Lee-Carter Model and Extensions to Forecast Thai Mortality Rate

Natthasurang Yasungnoen[1]

ABSTRACT

This paper aims to forecast Thai mortality rate using the Lee-Carter and two extension models: age-period-cohort model and the specific case of the age-period-cohort model. The parameters of models were estimated by minimizing the negative binomial deviance to fit the parameters. The assumptions of the death count were based on the equal-dispersion and the over-dispersion, a Poisson and Negative binomial distribution respectively. The Score Statistics were used to detect over-dispersion. The iterative fitting model was optimized by the Poisson maximum likelihood and Negative binomial maximum likelihood taking into consideration the associated deviance. The author fit the models into Thai population. The appropriate model was selected by Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). The Box-Jenkins methodology was applied to estimate and forecast time series index with the appropriate ARIMA time series model. The study found that Thai mortality rate tended to decrease according to increasing age in both male and female.

Keywords: Lee-Carter Model, Age-Period-Cohort Model, Mortality Rate

[1] Department of Computer Science, Faculty of Engineering, Vonchavalitkul University,
84, Ban Kho, Maung, Nakhonratchasima 30000, Thailand, Email: natthasurang_yas@vu.ac.th.

Introduction

At present, Thailand is entering the aged society. Thai population has a tendency to have a longer life because of advances in medical technology and healthcare system. Population is aware and cares about their own health more than in the past. Therefore, health quality of the population is better. Data regarding population structure are important to country development. In life insurance business, the trend of mortality rates is very useful since it involves the premium and reserve calculation for insurance company.

In the past, there were several models to forecast the mortality rate and different methods to estimate the parameter of the model. The important principles to fit the model were composed of the mortality model and the estimation method. The elegant mortality model was proposed in 1992 (Li, J. S.-H., Hardy, M. R., and Tan, K. S., 2009)which was the Lee-Carter model, the mortality model to fit and forecast the mortality rate by using the singular value decomposition, simple method and the time series applied to forecast the mortality rate. In 2002, Brouhns et al. (Brouhns, N., Denuit, M., and Vermunt, J. K,2002) presented the method of parameter estimation of the Lee-Carter model based on the assumption of the death count as the Poisson distribution using the maximum likelihood estimation to fit the parameters. Renshaw and Haberman (2006). proposed the age-period-cohort model to study the mortality rate forecast in the version of the Lee-Carter model. The fitting method used maximum likelihood estimates based on the Poisson error structure, minimizing the associated deviance. Delwards et al. (2007) and Li et al. (2009) proposed the negative binomial models assuming the death count of the population had negative binomial distribution which was called over dispersion for the Lee-Carter model. The Lee-Carter model and the age-period-cohort model are widely known to fit and forecast the mortality rate as shown in (Pittacco, E., Denuit, M., Haberman, S., and Olivieri, A., 2009) So in this paper, the author wishes to investigate the negative binomial models for the Lee-Carter model, age-period-cohort model, and the specific-age-period-cohort model.

The objective of this paper is to fit and forecast Thai mortality rate by comparing the Lee-Carter model, age-period-model and specific- age-period-cohort model. Regarding the parameter estimation method, the author considers the Poisson model and negative binomial model, a distribution of the death count of the population. Furthermore, the author wishes to seek for a suitable mortality mode to fit and forecast Thai mortality rate and life expectancy. This paper consists of the following sections: Section 2 presents

Data and Mortality Model. Section 3 represents Mortality Model. Section 4 briefly describes fitting Model. Section 5 and 6 present fitting model by the Poisson setting and negative binomial setting respectively. Section 7 is devoted to a forecast of mortality rate and life expectancy. Finally, Section 8 gives a conclusion of the results.

Data and Mortality Model

The Thai Ministry of Interior is the main source of the number of age-specific people in Thai population. The Bureau of Policy and Strategy, Ministry of Public Health is the main source of the number of death counts data during period of 1998-2012. The author computes the mortality $m_{x,t}$ rate by the Strategy, Ministry of Public Health is the main source of the number of death counts data during period of 1998-2012. The author computes the mortality $m_{x,t}$ rate by the following equation, $\widehat{m}_{x,t} = \frac{D_{x,t}}{ETR_{x,t}}$ when $\widehat{m}_{x,t}$ is the nonparametric estimation of $m_{x,t}$, $D_{x,t}$ is the observed number of deaths for age x at a year t and $ETR_{x,t}$ is the exposure-to-risk at the age x, last birthday during the year t.

