Optimizing Sri Lanka's Energy Mix to Meet Future Electricity Demand: A Case Study Using Linear Programming And Time Series Forecasting

For the Bachelor of Science Honours Degree in

Financial Mathematics and Industrial Statistics

By

D.M.B.D.DISSANAYAKE - SC/2021/12461

Supervisor:

Dr. D.M. Samarathunga

Department of Mathematics

University of Ruhuna

Matara

January 31, 2025

Declaration

I, D.M.B.D. Dissanayake, declare that the presented project report titled, Optimizing Sri Lanka's Energy Mix to Meet Future Electricity Demand: A Case Study Using Linear Programming And Time Series Forecasting is uniquely prepared by me based on the group project carried out under the supervision of Dr. D.M. Samarathunga, Department of Mathematics, Faculty of Science, University of Ruhuna, as a partial fulfillment of the requirements of the level III, Case Study II – MFM3151 of the Bachelor of Science Honours Degree in Financial Mathematics and Industrial Statistics in the Department of Mathematics, Faculty of Science, University of Ruhuna, Sri Lanka.

It has not been submitted to any other institution or study program by me for any other purpose.

Signature: ....

Date: ....

Supervisor's Recommendation

I certify that this study was carried out by D.M.B.D. Dissanayake under my supervision.

Signature: ....

Date: ....

Department of Mathematics
Faculty of Science
University of Ruhuna.

Acknowledgement

First, we thank our supervisor Dr. D.M. Samarathunga sir, who helped us to learn a lot about this case study. His ideas and comments helped to complete this study. Without his support and guidance, this study would not have been possible.

As well, we would like to thank our course coordinator, Ms. I.C. Liyanage madam, who gave us many more valuable instructions to complete our case study successfully.

And also, we extend our thanks to all the instructors who helped us to make this research a success. In addition, we express our sincere appreciation to all of our batchmates for sharing their knowledge with us.

Finally, we are thankful to everyone who played a part; big or small, to make this research a success.

D.M.B.D.DISSANAYAKE

SC/2021/12461

B.Sc. Honors Degree in Financial Mathematics and Industrial Statistics

Department of Mathematics

Faculty of Science

University of Ruhuna

Contents

List of Figures

List of Tables

Abstract

The research study examines how Sri Lanka can optimize its energy mix for future electricity requirements with affordable costs and renewable power sources integration. Through combining time series forecasting with linear programming models the research evaluates how renewable generation affects operational limitations and reduces costs within the power mix schedule. The research determines the best combination between hydroelectric power and wind and coal and oil generation and independent energy sources under differing weather patterns and demand levels. The research produces important conclusions about Sri Lanka's sustainable and cost-effective energy pathway by spotlighting hybrid energy systems alongside policy strategies.

Key words: Energy mix optimization, Time series forecasting, Linear programming

Chapter 1

Introduction

1.1 Background of the study

The forecast of electricity demand and the optimal energy mix remain key concerns for energy planners and policymakers in Sri Lanka, while this sector has rapidly growing electricity demand, a diverse portfolio of energy generation and increasing interest in integrating renewables with the goals of sustainable development. During recent years, Sri Lanka has undergone rapid development in economic growth and population growth, each of which has a significant impact on the way electricity is consumed in the residential, commercial and industrial sectors.]The country is in the process of converting to greener sources of energy as a result of international commitments to reduce the carbon footprint and reliance on fossil fuels.

It has traditionally depended on hydropower, coal, and oil-fired power plants for electricity generation.

Hitherto variable fuel prices, perceived environmental sustainability, and increased exposure to climatic changes, such as unpredictability of rainfall, have challenged energy security and efficiency. The above grievances have since compelled the government to further diversify its energy portfolio with renewable energy sources, like solar and wind, through increased investment in a bid to reduce carbon emissions. The precise anticipation of electricity demand is crucial for efficient energy planning, as it aids in predicting future consumption trends and guides decisions concerning the optimal distribution of resources for electricity generation. By comprehending the anticipated trajectory of demand, policymakers can mitigate the risks associated with either under or over-investing in generation capacity, thus maintaining a balance between supply and demand. Time series forecasting models, which evaluate historical patterns of electricity consumption and discern trends along with seasonal variations, serve as a valuable instrument for forecasting future demand.

On the other hand, linear programming methodologies are useful in elaborating an appropriate composition to meet projected demand economically. Considering the diversified energy resources in Sri Lanka, including hydropower, coal, natural gas, and renewable sources, optimization frameworks help identify the best mix of these components to reduce generation costs and meet policy goals, including scaling up the contribution from renewable energy sources and reducing environmental impacts associated with power generation. Adding renewable into the mix makes the process of optimization more complex because renewable energy is intermittent; therefore, it needs careful planning and backup options in the form of storage or other sources of energy.

The government of Sri Lanka has ambitious targets on renewable energy. For example, it is envisioned that 70% of its electricity should be generated via renewable sources by 2030. Accomplishment of this target requires fundamental changes in the composition of energy and, therefore calls for a well-informed data-driven approach which integrates projections of future demand against economic, environmental, and technical limitations.

Therefore, the study takes its cue from this and integrates time series demand forecasting with linear programming optimization to arrive at an optimal energy mix strategy for Sri Lanka under different future demand scenarios. The outcome of this study will add to the literature on electricity demand forecasting and optimization, besides actionable insight to policymakers, energy planners, and other stakeholders of Sri Lanka's electricity sector. The study will contribute to solving the challenges pertaining to the tri-compatibilities of demand, cost-effectiveness, and sustainability and help the country toward a more robust and environmentally sustainable electricity framework.

1.2 Problem Statement

- How can Sri Lanka meet its future electricity demand using an optimal energy mix that balances costs and renewable energy goals?

1.3 Objectives of the Study

- **Forecast Future Demand:** Use time series models to project Sri Lanka's electricity demand over the next 10-20 years under different economic, environmental, and policy scenarios.

- **Optimize Energy Mix:** Develop a linear programming model to find the optimal mix of energy sources (hydro, coal, solar, wind, etc.) to meet the projected demand while minimizing costs, emissions, or other constraints.

Chapter 2

Literature Review

Sri Lanka's electricity sector is an important part of the nation's infrastructure and economy. Over the past few decades, the country has made significant efforts to expand its sources of energy production, while also facing challenges related to cost and increasing consumption. This literature review discusses the key aspects of electricity production, cost management, and consumption trends in Sri Lanka, drawing from academic and policy-oriented studies.

Sri Lanka's electricity sector has been the subject of various studies focusing on demand forecasting and energy mix optimization. Rathnayaka et al. [2018] employed ARIMA models to predict electricity demand in Sri Lanka, identifying ARIMA(3,1,1) and ARIMA(1,1,1) as optimal models for forecasting electricity generation and consumption. Amarawickrama and Hunt (2007) explored econometric techniques for estimating electricity demand, while Silva and Samaliarachchi (2013) demonstrated the effectiveness of Artificial Neural Networks (ANNs) in forecasting daily peak demand. Other studies, such as Cooray and Peiris (2012), utilized state-space models for short-term demand predictions. These approaches highlight the need for accurate forecasting in energy planning, yet they often overlook optimization strategies for energy generation. This study integrates time series forecasting with linear programming optimization to address this gap, ensuring Sri Lanka can meet its electricity demand efficiently while incorporating renewable energy sources.

The study Rizvi et al. [2022] "The Power Play of Natural Gas and Crude Oil in the Move Towards the Financialization of the Energy Market" examines the financialization of natural gas and crude oil markets, leading to increased price volatility and unpredictable energy markets. Geopolitical factors and regulatory challenges remain, highlighting the need for effective oversight to ensure market integrity. The study highlights the complex interplay between energy sectors and global financial markets.

Weerasinghe and Jayasundara [2020] study focus fossil fuel power plant unit costs in Sri Lanka using statistical models. The findings help in financial and operational planning, improve fuel management and cost control for the national grid.

The study by Murshed et al. [2020] explores the relationship between electricity consumption and economic growth in Sri Lanka, highlighting the importance of structural breaks. The research highlights the need for comprehensive energy policies to address changing economic dynamics and contributes to sustainable development.

Thambawitage [2019] offers a good solution to meet hourly energy demand, utilizing a mix of solar, wind, biomass, and hydropower. However, this transition requires supportive policies, investments in renewable technologies, and improved grid infrastructure, ensuring sustainable energy development and climate change mitigation.

The article Williams [2013] emphasizes the importance of accurate forecasting methods for electricity generation and economic growth. It highlights the relationship between energy prices and economic performance, the risks of energy import dependency, and the need for a comprehensive energy management approach.

Chapter 3

Materials and Methods

3.1 Research Approach

The overall research approach for this study is quantitative research approach. **Quantitative research approach** Quantitative research is the process of collecting and analyzing numerical data. It can be used to find patterns and averages, make predictions, test causal relationships, and generalize results to wider populations. The multiple linear regression model is one of the tools of the quantitative approach. It is used to determine a mathematical relationship among several variables. In our case, dependent variable is continuous and we have several independent variables. Therefore, we selected multiple linear regression as our model. This study adopts a two-fold approach, combining time series forecasting and linear programming optimization to develop an optimal energy mix strategy for Sri Lanka's electricity sector.

