SEASONAL ADJUSTMENT OF HIGH-FREQUENCY TIME SERIES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF APPLIED MATHEMATICS
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

İLAYDA TOSUN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
FINANCIAL MATHEMATICS

APRIL 2023

Approval of the thesis:

SEASONAL ADJUSTMENT OF HIGH-FREQUENCY TIME SERIES

submitted by **İLAYDA TOSUN** in partial fulfillment of the requirements for the degree of **Master of Science in Financial Mathematics Department, Middle East Technical University** by,

Prof. Dr. A. Sevtap SELÇUK KESTEL
Dean, Graduate School of **Applied Mathematics**

Prof. Dr. A. Sevtap SELÇUK KESTEL
Head of Department, **Financial Mathematics**

Prof. Dr. Ceylan YOZGATLIGİL
Supervisor, **Statistics, METU**

Examining Committee Members:

Prof. Dr. A. Sevtap SELÇUK KESTEL
Department of Actuarial Sciences, METU

Prof. Dr. Ceylan YOZGATLIGİL
Department of Statistics, METU

Prof. Dr. Seher Nur SÜLKÜ
Department of Econometrics, Hacı Bayram Veli University

Date:

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: İLAYDA TOSUN

Signature :

ABSTRACT

SEASONAL ADJUSTMENT OF HIGH-FREQUENCY TIME SERIES

Tosun, İlayda
M.S., Department of Financial Mathematics
Supervisor : Prof. Dr. Ceylan YOZGATLIGİL

April 2023, 103 pages

Seasonal adjustment is a statistical methodology employed to eliminate the influence of regular cyclic patterns inherent in time series data. Seasonal adjustment is essential for accurately interpreting and analyzing time series, particularly in macroeconomics, where understanding long-term trends and patterns is crucial for making informed decisions. In recent years, there has been a substantial surge in the availability of high-frequency time series data, which pertains to data collected at very short intervals. Seasonal adjustment of high-frequency time series poses unique challenges due to their tendency to exhibit high levels of noise and volatility. As a result, accurately identifying and removing seasonal effects becomes more challenging. New methods are being developed on this subject. This study aims to assess a diverse range of data sets possessing different frequencies and characteristics, employing innovative approaches. Through the utilization of techniques such as naive, Multivariate Seasonal Trend Decomposition using Loess (MSTL), Seasonal-Trend Decomposition using Regression (STR), and Daily Seasonal Adjustment (DSA), the seasonal adjustment of time series exhibiting intricate seasonal patterns at hourly, daily, and weekly frequencies is undertaken. In particular, when analyzing hourly data, the STR method shows outstanding results, while for daily data, the DSA method performs better based on the four datasets used in the study. In general, both the MSTL and STR methods have shown impressive results.

Keywords: Time Series Analysis, Seasonal Adjustment, High-Frequency Time Series

ÖZ

YÜKSEK FREKANSLI ZAMAN SERİLERİNDE MEVSİMSEL DÜZELTME

Tosun, İlayda
Yüksek Lisans, Finansal Matematik Bölümü
Tez Yöneticisi : Prof. Dr. Ceylan YOZGATLIGİL

Nisan 2023, 103 sayfa

Mevsimsel düzeltme, zaman serilerinden düzenli, periyodik dalgalanmaların etkilerini kaldırmak için kullanılan istatistiksel bir tekniktir. Mevsimsel düzeltme, özellikle makroekonomide uzun vadeli eğilimleri ve desenleri anlamak, bilinçli kararlar almak için oldukça önemlidir, bu nedenle zaman serilerini doğru bir şekilde yorumlamak ve analiz etmek için gereklidir. Son yıllarda, çok kısa aralıklarla toplanan veriler olan yüksek frekanslı zaman serisi verilerinin sayısında önemli bir artış olmuştur. Yüksek frekanslı zaman serilerinin mevsimsel düzeltilmesi, yüksek düzeyde gürültü ve oynaklık sergileme eğilimleri nedeniyle zorluklar ortaya çıkarır. Sonuç olarak, mevsimsel etkileri doğru bir şekilde belirlemek ve ortadan kaldırmak daha zor hale gelir. Bu konuda yeni yöntemler geliştirilmektedir. Bu çalışma, yenilikçi yaklaşımlar kullanarak, farklı frekans ve özelliklere sahip çeşitli veri setlerini değerlendirmeyi amaçlamaktadır. Naif, Loess ile Çok Değişkenli Mevsim-Trend Ayrıştırma (MSTL), Regresyon ile Mevsim-Trend Ayrıştırma (STR) ve Günlük Mevsimsel Düzeltme (DSA) gibi teknikler kullanılarak, saatlik, günlük ve haftalık frekanslardaki kompleks mevsimsel yapıya sahip zaman serilerinin mevsimsel düzeltmesi gerçekleştirilmiştir. Özellikle saatlik veriler analiz edilirken, STR yöntemi güzel sonuçlar gösterirken, günlük veriler için DSA yöntemi, çalışmada kullanılan dört veri setine dayalı olarak daha iyi performans göstermiştir. Genel olarak, MSTL ve STR yöntemleri tüm serilerde etkileyici sonuçlar göstermiştir.

Anahtar Kelimeler: Zaman Serisi Analizi, Mevsimsel Düzeltme, Yüksek Frekanslı Seriler

To Yağız and Ali

ACKNOWLEDGMENTS

I would like to extend my sincere gratitude to my supervisor Prof. Dr. Ceylan Yozgatlıgil, for her invaluable contributions, including sharing her expertise, experience, valuable feedback, and constant support and patience throughout my study.

I must express my utmost appreciation to Dr. Özlem Yiğit, the head of the Methodology Department at TURKSTAT, for her guidance, advice, and support.

I also want to express my thanks to the head of the Data Analysis Techniques Group, Fatma Aydan Kocacan Nuray, for always supporting me.

I would like to express my gratitude to my colleagues, Dr. Gülsüm Merve Gökçin, Dr. Muhammed Fatih Tüzen, and Dr. Osman Sert, for their unwavering support, cooperation, and friendship throughout our work together. They assisted me in every way possible, and I could not have accomplished as much without their help. I am fortunate to have the privilege of collaborating with knowledgeable and kind-hearted coworkers.

I offer my endless gratitude to Sena Kihtır, Vuslat Nur Şahin Temel, and Esra Gölcük for their support and invaluable friendships throughout my life.

I would also like to heartfelt thank my colleagues, Ezgi Akçay, Tuğçe Seçil Tekten, Melike Karakoç, and Güvenç Gürbüz for their friendship. Ezgi's moral support was crucial in helping me complete my thesis.

I am grateful to my mother, Hüsniye Kuru, and my father, Sevet Kuru, for their support throughout my life. Also, my sister Ezgi is always there for me. I want to express my heartfelt gratitude to Ahmet Yağız and Ali Yiğit for simply being in my life and being my nephews. Their presence fills me with immense pride as their aunt.

I would like to thank my husband, Fatih Tosun, for his unwavering support and presence throughout my thesis. His presence in my life is a daily blessing; without him, this accomplishment would not have been possible. I feel honored that he chose me to be his partner for life.

And finally, I would like to thank everyone who supported and believed me in every aspect during the thesis.

TABLE OF CONTENTS

ABSTRACT vii

ÖZ ix

ACKNOWLEDGMENTS xi

TABLE OF CONTENTS xiii

LIST OF TABLES xvii

LIST OF FIGURES xviii

LIST OF ABBREVIATIONS xxiii

CHAPTERS

1 INTRODUCTION 1

2 LITERATURE REVIEW 5

2.1 Non-Parametric Methods 6

2.2 Semi-Parametric Methods 8

2.3 Parametric Methods 9

2.4 High-Frequency Methods 11

3 METHODOLOGY 15

3.1 Naive Method 15

3.2 MSTL 16
3.3 DSA (Daily Seasonal Adjustment) 18
3.4 STR (Seasonal-Trend decomposition using Regression) 19
3.5 Data Pre-Processing 20
3.5.1 Imputation 20
3.5.2 Anomaly Detection 21
3.5.3 Calendar Effect 21
3.5.4 Seasonal Test 21
4 ANALYSIS 23
4.1 Data Introduction 23
4.2 Data Pre-Processing 25
4.3 Analysis and Results 25
4.3.1 Weekly Time Series 25
4.3.1.1 Weekly Economic Index 26
4.3.2 Daily Time Series 33
4.3.2.1 Daily Electricity Consumption in Türkiye . . 33
4.3.3 Hourly Time Series 42
4.3.3.1 Hourly Electricity Consumption in Türkiye . . 42
4.3.4 Results 49
5 CONCLUSION AND FUTURE WORK 53
REFERENCES 57

6 APPENDIX 63

6.1 Weekly Time Series 63

6.1.1 The US Weekly Rail Traffic 63

6.1.2 Weekly GDP Tracker of Türkiye 70

6.2 Daily Time Series 76

6.2.1 PM10 in Ankara 76

6.2.2 CO in Bahçelievler 84

6.2.3 Bike Hiring in Washington 91

6.3 Hourly Time Series 98

6.3.1 Hourly Bike Hiring in London 98

LIST OF TABLES

Table 4.1 Description of Datasets Used in the Study 23

Table 4.2 Calender Effect on Daily Electricity Consumption in Türkiye 38

Table 4.3 Results of Seasonality Tests for Each Method on All Datasets 49

Table 4.4 Performance of Seasonal Adjustment Methods on All Datasets 51

Table 6.1 Calender Effect on PM10 in Ankara 80

Table 6.2 Calender Effect on CO in Bahçelievler 87

LIST OF FIGURES

Figure 2.1 Seasonal Adjustment Methods 5

Figure 3.1 Process of MSTL Phase One 17

Figure 3.2 Process of MSTL Phase Two 17

Figure 4.1 Time Series Plot of WEI 26

Figure 4.2 The ACF and PACF Graphs of WEI 27

Figure 4.3 The Box-Plots of WEI Data Monthly and Yearly 27

Figure 4.4 The Anomaly Detection Results of WEI 28

Figure 4.5 The Naive Method Results for WEI 28

Figure 4.6 The MSTL Method Results for WEI 29

Figure 4.7 The STR Method Results for WEI 30

Figure 4.8 The Seasonally Adjusted Series Result of All Methods for WEI 31

Figure 4.9 The Seasonally Adjusted Series ACF-PACF Plots for WEI 32

Figure 4.10 Time Series Plot of Daily Electricity Consumption in Türkiye 33

Figure 4.11 The ACF and PACF Graphs of Daily Electricity Consumption in Türkiye . 34

Figure 4.12 The Box-Plots of Daily Electricity Consumption in Türkiye Weekly, Monthly and Yearly 34

Figure 4.13 The Anomaly Detection Results of Daily Electricity Consumption in Türkiye 35

Figure 4.14 The Naive Method Results for Daily Electricity Consumption in Türkiye . 36

Figure 4.15 The MSTL Method Results for Daily Electricity Consumption in Türkiye . 36

Figure 4.16 The STR Method Results for Daily Electricity Consumption in Türkiye . . 37

Figure 4.17 The Calender Effect for Daily Electricity Consumption in Türkiye 38

Figure 4.18 The DSA Method Results for Daily Electricity Consumption in Türkiye . . 39

Figure 4.19 The Seasonally Adjusted Series Result of All Methods for Daily Electricity Consumption In Türkiye 40

Figure 4.20 The Seasonally Adjusted Series ACF-PACF Plots for Daily Electricity Consumption in Türkiye 41

Figure 4.21 Time Series Plot of Hourly Electricity Consumption in Türkiye 42

Figure 4.22 The ACF and PACF Graphs of Hourly Electricity Consumption in Türkiye 43

Figure 4.23 The Box-Plots of Hourly Electricity Consumption in Türkiye Daiy, Weekly, Monthly and Yearly 43

Figure 4.24 The Anomaly Detection Results of Hourly Electricity Consumption in Türkiye 44

Figure 4.25 The Naive Method Results for Hourly Electricity Consumption in Türkiye 45

Figure 4.26 The MSTL Method Results for Hourly Electricity Consumption in Türkiye 45

Figure 4.27 The STR Method Results for Hourly Electricity Consumption in Türkiye . 46

Figure 4.28 The Seasonally Adjusted Series Result of All Methods for Hourly Electricity Consumption In Türkiye 47

Figure 4.29 The Seasonally Adjusted Series ACF-PACF Plots for Hourly Electricity Consumption in Türkiye 48

Figure 6.1 Time Series Plot of the US Weekly Rail Traffic Data 64

Figure 6.2 The ACF and PACF Graphs of the US Weekly Rail Traffic Data 64

Figure 6.3 The Box-Plots of the US Weekly Rail Traffic Data Monthly and Yearly . . 65

Figure 6.4 The Anomaly Detection Results of the US Weekly Rail Traffic Data 65

Figure 6.5 The Naive Method Results for the US Weekly Rail Traffic Data 66

Figure 6.6 The MSTL Method Results for the US Weekly Rail Traffic Data 66

Figure 6.7 The STR Method Results for the US Weekly Rail Traffic Data 67

Figure 6.8 The Seasonally Adjusted Series Result of All Methods for the US Weekly Rail Traffic Data 68

Figure 6.9 The Seasonally Adjusted Series ACF-PACF Plots for the US Weekly Rail Traffic Data 69

Figure 6.10 Time Series Plot of the Weekly Tracker of GDP for Türkiye 70

Figure 6.11 The ACF and PACF Graphs of the Weekly Tracker of GDP for Türkiye . . 70

Figure 6.12 The Box-Plots of the Weekly Tracker of GDP for Türkiye as Monthly and Yearly 71

Figure 6.13 The Anomaly Detection Results of the Weekly Tracker of GDP for Türkiye 71

Figure 6.14 The Naive Method Results for the Weekly Tracker of GDP for Türkiye . . 72

Figure 6.15 The MSTL Method Results for the Weekly Tracker of GDP for Türkiye . . 72

Figure 6.16 The STR Method Results for the Weekly Tracker of GDP for Türkiye . . . 73

Figure 6.17 The Seasonally Adjusted Series Result of All Methods for the Weekly Tracker of GDP for Türkiye 74

Figure 6.18 The Seasonally Adjusted Series ACF-PACF Plots for the Weekly Tracker of GDP for Türkiye 75

Figure 6.19 Time Series Plot of PM10 in Ankara 76

Figure 6.20 The ACF and PACF Graphs of PM10 in Ankara 77

Figure 6.21 The Box-Plots of PM10 in Ankara as Weekly, Monthly and Yearly 77

Figure 6.22 The Anomaly Detection Results of PM10 in Ankara 78

Figure 6.23 The Naive Method Results for PM10 in Ankara 78

Figure 6.24 The MSTL Method Results for PM10 in Ankara 79

Figure 6.25 The STR Method Results for PM10 in Ankara 79

Figure 6.26 The Calender Effect for PM10 in Ankara 80

Figure 6.27 The DSA Method Results for PM10 in Ankara 81

Figure 6.28 The Seasonally Adjusted Series Result of All Methods for PM10 in Ankara 82

Figure 6.29 The Seasonally Adjusted Series ACF-PACF Plots for PM10 in Ankara . . . 83

Figure 6.30 Time Series Plot of CO in Bahçelievler 84

Figure 6.31 The ACF and PACF Graphs of CO in Bahçelievler 84

Figure 6.32 The Box-Plots of CO in Bahçelievler Weekly, Monthly and Yearly 85

Figure 6.33 The Anomaly Detection Results of CO in Bahçelievler 85

Figure 6.34 The Naive Method Results for CO in Bahçelievler 86

Figure 6.35 The MSTL Method Results for CO in Bahçelievler 86

Figure 6.36 The STR Method Results for CO in Bahçelievler 87

Figure 6.37 The Calender Effect for CO in Bahçelievler 88

Figure 6.38 The DSA Method Results for CO in Bahçelievler 88

Figure 6.39 The Seasonally Adjusted Series Result of All Methods for CO in Bahçelievler 90

Figure 6.40 The Seasonally Adjusted Series ACF-PACF Plots for CO in Bahçelievler . 91

Figure 6.41 Time Series Plot of Bike Hiring in Washington 91

Figure 6.42 The ACF and PACF Graphs of Bike Hiring in Washington 92

Figure 6.43 The Box-Plots of Bike Hiring in Washington Weekly, Monthly and Yearly . 92

Figure 6.44 The Anomaly Detection Results of Bike Hiring in Washington 93

Figure 6.45 The Naive Method Results for Bike Hiring in Washington 93

Figure 6.46 The MSTL Method Results for Bike Hiring in Washington 94

Figure 6.47 The STR Method Results for Bike Hiring in Washington 94

Figure 6.48 The DSA Method Results for Bike Hiring in Washington 95

Figure 6.49 The Seasonally Adjusted Series Result of All Methods for Bike Hiring in Washington 96

Figure 6.50 The Seasonally Adjusted Series ACF-PACF Plots for Bike Hiring in Washington 97

Figure 6.51 Time Series Plot of Hourly Bike Hiring in London 98

Figure 6.52 The ACF and PACF Graphs of Hourly Bike Hiring in London 98

Figure 6.53 The Box-Plots of Hourly Bike Hiring in London Daiy, Weekly, Monthly and Yearly 99

Figure 6.54 The Anomaly Detection Results of Hourly Bike Hiring in London 100

Figure 6.55 The Naive Method Results for Hourly Bike Hiring in London 100

Figure 6.56 The MSTL Method Results for Hourly Bike Hiring in London 101

Figure 6.57 The STR Method Results for Hourly Bike Hiring in London 101

Figure 6.58 The Seasonally Adjusted Series Result of All Methods for Hourly Bike Hiring in London 102

Figure 6.59 The Seasonally Adjusted Series ACF-PACF Plots for Hourly Bike Hiring in London 103

LIST OF ABBREVIATIONS

ACF	Autocorrelation Function
adj	Seasonal Adjusted
AMB	ARIMA Model-Based
ARIMA	AutoRegressive Integrated Moving Average
ATM	Automatic Teller Machine
CO	Carbon Monoxide
CPI	Consumer Price Index
DSA	Daily Seasonal Adjustment
ECB	European Central Bank
GADMM	Generalized Alternating Direction Method of Multipliers
GESD	Generalized Extreme Studentized Deviate Test
GDP	Gross Domestic Product
GLAS	Generalized Log-Additive Seasonal Decomposition
IQR	Inner Quartile Range
MBA	Model-Based Approaches
MSTL	Multivariate Seasonal Trend Decomposition using Loess
OECD	Organisation for Economic Cooperation and Development
PACF	Partial Autocorrelation Function
PM10	Particulate Matter with a Diameter of 10 Micrometers or Less
QS	Quenouille-Stockwell
RegARIMA	Regression with ARIMA Errors
Robust-STP	Seasonal-Trend Decomposition Method for Partial Periodic Time Series
SABL	Seasonal Adjustment at Bell Laboratories
SEASABS	SEASonal Analysis at Australian Bureau of Statistics
SEATS	Seasonal Adjustment with Exponential and Automatic Trend-Cycle-Seasonal-Irregular decomposition
STAMP	Structural Time Series Analysis of Multiple Phenomena
STL	Seasonal-Trend decomposition based on Loess

STD	Seasonal-Trend-Dispersion
STR	Seasonal-Trend Decomposition using Regression
STS	Structural Time Series
TRAMO-SEATS	Time Series Regression with ARIMA errors- Seasonal Adjustment Software
TBATS	Trikha, Bampslas, Allan, Taylor, and Surgery
TURKSTAT	Turkish Statistical Institute
UCARIMA	Unobserved Component ARIMA
unadj	Not Seasonal Adjusted
WEI	Weekly Economic Index

CHAPTER 1

INTRODUCTION

Time series exhibit seasonal movements that recur yearly and occur at certain times of the year. Seasonal fluctuations in time series are caused directly or indirectly by factors including weather, calendar influences, and the timing of administrative decisions. Such fluctuations are systematic throughout the year but not always regular [41]. Seasonal adjustment is a statistical technique to remove the effects of regular, periodic fluctuations from time series [33]. This process involves extracting the main signal or trend from the data and estimating any unobserved components that play a role in the overall data pattern. Additionally, seasonal adjustment can be used to make predictions about future data points in the series by using the estimated trends and patterns to forecast future values.

The main goal of seasonal adjustment is to eliminate regular fluctuations in a time series so that the data can be more easily understood and analyzed. Seasonal adjustment is essential for accurately interpreting and analyzing time series, particularly in macroeconomics, where understanding long-term trends and patterns is crucial for making informed decisions. Without removing the influence of seasonal factors, it cannot be easy to interpret the data and comprehend the underlying trends correctly.

According to Persons [65], it is possible to decompose the time series to its separate and unobserved elements. According to the statistical decomposition theory, a time series can be broken down into several components, each representing a different aspect of the data. Thus, once these components of the time series are figured out, the whole series can be better interpreted and analyzed.

Therefore, a time series, Y, can be formulated as;

$$Y = T + C + S + I$$

where,

(i) T is a long-term trend

(ii) C is a cyclical movement

(iii) S is a seasonal movement (regular periodic movement)

(iv) I is a short-term irregular movement

In decomposing a time series into its components, the chosen method will dictate whether the components are combined by adding or multiplying them. For example, in the case of additive decomposition, the individual components are added together to form the final time series. This is typically used when the magnitude of the components is relatively constant over time and the overall trend in the data is relatively smooth. On the other hand, in the case of multiplicative decomposition, the individual components are multiplied to form the final time series. This is typically used when the magnitude of components varies over time and the overall trend in the data is more complex. The choice of additive or multiplicative methods will depend on the characteristics of the data being analyzed and the goals of the analysis.

The concept of decomposition has a long history, with roots in fields such as astronomy and meteorology. Buys Ballot's study in 1987 is often cited as one of the earliest examples of the use of seasonal adjustment techniques [15]. In the years since, many other researchers have contributed to the development and advancement of these techniques, including [60], [18], [19], [20], [12], [8], [40], [52], [22]. These studies have contributed significantly to understanding decomposition methods and their applications in various fields.

Seasonal adjustment methods were initially utilized in the 1920s [47] and 1930s [50] to analyze seasonal economic variables and were based on non-parametric techniques such as moving averages. These early methods were primarily developed through empirical experimentation rather than being based on statistical models. After that, parametric methods for seasonal adjustment began to be developed, which are based on statistical models and make use of assumptions about the underlying data distribution. The primary studies on parametric methods are made in [39], [37], [13], [14], [38] and [30]. Consequently, these parametric methods have become more widely used over time due to their ability to provide more accurate and reliable results.

In recent years, there has been a significant increase in the amount of high-frequency time series, which refers to data collected at very short intervals, due to technological developments and increased awareness of the value of data. While traditional time series data is typically presented in daily, monthly, or yearly intervals, high-frequency time series include data collected at smaller time intervals. For example, data collected at an hourly, minute-level, or second-level frequency can be considered high-frequency time series. High-frequency time series find applications in various fields. Monitoring financial markets and analyzing financial data such as stock prices, exchange rates, and commodity prices are examples where high-frequency time series are extensively used. Additionally, they are used in other domains such as traffic flow, weather conditions, energy consumption, and telecommunications. High-frequency data can be beneficial for understanding short-term trends and patterns. Many fields, such as finance, economics, engineering, and scientific research, use high-frequency time series to capture and analyze detailed temporal dynamics. Since the observations are made frequently, high-frequency time series data is generated quickly. There is a large amount of data generated as a result. Time series with higher frequencies provide more detailed information about the observed process or phenomenon, making it possible to identify smaller patterns, trends, and relationships. High-frequency time series are more volatile and noisy than data recorded less often. It is more sensitive to short-term changes, random market movements, and measurement errors. Interpreting high-frequency time series data can be challenging due to complex dynamics, such as rapid oscillations, irregular patterns, and short-lived anomalies. Advanced statistical and mathematical methods are necessary to extract significant insights. Time series with high frequency are utilized for predictive modeling, including forecasting, discovering anomalies, and recognizing trading opportunities. Advanced statistical and machine learning techniques are utilized to extract valuable temporal information from the data.

However, high-frequency time series have unique challenges for seasonal adjustment. These data sets often have more than one seasonal pattern, noisier and higher volatility levels. This makes it harder to identify and remove seasonal effects accurately. Additionally, traditional seasonal adjustment methods may not be well-suited for high-frequency data, as the time-series modelling methods generally apply to data collected at longer intervals. In addition, for the decomposition of the seasonal component, analysis is performed by assuming that there is simple seasonality in the time series. As a result, specialized techniques and approaches may be needed to adjust seasonal high-frequency time series effectively.

Seasonal components of high-frequency time series commonly have complex patterns. The complex pattern in time series is usually caught in time series with the following features [24]:

(i) high-frequency seasonality

(ii) multiple seasonal periodst

(iii) non-integer seasonality

(iv) multiple calendar effects

These features make it harder to identify and remove seasonal effects accurately. Hence, several seasonal adjustment methods have been developed to remove just one or more seasonal patterns from the time series. Nowadays, for forecasting time series, many different machine learning algorithms are also used. Although some methods developed specifically to model the complicated trend and seasonal behaviours in the series, some of them suggest to remove the trend and seasonalities from the seris and then focus on the forecasting the series [76] [67] [68].

This thesis will examine complex seasonal time series and try to adjust them seasonally by comparing current methods, naive, MSTL, STR, and DSA methods. As far as our knowledge, there has yet to be a study comparing the performances of these seasonal adjustment methods on different seasonality periods such as hourly, daily, and weekly high-frequency data. The main goal of this thesis is to compare the performance of the selected methods for each frequency. This study will guide selecting the correct method for different seasonal patterns in the time series.

The first chapter serves as an introduction. Its main purpose is to present the study's objective. Chapter 2 is a literature review covering the history of seasonal adjustment methods, as well as a newly developing method for adjusting high-frequency time series. In Chapter 3, it can be found a comprehensive explanation of the methodology used, including the selected methods and pre-processing steps. In Chapter 4, the analysis results are presented and interpreted. Lastly, Chapter 5 summarizes the results obtained in this thesis.

CHAPTER 2

LITERATURE REVIEW

Seasonal adjustment is a statistical method to remove the effects of regular, predictable patterns from time series data. These patterns may be caused by factors such as changes in weather, holidays, or other events that occur at regular intervals. In the literature, there are three groups of seasonal adjustment methods which are non-parametric, parametric, and hybrid (semi-parametric) methods. The most common of these methods are shown in Figure 2.1 [23].

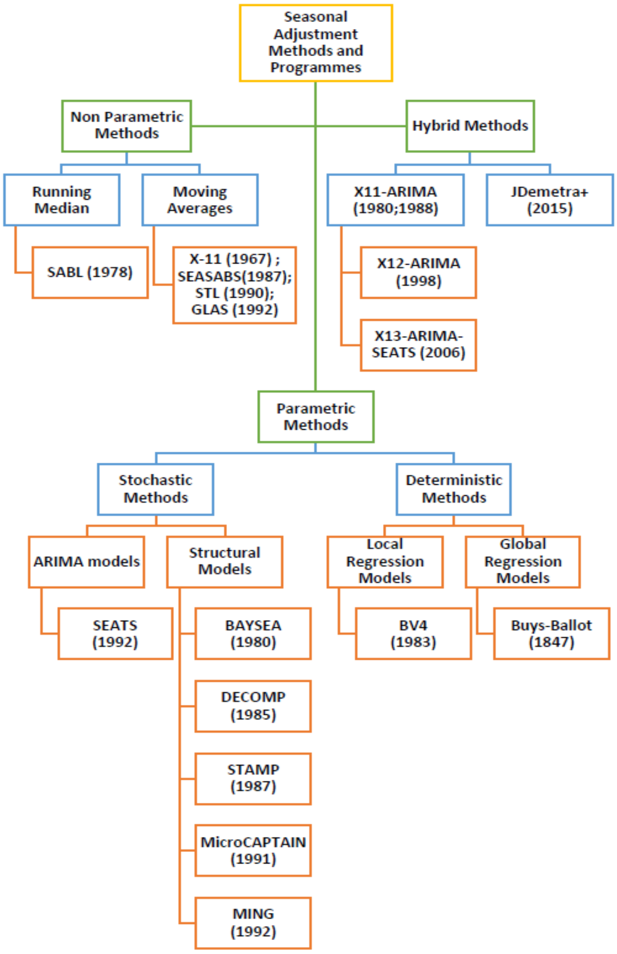


Figure 2.1: Seasonal Adjustment Methods

Seasonal adjustment has a long history dating back to the early 20th century when it was first developed to analyze economic data. One of the earliest methods for seasonal adjustment was the "moving average method," which involved taking the average of a series of data points over a certain period and then adjusting the data to remove the seasonal pattern [28], [48].

2.1 Non-Parametric Methods

In this part, we introduce an epidemic model which is called the vector-host model and apply the theory given within this chapter to analyze its dynamics.

One of the most well-known and widely used non-parametric seasonal adjustment methods is the X-11 method. It is developed by the U.S. Census Bureau in the 1960s [69]. It was used to adjust the Consumer Price Index (CPI) data for the first time. After that, it was used extensively by U.S. government agencies and international organizations to analyze time series and was later adopted by many other countries. X-11 utilizes a technique that combines moving averages and regression to separate the data into its seasonal, trend, and irregular components. Once these components have been estimated, the seasonal component can be removed from the data to obtain the trend-cycle component, which is a smoother representation of the underlying temporal patterns in the data.

This method has been widely used in economics and other fields. Still, it can be computationally intensive when applied to high-frequency data, and it assumes that the data is additive. The X-11 method has been the subject of many studies and improvements over the years, such as the X11-ARIMA, a more sophisticated version of the X-11 method, which is able to handle multiple seasonalities, outliers, and other characteristics of high-frequency time series data.

Another well-known non-parametric seasonal adjustment method is SEASABS (SEASonal Analysis at Australian Bureau of Statistics), developed by the Australian Bureau of Statistics in 1987. SEASABS is similar to X-11 in its use of moving averages and regressions, but it includes additional features such as the ability to handle irregularly spaced data and working day effects. One of the main enhancements is using a more robust algorithm for estimating the seasonal component of the data, which is better able to handle irregular patterns and outliers, like the Australian business survey data. Another enhancement is using a new way of handling missing data, which allows the program to better manage missing observations in the data.

GLAS (Generalized Log-Additive Seasonal Decomposition) is a seasonal adjustment software program developed by the European Central Bank (ECB) [75]. The program is similar to other seasonal adjustment software, such as X-11, X-12-ARIMA, and SEASABS. Still, it includes several unique features that make it well-suited for use with European economic time series data.

GLAS is based on the idea of decomposing a time series into its components using a specific algorithm. The method uses a combination of smoothing techniques, such as the Log-Additive decomposition method and regression analysis, to estimate the seasonal, trend, and irregular components of the data. Once these components have been calculated, the seasonal part can be removed from the data to obtain the trend-cycle component. One of the critical features of GLAS is the implementation of a minimum revision algorithm proposed by Lane [53], which is used to evaluate end-of-sample weights. This algorithm aims to minimize the revisions to the estimates of the trend and seasonal components, resulting in a more stable and reliable estimate.

Cleveland et al. [16] have presented the STL (Seasonal-Trend decomposition based on Loess) method. The concept of non-parametric regression based on the locally weighted average of the data is a core concept common in STL and GLAS. Notably, the trend is adjusted linearly or quadratically, but the seasonal component is adjusted linearly or constantly. STL is based on locally weighted polynomial regression, allowing for flexible trend and seasonal components estimation. It is robust to outliers and handles multiple seasonality, but it is computationally intensive.

Compared to other methods such as X11, X12, SEASABS, and GLAS, the STL method allows for a more flexible way of modelling the trend, seasonal and irregular components and does not require a priori knowledge of the data frequency or time frame. The method is utilized by various researchers and practitioners in the field of time series analysis, specifically in economics. It is also available as an R package called "stl". It is also included in other software and libraries such as "statsmodel" in Python and "fable" [61] and "forecast" [43] packages in R.

The STL method has been widely adopted in practice because of its ability to handle irregular, and non-stationary data and more flexible modelling of the seasonal component. It continues to be improved by the research community to handle more complex cases.

SABL (Seasonal Adjustment at Bell Laboratories) is another non-parametric method developed by Cleveland et al. [17] at Bell Laboratories. It is based on moving averages and regressions, with a non-linear filter based on M-estimation, before applying the linear filter to reach more robustness; it can handle anomalous data (outliers) and working days effect. SABL uses M-estimation, a technique developed by Huber [39], which is robust to outliers and allows for a more accurate estimation of the components of the time series. SABL was a state-of-the-art method when it was first introduced, but it has since been largely superseded by more advanced methods such as X12-ARIMA and SEASABS. These methods incorporate more recent advances in statistical techniques and are able to handle more complex and diverse types of time series data. However, SABL has a unique contribution to the development of seasonal adjustment methods, particularly in handling outliers and anomalies.

In summary, non-parametric methods are a class of techniques that use moving averages, regressions, and other non-parametric approaches to separate the trend and seasonal components of a time series. These methods are widely used in practice due to their simplicity and robustness. Each of these methods has its characteristics and assumptions, and it is vital to choose the appropriate method based on the characteristics of the data and the objective of the analysis. It is worth mentioning that, as technology and computing power have advanced, some of these methods have been improved or replaced by more sophisticated techniques.

2.2 Semi-Parametric Methods

X11-ARIMA, X12-ARIMA, and X13-ARIMA-SEATS are semi-parametric seasonal adjustment methods developed by the United States Census Bureau. These methods are based on the X-11 method, which is a non-parametric approach, but they also incorporate ARIMA (AutoRegressive Integrated Moving Average) models to account for the underlying structure of the time series data.

X11-ARIMA, the first of these methods and was introduced in 1975 by Dagum. The X11-ARIMA/80 is an automated version of the X11-ARIMA method, which Dagum developed in 1980 at Statistics Canada. This version of the method includes a user-friendly interface and additional features. It is based on the X-11 method but includes a pre-processing step that uses an ARIMA model to estimate and remove any underlying structure in the data. This improves the accuracy of the trend and seasonal estimates and reduces the number of revisions required.

What differentiates X11-ARIMA from X-11 is the inclusion of the ARIMA model to estimate the irregular component of the data. This allows X-11-ARIMA to handle missing data and better capture and remove the irregular fluctuations present in the data.

X12-ARIMA is an extension of X11-ARIMA that was introduced by Findley et al. [29] at the U.S. Census Bureau. It includes additional pre-processing steps, such as outlier detection and adjustment, and a more flexible ARIMA model that can handle multiple seasonal periods. It uses the ARIMA model to estimate the irregular and seasonal components, which is the main addition to the X11-ARIMA method. X12-ARIMA also includes additional post-processing steps, such as calculating uncertainty measures and adjusting the data for the effects of trading days and holidays.

X13-ARIMA-SEATS was developed by the United States Census Bureau in 2006. It is an extension of the X12-ARIMA method, and it includes a new seasonal adjustment procedure called SEATS (Seasonal Adjustment with Exponential and Automatic Trend-Cycle-Seasonal-Irregular decomposition). ARIMA is used to model the non-seasonal components of the data, such as trend and irregularity, while SEATS is used to extract the seasonal component of the data. The X13-ARIMA-SEATS method is commonly used in economics and finance to analyze time series data and is implemented in the X13-ARIMA-SEATS software developed by the U.S. Census Bureau [59].

In summary, X11-ARIMA, X12-ARIMA, and X13-ARIMA-SEATS are semi-parametric seasonal adjustment methods developed by the United States Census Bureau. X11-ARIMA and

X12-ARIMA are older methods, while X13-ARIMA-SEATS is the most recent method and combines the previous two methods. All three methods are implemented in the X13-ARIMA-SEATS software, which can be downloaded and used from the U.S. Census Bureau website. National statistical offices and central banks have widely used and accepted these methods and software.

2.3 Parametric Methods

In the 1930s, Fisher [31] and Mendershausen '[57] proposed a technique using polynomial regression models and the least squares method to remove the seasonal component of time series data. This marked the beginning of using model-based approaches (MBA) for seasonal adjustment.

These methods were based on modelling the original series and each component by simple parametric functions, such as polynomials, and estimating parameters using the ordinary least squares method. This allowed for a more accurate and flexible estimation of the different components, especially the seasonal component [36].

Buys Ballot [15] is considered one of the first authors to propose a deterministic seasonal adjustment method. He proposed a method based on global regression, which is a technique that uses a single set of regression coefficients to model the relationship between multiple variables. Buys Ballot's method was based on a trigonometric model that used sine and cosine functions to model the seasonal component of the data.

The European Commission developed DAINTIES in 1979 as a successor to the SEABIRD method. DAINTIES is based on moving regression methods, which use linear regression models with moving windows of data to estimate the seasonal and trend components of the time series. DAINTIES uses only asymmetric filters for filtering the series, which means that it does not require any revisions after the initial adjustment. Still, it does lead to drawbacks such as phase shifts and distortions of the estimator [30].

DAINTIES was, until recently, the official seasonal adjustment method of the European Comission and was widely used by Eurostat for seasonal adjustment of various economic time series data.

Apart from these deterministic methods, there are also parametric stochastic methods. There are two main types of approaches: the ARIMA model-based (AMB) approach and the structural time series (STS) approach. These methods are based on the specification of unobserved component ARIMA (UCARIMA) models and signal extraction techniques [11].

The AMB approach is based on fitting an ARIMA model to the data and using this model to estimate and remove the seasonal component. This approach is relatively simple, but it can be sensitive to the choice of model parameters and may not always provide accurate results for certain types of data. Notable figures in this field include Box et al. [13], Burman [14], Hillmer and Tiao [38], Bell [11], and Maravall and Pierce [56], who have made significant contributions to this field.

One of the main advantages of the AMB approach is its simplicity. Estimating the seasonal component can be done using simple regression techniques, and the seasonal component can be easily removed from the data by subtracting it from the original series.

However, the AMB approach does have some limitations. The choice of the ARIMA model is critical for the accuracy of the seasonal adjustment and can be sensitive to the specific characteristics of the data. In addition, the method may not always provide accurate results for certain types of data, such as data with non-linear trends or irregular patterns. Despite these limitations, the AMB approach is widely used in practice, particularly for series with solid seasonality patterns.

TRAMO-SEATS (Time Series Regression with ARIMA errors- Seasonal Adjustment Software) is a software program developed by the Bank of Spain in the 1990s [34] [35]. It is based on the combination of time series regression and ARIMA models.

TRAMO-SEATS uses a two-step process: the first step is called TRAMO, which is used for cleaning and pre-treatment of the data, and the second step, called SEATS, is used for seasonal adjustment. TRAMO-SEATS is considered one of the most effective and flexible software programs for seasonal adjustment of economic time series data. It is widely used by central banks, government agencies, international organizations, and research institutions. TURKSTAT (Turkish Statistical Institute) also uses this method for seasonal adjustment of statistics.

It is a potent tool for identifying and modeling various types of time series, including nonlinearity, irregularity, multiple breaks, and outliers, which are common characteristics of economic time series. The software is continuously updated and improved. It also has an R interface called "RJDemetra" [66]. With recent developments, the program also seasonal adjusts for the high-frequency time series.

The STS approach is based on a more general framework of structural time series models. It uses a combination of maximum likelihood estimation and Kalman filtering to estimate the parameters of the model and remove the seasonal component. This approach is more flexible and can handle a broader range of data types, but it can be more computationally intensive.

STAMP (Structural Time Series Analysis of Multiple Phenomena) is a software program used to analyze and model time series. It was developed by Koopman et al. [49] at the London School of Economics and Political Science. It uses a combination of maximum likelihood estimation and Kalman filtering to estimate the parameters of the model.

One of the main advantages of STAMP is its flexibility; it can be applied to a wide range of time series and can handle multiple phenomena. It also includes functionalities such as forecasting and simulation.

In conclusion, parametric seasonal adjustment methods have been widely used in time series analysis since their introduction in the 1930s. Using polynomial regression models and the least squares method to remove the seasonal component of time series data has become more effective.

Notable parametric methods are ARIMA model-based (AMB) approach and the structural time series (STS) approach. Both methods have advantages and limitations. The method chosen depends on the data's specific characteristics and the purpose of the analysis.

2.4 High-Frequency Methods

High-frequency data refers to time series data observed at a high rate, such as daily, weekly, or even higher frequencies. One of the main challenges of working with high-frequency data is the presence of strong seasonality patterns, which can make it difficult to identify underlying trends and patterns in the data.

MSTL (Multivariate Seasonal Trend Decomposition using Loess) is a method for decomposing a multivariate time series into its trend, seasonal, and irregular components. It was first proposed by Gebreselassie et al. [9].

The MSTL method applies a locally weighted regression method, precisely the Loess method, to estimate the trend, seasonal, and irregular components of a multivariate time series. The Loess method is a non-parametric regression method that uses a weighted least squares fitting to estimate the trend and seasonal components. It can be applied to both additive and multiplicative models, and it can handle multiple seasonal periods as well. MSTL has been widely used in various fields such as finance, econometrics, and environmental science and is particularly useful when dealing with multivariate time series data.

The forecast R package is a popular high-frequency seasonal adjustment tool [45] [42]. It includes "mstl()" functions that can accomplish high-frequency seasonal adjustment. Moreover, "auto.arima()" in the forecast package is also can use for seasonal adjustment. This function uses dynamic harmonic regression for seasonal adjustment [44].

TBATS (Trikha, Bampslas, Allan, Taylor, and Surgery) is a method for decomposing a univariate time series into its components and forecasting the time series. The TBATS method was first proposed by De Livera et al. in 2011 [24]. The TBATS method combines three different exponential smoothing methods (additive, multiplicative, and exponential) to handle different time series patterns, such as level changes and local time trends. It uses a state space framework that estimates the model's parameters and forecasting. The TBATS model also includes automatic procedures to identify the best model among possible models. It is implemented in open-source and commercial software such as R (in the forecast package), Python, and MATLAB.

The prophet is a package developed by Facebook for time series forecasting [71]. It is mainly designed for high-frequency time series data, such as daily or hourly data, and is available in the R and Python programming languages.