Mortality Model

3.1 The Lee-Carter Model

In 1992, Lee and Carter presented the mortality model, called Lee-Carter model, which is described as the logarithm of the mortality rate ($m_{x,t}$) :

$$ln(m_{x,t}) = \alpha_x + \beta_x k_t + \varepsilon_{x,t}, \qquad (1)$$

here $x = x_1, x_2, \dots, x_p$ as ages, and $t = t_1, t_2, \dots, t_n$ as the calendar years. Where α_x : the age profile averages overtime, β_x : age-specific represents the mortality rate which is changing at each age relating to the changing of k_t, k_t ,describing the time series for the general level of mortality rate. $\varepsilon_{x,t}$ is the error term. The constraints on the parameters are : $\sum_{t=t_1}^{t_n} k_t = 0$ and $\sum_{x=x_1}^{x_p} \beta_x = 1$.

3.2 The age-period-cohort model

In 2006, Renshaw and Haberman proposed the mortality model adding the cohort effect in the Lee-Carter model, cohort=period-age.

$$\ln(m_{x,t}) = \alpha_x + \beta_x^{(0)} \iota_{t-x} + \beta_x^{(1)} k_t + \varepsilon_{x,t}, \qquad (2)$$

where $\beta_x^{(0)}$, $\beta_x^{(1)}$ measures the corresponding interaction with age, ι_{t-x} is a random cohort effect that is a function of the year of birth, $t-x$, k_t is a period effect. The following restriction is used in order to estimate the parameter:

$$\sum_{x=x_1}^{x_p} \beta_x^{(0)} = 1, \sum_{x=x_1}^{x_p} \beta_x^{(1)} = 1, \sum_{t=t_1}^{t_n} k_t = 0, \sum_{t=t_1}^{t_n} \iota_{t-x} = 0. \quad (3)$$

3.3 The special case of age-period-cohort model (Specific-APC)

This model is a special case of Age-Period-Cohort model which $\beta_x^{(0)} = 1/n$ and $\beta_x^{(1)} = 1/n$, n the number of ages.

$$\ln(m_{x,t}) = \alpha_x + \frac{1}{n}\iota_{t-x} + \frac{1}{n}k_t + \varepsilon_{x,t}. \quad (4)$$

The following restriction is used in order to estimate the parameter, $\sum_{t=t_1}^{t_n} \iota_{t-x} = 0$, $\sum_{t=t_1}^{t_n} k_t = 0$.

Fitting Model

4.1 Poisson Setting

4.1.1 Model Fitting

The assumption of the death counts $D_{x,t}$, where $D_{x,t}$ is considered to have a Poisson distribution:

$$D_{x,t} \sim Poisson(ETR_{x,t}m_{x,t}).$$

The author fits mortality model by minimizing the Poisson deviance (Renshaw, A., and Haberman, S., 2003a; 2003b; 2006) with standard generalized modeling. The iterative fitting is optimized by the Poisson maximum likelihood taken into consideration the associated deviance (Pittacco, E., Denuit, M., Haberman, S., and Olivieri, A.,2009).

$$D(d_{x,t}, \hat{d}_{x,t}) = \sum_{x,t} dev(d_{x,t}, \hat{d}_{x,t})$$

$$= \sum_{x,t} 2\omega_{x,t} \begin{Bmatrix} d_{x,t} ln\left(\frac{d_{x,t}}{\hat{d}_{x,t}}\right) \\ -(d_{x,t} - \hat{d}_{x,t}) \end{Bmatrix}, \quad (5)$$

where $\hat{d}_{x,t} = ETR_{x,t}\widehat{m}_{x,t}$ (Mahidol University Institute for Population and Social Research.,2018).

4.1.2 Parameter Estimation

To update the parameter for each model, the author uses the iterative fitting procedure which was proposed by Goodman (Goodman, L. A.,1979)

$$update(\theta) = \theta - \frac{\frac{\partial D}{\partial \theta}}{\frac{\partial^2 D}{\partial \theta^2}}. \quad (6)$$

4.2 Negative Binomial Distribution

The unconditional distribution of $D_{x,t}$ is the negative binomial (Yasungnoen, N, 2015). :

$$D_{x,t} \sim NBin(\bar{\gamma}^{-1}, \frac{(ETR_{x,t}m_{x,t}\bar{\gamma})^{-1}}{1 + (ETR_{x,t}m_{x,t}\bar{\gamma})^{-1}})$$

$$= NBin\left(\bar{\gamma}^{-1}, \frac{1}{ETR_{x,t}m_{x,t}\bar{\gamma}+1}\right). \quad (7)$$

