Conference Paper No: PF-11

Time series modeling and forecasting of total primary energy consumption in Sri Lanka

P.A.D.S.P. Caldera*, N.N.D. Malshika, S.H.A.S. Nikapitiya, U.S.C.B. Udugedara and N.V. Chandrasekara

[1]Department of Statistics and Computer Science, University of Kelaniya, Sri Lanka
sandunicaldera97@gmail.com*

Abstract

Primary energy is the energy that is harvested directly from natural resources. Forecasting total primary energy consumption in Sri Lanka is significant as primary energy consumption worldwide is expected to continue increasing. This study aimed to model and forecast total primary energy consumption in Sri Lanka, which has not yet been analysed using Time Series Analysis. For this purpose, the annual data of total primary energy consumption in Sri Lanka from 1960 to 2019 in terawatt-hours was extracted from the world wide web and analysed with Auto-Regressive Integrated Moving-Average (ARIMA) model. The stationary of the series was tested using the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test, Phillips-Perron (PP) test, and Augmented Dickey-Fuller (ADF) test. The study revealed the ARIMA(4,2,1) model as a best-fitting model, which gave the minimum value of Akaike Information Criterion (AIC). Total primary energy consumption from 2008 to 2019 was forecasted using ARIMA(4,2,1) model as it satisfied the model diagnostics, which are ARCH test, autocorrelation function, and normality of residuals. With Mean Absolute Error (MAE) of 5.0283 and Root Mean Squared Error (RMSE) of 5.9216, the results illustrate that ARIMA(4,2,1) model captures the trend in total primary energy consumption accurately. Based on the results, the study suggests ARIMA(4,2,1) is more convenient in determining the trends and the patterns of the future in total primary energy consumption in Sri Lanka.

Keywords

ARIMA model, Energy consumption, Forecasting, Sri Lanka, Time series modeling

Introduction

Primary energy consumption is that the direct use or supply at the source of energy that has not been converted or transformed. Energy consumption in the world has been increasing over the years, that has been mainly due to the ever-increasing world population and the need to meet the energy needs of industries that are ever-expanding their capacity. Sri Lanka's primary energy supply mostly comes from oil and coal. The country imports crude oil and refined products with Liquefied Petroleum Gas (LPG) and coal. These global resources in Sri Lanka are used mainly for lighting, cooking in households, rail, road, air, and sea in transports, boilers in power generation. Forecasting is the process of making predictions of the future by gathering and analysing the past and the current data. Time series modeling is used widely in the process of forecasting (Fernando et al., 2017; Ozturk & Ozturk, 2018; Tian et al., 2019). Analysing and forecasting how the total primary energy consumption in Sri Lanka changes over time is very important as the government holds a strong interest in lowering energy usage. Total primary energy consumption in Sri Lanka has not yet been analysed and forecasted using time series modeling. Observing the past primary energy consumption patterns in Sri Lanka, the main goal of this study is to analyse and forecast it using Time Series Analysis.

This is further elaborated by identifying and analysing trends and patterns of the series, determining the relevant candidate models, and selecting the best fit time series model to forecast total primary energy consumption in Sri Lanka.

In 2012, two researchers Yasmeen and Sharif analyzed the monthly electricity consumption in Pakistan from January 1990 to December 2011 and obtained the forecast. The study concludes that ARIMA(3,1,2) model is the most appropriate model for forecasting the electricity consumption of Pakistan. Their results revealed that, electricity consumption is continuously increasing over time, and they have suggested that the government of Pakistan must take effective steps to increase the electricity production through different energy sources to restore the economic status of the country by meeting the demand for electricity in the country.

A study on forecasting hydroelectricity consumption in Pakistan based on the historical data of the past 53 years using Auto-Regressive Integrated Moving-Average (ARIMA) modeling was done by Jamil in 2020. Up to the year 2030, the hydroelectricity consumption was predicted, based on the developed forecasting equation. To validate the reliability of the forecasted data, the results were compared with the actual values and the results showed a good fit with minimum error.

Peiris and Kumarasinghe have presented a forecasting study of total annual tea production in Sri Lanka and major tea growing areas. They have fitted time series models by the Box and Jenkins ARIMA model approach, tested for stationary by the Augmented Dickey-Fuller test, and applied differencing techniques to make the series stationary. Model diagnostics were performed using the Ljung Box test and autocorrelation function of residuals. The ARIMA(2,2,1) model was identified as the most appropriate model for the total annual tea production. Their study concluded that, the total tea production increases in Sri Lanka by 2020, compared to the average production from 2011 to 2015.

