Chapter 6: Fitting and Forecasting Mortality for Assam population by Applying the Lee-Carter Model

6.1 Introduction

The attempt to forecasts mortality had a long history in demography and actuarial science. Traditionally, a parametric model was fitted to annual mortality rates. The most famous parametric models in the history were De Moivre [16], Gompertz [23], Makeham [46], and Weibull [62]. In the course of recent years, various new methodologies were produced for forecasting mortality using stochastic models, such as such as Alho, Alho and Spencer, McNown and Rogers [[44], [45]] Bell and Monsell and Wang [[5], [61]], Wong-Fuppy and Haberman [66] . Lee and Carter [40] introduced the first mortality model with stochastic estimate. In this model time dependent variable is modelled through time series models. The Lee and Carter model has been used for fitting and forecasting the mortality rates for many countries: US [40], Chile [41], China [42], Japan [65], the seven most economically developed nations (G7) [55], India [[54], [68]], the Nordic countries [36], Sri Lanka [1] and Thai [69]. Lee and Carter model is computationally easy to apply and it has given successful results for various countries.

In this chapter, a long-term study of mortality rates for Assam population has been concentrated for total, male and female. There are many advantages for choosing the Lee

Carter model for forecasting mortality. The Lee-Carter model performs one of the most influential recent developments in the field of mortality forecasts. Additionally, the essential component of this model is that for a precise value of the time index k_t, a complete set of death probabilities can be characterized that enable us to compute the all of the life table functions. The estimated values of the parameters, α_x and β_x, remain consistent and invariant through time. Subsequently, the estimated parameters can be utilized for any time of intrigue. Also, traditional projection models provide the forecaster with point estimates of future mortality rates. But, the Lee Carter model allows for uncertainty in forecasts.

6.2 Objectives

In this chapter, the main aim is to examine the feasibility of using the Lee-Carter model for forecasting the mortality of Assam population. The stochastic mortality model given by Lee and Carter is used for fitting and forecasting the human mortality of Assam for both the gender. The model has been fitted to the matrix of Assam death rates based on 15 years data separately for Assam male and female populations in the form of life tables for the period 1995-99 to 2009-13. The Singular Value Decomposition (SVD) methodology is applied to estimate the parameters of the model. A time-varying index of mortality is forecasted up to 2025 year using random walk drift model (RWD) and is used to generate projected life tables.

6.3 Materials and Methods

6.3.1 Data Description

The analysis is based on the data obtained from abridged life tables of Assam male and female populations for rural and urban area for 15 years from 1995-1999 to 2009-2013. The data source used for this study is "Sample Registration System" (SRS). The data which have been used is the death rate period and the population size is five-year age groups i.e., 0-5, 5-10, $\cdots$, 80-85. From these life tables, the age group specific central death rate $(m_{x,t})$ has been derived as:

$$m_{x,t} = \frac{1}{n}\left(\frac{2q_{x,t}}{2 - q_{x,t}}\right).$$

Where $q_{x,t}$ denotes probability of dying in the age group at time t.

6.3.2 The Lee-Carter Model

Lee-Carter (1992) developed the first stochastic mortality model for forecasting human mortality is a simple bilinear model in the variables x (age) and t (calendar year) as given by:

$$\ln(m_{x,t}) = a_x + b_x k_t + \varepsilon_{x,t}, \qquad (6.1)$$

where

$m_{x,t}$: observed central death rate at age x in year t,

a_x: average age-specific pattern of mortality,

b_x: pattern of deviations from the age of profile as the k_t varies,

k_t: a time-trend index of general mortality level,

$\varepsilon_{x,t}$: the residual term at age x and time t.

The time component k_t 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 k_t. The age component b_x 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 b_x need have the same sign, in which case movement in opposite directions could occur. In practice, all the do have the same sign, at least when the model is fit over fairly long periods. The model assumes that b_x is invariant over time.