The unconditional mean and variance are

$$E[D_{x,t}] = ETR_{x,t}m_{x,t} \quad (8)$$

and

$$Var[D_{x,t}] = ETR_{x,t}m_{x,t}\bar{\gamma}(E_{x,t}m_{x,t})^2. \quad (9)$$

4.2.1 Model Fitting

The death count variable $D_{x,t}$ assumes to be a negative binomial distribution as follows:

$$D_{x,t} \sim NB\left(\bar{\gamma}^{-1}, \frac{1}{ETR_{x,t}m_{x,t}\bar{\gamma}+1}\right). \quad (10)$$

Standard generalized linear modelling (GLM), fitting the parameters by maximizing the likelihood, corresponds to minimizing the deviance (Madsen, H., and Thyegod, P.,2010) Thus, the author fits the model by minimizing the negative binomial deviance (Yasungnoen. N.,2015). Then, the mortality model is fitted by minimizing the negative binomial deviance:

$$D(d_{x,t}, \hat{d}_{x,t}) = \sum_{x,t} dev(d_{x,t}, \hat{d}_{x,t}) \quad (11)$$

$$= \sum_{x,t} 2\omega_{x,t}\left\{\begin{matrix} d_{x,t}ln\left(\frac{d_{x,t}}{\hat{d}_{x,t}}\right) - \\ (d_{x,t} + \frac{1}{\bar{\gamma}})\ln(\frac{1+\bar{\gamma}d_{x,t}}{1+\bar{\gamma}\hat{d}_{x,t}}) \end{matrix}\right\}$$

with weight $\omega_{x,t} = \begin{cases} 1, ETR_{x,t} > 0 \\ 0, ETR_{x,t} = 0 \end{cases}$, where $\hat{d}_{x,t} = ETR_{x,t}\widehat{m}_{x,t}$.

4.2.2 Parameter Estimation

To estimate and update the parameter θ in the model, the author applies the iterative procedure proposed by Goodman (Goodman, L. A.,1979). In the iterative procedure, the author can stop the iteration when the value of deviance function increases a little (Yasungnoen, N.,2015).

5. Fitting model by the Poisson setting

The author computes the error values of the fitted mortality rate using the Lee-Carter model, the age-period-cohort model, and specific-age-period-cohort in terms of the number of deaths for the Poisson setting.

Table 1 The error values of the fitted mortality rate for male

Error	Male		
	Lee-Carter	APC	Specific-APC
MAD	0.002303028	0.006534078	**0.001846502**
MSE	**0.000030589**	0.005755655	0.000044380
RMSE	**0.005530742**	0.075866036	0.00666182
MAPE	**5.763018238**	29.9754896	8.422550608

Table 2 The error values of the fitted mortality rate for female

Error	Female		
	Lee-Carter	APC	Specific-APC
MAD	0.002390818	0.014034036	**0.002027081**
MSE	**0.000040359**	0.069088004	0.000095478
RMSE	**0.006352863**	0.26284597	0.009771269
MAPE	**6.701458816**	1361.395068	9.301141749

In Table 1 and 2, the entries in bold show the smallest error values of the fitted mortality rate with spanning period from 1998-2012. Next, the author studies a mean-variance restriction of the number of deaths by using an over dispersion test to detect overdispersion. The Score Statistics, proposed by Dean and Lawless (C.B. Dean.,1992). and Dean (Dean, C., and Lawless, J. F.,1989) is used in order to detect overdispersion. The variance function of the negative binomial distribution of the mortality model is applied as follows:

$$Var(d_{x,t}) = E[d_{x,t}] + \tau[E[d_{x,t}]]^2. \quad (12)$$

To test the negative binomial distribution data, the hypothesis $H_0 : \tau = 0$ against $H_1 : \tau > 0$. The Score Statistics Q can be shown as follows:

$$Q = \frac{\sum_{x,t}((d_{x,t} - \hat{d}_{x,t})^2 - d_{x,t})}{\sqrt{2\sum_{x,t}\hat{d}^2_{x,t}}} \sim N(0,1). \quad (13)$$

Table 4.2 The values of the Score Statistics

Gender	Lee-Carter	Age-period-cohort	Specific-APC
Male	182.94	5218.5	466.75
Female	209.96	87839.65	502.43

Table 4.2 shows the values of the Score Statistics and it is rejected because all p-values are less than 0.00001. These results showed that the mortality model allowed overdispersion. Then, for an alternative approach, the mortality model should be considered using the negative binomial distribution.

