

Mortality Forecasting Using Lee Carter Model Implemented to French Mortality Data

Md Hasinur Rahaman Khan*, Sadia Afrin and Mohammad Shahed Masud

Institute of Statistical Research and Training (ISRT), Dhaka University, Dhaka 1000, Bangladesh

(Received: 20 April 2015; Accepted: 29 March 2016)

Abstract

For the reason of simplicity the Lee and Carter (LC) method is getting widely adopted for long-run forecasts of age specific mortality rates. In this paper the LC model is applied to French mortality data to demonstrate the mortality results. The age-specific death rates are used for the period 1816 to 2006. The index of the level of mortality, and the shape and sensitivity coefficients for each age are obtained through the LC method. The autoregressive moving average and the singular value decomposition models are used to forecast the general index for a long period of time that goes from 2007 to 2056. The projection is useful since the projected mortality rates can be used to project life expectancy at birth which is the widely used social indicator in demography.

Keywords: Age-specific mortality rates; Forecasting; France; Lee-Carter model; Mortality.

I. Introduction

A parametric curve is usually fitted to annual mortality rates for any country's data and then simple graduation techniques is applied to obtain projected rates. The attempt to find such appropriate mortality curve has a long history in demography and actuarial sciences. Many researchers proposed many studies. Over the past few decades, a number of new approaches have been developed for forecasting mortality using stochastic models, such as Alho^{2,3}, Barinaga⁴, Delwarde et al.⁷. Booth et al.,⁵ provide overviews of the development and current state of stochastic mortality modelling. Models can be classified by the number of random period or cohort effects used to describe their dynamics or by whether they are formulated in discrete or continuous time.

However, studies conducted in the last twenty years revealed many errors in the forecasts. Keilman¹⁶ reported that the earlier forecasts have missed some important events such as the post second World War baby-boom and the decline in fertility in Greece and Spain after 1985. Old-age mortality decline was also underestimated and increases in life expectancy under-projected. Recently, the LC model has been more and more popular and has been applied for long-run forecasting the age specific mortality rates for many countries. This model is computationally simple to apply. It provides successful results for various countries for many years, for instance, U.S. (Lee and Carter¹⁸), Canada (Lee and Nault²⁰), Chile (Lee and Rofman¹⁶), Japan (Wilmoth²⁷). In contrast, the model did not succeed for some developed country's data, for example, Australian data (Booth et al.⁵) and United Kingdom (Renshaw and Haberman²²). French data has also been used for mortality projection in some studies. The LC model is used in many other studies (Koissi et al.¹⁷, Renshaw and Haberman²³, Wang and Liu²⁵, Haberman and Renshaw^{11,12}, Lee and Miller¹⁹).

Variations of the LC model have been used to forecast fertility rates and migration flows (Hrdle and Mykov¹⁰, Hyndman and Ullah¹⁵). Other applications of the LC method involve population forecasts (Hollmann et al.¹⁴), the assessment of longevity risk in life insurance contracts

(Antolin¹, Brouhns et al.⁶, Dowd et al.⁸). Two frameworks for incorporating the effects of external factors into the stochastic modelling of age-specific mortality rates have been proposed recently, both following the Bayesian paradigm. Girosi and King⁹ develop a Bayesian hierarchical model for forecasting different demographic variables that includes additional explanatory variables as proxies for systematic causes. They illustrate their method by studying, for example, mortality from transportation accidents in Argentina and Chile using GDP as a proxy for the level of infrastructure and effort invested in transportation safety. Reichmuth and Sarferaz²⁴ propose a Bayesian vector autoregressive model that allows studying the short-run interaction between a latent mortality variable and several covariates.

