Forecasting the Export Value of Oil and Gas in Indonesia using Autoregressive Integrated Moving Average (ARIMA)

Ansari Saleh Ahmar\textsuperscript{a,*}, Abdul Rahman\textsuperscript{b}, Parkhimenko Vladimir Anatolievich\textsuperscript{c}, Rusli\textsuperscript{b}, & Sitti Masyitah Meliyana R.\textsuperscript{a}

\textsuperscript{a}Department of Statistics, Universitas Negeri Makassar, Makassar, 90223, Indonesia
\textsuperscript{b}Faculty of Engineering and Economics, Belarusian State University of Informatics and Radioelectronics (BSUIR), Belarus
\textsuperscript{c}Department of Statistics, Universitas Negeri Makassar, Makassar, 90223, Indonesia

Abstract

This study aims to utilize the ARIMA method to predict the value of Indonesia's oil and gas exports. As quantitative research, it employs secondary data sourced from the Central Bureau of Statistics of the Republic of Indonesia's website. The data spans January 2010 to March 2022 and are presented on a monthly basis. Through the results and discussion, three ARIMA models were established, namely ARIMA (1,1,0), ARIMA (0,1,1), and ARIMA (1,1,1). Among these models, the ARIMA (0,1,1) model with an AIC value of 2047.65 was found to be the most suitable for forecasting Indonesia's oil and gas exports. The forecasted values for the next five periods were 1254.124 (April 2022), 1309.678 (May 2022), 1289.236 (June 2022), 1296.758 (July 2022), and 1293.990 (August 2022).

Keywords: forecasting; ARIMA; oil and gas.

Received: 4 April 2023 Revised: 13 July 2023 Accepted: 21 September 2023

1. Introduction

A country’s economic growth can be gauged by its productivity success rate. Both developed and developing nations recognize exports as a means of achieving economic growth. Exports sell products from within a country to foreign markets (Farina & Husaini, 2017). Engaging in export activities has numerous advantages such as increasing state revenue in the form of foreign exchange, creating numerous domestic jobs, and stimulating economic growth by expanding the market (Sihombing et al., 2021).

Generally, Indonesia's export products are divided into two categories: oil and gas exports and non-oil and gas exports. According to Indonesian Law No. 22 of 2021, oil and gas are crucial commodities that significantly contribute to the nation's economy, and their management must be optimized for the well-being of people. Unfortunately, over the past decade, from 2012 to 2021, there has been a shortfall in Indonesia's oil and gas export trade balance, indicating that the country's expenditures have exceeded its income.

The deficit in oil and gas exports has a detrimental effect on national development, as there is a lack of funds available for investment. To address this issue, the government implemented a foreign debt policy to obtain funding. However, relying heavily on foreign debt can reduce foreign exchange reserves. A decrease in foreign exchange reserves can negatively affect economic growth, causing the value of the national currency to depreciate.

To address the persistent oil and gas export deficit in Indonesia, careful planning and improvement measures are essential. Both short-term and long-term planning can serve as a guide for actions that must be taken to prevent future oil and gas export shortages. The planning process can be initiated by gathering information on the projected value of future oil and gas exports (Hayati et al. 2021). To accurately estimate the future value of oil and gas exports, a reliable

\* Corresponding author.
E-mail address: ansarisaleh@unm.ac.id
A forecasting method is required, which can be utilized by the National Export Development Agency (BPEN) and the Ministry of Trade of Indonesia (Yulisa et al., 2023) to formulate effective strategies.

Predicting future events involves both art and science, and the results of these predictions can be utilized by policymakers to create strategic policies (Utami & Darsyah, 2016; Singh et al., 2021). The time-series method involves examining the connections between the variables measured over time (Ahmar et al., 2022). A commonly used time series analysis method is the Box-Jenkins model, also known as the Autoregressive Integrated Moving Average (ARIMA) model, which is a type of time-series forecasting model that is based on the behavior of recorded variable data.

2. Methods

The objective of this quantitative study was to use secondary data as the primary source of information. Specifically, data were obtained from the website of the Central Bureau of Statistics of the Republic of Indonesia, which provides monthly data on the value of oil and gas exports in Indonesia from January 2010 to March 2022. To achieve the research objectives, the ARIMA model was employed for forecasting.

3. Result and Discussions

To initiate the forecasting process, it is crucial to visualize the data and identify their patterns and characteristics. This can be achieved by plotting the data, as shown in Fig. 1.

![Time-series plot of data on the value of oil and gas exports in Indonesia.](image)

From Figure 1, it is evident that the data pattern has not yet achieved stationarity on average. This conclusion is further supported by the Augmented Dickey-Fuller test results presented below.

```r
> adf.test(data_migas.ts)
Augmented Dickey-Fuller Test

data:  data_migas.ts
Dickey-Fuller = -2.548, Lag order = 5, p-value = 0.3482
alternative hypothesis: stationary
```

The Augmented Dickey-Fuller test suggests that the p-value of 0.3482, which is greater than the significance level of $\alpha = 0.05$, indicates that the data are not stationary on average.