3.2 Research design

3.2.1 Linear Programming

The projected demand will serve as parameters in a linear programming framework that optimizes the energy composition. This will aim to minimize generation costs, subject to meeting policy imperatives such as renewable energy targets, carbon emissions limits, and fuel availability. Different scenarios, including changes in fuel prices or the addition of renewable energy sources, will be evaluated for determining the most economically feasible and sustainable energy mix for future electricity supply. This will ensure that both the prospective demand and generation capacity are optimized to make a proper decision by energy planners.

Linear programming functions best when relationships between variables exhibit linear behavior because it translates cost and capacity and demand equations into linear representations. The core components of an LP model include:

1. Decision variables are modelable variables that conceivably optimize objective functions because they express options controlling the problem's results.
2. Objective Function : The specified measure gets measured through the quantitative objective function. A mathematical expression shows decision variables combined in a specific way which either minimizes or maximizes the function.
3. Constraints are typically expressed as mathematical inequalities or equalities that describe the feasible region within which the optimal solution must lie. Constraints ensure that the solution is realistic and practical, adhering to real-world limitations.

3.2.2 Time Series Analysis

Time series analysis involves the study of data points collected or recorded at specific intervals of time. It is a statistical method used to identify patterns, trends, seasonality, and other components within the data, enabling better understanding and forecasting of future values.

There are 4 key components of time series.

1. Trend: The long-term increase or decrease in the data over time.
2. Seasonality: Periodic fluctuations that occur at regular intervals, such as monthly or quarterly patterns.
3. Irregularity: Random variations not explained by the trend or seasonality.
4. Cyclical : Cyclical components refer to fluctuations that occur at irregular intervals.

Process of Time Series Analysis and Forecasting

- Trend and Seasonality Analysis: Decompose the time series into its components (trend, seasonal, and random) using classical decomposition methods. Visualize the decomposed components to understand the underlying structure of the data.

- Stationarity Testing:A time series is said to be stationary if there is no systematic change in its mean, variance, covariance, and standard deviation over time. In other words, a stationary time series does not have trends or seasonal components Use the Augmented Dickey-Fuller (ADF) test and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test to check if the time series is stationary.If the series is non-stationary, apply differencing to stabilize the mean and remove trends.

- Autocorrelation and Partial Autocorrelation Analysis: Generate Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots to identify the presence of seasonality and lag dependencies within the data.

Autocorrelation Function (ACF) - The Autocorrelation Function is a graphical representation of the correlation of a time series with itself at different lags. The correlation coefficient is a measure of how closely two variables are related.
Partial Autocorrelation Function (PACF) - The Partial Autocorrelation Function (PACF) is a graphical representation of the correlation of a time series with itself at different lags, after removing the effects of the previous lags. The PACF plot is used to identify the order of an MA model.

- Visualization and Interpretation: - Plot the forecasted time series data alongside historical data to illustrate future trends. - Analyze the forecast for patterns and potential applications in energy planning.

- Choose an appropriate time series forecasting model : There are so many forcasting methods like Moving Average, Exponential smoothing, ARIMA/SARIMA, Machine learning methods and etc.

Forecasting method of ARIMA/SARIMA

- ARIMA is suitable for non-seasonal time series.
- SARIMA (Seasonal ARIMA) includes seasonal terms to handle seasonality directly.

ARIMA stands for **Autoregressive Integrated Moving Average**. It is used for modeling and forecasting non-seasonal time series data by capturing trends, autocorrelation, and noise.

Components

ARIMA is described by the parameters (p, d, q):

- p: **Autoregressive (AR)** term — the number of lagged observations included in the model.

- d: **Differencing (I)** term — the number of times the data needs to be differenced to make it stationary.
- q: **Moving Average (MA)** term — the number of lagged forecast errors used in the model.

Equation

The ARIMA equation can be expressed as:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}$$

Where:

- y_t: Forecasted value.
- c: Constant term.
- ϕ: Autoregressive coefficients.
- θ: Moving average coefficients.
- ϵ: Error term.

SARIMA stands for **Seasonal ARIMA**. It extends ARIMA to handle **seasonality** by adding seasonal components.

Components

SARIMA is described by $(p, d, q) \times (P, D, Q, m)$:

- p, d, q: Non-seasonal ARIMA terms.
- P, D, Q: Seasonal counterparts of ARIMA terms:
  - P: Seasonal Autoregressive (SAR).
  - D: Seasonal Differencing.
  - Q: Seasonal Moving Average (SMA).
- m: Seasonality period (e.g., 12 for monthly data, 4 for quarterly data).

Equation

The SARIMA equation incorporates both non-seasonal and seasonal terms:

$$y_t = (\text{non-seasonal ARIMA terms}) + (\text{seasonal ARIMA terms based on } m) + \epsilon_t$$

Chapter 4

Data

Overview

In this chapter, we concern about the data set what we selected. This will gives brief introduction about data set and special characteristic related to data set.

4.1 Data Description

The Ceylon Electricity Board [2023] published by the Ceylon Electricity Board (CEB) is a primary data source for our analysis of Sri Lanka's future energy mix. This comprehensive document provides detailed projections on the country's electricity generation, infrastructure development, and energy policy for the next 20 years.

The dataset covers data from 1969 to 2023, hence over five decades of electricity-related data in the country. This period includes the changes in energy policy, infrastructural development, and fluctuations in the growth rate of the economy, which all make a difference in the demand and generation of electricity.

The data is on Electricity Sales, representing the quantity of electricity sold to different consumers in Sri Lanka over the period 1969-2023. Sales of Electricity to consumer segments such as residential, commercial, industrial, and governmental sectors. This attribute represents the demand side of electricity utilization in the country.

Electricity Generation, captures the overall amount of electricity generated by the CEB annually over the same period. Inclusion of generated power from various sources such as thermal - coal, oil, hydroelectric, wind, and other varied renewable sources. The above feature is representative of the supply side and is an epitome of the evolution of the strategies that Sri Lanka has applied in its energy production process.

4.2 Meta Data

The source of these data is Ceylon Electricity Board (CEB).

https://docs.google.com/spreadsheets/d/1V64u-eXhEM9dpy5q0oKDcSwRS09zfFC4kwSZ0Vd9nus/edi

4.3 Data Dictionary

1. YEAR
2. MONTH
3. TOTAL WITH LECO

4.4 Data set preparation

4.4.1 Check NA values

Then we check is this data set has NA values.

- is.na(data)

Here also found that this data set hasn't any NA values.

4.4.2 Check Outliers

We intend to follow some steps for handling outliers. First, we hope to plot boxplot to identify outliers.
This outliers can be ignored when compared to the previous one.