The LC model is useful in many ways. The model represents one of the most influential recent developments in the field of mortality forecasts in recent years. The important feature of this model is that for a precise value of the time index say \hat{k}_t , we can define a complete set of death probabilities that allow to calculate the life table. The main component of LC model is the age-specific death rate that is the ratio of the number of deaths within a specified age group in a given geographic area during a certain period of time to the corresponding population at risk of the same age group, in the same geographic area during a specified time period of study. In this paper, we focus on projecting mortality rates for a long-term for French population data. We also aim at giving an overview of the LC model through describing the basic method and evaluating the performance of the method. In addition, we will study how a forecaster could get a better performance of the model by selection of an optimal time period to fit the model.

II. Data and Methodology

The data source used for the study is based on human mortality database (HMD) (www.mortality.org). The HMD started their journey in 2002. The database provides detailed mortality and population data to researchers, students, journalists, policy analysts, and others interested in the

* Author for correspondence. e-mail: hasinur@isrt.ac.bd

history of human longevity and this database is free of charge. Currently, it contains detailed data for a collection of 28 countries. The information in the HMD is standardized and includes several features of population data—live birth counts, death counts, population size on January 1st, population exposed to risk of death, death rates (period and cohort), and life tables (period and cohort). All HMD data files are organized by sex, age and time. Population size is given for one-year and five-year age groups.

In 1992, based on a combination of statistical time series methods, Lee and Carter developed a new model known as Lee-Carter (LC) model for the extrapolation of trends and age patterns in mortality. By using age-specific death rates, LC estimates an index of the level of mortality called k . Such estimation is calculated for men, women, and the total population. ARIMA and state-space time series models are used to forecast each index. Both models are compared as regards their goodness of fit and predictive capacity. Once the index of the level of mortality is forecasted, it is possible to predict death rates and life expectancy.

The LC methodology for forecasting mortality rates is a simple bilinear model in the variables x (age) and t (calendar year). The model is defined as:

$$\ln(m_{x,t}) = a_x + b_x k_t + \epsilon_{x,t} \quad (1)$$

where $m_{x,t}$ is the observed central death rate at age x in year t , a_x is the average age-specific pattern of mortality, k_t is a time trend index of general mortality level, b_x is the pattern of deviations from the age of profile as the k_t varies, and $\epsilon_{x,t}$ is the residual term at age x and time t .

The time component captures the main time trend on the logarithmic scale in mortality rates at all ages. The model includes no assumption about the nature of the trend in. The age component modifies the main time trend according to whether change at a particular age is faster or slower than the main trend. In principle, not all the need have the same sign, in which case movement in opposite directions could occur. In practice, all they do have the same sign, at least when the model is fit over fairly long periods. The model assumes that is invariant over time. The key driver of mortality dynamics in the LC model is the index of the level of mortality k_t , which can be filtered from the matrix of age-specific mortality rates with singular value decomposition. This variable is characterized as the dominant temporal pattern in the decline of mortality (Tuljapurkar et al.²⁶), as a random period effect, or simply as a latent variable.

In order to obtain a unique solution for the system of equations of the model, is set equal to the averages over time, the square values of sum to unity, and values sum to zero, these are of the forms:

$$a_x = \frac{1}{T} \sum_x \ln(m_{x,t}), \quad \sum_x b_x^2 = 1, \quad \text{and} \quad \sum_t k_t = 0.$$

The parameters of the model (1) can be estimated with different methods such as the singular value decomposition (SVD), the maximum likelihood and the weighted least square. In this paper we used the SVD method for parameter estimation because of the computational simplicity of the SVD method.

