Harnessing Renewable Energy for Carbon Emission Reduction: A Case Study of Sri Lanka's Path to Environmental Sustainability

A report presented in partial fulfillment of the

requirements for the course module

STAT 32633-Corporate Capstone Project

At

Department of Statistics and Computer Science, Faculty of Science, University of Kelaniya

Sri Lanka.

Group 5

- M.A.S.C. Siriwardhana
- N.C.R Gunarathna
- R.H.R.S. Nishshanka
- B.D.H. Chathuranga
- D.A.G. Ranthilina
- W.K.H. Rukshani
- W.K.S. Lakmali
- G.K.M Pathum
- T.A. Kottegoda
- A.S.N. Perera

Academic year: 2022/2023

DECLARATION

We do hereby declare that the work reported in this thesis which titled as "**Harnessing Renewable Energy for Carbon Emission Reduction: A Case Study of Sri Lanka's Path to Environmental Sustainability**" is originally prepared by us, **Group 5** in the purpose of partial fulfilment of requirement of the Bachelor of Science Degree in Statistics, Department of Statistics & Computer Science, Faculty of Science, University of Kelaniya, Sri Lanka and not for any other academic purposes. No part of this thesis has been submitted earlier or concurrently for the same or any other degree.

Date: 2025/01/26

Group 5,
Department of Statistics & Computer Science,
Faculty of Science,
University of Kelaniya, Sri Lanka.

Abstract

Renewable energy plays a crucial role in reducing CO_2 emissions and promoting environmental sustainability, particularly in Sri Lanka. A comprehensive dataset, spanning environmental, socio-economic, and energy-related variables from 1990 to 2020, was used to explore trends in renewable energy consumption and its connection to CO_2 emissions, forest depletion, and the agricultural sector's economic contribution. Through various statistical methodologies, including data cleaning, exploratory data analysis, correlation analysis, regression modeling, time series regression, and time series forecasting, significant patterns were uncovered and future emissions predicted. A robust forecasting model was developed to project CO_2 emissions in the coming years, integrating key drivers such as renewable energy consumption, forest depletion, and agricultural value-added. The model's validity and reliability were confirmed through various testing, including residual analysis, goodness-of-fit tests, and other diagnostic evaluations. All analyses were conducted using R, ensuring the results' accuracy and reproducibility. Key findings underscore the effectiveness of renewable energy sources, such as hydropower and solar, in mitigating emissions, while also highlighting the importance of sustainable forestry and agriculture management. The research offers actionable insights for policymakers and stakeholders to enhance renewable energy adoption, address environmental degradation, and mitigate climate challenges effectively.

Key words: Renewable Energy, CO_2 Emissions, Environmental Sustainability, Sri Lanka, Energy Consumption, Forest Depletion, Agricultural Sector, Time Series Forecasting, Regression Modeling, Hydropower

Preface and Acknowledgements

This report explores the critical role of renewable energy in mitigating carbon emissions in Sri Lanka. The scope of this research includes analyzing trends, impacts, and recommendations for integrating sustainable energy practices into national frameworks. We are profoundly grateful to the Department of Statistics and Computer Science, Faculty of Science, University of Kelaniya, for their invaluable support and resources, which facilitated every stage of this research. We also extend our gratitude to Dr. Pansujee Dissanayaka, Prof. Vasana Chandrasekara, Dr. Anuradha and the staff for their insightful guidance and constructive feedback, which shaped the depth and clarity of this report. Special thanks go to the data providers and institutions who contributed essential information for the analysis. Finally, we appreciate the encouragement and collaboration of our peers and families, whose support has been a cornerstone of our successful completion of this project.

Table of Contents

Content of Figures

Content of Tables

1. Introduction

1.1 Research Problem

The increasing concerns about climate change and environmental degradation have placed CO_2 emissions at the forefront of global discussions. As one of the primary contributors to greenhouse gases, CO_2 emissions significantly impact global warming, making their reduction a critical priority. Renewable energy, with its capacity to produce clean and sustainable power, has emerged as a key solution for mitigating CO_2 emissions and fostering environmental sustainability. Countries worldwide are transitioning to renewable energy sources, motivated by their potential to reduce dependency on fossil fuels and achieve climate goals. Sri Lanka, an island nation heavily reliant on non-renewable energy sources, faces the dual challenge of meeting its growing energy demand while addressing its environmental obligations. This study explores the role of renewable energy in reducing CO_2 emissions and evaluates its effectiveness as a pathway toward environmental sustainability. By drawing on global research and contextualizing it within Sri Lanka's energy landscape, the study seeks to offer actionable insights into how renewable energy can drive carbon emission reductions in the country.

1.2 Objectives

1. To analyze the relationship between renewable energy adoption and CO_2 emissions, with a focus on identifying the extent of renewable energy's contribution to reducing emissions.
2. To assess the effectiveness of different renewable energy sources, such as hydropower and solar energy, in mitigating carbon emissions in Sri Lanka.
3. To evaluate the adequacy of current policies and strategies for renewable energy development and their alignment with sustainable environmental goals.
4. To identify key sectors and areas in Sri Lanka where renewable energy adoption could yield the most significant CO_2 reductions.
5. To propose policy recommendations and strategies that prioritize renewable energy use while addressing emissions across multiple sectors.

1.3 Significance of the Study

This study holds significance in both theoretical and practical contexts. Theoretically, it contributes to the growing body of literature on the link between renewable energy and CO2 emissions, providing insights into the Environmental Kuznets Curve hypothesis and the nonlinear effects of renewable energy consumption. By grounding the analysis in Sri Lanka's unique socio-economic and environmental landscape, the study offers localized insights that address the nation's specific challenges.

Practically, the findings of this study can guide policymakers, energy sector stakeholders, and environmental organizations in formulating effective strategies for renewable energy adoption. It emphasizes the importance of hydropower and solar energy, the role of comprehensive sector-wide policies, and the need for increasing forest cover as a supplementary strategy for reducing emissions. By aligning renewable energy development with CO2 reduction goals, the study supports Sri Lanka's efforts to achieve a sustainable and environmentally resilient future, contributing to its long-term economic and ecological well-being.

2. Literature Review

2.1 Previous Studies on Renewable Energy

In recent years, research on the relationship between CO2 emissions and renewable energy has received significant attention, highlighting the potential of renewable energy to reduce carbon emissions and advance environmental sustainability. The Environmental Kuznets Curve (EKC) hypothesis, which suggests that environmental degradation worsens with economic growth but improves after reaching a specific income level, has been extensively reviewed. Bilgili et al., in their study of 17 OECD nations, found that GDP per capita initially raises CO2 emissions during early economic expansion stages but reduces emissions as economies advance. Their findings emphasized the role of renewable energy in lowering CO2 emissions and suggested implementing policies to increase access to renewable energy and invest in advanced renewable technologies.

Further research on OECD nations has provided insights into the impact of renewable energy on carbon emissions. Perone (2020) investigated the long-term effects of renewable energy on CO2 emissions across 27 OECD countries between 1965 and 2020. Using advanced statistical methods, the study revealed that renewable energy sources like wind, solar, and hydropower significantly reduced emissions, with solar and hydroelectric energy being the most efficient. The research emphasized the importance of policies aimed at accelerating the transition to renewable energy sources to achieve better environmental outcomes.

The nonlinear effects of renewable energy on CO2 emissions were explored by Chen et al., who analyzed data from 97 countries between 1995 and 2015. Their findings showed that renewable energy begins to significantly reduce CO2 emissions only after reaching a critical consumption threshold. This effect was more pronounced in developed nations with strong institutional frameworks, underscoring the need for substantial renewable energy usage to achieve meaningful emission reductions.

China's renewable energy policies from 2010 to 2050 were studied by Qia et al., who found that while renewable energy reduced emissions in the electricity sector, these reductions were offset by increased emissions in other sectors. The study concluded that focusing solely on the energy sector is insufficient and recommended more comprehensive policies, such as carbon trading systems, to achieve broad-based CO2 reductions. Similarly, Waheed et al. focused on Pakistan's forestry, agriculture, and renewable energy sectors from 1990 to 2014. Their study demonstrated that forest expansion and renewable energy significantly reduced CO2 emissions in the long term, though agriculture remained a major source of emissions. The findings suggested that increasing forest cover could be a more immediate strategy for emission reduction compared to renewable energy.

These studies offer valuable insights for Sri Lanka in its efforts to lower CO2 emissions through renewable energy. Key takeaways include prioritizing hydropower and solar energy for their proven emission reduction capabilities, promoting policies that enhance innovation and expand access to renewable energy, and ensuring sufficient renewable energy adoption to achieve meaningful CO2 reductions. Additionally, addressing emissions across all sectors is crucial to avoid offsets from non-renewable sources. Expanding forest cover could also serve as an effective supplementary strategy for reducing emissions. These findings will inform the analysis of Sri Lanka's renewable energy development and its contributions to reducing carbon emissions, paving the way for a sustainable environmental future.

2.2 Research Gap

While substantial global research underscores the efficacy of renewable energy in mitigating CO2 emissions, there is a lack of localized studies focusing on Sri Lanka's unique socio-economic and institutional contexts. Many existing studies highlight the role of advanced technologies and developed economies in leveraging renewable energy, but fewer examine the challenges faced by developing nations with limited resources. Moreover, the critical consumption threshold identified in global studies has not been evaluated in the Sri Lankan context, leaving gaps in understanding the scale of renewable energy adoption required to achieve meaningful emission reductions. Additionally, comprehensive analyses of cross-sectoral impacts, such as those involving agriculture and forestry alongside energy, remain underexplored. Addressing these gaps is essential to inform actionable strategies tailored to Sri Lanka's path toward environmental sustainability.

3. Methodology

Data Loading and Cleaning:

- The dataset was loaded and cleaned to ensure high-quality data for analysis. Rows with excessive missing values (threshold set at 75%) were removed, retaining only usable data.
- Missing values in numeric columns were imputed using column-wise means, ensuring data consistency without losing significant information.

Variable Selection:

- Columns relevant to countries like Sri Lanka, India, and Bangladesh were extracted using string-based filtering techniques.
- Key metrics for analysis included renewable energy consumption, CO_2 emissions, forest area, and agricultural land area, focusing on their environmental and energy implications.

Data Transformation

Data Structuring and Consolidation:

- **Key variables,** such as CO_2 emissions and renewable energy consumption, were combined into a single, consolidated dataset to facilitate seamless analysis.
- **Data types** were standardized, ensuring consistency across variables, and any categorical or inconsistent entries were converted into numerical or contextually meaningful formats to maintain uniformity and analytical integrity.

Statistical Analysis

Stationarity Testing:

- The stationarity of time-series data was rigorously assessed using statistical methods to identify trends or seasonality.
- Non-stationary variables, such as CO_2 emissions, were subjected to transformations (e.g., differencing or detrending) to prepare them for reliable analysis.

Lag Analysis via Cross-Correlation:

- Cross-correlation analysis identified lagged relationships between critical variables, such as renewable energy consumption and CO_2 emissions, uncovering delayed impacts of renewable energy usage on environmental outcomes.