In order to obtain a unique solution the following constraints used by Lee and Carter (1992),

$$a_x = \frac{1}{T}\sum \ln\left(m_{x,t}\right), \tag{6.2}$$

$$\sum_x b_x = 1, \tag{6.3}$$

$$\sum_t k_t = 0. \tag{6.4}$$

6.3.3 Estimation of Parameters Using Singular Value Decomposition (SVD)

The Lee Carter model cannot be fitted by ordinary regression methods, because there are no given regressors; thus in order to find a least squares solution to the equation **(6.2**). The SVD methodology suggested by Lee and Carter [40] has been used for estimating the parameters assuming that the errors are homoschedastic. Here, the first stage SVD method is directly used. The different steps of the SVD method are as follows:

Step 1: From equation **(6.2**), the parameter a_x can be easily computed as the average over time of the logarithm of the central death rate as

$$\hat{a}_x = \frac{1}{T}\sum_{t=t_1}^{t_n} \ln(m_{x,t})\,, x = 0-1, 1-5, 5-10, \cdots, 80-85. \tag{6.5}$$

Step 2: Then for estimating the parameters $\hat{b}_x$ and $\hat{k}_t$, the SVD have been applied on the matrix $z_{x,t}$, where

$$z_{x,t} = \ln\left(m_{x,t}\right) - \hat{a}_x, \tag{6.6}$$

$$= \begin{pmatrix} ln\left(m_{0-1,1995-99}\right) - \hat{a}_0 & ln\left(m_{0-1,1996-2000}\right) - \hat{a}_{0-1} & \cdots & ln\left(m_{0-1,2009-13}\right) - \hat{a}_{0-1} \\ ln\left(m_{1-5,1995-99}\right) - \hat{a}_{1-5} & ln\left(m_{1-5,1996-2000}\right) - \hat{a}_{1-5} & \cdots & ln\left(m_{1-5,2009-13}\right) - \hat{a}_{1-5} \\ \vdots & \vdots & \vdots & \vdots \\ ln\left(m_{80-85,1995-99}\right) - \hat{a}_{80-85} & ln(m_{80-851996-2000}) - \hat{a}_{80-85} & \cdots & ln\left(m_{80-85,2009-13}\right) - \hat{a}_{80-85} \end{pmatrix}.$$

Step 3: Applying SVD to the matrix $z_{x,t}$ which decomposes the matrix of $z_{x,t}$ into the product of three matrices : $ULV^T = SVD\left(z_{x,t}\right) = L_1 U_{x1} V_{t1} + \cdots + L_x U_{xX} V_{tX}$, where

U representing the age component, L is the singular values and V representing the time component, where, $r = rank(z_{x,t})$, L_i $(i = 1, 2, \cdots r)$ are the singular values in increasing order with U_{xi} and V_{ti} $(i = 1, 2. \cdots, r)$ as the corresponding left and right singular vectors.

Step 4: The approximation to the first term gives the estimates $\hat{b}_x = U_{x1}$ and $\hat{k}_t = L_1 V_{t1}$. The proportion of variation explained by the LC model is $\frac{{L_1}^2}{\sum_{i=1}^{r} {L_1}^2}$.

Finally, estimate the logarithm of the central death rate, $\widehat{ln}\left(m_{x,t}\right) = \hat{a}_x + \hat{Z}_{x,t} = \hat{a}_x + \hat{b}_x \hat{k}_t$.

6.3.4 Forecasting of Mortality

One of the main advantages of using the Lee and Carter model is that once the data are fitted to the model and the estimations of the vectors $\hat{a}_x$, $\hat{b}_x$, and $\hat{k}_t$ are found, just the mortality index $\hat{k}_t$ needs to be predicted. Lee and Carter forecasted the mortality index $\hat{k}_t$ by using a standard univariate time series model autoregressive integrated moving average ARIMA (0, 1, 0). Here, an approach has been made to predict the mortality index $\hat{k}_t$ by fitting the random walk with drift model (RWD) The model is given by:

$$\hat{k}_t = \hat{k}_{t-1} + \hat{\theta} + \varepsilon_t,$$

where θ is known as the drift parameter and

$$\hat{\theta} = \frac{\hat{k}_T - \hat{k}_1}{T - 1},$$

where $\hat{\theta}$ depends on the first and last value of the estimated k_t's and ε_t is the error term.

For forecasting two periods ahead, simply put the definition of $\hat{k}_{t-1}$:

$$\hat{k}_t = \hat{k}_{t-1} + \hat{\theta} + \varepsilon_t,$$

$$= \left(\hat{k}_{t-2} + \hat{\theta} + \varepsilon_{t-1}\right) + \hat{\theta} + \varepsilon_t,$$

$$= \hat{k}_{t-2} + 2\hat{\theta} + (\varepsilon_t + \varepsilon_{t-1}).$$

To forecast $\hat{k}_t$ at time $T + \Delta t$, the same procedure is followed and iterate Δt times and obtain:

$$\hat{k}_{T+\Delta t} = \hat{k}_T + (\Delta t)\hat{\theta} + \sum_{n}^{(\Delta t)} \varepsilon_{T+n-1},$$

$$= \hat{k}_T + (\Delta t)\hat{\theta} + \sqrt{(\Delta t)}\varepsilon_t.$$

After ignoring the error term, one can forecast point estimates and is given by,

$$\hat{k}_{T+\Delta t} = \hat{k}_T + (\Delta t)\hat{\theta},$$

$$= \hat{k}_T + (\Delta t)\frac{\hat{k}_T - \hat{k}_1}{T-1},$$

which is a straight line as a function of (Δt), with slope $\hat{\theta}$.

