

Box-Jenkins Model for Forecasting Malaysia Life Expectancy

Wan Zakiyatussariroh Wan Husin, Nasuhar Ab. Aziz

Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Kelantan,
Malaysia

Email: wanzh@uitm.edu.my, nasuha978@uitm.edu.my

To Link this Article: <http://dx.doi.org/10.6007/IJARBSS/v11-i12/11980> DOI:10.6007/IJARBSS/v11-i12/11980

Published Date: 21 December 2021

Abstract

Life expectancy is an estimate of how long the average person might be expected to live and is most often quoted for an entire lifetime. Forecasting of future life expectancy is needed to plan for health and social services and pensions. This article attempts to propose the most appropriate time series model based on Box-Jenkins methodology to explain the behavior of Malaysia life expectancy at birth for the purpose of forecasting future life expectancy. Several autoregressive integrated moving average (ARIMA) models were developed to model Malaysia life expectancy on data collected from year 1966 to year 2016. The data which separated by gender are provided by Department of Statistics Malaysia (DOSM) on yearly basis. The results indicate that both ARIMA(1,1,1) and ARIMA(2,2,3) model performed well for both in-sample fitting and out-sample evaluation for male and female respectively with the least value of Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE).

Keywords: Box-Jenkins Methodology, Forecasting, Life Expectancy at Birth, Time Series Model.

Introduction

Life expectancy is an estimate of how long the average person might be expected to live and often quoted for an entire lifetime. Life expectancy at birth is the number of years that a newborn baby would live if they experienced the death rates of the local population at the time of their birth, throughout their life. It was calculated from the information of mortality date and year of birth of a human. Life expectancy is a key summary measure of the health and wellbeing of a population. A nation's life expectancy reflects its social and economic conditions and the quality of its public health and healthcare infrastructure, among other factors (Ho & Hendi, 2018). Globally the life expectancy increased from an average of 29 to 73 years in 2019 (Department of Economic and Social Affairs [DESA], 2019). It can be concluded that all humans generally have increased their longevity of life. Increasing life expectancy in global also affected in Malaysia trends which also showing increasing trend

(Department of Statistics Malaysia [DOSM] (2020)). A gradually increasing pattern of life expectancy in Malaysia also shows a decreasing trend of mortality rates in Malaysia. This shows that most Malaysian will live longer. For a long time period, this trend will lead to an increasing of ageing population. Malaysia is well on its way to becoming an ageing society by 2030 when it is projected that 15 per cent of its population will be 60 years old and above (DOSM 2020). As life expectancy increases, the mortality rate tends to decrease as both indexes have an inverse relationship to each other. Therefore, life expectancy is one of the important indicators in Sustainable Development Goal (SDG) number 3 which is Ensure healthy lives and promote well-being for all at all ages (Department of Economic and Social Affairs [DESA], 2020).

People are living healthier and wealthier years by years despite there are still part of the population that still need attention regarding improvement in their facilities and so on (DESA, 2020). Thus, it is important to forecast the life expectancy in Malaysia to verify the achievement of the third goals of Sustainable Development Goals (SDG) which is on the good health and well-being. The SDG is also known as a global goal to end poverty, protect the planet and ensure that all human enjoy peace and prosperity. The intention of this goals to ensure that all population live in good health and wellbeing in 2030 (DESA,2020).

Therefore, life expectancy forecasting is important to ensure a good human life planning. It was the essential of human resource management in deciding retiring age and health care allocation planning. Additionally, in measuring the performance of the SDG in achieving the target, forecasting of life expectancy is become essential. An accurate forecast value needed by government or any individuals concerned, such as financial institution for better planning. Forecasts of life expectancy are also become an important component of public policy that influences age-based entitlement programs such as social security and medicare. According to Mafauzy (2000), by not realizing the importance to forecast life expectancy it will cause a problem to new generation and government. The longer life expectancy can cause oversaturated job market that will result when young people without experience will be unemployed. If no forecasting made on life expectancy, government would have a problem in controlling all the ageing populations. A massive growing of ageing population will give government a burden in keeping their health (Mafauzy, 2000). This problem also will lead to a wrong prediction in making a future planning to the country.

