J. Ris. & Ap. Mat. Vol. 06 No. 01 (2022) pp. 73-79
Jurnal Riset dan Aplikasi Matematika
e-ISSN: 2581-0154
URL: journal.unesa.ac.id/index.php/jram

THE COMPARISON OF RANDOM WALK WITH DRIFT AND SES IN FORECASTING INDONESIAN MORTALITY RATE WITH LEE-CARTER MODEL

RAHMAT AL KAFI[1*], ANGGIA ABYGAIL SIHOMBING[2]

[1,2]Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Indonesia, Depok 16424, Indonesia

*rahmat.alkafi@sci.ui.ac.id

ABSTRACT

Nowadays, the demand of the mortality rate data in the future is needed by the insurance companies to determine the premium which has been paid by the insured. To do so, the insurance companies require a mathematical model which is able to represent problems in forecasting mortality rate. One of the models which has been acknowledged by the Society of Actuaries of Indonesia in forecasting mortality rate is Lee-Carter (LC) model. This study used LC model to forecast the Indonesian mortality rate based on total population. The mortality rate data from 1950 to 2015 is considered. The first stage is to estimate all parameters in the LC model using least square method and singular value decomposition method. The second stage is to project the value of the time-dependent parameter in LC model by considering two methods: (1) Random walk with drift and (2) SES. Afterwards the result of parameter projection from these two methods is compared by considering the Mean Absolute Error (MAE). The final result shows that the random walk with drift is more accurate than the SES in projecting the time-dependent parameter with MAE of 0.168730974. The projected parameter obtained by random walk with drift is then used to calculate the Indonesian mortality rate for one period ahead.

Keywords: Least square, Premium, Singular value decomposition

1 Introduction

People are always trying to get protection from the unexpected risks that can befall them, the people around them, and their property. If this happens, it will pose financial risks that can have an impact on the welfare of their lives. For this reason, insurance is needed to minimize the impact of the risks. One type of insurance that is sufficiently developed in Indonesia is life insurance. The amount of risk management costs incurred by the insurance company (the insurer) depends on the insurance premium paid by the policyholder (the insured), where the insurance premium is an amount of money determined by the insurer and agreed by the insured party to be paid based on the insurance agreement.

There are several factors used in the calculation of premiums, one of which is the mortality rate. In actuarial terminology, the mortality rate is defined as the probability of someone dying in a period [1]. Information about the insured's mortality rate in the future is needed by the insurance company so that the insurance company can plan and determine the amount of the premium charged to the insured party.

2010 *Mathematics Subject Classification*: 37M10
Tanggal Masuk: 23-02-22; direvisi: 01-05-22; diterima: 15-06-22

Future mortality rates can be determined by forecasting. Various mathematical models have been developed to predict the mortality rate in a country, one of which is the mathematical model proposed by Lee and Carter [2], which was eventually known as the Lee-Carter model. Mortality forecasting using the Lee-Carter model is a popular research topic conducted by many previous researchers, such as Shukur et al. with the case study on the mortality rate data by gender in Kirkuk City, Iraq [3], Safitri et al. used the Lee-Carter model to predict the mortality rate of the Indonesian population and concerned to analyze different parameter estimation methods [4], Nigri et al. with the case study on the mortality rate data by gender in six different countries [5], Hernandez and Sikov with the case study on Peruvian mortality rate [6], Hong et al. with the case study on Malaysian mortality rate [7]. However, all existing studies only considered one method in projecting the time-dependent parameter in the Lee-Carter model.

This study presents the forecasting of Indonesian mortality rate using the Lee-Carter model by considering two different forecasting methods in random walk with drift and simple exponential smoothing (SES). The SES itself has never been combined with Lee-Carter model by previous studies. In this study, both methods are used to project the time-dependent parameter in the Lee-Carter model. However, both methods have been used in some studies to project other time series data, such as [8], [9], [10], [11], [12], and [13]. The results of the parameter projection by these two methods are compared, and the best result is considered to calculate the value of the time-dependent parameter in the next period, which is then used to calculate the Indonesian mortality rate in one period ahead.