The next step involves conducting a Box-Cox test to determine whether the variance of the data is stationary. The results of the test are as follows.

```r
> BoxCox.lambda(data_migas.ts)
[1] 0.3679229
```

The Box-Cox test revealed that the value of $\lambda$ is 0.3679229, implying that the variance of the data is not yet stationary; therefore, transformation is required. The specific types of transformations employed are shown in Table 1.
Table 1. Type of Transformation Based on $\lambda$ Value

From Table 1, it can be seen that the type of transformation performed was $\sqrt{Z}$. In the next stage, the data were transformed as follows:

```r
> transzt <- sqrt(data_migas.ts)
> transzt
```

Following data transformation, the Box-Cox test was conducted once more to determine whether the variance of the data has become stationary. The results of the Box-Cox test are presented below.

```r
> BoxCox.lambda(transzt)
[1] 0.7096078
```

The Box-Cox test indicates that the value of $\lambda$ moves closer to 1, suggesting that the variance of the data has become stationary.

The next step involves stabilizing the data that have not been stationary, which is achieved through the process of differencing. The resulting difference data are presented below:

```r
> diff1 <- diff(transzt, differences = 1)
> diff1
```

1038
After performing the differencing process, the Augmented Dickey-Fuller test was repeated to assess whether the data had become stationary, on average. Otherwise, the differencing process is applied once more. The outcomes of the Augmented Dickey-Fuller test are as follows.

```r
adf.test(diff1)
```

Augmented Dickey-Fuller Test

```
data: diff1
Dickey-Fuller = -5.3466, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary
```

In the Augmented Dickey-Fuller test, the p-value is 0.01, which is smaller than the significance level of $\alpha = 0.05$, indicating that the data has been stationary on average. This finding is supported by the time-series plot for differencing presented in Figure 2, which reveals that the data has stabilized.

![Figure 2. Time series plot for diff1](image)

To identify the appropriate ARIMA model, we examined the ACF and PACF plots presented in Figures 3 and 4.

By examining figures 3 and 4, it is evident that the ACF plot exhibits a decline after the first lag, coinciding with the autocorrelation value at that lag, while the PACF plot diminishes and cuts off after the first lag. Consequently, the ARIMA models that have been established are ARIMA (1,1,0), ARIMA (0,1,1), and ARIMA (1,1,1).

The next step involves parameter estimation and conducting diagnostic checks on the model.

**a. Model ARIMA(1,1,0)**

```r
> fit1 <- arima(data_migas.ts, orde = c(1,1,0), method = "ML")
> fit1
```

Call:
```
arima(x = data_migas.ts, order = c(1, 1, 0), method = "ML")
```

Coefficients:
\begin{verbatim}
> coef(fit1)

Estimate  Std. Error    z value    Pr(>|z|) 
  ar1    -0.367964    0.077285    -4.7611    1.925e-06 *** 

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
\end{verbatim}

Figure 3. ACF plot of diff1

Figure 4. PACF plot of diff1
The ARIMA(1,1,0) analysis results reveal that the p-value or Pr(|z|) of the AR(1) parameter is less than 0.05, which indicates that the parameter is statistically significant, with a value of -0.367965. Additionally, the residuals of this model were examined using the Ljung–Box test to determine whether they fulfilled the white noise requirement.

```
> Box.test(fit1$residuals, type = "Ljung")
   Box-Ljung test

   data:  fit1$residuals
   X-squared = 0.27113, df = 1, p-value = 0.6026
```

According to the results of the Ljung–Box test, the p-value is 0.6026, which is larger than the significance level of $\alpha = 0.05$. Therefore, we were not reject $H_0$, indicating that the residuals were white noise.

The following test assesses whether the residuals exhibit a normal distribution by employing the Kolmogorov-Smirnov test.

```
> lillie.test(fit3$residuals)
   Lilliefors (Kolmogorov-Smirnov) normality test

   data:  fit3$residuals
   D = 0.054951, p-value = 0.3412
```

According to the Kolmogorov-Smirnov test, the p-value of 0.3412 is greater than the significance level $\alpha = 0.05$, indicating that the null hypothesis $H_0$ cannot be rejected. This suggests that the residuals followed a normal distribution.

b. Model ARIMA(0,1,1)

```
> fit2 <- arima(data_migas.ts, orde = c(0,1,1), method = "ML")
> fit2

Call:  arima(x = data_migas.ts, order = c(0, 1, 1), method = "ML")

coefficients:
   ma1  -0.4064
s.e.   0.0735

sigma^2 estimated as 70157:  log likelihood = -1021.83, aic = 2047.65
```

```
> coeftest(fit2)
    z test of coefficients:

     Estimate Std. Error z value  Pr(>|z|)    
ma1  -0.406430   0.073525  -5.5278 3.243e-08 ***

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The ARIMA(0,1,1) analysis results revealed that the p-value or Pr(|z|) for the MA(1) parameter was less than the significance level of 0.05, indicating that the parameter had a significant value of -0.406430. Additionally, the Ljung-Box test is employed to examine whether the residuals of this model adhere to the white noise criterion.