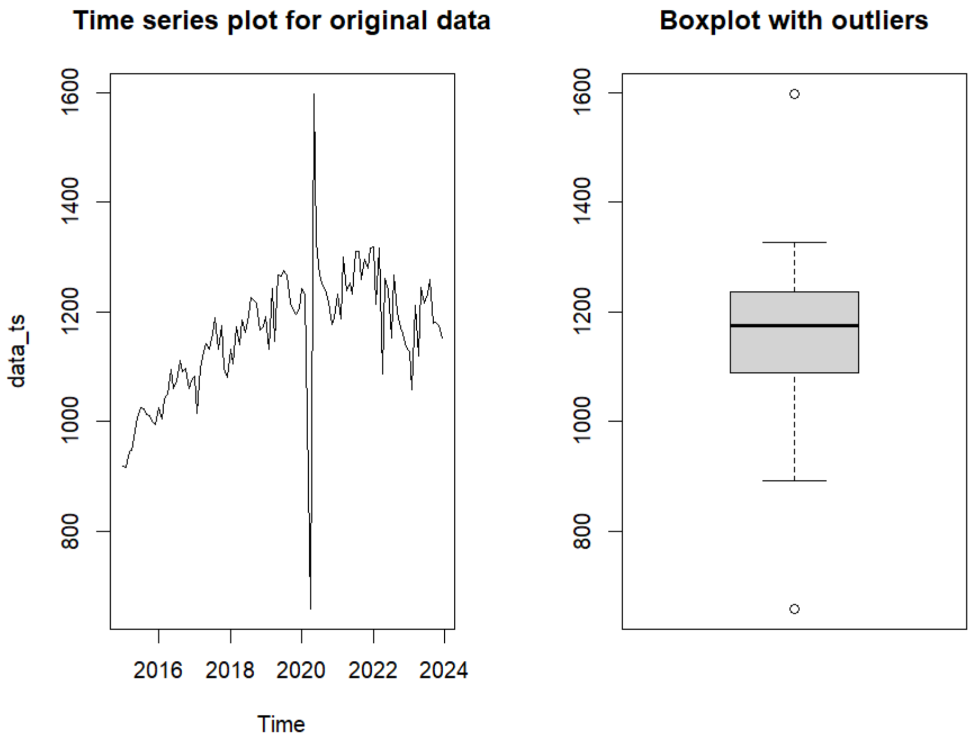

Figure 4.1: Time series plot for 1st differencing data

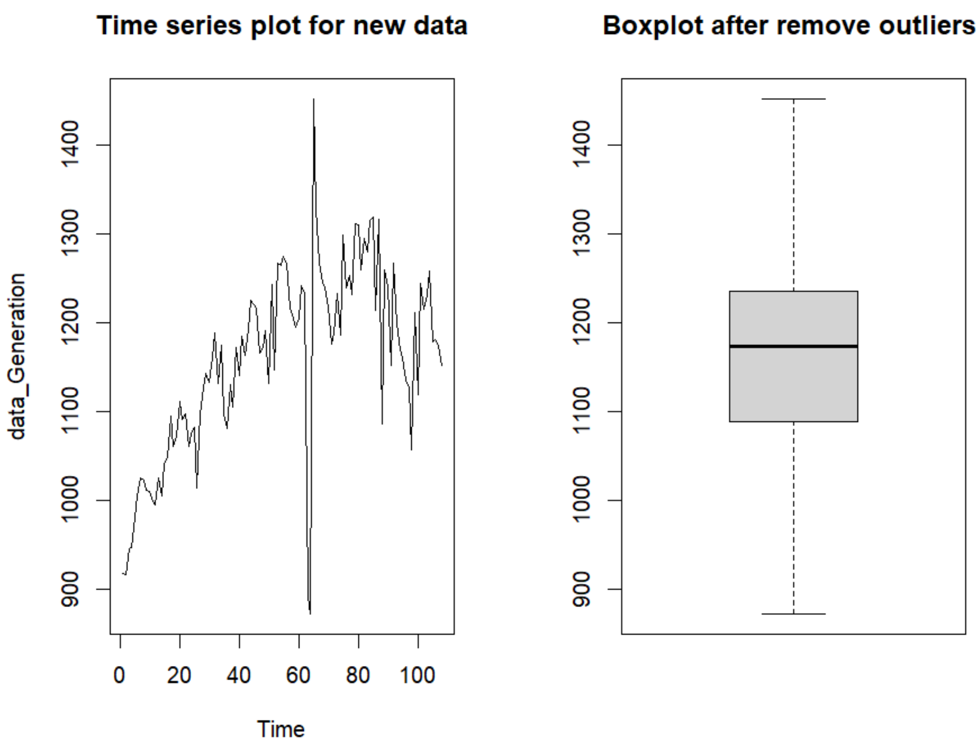

Figure 4.2: ACF and PACF plots for the original data

This outliers can be ignored when compared to the previous one.

Chapter 5

Results

5.1 Time Series Forecasting

The ARIMA (Autoregressive Integrated Moving Average) model is a widely used statistical approach for time series forecasting that captures trends, seasonality, and autocorrelations in data. Fitting an ARIMA model involves ensuring the dataset is stationary, often by applying differencing and confirming with tests like the Augmented Dickey-Fuller (ADF) test. The next step is identifying the model order (p, d, q)—where p represents autoregressive terms, d the level of differencing, and q the moving average terms—guided by ACF and PACF plots. The model is then fitted by estimating parameters using statistical tools such as R or Python, and diagnostic checks like residual analysis or the Ljung-Box test are performed to confirm the residuals behave as white noise. Once validated, the ARIMA model is used to forecast future values, with performance evaluated using error metrics like RMSE or MAE to ensure accuracy.

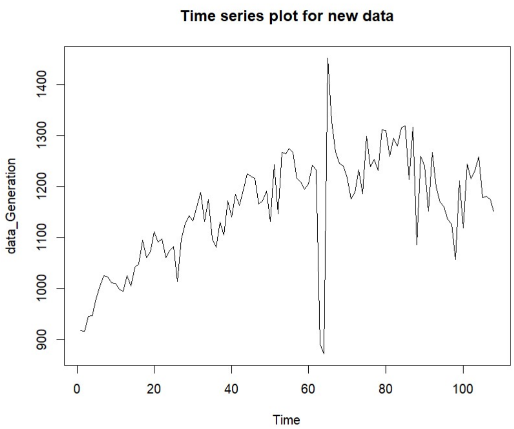

Figure 5.1: Time series plot for new data

Figure 5.1 shows that the time series appears to be non-stationary. This is because the data has a clear upward trend over time, which suggests that the mean and variance of the series are not constant over time. So, from that, we figure we can say there's an upward trend and no seasonality.

Then below figures shows the ACF and PACF plots for the original data set.

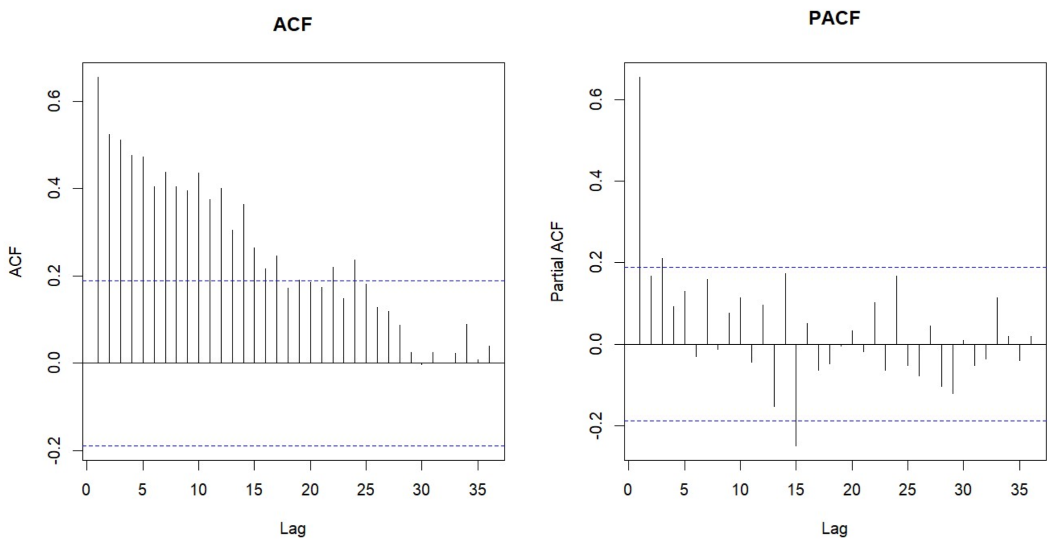

Figure 5.2: ACF and PACF plots for the original data

Test	**P-value**	**Conclusion**
ADF test	0.2647	Not stationary
KPSS test	0.01	Not stationary

Table 5.1: The results for the original time series data

Based on the results in Table 5.1, it is evident that our original dataset does not exhibit stationarity according to the outcomes of ADF and KPSS tests.
Therefore , before to get ARIMA model, we use differencing.
Differencing is a common technique used to make a time series stationary by removing trends or other patterns.

- 1st Difference

Check the 1st differencing data set is stationary or not using ACF and PACF.

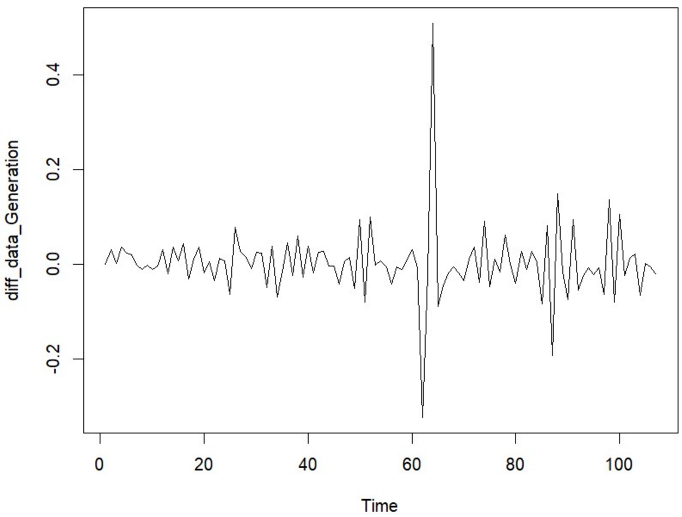

Figure 5.3: Time series plot for 1st differencing data

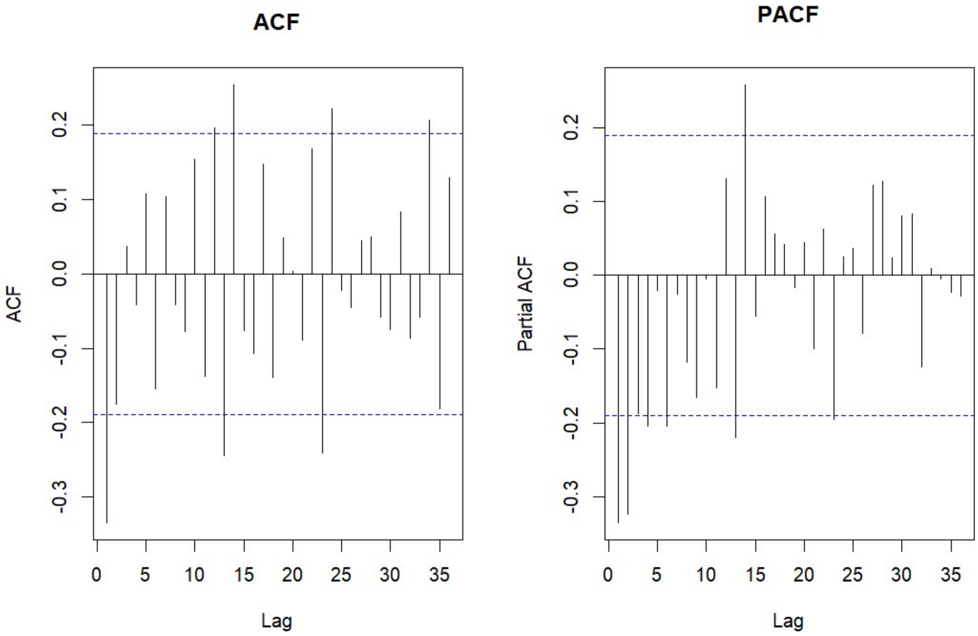

Figure 5.4: ACF and PACF plots for first differencing data

Test	**P-value**	**Conclusion**
ADF test	0.01	Stationary
KPSS test	0.1	Stationary

Table 5.2: The results for the original time series data

According to the Table 6.2, data set after 1st differencing was being stationary. But it had seasonality. Therefore, we want to remove the seasonality of 1st differencing data set.