Hence, there are a lot of research conducted by previous researchers (e.g: Dicker et al., 2019, Castillo et al., 2017; Shang, 2016; Bennett et al., 2015; Reza and Farad, 2015; Denton *et al.*, 2005) in forecasting life expectancy, globally. In the case of Malaysia, there is limited study found regarding to predicting life expectancy where the focus of the researchers (Husin & Abidin, 2020; Kamaruddin & Ismail, 2018; Shair et al., 2017; Husin et al., 2016; Ngataman et al., 2015; Husin et al., 2014) is more on forecasting mortality rates. However, currently, there are several studies such as (Shair et al., 2019). Islam *et al* (2017) which focus on forecasting life expectancy. Thus, this work aims to offer additional literature on forecasting Malaysia life expectancy in the future by using Box-jenkins model.

Method

The data on yearly Malaysia life expectancy at birth separated by gender used in this study was obtained from Department of Statistics Malaysia (DOSM). Life expectancy at birth is an

estimate of the average number of years a newborn baby is expected to live, if he or she were to experience the age-specific mortality rates of the reference period throughout his or her life. The data covered life expectancy at birth from 1970 to 2018 for male and female, respectively. In this study, the data were divided into two parts: observed data from year 1970 to year 2003 were used in developing the time series model and those from 2004 till 2018 were used for model validation.

The general Box-Jenkins models consist of the Autoregressive (AR) Models, Moving Average (MA) Model, combination of AR and MA (ARMA) model were used. The classical Box-Jenkins models assume the time series is stationary, that is, the mean and variance of the series are essentially constant through time. If the time series is not stationary, the transformation, by taking the first difference of the non-stationary time series values, is needed. In this study, the non-seasonal integrated ARMA (ARIMA) model has been considered because this study involves with yearly data. In order to tentatively identify a Box-Jenkins model, the series has been tested to verify whether it is stationary or not by plotting the series versus time.

The AR model of order p is written as:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad (1)$$

where $\phi_1, \phi_2, \dots, \phi_p$ are unknown parameters relating y_t to $y_{t-1}, y_{t-2}, \dots, y_{t-p}$. While, the MA model of order q is

$$y_t = \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (2)$$

where $\theta_1, \theta_2, \dots, \theta_p$ are unknown parameters relating y_t to $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$. The ARMA(p, q) models was written as:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (3)$$

where ϕ and θ are unknown parameters and ε are independent identically distributed error terms with zero mean. In the case of the original values y_1, y_2, \dots, y_n (life expectancy) nonstationary, the first differencing transformation is used.

$$y'_t = y_t - y_{t-1} \quad (4)$$

In the case of the series are still nonstationary, second transformation is used.

$$y'_t = (y_t - y_{t-1})(y_{t-1} - y_{t-2}) \quad (1)$$

Usually second transformation able to produce stationary time series values (Bowerman & O'Connell, 1993). Then, the non-seasonal ARIMA (p, d, q) model can be written as

$$y'_t = \mu + \phi_1 y'_{t-1} + \phi_2 y'_{t-2} + \dots + \phi_p y'_{t-p} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (4)$$

where y'_t is the differenced series.

This study follows the procedure suggested by Hyndman and Athanasopoulos (2018) in fitting the ARIMA models. The first step is investigating the data by constructing a simple plot and identifying any unusual observations exist in data set. Then, data was transformed by taking first differences of the data to achieve time series stationarity. Next step is examining the Autocorrelation function (ACF) and Partial Autocorrelation function (PACF). In identifying the possible ARIMA (p, d, q). Then, in next step, the data are used to estimate model parameters that have been identified in the previous step. Next, adequacy of each of the model was verified by Ljung-Box statistics that are useful for testing the randomness of the residuals by testing the null hypothesis of the residuals are white noise. Based on their adequate predictions (in-sample fitting), the most appropriate models have been chosen. However, in forecasting, a model that is the best in the in-sample fitting does not necessarily provide more accurate forecasts (Hyndman & Athanasopoulos, 2018). Thus, this study also used the performance of out-of-sample forecasts to aid the selection of a statistical model. Two accuracy measures were used which are mean squares error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE) to evaluate the performance of the models. In evaluating the models, the data have been divided into two sets: training set and validation set. The training set was used to develop the time series model (in-sample fitting) and the validation set was used to evaluate forecasting performance (out-of-sample forecast). Finally, the best final model can be used to forecast future life expectancy from year 2019 to 2030 in order to evaluate the performance of Malaysia life expectancy in achieving the third goals of SDG.

