Forecasting tuberculosis morbidity rate in Indonesia using autoregressive integrated moving average (ARIMA) method

W D Lesmono, S Mardiyati, D Lestari and A H A Zili

Department of Mathematics, Faculty of Mathematics and Natural Sciences (FMIPA), Universitas Indonesia, Depok 16424, Indonesia

Corresponding author’s email: sri_math@sci.ui.ac.id

Abstract. Tuberculosis is a disease that can affect socio-economic development. Based on data from the World Health Organization, there were 810,918 tuberculosis cases in Indonesia, which is noted as the third-highest number of tuberculosis cases in Asia in 2016. Prevention and control of tuberculosis are of considerable importance, especially in the insurance field, to cover the cost of treatment, so an accurate model of tuberculosis morbidity is needed. The method used in forecasting the tuberculosis morbidity rate is Autoregressive Integrated Moving Average (ARIMA) method. The ARIMA method is a time series method that is widely used to predict morbidity rates in the future. The data used in this study is the number of incidence morbidity tuberculosis rates that occurred in Indonesia from 2000 to 2017, which is obtained from the World Bank. The results showed that ARIMA (1, 2, 0) is the best and very accurate model to forecast the morbidity rate in Indonesia from 2018 to 2027, with the mean absolute percentage error (MAPE) is 0.1682 % and Akaike Information Criterion (AIC) values is -181.0120. The results of forecasting tuberculosis morbidity rate are expected to help insurance companies in determining the amount of premium paid by customers who suffer tuberculosis diseases.

Keywords: MAPE, premiums, stationarity, time series analysis

1. Introduction
Tuberculosis is an infectious respiratory disease caused by mycobacterium tuberculosis, which can cause sufferers to experience shortness of breath, persistent coughing, and sneezing [1]. Based on data from the World Health Organization in 2016, there were 810,918 cases of tuberculosis disease in Indonesia, which is the third highest number in Asia. Treatment for tuberculosis patients is expensive, so insurance is needed to cover the cost of treatment. The amount of medical expenses depends on the amount of premium paid by the customer. The main component in calculating premiums for tuberculosis customers is the level of tuberculosis morbidity [2]. In order to plan and set the amount of premium for this disease, the insurance company needs to determine the morbidity rate in the future. The morbidity rate in the future can be determined by forecasting.

The related research in forecasting the morbidity rate of a disease has been carried out by several researchers, including forecasting the morbidity rate of dengue using the support vector machine (SVM) method [3], forecasting the morbidity rate of lung cancer using the regression method [4], forecasting the morbidity rate of hepatitis using the grey model [5], and forecasting the morbidity rate of infectious gastrointestinal disease using the deep neural network (DNN) method [6]. In this study,
forecasting the morbidity rate of tuberculosis in Indonesia uses the autoregressive integrated moving average (ARIMA) method. ARIMA method is a time series forecasting method that uses the information of the value obtained in the prior period to predict the value in the next period [7]. This method is often used by many researchers to predict various morbidity cases of time series data, such as forecasting influenza [8], malaria [9] and pneumonia [10]. The ARIMA method also has been applied by several researchers to predict the tuberculosis morbidity rate in several other countries, such as China [11], Iran [12] and Benin [13].

Based on previous researches, the tuberculosis morbidity rate has a non-stationary time series pattern, which makes forecasting time series for the next period difficult. Therefore, the ARIMA method overcomes the problem of time series patterns because it has steps that can be carried out to be stationary time series and produces accurate forecasting for the next period [7]. Based on this background, this study aims to predict the tuberculosis morbidity rate for the next period in Indonesia.

