BAREKENG: Journal of Mathematics and Its Application
December 2022 Volume 16 Issue 4 Page 1241–1248
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol16iss4pp1241-1248

# LEE-CARTER MODELING FOR MORTALITY IN INDONESIA WITH A BAYESIAN APPROACH

**Muhammad Fajar**[1,2*], **Eko Fajariyanto**[3]

[1]*Department of Statistics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember*
*Jl. Teknik Mesin No. 175, Keputih, Sukolilo, Surabaya, Jawa Timur, 60115, Indonesia*
[2]*BPS-Statistics, Province of Banten, Indonesia*
*Jl. Syeh Nawawi Al Bantani Kav. H 1-2, Kec. Serang, Kota Serang, Banten, 42171, Indonesia*
[3]*BPS-Statistics Indonesia, Province of Bengkulu*
*Jl. Adam Malik Km. 8 Bengkulu, Bengkulu City, 38225, Indonesia*

*Corresponding author's e-mail: ¹* mfajar@bps.go.id*

***Abstract.*** *This study aims to model Lee-Carter mortality with a Bayesian approach, where the parameters in the model are assumed to be random variables. The data used in this study is data on mortality rates by age group from the period 1950–2015. The source of data was from the UN website. Age groups are categorized by age 0 years, 1-5 years, 6-10 years, 11-15 years, ..., 86-90 years. The results of this study are from Bayes estimation obtained information that the average infant mortality rate (population aged less than one year) is high, then at the age of toddlers (1-4 years) average mortality rate decreases. Furthermore, the average mortality rate for children, adolescents, young and older people has increased again. Meanwhile, the relative speed of the pattern of changes in mortality at infant age (less than one year) is high enough. At the age of toddlers (1 – 4 years), the pattern of changes in mortality has increased. Then, in the population of the next age group until the older age group, the mortality continues to decrease. The pattern of changes in mortality is lowest in the elderly population.*

***Keywords:*** *Lee-Carter, Bayesian Lee-Carter, Mortality.*

***Article info:***

*Submitted: 26th June 2022* *Accepted: 15th October 2022*

***How to cite this article:***
M. Fajar and E. Fajariyanto "LEE-CARTER MODELING FOR MORTALITY IN INDONESIA WITH A BAYESIAN APPROACH", *BAREKENG: J. Math. & App.*, vol. 16, iss. 4, pp. 1241-1248, Dec., 2022.

## 1. INTRODUCTION

Mortality is one of the elements of population growth factors, in addition to fertility and migration, which function as a component of natural population reduction. In demography, mortality measures consist of crude death rate (CDR), infant mortality rate (IMR), maternal mortality rate (MMR), and age-specific death rate (ASDR) [1]. Mortality is an indicator that can be used for monitoring the development of public health and as a monitoring tool to monitor the success of health services and programs. The higher the mortality rate in an area, the lower the level of public health in that area which reflects that the health services and programs that have been implemented have not been successful.

Mortality as a variable is an input material to demographic research (in this case modeling is carried out). Lee-Carter introduced a method for modeling and predicting mortality (in this case was ASDR) with a stochastic approach (then this method is known as the Lee-Carter model [2]. The Lee-Carter model has been successfully applied to mortality forecasting in several countries, including Norway [3], Malaysia [4], Indonesia [5] [6], the US [2], Nordic (Denmark, Finland, Norway, and Sweden) [7], Sweden [8], Peru [9], India [10], Hungary [11], Sri Lanka [12] and Mauritius [13]. Several studies that have already been mentioned previously refer to the frequency approach in the parameter estimation process. In this approach, the parameter is considered a constant. Meanwhile, on the other hand, there is a Bayesian approach in the estimation process. The Bayesian approach assumes that the parameter is a random variable (not a constant). Therefore, the authors examine the Lee-Carter mortality model in Indonesia using the Bayesian approach in this paper. Thus, this paper can add new information and model treasures in the field of demography. Then, the algorithm for estimating the Lee-Carter model with Bayesian in this study uses the Pedroza algorithm [14].