6.4 Results and Discussion

The Lee-Carter model has been fitted for estimating the parameters based on the period $1995 - 1999$ to $2009 - 2013$ at the age groups $0 - 1,\ 1 - 5,\ 5 - 10,\ \cdots,\ 80 - 85$. The estimated values of age-dependent parameters a_x and b_x are presented in **Table 6.1** and estimated values of time dependent parameter k_t is reported in **Table 6.2** for total, male and female populations in Assam. For SVD analysis, the Matlab program has been used. From SVD analysis, it is found that 81.78%, 71.27%, 76.93% variation explained by fitted LC model for Assam total, male and female mortality data respectively. It is observed that the fitted mortality rates are very close to observed (actual) mortality rates except for lower and higher ages.

Table 6.1: Estimated $\hat{\boldsymbol{a}}_{\boldsymbol{x}}$ and $\hat{\boldsymbol{b}}_{\boldsymbol{x}}$ for Assam total, male and female (1995-99 to 2009-13).

Age group (x)	Total		Male		Female	
	$\hat{a}_x$	$\hat{b}_x$	$\hat{a}_x$	$\hat{b}_x$	$\hat{a}_x$	$\hat{b}_x$
0-1	-2.65337	-0.2157	-2.67209	0.2367	-2.67558	-0.1889
1-5	-5.01268	-0.2417	-5.08341	0.3049	-4.99308	-0.0798

5-10	-6.1274	-0.6580	-6.15856	0.7378	-6.20371	-0.5486
10-15	-6.62297	-0.3144	-6.59279	0.2881	-6.71091	-0.2302
15-20	-6.11237	-0.1065	-6.30327	-0.0262	-5.98206	-0.1960
20-25	-6.00938	-0.2898	-6.07905	0.2395	-6.00998	-0.2342
25-30	-5.7975	-0.2405	-5.73282	0.1275	-5.89945	-0.3902
30-35	-5.64016	-0.1956	-5.64753	0.0758	-5.67571	-0.2753
35-40	-5.45808	-0.1889	-5.37817	0.2066	-5.58162	-0.1568
40-45	-5.15352	-0.1969	-5.00760	0.0775	-5.41493	-0.3761
45-50	-4.64295	-0.1015	-4.52169	0.1294	-4.84377	-0.0265
50-55	-4.3963	-0.1563	-4.27487	0.1555	-4.59294	-0.1339
55-60	-3.87734	-0.0894	-3.72246	0.0144	-4.1016	-0.1387
60-65	-3.60321	-0.2224	-3.46071	0.1954	-3.84318	-0.2288
65-70	-3.0389	-0.0360	-2.97170	0.0718	-3.11881	0.0025
70-75	-2.73238	-0.0360	-2.72231	0.0198	-2.73491	-0.0723
75-80	-2.34351	0.0162	-2.25616	-0.0820	-2.48302	-0.0454
80-85	-2.14713	-0.0484	-2.08505	-0.0574	-2.25139	-0.1384

It is seen that when b_x is large for some x then the death rate at age x varies significantly when the general level of mortality changes (again, as with $x = 0$ for infant mortality) and when b_x is small, then the death rate at that age varies little when the general level of mortality changes. Parameter $\hat{a}_x$ represents the general age shape of mortality. It is found that both females and males have upward trend of mortality in general, whereas the younger ages have a lower mortality and the older ages have a higher mortality.

Table 6.2: Estimated $\hat{\boldsymbol{k}}_{\boldsymbol{t}}$ for Assam total, male, female population based on decade-wise life tables (1995-99 to 2009-13).

