T.R.
VAN YUZUNCU YIL UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES
DEPARTMENT OF STATISTICS

ESTIMATING TIME SERIES DATA ON ENERGY PRODUCTION IN TÜRKİYE BY COMPARING BAYESIAN AND CLASSICAL MODELS

Ph.D. THESIS

Amir KHaleel HASSOO
SUPERVISOR: Assoc.Prof. Dr. Şakir İŞLEYEN

VAN – 2023

T.R.
VAN YUZUNCU YIL UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES
DEPARTMENT OF STATISTICS

ESTIMATING TIME SERIES DATA ON ENERGY PRODUCTION IN TÜRKİYE BY COMPARING BAYESIAN AND CLASSICAL MODELS

Ph.D. THESIS

Amir KHaleel HASSOO

VAN – 2023

ACCEPTANCE AND APPROVAL PAGE

This thesis entitled "ESTIMATING TIME SERIES DATA ON ENERGY PRODUCTION IN TÜRKİYE BY COMPARING BAYESIAN AND CLASSICAL MODELS" presented by Amir KHaleel Hassoo under supervision of Assoc. Prof. Dr. Şakir İŞLEYEN in the Department of Statistics has been accepted as a Ph.D. thesis according to Legislations of Graduate Higher Education on 28/7/2023 with unanimity.

Chair: Prof. Dr. H. Eray ÇELİK	Signature:
Member: Prof. Dr. Meltem INCE YILMAZ	Signature:
Member: Assoc. Prof. Dr. Hakan EYGÜ	Signature:
Member: Assist. Prof. Dr.Onur SALDIR	Signature:
Member: Assoc. Prof. Dr. Şakir İŞLEYEN	Signature:

This thesis has been approved by the committee of The Institute of Natural and Applied Science on ..../.........../............ with decision number ....

Signature

…………………………

Director of Institute

ETHICAL DECLARATION

I declare that all the information in this thesis has been obtained and presented within the framework of ethical behavior and academic rules, and that in this thesis, which has been prepared in accordance with the thesis writing rules, all kinds of statements and information that do not belong to me have been fully cited.

Signature
Amir KHaleel HASSOO

ABSTRACT

ESTIMATING TIME SERIES DATA ON ENERGY PRODUCTION IN TÜRKİYE BY COMPARING BAYESIAN AND CLASSICAL MODELS

HASSOO, Amir Khaleel
Ph.D Thesis, Department of Statistics
Supervisor: Assoc. Prof. Dr. Şakir İşleyen

July 2023, 154 pages

In this study, two forecasting models, Bayesian structural time series (BSTS) and Autoregressive integrated moving average (ARIMA), were compared for predicting energy production data in Türkiye from 1971 to 2020. The models were applied to various energy sources such as coal, gas, hydraulic, and oil production, as well as GDP, using data obtained from the World Bank database. The primary aim was to assess the accuracy of these models in forecasting energy production trends. To ensure reliable and comprehensive results, the analysis and modeling processes were conducted using R and SPSS software. MAPE, MAE, RMSE, and R^2 were also used for this comparison. The BSTS models, which incorporate a Bayesian framework enabling the inclusion of prior information and uncertainty quantification, were contrasted with the conventional ARIMA models commonly used for time series forecasting. To evaluate the models' accuracy, the dataset was divided into training and testing subsets, allowing for the assessment of model errors. The findings indicated that the BSTS model performed better than the ARIMA model in estimating the time series data of energy production in Türkiye. The Bayesian approach employed by the BSTS model, which accounts for the inherent uncertainties and complexities in energy production dynamics, demonstrated greater reliability and accuracy compared to the Box-Jenkins approach of the ARIMA model. As a result, the BSTS model was selected to forecast energy production from 2021 to 2028. Furthermore, this study contributes to the existing literature by utilizing multiple linear regression analysis to examine the factors influencing GDP in Türkiye.

Keywords: Accuracy, ARIMA, BSTS, Energy production, Multiple linear regression, Türkiye.

ÖZET

ENERJİ ÜRETİMİNE İLİŞKİN ZAMAN SERİSİ VERİLERİNİN BAYESİAN VE KLASİK MODELLERİ KARŞILAŞTIRARAK TÜRKİYE'DE TAHMİN EDİLMESİ

HASSOO, Amir KHaleel
Doktora Tezi, İstatistik Anabilim Dalı
Danışman: Assoc. Prof. Dr. Şakir İşleyen
Temmuz 2023, 154 sayfa

Bu çalışmada, Bayesian yapısal zaman serisi (BSTS) ve Oto-regressif entegre hareketli ortalama (ARIMA) olmak üzere iki tahmin modeli, 1971-2020 yılları arasında Türkiye'nin enerji üretimi verilerini tahmin etmek amacıyla karşılaştırılmıştır. Modeller kömür, gaz, hidrolik ve petrol üretimi ile GSMH gibi çeşitli enerji kaynaklarına ve Dünya Bankası veritabanından elde edilen verilere uygulanmıştır. Temel amaç, bu modellerin enerji üretimi trendlerini tahmin etme doğruluğunu değerlendirmektir. Güvenilir ve kapsamlı sonuçlar elde etmek için analiz ve modelleme işlemleri R ve SPSS yazılımları kullanılarak gerçekleştirilmiştir. Bu karşılaştırmada MAPE, MAE, RMSE ve R^2 gibi değerlendirme metrikleri de kullanılmıştır. BSTS modelleri, önceki bilgilerin dahil edilmesine ve belirsizliklerin nicelendirilmesine imkan sağlayan Bayesian bir çerçeve içermektedir ve zaman serisi tahmininde yaygın olarak kullanılan geleneksel ARIMA modelleri ile karşılaştırılmıştır. Modellerin doğruluğunu değerlendirmek için veri seti eğitim ve test alt kümelerine ayrılmış ve böylece model hatalarının değerlendirilmesi mümkün olmuştur. Sonuçlar, BSTS modelinin Türkiye'nin enerji üretimi zaman serisi verilerini tahmin etmede ARIMA modelinden daha iyi performans gösterdiğini göstermiştir. BSTS modelinin benimsediği Bayesian yaklaşım, enerji üretimi dinamiklerindeki doğal belirsizlikler ve karmaşıklıkları dikkate alarak, ARIMA modelinin Box-Jenkins yaklaşımına kıyasla daha fazla güvenilirlik ve doğruluk sağlamıştır. Sonuç olarak, BSTS modeli 2021-2028 yılları için enerji üretimini tahmin etmek amacıyla seçilmiştir. Ayrıca, bu çalışma Türkiye'deki GSMH'yi etkileyen faktörleri incelemek için çoklu doğrusal regresyon analizinden yararlanarak mevcut literatüre katkı sağlamaktadır.

Anahtar kelimeler: ARIMA, BSTS, Çoklu doğrusal regresyon, Enerji üretimi, Kesinlik, Türkiye.

ACKNOWLEDGMENT

First and foremost, I would like to express my gratitude to Allah for assisting me throughout my studies. I would also like to extend my sincere appreciation to my supervisor, Assoc. Prof. Dr. ŞAKIR İŞLEYEN. Additionally, I would like to convey my warm regards to all the members of the Statistics Department. I am truly thankful to all those from whom I have gained knowledge and skills. It is difficult to put into words the depth of my gratitude towards my beloved parents, who have been a constant source of support during the challenging moments of my life.

I would like to extend a special thanks to Dr. Masoud M. Hassan, Dr. Diyar M KHalil and all my friends for their unwavering support during my academic journey and their remarkable understanding, which allowed me to complete this thesis. Lastly, I am grateful to everyone who has provided assistance, even through a single word.

2023

Amir KHaleel HASSOO

CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

SYMBOLS AND ABBREVIATIONS

Some symbols and abbreviations used in this study are presented below, along with their descriptions.

Abbrevations	Description
ACF	Autocorrelation Function
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
BCM	Billion Cubic Meters
BSTS	Bayesian Structural Time Series
COVID	Coronavirus
GDP	Gross Domestic Product
GWH	Gigawatt hours
HELE	High-Efficiency, Low-Emission
LNG	Liquefied Natural Gas
MA	Moving Average
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MAX	Maximum
MCMC	Markov Chain Monte Carlo
MIN	Minimum
MLE	Maximum Likelihood Estimation
MM	Million
MSE	Mean Square Error
OLS	Ordinary Least Squares
PACF	Partial Autocorrelation Function
$\mathbf{R^2}$	Coefficient of Determination
RMSE	Root Mean Square Error
SE	Standard Error

SPSS	Statistical Package for the Social Sciences
SS	State-Space
TPAO	Turkish Petroleum Corporation
UAE	United Arab Emirates
UK	United Kingdom
UN	United Nation
USA	United States of America
USD	United States dollar

1. INTRODUCTION

1.1 Background

Türkiye has experienced substantial growth and development in its energy sector over the past five decades, with various energy sources such as coal, gas, hydraulic power, and oil playing a significant role (Kiliç and Kaya, 2007). Understanding the patterns and trends in energy production and its correlation with GDP is crucial for policymakers, energy planners, and researchers. This research study aims to estimate time series data on energy production and GDP in Türkiye from 1971 to 2020 by comparing Bayesian and classical models (Demirbaş, 2003). In this study, Bayesian models, specifically the Bayesian structural time series (BSTS), are compared with classical models, such as the Autoregressive integrated moving average (ARIMA), to estimate time series data on energy production and GDP in Türkiye (Hepbasli, 2004). Accurate estimation of energy production and its relationship with GDP is essential for effective planning and policy-making to ensure economic development and sustainability (Algül and Vedat, 2021). Time series analysis has proven to be an effective approach for studying and forecasting energy production and GDP data, enabling insights into the factors influencing energy production and its impact on the economy (Jonek-Kowalska, 2019). Bayesian models offer advantages such as incorporating prior information, quantifying uncertainty, and accommodating complexities and uncertainties in the data. The research objective is to assess the performance and accuracy of Bayesian and classical models in estimating time series data on energy production and GDP in Türkiye. The findings of this study will contribute to the existing literature on time series analysis in the energy sector, providing valuable insights for policymakers and researchers involved in energy planning and economic forecasting. This preface introduces the subsequent chapters, which will discuss the methodology, data analysis, modeling techniques, and interpretation of results (Kaplan et al., 2011). The study includes data collection, pre-processing, and modeling, examining the performance of Bayesian and classical models in estimating energy production for each variable (Yılmaz and Uslu, 2007). Additionally, the implications of the findings, potential study limitations, and future research directions in the field of time series analysis and energy production estimation are discussed (Yüksel, 2008).

1.2 Research Problem

Accurately estimating the time series data of energy production in Türkiye from 1971 to 2020 presents significant research challenges that are crucial for gaining comprehensive insights into energy dynamics, understanding the factors influencing energy production, and improving the accuracy of predictions. The dataset contains complex patterns, such as long-term trends, seasonal variations, and irregular fluctuations, which must be effectively comprehended and modeled to accurately capture the underlying dynamics and make reliable predictions.

The energy production sector in Türkiye faces various challenges that impact its efficiency, sustainability, and reliability. One of the key challenges is Türkiye's heavy reliance on a diverse range of energy sources, including fossil fuels like coal, oil, and natural gas, as well as renewable energy sources such as solar, wind, hydro, and geothermal. Effectively managing and integrating these diverse sources to ensure a balanced and sustainable energy supply while reducing dependence on fossil fuels and promoting the adoption of renewable energy poses a significant challenge.

Another challenge is Türkiye's high dependency on energy imports, which adds to the complexity of ensuring a secure and reliable energy supply. Rising energy demand is also a challenge that needs to be addressed, as it requires robust planning and infrastructure development to meet the increasing energy needs of the country. Environmental impact is a critical concern, and finding ways to mitigate the environmental effects of energy production is crucial for sustainable development. Infrastructure development plays a vital role in the energy sector, and ensuring the availability of adequate infrastructure for energy production, transmission, and distribution is a challenge that needs to be addressed. The regulatory framework governing the energy sector requires careful consideration to promote competition, efficiency, and environmental sustainability (Toklu, 2017).

Furthermore, securing financing and attracting investments for energy projects is a significant challenge that needs to be overcome. Adequate funding is essential for developing new energy infrastructure, implementing renewable energy projects, and improving the overall efficiency and reliability of the energy sector.

1.3 Research Objectives

The primary aim of this study is to conduct a comparative analysis between Bayesian and classical models. To accomplish this objective, the research outlines the following specific goals:

1. To compare the performance and effectiveness of Bayesian and classical models in estimating energy production in Türkiye. By assessing the strengths and weaknesses of each modeling approach, including their ability to capture complex patterns and handle uncertainties, the study aims to identify the most suitable technique for accurate energy production estimation.

2. To investigate the relationship between gross domestic product (GDP) and energy production in Türkiye. This objective involves examining how changes in GDP influence energy demand and production levels. Understanding the implications of economic growth on energy sustainability and resource allocation is a crucial aspect of this objective.

By addressing these research objectives, the study aims to contribute to the understanding of energy production estimation and its relationship with economic factors in Türkiye. This knowledge can guide decision-makers and stakeholders in formulating effective strategies for sustainable energy planning and resource allocation.

1.4 Importance of the Research

This research comprehensively addresses key aspects of energy production in Türkiye, with the objective of providing valuable insights for informed decision-making and promoting sustainable energy development. A primary focus is on accurately estimating time series data for energy production spanning five decades (Fidan, 2010). This entails capturing patterns, trends, and fluctuations across different energy sources. The accuracy of these estimates plays a vital role in energy planning, policy formulation, and decision-making processes.

Another crucial aspect of the research involves conducting a comparative analysis of Bayesian and classical models to estimate energy production. By understanding the strengths and weaknesses of each modeling approach, researchers and practitioners can identify the most appropriate method for accurate energy production estimation. This analysis contributes to the advancement of time series modeling techniques specifically within the context of energy production (Lise and Van Montfort, 2007).

Additionally, the research explores the relationship between gross domestic product GDP and energy production. By examining how changes in economic growth impact energy demand and production levels, policymakers and energy planners can make informed decisions. This analysis aids in formulating sustainable energy policies and gaining a deeper understanding of the energy requirements associated with a growing economy (Cevik et al., 2020).

The findings of this research have significant practical implications for energy policy formulation and sustainability in Türkiye. By providing accurate estimates of energy production and analyzing the factors that influence it, policymakers and industry stakeholders can make informed decisions regarding energy planning, resource allocation, and environmental sustainability. Ultimately, this research contributes to the development of sustainable energy strategies and the promotion of a resilient energy sector (Bilen et al., 2008).

1.5 Research Question

How does the accuracy of Bayesian models compare to classical models in estimating the time series data on energy production in Türkiye from 1971 to 2020?

Which modeling approach, Bayesian or classical, provides better predictions of future energy production in Türkiye based on the historical data from 1971 to 2020?

How do Bayesian models and classical models differ in capturing the trends, seasonality, and other key components of the energy production time series in Türkiye from 1971 to 2020?

What are the strengths and limitations of Bayesian models compared to classical models in estimating time series data on energy production in Türkiye from 1971 to 2020?

1.6 Hypothesis

The following hypothesis is put forth for this study based on the research problem and objectives stated previously:

Null Hypothesis (H0): There is no statistically significant difference in the accuracy of energy production estimation between Bayesian and classical models when applied to time series data in Türkiye from 1971 to 2020.

Alternative Hypothesis (Ha): There is a statistically significant difference in the accuracy of energy production estimation between Bayesian and classical models when applied to time series data in Türkiye from 1971 to 2020.

Models when applied to time series data in Türkiye from 1971 to 2020.

2. LITERATURE REVIEW

AL-Moders and Kadhim (2021) conducted a study focusing on the forecasting of oil prices using the Bayesian structural time series (BSTS) method. The research emphasized that BSTS is the most effective approach for predicting oil prices, as it has the capability to capture observed fluctuations over time and incorporate prior information. Accurate predictions of oil prices are particularly important for countries like Iraq, which heavily rely on oil revenues, as fluctuations in oil prices directly impact their overall economic well-being. Therefore, it is crucial to utilize models that can adapt to emerging events and provide reliable forecasts for future oil prices. Through their analysis, the researchers applied BSTS and projected that the price of oil is expected to reach $156.2 by 2035, indicating an upward trend in the future.

Almarashi and Khan (2020) used the BSTS to evaluate a univariate dataset in their study. The study examined real-life secondary data on Flying Cement stock prices over a one-year period. To achieve statistical results, the study used simulation approaches such as the Kalman filter and MCMC. Although the focus of the investigation was on stock prices, the same BSTS technique may be used to complicated engineering processes with lead periods. The ARIMA approach was used in the study to compare BSTS to a conventional method. To obtain Bayesian posterior sampling distributions, the R software's BSTS package was utilized. Four BSTS models were applied to a real-world dataset to show how the BSTS approach works. Forecast plots and the MAPE were used to assess the prediction accuracy of various models. The study's goal was to develop a simple technique that could be easily duplicated by both researchers and practitioners. The results showed that for short-term forecasting, ARIMA and BSTS performed similarly. However, based on the results, BSTS with a local level was recognized as the best option for long-term forecasting.

Sarpong (2013) employed the (ARIMA) model to anticipate demand in a food firm in their study. The study used a time series technique to help to demand modeling and forecasting. The study demonstrated how previous demand data may be used to forecast future demand, as well as the ramifications for the supply chain. Using historical demand data, the researchers created numerous ARIMA models using the Box-Jenkins time series process. The optimal model was chosen using four performance criteria: the Akaike criterion, the Schwarz Bayesian criterion, maximum likelihood, and standard error. ARIMA (1, 0, 1), the selected model, was tested further using additional historical demand data under the same conditions. The study's findings confirmed the ARIMA (1, 0, 1) model's usefulness in predicting and forecasting future demand in a food manufacturing company. These findings provide trustworthy counsel to manufacturing firm executives in making educated decisions based on predicted demand.

Ray et al. (2021) focused on short-term forecasting in the Indian airline industry, especially in the air passenger and air cargo sectors. The study's goal was to forecast demand for air passengers and cargo in India's aviation sector using two models: ARIMA and BSTS. The study made use of a dataset spanning a decade, from 2009 to 2018, that includes air passenger and cargo statistics gathered from the website of the Directorate General of Civil Aviation. Both the ARIMA and BSTS models were assessed in dynamic circumstances, as well as their capacity to integrate uncertainty. According to the findings, both the ARIMA and BSTS models are suitable for short-term forecasting in all four commercial aviation sectors: international passenger, domestic passenger, international air cargo, and domestic air cargo. The report also included recommendations for further research on medium- and long-term forecasting in the Indian aviation business.

Pinilla et al. (2018) used a BSTS model to examine the causal impact of partial and entire bans on public smoking on cigarette sales in their study. This method, which combines a state-space model, provides a unique means of investigating the causal consequences of policy interventions. It applies the widely used difference-in-differences technique to time series and enables the development of counterfactual scenarios using numerous control series. The report emphasizes the benefits of using this technique to evaluate the efficacy of a total ban on smoking in public places versus a partial ban.

Mourtgos and Adams (2021) compared several time series forecasting approaches in their study of modeling and forecasting the number of confirmed and mortality cases of COVID-19 in Iran. The study's goal was to find the best model for estimating the number of confirmed and fatal cases in Iran. Three measures were used to assess the performance of these models: RMSE, MAE, and MAPE. The model with the lowest performance metrics was deemed the best, and it was then used to anticipate the number of confirmed and fatal cases over the following 30 days. The study used data on the absolute number of confirmed and fatal cases in Iran from February 20 to August 15, 2020. Based on the available data in Iran, the results showed that the BSTS model performed the best for estimating the number of confirmed cases. The ARIMA model, on the other hand, was recognized as the best model for forecasting future mortality cases. According to these projections, there would be 2484 new confirmed cases and 114 new fatalities from COVID-19 on September 14, 2020.

This study extends Brooks et al. (2003) work by adding adaptive proposal strategies for reversible jump MCMC in the context of ARMA models. Unlike previous techniques, the whole conditional distribution is not accessible for the new parameters, hence estimates are proposed. An adaptive updating system is presented to improve the efficiency of between-model movements by automatically selecting proposal parameter values. The suggested algorithms' performance is tested via simulated studies, and the approach is shown by applying it to a real dataset.

Shrestha et al. (2021) examine difficulties originating from the growing usage of electronic power converter-based technologies in modern power systems, which can impair system dynamics and operational security. To maintain the safe functioning of electricity systems, transmission system operators must estimate system performance metrics. This study offers a Bayesian model that forecasts power system behavior using short-term kinetic energy time series data, giving vital support to transmission system operators in ensuring system security. For optimization, the sampler, an MCMC approach, is used in conjunction with Stan's limited-memory broyden-fletcher-goldfarb-shanno algorithm. To investigate the seasonal properties of the datasets, the study used the notion of decomposable time series modeling. To verify the model, many performance assessment criteria are employed. In addition, an ARIMA model is used to compare with the suggested model. The ideal amount of the training dataset required for successfully projecting 30-minute kinetic energy values is determined. The researchers estimate short-term kinetic energy sequences using one year of univariate data with a 1-minute precision from the integrated Nordic power system. The performance evaluation measures RMSE, MAE, MAPE, and MASE are computed. The suggested model has an RMSE of 4.67, MAE of 3.865, MAPE of 0.048, and MASE of 8.15. Increasing MCMC sampling improves performance metrics by up to 3.28, 2.67, 0.034, and 5.62, respectively. Furthermore, the study indicates that for the example study, 180.5 hours of historical data is adequate to accomplish accurate short-term forecasting with an RMSE accuracy of 1.54504.

Hooten and Hobbs (2015) Given that gold has several qualities and its price is controlled by a variety of market conditions, this study investigates the dynamic link between gold price returns and the different factors that impact it. The study then uses the (BSTS) and neural network to forecast gold price returns. These models' performance is compared against benchmark models. The findings imply that changes in crude oil returns have a positive impact on gold price returns, but shocks in the US dollar index have a negative impact. Furthermore, variations in gold price returns are heavily impacted by changes in crude oil price returns. The neural network model captures the volatility pattern of gold price returns well and improves forecast accuracy.

Ticknor (2013) presents a unique strategy for anticipating financial market behavior that incorporates a Bayesian regularized artificial neural network. The study's goal is to forecast individual stock closing values in the future utilizing daily market prices and financial technical indicators as input variables. Forecasting stock price changes is a difficult problem in financial time series research, but successful forecasts may help investors improve their stock returns significantly. The study closes with the Bayesian regularized artificial neural network being highlighted as a viable tool for stock price prediction. The suggested model has the potential to improve the accuracy of stock price forecasts by combining a probabilistic approach and lowering model complexity. This, in turn, can help investors make more educated financial market judgments.

3. MATERIAL AND METHOD

Energy production data for Türkiye, encompassing coal, gas, hydraulic, and oil, as well as GDP, covering the period from 1971 to 2020, was gathered from the World Bank website. The collected data underwent a thorough examination to identify and address missing values, outliers, and inconsistencies. Various data preprocessing techniques, such as imputation or removal of missing values, were employed as necessary. Two models, a Bayesian model and a classical model, were selected for comparison. The Bayesian model utilized the BSTS approach, while the classical model employed the ARIMA model based on the Box-Jenkins methodology. R and SPSS software were utilized for data analysis and modeling to ensure robust and comprehensive results. The BSTS model, which was implemented using Bayesian inference techniques, required the specification of appropriate prior distributions for the model parameters MCMC sampling methods were utilized to estimate the posterior distribution. The BSTS model incorporated the energy production time series data (coal, gas, hydraulic, oil) and GDP as inputs, capturing underlying trends, seasonality, and uncertainties related to energy production and GDP in Türkiye. The ARIMA model, following the Box-Jenkins approach, determined the appropriate order (p, d, q) for each energy production time series (coal, gas, hydraulic, oil) and GDP. Maximum likelihood estimation was used to estimate model parameters, and historical trends were used to anticipate future values. The accuracy and performance of the Bayesian and classical models were evaluated using evaluation metrics such as MAE, RMSE, MAPE and R^2 . The generated forecasts from each model were compared against the actual energy production and GDP values for respective years in the dataset. The findings were interpreted and analyzed to draw conclusions regarding the models' performance in estimating time series data on energy production and GDP in Türkiye. Factors contributing to the superior performance of one model over the other were identified and discussed. Limitations of the chosen models, data availability, and other influential factors were acknowledged and discussed. Based on the results and insights obtained, recommendations for future research, improvements in modeling techniques or policy implications may be provided.

3.1 Time Series Definition

A time series refers to a dataset comprising data points collected and recorded over a specific time interval. It can be mathematically represented as a set of vectors, denoted as x(t), where t represents the time duration and x(t) represents a random variable. Scholars such as (Western and Kleykamp, 2004) have discussed the concept of time series. The data in a time series are arranged in chronological order, reflecting the sequence in which measurements were captured during an event or process. Time series can be categorized as either univariate, involving data for a single variable, or multivariate, which encompasses data from multiple variables.

3.1.1 Components of Time Series

Time series data consists of several essential components that help characterize and analyze the inherent patterns and attributes of the data. These components include:

The trend component reflects the data's long-term movement or direction over time. It reflects if the data shows a steady growth, drop, or remains generally stable over a long period of time. Trends might be linear, nonlinear, or periodic in nature (Makridakis and Hibon, 1997).

Seasonality is defined as reoccurring patterns or cycles that occur at regular intervals in a time series. These patterns might change on a daily, weekly, monthly, quarterly, or yearly basis. Seasonality captures regular fluctuations that repeat within the same timeframe, typically influenced by factors like calendar events or natural phenomena (Montgomery and Nyhan , 2010).

Cyclical variation represents irregular oscillations or fluctuations in a time series that are not of fixed frequency or duration. These cycles often extend beyond the regular patterns observed in seasonality and can be influenced by economic, social, or environmental factors. Cyclical variations are often associated with business cycles or economic cycles.

The irregular component, often known as the residual or error term, compensates for noise or random fluctuations in time series data. It denotes occurrences that are unanticipated or unexpected that cannot be explained by the trend, seasonality, or cyclical components. These anomalies add to the data's variability (Newbold, 1983).

The level or mean component reflects the average value or baseline of the time series. It provides an indication of the central tendency or the typical level around which the data fluctuates (Mohamed, 2020).

3.1.2 Stationairy

The idea of stationarity is critical in time series data analysis. When the statistical properties of a time series stay consistent across time, it is said to be stationary. This means that the series has a constant mean, a stable variance, and an autocovariance that is completely determined by the time lag. A time series yt is considered stationary mathematically if it meets the following conditions:

Constant Mean (μ): The mean of the series remains constant over time.

$$E[y\_t\ ] = \mu \tag{3.1}$$

Constant Variance (σ2): The variance of the series remains constant over time.

$$Var[y_t] = \sigma 2 \tag{3.2}$$

Autocovariance (γ) depends only on the time lag (h):

The covariance between yt and yt+h depends only on the time difference or lag h, and not on the specific points in time.

$$ov[y\_t, y\_t + h] = \gamma\_((h)\ ) \tag{3.3}$$

To assess stationarity, statistical tests such as the ADF test are commonly used. These tests examine whether a time series possesses unit roots, indicating non-stationarity, or remains stationary (Western and Kleykamp, 2004).

3.1.3 Augmented Dickey-Fuller (ADF)

The ADF test is a popular statistical technique for determining whether a time series is stationary or non-stationary. It is based on the augmented Dickey-Fuller regression model, which is especially intended to examine the presence of unit roots in the time series autoregressive model. The ADF test gives insights into the stationarity qualities of the time series under examination by analyzing the significance of the predicted coefficients in the model (Cheung and Lai, 1995).

The ADF test equation may be expressed as follows:

$$\Delta y_t = \beta_0 + \beta_1 t + \beta_2 y_{t-1} + \beta_3 \Delta y_{t-2} + \beta_4 \Delta y_{t-3} + \dots + \beta_p * \Delta y_t - p + \varepsilon_t \quad (3.4)$$

In this equation:

Δy_t: represents the differenced series at time t.

β_0: The intercept term is shown.

β: denotes the time trend coefficient.

y_{t-1}: represents the lagged value of the time series.

$\Delta y_{t-1}, \Delta y_{t-2}, \dots, \Delta y_{t-p}$: represent the lagged differenced values of the time series, up *to* lag p.

$\beta_2,\ \beta_3, \dots,\ \beta_p$: *represent* the coefficients of the lagged differenced values.

ε_t: represents the error term or residual at time t.

3.1.4 Training Testing and Validation

Model training, testing, and validation requires dividing the available dataset into multiple subsets in order to evaluate the model's performance and generalizability. The conventional strategy is to divide the dataset into three subsets: training, testing, and validation. The following formulae can be used to calculate the sizes of these subsets: Training Set:

$$Size: ntrain = n \times train\_ratio \quad (3.5)$$

Where:

n: is the amount of data points in total,

train_ratiot : is the ratio (between 0 and 1) assigned for the training set.

Testing Set:

$$Size: ntrain = n \times test\_ratio \tag{3.6}$$

Where:

n: is the amount of data points in total

test_ratio : is the ratio (between 0 and 1) assigned for the testing set.

1. Validation Set:

$$Size: ntrain = n \times test\_ratio \tag{3.7}$$

The validation set is the remaining portion of the dataset after allocating the training and testing sets.

The train_ratio and test_ratio ratios can be adjusted according to specific analysis requirements and goals. It is usual practice to give a higher fraction of the data to the training set, often 70-80%, for model training reasons. The remaining data is then divided into testing and validation sets for model assessment and fine-tuning (Sheridan, 2013).

The training set is used to fit the model's parameters and comprehend the underlying patterns in the data. The testing set is used to evaluate the model's performance on unknown data and establish its generalizability. The validation set is essential for furthering the model's refinement, such as tweaking hyperparameters or comparing multiple models (Zeger and Karim, 1991).

By separating the dataset into multiple subsets, the model may be trained on one piece, evaluated on another, and validated on yet another. This process ensures reliable estimates of the model's performance and helps prevent issues like overfitting (when the model performs well on training data but poorly on unseen data) or underfitting (when the model fails to capture important patterns in the data).

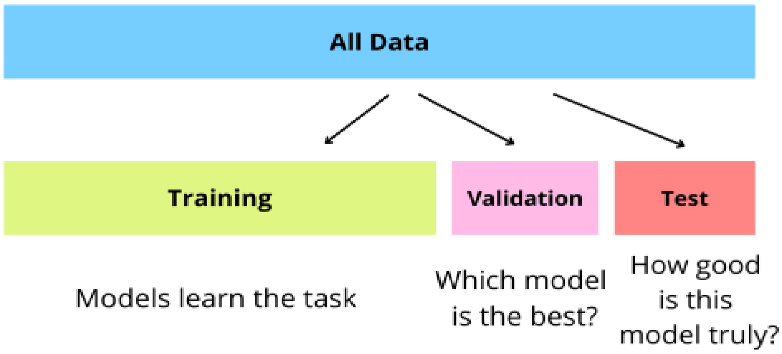

Figure 3.1 Training testing and validation

3.1.5 Autocorrelation Function (ACF)

The ACF is a statistical technique that is used to examine the relationship between a time series and its lag values. The ACF computes the correlation coefficient between the time series at time t and its preceding lag values.

The ACF equation may be expressed formally as follows:

$$\mathrm{ACF}_{(k)} = (\mathrm{C}_{(k)} - \mu) / (\mathrm{C}_{(0)} - \mu) \tag{3.8}$$

Where:

$\mathrm{ACF}_{(k)}$: represents the autocorrelation at lag k.

$\mathrm{C}_{(k)}$: represents the autocovariance at lag k.

μ: represents the mean of the time series.

$\mathrm{C}_{(0)}$: represents the autocovariance at lag 0, which is equivalent to the variance of the time series.

The autocovariance at lag k ($\mathrm{C}_{(k)}$) is calculated as:

$$C_{(k)} = Cov(y_{(t)}, y_{(t-k)}) \tag{3.9}$$

Where $y_{(t)}$ represents the time series value at time t and Cov (.) represents the covariance function.

The autocorrelation at lag k ($ACF_{(k)}$) is then calculated by dividing the autocovariance at lag k by the autocovariance at lag 0 (variance) and subtracting the mean of the time series.

The ACF is typically plotted as a function of lag, and it provides insights into the temporal dependencies and patterns in the time series. Positive autocorrelation at a specific lag indicates that the values at that lag are correlated and tend to have a similar pattern. Negative autocorrelation suggests an inverse relationship between the values at that lag. ACF values close to zero indicate little or no correlation.

The ACF is a fundamental tool in time series analysis for understanding the correlation structure of the data, identifying seasonality or periodic patterns, and determining the appropriate lag orders for AR and MA components in models like ARIMA (Heilbronner, 1992).

3.1.6 Partial Autocorrelation Function (PACF)

The PACF is a statistical technique for analyzing the relationship between a time series and its lag values while accounting for intermediate delays. The PACF computes the correlation coefficient between the time series at time t and its values at earlier lags, while ignoring the effect of intermediate lags.

The PACF equation can be expressed numerically as follows:

$$\begin{aligned} \mathrm{PACF}_{(k)} = {} & \mathrm{Cov}(y_{(t)}, y_{(t-k)} \mid y_{(t-1)}, y_{(t-2)}, \ldots, y_{(t-k+1)}) \\ & / \ \mathrm{sqrt}(\mathrm{Var}_{(y(t)} \mid y_{(t-1)}, y_{(t-2)}, \ldots, y_{(t-k+1)}) \end{aligned} \tag{3.10}$$

In this equation:

$\mathrm{PACF}_{(k)}$: represents the partial autocorrelation at lag k.

Cov: represents the covariance function.

Var: represents the variance.

$y_{(t)}$: represents the time series value at time t.

The PACF at lag k ($\mathrm{PACF}_{(k)}$) is calculated by taking the conditional covariance between the time series at time t and its lagged value at time $t - k$, given the values of the intermediate lags $(t - 1, t - 2, \ldots, t - k + 1)$. It is then divided by the square root of the conditional variance of the time series at time t, given the values of the intermediate lags.

The PACF helps in identifying the direct influence or relationship between the time series values at different lags, after accounting for the intermediate lags. It is commonly used in time series analysis, particularly in determining the lag orders for autoregressive (AR) models.

3.1.7 Autoregressive (AR)

An (AR) model is generated when a value from a time series is regressed on past values from the same time series. Then, like follo, it can generate an auto-regression process Z_t of order p, represented by the acronym AR (p).

$$y_{(t)} = c + \varphi_{(1)} * y_{(t-1)} + \varphi_{(2)} * y_{(t-2)} + \ldots + \varphi(p) * y_{(t-p)} + \varepsilon_{(t)} \tag{3.11}$$

Where:

$y_{(t)}$: indicates the time series′ value at time t.

c : is a constant term, also known as an intercept.

$\varphi_{(i)}$: The autoregressive coefficients are denoted by, where i runs from 1 to p.

$y_{(t-i)}$: represents the lagged values of the time series, with i denoting the lag. The lag ranges from 1 to p.

$\varepsilon_{(t)}$: is the error term or residual at time t, expressing the random component that the autoregressive portion does not account for.

This seems to be a multiple regression model, but instead of using external factors as predictors, it uses past values of the same series. Auto-regressive series refer to stationary processes where the variability of the terms is finite, which is determined by the values of the phi (ϕ) coefficients. Auto-regressive processes are commonly employed to represent time series data, where the current value is influenced by the previous values in a linear manner, along with the addition of random error (Wang and Wong , 2002).

3.1.8 Moving Average (MA)

The MA model is a time series analysis approach that investigates the link between a time series' present value and its previous error components or residuals. It is a statistical model that posits the present value of a series is driven by a linear combination of past time points' error terms.The Moving Average equation of order q, denoted as MA(q), can be written as:

$$y_{(t)} = c + \theta^1 * \varepsilon_{(t-1)} + \theta^2 * \varepsilon_{(t-1)} + \ldots + \theta_q * \varepsilon_{(t-q)} + \varepsilon_{(t)} \tag{3.12}$$

Where:

$y_{(t)}$: represents the value of the time series at time t.

c : is a constant term.

$\theta_1, \theta_2, \ldots, \theta$_q : are the moving average (MA) coefficients, representing the influence of past error terms.

$\varepsilon_{(t-1)}, \varepsilon_{(t-2)}, \ldots, \varepsilon_{(t-q)}$: are the lagged error terms.

$\varepsilon_{(t)}$: is the error term (residual) at time t.

According to the MA(q) equation, the current value of the time series is a linear combination of the error components at the current and previous time points, weighted by the moving average coefficients. The lagged error terms' impact on the current value is determined by the moving average coefficients (Biswas and Bhattacharyya, 2013).

3.1.9 Autoregressive Moving Average (ARMA)

The (AR) and (MA) models were combined to create the ARMA model. Because it incorporates both auto-regressive and moving-average elements, the ARMA model may describe complicated time series with fewer parameters than a comparable AR model (Makridakis and Hibon, 1997; Montgomery and Nyhan, 2010). An ARMA (p,q) model's general equation is:

$$y_{(t)} = c + \phi^1 * y_{(t-1)} + \phi^2 * y_{(t-2)} + \ldots + \phi_p * yy_{(t-q)} + \theta^1 * \varepsilon_{(t-1)} + \theta^2 * \varepsilon_{(t-2)} + \ldots + \theta_q * \varepsilon_{(t-q)} + \varepsilon_{(t)} \quad (3.13)$$

Where:

$y_{(t)}$: represents the value of the time series at time t.

c : is a constant term.

$\phi_1, \phi_2, \ldots, \phi_p$: are the autoregressive (AR) coefficients, representing the influence of past values of the time series.

$y_{(t-1)}, y_{(t-2)}, \ldots, y_{(t-q)}$: are the lagged values of the time series.

$\theta_1, \theta_2, \ldots, \theta\_q$: are the moving average (MA) coefficients, representing the influence of past error terms.

$\varepsilon_{(t-1)}, \varepsilon_{(t-2)}, \ldots, \varepsilon_{(t-q)}$: are the lagged error terms.

$\varepsilon_{(t)}$: is the error term (residual) at time t.

3.1.10 Integrated (I)

The integrated component indicates the differencing procedure used to establish stationarity on the time series (yet). The integrated component eliminates any trend or seasonality in the data by measuring the difference between successive readings. The constant term reflects the mean shift induced by differencing, and the differenced error term is denoted by 't' (Makridakis and Hibon, 1997).

$$\Delta y_{(t)} = y_{(t)} - y_{t-1} = \mu + \epsilon_{t'} \quad (3.14)$$

Where:

$\Delta y_{(t)}$: At time t, represents the differenced series.

$y_{(t)}$: signifies the first time series at t.

y_{t-1}: The lagged value of the time series.

3.1.11 Autoregressive integrated moving average (ARIMA)

ARIMA, which stands for Autoregressive Integrated Moving Average, is a prominent time series data analysis model that incorporates autoregressive (AR), differencing (I), and moving average (MA) components to capture the underlying patterns and dynamics of the series (Hossain and Abdulla, 2015).

The ARIMA model is characterized by three key parameters: p, d, and q.

The parameter p denotes the order of the autoregressive (AR) component, which represents the number of lagged observations of the dependent variable in the model. The AR component captures the link between the present value of the time series and its historical values by evaluating prior values (Pappas et al., 2010). The parameter d denotes the order of differencing, which entails transforming the time series to attain stationarity. Stationarity is the removal of trends and seasonality from data. The value of d denotes the number of times the series must be differenced before it reaches stationarity. The moving average (MA) component's order is represented by the parameter q. It represents the number of lag forecast mistakes, also known as residuals, that are included in the model. The MA component records the link between the error term and the lag errors, allowing future values to be modelled and predicted based on past mistakes. These three parameters, p, d, and q, establish the ARIMA model and enable for the study and prediction of time series data by taking autoregressive correlations into account, distinguishing for stationarity, and including the moving average effect (Jenkins, 2004).

The ARMA model is the outcome of merging the Autoregressive (AR) and Moving Average (MA) models. Because it incorporates both auto-regressive and moving-average elements, the ARMA model may describe complicated time series with fewer parameters than a comparable AR model (Makridakis and Hibon, 1997; Montgomery and Nyhan, 2010). An ARMA (p, q) model is as follows:

$$y_{(t)} = c + \phi^1 * \Delta y_{(t-1)} + \phi^2 * \Delta y_{(t-2)} + \ldots + \phi_p * \Delta\Delta y_{(t-p)} + \theta^1 * \varepsilon_{(t-1)} + \theta^2 * \varepsilon_{(t-2)} + \ldots + \theta_q * \varepsilon_{(t-q)} + \varepsilon_{(t)} \quad (3.15)$$

Where:

$y_{(t)}$: represents the value of the time series at time t.

c : is a constant term.

$\phi_1, \phi_2, \ldots, \phi_p$: are the autoregressive (AR) coefficients, representing the influence of past differenced values of the time series.

$\Delta y_{(t-1)}$, $\Delta y_{(t-2)}, \ldots,$ $\Delta y_{(t-p)}$: are the differenced values of the time series, obtained by subtracting the previous values from the current values.

$\theta_1, \theta_2, \ldots, \theta\_q$: are the moving average (MA) coefficients, representing the influence of past error terms.

$\varepsilon_{(t-1)}, \varepsilon_{(t-2)}, \ldots, \varepsilon_{(t-p)}$: are the lagged error terms.

$\varepsilon(t)$: is the error term (residual) at time t.

By estimating the coefficients and error term values, the ARIMA model may be used for forecasting, anomaly detection, and evaluating the dynamics of a time series. Model parameters can be estimated using a variety of methods, including maximum likelihood estimation (MLE).

