
	SVM	41110,8	152384,0	120226,9	-188,174	470,140	1,96327	0,40108
	XGBoost	72068,1	167241,7	121791,2	-202,141	313,970	2,78121	0,59907
Proposed Model		**23607.98**	**84654.96**	**66582.4**	**-3.41355**	**134.127**	**0.72070**	**0.027858**
Other Comb. Models	SA	42351.02	104748	79085.48	-265.05	320.7602	0.57996	-0.25954
	OLS	50137.12	113073.4	87264.95	-262.577	418.4049	1.33108	0.283218
	MED	72317.86	121157.7	88724.85	-270.565	337.9998	2.08044	0.473203
	TA	71016.62	119188.1	86665.65	-266.487	331.6697	2.02347	0.503087
	WA	71517.6	119900.9	87317.18	-268.057	333.8027	2.05576	0.492067
	CLS	constraints are inconsistent, no solution!						
	LAD	34178.6	115754.4	93065.17	-258.691	494.5812	1.26067	0.24991

Note: Points that perform worse than our combination model are painted with pink filler.

The results indicate that the proposed CILOS-TOPSIS hybrid MCDM-based forecast combination yielded similar outcomes to the CRITIC-TOPSIS-based combination, with a slight performance improvement. This can mean that neglecting the variation within the performance metrics improves the overall accuracy of the proposed forecast combination method.

4.2.4 Application of CILOS-ARAS hybrid MCDM technique

CILOS-TOPSIS-based predictor weight assessment for the combination gave robust results. How might the results differ if, instead of using TOPSIS and considering the closeness to the ideal solution of forecast models, we assess the utility function value using the ARAS technique? The elements of the decision matrix given in Table 4.16 are normalized by the normalization method given in the methodology section-ARAS method. The normalized decision matrix is created for the CILOS-ARAS method in Table 4.28.

Table 4.28: Normalized decision matrix for CILOS-ARAS.

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
STL	0.1114	0.1766	0.1709	0.1431	0.1251	0.1522	0.0014
ETS	0.1459	0.1751	0.1673	0.1123	0.1274	0.1326	0.0014
NNAR	0.0521	0.0880	0.0943	0.1597	0.1045	0.0430	0.0013
THETAM	0.0216	0.0123	0.0180	0.0084	0.0171	0.1997	0.4946
SMA(n=3)	0.0672	0.0994	0.1011	0.0423	0.0906	0.0551	0.0014
RF	0.1943	0.1022	0.1012	0.1084	0.1236	0.1059	0.0025
SVM	0.1487	0.0966	0.0937	0.0533	0.0971	0.0614	0.0015
XGBoost	0.0646	0.0731	0.0826	0.1863	0.1573	0.0504	0.0013
0-Optimal value	*0.1943*	*0.1766*	*0.1709*	*0.1863*	*0.1573*	*0.1997*	*0.4946*

For calculating the weighted decision matrix, the criteria weights determined by the CILOS technique (Table 4.25) are used. In Table 4.29, a weighted normalized decision matrix can be seen.

Table 4.29: Weighted normalized decision matrix for CILOS-ARAS.

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
STL	0.0106	0.0155	0.0157	0.0254	0.0184	0.0431	0.0002
ETS	0.0139	0.0154	0.0154	0.0199	0.0188	0.0375	0.0002
NNAR	0.0050	0.0077	0.0087	0.0283	0.0154	0.0122	0.0002
THETAM	0.0020	0.0011	0.0017	0.0015	0.0025	0.0565	0.0580
SMA(n=3)	0.0064	0.0087	0.0093	0.0075	0.0133	0.0156	0.0002
RF	0.0185	0.0090	0.0093	0.0192	0.0182	0.0300	0.0003
SVM	0.0141	0.0085	0.0086	0.0095	0.0143	0.0174	0.0002
XGBoost	0.0061	0.0064	0.0076	0.0331	0.0232	0.0143	0.0001
0-Optimal value	*0.0185*	*0.0155*	*0.0157*	*0.0331*	*0.0232*	*0.0565*	*0.0580*

The optimality S_i and the utility degree K_i are found for the final ranking of forecast models in Table 4.29 by using the values of the weighted normalized decision matrix for CILOS-ARAS.

Table 4.30: The optimality S_i and the utility degree K_i for each model and the optimal value.

	STL	ETS	NNAR	THETAM	SMA(n=3)	RF	SVM	XGBoost	*0-Optimal value*
S_i	0.13	0.12	0.08	0.12	0.06	0.10	0.07	0.09	0.22
K_i	0.58	0.55	0.35	0.56	0.28	0.47	0.33	0.41	3.54

Component forecast weights are found by normalizing the values of K_i (see Table 4.31).

Table 4.31: Component forecast weights determined using CILOS-ARAS.

	STL	ETS	NNAR	THETAM	SMA(n=3)	RF	SVM	XGBoost
Weights	16.5%	15.5%	9.9%	15.8%	7.8%	13.4%	9.3%	11.6%

Figure 4.19 shows the observed values for the validation and test sets, alongside the forecasted values generated by the proposed CILOS-ARAS hybrid MCDM-based forecast combination.

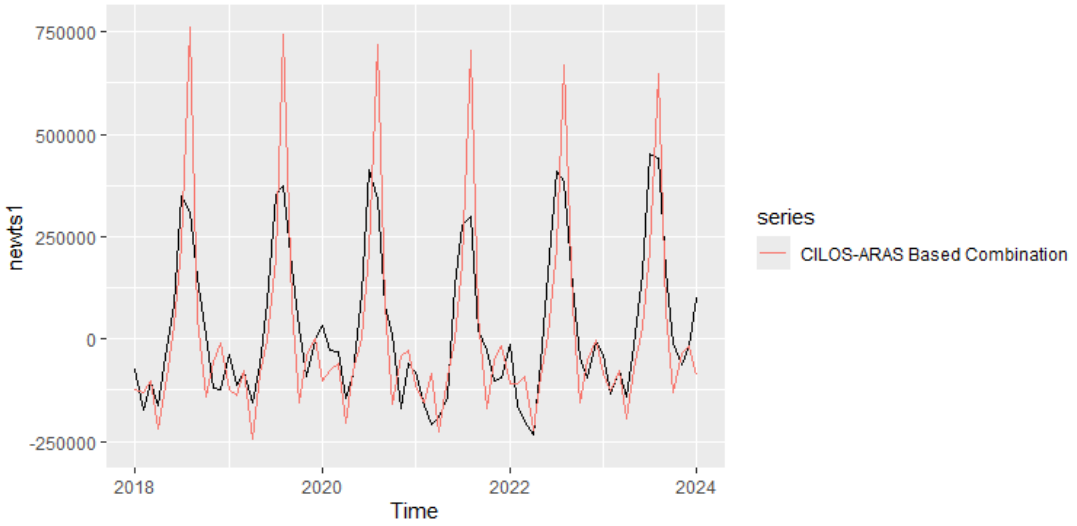

Figure 4.19: The forecasted values generated by the proposed CILOS-ARAS combination forecast, alongside the observed values for the validation and test sets.

4.2.4.1 Interpretation of findings

The new combination model, weighted using the results from the hybrid CILOS-TOPSIS MCDM model applied to the validation set, was compared against the

performance of individual models and other combination models on the test set, yielding the following results (see Table 4.32).

Table 4.32: Results of individual forecasting models, combination models, and the proposed combination method on the test set.