This study uses a long data period, namely 1950-2015. Neves and Migon [15] find that a short period (4 years) made it impossible to model mortality using Bayesian. Another research by Neves, Fernandes, and Hoeltgebaum. Another research by Neves et al. [16] has used the Lee-Carter model to estimate mortality rates in five types of distributions, namely Poisson, Binomial, Negative Binomial, Gaussian, and Beta. This study shows that the Negative Binomial extension of the Lee-Carter model is the most appropriate for forecasting mortality.

This paper has applied Lee-Carter mortality modeling with the structure of this paper includes the Introduction section, which contains the background of the problem. The content of the literature review section is the explanation of the Lee-Carter model and the estimation of the Lee-Carter model using the Bayesian approach. The content of the methodology is the data sources and analytical procedures used in this paper. And the content of the result and discussion section is an explanation and interpretation of the estimation results of the Lee-Carter model with the Bayesian approach. The Conclusion section contains the conclusions obtained from the results of the discussion.

## 2. RESEARCH METHODS

### 2.1 Literature Review

#### 2.1.1 Lee-Carter Model

Lee and Carter introduced a method for predicting death at a certain age using the model:

$$\boldsymbol{y}_t = \boldsymbol{\alpha} + \boldsymbol{\beta}\kappa_t + \boldsymbol{\varepsilon}_t, \tag{1}$$

where:

$y_{it}$ is the natural logarithm of the mortality rate in age group $i$ $(i = 1, \dots, p)$ in year $t$ $(t = 1, \dots, n)$, $\boldsymbol{y}_t = (y_{1t}, \dots, y_{pt})'$;

$\boldsymbol{\alpha}$ $(\boldsymbol{\alpha} = \alpha_1, \dots, \alpha_p)'$ is an intercept vector of size $p \times 1$ to measure the general pattern of mortality for each age group;

$\boldsymbol{\beta}$ $(\boldsymbol{\beta} = \beta_1, \dots, \beta_p)'$ is a vector of size p×1 for the relative change in mortality rate at each age;

$\kappa_t$ is a time series process constant;

$\boldsymbol{\varepsilon}_t$ is an error vector of size $p \times 1$.

To ensure that Equation (1) can be identified, then:

$$\sum_i \beta_i = 1, \quad \sum_t \kappa_t = 0 \tag{2}$$

To estimate Equation (1), Singular Value Decomposition (SVD) is using SVD applied to the corrected log-mortality mean $-\widehat{\boldsymbol{A}}$ :

$$\boldsymbol{Y} - \widehat{\boldsymbol{A}} = \boldsymbol{UDV}' \tag{3}$$

where:

$\boldsymbol{Y}$ is the natural logarithm matrix of mortality in age groups throughout the observation period $(p \times n\ dimension)$;

$\widehat{\boldsymbol{A}}$ is a matrix where each column contains repeatedly the vector $\widehat{\boldsymbol{\alpha}}$ $(p \times n\ dimension)$;

$\widehat{\boldsymbol{\alpha}}$ is the estimated vector of $\boldsymbol{\alpha}$ (the arithmetic mean of the natural logarithm of mortality for each particular age groups during the observation period);

$\boldsymbol{D}$ is a singular diagonal matrix value, $\boldsymbol{D} = diag(d_1, \dots, d_r)$;

$\boldsymbol{U}$ is an orthogonal matrix where the value depends on age;

$\boldsymbol{V}$ is an orthogonal matrix where its value is time-dependent;

$r$ is the rank of $\boldsymbol{Y} - \widehat{\boldsymbol{A}}$;

$\boldsymbol{\beta}$ is also the first column of $\boldsymbol{U}$ and $\kappa_t$ is the first column of $\boldsymbol{V}$.

$\widehat{\boldsymbol{\beta}}$ is the estimated vector of $\boldsymbol{\beta}$.

Estimation using the principal component method is applied to $\boldsymbol{Y} - \widehat{\boldsymbol{A}}$.