The SVD Method

The LC model cannot be fitted by ordinary regression methods, because there are no given regressors. Hence in order to find a least squares solution to the Equation (1) we use a close approximation, suggested by Lee and Carter¹⁸, to the singular value decomposition method, assuming that the errors are homoscedastic. The parameter vector a_x can be easily computed as the average over time of the logarithm of the central death rates. Then we apply the SVD to matrix $Z_{x,t} = \ln(m_{x,t}) - \hat{a}_x$, producing the matrices $ULV' = SVD(Z_{x,t}) = L_1 U_{x1} V_{t1} + \dots + L_x U_{xx} V_{tx}$. Approximation to the first term gives the estimates $\hat{b}_x = U_{x1}$ and $\hat{k}_t = L_1 V_{t1}$. The whole application of the SVD method is relative simple. Here an approximate a new matrix $\hat{Z}_{x,t}$ is obtained by the product of the estimated parameters \hat{b}_x and \hat{k}_t . That is, $\hat{Z}_{x1,t1} = \hat{b}_{x1} \hat{k}_{t1}$. In matrix format,

$$\hat{Z}_{x,t} = \begin{bmatrix} \hat{Z}_{x1,t1} & \hat{Z}_{x1,t2} & \dots & \hat{Z}_{x1,tn} \\ \hat{Z}_{x2,t1} & \hat{Z}_{x2,t2} & \dots & \hat{Z}_{x2,tn} \\ \dots & \dots & \dots & \dots \\ \hat{Z}_{xA,t1} & \hat{Z}_{xA,t2} & \dots & \hat{Z}_{xA,tn} \end{bmatrix} \hat{b}_x.$$

III. Data Analysis

We implement the LC model to age-specific mortality rates of France for the period 1816-2006 for ages that range from 0 to 110 years. The Singular Value Decomposition method gives the estimated values for logarithm of model components \hat{a}_x , \hat{b}_x and \hat{k}_t . One good property of the LC approach is that, once the data are fitted to the model and the values of the vectors a_x , b_x and k are found, only the mortality index k_t needs to be predicted. The estimates \hat{a}_x , and \hat{b}_x are presented in Table 1 (see in appendix) and the estimate \hat{k}_t is presented in Table 2 (see in appendix). The estimates are also plotted in Fig. 1. To study the efficiency of the LC model, the residual term on a logarithmic scale has been examined as well. The cohort effect does not appear very significant, and most of the residual terms show a similar lack of systematic pattern.

The age specific death rates a_x appear to be higher at ages 0 and 1 (Fig. 1). Then they decline until the age 20 and after age 20, the mortality rates again increase and that are continued to be higher until the year 100. The parameter, k declines at about the same pace during the second half. It also strikes that short run fluctuations in k do not appear much greater in the first part of the period than they do in the second, with the expectation of male in the first years. We can find that these results are consistent with the findings of LC¹⁸ in their analysis of the total French population. Both these features of k (it linearly declines with

Table 1. The estimates of Lee Carter model components, a_x and b_x .