6. Fitted mortality result of the negative binomial setting

The error values of the fitted mortality rate is calculated from the Lee-Carter model, the age-period-cohort model and specific-age-period-cohort model under the Negative Binomial setting for the death count. The mortality rate is fitted the period spanning from 1998-2012.

Table 3 The error values of the fitted mortality rate for male

Error	Male		
	Lee-Carter	APC	Specific-APC
MAD	0.002286686	**0.001581155**	0.002027999
MSE	0.0000301	**0.000016022**	0.000059520
RMSE	0.005488895	**0.004002734**	0.007714906
MAPE	5.677285696	**4.316380701**	7.819115045

Table 4 The error values of the fitted mortality rate for female

Error	Female		
	Lee-Carter	APC	Specific-APC
MAD	0.0022692	**0.001502511**	0.002396202
MSE	0.000036701	**0.000032307**	0.000119697
RMSE	0.006058168	**0.005683893**	0.010940605
MAPE	6.329610183	**4.240726242**	8.485033364

The results in Table 3 to 4 show that age-period-cohort model gives less error than the other models. Then, the age-period-cohort model is chosen to forecast mortality rate.

7. Forecasting Mortality Rate and Life Expectancy

7.1 Mortality rate forecasting

A time series model is used to forecast time mortality index. The Box-Jenkins method is used to forecast the mortality index k_t and ι_{t-x} with the appropriate ARIMA time series model. Several ARIMA models are compared by checking the diagnostic, the Ljung-Box Q-test, and then the appropriate model is chosen.

From the result, age-period-cohort model and the estimated model ARIMA(0,1,0) are to be the most appropriate model for k_t and ARIMA(0,1,0) with drift for ι_{t-x} for female data. The estimated model ARIMA(1,1,0) with drift is to be the most appropriate model for k_t and ARIMA(1,0,2) with drift for ι_{t-x} for male data.

Forecasting Model : the age-period-cohort model

Let $\dot{k}_{t_n+s}$, $s = 1,2,3,\ldots,10$ be the forecast mortality index and the estimated cohort effect, $\hat{\iota}_z$, $z\epsilon[t_1 - x_k, t_n - x_1]$ can be forecasted by the following equation (Renshaw, A., and Haberman, S.,2006):

$$\tilde{\iota}_{t_n-x+s} = \begin{cases} \hat{\iota}_{t_n-x+s}, & 0<s\leq x-x_1 \\ \dot{\iota}_{t_n-x+s}, & s>x-x_1 \end{cases} \quad (14)$$

The forecasted mortality rate is as follows:

$$\dot{m}_{x,t_n+s} = \quad (15)$$

$$\hat{m}_{x,t} exp \begin{Bmatrix} \hat{\beta}_x^{(0)}(\dot{k}_{t_n+s} - \hat{k}_{t_n}) \\ +\hat{\beta}_x^{(1)}(\tilde{\iota}_{t_n-x+s} - \hat{\iota}_{t_n-x}) \end{Bmatrix}.$$

The following figure shows the observed (1998-2012) and the forecasted (2013-2022) mortality rate by the age-period-cohort model in assumption of the number of deaths being negative binomial distribution.

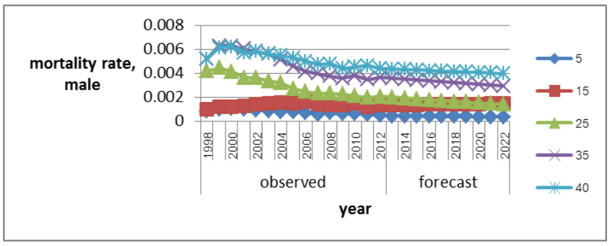

Figure 1: The observed (1998-2012) and the forecasted (2013-2022) mortality rate at age 5,15,25,35 and 40, for male

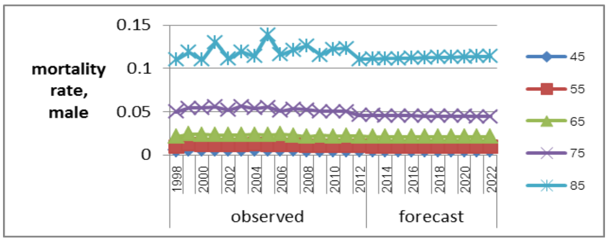

Figure 2: The observed (1998-2012) and the forecasted (2013-2022) mortality rate at age 45, 55, 65, 75, and 85 for male