Prophet uses a decomposable additive model with three main components: trend, seasonality, and holidays. The model fits the data using a Bayesian structural time series model. The package uses a Gaussian process to model the non-linearity of the trend component, which allows it to handle data with non-linear trends. The package also provides options for handling missing data and outliers and incorporating external regressors.

Prophet's main advantage is that it is designed to be easy to use and provides a black box seasonal adjustment method focused on forecasting. The package's automated feature engineering and model selection allows for quick and accurate seasonal adjustment of high-frequency time series data. However, the package may not always provide accurate results for specific data types, particularly if the data has complex patterns or features.

Another STL implementation is the *dsa* package in R developed by Bundesbank specifically for daily time series data [62]. The *dsa* package includes several additional functions specifically designed for handling daily data, such as missing values, outliers, and trend/seasonal breaks.

The *dsa* package uses a two-step process for seasonal adjustment: the first step is to decompose the series into its components using the STL function, and the second step is to remove the seasonal component from the original series. The package also includes options for adjusting for trading days and for handling multiple frequencies within the same series.

The main advantage of the *dsa* package is its ability to handle daily time series data specifically. It is particularly suitable for well-behaved series with strong seasonality patterns.

Timmermans et al. [73] used several seasonal adjustment methods, including the X13-ARIMA-SEATS, STL, and TBATS. They compared the performance of these methods. Each method has its strengths and weaknesses; the TBATS method performed the best overall regarding forecasting accuracy. Additionally, they highlight that the seasonal adjustment of high-frequency data is a challenging task.

Ladiray et al. [51] discussed the challenges of seasonal adjustment of daily data and presented a review of different methods for addressing these challenges. They reviewed several popular techniques for seasonal adjustment of daily data, including the X13-ARIMA-SEATS method, the STL method, and the TBATS method. They also discussed the strengths and weaknesses of each method and provided a comparison of their performance in out-of-sample forecast accuracy. While several options are available for seasonal adjustment of daily data, the choice of method will depend on the specific characteristics of the data and the purpose of the analysis.

The STD (Seasonal-Trend-Dispersion) decomposition method is a recent approach for seasonal adjustment of high-frequency time series data, proposed by Dudek [27]. This method decomposes the time series into seasonal, trend, and dispersion components. The decomposition uses a state-space model that accounts for the correlation between the components. Moreover, the model is estimated using Bayesian methods, which allow for flexible and adaptive modelling of the components over time. However, the STD approach may not be appropriate for short-term or noisy data since it needs a long time series to effectively estimate the seasonal and trend components. Additionally, Bayesian estimation can be computationally intensive and require advanced programming skills.

The Robust-STP (Seasonal-Trend Decomposition Method for Partial Periodic Time Series) is a method for seasonal adjustment of high-frequency time series designed to handle partial periodicity and robust to outliers [74]. This method is based on the classical STL method; however, unlike the STL method, Robust-STP can handle partial periodicity by allowing for a variable-length seasonal component. The Robust-STP method is beneficial for time series with partial periodicity and outliers, common in high-frequency data. However, the Robust-STP method can be computationally expensive, especially for large data sets, and may require detailed tuning of the parameters to obtain optimal results.

The OnlineSTL method is a scalable approach to seasonal adjustment of high-frequency time series data, which can process data up to 100 times faster than the classical STL method [58]. This method applies a sliding window to the time series data, which moves over the data in fixed intervals. This method also uses a hierarchical approach to update the seasonal component, which reduces the impact of data shifts and improves the accuracy of the seasonal adjustment. The OnlineSTL method is beneficial for the real-time processing of high-frequency data, such as stock prices, weather data, and social media data. However, the OnlineSTL method may require careful tuning of the window size and other parameters to obtain optimal results for different data types.

In addition, JDemetra+ is a widely used software for seasonal adjustment and time series analysis, particularly in national statistical offices such as the Turkish Statistical Institute. In the upcoming update of JDemetra+, a new module will be introduced, specifically designed for seasonal adjustment of high-frequency time series data. This new module will allow users to adjust time series data with different frequencies and include various seasonal adjustment methods, such as the STL and the Fractional Airline methods [70].

In conclusion, high-frequency data is a challenging area to work with due to the presence of strong seasonality patterns. There are several methods and tools available for performing high-frequency seasonal adjustment. Each method has its own strengths and limitations.

CHAPTER 3

METHODOLOGY

This section explains seasonal adjustment methods for the high-frequency time series used in this study. The choice of seasonal adjustment method depends on the data's nature and the analysis's purpose. In this thesis, the high-frequency time series are being analyzed. The aim is to remove the seasonal component from the data. The simple moving average method (Naive Method), The Multivariate Seasonal Trend Decomposition using Loess (MSTL), Daily Seasonal Adjustment (DSA), and Seasonal-Trend decomposition using Regression (STR) methods are used for seasonal adjustment.

The Naive Method is a straightforward process that takes the average of the data over a fixed window of time to evaluate the seasonal component. It is simple to implement and works well when the seasonal pattern is stable and predictable. The Multiplicative Seasonal-Trend decomposition using LOESS (MSTL) method is a robust and flexible approach that uses a combination of smoothing techniques and decomposition to estimate the seasonal, trend, and irregular components of the data. It is helpful when the seasonal pattern is complex or irregular. The Daily Seasonal Adjustment (DSA) method is similar to MSTL but is specifically designed for daily seasonality. The Seasonal-Trend decomposition using Regression (STR) method is a popular approach that uses regression models to estimate the seasonal and trend components of the data. It is particularly useful when the seasonal pattern is stable over time and can be modeled using regression.

Additionally, the data pre-processing part is explained in detail in this section.

3.1 Naive Method

The moving average method for seasonal adjustment of high-frequency time series is simple and widely used. It is a deterministic method that uses a moving average to remove the seasonal component from the original time series.

The basic idea behind this method is to calculate the average value of the series for each season (e.g., for each day or week) over a period of time, typically several years. This average value is then subtracted from the original series for each corresponding season, resulting in a seasonally adjusted series.

The moving average method can be applied to both additive and multiplicative models, depending on the nature of the data. The seasonal component is added or subtracted from the original series in an additive model. In contrast, the seasonal component is multiplied or divided by the original series in a multiplicative model [10].

One of the main advantages of the moving average method is its simplicity and ease of implementation. It is also relatively robust to small data changes and can handle multiple seasonal periods. However, the method does have some limitations. It can be sensitive to outliers and missing values. Furthermore, it may not always provide accurate results for specific data types, such as those with non-linear trends or irregular patterns.

3.2 MSTL

The Multivariate Seasonal Trend Decomposition using Loess (MSTL) method is a powerful technique for decomposing multivariate time series data into its components, specifically trend, seasonal, and irregular components. Moreover, the seasonal component includes different seasonal frequencies. The general representation of the additive time series is formulated as;

$$Y_t = T_t + S_t^{(1)} + S_t^{(2)} + ... + S_t^{(n)} + I_t$$

where T is the trend component, $S^{(i)}$ is seasonal components with different frequencies (daily, weekly, and yearly), and I is the irregular component.

The MSTL method was first proposed by Gebreselassie et al. in 2011 [9] and has since been widely used in various fields such as finance, econometrics, and environmental science.

The MSTL method is based on a locally weighted regression method, precisely the Loess method, which estimates the trend, seasonal, and irregular components of a multivariate time series. The Loess method is a non-parametric regression method that uses a weighted least squares fitting approach to estimate the trend and seasonal components. It can be applied to additive and multiplicative models and handle multiple seasonal periods.

For algorithm phases for the MSTL method can be summarized into four phases [9];

Phase one begins by specifying the periods of each seasonal component that wanted to extract (e.g., daily, weekly, etc.). For example, to extract the seasonal components from an hourly time series, the process begins by iterating through each period, starting with the shortest period (e.g. daily) and moving on to the longest period (e.g. yearly). On each iteration, use STL to extract the seasonal component. The extracted seasonal component is then subtracted from the time series, and the process is repeated until all seasonal components have been extracted. This results in a deseasonalized time series. Phase one is summarized in Figure 3.1.

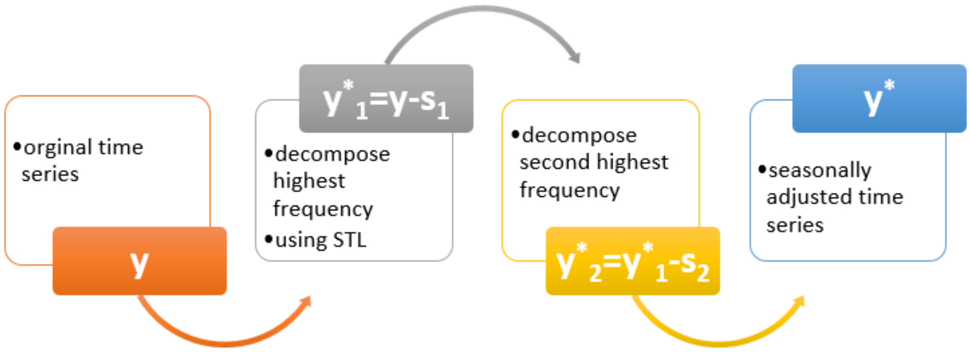


Figure 3.1: Process of MSTL Phase One

In the second phase, purify each extracted seasonal component by adding back to the fully deseasonalized time series from phase one. And then re-extract the same component using STL. This approach helps in refining the extracted seasonal components and ensures that STL is able to re-capture any parts of the seasonal component that might have been missed in the first phase. This is because, in phase one, it is possible that some of the seasonal components may interfere with each other and make it harder for STL to capture them all accurately. By isolating each seasonal component and re-extracting it separately, the algorithm is able to capture the details of each component. This results in more accurate estimates of the seasonal components and a more accurate overall time series decomposition. Phase two is summarized in Figure 3.2.

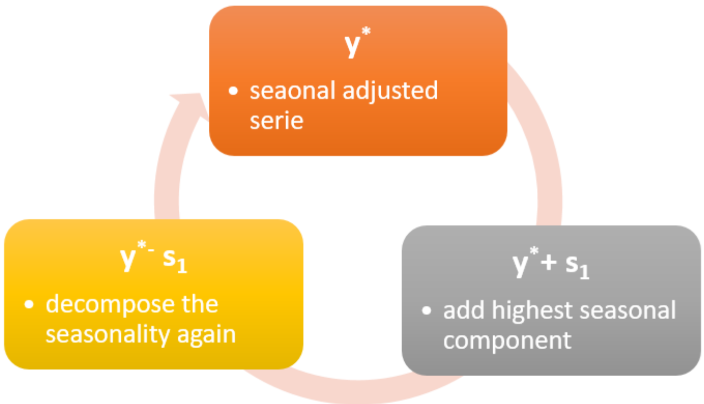


Figure 3.2: Process of MSTL Phase Two

After extracting and removing the seasonal components, the trend component is extracted in phase three. A smoothing technique, Loess, is used to extract the trend component. The deseasonalized time series is passed to the Loess function, which fits a smooth curve to the data. This smooth curve represents the estimated trend of the time series. The trend component is obtained by calculating the fitted values of the Loess function.

The trend component is subtracted from the deseasonalized time series in the fourth phase to obtain the residuals. The residuals are the remainder of the time series that is not explained by the trend and seasonal components.

The trend component, seasonal components, and residuals are then combined to reconstruct the original time series.

One of the main advantages of the MSTL method is its ability to handle multivariate time series data, which allows for the simultaneous analysis of multiple time series with similar patterns. Additionally, the MSTL method is robust to outliers and can handle multiple seasonal periods, making it a valuable tool for dealing with high-frequency time series data.

Additionally, the MSTL method has some advantages over other multivariate decomposition methods. One significant advantage is its flexibility, as it can handle both additive and multiplicative models and multiple seasonal periods. Furthermore, it also addresses some of the limitations of univariate decomposition methods, as it considers the cross-correlation among multiple variables in the decomposition.

Another advantage of the MSTL method is its robustness. It is less sensitive to outliers and data with missing values than other methods. Moreover, it is computationally efficient, as it only requires simple regression techniques, and it can be applied to large datasets and long time series with complex patterns.

3.3 DSA (Daily Seasonal Adjustment)

The Daily Seasonal Adjustment (DSA) method was created by Ollech [63] to adjust daily time series seasonally. This method uses an iterative STL-based seasonal adjustment routine combined with a RegARIMA model to estimate calendar and outlier effects.

The general representation of daily time series is

$$Y = T + S_7 + S_{31} + S + C + I$$

where T is the trend component, C is the calender effect and I is the irregular component.

Unlike other methods, the order of steps in this method is slightly different. The main steps are

1. Deseasonalize weekly effect with STL
2. Calendar and outlier adjustment RegArima
3. Deseasonalize monthly effect with STL
4. Deseasonalize annual effect with STL

It is crucial to account for the effects of the weekdays utilizing the complete data before any analysis. If the moving holiday effects are adjusted before the weekday effects, there is a risk that the estimated effects may contain some of the weekday patterns, leading to biased results.

The STL method has a smoothing parameter sensible for the analysis [16]. This parameter smoothest the sub-periodic seasonal time series, and also sets the number of observations included in the local regression to evaluate the seasonal factor.

Step 1 uses the STL method to get the weekly seasonal factor. For the STL method, a value for the parameter S_7 must be selected for calculating the weekly seasonal factor. As the time series has yet to be outlier adjusted, the robust version of STL is usually preferable.

In Step 2, since the ARIMA model is very computationally expensive for the daily time series, the RegArima model is used [33]. The order of the ARIMA model can be specified by the user or determined through automatic model detection. The Hyndman-Khandakar algorithm is used for automatic model detection [45].

A similar algorithm of the "tsoutlier()" function in R is used for the outlier detection [55]. This algorithm is developed for daily time series. The outlier detection algorithm consists of two parts. In the first part, possible outliers are identified at each time point by analyzing the residuals of an ARIMA model. In the second part, all potential outliers are included as regressors in a final ARIMA model, and those that do not exceed a pre-defined threshold for the t-statistics are removed from the list of candidates.

Step 3 uses the STL method to get the monthly seasonal factor. Since the number of days in a month changes between 28 and 31, the average is 30.4. For the STL method, a value for the parameter S_{31} must be selected to ensure a good fit for calculating the monthly seasonal factor.

Step 4 uses the STL method to get the annual seasonal factor. The number of days in a year is either 365 or 366. Since the leap year effect is included in Step 2, the parameter S_{365} must be selected to ensure a good fit for calculating the annual seasonal factor.

For this method, the "dsa" package in R is designed by Ollech [62]. Since it is designed for daily seasonality, it cannot be used for the series with other seasonal frequencies.

3.4 STR (Seasonal-Trend decomposition using Regression)

STR (Seasonal-Trend decomposition using regression) method is used for time series decomposition, which allows for incorporating multiple seasonal and cyclic components, covariates, and complex seasonal patterns with non-integer periods or complex topology. It is a robust method suitable for handling various data types and can produce more accurate and interpretable results than other decomposition techniques [25].

$$y_t = T_t + \sum_{i=1}^{I} S_t^{(i)} + \sum_{p=1}^{P} \phi_{p,t}, z_{t,p} + R_t$$

where T_t is a smoothly changing trend, $S_t^{(i)}$ are smoothly changing seasonal components with possibly complex topology, $z_{t,p}$ are covariates with coefficients $\phi_{p,t}$ which may be time-varying and even seasonal, R_t is the "remainder".

The STR method introduces an innovative approach by transforming the estimation process with matrix notations into a linear model similar to ridge regression.

This method also has smoothing parameters; the cross-validation formula in linear regression is used to estimate these parameters.

$$CV = \sum_{i=1}^{n} \left(\frac{y_i - \hat{y}_i}{1 - h_{ii}} \right)^2$$

Minimizing CV, the STR method locates the optimal λ and θ smoothing parameters where y_i is the ith element of the vector y, $\hat{y}_i$ is the i^{th} element of the vector $\hat{\mathbf{y}} = \mathbf{H}\mathbf{y}$, and h_{ii} is the i^{th} diagonal element of the hat matrix $H = X\left(X'X\right)^{-1}X'$.

To calculate this parameter "optim()" function from the stats package in R is used [72].

For this method, the "stR" package in R is designed by Dokumentov and Hyndman [26].

3.5 Data Pre-Processing

High-frequency time series have more difficulties than low-frequency data because of the high number of observations and volatility. It contains more missing observations and outliers. The calendar effect is also a more vital issue for high-frequency time series than low-frequency. The calendar effect is also a more crucial issue to high-frequency time series than low-frequency. Because calendar effects directly affect observations [64].

To satisfactorily seasonal adjust the time series, some pre-processing steps are necessary, such as imputation, anomaly detection, and calendar adjustment.

3.5.1 Imputation

The linear interpolation method is used to impute the missing observations by fitting a linear equation between known data points. It is a commonly used method for handling high-frequency time series data if the number of missing data is low.

The method assumes that the data points are evenly spaced in time and that the underlying trend is linear. The interpolated data points are calculated by fitting a straight line between the known data points and then estimating the values of the missing data points based on this line.

The linear interpolation method has several advantages, such as its simplicity and computational efficiency. It does not require any assumptions about the underlying distribution of the data and can be easily implemented using basic linear algebra techniques.

Time series with missing values in this study's data set impute by using Python.

3.5.2 Anomaly Detection

Anomaly detection is identifying abnormal or unusual observations in a dataset that differ from the usual patterns. It is an essential step in various applications. A well-known package for anomaly detection in R is "anomalize," which offers a set of functions for detecting anomalies in univariate and multivariate time [21]. The package comprises techniques like the moving median, moving average and Gaussian process. It also includes visualization tools to assist in identifying and analyzing the anomalies.

There are two methods for anomaly detection in this package: IQR (Inner Quartile Range) and GESD (Generalized Extreme Studentized Deviate Test). In this study, the GESD method, which introduced by Iglewicz and Hoaglin in 1993, is used [46].

There are two different parameters in the anomalize() function to decide which points are an outlier, "alpha" and "max_anoms". The "alpha" parameter controls the width of the confidence interval for identifying anomalies, while the max_anoms parameter controls the maximum percentage of data that can be identified as anomalies. Since high-frequency data is more volatile than monthly or quarterly data, the default value of these parameters is unsuitable for high-frequency data. In the analysis, these parameters are decreased for hourly time series.

3.5.3 Calendar Effect

The daily time series has a calendar effect, including the moving holiday effect. While national and public holidays vary across different countries and regions, most seasonal adjustment software typically includes commonly observed holidays in their respective areas, especially in Europe. But Türkiye has two moving holidays, Ramadan and Sacrifice Feasts, and different national holidays, April 23, May 19, etc. So, the calendar regressor for Türkiye starting from 2015 is created.

3.5.4 Seasonal Test

The Friedman Test

The Friedman Test is a statistical method utilized to determine whether a time series displays seasonal patterns. This test was developed by Milton Friedman in 1937 [32].

The Friedman test is a non-parametric method with the null hypothesis of no stable seasonality. The time series is ranked across all periods. By combining the ranks for a period across all years and dividing them by the total number of years, the mean rank for each period is calculated. Then test statistics and critical value are calculated. The p-value of the Friedman test can be calculated using the cumulative distribution function of the chi-squared distribution with k-1 degrees of freedom where k represents the number of groups or treatments being compared in the Friedman test. If the test statistic exceeds the critical value, then the null hypothesis (no seasonality in the data) is rejected.

Since the Friedman Seasonality Test is non-parametric, no presumptions are made regarding the data's underlying variance distribution. Additionally, it is a robust test, meaning it is unaffected by data with outliers or deviations from the normal distribution.

The QS Test

The QS (Quenouille-Stockwell) test is a statistical test used to evaluate the presence of seasonality in time series data. Specifically, this test checks whether there is significant positive autocorrelation in the data at seasonal lags, indicating that the time series exhibits a regular pattern of variation over time.

The QS (Quenouille-Stockwell) test is a variation of the Ljung-Box test used to assess seasonal autocorrelation in time series data [54]. Unlike the Ljung-Box test, which considers all lags of autocorrelation, the QS test only examines positive autocorrelations at seasonal lags.

Under the null hypothesis of independence of autocorrelation at seasonal lags, the test statistic follows a chi-square distribution.

CHAPTER 4

ANALYSIS

In this chapter some chosen high-frequency time series with different time frequencies and seasonalities are seasonally adjusted using the methods mentioned in Chapter 3. Python with version 3.10 and largely R Studio with version 4.2.2 is used for analysis.

The main steps for seasonal adjustment of high-frequency time series are abridged in order. Firstly, the time series data set is presented. Then, some pre-processing steps, like the imputation of missing values, anomaly detection, and calendar adjustment, are done. After that, time series are suitable for seasonal adjustment. Lastly, the residual seasonality is examined.

4.1 Data Introduction

High-frequency time series data refers to data that is collected at concise time intervals such as seconds, minutes, hours, or days. These data points are collected much faster than traditional time series data, usually monthly or quarterly. High-frequency time series data is commonly used in finance, telecommunications, and transportation, where the ability to capture rapidly changing events is critical. High-frequency time series data is characterized by its large volume, high velocity, and high variety.

In this study, time series at different frequencies were used for diversity. The time series in the data set with their frequency and number of observations are given in Table 4.1.

Table 4.1: Description of Datasets Used in the Study

Data Name	**Abbreviation**	**Start Date**	**End Date**	**Number of Oservation**	**Frequency**
Rail Traffic	rail_traffic	02.01.2010	25.02.2023	687	Weekly
Weekly Tracker	weekly_tracker	04.01.2004	19.02.2023	999	Weekly
Weekly Economic Index	WEI	05.01.2008	18.02.2023	790	Weekly
PM10	pm10	01.01.2010	01.01.2023	4749	Daily
Electricity Consumption	elec_cons_daily	01.01.2016	31.12.2022	2557	Daily
Bike Sharing	bike_sharing	01.01.2011	31.12.2012	731	Daily
CO	co	01.01.2010	09.08.2022	4604	Daily

Electricity Consumption	elec_cons_hourly	01.01.2016 00:00	31.12.2022 23:00	61368	Hourly
London Bike Sharing	london_bike	04.01.2015 00:00	03.01.2017 23:00	17544	Hourly

Electricity consumption data represents hourly real-time electricity consumption of Türkiye, which Turkey Electricity Transmission Inc. collects. In this study, this dataset is used both hourly and daily seasonality. Daily electricity consumption shows the average daily consumption since it is aggregated daily by average [6].

PM10 refers to particulate matter with a diameter of 10 micrometres or less. These particles are small enough to be inhaled into the lungs and can have adverse health effects, including respiratory and cardiovascular diseases. PM10 particles are emitted from various sources, including construction, transportation, and industrial activities, as well as natural sources, such as dust and wildfires. PM10 data is calculated from Ankara, Bahçelievler, and obtained from Turkey National Air Quality Monitoring Network [7].

CO in air pollution data refers to the concentration of carbon monoxide in the air. Carbon monoxide is a colorless and odorless gas produced by the incomplete combustion of fuels such as gasoline, natural gas, and coal. It harms human health as it can reduce the amount of oxygen in the bloodstream. CO data is calculated from Ankara, Bahçelievler, and obtained from Turkey National Air Quality Monitoring Network [7].

The bike-sharing dataset for Washington is a valuable resource for analyzing bike rental trends and patterns in the nation's capital. It includes daily data for two years (2011 and 2012), with information on bike rental usage. Bike-sharing dataset is obtained from the Center for Machine Learning and Intelligent Systems at the University of California [5].

Likely the bike-sharing dataset, the hourly bike-sharing dataset of London is used for this study. It is also available for two years (2015 and 2016). This data is obtained from Transport for London [2].

The US Weekly Rail Traffic dataset provides a broad view of the state of the US economy by tracking the volume of rail traffic across the country. This dataset includes data on weekly rail traffic for various commodities such as chemicals, metals, coal, and grain, as well as intermodal traffic, which consists of the movement of shipping containers and truck trailers on railroads [3].

The Weekly Tracker of GDP Growth dataset, provided by the Organisation for Economic Co-operation and Development (OECD), is an extensive collection of data that shows insights into the state of the global economy. The dataset follows the weekly growth of real gross domestic product (GDP) for a range of OECD countries, including the United States, Canada, Japan, and Türkiye [1].

The Weekly Economic Index dataset, supplied by the Federal Reserve Bank of New York, is a comprehensive resource for real-time monitoring of changes in the US economy. The dataset combines a range of indicators, including consumer behaviour, labor market conditions, and financial market activity, into a single index that provides a timely snapshot of economic activity. The dataset is updated weekly, providing up-to-date information on economic conditions, and can be used to monitor changes in economic activity as they happen [4].

4.2 Data Pre-Processing

In order to perform a successful seasonal adjustment on a time series, certain preliminary measures are required, including imputation, identification of anomalies, and adjustment for calendar effects.

Imputation

The linear interpolation technique replaces missing data points by fitting a linear equation between two known data points. This approach is frequently used for managing high-frequency time series data.

The time series with missing values are imputed using Python with *"scipy.interpolate"* libray.

Anomaly Detection

The GESD method was employed to detect anomalies in time series. The R package *"anomalize"* is a popular tool for identifying anomalies, and it provides a range of functions for detecting anomalies in both univariate and multivariate time series data.

Calendar Effect

In this study, the created calendar regressor for Türkiye, defined in Chapter 3, is used for calendar adjustment. Since it is a complicated analysis, the calendar adjustment is made only for the DSA method because the DSA method uses the RegARIMA method for calendar adjustment specifically.

4.3 Analysis and Results

In this section, the data sets and analysis are discussed. Subsections are divided according to time series frequencies in ascending frequency order, weekly to hourly.

4.3.1 Weekly Time Series

For the seasonal adjustment of the weekly time series naive method, the MSTL method and STR method are applied. Since the time series has a weekly period, the expected seasonal components are monthly and yearly. The monthly seasonal component refers to the week effect in a month, which appears four or five times a month, while the annual effect refers to the week effect in a year, which appears fifty-two times a year.

The analysis results other than weekly economic index (WEI) data are presented in the appendix for the comfort of reading. The rest of the applications with other data sets are given in Appendix 6.1.

4.3.1.1 Weekly Economic Index

The weekly economic index (WEI) is critical data that provide valuable insights into the state of the economy. This index tracks various economic indicators weekly, such as employment, consumer spending, and industrial production. Moreover, the weekly economic index can be handy for detecting changes in economic activity that may take time to be apparent from monthly or quarterly data.

Figure 4.1 displays a time series plot which is presented to aid in comprehending and interpreting the data.

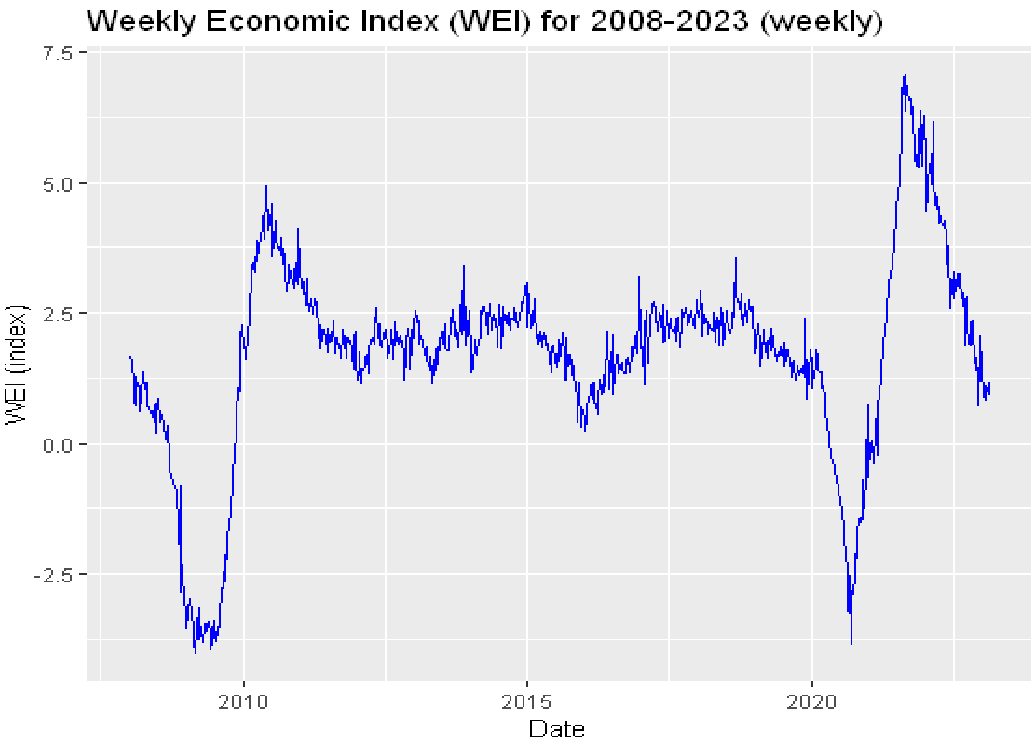


Figure 4.1: Time Series Plot of WEI

The graph shows WEI was affected by the Global Financial Crisis in 2008 and Covid-19 in 2020. Since WEI immediately reflects the impact of the economy and its observations are affected more than low-frequency series, there have been changes in the series' seasonal structure.

Understanding the data structure requires the use of ACF and PACF graphs, so the ACF and PACF graphs of WEI data are illustrated in Figure 4.2.

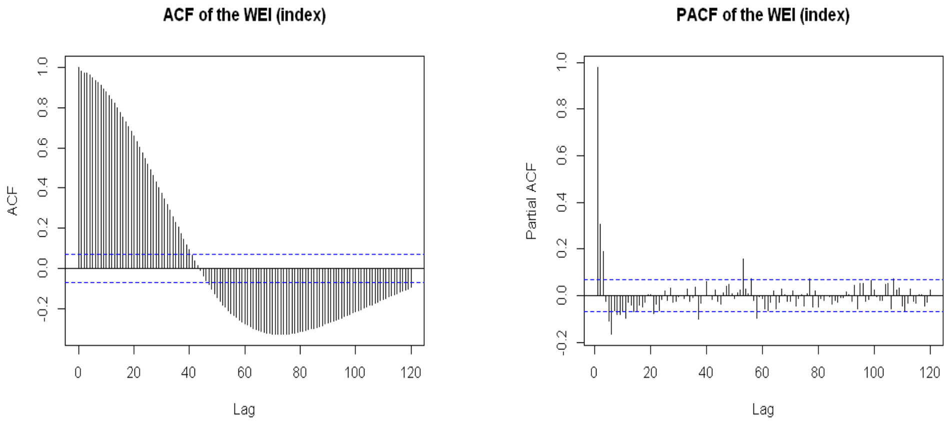


Figure 4.2: The ACF and PACF Graphs of WEI

According to PACF graphs, the lag 4 and its multiples have significant lags. Since one month includes 4 weeks, it may have monthly seasonality.

Box-Plots for monthly and yearly intervals can provide a better understanding of the seasonal data structure.

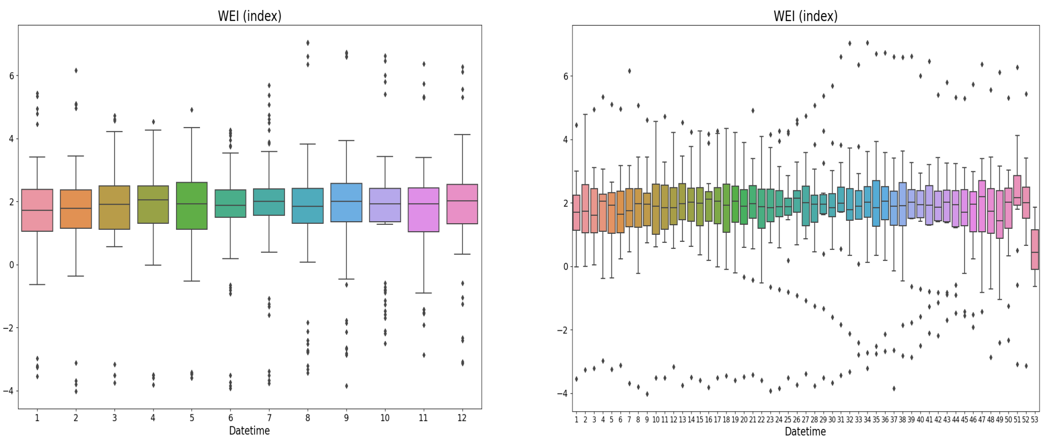


Figure 4.3: The Box-Plots of WEI Data Monthly and Yearly

Figure 4.3 the median of box plots and the upper quartile do not change much month by month, while the lower quartile change in some months. The median change during the year in the yearly plot. As seen in the time series plot, the effects that break the seasonal pattern are also seen in the box plots as outliers.

The data contains anomalies which are illustrated in the anomaly detection outcome displayed in Figure 4.4.

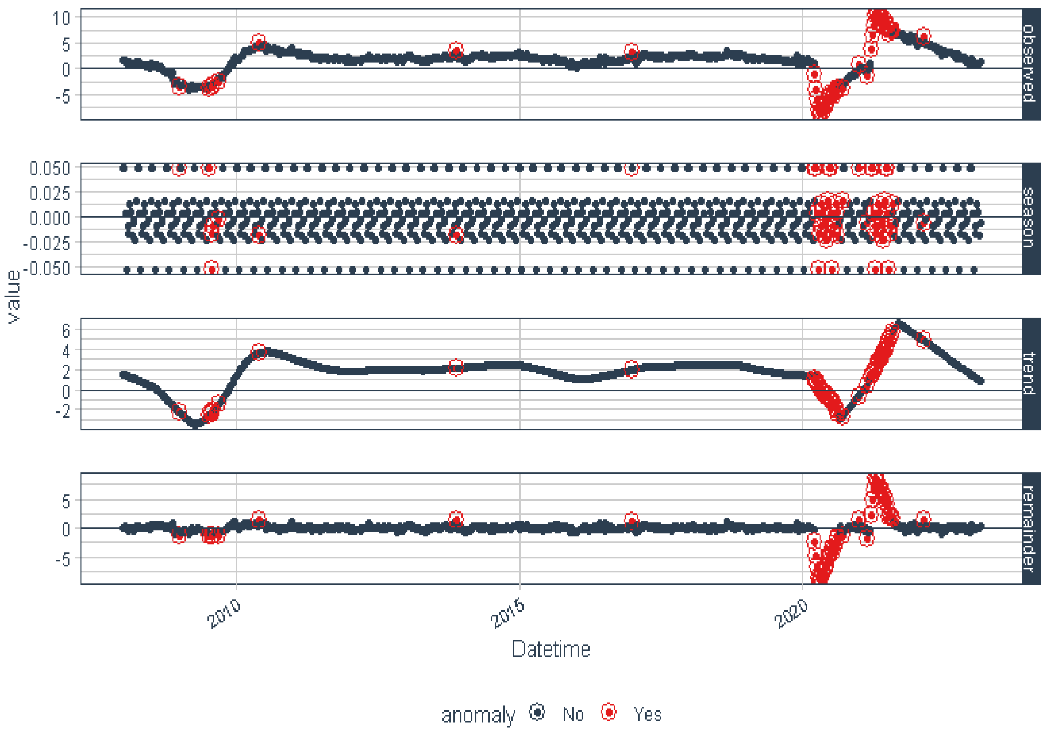


Figure 4.4: The Anomaly Detection Results of WEI

Naive Method Result

The seasonal periods of 4 for monthly and 52 for yearly are used. The outcomes of the naive method on WEI data are presented in Figure 4.5.

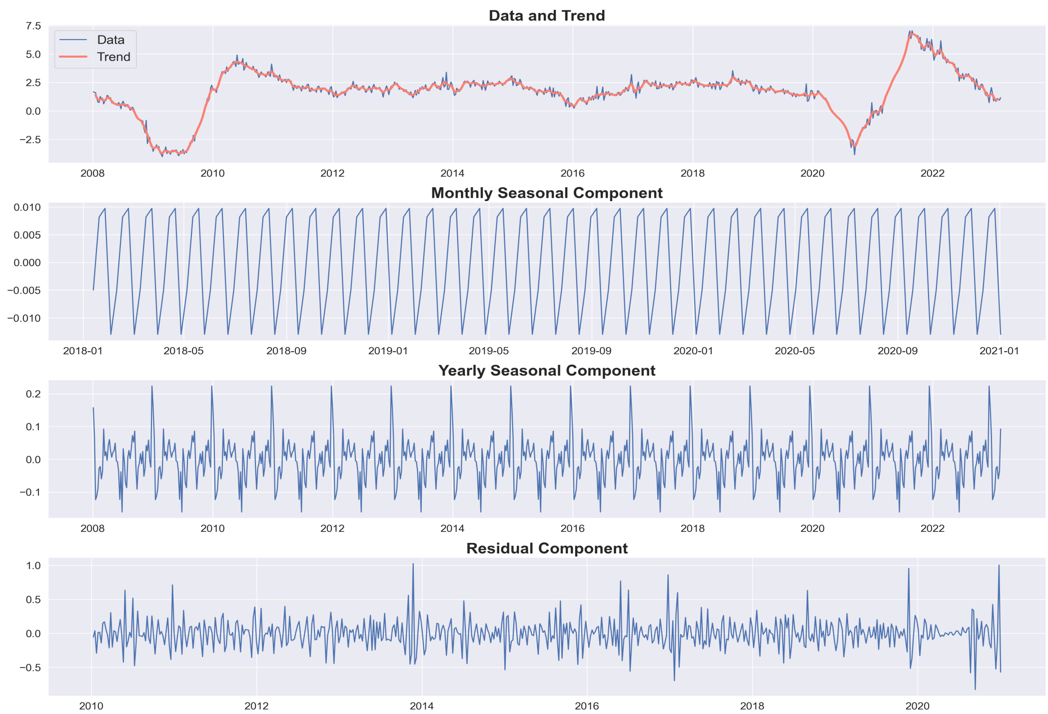


Figure 4.5: The Naive Method Results for WEI

Upon analyzing the outcomes, it is evident that the WEI data exhibits robust monthly and yearly seasonal trends. Specifically, the WEI data reaches its lowest point at the end of each month and peaks towards the end of each year.

MSTL Result

Under the details outlined in Chapter 3.2, the MSTL method is implemented on the data using monthly and yearly seasonal periods. The outcomes of the MSTL method on the WEI data are presented in Figure 4.6.

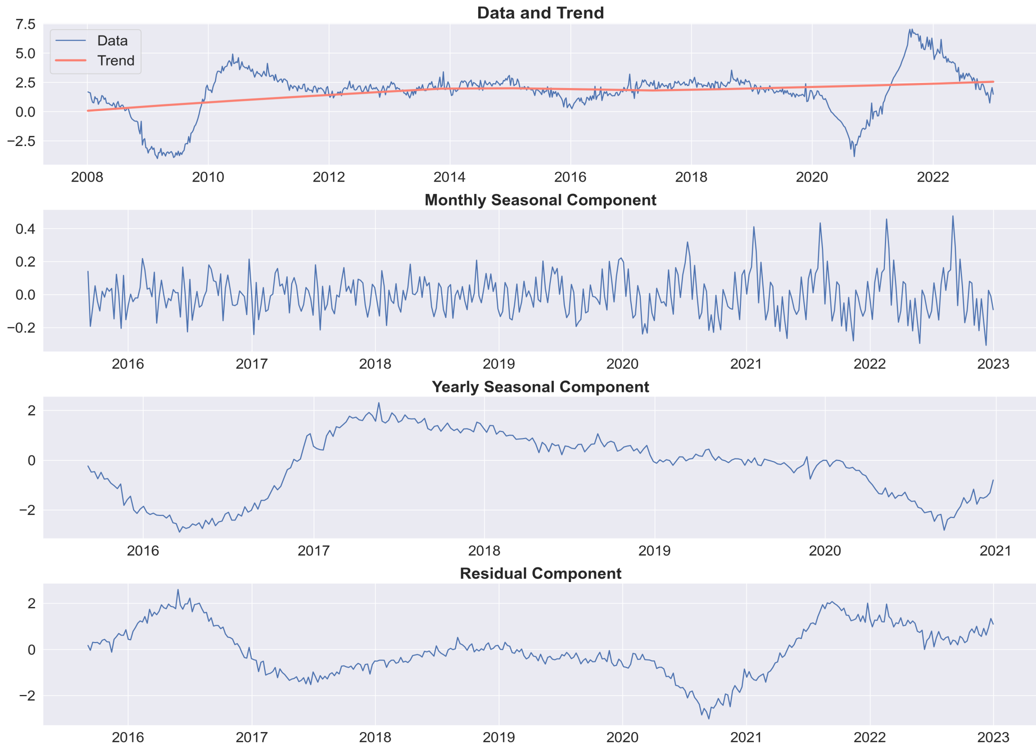


Figure 4.6: The MSTL Method Results for WEI

The time series plot reveals that the MSTL method also captures the effects that disrupt the seasonal pattern. The seasonal components display a varying structure over time. Especially the yearly seasonal component is not stable.

STR Result

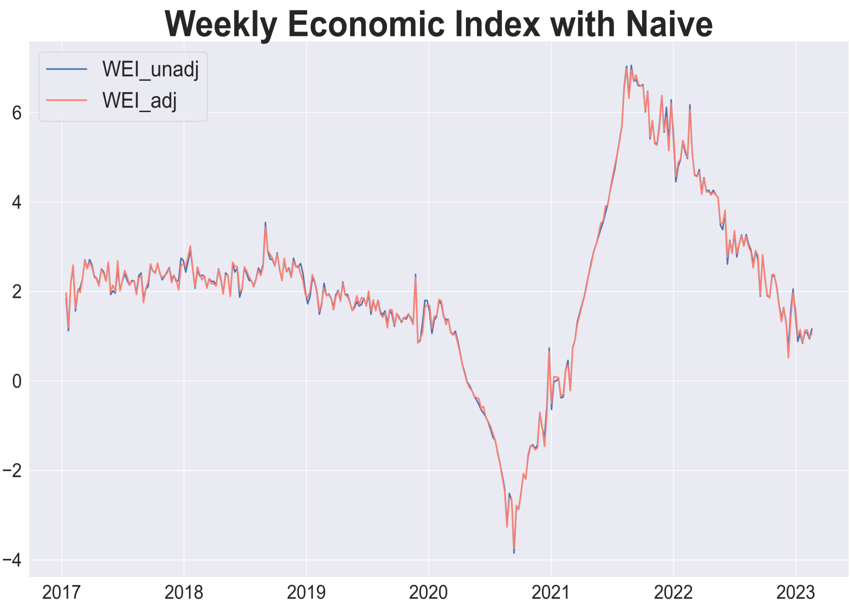
As specified in Chapter 3.4, the STR method is utilized on the data, with a "gapCV" parameter of 52 for WEI. The outcomes of the STR method on the WEI data are presented in Figure 4.7.