- 2nd Difference

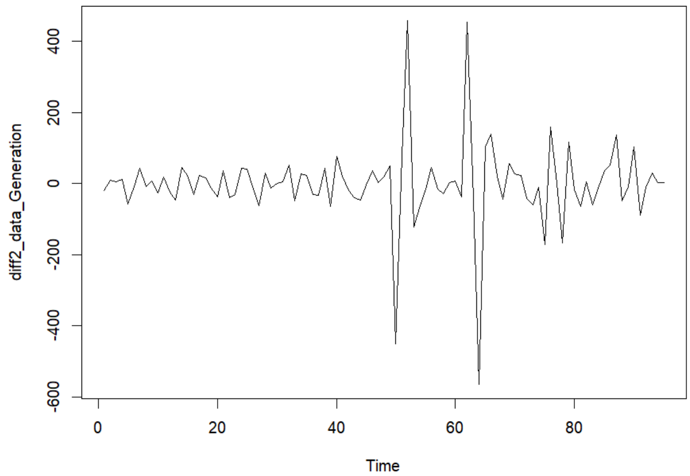

Figure 5.5: Time series plot for second differencing data

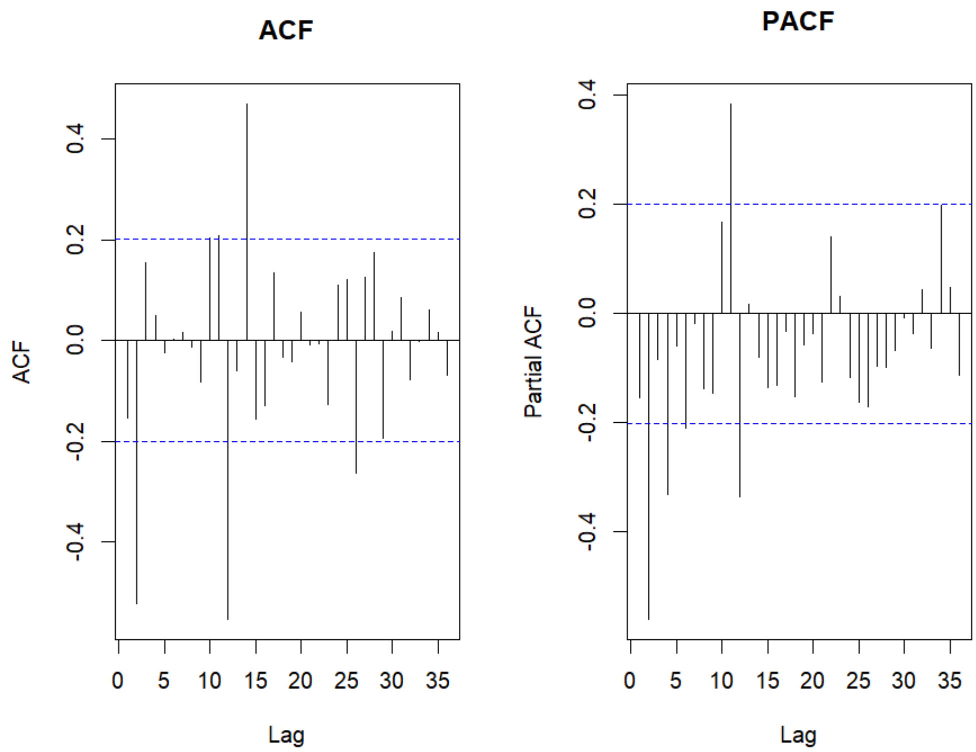

Figure 5.6: ACF and PACF plots for second differencing data

The examination of our dataset shows no evidence for substantial seasonal patterns since the corresponding findings came back negative from subsequent statistical tests. We selected the SARIMA model because of its strong foundation even though the data showed no seasonality pattern. SARIMA models bring together differencing capabilities and non-stationary correction features to make them suitable for measuring accuracy and reliability in forecasting our dataset.

5.1.1 Model Selection and Fit

In this section, we were selecting an appropriate model, which is crucial to ensure accuracy and reliability.The Seasonal AutoRegressive Integrated Moving Average (SARIMA) model received selection for this study because it addresses seasonal and non-seasonal components within the data set.

1) Manual SARIMA model

```
> summary(sarima_model)
Series: data_Generation
ARIMA(2,1,1)(2,1,2)[12]

Coefficients:
         ar1      ar2      ma1     sar1     sar2     sma1     sma2
      0.1706  -0.3645  -0.6934  -0.5884  -0.0179  -0.2511  -0.4267
s.e.  0.1240   0.1067   0.1043      NaN   0.1529      NaN      NaN

sigma^2 = 4721:  log likelihood = -539.77
AIC=1095.53   AICc=1097.21   BIC=1115.97

Training set error measures:
                    ME     RMSE      MAE       MPE     MAPE      MASE        ACF1
Training set -10.40464 62.02239 32.96326 -1.139506 2.894932 0.6804175 -0.01390933
```

Figure 5.7: Summary data for manual SARIMA model

2) SARIMA model created by using maximum no. of lags

```
> summary(final_model)
Series: data_Generation
ARIMA(0,1,3)(0,1,1)[12]

Coefficients:
          ma1      ma2     ma3     sma1
      -0.4846  -0.5423  0.2855  -0.7667
s.e.   0.1109   0.0959  0.1048   0.1376

sigma^2 = 4507:  log likelihood = -538.29
AIC=1086.58   AICc=1087.25   BIC=1099.35

Training set error measures:
                    ME     RMSE      MAE       MPE     MAPE      MASE         ACF1
Training set -9.839146 61.62344 34.05686 -1.084457 2.978508 0.7029913 -0.003565584
```

Figure 5.8: Summary data for manual SARIMA model

Models comparison

Model	AIC	BIC	MSE	MAPE
Manual model	1223.864	1234.518	5164.632	2.894932
Final model	1086.576	1099.346	3797.449	2.978508

Table 5.3: Model Comparison (AIC, BIC, MSE, MAPE)

Using AIC, BIC, MSE, MAPE values , we can say the best model for forecasting was final model. Because the final model demonstrates better fitting abilities through lower AIC and BIC values and enhanced predictive characteristics through lower MSE yet their slight higher MAPE indicates possible minor prediction deviations compared to the manual model. Analysis of lower AIC, BIC, and MSE values indicates that the final model provides an optimal blend of predictive accuracy along with model fit.

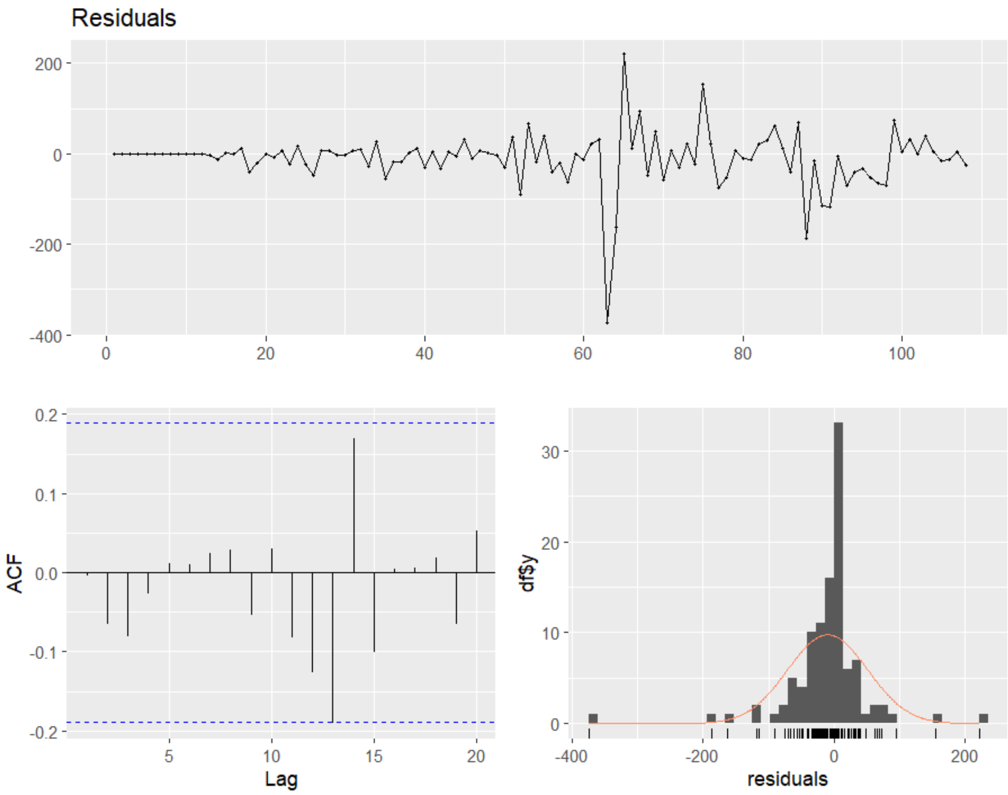

Figure 5.9: Residuals for final SARIMA model

```
> checkresiduals(final_model$residuals)

        Ljung-Box test

data:  Residuals
Q* = 1.9708, df = 10, p-value = 0.9966

Model df: 0.   Total lags used: 10

> shapiro.test(final_model$residuals)

        Shapiro-Wilk normality test

data:  final_model$residuals
W = 0.78009, p-value = 2.033e-11
```