After the estimation of $\kappa_t$ with the principal component is obtained, then for forecasting $\kappa_t$, the ARIMA (1,0,0) model is used, as follows:

$$\kappa_t = \delta + \kappa_{t-1} + \epsilon_t, \tag{4}$$

where $\delta$ is a constant (drift), and $\epsilon_t$ is an error, so the forecast for the h period ahead is:

$$\hat{\kappa}_{n+h} = \hat{\delta} + \hat{\kappa}_{n+h-1} \tag{5}$$

Equation (5) makes implications for forecasting using Equation (1) so that:

$$\hat{\boldsymbol{y}}_{n+h} = \widehat{\boldsymbol{\alpha}} + \widehat{\boldsymbol{\beta}}\kappa_{n+h} \tag{6}$$

### 2.1.2 Bayesian Lee-Carter

Pedroza [14] reformulates the Lee-Carter model into State-Space form as follows:

$$\boldsymbol{y}_t = \boldsymbol{\alpha} + \boldsymbol{\beta}\kappa_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \overset{iid}{\sim} \mathrm{N}_p(\boldsymbol{0}, \sigma_\varepsilon^2 \boldsymbol{I}) \tag{7}$$

$$\kappa_t = \delta + \kappa_{t-1} + \epsilon_t, \quad \epsilon_t \overset{iid}{\sim} \mathrm{N}\ (0, \sigma_\epsilon^2) \tag{8}$$

It is assumed that $\boldsymbol{\varepsilon}_t$ and $\epsilon_t$ are independent, and $\boldsymbol{I}$ is the identity matrix. In the Bayesian view, prior information is needed on the parameters, meaning that the parameters are considered random variables. In research using the same prior in reference, namely: $p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \delta) \propto 1, p(\sigma_\varepsilon^2) \propto 1/\sigma_\varepsilon^2$ , $p(\sigma_\epsilon^2) \propto 1/\sigma_\epsilon^2$ , and $\kappa_0 \sim \mathrm{N}(a_0, Q_0)$ with $\kappa_0$, $a_0$, $Q_0$ is the initial value that must be set first.

Algorithm for estimating Equations (7) and (8):

1. Run the Kalman filter with the updating equation:

$\boldsymbol{v}_t = \boldsymbol{y}_t - \boldsymbol{\alpha} - \boldsymbol{\beta} a_t, \quad Q_t = \boldsymbol{\beta} R_t \boldsymbol{\beta}' + \sigma_\varepsilon^2 \boldsymbol{I}, \quad K_t = R_t \boldsymbol{\beta}' Q_t^{-1}$

$a_{t+1} = a_t + \delta + K_t \boldsymbol{v}_t, \quad R_{t+1} = R_t(1 - K_t \boldsymbol{\beta}) + \sigma_\epsilon^2$

For $t = 1, \dots, n$ and save the result $a_t, Q_t,$ and $R_t$. Next, the state vector sample follows:

a. Generate sample from $\kappa_n | Y^n \sim \mathrm{N}(a_n, Q_n)$ where $Y^n$ is the observation data set from $t = 1, \dots, n$ then
b. For each $t = n-1, n-2, \dots, 1, 0$, generate the sample $\kappa_t$ of $p(\kappa_t | \kappa_{t+1}, Y^n)$ with:

$\kappa_t | \kappa_{t+1}, Y^n \sim \mathrm{N}(m_t, M_t)$

$m_t = a_t + B_t(\kappa_{t+1} - a_{t+1}), \quad M_t = Q_t - B_t R_{t+1} B_t', \quad \text{and} \quad B_t = Q_t R_{t+1}^{-1}$

2. Generate $\sigma_\varepsilon^2$ from

$$\sigma_\varepsilon^2 | Y^n, \boldsymbol{\kappa}, \boldsymbol{\alpha}, \boldsymbol{\beta} \sim \text{Invers Gamma}\left(\frac{pn}{2}, \frac{\sum_i \sum_t (y_{it} - \alpha_i - \beta_i \kappa_t)^2}{2}\right)$$