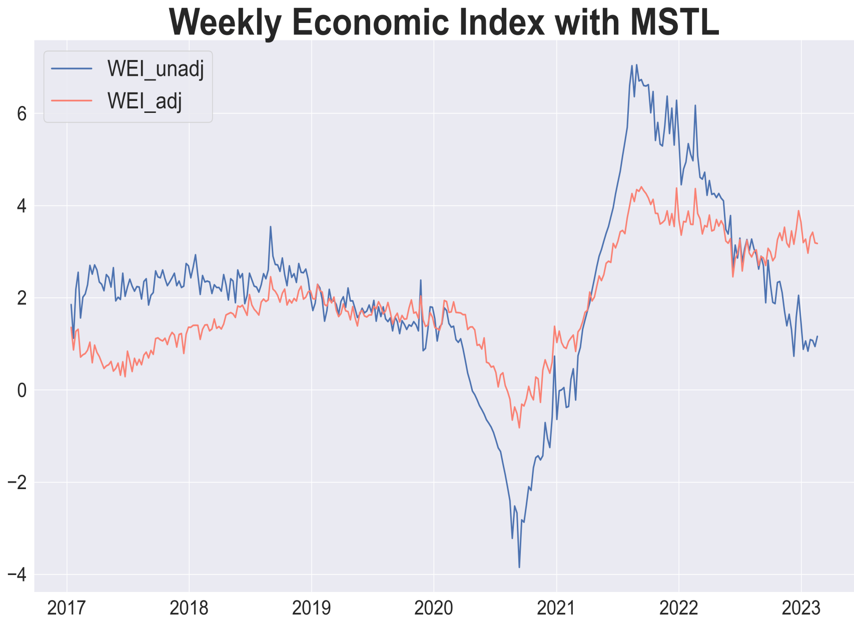
Figure 4.7: The STR Method Results for WEI

As with the MSTL method, the STR method can capture the effects that disrupt the seasonal pattern. Additionally, the STR method's monthly and yearly seasonality components are smoother than the MSTL method, along with the residual component.

Result for WEI

Figure 4.8 displays the ultimate seasonally adjusted outcomes for all methods.





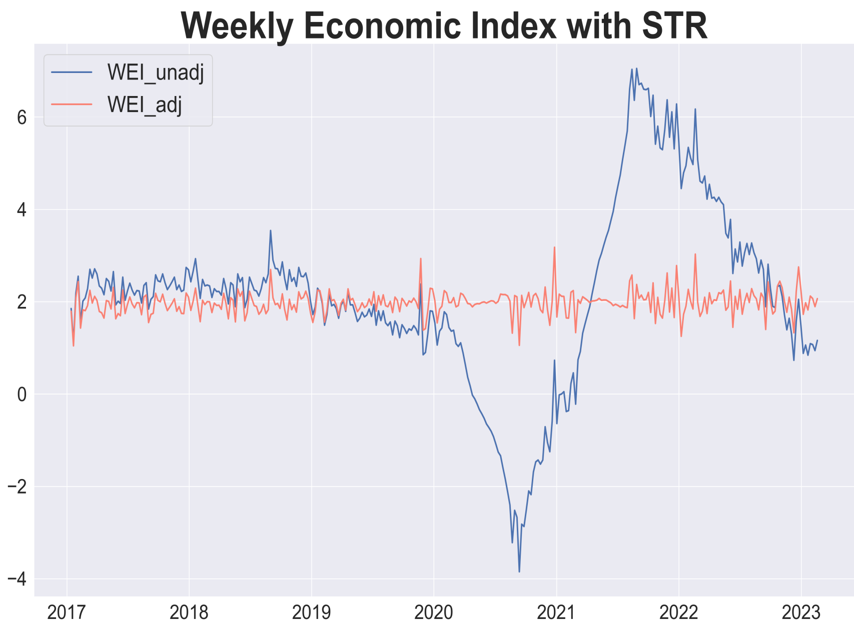


Figure 4.8: The Seasonally Adjusted Series Result of All Methods for WEI

Based on the visual outputs of the methods, it can be observed that the naive method has inadequate performance, whereas the STR method is excessively smooth. The STR approach cannot account for the Covid-19 impact, but the MSTL method successfully captures all the variations and anomalies.

The ACF and PACF plots for the final seasonally adjusted series are shown below.

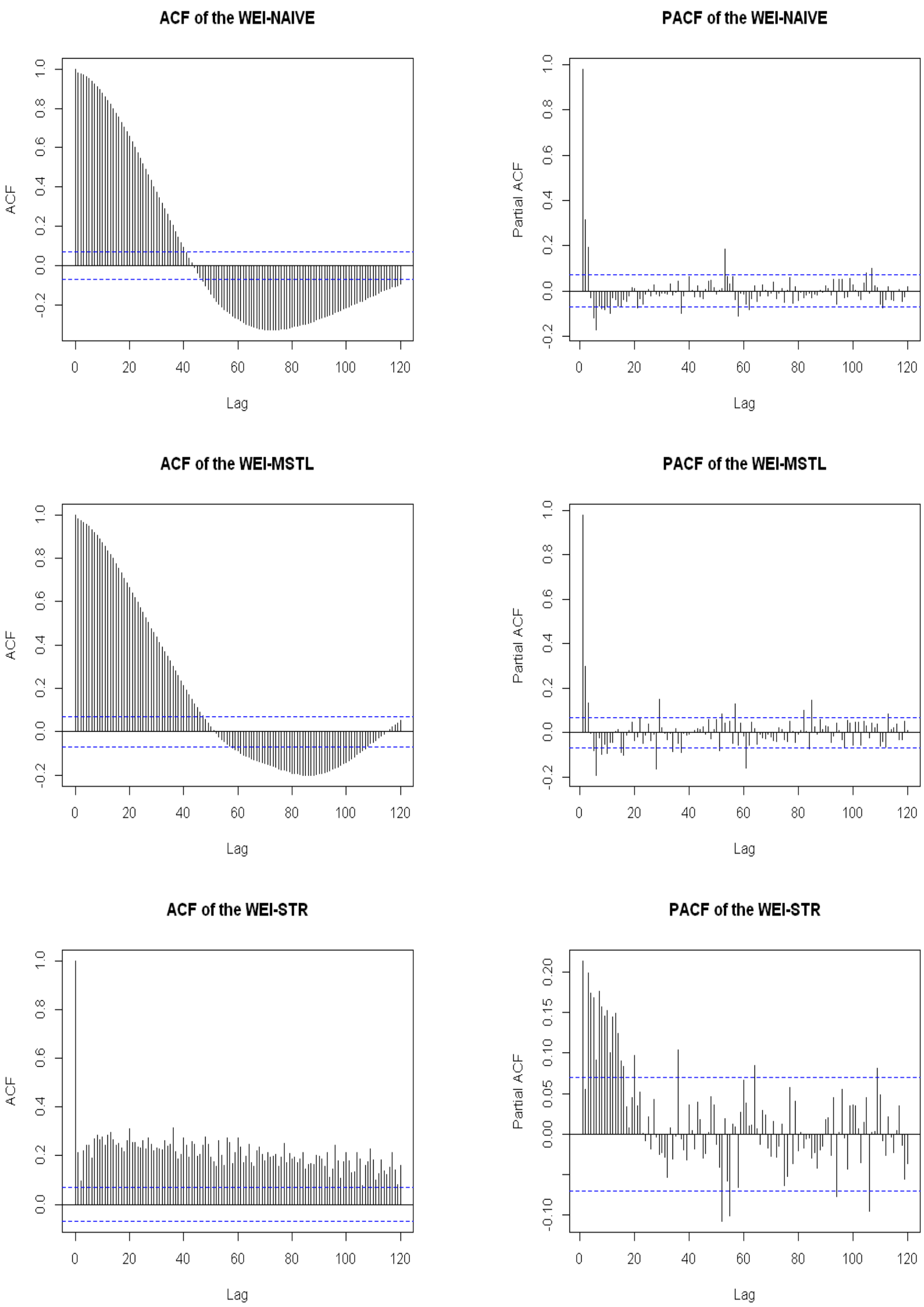


Figure 4.9: The Seasonally Adjusted Series ACF-PACF Plots for WEI

The ACF graph for the STR approach exhibits dissimilarities compared to the other methods. In contrast, the PACF graphs for the naive and MSTL approaches do not demonstrate any significant lags at the seasonal lags. However, the PACF graph for the STR technique illustrates significant lags at yearly seasonal lags.

4.3.2 Daily Time Series

The daily time series are subjected to seasonal adjustment using four different methods: naive method, MSTL method, DSA method, and STR method. Since the time series is of daily frequency, the expected seasonal components are weekly, monthly, and yearly. The weekly component is associated with the day-of-the-week effect, the monthly component is linked to the day-of-the-month effect, which is an average of 30 days, and the yearly component is related to the day-of-the-year effect, which occurs 365 days.

The analysis findings, excluding the daily electricity consumption data, have been included in the Appendix 6.2.

4.3.2.1 Daily Electricity Consumption in Türkiye

Electricity consumption is a crucial aspect of modern society, as it is an indispensable component of many daily activities. Thus, analyzing electricity consumption patterns is of significant interest to policymakers, energy suppliers, and researchers alike. In this context, the electricity consumption dataset provides an essential resource for investigating trends and patterns in electricity consumption over time. In this regard, the dataset can aid in developing effective energy management strategies, which can help reduce energy consumption and promote sustainability, moreover some economic clues like production in the industry.

In order to better understand and interpret the data, a time series plot is shown in Figure 4.10.

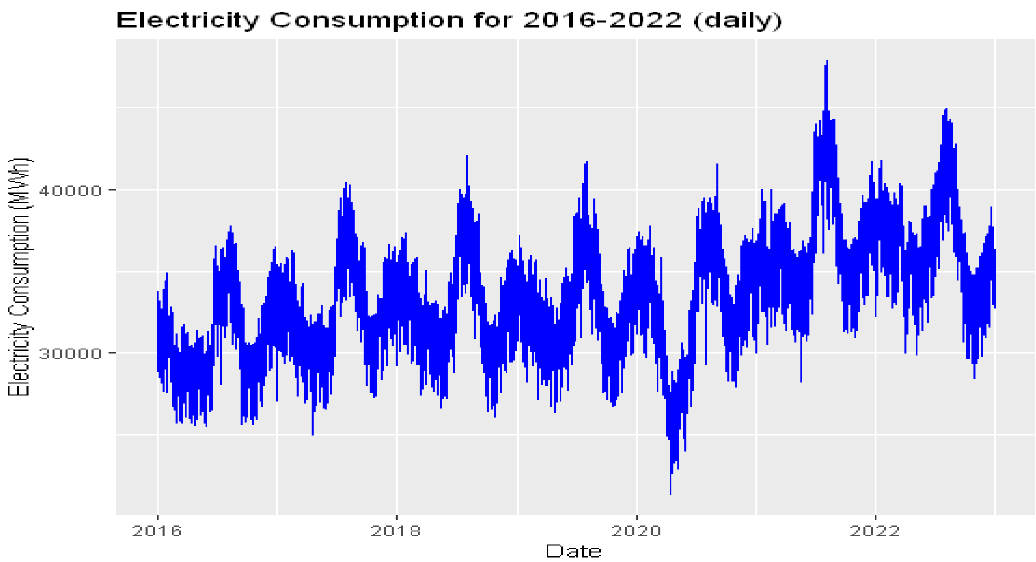


Figure 4.10: Time Series Plot of Daily Electricity Consumption in Türkiye

ACF and PACF graphs are also essential to understand the data structure. So, Figure 4.11 shows the ACF and PACF graphs of daily electricity consumption.

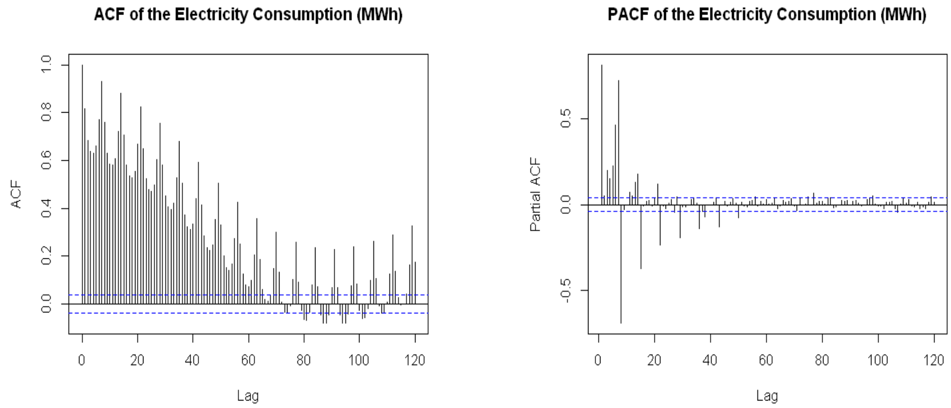


Figure 4.11: The ACF and PACF Graphs of Daily Electricity Consumption in Türkiye

Both ACF and PACF graphs show a strong seasonal pattern; there are significant lags at 7 and multiples. There is a high chance of multiple seasonality in the series.

To check multiple seasonality, the Box-Plots for the weekly, monthly and yearly intervals can help better sentiment to the seasonal data structure.

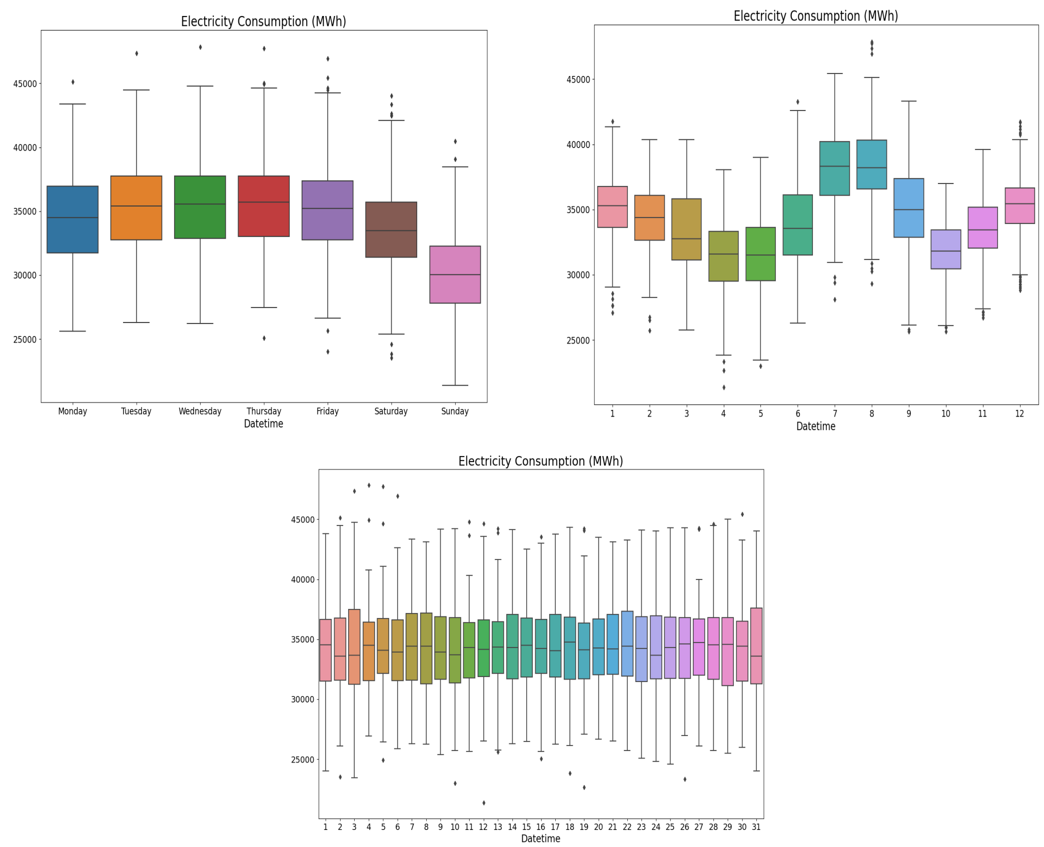


Figure 4.12: The Box-Plots of Daily Electricity Consumption in Türkiye Weekly, Monthly and Yearly

The weekly box plot shows that the median and upper quartile of weekends is below the weekdays. Since production is reduced in the industry on weekends, electricity consumption is decreased. Similarly, the monthly box plot also shows strong seasonal behavior. The yearly plot has a stable median.

The time series is imputed for the pre-processing step. The data also have anomalies. The anomaly detection result is shown in Figure 4.13.

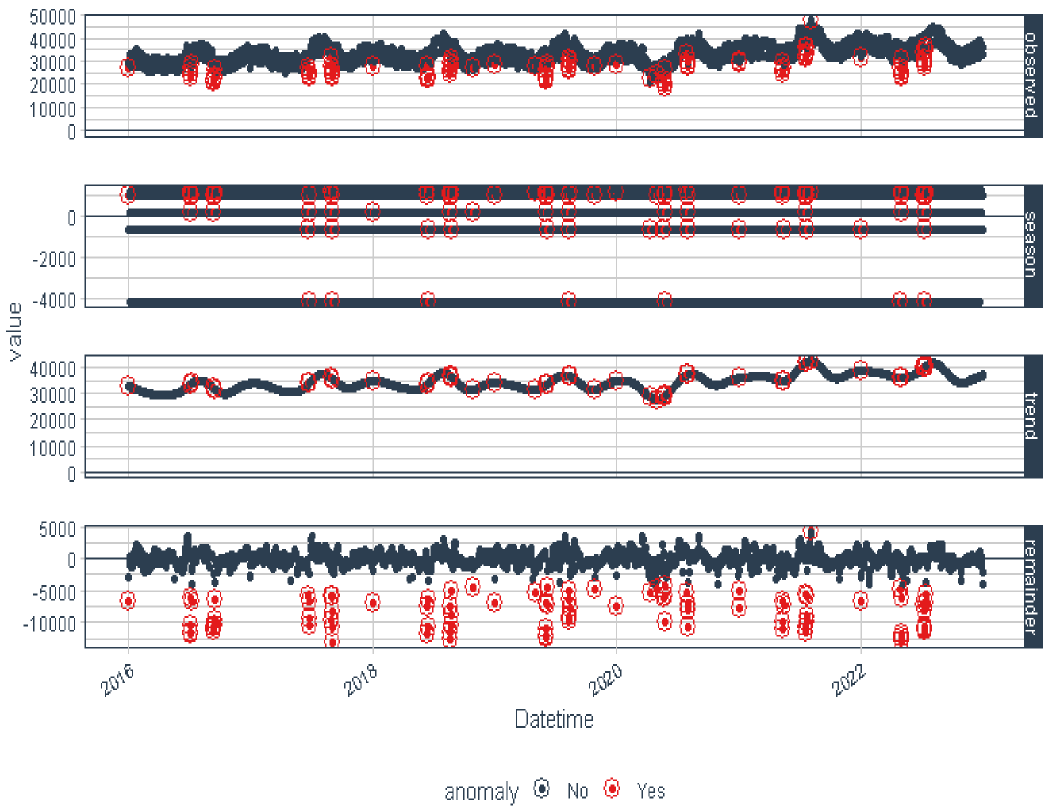


Figure 4.13: The Anomaly Detection Results of Daily Electricity Consumption in Türkiye

Naive Method Result

The naive method is applied to the data. It uses the basic moving average method, as mentioned in Chapter 3.1. The seasonal periods are operated weekly (7), monthly (30) and yearly (365).

The results of the naive method for daily electricity consumption are given in Figure 4.14.

The results show that the daily electricity consumption data exhibits significant weekly, monthly, and yearly seasonal patterns. The data shows a decline in consumption at the end of each week, while there is an increase in usage during the summer months.

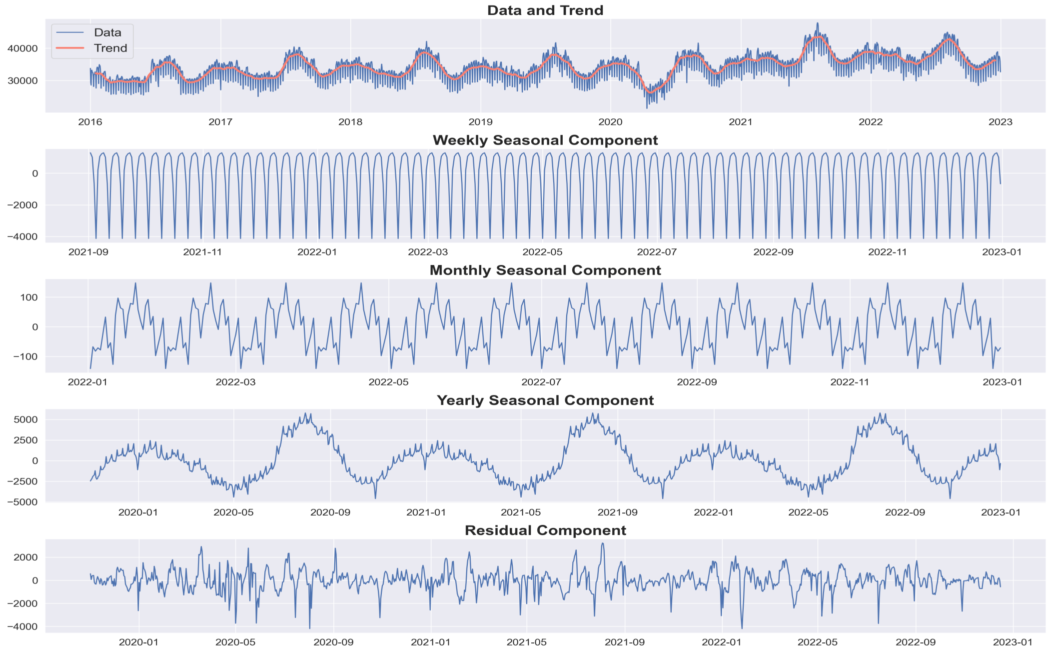


Figure 4.14: The Naive Method Results for Daily Electricity Consumption in Türkiye

MSTL Result

The MSTL method is applied to the data, the details are mentioned in Chapter 3.2. The seasonal periods are operated weekly, monthly and yearly.

The results of the MSTL method for daily electricity consumption are given in Figure 4.15.

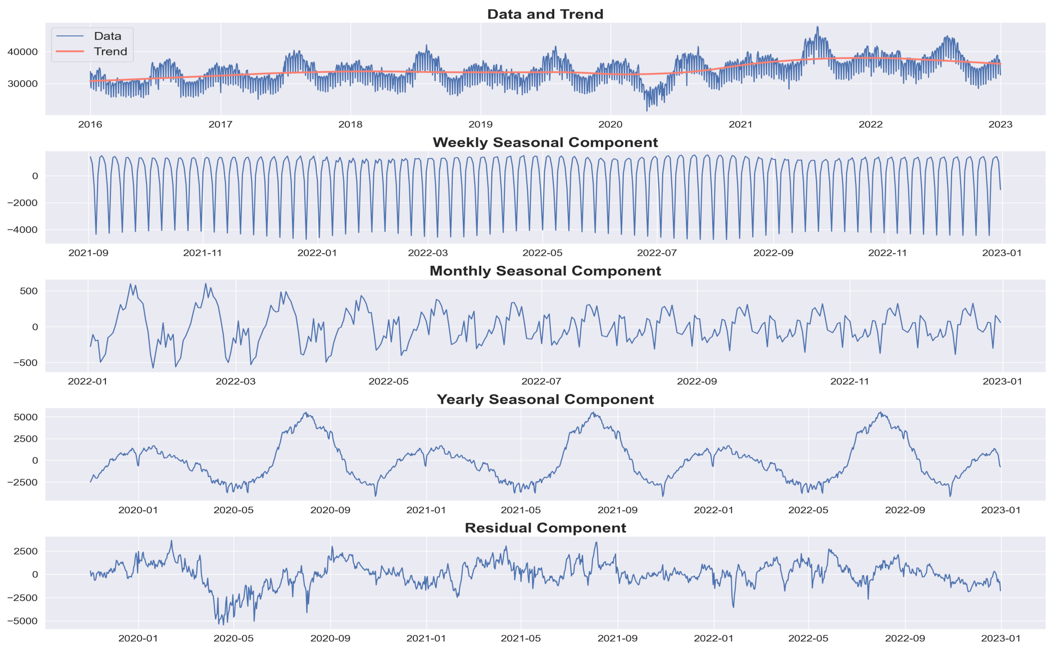


Figure 4.15: The MSTL Method Results for Daily Electricity Consumption in Türkiye

According to the MSTL method, there is a strong, complex seasonality. The monthly seasonal factor has fading seasonality.

STR Result

The STR method is applied to the data; the details are mentioned in Chapter 3.4. The STR method has a "gapCV" parameter set to 30 for daily electricity consumption.

The results of the STR method for daily electricity consumption data are given in Figure 4.16.

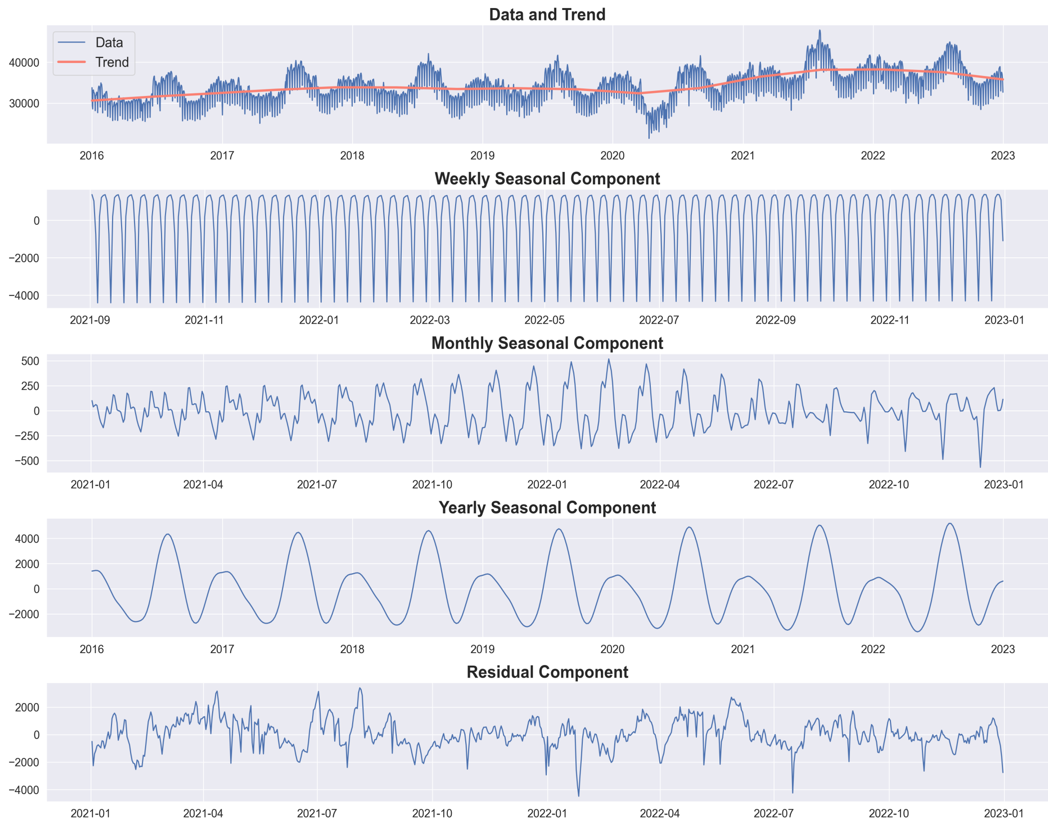


Figure 4.16: The STR Method Results for Daily Electricity Consumption in Türkiye

Similar to the MSTL method, the STR method identified multiple seasonality in the data. Compared to the MSTL method, all seasonal components, including monthly and yearly seasonality, appear smoother in the STR method. Additionally, the residual component of the STR method is smoother.

DSA Result

The DSA method is applied to the data, explained in Chapter 3.3. This method is used only for the daily time series. In this method, the daily electricity consumption is also calendar adjusted. The holiday regressors are created for Türkiye.

The effect of the calender is shown in the Table 4.2.

Table 4.2: Calender Effect on Daily Electricity Consumption in Türkiye

Holiday Names	**Coefficient**	**s.e**	**t_value**
Feast of Sacrifice	384.26	193.40	2.00
Feast of Ramadan	1478.25	196.07	7.50
New Year's Holiday	1560.59	184.13	8.50
April 23	-1866.15	176.90	-10.50
May 1	-1168.85	177.60	-6.60
May 19	-1643.64	176.94	-9.30
July 15	-1225.82	190.70	-6.40
August 30	-1508.90	176.90	-8.50
October 29	-2062.53	202.86	-10.20
Saturdays	-71.22	28.22	-2.50
Sundays	16.49	28.17	0.60

The calendar effect is also shown in Figure 4.17.

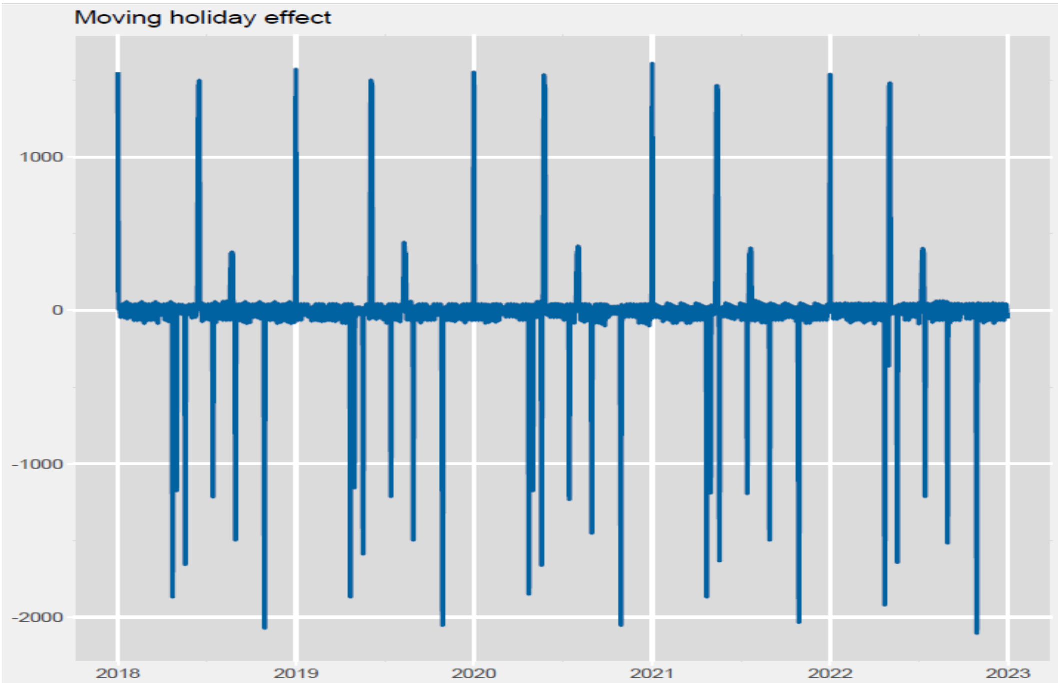


Figure 4.17: The Calender Effect for Daily Electricity Consumption in Türkiye

As a last step, the result of the DSA method is given in Figure 4.18.

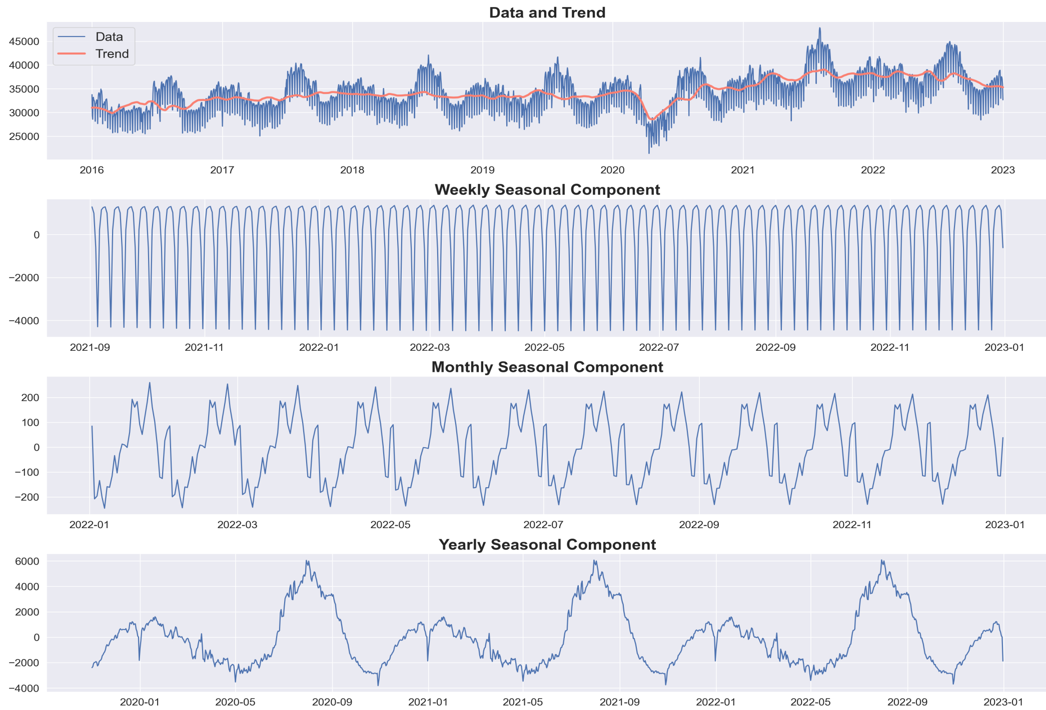
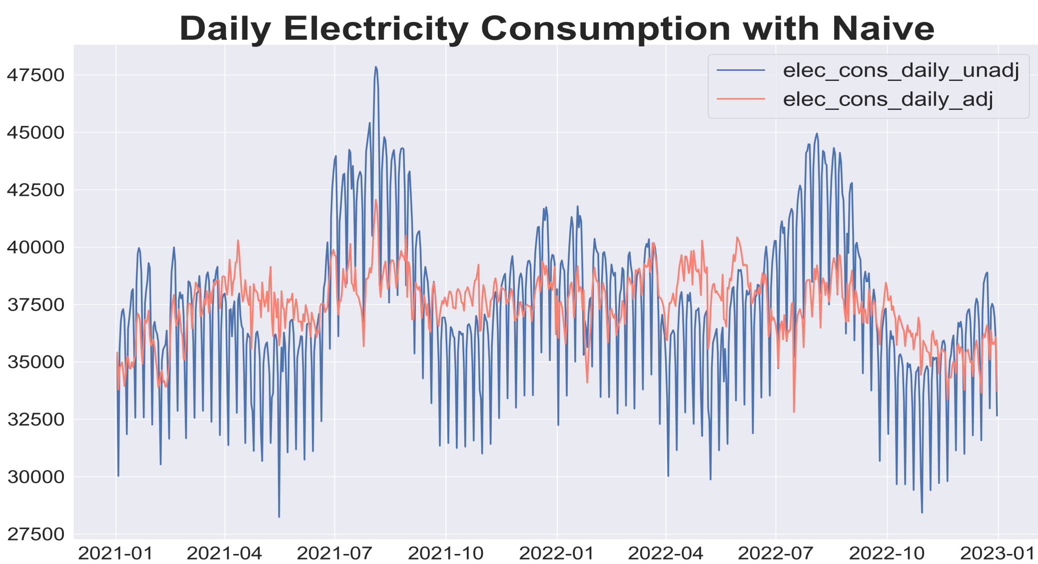


Figure 4.18: The DSA Method Results for Daily Electricity Consumption in Türkiye

The DSA method gives similar results to the STR method. But, the yearly seasonal component is more volatile than the STR method.

<u>Result for Daily Electricity Consumption in Türkiye</u>

Figure 4.19 depicts the graphs illustrating the results of all the seasonal adjustment techniques for the daily electricity consumption data.



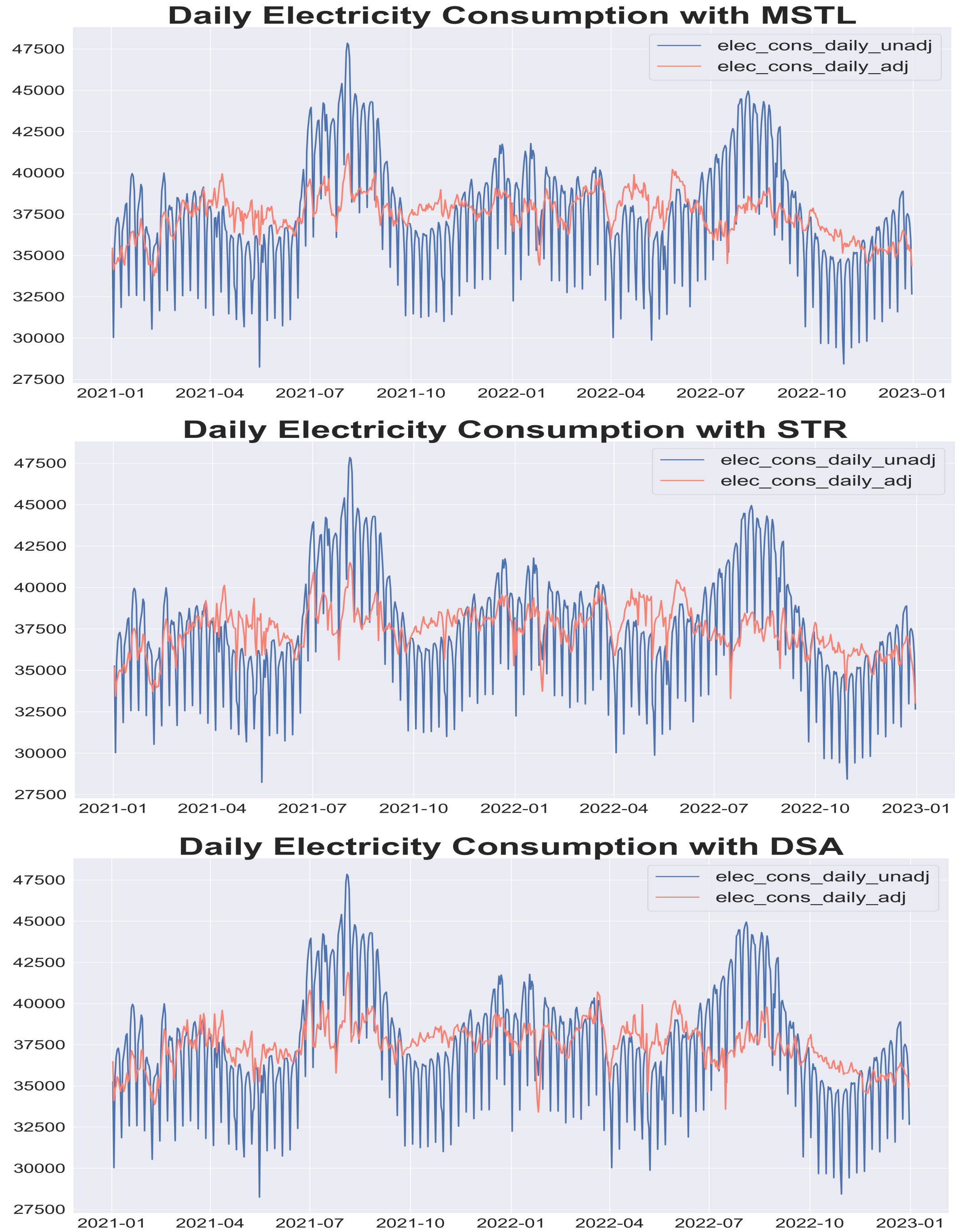


Figure 4.19: The Seasonally Adjusted Series Result of All Methods for Daily Electricity Consumption In Türkiye

Based on the visual outputs of the methods, it can be observed that the results series are pretty similar. But when examined in detail, DSA and MSTL methods are the smoothest.

The ACF and PACF plots for the final seasonally adjusted series are shown below.