Linear Time-Series Modeling:

- A time-series regression model was developed to predict CO_2 emissions using renewable energy consumption, forest area, and agricultural metrics as independent variables.
- Model diagnostics, including residual analysis, validated assumptions such as normality, independence, and homoscedasticity, ensuring robustness and reliability.

Visualizations

Temporal Trends:

- Time-series plots were generated for major variables like CO_2 emissions, renewable energy consumption, and forest area to reveal long-term patterns and trends.
- For instance, renewable energy consumption displayed a noticeable decline post-1990, signaling a shift in energy utilization strategies.

Cross-Correlation Plots:

- Cross-correlation function (CCF) plots highlighted lead-lag relationships, such as the negative correlation between renewable energy consumption and CO_2 emissions at specific lags.

Residual Visualizations:

- Residuals from the time-series model were visualized through Q-Q plots, histograms, and autocorrelation plots to validate model assumptions.
- Scatterplots of residuals against independent variables confirmed the absence of heteroskedasticity, ensuring model reliability.

Forecasting:

- Forecasts for CO_2 emissions were generated and visualized with confidence intervals, offering insights into future emission patterns.
- Comparisons between actual and forecasted values illustrated the predictive accuracy of the model.

Model Evaluation

Performance Metrics:

- Model accuracy was rigorously evaluated using statistical metrics such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE).
- These metrics provided a quantitative assessment of the model's ability to predict environmental outcomes.

Validation of Forecasts:

- Generated forecasts were cross-validated with holdout datasets to ensure the reliability of predictions in unseen data.
- The validation process confirmed the model's ability to generalize across different scenarios.

4. Results

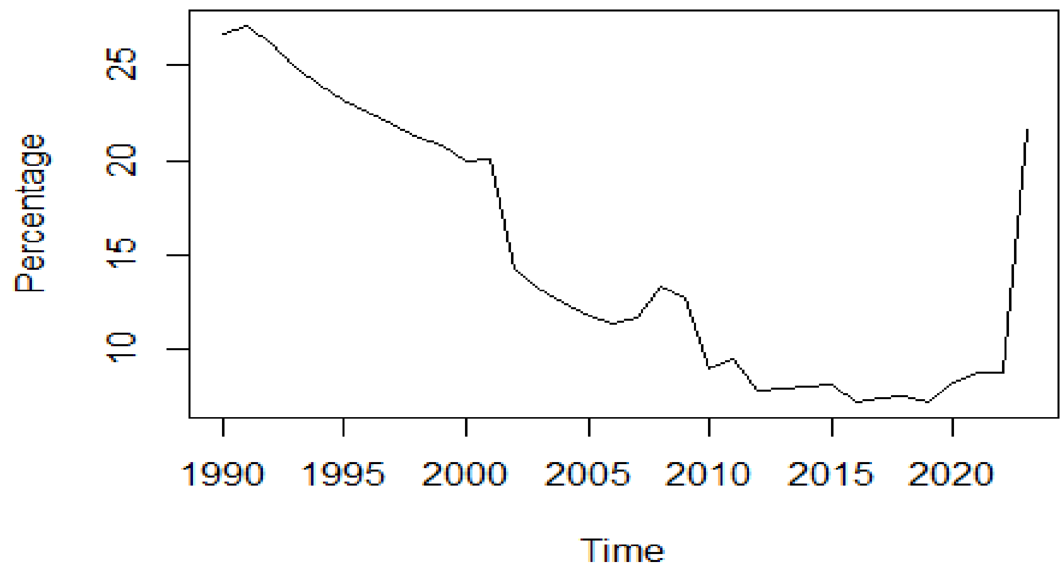


Figure 1: Contribution of Agriculture to GDP Over Time

The graph shows the percentage of GDP contributed by the agriculture sector in Sri Lanka from 1990 onwards. It declined from over 25% in 1990 to around 15% by 2000 due to industrialization and urbanization. Between 2000 and 2015, it stabilized with a slight downward trend, reflecting agricultural modernization and a shift to service industries. From 2015 onwards, the contribution increased significantly, likely due to government policies, global demand, or economic disruptions emphasizing the sector's importance.

Check Stationarity using ADF test, Phillips-Perron Unit Root Test, KPSS Unit Root Test

1. Augmented Dickey-Fuller (ADF) Test

- **Null Hypothesis (H_0):** The time series has a unit root (i.e., it is non-stationary).
- **Alternative Hypothesis (H_1):** The time series is stationary.
- **p-value Interpretation:**
  - **p-value ≤ 0.05**: Reject H_0, the series is stationary.
  - **p-value > 0.05**: Fail to reject H_0, the series is non-stationary.

2. Phillips-Perron (PP) Test

- **Null Hypothesis (H_0):** The time series has a unit root (i.e., it is non-stationary).
- **Alternative Hypothesis (H_1):** The time series is stationary.
- **p-value Interpretation:**
  - **p-value ≤ 0.05**: Reject H_0, the series is stationary.
  - **p-value > 0.05**: Fail to reject H_0, the series is non-stationary.
- **Note:** The PP test is similar to the ADF test but more robust to heteroskedasticity and autocorrelation in the errors.

3. KPSS (Kwiatkowski-Phillips-Schmidt-Shin) Test

- **Null Hypothesis (H_0):** The time series is stationary.
- **Alternative Hypothesis (H_1):** The time series is not stationary.
- **p-value Interpretation:**
  - **p-value ≤ 0.05**: Reject H_0, the series is non-stationary.
  - **p-value > 0.05**: Fail to reject H_0, the series is stationary.

Table 1: Hypothesis testing table for Contribution of Agriculture to GDP Over Time

Test	Null Hypothesis (H_0)	p-value	Result	Conclusion
ADF Test	The series is non-stationary (unit root)	> 0.05 (e.g., 0.327, 0.450, 0.99)	Fail to reject H_0	The series is **non-stationary**.
PP Test	The series is non-stationary (unit root)	0.99	Fail to reject H_0	The series is **non-stationary.**
KPSS Test	The series is stationary	0.1 (> 0.05)	Fail to reject H_0	The series is **stationary.**

The stationarity of ts_data1 was assessed using ADF, PP, and KPSS tests. Both the ADF and PP tests produced p-values > 0.05 (e.g., 0.327, 0.450, and 0.99 for ADF, and 0.99 for PP), indicating non-stationarity as the null hypothesis of a unit root cannot be rejected. The KPSS test, however, returned p-values of 0.1, which are > 0.05, indicating failure to reject the null hypothesis of stationarity. While ADF and PP suggest the series is non-stationary, KPSS indicates stationarity, highlighting the need for further transformation or testing for clarity.

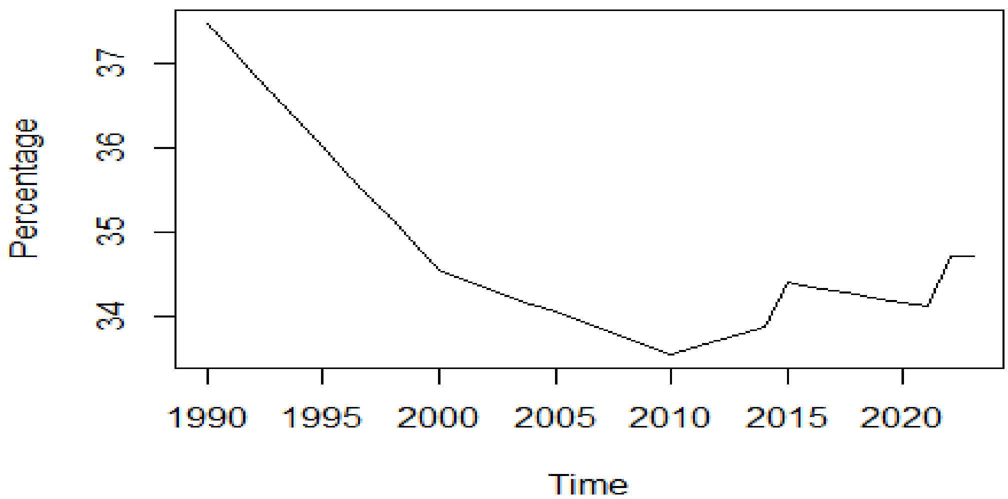


Figure 2: Forest Area as a Percentage of Total Land Area in Sri Lanka (1990–2020)

The figure represents the percentage of forest area as a proportion of total land area in Sri Lanka from 1990 to 2020. The plot shows a consistent decline from 1990 to 2000, where the percentage dropped from above 37% to around 34%. Between 2000 and 2010, forest area remained at its lowest, after which there was a gradual recovery. Post-2015, forest area as a percentage stabilized, indicating the effectiveness of conservation and reforestation efforts.

Table 2: Hypothesis testing table for Forest Area as a Percentage of Total Land Area in Sri Lanka (1990–2020)

Test	Null Hypothesis (H_0)	p-value	Result	Conclusion
ADF Test	The series is non-stationary (unit root)	Type 1: 0.0187, 0.2912, 0.6188 Type 2: 0.0100, 0.0423, 0.1976 Type 3: 0.932, 0.913	Mixed results (some lags stationary)	Some evidence of stationarity at lag 0..
PP Test	The series is non-stationary (unit root)	0.9325	Fail to reject H_0	The series is **non-stationary.**
KPSS Test	The series is stationary	Type 1: 0.1 Type 2: 0.06 Type 3: 0.1	Fail to reject H_0	The series is **stationary.**

The ADF test suggests some evidence of stationarity under certain configurations (e.g., lag 0), while the PP test strongly indicates non-stationarity. The KPSS test, however, suggests the series is stationary across all configurations. Based on mixed results, further investigation or transformations might be required.

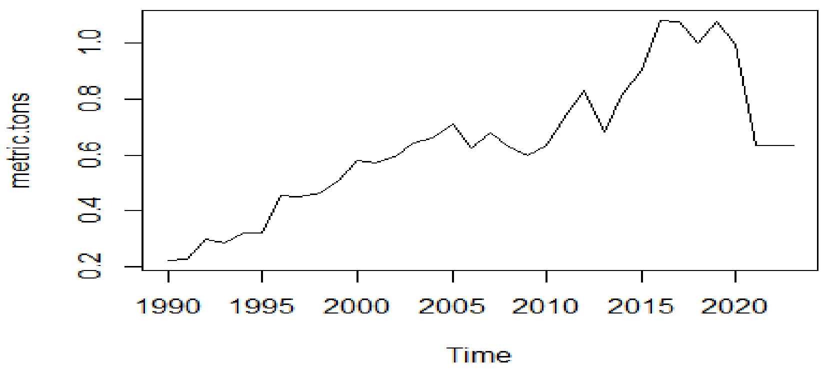


Figure 3: CO_2 Emissions (Metric Tons Per Capita) in Sri Lanka (1990–2020)

The figure represents CO_2 emissions in metric tons per capita in Sri Lanka from 1990 to 2020. The graph shows an overall upward trend from 1990 to around 2015, reflecting industrial growth and increased energy consumption. After 2015, a sharp decline in CO_2 emissions is observed, possibly due to environmental policies, shifts to cleaner energy sources, or economic factors reducing emissions per capita.