Figure 5.10: Ljung-Box test and Shapiro test for final SARIMA model

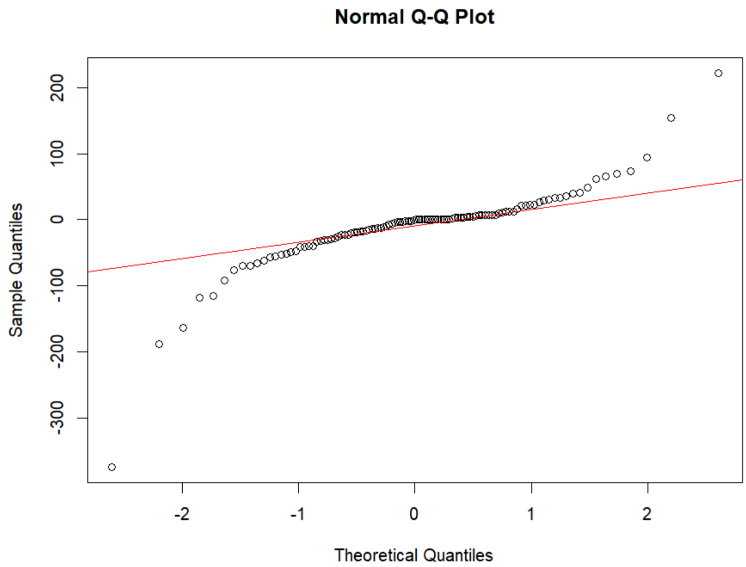

Figure 5.11: Normality Q-Q plot for final SARIMA model

By looking at this all graphs and test, residuals of final model was not normally distributed.

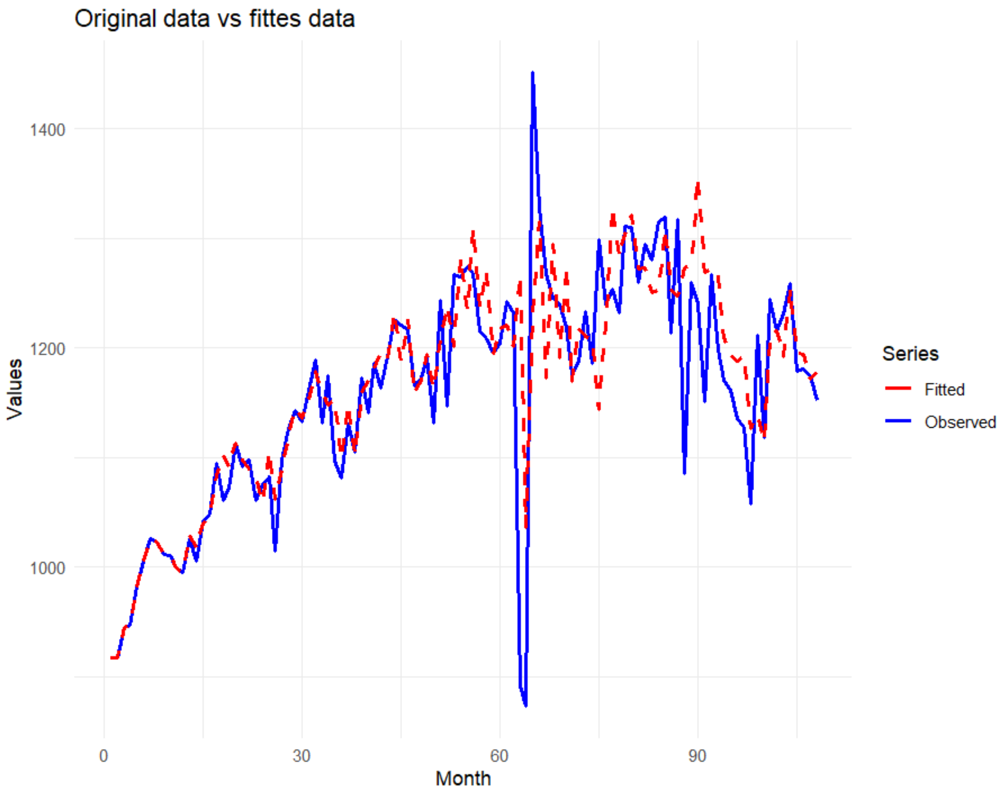

Figure 5.12: Original values Vs Fitted values

The statistical results indicate that the model successfully preserves the underlying pattern from the original data through its accurate representation of observed values with fitted values throughout most measurements. Notable spikes in observed values occur during time points 63 to 65 despite deviations present throughout the model. The model fails to capture effectively certain major deviations that seem to stem from data outliers or measurement errors or structural shifts outside its prediction range. The model exhibits limitations in detecting abrupt changes or rare events due to its regular pattern and stationary requirements. The model demonstrates strong performance across prevalent patterns yet needs additional enhancements to handle specific abnormal patterns given current results.

5.1.2 Forecasting process

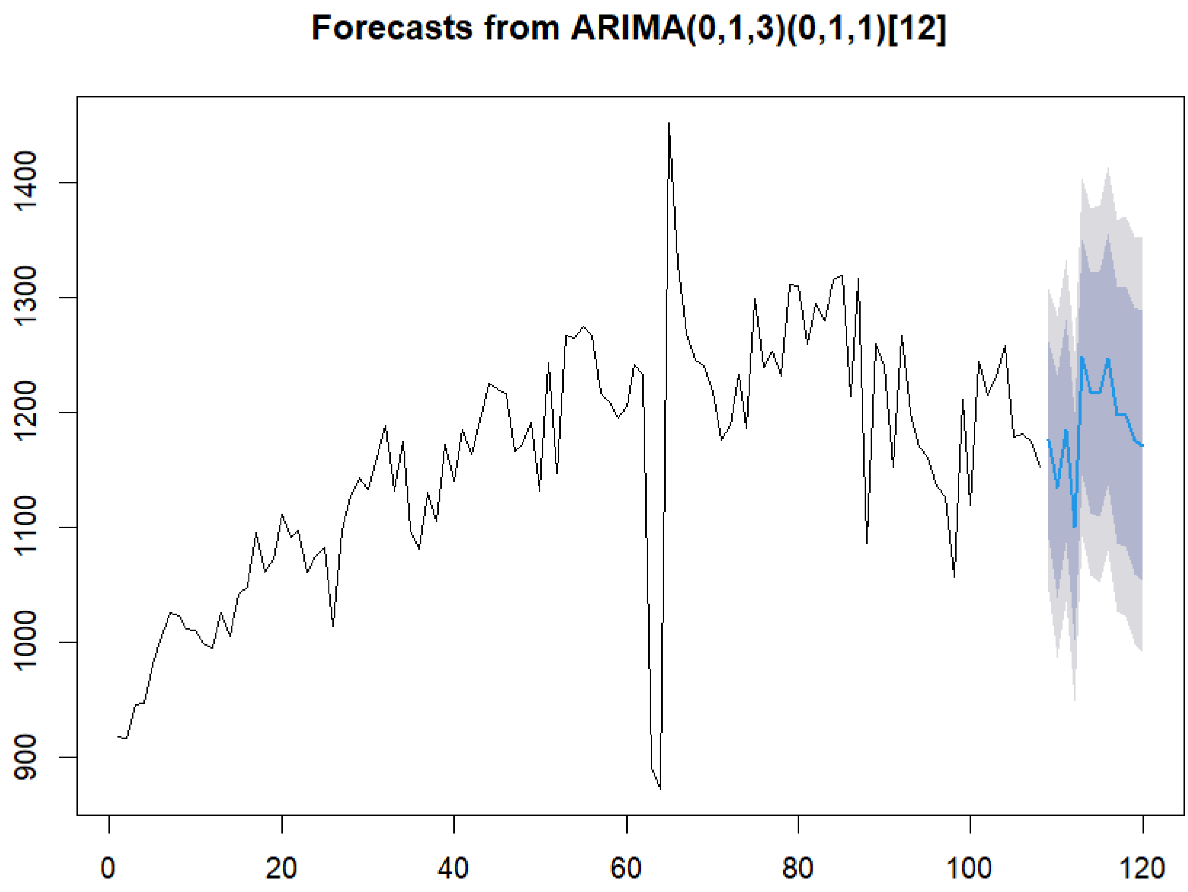

Figure 5.13: Forecasts from ARIMA(0,1,3)(0,1,1)[12]

The blue line represents the point forecasts, which indicate the energy consumption for the near future. The shaded blue area illustrates the confidence intervals. The model works as predicted since the forecast points fall within established thresholds and error margins. The data points remain within established testing limitations which demonstrates the strong reliability and quality of the forecasting method.According to the graph, forecast points were inside the infidence interval.