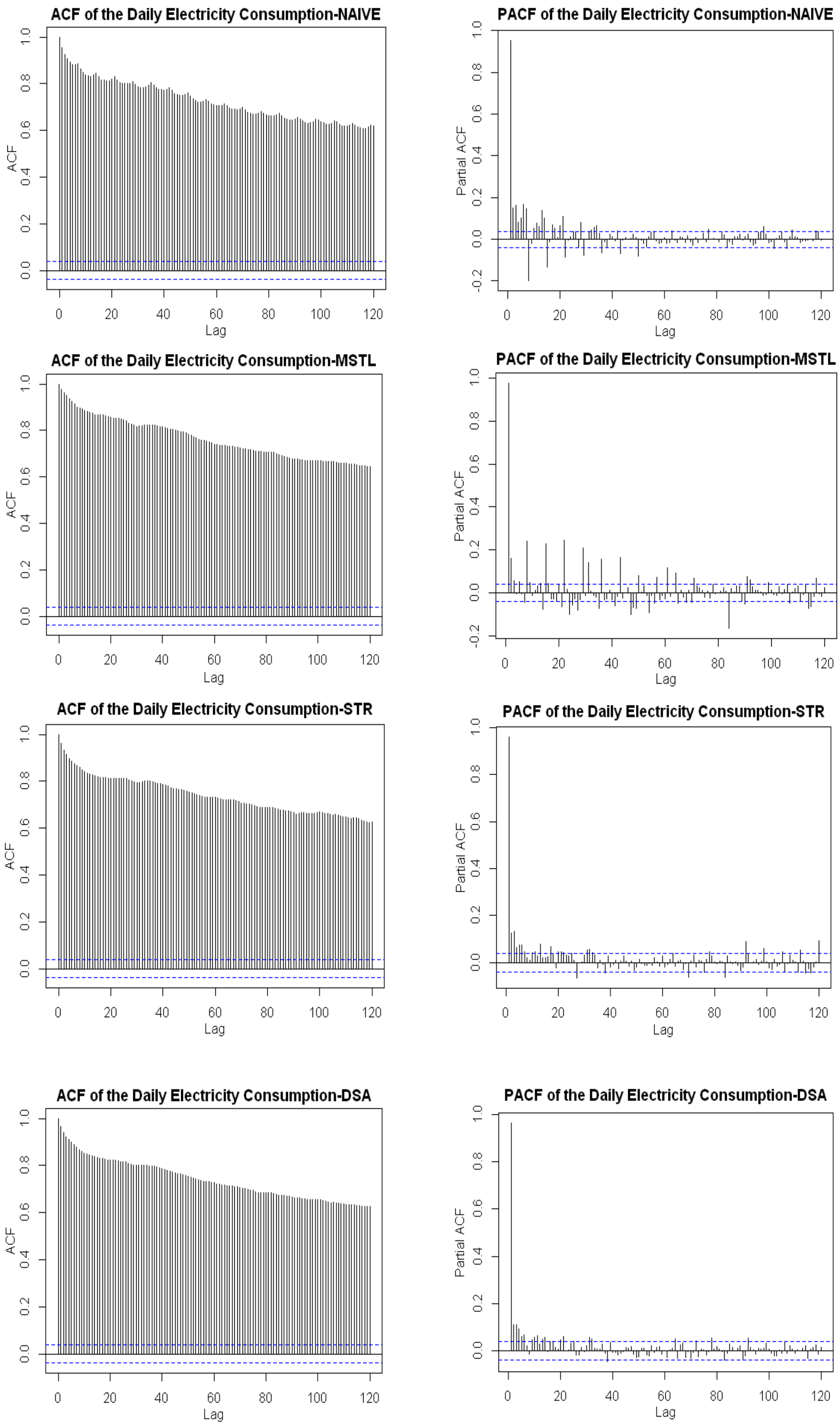


Figure 4.20: The Seasonally Adjusted Series ACF-PACF Plots for Daily Electricity Consumption in Türkiye

The ACF graphs have similar behavior. In contrast, the PACF graphs in the DSA method differ from others by having insignificant lags. Almost all lags are in the white noise band. To sum up, the DSA method perfoms better for daily time series.

4.3.3 Hourly Time Series

To perform a seasonal adjustment on the hourly time series, the naive, MSTL, and STR methods are applied. As the time series has an hourly period, the anticipated seasonal components are daily, weekly, monthly, and yearly.

The results of the London bike sharing data analysis have been incorporated into the Appendix 6.3.

4.3.3.1 Hourly Electricity Consumption in Türkiye

Hourly electricity consumption refers to the amount of electrical power used by households and businesses during each hour of the day. This data type is crucial for power grid operators and energy companies to forecast and manage their energy production and distribution. The hourly electricity consumption can exhibit strong seasonal patterns due to daily, weekly, monthly, and yearly factors. Analyzing and adjusting for these seasonal patterns is vital for accurately forecasting and managing energy supply and demand.

In order to better understand and interpret the data, a time series plot is shown in Figure 4.21.

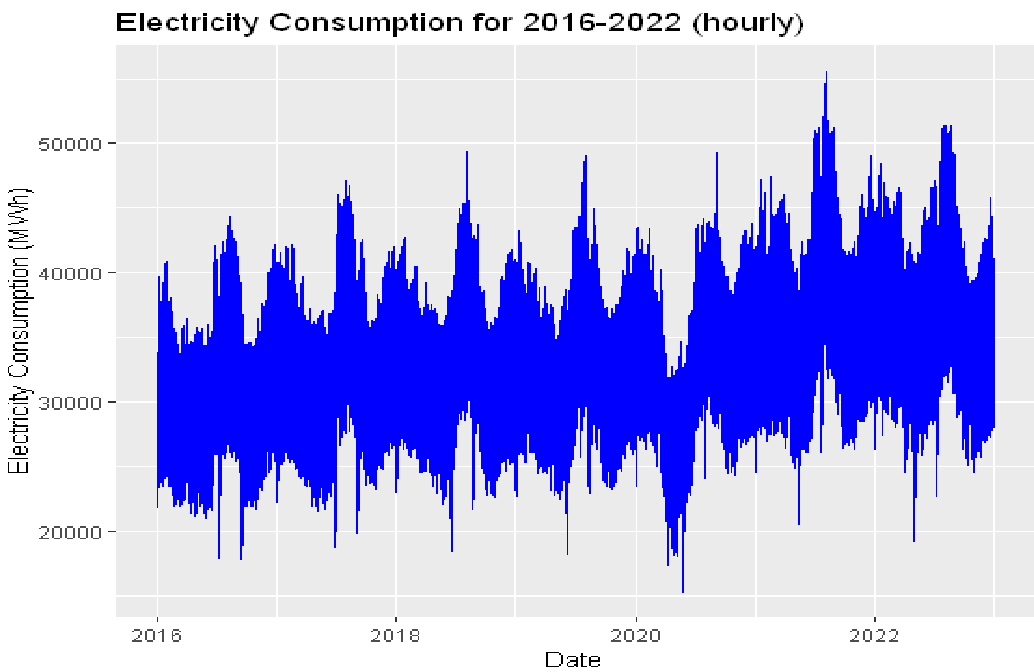


Figure 4.21: Time Series Plot of Hourly Electricity Consumption in Türkiye

Figure 4.21 illustrates that the hourly electricity consumption data exhibits significant periodic movements and multiple seasonal patterns.

To further understand the data structure, it is essential to examine the ACF and PACF graphs presented in Figure 4.22.

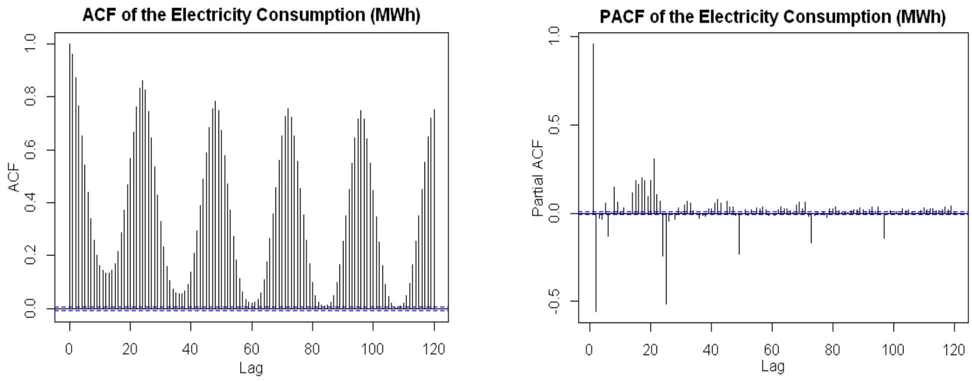


Figure 4.22: The ACF and PACF Graphs of Hourly Electricity Consumption in Türkiye

The ACF graphs indicate that the data has a strong seasonal behavior with significant lags. On the other hand, the PACF graph displays daily seasonality due to the presence of significant lags at 24 and its multiples.

To investigate multiple seasonality, analyzing the Box-Plots for daily, weekly, monthly, and yearly intervals can provide a clearer understanding of the seasonal data structure.

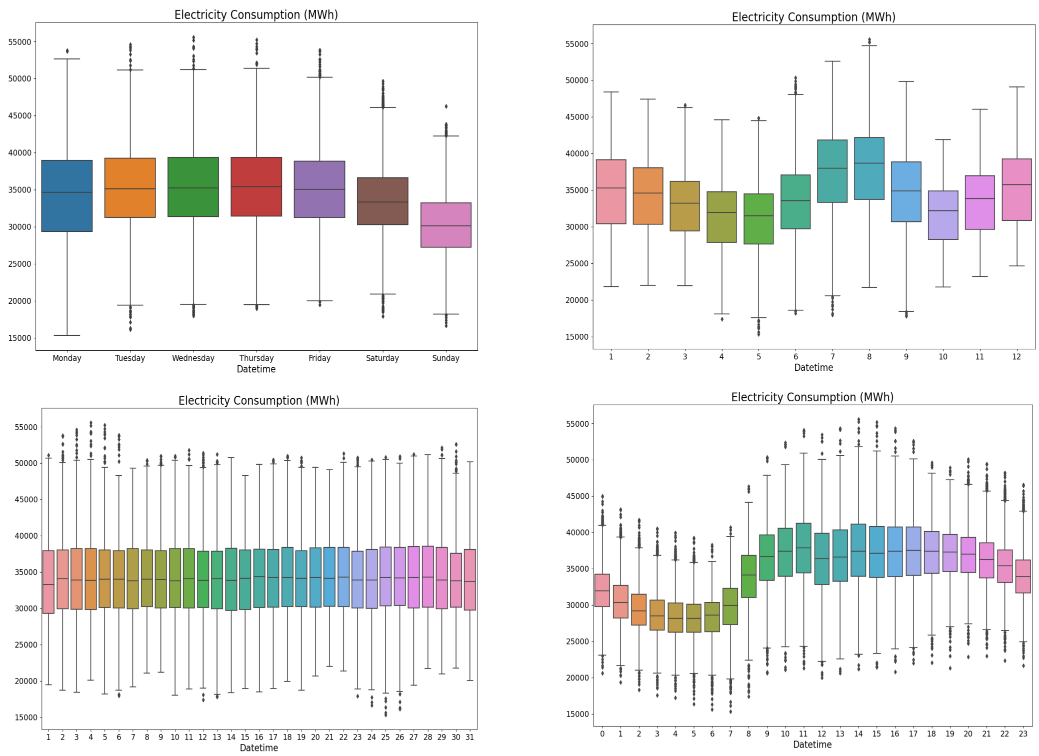


Figure 4.23: The Box-Plots of Hourly Electricity Consumption in Türkiye Daiy, Weekly, Monthly and Yearly

The box plot for hours a day indicates a noticeable daily seasonality, as the median decreases during night hours. This outcome is reasonable for electricity consumption data, as households and industries consume less energy at night. Similarly, the day-of-the-week effect appears seasonal, with the median for weekdays higher than for weekends. The yearly seasonality also exhibits a similar pattern, with the median being higher in autumn compared to other months.

The time series is imputed for the pre-processing step. The high frequency of the data makes it particularly sensitive to outliers, which can significantly impact the results. Classical methods used for monthly data often identify many outliers, which can lead to unreliable models. As a result, parameters are set explicitly for the hourly time series to account for this issue and improve the accuracy of the analysis.

The anomaly detection result is shown in Figure 4.24.

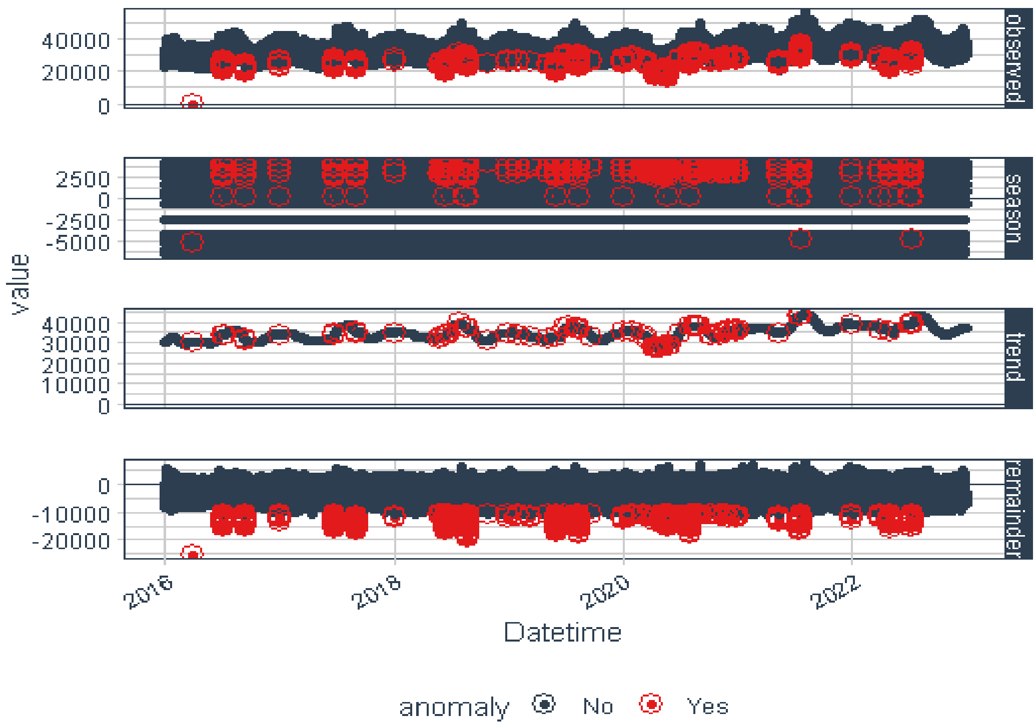


Figure 4.24: The Anomaly Detection Results of Hourly Electricity Consumption in Türkiye

<u>Naive Method Result</u>

The naive method is applied to the data. It uses the basic moving average method, as mentioned in Chapter 3.1. The seasonal periods are operated as;

- Daily (24)
- Weekly (7*24)
- Monthly (30*24)
- Yearly (365*24).

The results of the naive method for hourly electricity consumption are given in Figure 4.25.

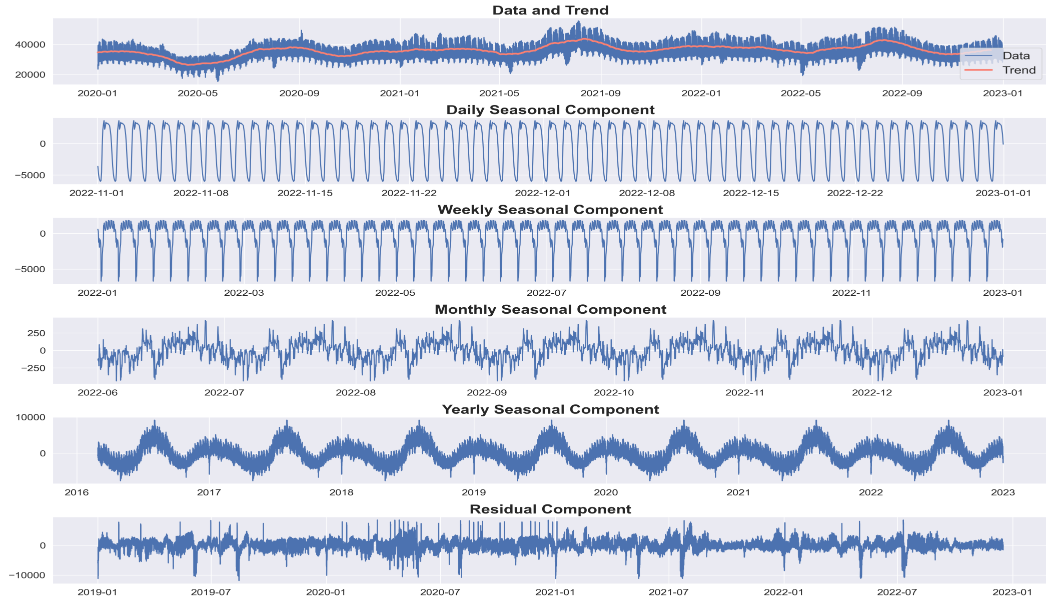


Figure 4.25: The Naive Method Results for Hourly Electricity Consumption in Türkiye

The analysis of the data reveals the presence of significant seasonal patterns in the hourly electricity consumption.

MSTL Result

The MSTL method is applied to the data, the details are mentioned in Chapter 3.2. The seasonal periods are operated daily, weekly, monthly and yearly.

The results of the MSTL method for rail traffic data are given in Figure 4.26.

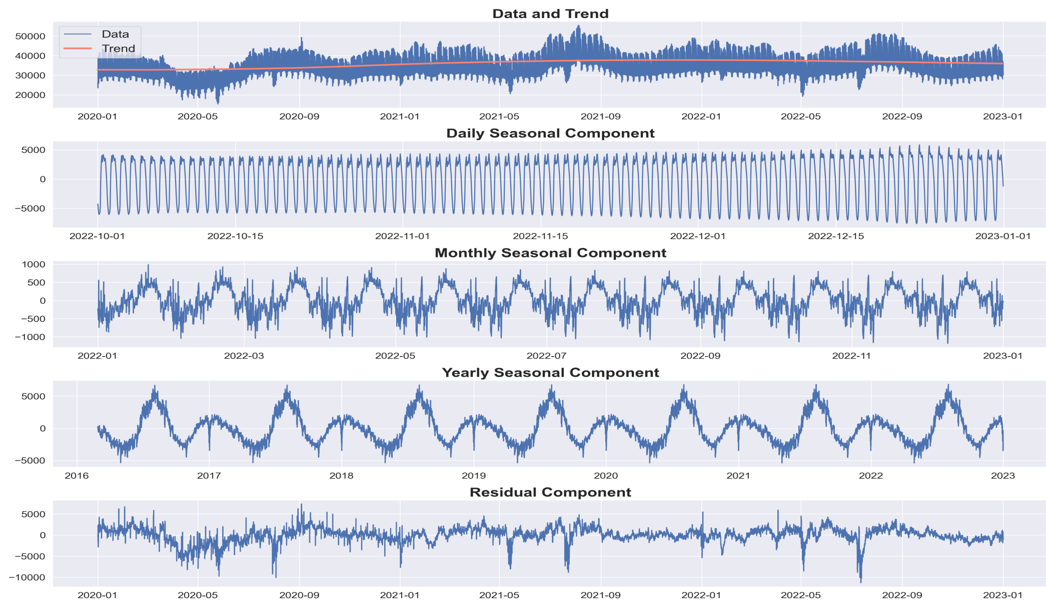


Figure 4.26: The MSTL Method Results for Hourly Electricity Consumption in Türkiye

The stable seasonal components of hourly electricity consumption can be observed in Figure 4.26. However, the weekly seasonality is not apparent in this particular method. Although the monthly seasonal component shows a weekly periodic pattern, the model cannot capture it.

STR Result

The electricity consumption data is processed using the STR method, as explained in Chapter 3.4, with the "gapCV" parameter set to 24. However, the STL method is not applicable due to a large number of observations. Only the last two years of the hourly electricity consumption data are used for the STL method.

Figure 4.27 presents the results of the STR method applied to hourly electricity consumption.

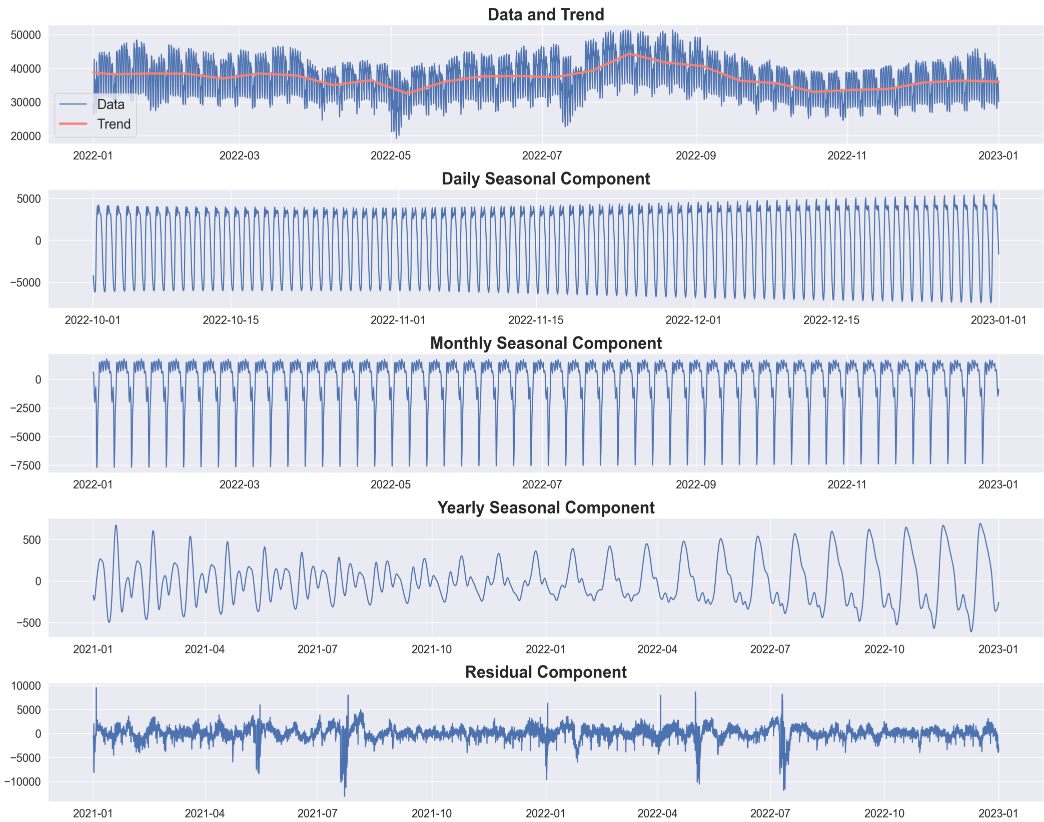


Figure 4.27: The STR Method Results for Hourly Electricity Consumption in Türkiye

The STR method is able to detect the weekly seasonality, which was not detected by the MSTL method. Additionally, the direction of movement in the monthly seasonality has changed, resulting in a change in the seasonal factor. The yearly seasonality in the STR method appears to be smoother compared to the MSTL method, along with the residual component.

Result for Hourly Electricity Consumption in Türkiye

The final seasonally adjusted results for all approaches is given below.

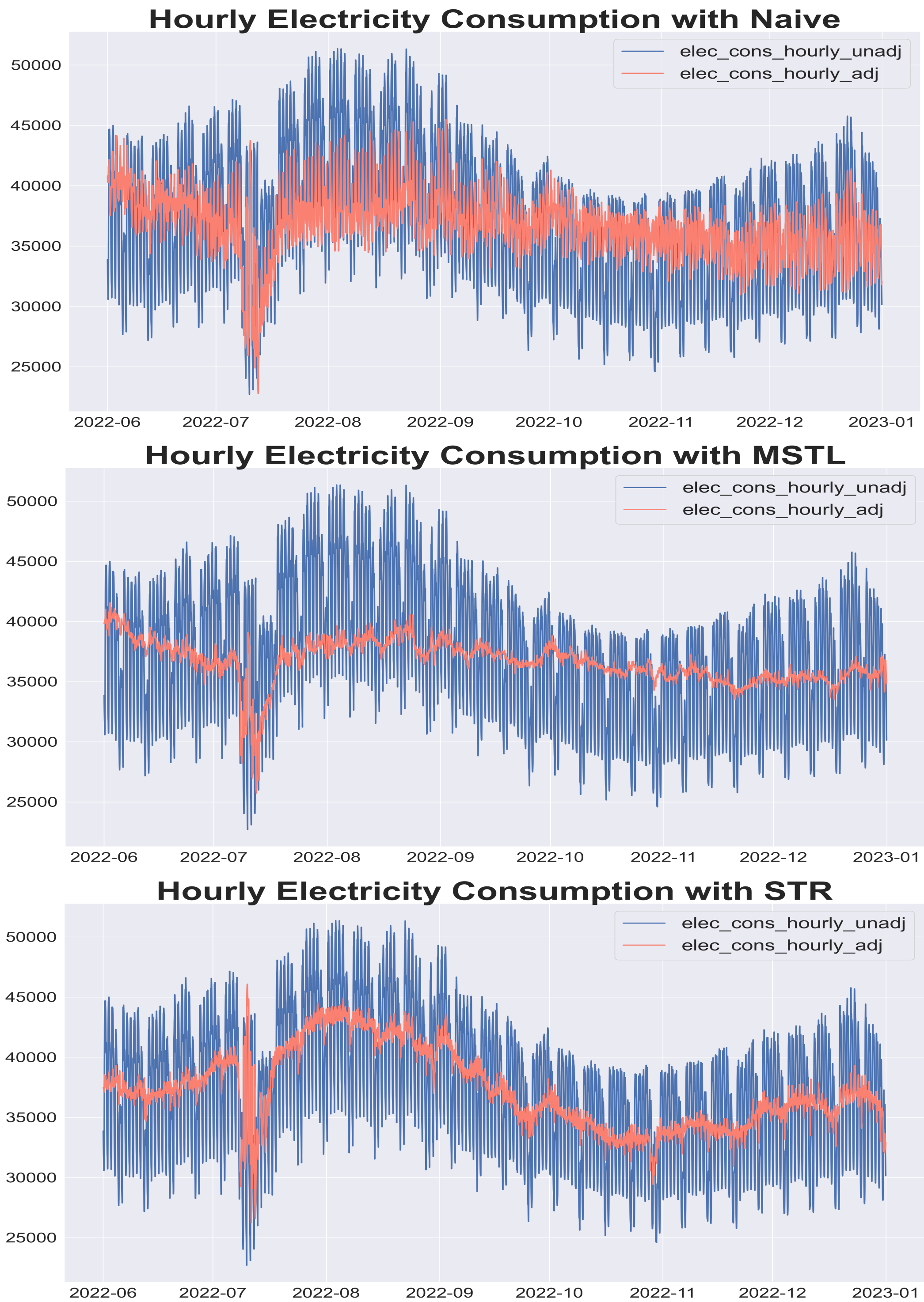


Figure 4.28: The Seasonally Adjusted Series Result of All Methods for Hourly Electricity Consumption In Türkiye

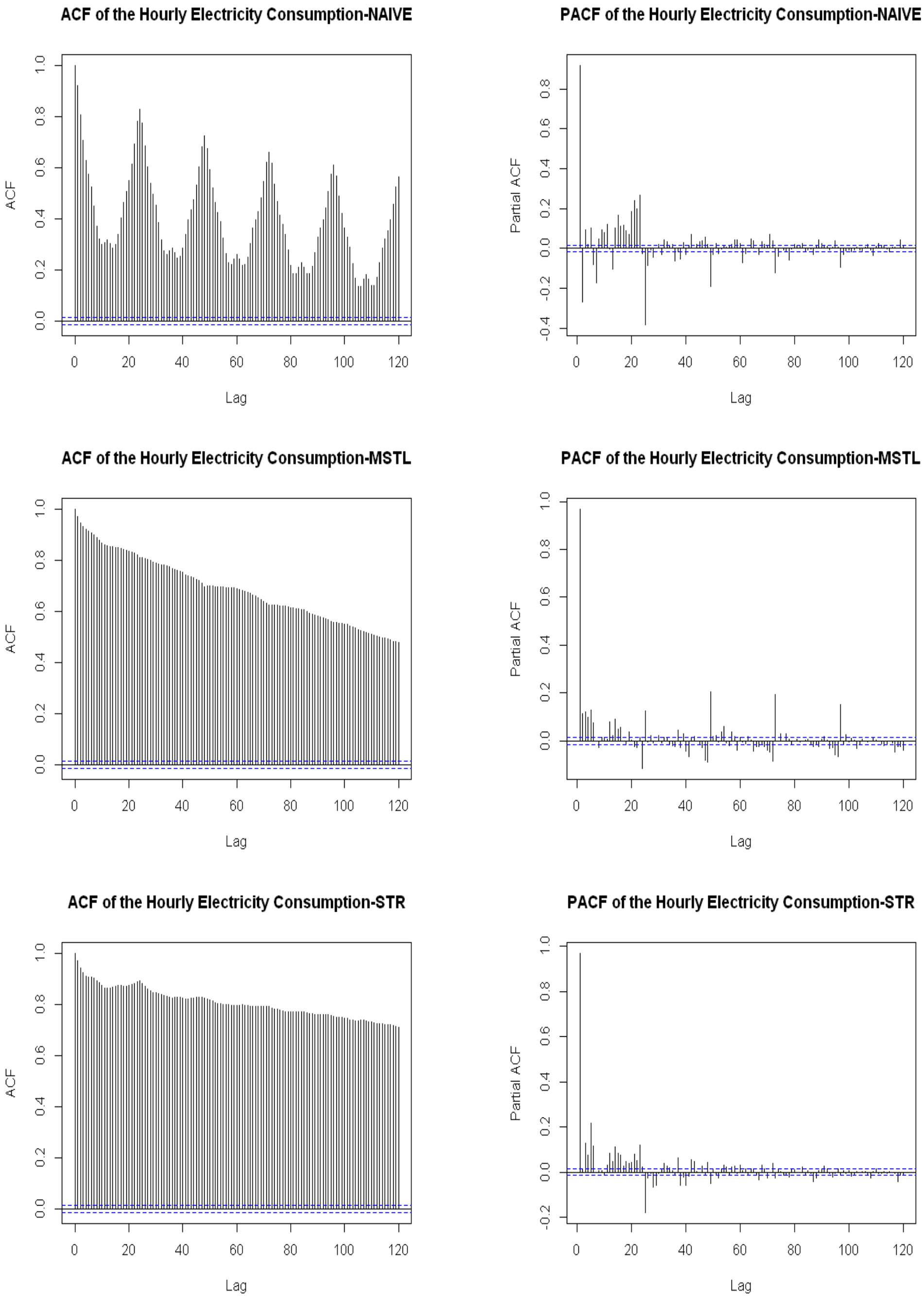


Figure 4.29: The Seasonally Adjusted Series ACF-PACF Plots for Hourly Electricity Consumption in Türkiye

The ACF graph of the Naive method has significant seasonal lags, while the other ACF graphs do not. PACF plots of the Naive method and MSTL method also have seasonal patterns. The STR method's PACF plot lags seem more in the white noise band and does not show any seasonla pattern anymore.

4.3.4 Results

Seasonal adjustment of high-frequency time series can be demanding, but ensuring accurate analysis and forecasting is essential. This analysis chapter explored several methods for seasonally adjusting high-frequency time series data, including Naive Method, MSTL, STR, and DSA. A thorough evaluation of the performance of each method should be examined to identify the most suitable approach for our dataset.

However, it is essential to note that the choice of method may depend on the characteristics of the data, and some methods may perform better for certain types of time series. There is no one method that is superior to the others.

Seasonal tests can provide valuable insights into a time series' seasonal patterns presence. In addition to graphical understanding, these tests can help confirm the existence of seasonality, identify the seasonal period, and assess the significance of the seasonal component. In chapter 3 of this thesis, the QS and the Friedman seasonal tests are described.

The QS test, also known as the non-parametric seasonal test, is a non-parametric approach for detecting seasonal patterns in time series data. This test checks the autocorrelation at seasonal lags.

On the other hand, the Friedman test is a non-parametric test used to detect stable seasonality in time series data. This test compares the median values of each seasonal period and whether they are equal.

The results of the seasonal tests for various techniques are presented in the table below.

Table 4.3: Results of Seasonality Tests for Each Method on All Datasets

WEI	Orj_Teststat	Orj_P-value	naive_sa_Teststat	naive_sa_P-value	mstl_sa_Teststat	mstl_sa_P-value	str_sa_Teststat	str_sa_P-value
Friedman365	437.2	0.005	413.6	0.037	436.9	0.04	429.2	0.01
QS365	13.5	0.001	10.3	0.006	17.2	0.1	12.3	0.002
Friedman12	10.4	0.499	3	0.99	2.5	0.996	16.3	0.13
QS12	0	1	0	1	0	1	0	1

Daily Electricity Consumption	Orj_Teststat	Orj_P-value	naive_sa_Teststat	naive_sa_P-value	mstl_sa_Teststat	mstl_sa_P-value	str_sa_Teststat	str_sa_P-value	dsa_sa_Teststat	dsa_sa_P-value
Friedman365	156.7	1	637.8	0	366.4	0.455	490	0	361	0.534
QS365	0	1	200.6	0	0	1	143.1	0	0	1
Friedman12	67.1	0	4.9	0.934	3.7	0.979	6.3	0.854	2.5	0.996
QS12	96.5	0	0	1	0	1	0	1	0	1
Friedman7	50.1	0	27	0	0.9	0.988	3.8	0.708	4	0.679
QS7	111.3	0	36.3	0	0	1	0	1	0	1

Hourly Electricity Consumption	Orj_Teststat	Orj_P-value	naive_sa_Teststat	naive_sa_P-value	mstl_sa_Teststat	mstl_sa_P-value	str_sa_Teststat	str_sa_P-value
Friedman365	132.6	1	288	0.999	273.1	1	283.4	1
QS365	0	1	0	0	0	1	0	1
Friedman12	61.7	0	5.6	0.899	3.7	0.978	4.8	0.9
QS12	83.4	0	0	1	0	1	0	1
Friedman7	5.7	0.0453	2.8	0.083	3.7	0.713	3.4	0.83
QS7	0	1	0	1	0	1	0	1
Friedman24	196.5	0	174.5	0	12.4	0.964	34.1	0.66
QS24	311.2	0	236.8	0	0	1	0	1

Weekly Tracker	Orj_Teststat	Orj_P-value	naive_sa_Teststat	naive_sa_P-value	mstl_sa_Teststat	mstl_sa_P-value	str_sa_Teststat	str_sa_P-value
Friedman365	403.1	0.047	401.3	0.016	423.9	0.16	399.9	0.095
QS365	0	1	0	1	11.2	0.004	0	1
Friedman12	7.8	0.073	3.9	0.972	4.1	0.966	5.9	0.878
QS12	1	0.037	0	1	2.2	0.333	0	1

Rail Traffic	Orj_Teststat	Orj_P-value	naive_sa_Teststat	naive_sa_P-value	mstl_sa_Teststat	mstl_sa_P-value	str_sa_Teststat	str_sa_P-value
Friedman365	364	0.49	359.2	0.49	396.1	0.001	354	0.009
QS365	0	1	0	1	8	0.003	0	1
Friedman12	69.79	0	16.37	0.128	0	0.966	7.2	0.79
QS12	36.1	0	0	1	2.1	0.08	0	1

CO	Orj_Teststat	Orj_P-value	naive_sa_Teststat	naive_sa_P-value	mstl_sa_Teststat	mstl_sa_P-value	str_sa_Teststat	str_sa_P-value	dsa_sa_Teststat	dsa_sa_P-value
Friedman365	385.2	0.213	376.8	0.311	317.1	0.963	358.8	0.567	495.4	0.42
QS365	6.8	0.033	6.2	0.044	0	1	0	1	10.6	0.005
Friedman12	48.8	0	2.4	0.997	2.8	0.993	3.8	0.975	6.2	0.858
QS12	23.7	0	0	1	0	1	0	1	0	1
Friedman7	13.5	0.035	4.5	0.604	1.2	0.979	10.3	0.113	6	0.428
QS7	0.3	0.85	0	1	0	1	1.3	0.516	0	1

PM10	Orj_Teststat	Orj_P-value	naive_sa_Teststat	naive_sa_P-value	mstl_sa_Teststat	mstl_sa_P-value	str_sa_Teststat	str_sa_P-value	dsa_sa_Teststat	dsa_sa_P-value
Friedman365	392.8	0.144	380.8	0.262	453	0.001	353.2	0.647	340.4	0.808
QS365	1.5	0.478	1	0.601	20.1	0	0	1	0	1
Friedman12	42.6	0	3.7	0.977	8.2	0.697	4.6	0.947	3.9	0.974
QS12	13.5	0.001	0	1	0	1	0	1	0	1
Friedman7	12.2	0.049	14.1	0.029	4.1	0.667	8	0.237	0.8	0.993
QS7	0	1	7	0.03	0	1	0	1	0	1

Bike Sharing	Orj_Teststat	Orj_P-value	naive_sa_Teststat	naive_sa_P-value	mstl_sa_Teststat	mstl_sa_P-value	str_sa_Teststat	str_sa_P-value	dsa_sa_Teststat	dsa_sa_P-value
Friedman365	384	0.049	354	0.027	345	0.01	333	0.48	364	0.049
QS365	15.5	0	0	1	0	1	0	1	0	1
Friedman12	17.77	0.087	7.4	0.032	3.2	0.87	11.2	0.424	4.1	0.968
QS12	24.51	0	0	1	5	1	0.2	1	0	1
Friedman7	11.19	0.083	8.3	0.169	7.7	0.03	9.1	0.02	6.3	0.386
QS7	2.16	0.339	0.3	0.15	2.97	0.23	0.3	0.857	0	1

London Bike Sharing	Orj_Teststat	Orj_P-value	naive_sa_Teststat	naive_sa_P-value	mstl_sa_Teststat	mstl_sa_P-value	str_sa_Teststat	str_sa_P-value
Friedman365	87.1	1	478.7	0	300.2	0.994	443.1	0.003
QS365	256.8	0	0	1	0	1	0	1
Friedman12	18.46	0.07	6.6	0.829	0.9	1	0.4	1
QS12	0	1	0	1	0	1	0	1
Friedman7	11.13	0.08	7.6	0.267	3.6	0.736	7.6	0.27
QS7	0	1	0	1	0	1	0	1
Friedman24	117.66	0	115.2	0	3.6	0	105.7	0.06
QS24	60.89	0	49.1	0	0	1	0	1

The table contains the results of seasonal tests performed for various seasonal frequencies to identify the presence of seasonality at different frequencies. Both tests have a null hypothesis that assumes the existence of seasonality. As the table indicates, p-values smaller than 0.05 are highlighted in red.

After analyzing the results of the seasonal tests, we have determined that the Naive method is the most appropriate for the rail traffic data. The MSTL method is effective for weekly series, weekly tracker, and WEI. It also works well for daily series, electricity consumption, CO, and hourly series, such as London bike-sharing data. The DSA method is ideal for daily time series. Meanwhile, the STR method effectively detects seasonality for hourly data, daily series, excluding electricity consumption, and weekly tracker data. Overall, the MSTL and STR methods are reliable for seasonal adjustment of high-frequency time series, while the

DSA method is useful for daily time series. It is important to choose the appropriate methodology for seasonality detection based on the specific features of the data being analyzed. The results of the all series are summarized in Table 4.4

Table 4.4: Performance of Seasonal Adjustment Methods on All Datasets

Data Name	**Frequency**	**Naive**	**MSTL**	**STR**	**DSA**
Rail Traffic	Weekly	**x**			
Weekly Tracker	Weekly		**x**	**x**	
Weekly Economic Index	Weekly		**x**		
PM10	Daily			**x**	**x**
Electricity Consumption Daily	Daily		**x**		**x**
Bike Sharing	Daily			**x**	**x**
CO	Daily		**x**	**x**	**x**
Electricity Consumption Hourly	Hourly			**x**	
London Bike Sharing	Hourly		**x**	**x**	

CHAPTER 5

CONCLUSION AND FUTURE WORK

Seasonal adjustment is a crucial step in analyzing time series data, as it identifies and removes seasonal fluctuations that may obscure the underlying trends and patterns in the data. Seasonal fluctuations can arise from various factors, including weather patterns, holidays, and other periodic events. By removing these fluctuations, seasonal adjustment enables researchers and analysts to understand the long-term trends of the data better and to make more accurate forecasts and projections.

One of the main benefits of seasonal adjustment is that it can reveal meaningful patterns and trends in the data that might otherwise be difficult to discern. For example, it can help identify whether a series is growing or declining over time or whether there are cyclical patterns that repeat at regular intervals. Removing the effects of seasonality can also reduce noise and variability in the data, making it easier to detect trends and changes over time.

Another critical benefit of seasonal adjustment is that it can improve the accuracy of statistical analyses and forecasts. When seasonality is present in a time series, failing to account for it can lead to inaccurate conclusions and forecasts. By removing seasonal fluctuations, seasonal adjustment can help analysts make more accurate forecasts and predictions about future trends and patterns in the data.

Seasonal adjustment is significant for high-frequency time series data, as it can help identify and remove multiple seasonal fluctuations present at shorter time intervals. With the development of technology, high-frequency time series data is increased and often used to follow changes in economic indicators such as stock prices, consumer spending, and employment rates.

Another vital benefit of seasonal adjustment for high-frequency time series data is that it can help to reduce the impact of noise and volatility in the data. High-frequency data is often subject to high randomness and volatility, which can obscure underlying trends and patterns. Seasonal adjustment can smooth out these fluctuations, making it easier to identify trends and patterns that may be present.

This thesis study examines the recent seasonal adjustment methods for high-frequency time series which has multiple seasonal patterns. Each approach has its set of strengths and weaknesses. Although there is no suitable method for every data, a method should be chosen according to the character of each data set.

The moving average method is primarily used for seasonal adjustment of high-frequency time series data. It involves calculating a moving average of the original data over a fixed time window to smooth out the seasonal fluctuations and identify the underlying trend. The method is simple to implement and effective for data with simple seasonal patterns.

However, the moving average method has limitations in dealing with more complex seasonal patterns and may need to be revised for data with irregularities or outliers. The moving average method is valuable for the seasonal adjustment of high-frequency time series data, but its effectiveness may be limited in certain situations. It is often used in combination with other techniques, such as regression or exponential smoothing, to improve the accuracy of seasonal adjustment and forecasting.

The MSTL method is a popular time series decomposition technique for analyzing and forecasting data with seasonal and trend components. Using STL methods, the MSTL method involves decomposing a time series into three components - trend, seasonal, and remainder.

The trend and seasonal components are estimated using local regression with Loess, while the remainder component is obtained by subtracting the trend and seasonal components from the original series. The method has been effective for a wide range of time series data, including non-linear trends and complex seasonal patterns.

The Daily Seasonal Adjustment (DSA) method is useful for adjusting daily time series seasonally. The method employs an iterative STL-based seasonal adjustment routine and a RegARIMA model to estimate calendar and outlier effects. The order of steps in this method differs from other methods, and it is crucial to account for the effects of weekdays using the complete data before any analysis. The STL method is used for deseasonalizing weekly, monthly, and annual effects and for outlier adjustment. The DSA method provides a robust and efficient way to handle seasonal adjustment in high-frequency daily time series data.