2. Materials and method

2.1. ARIMA method

ARIMA method is a time series method that is modeled in the form of an equation which is generally expressed in backward shift notation as follows [7]:

\[(1 - B)^d X_t = \mu + \phi_1(1 - B)^d X_{t-1} + \cdots + \phi_p(1 - B)^d X_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \cdots - \theta_q \epsilon_{t-q}\]  

with \(X_t\) is time series at time \(t\), \(\epsilon_t\) is an error at time \(t\) that is normally distributed with mean 0 and variance \(\sigma^2\), \(B\) is a backshift operator that shows time series \(X_t\) back down by one time period or \(B X_t = X_{t-1}\), \(\mu\) is a constant, \(d\) is the integrated order that is the number of differencing process on time series, \(\phi_p\) is an autoregressive parameter in \(p\)-order, dan \(\theta_q\) is a moving average parameter in \(q\)-order. ARIMA model with autoregressive order of \(p\), integrated order of \(d\), and moving average order of \(q\) in equation 1 can be written as ARIMA\((p, d, q)\).

2.2. Differencing process determination

The first step in using the ARIMA method is transforming non-stationary time series data into stationary time series data by performing the differencing process. In this study, the Dickey-Fuller (DF) test is applied to determine the necessity of using the differencing process on the time series. DF test on time series \(X_t\) can be done by estimating the parameters in the following regression equation [14]:

\[(1 - B)^d X_t = \mu + \gamma X_{t-1} + \epsilon_t\]  

The DF test tests the significance of the parameters \(\gamma\) with the null hypothesis that \(\gamma = 0\) (\(X_t\) not stationary) versus alternative hypothesis \(\gamma < 0\) (\(X_t\) stationary) with the statistical test as follows [14]:

\[T = \frac{\hat{\gamma}}{\sigma_{\hat{\gamma}}}\]  

where \(\hat{\gamma}\) is the coefficient of parameter estimation for \(X_{t-1}\), \(\sigma_{\hat{\gamma}}\) is standard error value from parameter estimation result of \(\hat{\gamma}\). The values of the statistical test \(T\) follow \(t\) distribution with a decision rule that if \(T\) is less than the critical value, then \(\gamma < 0\), which means the time series is not stationary. Therefore, it must be done by using the differencing process aim to change the time series that is non-stationary to stationary. The significance value that used in this research is 0.05.

2.3. Autocorrelation and partial autocorrelation order determination

The second step in using the ARIMA method is graphs the autocorrelation function and partial autocorrelation function on the data that has been transformed to identify the order of the possible
values for $p$ and $q$. The graph of the autocorrelation function can determine possible $q$ values. Meanwhile, the partial autocorrelation function graph to determine the possible values of $p$. The determination of the value of each order is reviewed from the significant value in each lag period. At this stage, the temporary model chosen from the identification results can be more than two, and an estimate of the temporary parameters model is carried out to determine the fitting and residuals value of the period for the next stage [7].

2.4. Feasibility test for temporary ARIMA method
Test the feasibility of temporary models by diagnosing residuals with a classical assumption test. The classic assumption test on time series residuals in the ARIMA model consists of a normality test, an autocorrelation test, and a heteroscedasticity test [15]. The normality test is used to test the null hypothesis that the resulting residuals spread normally. The statistical test for the normality test in this study uses the Jarque-Bera test with the statistical test values defined as follows [16]:

$$\chi^2_{JB} = \frac{n}{6} \left( \bar{s}^2 + \frac{(K - 3)^2}{4} \right)$$  \hspace{1cm} (4)

where $n$ is the length of time series, $S$ dan $K$ respectively are skewness and kurtosis from residual. The hypothesis tested in the Jarque-Bera test is the null hypothesis that the residuals generated from the ARIMA model are normally distributed. Jarque-Bera statistical test values follow the chi-square distribution with degree of freedom $v = 2$. If the Jarque-Bera statistical test value is greater than the critical value at the significance level $\alpha$, the residuals of the model indicate normality problems. The autocorrelation test is used to test the null hypothesis that the resulting residuals do not correlate between each period (autocorrelation). The statistical test for the autocorrelation test in this study uses the Ljung-box test with the statistical test values defined as follows [15]:

$$Q = N(N + 2) \sum_{k=1}^{l} \frac{\hat{\rho}_k^2}{n - k}$$  \hspace{1cm} (5)

where $N$ is the length of the time series after the differencing process, $\hat{\rho}_k^2$ is an autocorrelation sample from a time series at lag $k$, and $l$ is the number of lag being tested. The hypothesis tested in the Ljung-box test is the null hypothesis that the residuals of the ARIMA model formed do not contain autocorrelation problems, with the alternative hypothesis that the ARIMA model formed contains autocorrelation problems. The Ljung-box statistical test follows the chi-square distribution with degree of freedom $v = l - p - q$ with $l$ being the number of lag that being tested, $p$ and $q$ are the order of the AR (Auto Regressive) and MA (Moving Average) model, respectively. If test statistic $Q$ is greater than the critical value of signification level $\alpha$, the residuals of the model indicate normality problems. A heteroscedasticity test is used to test the null hypothesis that the resulting residuals are uniform between one period and another (homoscedasticity). The statistical test for the heteroscedasticity test in this study uses the Lagrange-multiplier test with the statistical test values defined as follows [17]:

$$\chi^2_{LM} = nR^2_{LM}$$  \hspace{1cm} (6)

with $R^2_{LM}$ being the coefficient of the determination value, which is obtained from the estimated regression parameters as follows:

$$\varepsilon_t = \beta_0 + \sum_{i=1}^{l} \beta_i \varepsilon_{t-i} + \nu_t$$  \hspace{1cm} (7)

with $\varepsilon_t$ being the residual from the ARIMA model at $t$ period, $\beta_i$ is a residual parameter from time series at $i$, and $\nu_t$ is the error time series at $t$. The Lagrange-multiplier statistical test follows the chi-square distribution with degree of freedom $v = l$. If test statistic $\chi^2_{LM}$ is more than critical value at significant level $\alpha$, then the residuals of the model indicate heteroscedasticity.
2.5. Best ARIMA model selection

Selection of the best ARIMA models is done using various temporary models that are feasible to be used from the previous stage using the Akaike information criteria (AIC) and mean absolute percentage error. AIC and MAPE values are respectively calculated by the following formula [18]:

\[
AIC = N \ln \left( \frac{\sum_{t=1}^{N} \varepsilon_t^2}{N} \right) + 2m
\]  
(8)

\[
MAPE = \frac{1}{N} \left( \sum_{t=1}^{N} \left( \frac{X_t - \hat{X}_t}{X_t} \right) \right) \times 100\%
\]  
(9)

with \( \hat{X}_t \) is the fitted value at period \( t \), \( N \) is the length of the time series after the differencing process, and \( m \) is the number of parameters that have been estimated. The model chosen to be the best model is the one with the lowest AIC and MAPE values. Accuracy of time series forecasting is said to be very good if the MAPE value is below 10, good if the MAPE value is between 10 to 20, adequate if the MAPE value is between 20 to 50, and poor if the MAPE value is more than 50 [18].

2.6. Data

The data used in this study are tuberculosis morbidity rate data in Indonesia from 2000 to 2017 obtained from the World Bank at https://data.worldbank.org/indicator/SH.TBS.INCD?end=2017&locations=ID&start=2000 [19]. Tuberculosis morbidity rate data in Indonesia from the World Bank are shown in table 1.

3. Results and discussion

3.1. Time series graph of tuberculosis morbidity rate

The time series graph of the tuberculosis morbidity rate in Indonesia from table 1 is illustrated in figure 1a, which shows that the movement of the tuberculosis morbidity rate in Indonesia has decreased every year. Although the number of cases and total population in Indonesia has increased every year, the ratio between tuberculosis cases and the population in Indonesia has decreased. This is because the Indonesian government has been responsive to tuberculosis’ spread throughout the world and is concerned with public health in Indonesia through a variety of health programs for the prevention of disease.