3. Generate $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ by regressing the OLS for each age group from $y_t$ over $\boldsymbol{\kappa}$. Let $\boldsymbol{X} = (\boldsymbol{1}, \boldsymbol{\kappa})$ where $\boldsymbol{1}$ is a vector of one, $\boldsymbol{\kappa}$ is a vector $\kappa_t$, and the conditional distributions $\alpha_i$ and $\beta_i$ for age group $\boldsymbol{i}$ are: $(\alpha_i, \beta_i) | Y^n, \boldsymbol{\kappa}, \sigma_\varepsilon^2 \sim \mathrm{N}((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}, \sigma_\varepsilon^2 (\boldsymbol{X}'\boldsymbol{X})^{-1})$
4. Generate sample $\delta$ from

$$\delta | \boldsymbol{\kappa}, \kappa_0, \sigma_\epsilon^2 \sim \mathrm{N}\left(\frac{\kappa_n - \kappa_0}{n}, \frac{\sigma_\epsilon^2}{n}\right)$$

5. Generate $\sigma_\epsilon^2$ from

$$\sigma_\epsilon^2 | \boldsymbol{\kappa}, \kappa_0, \delta \sim\sim \text{Invers Gamma}\left(\frac{n-1}{2}, \frac{\sum_t (\kappa_t - \kappa_{t-1} - \delta)^2}{2}\right)$$

Repeat step 1 to step 5 so that all values of $\boldsymbol{\alpha}, \boldsymbol{\beta}, \delta, \boldsymbol{\kappa}, \sigma_\varepsilon^2$, and $\sigma_\epsilon^2$ converge. In this study, 150,000 iterations were carried out with settingset.seed (19861008) and burn-in (a series of numbers with discarded iterations) of 1,000 (the results of the 1st to 1000th iterations were discarded).

### 2.2 Methodology

#### 2.2.1 Source of Data

The data used in this study is data on mortality rates by age group from the period 1950–2015. The data came from the UN website (https://esa.un.org/unpd/wpp/download/standard/mortality). Age groups are divided into 0 years, 1-5 years, 6-10 years, 11-15 years, ⋯, 86-90 years.

#### 2.2.2 Analytical Procedures

First, divide the age level data according to some age groups, namely 0 years, 1-5 years, 6-10 years, 11-15 years, 16-20 years, 21-25 years, 26-30 years, 31-35 years, 36-40 years, 41-45 years, 46-50 years, 51-55 years, 56-60 years, 61-65 years, 66-70 years, 71-75 years, 76-80 years, 81-85 years, and 86-90 years. The next steps are:

1. Estimation of parameter $\boldsymbol{\alpha}$ using the Bayesian Lee-Carter method,
2. Estimation of parameter $\boldsymbol{\beta}$ using the Bayesian Lee-Carter method,
3. Estimation of parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ is done through iteration until all the values of $\boldsymbol{\alpha}, \boldsymbol{\beta}, \delta, \boldsymbol{\kappa}, \sigma_\varepsilon^2$, and $\sigma_\epsilon^2$ converge.

## 3. RESULTS AND DISCUSSION

Figure 1 shows the development of mortality rates by age group over the observation period. It shows that the mortality rate of the population aged one year is high, meaning that infants less than one year are still susceptible to diseases due to their weak immune systems. However, the mortality rate decreased drastically

from age less than one year to 1-5 years and then 6-10 years. The population aged 6-10 to 51-55 years is relatively low (flat). Then, the death rate of the population aged 56-60 to 86–90 years has increased exponentially, shows that the health of the population of that age is starting to decrease and usually vulnerable to serious diseases such as heart disease, diabetes, cancer and the like. Historically, the pattern of the death rate has a similar pattern each period, the death rate decreases as the period increases. It shows that there are health improvements made by the government from time to time with the implementation of health programs such as programs to increase the number of health facilities and infrastructure, health insurance, and others.