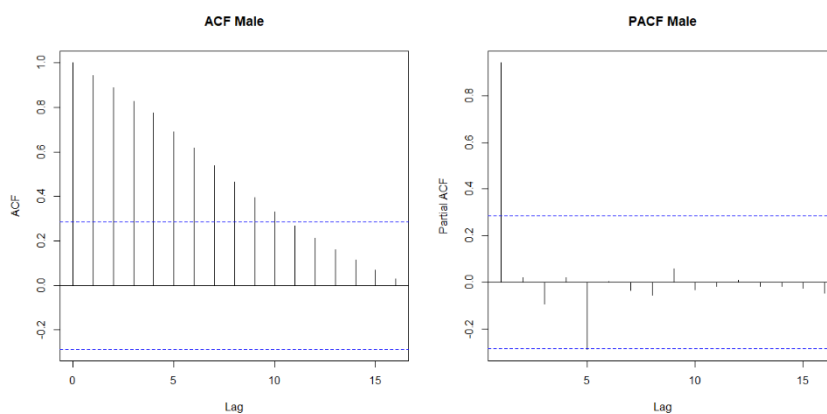
Results and Discussion

Fig.1 shows the Malaysia life expectancy at birth between 1970 and 2018 for male and female population respectively. It can be seen that the data for both genders show an upward trend toward the recent data. These could mean that the life expectancy of Malaysian is increasing over the years. With the highest value for male is 72.3 which is lower than female that is 77.3 in year 2018. Therefore, we can conclude that female have longer life expectancy than male in Malaysia. Observing Fig. 1, it shows that the data series is not stationary because it contains upward trend in male and female life expectancy. The ACF and PACF were also plotted as in Fig. 2 to collect more conclusive evidence on non-stationary data series. The ACF and PACF in Fig. 2 confirm the series is not stationary with the ACF of the original series values dies down extremely slowly with temporal dependence on dengue incidences. These findings confirmed the need to use a non-seasonal ARIMA (p, d, q) parameters in modeling life expectancy of Malaysian.



Fig.1: Time Series Plot of Life Expectancy for Male and Female Population.

Fig. 2 shows that there is no particular pattern in PACF, there are one peak that significant since they exceed the limit. The peaks are occurring at the lag 1. It also shows the normal decaying pattern of ACF. Since there are no exponential decaying pattern in both ACF and PACF figures, this means the Malaysia life expectancy for male and female population are not stationary, hence, AR (p) model and MA(q) model only are not suitable to be used to analyze life expectancy data of male Malaysian population. Since the data is not stationary as mention above, ARMA (p,q) model is also not suitable. Therefore, the data have been transformed using differencing method then ARIMA (p,d,q) model is appropriate to analyze the data. After the first differencing is done, Fig. 3 shows ACF and PACF for both genders. As seen in the Fig. 3, only the female's data has a downward trend toward the recent data. These mean that the data for female population is still not stationary. However, data for male population is seem to be stationary since the first difference values of the life expectancy seem to fluctuate around a constant mean. Further analysis using ACF and PACF are done on the data.