2 Basic Theory

The data used in this study is Indonesian mortality rate from 1950 to 2015 taken from [14], as given in the following Table 1. It shows that the data is decreasing as time passes for each age group.

Table 1: The actual Indonesian mortality rate

Age Group	**1950-1955**	**1955-1960**	**1960-1965**	⋯	**2000-2005**	**2005-2010**	**2010-2015**
0	0.22061	0.18733	0.15908	⋯	0.03766	0.03047	0.02553
1-4	0.03605	0.02804	0.02182	⋯	0.00237	0.00178	0.00136
5-9	0.00553	0.00472	0.00403	⋯	0.00092	0.00079	0.00066
10-14	0.00301	0.00267	0.00236	⋯	0.00074	0.00066	0.00058
15-19	0.00362	0.00331	0.00303	⋯	0.00136	0.00129	0.00118
20-24	0.00452	0.00414	0.00380	⋯	0.00179	0.00172	0.00159
25-29	0.00473	0.00434	0.00398	⋯	0.00193	0.00186	0.00172
30-34	0.00533	0.00490	0.00450	⋯	0.00226	0.00219	0.00204
35-39	0.00636	0.00589	0.00545	⋯	0.00294	0.00286	0.00269
40-44	0.00786	0.00734	0.00685	⋯	0.00404	0.00395	0.00375
45-49	0.00979	0.00927	0.00878	⋯	0.00591	0.00582	0.00560
50-54	0.01372	0.01305	0.01242	⋯	0.00887	0.00879	0.00849
55-59	0.01955	0.01868	0.01785	⋯	0.01352	0.01346	0.01308
60-64	0.03493	0.03320	0.03156	⋯	0.02329	0.02281	0.02227
65-69	0.05401	0.05149	0.04909	⋯	0.03682	0.03589	0.03480
70-74	0.08861	0.08450	0.08059	⋯	0.05931	0.05707	0.05527
75-79	0.13722	0.13180	0.12661	⋯	0.09624	0.09236	0.08945
80-84	0.20386	0.19746	0.19130	⋯	0.15194	0.14671	0.14275
85-89	0.30937	0.30388	0.29847	⋯	0.26106	0.25456	0.25093

2.1 Lee-Carter Model

In this study, the Lee-Carter model is applied to forecast Indonesian mortality rate. The model was first introduced by Lee and Carter. The Lee-Carter model is given by the following equation [2]:

$$\ln(m_{x,t}) = a_x + b_x\kappa_t \tag{1}$$

where $m_{x,t}$ is the mortality rate at age x and period t, the parameters a_x and b_x are parameters that depend on age x, and the parameter κ_t is the parameter that depends on the period t. This model also has the following constraints

$$\sum_{t=1}^{T} \kappa_t = 0 \text{ and } \sum_{x=1}^{N} b_x = 1 \tag{2}$$

Before the value of $m_{x,t}$ is projected, we need to find the estimated value of a_x, b_x, and κ_t in Equation (1). The parameter a_x is estimated using the Least Square method such that a closed form of a_x is obtained as follows

$$\hat{a}_x = \frac{\sum_{t=1}^{T} \ln m_{x,t}}{T} \tag{3}$$

Whereas the parameters κ_t and b_x are estimated by the Singular Value Decomposition method in such a way as to obtain the closed form of κ_t and b_x, respectively, as follows

$$\hat{\kappa}_t = \sum_{x=1}^{N} \left(\ln m_{x,t} - \hat{a}_x\right) \tag{4}$$

$$\hat{b}_x = \frac{\sum_{t=1}^{T} \hat{k}_t \left(\ln m_{x,t} - \hat{a}_x\right)}{\sum_{t=1}^{T} \hat{k}_t^2} \tag{5}$$

After obtaining the estimated value of the parameters a_x, b_x, and κ_t, then the value of the κ_t parameter is projected to calculate the forecast value of $m_{x,t}$. In this study, two methods are considered to project the value of κ_t.

