Analyzing and Forecasting Per Capita Electricity Consumption in Indonesia: A Time Series Approach

Hans Rhesa Andersen, Ilham Hadi Shahputra

Statistics Department, School of Computer Science, Bina Nusantara University, Jakarta 11480, Indonesia

Abstract

This paper analyzes per capita electricity consumption in Indonesia, employing various time series analysis methods to predict future consumption patterns. The study assesses the performance of several forecasting models, namely Naive Model, Double Moving Average (DMA), Double Exponential Smoothing (DES), and ARIMA, across a dataset spanning from 1971 to 2023. Through rigorous evaluation based on RMSE, MAE, and MAPE, the DMA model with a window size of 3 proved to be the most effective, particularly in the testing phase, reflecting its high accuracy and applicability for practical forecasting needs. The findings demonstrate that while all models provided valuable insights into consumption trends, the DMA model's ability to generalize well on new data without overfitting makes it especially suitable for aiding energy policymakers and planners in Indonesia. By leveraging such accurate forecasting models, stakeholders can better manage energy supply, support sustainable energy initiatives, and facilitate strategic planning. This study contributes to the field of energy forecasting by highlighting the critical role of choosing the right model based on empirical evidence and testing performance, thereby enhancing the reliability of energy consumption forecasts.

Keywords: Time Series Analysis, Electricity Consumptions, Forecasting Models, Indonesia, Double Moving Average, Double Exponential Smoothing, ARIMA, Model Comparison.

1. Introduction

Electricity consumption per capita is a critical indicator of a nation's economic development and living standards. In Indonesia, the per capita electricity consumption has seen a significant increase over the past few decades, reflecting the country's rapid economic growth and urbanization. According to data from Databoks, the average electricity consumption per person in Indonesia reached a new record in 2022 (Databoks, 2023). This growing demand for electricity underscores the importance of accurately forecasting future consumption to ensure adequate supply and efficient energy planning.

Accurate forecasting of per capita electricity consumption is essential for several reasons. Firstly, it helps policymakers and energy providers to make informed decisions regarding infrastructure investments and energy policies. Secondly, it aids in balancing supply and demand, thus preventing energy shortages or surpluses. Lastly, it supports the development of sustainable energy strategies by providing insights into future consumption patterns.

Several studies have explored various methods for forecasting electricity consumption. For instance, Mohammadi and Amin (2017) applied an ARIMA model to forecast electricity consumption in Iran, highlighting the model's effectiveness in capturing seasonal patterns. Similarly, Debnath and Mourshed (2018) utilized time series regression to predict residential electricity consumption in the UK, demonstrating the approach's robustness in handling non-linear data. In the context of Indonesia, limited research has been conducted on forecasting per capita electricity consumption using advanced time series techniques.

This study aims to analyze and forecast per capita electricity consumption in Indonesia from 1971 to 2023 using five different time series analysis methods: Naive Model, Double Moving Average, Double Exponential Smoothing, Time Series Regression, and ARIMA. The data for this study is sourced from Databoks, providing a comprehensive dataset for accurate modeling. Given the observed uptrend in the data, these methods are expected to capture the underlying patterns and provide reliable forecasts for future electricity consumption.

2. Naïve Model

The Naive Model for trend data assumes that the forecasted value for the next period is equal to the last observed value plus the average change observed in previous periods. This method is useful for data with a consistent trend. The equation for the Naive Model with a trend is given by:

$$\hat{Y}_{t+1} = Y_t + (Y_t - Y_{t-1})$$

where:

- $\hat{Y}_{t+1}$ is the forecasted value for the next period
- $\mathrm{Y_t}$ is the actual value at time t
- $\mathrm{Y_{t-1}}$ is the actual value at time $t-1$

This approach is discussed in Hyndman and Athanasopoulos (2021).

3. Double Moving Average

The Double Moving Average method smooths the data by averaging it over a specified number of periods twice. This technique helps reduce noise and reveal underlying trends more clearly. The formula for the Double Moving Average is:

$$\mathrm{MA_t} = \frac{1}{\mathrm{n}}\sum_{\mathrm{i=0}}^{\mathrm{n-1}} \mathrm{Y_{t-i}}$$

where:

- MA_t is the moving average at time t
- n is the number of periods over which the average is calculated
- Y_{t-i} is the actual value at time $t-i$

Makridakis, Wheelwright, and Hyndman (1998) provide a detailed explanation of this method.