	Performance Result / Forecasting Model	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Single Models	**STL**	74540,3	99709,6	78176,6	-214,246	263,446	1,07420	0,51081
	ETS	67309,5	98850,8	76377,7	-186,928	240,237	1,17603	0,46837
	NNAR	1.23495,7	223858,8	144827,6	-260,786	367,994	6,44352	0,59011
	Auto.arima	115974.11	134657.2	115974.1	-396.357	458.2368	1.68304	0.494684
	THETAM	-148631.5	929510,9	515490,8	-448,630	1227,260	0,52000	-0,03472
	NAÏVE	126833.401	142819.6	126833.4	-429.083	494.9744	1.22219	0.255235
	SMA(*n*=3)	87749,0	158351,0	110997,5	-285,622	442,194	2,34143	0,45119
	RF	13825,3	134342,0	110412,8	-255,887	475,126	1,27798	0,25941
	SVM	41110,8	152384,0	120226,9	-188,174	470,140	1,96327	0,40108
	XGBoost	72068,1	167241,7	121791,2	-202,141	313,970	2,78121	0,59907
Proposed Model		**4455.511**	**140416**	**99489.42**	**61.79982**	**176.1791**	**0.55075**	**-0.23121**
Other Comb. Models	**SA**	42351.02	104748	79085.48	-265.05	320.7602	0.57996	-0.25954
	OLS	50137.12	113073.4	87264.95	-262.577	418.4049	1.33108	0.283218
	MED	72317.86	121157.7	88724.85	-270.565	337.9998	2.08044	0.473203
	TA	71016.62	119188.1	86665.65	-266.487	331.6697	2.02347	0.503087
	WA	71517.6	119900.9	87317.18	-268.057	333.8027	2.05576	0.492067
	CLS	constraints are inconsistent, no solution!						
	LAD	34178.6	115754.4	93065.17	-258.691	494.5812	1.26067	0.24991

Note: Points that perform worse than our combination model are painted with pink filler.

The results indicate that the proposed CILOS-ARAS hybrid MCDM-based forecast combination gave much better results than TOPSIS-based combinations in terms of ME and MASE (especially ME reduced a lot), however, it gave worse results in terms of other performance measurements. The primary reason for this is that the CILOS-ARAS-based forecast combination assigned significantly more weight to THETAM compared to other hybrid models, which consequently influenced the overall results. When compared with the models in the literature, it can be seen that the proposed model continues to give more successful results than the single models, but the existing combination models have better results in terms of RMSE especially.

4.2.5 Application of CRITIC-ARAS hybrid MCDM technique

Under the methodology proposed in this thesis, to see the results also for the framework with CRITIC-ARAS, the ARAS technique's steps are applied to the decision matrix in Table 4.16 by using the criteria weights obtained by CRITIC in Table 4.17. The new CRITIC-ARAS-based weighted normalized decision matrix is in Table 4.33.

Table 4.33: Weighted normalized decision matrix for CRITIC-ARAS.

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
STL	0.0197	0.0211	0.0201	0.0167	0.0153	0.0252	0.0002
ETS	0.0258	0.0209	0.0196	0.0131	0.0156	0.0219	0.0003
NNAR	0.0092	0.0105	0.0111	0.0186	0.0128	0.0071	0.0002
THETAM	0.0038	0.0015	0.0021	0.0010	0.0021	0.0330	0.0899
SMA(n=3)	0.0119	0.0119	0.0119	0.0049	0.0111	0.0091	0.0003
RF	0.0343	0.0122	0.0119	0.0127	0.0151	0.0175	0.0005
SVM	0.0263	0.0115	0.0110	0.0062	0.0119	0.0102	0.0003
XGBoost	0.0114	0.0087	0.0097	0.0217	0.0192	0.0083	0.0002
0-Optimal value	*0.0343*	*0.0211*	*0.0201*	*0.0217*	*0.0192*	*0.0330*	*0.0899*

The optimality S_i and the utility degree K_i are found for the final ranking of forecast models in Table 4.34 by using the values of the weighted normalized decision matrix for CRITIC-ARAS.

Table 4.34: The optimality S_i and the utility degree K_i for each model and the optimal value.

	STL	ETS	NNAR	THETAM	SMA(n=3)	RF	SVM	XGBoost	*0-Optimal value*
S_i	0.12	0.12	0.07	0.13	0.06	0.10	0.08	0.12	0.12
K_i	0.49	0.49	0.29	0.56	0.25	0.44	0.32	0.49	0.49

Component forecast weights are found by normalizing the values of K_i (see Table 4.35).

Table 4.35: Component forecast weights determined using CRITIC-ARAS.

	STL	ETS	NNAR	THETAM	SMA(n=3)	RF	SVM	XGBoost
Weights	15.6%	15.4%	9.1%	17.5%	8.0%	13.7%	10.2%	10.4%

Figure 4.20 shows the observed values for the validation and test sets, alongside the forecasted values generated by the proposed CRITIC-ARAS hybrid MCDM-based forecast combination.

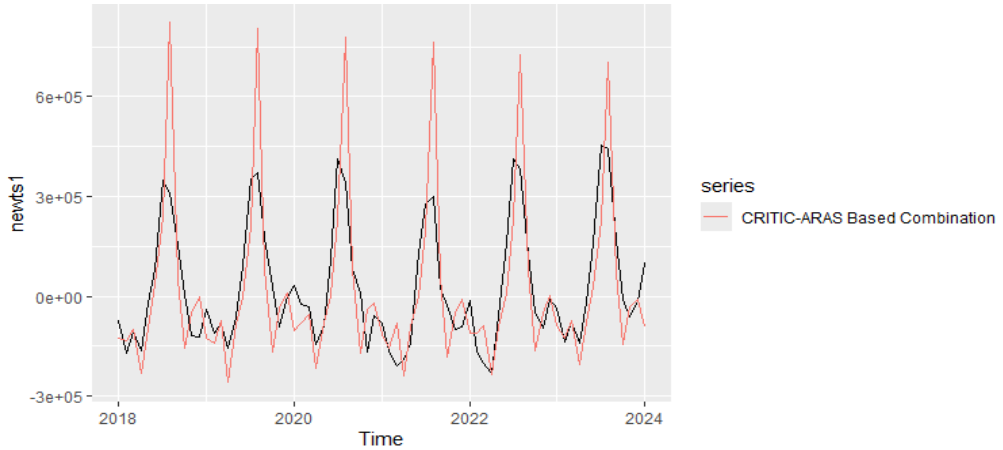

Figure 4.20: The forecasted values generated by the proposed CRITIC-ARAS combination forecast, alongside the observed values for the validation and test sets.

4.2.5.1 Interpretation of findings

The results indicate that the CRITIC-ARAS hybrid MCDM-based forecast combination yielded even smaller ME and MASE values (with ME experiencing a more significant reduction). However, it performed much worse in terms of other performance measures compared to the CILOS-ARAS model (see Table 4.36). The primary reason for this is that the CRITIC-ARAS-based forecast combination assigned even more weight to THETAM than the CILOS-based model.

Table 4.36: Results of individual forecasting models, combination models, and the proposed combination method on the test set.