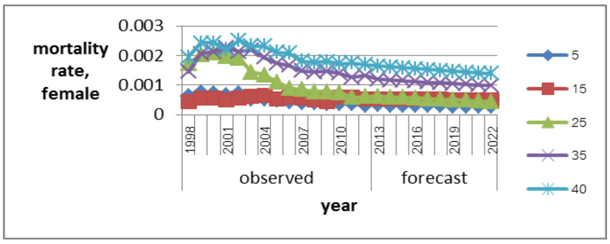

Figure 3: The observed (1998-2012) and the forecasted (2013-2022) mortality rate at age 5,15,25,35 and 40, for female

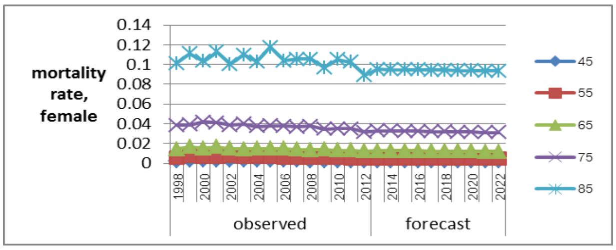

Figure 4: The observed (1998-2012) and the forecasted (2013-2022) mortality rate at age 45, 55, 65, 75, and 85 for female

7.2 Life Expectancy Forecasting

The forecasted mortality is applied to calculate the life expectancy. The author investigates the problem of forecasting from 2013 to 2022. Life expectancy at birth and life expectancy at the ages of 5, 10, 15… 80 can be found in Table 5 to 6. We note that the forecasted values of the life expectancy tend to increase from 2013 to 2022 both male. Moreover, the forecasted life expectancy all age of female are greater than males.

Table 5 The forecasted life expectancy, 2013-2022, males

Life Expectancy	Year									
	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
At birth	71.74	71.90	72.04	72.19	72.33	72.46	72.59	72.71	72.84	72.95
5	67.50	67.65	67.79	67.94	68.07	68.21	68.34	68.46	68.59	68.70
10	62.64	62.79	62.93	63.07	63.20	63.33	63.45	63.57	63.69	63.81
15	57.82	57.96	58.10	58.24	58.37	58.50	58.63	58.75	58.87	58.98
20	53.25	53.40	53.53	53.67	53.80	53.92	54.05	54.17	54.28	54.39
25	48.69	48.82	48.94	49.06	49.17	49.29	49.39	49.50	49.60	49.70
30	44.18	44.28	44.37	44.47	44.56	44.65	44.74	44.82	44.91	44.99
35	39.75	39.82	39.89	39.96	40.03	40.10	40.17	40.24	40.30	40.37
40	35.45	35.51	35.56	35.61	35.66	35.72	35.77	35.82	35.87	35.92
45	31.27	31.31	31.36	31.40	31.44	31.48	31.52	31.57	31.61	31.65
50	27.22	27.25	27.29	27.33	27.36	27.40	27.43	27.47	27.51	27.54
55	23.36	23.39	23.42	23.45	23.48	23.51	23.54	23.57	23.60	23.62
60	19.62	19.64	19.65	19.67	19.69	19.70	19.72	19.73	19.75	19.76
65	16.16	16.17	16.17	16.18	16.18	16.19	16.19	16.20	16.20	16.20
70	13.00	12.99	12.99	12.99	12.98	12.98	12.98	12.97	12.97	12.96
75	10.18	10.17	10.15	10.13	10.12	10.10	10.09	10.07	10.05	10.04
80	7.79	7.75	7.72	7.68	7.64	7.61	7.57	7.53	7.50	7.46