3.2. Differencing test

Forecasting using the ARIMA method is done by forming the ARIMA model using the box-Jenkins method with the first step carried out is to determine the number of integrated orders. The number of integrated orders can be determined by checking the non-stationary time series at the tuberculosis morbidity rate data using the DF test.

| Year | Morbidity rate (%) | Year | Morbidity rate (%) | Year | Morbidity rate (%) | Year | Morbidity rate (%) |
|------|--------------------|------|--------------------|------|--------------------|------|--------------------|
| 2000 | 0.370              | 2005 | 0.360              | 2010 | 0.342              | 2015 | 0.325              |
| 2001 | 0.369              | 2006 | 0.357              | 2011 | 0.338              | 2016 | 0.322              |
| 2002 | 0.367              | 2007 | 0.353              | 2012 | 0.335              | 2017 | 0.319              |
| 2003 | 0.366              | 2008 | 0.349              | 2013 | 0.332              |      |                    |
| 2004 | 0.363              | 2009 | 0.345              | 2014 | 0.329              |      |                    |
DF test is done by estimating the regression parameters in equation 2 and testing the significance of the parameters using equation 3. DF test results on tuberculosis morbidity rate data without differencing are illustrated in figure 1a, and the first differencing described in figure 1b shows that the data is non-stationary. These results are indicated by P-Value from DF test higher than the significance value of 0.05, which means the time series is not stationary. Time series tuberculosis morbidity rate in the second differencing, which is illustrated in figure 1c, indicates that the data is stationary. This result is indicated by P-Value from DF test less than the significance value of 0.05, which means the time series is stationary. Thus, the model with the second differencing is used for the formation of the ARIMA model. DF test results are shown in table 2.

3.3. Determination of the order of AR and MA
The next stage is to determine the order of AR and MA, respectively, using the plot of the partial autocorrelation function (PACF) and the autocorrelation function (ACF) using stationary time series data.

![Figure 1](image)

**Figure 1.** Tuberculosis morbidity rate in Indonesia from the year 2000 to 201, (a) without differencing, (b) after the first differencing process and (c) after the second differencing process.

| Time Series                  | $\hat{\gamma}$ | $\sigma_\theta$ | $T$  | P-value | Stationarity     |
|------------------------------|-----------------|-----------------|------|---------|------------------|
| Without differencing         | 0.0268          | 0.0132          | 2.0337| 0.9996  | Non-stationary   |
| First differencing           | -0.5714         | 0.1914          | -2.9859| 0.0578  | Non-stationary   |
| Second differencing          | -1.5227         | 0.2251          | -6.7654| 0.0001  | Stationary       |

**Table 2.** DF test result on time series without differencing, first differencing, and second differencing.
Based on the PACF and ACF plots using time series data on the second differencing in figure 2, it is found that the ACF plot had a significant lag in the 0th lag and the 1st lag indicated by the intersection of the needle lag with a significance line at 5% significance level. Besides, the PACF plot has a significant lag in lag 1. Thus, the possibility of candidates for the time series forecasting model using the ARIMA method includes the ARIMA (1, 2, 0), ARIMA (1, 2, 1) and ARIMA (0, 2, 1) models.

The parameters formed from the ARIMA (1, 2, 0) model is $\phi_1$. The parameters formed from the ARIMA (1, 2, 1) model are $\phi_1$ and $\theta_1$, and the parameters formed from the ARIMA (0, 2, 1) model is $\theta_1$. These parameters are estimated so that they can be used to calculate the fitting value of the tuberculosis morbidity rate from 2000 to 2017. The estimation parameter result of $\phi_1$ and $\theta_1$ on the three ARIMA model is shown in table 3. The fitting and residual values of tuberculosis morbidity in Indonesia from 2000 to 2017 for the ARIMA (1, 2, 0), ARIMA (1, 2, 1) and ARIMA (0, 2, 1) models are shown in table 4.