	Performance Result / Forecasting Model	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Single Models	**STL**	74540,3	99709,6	78176,6	-214,246	263,446	1,07420	0,51081
	ETS	67309,5	98850,8	76377,7	-186,928	240,237	1,17603	0,46837
	NNAR	1.23495,7	223858,8	144827,6	-260,786	367,994	6,44352	0,59011
	Auto.arima	115974.11	134657.2	115974.1	-396.357	458.2368	1.68304	0.494684
	THETAM	-148631.5	929510,9	515490,8	-448,630	1227,260	0,52000	-0,03472
	NAÏVE	126833.401	142819.6	126833.4	-429.083	494.9744	1.22219	0.255235
	SMA(*n*=3)	87749,0	158351,0	110997,5	-285,622	442,194	2,34143	0,45119
	RF	13825,3	134342,0	110412,8	-255,887	475,126	1,27798	0,25941
	SVM	41110,8	152384,0	120226,9	-188,174	470,140	1,96327	0,40108
	XGBoost	72068,1	167241,7	121791,2	-202,141	313,970	2,78121	0,59907
Proposed Model		**666.256**	**154883.2**	**106230.3**	**71.08263**	**193.6033**	**0.54132**	**-0.21669**
Other Comb. Models	**SA**	42351.02	104748	79085.48	-265.05	320.7602	0.57996	-0.25954
	OLS	50137.12	113073.4	87264.95	-262.577	418.4049	1.33108	0.283218
	MED	72317.86	121157.7	88724.85	-270.565	337.9998	2.08044	0.473203
	TA	71016.62	119188.1	86665.65	-266.487	331.6697	2.02347	0.503087
	WA	71517.6	119900.9	87317.18	-268.057	333.8027	2.05576	0.492067
	CLS	constraints are inconsistent, no solution!						
	LAD	34178.6	115754.4	93065.17	-258.691	494.5812	1.26067	0.24991

Note: Points that perform worse than our combination model are painted with pink filler.

4.3 Comparison of MCDM-Based Forecast Combination Methods Developed Using the Methodology Proposed in This Thesis

Table 4.37 presents the performance metrics for the MCDM-based forecast combination methods developed using the methodology proposed in this thesis, which were previously discussed under their respective headings. The best values for each performance metric were written in bold.

Table 4.37: The performance metrics for the MCDM-based forecast combination methods developed using the methodology proposed in this thesis.

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
CRITIC-TOPSIS	26318.2	84862.2	**64488.1**	-10.2	**130.49**	0.78	0.14
CILOS-TOPSIS	23608.0	**84655.0**	66582.4	**-3.41**	134.13	**0.72**	**0.03**
CILOS-ARAS	4455.5	140416.0	99489.4	61.80	176.18	0.55	-0.23
CRITIC-ARAS	**666.3**	154883.2	106230.3	71.08	193.60	0.54	-0.22

5. CONCLUSIONS AND RECOMMENDATIONS

The hybrid MCDM-based new forecast combination method proposed within the scope of this thesis, demonstrated enhanced overall performance in comparison to both key benchmark individual forecasting models and combination approaches reported in the literature. The new methodology demonstrated better performance on the EIA's dataset characterized by more regular and diverse patterns, as well as low-frequency noise. This inference is in line with other combination techniques because such datasets enable different forecasting models to capture varying effects during the combination process, thereby enhancing the overall success of the combination models. However, for the dataset with irregular patterns, high-frequency noise, and significant volatility (D2035 dataset), the proposed model and the other forecasting models produced the prediction results tend to average-based approach, and also the chance of combination models to show superior results over individual models is generally reduced. Still, it can be stated that the proposed model showed promising results compared to benchmark models for this dataset too.

The newly proposed approach enables the development of a combination model that incorporates both traditional and machine learning forecasting methods as components, while the assignment of component weights is carried out by objectively and simultaneously considering various performance metrics, a methodological approach not previously tested in the literature before.

The proposed approach was tested using the EIA dataset by developing and applying CRITIC-TOPSIS, CILOS-TOPSIS, and CILOS-ARAS-based hybrid MCDM techniques. For the application of all of them, first, the decision matrix of predictors versus performance metrics was created. By using CRITIC and CILOS for the determination of performance metrics' weights, a comparative analysis was conducted to evaluate how varying weight assignments for accuracy measurements, based on the inclusion of the variation (spread) within criteria, impact the success of the forecast combination. Because, while CRITIC considers both variability and redundancy, CILOS focuses exclusively on minimizing redundancy and does not directly take variability within criteria into account. In the scenario where both criteria weighting

methods were applied to the same dataset in conjunction with TOPSIS, it was observed that the methodology using CILOS criteria weighting produced relatively better results than the one using CRITIC criteria weighting. Based on this dataset and the component forecast models employed, it can be concluded that using a methodology that does not account for variation within criteria enhances the success of the combination. However, it is important to note that this may yield different results depending on other datasets and component forecast models, making it inappropriate to draw a general conclusion.

Subsequently, the CILOS-TOPSIS and CILOS-ARAS methods were applied for performance evaluation, and the component forecast weights were determined based on their outcomes. In this process, TOPSIS assigned a proximity measure to the ideal solution for each component forecast model, while ARAS assigned utility function values. The findings show that in terms of ME and MASE (with a particularly notable decrease in ME), the suggested CILOS-ARAS hybrid MCDM-based forecast combination performed better than the TOPSIS-based combination. For other performance metrics, though, it produced worse outcomes. Based on the dataset used, considering utility function values reduced ME and MASE; however, in a multi-dimensional context, the overall accuracy decreased. Since the primary goal of the proposed approach is to enhance overall accuracy in a multi-dimensional structure, it can be concluded that, for the dataset we worked with, the CILOS-TOPSIS-based forecast combination approach was the most preferable among all the MCDM-based models developed within the proposed framework, as well as compared to other existing models in the literature.

All in all, the proposed hybrid MCDM-based forecast combination method (especially, the CILOS-TOPSIS-based model) indicated superior results compared to all its key counterparts existing in the literature for the EIA's dataset we worked with. Even when compared with the simple arithmetic mean method, considered the hardest to beat and the most effective combination model in many sources in the literature on time series combination models, the proposed methodology performed significantly better or nearly so across all performance criteria. In this context, we believe that MCDM methods have the potential to introduce a novel perspective to the literature on time series forecasting combinations and pave the way for the integration of additional models within the MCDM framework.

In addition, during the determination and testing of the datasets to be studied in this thesis, combination models were tested on many datasets and the following conclusions were reached:

1) Datasets with high irregularity and volatility, such as stock prices in financial markets, are either white noise which leads to direct mean base modeling, or the mean modeling is insufficient itself which leads to the need for volatility modeling in addition to the mean model. White noise datasets are not suitable for applying the method proposed in this thesis. And, volatility modeling is beyond the scope of the method proposed in this thesis. However, future studies may also try to integrate volatility modeling into the methodology proposed in this thesis.

2) The scope of the method proposed in this thesis can better serve datasets with more regular patterns, low volatility/irregularity, and sufficient size for time series analysis. In this context, energy consumption and production, weather and climate data, manufacturing, retail and sales data for some scenarios, agriculture and food production, sports analytics, tourism analytics, or other similar subjects can be studied by using the proposed model in this thesis with the expectation of more efficient results.

3) The selection of component forecast models within the decision matrix significantly impacted the results' success. As Timmermann (2006), Wang et al. (2023), Lichtendahl and Winkler (2020), and some other sources in the combination literature had already mentioned, the combination's success depended largely on selecting the right forecasts to combine. The diversity among the selected forecast models affected the success of the combination notably under the model proposed in this thesis, as it helps counterbalance errors; however, this was challenging due to predictors being trained on the same data and similar techniques, which can limit the potential for improved accuracy.