Table 5.4: Forecast Data with Confidence Intervals

Time	**Forecast**	**Lower 80%**	**Upper 80%**	**Lower 95%**	**Upper 95%**
Jan 2024	1176.126	1089.887	1262.365	1044.2346	1308.017
Feb 2024	1134.785	1037.771	1231.799	986.4151	1283.154
Mar 2024	1184.463	1087.431	1281.494	1036.0660	1332.859
Apr 2024	1100.511	1000.951	1200.070	948.2479	1252.774
May 2024	1248.133	1146.108	1350.158	1092.0996	1404.166
June 2024	1217.716	1113.284	1322.148	1058.0009	1377.431
July 2024	1216.087	1109.303	1322.872	1052.7741	1379.401
Aug 2024	1246.552	1137.465	1355.639	1079.7179	1413.387
Sep 2024	1197.261	1085.919	1308.603	1026.9786	1367.544
Oct 2024	1197.011	1083.459	1310.563	1023.3489	1370.673
Nov 2024	1175.720	1060.014	1291.425	998.7637	1352.675
Dec 2024	1171.317	1053.475	1289.160	991.0928	1351.542

This is a table for 1 year forecasting data.

5.2 Linear Programming

By using forecasting data , we were model a energy mix to meet future electricity demand. According to the research analysts applied linear programming (LP) to create optimized monthly electricity-generation strategies for Sri Lanka's power sector. The optimization goal focused on minimizing generation costs and maintaining both system demands and operational framework limits. The LP model was built with the following components:

- Decision variables

  The decision variables represent the energy generation from each source for each month:

$$x_{i,t} \quad \forall\, i \in \{1, 2, 3, 4, 5\},\, t \in \{1, 2, \ldots, 12\}$$

  Where:

  - i: Energy source (1: Hydro, 2: Oil, 3: Wind, 4: Coal, 5: Private).
  - t: Month (January to December).

Each $x_{i,t}$ represents the **amount of energy (in GWh)** generated by source i in month t.

- Objective Function

  The objective is to **minimize the total cost of energy generation** over 12 months:

$$\text{Minimize: } \sum_{t=1}^{12} \sum_{i=1}^{5} c_i \cdot x_{i,t}$$

  Where:

  - c_i: Cost per unit energy (LKR/GWh) for source i:

    $$c_1 = 2,170,000,\ c_2 = 56,860,000,\ c_3 = 24,750,000,\ c_4 = 26,810,000,\ c_5 = 69,230,000.$$

  - $x_{i,t}$: Decision variable for energy generation.

- Constraints

a. Demand Satisfaction Constraint

The total energy generated by all sources in each month must meet the adjusted energy demand including grid losses(10%). Because the energy lost during transmission and distribution from the power generation source to the end-users.Therefore, the energy generation must be 1.1 times the consumer demand.:

$$\sum_{i=1}^{5} x_{i,t} = D_t \quad \forall t \in \{1, 2, \ldots, 12\}$$

Where D_t is the monthly demand (adjusted for 10% grid losses).
D_t = Consumer Demand × (1 + Grid Loss Percentage)

b. Source-Specific Capacity Constraints

Each energy source obtains its monthly power generation from its maximum operational capacity through the Source-Specific Capacity Constraints. The wind capacity constraints for wind farms stem from wind availability and infrastructure requirements while coal plants encounter mechanical operational barriers.

$$0 \leq x_{i,t} \leq \text{Max Capacity}_{i,t} \quad \forall i, t$$

c. Base-Load Minimum Generation Constraints

To ensure grid stability the Base-Load Minimum Generation Constraints mandate that sources including oil and coal deliver a minimum output during specific operating periods. Reliable base-load power comes from coal and oil plants which need to operate at minimum generation levels between times when renewable sources become insufficient during non-monsoonal periods. The technical constraints element captures the real-world operational needs of power plants requiring stable maintenance costs throughout operation in order to deliver continuous system reliability even during periods of low renewable energy availability.

$$x_{\text{Coal},t} \geq \begin{cases} 432, & \text{Non-Monsoon Months} \\ 216, & \text{Monsoon Months} \end{cases}$$

$$x_{\text{Oil},t} \geq 59.04 \quad \forall t$$

d. Hydro Priority in Monsoon Months

Hydroelectric Power Placement in Monsoon Months(May, June, July, October, November, December) maintains priority status for both peak water resources utilization and the execution of sustainable power generation efforts while decreasing dependency on pricey fossil fuels including coal and oil.

$$x_{\text{Hydro},t} \geq x_{\text{Oil},t} + x_{\text{Wind},t} + x_{\text{Coal},t} + x_{\text{Private},t} \quad \forall t \in \text{Monsoon Months.}$$

e. Hydro Limit in Non-Monsoon Months

During non-monsoonal periods power generation from hydroelectric facilities reaches only 40% of their peak output which results in water scarcity. Water protection measures combined with critical supply management enable alternate power systems and preserve essential water reserves.

$$x_{\text{Hydro},t} \leq 0.4 \cdot \text{Max Capacity}_{\text{Hydro}} \quad \forall t \in \text{Non-Monsoon Months.}$$

f. Non-Negativity Constraints

All decision variables must be non-negative:

$$x_{i,t} \geq 0 \quad \forall i, t$$

5.2.1 Optimal Solutions

This was a optimal solution for minimizing cost using energy mixup on each month of year.

Optimal solution found.

Optimal energy generation (GWh by source and month):

	Month_1	Month_2	Month_3	Month_4	Month_5	Month_6	Month_7	Month_8	Month_9	Month_10	Month_11	Month_12
Hydro	407.05	407.05	407.05	407.05	687.2	670.4	669.04	407.05	407.05	658.07	646.74	644.15
Oil	453.33	115.36	59.04	371.67	59.04	238.4	59.04	59.04	477.9	59.04	214.74	59.04
Wind	0	74.52	0	0	0	0	74.52	0	0	0	0	74.52
Coal	432	648	432	432	628.16	432	535.48	432	432	599.03	432	510.59
Private	0	0	407.16	0	0	0	0	474.59	0	0	0	0

Optimal cost for each month:

	Month_1	Month_2	Month_3	Month_4	Month_5	Month_6	Month_7	Month_8	Month_9	Month_10	Month_11	Month_12
Total_Cost	3.8241e+10	2.666e+10	4.401e+10	3.3599e+10	2.1689e+10	2.6592e+10	2.1009e+10	4.8678e+10	3.9639e+10	2.0845e+10	2.5195e+10	2.0288e+10

Total cost: 485747261770 LKR

Figure 5.14: Results in Matlab

Below graph graphically represent the optimal solution.

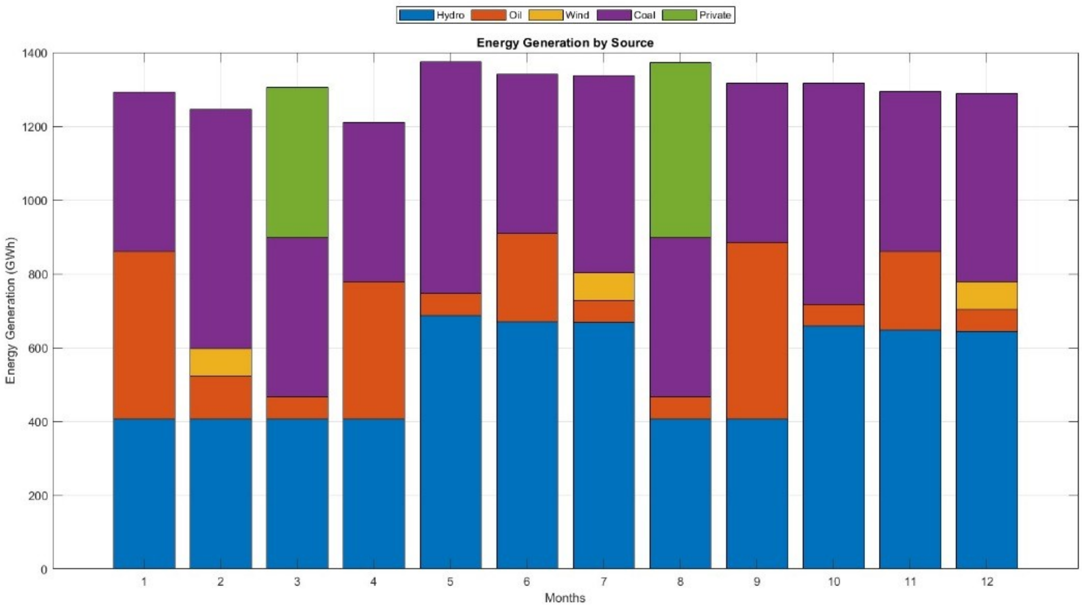

Figure 5.15: Results in Matlab

Chapter 6

Discussion and conclusion

6.1 Discussion

The main purpose of this study focused on finding the best combination of Sri Lanka's energy sources which would match future electricity requirements and reduce expenses alongside renewable energy integration. The research combined time series forecasting methods with linear programming optimization techniques. The research used multidimensional operational data across several consecutive years that contained no absent or incomplete records which facilitated strong modeling capabilities.

The observed variations in the time series pattern stood out after we executed the data visualization. Seasonal patterns alongside economic developments and COVID-19 pandemic disruptions serve as the main drivers behind these changes in the data. COVID-19 lockdowns triggered sudden decreases in electricity consumption which later translated into increased usage when economies began reopening. The time series achieved stabilization by differencing while both the ADF and KPSS tests validated stationarity.