3.4. Classic assumption test

In this stage, we performed a classical assumption test of the residuals, which is generated from table 4 to the ARIMA model candidates. The classic assumption test in this study was conducted to test the feasibility of the temporary ARIMA model from the results of the previous stages to forecast the tuberculosis morbidity rate in Indonesia in the next period. Based on the results of the normality test using the Jarque-Bera test method in equation 4, autocorrelation test with Ljung-box in equation 5, and heteroscedasticity test with the Lagrange-multiplier test method in equation 6 shown in table 5, it is found that the residuals of the three ARIMA model candidates that were formed successively did not show any problems of normality, autocorrelation, and heteroscedasticity. This result is indicated by P-Value higher than the significance value of 0.05, which means the null hypothesis is accepted for each classic assumption test for each ARIMA model. This result also shows that forecasting the tuberculosis morbidity rate in Indonesia using the ARIMA method produces residuals that spread normally, uniform, and do not correlate between each period so that the ARIMA method is feasible to be used to forecasting the tuberculosis morbidity rate in Indonesia.

$$
\begin{align*}
\phi_1 & = -0.5110 \\
\theta_1 & = -0.0502
\end{align*}
$$

![Figure 2. Time series PACF (a) and ACF (b) plot after second differencing.](image)

| Parameter | ARIMA (1, 2, 0) | ARIMA (1, 2, 1) | ARIMA (0, 2, 1) |
|-----------|----------------|----------------|----------------|
| $\phi_1$  | -0.5110        | -0.5489        | -              |
| $\theta_1$| -              | 0.0502         | -0.3834        |
Table 4. Fitting and residual value of tuberculosis morbidity rate in Indonesia from 2000 to 2017 for ARIMA (1, 2, 0), ARIMA (1, 2, 1) and ARIMA (0, 2, 1).

| Year | Actual value (%) | ARIMA (1, 2, 0) | ARIMA (1, 2, 1) | ARIMA (0, 2, 1) |
|------|------------------|-----------------|-----------------|-----------------|
|      | Fitted value (%) | Residual value  | Fitted value (%) | Residual value  | Fitted value (%) | Residual value  |
| 2000 | 0.370            | 0.36983         | 0.00017         | 0.36983         | 0.00017         | 0.36983         |
| 2001 | 0.369            | 0.36950         | -0.00050        | 0.36950         | -0.00050        | 0.36950         |
| 2002 | 0.367            | 0.36786         | 0.00086         | 0.36786         | 0.00086         | 0.36793         |
| 2003 | 0.366            | 0.36551         | 0.00049         | 0.36551         | 0.00049         | 0.36534         |
| 2004 | 0.363            | 0.36449         | 0.00149         | 0.36448         | 0.00148         | 0.36475         |
| 2005 | 0.360            | 0.36102         | -0.00102        | 0.36102         | -0.00102        | 0.36067         |
| 2006 | 0.357            | 0.35700         | 0.00000         | 0.35695         | 0.00005         | 0.35726         |
| 2007 | 0.353            | 0.35400         | 0.00100         | 0.35400         | 0.00100         | 0.35410         |
| 2008 | 0.349            | 0.34951         | 0.00051         | 0.34950         | 0.00050         | 0.32942         |
| 2009 | 0.345            | 0.34500         | 0.00000         | 0.34497         | 0.00003         | 0.34516         |
| 2010 | 0.342            | 0.34100         | 0.00100         | 0.34100         | 0.00100         | 0.34106         |
| 2011 | 0.338            | 0.33849         | 0.00049         | 0.33850         | 0.00050         | 0.33864         |
| 2012 | 0.335            | 0.33451         | 0.00049         | 0.33452         | 0.00048         | 0.33425         |
| 2013 | 0.332            | 0.33149         | 0.00051         | 0.33148         | 0.00052         | 0.33171         |
| 2014 | 0.329            | 0.32900         | 0.00000         | 0.32903         | 0.00003         | 0.32889         |
| 2015 | 0.325            | 0.32600         | 0.00100         | 0.32600         | 0.00100         | 0.32596         |
| 2016 | 0.322            | 0.32151         | 0.00049         | 0.32150         | 0.00050         | 0.32137         |
| 2017 | 0.319            | 0.31849         | 0.00051         | 0.31848         | 0.00052         | 0.31876         |

