Fuzzy Formulation of the Lee-Carter Model
Forecasting With Age-specific Enhancement

Yefu Kou
China Institute for Actuarial Science (CIAS)
Central University of Finance and Economics (CUFE)
39 South College Road, Haidian, Beijing 100081, China
kouyefu@sina.com, kouyefu@163.com
Outline

- Introduction
  - Explains the fuzzy formulation of the modified Lee-Carter model
  - Implements the fuzzy formulation of the classical Lee-Carter model and the modified Lee-Carter model with age-enhancement on China population data
- Conclusions
Introduction

- With rapid growth in aging and the trends in improving mortality among the elderly, there exerts significant challenges to the public pension plans as well as private pension funds and life insurers.

- There is a demand in a stochastic mortality model which adequately projects the mortality/longevity trends of the China population.

- In this paper there are two main contributions:
  - The first contribution is to consider a fuzzy formulation of the modified Lee-Carter (1992) model analyzed in Renshaw and Haberman (2003), thus extending the work of Koissi and Shapiro (2006) which only consider the fuzzy formulation of the basic Lee-Carter model.
Introduction

- In addition to mortality modeling, the fuzzy set theory has also found its usefulness in a variety of other insurance applications.

- The second contribution of the paper is to implement the fuzzy formulation of the basic Lee-Carter model and the extended Lee-Carter model (with age-specific enhancement) on the China data.

- The comparative advantages of our proposed fuzzy formulation of the extended Lee-Carter model, relative to the classical Lee-Carter model, are analyzed and discussed.
The basic Lee-Carter (LC) model

$$\ln m_{x,t} = a_x + b_x k_{t} + \varepsilon_{x,t}, \quad x = x_1, \ldots, x_N \text{ and } t = t_1, t_1 + 1, \ldots, t_1 + T - 1$$

(2.1)

An extension of the LC model is possible by including higher order terms

$$\log m_{x,t} = a_x + b_x^{(1)} k_{t}^{(1)} + b_x^{(2)} k_{t}^{(2)} + \ldots + b_x^{(r)} k_{t}^{(r)} + \varepsilon_{x,t}$$

(2.2)

The age-time interaction term $b_x^{(i)} k_{t}^{(i)}$ is referred to as the $i$-th term of the rank $r$ approximation (see Booth et al. 2001)

Renshaw and Haberman (2003) investigate the above modified Lee-Carter model with age-specific enhancement for mortality forecasts by considering $r=2$. 
Review of the Lee-Carter model and Fuzzy Set Theory

Fuzzy set and system

The description below is drawn largely from Koissi and Shapiro (2006).

**Definition 1:** A fuzzy subset $A$ (over a reference set $X$) is a function on $X$ that takes values in the unit-interval $[0,1]$

$$\mu_A : X \rightarrow [0,1]$$

**Definition 2:** (Zimmermann, 1996) Let $\tilde{A} = (a, l_a, r_a)$ be a triangular fuzzy number with center $a \in R$ and left and right spreads $(l_a, r_a)$. 
Review of the Lee-Carter model and Fuzzy Set Theory

Then its characteristic can be denoted by a membership function

$$\mu_A(x) = \begin{cases} 
\frac{x-a+l_a}{l_a}, & a-l_a < x \leq a \\
\frac{a+r_a-x}{r_a}, & a < x \leq a+r_a \\
0, & \text{otherwise}
\end{cases}$$

Figure 2.1 Triangular Fuzzy Number: $\tilde{A} = (a, l_a, r_a)$
3. Fuzzy formulation of the LC model with age-specific enhancement

A fuzzy formulation of the Lee-Carter model is

\[
\tilde{Y}_{x,t} = \tilde{A}_x \oplus \sum_{i=1}^{2} (\tilde{B}_x^{(i)} \otimes \tilde{K}_t^{(i)}), \quad x = x_1, \ldots, x_N, \quad t = t_1, t_1 + 1, \ldots, t_1 + T - 1
\] (3.1)

where \( \tilde{Y}_{x,t} \) are known fuzzy log-central death rates and \( \tilde{A}_x, \tilde{B}_x^{(i)}, \tilde{K}_t^{(i)} \) are unknowns. \( \tilde{K}_t^{(i)} \) represents the fuzzy formulation of the general mortality level.

\( \tilde{B}_x^{(i)} \) captures the decline in mortality at age \( x \).