4) The framework of MCDM can be used to study for developing other novel approaches for the time series forecast combination literature to improve multi-dimensional accuracy and interpret to relations of performance measurements and predictors among themselves and with each other.

REFERENCES

Aiolfi, Marco, and Allan Timmermann. (2006). 'Persistence in Forecasting Performance and Conditional Combination Strategies'. *Journal of Econometrics* 135 (1–2): 31–53. https://doi.org/10.1016/j.jeconom.2005.07.015.

A**linezhad, Alireza, and Javad Khalili.** (2019). *New Methods and Applications in Multiple Attribute Decision Making (MADM)*. Vol. 277. International Series in Operations Research & Management Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-15009-9.

Alonso, Andre'es M., Daniel Peña, and Juan Romo. (2002). 'Forecasting Time Series with Sieve Bootstrap'. *Journal of Statistical Planning and Inference* 100 (1): 1–11. https://doi.org/10.1016/S0378-3758(01)00092-1.

Armstrong, J. Scott, ed. (2007). *Principles of Forecasting: A Handbook for Researchers and Practitioners*. 7. print. International Series in Operations Research & Management Science 30. New York: Springer.

Armstrong, J.Scott. (1989). 'Combining Forecasts: The End of the Beginning or the Beginning of the End?' *International Journal of Forecasting* 5 (4): 585–88. https://doi.org/10.1016/0169-2070(89)90013-7.

Assimakopoulos, V., and K. Nikolopoulos. (2000). 'The Theta Model: A Decomposition Approach to Forecasting'. *International Journal of Forecasting* 16 (4): 521–30. https://doi.org/10.1016/S0169-2070(00)00066-2.

Azevedo, Vitor G., and Lucila M.S. Campos. (2016). 'Combination of Forecasts for the Price of Crude Oil on the Spot Market'. *International Journal of Production Research* 54 (17): 5219–35. https://doi.org/10.1080/00207543.2016.1162340.

Babikir, Ali, and Henry Mwambi. (2016). 'Evaluating the Combined Forecasts of the Dynamic Factor Model and the Artificial Neural Network Model Using Linear and Nonlinear Combining Methods'. *Empirical Economics* 51 (4): 1541–56. https://doi.org/10.1007/s00181-015-1049-1.

Badulescu, Yvonne, Ari-Pekka Hameri, and Naoufel Cheikhrouhou. (2021). 'Evaluating Demand Forecasting Models Using Multi-Criteria Decision-Making Approach'. *Journal of Advances in Management Research* 18 (5): 661–83. https://doi.org/10.1108/JAMR-05-2020-0080.

Bates, J M, and C W J Granger. (1969). 'The Combination of Forecasts'. *Operational Research Society* 20 (4): 451–68.

Blanc, Sebastian M., and Thomas Setzer. (2016). 'When to Choose the Simple Average in Forecast Combination'. *Journal of Business Research* 69 (10): 3951–62. https://doi.org/10.1016/j.jbusres.2016.05.013.

Box, G. E. P., and D. R. Cox. (1964). 'An Analysis of Transformations'. *Journal of the Royal Statistical Society: Series B (Methodological)* 26 (2): 211–43. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x.

Box, George E. P., and Gwilym M. Jenkins. (1976). *Time Series Analysis: Forecasting and Control*. Holden-Day.

Box, George E. P., Gwilym M. Jenkins, Gregory C. Reinsel, and Greta M. Ljung. (2016). *Time Series Analysis: Forecasting and Control*. Fifth edition. Wiley Series in Probability and Statistics. Hoboken, New Jersey: Wiley.

Breiman, Leo. (2001). 'Random Forests'. *Machine Learning* 45 (1): 5–32. https://doi.org/10.1023/A:1010933404324.

Brown, R.G. (1959). *Statistical Forecasting for Inventory Control*. McGraw-Hill. https://books.google.com.tr/books?id=oKI8AAAAIAAJ.

Chen, Tianqi, and Carlos Guestrin. (2016). 'XGBoost: A Scalable Tree Boosting System'. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–94. San Francisco California USA: ACM. https://doi.org/10.1145/2939672.2939785.

Clemen, Robert T. (1989). 'Combining Forecasts: A Review and Annotated Bibliography'. *International Journal of Forecasting* 5 (4): 559–83. https://doi.org/10.1016/0169-2070(89)90012-5.

Clemen, T. (1989). 'Combining Forecasts: A Review and Annotated Bibliography'. *International Journal of Forecasting* 5 (4): 559–83.

Cleveland, RB. (1990). 'STL : A Seasonal-Trend Decomposition Procedure Based on Loess'. *J Off Stat* 6:3–73.

Crane, Dwight B., and James R. Crotty. (1967). 'A Two-Stage Forecasting Model: Exponential Smoothing and Multiple Regression'. *Management Science* 13 (8): B-501-B-507. https://doi.org/10.1287/mnsc.13.8.B501.

Davidescu, Adriana AnaMaria, Simona-Andreea Apostu, and Andreea Paul. (2021). 'Comparative Analysis of Different Univariate Forecasting Methods in Modelling and Predicting the Romanian Unemployment Rate for the Period 2021–2022'. *Entropy* 23 (3): 325. https://doi.org/10.3390/e23030325.

Davydenko, Andrey, and Robert Fildes. (2013). 'Measuring Forecasting Accuracy: The Case of Judgmental Adjustments to SKU-Level Demand Forecasts'. *International Journal of Forecasting* 29 (3): 510–22. https://doi.org/10.1016/j.ijforecast.2012.09.002.

De Livera, Alysha M., Rob J. Hyndman, and Ralph D. Snyder. (2011). 'Forecasting Time Series With Complex Seasonal Patterns Using Exponential Smoothing'. *Journal of the American Statistical Association* 106 (496): 1513–27. https://doi.org/10.1198/jasa.2011.tm09771.

De Menezes, Lilian M, Derek W. Bunn, and James W Taylor. (2000). 'Review of Guidelines for the Use of Combined Forecasts'. *European Journal of Operational Research* 120 (1): 190–204. https://doi.org/10.1016/S0377-2217(98)00380-4.

Diakoulaki, D., G. Mavrotas, and L. Papayannakis. (1995). 'Determining Objective Weights in Multiple Criteria Problems: The Critic Method'.

Computers & Operations Research 22 (7): 763–70. https://doi.org/10.1016/0305-0548(94)00059-H.

Dickey, David A., and Wayne A. Fuller. (1979). 'Distribution of the Estimators for Autoregressive Time Series With a Unit Root'. *Journal of the American Statistical Association* 74 (366): 427. https://doi.org/10.2307/2286348.

Diebold, Francis X. 2007. *Elements of Forecasting*. 4th ed. Mason, Ohio: Thomson/South-Western.

Donaldson, R. Glen, and Mark Kamstra. (1996). 'Forecast Combining with Neural Networks'. *Journal of Forecasting* 15 (1): 49–61. https://doi.org/10.1002/(SICI)1099-131X(199601)15:1<49::AID-FOR604>3.0.CO;2-2.

Elliott, Graham, Antonio Gargano, and Allan Timmermann. (2013). 'Complete Subset Regressions'. *Journal of Econometrics* 177 (2): 357–73. https://doi.org/10.1016/j.jeconom.2013.04.017.

Franses, Philip Hans. (2011). 'Model Selection for Forecast Combination'. *Applied Economics* 43 (14): 1721–27. https://doi.org/10.1080/00036840902762753.