2.2 Random Walk with Drift

Random walk with drift is one of the time series methods. In this study, random walk with drift is used to project the value of κ_t expressed as the following equation [15]

$$\kappa_t = \kappa_{t-1} + d + e_t, \qquad t = 2,3, \dots, T \tag{6}$$

with d denotes the drift parameter and e_t denotes the error that is assumed to be normally distributed with mean 0 and variance σ_e^2. In this study, the parameter d in Equation (6) is estimated using the Least Square method such that a closed form of d is obtained as follows

$$\hat{d} = \frac{\kappa_T - \kappa_1}{T - 1} \tag{7}$$

with T represents the number of κ_t in data.

Equation (6) and Equation (7) lead to the following equation

$$\hat{\kappa}_{T+n} = \kappa_T + n\hat{d}, \qquad n = 1,2,3,\ldots \tag{8}$$

Equation (8) is used to project the value of κ_t for the next n periods.

2.3 Simple Exponential Smoothing

The Simple Exponential Smoothing (SES) is a forecasting method proposed by Brown in 1956. The forecast equation for the SES method is as follows [11]:

$$F_t = F_{t-1} + \alpha(A_{t-1} - F_{t-1}), \qquad t = 2,3,\ldots,T \tag{9}$$

with α represents the smoothing parameter, $\alpha \in [0,1]$. A_t denotes the actual value of κ at time t, and F_t denotes the fitted value of κ at time t. This method is categorized as short-term forecasting method.

3 Results and Discussion

From the values of $m_{x,t}$ in Table 1, we estimate the parameters a_x, b_x and κ_t in the Lee-Carter model. The estimated value of a_x is obtained by using Equation (3), while the estimated value of κ_t and b_x are obtained by using Equation (4) and Equation (5), respectively. The estimation results of a_x and b_x are listed in Table 2, while the estimation results of κ_t is listed in Table 3.

Table 2: The estimated values of $\boldsymbol{a_x}$ and $\boldsymbol{b_x}$

Age Group	$\hat{a}_x$	$\hat{b}_x$
0	-2.5353080	0.1105147
1-4	-4.9034110	0.1693437
5-9	-6.2581360	0.1128511
10-14	-6.6424840	0.0889009
15-19	-6.2278520	0.0617332
20-24	-5.9850260	0.0581740
25-29	-5.9308560	0.0563971
30-34	-5.7888620	0.0536559
35-39	-5.5576820	0.0479941
40-44	-5.2806140	0.0411830
45-49	-4.9614250	0.0310795
50-54	-4.5803840	0.0263810
55-59	-4.1833510	0.0217667
60-64	-3.6217730	0.0237634
65-69	-3.1671940	0.0225537
70-74	-2.6762930	0.0239695
75-79	-2.2043660	0.0214453
80-84	-1.7667250	0.0178535
85-89	-1.2736680	0.0104398

Table 3: The estimated values of $\boldsymbol{\kappa_t}$

t	Period	$\hat{\kappa}_t$
1	1950-1955	9.5980666
2	1955-1960	8.1037331
3	1960-1965	6.6080300
4	1965-1970	5.1115057
5	1970-1975	3.4717434
6	1975-1980	1.5394030
7	1980-1985	-0.3147917
8	1985-1990	-1.8021134
9	1990-1995	-3.5143535
10	1995-2000	-5.2685643
11	2000-2005	-6.4172965
12	2005-2010	-7.8005591
13	2010-2015	-9.3148033

It is clear that the estimated values of κ_t and b_x satisfy the constraints as in (2). After obtaining the estimated values of a_x, b_x, and κ_t, the next step is to project the values of κ_t using random walk with drift and SES method. Projecting the values of κ_t is started by fitting the values of κ_t to both methods, which aim to see the accuracy of the two methods in fitting the set of κ_t on Table 3. The method with the best accuracy will be used to project the value of κ_t for subsequent periods.