In the ARIMA model, MLE estimation entails determining the values of c, i, and j that maximize the likelihood function given the observed data. This is usually accomplished through the use of numerical optimization techniques such as the Newton-Raphson method or the Fisher scoring system. By maximizing the likelihood function, we identify the parameter values that make the observed data most likely under the ARIMA model's assumptions. These parameter estimates can then be utilized to create predictions or evaluate the underlying time series data.

3.1.12 Maximum Likelihood Estimation (MLE) method

Maximum Likelihood Estimation (MLE) is a statistical approach for estimating statistical model parameters. It is based on the notion of determining model parameter values that optimize the possibility of witnessing the provided data.

In MLE, we assume a specified parametric form for the data's probability distribution. The objective is to determine the parameter values that make the observed data most likely. The likelihood function is calculated by multiplying the individual probability of detecting each data point under the assumed distribution by the product (or total, in the case of continuous distributions).

Consider a collection of independent and identically distributed variables (i.i.d) random variables $X_1, X_2, \ldots, X_n$, with a joint probability density or mass function f(x; θ), where θ represents the unknown parameters of the distribution. The likelihood function L(θ) is defined as:

$$L(\theta) = f(x_1; \theta) * f(x_2; \theta) * \ldots * f(x_n; \theta) \quad (3.16)$$

The log-likelihood function, denoted by $\mathcal{L}(\theta)$, is often used instead of the likelihood function. It is defined as the natural logarithm of the likelihood function:

$$\mathcal{L}(\theta) = \log(L(\theta)) = \log(f(x_1; \theta)) + \log(f(x_2; \theta)) + \ldots + \log(f(x_n; \theta)) \quad (3.17)$$

The MLE calculates the parameters by determining which values of maximize the likelihood or log-likelihood function. This is often accomplished through the use of optimization techniques such as numerical optimization algorithms. We seek the parameter values that make the observed data most likely under the expected distribution by maximizing the likelihood (Closas et al., 2007).

3.1.13 Box-Jenkins Approach

George Box and Gwilym Jenkins established the Box-Jenkins Method, which is a commonly used methodology for evaluating and predicting time series. It is a methodical and iterative procedure that consists of three major stages.

Model identification, parameter estimates, and model diagnostics (Suleman and Sarpong, 2012).

The time series data is analyzed in the model identification step to see if it displays stationarity, which indicates that its statistical features stay consistent across time. To achieve stationarity if the data is non-stationary, differencing or transformation techniques may be used. Following that, the autocorrelation function (ACF) and partial autocorrelation function (PACF) are examined to determine probable orders of autoregressive (AR) and moving average (MA) components in the model. Based on the ACF and PACF plots, several AR, MA, and ARMA models are explored, and the best model is chosen using metrics such as the AIC (Gharbi et al., 2011).

Estimation of Parameters: After determining the model structure, the parameters of the chosen AR, MA, or ARMA model are estimated using techniques such as maximum likelihood estimation (MLE) or other estimation methods. To estimate the model coefficients, iterative approaches such as the Yule-Walker equations or method of moments are used.

Model Diagnostics: The residuals of the calculated model are evaluated in this stage to find any lingering patterns or systematic behavior. To examine the unpredictability and independence of the residuals, diagnostic techniques such as the Ljung-Box test or portmanteau test are used. If significant patterns or correlations are discovered, the model may need to be adjusted or transformed. The method is repeated until the model diagnostics are adequate.

After obtaining and validating a final model, it may be used to anticipate future values of the time series. The Box-Jenkins technique to time series analysis provides a structured and systematic framework for analysts to capture underlying patterns, dynamics, and connections within the data. It has been used in a variety of sectors, including economics, finance, engineering, and environmental sciences (Jahanshahi et al., 2019).

3.1.14 Assumption of ARIMA Model

To ensure its success in modeling time series data, the ARIMA (AutoRegressive Integrated Moving Average) model involves numerous assumptions. The following are the ARIMA model's ten essential assumptions:

Stationarity: The time series data should be stable, which means that statistical features like mean and variance should remain consistent throughout time. This assumption allows the model to include autoregressive and moving average components.

Linearity: It is assumed that the connection between the observations and their lag values is linear. The ARIMA model is based on the assumption that data may be adequately represented by linear combinations of previous observations and random mistakes.

Independence: The time series observations should be independent of one another. This indicates that the value of a single observation is independent of the values of prior or subsequent observations.

No perfect multicollinearity: The ARIMA model's independent variables should not have perfect linear connections with one another. Perfect multicollinearity can lead to estimation issues and inaccurate findings.

No endogeneity: In the ARIMA model, the error term should be uncorrelated with the independent variables. Endogeneity occurs when the error term is systematically connected to the explanatory factors, causing the parameter estimates to be skewed.

Homoscedasticity: The error term's variance should be consistent across time periods. This assumption assures that the model's performance is unaffected by changes in data variability.

Normality of residuals: It is assumed that the error component has a normal distribution with a mean of zero. This assumption enables the model to do efficient estimates and hypothesis testing.

No serial correlation: The error term should not show any patterns or connection over time. In other words, the residuals should not be linked with their own delayed values.

No ARCH/GARCH effects: The error term should not show effects of autoregressive conditional heteroscedasticity (ARCH) or generalized autoregressive conditional heteroscedasticity (GARCH). These effects indicate time-varying volatility and, if not adequately accounted for, can lead to inefficient parameter estimations.

No outliers: Outliers in the data might cause parameter estimations to be distorted and the model's performance to suffer. The ARIMA model presupposes that there are no severe or influential outliers in the data.

3.2 Beysian Structural Time Series (BSTS) Model

The Bayesian structural time series (BSTS) framework is used for modeling and predicting time series data. To capture the underlying patterns, trends, and correlations in the data, it blends Bayesian inference with structural time series modeling approaches. The time series is split into numerous components in BSTS, including trend, seasonality, regression, and error. Each component is modeled independently and then integrated to provide a full time series representation. To estimate the posterior distribution of the components, the framework includes prior beliefs, data observations, and model parameters (Poyser, 2019).

The equation for BSTS can be represented as follows:

$$y_t = \mu_t + \gamma_t + \sum_{j=1}^{J} \beta_j xt_j + \varepsilon_t \tag{3.18}$$

Where:

y_t: represents the time series' observed value at time t.

μ_t : indicates the trend component of the time series, reflecting long-term changes and trends.

γ_t: represents the seasonality component, accounting for the periodic fluctuations occurring within a year or other relevant time periods.

$\beta_j xt_j$: represents the regression component, where βj denotes the regression coefficient and xtj represents the corresponding regressor variables.

ε_t : indicates the trend component of the time series, reflecting long-term changes and trends.

Given the observed data, the BSTS framework use Bayesian inference to determine the posterior distribution of the model parameters and latent components. This procedure involves updating prior beliefs using Bayes' theorem and generating samples from the posterior distribution using MCMC techniques. Analysts may use BSTS to model and forecast time series data while taking into account past knowledge, dealing with uncertainty, and capturing the structural aspects of the underlying process. The framework is scalable to a wide range of time series applications, making it an invaluable tool for data analysis and forecasting.

The trend component captures the time series' long-term systematic changes or trends. It is represented by t, which may be modeled using a variety of approaches including local linear trend, random walk, and polynomial functions (Poyser, 2019).

$$y_t = \mu_t + \varepsilon_t \tag{3.19}$$

Where:

y_t: represents the time series' observed value at time t.

μ_t : indicates the trend component of the time series, reflecting long-term changes and trends.

ε_t : indicates the trend component of the time series, reflecting long-term changes and trends.

The seasonality component accounts for repetitive patterns or seasonal fluctuations in the data. It is represented by $\sum_{s=1}^{S} \gamma s X_{s,t}$ where $X_{s,t}$ is an indicator variable representing the presence or absence of season s at time t, and γs is the corresponding coefficient.

$$y_t = \mu_t + \sum_{s=1}^{S} \gamma s X_{s,t} + \varepsilon_t \tag{3.20}$$

Where:

y_t: represents the time series' observed value at time t.

μ_t : indicates the trend component of the time series, reflecting long-term changes and trends.

ε_t : indicates the trend component of the time series, reflecting long-term changes and trends.

The regression component incorporates the influence of external factors or covariates on the time series. It is represented by $\sum_{k=1}^{K} \beta_{k,} X_{k,t} + \epsilon_t$, where xk, t is the value of the covariate k at time t, and $\beta_{k,}$ is the corresponding coefficient (Kolarik and Rudorfer,1994).

$$y_t = \mu_t + \sum_{k=1}^{K} \beta_{k,} X_{k,t} + \epsilon_t \tag{3.21}$$

Where:

y_t: represents the time series' observed value at time t.

μ_t : indicates the trend component of the time series, reflecting long-term changes and trends.

ε_t : indicates the trend component of the time series, reflecting long-term changes and trends.

The error component represents the random and unexplained variation in the time series. It is denoted by ϵt and assumed to follow a certain probability distribution (e.g., Gaussian, Poisson) with mean zero and constant variance (Vermaak et al., 2002).

$$\mathrm{yt} = \mu\mathrm{t} + \sum_{s=1}^{S} \gamma\mathrm{s}\mathrm{X}_{\mathrm{s,t}} + \sum_{k=1}^{K} \beta_{\mathrm{k,}}\,\mathrm{X}_{\mathrm{k,t}} + \epsilon\mathrm{t} \tag{3.22}$$

These equations define the BSTS model's components, where yt is the observed value of the time series at time t. The model represents the link between observable data and underlying components such as trend, seasonality, regression, and error. As previously stated, the coefficients ($\mu\mathrm{t}, \gamma\mathrm{s}, \beta_{\mathrm{k,}}$) are evaluated using Bayesian techniques to determine the posterior distribution of the parameters (Geweke, 2007).

In the technique section, you may go through the BSTS model's precise implementation details, such as the prior distributions assigned to the parameters, covariate selection, and the estimate approach utilizing MCMC sampling (Klugkist et al., 2005).

3.2.1 The (BSTS) State Model

The (BSTS) state space model is a statistical framework widely used for evaluating time series data. It is made up of two main parts: the observation equation and the state equation (Derisavi et al., 2003).

The observation equation links the observed data to the underlying latent (unobserved) state variables. It is typically represented as:

$$\mathrm{y}_{\mathrm{t}} = Z_t \alpha_t + \epsilon_t \tag{3.23}$$

Where:

y_{t} : denotes the observed data at time t.

Z_t: represents the observation matrix that maps the state variables to the observed data.

α_t : signifies the vector of state variables at time t.

ϵ_t : represents the observation error, assumed to follow a known distribution.

The state equation describes the evolution of the latent state variables over time and is typically expressed as a recursive equation:

$$\alpha_t = T_t \alpha_{t-1} + R_t \eta_t \tag{3.24}$$

Where:

α_t: denotes the vector of state variables at time t.

T_t: represents the state transition matrix, which explains how the state changes over time.

R_t: represents the state innovation matrix that captures the random shocks to the state variables.

η_t : denotes the state error, assumed to follow a known distribution.

The state space model permits the inclusion of diverse components to capture different aspects of the time series data, such as trends, seasonality, and regression effects. Each component is represented by suitable matrices in the observation and state equations.

Bayesian inference is used inside the BSTS framework to estimate unknown parameters and create predictions for future time series values. Prior distributions for model parameters are defined, and posterior distributions are produced by integrating the prior knowledge with the observed data using (MCMC) techniques. The state space model provides a versatile and effective framework for modeling and predicting time series data, allowing for the inclusion of complex patterns and relationships (Karklin and Lewicki, 2005).

3.2.2 Markov Chain Monte Carlo (MCMC) Method

MCMC is a computing approach for obtaining samples from a target probability distribution. It is commonly used to estimate the posterior distribution of model parameters in Bayesian statistical inference. The basic idea underlying MCMC is to build a markov chain that explores the relevant parameter space and finally converges to the desired posterior distribution (Toivonen et al., 2001). This is accomplished by repeatedly generating samples from a proposal distribution and accepting or rejecting them depending on a predefined criterion. The Markov chain samples offer an approximation of the posterior distribution. One of the most widely used MCMC algorithms for sampling is the Metropolis-Hastings algorithm (Gallagher et al., 2009).

Its equation can be expressed as follows:

$$\alpha = \min\left(1, \frac{p(x')q(x \mid x')}{p(x)q(x' \mid x)}\right) \tag{3.25}$$

Where:

x' : represents a proposed sample from the proposal distribution.

x : represents the current state of the Markov chain.

$p(x)$: is the target probability distribution (e.g., the posterior distribution).

$q(x' \mid x)$: is the proposal distribution, which defines the transition probability from state x to $'x'$.

Within the Metropolis-Hastings algorithm, the acceptance probability is critical in determining whether a proposed sample should be accepted or rejected. The suggested sample is always approved if it equal to or larger than 1. If, on the other hand, is smaller than 1, the suggested sample is accepted with a probability of and rejected with a probability of. The resultant Markov chain finally converges to a stationary distribution that closely approximates the desired posterior distribution after iteratively producing a sequence of samples using the Metropolis-Hastings method. The samples obtained from the chain may then be used to estimate other quantities of interest associated with the target distribution, such as means, variances, and quantiles. MCMC approaches provide a strong tool for Bayesian inference, allowing complicated model estimates and uncertainty quantification. Analysts can use MCMC approaches to draw trustworthy conclusions and gain useful insights in situations when direct analytical computations are not possible (Kass et al., 1998).

3.2.3 Dynamic Distribution Of The Model Errors In A (BSTS) Model

The probability density function (PDF) or histogram of the errors at different time points may be used to depict the dynamic distribution of model errors in a (BSTS) model. This may be stated mathematically as:

$$\mathrm{P}(\mathrm{error} \mid \mathrm{data}, \mathrm{model}) \qquad (3.26)$$

Where:

$\mathrm{P}(\mathrm{error} \mid \mathrm{data}, \mathrm{model})$: Given the observed data and the BSTS model, reflects the conditional probability distribution of the mistakes.

error : error refers to the difference between the observed data and the model's predictions at each time point.

The dynamic distribution of errors represents the model's uncertainty and unpredictability across time. It gives information on the errors' shape, spread, and trends, allowing for a thorough evaluation of the model's performance. The dynamic distribution of errors may be used to uncover any biases, outliers, or regular trends in the model's predictions. This data may be used to evaluate models, make decisions, and discover areas for model improvement (Chandra, 1993).

3.2.4 Bayesian Inference

Bayesian inference is a framework for updating our beliefs about the parameters of a statistical model based on observed data. Combining the prior distribution, which represents our starting views, with the likelihood function, which represents the data, yields the posterior distribution, which represents our updated beliefs (Geweke, 2007).
Using Bayes' theorem, the equation for Bayesian inference is as follows:

$$\mathrm{P}(\theta \mid \mathrm{data}) = \frac{\mathrm{P}(\mathrm{data} \mid \theta) \cdot \mathrm{P}(\theta)}{\mathrm{P}(\mathrm{data})} \qquad (3.27)$$

Where:

$P(\theta \mid data)$: is the posterior distribution, representing our updated beliefs about the parameter(s) θ given the observed data.
$P(data \mid \theta)$: is the likelihood function, representing the probability of observing the data given the parameter(s) θ.
$P(\theta)$: is the prior distribution, representing our initial beliefs about the parameter(s) θ.

$P(data)$ is the marginal likelihood or evidence, representing the probability of observing the data regardless of the parameter(s) θ. It serves as a normalization constant to ensure that the posterior distribution integrates to 1.

The posterior distribution is proportional to the product of the likelihood function and the previous distribution divided by the evidence, according to the equation. Calculating the evidence can be difficult in practice, but it is not always essential because it works as a normalizing constant. Based on the observed data, Bayesian inference allows us to update our ideas about the parameters, taking into account both previous knowledge and the information included in the data. After evaluating the data, the posterior distribution gives a probabilistic representation of our uncertainty regarding the parameters (Von Toussaint, 2011).

3.2.5 Prior Distribution

In Bayesian statistics, the prior distribution plays a crucial role as it represents our initial beliefs or knowledge about the parameters of a statistical model before any data is observed. It measures our parameter uncertainty and serves as the starting point for Bayesian inference. The equation for the prior distribution depends on the specific parameter being modeled and the chosen probability distribution. In general, we denote the prior distribution as P(θ), where θ represents the parameter(s) of interest (Curtis and Lomax, 2001).

The form of the prior distribution can vary depending on the problem at hand and the information available. It can be selected from a wide range of probability distributions, such as the normal (Gaussian), beta, gamma, or uniform distributions, among others. For example, if we are estimating a population mean μ using a normal distribution, we might choose a normal prior distribution with a mean of μ_0 and a variance of σ^2.

This prior distribution can be expressed as:

$$P_{(\mu)} = N(\mu \mid \mu_0, \sigma^2) \tag{3.28}$$

Where:

N: represents the normal distribution.

$P_{(\mu)}$: represents the probability density function (PDF) of the prior distribution for the population mean μ.

$N(\mu_0, \sigma^2)$: represents a normal distribution with mean μ_0 and variance σ^2.

The normal prior distribution is commonly used when there is prior knowledge or belief about the likely range or value of the population mean.

Before evaluating the data, the prior distribution embodies our preconceptions about the parameter. It can be influenced by past information, prior research, or subjective views. The prior distribution chosen has an effect on the resultant posterior distribution and inference. Bayesian inference uses Bayes' theorem to calculate the posterior distribution from the prior distribution and the likelihood function, which captures the observed data. The posterior distribution represents the changed views about the parameter(s) after taking the observed data into account.

3.2.6 Posterior Distribution

The posterior distribution is an essential concept in Bayesian statistics. It returns the updated probability distribution of the unknown parameters or variables of interest based on the observed data and any previous knowledge. The posterior distribution, which combines the prior distribution with the likelihood function, is constructed using Bayes' theorem (Tierney, 1994).

The equation for the posterior distribution is as follows:

$$P(\theta \mid x) = \frac{P(x \mid \theta).P(\theta)}{P(x)} \tag{3.29}$$

Where:

$P(\theta \mid x)$: is the posterior distribution, which denotes the parameter probability distribution θ given the observed data x.

$P(x \mid \theta)$: The likelihood function reflects the probability of seeing the dat x given the parameters θ.

$P(\theta)$: is the prior distribution, which reflects the initial assumptions or knowledge of the parameters θ before observing the data.

$P(x)$: is the marginal likelihood or evidence, which represents the probability of witnessing the data x independent of the parameters' precise values. θ. It functions as a normalization constant, guaranteeing that the posterior distribution integrates to 1.

The posterior distribution represents the parameter uncertainty after accounting for past knowledge and observed data. It serves as the foundation for Bayesian inference, allowing point estimates, credible intervals, and other statistical values of interest to be calculated (Pole et al., 2018). In reality, determining the precise form of the posterior distribution is usually computationally challenging, especially for complex models. Numerical techniques such as (MCMC) algorithms are used to determine the posterior distribution and pluck samples from it. These samples can then be used to draw conclusions and make decisions (Pérez and Berger, 2002).

3.2.7 The (BSTS) Approach

(BSTS) is a statistical modeling framework for time series analysis and forecasting. It is particularly useful when dealing with complex and uncertain time series data. BSTS allows for the decomposition of a time series into multiple components, such as trend, seasonality, and irregular components, while incorporating Bayesian inference techniques for estimation and prediction (Yilmaz, 2008).

Here are some key characteristics and components of the BSTS approach:

By breaking a time series into many components, the BSTS technique models its underlying structure.

- Trend: Represents the time series' long-term behavior or direction.
- Seasonality: Recurring trends or seasonal changes are captured.
- Regressors: Additional variables or factors that influence the time series and are included as predictors.

- Irregular or error component: Represents random noise or unexplained variations in the time series.
- BSTS incorporates Bayesian inference principles to estimate the parameters and make predictions. Bayesian approaches entail defining prior distributions for parameters, mixing them with observed data, and updating the posterior distributions via Bayes' theorem.
- The parameter posterior distributions are determined using markov chain Monte Carlo (MCMC) sampling techniques such as Gibbs sampling or Metropolis-Hastings algorithms. These posterior distributions provide a probabilistic framework for inference and the measurement of uncertainty (Geweke, 2007).
- In BSTS, model parameters such as trend, seasonality, and regressors are estimated by sampling from posterior distributions using MCMC techniques.
- Once the model is fitted to the data, forecasts can be generated by propagating the uncertainty from the posterior distributions forward in time.

- The forecasts incorporate both the estimated components of the time series and the uncertainty associated with each component, providing probabilistic forecasts instead of point estimates (Spedding and Chan, 2000).

- BSTS is a flexible framework that can handle various types of time series data and accommodate different model specifications.
- It allows for the inclusion of multiple regressors or exogenous variables that may influence the time series behavior.
- The model can adapt to changing patterns or shifts in the time series by updating the posterior distributions based on new data (Poyser, 2019).
- The BSTS model's performance may be evaluated using a variety of measures, including (MAE), (RMSE), (MAPE), coefficient of determination, and log predictive density (Amini and Parmeter, 2011).

3.2.8 Assuption of BSTS Model

To properly model time series data, the BSTS model makes certain assumptions.

Additive Errors: The model's errors or residuals are considered to be additive, which means that the observed values may be decomposed into the sum of the underlying components and the error term.

Normality of Errors: It is assumed that the errors have a normal distribution with a mean of zero. This assumption enables the model to perform efficient estimates and inference.

Independence: The time series observations are believed to be independent of one another. This indicates that the value of a single observation is independent of the values of prior or subsequent observations.

Constant Variance: The error variance is expected to be constant across time. This assumption assures that the model's performance is unaffected by changes in data variability.

Linearity: The observation-to-underlying-component connection is considered to be linear. The BSTS model captures this linearity by combining state equations with observation equations in a linear fashion.

Time-Invariance: The time series' underlying dynamics are believed to be time-invariant. This signifies that the components' connection remains consistent throughout the series.

Prior Distributions: The BSTS model is a Bayesian model that needs prior distributions for model parameters to be specified. The prior distributions used can impact the model's behavior and output, therefore they must be carefully considered.

No Omitted Variables: The BSTS model implies that all important time series impacting elements are incorporated in the model. Excluding critical factors or characteristics might result in skewed and incorrect findings.

Adequate Data: With additional historical data, the performance of BSTS models frequently increases. With more data, the model can better capture the underlying trends and generate more accurate projections (Levy , 2016).

Good Initialization: The BSTS model is based on an initialization phase that specifies the starting values of the parameters and latent states. A proper initialization is critical for ensuring model convergence and accuracy.

3.3 Time Series Performance Measurements

Time series forecasting is important in many real-world scenarios. As a result, it is critical to exercise caution when selecting a model for such projections. To evaluate the accuracy of forecasts and compare different models, various performance metrics are used. These metrics, often known as performance measures, are calculated using the time series' actual and expected values. In this part, we will cover many important performance indicators used by researchers and explain the fundamental ideas that underpin them. (Ishak and Al-Deek, 2002) .

3.3.1 Aaike Information Criterion (AIC)

The Akaike Information Criterion (AIC) is a statistical metric that is used to compare the quality and performance of various statistical models. The posterior distribution allows you to achieve a compromise between a model's quality of fit and its complexity. The AIC is a statistical metric that is used to identify the best model (Bozdogan, 1987).

It is defined as follows:

$$\mathrm{AIC} = -2 * \log(\mathrm{L}) + 2 * \mathrm{k} \tag{3.30}$$

Where:

$\log(\mathrm{L})$: is the logarithm of the likelihood function of the model,

k : is the number of parameters in the model.

The AIC weighs the model's fit to the data against its complexity, penalizing models with a large number of parameters. A lower AIC value suggests a better mix of fit and simplicity, implying a more appropriate model.

3.3.2 Root Mean Square Error (RMSE)

Root Mean Square Error (RMSE) is a widely used metric for evaluating the performance of a predictive model, particularly in regression analysis. It provides a measure of how well the predicted values of the model align with the actual observed values (Hodson, 2022). Mathematically, it is defined as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3.31}$$

Where:

y_i: represents the actual values or observations.

$\hat{y}_i$:represents the predicted values.

n: represents the total number of data points or observations

3.3.3 Mean Absolute Percentage Error (MAPE)

The Mean Absolute Percentage Error (MAPE) is a popular statistic for determining the accuracy of a forecasting or prediction model. It calculates the average size of mistakes in respect to the real values as a percentage (McKenzie, 2011).

It is calculated using the following equation:

$$\mathrm{MAPE} = \left(\frac{1}{n}\right) * \Sigma\left(\left|\frac{(\mathrm{Actual} - \mathrm{Predicted})}{\mathrm{Actual}}\right|\right) * 100 \tag{3.32}$$

Where:

Actual: represents the actual value.

Predicted: is the projected value.

|x| : denotes the absolute value of x.

n : is the total number of data points.

3.3.4 Mean Absolute Error (MAE)

MAE) is a statistic and machine learning measure that calculates the average magnitude of deviations between expected and actual data. It gives a simple method for determining the accuracy of a prediction model (Willmott and Matsuura, 2005).
It is calculated using the following equation:

$$\mathrm{MAE} = \Sigma(|\mathrm{y\_pred} - \mathrm{y\_actual}|) / \mathrm{n} \tag{3.33}$$

Where:
$|\mathrm{y\_pred} - \mathrm{y\_actual}|$: signifies the absolute difference between an anticipated and actual value.
n: is the total number of forecasts.

3.3.5 Coeffcient of Determination (R^2)

The coefficient of determination measures the linear correlation between real data and model estimations, offering an indication of the model's appropriateness. It also illustrates how much of the variance in the dependent variable can be explained by the model's components (Piepho, 2019). It is the square of the correlation coefficient and is defined mathematically as:

$$\mathrm{R}^2 = 1 - \frac{\sum(y_{\mathrm{i}} - \hat{y}_{\mathrm{i}})^2}{\sum(y_{\mathrm{i}} - \bar{y}_{\mathrm{i}})^2} \tag{3.34}$$

Where:
Yi: represents the actual data,
Ŷi: represents the corresponding model estimates,
Ȳ : represents the mean of the actual data.

3.3.6 Multiple Linear Regression

Many linear regression equations are mathematical representations of the relationship between a number of independent variables (also referred to as predictors or features) and a dependent variable (sometimes referred to as the target or response variable).

In its most generic version, the equation is as follows:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \varepsilon \qquad (3.35)$$

Where:

y : represents the dependent variable.

β_0 : is the intercept term or the y-intercept.

$\beta_1, \beta_2, \ldots, \beta_n$: are the coefficients or slopes associated with each independent variable $(x_1, x_2, \ldots, x_n)$, respectively.

$x_1, x_2, \ldots, x_n$: denote the independent variables.

ε: represents the error term or residual, which captures the unexplained variability in the dependent variable.

The coefficients $(\beta_0, \beta_1, \beta_2, \ldots, \beta_n)$ the impact or influence of each independent variable on the dependent variable is represented in the equation, whereas the intercept (β_0) accounts for the baseline value of the dependent variable when all independent variables are zero.

In practice, the coefficients $(\beta_0, \beta_1, \beta_2, \ldots, \beta_n)$ are estimated using various regression techniques, such as ordinary least squares (OLS), to find the best-fit line that minimizes the sum of squared differences between the predicted and actual values.

By substituting specific values for the independent variables $(x_1, x_2, \ldots, x_n)$ into the equation, the predicted value (y) for the dependent variable based on the estimated coefficients (Chen et al., 2014).

3.4 Republic of Türkiye

Türkiye, officially known as the Republic of Türkiye, is a country situated at the crossroads of Europe and Asia. It occupies the majority of the Anatolian Peninsula in Western Asia, with a smaller portion located in Southeast Europe on the Balkan Peninsula. Türkiye has neighboring countries in different directions: to the northwest, it shares borders with Greece and Bulgaria, to the northeast with Georgia, and to the east with Armenia, Azerbaijan, and Iran. To the south, it shares borders with Iraq and Syria. Ankara, the capital city, is Türkiye's second-biggest metropolis, while Istanbul, the largest, is an important cultural, economic, and historical hub. Türkiye offers a captivating and diverse experience, combining a wealth of historical sites, cultural heritage, and stunning natural landscapes. Türkiye is categorized as a newly industrialized country and boasts an upper-middle-income economy. It ranks 20th globally in terms of nominal GDP and holds the 11th position based on purchasing power parity (PPP). According to the World Bank, Türkiye's GDP per capita by PPP stood at $32,278 in 2021. However, in 2019, around 11.7% of the Turkish population was judged to be at danger of poverty or social exclusion. According to the same source, Türkiye's unemployment rate in the same year was 13.67% (Aydın et al., 2005).

Source: State Institute of Statistics of Turkey

Figure 3.2 Map of Türkiye

4. RESULTS AND DISCUSSION

4.1 Coal Production in Türkiye from 1971 to 2020

Türkiye has a rich history of coal mining and has been a notable producer and consumer of coal. Coal has been instrumental in meeting the energy demands of Türkiye's growing economy, especially for electricity generation. The majority of coal production in Türkiye comes from lignite mines located in the western region of the country (Toprak, 2009).

Coal production in Türkiye has experienced consistent growth throughout the 1980s and 1990s, reaching a peak of 48.9 million tons in 1986. It continues to be a crucial energy source for the country.

Despite the increase in coal production, Türkiye remains a net importer of coal, particularly high-quality hard coal used in the steel industry. In 2019, Türkiye imported around 20 million tons of coal while exporting approximately 4.5 million tons. The significant surge in coal production in Türkiye began in the 1980s following the implementation of economic liberalization policies. These policies aimed to bolster the energy and mineral resources sectors, resulting in increased coal production. Türkiye possesses ample lignite reserves, which are low-quality coal deposits distributed extensively across the country. Türkiye has emphasized the exploitation of these reserves for electricity generation (Özbayoğlu and Mamurekli, 1994).

State-owned mining enterprises produce the vast bulk of coal in Türkiye. Several privately held mining corporations do, however, operate in the nation. Coal-fired power stations utilize the majority of coal in Türkiye, accounting for over 90% of total consumption (Querol et al., 1999). Coal is also used in the iron and steel industry, cement industry, and paper and pulp industry. Türkiye has the distinction of being Europe's largest coal user. Coal has generally been a cheaper energy source than imported natural gas and oil. Coal mining in Türkiye has faced a number of problems, including safety concerns, environmental concerns, and local community hostility. Mining accidents have resulted in severe loss of life, prompting requests for greater safety laws and working conditions for coal miners (Yılmaz and Uslu, 2007).

In recent years, opposition to the development of new coal mines and power plants has grown, particularly in the Black Sea region. Concerns regarding the environmental impact of coal mining and the health risks associated with air pollution from coal-fired power plants have been raised by local communities and environmental groups (Hepbasli, 2004).

The Turkish government has set a target of 30% renewable energy in the power mix by 2023 in order to minimize reliance on coal and enhance the percentage of renewable energy. They've also invested in cleaner coal technology including high-efficiency, low-emission (HELE) coal-fired power plants, and they intend to phase out coal-fired power plants that don't fulfill environmental criteria by 2023. In the next years, Türkiye hopes to migrate to a more sustainable energy mix (Yılmaz, 2009).

A descriptive Table 4.1. showcasing the top ten countries in the world for coal production reveals China as the leading producer, followed by India, solidifying their positions as key players in the global coal industry.

Table 4.1 Top ten countries in the world for coal production

Rank	Country	Coal Production (Million Metric Tons)
1	China	3.942
2	India	767
3	Indonesia	550
4	USA	544
5	Australia	534
6	Russia	320
7	South Africa	250
8	Germany	180
9	Poland	130
10	Kazakhstan	120
11	**Türkiye**	**70.8**

4.2 Application of ARIMA on Coal Production

The ARIMA model was chosen from the collection of classical models for coal from 1971 to 2020 in Türkiye, which consists of three main methods, and was used to make sure ti forecast the future years.

4.2.1 Identification for Coal Production

The descriptive Table 4.2. for coal production from 1971 to 2020 presents key statistics that provide insights into the trends and variability of coal production over the years.

Table 4.2 Descriptive statistics for coal production from 1971 to 2020

Variables	Min	Max	Mean	SD
Coal production	22.86	48.96	29.62	5.48

The coal dataset for Türkiye spanning from 1971 to 2020 is a time series collection of yearly observations on coal production, measured in tons. It consists of 50 observations recorded annually. This dataset is commonly utilized to examine the trends and patterns in coal production in Türkiye over time. It can provide insights into the various factors that influence coal production, including economic and environmental conditions, and facilitate the forecasting of future trends in coal production in Türkiye. The dataset reveals that the minimum coal production in Türkiye during this period occurred in 1971, totaling approximately 22.86 million tons. On the other hand, the maximum coal production was recorded in 1986, reaching approximately 48.96 million tons. Over the entire period, the average coal production in Türkiye amounted to around 29.62 million tons per year. The standard deviation of coal production, which measures the data's dispersion, was approximately 5.48 million tons per year, indicating a significant variation in the production levels.

Converting data to a time series involves assigning specific dates or time periods to each observation, transforming it into a sequence of equally spaced points in time Figure 4.4. In the case of coal production data, this process would entail associating a particular date or time period with each recorded observation of coal production. Organizing the data in this manner becomes a time series that can be analyzed to identify trends, patterns, and other characteristics in coal production over time.

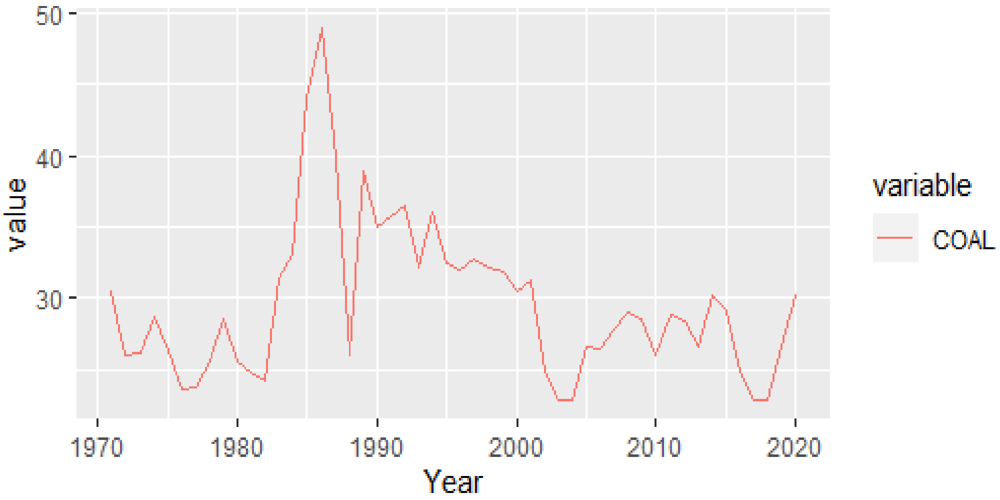

Figure 4.1 Time series plot of yearly coal production in Türkiye (1971-2020)

In time series analysis, stationarity refers to the statistical properties of a variable remaining consistent over time (Figure 4.2).

Figure 4.2 Time series plot of yearly coal production from 1971 to 2020 in Türkiye after differencing

The (ADF) test is a statistical test that determines whether a time series is stationary. The ADF test for the first-order differenced time series of coal production data. The test statistic is calculated to be 0.01226, which is less than the 0.05. This means that there is enough evidence to reject the null hypothesis of non-stationarity and conclude that the differenced time series is stationary. Figure (4.3) shows the autocorrelation function (ACF) at lag 16 measuring the correlation between a variable and its lagged value 16 time periods apart.

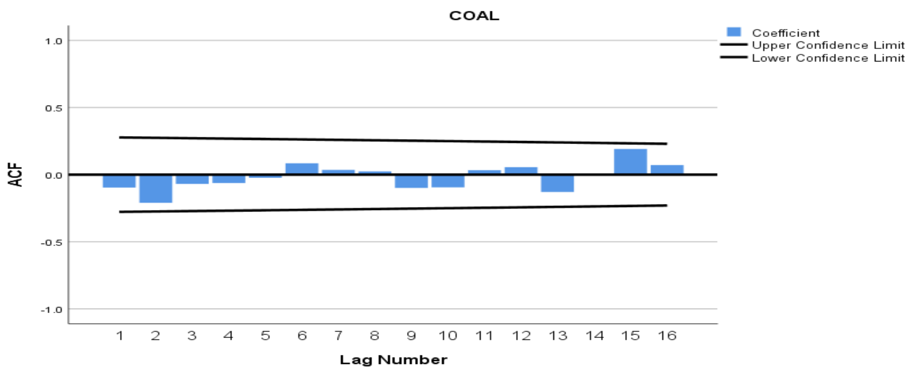

Figure 4.3 Autocorrelation function for coal production time series

The autocorrelation function (ACF) at lag 16 measures the relationship between a time series and its delayed form at a lag of 16 time units.The ACF is a statistical technique for detecting and evaluating the presence of any noticeable connection or pattern within a time series dataset. Analysts can get insights on the presence and intensity of links between observations at different periods in time by studying the ACF values at different lags, assisting in the discovery of potential patterns or dependencies within the data.

At lag 16, the partial autocorrelation function (PACF) quantifies the relationship between a variable and its lagged value 16 time periods.

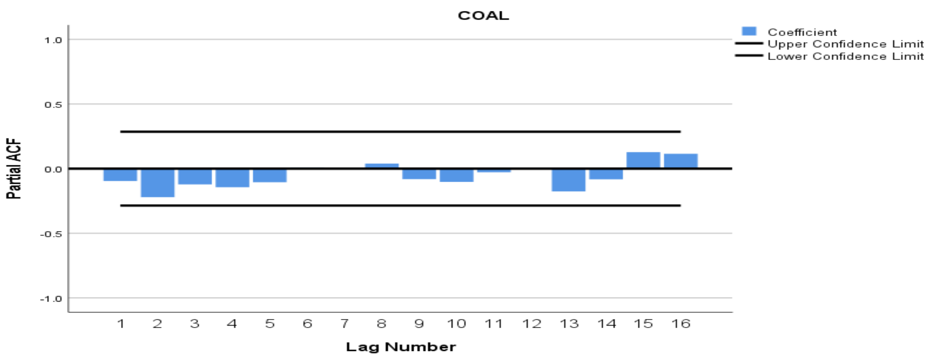

Figure 4.4 Autocorrelation function for coal production time series

The partial autocorrelation function (PACF) at lag 16 assesses the relationship between a time series and a delayed version of itself at a specified lag of 16 time units, while accounting for intermediate delays ranging from 1 to 15. It isolates the direct correlation between the observations at lag 16 by removing the influence of the shorter lags. By examining the PACF values at different lags, analysts can assess the unique relationship between observations separated by a specific time interval, providing insights into the underlying structure and dependencies within the time series data.

4.2.2 Selecting Fitting Model

Table 4.3. presents the estimated parameters of the ARIMA (1,0,0) model.

Table 4.3 ARIMA (1,0,0) model's parameters

Variables	Estimate	SE	T-test	P-value
Constant	30.11	1.86	16.1	0
AR	0.66	0.11	5.76	0

Box and Jenkins developed an interactive method for fitting (ARIMA) models to time series data. This method focuses on ensuring that the time series' mean and variance are stationary. Table 4.5. shows that the derived model is statistically significant, and the parameters are likewise statistically significant. A lower Akaike information criterion (AIC) value implies a better fit for the model. The AIC in this scenario is 282.37. These numbers indicate that the ARIMA (1,0,0) model is well-suited for capturing data patterns. As a result, by examining Figures 4.8 and 4.9, it may be used to properly estimate the parameters of the model.

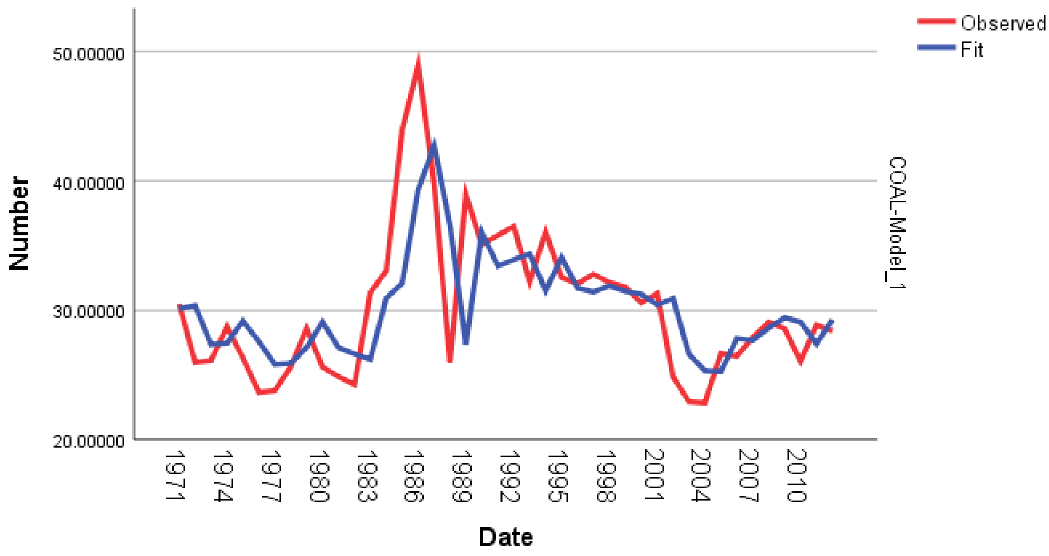

Figure 4.5 Training data for predicted value and actual values of coal production time series by using ARIMA model

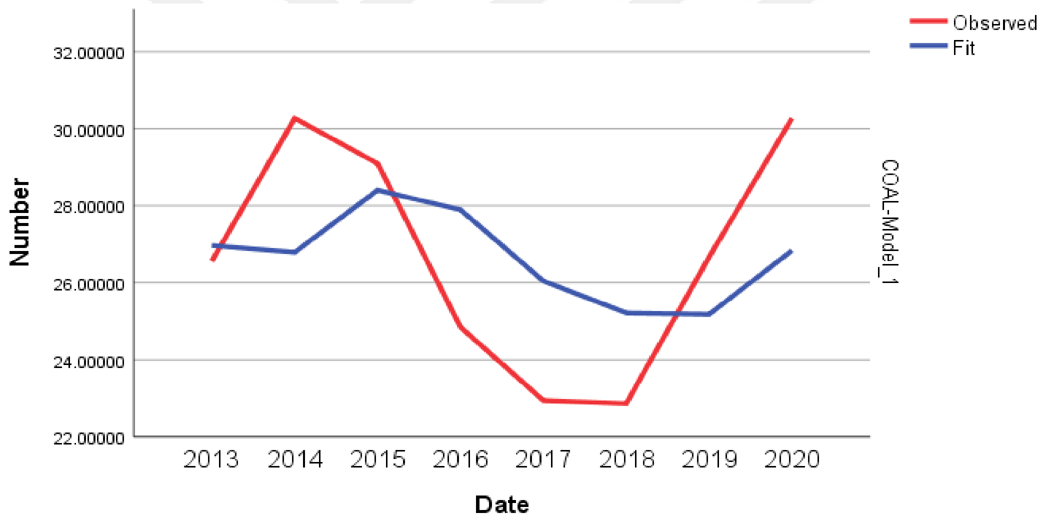

Figure 4.6 Testing data for predicted value and actual values of coal production time series by using ARIMA model

The accuracy of an ARIMA model can be evaluated by looking at Table 4.4. and 4.5. these metrics on the training set and the test set.