This article is organized as follows: In the introduction, the background, objective, and significance of this study are specified, and a brief literature review is included; the methods used in this study are introduced in the methodology section; in Results and Discussion, total primary energy consumption in Sri Lanka is analyzed, predicted with ARIMA model and the results are compared; finally, the conclusions of this study are made in the Conclusion. References concludes the article.

Methodology

Initially, the dataset was divided into two subsets, as one contains the latest 20% data in the dataset and the other contains the remaining 80%. Next, an exploratory data analysis was performed for the sub dataset contains 80% data of total primary energy consumption in Sri Lanka, and forecasting was done on the latest 20%, using the selected model.

A. Step by step Methodology

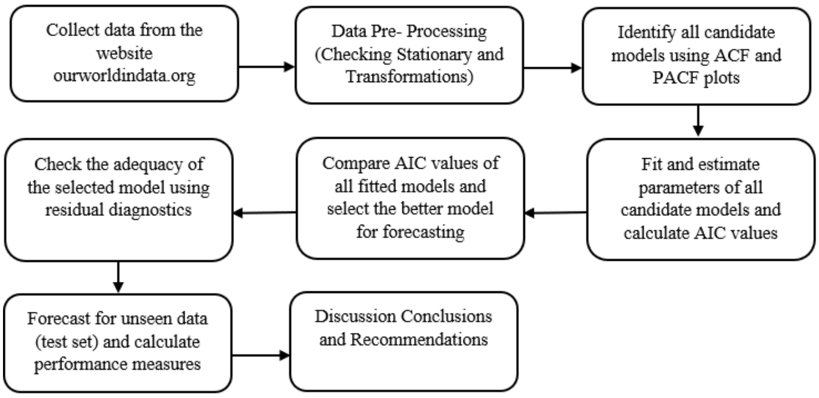

Figure 1. *Step by step procedure followed in the study.*

In this study, time-series analysis was used to model the time-dependent structure of total primary energy consumption in Sri Lanka. Figure 1 illustrates the step-by-step methodology employed in the study.

B. Theories

Time series: A time series is a set of observations x_t, each one being recorded at a specific time t.

The three statistical tests used to assess the stationarity of the time series, namely, Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test, Phillips-Perron (PP) test and Augmented Dickey-Fuller (ADF) test, have the following null hypothesis and alternative hypothesis.

ADF test - H_0: The series is not stationary. H_1: The series is stationary.

PP test - H_0: The series is not stationary. H_1: The series is stationary.

KPSS test - H_0: The series is stationary. H_1: The series is not stationary.

In both ADF and PP tests, if the p value is less than 0.05, then H_0 will be rejected. In KPSS test, if the Test Statistic < Critical value at 5% level, then H_0 will be rejected.

Moving Average process of order q, MA(q) : $y_t = \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} - \cdots - \theta_q\varepsilon_{t-q}$ (1)

Auto Regressive process of order p, AR(p) : $y_t = \delta + \emptyset_1 y_{t-1} + \cdots + \emptyset_p y_{t-p} + \varepsilon_t$ (2)

where ε_t is white noise.

A mixed Autoregressive/ Moving Average process containing p AR terms and q MA terms is said to be an ARMA process of order (p, q), which is given by,

$$y_t = \delta + \emptyset_1 y_{t-1} + \cdots + \emptyset_p y_{t-p} + \varepsilon_t - \theta_1\varepsilon_{t-1} - \cdots - \theta_q\varepsilon_{t-q} \quad (3)$$

ARIMA Model is one of the most successful time series forecasting techniques. According to Box-Jenkins (1976), a non-seasonal ARIMA model is denoted by ARIMA (p, d, q). This model is a combination of Auto-Regressive (AR) and Moving Average (MA), where the number of AR terms, the number of non-seasonal differences needed for stationarity, and the number of MA terms are denoted by p, d, and q, respectively. The ARIMA is the generalization form of the ARMA approach.

The residual diagnostics: i.e., the heteroscedasticity, autocorrelation, and the normality of residuals, are tested using the ARCH test, Ljung–Box Q test, and the Jarque-Bera test, respectively. Those residual diagnostics have the following null and alternative hypotheses regarding the residuals.

Heteroscedasticity	- H_0: There is no heteroscedasticity.	H_1:There is heteroscedasticity.
Autocorrelation	- H_0: There is no autocorrelation.	H_1: There is autocorrelation.
Normality	- H_0: Normally distributed.	H_1: Not normally distributed.