The STR method is a robust and flexible method for seasonal adjustment of high-frequency time series. It is a generalization of the classical STL method, combining it with regression techniques to handle trends and other factors. The STR method decomposes a time series into trend, seasonal, and remainder components. This decomposition is based on fitting a regression model to the time series. The method has several advantages, such as handling complex seasonal patterns, accommodating multiple seasonal cycles, and handling missing values and outliers robustly. The main steps of the STR method involve fitting a linear regression model with the time series as the response variable and covariates representing the trend, seasonality, and any additional effects. Using the STL method, the regression model is then used to obtain residuals, which are further decomposed into trend and seasonal components.

The STR method has a longer processing time than other methods, and it is unsuitable for analyzing large datasets. Specifically, the STR method cannot handle data spanning more than two years for hourly time series data, whereas other methods, such as MSTL and DSA, can handle long time series. Therefore, when dealing with large datasets, it is crucial to consider the limitations of the STR method and choose alternative methods better suited for the task.

To sum up, seasonal adjustment of high-frequency time series with multiple seasonal patterns is essential for the success of other time series analysis methods, such as forecasting, because seasonal adjustment is a pre-adjustment step in time series analysis.

In conclusion, seasonal adjustment is essential in analyzing time series data, especially high-frequency data, to remove seasonal fluctuations and reveal underlying trends and patterns. The choice of seasonal adjustment method depends on the characteristics of the data, with each method having its unique strengths and weaknesses. The moving average, MSTL, DSA, and STR methods are among the popular techniques used for seasonal adjustment of high-frequency time series data. Overall, the appropriate seasonal adjustment technique can significantly improve the accuracy of statistical analyses and forecasting of high-frequency time series data.

In future work, Deep Learning models, such as LSTM, GRU, and other deep learning models can be used for seasonal adjustment. Since high-frequency data generally have many observations, the Deep Learning models can work better. They can capture complex patterns in the data and can handle large amounts of data. State Space Models are also used to detect the seasonal patterns of the data. Wavelet transform can also be used for the seasonal adjustment of high-frequency data. These methods can be studied in the future.

REFERENCES

[1] *OECD*, 2023. Weekly Tracker of GDP Growth, OECD Economic Outlook, OECD Publishing, Paris, https://www.oecd.org/economy/weekly-tracker-of-gdp-growth.htm.

[2] *Transport for London (TfL)*, 2023. https://tfl.gov.uk.

[3] *US Weekly Rail Traffic*, 2023. YCharts.https://ycharts.com/indicators/us-weekly-rail-traffic.

[4] *Weekly Economic Index*, 2023. Federal Reserve Bank of New York. https://www.newyorkfed.org/research/policy/weekly-economic-index.

[5] *Bike Sharing Dataset.*, 31 December 2022. University of California, Irvine, Machine Learning Repository. https://archive.ics.uci.edu/ml/datasets/bike+sharing+dataset.

[6] *EPİAŞ Şeffaflık Platformu, Tüketim - Gerçekleşen Tüketim*, 31 December 2022. https://seffaflik.epias.com.tr/transparency/tuketim/gerceklesen-tuketim/gercek-zamanli-tuketim.xhtml.

[7] *T.C. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı, SİM-Sürekli İzleme Merkezi*, 31 December 2022. https://sim.csb.gov.tr/Services/AirQuality.

[8] M. Armatte. *Conjonctions, Conjoncture et Conjecture*, 1992. Les baromÃ´ltres Ãl'conomiques (1885-1930), Histoire et Mesure, 7 (1), 99-149 (in french).

[9] K. Bandara, R. J. Hyndman, and C. Bergmeir. *MSTL: A Seasonal-Trend Decomposition Algorithm for Time Series with Multiple Seasonal Patterns*, July 2021. arXiv preprint arXiv:2107.13462.

[10] G. A. Barnard, G. M. Jenkins, and C. B. Winsten. Exact likelihood inference and time series. *Journal of the Royal Statistical Society: Series A (General)*, 125(3):321–352, 1962.

[11] W. Bell. Signal extraction for nonstationary time series. *Annals of Statistics*, 12(1):646–664, 1984.

[12] W. R. Bell and S. C. Hillmer. Issues involved with the seasonal adjustment of economic time series. *Journal of Business and Economic Statistics*, 2(4):291–320, 1984.

[13] G. E. P. Box, S. C. Hillmer, and G. C. Tiao. *Analysis and Modeling of Seasonal Time Series*, 1978. in Seasonal Analysis of Economic Time Series, Zellner (ed.), US Bureau of the Census, Washington DC.

[14] J. P. Burman. Seasonal adjustment by signal extraction. *Journal of the Royal Statistical Society: Series A (General)*, 143(3):321–337, 1980.

[15] C. H. D. Buys-Ballot. *Les Changements Périodiques de Température*, 1847. Kemink and Fils, Utrecht.

[16] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. J. Terpenning. STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6(1):3–33, 1990.

[17] W. S. Cleveland, D. M. Dunn, and I. J. Terpenning. *SABL: A Resistant Seasonal Adjustment Procedure with Graphical Methods for Interpretation and Diagnosis,*. dans Seasonal Analysis of Economic Time Series, Zellner (ed.), US Bureau of the Census, Washington DC., 1978.

[18] E. B. Dagum. Ajustement saisonnier et problèmes de sÃl'ries temporelles. *Economie appliquée Paris*, 32(1):5–129, 1979.

[19] E. B. Dagum. *Seasonality*, 1986. In Encyclopedia of Statistical Sciences. Kots and Johnson (eds.), Vol. 5, John Wiley and Sons, New York.

[20] E. B. Dagum. Time series: Seasonal adjustment. *Journal of the Royal Statistical Society Series A*, 143:321–337, 2001.

[21] M. Dancho and D. Vaughan. *Package 'anomalize'*, 2018. R package version 0.2.2.

[22] O. Darné. *La désaisonnalisation des chroniques économiques : Analyse des conditions conjoncturelles*, 2003. PhD Thesis, Department of Economics, University of Montpellier I (in french).

[23] O. Darné, L. Ferrara, and D. Ladiray. *A Brief History of Seasonal Adjustment Methods and Software Tools*. June 2018.

[24] A. M. De Livera, R. J. Hyndman, and R.D. Snyder. Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American statistical association*, 106(496):1513–1527, 2011.

[25] A. Dokumentov and R. J. Hyndman. *STR: Seasonal-Trend Decomposition Using Regression*, 2021. arXiv preprint arXiv:2009.05894.

[26] A. Dokumentov and R. J. Hyndman. *stR: STR Decomposition*, 2022. R package version 0.5.

[27] G. Dudek. *STD: A Seasonal Trend Dispersion Decomposition of Time Series*, 2022. arXiv preprint arXiv:2204.10398.

[28] H. D. Falkner. The measurement of seasonal variation. *Journal of the American Statistical Association*, 19(146):167–179, 1924.

[29] D. F. Findley, B. C. Monsell, W. R. Bell, M. C. Otto, and B. C. Chen. New capabilities and methods of the x-12-arima seasonal adjustment program. *Journal of Business and Economic Statistics*, 16(2):127–152, March 1998.

[30] B. Fischer. *Decomposition of Time Series: Comparing Different Methods in Theory and Practice*, 1995. Mimeo, Eurostat, Luxembourg.

[31] A. Fisher. A brief note on seasonal variation. *Journal of Accountancy*, 64(1):54–59, 1937.

[32] M. Friedman. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association (American Statistical Association)*, 32(200):675–701, December 1937.

[33] E. Ghysels and D. R. Osborn. *The Econometric Analysis of Seasonal Time Series*. Cambridge University Press: Cambridge, 2001.

[34] V. Gómez and A. Maravall. *Time series regression with ARIMA noise and missing observations: Program TRAM*, 1992. Working Paper ECO No 92-81, Department of Economics, European University Institute.

[35] V. Gómez and A. Maravall. *Programs TRAMO and SEATS: Instructions for the user (beta version: June 1997)*, 1997. Working Paper No 97001, Dirección General de Anélisis y ProgramaciÃ¸sn Presupuestaria, Ministerio de Economía y Hacienda, Madrid.

[36] E. J. Hannan. The estimation of seasonal variation. *Australian Journal of Statistics*, 2(1):1–15, 1960.

[37] E. J. Hannan. Measurement of a wandering signal amid noise. *Journal of Applied Probability*, 4:90–102, 1967.

[38] S. C. Hillmer and G. C Tiao. An arima-model based approach to seasonal adjustment. *Journal of the American Statistical Association*, 77:63–70, 1982.

[39] P. J. Huber. Robust estimation of a location parameter. *Annals of Mathematical Statistics*, 35(1):73–101, March 1964.

[40] S. Hylleberg. *Modelling Seasonality*. Oxford University Press, New York, 1992.

[41] S. Hylleberg. Econometrics and the philosophy of economics. *The Scandinavian Journal of Economics*, 108(2):357–359, July 2006.

[42] R. Hyndman, G. Athanasopoulos, C. Bergmeir, G. Caceres, L. Chhay, K. Kuroptev, M. OHara-Wild, F. Petropoulos, S. Razbash, E. Wang, and F. Yasmeen. *forecast: Forecasting Functions for Time Series and Linear Models*, 2020. R package version 8.13.

[43] R. Hyndman, G. Athanasopoulos, C. Bergmeir, G. Caceres, L. M. Chhay, F. O'Hara-Wild, Petropoulos, S. Razbash, E. Wang, and F. Yasmeen. *fore- cast: Forecasting functions for time series and linear models*, 2022. R package version 8.9.

[44] R. J. Hyndman and G. Athanasopoulos. *Forecasting: Principles and Practice*, 2018.

[45] R. J. Hyndman and Y. Khandakar. Automatic time series forecasting: The forecast package for R. *Journal of statistical software*, 27:1–22, 2008.

[46] B. Iglewicz and D. C. Hoaglin. How to detect and handle outliers. *Quality Press*, 16, 1993.

[47] A. Joy and W. Thomas. The use of moving averages in the measurement of seasonal variations. *Journal of the American Statistical Association*, 23(163):241–252, September 1928.

[48] W. W. I. King. An improved method for measuring the seasonal factor. *Journal of the American Statistical Association*, 19(147):301–313, 1924.

[49] S. J. Koopman, A. C. Harvey, J. A. Doornik, and N. Shephard. *STAMP 5.0 Structural Time Series Analyser*, 1995. Modeller and Predictor, Chapman and Hall, London.

[50] S. Kuznets. Seasonal pattern and seasonal amplitude: Measurement of their short-time variations. *Journal of the American Statistical Association*, 27(177):9–20, March 1932.

[51] D. Ladiray, J. Palate, G. Mazzi, and T. Proietti. *Seasonal Adjustment of Daily Data*, 2018. 16th Conference of the International Association of Official Statisticians (IAOS), Paris, France.

[52] D. Ladiray and B. Quenneville. *Seasonal Adjustment with the X-11 Method*, 1999 and 2001. Lecture Notes in Statistics No 158, Springer, New York.

[53] R. O. D. Lane. *Minimal-Revisions Trend Estimates*. Statistica Office, London, Research Exercise Note 8/72, 1972.

[54] G. M. Ljunk and G. E. P. Box. On a measure of a lack of fit in time series models. *Biometrika*, 65(2):297–303, 1978.

[55] J. L´opez. *tsoutliers: R Package for Automatic Detection of Outliers in Time Series*, 2015.

[56] A. Maravall and D. A. Pierce. A prototypical seasonal adjustment model. *Journal of Time Series Analysis*, 8(2):177–193, 1987.

[57] H. Mendershausen. Eliminating changing seasonals by multiple regression analysis. *The Review of Economic Studies*, 21(1):171–177, 1939.

[58] A. Mishra, R. Sriharsha, and S. Zhong. *OnlineSTL: Scaling Time Series Decomposition by 100x*, 2021. arXiv preprint arXiv:2107.09110.

[59] B. C. Monsell. *The X-13A-S Seasonal Adjustment Program*, 2007. Proceedings of the 2007 Federal Committee On Statistical Methodology Research Conference.

[60] M. Nerlove, D. M. Grether, and J. L. Carvalho. *Analysis of Economic Time Series: A Synthesis*. York Academic Press, New York, 1979.

[61] M. O'Haara-Wild, R. Hyndman, and E. Wang. *fable: Forecasting Models for Tidy Time Series*, 2022. R package version 0.3.2.

[62] D. Ollech. *dsa: Seasonal Adjustment of Daily Time Series*, 2021. R package version 1.0.12.

[63] D. Ollech. Seasonal adjustment of daily time series. *Journal of Time Series Econometrics*, 13(2):235–264, 2021.

[64] D. Ollech and D. Bundesbank. Economic analysis using higher-frequency time series: Challenges for seasonal adjustment. *Empirical Economics*, pages 1–24, 2022.

[65] W. M. Persons. Indices of business conditions: An index of general business conditions. *The Review of Economic Statistics*, 1(1):5–107, 1919.

[66] A. Quartier-la Tente, A. Michalek, J. Palate, and R. Baeyens. *RJDemetra: Interface to 'JDemetra+' Seasonal Adjustment Software*, 2022. R package version 0.2.1.

[67] S. Sharma, V. Rana, and Kumar V. Deep learning based semantic personalized recommendation system. *International Journal of Information Management Data Insights*, 1(2):501–514, 2021.

[68] D. Shibitov and M. Mamedli. *Forecasting Russian CPI with Data Vintages and Machine Learning Techniques*, 2021. IFC-Bank of Italy Workshop on "Machine learning in central banking", 19-22 October 2021, Rome / virtual event.

[69] J. Shiskin. *The X11 Variant of the Census Method II Seasonal Adjustment Program*. Technical Paper No 15, US Bureau of the Census, Washington DC, February 1967.

[70] A. Smyk, A. Quartier-la Tente, Barthelemy T., and Webel K. *JDemetra+ documentation*, 2023.

[71] S. J. Taylor and B. Letham. Forecasting at scale. *The American Statistician*, 72(1):37–45, 2018.

[72] R Core Team. *R: A Language and Environment for Statistical Computing*, 2020. R Foundation for Statistical Computing, Vienna, Austria.

[73] M. Timmermans, R. Heijmans, and H. Daniels. *Cyclical Patterns in Risk Indicators Based on Financial Market Infrastructure Transaction Data*, 2017. De Nederlandsche Bank Working Paper 558.

[74] H. Xu, X. Zhang, D. Liang, and G. Wang. Robust-STP: A robust seasonal-trend decomposition method for partial periodic time series. *In 2022 IEEE 8th International Conference on Cloud Computing and Intelligent Systems (CCIS)*, pages 691–695, November 2022.

[75] T. Young. *Seasonal Adjusting of Flow of Fund Matrices*. Internal Mimeo, Bank of England, 1992.

[76] G. P. Zhang and M. Qi. Neural network forecasting for seasonal and trend time series. *European Journal of Operational Research*, 160(2):501–514, 2005.

CHAPTER 6

APPENDIX

Given the large dataset used in this thesis, it is only feasible to include some of the analysis results in the main text. Therefore, some analysis results are provided in the appendix for curious readers. The appendix contains a selection of results from the seasonal adjustment of the data using different methods.

6.1 Weekly Time Series

6.1.1 The US Weekly Rail Traffic

The US Weekly Rail Traffic tracks the volume of goods transported by rail in the US weekly. It is viewed as a trustworthy indicator of economic activity because it images the movement of raw materials, commodities, and finished goods across the country. The US Weekly Rail Traffic data has become increasingly important in recent years due to its effectiveness in predicting economic growth and identifying potential financial risks. For instance, changes in rail traffic volume can predict shifts in consumer demand or changes in industrial production, which can affect GDP growth and other economic indicators. The correct analysis and interpretation of this series are of great importance to economists, policymakers, and analysts. The decisions on economy can be mistakenly taken for underlying trends if an increase or decrease due to seasonality is misinterpreted. Being able to decompose the series properly will also be helpful for forecasting.

In order to better understand and interpret the data, a time series plot is shown in Figure 6.1.

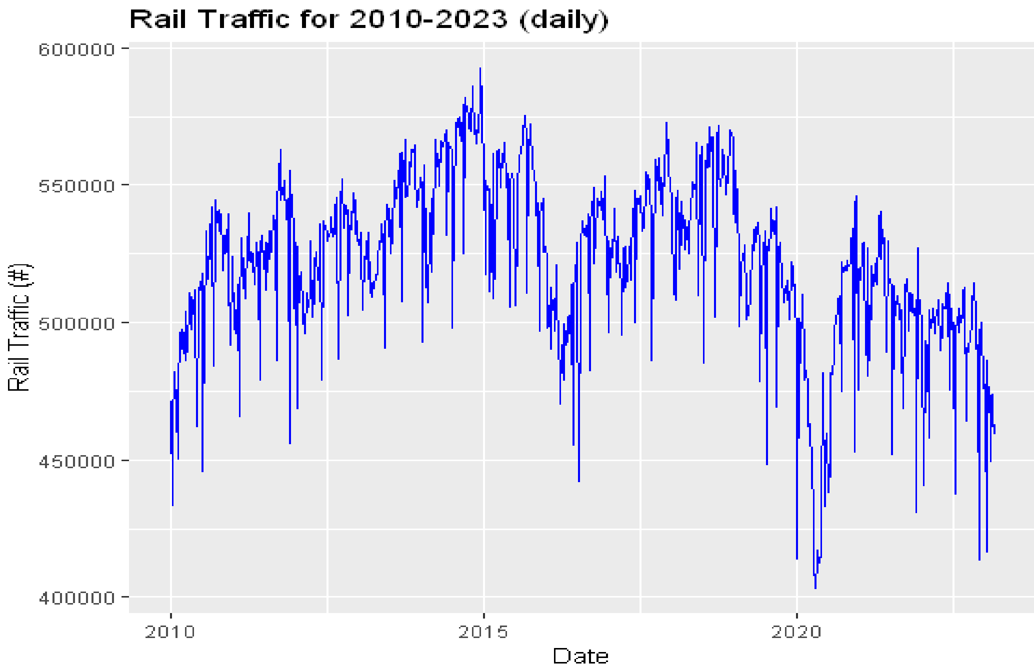


Figure 6.1: Time Series Plot of the US Weekly Rail Traffic Data

As seen in Figure 6.1, the data has periodic movements. It may have multiple seasonal patterns.

ACF and PACF graphs are also essential to understand the data structure. For this reason, Figure 6.2 shows the ACF and PACF graphs of rail traffic data.

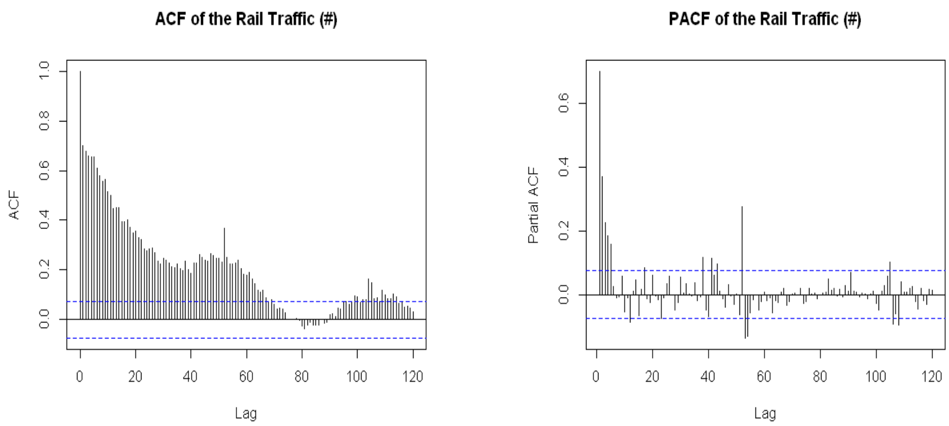


Figure 6.2: The ACF and PACF Graphs of the US Weekly Rail Traffic Data

According to ACF and PACF graphs, the lag 52 and its multiples have significant lags. Since one year includes 52 weeks, it may have yearly seasonality.

To check multiple seasonality, the Box-Plots for the monthly and yearly intervals can help better sentiment to the seasonal data structure.

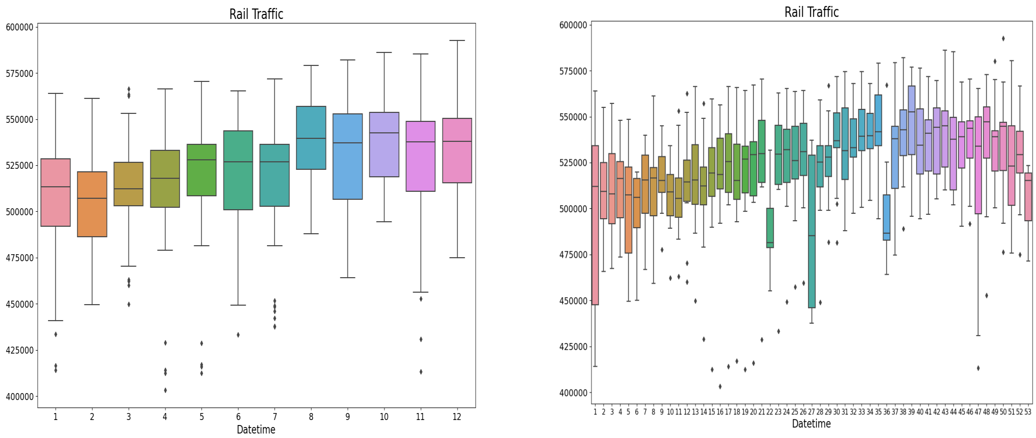


Figure 6.3: The Box-Plots of the US Weekly Rail Traffic Data Monthly and Yearly

Figure 6.3 the median of box plots changed in August, while the median was stable for the first seven months. So, there may be periodic movement in summer and winter. The median and the lower quartile change during the year in the yearly plot. The upper quartile is below the remaining weeks in some weeks, such as 22, 27, and 36.

The time series is imputed for the pre-processing step, as mentioned in Chapter 3.5. The data also have anomalies. The anomaly detection result is shown in Figure 6.4.

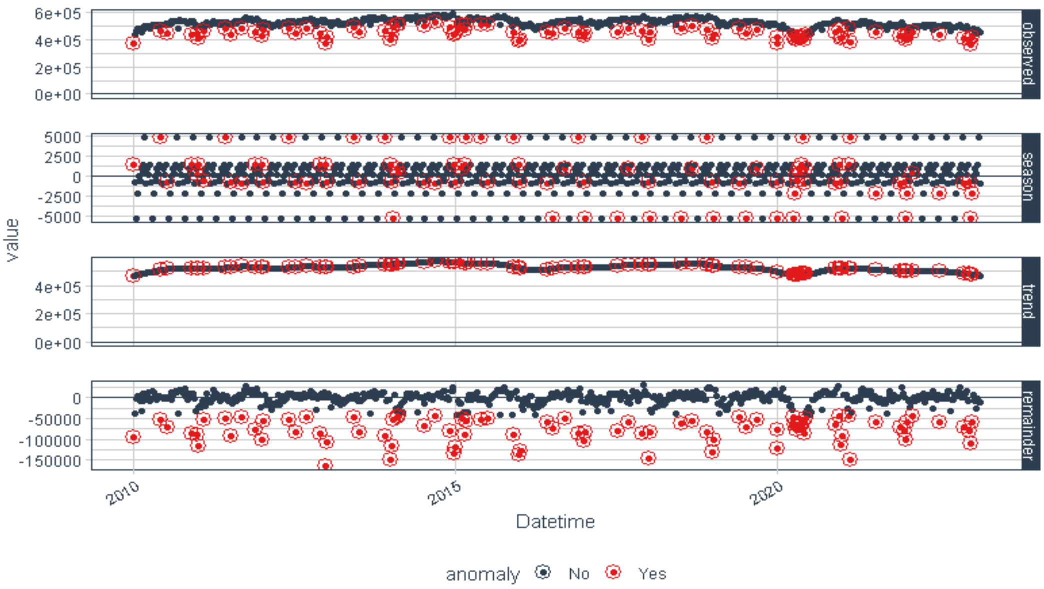


Figure 6.4: The Anomaly Detection Results of the US Weekly Rail Traffic Data

Naive Method Result

The naive method is applied to the data. It uses the basic moving average method, as mentioned in Chapter 3.1. The seasonal periods are operated monthly $(30/7 = 4)$ and yearly $(365/7 = 52)$.

The results of the naive method for rail traffic data are given in Figure 6.5.

Figure 6.5: The Naive Method Results for the US Weekly Rail Traffic Data

When examining the results, the rail traffic data shows strong monthly and yearly seasonality. The rail traffic data peaks at the end of the month while it troughs at the end of the year. Since the moving average method assumes the existence of seasonality, the decomposition always has strong seasonal patterns. In the residual component, there are similar movements year by year. There are two possible reasons for that. There may be residual seasonality or this similar movement caused by the cycle component of the time series.

<u>MSTL Result</u>

The MSTL method is applied to the data, the details are mentioned in Chapter 3.2. The seasonal periods are operated monthly and yearly. The results of the MSTL method for rail traffic data are given in Figure 6.6.

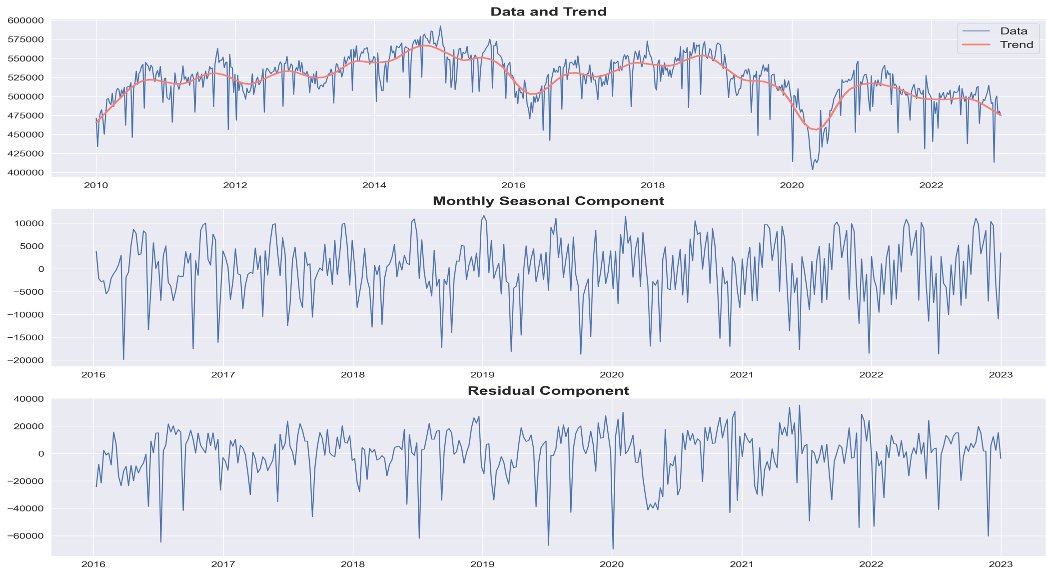


Figure 6.6: The MSTL Method Results for the US Weekly Rail Traffic Data

The MSTL method says there is no multiple seasonality. So, there is only one seasonal component. Moreover, when the residual component is examined, there is no evidence of any periodic pattern. It is more accurate than the naïve method.

STR Result

As specified in Chapter 3.4, the STR method is applied to the data. The STR method has a "gapCV" parameter set to 52 for rail traffic data.

The results of the STR method for rail traffic data are given in Figure 6.7.

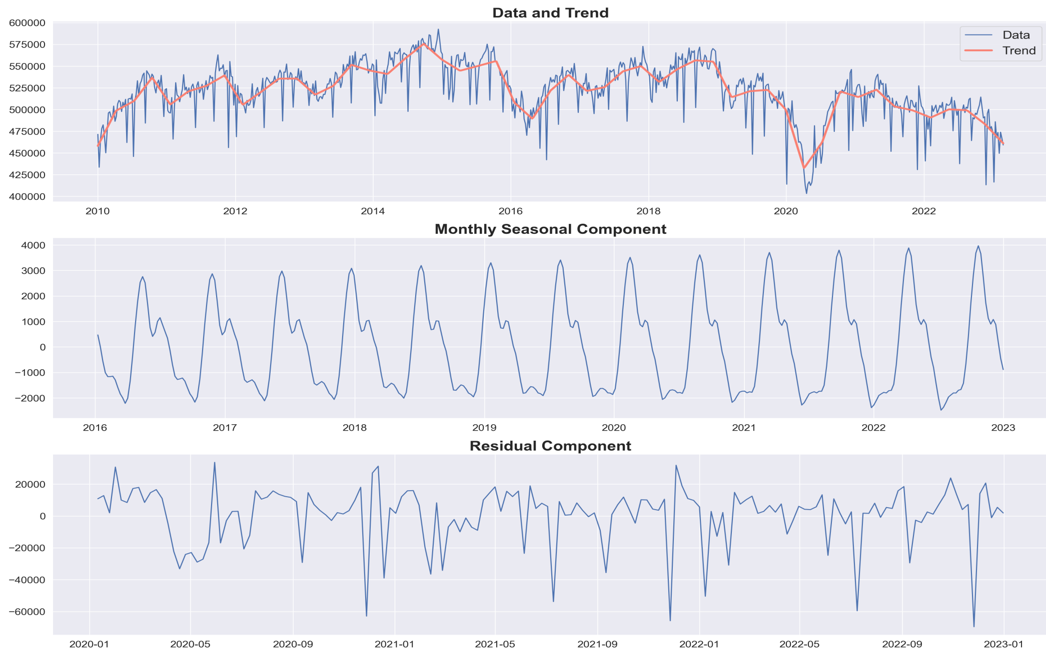


Figure 6.7: The STR Method Results for the US Weekly Rail Traffic Data

Like the MSTL method, the STR method couldn't find the multiple seasonality. Monthly seasonality is smoother than the MSTL method and also the residual component.

Result for the US Weekly Rail Traffic

Figure 6.8 shows the results of all the seasonal adjustment methods for rail traffic.

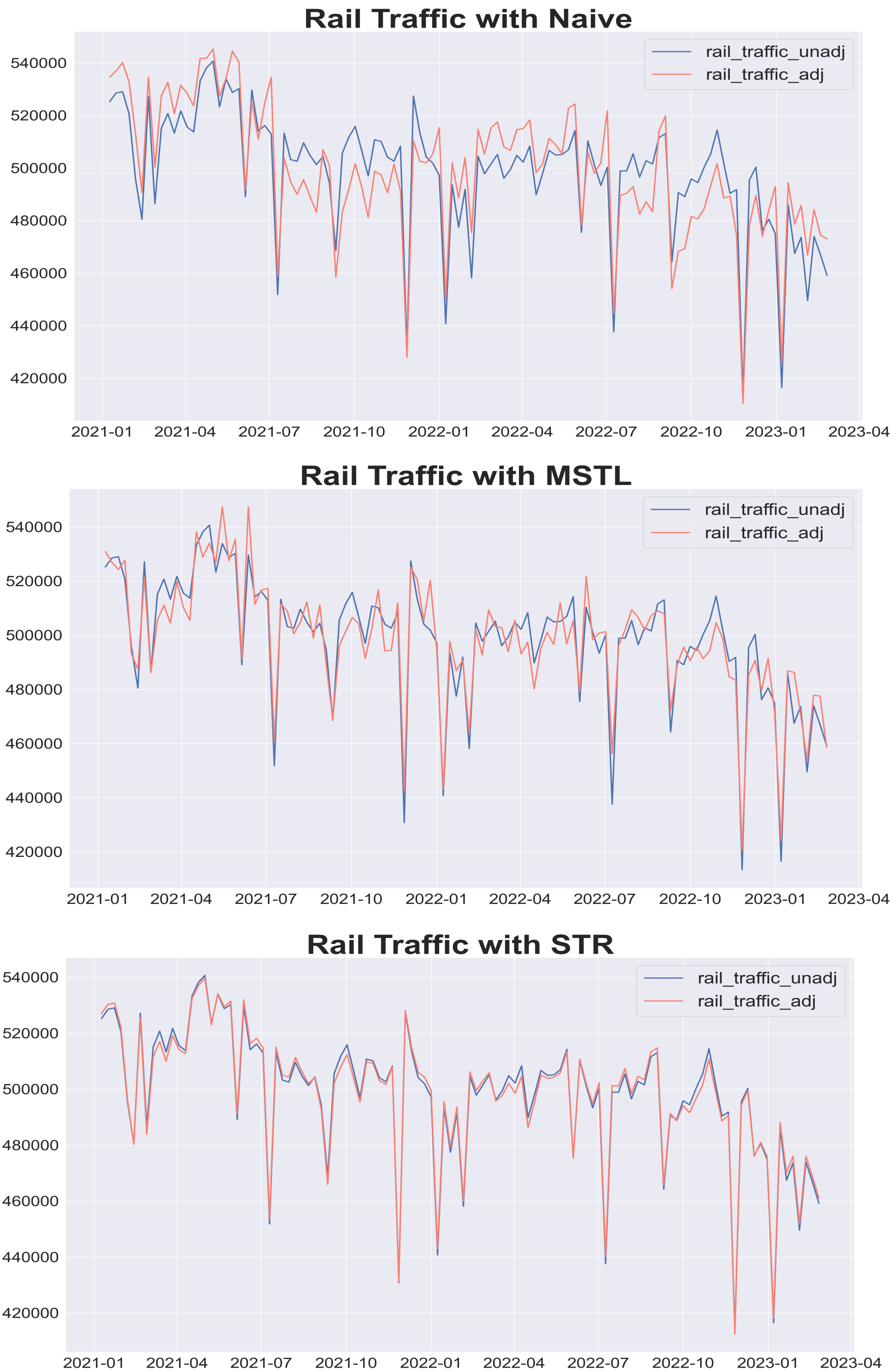


Figure 6.8: The Seasonally Adjusted Series Result of All Methods for the US Weekly Rail Traffic Data

Based on the graphical representations, it can be observed that the STR method failed to decompose the seasonal effect of the data accurately. In contrast, the results of the other methods were similar.

The ACF and PACF plots for the seasonally adjusted series are shown below.

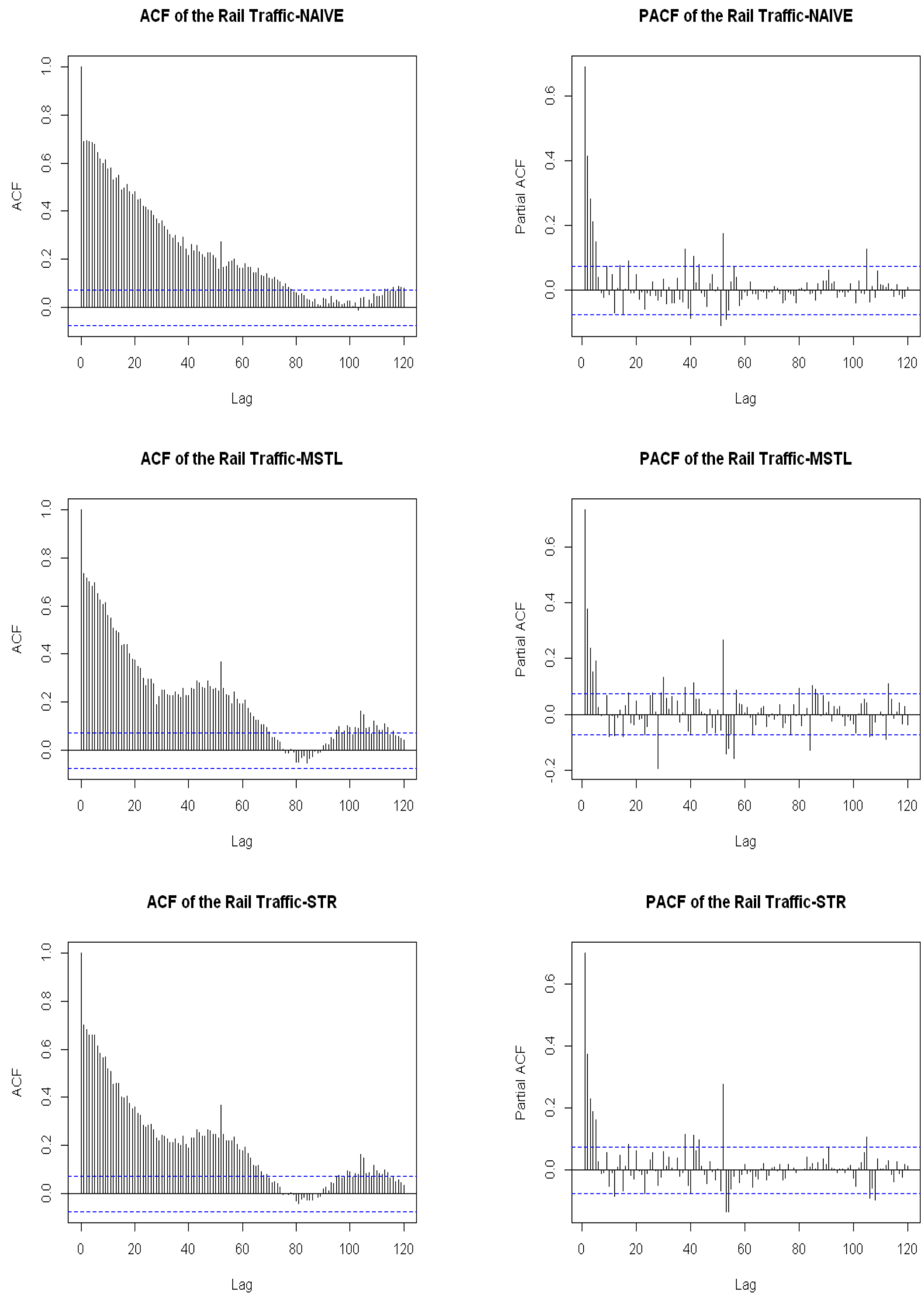


Figure 6.9: The Seasonally Adjusted Series ACF-PACF Plots for the US Weekly Rail Traffic Data

The ACF plots exhibit similar behaviors for both the MSTL and STR methods, but the PACF plot of the MSTL method displays more significant lags at seasonal lags.

6.1.2 Weekly GDP Tracker of Türkiye

OECD releases a weekly tracker of GDP for select countries, including Türkiye. This data provides valuable insights into these countries' real-time economic activity and growth. The weekly tracker offers an up-to-date picture of the economy and is a valuable tool for policymakers, economists, and investors to monitor financial performance. By providing timely and accurate information, the OECD's weekly tracker of GDP plays an essential role in understanding the economic health of the countries.

Figure 6.10 provides a graphical representation of the data over time.

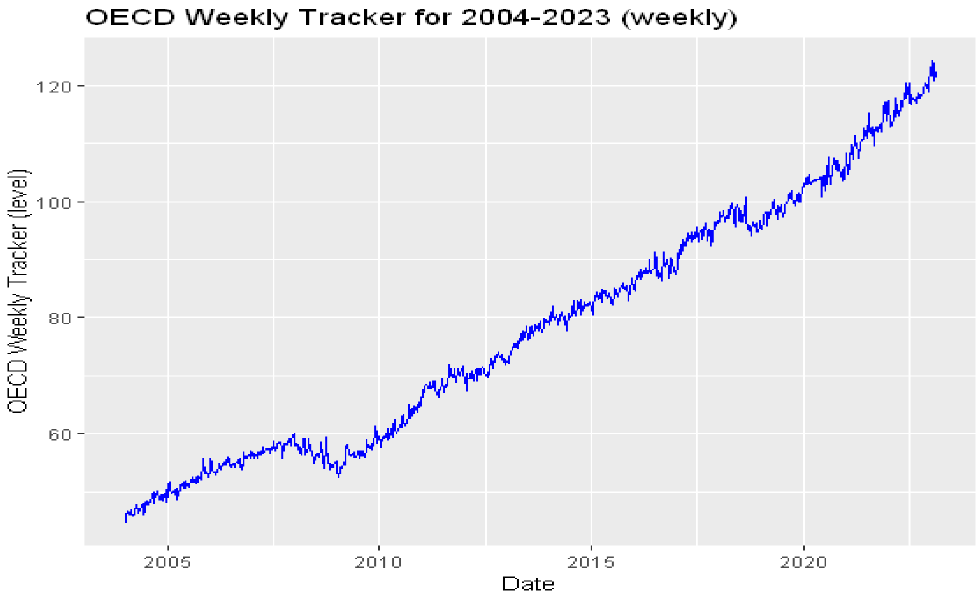


Figure 6.10: Time Series Plot of the Weekly Tracker of GDP for Türkiye

The time series has an increasing trend. There are multiple seasonal behaviors.

To comprehend the data structure, it is necessary to examine the ACF and PACF graphs. Thus, in Figure 6.11, the ACF and PACF graphs of Weekly Tracker are presented.

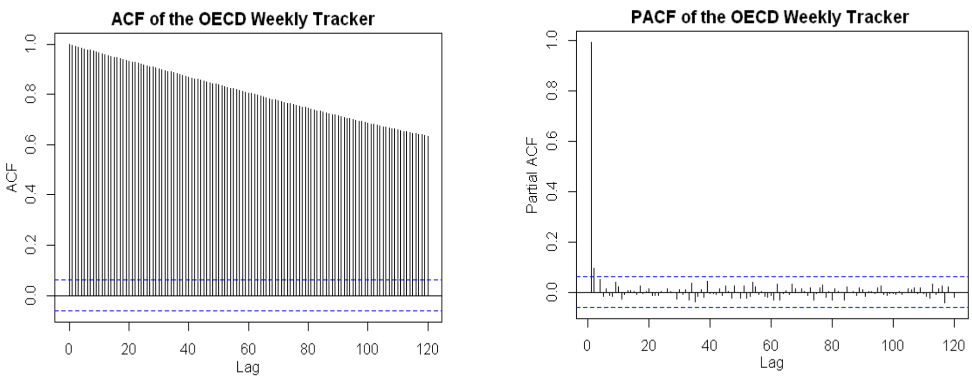


Figure 6.11: The ACF and PACF Graphs of the Weekly Tracker of GDP for Türkiye

According to the PACF graph, there are no significant lags where the ACF graph slow decay.