Age	\hat{a}_x	\hat{b}_x	Age	\hat{a}_x	\hat{b}_x	Age	\hat{a}_x	\hat{b}_x	Age	\hat{a}_x	\hat{b}_x
0	-2.62	0.02	25	-5.37	0.02	50	-4.49	0.01	75	-2.53	0.01
1	-4.38	0.03	26	-5.38	0.01	51	-4.43	0.01	76	-2.45	0.01
2	-4.96	0.03	27	-5.39	0.01	52	-4.37	0.01	77	-2.36	0.01
3	-5.34	0.03	28	-5.39	0.01	53	-4.31	0.01	78	-2.27	0.01
4	-5.62	0.02	29	-5.37	0.01	54	-4.25	0.01	79	-2.17	0.01
5	-5.82	0.02	30	-5.35	0.01	55	-4.20	0.01	80	-2.07	0.01
6	-6.00	0.02	31	-5.34	0.01	56	-4.14	0.01	81	-1.98	0.01
7	-6.17	0.02	32	-5.31	0.01	57	-4.07	0.01	82	-1.87	0.00
8	-6.30	0.02	33	-5.28	0.01	58	-3.99	0.01	83	-1.78	0.00
9	-6.41	0.02	34	-5.25	0.01	59	-3.91	0.01	84	-1.69	0.00
10	-6.47	0.02	35	-5.22	0.01	60	-3.82	0.01	85	-1.61	0.00
11	-6.49	0.02	36	-5.19	0.01	61	-3.74	0.01	86	-1.54	0.00
12	-6.46	0.02	37	-5.15	0.01	62	-3.66	0.01	87	-1.47	0.00
13	-6.41	0.02	38	-5.11	0.01	63	-3.58	0.01	88	-1.41	0.00
14	-6.27	0.02	39	-5.07	0.01	64	-3.51	0.01	89	-1.35	0.00
15	-6.12	0.02	40	-5.01	0.01	65	-3.43	0.01	90	-1.28	0.00
16	-5.97	0.02	41	-4.98	0.01	66	-3.36	0.01	91	-1.21	0.00
17	-5.82	0.01	42	-4.92	0.01	67	-3.28	0.01	92	-1.15	0.00
18	-5.64	0.01	43	-4.88	0.01	68	-3.19	0.01	93	-1.09	0.00
19	-5.53	0.01	44	-4.83	0.01	69	-3.09	0.01	94	-1.04	0.00
20	-5.45	0.02	45	-4.78	0.01	70	-2.99	0.01	95	-0.96	0.00
21	-5.39	0.02	46	-4.73	0.01	71	-2.90	0.01	96	-0.89	0.00
22	-5.36	0.02	47	-4.68	0.01	72	-2.80	0.01	97	-0.81	0.00
23	-5.35	0.02	48	-4.62	0.01	73	-2.71	0.01	98	-0.73	0.00
24	-5.36	0.02	49	-4.56	0.01	74	-2.62	0.01	99	-0.64	0.00
									100	-0.54	0.00

Table 2. The estimates of Lee Carter model component k_t .

Year	\hat{k}_t	Year	\hat{k}_t	Year	\hat{k}_t	Year	\hat{k}_t	Year	\hat{k}_t
0	57.84	46	51.41	92	33.59	138	-47.15	184	-154.97
1	59.91	47	53.79	93	31.53	139	-48.83	185	-157.49
2	61.39	48	53.15	94	27.38	140	-49.01	186	-160.33
3	64.36	49	57.67	95	36.87	141	-51.93	187	-160.38
4	59.79	50	53.56	96	26.64	142	-60.66	188	-174.61
5	58.35	51	52.30	97	27.45	143	-60.99	189	-174.23
6	61.95	52	57.46	98	62.21	144	-62.61	190	-180.90
7	57.99	53	55.23	99	66.69	145	-67.12		
8	59.97	54	66.39	100	57.44	146	-63.39		
9	61.44	55	81.38	101	50.39	147	-62.21		
10	63.24	56	51.52	102	69.89	148	-69.73		
11	59.49	57	53.67	103	38.37	149	-68.33		
12	63.71	58	47.24	104	26.77	150	-71.64		
13	59.39	59	50.26	105	23.31	151	-71.55		
14	58.94	60	49.28	106	15.87	152	-71.42		
15	58.69	61	47.29	107	16.78	153	-69.13		
16	67.65	62	49.79	108	14.62	154	-76.57		
17	60.71	63	48.16	109	17.66	155	-76.31		
18	69.44	64	51.36	110	18.88	156	-78.44		
19	59.20	65	49.34	111	12.74	157	-79.78		
20	53.18	66	50.38	112	14.03	158	-82.81		
21	59.43	67	50.06	113	18.01	159	-83.61		
22	59.56	68	52.01	114	8.74	160	-85.8		