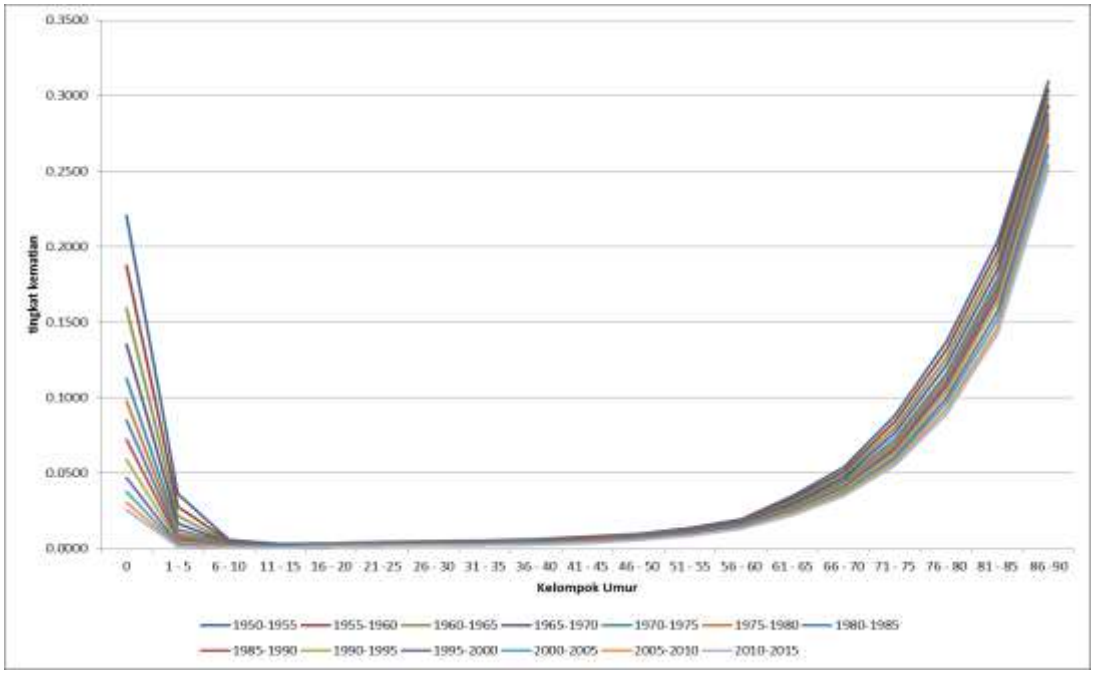


**Figure 1. Development of Mortality Rate by Age Group in Indonesia**

**Table 1. Bayesian Estimation Results Coefficient $\alpha$ and $\beta$**

| Age Group | A | B |
|---|---|---|
| 0 | -2.535 | 0.306 |
| 1 – 5 | -4.903 | 0.490 |
| 6 – 10 | -6.258 | 0.346 |
| 11 – 15 | -6.642 | 0.291 |
| 16 – 20 | -6.228 | 0.224 |
| 21 – 25 | -5.985 | 0.225 |
| 26 – 30 | -5.931 | 0.217 |
| 31 – 35 | -5.789 | 0.210 |
| 36 – 40 | -5.558 | 0.186 |
| 41 – 45 | -5.281 | 0.162 |
| 46 – 50 | -4.961 | 0.123 |
| 51 – 55 | -4.580 | 0.105 |
| 56 – 60 | -4.183 | 0.085 |
| 61 – 65 | -3.622 | 0.086 |
| 66 – 70 | -3.167 | 0.077 |
| 71 – 75 | -2.676 | 0.076 |
| 76 – 80 | -2.204 | 0.065 |
| 81 – 85 | -1.767 | 0.052 |
| 86 – 90 | -1.274 | 0.013 |

The estimation results using the Bayesian method for parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are presented in Table 1, and parameter $\kappa_t$ is presented in Table 2. The parameter estimation reflects the average mortality rate in general by age. Table 1 and Figure 2 show that the average infant mortality rate (population aged less than one year) is high, then at the age of toddlers (1–4 years old), an average mortality rate decreases. Furthermore, the average mortality rate for children, adolescents, young and old age has increased again.