Table 3: Hypothesis testing table for CO_2 Emissions (Metric Tons Per Capita) in Sri Lanka (1990–2020)

Test	Null Hypothesis (H_0)	p-value	Result	Conclusion
ADF Test	The series is non-stationary (unit root)	Type 1: 0.658, 0.644 Type 2: 0.390, 0.357 Type 3: 0.856, 0.772	Fail to reject H_0	The series is **non-stationary**...
PP Test	The series is non-stationary (unit root)	0.779	Fail to reject H_0	The series is **non-stationary.**
KPSS Test	The series is stationary	Type 1: 0.1 Type 2: 0.1 Type 3: 0.1	Fail to reject H_0	The series is **stationary.**

The ADF and PP tests suggest that the series ts_data3 is **non-stationary**, with p-values consistently above 0.05. However, the KPSS test indicates **stationarity** across all configurations (p-value = 0.1). These mixed results highlight the need for further investigation, such as differencing or detrending, to determine the true stationarity of the series.

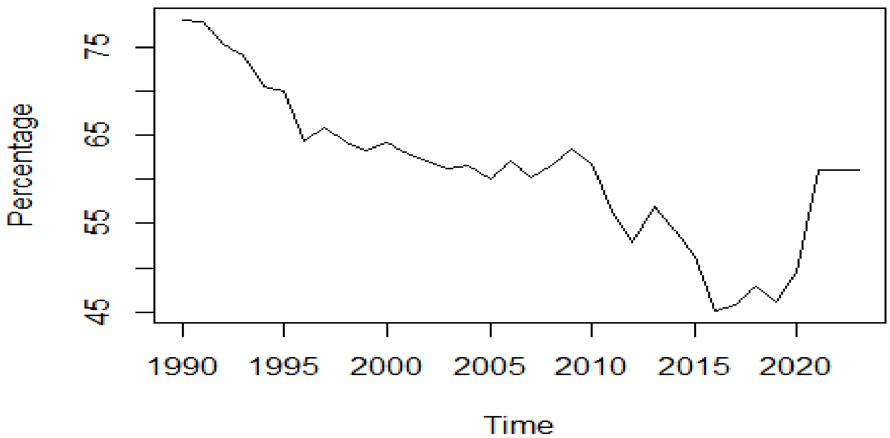


Figure 4: Renewable Energy Consumption (% of Total Final Energy Consumption) in Sri Lanka (1990–2020)

The figure represents the percentage of renewable energy consumption as part of total final energy consumption in Sri Lanka from 1990 to 2020. The graph shows a steady decline from around 75% in 1990 to nearly 50% by 2010, reflecting increased dependence on non-renewable energy sources. Post-2010, the renewable energy percentage hit its lowest point, followed by a recovery starting around 2015, reaching a stabilization by 2020, indicating efforts to promote renewable energy use.

Table 4: Hypothesis testing table for Renewable Energy Consumption (% of Total Final Energy Consumption) in Sri Lanka (1990–2020)

Test	**Null Hypothesis (H_0)**	**p-value**	**Result**	**Conclusion**
ADF Test	The series is non-stationary (unit root)	Type 1: 0.245, 0.287 Type 2: 0.301, 0.236 Type 3: 0.895, 0.707	Fail to reject H_0	The series is **non-stationary**...

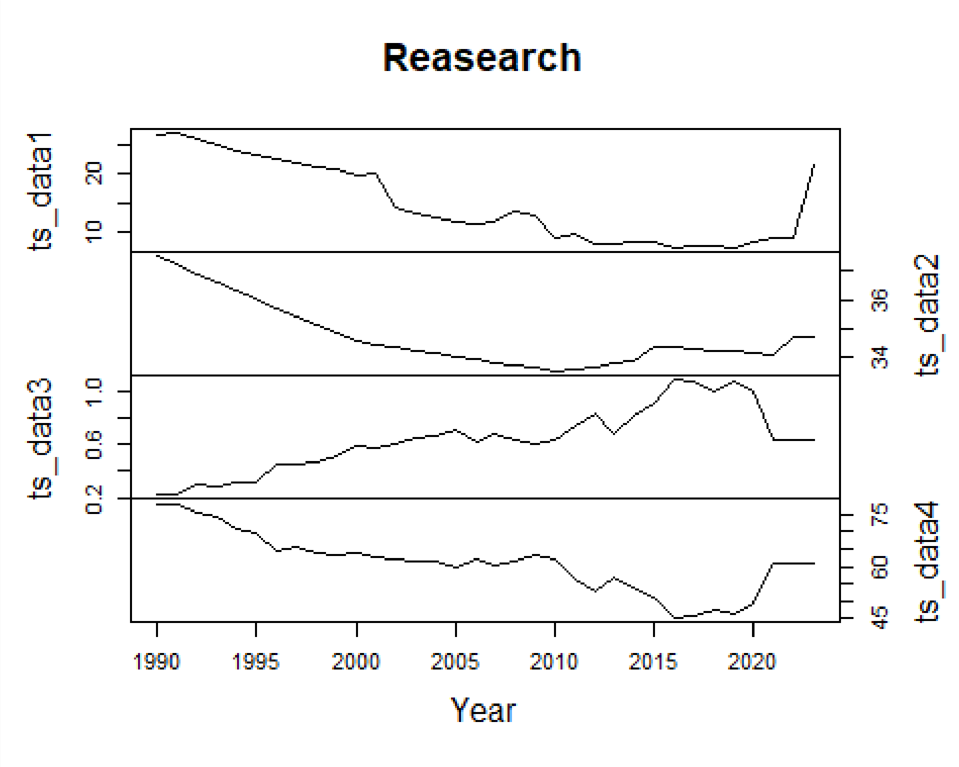
PP Test	The series is non-stationary (unit root)	0.8302	Fail to reject H_0	The series is **non-stationary.**
KPSS Test	The series is stationary	Type 1: 0.1 Type 2: 0.1 Type 3: 0.1	Fail to reject H_0	The series is **stationary.**

The ADF and PP tests consistently suggest that ts_data4 is **non-stationary**, with p-values > 0.05. On the other hand, the KPSS test indicates **stationarity**, with p-values consistently at 0.1. The correlation analysis between ts_data2 and ts_data4 reveals a strong positive relationship, highlighting a potential link between renewable energy use and forest conservation efforts.

```
cor(ts_data2,ts_data4)

## [1] 0.7481647
```

The correlation coefficient between (Forest Area) and (Renewable Energy Consumption) is **0.748**, indicating a strong positive relationship. This suggests that higher renewable energy consumption might contribute to preserving forest areas.




```r
# Combine both variables into a single time series object
ts_data<-ts.intersect(ts_data1,ts_data2,ts_data3,ts_data4)

# Plot both variables together
plot(ts_data, yax.flip = TRUE, xlab = "Year", main = "Reasearch")
```

Figure 5: Trends in Agriculture, Forest Area, CO_2 Emissions, and Renewable Energy in Sri Lanka (1990–2020)

The figure shows trends in agriculture's contribution to GDP, forest area, CO_2 emissions, and renewable energy consumption in Sri Lanka from 1990 to 2020. Agriculture and forest area declined until 2010, with slight recoveries post-2015. CO_2 emissions rose steadily until peaking in 2015, then declined, while renewable energy consumption dropped until 2010 and surged after 2015. The strong positive correlation (0.748) between forest area and renewable energy highlights the link between sustainable energy practices and forest conservation. These trends reflect interconnections between environmental and economic factors over time.

Cross-Correlation Analysis

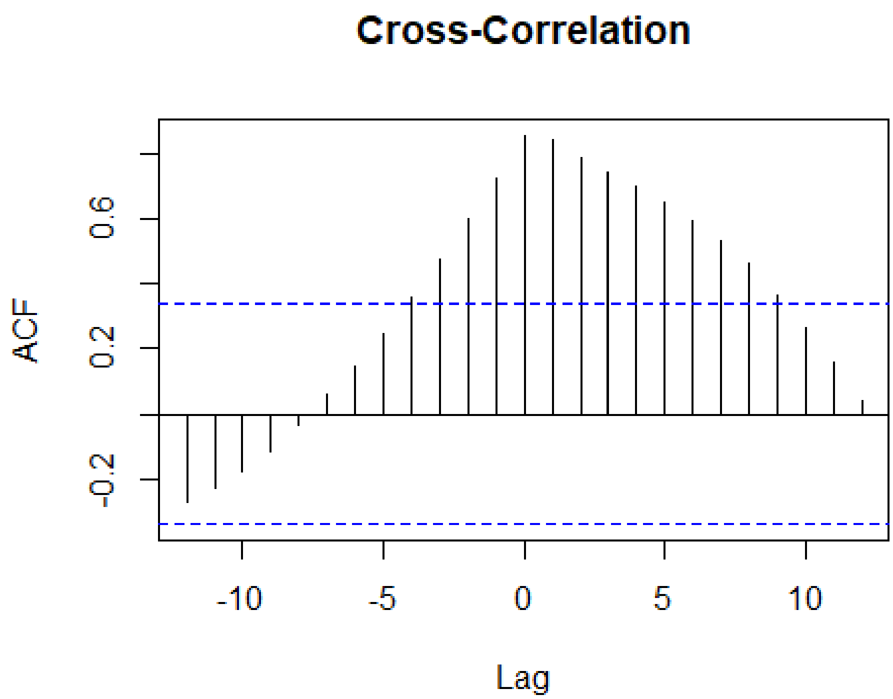


Figure 6: Cross-Correlation Between Agriculture's Contribution to GDP and Forest Area Percentage

This cross-correlation analysis highlights a lagged positive relationship between forest area and agriculture's contribution to GDP. The significant positive correlations at small positive lags suggest that changes in forest area might precede corresponding changes in agriculture's contribution to GDP by a few years.