Gardner, Everette S. (1985). 'Exponential Smoothing: The State of the Art'. *Journal of Forecasting* 4 (1): 1–28. https://doi.org/10.1002/for.3980040103.

Gaur, Sulakshya, Satyanarayana Dosapati, and Abhay Tawalare. (2023). 'Stakeholder Assessment in Construction Projects Using a CRITIC-TOPSIS Approach'. *Built Environment Project and Asset Management* 13 (2): 217–37. https://doi.org/10.1108/BEPAM-10-2021-0122.

Géron, Aurélien. (2019). *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. Second edition. Covid-19 Collection. Beijing Boston Farnham Sebastopol Tokyo: O'Reilly.

Granger, Clive W. J., and Ramu Ramanathan. (1984). 'Improved Methods of Combining Forecasts'. *Journal of Forecasting* 3 (2): 197–204. https://doi.org/10.1002/for.3980030207.

Granger, C.W.J., and P. Newbold. (1974). 'Spurious Regressions in Econometrics'. *Journal of Econometrics* 2 (2): 111–20. https://doi.org/10.1016/0304-4076(74)90034-7.

Hendry, David F., and Michael P. Clements. (2002). 'Pooling of Forecasts'. *The Econometrics Journal* 7 (1): 1–31. https://doi.org/10.1111/j.1368-423X.2004.00119.x.

Holt, Charles C. (2004). 'Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages'. *International Journal of Forecasting* 20 (1): 5–10. https://doi.org/10.1016/j.ijforecast.2003.09.015.

***Homepage - U.S. Energy Information Administration (EIA)**. (n.d). Retrieved September 12, 2024 from https://www.eia.gov/index.php.*

Hsiao, Cheng, and Shui Ki Wan. (2014). 'Is There an Optimal Forecast Combination?' *Journal of Econometrics* 178 (January):294–309. https://doi.org/10.1016/j.jeconom.2013.11.003.

Hwang, Ching-Lai, and Kwangsun Yoon. (1981). *Multiple Attribute Decision Making*. Vol. 186. Lecture Notes in Economics and Mathematical Systems. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-48318-9.

Hyndman, Rob J, and George Athanasopoulos. (2018). 'Forecasting: Principles and Practice', April.

Hyndman, Rob J., and Anne B. Koehler. (2006). 'Another Look at Measures of Forecast Accuracy'. *International Journal of Forecasting* 22 (4): 679–88. https://doi.org/10.1016/j.ijforecast.2006.03.001.

Hyndman, Rob, Anne Koehler, Keith Ord, and Ralph Snyder. (2008). *Forecasting with Exponential Smoothing*. Springer Series in Statistics. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-71918-2.

Isikli, Erkan, and Seyda Serdarasan. (2023). 'Correction to: The Power of Combination Models in Energy Demand Forecasting'. In *Decision Making Using AI in Energy and Sustainability*, edited by Gülgün Kayakutlu and M. Özgür Kayalica, C1–C1. Applied Innovation and Technology Management. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-38387-8_19.

Jurman, Giuseppe, Samantha Riccadonna, and Cesare Furlanello. (2012). 'A Comparison of MCC and CEN Error Measures in Multi-Class Prediction'. Edited by Giuseppe Biondi-Zoccai. *PLoS ONE* 7 (8): e41882. https://doi.org/10.1371/journal.pone.0041882.

Kapetanios, George, Vincent Labhard, and Simon Price. (2008). 'Forecast Combination and the Bank of England's Suite of Statistical Forecasting Models'. *Economic Modelling* 25 (4): 772–92. https://doi.org/10.1016/j.econmod.2007.11.004.

Kemal Burç Ülengin. (2024). 'Financial Econometrics Lecture Notes'.

Koning, Alex J., Philip Hans Franses, Michèle Hibon, and H.O. Stekler. (2005). 'The M3 Competition: Statistical Tests of the Results'. *International Journal of Forecasting* 21 (3): 397–409. https://doi.org/10.1016/j.ijforecast.2004.10.003.

Krasnopolsky, Vladimir M., and Ying Lin. (2012). 'A Neural Network Nonlinear Multimodel Ensemble to Improve Precipitation Forecasts over Continental US'. *Advances in Meteorology* 2012:1–11. https://doi.org/10.1155/2012/649450.

Kumar, Ajay, and Kamakleep Kaur. (2021). 'A Hybrid SOM-Fuzzy Time Series (SOMFTS) Technique for Future Forecasting of COVID-19 Cases and MCDM Based Evaluation of COVID-19 Forecasting Models'. In *2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, 612–17. Greater Noida, India: IEEE. https://doi.org/10.1109/ICCCIS51004.2021.9397216.

Kwiatkowski, Denis, Peter C.B. Phillips, Peter Schmidt, and Yongcheol Shin. (1992). 'Testing the Null Hypothesis of Stationarity against the Alternative of

a Unit Root'. *Journal of Econometrics* 54 (1–3): 159–78. https://doi.org/10.1016/0304-4076(92)90104-Y.

Lichtendahl, Kenneth C., and Robert L. Winkler. (2020). 'Why Do Some Combinations Perform Better than Others?' *International Journal of Forecasting* 36 (1): 142–49. https://doi.org/10.1016/j.ijforecast.2019.03.027.

M4-Methods/Dataset/M4-Info.Csv at Master · Mcompetitions/M4-Methods. *(n.d.) GitHub. Retrieved November, 2, 2024, from https://github.com/Mcompetitions/M4-methods/blob/master/Dataset/M4-info.csv.*

Makridakis Competitions. (2023). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Makridakis_Competitions&oldid=1190322521.

Makridakis, S., A. Andersen, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen, and R. Winkler. (1982). 'The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition'. *Journal of Forecasting* 1 (2): 111–53. https://doi.org/10.1002/for.3980010202.

Makridakis, Spyros G., Steven C. Wheelwright, and Rob J. Hyndman. (1998). *Forecasting: Methods and Applications*. 3rd ed. New York: John Wiley & Sons.

Makridakis, Spyros, and Michèle Hibon. (2000). 'The M3-Competition: Results, Conclusions and Implications'. *International Journal of Forecasting* 16 (4): 451–76. https://doi.org/10.1016/S0169-2070(00)00057-1.

Makridakis, Spyros, Evangelos Spiliotis, and Vassilios Assimakopoulos. (2020). 'The M4 Competition: 100,000 Time Series and 61 Forecasting Methods'. *International Journal of Forecasting* 36 (1): 54–74. https://doi.org/10.1016/j.ijforecast.2019.04.014.

———. (2022). 'M5 Accuracy Competition: Results, Findings, and Conclusions'. *International Journal of Forecasting* 38 (4): 1346–64. https://doi.org/10.1016/j.ijforecast.2021.11.013.

Makridakis, Spyros, Evangelos Spiliotis, Vassilios Assimakopoulos, Zhi Chen, Anil Gaba, Ilia Tsetlin, and Robert L. Winkler. (2022). 'The M5 Uncertainty Competition: Results, Findings and Conclusions'. *International Journal of Forecasting* 38 (4): 1365–85. https://doi.org/10.1016/j.ijforecast.2021.10.009.

Makridakis, Spyros, Evangelos Spiliotis, Ross Hollyman, Fotios Petropoulos, Norman Swanson, and Anil Gaba. (2023). 'The M6 Forecasting Competition: Bridging the Gap between Forecasting and Investment Decisions'. arXiv. https://doi.org/10.48550/ARXIV.2310.13357.