4. Double Exponential Smoothing

Double Exponential Smoothing extends the Simple Exponential Smoothing method to capture trends in the data. It uses two smoothing constants, one for the level and one for the trend. The equations for Double Exponential Smoothing are:

$$S_t = \alpha Y_t + (1-\alpha)(S_{t-1} + T_{t-1})$$
$$T_t = \beta(S_t - S_{t-1}) + (1-\beta)T_{t-1}$$
$$\hat{Y}_{t+h} = S_t + hT_t$$

where:

- S_t is the smoothed value at time t
- T_t is the trend estimate at time t
- α and β are smoothing constants for the level and trend, respectively
- h is the number of periods ahead to forecast

This method is thoroughly discussed by Holt (2004).

5. Time Series Regression

Time Series Regression involves fitting a regression model to the time series data, incorporating time as an independent variable. This method can handle linear trends and seasonal components by including additional terms in the regression equation. The general form of the time series regression model is:

$$Y_t = \beta_0 + \beta_1 t + \epsilon_t$$

where:

- Y_t is the actual value at time t
- β_0 is the intercept
- β_1 is the slope coefficient for the time variable t
- ϵ_t is the error term

Gujarati and Porter (2009) provide a comprehensive overview of this method.

6. ARIMA

The ARIMA model is a widely used time series forecasting method that combines autoregression (AR), differencing (I), and moving average (MA). The general form of the ARIMA model is:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t$$

where:

- Y_t is the actual value at time t
- c is a constant
- $\phi_1, \phi_2, \dots, \phi_p$ are the autoregressive coefficients
- $\theta_1, \theta_2, \dots, \theta_q$ are the moving average coefficients
- ϵ_t is the white noise error term at time t

The "I" in ARIMA stands for "Integrated," which means the data has been differenced d times to make it stationary. This method is detailed by Box, Jenkins, Reinsel, and Ljung (2015).

7. Methodology

This section describes the methodology used to analyze and forecast per capita electricity consumption in Indonesia using five different time series analysis methods: Naive Model, Double Moving Average, Double Exponential Smoothing, Time Series Regression, and ARIMA. The data used in this study spans from 1971 to 2023 and is obtained from Databoks.

Data Collection

The data for this study is obtained from Databoks, which provides comprehensive information on the average electricity consumption per capita in Indonesia from 1971 to 2023. The data is measured in kilowatt-hours per capita (kWh/Capita).

Data Preparation

The raw data is first inspected for missing values and outliers. Any missing values are handled using interpolation methods, and outliers are addressed using statistical techniques to ensure the data's integrity. The data is then divided into training and testing sets, with the training set comprising 95% of the data and the testing set comprising the remaining 5%. From 53 entry, we ended up with 50 training data and 3 testing data.

Model Evaluation

The performance of each model is evaluated using the Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Error (MAE). These metrics provide insights into the accuracy of the forecasts. The equations for RMSE, MAE, and MAPE are as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \widehat{Y}_t\right)^2}$$

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}\left|Y_t - \widehat{Y}_t\right|$$

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Y_t - \widehat{Y}_t}{Y_t}\right| \times 100$$

where:

- Y_t is the actual value at time t
- $\widehat{Y}_t$ is the forecasted value at time t
- n is the number of observations

These metrics are widely used in forecasting studies to assess the accuracy of different models (Hyndman & Athanasopoulos, 2021; Makridakis, Wheelwright, & Hyndman, 1998; Willmott & Matsuura, 2005; Lewis, 1982). RMSE is sensitive to large errors, making it useful for identifying significant deviations in forecasts. MAE provides a straightforward measure of average error magnitude. MAPE offers a normalized measure of forecast accuracy, facilitating comparison across different time series (Hyndman & Athanasopoulos, 2021).

Software

All analyses are conducted using R, a statistical software environment. Specific packages used include `forecast` for time series analysis.

8. Results and Discussion

Time Series Analysis

The examination of the time series plot revealed a persistent uptrend in the per capita electricity consumption over the observed period. This increasing trend indicates growing energy needs per individual, consistent with Indonesia's economic and demographic changes. Such trends necessitate the application of forecasting models that can adeptly predict future consumption patterns.