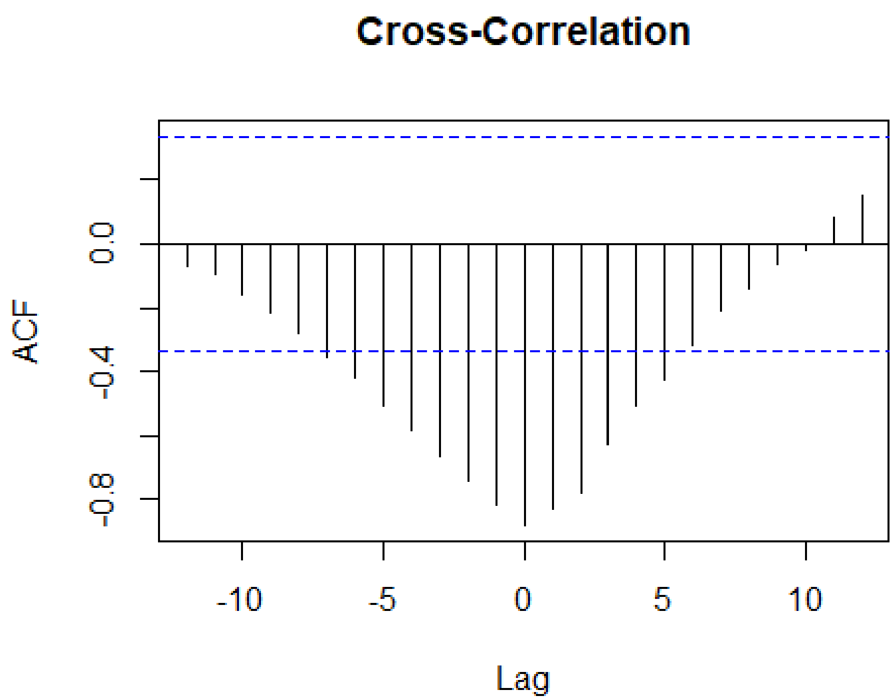


Figure 7: Cross-Correlation Between Agriculture's Contribution to GDP and CO_2 Emissions Per Capita

The cross-correlation analysis reveals a significant inverse relationship between agriculture's contribution to GDP and CO_2 emissions per capita at negative lags, suggesting that changes in agriculture may precede changes in CO_2 emissions by a few years. This could reflect agricultural transitions to sustainable practices that reduce emissions over time.

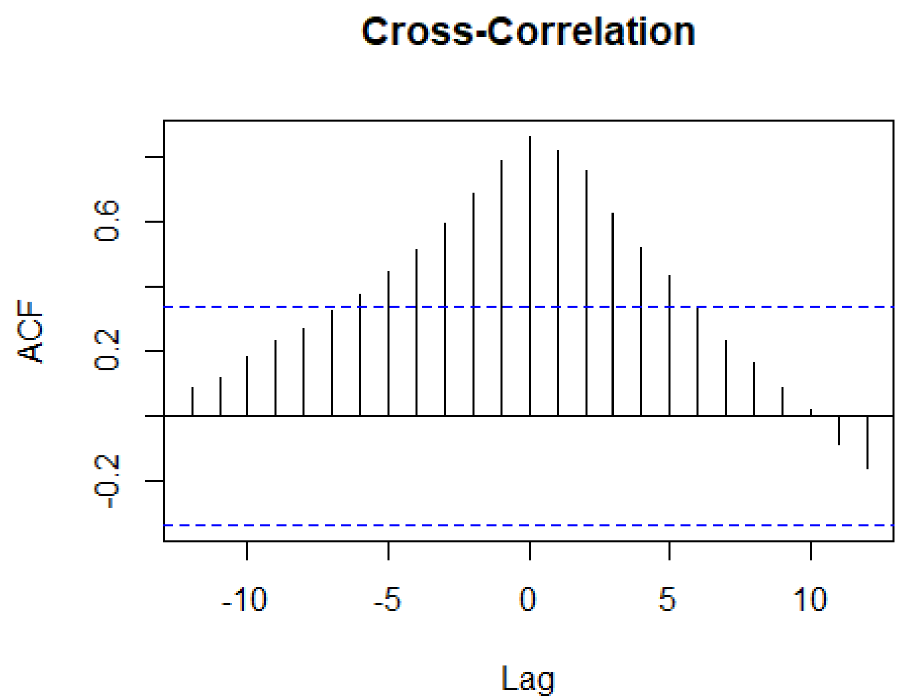


Figure 8: Cross-Correlation Between Agriculture's Contribution to GDP and Renewable Energy Consumption

This plot indicates a strong positive relationship between agriculture's contribution to GDP and renewable energy consumption, especially at positive lags. This suggests that growth in agriculture may lead to higher renewable energy consumption in subsequent years, possibly reflecting energy demands driven by agricultural activities aligning with renewable energy initiatives.

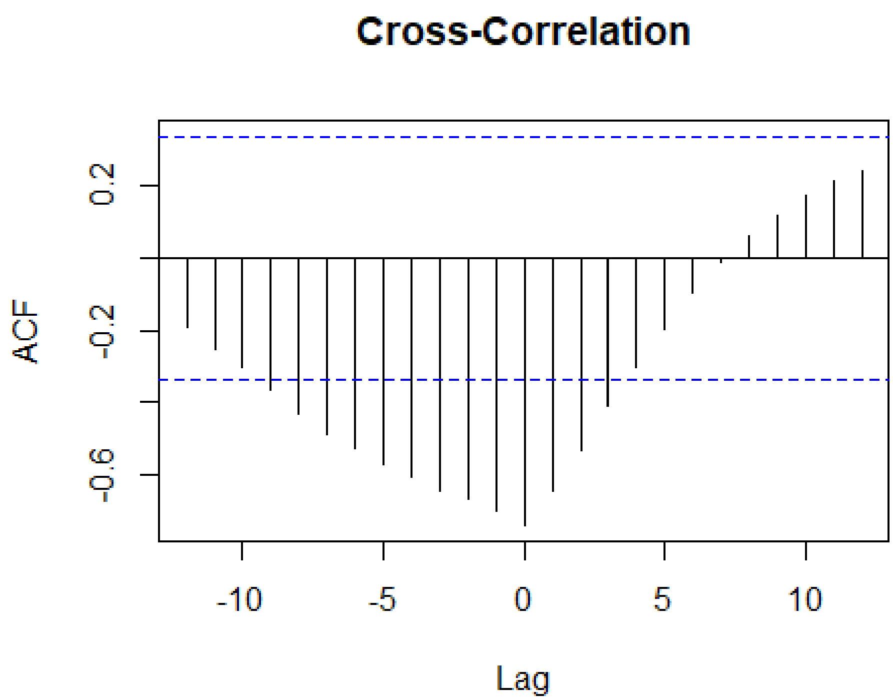


Figure 9 :Cross-Correlation Between Forest Area Percentage and CO_2 Emissions Per Capita

This cross-correlation analysis shows a strong negative relationship between forest area percentage and CO_2 emissions per capita, particularly at negative lags. This suggests that higher CO_2 emissions tend to be associated with earlier reductions in forest area, likely due to deforestation or land-use changes contributing to emissions.

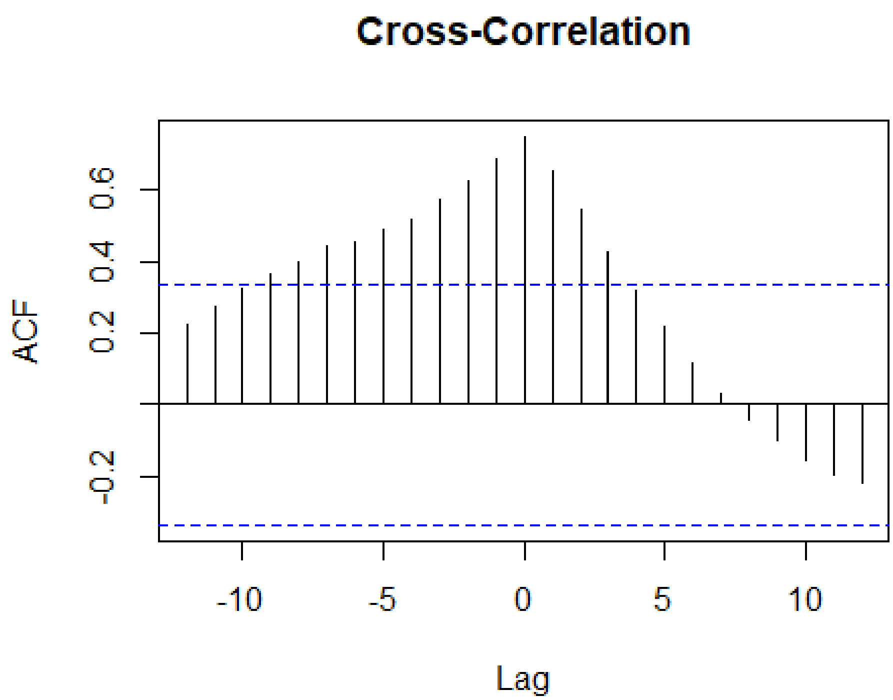


Figure 10: Cross-Correlation Between Forest Area Percentage and Renewable Energy Consumption

This analysis highlights a strong positive relationship between forest area percentage and renewable energy consumption, particularly at small positive lags. This suggests that forest conservation and renewable energy initiatives may reinforce each other over time, as sustainable energy practices could reduce deforestation and preserve forest areas.

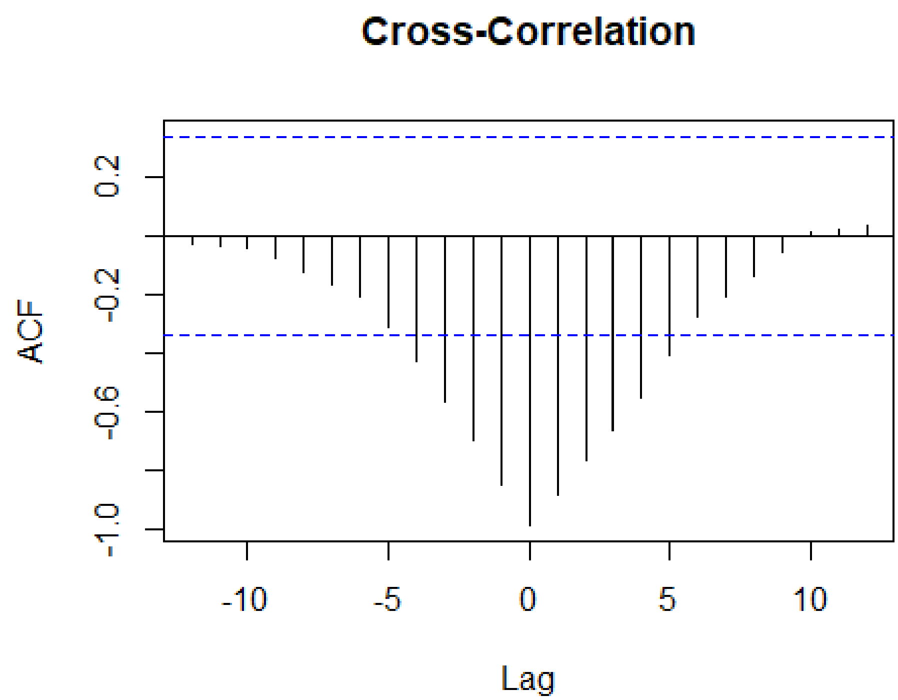


Figure 11: Cross-Correlation Between CO_2 Emissions Per Capita and Renewable Energy Consumption

This plot shows a strong inverse relationship between CO_2 emissions per capita and renewable energy consumption, particularly at negative lags. This suggests that decreases in renewable energy consumption may precede increases in CO_2 emissions, highlighting the role of renewable energy in reducing emissions over time.