The first fitting of κ_t begins with random walk with drift. By using Equation (7) and κ_1 and κ_{13} (on Table 3) as its input, the estimated value of d is -1.576072492. By using Equation (6) and $d = -1.576072492$, the fitted values of κ_t for $t = 2,3,\dots,13$ are obtained and presented in Table 4. The next fitting of κ_t is by using SES method. The optimal value of α is 1 which is confirmed by trial and error among α of 0, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95 and 1. By using Equation (9) and $\alpha = 1$, the fitted values of κ_t for $t = 2,3,\dots,13$ are obtained and presented in Table 4.

Table 4: The actual and fitted value of $\boldsymbol{\kappa_t}$ for $\boldsymbol{t = 2, 3, \dots, 13}$ and its absolute error

t	Random Walk with Drift			SES		
	κ_t (actual)	κ_t (fitted)	Absolute Error	κ_t (actual)	κ_t (fitted)	Absolute Error
2	8.1037331	8.0219941	0.0817390	8.1037331	9.5980666	1.4943335
3	6.6080300	6.5276606	0.0803694	6.6080300	8.1037331	1.4957031
4	5.1115057	5.0319575	0.0795482	5.1115057	6.60803	1.4965243
5	3.4717434	3.5354332	0.0636898	3.4717434	5.1115057	1.6397623
6	1.5394030	1.8956709	0.3562679	1.5394030	3.4717434	1.9323404
7	-0.3147917	-0.0366695	0.2781222	-0.3147917	1.539403	1.8541947
8	-1.8021134	-1.8908642	0.0887508	-1.8021134	-0.3147917	1.4873217
9	-3.5143535	-3.3781859	0.1361676	-3.5143535	-1.8021134	1.7122401
10	-5.2685643	-5.0904260	0.1781383	-5.2685643	-3.5143535	1.7542108
11	-6.4172965	-6.8446368	0.4273403	-6.4172965	-5.2685643	1.1487322
12	-7.8005591	-7.9933690	0.1928099	-7.8005591	-6.4172965	1.3832626
13	-9.3148033	-9.3766316	0.0618283	-9.3148033	-7.8005591	1.5142442

Table 4 leads to the Mean Absolute Error (MAE) for both fitted result. The MAE of random walk with drift and SES are 0.168730974 and 1.576072492, respectively. Hence, the highest accuracy is gained by random walk with drift. Therefore, we use the random walk with drift to project the value of κ_t for the period of 2015-2020 ($t = 14$). The value of κ_{14} by random walk with drift is -10.89087579. Therefore, the Indonesian mortality rate in period of 2015-2020 for each age group can be calculated using Equation (1) which can be written as Equation (9)

$$m_{x,t} = e^{a_x + b_x \kappa_t} \tag{9}$$

with the values of a_x and b_x as in Table 2 and $\kappa_{14} = -10.89087579$. The result is listed in Table 5.

Table 5: The forecasted value of $\boldsymbol{m_{x,t}}$ for the period of 2015-2020

Age Group	$\boldsymbol{m_{x,t}}$ **(2015-2020)**
0	0.023780012
1-4	0.001173561
5-9	0.000560219
10-14	0.000495129
15-19	0.001007603
20-24	0.00133531
25-29	0.001437184
30-34	0.001706652
35-39	0.002287308
40-44	0.003249894
45-49	0.004992053
50-54	0.007691064
55-59	0.012029306
60-64	0.020638908
65-69	0.032948039
70-74	0.053006441
75-79	0.087341977
80-84	0.140694381
85-89	0.249732093

Table 5 shows that, in the period of 2015-2020, the mortality rate is higher for elderly, but the mortality rates for all ages in this period is lower than the previous periods. This finding in accordance with the previous research that conclude that the mortality rate will decrease as the time passes, and the highest mortality occur on the high ages (elderly). This finding can be useful for insurance company to estimate the amount of premium for policyholder based on age.