For all three tests, if the p value is less than 0.05 (at 5% level of significance) then H_0 will be rejected.

Results and Discussion

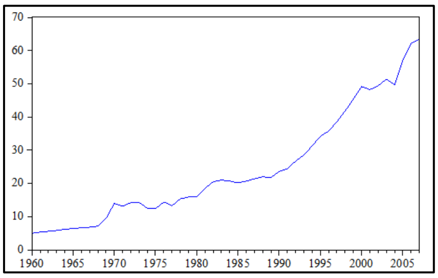

Figure 2. *Time Series plot of total primary energy consumption in Sri Lanka.*

This section illustrates the results of the analysis and the discussion. The time series plot which was generated with observations from 1960 to 2007 illustrates in Figure 2 and it exhibits an upward trend. It seems that the mean function of the series, $E[X_t]$ increases with time. Therefore, it can be concluded that the series is not stationary. Then stationary conditions of primary energy consumption were measured using three Unit Root tests. The tests with their estimated results are presented in Table 1.

Table 1. *The results of Stationary tests.*

Total primary energy consumption	**Statistical test**		**Stationarity at 5% level of significance**
Level data	ADF PP KPSS	p-value: 1.0000 p-value: 1.0000 Test statistic: 0.84803 Critical value at 5%: 0.463	Not stationary
First differenced data	ADF PP KPSS	p-value: 0.0000 p-value: 0.0000 Test statistic: 0.60337 Critical value at 5%: 0.463	Not stationary
Second differenced data	ADF PP KPSS	p-value: 0.0001 p-value: 0.0000 Test statistic: 0.03728 Critical value at 5%: 0.463	Stationary

All three tests imply that the original series of total primary energy consumption in Sri Lanka is not stationary at a 0.05 level of significance. Thus, to make the series stationary by reducing the trend component, the first differenced data were considered. According to the results shown in Table 1, the first differenced data of primary energy consumption were not stationary. Therefore, the second differenced data were considered.

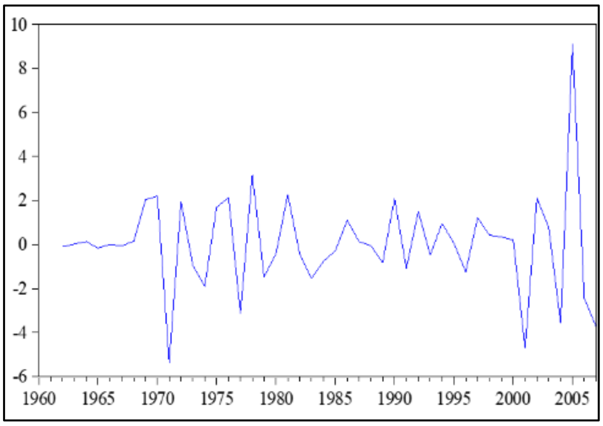

***Figure 3.** Time Series plot of second differenced data of the series.*

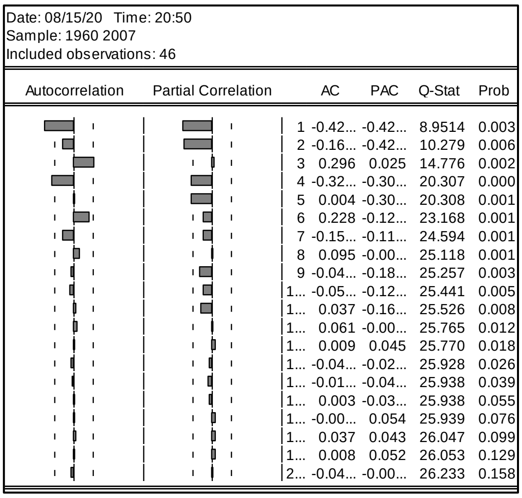

***Figure 4.** Correlogram of the stationary series.*

As exhibits in Figure 3, the series of the second differenced data fluctuate around a horizontal line. It indicates, the series seems to be stationary. By the results in Table 1, all three tests imply that the second differenced data of Primary energy consumption were stationary under the 0.05 level of significance.

Figure 4 shows the ACF and PACF plots of the stationary series. ACF cuts off at lags 1, 3, 4 and PACF cuts off at lags 1, 2, 4, 5. According to the cut-off lags, MA, AR, and ARIMA candidate models were suggested and fitted. Table 2 contains the candidate models with their respective AIC values.