Time (year)	Total	Male	Female
1995-99	-0.853	0.74	-1.151
96-2000	-0.475	0.072	-1.151
97-01	-0.83	-0.495	-0.765
98-02	0.49	-1.19	0.861
99-03	-0.136	-0.609	-0.606

2000-04	-0.075	-0.376	0.006
01-05	-0.695	0.188	0.552
02-06	-0.102	-0.761	0.123
03-07	1.18	-0.236	0.267
04-08	-0.565	0.614	0.358
05-09	0.641	0.461	-0.697
60-10	2.434	0.293	-0.153
07-11	4.377	0.111	0.437
08-12	6.469	-0.084	1.072
09-13	8.71	-0.294	1.752

The mortality index, $\hat{k}_t$, captures the main time trend on the logarithmic scale in death rates at all ages. Parameter $\hat{b}_x$ describes the tendency of mortality at age x to change as the general level of mortality ($\hat{k}_t$) changes. This indicates that when $\hat{b}_x$ is large for some x, the death rate at age x varies a lot than the general level of mortality change and when $\hat{b}_x$ is small, then the death rate at that age varies a little. The Assam mortality for the prediction period of $2014-2025$ has been projected. **Table 6.3, Table 6.4** and **Table 6.5** represents the projected life expectancies at the age groups $0-1, 1-5,5-10, 10-15 \cdots 80-85$ separately for total, male and female.

Table 6.3: Projected life expectancy of total 2014-2025.

age	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025
0-1	62.5	64.9	67.1	68.9	70.5	71.9	73.0	74.0	74.8	75.5	76.1	76.6
1-5	65.2	67.0	68.5	69.8	71.0	72.0	72.9	73.7	74.3	74.9	75.4	75.8
5-10	62.5	64.0	65.3	66.5	67.5	68.4	69.2	69.9	70.5	71.0	71.4	71.9
10-15	57.9	59.2	60.4	61.5	62.5	63.4	64.2	64.9	65.5	66.0	66.4	66.9
15-20	53.1	54.4	55.6	56.6	57.6	58.4	59.2	59.9	60.5	61.0	61.5	61.9
20-25	48.6	49.9	51.0	52.0	52.9	53.8	54.5	55.1	55.7	56.2	56.6	57.0
25-30	44.1	45.2	46.2	47.2	48.1	48.8	49.5	50.2	50.7	51.2	51.6	52.0
30-35	39.6	40.6	41.6	42.4	43.2	44.0	44.6	45.2	45.8	46.2	46.6	47.0
35-40	35.1	36.1	36.9	37.7	38.5	39.2	39.8	40.3	40.8	41.3	41.7	42.0
40-45	30.7	31.5	32.3	33.1	33.8	34.4	35.0	35.5	36.0	36.4	36.7	37.1
45-50	26.4	27.1	27.8	28.5	29.1	29.6	30.2	30.6	31.1	31.5	31.8	32.1

50-55	22.5	23.1	23.7	24.3	24.8	25.3	25.7	26.2	26.5	26.8	27.1	27.4
55-60	18.5	19.1	19.5	20.0	20.4	20.8	21.2	21.5	21.8	22.1	22.3	22.5
60-65	15.1	15.5	15.9	16.3	16.6	16.9	17.1	17.3	17.5	17.7	17.9	18.1
65-70	11.6	11.7	11.9	12.0	12.2	12.3	12.4	12.6	12.7	12.9	13.0	13.1
70-75	9.0	9.1	9.1	9.2	9.3	9.3	9.4	9.5	9.5	9.6	9.6	9.7
75-80	6.4	6.4	6.4	6.4	6.3	6.3	6.3	6.3	6.3	6.2	6.2	6.1
80-85	3.9	4.0	4.0	4.0	4.1	4.1	4.2	4.2	4.3	4.3	4.4	4.4

Table 6.4: Projected life expectancy of male 2014-2025.