4 Conclusion

In this study, random walk with drift and SES are compared in term of projecting time-dependent parameter of Lee-Carter model, and the best result will be considered to project Indonesia mortality rate. From the simulation result, the mean absolute error in fitting the value of κ_t for the random walk with drift is smaller than the mean absolute error for the SES method. Hence the random walk with drift is more reliable to project the value of κ_t for the period of 2015-2020. The result of κ_t projection for that period is then used in the calculation of Indonesian mortality rate for the period of 2015-2020 and each age group which values are lower than the mortality rates on the previous periods.

References

[1] D. Dickson, M. Hardy, and H. Waters, *Actuarial Mathematics for Life Contingent Risks*, New York: Cambridge University Press, 2009.

[2] N. Ngataman, R. Ibrahim, and M. Yusuf, 'Forecasting the mortality rates of Malaysian population using Lee-Carter method', *AIP Conf. Proc.*, vol. 1750, 2016.

[3] R. H. Shukur, M. A. Mohammed, and H. H. Shukr, 'Forecasting and analysis mortality using Lee-Carter model with application', *Int. J. Curr. Res.*, vol. 10, pp. 74529–74537, 2018.

[4] L. Safitri, S. Mardiyati, and H. Rahim, 'Comparison of Lee-Carter's classic and general model for forecasting mortality rate in Indonesia', *International Journal of GEOMATE*, vol. 16, pp. 119–124, 2019.

[5] A. Nigri, S. Levantesi, M. Marino, S. Scognamiglio, and F. Perla, 'A deep learning integrated Lee-Carter model', *Risks MDPI*, vol. 7, pp. 1–16, 2019.

[6] J. C. Hernandez and A. Sikov, 'Lee-Carter method for forecasting mortality for Peruvian population, *Selecciones Matemáticas*, vol. 8, pp. 52–65, 2021.

[7] W. H. Hong, J. H. Yap, G. Selvachandran, P. H. Thong, and L. H. Son, 'Forecasting mortality rates using hybrid Lee–Carter model, artificial neural network and random forest', *Complex Intell. Syst.*, vol. 7, pp. 163–189, 2021.

[8] I. Naim and T. Mahara, 'Comparative Analysis of Univariate Forecasting Techniques for Industrial Natural Gas Consumption', *Int. j. image graph. signal process.*, vol. 5, pp. 33–44, 2018.

[9] A. H. A. Zili, S. Mardiyati, and D. Lestari, 'Forecasting Indonesian mortality rates using the Lee-Carter model and ARIMA method', *AIP Conf. Proc.*, vol. 2023, 2018.

[10] F. Sidqi and I. D. Sumitra, 'Forecasting product selling using single exponential smoothing and double exponential smoothing methods', *IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 662, 2019.

[11] S. Zafar, M. Kashif, M. I. Khan, and H. Nida, 'Application of simple exponential smoothing method for temperature forecasting in two major cities of the Punjab, Pakistan', *Agrobiological Records*, vol. 4, pp. 64–68, 2020.

[12] R. Darwas, R. Rahimullaily, and N. Abdi, 'Good Sales Forecasting Information System using Single Exponential Smoothing Method', *JTIP*, vol. 14, pp. 77–82, 2021.

[13] N. P. L. Santiari and I. G. S. Rahayuda, 'Analisis perbandingan metode single exponential smoothing dan single moving average dalam peramalan pemesanan', *Jurnal Informatika Universitas Pamulang*, vol. 6, pp. 312–318, 2021.

[14] United Nations, *World population prospects*. 2019. Accessed: January 2022

[15] Y. R. Safitri, S. Mardiyati, and M. Malik, 'The Cairns-Blake-Dowd model to forecast Indonesian mortality rates', *AIP Conf. Proc.*, vol. 2168, 2019.