***Table 2.** Candidate models with AIC values.*

Model	AIC	Model	AIC
AR(1)	4.4147	ARIMA(2,2,1)	4.181
AR(2)	4.1751	ARIMA(2,2,3)	4.0492
AR(4)	4.1529	ARIMA(2,2,4)	4.0624
AR(5)	4.0689	**ARIMA(4,2,1)**	**3.9975**
MA(1)	4.0277	ARIMA(4,2,3)	4.0705
MA(3)	4.0914	ARIMA(4,2,4)	4.1144
MA(4)	4.114	ARIMA(5,2,1)	4.1414
ARIMA(1,2,1)	4.0705	ARIMA(5,2,3)	4.181
ARIMA(1,2,3)	4.1144	ARIMA(5,2,4)	4.0492
ARIMA(1,2,4)	4.1414		

Dependent Variable: D(PRIMARY_ENERGY_CONSUMPTI,2)
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 08/16/20 Time: 10:05
Sample: 1962 2007
Included observations: 46
Failure to improve objective (non-zero gradients) after 427 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.055729	0.014413	3.866622	0.0004
AR(1)	0.085938	0.127494	0.674055	0.5043
AR(2)	-0.067162	0.196886	-0.341121	0.7348
AR(3)	0.234006	0.177363	1.319363	0.1947
AR(4)	-0.470917	0.187446	-2.512279	0.0162
MA(1)	-0.999989	3077.206	-0.000325	0.9997
SIGMASQ	2.091331	468.8527	0.004461	0.9965

R-squared	0.601353	Mean dependent var	0.019566
Adjusted R-squared	0.540022	S.D. dependent var	2.315739
S.E. of regression	1.570572	Akaike info criterion	3.997485
Sum squared resid	96.20121	Schwarz criterion	4.275756
Log likelihood	-84.94215	Hannan-Quinn criter.	4.101727
F-statistic	9.805142	Durbin-Watson stat	2.029529
Prob(F-statistic)	0.000001		

Inverted AR Roots	.59-.51i	.59+.51i	-.55+.68i	-.55-.68i
Inverted MA Roots	1.00			

Figure 5. *ARIMA(4,2,1) model.*

According to Table 2, ARIMA(4,2,1) model was chosen as a better model to forecast the series of total primary energy consumption of Sri Lanka, which exhibits a minimum AIC of 3.9975 among all candidate models.

In the fitted ARIMA(4,2,1) model, the coefficients of AR(1), AR(2), AR(3), MA(1) were insignificant. Thus, corresponding model equation is represented in the Equation 4:

$$y_t = 0.0557 - 0.4709y_{t-4} + \varepsilon_t \tag{4}$$

Figure 5 illustrates the coefficients, probabilities, and standard errors of parameters of the ARIMA(4,2,1) model. As the next phase of the study, the adequacy of the ARIMA(4,2,1) model was checked as follows.

Table 3. *Results of residual diagnostic tests.*

Tests	**ARCH test: (Heteroskedasticity)**	**Ljung–Box Q test: (Autocorrelation)**	**Jarque-Bera test: (Normality)**
p-value	0.2150	All the p-values > 0.05	0.2582

As shown in Table 3, since the p-values for each test were greater than 0.05, study concluded that homoscedasticity in the residuals and the residuals are independently, normally distributed. Hence, all three assumptions were satisfied by the model ARIMA(4,2,1) and the model was used for forecasting total primary energy consumption in Sri Lanka from 2008 to 2019.

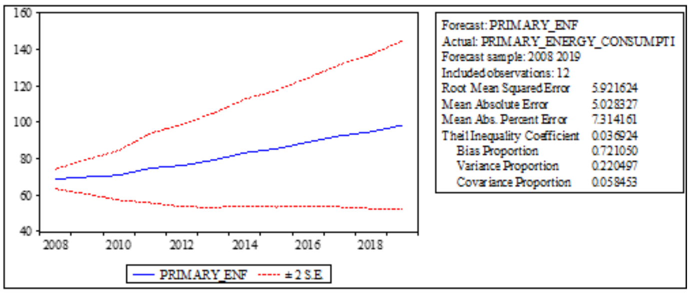

Figure 6. *Forecasted data from 2008 to 2019 using ARIMA(4,2,1) model.*

Figure 6 exhibits that Mean Absolute Error (MAE) is 5.0283 and Root Mean Squared Error (RMSE) is 5.9216 of the forecasted series using ARIMA(4,2,1) model. Since RMSE and MAE were considerably low, ARIMA(4,2,1) model could be used for forecasting total primary energy consumption in Sri Lanka.