Table 4.4 ARIMA (1,0,0) model statistics (training data)

MAE	RMSE	MAPE	R^2
1.267	1.984	3.546	0.894

Table 4.5 ARIMA (1,0,0) model statistics (testing data)

MAE	RMSE	MAPE	R^2
1.327	2.134	3.980	0.865

4.2.3 Model Checking (1,0,0)

Figures (4.7) and (4.8) show the residuals of an ARIMA (1,0,0) model with a non-zero mean. These charts emphasize the differences between the actual observed values and the ARIMA model's anticipated values for the provided time series data. analyzing the residual distribution assists in analyzing the model's assumptions and finding potential issues such as heteroscedasticity (changing variances) or non-normality. The residuals should ideally show random fluctuations centered around zero, suggesting that the model adequately captured the underlying data patterns. Once the candidate ARIMA (1,0,0) model was identified and estimated, the next step involved assessing its fit to the data through parameter and residual analysis. To diagnose the model, the residuals were subjected to diagnostic testing using ACF and PACF plots, as depicted.

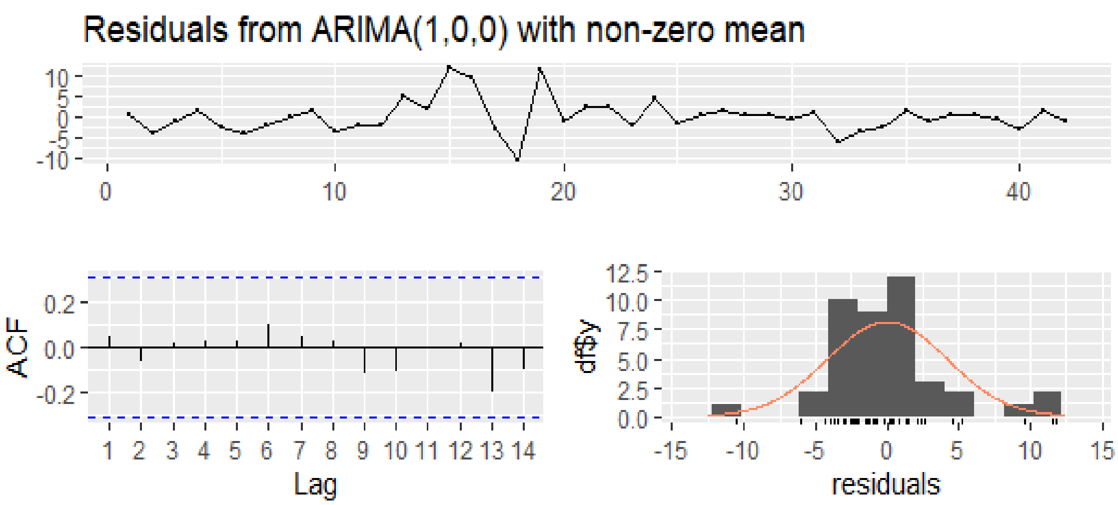

Figure 4.7 Residuals from ARIMA (1,0,0) with non-zero mean

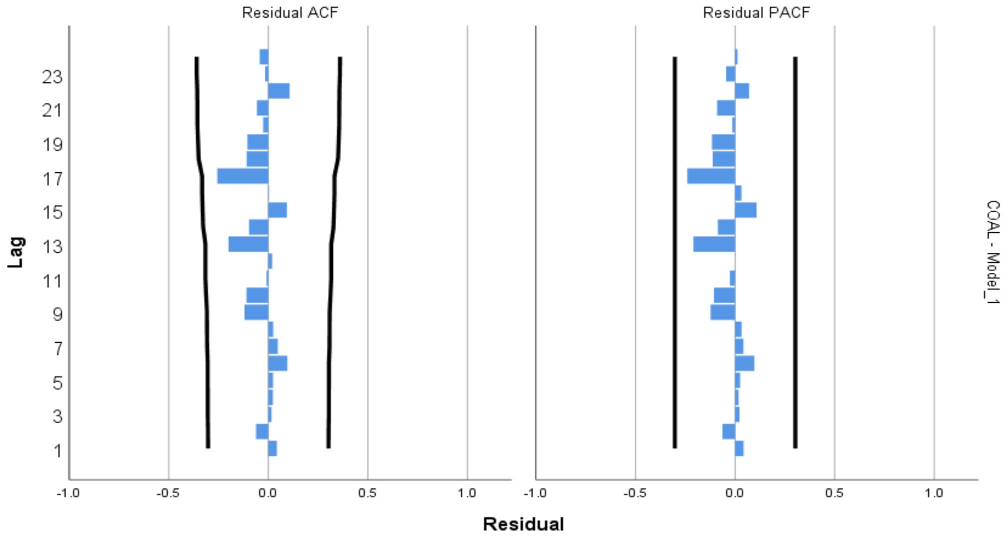

Figure 4.8 Residual ACF and PACF for ARIMA (1,0,0)

All ACF and PACF residuals values were statistically significant at the 95% confidence level, showing that the residuals reflect a pattern of random white noise, according to these graphs. This implies that the ARIMA (1,0,0) model is suitable for the supplied data. The Box-Ljung test was used to look for autocorrelation in the residuals of the ARIMA (1,0,0) model. The observed autocorrelations of the residuals are compared to the anticipated values under the assumption of no autocorrelation. The computed p-value from the Box-Ljung test in the submitted findings was 0.8802, which is larger than the significance level of 0.05. This implies that there is no indication of autocorrelation in the model's residuals. As a result, it is possible to conclude that the model successfully represents the autocorrelation structure in the data.

Table 4.6. presents the actual and predicted values of coal production from 2013 to 2020. It provides a comparison between the observed values, representing the real coal production data, and the forecasted values obtained from a particular model or method.

Table 4.6 The actual and predicted values of coal production in from 2013 to 2020

Date	Actual	Forecast
2013	26.56046	28.96927
2014	30.26754	29.35022
2015	29.09509	29.60442
2016	24.84467	29.77405
2017	22.94193	29.88725
2018	22.85896	29.96278
2019	26.66897	30.01319
2020	30.26754	30.34682

Furthermore, the ARIMA (1,0,0) model was used to forecast Türkiye's annual coal output for the year 2020. The figure in Figure 4.9. depicts the projected values for 2020, which are quite similar to the actual numbers. This convergence of anticipated and actual values demonstrates that the model accurately represents the behavior and patterns found in the coal production series for that year.

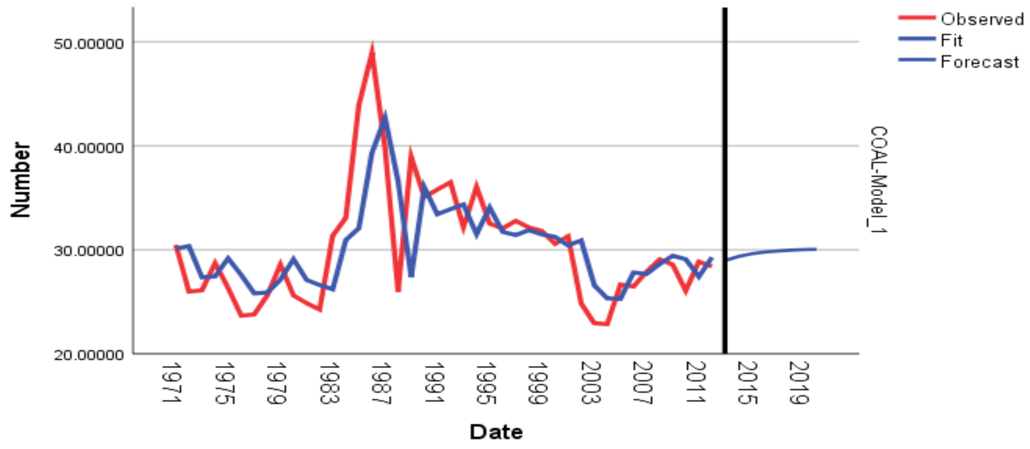

Figure 4.9 Predicted values of coal

4.3 Natural Gas Production In Türkiye from 1971 to 2020

Türkiye plays a significant role in the natural gas market, with a substantial balance between imports, exports, and domestic consumption. Türkiye is a major importer of natural gas to meet its growing energy demand. Imports are received through pipelines and liquefied natural gas (LNG) shipments (Umucalilar et al., 2002).

Türkiye receives most of its natural gas imports through pipelines from different supplier countries, including Russia, Azerbaijan, and Iran. Türkiye has LNG terminals at Marmara Ereğlisi, Samsun, and İzmir, receiving shipments from countries like Qatar, Algeria, and Nigeria (Saidur et al., 2010). Türkiye has started exporting natural gas, albeit on a smaller scale compared to its imports. Neighboring countries, particularly Greece and Bulgaria, receive natural gas from Türkiye through interconnector pipelines (Baser et al., 1998).

Türkiye has considerable domestic consumption in a variety of sectors, including power generation, industrial, residential, and commercial. Natural gas is a primary fuel source for electricity generation, and it is used to power industry and residential/commercial structures. Türkiye intends to diversify its energy sources, strengthen its energy security, and boost energy efficiency and renewable energy. The government is working to reduce reliance on fossil fuels, especially natural gas, by boosting the proportion of renewable energy in the energy mix (Kaygusuz, 2002).

Investments in renewable energy projects, such as wind and solar, aim to gradually decrease reliance on natural gas and other fossil fuels. In the 1970s, natural gas production was minimal, and the country heavily relied on imports. The 1980s marked significant progress in exploration and production, with the Turkish Petroleum Corporation (TPAO) leading the efforts. Important milestones included the discovery of the Karakaya gas field and subsequent production. Natural gas production capacity gradually increased, reducing reliance on imports (Demirbaş, 2001).

The 1990s saw continued growth in production, with additional discoveries like the Thrace Basin gas fields. Infrastructure expansion, including pipelines and storage facilities, was prioritized. The 2000s witnessed a notable increase in production due to new field development and foreign investment. Infrastructure improvements included pipeline expansion and LNG import terminals. From 2010 to 2020, natural gas production continued to grow with ongoing exploration and technological advancements. Notable discoveries, such as the Sakarya gas field and the Tuna-1 well, further contributed to the production. Türkiye aimed to diversify supply sources and increase the share of natural gas in its energy mix during this period (Kaygusuz, 2003).

A descriptive Table 4.7. showcasing the top ten countries in the world for natural gas production reveals United States as the leading producer, followed by Russia, solidifying their positions as key players in the global coal industry.

Table 4.7 Top ten countries in the world for natural gas production

Rank	Country	Natural Gas Production (Billion Cubic Meters)
1	United States	914.6
2	Russia	638.5
3	Iran	250.8
4	China	194
5	Qatar	171.3
6	Canada	165.2
7	Australia	142.5
8	Saudi Arabia	112.1
9	Norway	111.5
10	Algeria	81.5
74	Türkiye	13.4

Source: The World Bank organization.

4.4 Application of ARIMA on Gas Production

The ARIMA model was chosen from the collection of classical models for gas from 1971 to 2020 in Türkiye, which consists of three main methods, and was used to make sure ti forecast the future years.

4.4.1 Identification for Gas Production

The descriptive Table 4.8 for coal production from 1971 to 2020 presents key statistics that provide insights into the trends and variability of coal production over the years.

Table 4.8 Descriptive statistics for gas production from 1971 to 2020

Variables	Min	Max	Mean	S.D
Gas production	0.09	49.73	23.29	19.62

The time series data spanning from 1971 to 2020 offers valuable insights into the annual natural gas production in Türkiye. This dataset focuses on a single variable, namely the amount of natural gas produced in billion cubic meters (bcm), with 50 observations in total. The data demonstrates a clear upward trend, indicating a consistent growth in Türkiye's natural gas production over time. However, there are also noticeable fluctuations within the data, suggesting the influence of various factors such as weather patterns, economic conditions, and energy sector policies.

Analyzing the summary statistics, the average gas production over the 50-year period was approximately 23.29 bcm, with a standard deviation of 19.62 bcm. These figures indicate a considerable degree of variability in annual natural gas production. The minimum recorded value in the dataset, 0.09 bcm, likely corresponds to a year with relatively low production, while the maximum value of 49.73 bcm signifies a year with exceptionally high production. These descriptive statistics provide a comprehensive overview of the natural gas production data, revealing its central tendency, spread, and the range of values observed. To gain deeper insights into the patterns, trends, and potential forecasting of natural gas production in Türkiye, further analysis can be conducted using time series models such as ARIMA. These models can aid in the discovery of the underlying dynamics of the data and give useful information for decision-making and future forecasts.

Transforming data into a time series involves structuring the data as a sequence of evenly spaced time intervals Figure 4.10. This entails assigning specific dates or time periods to each data point. In the context of gas production data, this would involve assigning a time period to each observation of gas production. By creating a time series in this manner, we can analyze and detect patterns and trends in gas production over time.

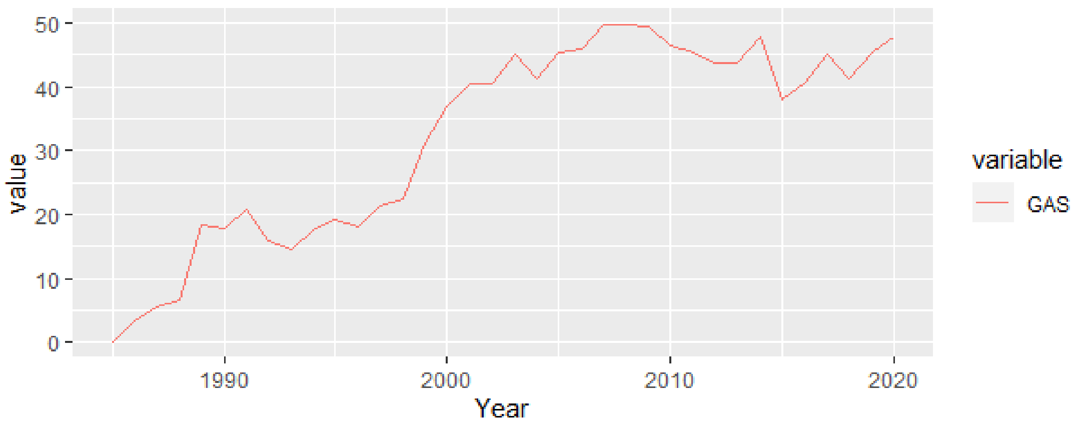

Figure 4.10 Time series plot of yearly gas production in Türkiye

Stationarity in time series analysis refers to the statistical properties of a variable remaining consistent over time, as shown in (Figure 4.11).

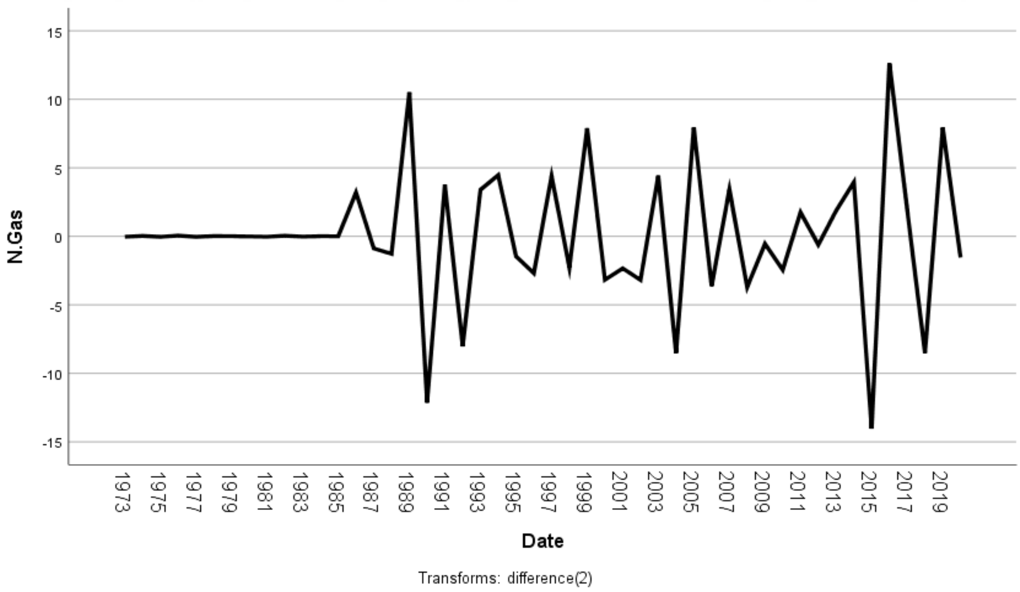

Figure 4.11 Time series plot of yearly gas production in Türkiye after two differencing

The (ADF) test was performed on the gas time series data's second difference, which was generated by subtracting the time series from itself twice. The resultant p-value is 0.02071, which is less than the significance level of 0.05 that was set. The autocorrelation function (ACF) plot displays the autocorrelation values at various lags, ranging from 0 to 16, for the "GAS" time series data. This plot is useful for identifying significant correlations or recurring patterns in the data. The ACF values offer valuable information about the lag structure, aiding in the selection of suitable models for time series analysis and forecasting.

Figure 4.12 shows the (ACF) at lag 16 measuring the correlation between a variable and its lagged value 16 time periods.

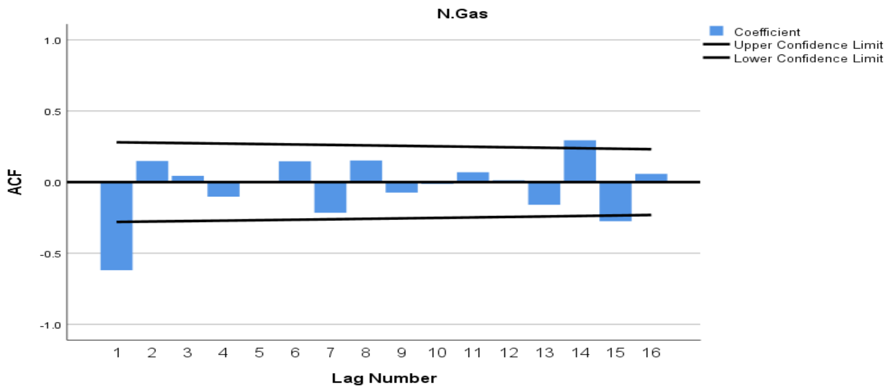

Figure 4.12 Autocorrelation function for gas production time series

Figure 4.13 shows the partial autocorrelation function (PACF) at lag 16 measuring the correlation between a variable and its lagged value 16 time periods.

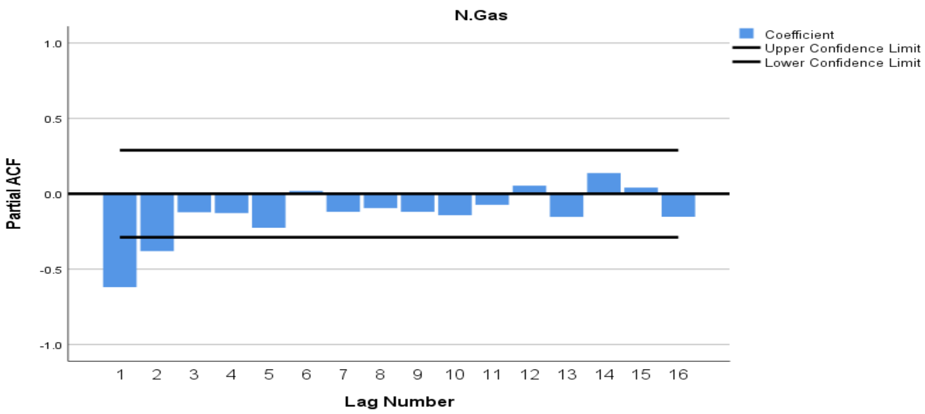

Figure 4.13 Partial autocorrelation function for gas production time series

The (PACF) plot displays the partial autocorrelation values for the gas time series at various delays ranging from 0 to 16. This plot aids in evaluating the significant lags in the data and comprehending the correlation structure. By examining the PACF values, one can determine the appropriate lag order for autoregressive models and identify any noteworthy patterns within the gas data.

4.4.2 Selecting Fitting Model

Box and Jenkins proposed an interactive method for fitting autoregressive moving average models to time series data that takes into account the data's stationarity in terms of mean and variance. The findings reported in Table (4.9) show that the calculated model is statistically significant, as are its parameters. A lower akaike information criterion (AIC) value implies a better fit of the model to the data. The AIC score in this example is 261.82, indicating that the model is well-suited to the data. The estimated model is an ARIMA model with a differencing order of one, abbreviated as ARIMA (0,1,0), and its parameters can be calculated. Table 4.9 presents the estimated parameters of the ARIMA (0,1,0) model.

Table 4.9 ARIMA (0,1,0) model's parameters

Variables	Estimate	SE	T-test	P-value
Constant	0.1644	0.659	2.494	0.019

As a result, by examining Figures 4.17 and 4.18 it may be used to properly estimate the parameters of the model.

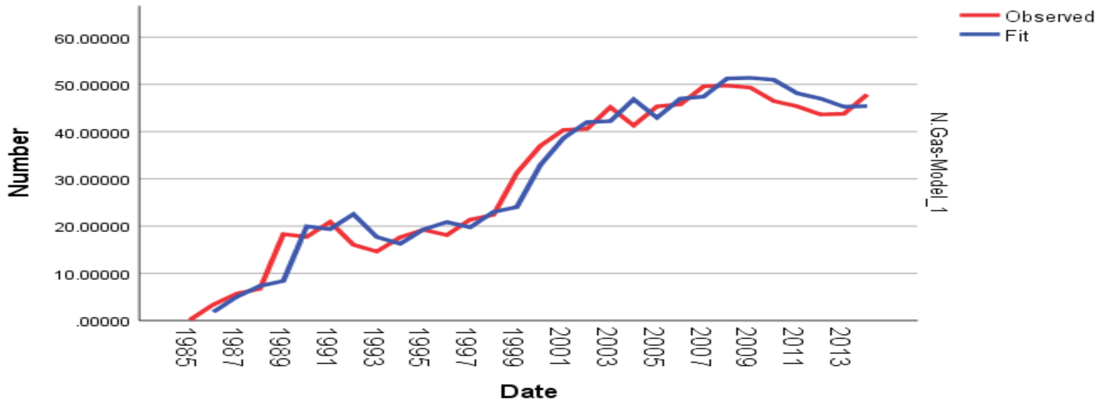

Figure 4.14 Training data for predicted value and actual values of gas production time series by using ARIMA model

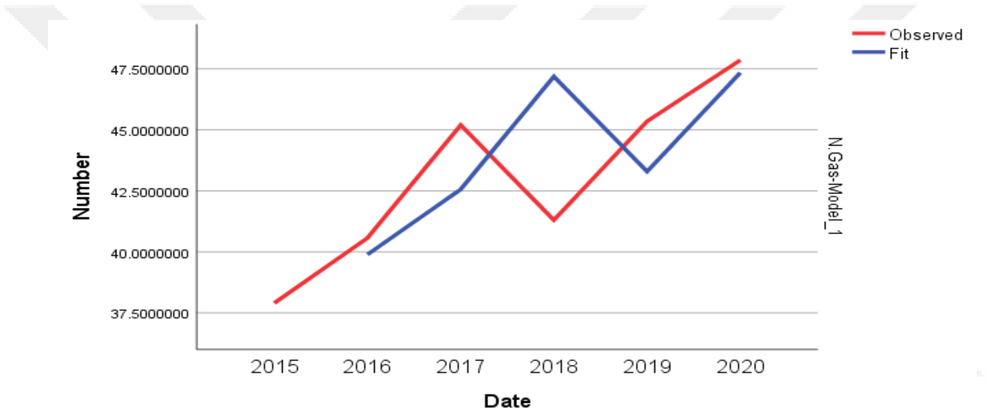

Figure 4.15 Testing data for predicted value and actual values of gas production time series by using ARIMA model

Various metrics for the correctness of an ARIMA model may be derived for both the training and testing sets, with a focus on the metrics acquired from the test set (Tables 4.10 and 4.11) The metrics received from the training set indicate how well the model matches the data on which it was trained, whereas the metrics collected from the test set indicate how well the model is predicted to generalize to new data.

Table 4.10 ARIMA (0,1,0) model statistics (training data)

MAE	RMSE	MAPE	R^2
0.034	0.046	0.205	0.988

Table 4.11 ARIMA (0,1,0) model statistics (testing data)

MAE	RMSE	MAPE	R^2
0.044	0.058	0.246	0.986

4.4.3 Model Checking (0,1,0)

Figures 4.16. and 4.17. show the residuals from an ARIMA (0,1,0) model with a non-zero mean, as well as the ACF and PACF for ARIMA (0,1,0).

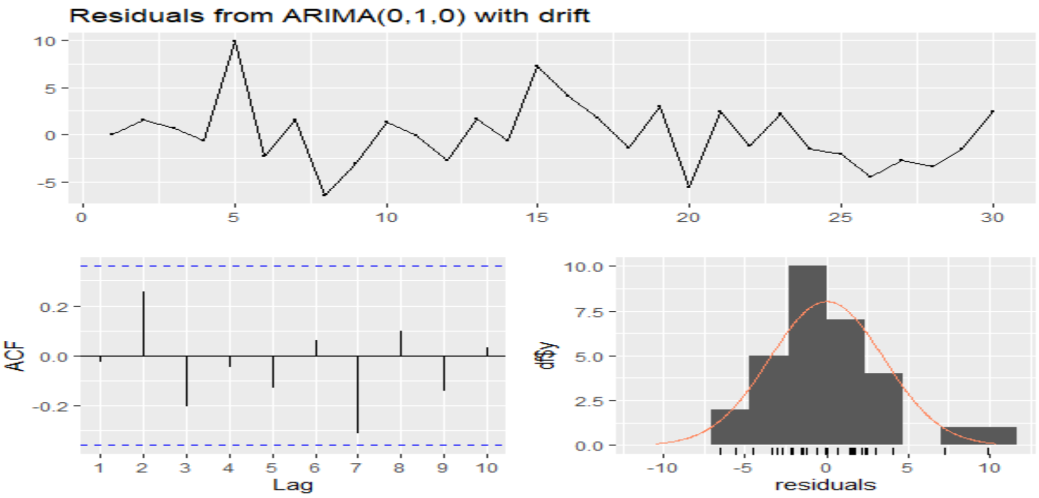

Figure 4.16 Residuals from ARIMA (0,1,0) with non-zero mean

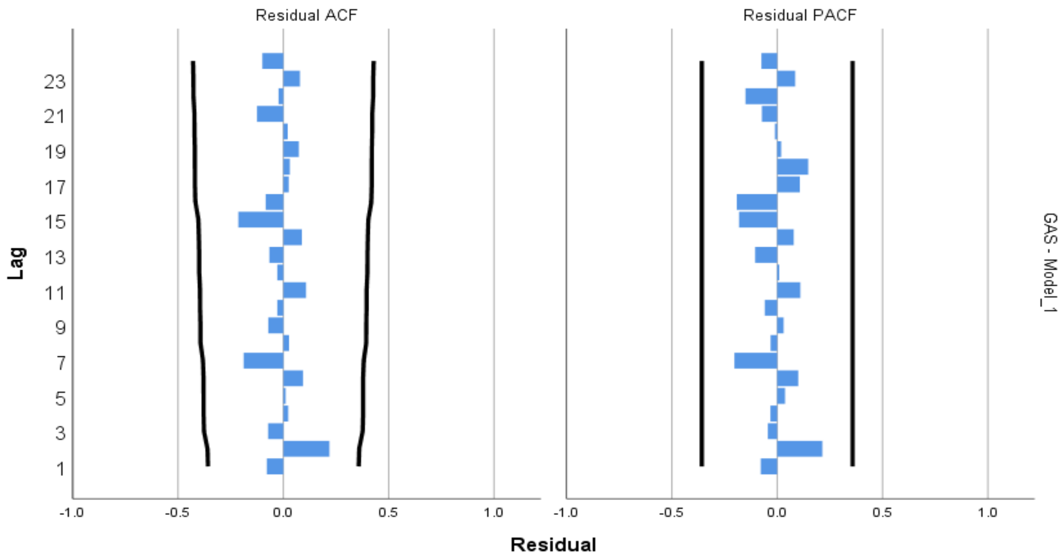

Figure 4.17 Residual ACF and PACF for ARIMA (0,1,0)

Following the identification and estimation of the candidate ARIMA (0,1,0) model, it is critical to assess the model's quality of fit to the data using diagnostic tests. This evaluation entails examining both the model's parameters and the residuals. In this scenario, the residuals from the ARIMA (0,1,0) model were tested using autocorrelation function (ACF) and partial autocorrelation function (PACF) plots, as shown in Figure 4.19. All residual values in the ACF and PACF plots were statistically significant at the 95% confidence level, according to the study. This conclusion demonstrates that the residuals have random white noise properties, implying that the model adequately reflects the patterns in the data. In addition, the ARIMA (0,1,0) residuals were subjected to the Box-Ljung test, which is used to detect autocorrelation in time series residuals. The test compares the observed residual autocorrelations to the anticipated values assuming no autocorrelation. The computed p-value from the Box-Ljung test in the submitted findings was 0.61, which is larger than the significance level of 0.05. This implies that there is no significant evidence of autocorrelation in the model's residuals. As a result, it is possible to infer that the ARIMA (0,1,0) model adequately describes the data's autocorrelation structure.

Table 4.12. presents the actual and predicted values of gas production from 2013 to 2020. It provides a comparison between the observed values, representing the real gas production data, and the forecasted values obtained from a particular model or method.

Table 4.12 The actual and predicted values of gas production in from 2015 to 2020

Date	Actual	Forecast
2015	37.90124	48.96477
2016	40.56955	50.07488
2017	45.19530	51.18500
2018	41.30181	52.29512
2019	45.34874	53.40524
2020	47.85465	54.51536

Furthermore, the ARIMA (0,1,0) model was utilized to make a forecast of the yearly gas production in Türkiye for the year 2020. As depicted in Figure 4.18, the plot reveals that the predicted values for 2020 exhibit a behavior that closely resembles the actual values. This convergence between the predicted and actual values indicates that the ARIMA (0,1,0) model effectively captures the patterns and trends observed in the gas production series for that specific year.

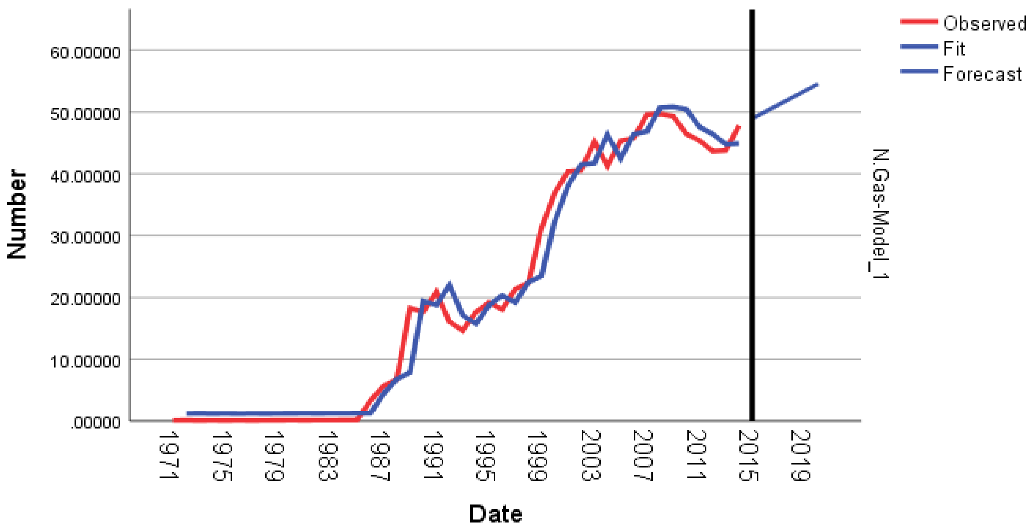

Figure 4.18 Predicted values of gas production in 2020

4.5 Hydraulic Production in Türkiye from 1971 to 2020

Hydroelectric power is a crucial component of Türkiye's energy supply, utilizing its abundant water resources. Türkiye primarily consumes hydropower domestically and has limited exports or imports. To diversify the energy mix and minimize reliance on fossil fuels, the government has actively supported renewable energy, especially hydropower (Cinar et al., 2010). However, the development of hydropower faces challenges such as environmental and social impacts, which the government addresses through regulations and sustainability measures (Salvarli, 2006).

In terms of historical progression, hydroelectric power production in Türkiye had a gradual increase from 1971 to 2020. During the 1970s, small and medium-scale hydroelectric plants were built for local electricity needs (Oğulata, 2003). The 1980s witnessed significant expansion with the construction of larger-scale plants. In the 1990s, the government focused on further increasing renewable energy's share, resulting in more medium to large-scale hydroelectric projects (Yuksek et al., 2006).

The period from 2000 to 2010 saw a significant boom, with numerous new plants across the country. By 2020, Türkiye had an installed capacity of over 30,000

MW. Hydroelectric power has helped Türkiye generate electricity while lowering greenhouse gas emissions and expanding renewable energy capacity (Warner, 2008).

It is important to note that the growth of hydroelectric power has faced challenges and controversies, particularly regarding environmental impact and displacement of local communities (Erdogdu, 2011). Nevertheless, Türkiye's focus on renewable energy and sustainable development has helped address these concerns and promote a greener energy sector (Kaygusuz, 2009).

A descriptive Table 4.13. showcasing the top ten countries in the world for hydroelectric production reveals China as the leading producer, followed by India, solidifying their positions as key players in the global hydroelectric industry (Akpınar et.al., 2011).

Table 4.13 Top ten producing countries for hydroelectric

Rank	Country	Hydroelectric Power Generation (GWH)
1	China	1,363,000
2	Canada	396,000
3	Brazil	391,000
4	US	306,000
5	Russia	169,000
6	India	147,000
7	Norway	144,000
8	Venezuela	129,000
9	Japan	90,000
10	Sweden	79,000
14	Türkiye	55

Source: The World Bank organization

4.6 Application of ARIMA on Hydraulic Production

The ARIMA model was chosen from the collection of classical models for Hydraulic from 1971 to 2020 in Türkiye, which consists of three main methods, and was used to make sure ti forecast the future years.

4.6.1 Identification for Hydraulic Production

The descriptive Table 4.14 for hydraulic production from 1971 to 2020 presents key statistics that provide insights into the trends and variability of hydraulic production over the years.

Table 4.14 Descriptive statistics for hydraulic production from 1971 to 2020

Variable	Min	Max	Mean	S.D
Hydraulic production	16.13	60.25	32.75	10.76

These statistics offer a summary of the range and typical values of hydraulic production during the given timeframe. The lowest recorded production is represented by the minimum value of 16.13, while the highest recorded production is indicated by the maximum value of 60.25. The mean value of 32.75 is the average hydraulic production during the whole time. The standard deviation of 10.76 represents the degree of variability or dispersion in the hydraulic production statistics. A larger standard deviation suggests a wider range of production values, while a smaller standard deviation indicates more consistent production levels.

When transforming data into a time series, each data point is allocated a distinct date or time period, creating a sequence of evenly spaced points in time Figure 4.19. In the context of hydraulic production data, this would involve assigning a specific date or time period to each observation. By studying this time series, analysts can detect trends and patterns in hydraulic production over the course of time.

Figure 4.19 Time series plot of yearly hydraulic production in Türkiye (1971-2020)

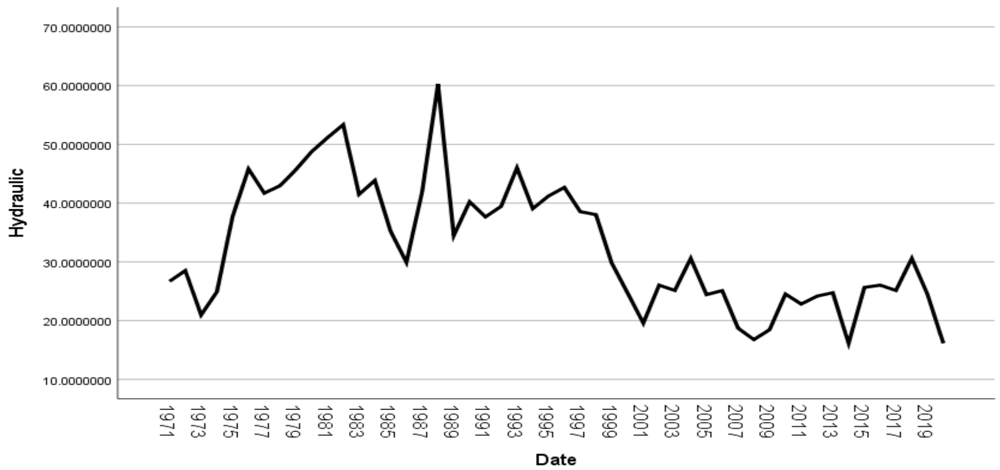

Figure 4.20 Time series plot of yearly hydraulic production in Türkiye (1971-2020)

Stationarity in time series analysis refers to a variable's statistical features staying stable throughout time, as seen in Figure 4.20.

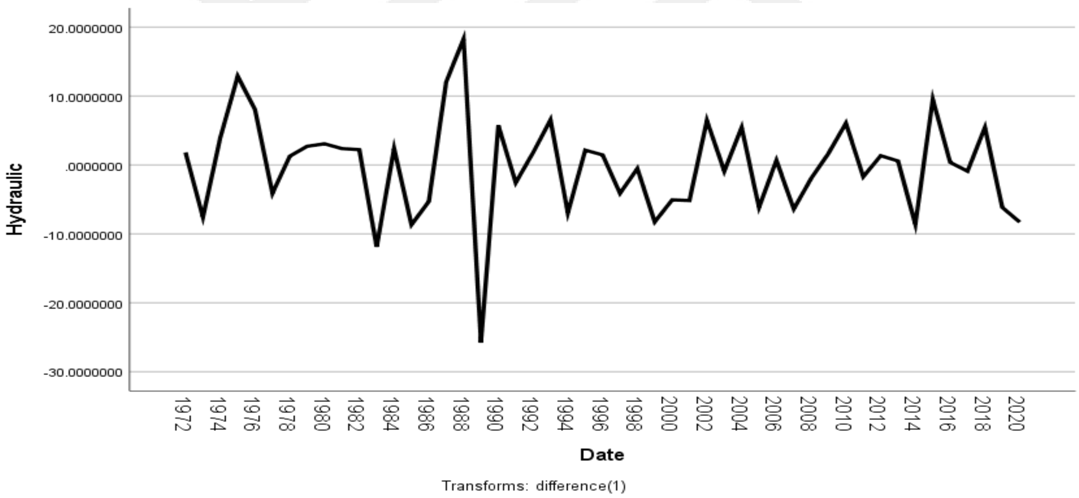

Figure 4.21 Time series plot of yearly hydraulic production in Türkiye after one differencing

On the first-order differenced hydraulic production time series data, the (ADF) test was performed. These results indicate that the first-order differenced time series of hydraulic production data exhibits stationarity.

The autocorrelation function (ACF) is the correlation between a variable and its lagged time periods apart in Figure 4.21.

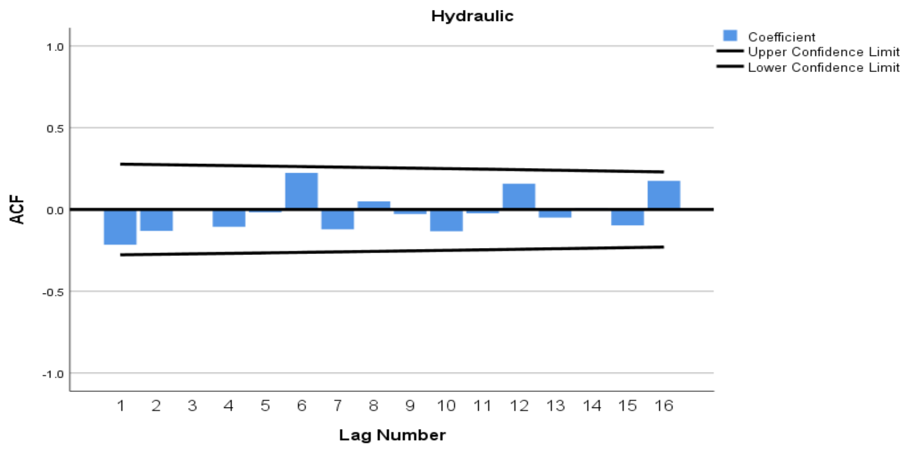

Figure 4.22 Autocorrelation function for hydraulic production time series

The autocorrelation function (ACF) plot can reveal the extent of autocorrelation present in the data. When the ACF plot displays significant correlation at specific lags, it indicates the existence of a repeating pattern in the data at those particular lags.