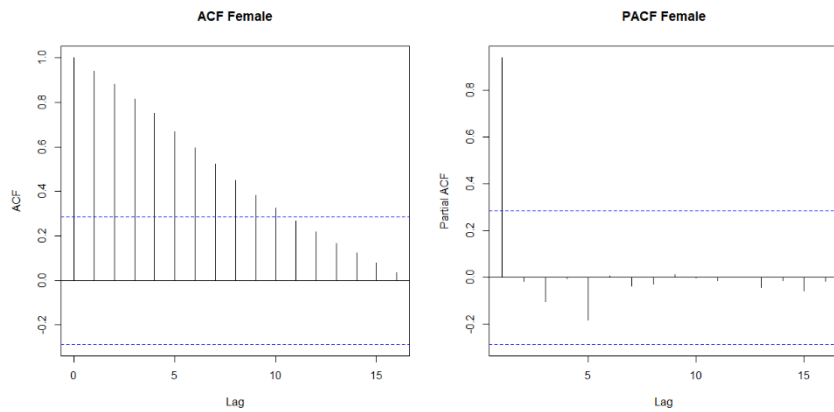


Fig.2: ACF and PACF for Life Expectancy of Male and Female Population.

Fig. 3 shows that there is no particular pattern in ACF and PACF for both genders. There are no peak that significant since none exceed the limit. The information from ACF, PACF and time series plot cannot be used to devise a model. Therefore, second differencing has to be done on the data. Fig. 4 shows the ACF and PACF plot for female data after second differencing.

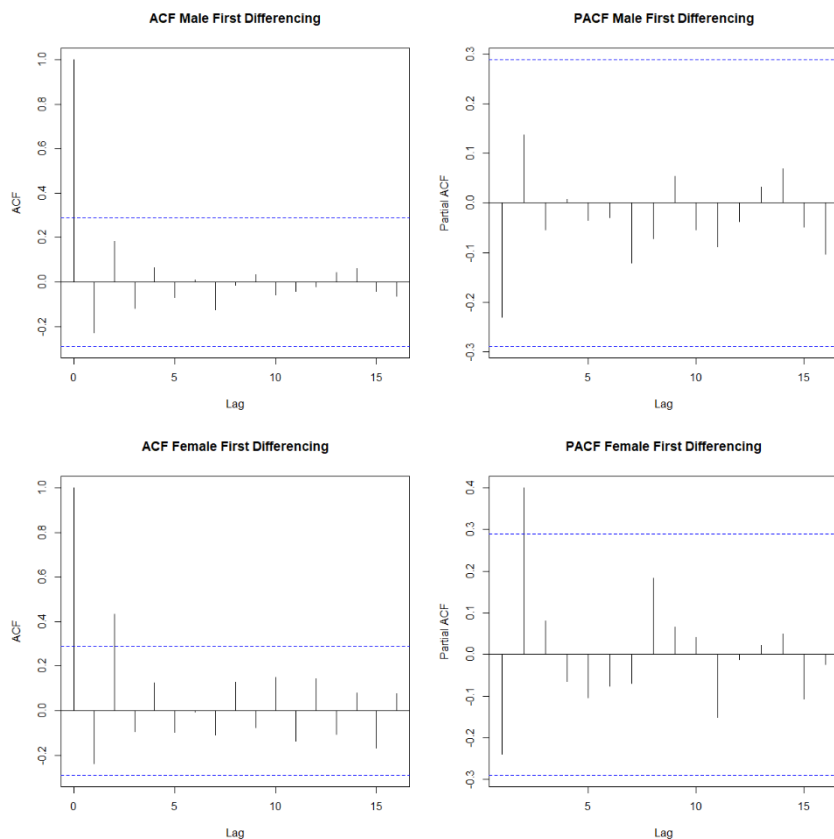


Fig. 3: ACF and PACF of First Differencing for Life Expectancy of Male and Female Population. Further analysis was done using ACF and PACF figure to confirm the status of the data. Based on the ACF and PACF Fig.4, none of the figure have pattern and have significant spikes. The information from ACF, PACF and time series plot shows that the female’s data is now stationary. Fig. 4 shows there is one significant peak that exceeds the limit in both ACF and PACF plot for male and female population. From this result, it can be deduced that ARIMA(1,2,1) is the suitable model for both male and female population data. However, Lazim, (2014), stated that we cannot exactly specify the model class based on the evidence

provide by the ACF and PACF only since we do not know the impact of the magnitudes of the spikes on the model. Therefore, this study come out with four possible ARIMA model by referring to the number of peak in the Fig. 4 and perform necessary validation test on the model as a way out.