The box-plots provide a graphical representation of data distribution across different periods, allowing for easy identification of seasonal patterns and anomalies.

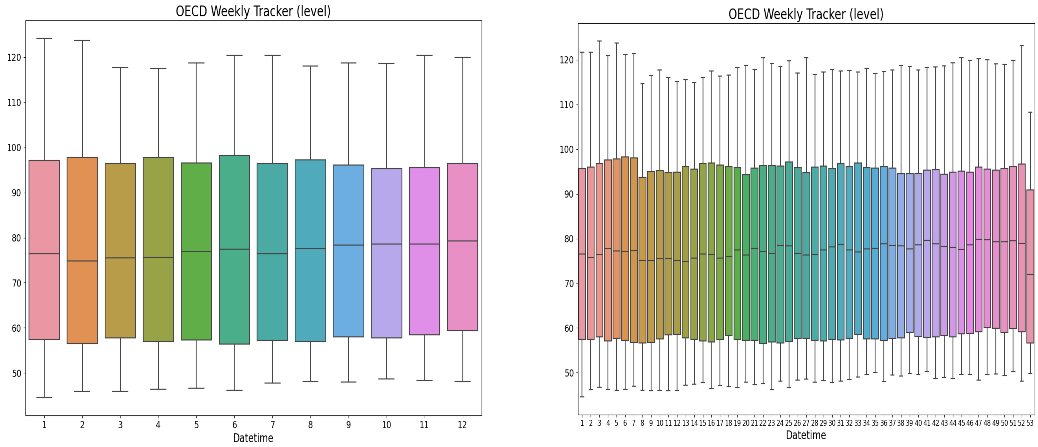


Figure 6.12: The Box-Plots of the Weekly Tracker of GDP for Türkiye as Monthly and Yearly

In Figure 6.12, it can be observed that the median and upper quartile of the box plots do not exhibit significant changes monthly, and the plots appear to be highly symmetric. However, yearly, the median does show some variation over time. Additionally, the length of the upper quartile is higher during the first few weeks of the year compared to the other weeks.

The data contains anomalies which are illustrated in the anomaly detection outcome displayed in Figure 6.13.

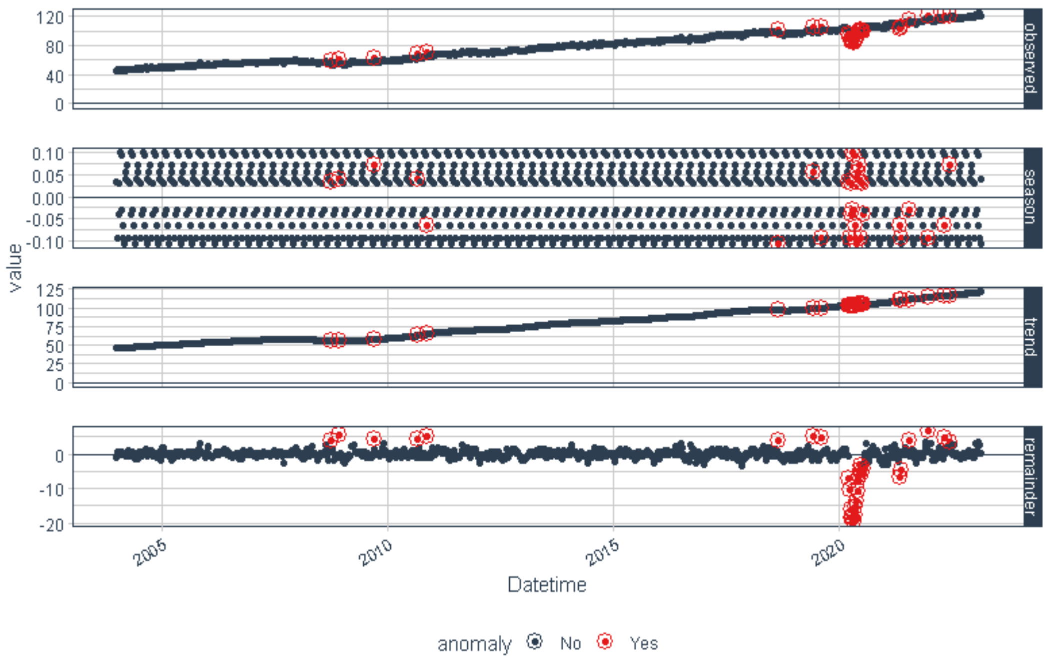


Figure 6.13: The Anomaly Detection Results of the Weekly Tracker of GDP for Türkiye

Naive Method Result

The seasonal periods of 4 for monthly and 52 for yearly are used. The outcomes of the naive method on weekly tracker data are presented in Figure 6.14.

Figure 6.14: The Naive Method Results for the Weekly Tracker of GDP for Türkiye

After examining the results, it can be observed that the weekly tracker data shows strong monthly and yearly seasonality patterns. In particular, the weekly tracker data hits its minimum point at the start of each month and the beginning of each year.

MSTL Result

Figure 6.15 displays the results obtained by applying the MSTL method to the weekly tracker data.

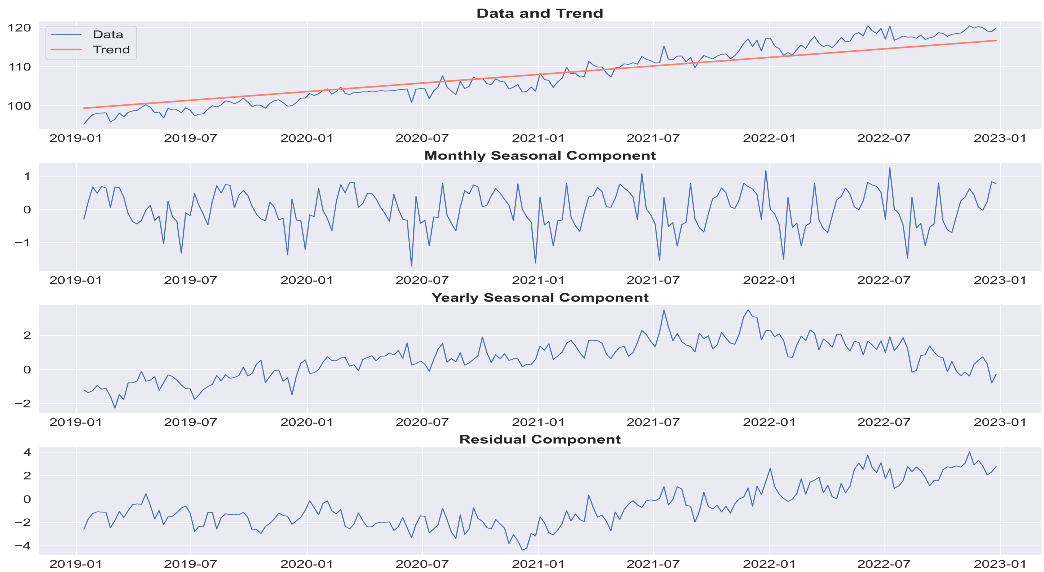


Figure 6.15: The MSTL Method Results for the Weekly Tracker of GDP for Türkiye

The monthly seasonal component in the MSTL results exhibits a strong pattern, but the yearly seasonal component appears unstable. A newly seasonal effect is formed at the end of the residual component.

STR Result

The outcomes of the STR method with a "gapCV" parameter of 52 on the WEI data are presented in Figure 6.16.

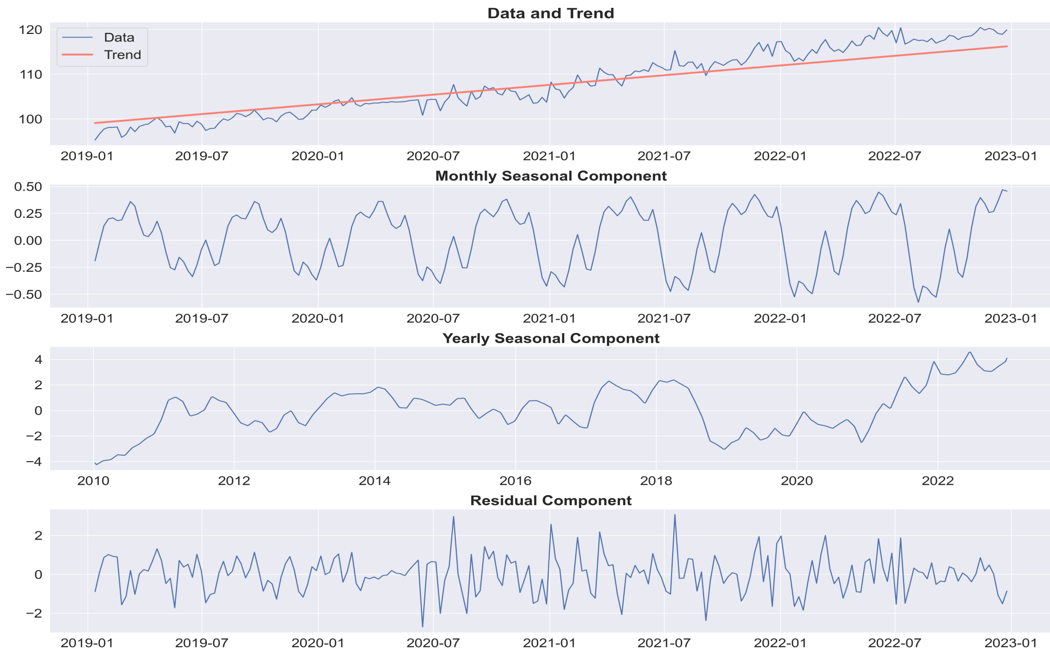


Figure 6.16: The STR Method Results for the Weekly Tracker of GDP for Türkiye

Like the MSTL method, the STR method can decompose irregular yearly seasonal patterns. Moreover, the residual component of the STR method is free from seasonal fluctuations.

Result for the Weekly GDP Tracker of Türkiye

Figure 6.17 presents the results of all the seasonal adjustment methods for weekly tracker.

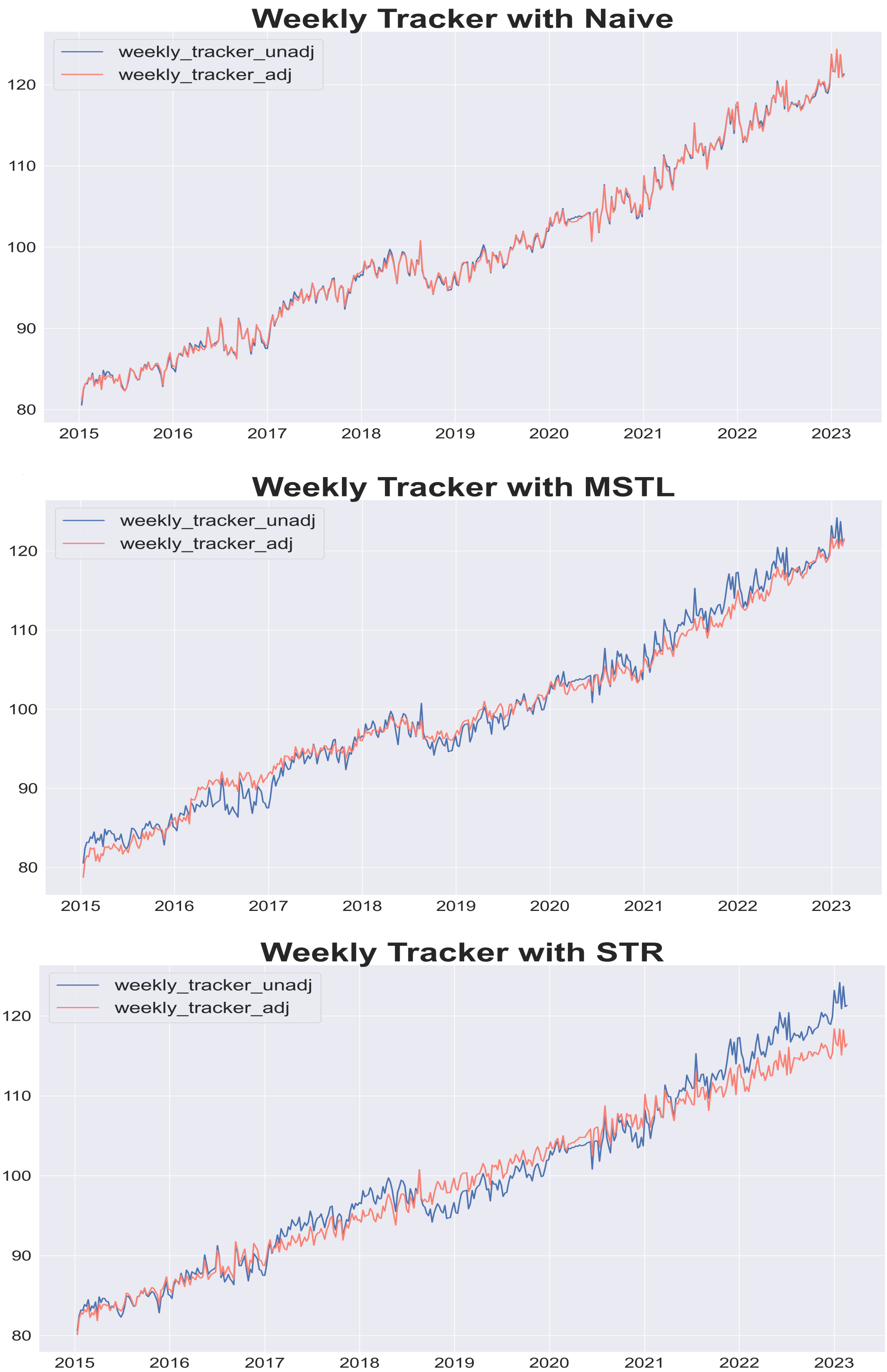


Figure 6.17: The Seasonally Adjusted Series Result of All Methods for the Weekly Tracker of GDP for Türkiye

The outcomes of the different methods appear comparable except for the naive method. The

STR method is ineffective in decomposing the data's seasonal effect.

The ACF and PACF graphs of the seasonally adjusted series are depicted below.

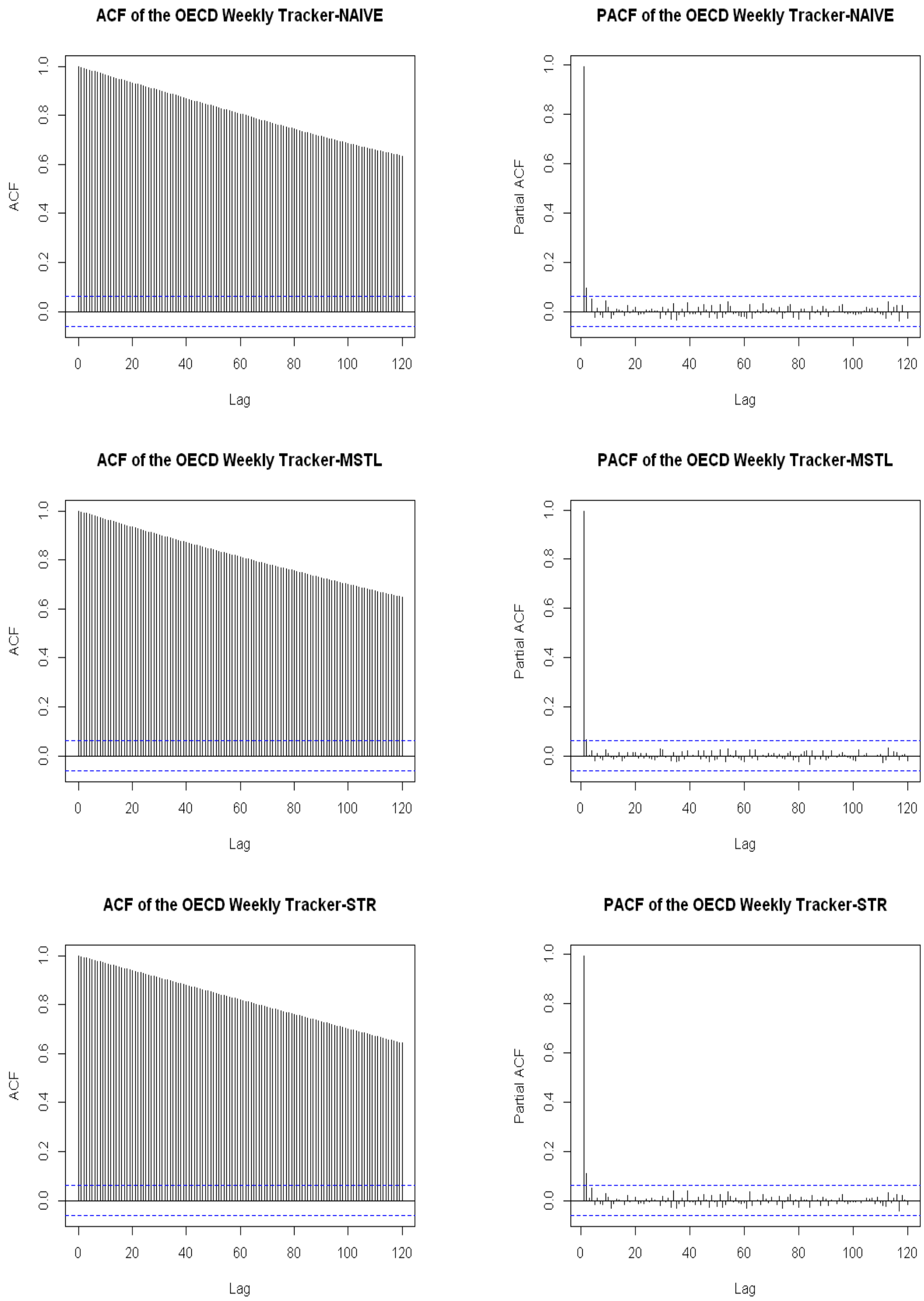


Figure 6.18: The Seasonally Adjusted Series ACF-PACF Plots for the Weekly Tracker of GDP for Türkiye

The ACF and PACF graphs exhibit the same patterns.

6.2 Daily Time Series

6.2.1 PM10 in Ankara

PM10 is a type of air pollutant that consists of fine particles with a diameter of 10 micrometers or less. These particles are known to harm human health and the environment. Therefore, the monitoring and analysis of PM10 levels are critical for public health, as exposure to high levels of PM10 can lead to respiratory and cardiovascular issues. PM10 data has been collected and analyzed to identify patterns and trends that can aid in policymaking and reduce air pollution's negative impact.

Figure 6.19 displays a time series plot which is presented to aid in comprehending and interpreting the data.

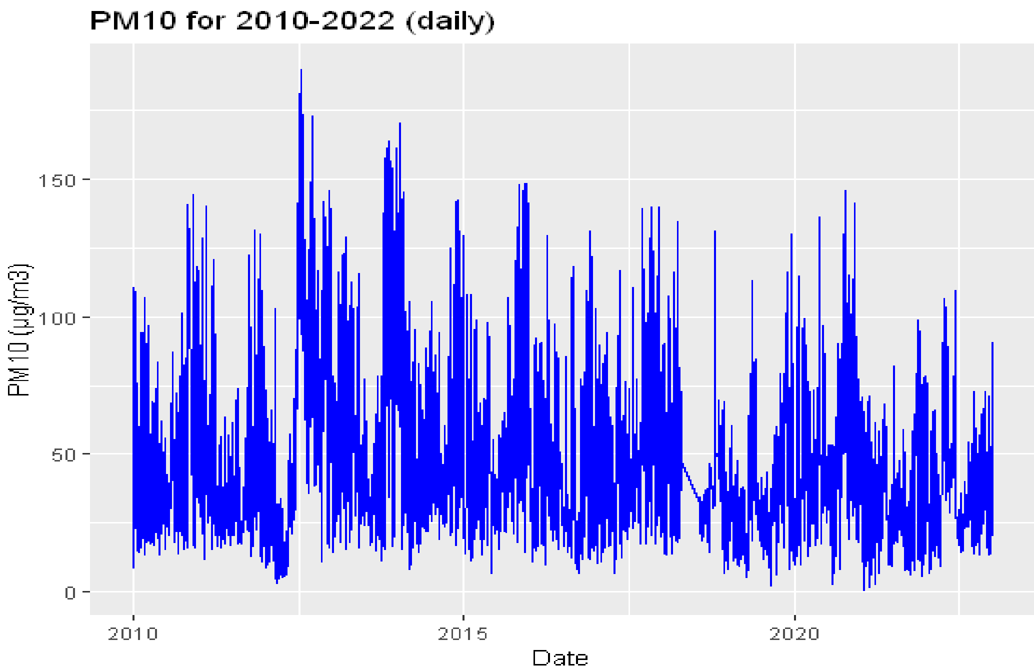


Figure 6.19: Time Series Plot of PM10 in Ankara

The graph shows PM10 is very volatile and may have multiple seasonal patterns.

Understanding the data structure requires the use of ACF and PACF graphs, so the ACF and PACF graphs of PM10 are illustrated in Figure 6.20.

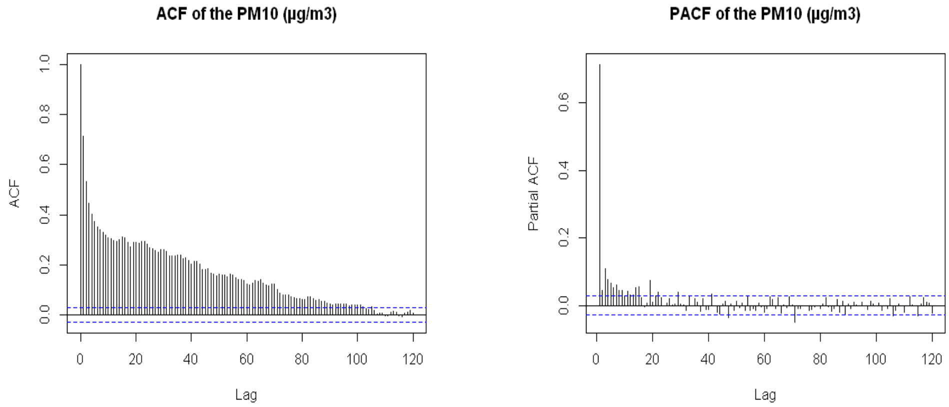


Figure 6.20: The ACF and PACF Graphs of PM10 in Ankara

The time series exhibits significant lags in both the ACF and PACF graphs, indicating the possibility of multiple seasonal patterns in the data.

Box-Plots for weekly, monthly and yearly intervals can provide a better understanding of the seasonal data structure.

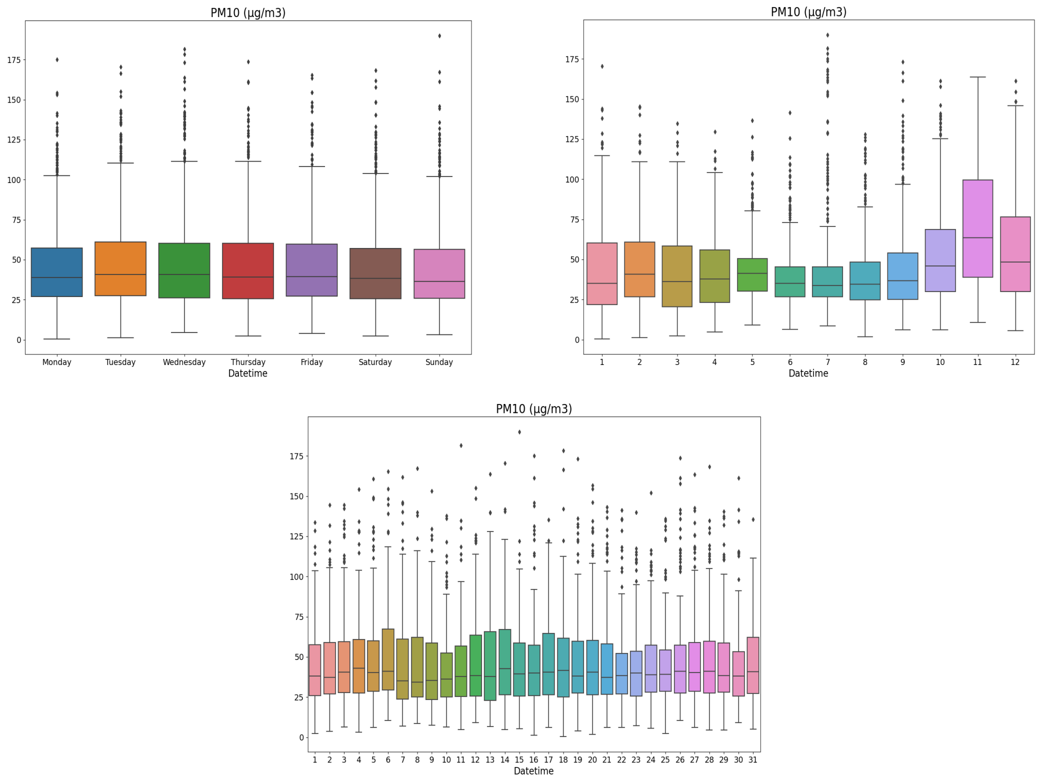


Figure 6.21: The Box-Plots of PM10 in Ankara as Weekly, Monthly and Yearly

The box plot for the monthly data displays a higher median and upper quartile for the final months of the year in comparison to the other months. Similarly, the box plot for the yearly data indicates a strong seasonal pattern. However, the box plot for the weekly data shows a consistent median value.

The data contains anomalies which are illustrated in the anomaly detection outcome displayed in Figure 6.22.

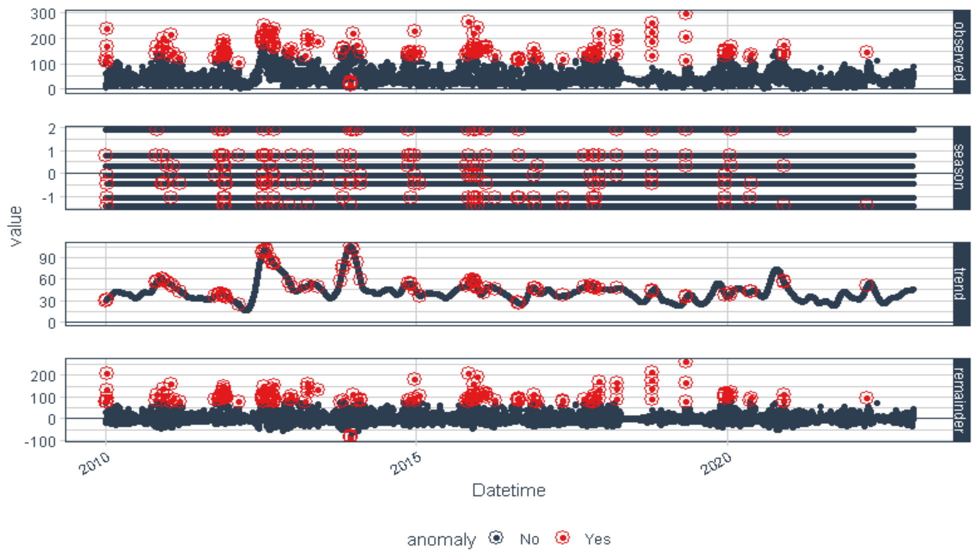


Figure 6.22: The Anomaly Detection Results of PM10 in Ankara

Naive Method Result

The seasonal periods of 7 for weekly, 30 for monthly and 365 for yearly are used. The outcomes of the naive method on PM10 data are presented in Figure 6.23.

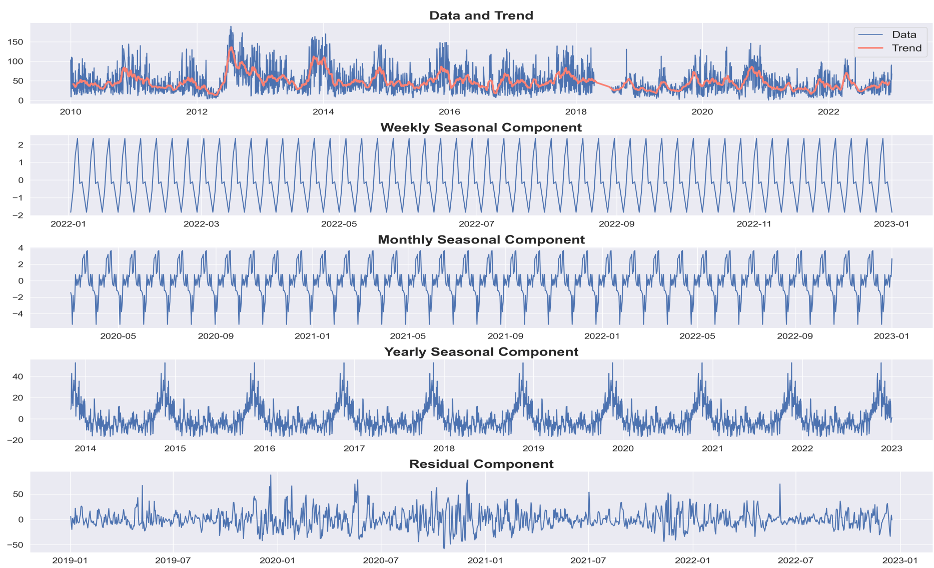


Figure 6.23: The Naive Method Results for PM10 in Ankara

The analysis of the PM10 data reveals the presence of noticeable seasonal patterns on a weekly, monthly, and yearly basis. Specifically, the data indicate a decrease in PM10 levels towards the end of each month. Furthermore, while PM10 levels tend to increase during autumn, they decrease during spring.

MSTL Result

The MSTL method is implemented on the data using weekly, monthly and yearly seasonal periods. The outcomes of the MSTL method on the PM10 data are presented in Figure 6.24.

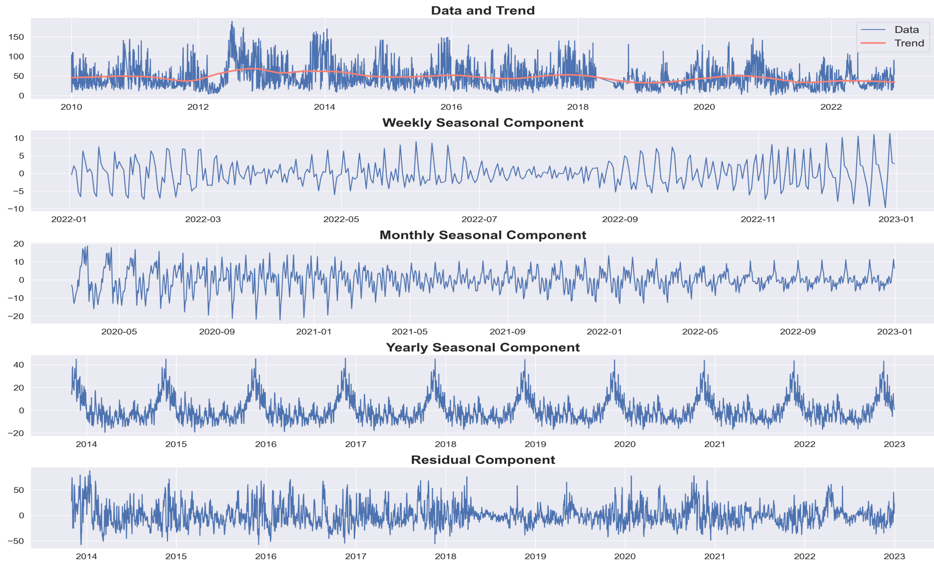


Figure 6.24: The MSTL Method Results for PM10 in Ankara

Based on the MSTL method, the PM10 data shows an intense and complex seasonality, with fluctuating weekly seasonal factors. Meanwhile, the monthly seasonal factor is getting very small but smoother.

STR Result

The STR method is utilized on the data, with a "gapCV" parameter of 30 for PM10. The outcomes of the STR method on the PM10 data are presented in Figure 6.25.

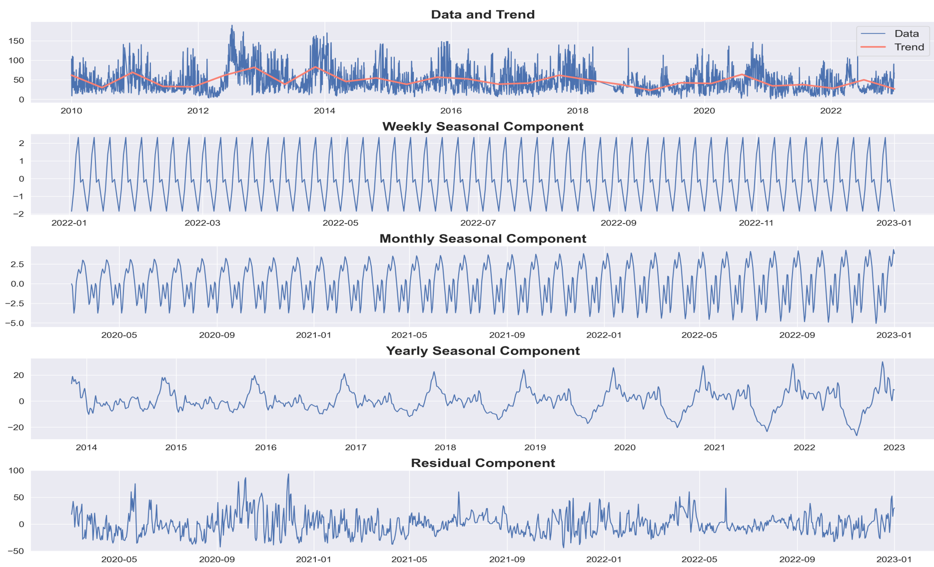


Figure 6.25: The STR Method Results for PM10 in Ankara

Like the MSTL method, the STR method detected multiple seasonal patterns in the PM10 data. However, the seasonal components identified by the STR method appear smoother than those determined by the MSTL method. Moreover, the weekly and monthly seasonal patterns in the PM10 data appear more stable in the STR method than in the MSTL method.

DSA Result

The PM10 data is subjected to the DSA method, which involves adjusting the data for calendar effects. In particular, holiday regressors specific to Türkiye are incorporated to account for any seasonal effects associated with public holidays. Table 6.1 presents the impact of holidays.

Table 6.1: Calender Effect on PM10 in Ankara

Holiday Names	Coefficient	s.e	t_value
Feast of Sacrifice	-6.72	4.28	-1.60
Feast of Ramadan	-6.20	4.52	-1.40
New Year's Holiday	2.03	4.64	0.40
April 23	4.92	4.69	1.00
May 1	-3.64	4.70	-0.80
May 19	-5.13	4.69	-1.10
July 15	-4.14	6.87	-0.60
August 30	0.36	4.70	0.10
October 29	0.12	5.40	0.00
Saturdays	0.27	0.75	0.40
Sundays	-0.10	0.75	-0.10

The calendar effect is also shown in Figure 6.26.

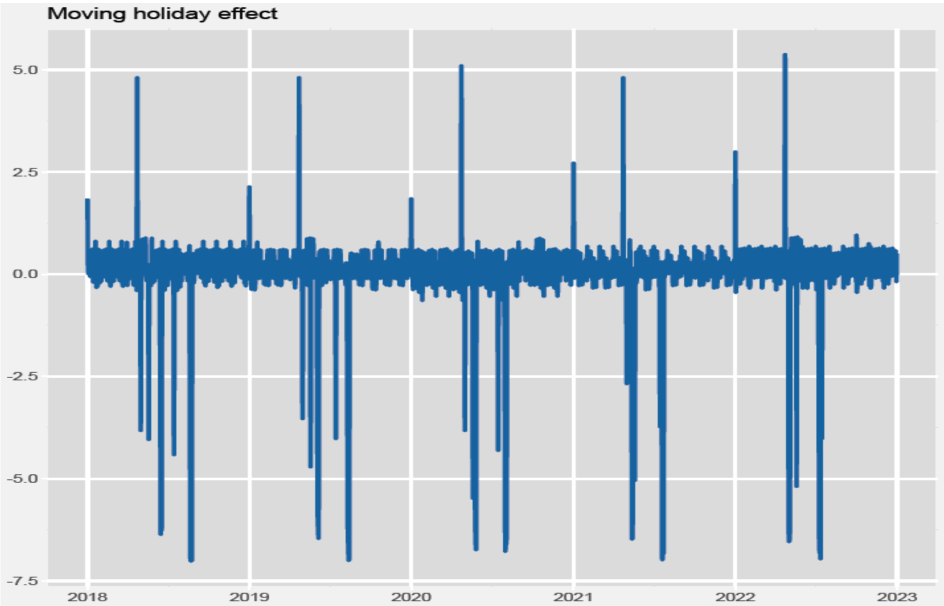


Figure 6.26: The Calender Effect for PM10 in Ankara

Figure 6.27 displays the final output of the DSA method after calendar adjustment.

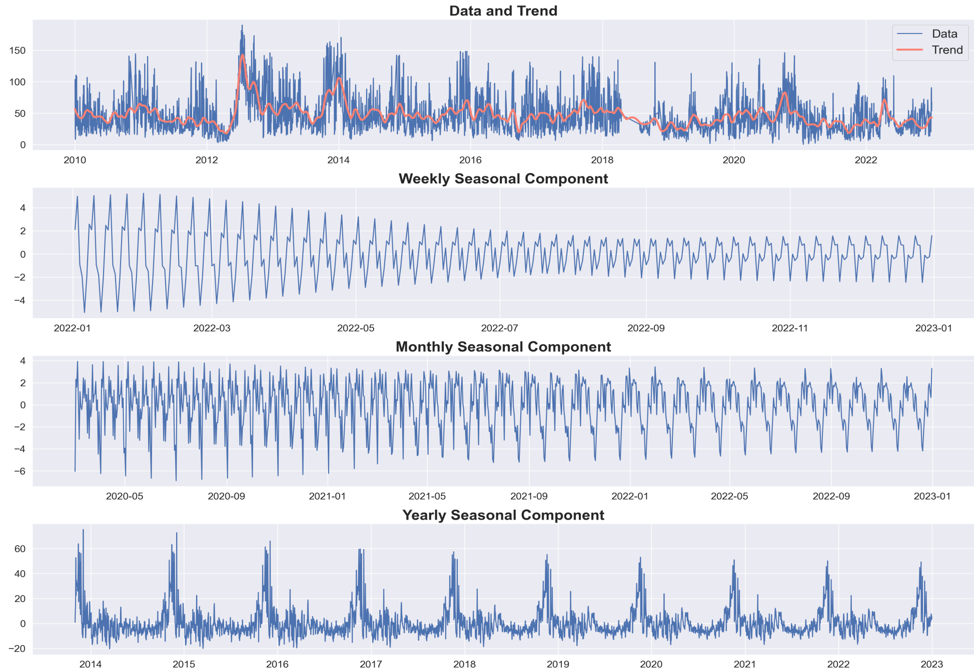
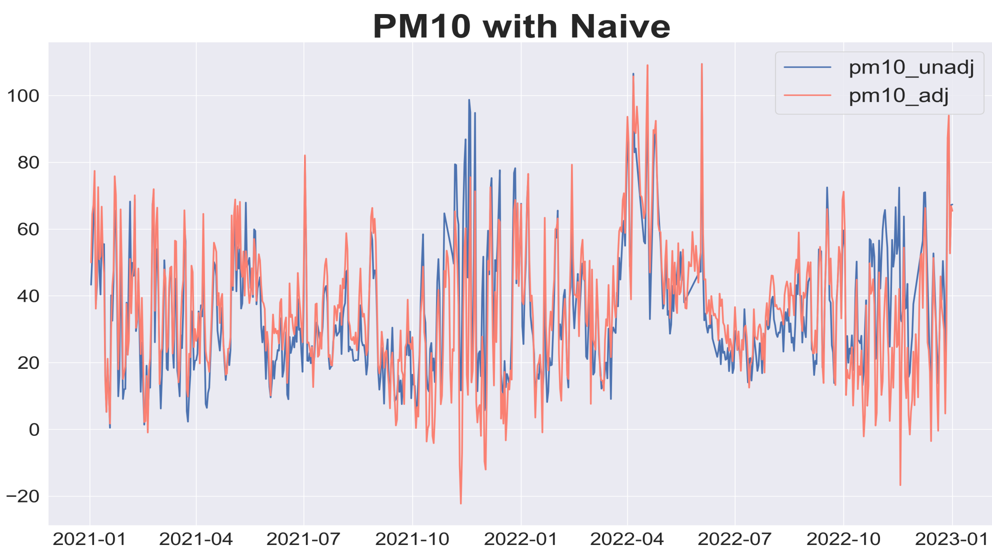


Figure 6.27: The DSA Method Results for PM10 in Ankara

Like the MSTL and STR methods, the DSA method also identifies decreasing seasonal components. The seasonal patterns of the PM10 data vary over time. However, compared to the STR method, the DSA results are more volatile.

<u>Result for PM10 in Ankara</u>

Figure 6.28 presents the results of all the seasonal adjustment methods for PM10.