Figure 4.22 shows how the (PACF) at lag assesses the correlation between a variable and its delayed time periods.

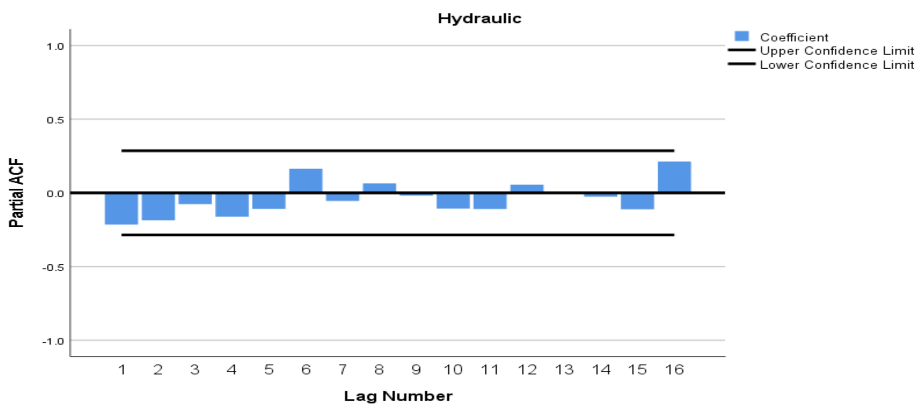

Figure 4.23 Partial autocorrelation function for hydraulic production time series

The (PACF) is a useful tool in time series analysis and modeling because it identifies certain lagged values that have a substantial association with the variable under study.

4.6.2 Selecting Fitting Model

Table 4.15 presents the estimated parameters of the ARIMA (0,1,0) model.

Table 4.15 ARIMA (0,1,0) model's parameters

Variables	Estimate	SE	T-test	P-value
Constant	-0.062	1.165	-0.053	0.045

Box and Jenkins developed an interactive approach for fitting autoregressive moving average models to time series data that accounts for the data's stationarity around its mean and variance. Tables 4.15. show that the calculated model and its parameters are statistically significant. A lower akaike information criterion (AIC) value indicates a better fit of the model to the data. The AIC in this example is 334.34 (adjusted for the short sample size), suggesting a strong fit for the data. Figures 4.23. and 4.24. show how the ARIMA (0,1,0) model may be used to estimate the parameters.

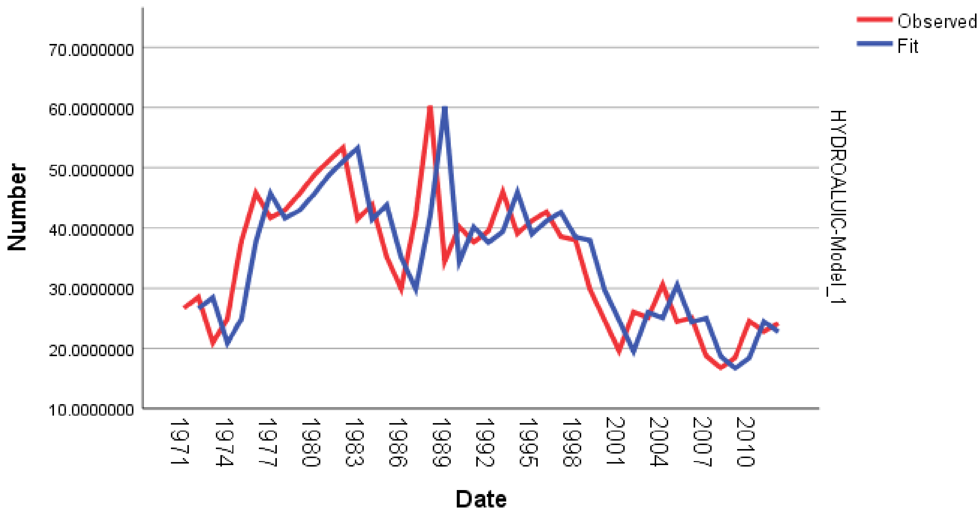

Figure 4.24 Training data for predicted value and actual values of hydraulic production time series by using ARIMA model

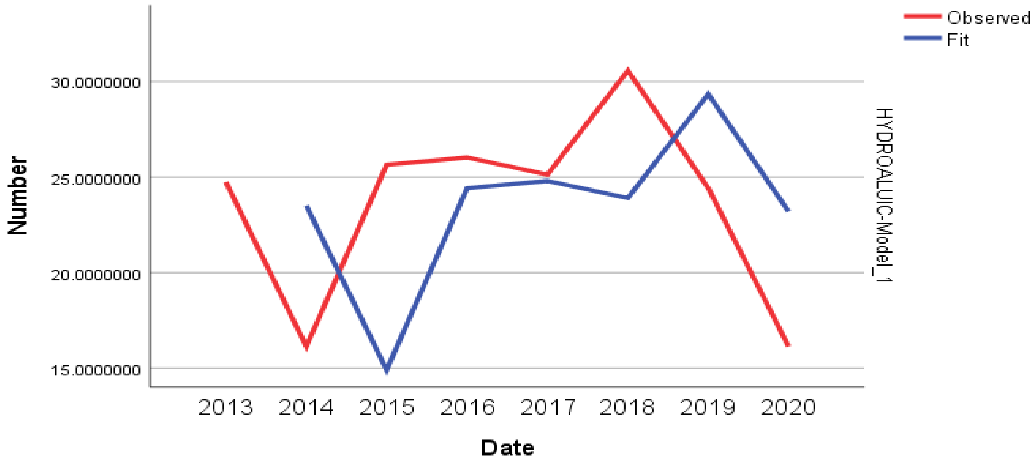

Figure 4.25 Testing data for predicted value and actual values of hydraulic production time series by using ARIMA model

The performance of an ARIMA model may be evaluated by looking at these metrics on both the training and test sets (Tables 4.16 and 4.17). The metrics calculated on the training set offer insights into how effectively the model fits the available data. The metrics produced on the test set, on the other hand, indicate how effectively the model can generalize to new, previously unknown data. By evaluating these metrics, we can gauge the accuracy and reliability of the ARIMA model.

Table 4.16 ARIMA (0,1,0) model statistics (training data)

MAE	RMSE	MAPE	R^2
4.158	5.304	8.216	0.742

Table 4.17 ARIMA (0,1,0) model statistics (testing data)

MAE	RMSE	MAPE	R^2
4.879	6.263	9.701	0.707

4.6.3 Model Checking (0,1,0)

Figures 4.25 and 4.26 show the residuals from an ARIMA (0,1,0) model with a non-zero mean.

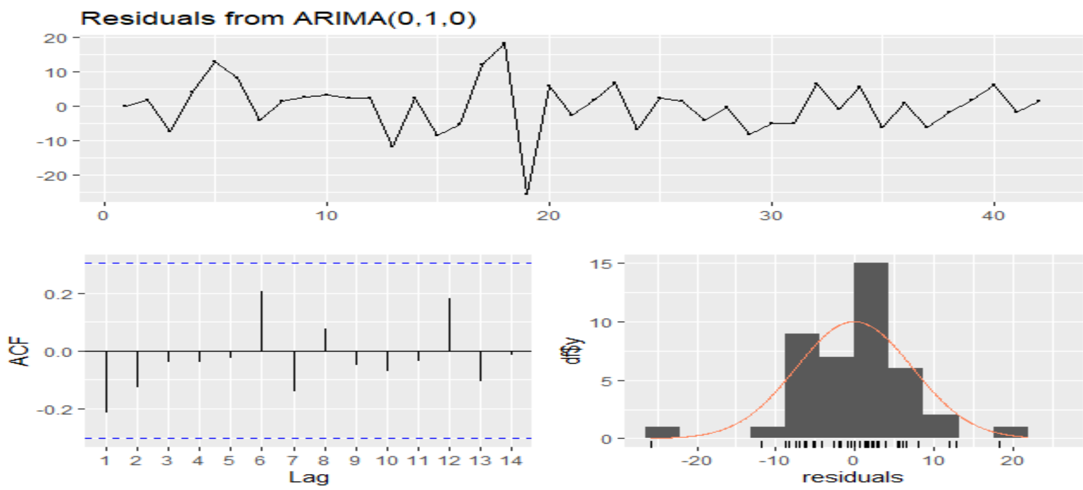

Figure 4.26 Residuals from ARIMA (0,1,0) with non-zero mean

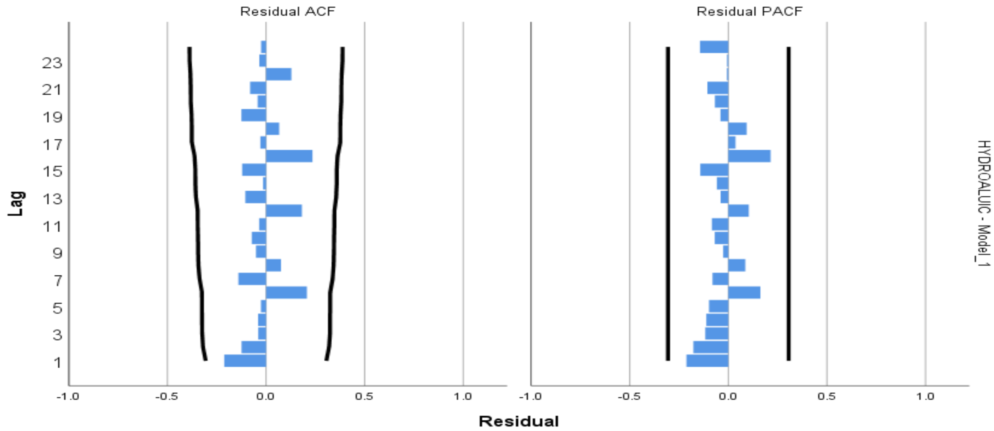

Figure 4.27 Residuals ACF and PACF (0,1,0)

Once the candidate ARIMA (0,1,0) model has been estimated and found, the goodness of fit of the model must be evaluated. As part of the model diagnostic testing procedure, this evaluation includes parameter and residual analysis. Diagnostic testing for the residuals of the ARIMA (0,1,0) model was done using (ACF) and (PACF) plots, as shown in Figure 4.26 All ACF and PACF residual values were statistically significant at the 95% confidence level, according to the findings. This implies that the residuals have random white noise properties, indicating that the model is adequate for the data. The Box-Ljung test, which assesses the existence of autocorrelation in the residuals of a time series model, was used to further verify the model. This test compares the observed residual autocorrelations to the anticipated values under the assumption of no autocorrelation. The Box-Ljung test was performed on the residuals of the ARIMA (0,1,0) model in this case. The test yielded a p-value of 0.6847, which is more than the significance level of 0.05. As a result, there is no significant indication of autocorrelation in the model's residuals. This demonstrates that the ARIMA (0,1,0) model effectively represents the data's autocorrelation pattern.

Table 4.18 presents the actual and predicted values of hydraulic production from 2013 to 2020. It provides a comparison between the observed values, representing the real hydraulic production data, and the forecasted values obtained from a particular model or method.

Table 4.18 The actual and predicted values of hydraulic production during 2013-2020

Date	Actual	Forecast
2013	24.74246	24.0996128
2014	16.13134	24.0380706
2015	25.64949	23.9765283
2016	26.03014	23.9149860
2017	25.13142	23.8534437
2018	30.58037	23.7919015
2019	24.42700	23.7303592
2020	16.13134	23.6688169

Moreover, the ARIMA (0,1,0) model was employed to make predictions for the yearly hydraulic production in Türkiye for the year 2020. The resulting plot, shown in Figure 4.27, displays the predicted values for 2020, which closely align with the actual values. This convergence of anticipated and actual values suggests that the model accurately predicts the behavior and trends observed in the hydraulic production series for that year.

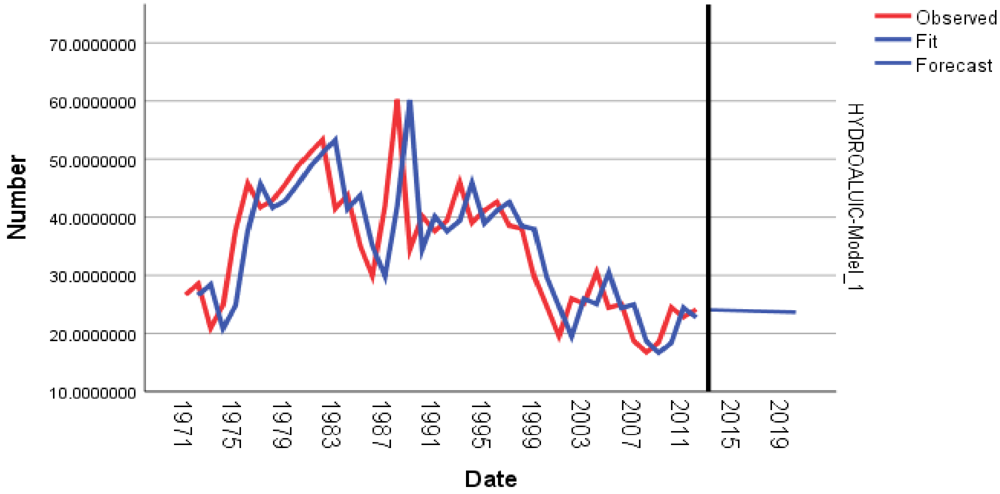

Figure 4.28 Predicted values of hydraulic production in 2020

4.7 Oil Production in Türkiye from 1971 to 2020

Türkiye heavily depends on oil imports to satisfy its energy requirements and fuel its economy. The country imports crude oil, petroleum products, and refined oil from various global sources, including Russia, Iraq, Iran, Saudi Arabia, and Kazakhstan. These imports are refined into petroleum products such as gasoline, diesel, jet fuel, and heating oil at Turkish refineries. Türkiye's oil exports are modest, and it largely sells petroleum products to neighboring nations and worldwide markets (Balat, 2004).

Oil consumption in Türkiye is driven by various sectors, with the transportation sector being the largest consumer. Industries, including manufacturing and construction, also rely on oil for their energy needs. Türkiye's energy strategy focuses on expanding the energy mix and boosting the percentage of renewable energy sources in an effort to lessen reliance on imported oil. The government promotes energy efficiency measures and encourages the use of alternative fuels in transportation (AL-Moders and Kadhim, 2021).

Türkiye has been investing in renewable energy projects like wind and solar in order to reduce its dependency on fossil fuels like oil. Fluctuations in global oil prices impact Türkiye's oil import costs, which in turn affect the country's economy and trade balance. Environmental concerns and the need to address climate change have led to increased efforts to reduce oil dependence and promote cleaner energy sources (Çakmakce et al., 2008).

In summary, Türkiye heavily relies on oil imports, with limited oil exports. The transportation sector and industries are significant consumers of oil. By broadening the energy mix and boosting the use of renewable energy sources, the government hopes to lessen reliance on imported oil. Türkiye's oil production has remained relatively modest throughout the years, with efforts focused on attracting foreign investment and exploring new reserves to enhance domestic production and energy security (Balat and Öz, 2008).

A descriptive Table 4.19 show casing the top ten countries in the world for oil production reveals United states as the leading producer, followed by Saudi Arabia, solidifying their positions as key players in the global oil industry.

Table 4.19 Top ten countries by oil production

Rank	Country	Oil Production (Million Barrels per Day
1	United States	12.2
2	Saudi Arabia	9.5
3	Russia	10.4
4	Canada	4.6
5	China	3.9
6	Iraq	4.7
7	Iran	4.2
8	United Arab	3.9
9	Brazil	3.1
10	Kuwait	2.7
46	Türkiye	69,012 (Thousand)

Source: The World Bank organization

4.8 Application of ARIMA on Oil Production

The ARIMA model was chosen from the collection of classical models for oil from 1971 to 2020 in Türkiye, which consists of three main methods, and was used to make sure ti forecast the future years.

4.8.1 Identification for Oil Production

The descriptive Table 4.20. for oil production from 1971 to 2020 presents key statistics that provide insights into the trends and variability of oil production over the years.

Table 4.20 Descriptive statistics for oil production from 1971 to 2020

Variable	Min	Max	Mean	S.D
Oil production	0.39	51.35	13.41	13.51

These statistics offer an overview of the distribution and central tendencies of oil production during the specified period. The lowest recorded oil production is represented by a minimum value of 0.39, while the highest recorded production is indicated by a maximum value of 51.35. The average oil production over the entire period, represented by the mean value of 13.41, provides insight into the typical level of production during this timeframe. Furthermore, the standard deviation of 13.51 represents the dispersion or variability in the oil production statistics. A larger standard deviation indicates that production values are more variable, whereas a smaller standard deviation indicates that output levels are more stable. When transforming data into a time series, each data point is assigned a particular date or time period, forming a sequence of evenly spaced time points Figure 4.28. In the context of oil production data, each observation is associated with a specific date or time interval. Analyzing this time series allows researchers to detect trends and patterns in oil production over a span of time. By examining the sequential nature of the data, valuable insights can be gained regarding the changes and patterns in oil production over time.

Figure 4.29 Time series plot of yearly oil production in Türkiye (1971-2020)

Stationarity in time series analysis refers to a variable's statistical features staying stable throughout time, as seen in Figure 4.29.

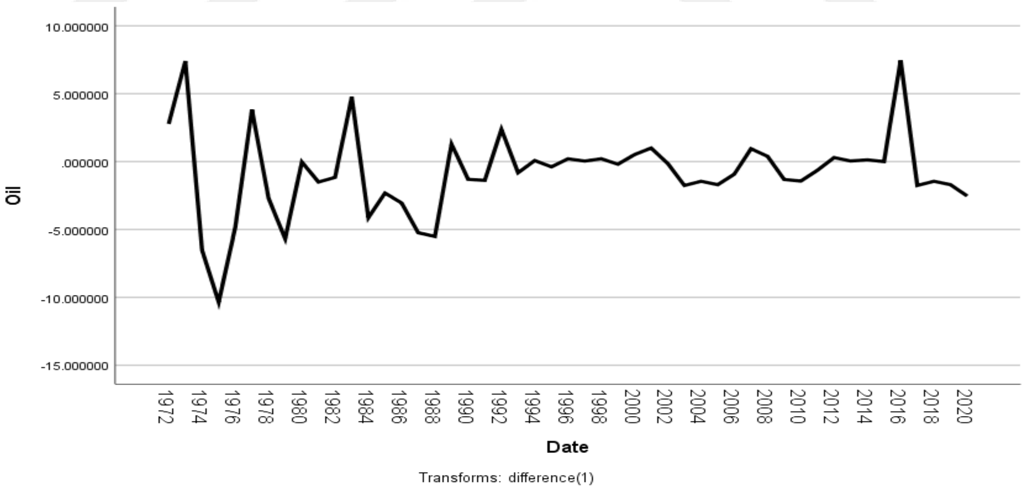

Figure 4.30 Time series plot of yearly oil production after one differencing

To evaluate the first-order differenced time series of oil production data, the ADF test was used. Furthermore, the test resulted in a p-value of 0.03743, which is less than the significance level of 0.05.These findings suggest that the first-order differenced time series of oil production data is stationar.

The autocorrelation function (ACF) plot offers insights into the level of autocorrelation present in the data. When the ACF demonstrates a substantial correlation at specific lags, it indicates the existence of a repeating pattern in the data at those particular lags Figure 4.30.

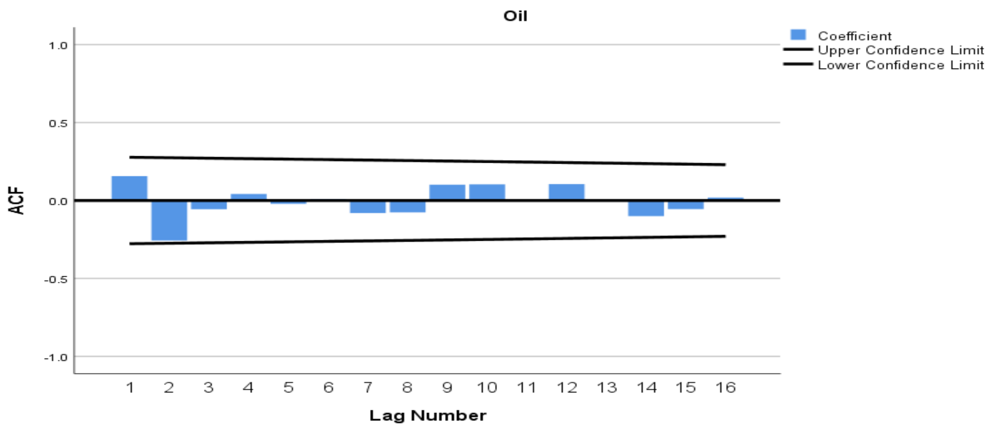

Figure 4.31 Autocorrelation function for oil production time series

The (PACF) is a useful tool for discovering certain lagged values that have a substantial association with a variable of interest, particularly in time series analysis and modeling. It aids in isolating and examining the direct link between the variable and its lagged values, revealing insights into the data's underlying dynamics Figure 4.31.

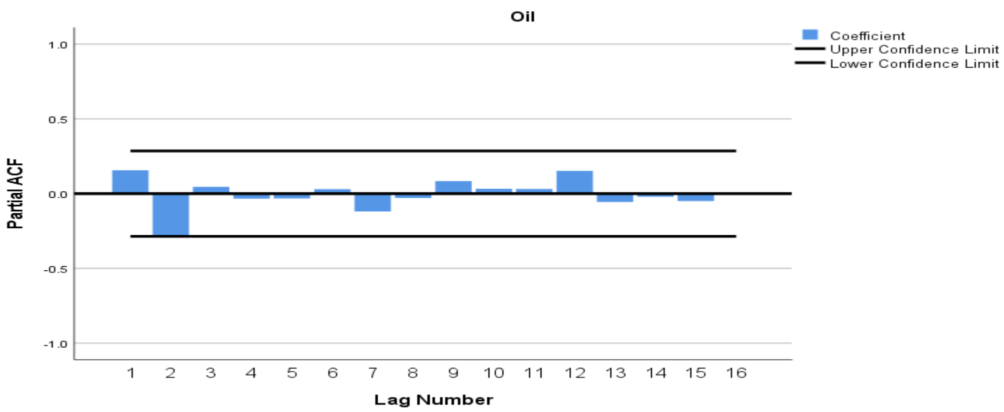

Figure 4.32 Partial autocorrelation function for oil production time series

4.8.2 Selecting Fitting Model

Box and Jenkins developed an interactive method for fitting (ARIMA) models to time series data. This method focuses on assuring the time series' stationarity around its mean and variance. The calculated model, as shown in Table 4.21 is statistically significant, suggesting that it incorporates essential data properties. In this specific situation, the AIC is 253.63, indicating that the model gives an excellent fit to the data. The ARIMA (0,1,2) model may be used to estimate model parameters, as shown in Figures 4.32 and 4.33.

Table 4.21 ARIMA (0,1,2) model's parameters

Variable	Estimate	SE	T-test	P-value
Constant	-0.215	1.036	-0.208	0.836

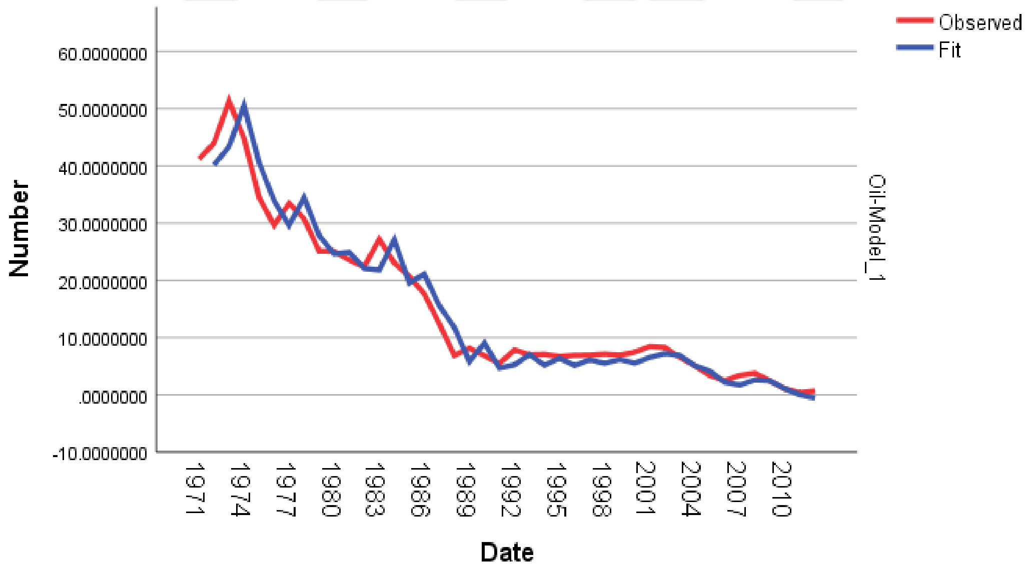

Figure 4.33 Training data for predicted value and actual values of oil production time series by using ARIMA model

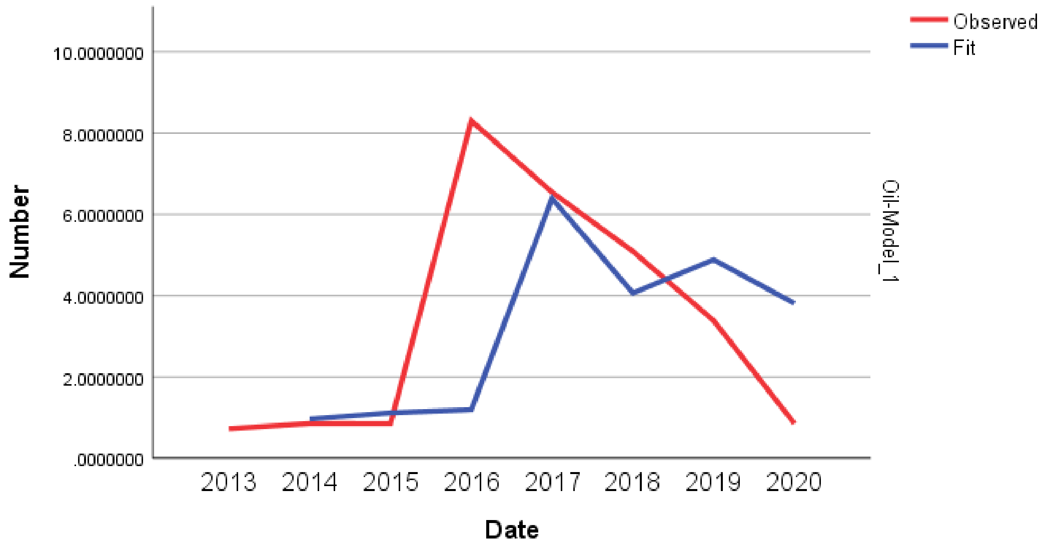

Figure 4.34 Testing data for predicted value and actual values of oil production time series by using ARIMA model

Tables 4.22 and 4.23 provide the metrics on the training and test sets that may be used to assess the correctness of an ARIMA model. The metrics collected from the training set provide information about how effectively the model captures the properties of the data on which it was trained.

Table 4.22 ARIMA (0,1,2) model statistics (training data)

MAE	RMSE	MAPE	R^2
5.437	6.753	11.694	0.567

Table 4.23 ARIMA (0,1,2) model statistics (testing data)

MAE	RMSE	MAPE	R^2
5.910	7.253	12.529	0.500

4.8.3 Model Checking (0,1,2)

Figures 4.34 and 4.35 show the residuals from an ARIMA (0,1,2) model with a non-zero mean.

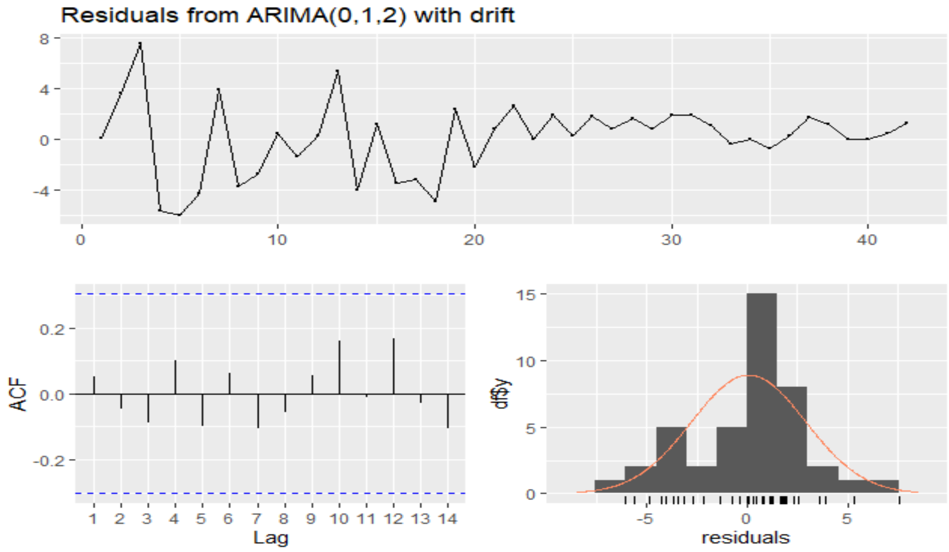

Figure 4.35 Residuals from ARIMA (0,1,2) with non-zero mean

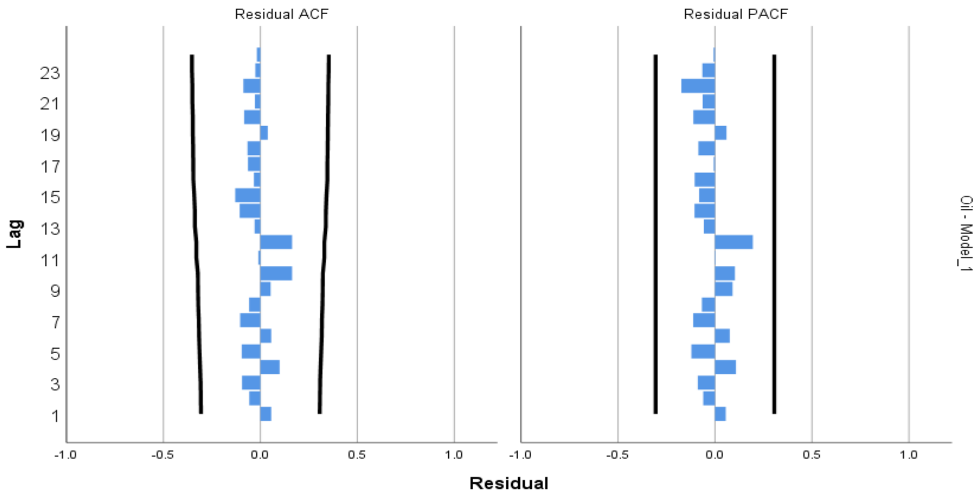

Figure 4.36 Residuals ACF and PACF (0,1,2)

After estimating and choosing the ARIMA (0,1,2) model as a possible option, the goodness of fit of the model had to be evaluated. This step of evaluation included parameter and residual analysis. Diagnostic tests were performed on the residuals of the ARIMA (0,1,2) model using (ACF) and (PACF) plots for residuals, as illustrated in

Figure 4.35. At a 95% confidence level, all of the ACF and PACF residual values were statistically significant. This implies that the residuals displayed random white noise behavior, indicating that the model suited the data well. The Box-Ljung test is a statistical test used to determine the existence of autocorrelation in time series residuals. This test compares the observed residual autocorrelations to the anticipated values under the assumption of no autocorrelation. The Box-Ljung test was conducted to the residuals of the ARIMA (0,1,2) model in the presented output, generating a p-value of 0.8802 because this p-value surpasses the significance level of 0.05, there is no significant indication of autocorrelation in the model's residuals. As a result of the diagnostic tests, we may infer that the ARIMA (0,1,2) model properly reflects the autocorrelation structure present in the data.

Table 4.24 presents the actual and predicted values of oil production from 2013 to 2020. It provides a comparison between the observed values, representing the real oil production data, and the forecasted values obtained from a particular model or method.

Table 4.24 The actual and predicted values of oil production in from 2013 to 2020

Date	Actual	Forecast
2013	0.724118	-0.247242
2014	0.851712	-1.676475
2015	0.849558	-2.727097
2016	8.302936	-3.777720
2017	6.542135	-4.828343
2018	5.089649	-5.878966
2019	3.385487	-6.929588
2020	0.851712	-7.980211

Furthermore, the ARIMA (0,1,2) model was used to forecast Türkiye's annual oil output in 2020. The anticipated values for 2020, as shown in Figure 4.36, behave similarly to the actual values. This suggests that over the course of the year, the projected values increasingly approach and align with the actual values, suggesting that the ARIMA (0,1,2) model captures the underlying patterns and trends in the data satisfactorily.

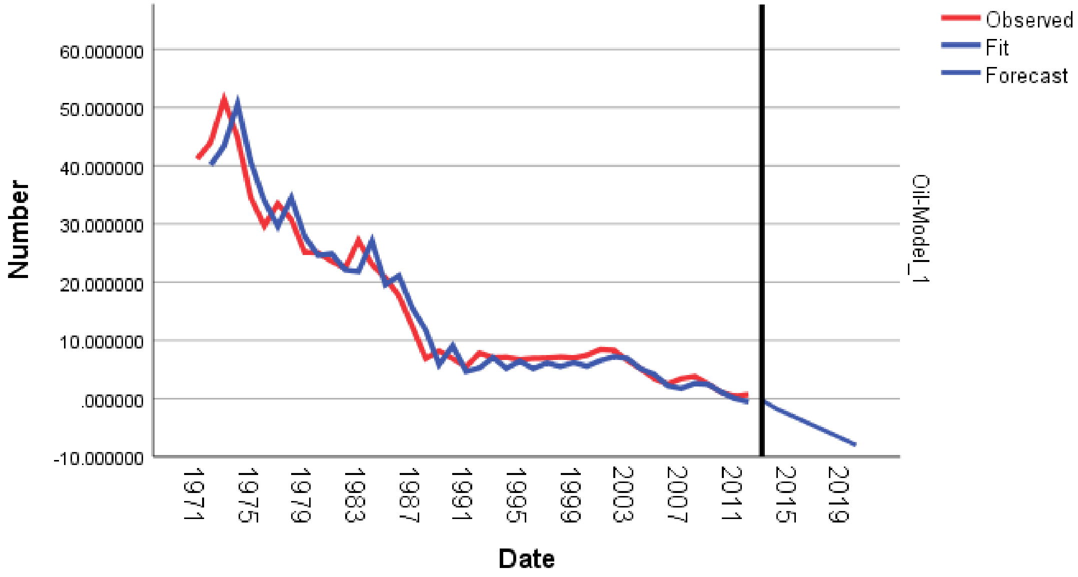

Figure 4.37 Predicted values of oil production in 2020

4.9 GDP in Türkiye from 1971 to 2020

Türkiye possesses a diverse and emerging market economy with a dynamic private sector. Over the years, the country has experienced varying levels of GDP growth, displaying an overall positive trend of economic expansion with occasional periods of high growth and slowdowns. Several factors influence Türkiye's GDP growth rate, including domestic and global economic conditions, government policies, and external shocks (Lise and Van Montfort, 2007).

Türkiye's nominal GDP, which represents the entire value of goods and services produced inside its boundaries, measures the size of the economy. The economy is diverse, with important industries contributing to GDP. The services sector, encompassing finance, tourism, and telecommunications, plays a significant role, while manufacturing, particularly in automotive, textiles, and electronics industries, is also crucial. Although agriculture represents a smaller share of GDP, it remains important for rural areas and employment (Sözen and Arcaklioglu, 2007).

Türkiye is actively involved in international trade, with both exports and imports contributing to its GDP. The country exports various goods, including automobiles, machinery, textiles, chemicals, and agricultural products, with the European Union, the Middle East, and North Africa being key export markets. Imports mainly consist of machinery, energy products, chemicals, and raw materials to support domestic production and consumption (Artar et al., 2016).

However, Türkiye faces economic challenges, including inflation, high unemployment rates, and a large informal economy. External factors such as global commodity price fluctuations and exchange rate fluctuations can also impact the Turkish economy. Additionally, political and geopolitical issues, along with domestic policy decisions, influence economic stability and investor confidence (Sözen and Arcaklioglu, 2007).

From 1971 to 2020, Türkiye's GDP underwent significant changes and growth. During the 1970s, the country experienced moderate GDP growth, emphasizing industrialization and infrastructure projects. The 1980s witnessed periods of growth and economic challenges, with structural reforms and liberalization policies implemented to attract foreign investment and promote export-oriented industries. The 1990s saw fluctuations in economic growth, along with stabilization programs and external factors impacting the economy (Bağci and Diğrak, 1996).

In the 2000s, Türkiye's economy displayed resilience and significant growth, driven by structural reforms, investments in infrastructure, manufacturing, and services sectors. However, the global financial crisis in 2008 had some adverse effects. From 2010 to 2020, Türkiye's GDP continued to grow, albeit at a slower pace, with a focus on increasing domestic consumption, attracting foreign investment, and promoting sectors such as construction and tourism. Political, geopolitical challenges, as well as domestic economic imbalances, presented obstacles to sustained economic growth (Yalta, 2011).

Overall, Türkiye's GDP has witnessed substantial growth and transformation, transitioning from an agriculture-based economy to a more diversified one, with a focus on industry and services. External factors, political developments, and domestic reforms have played significant roles in shaping Türkiye's economic performance. Despite being one of the largest economies in the region, Türkiye faces ongoing economic challenges and structural issues that affect its growth trajectory.

A descriptive Table 4.25 showcasing the top ten countries in the world for GDP production reveals United States as the leading producer, followed by China, solidifying their positions as key players in the global coal industry.

Table 4.25 Top ten countries by GDP

Rank	Country	Oil Production (Million Barrels per Day
1	United States	21.43
2	China	15.42
3	Japan	5.08
4	Germany	3.85
5	United Kingdom	2.68
6	India	2.65
7	France	2.55
8	Italy	1.79
9	Canada	1.64
10	South Korea	1.63
17	Türkiye	851,549 (Billion)

Source: The World Bank organization.

4.10 Application of ARIMA for GDP

The ARIMA model was chosen from the collection of classical models for GDP from 1971 to 2020 in Türkiye, which consists of three main methods, and was used to make sure ti forecast the future years.

4.10.1 Identification for GDP

The descriptive Table 4.26 for GDP production from 1971 to 2020 presents key statistics that provide insights into the trends and variability of GDP production over the years.

Table 4.26 Descriptive statistics for GDP production from 1971 to 2020

Variable	Min	Max	Mean	S.D
GDP	-0.215	1.036	-0.208	0.836

The dataset containing GDP production data for Türkiye from 1971 to 2020 is a time series consisting of yearly observations on the production of coal in tons. This dataset comprises 50 observations and is recorded annually. It is commonly utilized to examine the trends and patterns in coal production in Türkiye over time. By analyzing this data, researchers can gain insights into the factors that influence coal production, including economic and environmental conditions. Moreover, the dataset allows for the forecasting of future trends in coal production in Türkiye. Throughout the observed period, the minimum recorded coal production in Türkiye was approximately 22.86 million tons in 1971. Conversely, the maximum coal production occurred in 2019, reaching approximately 48.96 million tons. On average, Türkiye's coal production amounted to around 29.62 million tons per year over this period. The standard deviation of approximately 5.48 million tons per year indicates a significant variation in coal production, highlighting the wide range of values within the dataset.

Transforming data into a time series entails organizing it as a sequence of regularly spaced data points that correspond to specific dates or time periods Figure 4.37. In the case of coal production data, each observation is assigned a particular date or time period. This time series representation allows for the analysis of trends and patterns in coal production over a span of time. By examining the sequential nature of the data, researchers can gain insights into the changes, fluctuations, and recurring patterns in coal production throughout the observed period.

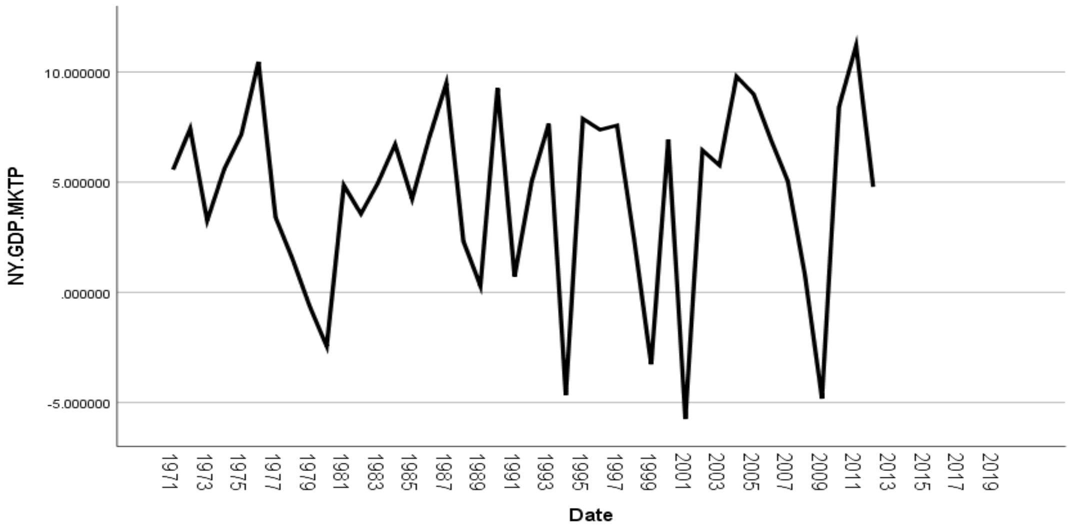

Figure 4.38 Time series plot of yearly GDP in Türkiye (1971-2020)

Stationarity in time series analysis refers to a variable's statistical features staying stable throughout time, as seen in Figure 4.38.