Table 5. Classic assumption test of residual time series result on three ARIMA model candidates

| ARIMA model | Normality test (jarque-bera test) | Autocorrelation test (ljung-box test) | Heteroscedasticity test (lagrange-multiplier test) |
|-------------|-----------------------------------|---------------------------------------|--------------------------------------------------|
|             | Statistics test $\chi^2_{JB}$ | P-value | Statistics test $Q$ | P-value | Statistics test $\chi^2_{LM}$ | P-value |
| ARIMA(1, 2, 0) | 0.9879 | 0.6102 | 3.2487 | 0.3548 | 0.2033 | 0.977 |
| ARIMA(1, 2, 1) | 1.0328 | 0.5967 | 6.5313 | 0.0884 | 0.1374 | 0.987 |
| ARIMA(0, 2, 1) | 0.5593 | 0.7560 | 2.3645 | 0.5003 | 1.0825 | 0.781 |

Therefore, the three ARIMA models were feasible to be used for forecasting the tuberculosis morbidity rate in Indonesia for the next period. Because the three ARIMA models are feasible to be used to forecast the tuberculosis morbidity rate in Indonesia for the next period, the selection of the best ARIMA model is needed in the next stage to get the best ARIMA model in forecasting tuberculosis morbidity in Indonesia for the next period.
3.5. Selection of the best ARIMA model
The last stage is to determine the best ARIMA model that can predict the movement of tuberculosis morbidity in Indonesia for the following year by comparing the AIC and MAPE values of the obtained ARIMA models in the previous stage using the results of fitting and residual values from table 4. The AIC calculation of the three candidates for the ARIMA model using equation 7 shows that ARIMA (1, 2, 0) is the ARIMA model with the lowest AIC value. Besides, the calculation of MAPE using equation 8 shows that the forecast accuracy of the three ARIMA models is very good. It is because the three ARIMA models have MAPE values below 10%. The lowest MAPE value of the three ARIMA model candidates is ARIMA (1, 2, 0). Thus, it can be concluded that the ARIMA model (1, 2, 0) is the best model of the three candidates for time series forecasting models to predict tuberculosis morbidity data in Indonesia. The results of the calculation of the AIC and MAPE values are shown in table 6.

3.6. Forecasting Result of ARIMA (1, 2, 0)
ARIMA model is chosen as the next best ARIMA model used for forecasting tuberculosis morbidity rates for future periods. ARIMA model (1,2,0) is used to forecast tuberculosis morbidity rate in Indonesia for future periods based on the best AIC and MAPE values from previous stage. The results of forecasting the tuberculosis morbidity rate in Indonesia from 2000 to 2027 and forecasting for the next ten years are shown in table 7 and illustrated in figure 3. Figure 3 shows that the tuberculosis morbidity rate in Indonesia will decrease until 2027, which is plotted with a blue diamond line. The results of forecasting tuberculosis morbidity in Indonesia for the next ten years from 2018 to 2027 are carried out to provide information for insurance companies in calculating the amount of premium to be paid to customers from 2018 to 2027.