age	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025
0-1	61.4	63.3	65.0	66.4	67.6	68.7	69.6	70.3	70.9	71.3	71.7	71.9
1-5	63.9	65.2	66.4	67.3	68.2	68.9	69.5	70.0	70.4	70.7	70.9	71.1
5-10	61.1	62.2	63.1	63.9	64.5	65.1	65.7	66.1	66.5	66.8	67.0	67.1
10-15	56.5	57.3	58.2	58.9	59.6	60.2	60.7	61.1	61.5	61.8	62.0	62.1
15-20	51.8	52.6	53.3	54.0	54.7	55.2	55.7	56.1	56.5	56.8	57.0	57.2
20-25	47.2	48.0	48.8	49.5	50.2	50.8	51.3	51.8	52.1	52.4	52.7	52.9
25-30	42.6	43.4	44.1	44.8	45.4	45.9	46.4	46.8	47.2	47.5	47.7	47.9
30-35	38.2	38.9	39.6	40.2	40.8	41.3	41.7	42.1	42.4	42.7	42.9	43.0
35-40	33.8	34.5	35.1	35.7	36.2	36.7	37.1	37.5	37.7	38.0	38.2	38.3
40-45	29.4	30.0	30.6	31.1	31.5	31.9	32.3	32.6	32.9	33.1	33.2	33.3
45-50	25.3	25.8	26.3	26.8	27.2	27.6	27.9	28.1	28.3	28.5	28.6	28.7
50-55	21.4	21.9	22.3	22.6	22.9	23.2	23.4	23.6	23.8	23.9	23.9	23.9
55-60	17.6	17.9	18.2	18.4	18.6	18.8	18.9	19.0	19.1	19.1	19.1	19.1
60-65	14.5	14.8	15.1	15.4	15.6	15.7	15.8	15.9	15.9	15.9	15.9	15.8
65-70	11.3	11.3	11.4	11.4	11.4	11.4	11.4	11.3	11.3	11.2	11.1	11.0
70-75	8.7	8.6	8.4	8.3	8.2	8.0	7.9	7.7	7.5	7.3	7.1	6.9
75-80	6.0	5.8	5.6	5.4	5.2	4.9	4.7	4.4	4.1	3.8	3.4	3.1
80-85	3.8	3.7	3.7	3.6	3.6	3.5	3.4	3.3	3.3	3.2	3.1	3.0

Table 6.5: Projected life expectancy of female 2014-2025.

age	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025
0-1	64.4	67.1	69.4	71.3	72.8	74.1	75.2	76.1	76.8	77.4	78.0	78.4
1-5	67.0	69.0	70.6	71.9	73.1	74.0	74.9	75.5	76.1	76.7	77.1	77.5

5-10	64.6	66.4	67.9	69.2	70.2	71.0	71.7	72.3	72.8	73.2	73.5	73.9
10-15	59.9	61.6	63.0	64.2	65.2	66.0	66.7	67.3	67.8	68.2	68.5	68.9
15-20	55.1	56.8	58.1	59.3	60.2	61.0	61.7	62.3	62.8	63.2	63.6	63.9
20-25	50.7	52.2	53.4	54.5	55.4	56.2	56.8	57.4	57.8	58.2	58.6	58.9
25-30	46.1	47.5	48.7	49.7	50.5	51.2	51.9	52.4	52.8	53.2	53.6	53.9
30-35	41.4	42.7	43.8	44.8	45.6	46.3	46.9	47.4	47.8	48.2	48.6	48.9
35-40	36.9	38.0	39.0	39.9	40.7	41.3	41.9	42.4	42.9	43.2	43.6	43.9
40-45	32.4	33.5	34.4	35.2	35.9	36.5	37.0	37.5	37.9	38.3	38.6	38.9
45-50	27.9	28.7	29.5	30.3	30.9	31.5	32.1	32.5	32.9	33.3	33.6	33.9
50-55	23.8	24.7	25.5	26.2	26.9	27.4	27.9	28.4	28.8	29.1	29.4	29.6
55-60	19.8	20.5	21.2	21.8	22.3	22.8	23.2	23.6	23.9	24.2	24.5	24.7
60-65	16.0	16.6	17.1	17.5	17.9	18.3	18.6	18.9	19.2	19.4	19.6	19.8
65-70	12.1	12.4	12.7	13.0	13.2	13.5	13.8	14.0	14.2	14.4	14.6	14.8
70-75	9.5	9.9	10.2	10.6	11.0	11.3	11.6	12.0	12.3	12.6	12.8	13.1
75-80	6.9	7.1	7.3	7.5	7.7	7.9	8.1	8.3	8.4	8.6	8.7	8.8
80-85	4.1	4.2	4.4	4.5	4.6	4.7	4.7	4.8	4.8	4.9	4.9	4.9

6.5 Conclusion

An attempt has been made to identify the common trend of mortality change by fitting a standard Lee-Carter model to Assam population data. The parameters of the model are estimated by using the Singular Value Decomposition (SVD) based on the data from time series of 1995-99 to 2009-13. Based on our results it is observed that the general pattern of mortality ($\hat{a}_x$) for both male and female populations of Assam have high infant mortality, an accidental hump around ages 20 years and almost exponential increase at older ages. The sensitivity of mortality ($\hat{b}_x$) demonstrated mortality decay at high rate for ages 25-34 years for female and for ages 15-24 years for male population than other ages. It is also observed that life expectancy of female is high than male. Improvement in female mortality is larger than male mortality. The Lee-Carter Model is one of the most popular methodologies for forecasting mortality rates.

*****