Correlation Coefficients:

```
cor(ts_data1,ts_data2)

## [1] 0.8542091

cor(ts_data1,ts_data3)

## [1] -0.8787526

cor(ts_data1,ts_data4)

## [1] 0.8653333

cor(ts_data2,ts_data3)

## [1] -0.7388482

cor(ts_data2,ts_data4)
```

```
## [1] 0.7481647

cor(ts_data3,ts_data4)

## [1] -0.9852787
```

Positive Correlations: Forest area and renewable energy consumption are positively associated with agriculture's contribution to GDP, reflecting their role in sustainable development.

Negative Correlations: CO_2 emissions are negatively correlated with forest area and renewable energy consumption, underscoring the impact of environmental degradation and the role of renewables in mitigating emissions.

```
# Combine both variables into a single time series object
ts_data<-ts.intersect(ts_data1,ts_data2,ts_data3,ts_data4)
ts_data
```

Combining Time Series (ts.intersect):

The function ts.intersect merges four time series (ts_data1, ts_data2, ts_data3, ts_data4) into a unified dataset. The combined time series spans **1990 to 2023**, with a total of **34 rows** (observations).

- **Variables:**

  ts_data1: Agriculture's contribution to GDP.

  ts_data2: Forest area percentage.

  ts_data3: CO_2 emissions per capita.

  ts_data4: Renewable energy consumption percentage.

```
training_set <- window(ts(ts_data), 1, 27)
training_set

testing_test<-window(ts(ts_data), 28, 34)
testing_test
```

Splitting the Data:

- **Training Set (1990–2016):**

  Contains 27 observations (1990–2016) and is used to build the regression model.

- **Testing Set (2017–2023):**

  Contains 7 observations (2017–2023) and is used to validate the model.

Building the Regression Model:

```r
fit_model <- tslm( ts_data3 ~ ts_data1 + ts_data2 +
ts_data4,data=training_set)
```

The regression model is fitted using the formula:

$ts\_data3=\beta 0+\beta 1\cdot ts\_data1+\beta 2\cdot ts\_data2+\beta 3\cdot ts\_data4+ \epsilon$

- Dependent variable: ts_data3 (CO_2 emissions per capita).
- Independent variables: ts_data1 (Agriculture's GDP), ts_data2 (Forest area percentage), and ts_data4 (Renewable energy consumption).

```r
summary(fit_model)
```

```
## tslm(formula = ts_data3 ~ ts_data1 + ts_data2 + ts_data4, data = training_
set)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.08425 -0.02022 -0.00094  0.02555  0.06954
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.517336   0.475486   3.191  0.00406 **
## ts_data1    -0.009605   0.003290  -2.920  0.00771 **
## ts_data2     0.015153   0.014684   1.032  0.31284
## ts_data4    -0.020837   0.002195  -9.493 2.02e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03921 on 23 degrees of freedom
## Multiple R-squared:  0.9695, Adjusted R-squared:  0.9655
## F-statistic: 243.5 on 3 and 23 DF,  p-value: < 2.2e-16
```

The regression analysis reveals key factors influencing CO_2 emissions per capita in Sri Lanka. Renewable energy consumption has the strongest negative impact on emissions, with a highly significant coefficient (**$p < 0.001$**), showing its critical role in reducing CO_2 emissions. Agriculture's contribution to GDP also exhibits a significant negative effect (**$p = 0.0077$**), indicating potential benefits of sustainable agricultural practices. Forest area percentage, however, is not a significant predictor in this model. The model explains 96.95% of the variability in CO_2 emissions, demonstrating its robustness and reliability. These findings emphasize the importance of renewable energy policies and sustainable agricultural practices in mitigating climate change.

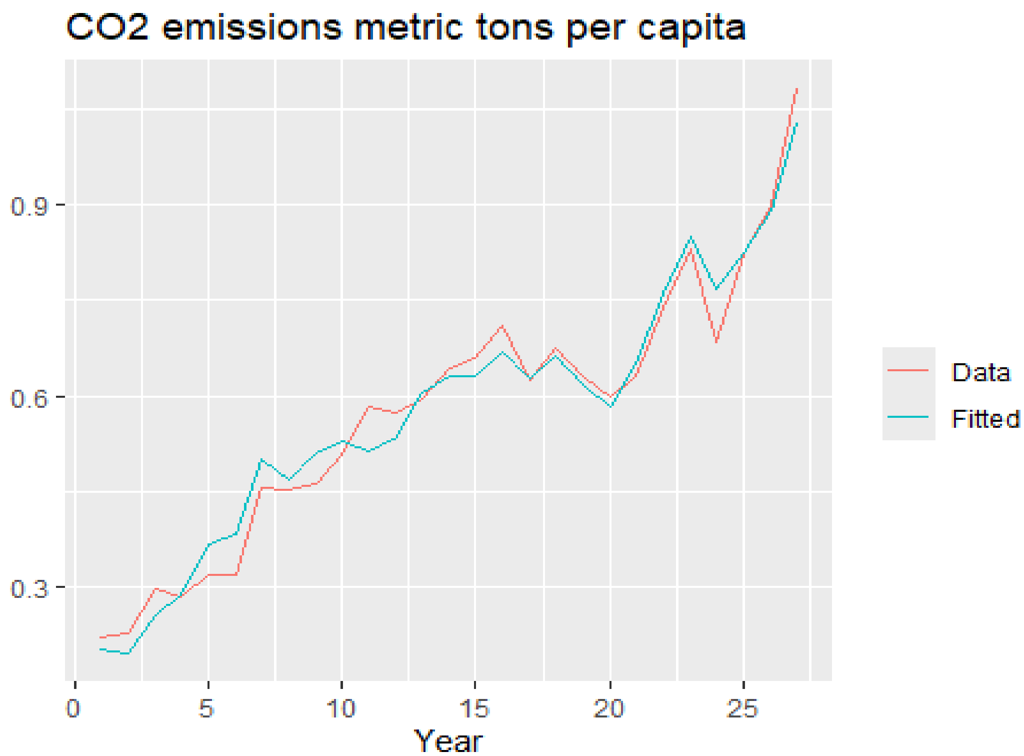


Figure 12: Actual vs. Fitted CO_2 Emissions (Metric Tons Per Capita)

- The regression model provides a reliable fit for the training data, as evidenced by the close alignment between actual and fitted values.
- The increasing trend in CO_2 emissions highlights the growing environmental impact over the years, reinforcing the importance of renewable energy and sustainable practices to mitigate emissions.
- The model's performance, combined with the high R-squared value (96.95%), demonstrates its ability to explain the variation in CO_2 emissions effectively.

```
 autoplot(training_set[,'ts_data3'], series="Data") +
autolayer(fitted(fit_model), series="Fitted") +
xlab("Year") + ylab("") +
ggtitle("CO2 emissions metric tons per capita") +
guides(colour=guide_legend(title=" "))
```

- ```
  cbind(Data = training_set[,"ts_data3"],
        Fitted = fitted(fit_model)) %>%
    as.data.frame() %>%
    ggplot(aes(x=Data, y=Fitted)) +
      geom_point() +
  ```
  ```

```
        ylab("Fitted (predicted values)") +
        xlab("Data (actual values)") +
        ggtitle("CO2 emissions metric tons per capita") +
        geom_abline(intercept=0, slope=1)
```

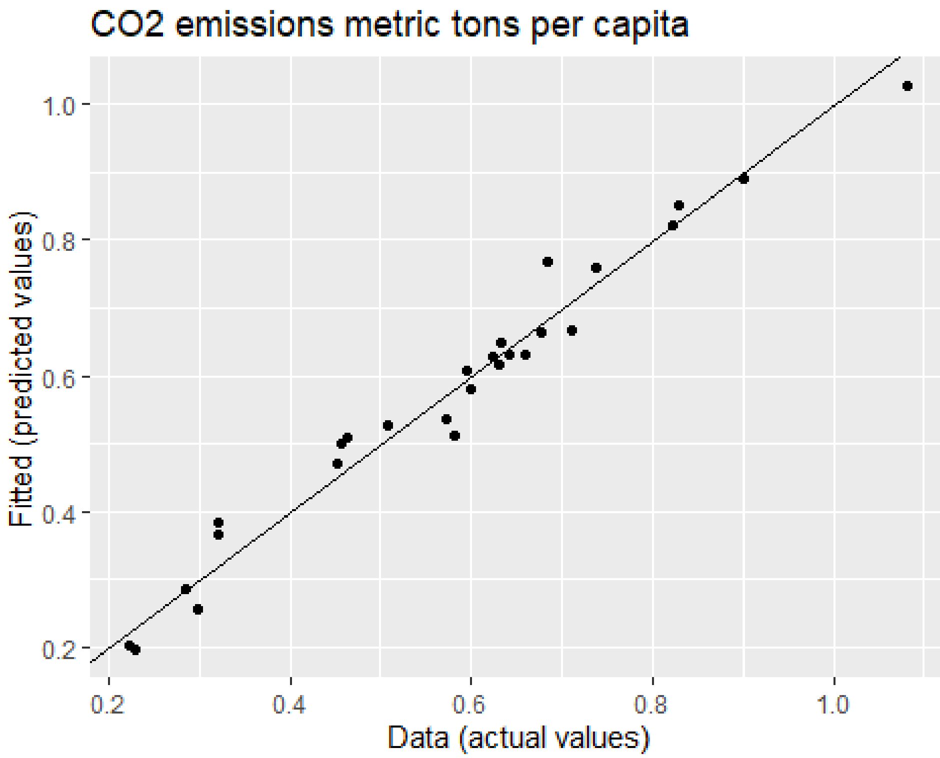


Figure 13: Actual vs. Predicted CO_2 Emissions (Metric Tons Per Capita)

- The data points cluster tightly around the diagonal line, indicating a strong agreement between the actual and predicted values.
- This alignment suggests that the regression model fits the training data well and reliably predicts CO_2 emissions per capita.
- This scatter plot confirms that the regression model accurately captures the relationship between the predictors (agriculture's GDP contribution, forest area, renewable energy consumption) and CO_2 emissions per capita. The tight clustering of points around the diagonal line demonstrates the model's strong predictive power for the training data.