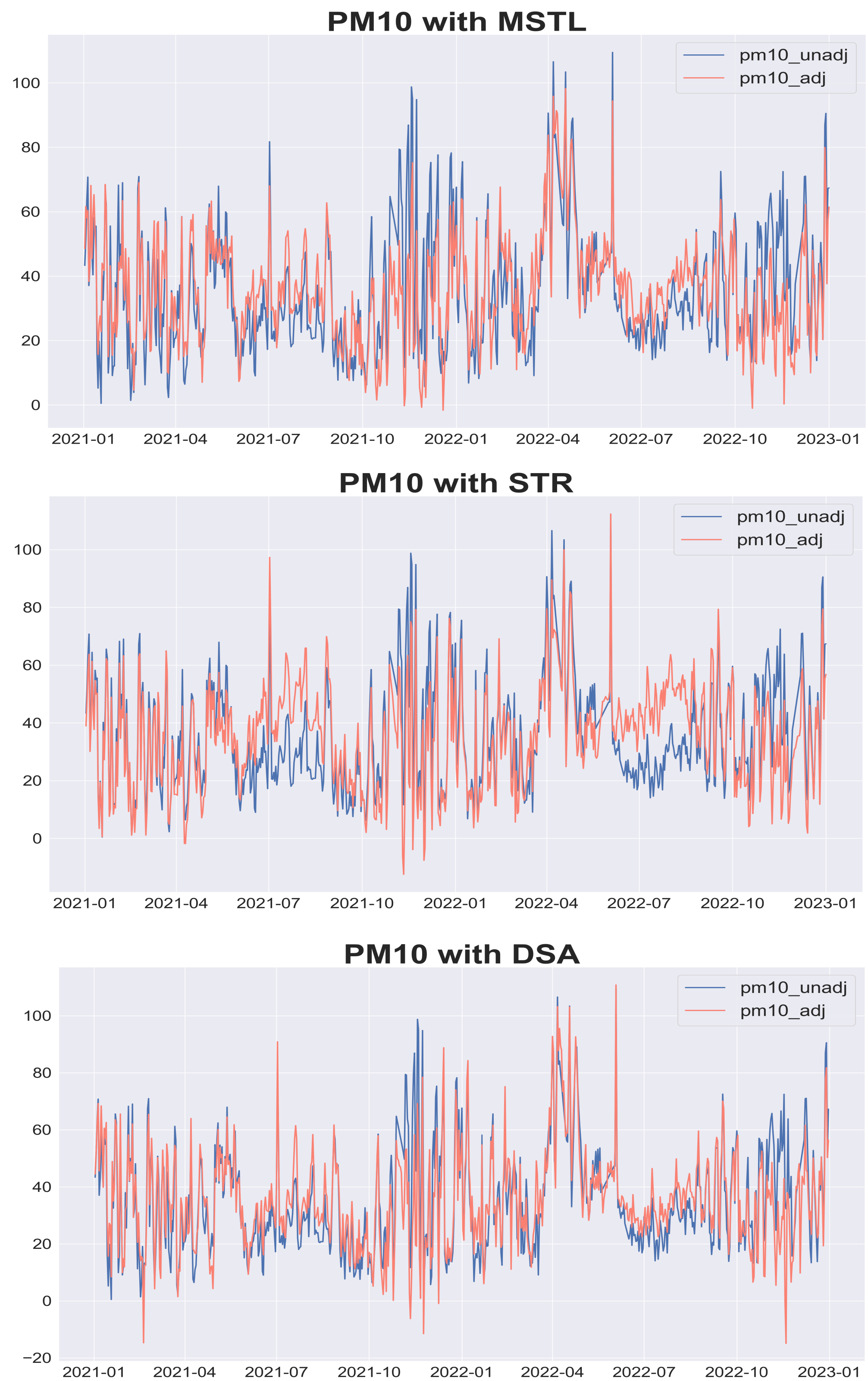


Figure 6.28: The Seasonally Adjusted Series Result of All Methods for PM10 in Ankara

The outcomes of the different methods appear comparable except for the naive method, which has a volatile result.

The ACF and PACF graphs of the seasonally adjusted series are depicted below.

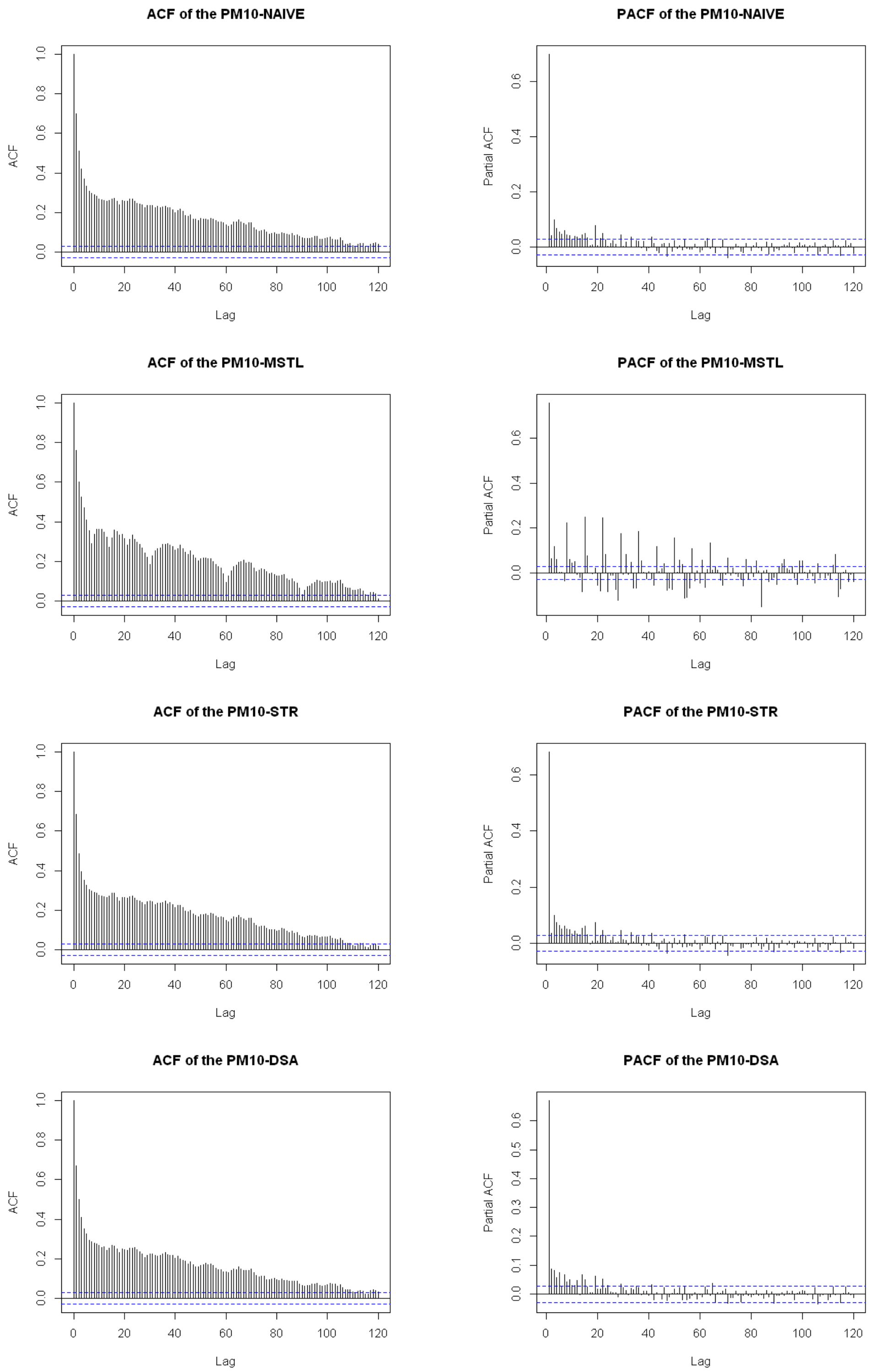


Figure 6.29: The Seasonally Adjusted Series ACF-PACF Plots for PM10 in Ankara

The ACF and PACF plots indicate that the MSTL method detects significant seasonal patterns. On the other hand, the ACF and PACF plots for the Naive and DSA methods show better results.

6.2.2 CO in Bahçelievler

The CO in the air dataset is a subset of the CO dataset that focuses on measuring the concentration of carbon monoxide (CO) in the atmosphere of Bahçelievler/Ankara. The dataset includes daily measurements of CO levels obtained from several monitoring stations located in different parts of Ankara.

In order to better understand and interpret the data, a time series plot is shown in Figure 6.30.

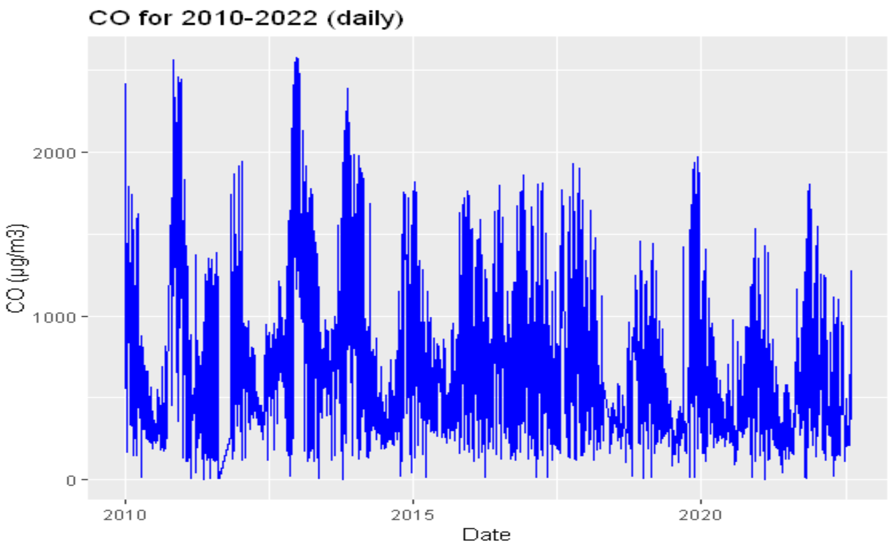


Figure 6.30: Time Series Plot of CO in Bahçelievler

The seasonal pattern data is presented in Figure 6.30. On the other hand, Figure 6.31 illustrates the ACF and PACF graphs of the CO dataset.

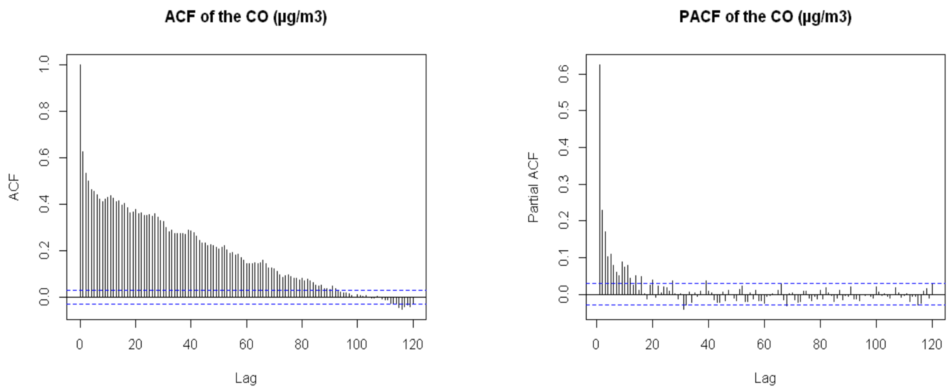


Figure 6.31: The ACF and PACF Graphs of CO in Bahçelievler

The ACF plot demonstrates a clear seasonal pattern, indicating a high probability of the presence of multiple seasonality in the series.

Box-Plots can be utilized to examine the seasonal data structure further and identify potential multiple seasonalities.

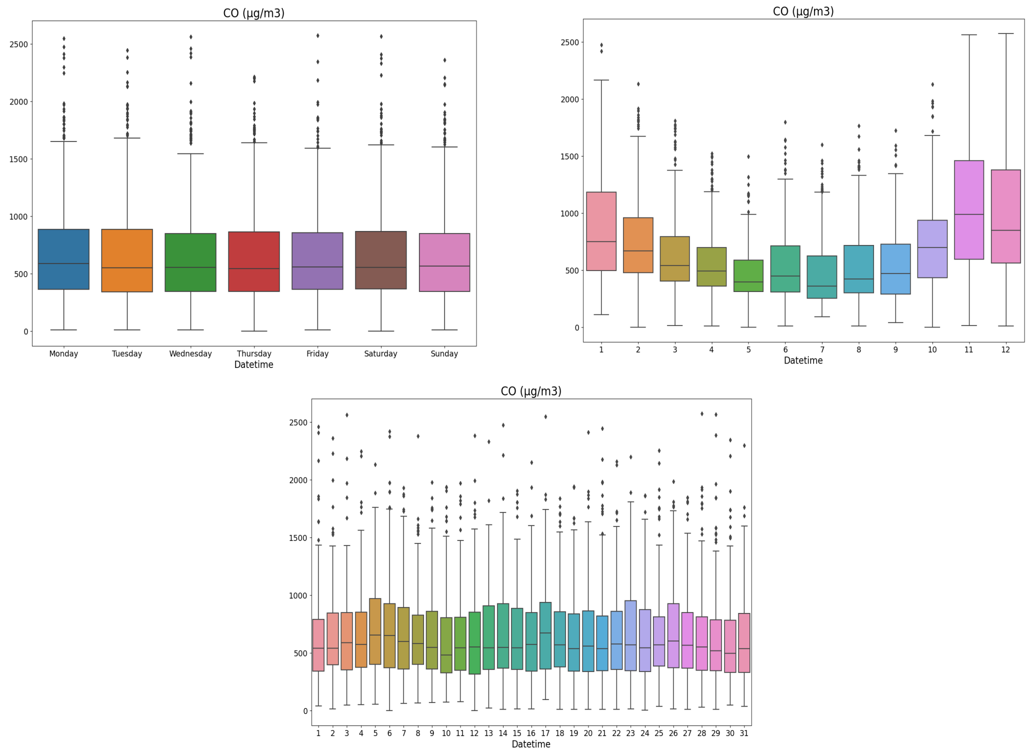


Figure 6.32: The Box-Plots of CO in Bahçelievler Weekly, Monthly and Yearly

The yearly plot demonstrates a prominent seasonal behavior, with a decrease in CO levels during the summer months and an increase during winter, which is attributed to heating sources in winter. However, the median values in the weekly and monthly plots remain relatively stable.

The time series data is processed with imputation as a pre-processing step, and the anomaly detection results are presented in Figure 6.33.

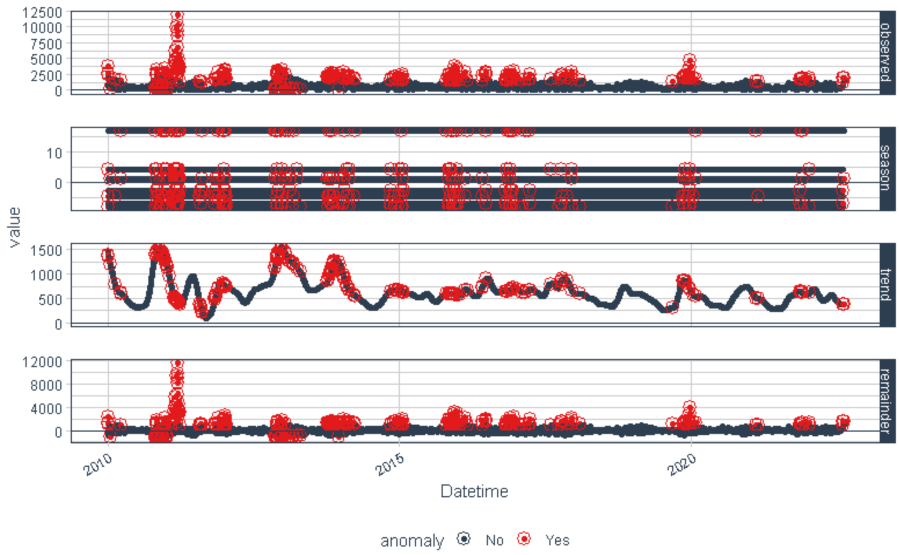


Figure 6.33: The Anomaly Detection Results of CO in Bahçelievler

Naive Method Result

The data is processed using the naive method, and the outcomes of the naive method applied to CO are presented in Figure 6.34.

Figure 6.34: The Naive Method Results for CO in Bahçelievler

The results indicate that the CO dataset displays notable seasonal patterns on a weekly, monthly, and yearly basis.

MSTL Result

The MSTL method is utilized to process the data, and the results of applying the MSTL method to CO are presented in Figure 6.35.

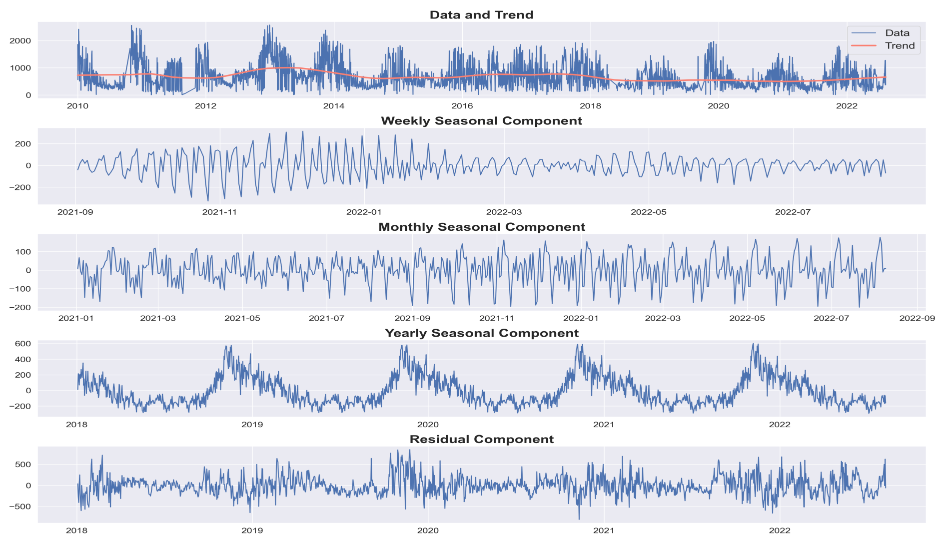


Figure 6.35: The MSTL Method Results for CO in Bahçelievler

Intense and complex seasonality is indicated by the MSTL method, and the weekly seasonal component displays a declining seasonal trend.

STR Result

The STR method is applied to the data, and the results of the STR method are presented in Figure 6.36.

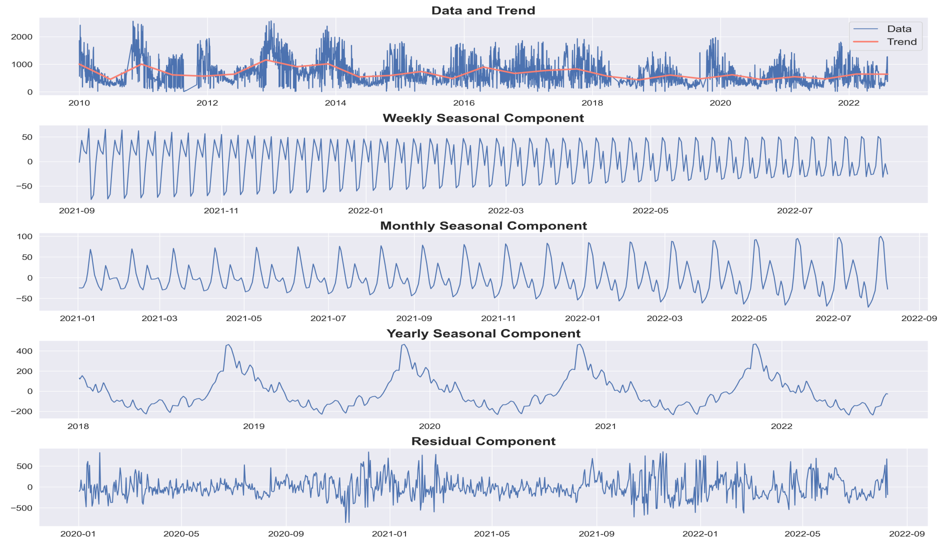


Figure 6.36: The STR Method Results for CO in Bahçelievler

The seasonal components, including the monthly and yearly seasonality, have a smoother formation in the results of the STR method compared to those of the MSTL method.

DSA Result

The DSA method is applied to the data. In this method, CO is also calendar adjusted with the holiday regressors are created for Türkiye.

The effect of the holiday effect is shown in Table 6.2.

Table 6.2: Calender Effect on CO in Bahçelievler

Holiday Names	Coefficient	s.e	t_value
Feast of Sacrifice	39.2166	56.18085	0.7
Feast of Ramadan	-127.47827	60.74086	-2.1
New Year’s Holiday	74.67858	81.98897	0.9
April 23	118.6269	81.3787	1.5
May 1	-80.35244	81.40483	-1
May 19	-48.36273	81.38308	-0.6
July 15	-6.03361	118.30115	-0.1
August 30	18.17256	84.75209	0.2
October 29	80.19258	88.30855	0.9
Saturdays	10.27951	12.20297	0.8
Sundays	7.81526	12.19325	0.6

The calendar effect is also shown in Figure 6.37.

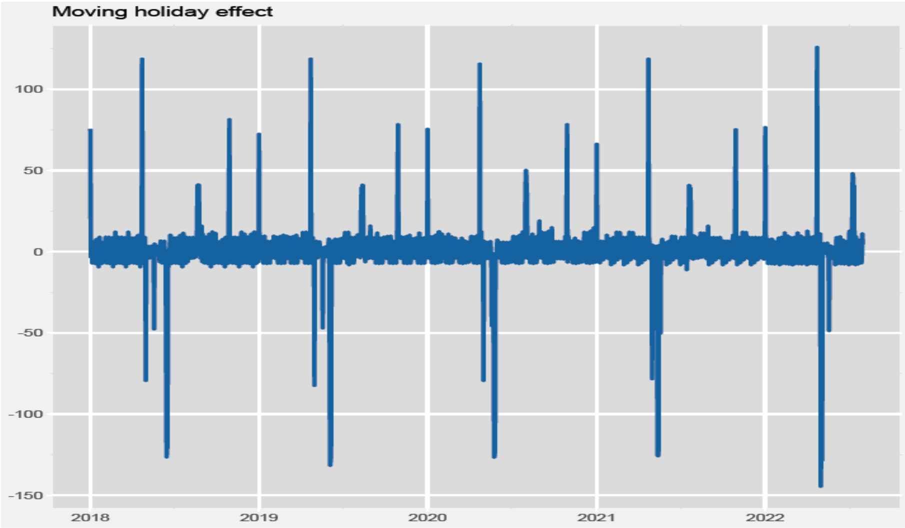


Figure 6.37: The Calender Effect for CO in Bahçelievler

The result of the DSA method is given in Figure 6.38.

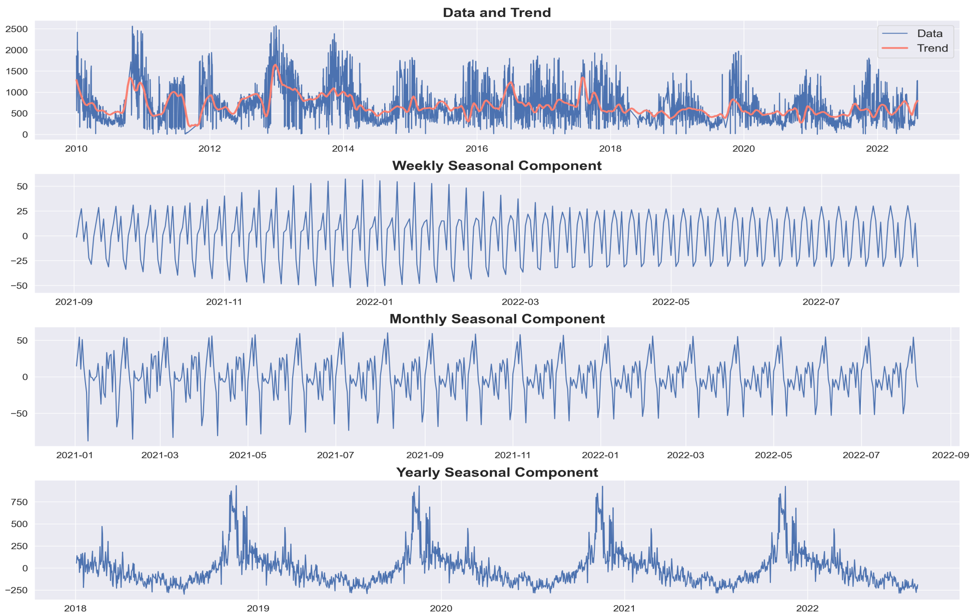
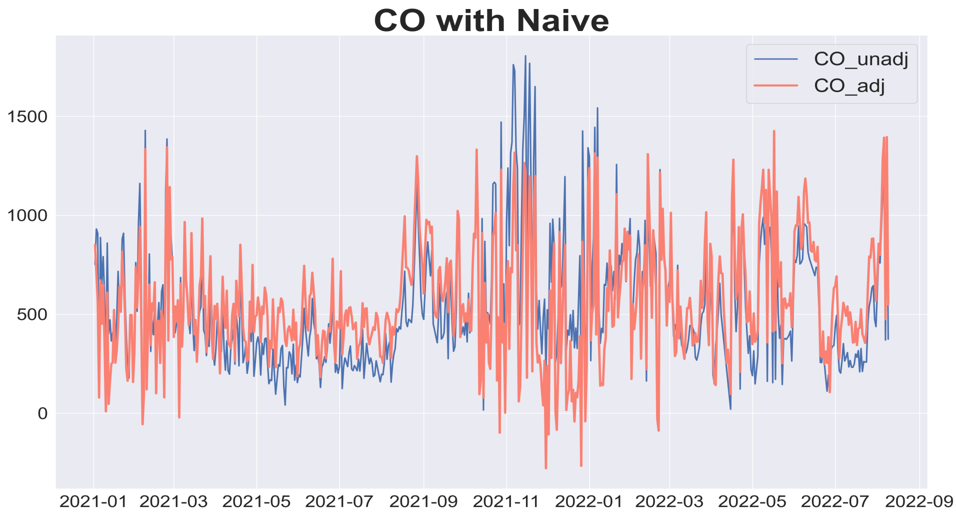


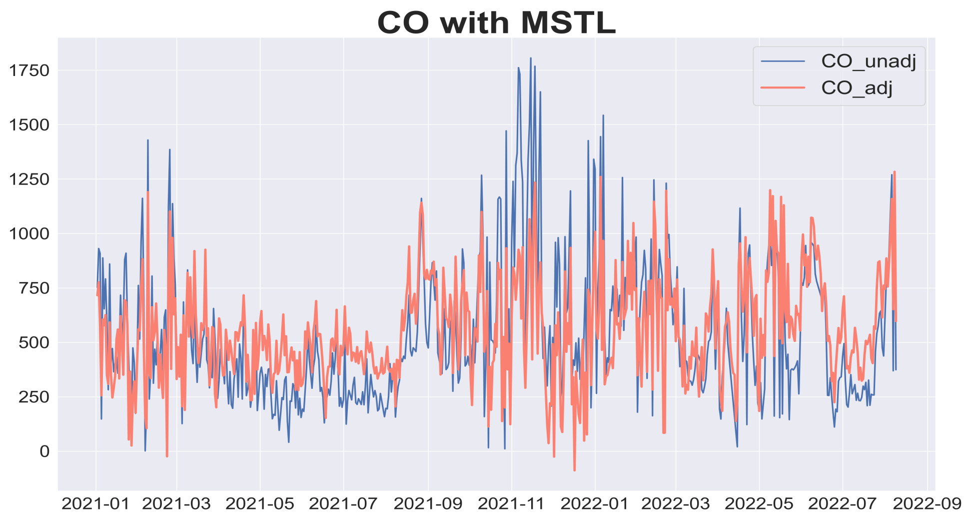
Figure 6.38: The DSA Method Results for CO in Bahçelievler

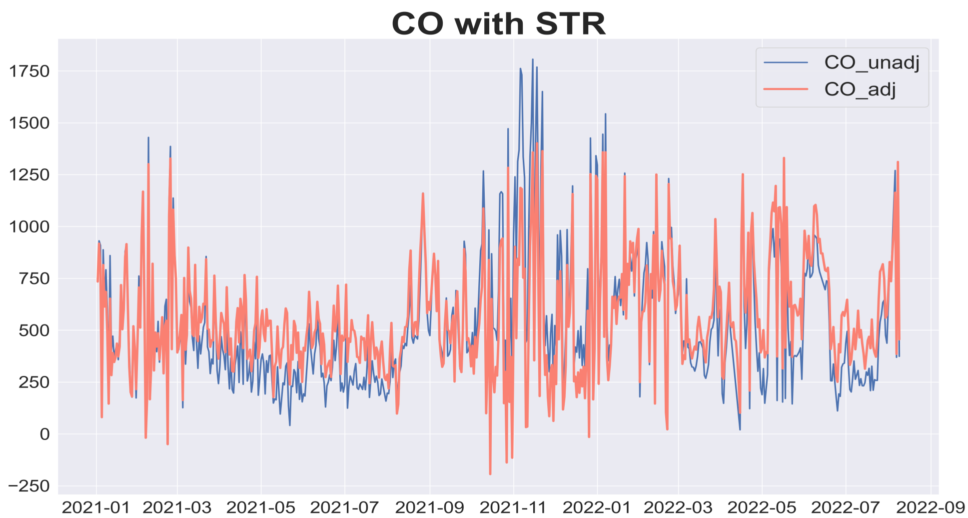
The DSA method gives similar results to the MSTL method. However, the weekly seasonal component is more stable than the MSTL method.

<u>Result for CO in Bahçelievler</u>

Figure 6.39 shows the results of all the seasonal adjustment methods for the CO.


CO with Naive
CO_unadj
CO_adj
1500
1000
500
0
2021-01
2021-03
2021-05
2021-07
2021-09
2021-11
2022-01
2022-03
2022-05
2022-07
2022-09


CO with MSTL
CO_unadj
CO_adj
1750
1500
1250
1000
750
500
250
0
2021-01
2021-03
2021-05
2021-07
2021-09
2021-11
2022-01
2022-03
2022-05
2022-07
2022-09


CO with STR
CO_unadj
CO_adj
1750
1500
1250
1000
750
500
250
0
−250
2021-01
2021-03
2021-05
2021-07
2021-09
2021-11
2022-01
2022-03
2022-05
2022-07
2022-09

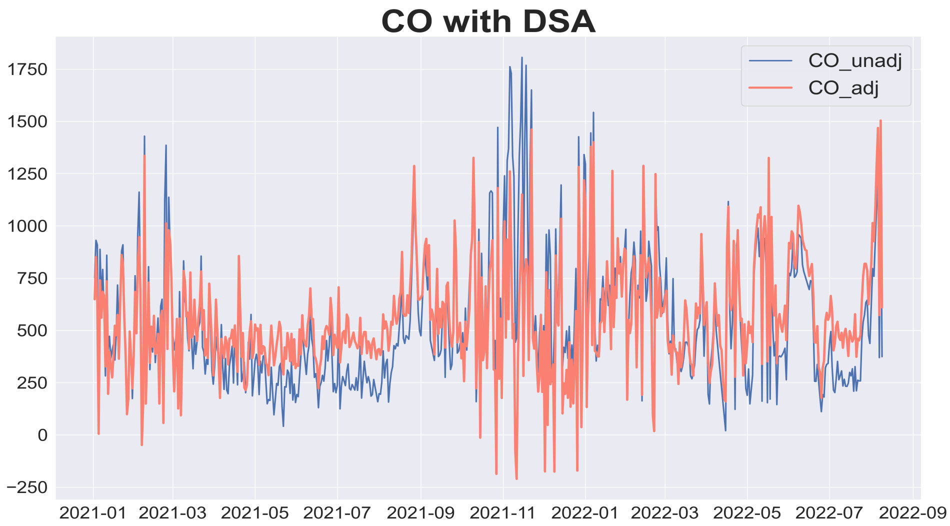
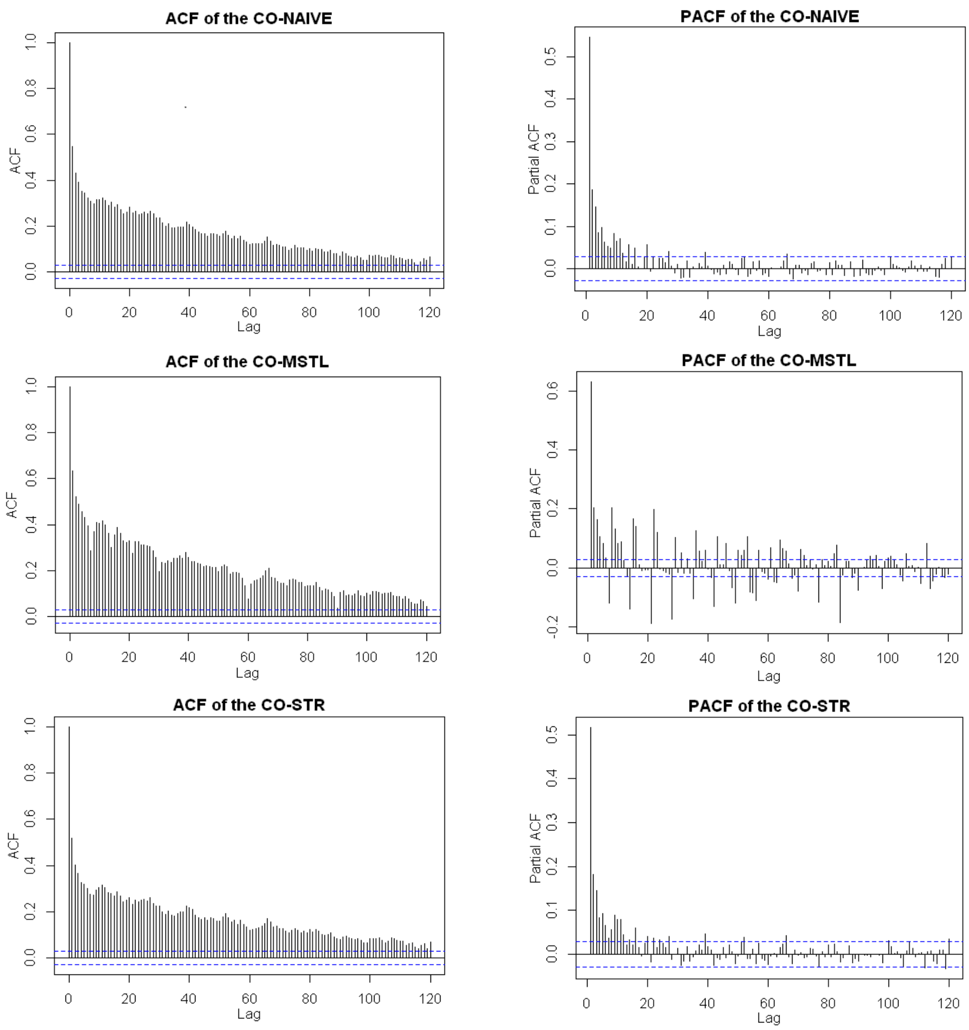


Figure 6.39: The Seasonally Adjusted Series Result of All Methods for CO in Bahçelievler

The visual outputs of the methods show that the resulting series are comparable.

The ACF and PACF plots for the final seasonally adjusted series are shown below.



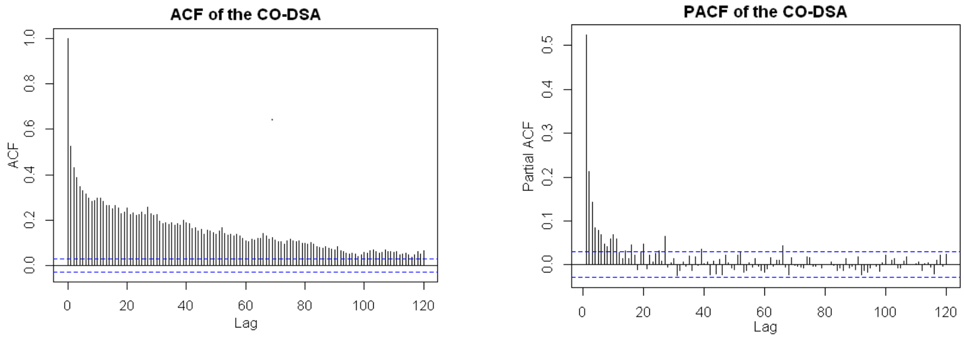


Figure 6.40: The Seasonally Adjusted Series ACF-PACF Plots for CO in Bahçelievler

The ACF graphs exhibit similar patterns, while the PACF graphs in the STR method show dissimilarities by having insignificant lags. Almost all the lags fall within the white noise band.

6.2.3 Bike Hiring in Washington

The daily bike hiring dataset of the U.S. contains the number of bikes rented daily in Washington, D.C. Figure 6.41 provides a visualization of the time series data to facilitate data interpretation.

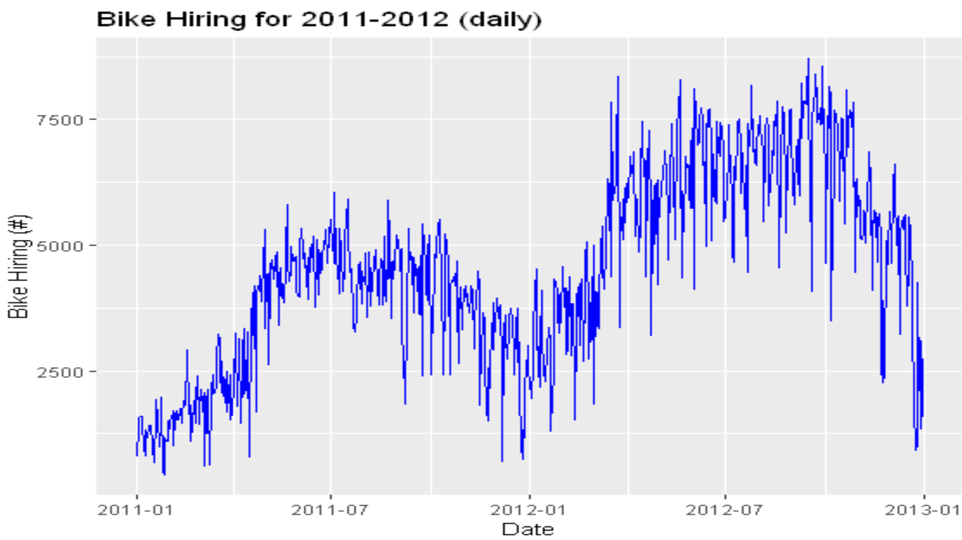


Figure 6.41: Time Series Plot of Bike Hiring in Washington

The plot displays that the bike hiring data exhibits multiple seasonal patterns.

The ACF and PACF graphs of bike hiring are illustrated in Figure 6.42.

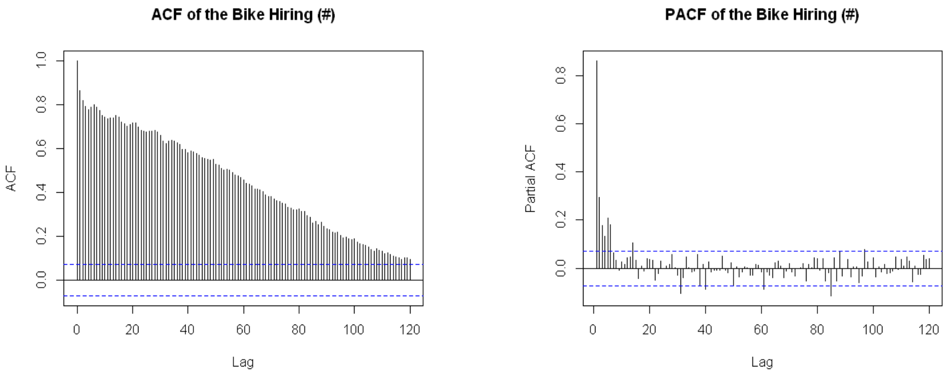


Figure 6.42: The ACF and PACF Graphs of Bike Hiring in Washington

The time series exhibits significant lags in both the ACF and PACF graphs. PACF graph shows the seasonal pattern in the monthly seasonal lags.

Box-Plots for weekly, monthly and yearly intervals can provide a better understanding of the seasonal data structure.

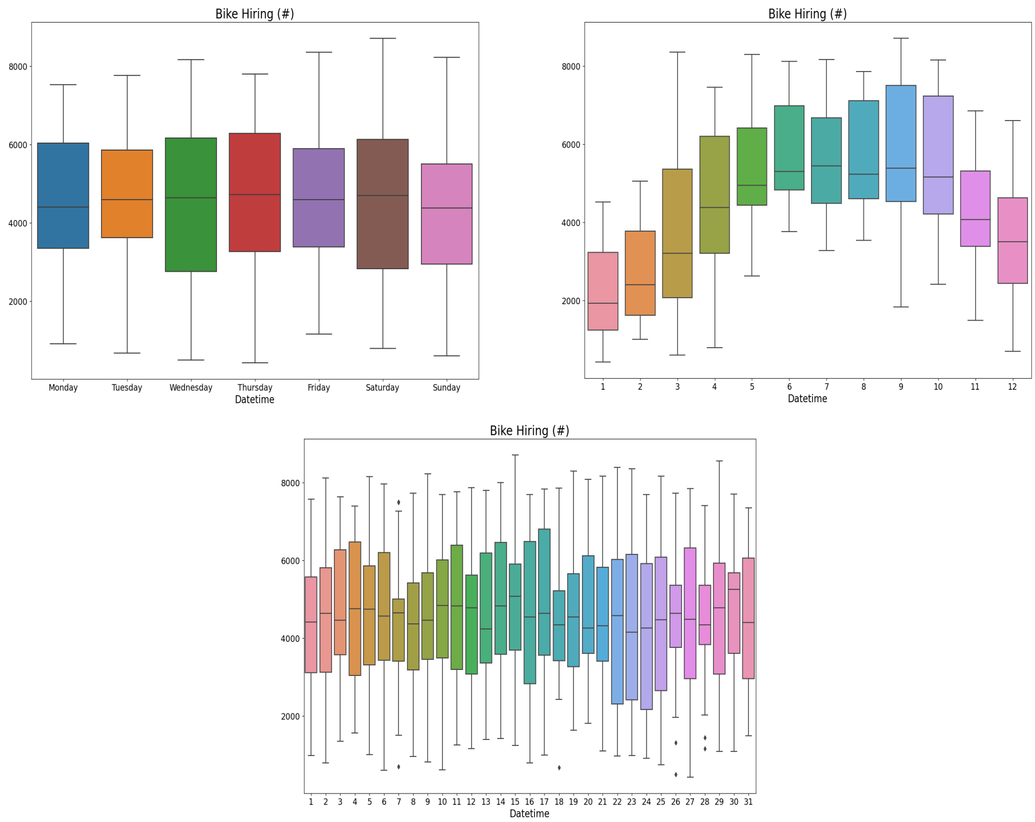


Figure 6.43: The Box-Plots of Bike Hiring in Washington Weekly, Monthly and Yearly

The box plot for the yearly data displays a higher median and upper quartile in summer. Similarly, the box plot for the yearly data indicates a seasonal pattern. However, the box plot for the weekly data shows a consistent median value.