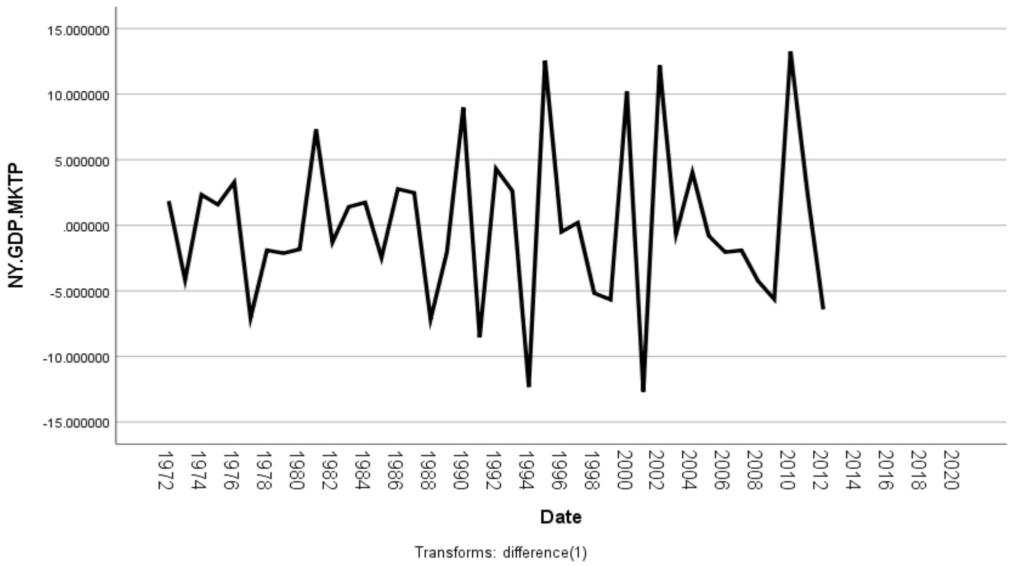

Figure 4.39 Time series plot of yearly GDP from 1971 to 2020 in Türkiye after differencing

The ADF test is a statistical test used to determine a time series' stationarity. The ADF test is being utilized in this case to investigate the first-order differenced time series of coal production data. The test results show that the computed Dickey-Fuller statistic is -4.1147, which is less than the crucial value at a 5% level of significance. Furthermore, the test's p-value is 0.01226, which is less than the significance limit of 0.05.

Figure 4.39 shows that as the lag grows, the (ACF) values rapidly decrease and finally approach zero. Furthermore, all ACF values are inside the confidence interval ranges. This suggests that there is no significant autocorrelation in the data.

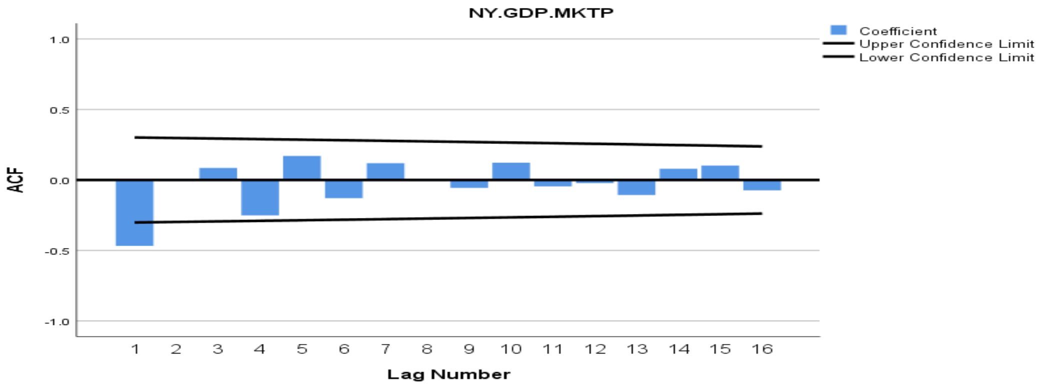

Figure 4.40 Autocorrelation function for GDP production time series

The (PACF) is a valuable tool for discovering strong correlations between the present value of GDP output and lagged values, as shown in Figure 4.40. A high and statistically significant PACF value at a certain lag indicates a strong relationship between the present value and the value at that lag.

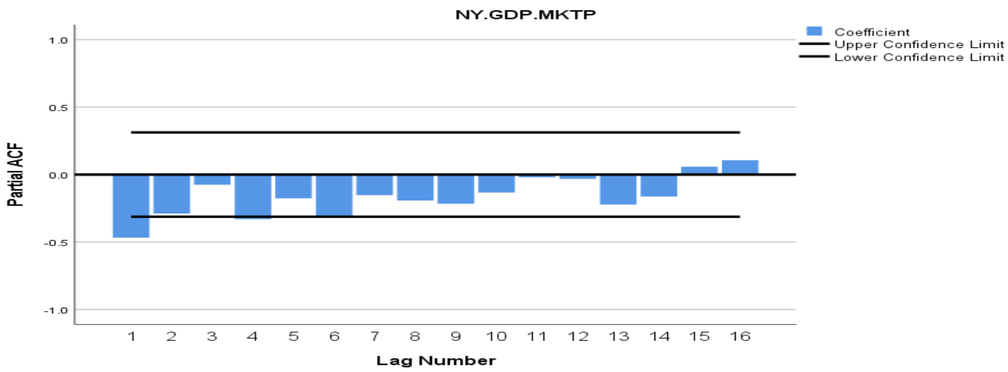

Figure 4.41 Partial autocorrelation function for GDP production time series

4.10.2 Selecting Fitting Model

Table 4.27 presents the estimated parameters of the ARIMA (0,0,0) model.

Table 4.27 ARIMA (0,0,0) model's parameters

Variable	Estimate	SE	T-test	P-value
AR	0.019	0.955	0.20	0.01

Box and Jenkins developed an interactive approach for fitting autoregressive moving average models to time series data. This approach focuses on ensuring that the time series remains stationary around its mean and variance. The study in Tables 4.28 and 4.3 shows that the calculated model is substantial, and the parameters have statistical significance. The Akaike Information Criterion (AIC) score is 285.28, indicating that the ARIMA (0,0,0) model is an excellent match for the data. As a result, it may be used to estimate the model's parameters. As a result, by examining Figures 4.41 and 4.42, it may be used to efficiently estimate the parameters of the model.

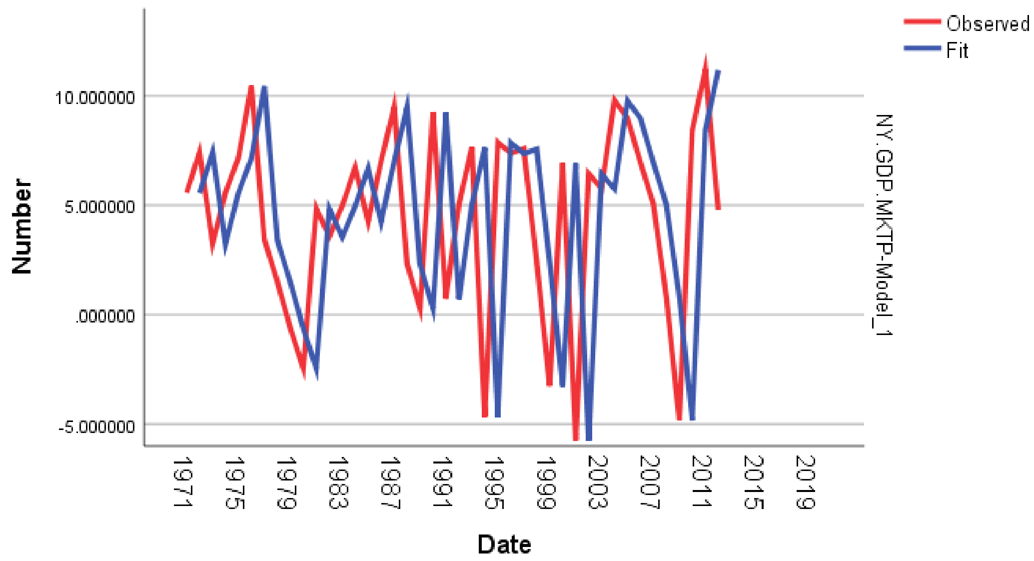

Figure 4.42 Training data for predicted value and actual values of GDP production time series by using ARIMA model

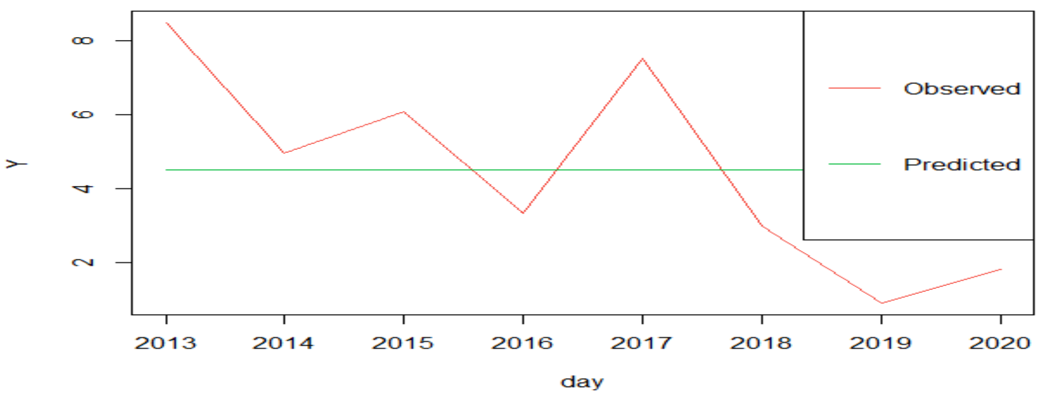

Figure 4.43 Testing data for predicted value and actual values of GDP production time series by using ARIMA model

Tables 4.28 and 4.29 provide the metrics on the training and test sets that may be used to assess the correctness of an ARIMA model.

Table 4.28 ARIMA (0,0,0) model statistics (training data)

MAE	RMSE	MAPE	R^2
1.254	1.542	2.982	0.823

Table 4.29 ARIMA (0,0,0) model statistics (testing data)

MAE	RMSE	MAPE	R^2
5.910	7.253	12.529	0.500

4.10.3 Model Checking (0,0,0)

Figure 4.43 and Figure 4.44 exhibit the residuals derived from an ARIMA (0,0,0) model with a non-zero mean.

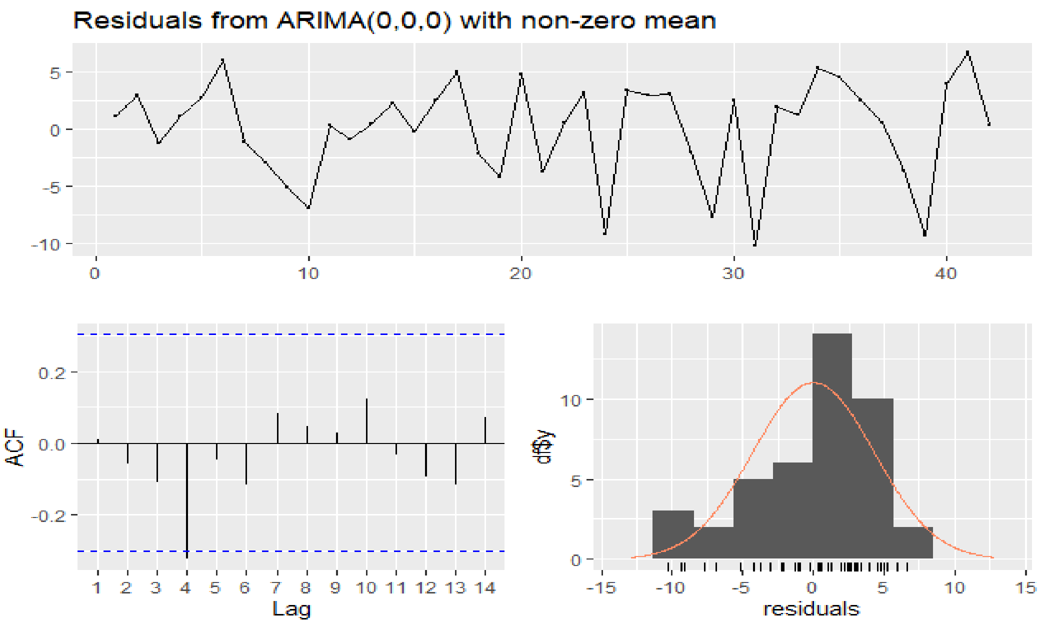

Figure 4.44 Residuals from ARIMA (0,0,0) with non-zero mean

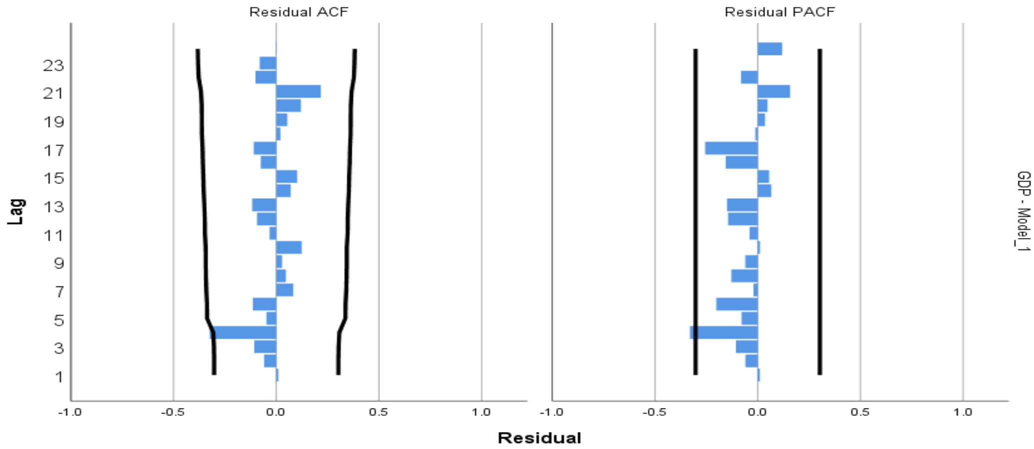

Figure 4.45 Residuals ACF and PACF (0,0,0)

The fit of the ARIMA (0,0,0) model to the data was tested once it was identified and estimated. This stage entailed evaluating the model's parameters as well as its residuals. Diagnostic testing was performed using (ACF) and (PACF) plots to analyze the residuals of the ARIMA (0,0,0) model, as shown in Figure 4.44. At the 95% confidence level, all ACF and PACF residual values were statistically significant. This means that the residuals have a random white noise pattern, indicating that the model is appropriate for the data.

The ARIMA (0,0,0) residuals were subjected to the Box-Ljung test, a statistical test designed to detect autocorrelation in time series residuals. The computed p-value of 0.8638 is above than the 0.05 threshold of significance. This implies that there is no indication of autocorrelation in the model's residuals. As a result, we may infer that the ARIMA (0,0,0) model adequately describes the data's autocorrelation pattern.

Table 4.30. presents the actual and predicted values of GDP production from 2013 to 2020. It provides a comparison between the observed values, representing the real GDP production data, and the forecasted values obtained from a particular model or method.

Table 4.30 The actual and predicted values of GDP from 2013 to 2020

Date	Actual	Forecast
2013	8.485817	4.769512
2014	4.939715	4.750532
2015	6.084486	4.731551
2016	3.323084	4.712571
2017	7.501997	4.693590
2018	2.979885	4.674610
2019	0.889585	4.655629
2020	1.793551	4.636648

Furthermore, the ARIMA (0,0,0) model was used to forecast Türkiye's annual GDP for 2020. Figure 4.45 depicts the findings, which clearly indicate that the anticipated values for 2020 roughly agree with the actual values. This convergence implies that the model's projections for that year are in strong agreement with the actual data, implying that the ARIMA (0,0,0) model successfully represents Türkiye's underlying GDP patterns and trends.

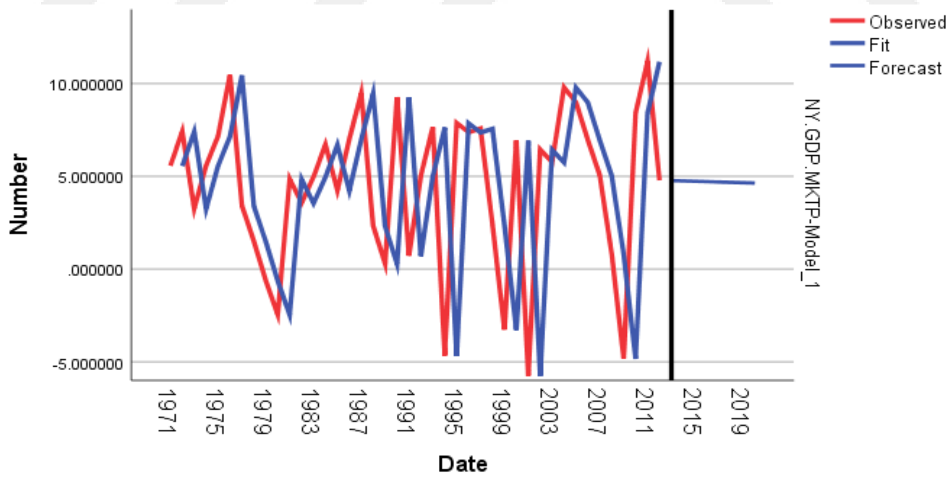

Figure 4.46 Predicted values of GDP production in 2020

4.11 Application of BSTS on Coal Production

A BSTS (Bayesian Structural Time Series) model's goal is to separate time series into discrete aspects such as trend, seasonality, regression, and error. The trend component accounts for the long-term patterns and changes observed in the time series, while the seasonality component captures the recurring fluctuations occurring at regular intervals. The regression component allows for modeling the connection between the time series and additional predictor variables, while the error component represents the random variations present in the data (Figure 4.46).

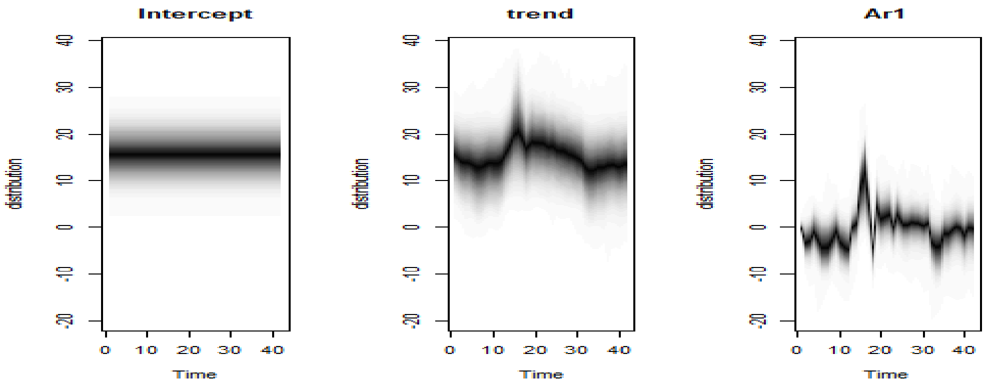

Figure 4.47 Components coal of BSTS Model

4.11.1 Select Fitting Model

By looking at the Figure 4.47 and Figure 4.48 These values suggest that the model is well-suited for capturing the patterns in the data. As a result, it may be used to successfully estimate the model's parameters.

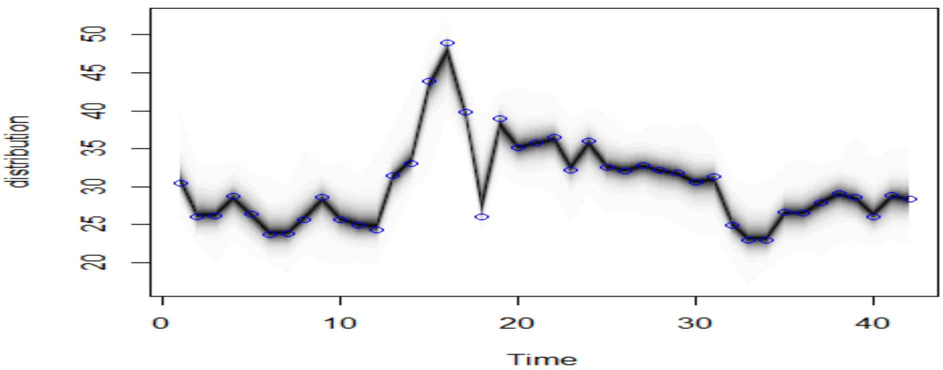

Figure 4.48 Training data for predicted value and actual values of coal production time series by using BSTS

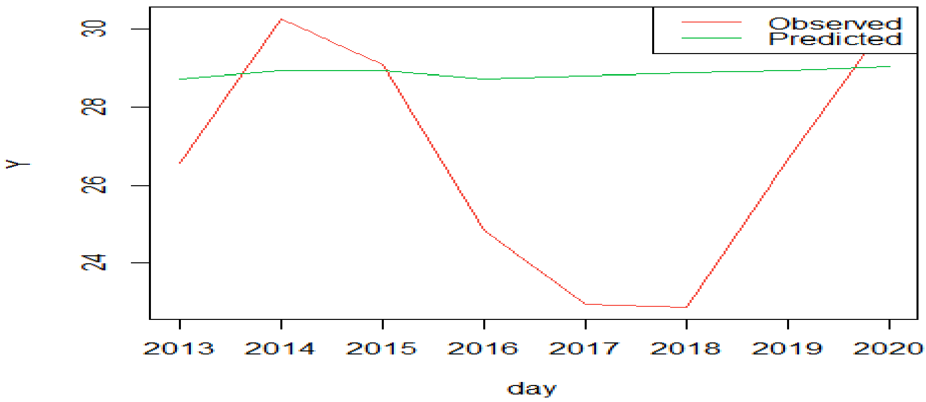

Figure 4.49 Testing data for predicted value and actual values of coal production time series by using BSTS

The residual standard deviation (1.396218) of the model represents the variability between the predicted and observed values. A higher value indicates greater variability between the predictions and the actual data. The standard deviation of the predicted values (6.368165) indicates the uncertainty associated with the model's forecasts. A larger value suggests more uncertainty in predicting future values. The coefficient of determination (0.9398547) quantifies the amount of variation in the observed data that the model explains. A greater score suggests a better fit, suggesting that the model accounts for a considerable percentage of the variability in the data. The relative goodness-of-fit (-1.018128) compares the data's actual log-likelihood to the model's anticipated log-likelihood. A negative number indicates that the BSTS model fits better than the null model. Overall, these output metrics show that the BSTS model fits the data well and accurately depicts the time series' underlying dependency structure. The model parameters are estimated using Bayesian inference and the MCMC method throughout the fitting procedure. More iterations (in this example, 1000) result in more accurate parameter estimates. The output acts as a progress report, displaying iteration numbers and timestamps to illustrate the algorithm's convergence and successful model estimate.

4.11.2 Errors in Sample

By looking at the Figure 4.49 in-sample errors will further ensure that the model fits the data.

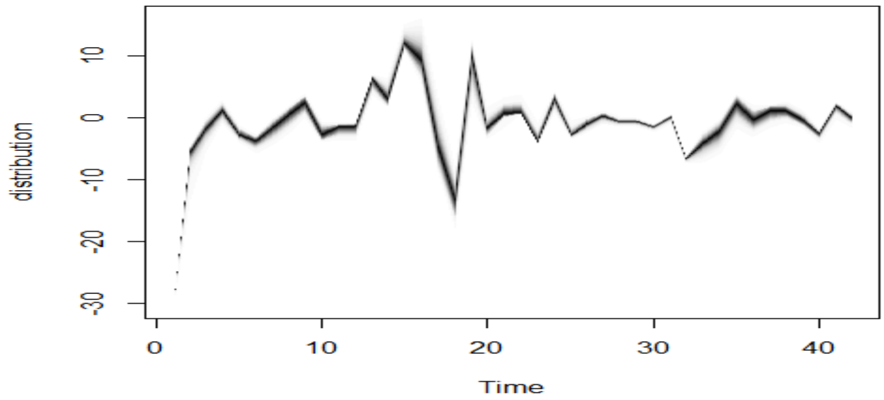

Figure 4.50 Plot dynamic distribution errors in sample

The prediction errors for the BSTS model are computed after a burn-in period of 10 iterations. These errors offer valuable information about the model's performance at different time points. The burn-in period, which is often employed in Bayesian modeling, includes discarding the MCMC algorithm's early iterations. This helps the algorithm to more effectively search the parameter space and converge to the genuine posterior distribution. It is critical to include the uncertainty associated with the estimated model parameters when examining prediction errors during the burn-in period. The burn-in duration is chosen to achieve a compromise between computing efficiency and accurate parameter estimation.

To display the distribution of in-sample prediction mistakes, the dynamic distribution plot is used. This graphic gives insights into the patterns and features of the mistakes, allowing systematic departures from anticipated behavior to be identified. By studying the prediction errors and their distribution, it is feasible to evaluate the BSTS model's performance and accuracy in capturing the underlying patterns and variability in the training data. The prediction errors are represented by a 41580-length numeric vector that corresponds to the in-sample disparities between the observed and predicted values for the training data. A statistical overview of these in-sample prediction errors, including variables such as mean, median, minimum, maximum, and quartiles, is also presented. This summary serves in evaluating the BSTS model's performance in fitting the training data.

Tables 4.31 and 4.32 provide the metrics on the training and test sets that may be used to assess the correctness of an ARIMA model.

Table 4.31 BSTS model evaluation for coal production (training data)

MAE	RMSE	MAPE	R^2
0.953	1.567	2.654	0.926

Table 4.32 BSTS model evaluation for coal production (testing data)

MAE	RMSE	MAPE	R^2
1.041	1.651	3.170	0.907

4.11.3 Model Checking

Figure 4.50 exhibit the residuals derived from an BSTS model. Once the BSTS model was identified and estimated, it was important to assess how well the model fit the data. This critical step in the model diagnostic process involved analyzing both the model's parameters and its residuals. The residuals of the BSTS model were examined using ACF and PACF plots, as shown in Figure 4.51, and all ACF and PACF values of the residuals were statistically significant at the 95% confidence level. This implies that the residuals have random white noise properties, confirming that the model is appropriate for the given data.

The Box-Ljung test, a statistical test designed to assess the presence of autocorrelation in time series residuals, was performed on the BSTS model residuals using the given output. The obtained p-value of 0.9585 was more than the 0.05 criterion of significance. This means that there isn't enough data to justify the presence of autocorrelation in the model's residuals. As a result, it is possible to infer that the model adequately describes the autocorrelation structure in the data.

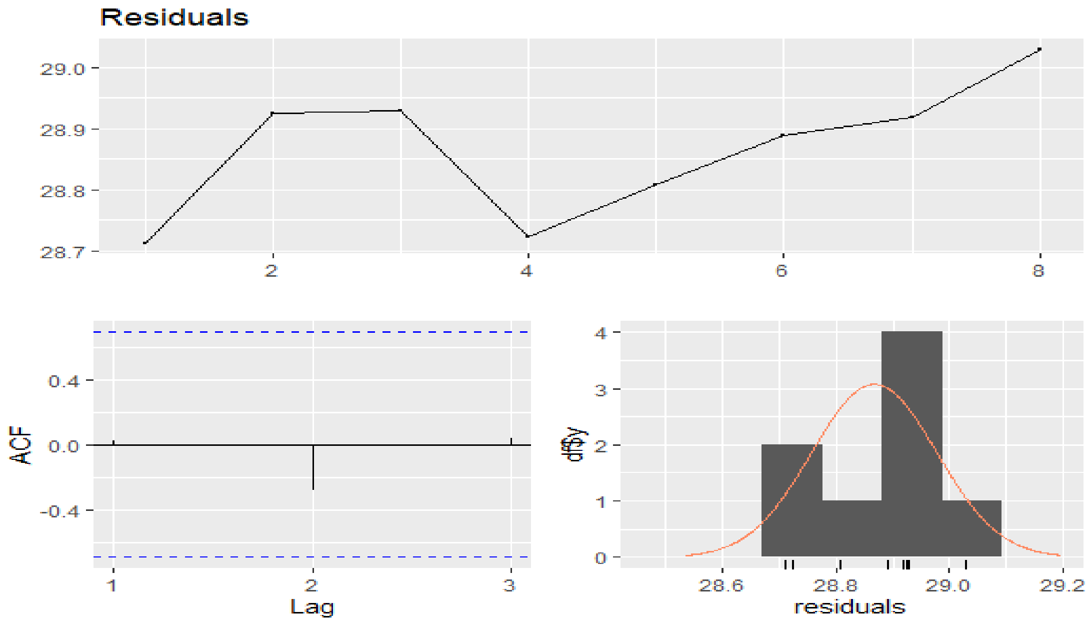

Figure 4.51 Residuals from BSTS model in sample

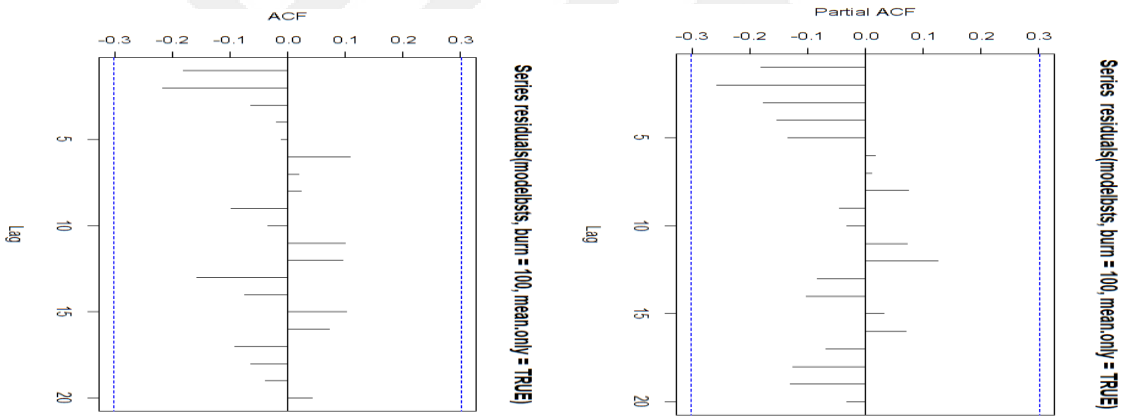

Figure 4.52 Residuals ACF and PACF BSTS model for coal production

According to the Table 4.33 in 2013, the actual coal production was 26.56046, while the forecasted value was 28.71166. In 2020, the actual coal production was 30.26754, and the forecasted value was 29.03021. Similarly can compare the actual and forecasted values for the remaining years.

Table 4.33 The actual and predicted values of coal production in from 2013 to 2020

Date	Actual	Forecast
2013	26.56046	2871166
2014	30.26754	28.92535
2015	29.09509	28.92865
2016	24.84467	28.72390
2017	22.94193	28.80888
2018	22.85896	28.88975
2019	26.66897	28.91793
2020	30.26754	29.03021

Furthermore, the BSTS model was used to forecast Türkiye's annual coal output for the year 2020. The figure, as shown in Figure 4.52, demonstrates that the anticipated values for 2020 roughly coincide with the actual values, suggesting convergence between the expected and observed series.

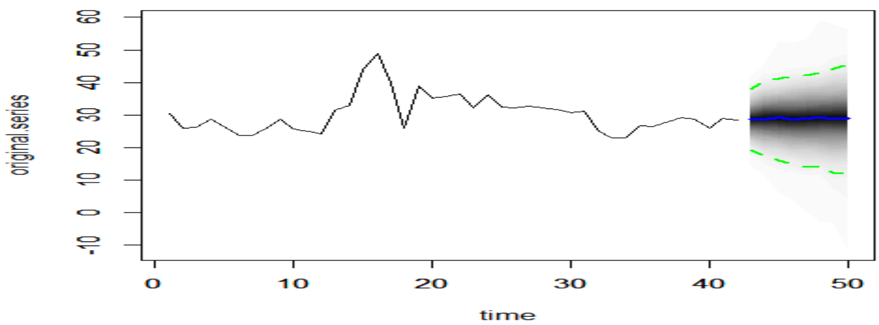

Figure 4.53 Predicted values of coal production in 2020

4.12 Application of BSTS on Gas Production

A BSTS model's main purpose is to breakdown a time series into multiple components such as trend, seasonality, regression, and error. The trend component records the time series' long-term changes, whereas the regression component represents the time series' connection with additional predictor variables. Finally, as shown in Figure 4.53, the error component compensates for the random fluctuations or variances in the data.

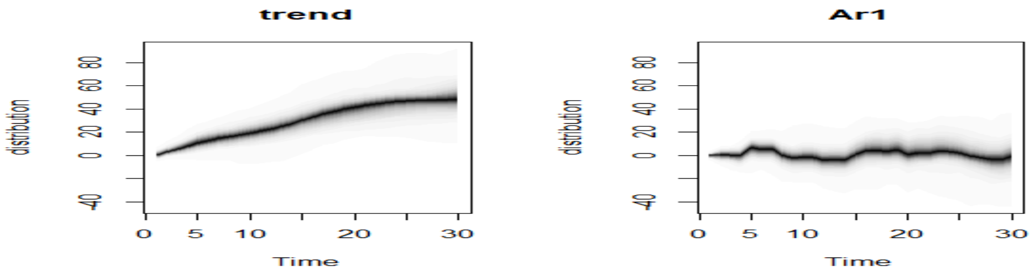

Figure 4.54 Components gas of BSTS model

4.12.1 Select Fitting Model

In accordance with Figures 4.54 and 4.55 these results indicate that the model is well-suited to collecting data patterns. As a result, it may be used to successfully estimate the model's parameters.

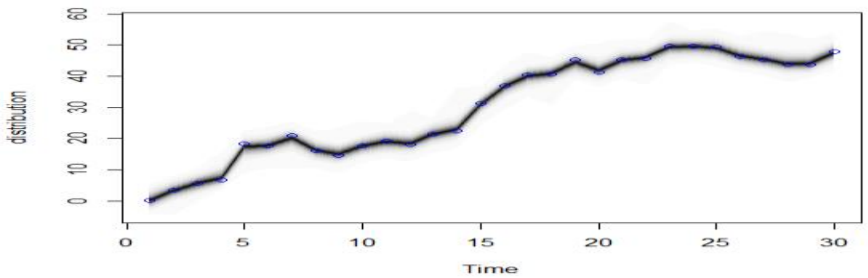

Figure 4.55   Training data for predicted value and actual values of gas production time series by using BSTS

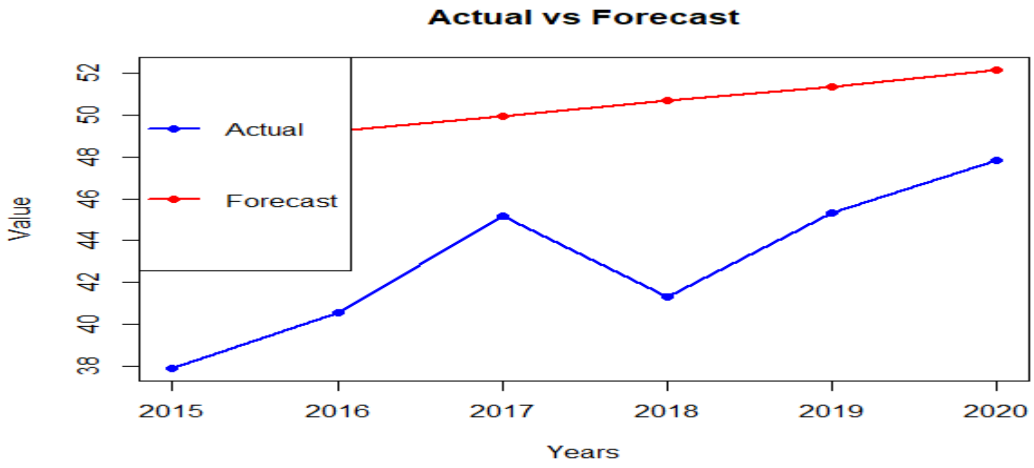

Figure 4.56 Testing data for predicted value and actual values of gas production time series by using BSTS

The performance and fit of the BSTS model are evaluated through various output metrics. The standard deviation of the model residuals (1.396218) measures the variability between the predicted and observed values, where a higher value suggests greater variability. Similarly, the standard deviation of the predicted values (6.368165) indicates the uncertainty associated with the model's forecasts, with a larger value implying higher uncertainty in predicting future values. The coefficient of determination (0.9398547) measures the amount of variation in the observed data that the model can explain, with a larger number suggesting a better fit and the capacity to capture a significant portion of the data's variability. The relative goodness-of-fit (-1.018128) compares the data's actual log-likelihood to the model's anticipated log-likelihood. A negative number indicates that the BSTS model performs better than the null model. Overall, these output metrics show that the BSTS model is well-suited to the data, adequately representing the time series' underlying dependency structure. The fitting process involves utilizing Bayesian inference and the MCMC algorithm to estimate model parameters. Running a larger number of iterations (in this case, 1000) generally results in more accurate parameter estimates. The provided output represents a progress report, with iteration numbers and timestamps indicating the convergence of the algorithm and the accurate estimation of the model.

4.12.2 Errors in Sample

In-sample errors are shown in Figure 4.56 to check that the model matches the data. The BSTS model computes prediction errors over a 10-iteration burn-in time. These mistakes provide useful information about the model's performance at various time intervals. The burn-in period is a technique employed in Bayesian modeling to discard initial iterations of the MCMC algorithm. This method enables the algorithm to efficiently search the parameter space and converge on the real posterior distribution. It is critical to include the uncertainty associated with the estimated model parameters when examining prediction errors during the burn-in period. The selection of an appropriate burn-in period must strike a balance between computational efficiency and the accuracy of parameter estimates.

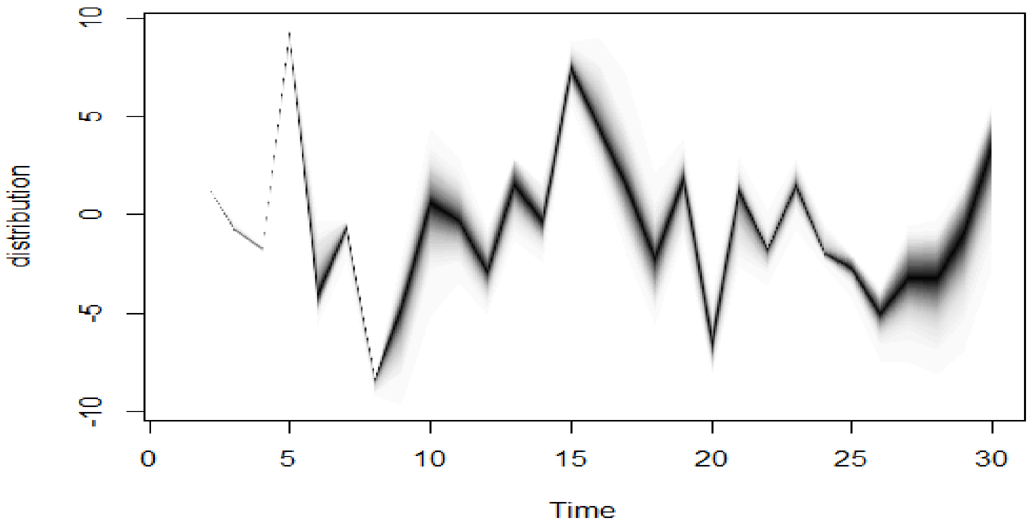

Figure 4.57 Plot dynamic distribution errors in sample

The accuracy of an ARIMA model can be evaluated by looking at Table 4.34 and Table 4.35 these metrics on the training set and the test set.