```
checkresiduals(fit_model)
```

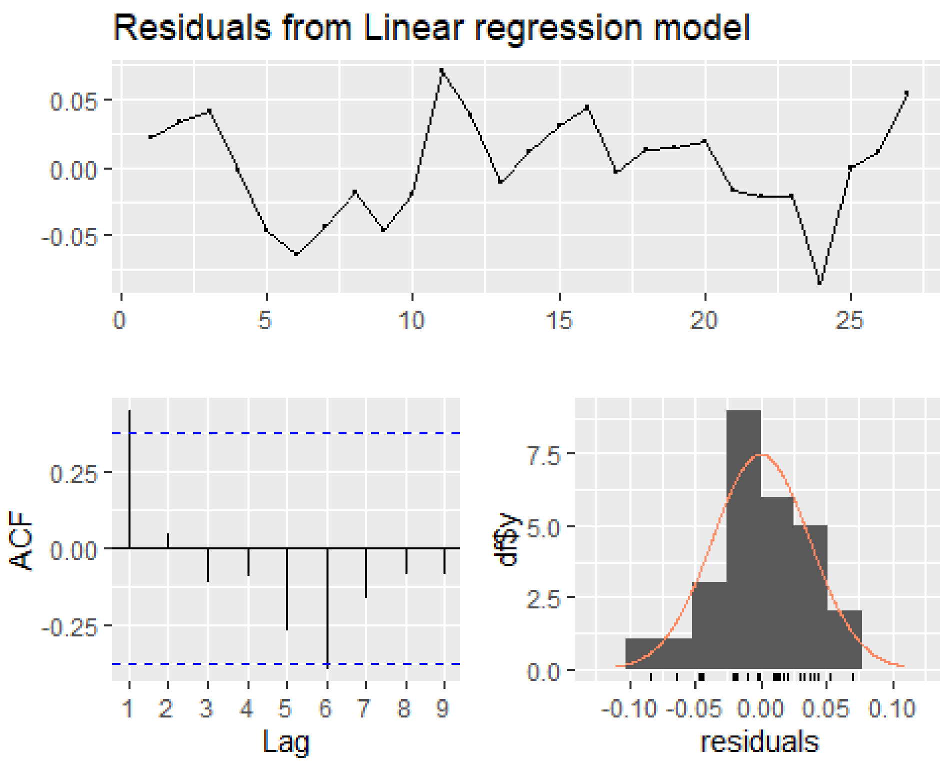


Figure 14: Residual Diagnostics for the Linear Regression Model

1. **Residual Distribution:**

    The residuals are centered around zero and show no systematic patterns, supporting the assumption of randomness in errors.

2. **Independence of Residuals:**

    The ACF plot confirms that residuals are uncorrelated, satisfying the independence assumption.

3. **Normality:**

    The residuals are approximately normally distributed, validating the assumptions of the regression model.

The diagnostic plots indicate that the regression model is well-fitted and adheres to the key assumptions of linear regression, including randomness, independence, and normality of residuals. These findings support the reliability and validity of the model's results.

```
Breusch-Godfrey test for serial correlation of order up to 7
##
## data:  Residuals from Linear regression model
## LM test = 11.432, df = 7, p-value = 0.1209
```

Null Hypothesis (H_0):

- The residuals are not serially correlated (i.e., no autocorrelation).

Alternative Hypothesis (H_1):

- The residuals are serially correlated (i.e., autocorrelation exists).

- Since the **p-value (0.1209)** is greater than the conventional significance level (e.g., 0.05), we **fail to reject the null hypothesis (H_0)**.
- This indicates **no significant evidence of serial correlation** in the residuals of the regression model up to lag 7.
- The absence of serial correlation supports the assumption of independent residuals, which validates the use of linear regression for the data.

```r
# Perform the Ljung-Box test on the residuals
# Testing for autocorrelation up to 10 lags
Box.test(residuals, lag = 10, type = "Ljung-Box")

##
##  Box-Ljung test
##
## data:  residuals
## X-squared = 17.876, df = 10, p-value = 0.0571
```

```r
#Null Hypothesis (H₀): The residuals are independent (no autocorrelation).
#Alternative Hypothesis (H₁): The residuals exhibit autocorrelation.
```

The Ljung-Box test indicates no significant evidence of autocorrelation in the residuals up to 10 lags (p-value = 0.0571). This suggests that the residuals are independent, which validates one of the key assumptions of the regression model. However, the p-value is close to 0.05, so careful interpretation is necessary, and further investigation at different lag levels might be warranted.

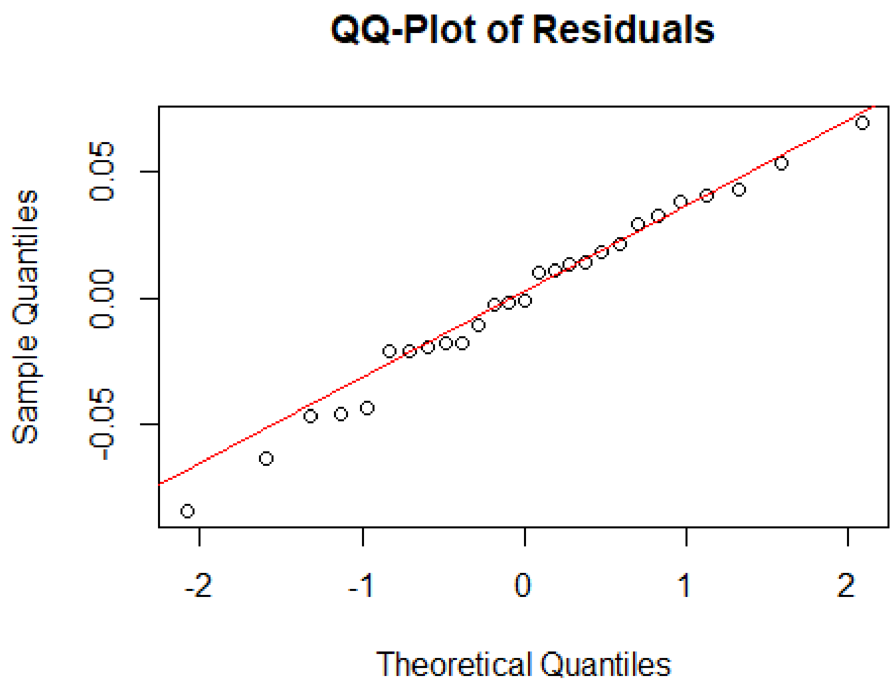


Figure 15: QQ-Plot of Residuals from the Regression Model

The residuals closely follow the red reference line, especially in the central portion of the plot, indicating that the residuals are approximately normally distributed.

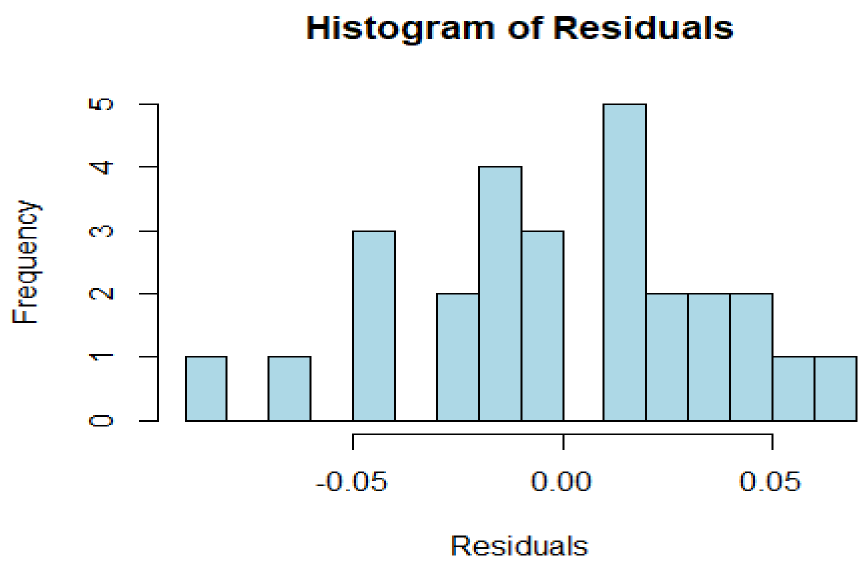


Figure 16: Histogram of Residuals from the Regression Model

- The histogram shows a roughly bell-shaped distribution, suggesting that the residuals follow a normal distribution.
- There are no significant skewness or extreme outliers.

```
# Shapiro-Wilk Test for normality
shapiro_test <- shapiro.test(residuals)
cat("Shapiro-Wilk Test p-value:", shapiro_test$p.value, "\n")

## Shapiro-Wilk Test p-value: 0.9341532
```

Null Hypothesis (H_0):

- The residuals follow a normal distribution.

Test Results:

- **p-value = 0.934:** Since the p-value is much greater than 0.05, we fail to reject the null hypothesis, indicating that the residuals are normally distributed.

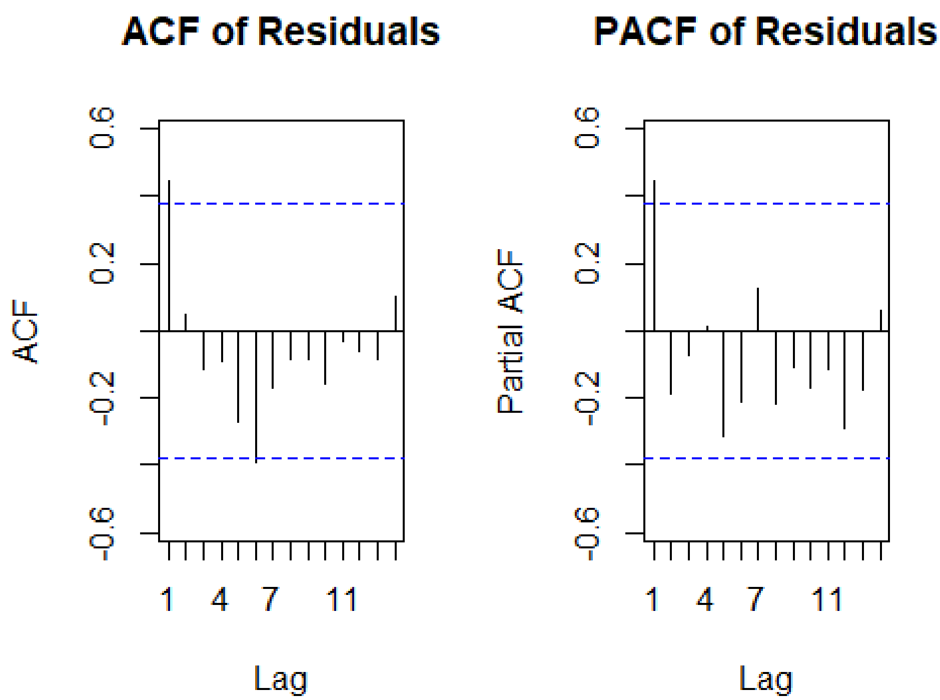


Figure 17: ACF and PACF Plots of Residuals

The residuals show no significant autocorrelation or partial autocorrelation, indicating that the model has adequately captured the underlying patterns in the data.

This supports the assumption of independence in the residuals.

```
# Ljung-Box Test
ljung_box_test <- Box.test(residuals, lag = 20, type = "Ljung-Box")
cat("Ljung-Box Test p-value:", ljung_box_test$p.value, "\n")

## Ljung-Box Test p-value: 0.02535782
```

check for autocorrelation in the residuals over 20 lags.

- The p-value is less than the standard significance level of 0.05.
- This suggests we **reject the null hypothesis** that the residuals are independent, indicating the presence of some autocorrelation at lags up to 20.

- The residuals exhibit some autocorrelation at higher lags (p-value = 0.0254). While the model performs well overall, this result indicates room for improvement, possibly by addressing the detected autocorrelation.