During our analysis we examined multiple transformation(Log transformation) procedures along with machine learning techniques alongside advanced algorithms including XGBoost and LSTM which predicted energy consumption levels. Several technical methods demonstrated potential yet needed intense parameter shakes to reach their intended effect which required them to remain elaborate. Other supplementary techniques failed to maintain reliable identification of systematic seasonal patterns and trend metrics embedded within the data.

Therefore, we used the SARIMA model to analyze timeseries data. A residual analysis of the model forecasted differences between model predictions and actual observations during times when significant changes occurred. Sudden structural shifts like those from the pandemic alongside unexpected policy changes produced these differences because the modeling framework has limitations with non-linear patterns and breaks in stability. The SARIMA method achieved solid performance that was supported by its low AIC and BIC measurement values.

These were the 2 models we ckecked.

- Manual model - ARIMA(2,1,1)(2,1,2)[12]
- Final model - ARIMA(0,1,3)(0,1,1)[12]

Model	**AIC**	**BIC**	**MSE**	**MAPE**
Manual model	1223.864	1234.518	5164.632	2.894932
Final model	1086.576	1099.346	3797.449	2.978508

Table 6.1: Model Comparison (AIC, BIC, MSE, MAPE)

Using Table 6.1, we chose final model as best fit model for forecasting.

In optimization phases the research minimized energy costs but maintained operational constraints which included satisfying demand levels along with capacity limitations from energy sources and national policy targets. The linear programming model determined the ideal energy combination by selecting hydroelectric power during the monsoon months while maintaining foundational power stability through coal and oil operations outside of the monsoon season.

6.2 Conclusion

A proposed optimization model serves as a practical framework which supports energy planning through equitable considerations of cost requirements together with sustainability targets and demand predictions. The analysis demonstrates how hydro and coal infrastructures contribute to power grid stability during clean energy integration.

SARIMA-based demand predictions provide dependable patterns for establishing future requirements. The observed fluctuations in fitted data values together with the effects of pandemic disruptions emphasize the necessity of forecasting models to incorporate external factors including policy adjustments and economic activities as well as public health emergencies.

To reach long-term renewable energy targets the necessary steps include storage technology investments and infrastructure enhancements together with regulatory modifications to manage both intermittent energy supply and steady base-load demands.

Sri Lankan policymakers together with energy planners can use the research findings to build a sustainable and resilient power grid through this study. The development of new hybrid optimization systems using machine learning approaches and scenario analysis methods should become research priorities for addressing ongoing pandemic-related irregularities.

Chapter 7

Appendix

7.1 Linear Programming

```

%% Clear Everything
clc, clear, close all

%% Defining Parameters

% Sources
months = 12;
sources = {'Hydro', 'Oil', 'Wind', 'Coal',  'Private'};
n_sources = length(sources);

% Avg. Costs (LKR/GWh)
costs = [2170000, 56860000, 24750000, 26810000, 69230000];

% Max capacity (GWh) for each source for each month
max_capacity = [
    1017.63, 1017.63, 1017.63, 1017.63, 1017.63, 1017.63, 1017.63,
       1017.63, 1017.63, 1017.63, 1017.63, 1017.63;        % Hydro
    576.72, 576.72, 576.72, 576.72, 576.72, 576.72, 576.72, 576.72,
       576.72, 576.72, 576.72, 576.72                      % Oil
    74.52, 74.52, 74.52, 74.52, 74.52, 74.52, 74.52, 74.52, 74.52,
       74.52, 74.52, 74.52;                                % Wind
    648, 648, 648, 648, 648, 648, 648, 648, 648, 648, 648, 648;
                                                           % Coal
    837.44, 837.44, 837.44, 837.44, 837.44, 837.44, 837.44, 837.44,
       837.44, 837.44, 837.44, 837.44;                     % Private
];

% Monthly energy demand (GWh)
```

```
demand = [1174.89, 1131.76, 1186.59, 1100.66, 1249.45, 1218.90,
    1216.43, 1247.89, 1197.23, 1196.49, 1175.89, 1171.19];

% Grid losses (10% loss included in demand)
% Adjust for 10% losses
demand = demand * 1.1;

% Minimum generation (Coal and Oil)
min_coal = 432;  % GWh
min_oil = 59.04; % GWh

% Monsoon months and corresponding coal minimum generation
monsoon_months = [5, 6, 7, 10, 11, 12]; % May, June, July, October,
    November, December
non_monsoon_months = setdiff(1:12, monsoon_months);

%% Defining Variables

% Decision variables: Each source's generation for 12 months
% x = [x_hydro, x_oil, x_wind, x_coal, x_private]
x_size = n_sources * months;

% Upper and lower bounds
lb = zeros(x_size, 1);
ub = reshape(max_capacity', [], 1);

% Minimum coal generation bounds
coal_idx = 3 * months + (1:months);      % Indices for Coal
lb(coal_idx(non_monsoon_months)) = 432; % Minimum coal generation in
    non-monsoon months
lb(coal_idx(monsoon_months)) = 216;      % Minimum coal generation in
    monsoon months

% Minimum oil generation bounds
oil_idx = months + (1:months);       % Indices for Oil
lb(oil_idx) = min_oil;

%% Objective Function

% Objective function: Minimize cost
c = repmat(costs, 1, months);

%% Constraints

% Equality constraints (demand satisfaction)
Aeq = zeros(months, x_size);
for t = 1:months
```

```
    Aeq(t, t:months:end) = 1;
end
beq = demand';

% Enforce minimum generation for base load sources (Coal and Oil)
coal_idx = 3 * months + (1:months); % Indices for Coal
oil_idx = months + (1:months);      % Indices for Oil
lb(coal_idx) = min_coal;
lb(oil_idx) = min_oil;

% Hydro Priority in Monsoon Months
A_hydro_priority = zeros(length(monsoon_months), x_size);

% Loop over monsoon months and create constraints for Hydro priority
for t = 1:length(monsoon_months)
    month_idx = monsoon_months(t);

    % Hydro generation should be maximized first for monsoon months
    A_hydro_priority(t, month_idx) = -1;                  % Hydro (negative because we want it to be >= 0)
    A_hydro_priority(t, month_idx + months) = 1;          % Oil
    A_hydro_priority(t, month_idx + 2*months) = 1;        % Wind
    A_hydro_priority(t, month_idx + 3*months) = 1;        % Coal
    A_hydro_priority(t, month_idx + 4*months) = 1;        % Private
end

% The right-hand side for this constraint will be 0 (Hydro should be maximized before others)
b_hydro_priority = zeros(length(monsoon_months), 1);

% Combine the new priority constraint into the existing ones
A = [A_hydro_priority; Aeq];
b = [b_hydro_priority; beq];

% Hydro Generation Limit in Non-Monsoon Months (40% of Max Capacity)

A_hydro_limit = zeros(length(non_monsoon_months), x_size);

% Set Hydro limit to 40% of its max capacity for non-monsoon months
for t = 1:length(non_monsoon_months)
    month_idx = non_monsoon_months(t);
    A_hydro_limit(t, month_idx) = 1;                     % Hydro generation
end

b_hydro_limit = repmat(0.4 * max_capacity(1, 1), length(non_monsoon_months), 1);   %40% of max capacity
```

```

% Add these new constraints to the existing constraints
A = [A; A_hydro_limit];
b = [b; b_hydro_limit];

%% Solving the Problem

options = optimoptions('linprog', 'Algorithm', 'dual-simplex', 'Display', 'iter');
[x, fval, exitflag, output] = linprog(c, [], [], A, b, lb, ub, options);

% Reshape the Solution for Interpretation
if isempty(x)
    error('Optimization failed: No feasible solution found.');
else
    x_reshaped = reshape(x, months, n_sources)'; % Reshape the solution
end

%% Display Results

disp('Optimal energy generation (GWh by source and month):');
disp(array2table(x_reshaped, 'RowNames', sources, 'VariableNames', ..733
    arrayfun(@(x) sprintf('Month_%d', x), 1:months, 'UniformOutput', false)));

% Calculate the monthly costs (sum of generation for each source times its cost)
monthly_costs = sum(x_reshaped .* costs', 1); % Multiply each month's generation with its corresponding source cost

disp('Optimal cost for each month:');
disp(array2table(monthly_costs, 'RowNames', {'Total_Cost'}, 'VariableNames', ...
    arrayfun(@(x) sprintf('Month_%d', x), 1:months, 'UniformOutput', false)));

disp(['Total cost: ', num2str(fval), ' LKR']);

% Plotting the Energy Mix
figure;
bar(1:months, x_reshaped', 'stacked');
xlabel('Months');
ylabel('Energy Generation (GWh)');
title('Energy Generation by Source');
```

```
legend(sources, 'Location', 'northoutside', 'Orientation', '
    horizontal');
colormap(lines(n_sources));
grid on;
```