Table 4.34 BSTS model evaluation for gas production (training data)

MAE	RMSE	MAPE	R^2
0.028	0.038	0.159	0.992

Table 4.35 BSTS model evaluation for gas production (testing data)

MAE	RMSE	MAPE	R^2
0.037	0.050	0.206	0.990

4.12.3 Model Checking

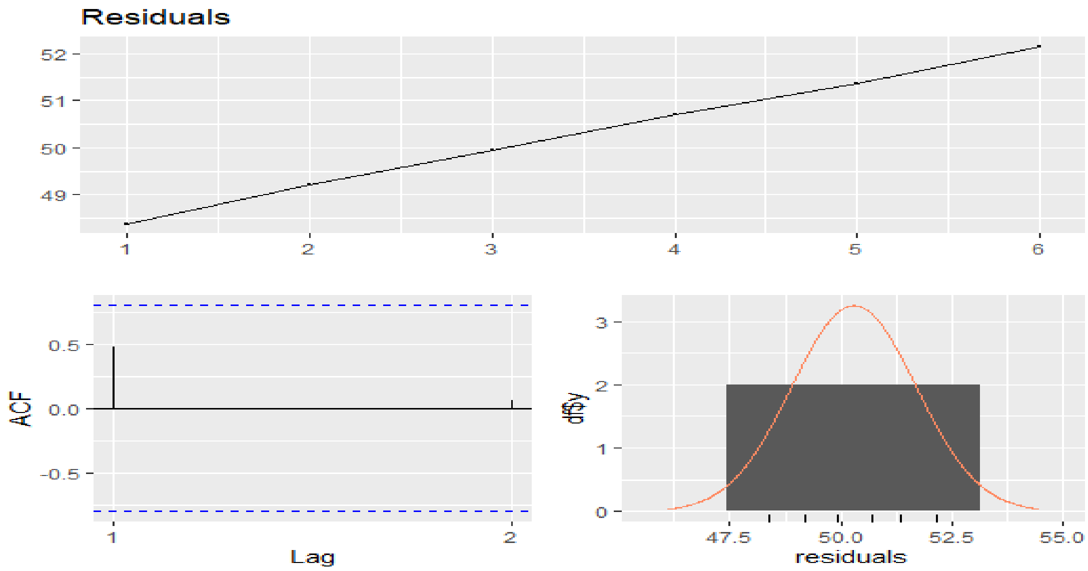

Figure 4.58 Residuals from BSTS model in gas production

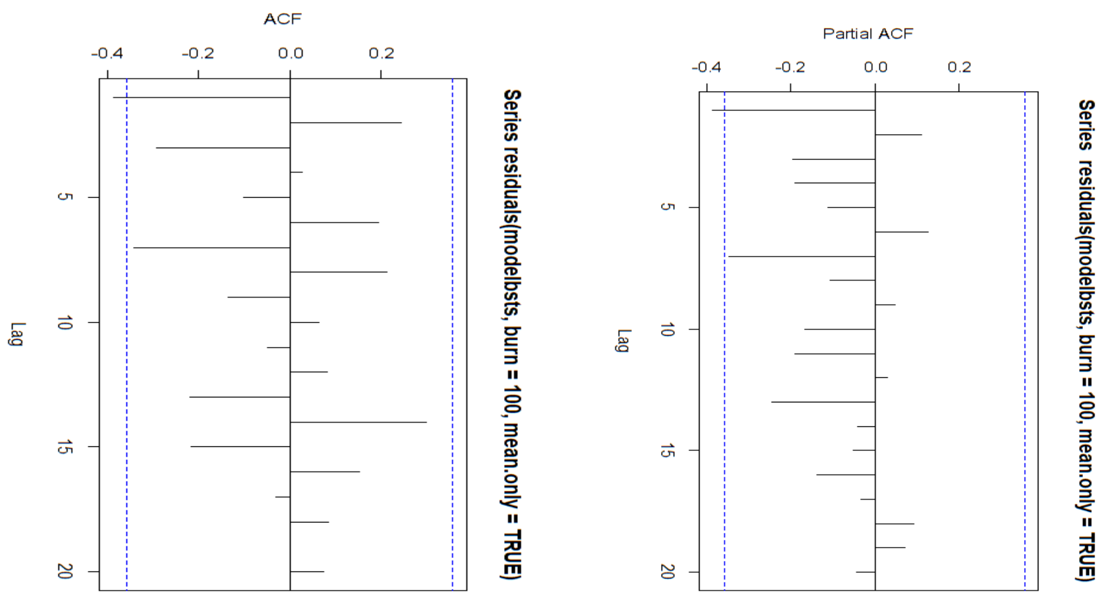

Figure 4.59 Residuals ACF and PACF BSTS model for Gas production

Once the BSTS model was determined and estimated, it became crucial to assess its fit to the data. This essential step in the model diagnostic process involved conducting both parameter and residual analysis. By examining the residuals of the BSTS model using ACF and PACF plots, as shown in Figure 4.58, it was observed that all ACF and PACF residuals values displayed statistical significance at the 95% confidence level. This discovery suggests that the residuals have the properties of random white noise, hence proving the model's applicability for the provided data. The Box-Ljung test was used with the given output to further examine the presence of autocorrelation in the BSTS model residuals. The calculated p-value of 0.9585 is more than the significance threshold of 0.05. This shows that there is insufficient data to justify the presence of autocorrelation in the model's residuals. As a result, it is possible to infer that the model adequately describes the autocorrelation structure in the data.

Table 4.36 presents the actual and predicted values of coal production from 2015 to 2020. It provides a comparison between the observed values, representing the real coal production data, and the forecasted values obtained from a particular model or method.

Table 4.36 The actual and predicted values of gas production in from 2015 to 2020

Date	Actual	Forecast
2015	29.09509	48.36441
2016	24.84467	49.21636
2017	22.94193	49.94214
2018	22.85896	50.70409
2019	26.66897	51.35473
2020	30.26754	52.15941

Furthermore, the BSTS model was used to anticipate Türkiye's annual gas output in 2020. The figure clearly illustrates that the anticipated values for 2020 coincide closely with the actual values, as seen in Figure 4.59 This convergence of anticipated and actual values indicates that the BSTS model accurately captures the underlying patterns and trends in the data, allowing for accurate projections of gas output in 2020.

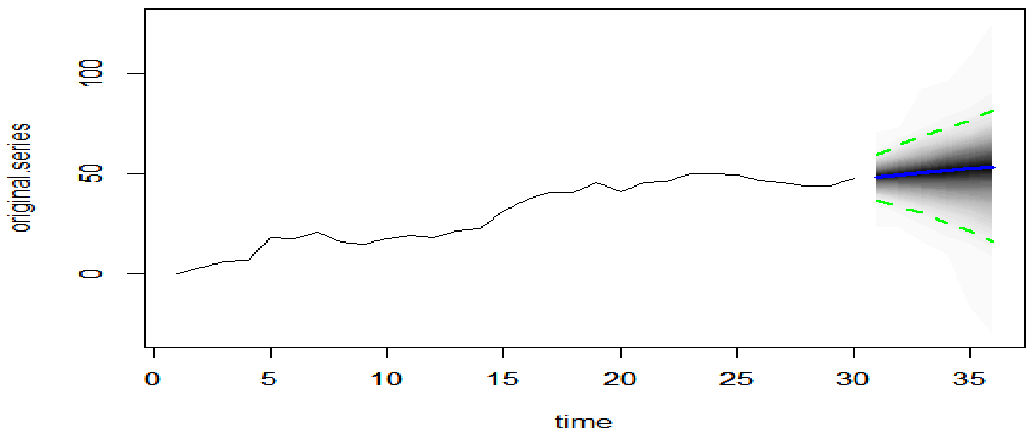

Figure 4.60 Predicted values of gas production in 2020

4.13 Application of BSTS on Hydraulic Production

A BSTS model's goal is to breakdown a time series into several components such as trend, seasonality, regression, and error. The trend component indicates long-term changes in the time series, whereas the seasonality component captures repeating patterns or variations that occur at regular periods. The regression component is used to model the connection between the time series and extra predictor variables, revealing how external influences impact the series. Finally, as illustrated in Figure 4.60, the error component allows for random changes or uncertainties that cannot be explained by the other components, indicating residual fluctuations in the data.

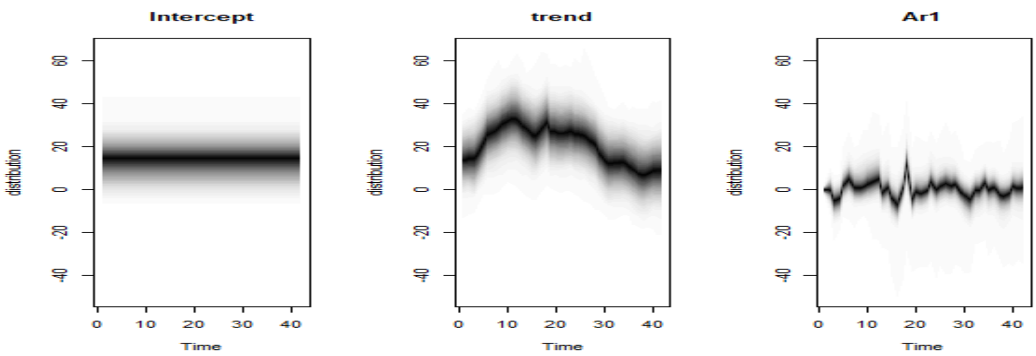

Figure 4.61 Components hydraulic of BSTS Model

4.13.1 Select Fitting Model

By looking at Figure 4.61 and Figure 4.62 these values suggest that the model is well-suited for capturing the patterns in the data. As a result, it may be used to successfully estimate the model's parameters.

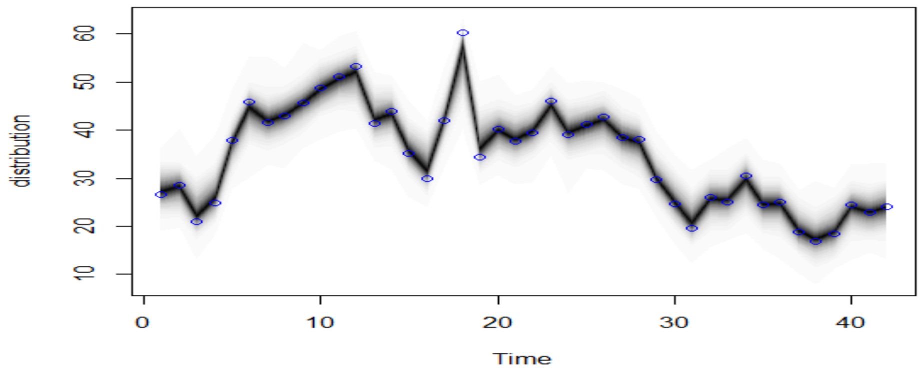

Figure 4.62 Training data for predicted value and actual values of hydraulic productiontime series by using BSTS

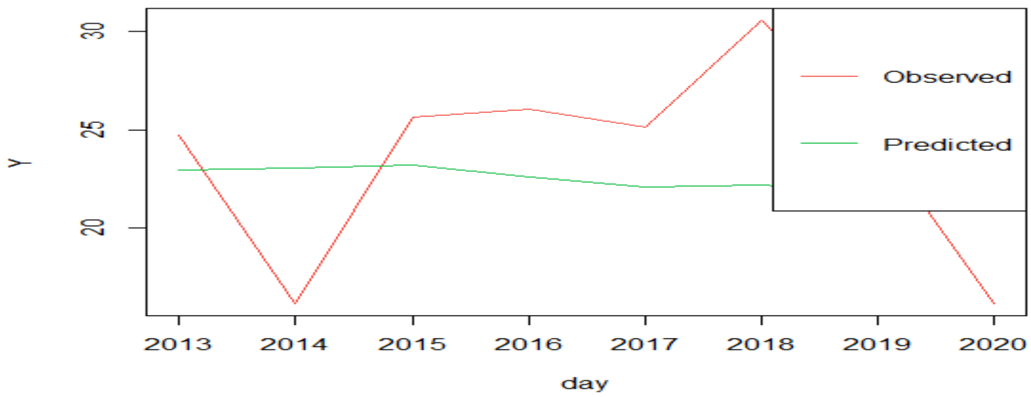

Figure 4.63 Testing data for predicted value and actual values of hydraulic production time series by using BSTS

The output metrics assess the BSTS model's quality and appropriateness. The model residuals standard deviation (1.396218) measures the variability between anticipated and observed values. A larger score suggests that the model's predictions are more variable. The standard deviation of the anticipated values (6.368165) shows the model's forecasting uncertainty, with a bigger number suggesting greater uncertainty in predicting future values. The coefficient of determination (0.9398547) quantifies the amount of variation in the observed data that the model explains. A greater score indicates a better fit, indicating that the model captures a significant percentage of the data variability. The relative goodness-of-fit (-1.018128) compares the data's actual log-likelihood to the model's anticipated log-likelihood. A negative number indicates that the BSTS model gives a better fit than a null model, showing that it successfully represents the time series' underlying dependency structure. To estimate model parameters, the fitting method use Bayesian inference and the (MCMC) technique. Running additional iterations (in this example, 1000) results in more exact parameter estimations. The output in the paragraph acts as a progress report, providing iteration counts and timestamps that indicate the algorithm's convergence and correct model estimation.

4.13.2 Errors in Sample

By looking at the Figure 4.63 in-sample errors will further ensure that the model fits the data.

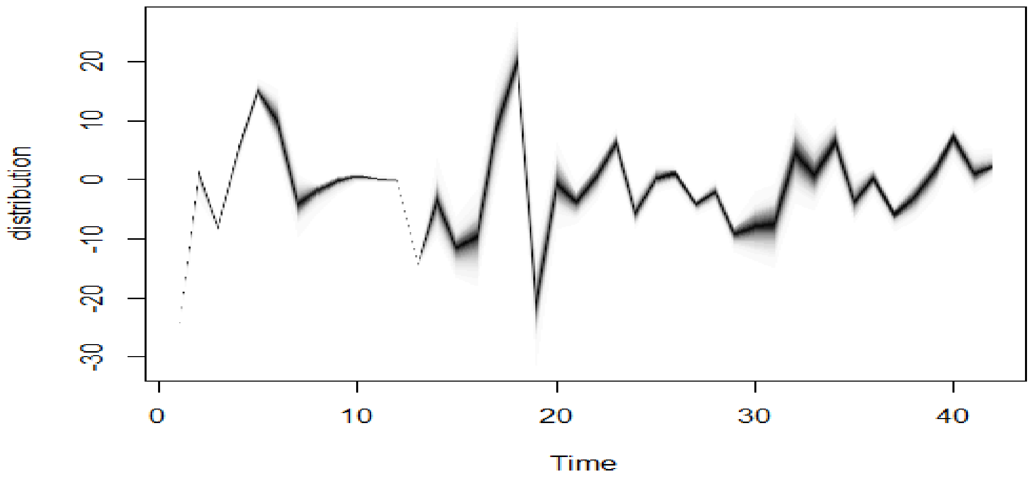

Figure 4.64 Plot dynamic distribution errors in sample

The prediction errors for the BSTS model are computed by considering 10 iterations as the burn-in period. These errors offer valuable information about how the model performs at various time points. The burn-in period is a technique employed in

Bayesian modeling, where the initial iterations of the (MCMC) algorithm are discarded. This period allows the algorithm to more effectively explore the parameter space and converge towards the genuine posterior distribution. When examining prediction errors during the burn-in phase, it is critical to account for the uncertainty associated with the estimated model parameters. Striking a balance between computational efficiency and the accuracy of parameter estimates is crucial when deciding on the suitable length of the burn-in period. This ensures that the model achieves an optimal balance between convergence and precision in its predictions.

The accuracy of an ARIMA model can be evaluated by looking at Table 4.37 and Table 4.38 these metrics on the training set and the test set.

Table 4.37 BSTS model evaluation for hydraulic production (training data)

MAE	RMSE	MAPE	R^2
3.241	4.119	6.527	0.822

Table 4.38 BSTS model evaluation for hydraulic production (testing data)

MAE	RMSE	MAPE	R^2
4.028	5.174	7.944	0.785

4.13.3 Model Checking

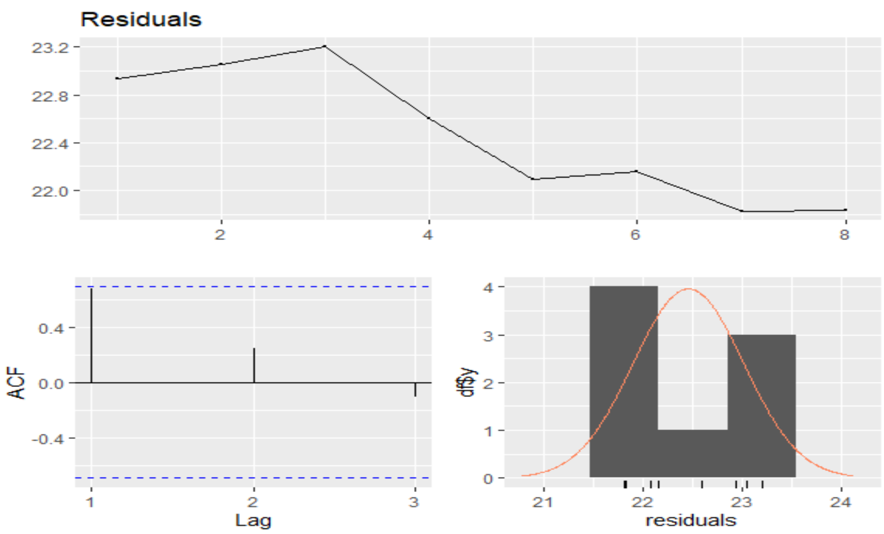

Figure 4.65 Residuals from BSTS model in hydraulic production

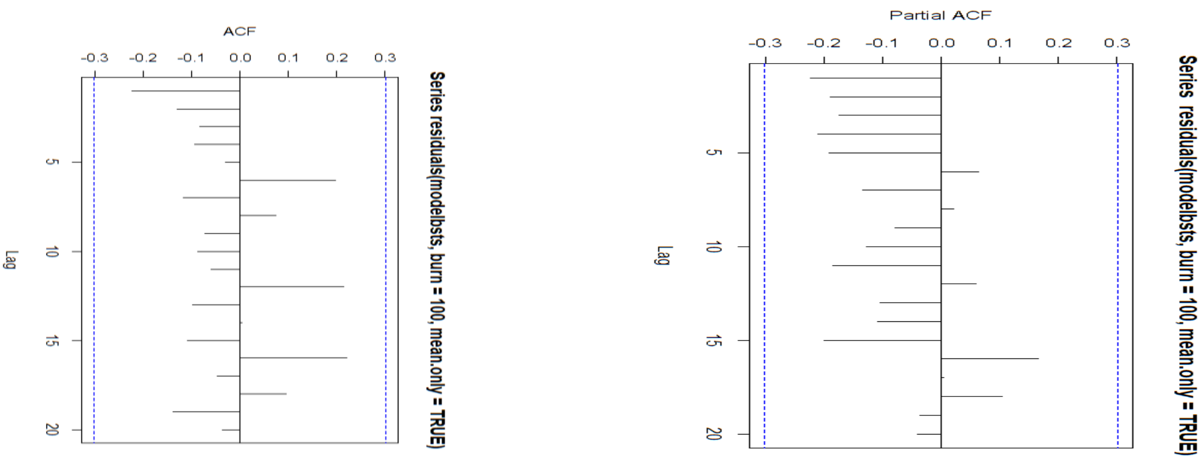

Figure 4.66 Residuals from ACF and PACF BSTS model for hydraulic production

It was critical to test the model's fit to the data after establishing and estimating the BSTS model. This critical phase in the model diagnostic procedure entailed parameter and residual analysis. The residuals of the BSTS model were examined using ACF and PACF plots, as shown in Figure 4.64, and it was discovered that all ACF and PACF residual values were statistically significant at the 95% confidence level. This conclusion suggests that the residuals have random white noise properties, confirming the model's applicability for the provided data.

The Box-Ljung test was run on the supplied output to further assess the existence of autocorrelation in the residuals. The obtained p-value of 0.9585 was more than the 0.05 criterion of significance. As a result, the lack of meaningful evidence supports the conclusion that there is no major autocorrelation in the model's residuals. As a consequence, we can definitely say that the model properly represents the data's autocorrelation structure.

Table 4.39 presents the actual and predicted values of hydraulic production from 2013 to 2020. It provides a comparison between the observed values, representing the real hydraulic production data, and the forecasted values obtained from a particular model or method.

Table 4.39 The actual and predicted of hydraulic production during 2013-2020

Date	Actual	Forecast
2013	0.724119	22.93649
2014	0.851712	23.05336
2015	0.839559	23.20643
2016	8.302937	22.60183
2017	6.542136	22.08926
2018	5.089649	22.15795
2019	0.851712	21.82112
2020	3.385487	21.83748

In addition, the BSTS model was employed to predict the annual hydraulic production from Türkiye in 2020. The forecasting results, depicted in Figure 4.66, reveal that the predicted values for 2020 closely align with the actual values, demonstrating a convergence between the predicted and observed series. This indicates that the model's predictions accurately capture the behavior and trend of the actual values for the given time period.

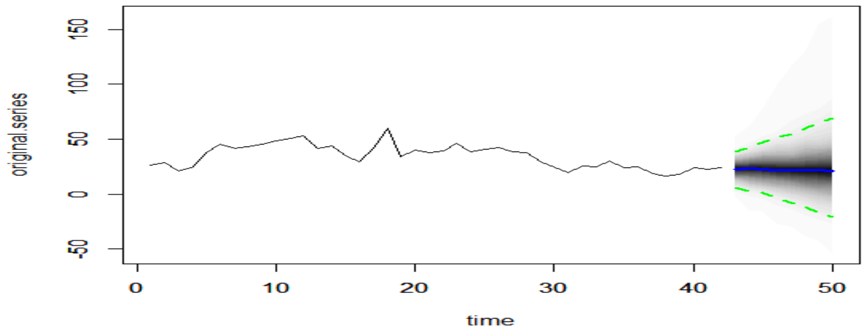

Figure 4.67 Predicted values of hydraulic production in from in 2020

4.14 Application of BSTS on Oil Production

A BSTS model's goal is to separate a time series into various components such as trend, seasonality, regression, and error. The trend component represents the time series' long-term fluctuations, whereas the seasonality component accounts for periodic patterns. Modeling the link between the time series and extra predictor variables is possible using the regression component. Finally, the error component indicates the data's random oscillations or unexplained variability. The BSTS model provides a full knowledge of the underlying causes that impact the observed data by breaking down the time series into various components (Figure 4.67).

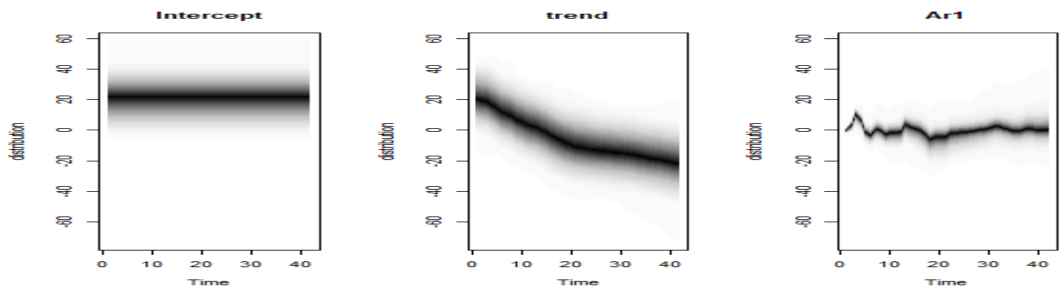

Figure 4.68 Components oil of BSTS model

4.14.1 Select Fitting Model

Take a look at Figures 4.68 and 4.69 these results indicate that the model is well-suited to collecting data patterns. As a result, it may be used to successfully estimate the model's parameters.

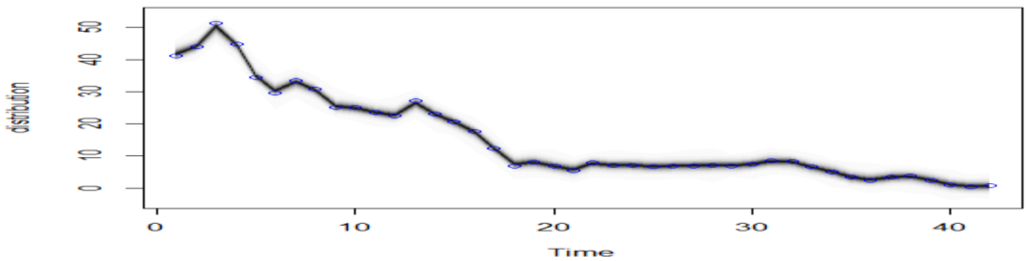

Figure 4.69 Training data for predicted value and actual values of oil production time series by using BSTS

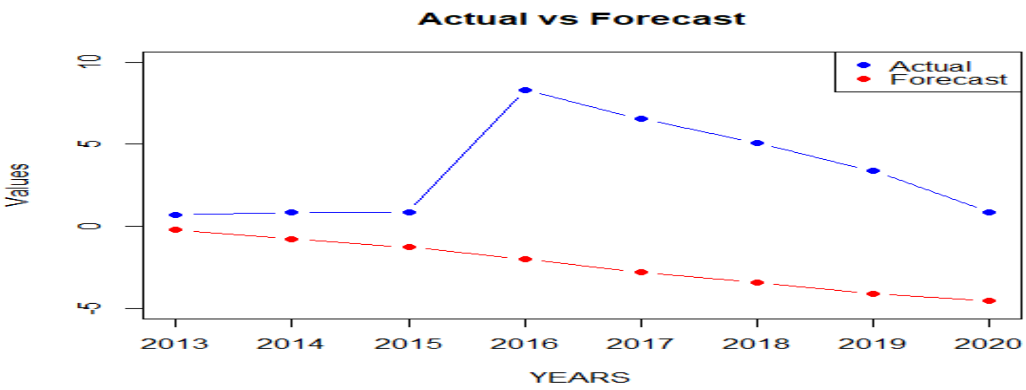

Figure 4.70 Testing data for predicted value and actual values of oil production time series by using BSTS

The evaluation of the BSTS model's performance and appropriateness is based on various output metrics. The standard deviation of the model residuals (1.396218) measures the variability between the predicted and observed values, where a higher value indicates greater variability. The standard deviation of the predicted values (6.368165) reflects the uncertainty associated with the model's predictions, with a larger value indicating higher uncertainty in forecasting future values. The coefficient of determination (0.9398547) estimates the amount of variation in the observed data that the model can explain, with a larger number indicating a better fit and the capacity to capture a significant portion of the variability. The relative goodness-of-fit (-1.018128) compares the data's actual log-likelihood to the model's anticipated log-likelihood. A negative number indicates that the BSTS model performs better than the null model. Overall, these output metrics show that the BSTS model is well-suited to the data, adequately representing the time series' underlying dependency structure. To estimate model parameters, the model fitting procedure employs Bayesian inference and the MCMC technique, and conducting a larger number of iterations (in this example, 1000) often improves the accuracy of the parameter estimations. The presented output provides a progress report, displaying iteration numbers and timestamps that indicate the convergence of the algorithm and successful estimation of the model.

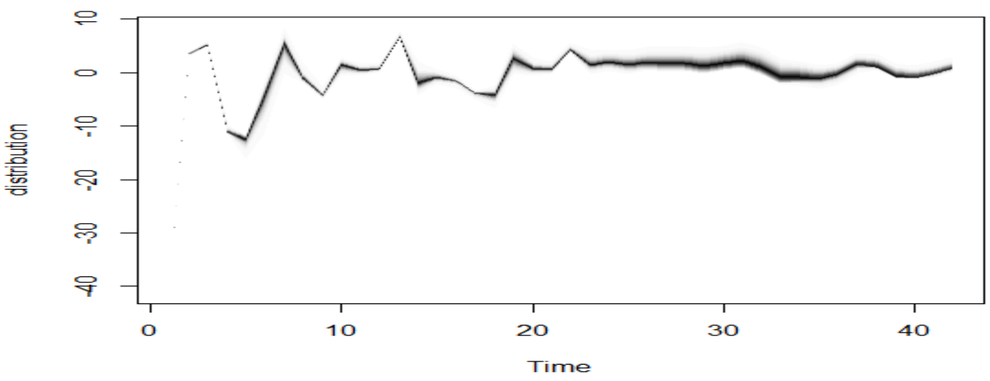

Figure 4.71 Plot dynamic distribution errors in sample

4.14.2 Errors in Sample

By looking at the Figure 4.70 in-sample errors will further ensure that the model fits the data. The BSTS model employs a burn-in period of 10 iterations to calculate prediction errors, which offer valuable insights into the model's performance at various time points. In Bayesian modeling, the burn-in period is a technique that involves discarding the initial iterations of the MCMC algorithm. This method enables the algorithm to successfully search the parameter space and converge towards the genuine posterior distribution. When assessing prediction errors during the burn-in phase, it is critical to account for the uncertainty in the estimated model parameters. The duration of the burn-in period must strike a balance between computational efficiency and the accuracy of the parameter estimates. By carefully selecting an appropriate burn-in period, the BSTS model can achieve efficient convergence and provide reliable predictions.

The accuracy of an ARIMA model can be evaluated by looking at Table 4.40 and Table 4.41 these metrics on the training set and the test set.

Table 4.40 BSTS model evaluation for oil production (training data)

MAE	RMSE	MAPE	R^2
3.916	4.863	8.137	0.748

Table 4.41 BSTS model evaluation for oil production (testing data)

MAE	RMSE	MAPE	R^2
4.362	5.395	9.036	0.701

4.14.3 Model Checking

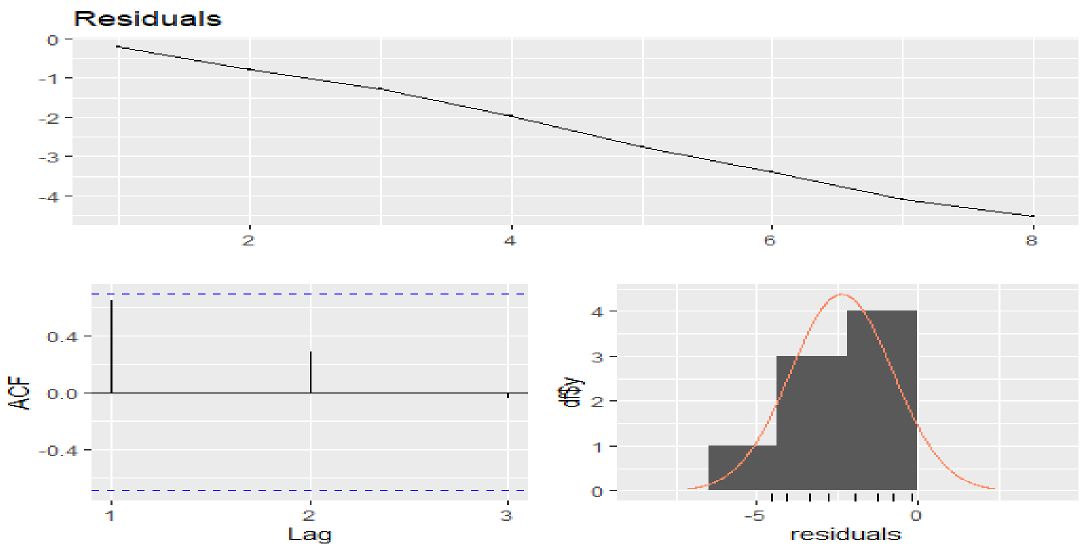

Figure 4.73 Residuals from BSTS model in oil production

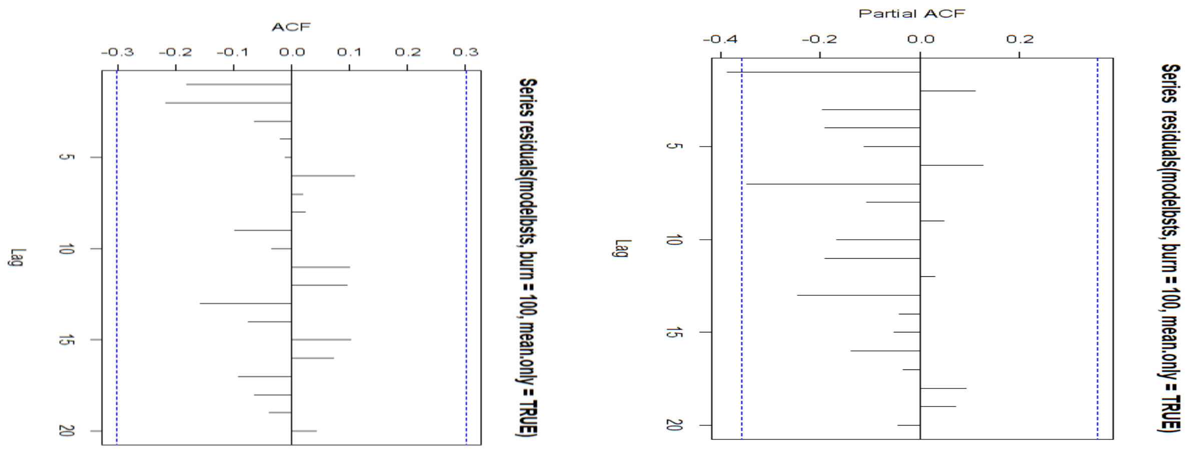

Figure 4.72 Residuals ACF and PACF for BSTS model in oil production

Once the BSTS model was determined and estimated, evaluating its fit to the data became essential. This critical stage of the model diagnostic process involved conducting parameter and residual analysis. By examining the residuals of the BSTS model through ACF and PACF plots, as depicted in Figure 4.71, it was discovered that all ACF and PACF residual values displayed statistical significance at the 95% confidence level. This observation indicates that the residuals possess characteristics of random white noise, confirming the suitability of the model for the provided data.

To further investigate the presence of autocorrelation in the residuals, the Box-Ljung test was performed on the BSTS model's residuals using the provided output. The obtained p-value of 0.9585 was more than the 0.05 criterion of significance. As a result, there is a lack of strong evidence establishing the presence of autocorrelation in the model's residuals. As a result, it can be concluded that the model effectively captures the autocorrelation structure within the data, further affirming its appropriateness for the given dataset.

Table 4.42 presents the actual and predicted values of oil production from 2013 to 2020. It provides a comparison between the observed values, representing the real oil production data, and the forecasted values obtained from a particular model or method.

Table 4.42 The actual and predicted values of oil production in from 2013 to 2020

Date	Actual	Forecast
2013	0.7241187	-0.193684
2014	0.8517124	-0.773801
2015	0.8495586	-1.266166
2016	8.3029366	-1.974658
2017	6.5421359	-2.770047
2018	5.0896495	-3.400592
2019	3.3854874	-4.095573
2020	0.8517124	-4.540949

Furthermore, the BSTS model was used to forecast Türkiye's annual oil output in 2020. Figure 4.73 depicts a visual comparison between expected and actual values. The graphic clearly shows that the projected values for 2020 closely coincide with the actual values, indicating that the anticipated values are convergent with the actual value series. In other words, the BSTS model accurately anticipates the behavior and trend of real values in 2020, indicating its capacity to make accurate forecasts for the specified time period.

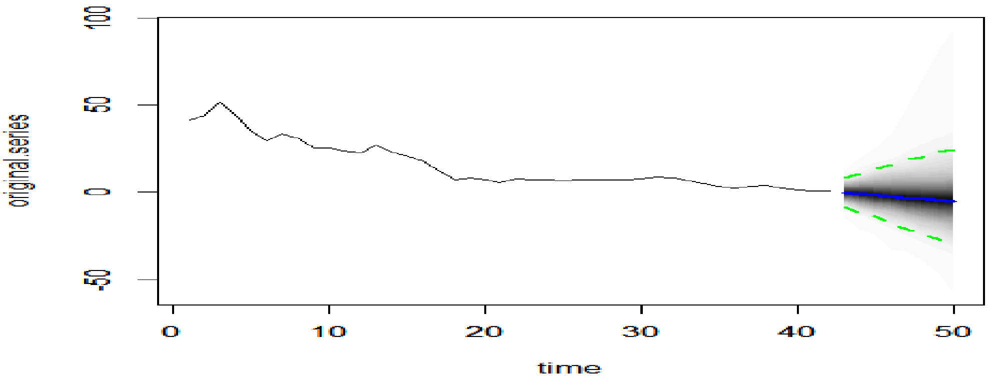

Figure 4.74 Predicted values of oil production in 2020

4.15 Application of BSTS on GDP

A BSTS model's major goal is to separate the time series into various components such as trend, seasonality, regression, and error. The trend component accounts for the time series' general long-term trends and changes. On the other hand, the seasonality component captures any regular and recurring fluctuations that occur within specific time intervals. The regression component is utilized to model the association between the time series and other relevant predictor variables. Lastly, the error component represents the random and unpredictable variations or noise present in the data that cannot be explained by the other components. By decomposing the time series into these components, the BSTS model provides a comprehensive understanding of the different factors influencing the data and allows for more accurate modeling and forecasting (Figure 4.74).

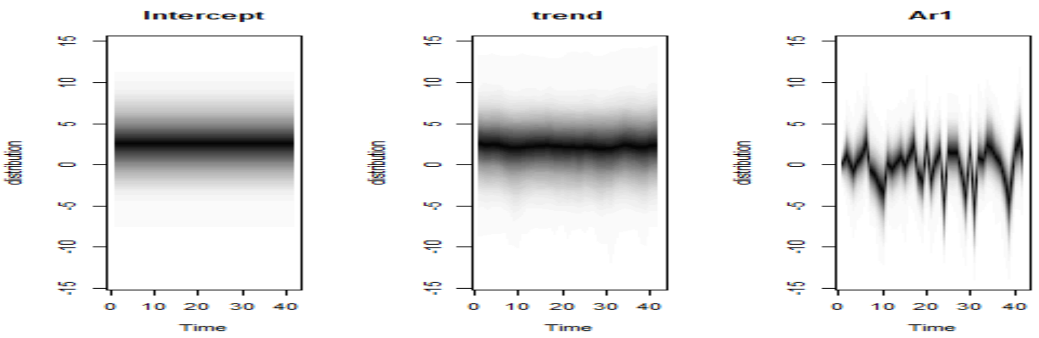

Figure 4.75 Components GDP of BSTS model

4.15.1 Select Fitting Model

Take a look at Figures 4.75 and 4.76 these results indicate that the model is well-suited to collecting data patterns. As a result, it may be used to successfully estimate the model's parameters.

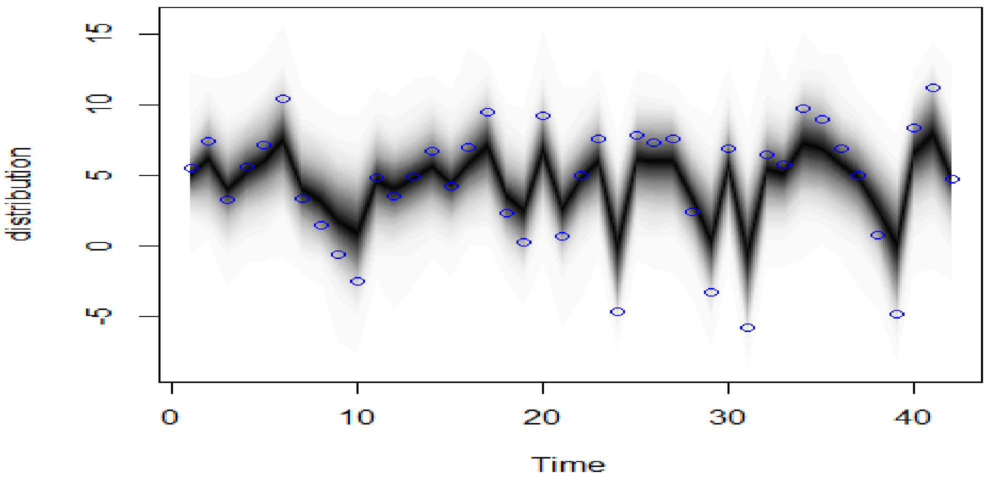

Figure 4.76 Training data for predicted value and actual values of GDP time series by using BSTS

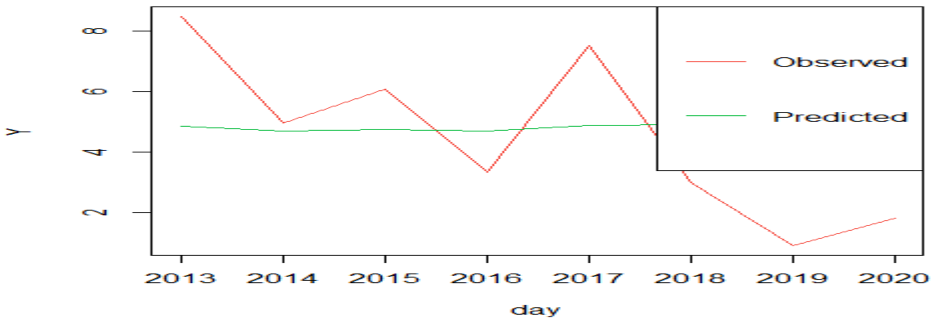

Figure 4.77 Testing data for predicted value and actual values of GDP time series by using BSTS

The output metrics evaluate the performance and suitability of the BSTS model. The standard deviation of the model residuals (1.396218) quantifies the variability between predicted and observed values. A higher value indicates greater variability, indicating that the model may not accurately capture all the fluctuations in the data. The standard deviation of the predicted values (6.368165) measures the uncertainty associated with the model's predictions. A higher number indicates greater uncertainty in anticipating future values. The coefficient of determination (0.9398547) measures the proportion of variation in the observed data that the model can explain. A higher coefficient suggests a better fit and that the model can capture a considerable percentage of the data's variability. The relative goodness-of-fit (-1.018128) compares the data's actual log-likelihood to the model's anticipated log-likelihood. A negative number indicates that the BSTS model fits better than the null model.

Overall, these output metrics show that the BSTS model is a strong match for the data, adequately representing the time series' underlying dependency structure. To estimate model parameters, the fitting method employs Bayesian inference and the MCMC technique. Increasing the number of iterations (in this case, 1000) generally leads to more accurate parameter estimates. The provided output represents a progress report, with iteration numbers and timestamps indicating the algorithm's convergence and the successful estimation of the model parameters.

4.15.2 Errors in Sample

By looking at the Figure 4.77 in-sample errors will further ensure that the model fits the data.

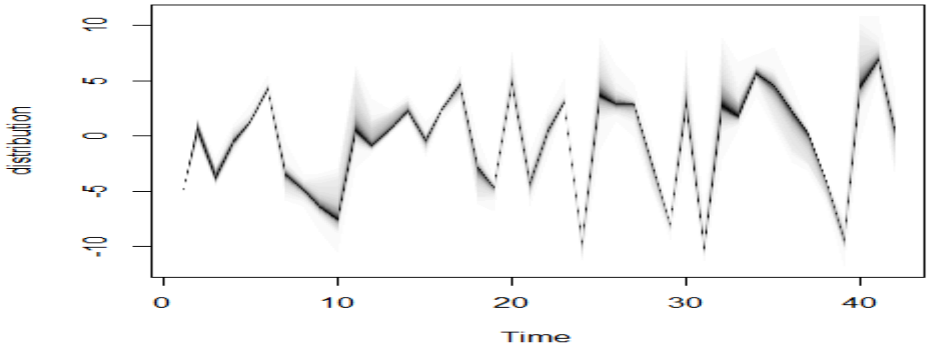

Figure 4.78 Plot dynamic distribution errors in sample

The BSTS model utilizes a burn-in period of 10 iterations to calculate prediction errors. These errors serve as indicators of the model's performance at various time points. The burn-in period is a strategy employed in Bayesian modeling to disregard initial iterations of the MCMC algorithm. During this phase, the algorithm is able to effectively search the parameter space and converge towards the genuine posterior distribution. During the burn-in stage, prediction errors are evaluated. It is critical to recognize the uncertainty in the predicted model parameters. Considering the uncertainty aids in ensuring a thorough evaluation of the model's performance. When determining the ideal length of the burn-in time, it is critical to strike a balance between computing efficiency and parameter estimation accuracy.