23	55.75	69	48.29	115	8.53	161	-91.41		
24	57.05	70	50.33	116	7.30	162	-92.54		
25	56.32	71	48.83	117	5.60	163	-95.34		
26	57.98	72	47.85	118	3.03	164	-96.39		
27	56.14	73	44.12	119	3.14	165	-98.02		
28	52.34	74	49.77	120	1.18	166	-101.84		
29	49.11	75	47.85	121	-0.28	167	-101.54		
30	56.37	76	49.29	122	0.51	168	-106.69		
31	56.96	77	49.28	123	-2.14	169	-107.58		
32	57.26	78	44.23	124	32.76	170	-110.56		
33	67.46	79	45.09	125	5.61	171	-117.03		
34	49.87	80	38.66	126	6.56	172	-119.96		
35	53.66	81	37.63	127	20.99	173	-121.71		
36	54.78	82	42.96	128	39.46	174	-125.25		
37	51.78	83	45.03	129	15.57	175	-127.44		
38	66.98	84	45.41	130	-14.47	176	-130.97		
39	63.77	85	40.21	131	-22.35	177	-131.66		
40	57.59	86	37.36	132	-32.46	178	-137.12		
41	59.02	87	36.17	133	-27.41	179	-138.43		
42	58.75	88	37.14	134	-35.81	180	-141.03		
43	68.92	89	36.38	135	-34.22	181	-145.98		
44	49.78	90	38.13	136	-42.02	182	-148.30		
45	56.96	91	36.56	137	-41.54	183	-150.66		

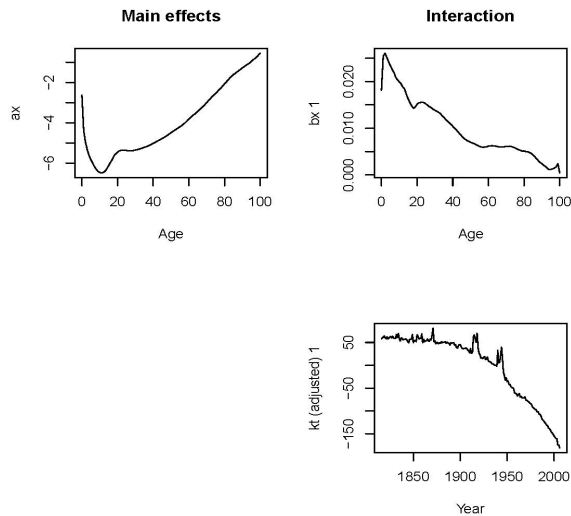


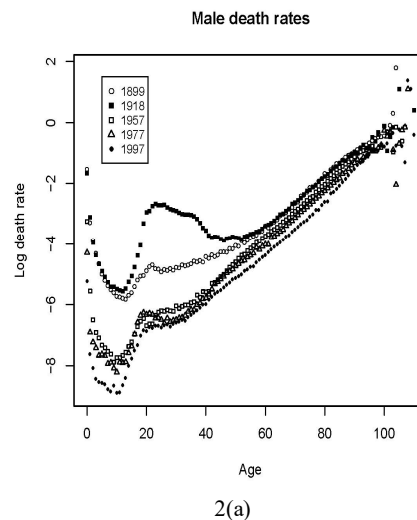
Fig 1. Graph of main and interaction effect and value of k_t of overall years.

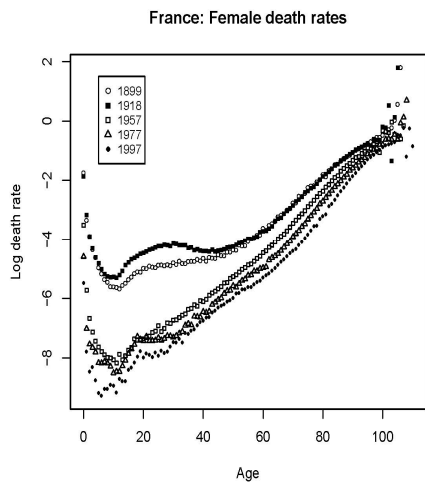
relatively constant variance) are very convenient for forecasting purpose. We can see from Fig. 2 that, the mortality pattern (log-mortality rates) is almost same in the year 1899, 1957, 1977 and 1997. Different scenario is seen in the year 1918 when there was running the 1st world war and also there were some epic broken like influenza, cholera etc. In both curves of Fig. 2 it shows that between the ages 20 to 40 the effect of mortality increases because of the First World War. After age 40 the mortality rates decrease with almost constant rates.