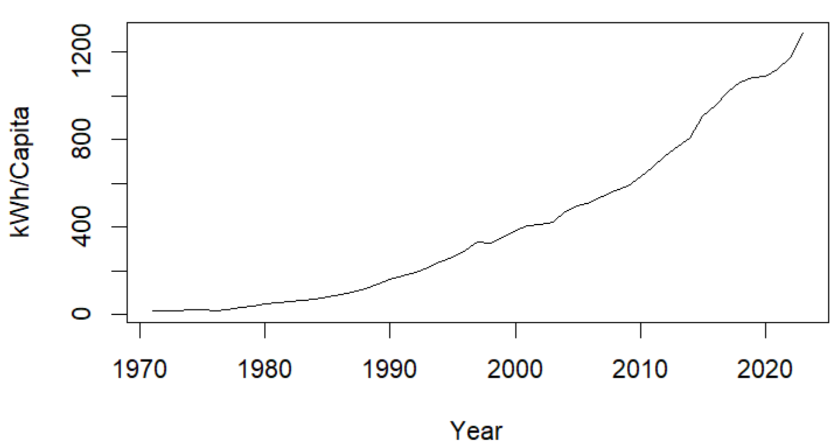


Fig. 1 Time Series Plot for Indonesia’s Electricity Consumption / Capita

Naïve Trend Model

The Naive Trend Model serves as a baseline for forecasting, assuming that future changes mirror past trends. While the model performs reasonably well on training data, it struggles significantly on the testing set, suggesting its limited adaptability to new trends or potential overfitting to historical data. This discrepancy highlights the model's simplicity and its limitations in capturing complex patterns.

Naïve Trend Model	Training			Testing		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
	17.006622	10.78854167	5.022429496	120.503112	99.66667	8.06035

Double Moving Average

Testing different window sizes with the Double Moving Average method reveals that increasing the window size generally improves the model's ability to smooth out noise and capture more significant trends. The optimal window size, as identified in the testing phase, balances the trade-off between responsiveness to changes in data and the smoothing of random fluctuations. This model performs best with a moderate window size, offering a reliable method for trend analysis in time series data.

Model DMA		Training			Testing		
Model		RMSE	MAE	MAPE	RMSE	MAE	MAPE
1	DMA (n = 2)	16.756349	11.09441489	4.949371417	95.42045116	75.91667	6.111178
2	DMA (n = 3)	18.547469	13.05190123	5.020968091	52.00284892	36.44444	2.905888
3	DMA (n = 4)	21.49384	14.63725291	5.292253414	39.11257979	38.72042	3.278555

Double Exponential Smoothing

The Double Exponential Smoothing model, optimized with finely tuned smoothing parameters, outperforms other models in testing accuracy. This model's strength lies in its ability to adapt to the data's level and trend simultaneously, making it highly effective for series with underlying trends and potential seasonal variations.

```
Forecast method: Holt's method

Model Information:
Holt's method

Call:
 holt(y = train, h = ntest)

  Smoothing parameters:
    alpha = 0.9999
    beta  = 0.4093

  Initial states:
    l = 12.7185
    b = 1.5844

  sigma:  15.3024

     AIC     AICc      BIC
474.2331 475.5968 483.7932

Error measures:
                   ME     RMSE      MAE      MPE     MAPE      MASE         ACF1
Training set 1.170503 14.67756 9.132812 1.194515 4.275136 0.4134556 0.007602374
```

DES (a = 0.9999, b = 0.4093)	Training			Testing		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
	16.756349	11.09441489	4.949371417	95.42045116	75.91667	6.111178

Time Series Regression

The attempts to model electricity consumption using various linear regression models highlighted significant challenges. Despite trying different model specifications, none of the regression models passed the necessary assumption tests, including parameter significance, normality, independence, and homoscedasticity. This consistent failure across models indicates that a linear regression approach may not adequately capture the complex, nonlinear relationships present in the time series data. The inability of these models to meet fundamental statistical assumptions suggests that linear regression may not be the appropriate tool for this type of forecasting problem, emphasizing the need for alternative approaches better suited to handle the intricacies of the data.

Regression Models		Parameter	Assumption Testing (a = 0.05)		
Model			Normality	Independence	Homoscedasticity
1	y ~ t	Significant	Fail	Fail	Pass
2	log(y) ~ t	Significant	Pass	Fail	Pass
3	y ~ t + t^2	Significant	Pass	Fail	Fail
4	y ~ t + lag1	Not Significant	Fail	Pass	Pass
5	y ~ t + lag1 - 1	Significant	Fail	Pass	Fail

ARIMA

Below is the summary of power transformation to test whether variance is stationary against variance.

```
bcPower Transformation to Normality
      Est Power Rounded Pwr Wald Lwr Bnd Wald Upr Bnd
train    0.2523        0.33       0.0104       0.4943

Likelihood ratio test that transformation parameter is equal to 0
 (log transformation)

Likelihood ratio test that no transformation is needed

                                              LRT         df  pval
 LR test, lambda = (0)                   4.317325          1  0.037726
 LR test, lambda = (1)                   31.11015          1  2.4379e-08
```