The data contains anomalies which are illustrated in the anomaly detection outcome displayed in Figure 6.44.

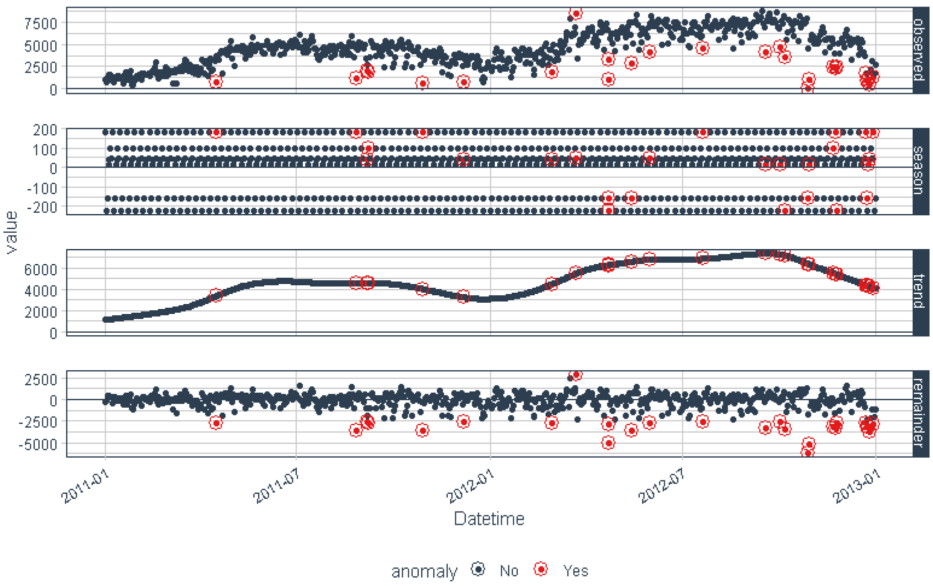


Figure 6.44: The Anomaly Detection Results of Bike Hiring in Washington

<u>Naive Method Result</u>

The results of the naive method on bike hiring data are presented in Figure 6.45.

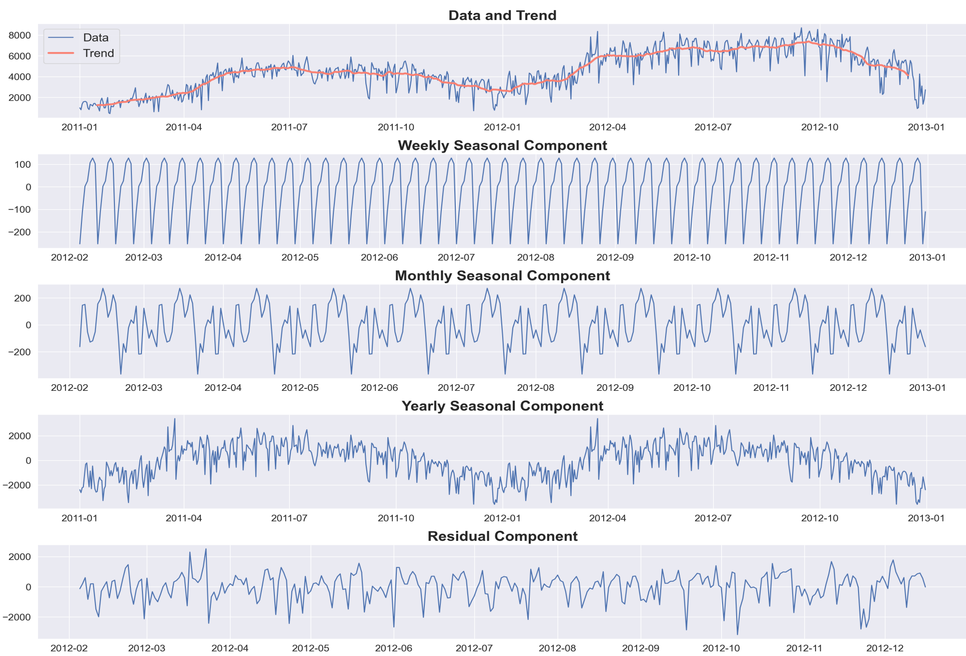


Figure 6.45: The Naive Method Results for Bike Hiring in Washington

The examination of the bike hiring data indicates the existence of significant seasonal patterns occurring weekly, monthly, and yearly.

MSTL Result

The results of the MSTL method on the bike hiring data are shown in Figure 6.46.

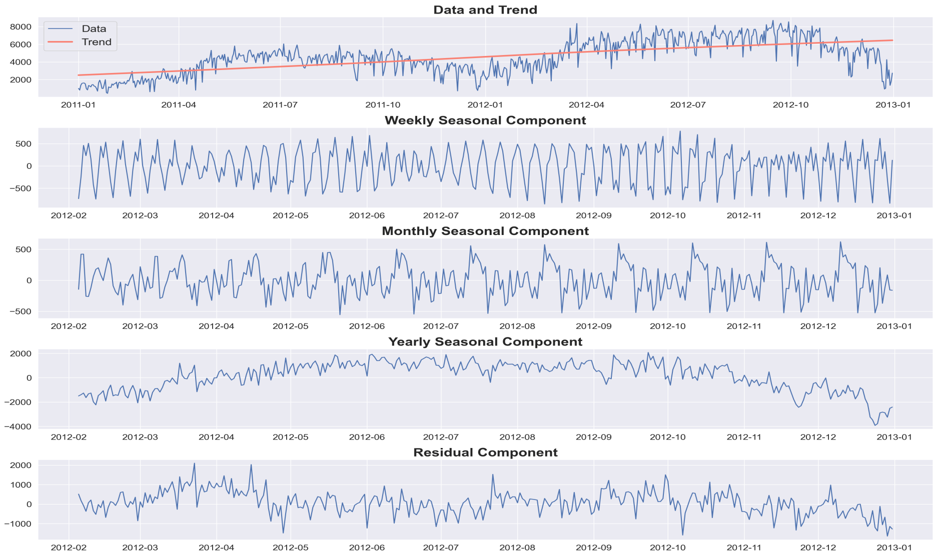


Figure 6.46: The MSTL Method Results for Bike Hiring in Washington

According to the results obtained using the MSTL method, the weekly seasonal factor of the bike hiring data exhibits fluctuations while the monthly seasonal factor remains stable.

STR Result

The outcomes of the STR method on the bike hiring data are presented in Figure 6.47.

Figure 6.47: The STR Method Results for Bike Hiring in Washington

Like the MSTL method, the STR method found multiple seasonal patterns in the bike hiring data. However, the seasonal components detected by the STR method seem smoother compared to those identified by the MSTL method.

DSA Result

The DSA method is applied to the bike hiring data; the outcome is presented in Figure 6.48.

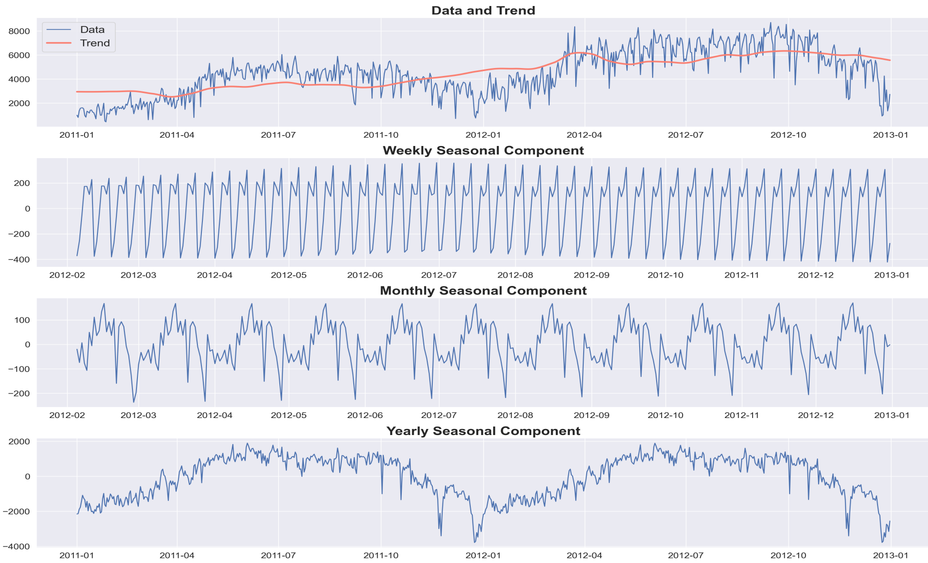
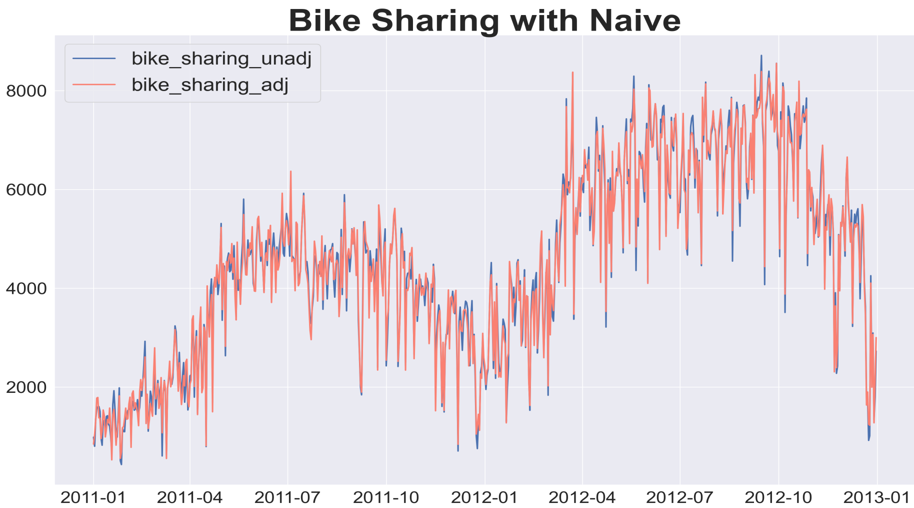


Figure 6.48: The DSA Method Results for Bike Hiring in Washington

The seasonal components identified by the DSA method are smoother and more stable compared to other methods. In addition, the trend component is able to capture all data movements more accurately.

Result for Bike Hiring in Washington

Figure 6.49 shows the results of all the seasonal adjustment methods for bike hiring.



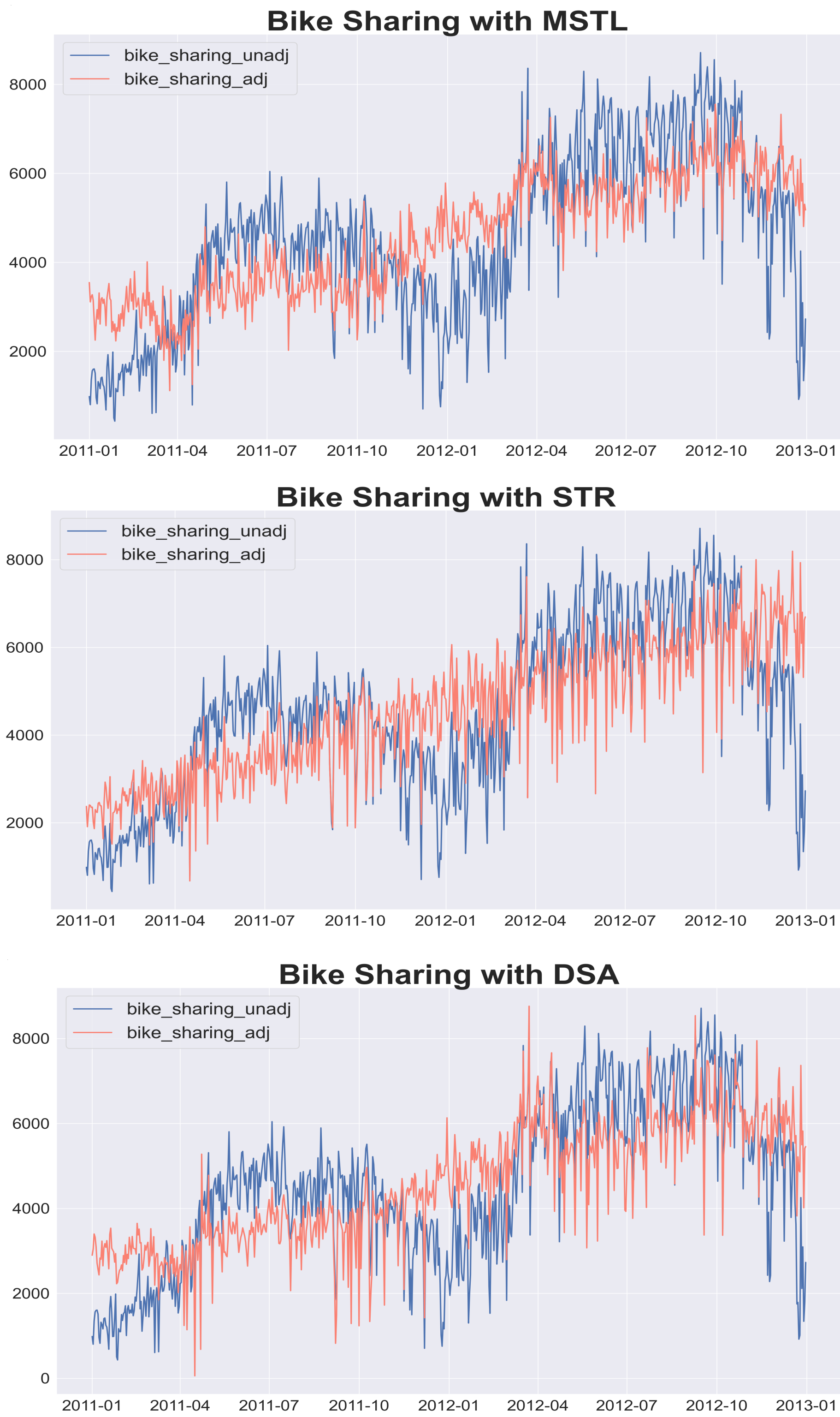


Figure 6.49: The Seasonally Adjusted Series Result of All Methods for Bike Hiring in Washington

The graphical representations of the methods indicate that the seasonal effect of the data could not be accurately decomposed by the naive method, and the outcomes of the other methods are similar.

The ACF and PACF plots for the final seasonally adjusted series are shown below.

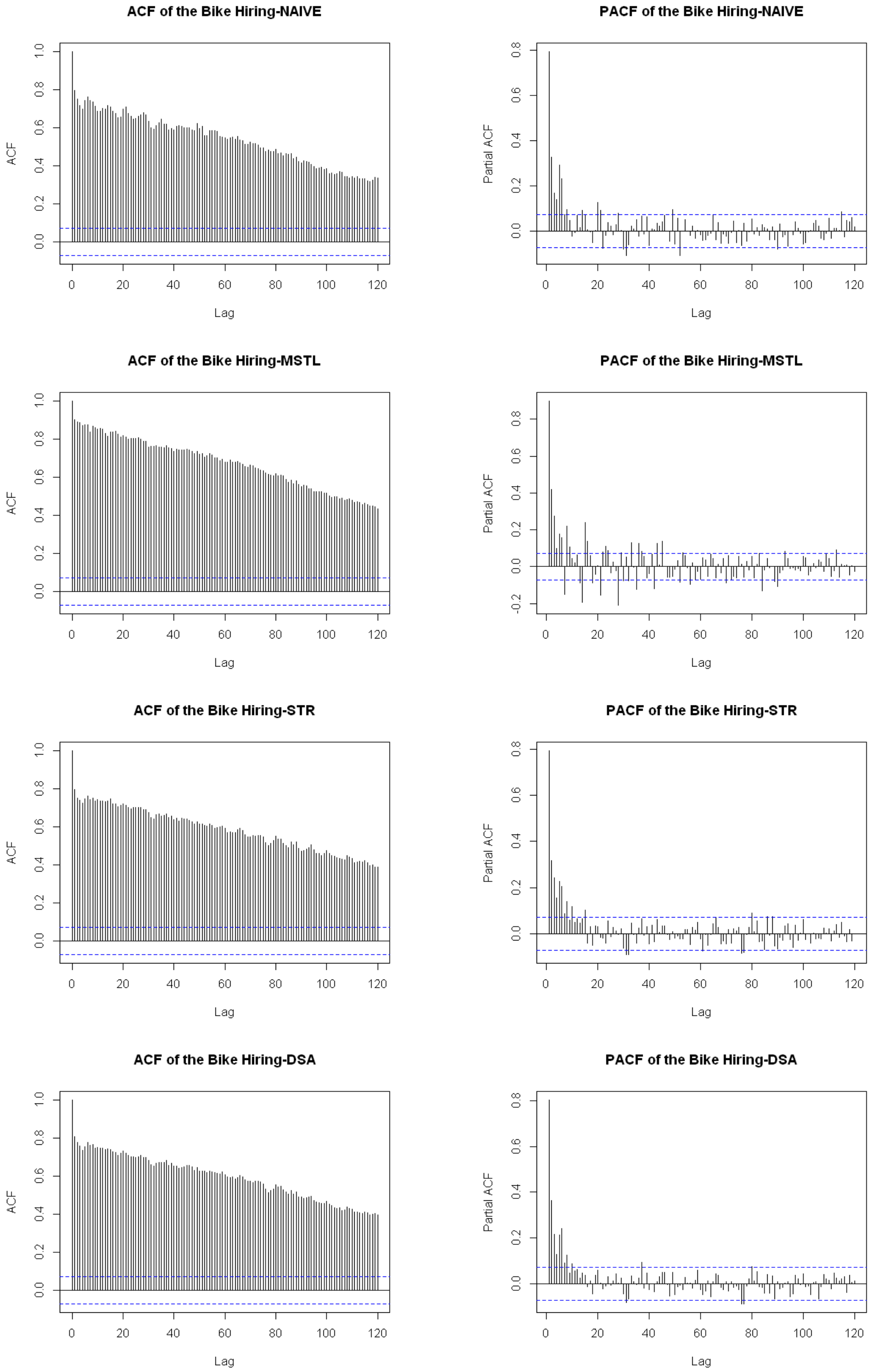


Figure 6.50: The Seasonally Adjusted Series ACF-PACF Plots for Bike Hiring in Washington

The ACF and PACF plots show relative movements. However, the DSA method shows more non-significant lags compared to the other methods.

6.3 Hourly Time Series

6.3.1 Hourly Bike Hiring in London

The bike hiring in London dataset has information about the London, UK bike hiring system. It is an hourly data set that contains two years of data.

In order to better understand and interpret the data, a time series plot is shown in Figure 6.51.

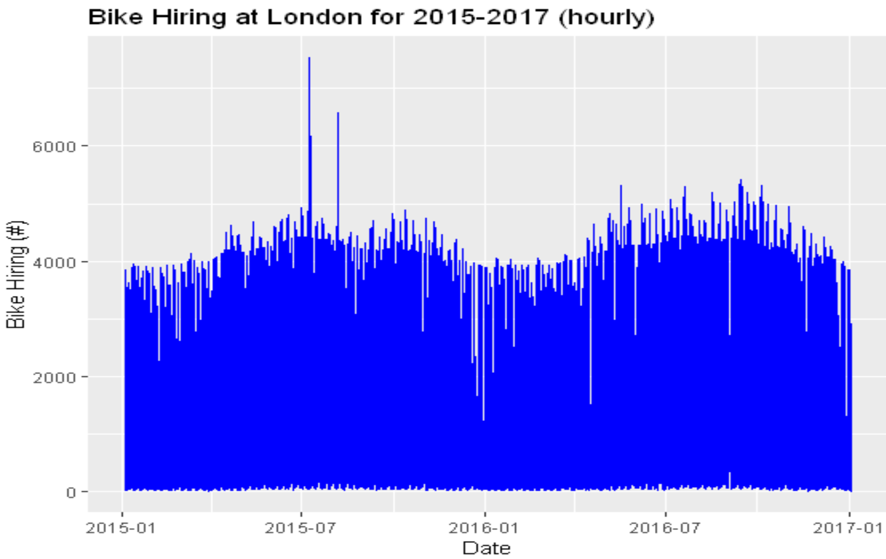


Figure 6.51: Time Series Plot of Hourly Bike Hiring in London

Figure 6.51 shows that the bike sharing data exhibit significant periodic movements and multiple seasonal patterns.

In order to gain a better comprehension of the data arrangement, it is crucial to analyze the ACF and PACF charts illustrated in Figure 6.52.

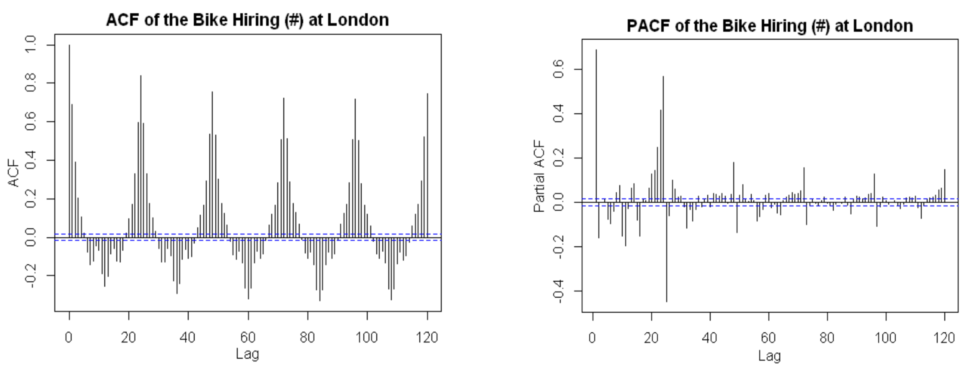


Figure 6.52: The ACF and PACF Graphs of Hourly Bike Hiring in London

The ACF and PACF plots suggest a pronounced seasonal pattern in the data, with significant lags. Specifically, the PACF plot reveals a daily seasonality, evidenced by the significant lags occurring at intervals of 24 hours and its multiples.

To explore the presence of multiple seasonal patterns, investigating the Box-Plots at daily, weekly, monthly, and yearly intervals can offer a more specific understanding of the seasonal structure of the data.

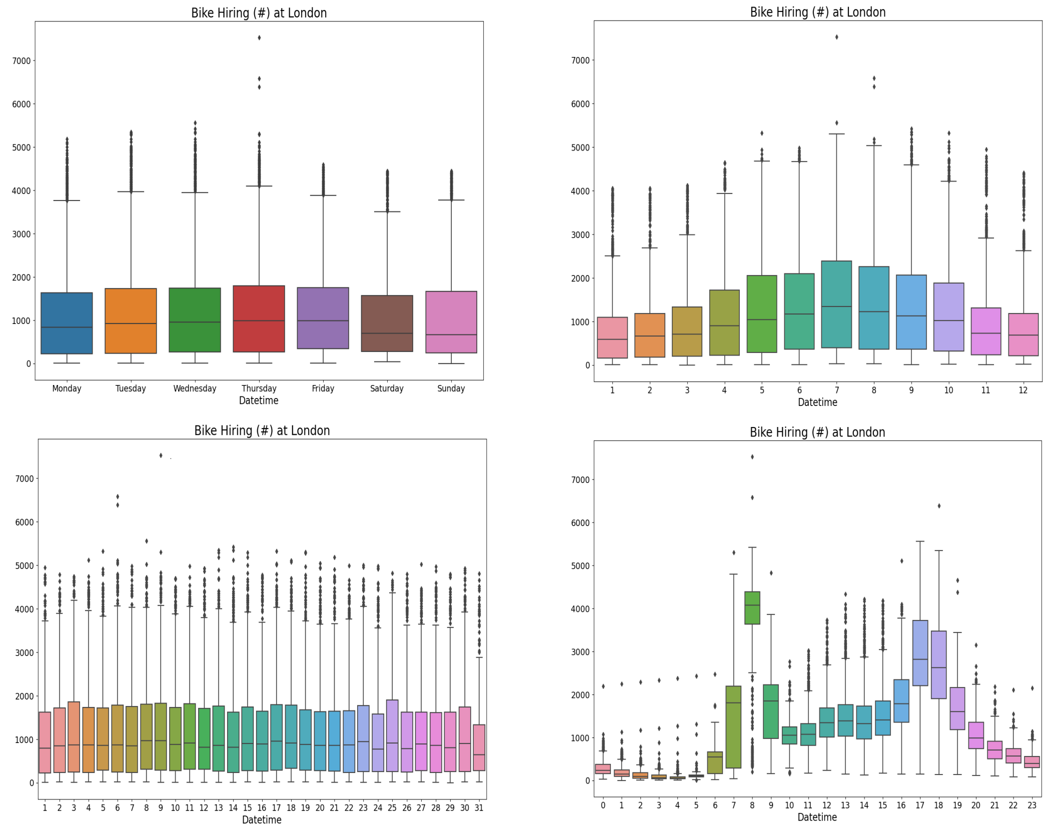


Figure 6.53: The Box-Plots of Hourly Bike Hiring in London Daiy, Weekly, Monthly and Yearly

The Box-Plot for hourly intervals demonstrates a clear daily seasonal pattern, as evidenced by the increase in the median values at the beginning and end of the working hours. This observation is logical, considering that peoples often use bikes for commuting to work during these periods. Consequently, at the weekends, the number of bike hirings decreases. The monthly plot has a seasonal pattern in spring and summer due to the good weather conditions.

For the preprocessing stage, the time series was imputed. The outcome of the anomaly detection is presented in Figure 6.54.

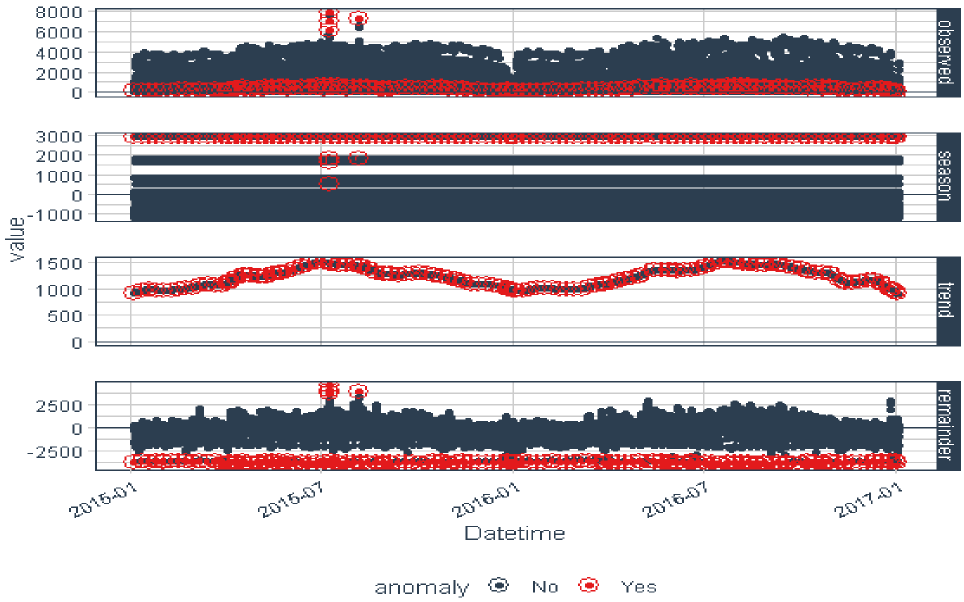


Figure 6.54: The Anomaly Detection Results of Hourly Bike Hiring in London

Naive Method Result

The data was subjected to the naive method, and the output of the hourly bike hiring predictions is illustrated in Figure 6.55.

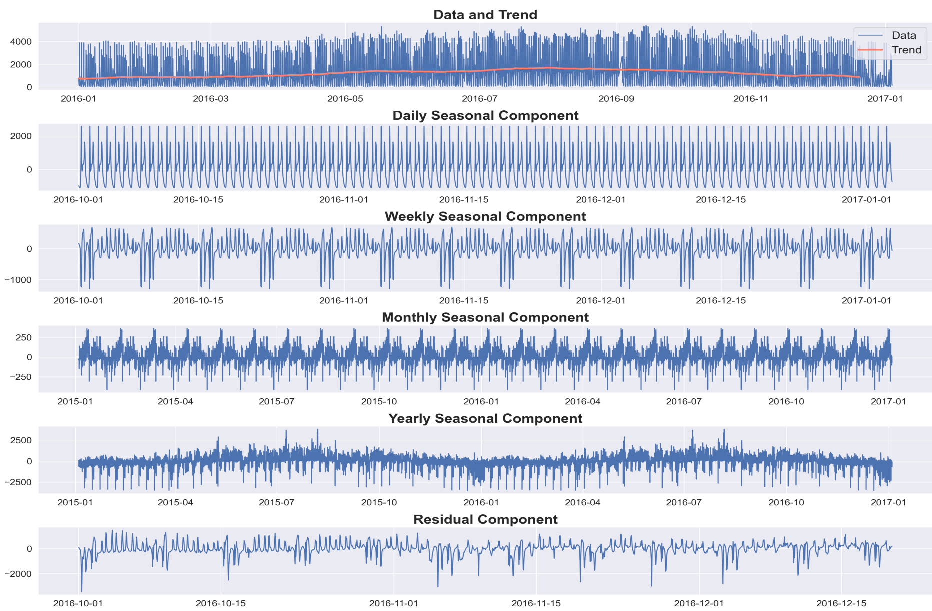


Figure 6.55: The Naive Method Results for Hourly Bike Hiring in London

The resulting graphic indicates the presence of significant seasonal patterns in hourly bike hiring. A seasonal movement is also seen in the last observations of the residual component.

MSTL Result

The MSTL method is applied to the data, and the results for hourly bike hiring are given in Figure 6.56.

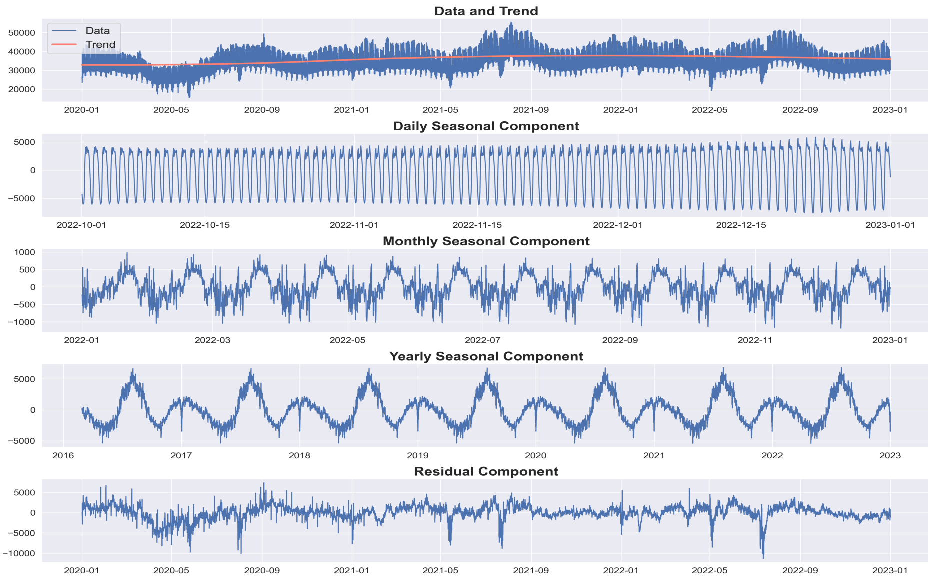


Figure 6.56: The MSTL Method Results for Hourly Bike Hiring in London

All of the anticipated elements exhibit stable seasonal patterns, with the impact of seasonality decreasing towards the end of the data for the daily and weekly seasonal components.

STR Result

The bike hiring data is processed using the STR method. The outcomes of applying the STR method to the hourly bike hiring are depicted in Figure 6.57.

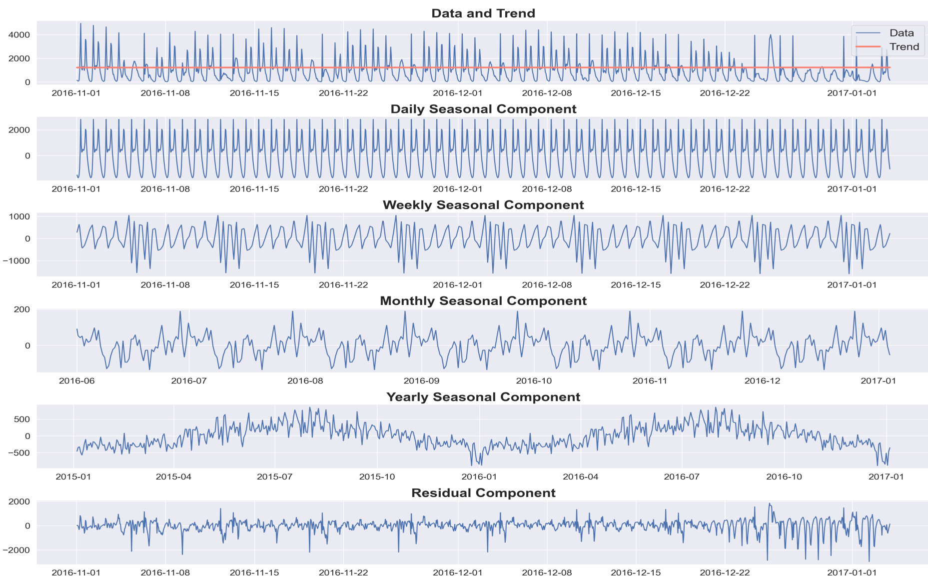


Figure 6.57: The STR Method Results for Hourly Bike Hiring in London

Compared to the results obtained from the MSTL method, the seasonal components generated by the STR method are smoother and more stable.

Result for Hourly Bike Hiring in London

The final seasonally adjusted results for all approaches is given below.

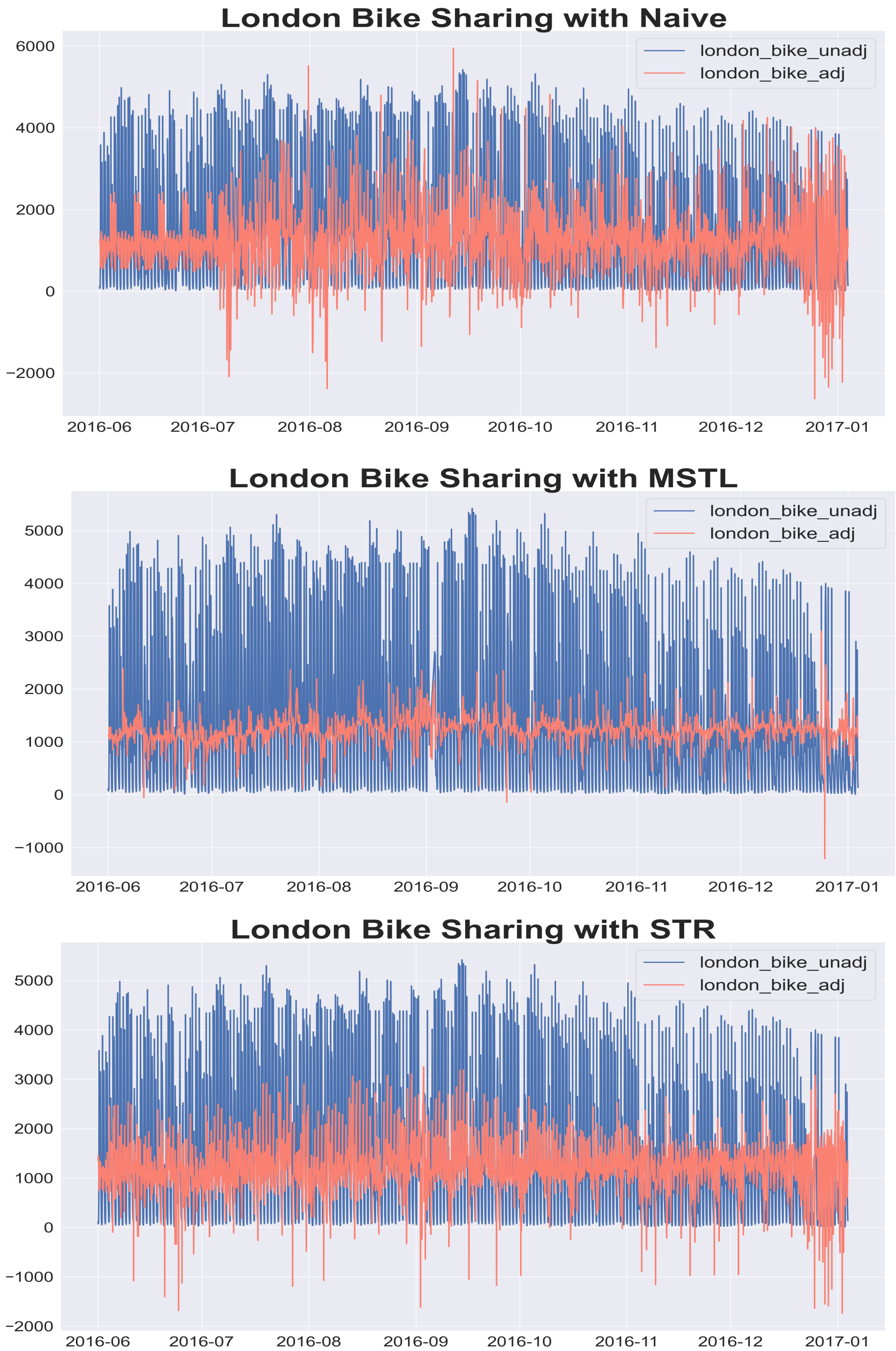


Figure 6.58: The Seasonally Adjusted Series Result of All Methods for Hourly Bike Hiring in London

Upon examining the graphical output, it appears that the results obtained from the MSTL method are smoother than the others.

The seasonally adjusted series' ACF and PACF plots are presented below.

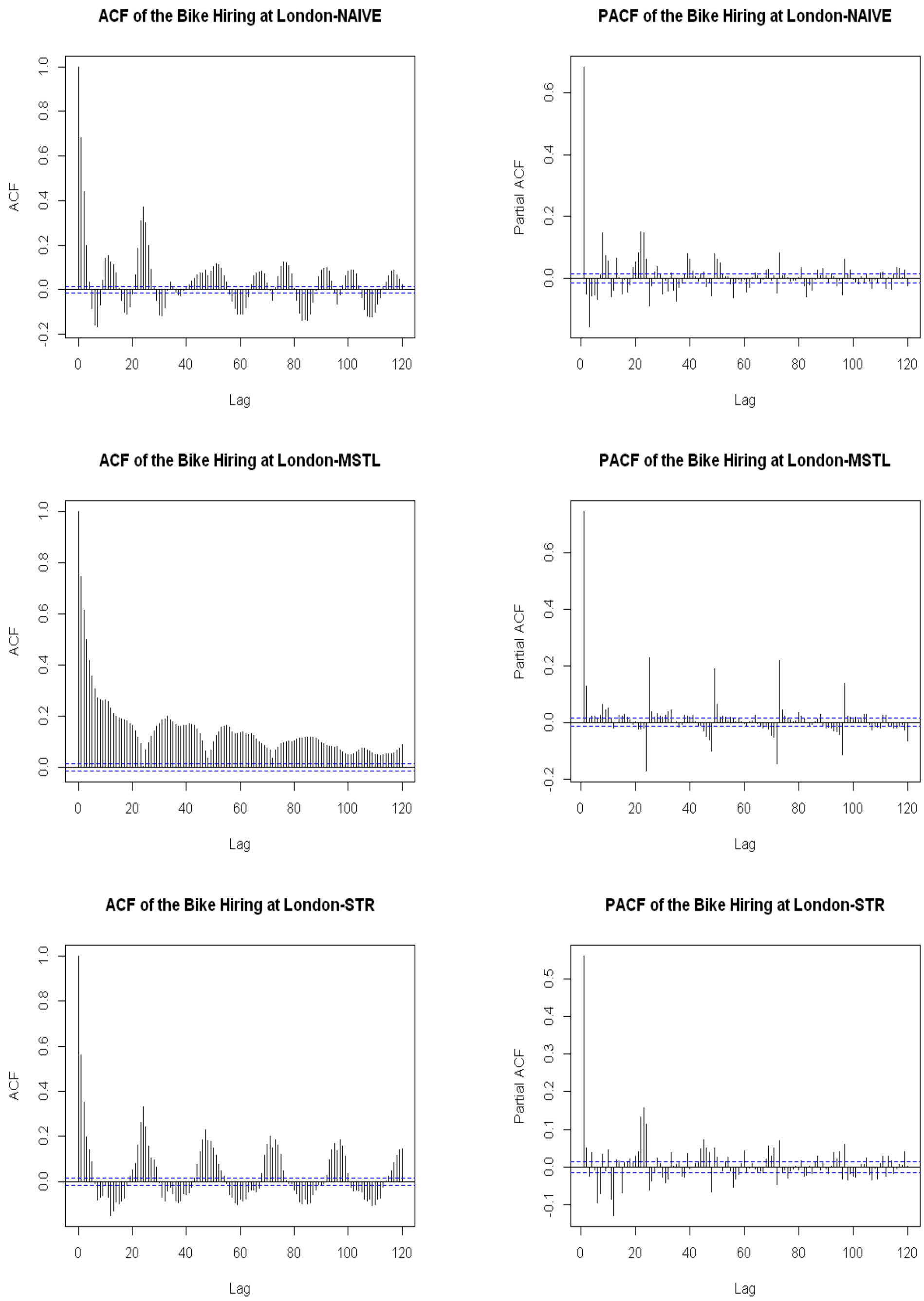


Figure 6.59: The Seasonally Adjusted Series ACF-PACF Plots for Hourly Bike Hiring in London

The ACF plot of the MSTL method deviates from the others, whereas the PACF plot exhibits significant lags at daily intervals. The graphical outputs of both the Naive and STR methods also reveal seasonal patterns.