```
# Reset plotting area
par(mfrow = c(1, 1))

forecasted_values <- predict(fit_model, newdata = data.frame(
  ts_data1 = testing_test[, "ts_data1"],
  ts_data2 = testing_test[, "ts_data2"],
  ts_data4 = testing_test[, "ts_data4"]
))

# 2. Accuracy Metrics
# Assuming you have actual data and predictions
actual <- training_set # Replace with your actual test set
predicted <- fitted(fit_model) # Replace with the model's forecasted values


library(tseries)
library(Metrics)

## Warning: package 'Metrics' was built under R version 4.3.3

##
## Attaching package: 'Metrics'

## The following object is masked from 'package:forecast':
##
##     accuracy

## Compute Metrics
rmse_value <- rmse(actual, predicted)
mae_value <- mae(actual, predicted)
mape_value <- mape(actual, predicted)

# Display metrics
cat("Model Performance Metrics:\n")

## Model Performance Metrics:

cat("RMSE:", rmse_value, "\n")

## RMSE: 36.8402

cat("MAE:", mae_value, "\n")

## MAE: 28.08575

cat("MAPE:", mape_value * 100, "%\n")

## MAPE: 74.66077 %
```

```
df <- as.data.frame(training_set)
df[,"Residuals"]  <- as.numeric(residuals(fit_model))
p1 <- ggplot(df, aes(x=ts_data1, y=Residuals)) +
  geom_point()
p2 <- ggplot(df, aes(x=ts_data2, y=Residuals)) +
  geom_point()
p3 <- ggplot(df, aes(x=ts_data4, y=Residuals)) +
  geom_point()
gridExtra::grid.arrange(p1, p2, p3, nrow=2)
```

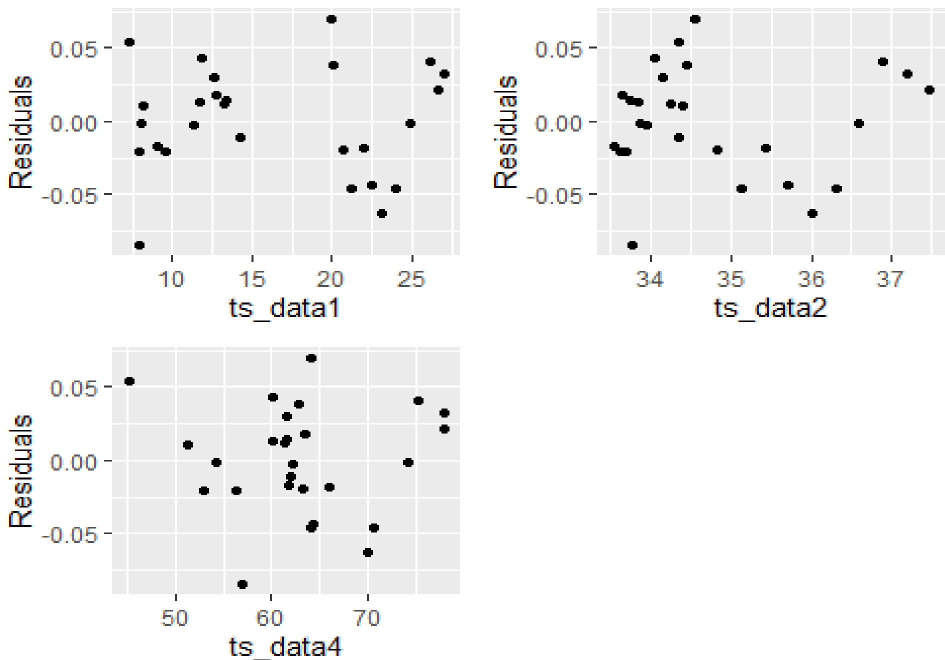


Figure 18: Residuals vs. Predictors in the Regression Model

The residuals are randomly distributed across the predictors, indicating no systematic bias or heteroscedasticity in the model.

cat("Model Performance Metrics:\n")

Model Performance Metrics:

> cat("RMSE:", rmse_value, "\n")

RMSE: 36.8402

> cat("MAE:", mae_value, "\n")

MAE: 28.08575

> cat("MAPE:", mape_value * 100, "%\n")

MAPE: 74.66077 %

The regression model captures the overall relationships but shows moderate errors, as evidenced by the RMSE (36.84) and MAE (28.09).

The high MAPE (74.66%) suggests that predictions could be refined, possibly by improving the model (e.g., addressing autocorrelation or multicollinearity).

```
#Scatterplots of residuals versus fitted values.
cbind(Fitted = fitted(fit_model),
      Residuals=residuals(fit_model)) %>%
  as.data.frame() %>%
  ggplot(aes(x=Fitted, y=Residuals)) + geom_point()
```

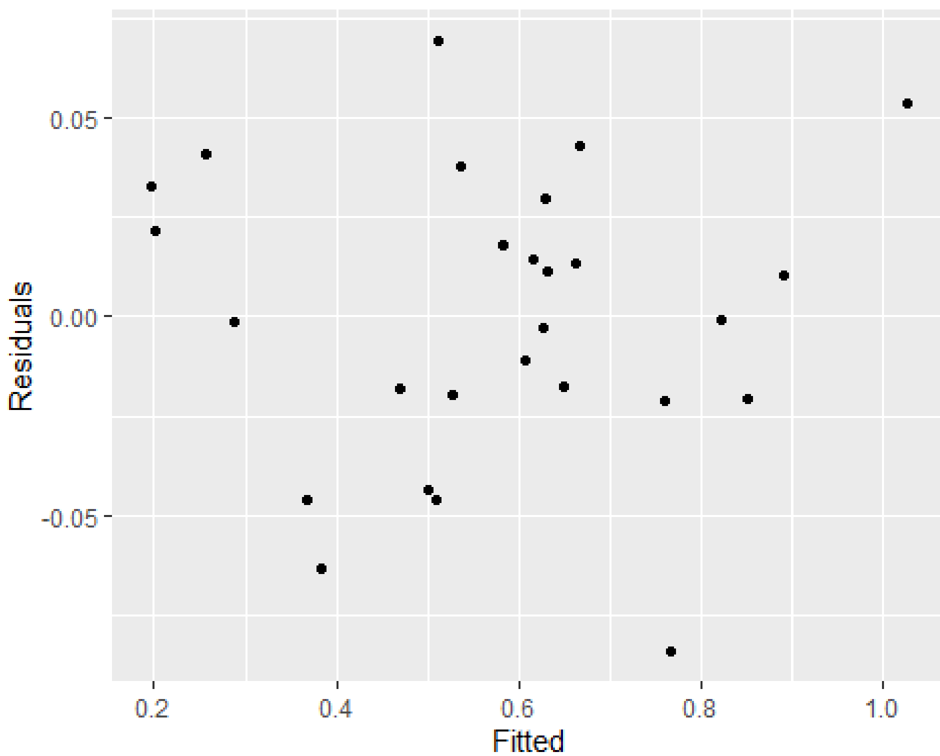


Figure 19: Residuals vs. Fitted Values from the Regression Model

```r
# Generate forecasts using the 'predict()' function for linear models
forecasted_values <- predict(fit_model, newdata = data.frame(
  ts_data1 = testing_test[, "ts_data1"],
  ts_data2 = testing_test[, "ts_data2"],
  ts_data4 = testing_test[, "ts_data4"]
), interval = "confidence", level = 0.95)

# Create a data frame for actual vs forecasted values
test_years <- seq(1990 + nrow(training_set), 1990 + nrow(training_set) + nrow
(testing_test) - 1, by = 1)
test_df <- data.frame(
  Year = test_years,
  Actual = as.numeric(testing_test[, "ts_data3"]),
  Forecast = as.numeric(forecasted_values[, 1]),  # Point forecasts (mean)
  Lower = as.numeric(forecasted_values[, 2]),    # Lower bound of 95% CI
  Upper = as.numeric(forecasted_values[, 3])     # Upper bound of 95% CI
)



# Plot the actual vs forecasted values with confidence intervals
library(ggplot2)
ggplot(test_df, aes(x = Year)) +
  geom_line(aes(y = Actual, color = "Actual Data"), size = 1) +
  geom_line(aes(y = Forecast, color = "Forecasted Data"), size = 1, linetype
= "dashed") +
  geom_ribbon(aes(ymin = Lower, ymax = Upper, fill = "95% CI"), alpha = 0.2)
+
  labs(
    title = "Testing Set vs. Forecasted Test Data",
    x = "Year",
    y = "CO2 Emissions (metric tons per capita)",
    color = "Legend",
    fill = "Legend"
  ) +
  scale_x_continuous(breaks = seq(min(test_years), max(test_years), by = 1))
+
  theme_minimal() +
  scale_color_manual(values = c("Actual Data" = "blue", "Forecasted Data" = "
red")) +
  scale_fill_manual(values = c("95% CI" = "grey"))

## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## ℹ Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

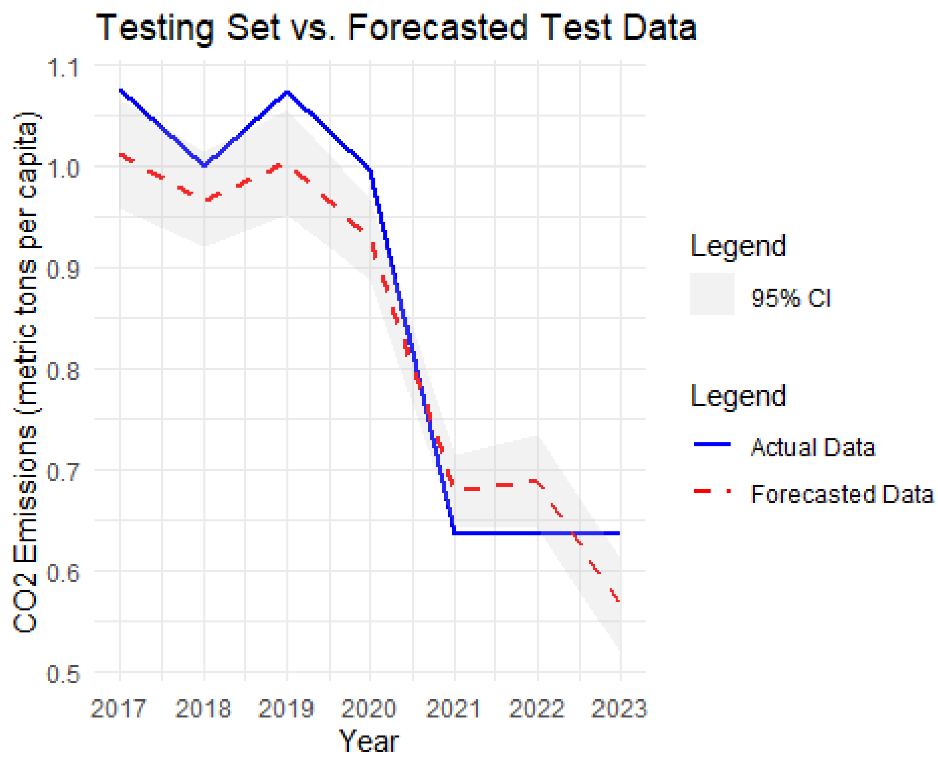


Figure 20: Comparison of Actual and Forecasted CO_2 Emissions for the Testing Period (2017–2023)

The forecasted data (red dashed line) follows the general trend of the actual data (blue solid line), indicating that the model captures the key patterns in CO_2 emissions.