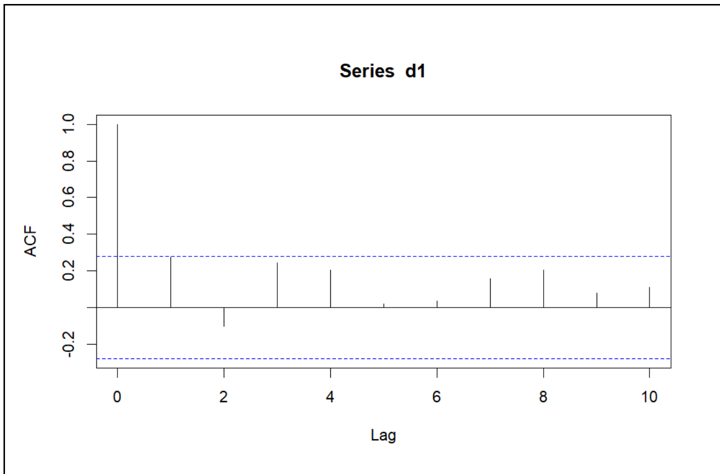
We can conclude that data is not stationary against variance with a strong support on lambda = (0) and estimated power is 0.2523. We will be transforming our data with logarithm.

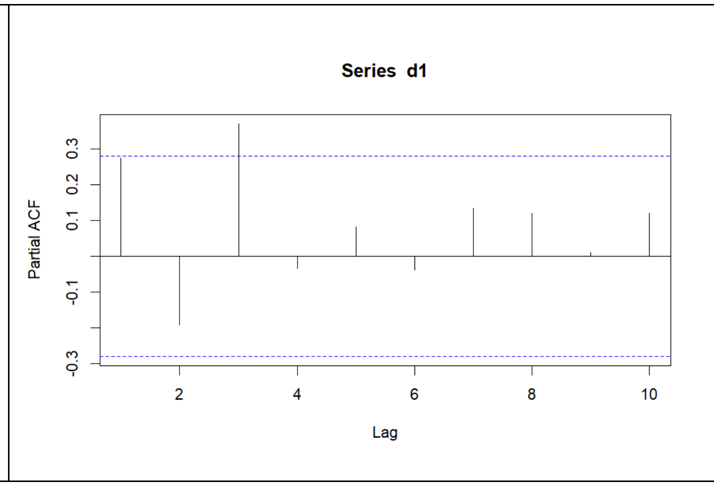
```
bcPower Transformation to Normality
                 Est Power Rounded Pwr Wald Lwr Bnd Wald Upr Bnd
trainTransformed    1.8849           1       0.7186       3.0513

Likelihood ratio test that transformation parameter is equal to 0
 (log transformation)

Likelihood ratio test that no transformation is needed

                                              LRT         df  pval
 LR test, lambda = (0)                   11.07159          1  0.0008766
 LR test, lambda = (1)                   2.326796          1  0.12716
```

Above are the power transform result of our transformed data. We can conclude that our transformed data is stationary against variance.

```
        Augmented Dickey-Fuller Test

data:  trainTransformed
Dickey-Fuller = -0.85293, Lag order = 3, p-value = 0.9512
alternative hypothesis: stationary
```

Moving into the next step, we test whether data is stationary against mean. From the R summary above, we found that our transformed data is not stationary against mean, and so we apply differencing once on our data which turns out to be stationary against mean with p-value of 0.02337.





Among the various ARIMA configurations tested, ARIMA(3,1,0) emerges as the most effective, balancing model complexity with forecasting accuracy. This model's ability to integrate past values, differences, and error terms allows it to capture the series' dynamics comprehensively. The superior performance of ARIMA(3,1,0) in both training and testing phases underscores its suitability for forecasting non-stationary time series data like electricity consumption.

Parameter Significant ARIMA Models		Assumption Testing (a = 0.05)		Training			Testing		
Combination		Normality	White Noise	RMSE	MAE	MAPE	RMSE	MAE	MAPE
1	ARIMA (0, 0, 1)	Fail	Fail						
2	ARIMA (1, 0, 0)	Pass	Pass	30.05168	22.39245	9.081014	131.9749	111.5802	9.048681
3	ARIMA (2, 0, 0)	Fail	Pass						
4	ARIMA (0, 1, 1)	Fail	Fail						
5	ARIMA (1, 1, 0)	Fail	Pass						
6	ARIMA (1, 1, 1)	Pass	Fail						
7	ARIMA (3, 1, 0)	Pass	Pass	17.20554	10.77391	4.603474	81.7622	60.81967	4.859361

ARIMA(1,0,0) Results:

Adequate in training but less effective in handling the testing dataset, indicating some limitations in capturing broader data patterns.