Forecasted mortality index are used to compute forecasted death rates using

$$\ln(m_{x,2006+t}) \approx \hat{a}_x + \hat{b}_x \hat{k}_{2006+t}$$

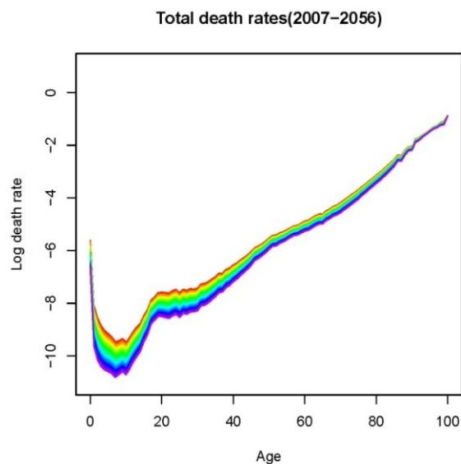
Fig. 3 (top panel) shows the fitted and forecasted mortality rates in log scale for future 50 years, obtained with the LC method. Fig. 3 (bottom panel) gives also the long-term forecasting scenario up to year 2056 using the random walk with drift where the drift is zero.



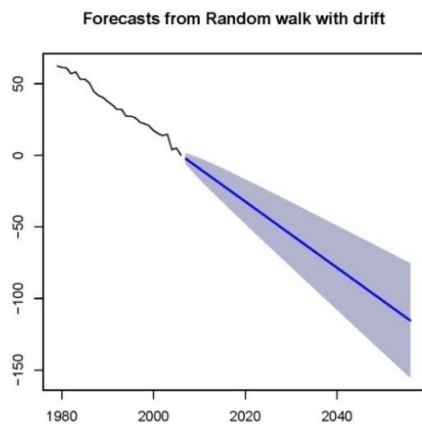


2(b)

Fig 2. Male (top panel) and female (bottom panel) death rates for several years.



3(a)



3(b)

Fig 3. Total death pattern forecast of 2007 to 2056 (top panel) and forecasted death rates with random walk with drift zero (bottom panel).

IV. Conclusions

This paper presented an application of the model underpinning the Lee-Carter methodology for forecasting vital rates. To understand how the Lee-Carter model is efficient, the residual term on a logarithmic scale was examined and found no lack of goodness of fit. In particular we focused on forecasting mortality rates on a period basis. Interesting results are found. There a priori assumption would be that the mortality rates would be different, and this is what we found in our analysis. The efficiency of the Lee-Carter method was examined with implementing the singular valued decomposition to the underlying mortality rates. The results showed that under an easy and popular estimation period, the model fit the observed death rates quite well with the singular valued decomposition. The estimates for the age parameters a and b are almost alike, while there is some variation in the estimates of the time-dependent mortality index k . This methodology could further easily be used to Bangladeshi data if there was availability of data for many years.