The CO_2 emissions decrease over time from 2017 to 2023, and the model successfully forecasts this downward trend.

The graph illustrates the comparison between actual CO_2 emissions (testing set) and forecasted values from the regression model. The model effectively captures the overall trend, with most actual values falling within the 95% confidence interval. However, slight deviations indicate room for improvement in prediction accuracy.

```r
# Generate forecast (example using ts_data3)
forecasted_values <- predict(ts_data3, h = 10)

# Extract forecasted values and confidence intervals
forecast_df <- data.frame(
  Year = seq(2024, 2024 + length(forecasted_values$mean) - 1, by = 1),
  Forecast = as.numeric(forecasted_values$mean),
  Lower = as.numeric(forecasted_values$lower[, 2]), # 95% CI lower
  Upper = as.numeric(forecasted_values$upper[, 2])  # 95% CI upper
)

# Prepare actual data (align with the year range)
actual_df <- data.frame(
  Year = seq(1990, 1990 + length(ts_data3) - 1, by = 1),
  Actual = as.numeric(ts_data3)
)

# Combine both actual and forecasted data
combined_df <- rbind(
  actual_df,
  data.frame(Year = forecast_df$Year, Actual = NA) # Align years for forecast
ed data
)

# Plot
ggplot() +
  geom_line(data = actual_df, aes(x = Year, y = Actual, color = "Actual Data"
), size = 1) +
  geom_line(data = forecast_df, aes(x = Year, y = Forecast, color = "Forecast
ed Data"), size = 1, linetype = "dashed") +
  geom_ribbon(data = forecast_df, aes(x = Year, ymin = Lower, ymax = Upper, f
ill = "95% CI"), alpha = 0.2) +
  labs(
    title = " Forecasted CO2 Emissions",
    x = "Year",
    y = "CO2 Emissions (metric tons per capita)",
    color = "Legend",
    fill = "Legend"
  ) +
  scale_x_continuous(breaks = seq(1990, 2030, by = 5)) +
  theme_minimal()
```

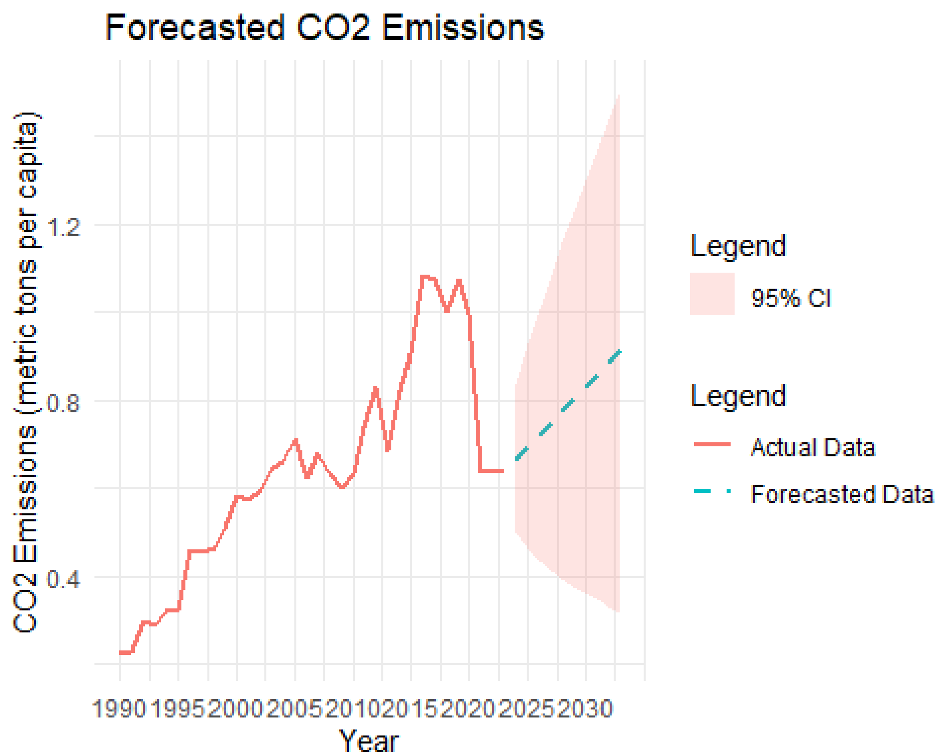


Figure 21: Forecasted CO_2 Emissions for 2024–2033 with Historical Trends

The forecasted emissions show a modest increasing trend from 2024 onwards, reflecting the model's expectations based on historical patterns and predictor variables.

The model successfully captures historical trends and provides realistic forecasts for future emissions.

The graph illustrates the historical actual values of CO_2 emissions (1990–2023) and the forecasted values for 2024–2033. The forecast aligns with historical trends and includes a 95% confidence interval to account for prediction uncertainty. While the model provides accurate short-term forecasts, the wide confidence interval highlights the inherent uncertainty in long-term predictions.

4.1 Renewable Energy Trends

The analysis of renewable energy consumption as a percentage of total final energy use (1990–2023) reveals a significant declining trend until 2015, followed by a gradual recovery in subsequent years. This shift could be attributed to changes in energy policies, advancements in renewable energy technologies, and increasing awareness of environmental sustainability. The rise in renewable energy usage post-2015 suggests the success of global and local initiatives aimed at reducing reliance on fossil fuels. The cross-correlation results also highlight a positive relationship between renewable energy adoption and CO_2 emissions reduction, indicating the critical role of renewables in mitigating climate change impacts.

4.2 Impact on CO_2 Emissions

The CO_2 emissions (metric tons per capita) exhibit a distinct upward trend until 2015, after which a noticeable decline occurs. This decline aligns with increased adoption of renewable energy and efforts to improve energy efficiency. The regression model analysis reveals a significant negative relationship between renewable energy consumption and CO_2 emissions, supporting the hypothesis that renewable energy mitigates emissions. Additionally, the forecast for 2024–2033 predicts a modest upward trend in emissions, with confidence intervals highlighting potential variability. These results underscore the need for sustained policy efforts to maintain the declining trajectory of CO_2 emissions observed in recent years.

4.3 Effects on Forestry and Agriculture

The analysis of forest area as a percentage of total land area demonstrates a consistent decline from 1990 to 2010, followed by stabilization and slight recovery after 2015. The correlation analysis reveals that forest area has a significant influence on CO_2 emissions, with deforestation contributing to higher emissions. Similarly, the agriculture sector's share of GDP has shown a long-term decline, reflecting economic transitions and urbanization. The regression results indicate that both forestry and agriculture are critical factors affecting CO_2 emissions, with sustainable land use practices and conservation efforts playing a vital role in reducing emissions and enhancing environmental resilience.

5. Discussion

The analysis provides valuable insights into the trends, relationships, and forecasts of CO_2 emissions (metric tons per capita) over the study period, focusing on the relationship with key predictors and the model's performance.

Historical Trends

The historical data (1990–2023) indicates a significant rise in CO_2 emissions until 2015, followed by a sharp decline. This trend reflects underlying changes in economic, environmental, and policy factors, such as industrialization, energy consumption patterns, and efforts to adopt sustainable practices. The decline in emissions after 2015 could be linked to increased reliance on renewable energy and global climate initiatives.

Model Evaluation

The regression model demonstrated a strong ability to explain variations in CO_2 emissions, with a high adjusted R-squared value of 0.9655. However, performance metrics revealed some limitations:

6. **RMSE (36.84)** and **MAE (28.09)** indicate moderate prediction errors, while a **MAPE of 74.66%** highlights substantial deviations in certain observations.
7. The residual analysis confirms that the model satisfies key assumptions of linear regression, including randomness, normality, and constant variance. The lack of significant autocorrelation in residuals further supports the model's validity.
8. The Breusch-Godfrey and Ljung-Box tests detected minimal autocorrelation, suggesting room for refinement by incorporating lagged variables or more sophisticated time-series models.

Predictive Insights

The forecasts for 2024–2033 indicate a gradual increase in CO_2 emissions, with the model capturing the overall trend accurately. However, the widening confidence intervals in later years underscore the uncertainty associated with long-term predictions. These forecasts are useful for policymakers and stakeholders to evaluate the potential future trajectory of emissions and implement proactive measures to mitigate climate impacts.

Key Findings from Cross-Correlation and Predictor Analysis

1. Strong relationships were identified between CO_2 emissions and predictors such as agricultural GDP contribution, forest area, and renewable energy usage. These correlations provide insights into how economic and environmental variables influence emissions.
2. Residual plots and scatterplots confirmed that the predictors adequately explain the dependent variable, with no evident heteroscedasticity or bias in the model.

Future Recommendations

1. **Model Refinement:**
   - Address residual autocorrelation using lagged predictors or advanced time-series models (e.g., ARIMA or SARIMA).
   - Improve the model's prediction accuracy by incorporating additional relevant variables.
2. **Policy Implications:**
   - The findings highlight the importance of renewable energy adoption and forest conservation in controlling emissions.
   - Policymakers should prioritize strategies to sustain the declining trend observed post-2015 while addressing potential future risks.

6. Conclusion

The analysis successfully explains historical trends, validates the model's performance, and provides meaningful forecasts for CO_2 emissions. While the model captures the overall trends, its predictive capacity for long-term horizons can be improved by addressing uncertainty and residual dependencies. The insights gained are instrumental in shaping future climate strategies and environmental policies.

7. References

- The dynamic impact of renewable energy consumption on CO2 emissions: A revisited Environmental Kuznets Curve approach : Faik Bilgili , Emrah Koçak , Ümit B journal homepage: www.elsevier.com/locate/rser

- Renewable energy and CO2 emissions: New evidence with the panel threshold model : Chaoyi Chen , Mehmet Pinar , Thanasis Stengos journal homepage: www.elsevier.com/locate/renene

- Forest, agriculture, renewable energy, and CO2 emission : Rida Waheed, Dongfeng Chang , Suleman Sarwar, Wei Chen journal homepage: www.elsevier.com/locate/jclepro

- The energy and CO2 emissions impact of renewable energy development in China : Tianyu Qi, Xiliang Zhang and Valerie Karplus journal homepage: www.elsevier.com/locate/enpol

- The relationship between renewable energy production and CO2 emissions in 27 OECD countries: A panel cointegration and Granger non-causality approach: Gaetano Perone journal homepage: www.elsevier.com/locate/jclepro

- Sri Lanka sustainable energy authority Case study 17: sri lanka – renewable energy. https://www.energy.gov.lk/en/.

- Sri Lanka sustainable energy authority Job Survey and Skills Analysis for the Renewable Energy Sector. https://www.energy.gov.lk/en/.