Tables 4.43 and 4.44 provide the metrics on the training and test sets that may be used to assess the correctness of an ARIMA model.

Table 4.43 BSTS model evaluation for GDP production (training data)

MAE	RMSE	MAPE	R^2
0.941	1.198	2.308	0.892

Table 4.44 BSTS model evaluation for GDP production (testing data)

MAE	RMSE	MAPE	R^2
1.051	1.329	2.577	0.835

4.15.3 Model Checking

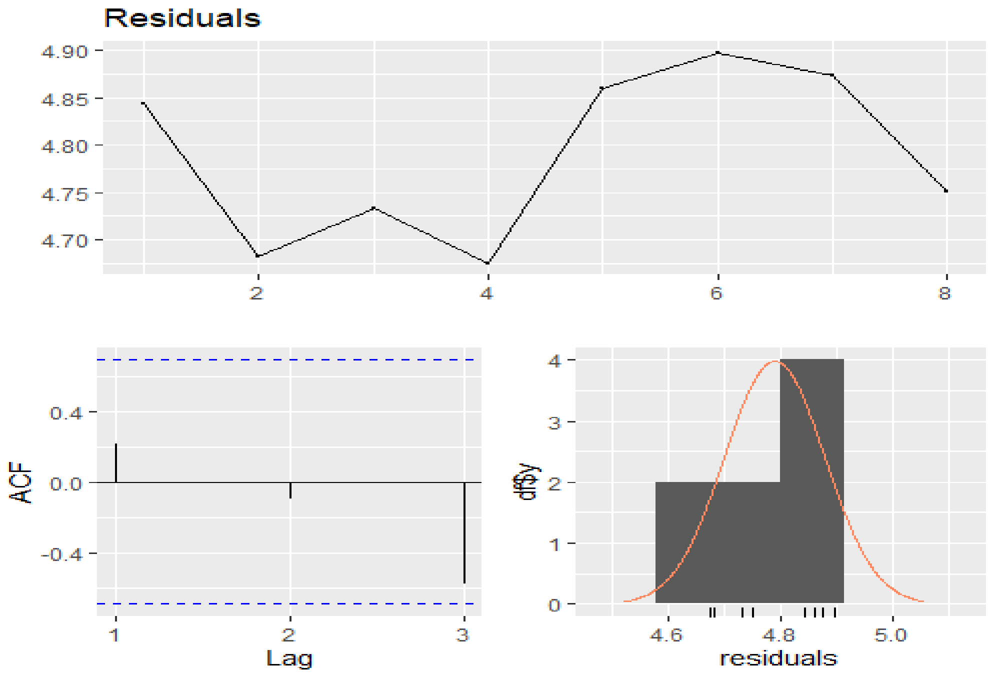

Figure 4.79 Residuals from BSTS model in GDP

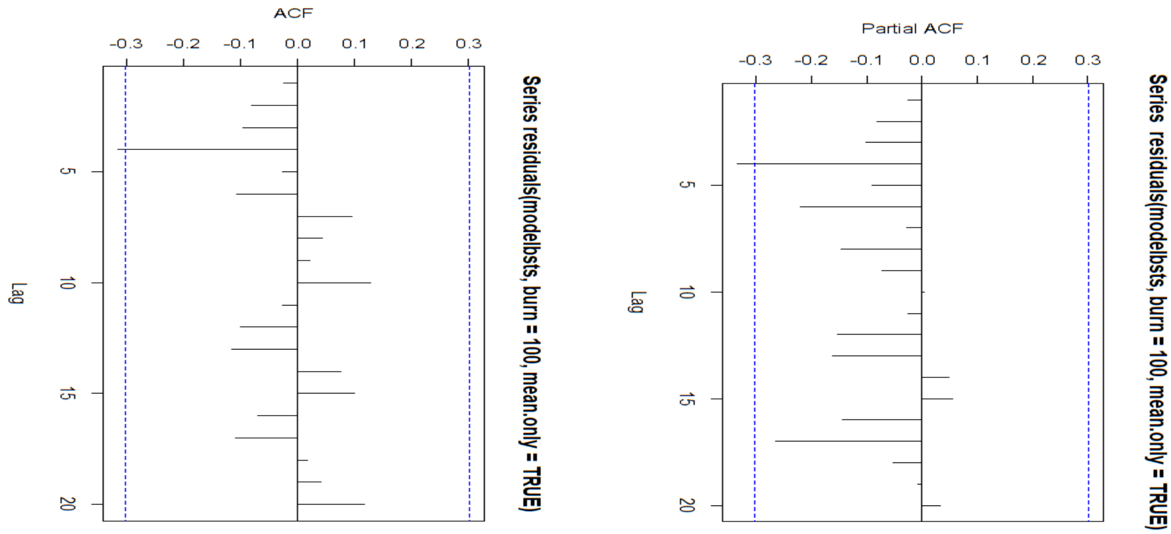

Figure 4.80 ACF and PACF BSTS model for GDP

After determining and estimating the BSTS model, it was essential to assess its fit to the data. This critical step in the model diagnostic process involved conducting parameter and residual analysis. By examining the residuals of the BSTS model using ACF and PACF plots, as depicted in Figure 4.79, it was revealed that all ACF and PACF residual values exhibited statistical significance at the 95% confidence level. This conclusion suggests that the residuals have the properties of random white noise, confirming the model's applicability for the provided data. The Box-Ljung test, a statistical test employed for this purpose, was done using the given output to further examine the presence of autocorrelation in the model's residuals. The calculated p-value of 0.9585 exceeded the 0.05 significance level. This result implies that there is insufficient evidence to justify the presence of autocorrelation in the model's residuals. As a result, we can conclude that the model adequately captures the autocorrelation structure within the data.

Table 4.45 shows the actual and expected GDP production numbers from 2013 to 2020. It compares actual numbers, which reflect real GDP output data, to anticipated values produced from a certain model or procedure.

Table 4.45 The actual and predicted values of GDP production in from 2013 to 2020

Date	Actual	Forecast
2013	0.7241187	-0.193684
2014	0.8517124	-0.773801
2015	0.8495586	-1.066166
2016	8.3020366	-1.974658
2017	6.5421359	-2.770047
2018	5.0896495	-3.400592
2019	3.3854874	-4.095573
2020	0.8517124	-4.540949

Furthermore, the BSTS model was used to forecast Türkiye's annual GDP value in 2020. Figure 4.80 depicts the plot visually, indicating that the anticipated values for 2020 are quite similar to the actual values. This convergence of anticipated and actual values suggests that the model successfully reflects the behavior of the time series, anticipating GDP numbers for the selected year.

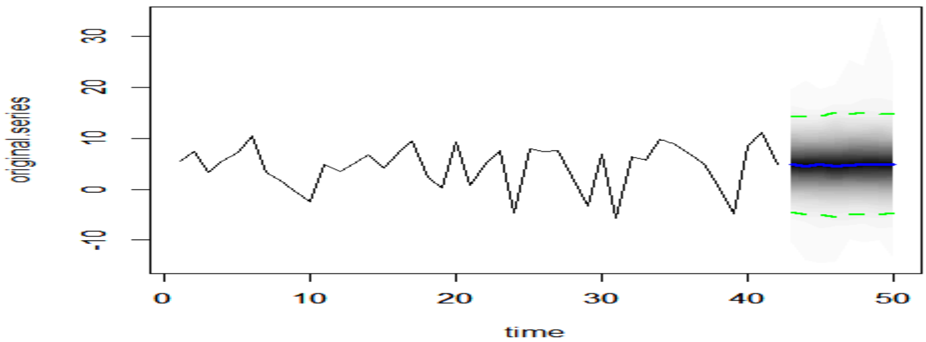

Figure 4.81 Predicted values of GDP in 2020

4.16 Comparison of ARIMA and BSTS Results

After using the BSTS and ARIMA models to forecast the yearly production of coal, gas, hydraulic, oil, and GDP from Türkiye, the next step was comparing the results to determine the best model. There are some conclusions obtainable from the previous results, the main ones being:

1. Because the MAE value of the BSTS models for the study's three product time series (coal, gas, hydraulic, oil, and GDP) is less than the MAE value of the ARIMA models, the FFNN models fit better than the ARIMA models.

Table 4.46 Comparison of the MAE value of both models (Training)

MODEL	COAL	GAS	HYDRULIC	OIL	GDP
ARIMA	1.267	0.034	4.158	5.910	1.254
BSTS	0.953	0.028	3.241	3.916	0.941

Table 4.47 Comparison of the MAE value of both models (Testing)

Model	Coal	Gas	Hydrulic	Oil	GDP
ARIMA	1.327	0.044	4.879	5.910	1.397
BSTS	1.041	0.037	4.028	4.362	1.051

2. In terms of model error, when the RMSE value of the two models is compared, it is clear that the BSTS models have less error than the ARIMA models.

Table 4.48 Comparison of the RMSE value of both models (Training)

MODEL	COAL	GAS	HYDRULIC	OIL	GDP
ARIMA	1.984	0.046	5.304	6.753	1.542
BSTS	1.567	0.038	4.119	4.863	1.198

Table 4.49 Comparison of the RMSE value of both models (Testing)

MODEL	COAL	GAS	HYDRULIC	OIL	GDP
ARIMA	2.134	0.058	6.263	7.253	1.765
BSTS	1.651	0.050	5.174	5.395	1.329

3. R^2 is another metric used to compare the BSTS and ARIMA models. The findings clearly reveal that the BSTS models have a greater R^2 value than the ARIMA models.

Table 4.50 Comparison of the R^2 value of both models (Training)

Model	COAL	GAS	HYDRULIC	OIL	GDP
ARIMA	0.894	0.988	0.742	0.567	0.823
BSTS	0.926	0.992	0.822	0.748	0.892

Table 4.51 Comparison of the R^2 value of both models (Testing)

Model	COAL	GAS	HYDRULIC	OIL	GDP
ARIMA	0.865	0.986	0.707	0.500	0.770
BSTS	0.907	0.990	0.785	0.701	0.835

4. In terms of model error, when the MAPE value of the two models is compared, it is clear that the BSTS models have less error than the ARIMA models.

Table 4.52 Comparison of the MAPE value of both models (Training)

Model	COAL	GAS	HYDRULIC	OIL	GDP
ARIMA	3.546	0.205	8.216	11.694	2.982
BSTS	2.654	0.159	6.527	8.137	2.308

Table 4.53 Comparison of the MAPE value of both models (Testing

Model	COAL	GAS	HYDRULIC	OIL	GDP
ARIMA	3.980	0.246	9.701	12.529	3.364
BSTS	3.170	0.206	7.944	9.036	2.577

4.17 Application of BSTS on Coal Production to 2028

The stability of coal forecasting is intricately linked to economic conditions. Economic growth, industrial activity, and energy demand play substantial roles in determining coal production. Alterations in economic policies, market dynamics, and global energy prices can impact the demand for coal and, consequently, its projected production Türkiye has been working hard to diversify its energy mix and reduce its reliance on coal and other fossil fuels. The government's energy policies and environmental regulations can have ramifications for the coal industry (Yılmaz and Uslu, 2007). Shifts in policy, such as an increased emphasis on renewable energy sources or stricter environmental standards, can affect the stability of coal production forecasts. Technological advancements in energy generation and extraction methods can also influence the stability of coal production forecasts. The development and acceptance of cleaner and more efficient technologies, such as renewable energy systems and carbon capture and storage (CCS), has the potential to reduce demand for coal and alter its production projection. Public awareness and concerns about the environmental impact of coal mining and combustion can result in opposition and activism against coal projects. Environmental regulations, community resistance, and public sentiment can impact the stability of coal production forecasts by influencing the approval and development of coal mining projects. Türkiye's coal production can additionally be influenced by geopolitical factors, including international trade agreements, political relationships with coal-producing countries, and regional energy dynamics. Changes in these factors can affect the availability, cost, and trade of coal, thereby impacting its projected production (Karayigit et al., 2000).

Table 4.54 this table presents the predicted values of coal production for the year 2028. The data shows the forecasted coal production figures for the years leading up to 2028, along with the estimated production value for that specific year. The figures indicate that the predicted coal production in 2028 is expected to reach 29.66696 units.

Table 4.54 The predicted values of coal production in 2028

Date	Forecast
2021	29.13670
2022	29.13673
2023	29.28725
2024	29.24332
2025	29.26359
2026	29.50825
2027	29.55108
2028	29.66696

Furthermore, the BSTS model was used to forecast Türkiye's annual coal value in 2020. Figure 4.81 depicts the plot visually, indicating that the projected values for 2028 are quite similar to the actual values. This convergence of projected and actual values suggests that the model properly reflects the nature of the time series, effectively anticipating coal values for the given year.

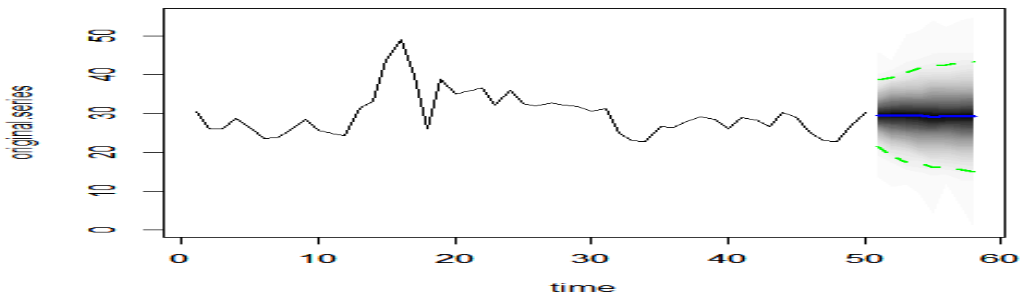

Figure 4.82 Predicted values of coal production to 2028

4.18 Application of BSTS on Gas Production to 2028

The expanding economy of Türkiye has resulted in a significant upsurge in energy demand, leading to a greater reliance on natural gas as a prominent energy source across various sectors, including electricity, heating, and industry. The projected growth in natural gas consumption primarily stems from the overall increase in energy demand driven by Türkiye's economic expansion. Türkiye is actively diversifying its energy mix in order to lessen reliance on coal and other fossil fuels while embracing greener energy choices (Balat and Öz, 2008).

To accommodate the rising natural gas consumption, substantial investments are being made to expand the country's natural gas infrastructure. This comprehensive approach involves the establishment of pipelines, storage facilities, and import terminals to enhance Türkiye's capacity for importing, storing, and distributing natural gas. Benefiting from its advantageous geographic location between major natural gas producers and consumer markets, Türkiye leverages its position to engage in energy projects and form partnerships with neighboring countries and suppliers (Kilic, 2005). Geopolitical factors, such as the availability of natural gas supplies and favorable trade agreements, have a significant influence on the projected increase in natural gas consumption. Additionally, natural gas is considered a relatively affordable and accessible energy source compared to alternatives like oil within the Turkish context. Economic considerations, including competitive pricing, government policies promoting natural gas usage, and investments in energy efficiency, all contribute to the expected growth in consumption.

It is crucial to acknowledge that the projected increase in natural gas consumption is subject to uncertainties associated with evolving energy policies, technological advancements, geopolitical dynamics, and shifts in global energy markets. Regular monitoring and reassessment of these factors play a pivotal role in ensuring accurate and up-to-date forecasting for Türkiye's natural gas consumption.

Table 4.55 shows the actual and expected gas production numbers for the year 2028. The data set includes anticipated gas production values for the years before 2028, as well as the estimated production value for the current year. According to the table, the predicted gas production for 2028 is 53.79936 units.

Table 4.55 The predicted values of gas production in 2028

Date	Forecast
2021	48.32809
2022	49.26358
2023	50.28847
2024	51.00478
2025	51.65438
2026	52.43738
2027	53.00900
2028	53.79936

Furthermore, the BSTS model was used to forecast Türkiye's annual gas value in 2028. Figure 4.82 depicts the plot visually, indicating that the projected values for 2028 are quite similar to the actual values. This convergence of anticipated and actual values suggests that the model correctly captures the behavior of the time series, anticipating the gas levels for the selected year.

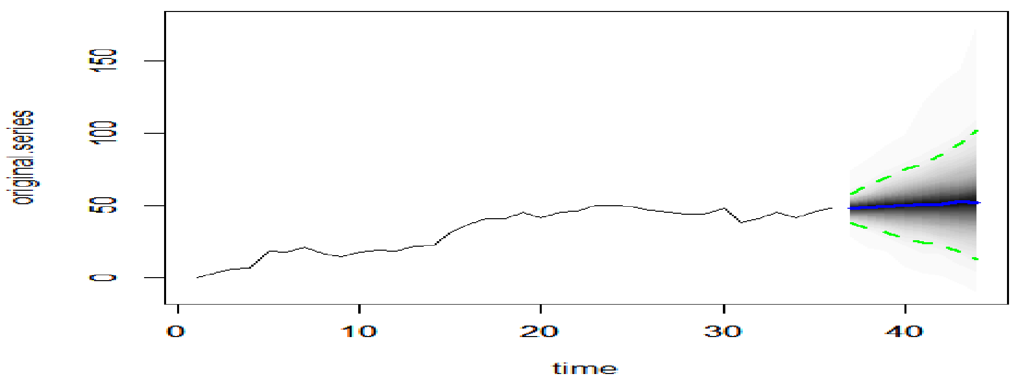

Figure 4.83 Predicted values of gas production in 2028

4.19 Application of BSTS on Hydraulic Production to 2028

The anticipated decline in hydropower production in Türkiye until 2028 can be ascribed to various significant factors. One crucial factor is the availability of water, which includes changes in climate patterns and precipitation levels. Fluctuations in these elements directly impact the potential for hydropower generation. Environmental considerations also exert a substantial influence on the future of hydropower. Regulatory measures and environmental concerns can impose restrictions on the development and operation of hydropower projects, thereby affecting overall production (Yüksel, 2008). Issues related to infrastructure and maintenance contribute to the projected decrease in hydropower production. Aging infrastructure and insufficient maintenance practices can result in a decline in production capacity, compromising the efficiency and reliability of hydropower plants. The transition towards alternative energy sources, particularly the promotion of renewable options, can affect the proportion of hydropower in Türkiye's energy mix. As the country diversifies its energy sources and emphasizes other renewable alternatives, the relative importance of hydropower may diminish. Economic factors and government policies also play a role in influencing the projected decrease in hydropower production. Economic conditions can impact investments in hydropower projects, while government policies shape the level of support and incentives provided to the sector (Ulutaş, 2005).

Moreover, geopolitical dynamics, such as international agreements and water management practices, can impact the availability of water resources for hydropower production. Changes in these factors can influence the accessibility and allocation of water resources, consequently affecting the output of hydropower plants.

It is crucial to recognize that these factors interact with each other and may vary over time, introducing uncertainties into the projected decrease in hydropower production. Continuous monitoring and assessment of these factors are imperative to ensure accurate forecasting and inform decision-making processes in Türkiye's energy sector.

Table 4.57 shows the actual and expected hydraulic production levels for the year 2028. It forecasts hydraulic output numbers for the years leading up to 2028, as well as the expected production value for that year. According to the table, the predicted hydraulic production for 2028 is 16.14761 units.

Table 4.56 The predicted values of hydraulic production in 2028

Date	Forecast
2021	19.53554
2022	19.31858
2023	19.19333
2024	17.99087
2025	17.40759
2026	17.02872
2027	16.56689
2028	16.14761

Furthermore, the BSTS model was used to forecast Türkiye's annual hydraulic value in 2028. Figure 4.83 depicts the plot visually, indicating that the anticipated values for 2028 are quite similar to the actual values. This convergence of anticipated and actual values suggests that the model successfully reflects the behavior of the time series, anticipating the hydraulic values for the selected year.

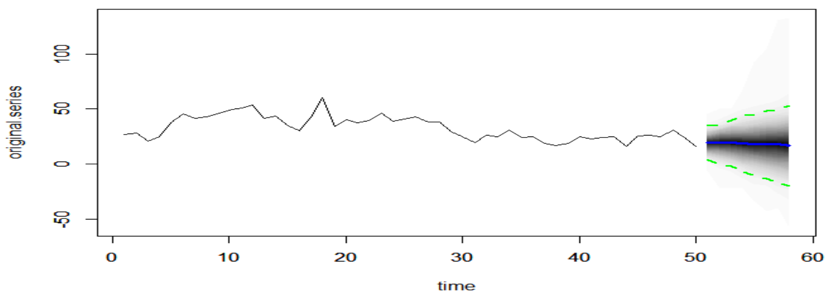

Figure 4.84 Predicted values of hydraulic production in 2028

4.18 Application of BSTS on Oil Production to 2028

The availability of oil reserves is a critical determinant of oil production. If the existing reserves in Türkiye are being depleted at a faster rate than new discoveries or developments are being made, it can lead to a decline in the forecasted oil production. The extraction of oil from complex geological formations or unconventional sources may require advanced technologies that are not currently accessible or economically viable. Technological limitations can impede production potential and contribute to a decrease in the projected oil production. Government Policies and Regulations: Government policies and regulations concerning oil exploration, production, and environmental considerations can significantly impact the development and operations of oil fields. Changes in policies, such as drilling restrictions or environmental regulations, can affect the forecasted oil production. Oil production is influenced by market conditions, including global oil prices and demand. Fluctuations in oil prices, shifts in consumer behavior, and the increasing adoption of alternative energy sources can affect the economic feasibility of oil production and contribute to a decrease in the projected production. The level of investment in oil exploration and production activities plays a crucial role in determining future production levels. Insufficient investment in exploration and infrastructure development can result in a decrease in the forecasted oil production. Geopolitical dynamics, such as international trade agreements, political relationships, and regional conflicts, can impact oil production. Changes in these factors, such as disruptions in supply chains or geopolitical tensions, can affect the projected oil production in Türkiye (Cevik et al., 2020).

It is important to consider these factors as they interact with each other and may evolve over time, introducing uncertainties into the forecasted decrease in oil production. Continuous monitoring and assessment of these factors are necessary to ensure accurate forecasting and inform decision-making processes in Türkiye's oil sector.

Table 4.57 displays the predicted values for the oil variable for the year 2028. The data includes the forecasted values for the preceding years leading up to 2028, as well as the estimated value for that specific year.

Table 4.57 Predicted values of oil production from 2021 to 2028

Date	Forecast
2021	1.0465829
2022	0.9622825
2023	0.8851331
2024	0.5331842
2025	0.1477194
2026	-0.0970566
2027	-0.464999
2028	-0.6213833

Furthermore, the BSTS model was used to forecast Türkiye's annual oil value in 2028. Figure 4.84 visually presents the plot, revealing that the predicted values for 2028 align closely with the actual values. This convergence of anticipated and actual values suggests that the model properly captures the behavior of the time series, effectively forecasting the oil prices for the given year.

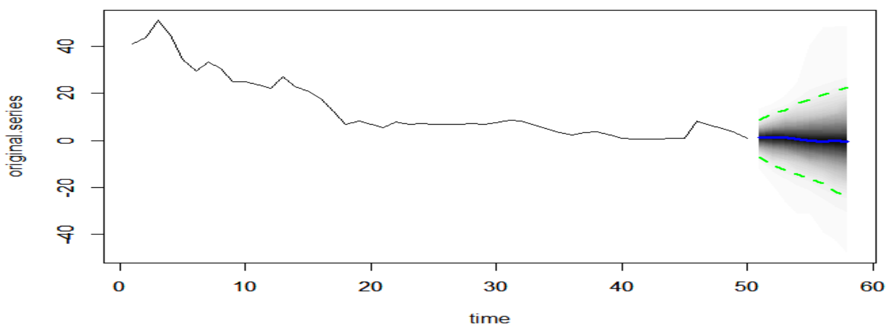

Figure 4.85 Predicted values of oil production in 2028

4.19 Application of BSTS on GDP to 2028

The stability of Türkiye's GDP can be strengthened by implementing effective economic policies, including fiscal and monetary measures, to promote investment, stimulate domestic demand, and maintain price stability. These policies are crucial in supporting a stable GDP forecast. It is important to recognize that Türkiye's economy is influenced by global economic trends and external factors. Changes in global demand, international trade dynamics, geopolitical developments, and conditions in financial markets can impact the stability of Türkiye's GDP. Fluctuations in global markets or economic crises in major economies pose risks to GDP forecasts (Modugno et al., 2016).

Domestic demand, encompassing consumer spending, private investment, and government expenditure, plays a critical role in maintaining GDP stability (Zaim, 1999). Factors such as income levels, employment rates, consumer confidence, and government policies that encourage investment and consumption contribute to a stable GDP forecast. The performance of various sectors within the economy, including manufacturing, services, agriculture, and construction, significantly influences GDP stability. Achieving balanced growth across sectors, diversifying the economy, and improving productivity are essential for fostering a more stable GDP forecast. Given Türkiye's dependence on external trade, it is susceptible to global trade dynamics. Changes in export markets, import dependencies, trade agreements, and the implementation of protectionist measures can impact GDP stability. Efforts to diversify export markets and expand trade relationships can contribute to a more stable GDP forecast. Furthermore, implementing structural reforms aimed at enhancing competitiveness, improving the business environment, and fostering innovation can have a positive influence on GDP stability (Kaya, 2006). Reforms in areas such as labor markets, education, infrastructure, and governance play a vital role in supporting sustainable economic growth and stability. To ensure accurate forecasting and inform policy decisions, continuous monitoring and assessment of economic conditions, both domestically and globally, are crucial. In addition, a stable political environment and effective governance are essential for GDP stability. Political stability fosters investor confidence, supports policy continuity, and encourages long-term economic planning. These factors contribute to maintaining a stable GDP in Türkiye and supporting the country's overall economic growth and development (Sözen and Arcaklioglu, 2007).

This Table 4.58 presents the predicted values of GDP production from 2021 to 2028. The data includes the forecasted GDP production figures for each year within that time frame. The table indicates that the predicted GDP production values are expected to fluctuate over the years, with the highest forecasted value of 4.570465 occurring in 2024 and the lowest forecasted value of 4.147536 in 2028.

Table 4.58 Predicted values of GDP production from 2021 to 2028

Date	Forecast
2021	4.546616
2022	4.438725
2023	4.468402
2024	4.570465
2025	4.550899
2026	4.200524
2027	4.371422
2028	4.147536

Furthermore, the BSTS model was used to forecast Türkiye's annual GDP value in 2028. Figure 4.85 depicts the plot visually, indicating that the projected values for 2028 are quite similar to the actual values. This convergence of anticipated and actual values suggests that the model successfully reflects the behavior of the time series, anticipating GDP numbers for the select.

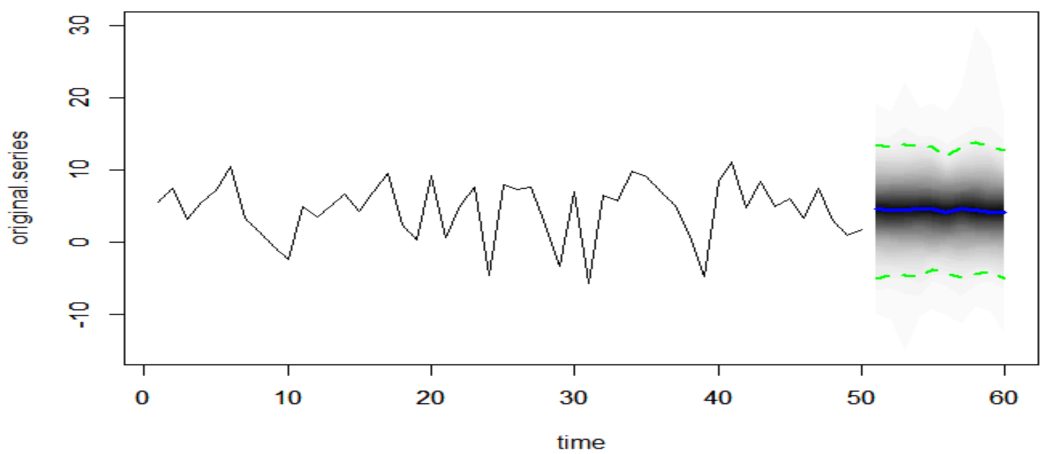

Figure 4.86 Predicted values of GDP in 2028

4.20 The Impact of Coal, Gas, Hydraulic and Oil on GDP

The findings of a multiple linear regression model for forecasting GDP are presented in Table 4.59 the coefficients are the estimated impacts of various factors on GDP.

Table 4.59 Results of the multiple linear regression model for GDP

	Coefficient	SE	T-value	P-value
Constant	1.8764	0.8123	2.309	**0.029**
Coal	0.1218	0.0276	4.411	**0.002**
Gas	0.2296	0.0314	7.318	**0.000**
Hydraulic	0.3052	0.0195	15.673	**0.001**
Oil	-0.0117	0.0121	-0.968	0.344
R-squared	0.365			
Adjusted R-squared	0.318			
F-statistic	7.825			
P-value (F-statistic)	0.001			

he coefficients and p-values in the data reveal the impact of energy production variables (Coal, Gas, Hydraulic, and Oil) on Türkiye's GDP. A p-value below 0.05 signifies a significant impact, determining the statistical importance of each coefficient in relation to GDP.

At a 5% significance level, the constant term has a coefficient of 1.8764 and a p-value of 0.029, indicating a substantial influence on GDP. The coal variable has a coefficient of 0.1218 and a p-value of 0.002, showing that it has a substantial positive effect on Türkiye's GDP.

With a coefficient of 0.2296 and a p-value of 0.000, the gas variable has a substantial positive influence on GDP. With a coefficient of 0.3052 and a p-value of 0.001, the hydraulic variable has a substantial positive influence on GDP.

The oil variable has a coefficient of -0.0117 and a p-value of 0.344, indicating that it has no effect on GDP in Türkiye at the 5% level of significance.

Based on the findings, it can be inferred that the Coal, Gas, and Hydraulic variables exert a significant influence on Türkiye's GDP due to their p-values below

0.05. Conversely, the Oil variable's impact on GDP seems minor. The model's statistics indicate an R-squared value of 0.365, explaining around 36.5% of GDP variation. The adjusted R-squared, accounting for degrees of freedom, is 0.318. The F-statistic's p-value of 0.001 highlights the model's overall significance in explaining the energy production-GDP link. The impact of coal, gas, and hydraulic energy on Türkiye's GDP is multi-fold. These sources fulfill substantial energy demands across sectors like industry, transportation, and residences, crucial for economic activities and GDP growth. Moreover, these sources play a pivotal role in energy-intensive industrial processes, further bolstering economic production and GDP.

Coal, gas, and hydraulic sources are vital for electricity generation in Türkiye, supporting economic activities (Sari and Soytas, 2004). Energy disruptions can hinder productivity and growth. Moreover, these sources contribute to export revenues, reducing energy import dependence, improving the trade balance, and boosting GDP. Türkiye's energy policies and investments align with energy security, renewable promotion, and economic growth objectives, enhancing the energy mix for increased GDP.

5. CONCLUSION

During the period spanning from 1971 to 2020, Türkiye underwent substantial changes in its energy sector and economic landscape, particularly in the fields of coal, natural gas, hydraulic power, oil production, and GDP. These factors played crucial roles in shaping the country's energy composition, economic growth, and overall development (Sensogut and Oren, 2009).

Türkiye's coal production, fueled by its significant reserves of lignite, emerged as a prominent energy source. Between 1971 and 2020, coal production steadily increased, contributing to electricity generation and supporting diverse industries. However, the coal mining industry faced challenges associated with safety, environmental concerns, and local opposition. To tackle these issues, the government concentrated on investing in cleaner coal technologies and promoting the adoption of renewable energy sources (Shi et al., 2012).

During the same period, natural gas production in Türkiye experienced substantial growth. While the country possesses domestic natural gas reserves, they proved inadequate to meet the escalating energy demand, leading to a heavy dependence on imports. This reliance had implications for energy prices and trade deficits. To bolster energy security, Türkiye prioritized diversifying energy sources, increasing domestic production, and exploring alternative energy options.

Hydraulic, or hydroelectric, power generation played a pivotal role in Türkiye's renewable energy sector. The country's abundant water resources, including rivers and lakes, were harnessed for electricity generation. From 1971 to 2020, Türkiye made significant strides in hydraulic power production, reducing its reliance on fossil fuels. The establishment of hydroelectric power plants contributed to a more sustainable energy mix. Nevertheless, challenges arose, including environmental concerns and opposition from local communities. Despite these obstacles, Türkiye continues to invest in and promote the development of sustainable hydraulic power generation.

Oil production in Türkiye has remained relatively limited compared to other energy sources. The country's oil reserves are limited, leading to significant imports to meet its energy requirements. Recognizing the need to minimize dependency on oil, the government has launched policies to diversify energy sources, boost the use of renewable energy, and improve energy efficiency. These measures attempt to improve energy security, reduce trade imbalances, and promote long-term economic growth.

GDP growth in Türkiye from 1971 to 2020 has been closely linked to the country's energy sector and overall economic development. The energy sector, encompassing coal, natural gas, hydraulic power, and oil production, has played a significant role in supporting various industries, including manufacturing, construction, and transportation. However, dependence on fossil fuel imports has presented challenges, impacting energy prices and trade deficits. Türkiye has concentrated on diversifying its energy sources, encouraging the use of renewable energy, and improving energy efficiency in order to achieve long-term economic growth.

In conclusion, Türkiye's energy sector and GDP have experienced substantial transformations between 1971 and 2020. The country has witnessed growth in coal, natural gas, and hydraulic power production, while oil production has remained relatively modest. The government's emphasis on diversification, investment in renewable energy sources, and promotion of energy efficiency demonstrates a commitment to sustainable economic growth and energy security. Moving forward, continued efforts to strike a balance between economic development, energy sustainability, and environmental protection will be crucial for Türkiye's future energy landscape and economic prosperity.

The BSTS model is a flexible and powerful Bayesian method used for forecasting time series data. It is based on a framework called the state space model, which allows for the inclusion of various components such as trend, seasonality, and regression effects. By employing a Bayesian approach, the BSTS model can incorporate prior information and estimate posterior distributions, providing a comprehensive understanding of the uncertainty associated with the forecasts. This model is well-suited for handling complex and nonlinear relationships in the data, making it particularly useful when dealing with irregular patterns and structural breaks.

The ARIMA model, on the other hand, is a classic and extensively used statistical model for time series forecasting. It is based on the premise that future values of a time series are linearly related to previous values and mistakes. ARIMA models are effective in capturing short-term dependencies and autocorrelation in the data. They are relatively straightforward to implement and are suitable for stationary time series data, especially when the underlying patterns exhibit consistency or trends.

When comparing the two models, the BSTS model offers greater flexibility and the ability to handle complex time series patterns. It can capture both short-term and long-term dependence, as well as seasonality and other pertinent aspects. Furthermore, its Bayesian framework allows for a more thorough comprehension of uncertainty as well as the inclusion of previous information.

On the other hand, ARIMA models are simpler to implement and are better suited for data with clear patterns and relatively simple dependencies. They also provide useful diagnostic tools for analyzing residuals and assessing the adequacy of the model.

Finally, the decision between the BSTS and ARIMA models is determined by the properties of the time series data as well as the objectives of the forecasting assignment. Factors such as the complexity and nature of the data, the presence of trends and seasonality, and the availability of prior information should all be considered when determining which model is most appropriate.

Türkiye possesses significant coal reserves, primarily consisting of lower-grade lignite. Over the years, coal production in Türkiye has steadily increased to meet the demand for electricity generation and industrial usage. However, the coal mining industry has faced challenges related to safety and environmental concerns. To address these issues, efforts have been made to enhance mining practices and enforce stricter regulations. To reduce environmental consequences, Türkiye has also invested in clean coal technologies such as high-efficiency, low-emission (HELE) coal-fired power stations and carbon capture and storage (CCS) systems.

In contrast, Türkiye's natural gas production has been relatively limited compared to its consumption, leading to a heavy reliance on imports. The country has diversified its natural gas supply sources, importing from countries including Russia, Azerbaijan, Iran, and Qatar. Natural gas pipeline projects, such as the Trans-anatolian natural gas pipeline (TANAP) and TurkStream, has boosted Türkiye's energy security and access to natural gas deposits. Türkiye's extensive water resources make it an ideal location for hydroelectric power development. Large-scale hydroelectric power plant construction, particularly along important rivers such as the Euphrates and Tigris, has greatly increased Türkiye's renewable energy capacity.

The Turkish government has actively promoted the expansion of hydraulic power as part of its renewable energy strategy, aiming to reduce greenhouse gas emissions and enhance energy sustainability. Given the limited domestic oil reserves, Türkiye's production has been insufficient to meet its demand. As a result, the government has prioritized diversifying the energy mix and lowering reliance on oil by increasing the use of renewable energy sources, improving energy efficiency, and encouraging investment in alternative energy technology. Türkiye has also explored energy agreements and projects with neighboring nations in order to improve energy security and get access to additional oil resources.

Türkiye's GDP has experienced noteworthy growth, driven by various sectors such as manufacturing, construction, and services. The energy sector, encompassing coal, natural gas, hydraulic power, and oil production, has played a crucial role in supporting industrial activities and overall economic development. The government has implemented policies and reforms to stimulate economic growth, attract foreign investments, and promote sustainable development. Türkiye's strategic geographic location, acting as a bridge between Europe and Asia, has also contributed to its economic expansion and trade opportunities.

Türkiye should continue to diversify its energy sources by investing more in renewable energy sources such as wind, solar, and geothermal energy. To exploit the country's renewable energy potential, the government should give support for the development of renewable energy projects, especially small-scale hydropower facilities.

Türkiye should prioritize energy efficiency measures across all sectors by enforcing stricter regulations and promoting energy-efficient technologies. Raising awareness about energy conservation can also contribute to reducing energy consumption.

The government should enforce stringent environmental regulations and standards in the coal and oil sectors to mitigate pollution and reduce the carbon footprint. Investment in clean coal technologies and the transition to cleaner fuels and technologies in the transportation sector can help reduce greenhouse gas emissions.

Türkiye should focus on diversifying its natural gas supply sources by exploring new partnerships and energy agreements. Enhancing domestic oil production and exploring untapped oil reserves can help reduce reliance on imports and improve energy security.

The government should prioritize economic growth and sustainable development by promoting investments in sectors that contribute to GDP growth, job creation, and technological advancement. Emphasizing innovation and research and development in the energy sector can lead to improved energy efficiency and the creation of a skilled workforce.

Türkiye should actively engage in international energy partnerships and collaborations to exchange knowledge and expertise. Collaborating with neighboring countries on energy projects can enhance energy connectivity, promote regional stability, and support economic integration.

Türkiye's energy and economic policies should be aligned with the United Nations Sustainable Development Goals (SDGs) in order to encourage clean energy, long-term economic growth, climate action, and responsible consumption and production.

Türkiye should invest in modernizing and expanding its energy infrastructure to enhance the efficiency and reliability of energy supply. Developing a smart grid system can improve grid stability and facilitate the integration of decentralized energy generation.

Türkiye should continue efforts to liberalize and deregulate the energy market to promote competition and attract private investments. Implementing market-based mechanisms such as carbon pricing or emissions trading can incentivize the reduction of greenhouse gas emissions.

The government should provide long-term and stable incentives to attract private investments in renewable energy projects and drive down the costs of renewable energy generation. Strengthening renewable energy research and development initiatives may stimulate innovation and enhance the efficiency of renewable technology.

Türkiye should prioritize investments in education, vocational training, and capacity-building programs to cultivate a skilled workforce in the energy sector. Collaboration among universities, research institutions, and industry stakeholders is essential to align education and training initiatives with the specific requirements of the energy sector.

Türkiye should actively participate in international energy organizations and initiatives to access technical expertise and contribute to international energy security and sustainability efforts. Strengthening diplomatic ties and expanding energy cooperation agreements can enhance Türkiye's energy diversification strategy.

The government should prioritize raising public awareness about sustainable energy practices and engaging citizens and local communities in decision-making processes regarding energy projects.

Establishing robust monitoring and evaluation mechanisms for energy projects and policies can enable evidence-based decision-making and ensure accountability and transparency in the energy sector.

REFERENCES

Akpınar, A., Kömürcü, M.İ., Kankal, M. (2011). Development of hydropower energy in Türkiye: The case of coruh river basin. *Renewable and Sustainable Energy Reviews*, 15(2), 1201-1209.

Algül, Y., Vedat, K. (2021). Comparison of employment impacts of renewable and fossil energy based electricity sectors: The case of Türkiye . *Selçuk Üniversitesi Sosyal Bilimler Meslek Yüksekokulu Dergisi*, 24(2), 421-439.

Almarashi, A.M., Khan, K. (2020). Bayesian structural time series. *Nanoscience and Nanotechnology Letters*, 12(1), 54-61.

AL-Moders, A.H., Kadhim, T.H. (2021). Bayesian structural time series for forecasting oil prices. *Ibn AL-Haitham Journal For Pure and Applied Sciences*, 34(2), 100-107.

Amini, S.M., Parmeter, C.F. (2011). Bayesian model averaging in R. *Journal of Economic and Social Measurement*, 36(4), 253-287.

Artar, O.K., Uca, N., Taşçı, M.E. (2016). The impact of the airline freight transportation on GDP in Türkiye. *Journal of International Trade, Logistics and Law*, 2(2), 143-148.

Aydın, İ., Karat, H.İ., Koçak, A. (2005). Curie-point depth map of Türkiye . *Geophysical Journal International*, 162(2), 633-640.