7.2 Time series and forecasting

```
# Load libraries
library(tidyverse)
library(lubridate)
library(forecast)
library(tseries)
library(xgboost)
library(seastests)
library(seasonal)
library(keras)
library(TSA)
library(fpp2)
library(fable)
library(dplyr)
library(ggplot2)
library(tensorFlow)

# Load data
data = Energy_Consumption_2015_to_2023_NEW


# Data Preparation
data$Date <- as.Date(paste(data$YEAR, data$MONTH, "1", sep = "-"),
    format = "%Y-%m-%d")
missing_data <- sum(is.na(data$TOTAL.WITH.LECO))
if (missing_data == 0) print("No Missing Values") else print("There
    is Missing Values")
par(mfrow = c(1,2))
data_ts <- ts(data$TOTAL.WITH.LECO, start=c(2015, 1), frequency=12)
data_Generation_with_outliers <- as.numeric(data_ts)
Q1 <- quantile(data_Generation_with_outliers, 0.25)
Q3 <- quantile(data_Generation_with_outliers, 0.75)
IQR <- Q3 - Q1
lower_limit <- Q1 - 1.5 * IQR
upper_limit <- Q3 + 1.5 * IQR
cat("Lower Limit:", lower_limit, "\n")
cat("Upper Limit:", upper_limit, "\n")
```

```r
outliers <- data_Generation_with_outliers[data_Generation_with_outliers < lower_limit | data_Generation_with_outliers > upper_limit]
cat("Outliers:", outliers, "\n")
data_Generation_with_outliers[data_Generation_with_outliers < lower_limit] <- lower_limit
data_Generation_with_outliers[data_Generation_with_outliers > upper_limit] <- upper_limit
boxplot(data_Generation_with_outliers, main="Boxplot with Outliers", ylab="Generation", col="skyblue", horizontal=FALSE)
abline(h=c(lower_limit, upper_limit), col="red", lty=2)

# Plot the data
plot(data_ts, main="Energy Generation (2015-2023)", ylab="Energy (GWh)", xlab="Year")
data_ts_after_remove_outliers <- ts(data_Generation_with_outliers, start=c(2015, 1), frequency=12)
plot(data_ts_after_remove_outliers, main="Energy Generation (2015-2023)", ylab="Energy (GWh)", xlab="Year")
data_Generation <- as.numeric(data_ts_after_remove_outliers)

# ARIMA/SARIMA Model
# ----------------------------------------------------------------------------------------------------

# Check stationarity using ADF test
adf_test_results <- adf.test(data_Generation)
if (adf_test_results$p.value <= 0.05) print("Data is stationary") else print("Data is not stationary")

# Regular differencing (to remove trend)
d <- ndiffs(data_Generation)
cat("Number of regular differences needed: ", d, "\n")

# Seasonal differencing (to remove seasonality)
D <- nsdiffs(data_ts_after_remove_outliers)
cat("Number of seasonal differences needed: ", D, "\n")

# Differencing the series
diff_data_Generation <- diff(data_Generation)
plot.ts(diff_data_Generation)

# Check Seasonality
if (isSeasonal(diff_data_Generation, freq = 12)) print("Data have Seasonality") else print("Data do not have Seasonality")
#nsdiffs(ts(diff_data_Generation, frequency = 12))

```

```
# Plot ACF and PACF
acf(diff_data_Generation, main="ACF", lag.max = 36)
pacf(diff_data_Generation, main="PACF")

# Differencing the series
diff2_data_Generation <- diff(diff_data_Generation, lag = 12)
plot.ts(diff2_data_Generation)

# Check Seasonality
if (isSeasonal(diff2_data_Generation, freq = 12)) print("Data have
    Seasonality") else print("Data do not have Seasonality")

# Plot ACF and PACF
acf(diff2_data_Generation, main="ACF", lag.max = 36)
pacf(diff2_data_Generation, main="PACF")
par(mfrow = c(1,2))
acf(diff_data_Generation, main="ACF", lag.max = 36)
pacf(diff2_data_Generation, main="ACF", lag.max = 36)
dev.off()

# SARIMA Implementation
# Auto-select SARIMA parameters
auto_sarima_model <- auto.arima(data_Generation, seasonal= TRUE)
summary(auto_sarima_model)

# Manually-select SARIMA parameters
sarima_model <- Arima(data_Generation,order = c(2,1,1), seasonal =
    list(order=c(2,1,2), period=12))
summary(sarima_model)

# Define maximum values for p, q, P, Q
max_p <- 3
max_P <- 3
max_q <- 3
max_Q <- 3

# Placeholder to store the results
results <- data.frame(p = integer(), d = integer(), q = integer(),
                      P = integer(), D = integer(), Q = integer(),
                      AIC = numeric())

# Loop through all combinations of p, q, P, Q
for (p in 0:max_p) {
  for (q in 0:max_q) {
    for (P in 0:max_P) {
      for (Q in 0:max_Q) {
        # Try fitting the ARIMA model
```

```
        tryCatch({
          model <- Arima(data_Generation, order = c(p, 1, q),
                          seasonal = list(order = c(P, 1, Q), period =
                             12))
          # Store the results
          results <- rbind(results, data.frame(p = p, d = 1, q = q,
                                                P = P, D = 1, Q = Q,
                                                AIC = AIC(model)))
        }, error = function(e) {
          # Ignore models that fail
          NULL
        })
      }
    }
  }
}

# Sort the results by AIC to find the best model
best_model <- results[which.min(results$AIC), ]
print("Best ARIMA Model based on AIC:")
print(best_model)

comparison_table <- data.frame(Model = c("sarima_model","final_model"
   ),
  AIC = c(AIC(sarima_model),AIC(final_model)),
  BIC = c(BIC(sarima_model),BIC(final_model)),
  MSE = c(mean(residuals(sarima_model)^2) ,mean(residuals(final_model
     )^2) )
)

print(comparison_table)

# Fit the best model for further analysis
final_model <- Arima(data_Generation, order = c(0, 1, 3),
                     seasonal = list(order = c(0, 1, 1), period = 12)
                        )
summary(final_model)
fitted_values <- fitted(final_model)

observed_data <- data.frame(Time = as.numeric(time(data_Generation)),
     Values = as.numeric(data_Generation))
fitted_data <- data.frame(Time = as.numeric(time(data_Generation)),
    Values = as.numeric(fitted_values))

ggplot() +
  geom_line(data = observed_data, aes(x = Time, y = Values, color = "
     Observed"), linewidth = 1) +
```

```
  geom_line(data = fitted_data, aes(x = Time, y = Values, color = "Fitted"), linewidth = 1, linetype = "dashed") +
  scale_color_manual(name = "Series", values = c("Observed" = "blue", "Fitted" = "red")) +
  ggtitle("Original data vs fittes data") +
  xlab("Month") +
  ylab("Values") +
  theme_minimal()

############################
forecasted <- forecast(final_model, h = 12)
plot(forecasted)

# Extract the forecast data set
forecast_data <- data.frame(
  Time = time(forecasted$mean),            # Time periods for forecasts
  Forecast = as.numeric(forecasted$mean), # Point forecasts
  Lower_80 = forecasted$lower[, 1],       # 80% lower bound
  Upper_80 = forecasted$upper[, 1],       # 80% upper bound
  Lower_95 = forecasted$lower[, 2],       # 95% lower bound
  Upper_95 = forecasted$upper[, 2]        # 95% upper bound
)
forecast_data

plot(forecasted, main = "Time Series with SARIMA and Forecast", xlab = "Month", ylab = "Values", col = "blue")
lines(data_Generation, col = "black", lwd = 2)
lines(fitted_values, col = "red", lwd = 2, lty=2)
legend("topleft", legend = c("Original Data", "Fitted (SARIMA)", "Forecast"),
       col = c("black", "red", "blue"), lty = c(1,2,1), lwd = 2)
grid()

# View the forecast data set
print(forecast_data)

# residual check
checkresiduals(final_model$residuals)
qqnorm(final_model$residuals)
qqline(final_model$residuals,col="red")
normility_check_residual <- shapiro.test(final_model$residuals)
if(normility_check_residual$p.value <= 0.05) print("Residuals are not normally distributed") else print("Residuals are normally distributed")
```

Bibliography

Ceylon Electricity Board. Long Term Generation Expansion Plan 2023-2042, 2023. URL `https://www.pucsl.gov.lk/wp-content/uploads/2023/03/LTGEP-2023-2042-Final.pdf`. Accessed: 2025-01-29.

M. Murshed, H. Mahmood, T. T. Y. Alkhateeb, and M. Bassim. The impacts of energy consumption, energy prices and energy import-dependency on gross and sectoral value-added in sri lanka. *Energies*, 13(24):6565, 2020.

R. M. K. T. Rathnayaka, J. L. L. S. Fernando, J. R. N. A. Gunawardana, et al. Time series modelling approach for forecasting electricity demand in sri lanka. In *Conference Proceedings*, September 2018. URL `https://www.researchgate.net/publication/327764131`.

S. K. A. Rizvi, B. Naqvi, S. Boubaker, and N. Mirza. The power play of natural gas and crude oil in the move towards the financialization of the energy market. *Energy Economics*, 112:106131, 2022. ISSN 0140-9883. doi: https://doi.org/10.1016/j.eneco.2022.106131. URL `https://www.sciencedirect.com/science/article/pii/S0140988322002870`.

A. U. Thambawitage. Renewable energy in sri lanka: An overview on the possible energy sources to fulfill the hourly energy demand. *Independent thesis Advanced level, Master's Programme in Renewable Energy Systems*, 2019.

W. Weerasinghe and D. Jayasundara. Modelling and forecasting the unit cost of electricity generated by fossil fuel power plants in sri lanka. *International Journal of Sciences: Basic and Applied Research (IJSBAR)*, 51(2):163–178, 2020.

H. P. Williams. *Model building in mathematical programming*. Wiley, Hoboken, N.J., 2013. ISBN 9781118443330 1118443330. URL `https://ebookcentral.proquest.com/lib/uvic/detail.action?docID=1120846`.