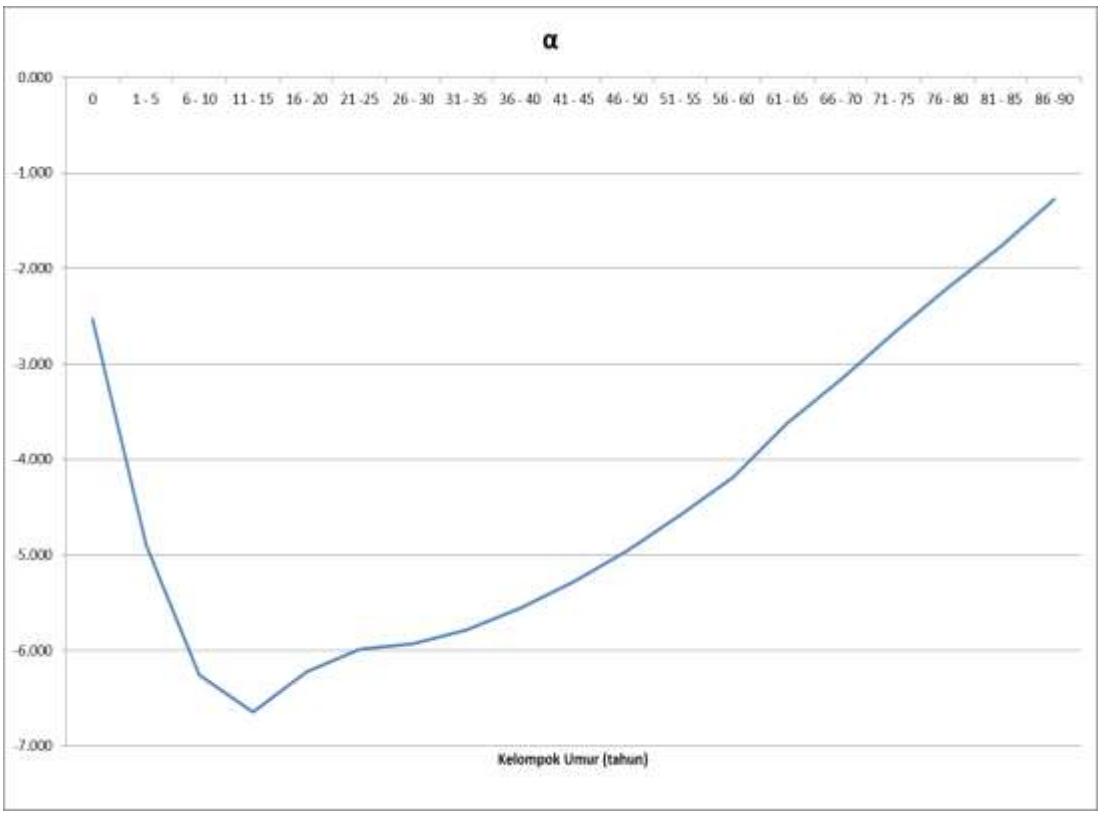


**Figure 2. The Development of Parameter $\alpha$ by Age Group in Indonesia**

Parameter $\boldsymbol{\beta}$ describes the pattern of changes in mortality rates according to age groups due to changes in the value of parameter $\boldsymbol{\kappa}$. Table 1 and Figure 3 show the relatively high rate of change in the pattern of mortality in infants (less than one year of age). At the age of toddlers (1–4 years), the pattern of changes in mortality has increased. Then in the population of the next age group until the population of the old age group, the mortality again continues to decrease. The pattern of changes in mortality is lowest in the elderly population.