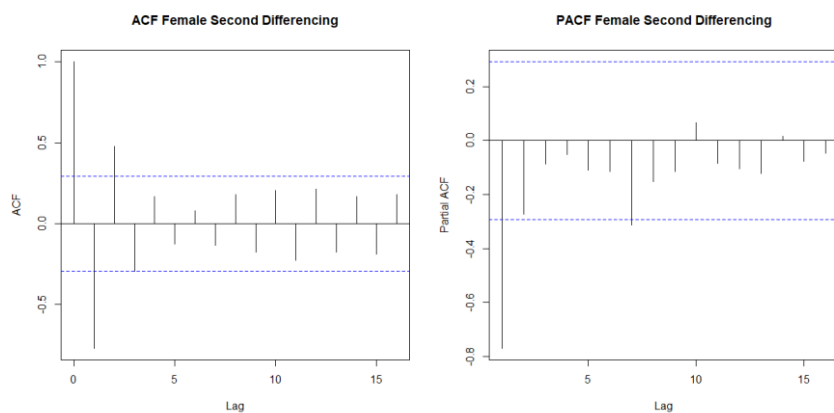


Fig 4: ACF and PACF of Second Differencing for Life Expectancy of Female Population.

Although thorough observation was made on the ACF and PACF plot, it cannot be perfectly certain of the value of respective p and q that should be assigned to the model. Hence, to ensure that a well specified model had been formed, several combinations of model formulations were estimated. Based on the identified spikes, four different possible ARIMA models have been fitted for male and female population respectively to find the best fitting model as in Table 1 and Table 2. The model with the lowest MAE, MAPE and RMSE values has been identified as the best fitted model which ARIMA (1,1,1) and ARIMA (2,2,3) was the best fitted model for male population and for female population . In addition, it was found that all the ARIMA models were capable of representing the life expectancy in a subsequent year with relative precision when these models have been used to produce out-of-sample prediction of the Malaysian life expectancy. Here, based on the lowest values of MAE, MAPE and RMSE out-of-sample evaluation shows that ARIMA (1,1,1) model is the most accurate model for male population, while ARIMA (2,2,3) is the most accurate model for female population. Therefore, ARIMA (1,1,1) model and ARIMA (2,2,3) has been used to forecast future life expectancy for Malaysian male and female population respectively.

Table 1:

Accuracy Measures for In-Sample Estimation and Out-of Sample Evaluation for Male Population

Model	Estimation			Evaluation		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE
ARIMA (0,1,0)	0.293	0.440	0.373	0.014	1.397	1.087
ARIMA (1,1,0)	0.196	0.294	0.255	0.010	1.001	0.798
ARIMA (0,1,1)	0.219	0.331	0.299	0.013	1.305	1.026
ARIMA (1,1,1)	0.191	0.285	0.237	0.007	0.607	0.645

Table 2:

Accuracy Measures for In-Sample Estimation and Out-of Sample Evaluation for Female Population

Model	Estimation			Evaluation		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE
ARIMA (2,2,2)	0.145	0.202	0.187	0.011	1.126	1.072
ARIMA (2,2,3)	0.145	0.202	0.172	0.009	0.911	0.870
ARIMA (2,2,1)	0.157	0.219	0.193	0.011	1.078	1.032
ARIMA (1,2,1)	0.157	0.219	0.195	0.013	1.296	1.228

Forecast value generated until years 2030 will aid in outlining relevance effort especially in making Malaysian citizen as well-being societies with high living standard as highlighted in sustainable development goal (SDG) number 3 where the target is ensuring healthy lives and promote well-being for all at all ages by year 2030 (United Nation, 2020). After finding the best model to forecast life expectancy of Malaysian, it's time to actually generate the forecast value. The forecast values were a fifteen-step ahead forecast which are from 2019 till 2030. These forecast values were generated using ARIMA (1,1,1) for male population and ARIMA (2,2,3) for female population. The forecast values for life expectancy of Malaysian's male and female population from year 2019 until 2030 are presented in Table 3. Based on forecast values, the trend of forecast value of life expectancy for male and female population is increasing slightly from year 2019 to 2030. As the result shown in the Table 3, the forecasted life expectancy at birth of male Malaysian population for year 2030 is 75 years old while the forecasted life expectancy at birth of female Malaysian population for year 2030 is 77 years old.