References

1. Antolin, P., 2007. *Longevity Risk and Private Pensions*. OECD Working Papers on Insurance and Private Pensions 3. OECD Publishing.
2. Alho, J., 1998. *A Stochastic Forecast of the Population of Finland*. Tilastokeskus, reviews 1998/4.
3. Alho, J., 2000. *Discussion of Lee Ronald D*. North American Actuarial Journal. **4** (1): 91-93.
4. Barinaga, M., 1991. *How long is the human lifespan*. Science, 254: 936-938.
5. Booth H., J. Maindonald and L. Smith, 2002. *Applying Lee-Carter under conditions of variable mortality decline*, Population Studies **56**: 325-336.
6. Brouhns, N., M. Denuit, V. K. Vermunt, 2002. *Measuring the Longevity risk in mortality projections*. Bulletin of the Swiss Association of Actuaries, 105-130.
7. Delwarde, A., M. Denuit, and P. Eilers, 2007. *Smoothing the Lee-Carter and Poisson log-bilinear models for mortality forecasting*. Statist. Model, 7: 29-48.
8. Dowd, K., D. Blake, A. J. G. Cairns, P. H. C. Eilers, and B. D. Marx, 2010. *Facing up to uncertain life expectancy: The longevity fan charts*. Demography, 47: 67-78.
9. Girosi, F. and G. King. 2008. *Demographic Forecasting*. Princeton: Princeton University Press.
10. Hrdle, W.K. and A. Mykov, 2009. *Stochastic Population Forecast for Germany and Its Consequence for the German Pension System*. SFB 649 Discussion Paper 2009. Humboldt-Universit zu Berlin.
11. Haberman, S. and A. E. Renshaw, 2003. *On the forecasting of mortality reduction factors*. Insurance Math. Econom. 32: 379-401.
12. Haberman, S. and A. E. Renshaw, 2008. *On simulation-based approaches to risk measurement in mortality with specific reference to Poisson Lee-Carter modelling*. Insurance Math. Econom. **42**: 797-816.
13. Himes, C. L., 1994. *Age Patterns of Mortality and Cause of Death Structures in Sweden, Japan, and the United States*. Demography **31** (4): 633-650.

14. Hollmann, F.W., T. J. Mulder, and J. E. Kallan, 2000. *Methodology and Assumptions for the Population Projections of the United States: 1999 to 2100. Working Paper 38. Population Division, U.S. Bureau of the Census.*
15. Hyndman, R. J. and S. Ullah, 2007. *Robust forecasting of mortality and fertility rates: A functional data approach.* Comput. Statist. and Data Anal. 51:4942-4956.
16. Keilman, N., 1998. *How accurate are the United Nations world population projections?* Population and Development Review 24 (supplement): 15-41.
17. Koissi, M. C., A. F. Shapiro, and G. Hognas, 2006. *Evaluating and extending the Lee Carter model for mortality forecasting: Bootstrap confidence interval.* Insurance Math. Econom. 26: 1-20.
18. Lee, R. D. and L. R. Carter, 1992. *Modelling and forecasting US mortality.* J. Amer. Statist. Assoc. 87: 659-671.
19. Lee R. D. and T. Miller, 2001. *Evaluating the performance of the Lee Carter method for forecasting mortality.* Demography, **38(4)**: 537-549.
20. Lee, R. D. and F. Nault, 1993. *Modeling and Forecasting Provincial Mortality in Canada*, paper presented at the World Congress of the International Union for the Scientific Study of Population, Montreal.
21. Lee, R. D. and R. Rofman, 1994. *Modelacion y Proyeccion de la Mortalidad en Chile*, NOTAS 22, no 59: 182-213.
22. Renshaw, A. E. and S. Haberman, 2003a. *Lee Carter mortality forecasting: A parallel generalised linear modelling approach for England and Wales mortality projections.* Appl. Statist. 52 119-137.
23. Renshaw, A. E. and S. Haberman, 2003b. *Lee Carter mortality forecasting with age specific enhancement.* Insurance Math. Econom. 33: 255-272.
24. Reichmuth, W. H. and S. Sarferaz., 2008. *The Influence of the Business Cycle on Mortality. SFB 649 Discussion Paper 2008-059. Humboldt-Universitt zu Berlin.*
25. Wang, C. and Y. Liu, 2010. *Comparison of mortality modelling and forecasting empirical evidence from Taiwan. Lee Carter mortality forecasting with age specific enhancement.* Int. Res. J. Fin. Econ. 37.
26. Tuljapurkar, S., and C. Boe., 1998. *Sex Mortality Differences in the United States: The Role of Cohort Smoking Patterns.* Demography, **43 (4)**: 631-646.
27. Wilmoth, J. R., 1998. *Is the pace of Japanese mortality decline converging toward international trends?* Population and Development Review **24(3)**:593-600.