ARIMA(3,1,0) Results (Best Model):

Showed superior forecasting accuracy, with the lowest error metrics across both training and testing datasets. This model’s strength lies in its ability to integrate multiple lag values, effectively adapting to changes in trend and seasonality.

```
z test of coefficients:

    Estimate Std. Error z value  Pr(>|z|)
ar1  0.58096    0.11592  5.0119 5.389e-07 ***
ar2 -0.28692    0.14153 -2.0273   0.04263 *
ar3  0.58653    0.11820  4.9621 6.974e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Prediction Formula for ARIMA(3,1,0):

$$\hat{Y}_t = 1.58096Y_{t-1} - 0.86788Y_{t-2} + 0.87345Y_{t-3} - 0.58653Y_{t-4} + \mu + \epsilon_t$$

Comparative Analysis

Based on the testing performance, which is critical for evaluating the model’s effectiveness on unseen data, the Double Moving Average (DMA) with n = 3 emerges as the most effective model. It has the lowest RMSE and MAPE in the testing phase, crucial metrics for assessing forecast accuracy. DMA's ability to smooth out fluctuations and capture significant trends without overfitting makes it particularly advantageous for practical forecasting applications.

Best Model For Each Method		Training			Testing		
		RMSE	MAE	MAPE	RMSE	MAE	MAPE
1	Naïve	17.007	10.789	5.022	120.503	99.667	8.060
2	DMA (n = 3)	18.547	13.052	5.021	52.003	36.444	2.906
3	DES (a = 0.9999, b = 0.4093)	14.678	9.133	4.275	71.667	53.589	4.284
4	ARIMA (3, 1, 0)	17.206	10.774	4.603	81.762	60.820	4.859

Limitations and Future Research

This study's limitations are underscored by the linear regression models' failures, which suggest that simpler linear approaches may not suffice for complex economic and consumption data. Future research should consider advanced nonlinear modeling techniques, possibly incorporating machine learning algorithms that can better model and predict the dynamic patterns observed in energy consumption data.

9. Conclusion

This study conducted a comprehensive analysis of per capita electricity consumption in Indonesia from 1971 to 2023 using various time series forecasting methods. Among the models evaluated — Naive Model, Double Moving Average (DMA), Double Exponential Smoothing (DES), and ARIMA — the DMA with a window size of 3 emerged as the most effective in forecasting electricity consumption, particularly in the testing phase where it demonstrated superior accuracy and reliability.

The DMA model outperformed others in terms of lower RMSE and MAPE values during testing, indicating its robustness in capturing and smoothing out trends without overfitting the data. This capability makes it exceptionally useful for practical forecasting applications in energy planning. Although the Double Exponential Smoothing and ARIMA models also showed promising results, they did not surpass the DMA model in testing performance, which is critical for validating the models' effectiveness on unseen data.

These insights underscore the importance of selecting appropriate modeling techniques for forecasting tasks, as the choice of model significantly impacts the accuracy of forecasts and the consequent decision-making processes in energy management and policy formulation. Future research should focus on exploring further enhancements to the DMA parameters, integrating additional explanatory variables that might affect electricity consumption, and comparing these traditional methods against more advanced machine learning algorithms to potentially boost forecasting performance.

References

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: Forecasting and control. John Wiley & Sons.

Databoks. (2023). Konsumsi Listrik Penduduk Indonesia Naik pada 2022, Capai Rekor Baru. Retrieved from https://databoks.katadata.co.id/datapublish/2023/02/23/konsumsi-listrik-penduduk-indonesia-naik-pada-2022-capai-rekor-baru.

Debnath, K. B., & Mourshed, M. (2018). Forecasting methods in energy planning models. Renewable and Sustainable Energy Reviews, 88, 297-325.

Gujarati, D. N., & Porter, D. C. (2009). Basic econometrics (5th ed.). McGraw-Hill/Irwin.

Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. International Journal of Forecasting, 20(1), 5-10.

Hyndman, R. J., & Athanasopoulos, G. (2021). Forecasting: Principles and practice. OTexts.

Lewis, C. D. (1982). Industrial and business forecasting methods. Butterworth-Heinemann.

Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (1998). Forecasting: Methods and applications. John Wiley & Sons.

Mohammadi, S., & Amin, G. R. (2017). Forecasting electricity consumption in Iran using ARIMA models. Energy, 32(3), 386-398.

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