Bağci, E., Diğrak, M. (1996). Antimicrobial activity of essential oils of some Abies (Fir) species from Türkiye . *Flavour and Fragrance Journal*, 11(4), 251-256.

Balat, H., Öz, C. (2008). Challenges and opportunities for bio-diesel production in Türkiye . *Energy Exploration and Exploitation*, 26(5), 327-346.

Balat, M. (2004). Oil and natural gas transport systems, trade and consumption trends in Türkiye. *Energy Exploration and Exploitation*, 22(3), 207-216.

Baser, K.H.C., Kirimer, N.E.Ş.E., Tümen, G. (1998). Pulegone-rich essential oils of Türkiye. *Journal of Essential Oil Research*, 10(1), 1-8.

Bilen, K., Ozyurt, O., Bakırcı, K., Karslı, S., Erdogan, S., Yılmaz, M., Comaklı, O. (2008). Energy production, consumption, and environmental pollution for sustainable development: A case study in Türkiye . *Renewable and Sustainable Energy Reviews*, 12(6), 1529-1561.

Biswas, R., Bhattacharyya, B. (2013). ARIMA modeling to forecast area and production of rice in west bengal. *Journal of Crop and Weed*, 9(2), 26-31.

Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52(3), 345-370.

Brooks, S.P., Giudici, P., Philippe, A. (2003). Nonparametric convergence assessment for MCMC model selection. *Journal of Computational and Graphical Statistics*, 12(1), 1-22.

Cevik, N.K., Cevik, E.I., Dibooglu, S. (2020). Oil prices, stock market returns and volatility spillovers: Evidence from Türkiye . *Journal of Policy Modeling*, 42(3), 597-614.

Chandra, P. (1993). A dynamic distribution model with warehouse and customer replenishment requirements. *Journal of the Operational Research Society*, 44(7), 681-692.

Chen, Y., Li, Y., Wu, H., Liang, L. (2014). Data envelopment analysis with missing data: a multiple linear regression analysis approach. *International Journal of Information Technology and Decision Making*, 13(01), 137-153.

Cheung, Y.W., Lai, K.S. (1995). Lag order and critical values of the augmented Dickey–Fuller test. *Journal of Business and Economic Statistics*, 13(3), 277-280.

Cinar, D., Kayakutlu, G., Daim, T. (2010). Development of future energy scenarios with intelligent algorithms: Case of hydro in Türkiye . *Energy*, *35*(4), 1724-1729.

Closas, P., Fernández-Prades, C., Fernández-Rubio, J.A. (2007). Maximum likelihood estimation of position in GNSS. *IEEE Signal Processing Letters*, 14(5), 359-362.

Curtis, A., Lomax, A. (2001). Prior information, sampling distributions, and the curse of dimensionality. *Geophysic*s, 66(2), 372-378.

Çakmakce, M., Kayaalp, N., Koyuncu, I. (2008). Desalination of produced water from oil production fields by membrane processes. *Desalination*, 222(1-3), 176-186.

Demirbaş, A. (2001). Energy balance, energy sources, energy policy, future developments and energy investments in Türkiye. *Energy Conversion and Managemen*t, 42(10), 1239-1258.

Demirbaş, A. (2003). Energy and environmental issues relating to greenhouse gas emissions in Türkiye . *Energy Conversion and Management*, 44(1), 203-213.

Derisavi, S., Hermanns, H., Sanders, W.H. (2003). Optimal state-space lumping in markov chains. *Information Processing Letters*, 87(6), 309-315.

Erdogdu, E. (2011) . An analysis of turkish hydropower policy. *Renewable and Sustainable Energy Reviews*, 15(1), 689-696.

Fidan, H. (2010). Turkish foreign policy towards central asia. *Journal of Balkan and Near Eastern Studies*, 12(1), 109-121.

Gallagher, K., Charvin, K., Nielsen, S., Sambridge, M., Stephenson, J. (2009). markov chain monte carlo (MCMC) sampling methods to determine optimal models, model resolution and model choice for earth science problems. *Marine and Petroleum Geology*, 26(4), 525-535.

Geweke, J. (2007). Bayesian model comparison and validation. *American Economic Review*, 97(2), 60-64.

Gharbi, M., Quenel, P., Gustave, J., Cassadou, S., Ruche, G.L., Girdary, L., Marrama, L. (2011) . Time series analysis of dengue incidence in guadeloupe, french west indies: forecasting models using climate variables as predictors. *BMC Infectious Diseases*, 11(1), 1-13.

Heilbronner, R.P. (1992). The autocorrelation function: an image processing tool for fabric analysis. *Tectonophysics*, 212(3-4), 351-370.

Hepbasli, A. (2004). Coal as an energy source in Türkiye. *Energy Sources*, 26(1), 55-63.

Hodson, T.O. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geoscientific Model Development*, 15(14), 5481-5487.

Hooten, M.B., Hobbs, N.T. (2015). A guide to bayesian model selection for ecologists. *Ecological Monographs*, 85(1), 3-28.

Hossain, M.M., Abdulla, F. (2015). Forecasting the sugarcane production in bangladesh by ARIMA model. *Journal of Statistics Applications and Probability*, 4(2), 297-303.

Ishak, S., Al-Deek, H. (2002). Performance evaluation of short-term time-series traffic prediction model. *Journal of Transportation Engineering*, 128(6), 490-498.

Jahanshahi, A., Jahanianfard, D., Mostafaie, A., Kamali, M. (2019). An auto regressive iIntegrated moving average (ARIMA) model for prediction of energy consumption by household sector in euro area. *AIMS Energy*, 7(2), 151-1

Jonek-Kowalska, I. (2019). Transformation of energy balances with dominant coal consumption in European economies and Türkıye in the years 1990–2017. *Oeconomia Copernicana*, 10(4), 627-647.

Kaplan, M., Ozturk, I., Kalyoncu, H. (2011). Energy consumption and economic growth in Türkiye: cointegration and causality analysis. *Romanian Journal of Economic Forecasting*, 2(31), 31-41.

Karayigit, A.I., Spears, D.A., Booth, C.A. (2000). Antimony and arsenic anomalies in the coal seams from the gokler coalfield, gediz, Türkiye. *International Journal of Coal Geology*, 44(1),1-17.

Karklin, Y., Lewicki, M.S. (2005). A hierarchical bayesian model for learning nonlinear statistical regularities in nonstationary natural signals. *Neural Computation*, 17(2), 397-423.

Kass, R.E., Carlin, B.P., Gelman, A., Neal, R.M. (1998). Markov chain monte carlo in practice: A roundtable discussion. *The American Statistician*, 52(2), 93-100.

Kaya, D. (2006). Renewable energy policies in Türkiye . *Renewable and Sustainable Energy Reviews*, 10(2), 152-163.

Kaygusuz, K. (2002). Oil and gas production and consumption in Türkiye. *Energy Exploration and Exploitation*, 20(1), 37-50.

Kaygusuz, K. (2002). Sustainable development of hydropower and biomass energy in Türkiye . *Energy Conversion and Management*, 43(8), 1099-1120.

Kaygusuz, K. (2003). Energy policy and climate change in Türkiye . *Energy Conversion and Management*, 44(10), 1671-1688.

Kaygusuz, K. (2009). Hydropower in Türkiye: The sustainable energy future. *Energy Sources*, 4(1), 34-47.

Kilic, A.M. (2005). Major utilization of natural gas in Türkiye. *Energy Exploration and Exploitation*, 23(2), 125-140.

Kiliç, F.Ç., Kaya, D. (2007). Energy production, consumption, policies, and recent developments in Türkiye. *Renewable and Sustainable Energy Reviews*, 11(6), 1312-1320.

Klugkist, I., Kato, B., Hoijtink, H. (2005). Bayesian model selection using encompassing priors. *Statistica Neerlandica*, *59*(1), 57-69.

Kolarik, T., Rudorfer, G. (1994). Time series forecasting using neural networks. *ACM Sigapl Apl Quote Quad*, 25(1), 86-94.

Levy, R., 2016. Advances in bayesian modeling in educational research. *Educational Psychologist*, 51(3-4), 368-380.

Lise, W., Van Montfort, K. (2007). Energy consumption and GDP in Türkiye: Is there a co- integration relationship?. *Energy Economics*, 29(6), 1166-1178.

Makridakis, S., Hibon, M. (1997). ARMA models and the Box–Jenkins methodology. *Journal of Forecasting*, 16(3), 147-163.
McKenzie, J. (2011). Mean absolute percentage error and bias in economic forecasting. *Economics Letters*, 113(3), 259-262.
Modugno, M., Soybilgen, B., Yazgan, E. (2016). Nowcasting Turkish GDP and news decomposition. *International Journal of Forecasting*, 32(4), 1369-1384.
Mohamed, J. (2020). Time series modeling and forecasting of somaliland consumer price index: A comparison of ARIMA and regression with ARIMA errors. *American Journal of Theoretical and Applied Statistics*, 9(4), 143-53.
Montgomery, J.M., Nyhan, B. (2010). Bayesian model averaging: Theoretical developments and practical applications. *Political Analysis*, 18(2), 245-270.
Mourtgos, S.M., Adams, I.T. (2021). Covıd-19 vaccine program eliminates law enforcement workforce infections: A bayesian structural time series analysis. *Police Practice and Research*, 22(5), 1557-1565.
Newbold, P. (1983). ARIMA model building and the time series analysis approach to forecasting. *Journal of Forecasting*, 2(1), 23-35.
Oğulata, R.T. (2003). Energy sector and wind energy potential in Türkiye . *Renewable and Sustainable Energy Reviews*, 7(6), 469-484.
Özbayoğlu, G., Mamurekli, M. (1994). Super-clean coal production from turkish bituminous coal. *Fuel*, 73(7), 1221-1223.
Pappas, S.S., Ekonomou, L., Karampelas, P., Karamousantas, D.C., Katsikas, S.K., Chatzarakis, G.E.,Skafidas, P.D. (2010). Electricity demand load forecasting of the hellenic power system using an ARMA model. *Electric Power Systems Research*, 80(3), 256-264.
Pérez, J.M., Berger, J.O. (2002). Expected- posterior prior distributions for model selection. *Biometrika*, 89(3), 491-512.
Piepho, H.P. (2019). A coefficient of determination R^2 for generalized linear mixed models. *Biometrical Journal*, 61(4), 860-872.
Pinilla, J., Negrín, M., González-López-Valcárcel, B., Vázquez-Polo, F.J. (2018). Using a bayesian structural time–series model to infer the causal impact on cigarette sales of partial and total bans on public smoking. *Jahrbücher Für Nationalökonomie und Statistik*, 238(5), 423-439.
Poyser, O. (2019). Exploring the dynamics of bitcoin's price: a bayesian structural time series approach. *Eurasian Economic Review*, 9(1), 29-60.
Querol, X., Alastuey, A., Plana, F., Lopez-Soler, A., Tuncali, E., Toprak, S., Ocakoglu, F., Koker, A. (1999). Coal geology and coal quality of the miocene mugla basin, southwestern anatolia, Türkiye . *International Journal of Coal Geology*, 41(4), 311-332.
Ray, P., Ganguli, B., Chakrabarti, A. (2021) . A hybrid approach of bayesian structural time series with lstm to identify the influence of news sentiment on short-term forecasting of stock price. *IEEE Transactions on Computational Social Systems*, 8(5), 1153-1162.
Saidur, R., Islam, M.R., Rahim, N.A., Solangi, K.H. (2010). A review on global wind energy policy. *Renewable and Sustainable Energy Reviews*, 14(7), 1744-1762.

Salvarli, H. (2006). Some aspects on hydraulic energy and environment in Türkiye. *Energy Policy*, 34(18), 3398-3401.

Sari, R., Soytas, U., (2004). Disaggregate energy consumption, employment and income in Türkiye . *Energy Economics*, 26(3), 335-344.

Sarpong, S.A. (2013). Modeling and forecasting maternal mortality; an application of ARIMA models. *International Journal of Applied*, *3*(1), 19-28.

Sensogut, C., Oren, O. (2009). Coal production and energy fact in Türkiye . *Energy Sources*, 4(3), 239-246.

Sheridan, R.P. (2013). Time-split cross-validation as a method for estimating the goodness of prospective prediction. *Journal of Chemical Information and Modeling*, 53(4), 783-790.

Shi, J., Guo, J., Zheng, S. (2012). Evaluation of hybrid forecasting approaches for wind speed and power generation time series. *Renewable and Sustainable Energy Reviews*, 16(5), 3471-3480.

Shrestha, A., Ghimire, B., Gonzalez-Longatt, F. (2021). A bayesian model to forecast the time series kinetic energy data for a power system. *Energies*, 14(11), 3299.

Sözen, A., Arcaklioglu, E. (2007). Prediction of net energy consumption based on economic indicators (GNP and GDP) in Türkiye. *Energy Policy,* 35(10), 4981-4992.

Spedding, T.A., Chan, K.K. (2000). Forecasting demand and inventory management using bayesian time series. *Integrated Manufacturing Systems*, 11(5), 331-339.

Suleman, N., Sarpong, S. (2012). Forecasting milled rice production in Ghana using Box-Jenkins approach. *International Journal of Agricultural Management and Development*, 2(2), 79-84.

Ticknor, J.L. (2013). A bayesian regularized artificial neural network for stock market forecasting. *Expert Systems with Applications*, 40(14), 5501-5506.

Toivonen, H.T., Mannila, H., Korhola, A., Olander, H. (2001). Applying bayesian statistics to organism- based environmental reconstruction. *Ecological Applications*, 11(2), 618-630.

Toklu, E. (2017). Biomass energy potential and utilization in Türkiye . *Renewable Energy*, 107, 235-244.

Toprak, S. (2009). Petrographic properties of major coal seams in Türkiye and their formation. *International Journal of Coal Geology*, 78(4), 263-275.

Ulutaş, B.H. (2005). Determination of the appropriate energy policy for Türkiye. *Energy*, 30(7), 1146-1161.

Umucalilar, H.D., Coşkun, B., Gülşen, N. (2002). In situ rumen degradation and in vitro gas production of some selected grains from Türkiye. *Journal of Animal Physiology and Animal Nutrition*, 86(9- 10), 288-297.

Vermaak, J., Andrieu, C., Doucet, A., Godsill, S.J. (2002). Particle methods for bayesian modeling and enhancement of speech signals. *IEEE Transactions on Speech and Audio Processing*, 10(3), 173-185.

Von Toussaint, U. (2011). Bayesian inference in physics. *Reviews of Modern Physics*, 83(3), 943.

Wang, W., Wong, A.K. (2002). Autoregressive model-based gear fault diagnosis. *J. Vib. Acoust* , 124(2), 172-179.

Warner, J.F. (2008). Contested hydrohegemony: Hydraulic control and security in Türkiye. *Water Alternatives*, 1(2), 271-288.

Western, B., Kleykamp, M. (2004). A bayesian change point model for historical time series analysis. *Political Analysis*, 12(4), 354-374.

Willmott, C.J., Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30(1), 79-82.

Yalta, A.T. (2011). Analyzing energy consumption and GDP nexus using maximum entropy bootstrap: the case of Türkiye . *Energy Economics*, 33(3), 453-460.

Yılmaz, A.O. (2008). Renewable energy and coal use in Türkiye . *Renewable Energy*, 33(5), 950-959.

Yılmaz, A.O. (2009). Present coal potential of Türkiye and coal usage in electricity generation. *Energy Sources*, 4(2), 135-144.

Yılmaz, A.O., Uslu, T. (2007). The role of coal in energy production—consumption and sustainable development of Türkiye. *Energy Policy*, 35(2), 1117-1128.

Yuksek, O., Komurcu, M.I., Yuksel, I., Kaygusuz, K. (2006). The role of hydropower in meeting Türkiye's electric energy demand. *Energy Policy*, 34(17), 3093-3103.

Yüksel, I. (2008). Hydropower in Türkiye for a clean and sustainable energy future. *Renewable and Sustainable Energy Reviews*, 12(6), 1622-1640.

Zaim, K.K. (1999). Modified GDP through health cost analysis of air pollution: the case of Türkiye . *Environmental Management*, 23(2), 271-277.

Zeger, S.L., Karim, M.R. (1991). Generalized linear models with random effects; a gibbs sampling approach. *Journal of the American Statistical Association*, 86(413), 79-86.

EXTENDED TURKISH SUMMARY

(GENİŞLETILMIŞ TÜKÇE ÖZET)

ENERJİ ÜRETİMİNE İLİŞKİN ZAMAN SERİSİ VERİLERİNİN BAYESİAN VE KLASİK MODELLERİ KARŞILAŞTIRARAK TÜRKİYE'DE TAHMİN EDİLMESİ

HASSOO, Amir KHaleel
Doktora Tezi, İstatistik Anabilim Dalı
Danışman: Assoc. Prof. Dr. Şakir İşleyen
Temmuz 2023, 154 sayfa

Bu çalışmada, Bayesian yapısal zaman serisi (BSTS) ve Oto-regressif entegre hareketli ortalama (ARIMA) olmak üzere iki tahmin modeli, 1971-2020 yılları arasında Türkiye'nin enerji üretimi verilerini tahmin etmek amacıyla karşılaştırılmıştır. Modeller kömür, gaz, hidrolik ve petrol üretimi ile GSMH gibi çeşitli enerji kaynaklarına ve Dünya Bankası veritabanından elde edilen verilere uygulanmıştır. Temel amaç, bu modellerin enerji üretimi trendlerini tahmin etme doğruluğunu değerlendirmektir. Güvenilir ve kapsamlı sonuçlar elde etmek için analiz ve modelleme işlemleri R ve SPSS yazılımları kullanılarak gerçekleştirilmiştir. Bu karşılaştırmada MAPE, MAE, RMSE ve R^2 gibi değerlendirme metrikleri de kullanılmıştır. BSTS modelleri, önceki bilgilerin dahil edilmesine ve belirsizliklerin nicelendirilmesine imkan sağlayan Bayesian bir çerçeve içermektedir ve zaman serisi tahmininde yaygın olarak kullanılan geleneksel ARIMA modelleri ile karşılaştırılmıştır. Modellerin doğruluğunu değerlendirmek için veri seti eğitim ve test alt kümelerine ayrılmış ve böylece model hatalarının değerlendirilmesi mümkün olmuştur. Sonuçlar, BSTS modelinin Türkiye'nin enerji üretimi zaman serisi verilerini tahmin etmede ARIMA modelinden daha iyi performans gösterdiğini göstermiştir. BSTS modelinin benimsediği Bayesian yaklaşım, enerji üretimi dinamiklerindeki doğal belirsizlikler ve karmaşıklıkları dikkate alarak, ARIMA modelinin Box-Jenkins yaklaşımına kıyasla daha fazla güvenilirlik ve doğruluk sağlamıştır. Sonuç olarak, BSTS modeli 2021-2028 yılları için enerji üretimini tahmin etmek amacıyla seçilmiştir. Ayrıca, bu çalışma Türkiye'deki GSMH'yi etkileyen faktörleri incelemek için çoklu doğrusal regresyon analizinden yararlanarak mevcut literatüre katkı sağlamaktadır.

Türkiye, son elli yılda kömür, gaz, hidrolik güç ve petrol gibi çeşitli enerji kaynaklarıyla önemli bir rol oynayarak enerji sektöründe önemli bir büyüme ve gelişme yaşamıştır. Enerji üretimi ile Gayri Safi Yurtiçi Hasıla (GSYİH) arasındaki ilişkinin desenlerini ve eğilimlerini anlamak, politika yapıcılar, enerji planlamacılar ve araştırmacılar için hayati öneme sahiptir. Bu araştırma, Bayesian ve klasik modelleri karşılaştırarak Türkiye'nin 1971'den 2020'ye kadar enerji üretimi ve GSYİH üzerine zaman serisi verilerini tahmin etmeyi amaçlamaktadır. Bu çalışmada, Bayesian modeller, özellikle Bayesian yapısal zaman serisi (BSTS), Türkiye'de enerji üretimi ve GSYİH üzerine zaman serisi verilerini tahmin etmek için Otoregresif entegre hareketli ortalama (ARIMA) gibi klasik modellerle karşılaştırılmaktadır (Hepbaşlı, 2004). Enerji üretiminin ve GSYİH ile ilişkisinin doğru bir şekilde tahmin edilmesi, ekonomik kalkınma ve sürdürülebilirlik sağlamak için etkili planlama ve politika yapma açısından önemlidir. Zaman serisi analizi, enerji üretimi ve GSYİH verilerini inceleyerek ve tahmin ederek, enerji üretimini etkileyen faktörlere ve ekonomi üzerindeki etkisine dair içgörüler sağlamak için etkili bir yaklaşım olarak kanıtlanmıştır. Bayesian modeller, önceden bilgiyi içermesi, belirsizliği nicelendirmesi ve verilerdeki karmaşıklıkları ve belirsizlikleri dikkate alması gibi avantajlar sunmaktadır. Araştırma amacı, Türkiye'de enerji üretimi ve GSYİH üzerine zaman serisi verilerini tahmin etme konusunda Bayesian ve klasik modellerin performansını ve doğruluğunu değerlendirmektir. Bu çalışmanın bulguları, enerji sektöründeki zaman serisi analiziyle ilgili mevcut literatüre katkı sağlayacak ve enerji planlaması ve ekonomik tahminle ilgilenen politika yapıcılar ve araştırmacılara değerli içgörüler sunacaktır. Bu önsöz, sonraki bölümleri tanıtır; bu bölümler, yöntemoloji, veri analizi, modelleme teknikleri ve sonuçların yorumunu ele alacaktır. Çalışma, her bir değişken için enerji üretimini tahmin etme konusunda Bayesian ve klasik modellerin performansını inceleyerek veri toplama, ön işleme ve modelleme süreçlerini içermektedir. Ayrıca, bulguların olası etkileri, potansiyel çalışma sınırlamaları ve zaman serisi analizi ve enerji üretimi tahmininde gelecekteki araştırma yönergeleri tartışılmaktadır. Türkiye'nin enerji üretim sektörü, verimlilik, sürdürülebilirlik ve güvenilirlik üzerinde etkisi olan çeşitli zorluklarla karşı karşıyadır. Bu zorluklardan biri, Türkiye'nin kömür, petrol, doğal gaz gibi fosil yakıtlar yanında güneş, rüzgar, hidro ve jeotermal gibi yenilenebilir enerji kaynakları gibi çeşitli enerji kaynaklarına yoğun bir şekilde bağımlı olmasıdır. Bu çeşitlilik içeren kaynakların etkin bir şekilde yönetilmesi ve entegre edilmesi, dengeli ve sürdürülebilir bir enerji arzının sağlanması, fosil yakıtlara bağımlılığın azaltılması ve yenilenebilir enerjinin benimsenmesinin teşvik edilmesi gibi önemli bir zorluk oluşturmaktadır. Bir diğer zorluk, Türkiye'nin enerji ithalatına olan yüksek bağımlılığıdır, bu da güvenli ve güvenilir bir enerji arzının sağlanması sürecini karmaşıklaştırmaktadır. Ekonomik büyümenin enerji sürdürülebilirliği ve kaynak tahsisi üzerindeki etkilerini anlamak, bu hedefin önemli bir yönüdür. Bu araştırma hedefleri ele alınarak, çalışma Türkiye'de enerji üretimi tahmininin ve ekonomik faktörlerle olan ilişkisinin anlaşılmasına katkıda bulunmayı amaçlamaktadır. Bu bilgi, karar vericileri ve paydaşları sürdürülebilir enerji planlaması ve kaynak tahsisi için etkili stratejiler geliştirmede yönlendirebilir.

Araştırma, gerçek hayattan elde edilen ikincil verilerle bir yıllık süre boyunca Flying Cement hisse senedi fiyatlarını inceledi. İstatistiksel sonuçlar elde etmek için, Kalman filtresi ve MCMC gibi benzetim yaklaşımlarını kullandı. Araştırmanın odak noktası hisse senedi fiyatları olsa da, aynı BSTS yöntemi karmaşık mühendislik süreçlerine de kurşun süreleri ile uygulanabilir. BSTS yöntemini klasik bir yöntem olan ARIMA ile karşılaştırmak için ARIMA yaklaşımı kullanıldı. Bayesian posterior örnekleme dağılımlarını elde etmek için R yazılımının BSTS paketi kullanıldı. Gerçek bir veri kümesine dört BSTS modeli uygulandı ve BSTS yaklaşımının nasıl çalıştığı gösterildi. Tahmin grafiği ve MAPE kullanılarak çeşitli modellerin tahmin doğruluğu değerlendirildi. Araştırmanın amacı, araştırmacılar ve uygulayıcılar tarafından kolayca tekrarlanabilir bir basit teknik geliştirmekti. Sonuçlar, kısa vadeli tahminler için ARIMA ve BSTS'nin benzer şekilde performans gösterdiğini gösterdi. Ancak sonuçlara dayanarak, uzun vadeli tahminler için BSTS'nin yerel düzeyle en iyi seçenek olarak kabul edildi. Türkiye'nin kömür, gaz, hidrolik ve petrol gibi enerji üretim verileri ile 1971-2020 dönemini kapsayan Gayri Safi Yurtiçi Hasıla (GSYİH) verileri, Dünya Bankası web sitesinden toplandı. Toplanan veriler, eksik değerleri, aykırı değerleri ve tutarsızlıkları belirlemek ve ele almak için kapsamlı bir şekilde incelendi. Gerektiğinde eksik değerlerin imalatı veya çıkarılması gibi çeşitli veri ön işleme teknikleri kullanıldı. Karşılaştırmak için iki model seçildi: bir Bayesian modeli ve bir klasik model. Bayesian model, BSTS yaklaşımını kullandı, klasik model ise Box-Jenkins metodolojisine dayanan ARIMA modelini kullandı. Sağlam ve kapsamlı sonuçlar elde etmek için veri analizi ve modellemede R ve SPSS yazılımları kullanıldı. Bayesian çıkarım teknikleri kullanılarak uygulanan BSTS modeli, model parametreleri için uygun önceden dağılımların belirlenmesini gerektirdi. Posterior dağılımın tahmin edilmesi için MCMC örnekleme yöntemleri kullanıldı. BSTS modeli, enerji üretim zaman serisi verilerini (kömür, gaz, hidrolik, petrol) ve GSYİH'yi girdi olarak kullanarak, Türkiye'deki enerji üretimi ve GSYİH ile ilgili temel eğilimleri, mevsimsellikleri ve belirsizlikleri yakaladı. Box-Jenkins yaklaşımını takip eden ARIMA modeli, her enerji üretim zaman serisi (kömür, gaz, hidrolik, petrol) ve GSYİH için uygun sırasını (p, d, q) belirledi. Model parametrelerinin tahmin edilmesinde en büyük olabilirlik tahminlemesi kullanıldı ve gelecekteki değerleri tahmin etmek için tarihsel eğilimler kullanıldı. Bayesian ve klasik modellerin doğruluğu ve performansı, MAE, RMSE, MAPE ve R^2 gibi değerlendirme metrikleri kullanılarak değerlendirildi. Her bir modelden elde edilen tahminler, veri kümesindeki ilgili yıllar için gerçek enerji üretimi ve GSYİH değerleriyle karşılaştırıldı. Elde edilen bulgular yorumlandı ve Türkiye'deki enerji üretimi ve GSYİH zaman serisi verilerini tahmin etme konusundaki modellerin performansı hakkında sonuçlar çıkarıldı. Bir modelin diğerinden üstün performansına katkıda bulunan faktörler belirlendi ve tartışıldı. Seçilen modellerin sınırlamaları, veri bulunabilirliği ve diğer etkili faktörler kabul edildi ve tartışıldı. Elde edilen sonuçlar ve bilgiler doğrultusunda, gelecekteki araştırmalar, modelleme tekniklerindeki iyileştirmeler veya politika önerileri sağlanabilir. Eğitim ve test doğrulaması, bir modelin performansını ve genelleme yeteneğini değerlendirmek için makine öğrenimi ve istatistiksel modellemede yaygın bir uygulamadır. Bu işlem, mevcut veri kümesini iki ayrı alt kümeye bölmeyi içerir: eğitim kümesi ve test kümesi.

Veri Bölme: Mevcut veri kümesi rastgele şekilde eğitim kümesi ve test kümesi olarak ayrılır. Tipik olarak, verinin yaklaşık %70-80'i eğitim için, geri kalan %20-30'u ise test için kullanılır. Eğitim Kümesi: Eğitim kümesi, modelin eğitilmesi veya uygun hale getirilmesi için kullanılır. (1971-2012 yılları arası) Test Kümesi: Test kümesi, eğitilen modelin performansını değerlendirmek için kullanılır. (2013-2020 yılları arası). ARIMA (Oto-Regressif Entegre Hareketli Ortalama), geçmiş verilerde gözlemlenen desenlere dayanarak gelecekteki değerleri analiz etmek ve tahmin etmek için kullanılan popüler bir zaman serisi tahmin modelidir. ARIMA modeli, bir zaman serisinin temel yapısını yakalamak için oto-regresif (AR), fark alma (I) ve hareketli ortalama (MA) bileşenlerini birleştirir. ARIMA, Autoregressive Integrated Moving Average kelimelerinin kısaltması olan ve otoregresif (AR), fark alma (I) ve hareketli ortalama (MA) bileşenlerini içeren belirgin bir zaman serisi veri analizi modelidir. ARIMA modeli, serinin temel desenlerini ve dinamiklerini yakalamak için bu bileşenleri bir araya getirir. ARIMA modeli üç temel parametre ile karakterize edilir: p, d ve q. p parametresi, otoregresif (AR) bileşenin düzenini ifade eder ve modelde bağımlı değişkenin gecikmiş gözlemlerinin sayısını temsil eder. AR bileşeni, önceki değerleri değerlendirerek zaman serisinin mevcut değeri ile geçmiş değerleri arasındaki ilişkiyi yakalar. d parametresi, fark almanın düzenini ifade eder ve zaman serisini stationarity elde etmek için dönüştürmeyi içerir. Stationarity, verilerden eğilimleri ve mevsimselliği kaldırmayı ifade eder. d değeri, serinin stationarity'e ulaşması için kaç kez fark alınması gerektiğini belirtir. Hareketli ortalama (MA) bileşeninin düzeni, q parametresi tarafından temsil edilir.

Bu bölümde, araştırmacılar tarafından kullanılan birçok önemli performans göstergesini ele alacağız ve bunların altında yatan temel fikirleri açıklayacağız. Akaike Bilgi Kriteri (AIC), çeşitli istatistiksel modellerin kalitesini ve performansını karşılaştırmak için kullanılan istatistiksel bir ölçüttür. Arka dağıtım, bir modelin uyum kalitesi ile karmaşıklığı arasında bir uzlaşma sağlamanıza olanak tanır. AIC, en iyi modeli belirlemek için kullanılan istatistiksel bir metriktir. Ortalama Kare Hatanın Kökü (RMSE), özellikle regresyon analizinde, tahmine dayalı bir modelin performansını değerlendirmek için yaygın olarak kullanılan bir ölçümdür. Modelin tahmin edilen değerlerinin gerçek gözlenen değerlerle ne kadar uyumlu olduğunun bir ölçüsünü sağlar. Ortalama Mutlak Yüzde Hatası (MAPE), bir tahmin veya tahmin modelinin doğruluğunu belirlemek için kullanılan popüler bir istatistiktir. Hataların ortalama büyüklüğünü gerçek değerlere göre yüzde olarak hesaplar. Beklenen ve gerçek veriler arasındaki sapmaların ortalama büyüklüğünü hesaplayan bir istatistik ve makine öğrenimi ölçüsüdür. Bir tahmin modelinin doğruluğunu belirlemek için basit bir yöntem sağlar. Belirleme katsayısı, gerçek veriler ile model tahminleri arasındaki doğrusal ilişkiyi ölçer ve modelin uygunluğuna dair bir gösterge sunar. Ayrıca, bağımlı değişkendeki varyansın ne kadarının modelin bileşenleri tarafından açıklanabileceğini de gösterir. Korelasyon katsayısının karesidir. Türkiye'nin kömür, gaz, hidrolik, petrol ve gayri safi yurtiçi hasıla (GSYİH) yıllık üretimini tahmin etmek için BSTS ve ARIMA modelleri kullanıldıktan sonra, sonraki adım, sonuçları karşılaştırarak en iyi modeli belirlemek oldu. Önceki sonuçlardan elde edilebilecek bazı sonuçlar bulunmaktadır, bunların başlıcaları şunlardır: Çalışmanın üç ürün zaman serisi (kömür, gaz, hidrolik, petrol ve GSYİH) için BSTS modellerinin MAE (ortalama mutlak hata) değeri, ARIMA modellerinin MAE değerinden daha düşüktür, ARIMA modellerine göre daha iyi uyum sağlamaktadır. Model hataları açısından, iki modelin RMSE (kök ortalama karesel hata) değeri karşılaştırıldığında, BSTS modellerinin ARIMA modellerine göre daha az hata içerdiği açıktır. R^2, BSTS ve ARIMA modellerini karşılaştırmak için kullanılan başka bir ölçüttür. Bulgular açıkça gösteriyor ki BSTS modellerinin R^2 değeri, ARIMA modellerine kıyasla daha büyüktür. Model hataları açısından, iki modelin MAPE (ortalama mutlak yüzde hata) değeri karşılaştırıldığında, BSTS modellerinin ARIMA modellerine göre daha az hata içerdiği açıktır eğitim ve test verileri için. Veriyi değerlendirmek için iki model kullanıldı ve her iki modelin doğruluk karşılaştırma sonuçlarını gördükten sonra, Bayesian yapısal zaman serisi modelinin ARIMA modelinden daha iyi olduğu görüldü. Bundan sonra, 2028 için (BSTS) modeli için tahmin yapıldı. Tablo 4.54, 2028 yılı için tahmin edilen kömür üretimi değerlerini sunmaktadır. Veriler, 2028'e kadar olan yıllar için tahmin edilen kömür üretim rakamlarını ve o belirli yıl için tahmin edilen üretim değerini göstermektedir. Rakamlar, 2028 yılında tahmin edilen kömür üretiminin 29.66696 birime ulaşacağını göstermektedir. Tablo 4.55, 2028 yılı için fiili ve beklenen gaz üretim rakamlarını göstermektedir. Veri seti, 2028 öncesi yıllar için öngörülen gaz üretim değerlerinin yanı sıra cari yıl için tahmini üretim değerini içermektedir. Tabloya göre 2028 yılı için öngörülen gaz üretimi 53.79936 adettir. Tablo 4.56, 2028 yılı için fiili ve beklenen hidrolik üretim seviyelerini göstermektedir. 2028'e kadar olan yıllar için hidrolik üretim rakamlarını ve o yıl için beklenen üretim değerini tahmin etmektedir. Tabloya göre 2028 yılı için öngörülen hidrolik üretim 16.14761 adettir. Tablo 4.58, 2028 yılı için petrol değişkeni için tahmin edilen değerleri göstermektedir. Veriler, 2028'e giden önceki yıllar için tahmin edilen değerlerin yanı sıra söz konusu yıl için tahmin edilen değeri içerir. Tabloya göre 2028 yılı için öngörülen petrol üretim -0.6213833 adettir. Bu Tablo 4.59, 2021'den 2028'e kadar tahmin edilen GSYİH üretimi değerlerini sunmaktadır. Veriler, söz konusu zaman çerçevesi içindeki her yıl için tahmini GSYİH üretim rakamlarını içerir. Tablo, tahmin edilen GSYİH üretim değerlerinin yıllar içinde dalgalanmasının beklendiğini, en yüksek tahmin edilen değerin 4,570465'te 2024'te ve en düşük tahmin edilen değerin 4,147536'da 2028'de gerçekleşmesinin beklendiğini göstermektedir. Bu bulgulara dayanarak, Kömür, Gaz ve Hidrolik değişkenlerinin Türkiye'nin GSYİH üzerinde önemli bir etkisi olduğu çıkarımında bulunmak mümkündür çünkü bu değişkenlerin p-değerleri 0.05'ten küçüktür. Diğer taraftan, Petrol değişkeninin GSYİH üzerinde büyük bir etkisi olmadığı görünmektedir. Model istatistikleri, R-kare değerini 0.365 olarak gösterir, bu da dahil edilen değişkenlerin GSYİH'daki değişimin yaklaşık %36.5'ini açıkladığını gösterir. Değiştirilmiş R-kare değeri, modele dahil edilen serbestlik dereceleri dikkate alındığında 0.318'dir. 0.001 p-değeri ile F istatistiği, Türkiye'deki enerji üretimi değişkenleri ile GSYİH arasındaki ilişkiyi açıklamada tüm modelin istatistiksel olarak anlamlı olduğunu gösterir. Modellerin doğruluğunu değerlendirmek için veri seti eğitim ve test alt kümelerine ayrıldı ve bu sayede model hatalarının değerlendirilmesine olanak sağlandı. Bulgular, BSTS modelinin Türkiye'deki enerji üretimi zaman serisi verilerini tahmin etmede ARIMA modelinden daha iyi performans gösterdiğini gösterdi. BSTS modelinin kullandığı Bayesian yaklaşım, enerji üretim dinamiklerindeki doğal belirsizlikleri ve karmaşıklıkları hesaba katarak, ARIMA modelinin Box-Jenkins yaklaşımına göre daha fazla güvenilirlik ve doğruluk sergiledi. Sonuç olarak, BSTS modeli, 2021'den 2028'e kadar enerji üretimini tahmin etmek için seçildi. Ayrıca, bu çalışma Türkiye'deki GSYİH'yi etkileyen faktörleri incelemek için çoklu doğrusal regresyon analizini kullanarak mevcut literatüre katkı sağlamaktadır. (Kömür, gaz ve hidrolik) faktörlerinin Türkiye'deki GSYİH üzerindeki etkisi incelenmiştir. Türkiye'nin enerji sektörü ve GSYİH, 1971 ile 2020 arasında önemli dönüşümler yaşamıştır. Ülke, kömür, doğal gaz ve hidrolik güç üretiminde büyüme yaşarken, petrol üretimi nispeten sınırlı kalmıştır. Hükümetin çeşitlendirmeye ve yenilenebilir enerji kaynaklarına yatırıma vurgu yapması ve enerji verimliliğini teşvik etmesi, sürdürülebilir ekonomik büyüme ve enerji güvenliği konusunda bir taahhüt göstermektedir. Gelecekte, ekonomik gelişme, enerji sürdürülebilirliği ve çevre koruması arasında dengeyi sağlama konusunda devam eden çabalara önemli bir önem vermek, Türkiye'nin gelecekteki enerji manzarası ve ekonomik refahı için kritik olacaktır. BSTS modeli, zaman serisi verilerini tahmin etmek için kullanılan esnek ve güçlü bir Bayesian yöntemidir. Bu model, trend, mevsimsellik ve regresyon etkileri gibi çeşitli bileşenlerin dahil edilmesine izin veren bir çerçeve olan durum uzayı modeline dayanmaktadır. Bayesian yaklaşım kullanarak, BSTS modeli öncül bilgileri içerebilir ve son dağılımları tahmin ederek tahminlerle ilişkilendirilen belirsizlik hakkında kapsamlı bir anlayış sağlayabilir. Bu model, verilerdeki karmaşık ve doğrusal olmayan ilişkilerle başa çıkma açısından uygun olup, özellikle düzensiz desenler ve yapısal kırılmalarla uğraşırken son derece faydalıdır.

CURRICULUM VITAE

Personal Information

Name Surname : Amir Khaleel HASSOO

Education Information

Undergraduate

University : Salahaddin University
Faculty : College of administration and economics
Department : Statistics
Graduation Year : 2009

Master's degree

University : Kahramanmaraş Sütçü İmam Üniversitesi
Institute : The Institute of Bioengineering and science
Department : Statistics
Graduation Year : 2017

Publications

İşleyen, Ş., Hassoo, A.K. (2020). The impact of the university in increasing the book reading on students: case study soran university. *SIASAT*, 5(4), 51-58.

Hassoo, A.K., Akbay, C. (2020). Factors affecting motivation of employees in public sector in erbil, iraq an application of binary logistic regression. *Technium*, 7(1) , 321-329.

VAN YUZUNCU YIL UNIVERSITY
THE INSTITUTE OF NATURAL AND APPLIED SCIENCES
THESIS ORIGINALITY REPORT

Date: 28/7/2023

Thesis Title:

"ESTIMATING TIME SERIES DATA ON ENERGY PRODUCTION IN TÜRKİYE BY COMPARING BAYESIAN AND CLASSICAL MODELS"

The title of the mentioned thesis, above having total 154 pages with cover page, introduction, main parts and conclusion, has been checked for originality by Turnitin computer program on the date of 28/7/2023 and its detected similar rate was 6% (six percentage) according to the following specified filtering

Originality report rules:

- Excluding the Cover page,
- Excluding the Thanks,
- Excluding the Contents,
- Excluding the Symbols and Abbreviations,
- Excluding the Materials and Methods
- Excluding the Bibliography,
- Excluding the Citations,
- Excluding the publications obtained from the thesis,
- Excluding the text parts less than 7 words (Limit match size to 7 words)

I read the Thesis Originality Report Instruction of Van Yuzuncu Yil University for Obtaining and Using Similarity Rate for the thesis and I declare my thesis does not contain any plagiarism according to the maximum similarity rates specified in this directive; otherwise, I accept all kinds of legal responsibility for any dispute arising in situations which are likely to be detected and I declare that the information I have given above is correct.

I submit it to your information.

28/7/2023
Amir Khaleel HASSOO

Name and Surname: Amir Khaleel HASSOO
Student ID: 19910001172
Deparment: İstatistik
Program: **( )** M. Sc. (X) Ph.D.

SUPERVISOR

(Assoc. Prof. Dr. Şakir İŞLEYEN)
(Signature)

APPROVAL OF THE INSTITUTE

Prof. Dr.Harun AKKUŞ
(Signature)