Mariñas-Collado, Irene, Ana E. Sipols, M. Teresa Santos-Martín, and Elisa Frutos-Bernal. (2022). 'Clustering and Forecasting Urban Bus Passenger Demand with a Combination of Time Series Models'. *Mathematics* 10 (15): 2670. https://doi.org/10.3390/math10152670.

Masini, Ricardo P., Marcelo C. Medeiros, and Eduardo F. Mendes. (2023). 'Machine Learning Advances for Time Series Forecasting'. *Journal of Economic Surveys* 37 (1): 76–111. https://doi.org/10.1111/joes.12429.

Maté, Carlos. (2021). 'Combining Interval Time Series Forecasts. A First Step in a Long Way (Research Agenda)'. *Revista Colombiana de Estadística* 44 (1): 123–57. https://doi.org/10.15446/rce.v44n1.85116.

Mehdiyev, Nijat, David Enke, Peter Fettke, and Peter Loos. (2016). 'Evaluating Forecasting Methods by Considering Different Accuracy Measures'. *Procedia Computer Science* 95:264–71. https://doi.org/10.1016/j.procs.2016.09.332.

Meng, Ge, Jian Liu, and Rui Feng. (2022). 'Prediction of Construction and Production Safety Accidents in China Based on Time Series Analysis Combination Model'. *Applied Sciences* 12 (21): 11124. https://doi.org/10.3390/app122111124.

Min, Chung-ki, and Arnold Zellner. (1993). 'Bayesian and Non-Bayesian Methods for Combining Models and Forecasts with Applications to Forecasting International Growth Rates'. *Journal of Econometrics* 56 (1–2): 89–118. https://doi.org/10.1016/0304-4076(93)90102-B.

Mirkin BG. Problema Grupovogo Vibora. Moscow, Russia: Nauka; 1974

Nowotarski, Jakub, Eran Raviv, Stefan Trück, and Rafał Weron. (2014). 'An Empirical Comparison of Alternative Schemes for Combining Electricity Spot Price Forecasts'. *Energy Economics* 46 (November):395–412. https://doi.org/10.1016/j.eneco.2014.07.014.

Palm, Franz C., and Arnold Zellner. (1992). 'To Combine or Not to Combine? Issues of Combining Forecasts'. *Journal of Forecasting* 11 (8): 687–701. https://doi.org/10.1002/for.3980110806.

Pawlikowski, Maciej, and Agata Chorowska. (2020). 'Weighted Ensemble of Statistical Models'. *International Journal of Forecasting* 36 (1): 93–97. https://doi.org/10.1016/j.ijforecast.2019.03.019.

Petropoulos, Fotios, Rob J. Hyndman, and Christoph Bergmeir. (2018). 'Exploring the Sources of Uncertainty: Why Does Bagging for Time Series Forecasting Work?' *European Journal of Operational Research* 268 (2): 545–54. https://doi.org/10.1016/j.ejor.2018.01.045.

Rob Hyndman, G. Athanasopoulos. (2021). *Forecasting: Principles and Practice (3rd Ed)*. 3rd ed. OTexts. https://otexts.com/fpp3/#.

Ryan, Oisín, Jonas M B Haslbeck, and Lourens Waldorp. (2023). 'Non-Stationarity in Time-Series Analysis: Modeling Stochastic and Deterministic Trends'. https://doi.org/10.31234/osf.io/z7ja2.

Savun-Hekimoğlu, Basak, Barbaros Erbay, Mustafa Hekimoğlu, and Selmin Burak. (2021). 'Evaluation of Water Supply Alternatives for Istanbul Using Forecasting and Multi-Criteria Decision Making Methods'. *Journal of Cleaner Production* 287 (March):125080. https://doi.org/10.1016/j.jclepro.2020.125080.

Schölkopf, Bernhard, and Alexander Johannes Smola. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and*

Beyond. Reprint. Adaptive Computation and Machine Learning Series. Cambridge, Mass.: MIT Press.

Siegel, Andrew F. (2016). 'Time Series'. In *Practical Business Statistics*, 431–66. Elsevier. https://doi.org/10.1016/B978-0-12-804250-2.00014-6.

Stock, James H., and Mark W. Watson. (2004). 'Combination Forecasts of Output Growth in a Seven-country Data Set'. *Journal of Forecasting* 23 (6): 405–30. https://doi.org/10.1002/for.928.

Stock, James, and Mark Watson. (1998). 'A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series'. w6607. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w6607.

Su, Ying, Morgan C. Wang, and Shuai Liu. (2024). 'Automated Machine Learning Algorithm Using Recurrent Neural Network to Perform Long-Term Time Series Forecasting'. *Computers, Materials & Continua* 78 (3): 3529–49. https://doi.org/10.32604/cmc.2024.047189.

Swanson, Norman R., and Tian Zeng. (2001). 'Choosing among Competing Econometric Forecasts: Regression-based Forecast Combination Using Model Selection'. *Journal of Forecasting* 20 (6): 425–40. https://doi.org/10.1002/for.784.

Tao, Lili, Yan Chen, Xiaodong Liu, and Xin Wang. (2012). 'An Integrated Multiple Criteria Decision Making Model Applying Axiomatic Fuzzy Set Theory'. *Applied Mathematical Modelling* 36 (10): 5046–58. https://doi.org/10.1016/j.apm.2011.12.042.

Theodosiou, Marina. (2011). 'Forecasting Monthly and Quarterly Time Series Using STL Decomposition'. *International Journal of Forecasting* 27 (4): 1178–95. https://doi.org/10.1016/j.ijforecast.2010.11.002.

Timmermann, Allan. (2006). 'Chapter 4 Forecast Combinations'. In *Handbook of Economic Forecasting*, 1:135–96. Elsevier. https://doi.org/10.1016/S1574-0706(05)01004-9.

Velicer, Wayne F., and Joseph L. Fava. (2003). 'Time Series Analysis'. In *Handbook of Psychology*, edited by Irving B. Weiner, 1st ed., 581–606. Wiley. https://doi.org/10.1002/0471264385.wei0223.

Wang, Xiaoqian, Rob J. Hyndman, Feng Li, and Yanfei Kang. (2023). 'Forecast Combinations: An over 50-Year Review'. *International Journal of Forecasting* 39 (4): 1518–47. https://doi.org/10.1016/j.ijforecast.2022.11.005.

Wei, Xiaoqiao, and Yuhong Yang. (2012). 'Robust Forecast Combinations'. *Journal of Econometrics* 166 (2): 224–36. https://doi.org/10.1016/j.jeconom.2011.09.035.

Weiss, Christoph, E., Eran Raviv, and Gernot Roetzer. (2019). 'Forecast Combinations in R Using the ForecastComb Package'. *The R Journal* 10 (2): 262. https://doi.org/10.32614/RJ-2018-052.

Widodo, Agus, and Indra Budi. 2011. 'Model Selection for Time Series Forecasting Using Similarity Measure'.

Winters, Peter R. (1960). ‘Forecasting Sales by Exponentially Weighted Moving Averages’. *Management Science* 6 (3): 324–42. https://doi.org/10.1287/mnsc.6.3.324.

Xu, Bing, and Jamal Ouenniche. (2012). ‘Performance Evaluation of Competing Forecasting Models: A Multidimensional Framework Based on MCDA’. *Expert Systems with Applications* 39 (9): 8312–24. https://doi.org/10.1016/j.eswa.2012.01.167.