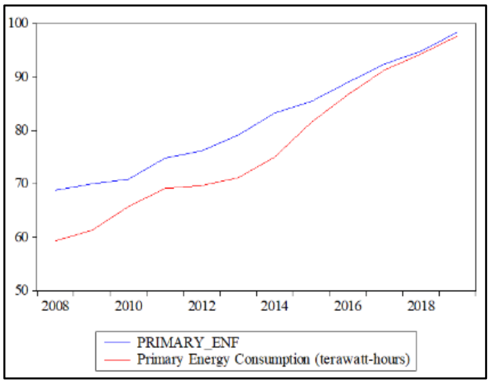

Figure 7. *Actual values vs forecasted values from 2008-2019.*

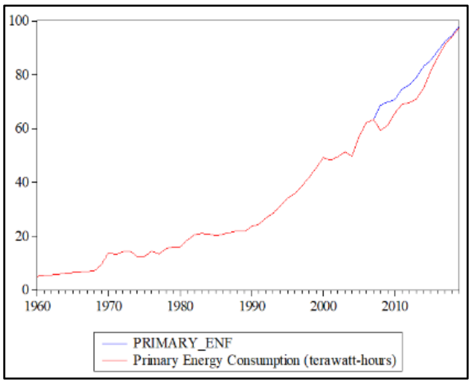

Figure 8. *Total primary energy consumption in Sri Lanka from 1960 to 2019 with forecasted series from 2008 to 2019.*

As displayed in Figure 7, the differences between values of actual and forecasted data from 2008 to 2019 were considerably low. Figure 8 demonstrates the trend of the fitted values is generally consistent with that of the actual values. These findings suggested that ARIMA(4,2,1) model can capture the future movements of total primary energy consumption in Sri Lanka. Hence, the forecast using the model ARIMA(4,2,1) can help the decision-makers to know the volume and trend of the future primary energy consumption to better schedule and plan the operations of the supply system. Since this study is a univariate time series forecasting, it can further be improved by applying multivariate time series models.

Conclusion

This study takes up the modeling and forecasting of total primary energy consumption in Sri Lanka using Time Series Analysis. After applying the tests of stationary and the data was stationary at the second difference. Subsequently, the suggested models are identified using cut-off lags of ACF and PACF plots and fitted. Using minimum AIC criteria ARIMA(4,2,1) model was selected as a better model among all the candidate models. Since all the performed residual diagnostics were satisfied by the ARIMA(4,2,1) model, total primary energy consumption in Sri Lanka from 2008-2019 is forecasted using it. The forecasted series captures the increasing trend and patterns of the actual series more accurately with considerably low forecasted errors: an MAE of 5.0283 and an RMSE of 5.9216. Thus, the study concludes that ARIMA(4,2,1) model is the most appropriate model for forecasting total primary energy consumption in Sri Lanka. As the forecasted values show an increasing trend of total primary energy consumption in Sri Lanka, it suggests that the government must go for alternative energy sources that will not run out instead of using non-renewable primary energy resources to meet the demand for primary energy in the country.

References

Fernando, J. L. L. S., Gunawardana, J. R. N. A., Perera, K. A. I. T., Perera, M. L. D. M., Shashikala, M. A. G., Jayasundara, D. D. M., & Rathnayaka, R. M. (2017). Time Series Modelling approach for forecasting Electricity Demand in Sri Lanka.

Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015). Introduction to time series analysis and forecasting. John Wiley & Sons.

Ozturk, S., & Ozturk, F. (2018). Forecasting energy consumption of Turkey by Arima model. Journal of Asian Scientific Research, 8(2), 52.

Rahman, A., & Ahmar, A. S. (2017, September). Forecasting of primary energy consumption data in the United States: A comparison between ARIMA and Holter-Winters models. In AIP Conference Proceedings (Vol. 1885, No. 1, p. 020163). AIP Publishing LLC.

Ritchie, H., & Roser, M. (2020). Energy. Published online at OurWorldInData. org. https://ourworldindata.org/energy.

Tian, C. W., Wang, H., & Luo, X. M. (2019). Time-series modelling and forecasting of hand, foot and mouth disease cases in China from 2008 to 2018. Epidemiology & Infection, 147.