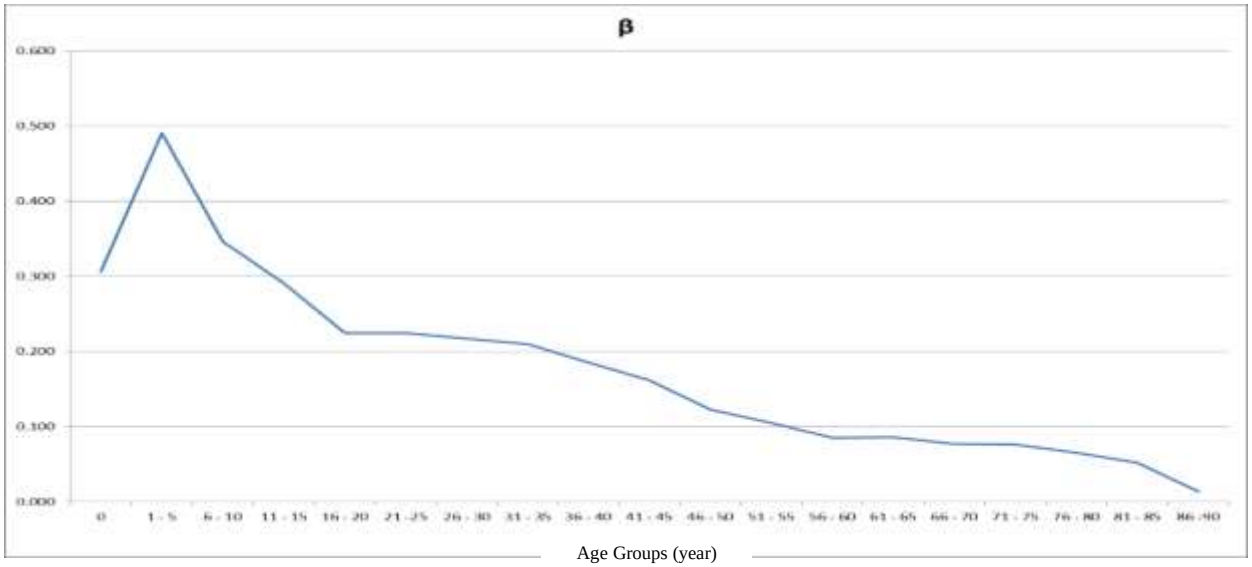


**Figure 3. The Development of Parameter $\beta$ by Age Group in Indonesia**

Parameter $\boldsymbol{\kappa}$ describes the population mortality index in year $t$. In Table 2 it shows that the estimated value of $\boldsymbol{\kappa}$ has decreased every year. It indicates that population mortality has decreased every year, so it means health services and programs have succeeded in improving the health status of the community.

**Table 2. Bayesian Estimation Results Coefficient $\kappa$**

| Year | $\kappa$ |
|---|---|
| 1950 – 1955 | 2.732 |
| 1955 – 1960 | 2.347 |
| 1960 – 1965 | 1.863 |
| 1965 – 1970 | 1.410 |
| 1970 – 1975 | 0.899 |
| 1975 – 1980 | 0.319 |
| 1980 – 1985 | -0.351 |
| 1985 – 1990 | -0.510 |
| 1990 – 1995 | -0.880 |
| 1995 – 2000 | -1.309 |
| 2000 – 2005 | -1.740 |
| 2005 – 2010 | -2.138 |
| 2010 – 2015 | -2.632 |

## 4. CONCLUSIONS

This paper has applied Lee-Carter mortality modeling with the Bayesian approach. Bayesian estimation results obtained information that the average infant mortality rate (population aged less than one year) is high, then at the age of toddlers (1–4 years), the average mortality rate decreases. Furthermore, the average mortality rate for children, adolescents, young and old age has increased again. Meanwhile, the relative speed of the pattern of changes in mortality at the age of infants (less than one year) is high enough. At the age of toddlers (1–4 years), the pattern of changes in mortality has increased. Then, in the population of the next age group until the old age group, the change in mortality continues to decrease. The pattern of changes in mortality is lowest in the elderly population.

## AKNOWLEDGEMENT

Thank you to BPS-Statistics Indonesia for always providing support and encouragement to conduct our research.