Yang, Yuhong. (2004). ‘Combining Forecasting Procedures: Some Theoretical Results’. *Econometric Theory* 20 (01). https://doi.org/10.1017/S0266466604201086.

Zavadskas, Edmundas Kazimieras, and Valentinas Podvezko. (2016). ‘Integrated Determination of Objective Criteria Weights in MCDM’. *International Journal of Information Technology & Decision Making* 15 (02): 267–83. https://doi.org/10.1142/S0219622016500036.

Zavadskas, Edmundas Kazimieras, and Zenonas Turskis. (2010). ‘A New Additive Ratio Assessment (ARAS) Method In Multicriteria Decision-Making / Naujas Adıtyvınıs Krıterıjų Santykių Įvertınımo Metodas (Aras) Daugıakrıterınıams Uždavınıams Spręstı’. *Technological and Economic Development of Economy* 16 (2): 159–72. https://doi.org/10.3846/tede.2010.10.

Ziegel, Eric R. (1995). ‘Applied Econometric Time Series’. *Technometrics* 37 (4): 469–70. https://doi.org/10.1080/00401706.1995.10484400.

APPENDICES

APPENDIX A: Tests for the D2035 dataset.
APPENDIX B: Tests for the EIA's dataset.

APPENDIX A

```
> library(urca)
> T<-length(myts)
> k_max<-floor(12*(T/100)^(1/4))
> k_max
[1] 30
```

Figure A.1: The optimum maximum lag length determination based on Schwert's Rule for D2035 at R.

(a)

Augmented Dickey-Fuller Unit Root Test on LOGDATA

Null Hypothesis: LOGDATA has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 5 (Automatic - based on AIC, maxlag=30)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.078594	0.5571
Test critical values:	1% level	-3.960228	
	5% level	-3.410877	
	10% level	-3.127241	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(LOGDATA)
Method: Least Squares
Date: 10/29/24 Time: 23:41
Sample (adjusted): 4/05/2001 9/24/2012
Included observations: 4191 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOGDATA(-1)	-0.002495	0.001201	-2.078594	0.0377
D(LOGDATA(-1))	-0.039317	0.015458	-2.543481	0.0110
D(LOGDATA(-2))	0.004865	0.015469	0.314473	0.7532
D(LOGDATA(-3))	-0.010165	0.015462	-0.657416	0.5109
D(LOGDATA(-4))	-0.022676	0.015460	-1.466756	0.1425
D(LOGDATA(-5))	-0.035935	0.015450	-2.325873	0.0201
C	0.019413	0.009101	2.132944	0.0330
@TREND("3/30/2001")	1.20E-06	6.07E-07	1.971237	0.0488

R-squared	0.004650	Mean dependent var	0.000398
Adjusted R-squared	0.002984	S.D. dependent var	0.010891
S.E. of regression	0.010875	Akaike info criterion	-6.202831
Sum squared resid	0.494687	Schwarz criterion	-6.190727
Log likelihood	13006.03	Hannan-Quinn criter.	-6.198551
F-statistic	2.791592	Durbin-Watson stat	1.997431
Prob(F-statistic)	0.006738		

(b)

KPSS Unit Root Test on LOGDATA

Null Hypothesis: LOGDATA is stationary
Exogenous: Constant, Linear Trend
Bandwidth: 51 (Newey-West automatic) using Bartlett kernel

		LM-Stat.
Kwiatkowski-Phillips-Schmidt-Shin test statistic		0.696996
Asymptotic critical values*:	1% level	0.216000
	5% level	0.146000
	10% level	0.119000

*Kwiatkowski-Phillips-Schmidt-Shin (1992, Table 1)

Residual variance (no correction)	0.019757
HAC corrected variance (Bartlett kernel)	0.971292

KPSS Test Equation
Dependent Variable: LOGDATA
Method: Least Squares
Date: 10/29/24 Time: 23:42
Sample: 3/30/2001 9/24/2012
Included observations: 4197

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.578287	0.004340	1746.300	0.0000
@TREND("3/30/2001")	0.000491	1.79E-06	274.3629	0.0000

R-squared	0.947213	Mean dependent var	8.609343
Adjusted R-squared	0.947200	S.D. dependent var	0.611861
S.E. of regression	0.140595	Akaike info criterion	-1.085393
Sum squared resid	82.92211	Schwarz criterion	-1.082371
Log likelihood	2279.698	Hannan-Quinn criter.	-1.084325
F-statistic	75275.01	Durbin-Watson stat	0.006001
Prob(F-statistic)	0.000000		

Figure A.2: Stationarity tests for log-transformed D2035 dataset: (a) ADF test. (b) KPSS test.

(a)

Augmented Dickey-Fuller Unit Root Test on LOGDIFF

Null Hypothesis: LOGDIFF has a unit root
Exogenous: None
Lag Length: 4 (Automatic - based on AIC, maxlag=30)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-30.83233	0.0000
Test critical values:	1% level	-2.565514	
	5% level	-1.940900	
	10% level	-1.616649	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(LOGDIFF)
Method: Least Squares
Date: 12/15/24 Time: 20:57
Sample (adjusted): 4/05/2001 9/24/2012
Included observations: 4191 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOGDIFF(-1)	-1.102088	0.035745	-30.83233	0.0000
D(LOGDIFF(-1))	0.062939	0.031757	1.981910	0.0476
D(LOGDIFF(-2))	0.068038	0.027369	2.485943	0.0130
D(LOGDIFF(-3))	0.058081	0.022253	2.609989	0.0091
D(LOGDIFF(-4))	0.035649	0.015443	2.308448	0.0210

R-squared	0.520120	Mean dependent var	3.29E-07
Adjusted R-squared	0.519661	S.D. dependent var	0.015706
S.E. of regression	0.010885	Akaike info criterion	-6.201584
Sum squared resid	0.496014	Schwarz criterion	-6.194019
Log likelihood	13000.42	Hannan-Quinn criter.	-6.198908
Durbin-Watson stat	1.997411		

(b)

KPSS Unit Root Test on LOGDIFF

Null Hypothesis: LOGDIFF is stationary
Exogenous: Constant
Bandwidth: 7 (Newey-West automatic) using Bartlett kernel

		LM-Stat.
Kwiatkowski-Phillips-Schmidt-Shin test statistic		0.081050
Asymptotic critical values*:	1% level	0.739000
	5% level	0.463000
	10% level	0.347000

*Kwiatkowski-Phillips-Schmidt-Shin (1992, Table 1)

Residual variance (no correction)	0.000119
HAC corrected variance (Bartlett kernel)	0.000105

KPSS Test Equation
Dependent Variable: LOGDIFF
Method: Least Squares
Date: 12/15/24 Time: 20:58
Sample (adjusted): 3/31/2001 9/24/2012
Included observations: 4196 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000389	0.000168	2.311592	0.0208

R-squared	0.000000	Mean dependent var	0.000389
Adjusted R-squared	0.000000	S.D. dependent var	0.010890
S.E. of regression	0.010890	Akaike info criterion	-6.201622
Sum squared resid	0.497535	Schwarz criterion	-6.200111
Log likelihood	13012.00	Hannan-Quinn criter.	-6.201088
Durbin-Watson stat	2.078133		

Figure A. Stationarity tests for log-differenced D2035 dataset: (a) ADF test. (b) KPSS test.