Table 6 The forecasted life expectancy, 2013-2022, females

Life Expectancy	Year									
	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
At birth	78.65	78.73	78.82	78.90	78.98	79.06	79.14	79.22	79.29	79.37
5	74.27	74.35	74.43	74.51	74.59	74.67	74.75	74.82	74.90	74.97
10	69.38	69.46	69.54	69.62	69.70	69.77	69.85	69.92	70.00	70.07
15	64.49	64.57	64.65	64.73	64.80	64.88	64.95	65.03	65.10	65.17
20	59.65	59.73	59.80	59.88	59.96	60.03	60.10	60.18	60.25	60.32
25	54.81	54.89	54.96	55.03	55.10	55.17	55.24	55.31	55.37	55.44
30	50.00	50.07	50.13	50.20	50.27	50.33	50.40	50.46	50.52	50.59
35	45.23	45.29	45.35	45.41	45.47	45.53	45.59	45.65	45.71	45.77
40	40.52	40.57	40.63	40.68	40.74	40.79	40.85	40.90	40.95	41.00
45	35.88	35.93	35.98	36.03	36.08	36.13	36.18	36.23	36.28	36.33
50	31.35	31.39	31.44	31.48	31.52	31.57	31.61	31.66	31.70	31.74
55	26.94	26.98	27.02	27.06	27.10	27.14	27.18	27.22	27.26	27.30
60	22.67	22.71	22.74	22.78	22.81	22.85	22.88	22.92	22.95	22.98
65	18.66	18.69	18.72	18.75	18.77	18.80	18.83	18.86	18.88	18.91
70	14.94	14.96	14.98	15.00	15.02	15.04	15.06	15.08	15.10	15.12
75	11.55	11.56	11.57	11.58	11.59	11.60	11.61	11.63	11.64	11.65
80	8.67	8.67	8.67	8.67	8.67	8.67	8.67	8.67	8.68	8.68

8. Conclusion

This paper investigates the parameter estimation the Lee-carter, age-period-cohort and specific-age-period-cohort model on two distribution of the death count. The methods of parameter estimation are used by minimizing the Poisson deviance and negative binomial deviance for the death counts as a Poisson distribution and Negative binomial distribution respectively. The results of error show that the fitted mortality rate by the age-period-cohort model give the smallest value. Therefore, the suitable model for Thailand data is the age-period-cohort model underlying the negative binomial distribution of the number of death by the Score Statistic. It is evidence to insist the data as the Negative binomial distribution. The forecasted mortality rate shows a decreasing trend from past to the future. Moreover, the forecasted life expectancy trend to increase both gender. The results confirm that Thailand is entering the aged society. Thai population has a tendency to longer life in the future.

References

Brouhns, N., Denuit, M., and Vermunt, J. K (2002). "A Poisson log-bilinear regression approach to the construction of projected lifetable". **Insurance: Mathematics and Economics.** 373-393.

C.B. Dean (1992). "Testing for overdispersion in Poisson and binomial regression". **Journal of the American Statistical Association**. 87: 451-457.

Dean, C., and Lawless, J. F. (1989). "Tests for detecting overdispersion in Poisson". **Journal of the American Statistical Association.** 84: 467-472.

Delwarde, A., Denuit, M., & Partrat, C. (2007). "Negative binomial version of the Lee-Carter model for mortality forecasting". **Applied Stochastic Models in Business and Industry.** 23:385-401.

Goodman, L. A. (1979). "Simple Models for the Analysis of Associate in Cross-Classification having Ordered categories". **Journal of the American Statistical Association.** 74(367): 537-552.

Lee, R. D., and Carter, L. R. (1992). "Modeling and Forecasting U.S. Mortality". **Journal of the American Statistical Association.** 87(419): 659-671.

Li, J. S.-H., Hardy, M. R., and Tan, K. S. (2009). "Uncertainty in mortality forecasting: an extension to the classical Lee-Carter approach". **Astin Bulletin.** 137-164.

Madsen, H., and Thyegod, P. (2010). **Introduction to general and generalized linear models.** Taylor & Francis Group, LLC.

Pittacco, E., Denuit, M., Haberman, S., and Olivieri, A. (2009). **Modeling Longevity Dynamics for Pensions and Annuity Business.** The United States: Oxford University Press Inc., New York.

Renshaw, A., and Haberman, S. (2003a). "Lee-Carter mortality forecasting with age-specific enhancement". **Insurance: Mathematics and Economics.** 33: 255-272.

Renshaw, A., and Haberman, S. (2003b). "Lee-Carter mortality forecasting: a parallel generalized linear modelling approach for English and Wales mortality projection". **Applied Statistics.** 52: 119-137.

Renshaw, A., and Haberman, S. (2006). “A cohort-based extension to the Lee-Carter model for mortality reduction factors”. **Insurance: Mathematics and Economics.** Vol 38: 556-570.

Mahidol University Institute for Population and Social Research. (2018). “Population of Thailand, 2018”. **Mahidol Population Gazette.** Vol. 27.

Natthasurang Yasungnoen. (2015). **Forecasting Thai Mortality by using a Modified Lee-Carter Model.** Doctor of Philosophy in Applied Mathematics. Suranaree University of Technology.