| Table 6. AIC and MAPE value for three ARIMA model candidates in this study. |
|-----------------------------------------------|
| ARIMA model | AIC       | MAPE     |
|--------------|----------|----------|
| ARIMA(1, 2, 0) | -181.0120 | 0.1682 % |
| ARIMA(1, 2, 1) | -179.0234 | 0.1701 % |
| ARIMA(0, 2, 1) | -179.6326 | 0.1783 % |

| Table 7. The forecasted value of tuberculosis morbidity rate in Indonesia on the ARIMA (1, 2, 0). |
|-----------------------------------------------|
| Year | Forecasted value (%) | Year | Forecasted value (%) |
|------|----------------------|------|----------------------|
| 2018 | 0.316                | 2023 | 0.301                |
| 2019 | 0.313                | 2024 | 0.298                |
| 2020 | 0.310                | 2025 | 0.295                |
| 2021 | 0.307                | 2026 | 0.292                |
| 2022 | 0.304                | 2027 | 0.289                |
4. Conclusion
The construction of the time series model of tuberculosis morbidity rate in Indonesia with the ARIMA method can be used to forecast the morbidity rate for the next period. Based on the results of data analysis, it is found that the three candidate models of ARIMA (1, 2, 0), ARIMA (1, 2, 1), and ARIMA (0, 2, 1) are suitable to be used as forecasting tuberculosis morbidity rate in Indonesia. From the AIC and MAPE values, it is found that the ARIMA model (1, 2, 0) is the best model. Therefore, the ARIMA model (1, 2, 0) is used to predict the tuberculosis morbidity rate in Indonesia from 2018 to 2027. The results of forecasting tuberculosis morbidity rates are expected to help insurance companies in determining the amount of premiums paid by customers so that insurance companies can make financial plans and provide information to customers to pay premiums in the next period.

Acknowledgments
This work was financially supported by Universitas Indonesia under research grant PITTA with grant contract number NKB-0668/UN2.R3.1/HKP.05.00/2019.

References
[1] Koch A and Mizrahi V 2018 Trends Microbiol. 26 555-6
[2] Yue J C, Wang H C, Leong Y Y and Su W P 2018 Insur. Math. Econ. 78 316-24
[3] Kesorn K, Ongruk P, Chompoosri J, Phumee A, Thavara U, Tawatsin A and Siriyasatien P 2015 PLoS One 10 1-16
[4] Costa G J, Mello M R G, Ferreira C G, Bergmann A and Thuler L C S 2018 Lung Cancer 125 77-85
[5] Zhang L, Zheng Y, Wang K, Zhang X and Zheng Y 2014 Comput. Biol. Med. 49 67-73
[6] Song Q, Zhao M, Zhou X H and Zheng Y J 2017 Ecol. Indic. 82 76-81
[7] Montgomery D C, Lennings C L and Kulahci M 2015 Introduction to Time Series Analysis and Forecasting (Canada: John Wiley & Sons)
[8] Wang C, Li Y, Feng W, Liu K, Zhang S, Hu F, Jiao S, Lao X, Ni H and Xu G 2017 Int. J. Environ. Res. Public Health 14 1-10
[9] Olawale A N and Donaldson E I 2018 Am. J. Appl. Math. Stat. 6 170-175
[10] Lim C, Chen M and Chen M 2018 Online J. Public Health Inform. 10 1
[11] Rios M, Garcia J M, Sánchez J A and Pérez D 2000 Eur. J. Epidemiol. 16 483-8
[12] Moosazadeh M, Khanjani N, Nasehi M and Bahrampour A 2015 Iran J Public Heal. 44 1526-34
[13] Ade S, Békou W, Adjobimey M, Adjibode O, Ade G, Harries A D and Anagonou S 2016 Tuberc. Res. Treat. 2016 1-9
[14] Cheung Y W and La K S 1995 J. Bus. Econ. Stat. 13 277-80
[15] Yüreki K, Kurunç A and Öztürk F 2005 Turkish J. Eng. Environ. Sci. 29 61-74
[16] Yap B W and Sim C H 2011 J. Stat. Comput. Sim. 81 2141-55
[17] Engle R F 1982 Econometrica 50 987-1008
[18] Lewis C D 1982 Industrial and Business Forecasting Methods (London: Butterworths)
[19] The World Bank 2019 available at https://data.worldbank.org/indicator/SH.TBS.INCD?end =2017&locations=ID&start=2000