```
> Box.test(fit2$residuals, type = "Ljung")
   Box-Ljung test

   data:  fit2$residuals
   X-squared = 0.039669, df = 1, p-value = 0.8421
```

The Ljung-Box test results indicate that the p-value is 0.8421, which is greater than the significance level of $\alpha = 0.05$. 

1041
Therefore, the null hypothesis H0 cannot be rejected, suggesting that the residuals are white noise.

The next step is to determine whether the residuals from the model follow a normal distribution using the Kolmogorov-Smirnov test.

```r
> lillie.test(fit2$residuals)
Lilliefors (Kolmogorov-Smirnov) normality test
data: fit2$residuals
D = 0.067499, p-value = 0.09808
```

The Kolmogorov-Smirnov test yielded a p-value of 0.09808, which was larger than the significance level of α = 0.05. Consequently, we do not have sufficient evidence to reject the null hypothesis, which posits that the residuals are normally distributed.

c. Model ARIMA(1,1,1)

```r
> fit3 <- arima(data_migas.ts, orde = c(1,1,1), method = "ML")
> fit3
Call:
arima(x = data_migas.ts, order = c(1, 1, 1), method = "ML")
Coefficients:
ar1   ma1
-0.0695  -0.3487
s.e. 0.2045  0.1924
sigma^2 estimated as 70100:  log likelihood = -1021.77, aic = 2049.53
> coeftest(fit3)
z test of coefficients:

| Estimate | Std. Error | z value | Pr(>|z|) |
|----------|------------|---------|----------|
| ar1      | -0.069495  | -0.3398 | 0.73400  |
| ma1      | -0.348655  | -1.8122 | 0.06996  |
```

The ARIMA(1,1,1) analysis results showed that the p-value or Pr(>|z|) for the AR(1) and MA(1) parameters was greater than α = 0.05, indicating that the parameters were not statistically significant.

Table 2 presents a comprehensive overview of the outcomes of the parameter estimation and model diagnostic assessments for every conceivable model.

<table>
<thead>
<tr>
<th>Model</th>
<th>Assumptions Test</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Parameter Significance</td>
<td>White Noise (Residual)</td>
</tr>
<tr>
<td>ARIMA(1,1,0)</td>
<td>Significantly</td>
<td>Significantly</td>
</tr>
<tr>
<td>ARIMA(0,1,1)</td>
<td>Significantly</td>
<td>Significantly</td>
</tr>
<tr>
<td>ARIMA(1,1,1)</td>
<td>Not significantly</td>
<td>-</td>
</tr>
</tbody>
</table>

To evaluate the models presented in Table 2, it is evident that only two meet the ARIMA (1,1,0) and ARIMA (0,1,1) criteria. To determine the best model among these two, the AIC value was compared with the model with the smallest AIC being considered the best. As shown in Table 2, the most suitable and superior model is ARIMA (0,1,1).

The final step in the process is forecasting, which involves utilizing data from five subsequent periods to generate predictions. The forecast results are then presented in an easily understandable format.

```r
> ramal <- forecast(data_migas.ts, model = fit1, h=5)
> ramal
```

<table>
<thead>
<tr>
<th>Point Forecast</th>
<th>Lo 80</th>
<th>Hi 80</th>
<th>Lo 95</th>
<th>Hi 95</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apr 2022</td>
<td>1254.12</td>
<td>912.0887</td>
<td>1596.160</td>
<td>731.0259</td>
</tr>
<tr>
<td>May 2022</td>
<td>1309.68</td>
<td>905.0525</td>
<td>1714.303</td>
<td>690.8568</td>
</tr>
<tr>
<td>Jun 2022</td>
<td>1289.24</td>
<td>806.9266</td>
<td>1771.546</td>
<td>551.6074</td>
</tr>
<tr>
<td>Jul 2022</td>
<td>1296.76</td>
<td>753.5855</td>
<td>1837.931</td>
<td>469.1060</td>
</tr>
<tr>
<td>Aug 2022</td>
<td>1293.99</td>
<td>697.1399</td>
<td>1890.841</td>
<td>381.1864</td>
</tr>
</tbody>
</table>

Figure 5. Forecasting results for the next 5 periods.

4. Conclusions

Based on the analysis and discussion, it can be concluded that there are three ARIMA models that have been formed, namely ARIMA (1,1,0), ARIMA (0,1,1), and ARIMA (1,1,1). Among these three models, the ARIMA (0,1,1) model is the most suitable for forecasting the value of oil and gas exports in Indonesia, as indicated by its AIC value of 2047.65. The forecasted values for the next five periods were 1254.124 (April 2022), 1309.678 (May 2022), 1289.236 (June 2022), 1296.758 (July 2022), and 1293.990 (August 2022).

References