APPENDIX B

```
> library(urca)
> T <- length(data)
> T <- length(myts)
> k_max <- floor(12 * (T / 100)^(1/4))
> k_max
[1] 15
```

Figure B.1: The optimum maximum lag length determination based on Schwert's Rule for EIA dataset at R.

(a)

Augmented Dickey-Fuller Unit Root Test on USGAS

Null Hypothesis: USGAS has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 14 (Automatic - based on HQ, maxlag=15)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-4.328321	0.0033
Test critical values:	1% level	-3.992933	
	5% level	-3.426809	
	10% level	-3.136666	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(USGAS)
Method: Least Squares
Date: 11/04/24 Time: 15:33
Sample (adjusted): 2002M04 2024M05
Included observations: 266 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
USGAS(-1)	-0.370340	0.085562	-4.328321	0.0000
D(USGAS(-1))	-0.030196	0.093361	-0.323436	0.7466
D(USGAS(-2))	-0.001260	0.093121	-0.013534	0.9892
D(USGAS(-3))	0.038236	0.091987	0.415674	0.6780
D(USGAS(-4))	0.057215	0.089898	0.636437	0.5251
D(USGAS(-5))	0.077811	0.084863	0.916903	0.3601
D(USGAS(-6))	-0.023637	0.079191	-0.298479	0.7656
D(USGAS(-7))	0.003942	0.074251	0.053094	0.9577
D(USGAS(-8))	-0.057986	0.071021	-0.816466	0.4150
D(USGAS(-9))	-0.155137	0.066592	-2.329654	0.0206
D(USGAS(-10))	-0.089237	0.065040	-1.372032	0.1713
D(USGAS(-11))	-0.028393	0.062201	-0.456481	0.6484
D(USGAS(-12))	0.612313	0.059212	10.34097	0.0000
D(USGAS(-13))	0.294901	0.066789	4.415415	0.0000
D(USGAS(-14))	0.152153	0.064519	2.358264	0.0191
C	128522.1	30070.55	4.274018	0.0000
@TREND("2001M01")	936.0603	212.2286	4.410624	0.0000

R-squared	0.823168	Mean dependent var	2380.925
Adjusted R-squared	0.811805	S.D. dependent var	115215.8
S.E. of regression	49982.27	Akaike info criterion	24.53850
Sum squared resid	6.22E+11	Schwarz criterion	24.76752
Log likelihood	-3246.621	Hannan-Quinn criter.	24.63051
F-statistic	72.44462	Durbin-Watson stat	2.018112
Prob(F-statistic)	0.000000		

(b)

KPSS Unit Root Test on USGAS

Null Hypothesis: USGAS is stationary
Exogenous: Constant, Linear Trend
Bandwidth: 17 (Newey-West automatic) using Bartlett kernel

		LM-Stat.
Kwiatkowski-Phillips-Schmidt-Shin test statistic		0.103177
Asymptotic critical values*:	1% level	0.216000
	5% level	0.146000
	10% level	0.119000

*Kwiatkowski-Phillips-Schmidt-Shin (1992, Table 1)

Residual variance (no correction)	2.35E+10
HAC corrected variance (Bartlett kernel)	3.17E+10

KPSS Test Equation
Dependent Variable: USGAS
Method: Least Squares
Date: 11/04/24 Time: 15:36
Sample: 2001M01 2024M05
Included observations: 281

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	367570.2	18289.86	20.09694	0.0000
@TREND("2001M01")	2408.260	113.0383	21.30481	0.0000

R-squared	0.619318	Mean dependent var	704726.6
Adjusted R-squared	0.617953	S.D. dependent var	248675.8
S.E. of regression	153706.3	Akaike info criterion	26.73057
Sum squared resid	6.59E+12	Schwarz criterion	26.75646
Log likelihood	-3753.645	Hannan-Quinn criter.	26.74095
F-statistic	453.8949	Durbin-Watson stat	0.546482
Prob(F-statistic)	0.000000		

Figure B.2: Stationarity tests for original EIA dataset: (a) ADF test. (b) KPSS test.

```
Forecast method: ETS(A,N,A)

Model Information:
ETS(A,N,A)

Call:
ets(y = train)

  Smoothing parameters:
    alpha = 0.7005
    gamma = 1e-04

  Initial states:
    l = 63658.8589
    s = -83824.72 -114364.7 -30416.5 70959.4 263320.5 245713
           81097.9 -38534.68 -96335.43 -93802.54 -125925 -77887.25

  sigma:  40479.58

     AIC     AICc      BIC
5428.683 5431.236 5478.454
```

Figure B.3: ETS model output for detrended EIA dataset from R.

```
Forecast method: ARIMA(1,0,0)(0,1,1)[12] with drift

Model Information:
Series: train
ARIMA(1,0,0)(0,1,1)[12] with drift

Coefficients:
         ar1     sma1      drift
      0.7180  -0.8787  -332.7412
s.e.  0.0504   0.0710   187.7894

sigma^2 = 1.626e+09:  log likelihood = -2316.15
AIC=4640.31   AICc=4640.52   BIC=4653.34
```

Figure B.4: ARIMA model output for detrended EIA dataset from R.

CURRICULUM VITAE

Name Surname: Tuğba Yasemin Karagöz

EDUCATION:

- **B.Sc:** 2019, Istanbul Sabahattin Zaim University, Industrial Engineering

PROFESSIONAL EXPERIENCE AND REWARDS:

- Production Engineer, Erkul Cosmetics, Istanbul (2020-2021)
- Resarch Asisstant, Uskudar University, Industrial Engineering Department (2021-2022)
- Research Assistant, Istanbul Technical University, Management Engineering Department (2022-present)
- Researcher, EELISA innoCORE (EELISA INNOvation and COmmon REsearch Strategy) Project (Support for the Research and Innovation Dimension of European Universities" (Horizon 2020 – Science with and for Society program) (2023-2024)
- Researcher, European Engineering Learning Innovation and Science Alliance-EELISA (European Union Erasmus+ Program "Key Action-2 Innovation and Good Practices Cooperation for Exchange-European Universities") (2024-present)

PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:

- **T. Y. Karagöz** and **E. Işıklı**, "Zaman Serisi Tahmin Kombinasyonları İçin Çok Kriterli Karar Verme Yöntemine Dayalı Yeni Bir Yaklaşım" presented at the International Data Science and Statistics Congress IDSSC 2024, Ankara and the study was published in the proceeding book of the relevant congress.

OTHER PUBLICATIONS, PRESENTATIONS AND PATENTS:

- Reyhanoğlu İ., Sidal, H. B., & **Karagöz, T. Y.,** Emgili D. (2023). "Fuzzy Cognitive Mapping Approach for Project Evaluation Considering Sustainability: Application in Consumer Electronic Sector", *Opportunities and Challenges of Engineering Applications in the Digital Age*, Nobel Akademik Yayıncılık, 305-326.
- Çiftçi, F. S., **Karagöz, T. Y.**, Bulak, M. E., Taşer, S., & Kutty, A. A. (2024). "Çok Boyutlu Yaklaşım ile Ürün Performans Ölçümü: Deneysel Bir Çalışma." Düzce Üniversitesi Bilim ve Teknoloji Dergisi, 12(3), 1250-1266.
- Karabacak, S., Sidal, H. B., **Karagöz, T. Y.** (2024). "Capacity Needs For Carbon Fiber Production In Turkey Considering Future Demand For Offshore Wind Turbines." Osmaniye Korkut Ata Üniversitesi İktisadi Ve İdari Bilimler Fakültesi Dergisi, 8(2), 1-18.