Table 3:

Forecast Value for Life Expectancy of Male and Female Population

Year	Male Population			Female Population		
	Forecast	95% Confidence Interval		Forecast	95% Confidence Interval	
		Lower	Upper		Lower	Upper
2019	72.7506	71.5049	73.9962	77.1391	76.1522	78.1260
2020	72.9305	71.5151	74.3459	77.1469	75.9942	78.2996
2021	73.1121	71.5444	74.6798	77.1848	75.8494	78.5202
2022	73.2934	71.5870	74.9997	77.1948	75.7123	78.6772
2023	73.4747	71.6402	75.3092	77.2125	75.5819	78.8430
2024	73.6560	71.7017	75.6103	77.2157	75.4517	78.9796
2024	73.8373	71.7702	75.9045	77.2181	75.3231	79.1130
2025	74.0186	71.8445	76.1928	77.2112	75.1925	79.2299
2026	74.2000	71.9238	76.4761	77.2003	75.0606	79.3400
2027	74.3813	72.0075	76.7551	77.1821	74.9257	79.4386
2028	74.5626	72.0950	77.0301	77.1587	74.7878	79.5296
2029	74.7439	72.1860	77.3018	77.1288	74.6462	79.6114

Year	Male Population			Female Population		
	Forecast	95% Confidence Interval		Forecast	95% Confidence Interval	
		Lower	Upper		Lower	Upper
2030	74.9252	72.2801	77.5703	77.0932	74.5007	79.6856

Conclusion

This research's main aim is to forecast life expectancy of Malaysian at birth. The Box-Jenkins model was used in modeling and forecasting the Malaysian life expectancy based on yearly data. The model prediction was developed by using ARIMA model based on historical data. Several ARIMA models were investigated to identify the most appropriate model which can provide accurate prediction values. As can be seen the life expectancy of Malaysia male and female population shows increasing pattern from year 1966 to year 2016 represent the growth of longevity of Malaysian life, over time. The result shows that ARIMA (0,1,0) was the best fitted model for male population with the smallest values of AIC and SBC that well reflect the trend in the life expectancy in Malaysia where the model was able to fit the historical data adequately. However, for female population, the ARIMA (2,2,1) is the best if based on AIC but if based on BIC, ARIMA (1,2,1) is become the best model. In addition, it was found that all the ARIMA models were capable of representing the life expectancy in a subsequent year with relative precision when these models have been used to produce out-of-sample prediction of the Malaysian life expectancy. Here, out-of-sample evaluation shows that ARIMA (1,1,0) model is the most accurate model for male population, while ARIMA (2,2,3) is the most accurate model for female population. Both models are expected to generate an accurate prediction for the Malaysian life expectancy for male and female population respectively. In addition, the forecast values show that the trend Malaysian of life expectancy for male and female population is increasing from year 2016 to 2030. The expected life expectancy at birth of male Malaysian population for year 2030 is 74 years old while the forecast life expectancy at birth of female Malaysian population for year 2030 is 77 years old. Therefore, we would recommend future researcher to use more advance technique such as Lee-Carter technique. This study forecasting style also only used one dependent variable which is not most suitable for the topic since life expectancy was affected by a lot factors such as life style, environment and common knowledge of resident on life expectancy. Booth et al. (2002), revealed that forecasting the life expectancy can be found by using ARIMA or other time series extrapolation procedures. It has the advantage of simplicity, but it cannot account for changes in mortality at specific ages, and the forecast number cannot be used to derive other life table results. Therefore, future researcher may consider to use more advance technique such as Lee-Carter technique which able to model age-specific mortality rates.

Acknowledgements

We thank the Malaysian Department of Statistics (DOSM) for the data on Malaysian life expectancy at birth. The authors are grateful to the Editor, the Associate Editor and anonymous referee for insightful comments and suggestions.