## REFERENCES

[1] S. H. Preston, P. Heuveline, and M. Guillot, *Demography Measuring and Modeling Population Processes*. United Kingdom: Blackwell Publishing, 2004.
[2] R. D. Lee and L. R. Carter, "Modeling and Forecasting U. S. Mortality," *J. Am. Stat. Assoc.*, vol. 87, no. 419, p. 659, 1992, doi: 10.2307/2290201.
[3] N. LS, "Modeling and Forecasting Norway Mortality Rates using the Lee Carter Model," *Biometrics Biostat. Int. J.*, vol. 6, no. 1, pp. 1–10, 2017, doi: 10.15406/bbij.2017.06.00158.
[4] N. Ngataman, R. I. Ibrahim, and M. M. Yusuf, "Forecasting the mortality rates of Malaysian population using Lee-Carter method," *AIP Conf. Proc.*, vol. 1750, no. June 2016, 2016, doi: 10.1063/1.4954522.
[5] L. Safitri, S. Mardiyati, and H. Rahim, "Estimation of mortality rate in Indonesia with Lee-Carter model," *AIP Conf. Proc.*, vol. 2023, no. October 2018, 2018, doi: 10.1063/1.5064207.
[6] A. H. A. Zili, S. Mardiyati, and D. Lestari, "Forecasting Indonesian Mortality Rates Using the Lee-Carter Model and ARIMA Method," vol. 02021, 2018, doi: 10.1063/1.5064209.
[7] M. C. Koissi, "Fitting and Forecasting Mortality Rates for Nordic Countries Using the Lee-Carter method," pp. 1–21.
[8] H. Lundstr and J. Qvist, "Mortality Forecasting and Trend Shifts : an Application of the Lee – Carter Model to Swedish Mortality Data £," pp. 37–50, 2004.
[9] J. C. Hernandez and A. Sikov, "Lee-Carter method for forecasting mortality for Peruvian Population," *Univ. Nac. Trujillo*, vol. 8(1), pp. 52–65, 2021, doi: dx.doi.org/10.17268/sel.mat.2021.01.05.
[10] A. Yadav, S. Yadav, and R. Kesarwani, "Decelerating Mortality Rates in Older Ages and its Prospects through Lee-Carter Approach," *PLoS One*, vol. 7, no. 12, 2012, doi: 10.1371/journal.pone.0050941.
[11] S. Baran, M. Isp, and G. Pap, "Forecasting Hungarian Mortality Rates Using The Lee – Carter Method," vol. 57, no. 1, pp.

21–34, 2007, doi: 10.1556/AOecon.57.2007.1.3.
[12] W. Aberathna, L. Alles, W. N. Wickremasinghe, and I. Hewapathirana, "Modeling and Forecasting Mortality in Sri Lanka," *Sri Lankan J. Appl. Stat.*, vol. 15–3, pp. 141–170, 2014, doi: 10.4038/sljstats.v15i3.7794.
[13] W. Dhandevi, H. M. Kang, and R. R. Ponnusamy, "Lee-Carter Mortality Forecasting : Application to Mauritian Population," no. 5, pp. 169–175, 2019.
[14] C. Pedroza, "A Bayesian forecasting model: Predicting U.S. male mortality," *Biostatistics*, vol. 7, no. 4, pp. 530–550, 2006, doi: 10.1093/biostatistics/kxj024.
[15] C. da Rocha Neves and H. S. Migon, "Bayesian graduation of mortality rates: An application to reserve evaluation," *Insur. Math. Econ.*, vol. 40, no. 3, pp. 424–434, 2007, doi: 10.1016/j.insmatheco.2006.06.005.
[16] C. Neves, C. Fernandes, and H. Hoeltgebaum, "Five different distributions for the Lee–Carter model of mortality forecasting: A comparison using GAS models," *Insur. Math. Econ.*, vol. 75, pp. 48–57, 2017, doi: 10.1016/j.insmatheco.2017.04.004.