References

- Bennett, J. E., Li, G., Foreman, K., Best, N., Kontis, V., Pearson, C., & Ezzati, M. (2015). The future of life expectancy and life expectancy inequalities in England and Wales: Bayesian spatiotemporal forecasting. *The Lancet*, 386(9989), 163-170.
- Denton, F. T., Feaver, C. H., & Spencer, B. G. (2005). Time series analysis and stochastic forecasting: An econometric study of mortality and life expectancy. *Journal of Population Economics*, 18(2), 203-227.
- Department of Statistics Malaysia. (2020), Retrieved from https://www.dosm.gov.my/v1/index.php?r=column/cthemedByCat&cat=155&bul_id=OVByWjg5YkQ3MWFZRTN5bDJiaEVhZz09&menu_id=L0pheU43NWJwRWVSzklWdzQ4TlhUUT09
- Department of Economic and Social Affairs. (2020), United Nations. The Sustainable Development Goals Report 2020. Retrieved from <https://unstats.un.org/sdgs/report/2020/>
- Department of Economic and Social Affairs. (2019) United Nations. World population prospects 2019: Highlights. *New York (US): United Nations Department for Economic and Social Affairs*.
- Dicker, D., Nguyen, G., Abate, D., Abate, K. H., Abay, S. M., Abbafati, C., ... & Abdelalim, A. (2018). Global, regional, and national age-sex-specific mortality and life expectancy, 1950–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The lancet*, 392(10159), 1684-1735.
- Husin, W. Z. W., & Abidin, N. S. Z. (2020). Point forecasts of mortality rates in Malaysia: A comparison of principal component methods. *Indonesian Journal of Electrical Engineering and Computer Science*, 19(3), 1661-1669.
- Husin, W. Z. W., Zainol, M. S., & Ramli, N. M. (2016). Common Factor Model with Multiple Trends for Forecasting Short Term Mortality. *Engineering Letters*, 24(1).
- Husin, W. Z. W., Zainol, M. S., & Ramli, N. M. (2014, December). Evaluating the performance of the Lee-Carter method and its variants in modelling and forecasting Malaysian mortality. In *AIP Conference Proceedings* (Vol. 1635, No. 1, pp. 762-769). American Institute of Physics.
- Ho, J. Y., & Hendi, A. S. (2018). Recent trends in life expectancy across high income countries: retrospective observational study. *BMJ*, 362, k2562
- Hyndman, R. J., & Athanasopoulos, G. (2018) *Forecasting: principles and practice*, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2.
- Islam, M., Shitan, M., & Mondal, I. M. (2017). Projection of Life Expectancy at Birth of Malaysian Population: A Modelling Approach. *Journal of the Indian Academy of Geriatrics*, 13(2).
- Kamaruddin, H. S., & Ismail, N. (2018). Forecasting selected specific age mortality rate of Malaysia by using Lee-Carter model. In *Paper presented at the Journal of Physics: Conference Series*.
- Mafauzy, M. (2000). The Problems and Challenges of the Aging Population of Malaysia. *The Malaysia Journal of Medical Sciences*, 1-3
- Ngataman, N., Ibrahim, R. I., & Yusuf, M. M. (2016). Forecasting the mortality rates of Malaysian population using Lee-Carter method. In *AIP Conference Proceedings* (Vol. 1750, No. 1, p. 020009). AIP Publishing LLC.

- Reza, H., & Farad, S. (2015). Forecasting life expectancy in couples based on the self-compassion variables and perceived social support. *International Academic Journal of Humanities*, 19-30.
- Shair, S. N., Zolkifi, N. A., Zulkefi, N. F., & Murad, A. (2019). A Functional Data Approach to the Estimation of Mortality and Life Expectancy at Birth in Developing Countries. *Pertanika Journal of Science & Technology*, 27(2).
- Shair, S., Purcal, S., & Parr, N. (2017). Evaluating extensions to coherent mortality forecasting models. *Risks*, 5(1), 16.
- Shang, H. L. (2016). Mortality and life expectancy forecasting for a group of populations in developed countries: a multilevel functional data method. *The Annals of Applied Statistics*, 10(3), 1639-1672.