\( \tilde{A}_x, \tilde{B}_x^{(i)}, \tilde{K}_t^{(i)} \) are symmetric triangular fuzzy numbers.
Fuzzy formulation of the LC model with age-specific enhancement

The fuzzy “addition” and “multiplication” reduce to

\[
\tilde{A}_x \oplus \tilde{B}_x^{(i)} = (a_x + b_x^{(i)}, \max(\alpha_x, \beta_x^{(i)}))
\]

\[
\tilde{A}_x \otimes \tilde{B}_x^{(i)} = (a_x b_x^{(i)}, \max(\alpha_x \mid b_x^{(i)} \mid, \beta_x^{(i)} \mid a_x \mid))
\]

Consequently (3.1) becomes

\[
\tilde{Y}_{x,t} = \tilde{A}_x \oplus \sum_{i=1}^{2} (\tilde{B}_x^{(i)} \otimes \tilde{K}_t^{(i)}) = (a_x + b_x^{(1)} k_t^{(1)}, \max(\alpha_x, \mid b_x^{(1)} \mid \delta_t^{(1)}, \beta_x^{(1)} \mid k_t^{(1)} \mid)) \oplus (\tilde{B}_x^{(2)} \otimes \tilde{K}_t^{(2)})
\]

\[
= (a_x + b_x^{(1)} k_t^{(1)} + b_x^{(2)} k_t^{(2)}, \max(\alpha_x, \mid b_x^{(1)} \mid \delta_t^{(1)}, \beta_x^{(1)} \mid k_t^{(1)} \mid, \mid b_x^{(2)} \mid \delta_t^{(2)}, \beta_x^{(2)} \mid k_t^{(2)} \mid)). \quad (3.3)
\]
Fuzzy formulation of the LC model with age-specific enhancement

This is the formulation of Fuzzy-LC with age-specific enhancement.

The log-central death rate for age-group $x$ in year $t$ is a symmetric triangular fuzzy number, instead of exactly as in the traditional non-fuzzy formulation.

Step 1: Fuzzification of the log-central death rate

We assume that the log-center death rate be captured by the symmetric triangular membership function

By introducing $\tilde{Y}_{x,t} = (y_{x,t}, e_{x,t})$, $\tilde{A}_1 = (c_{0x}, s_{0x})$, $\tilde{A}_1 = (c_{1x}, s_{1x})$, and $\tilde{A}_2 = (c_{2x}, s_{2x})$, we have

$$(y_{x,t}, e_{x,t}) = (c_{0x}, s_{0x}) + (c_{1x}, s_{1x}) \times t^{(1)} + (c_{2x}, s_{2x}) \times t^{(2)}.$$
Fuzzy formulation of the LC model with age-specific enhancement

The above formulation can further be simplified by noting \( t^{(1)} = t^{(2)} \)

So that

\[
(y_{x,t}, e_{x,t}) = (c_{0x}, s_{0x}) + [(c_{1x}, s_{1x}) + (c_{2x}, s_{2x})] \times t
\]

\[
= (c_{0x}, s_{0x}) + (c_{3x}, s_{3x}) \times t
\]

Where

\[
(c_{3x}, s_{3x}) = (c_{1x}, s_{1x}) + (c_{2x}, s_{2x})
\]

As argued in Koissi and Shapiro (2006), the centers \( c_{0x}, c_{3x} \) are easily found by fitting the ordinary least-squares regression to

\[
Y_{x,t} = c_{0x} + c_{3x} \times t
\]
Fuzzy formulation of the LC model with age-specific enhancement

The spreads $S_{0x}$, $S_{3x}$ are then obtained based on minimum fuzziness criterion

$$\hat{Y}_{x,t} = (\hat{c}_{0x}, \hat{s}_{0x}) + (\hat{c}_{3x}, \hat{s}_{3x}) \times t$$

It is required that the $\mu(\tilde{Y}_{x,t} \subseteq \hat{Y}_{x,t}) \geq h, h \in [0,1]$.

The level $h$, which measures the degree of fit of the estimated model to the given data, is a user input.
Fuzzy formulation of the LC model with age-specific enhancement

- Step 2: Estimating the parameters of fuzzy-LC with age-specific enhancement

- The objective of this step is to determine

  \[ \tilde{A}_x, \tilde{B}_x^{(i)}, \tilde{K}_t^{(i)}, i = 1, 2 \]

- We achieve this task by minimizing the square of the distance between

  \[ \tilde{A}_x \oplus \sum_{i=1}^{2} (\tilde{B}_x^{(i)} \otimes \tilde{K}_t^{(i)}) \]

And  \[ \tilde{Y}_{x,t} \]
Fuzzy formulation of the LC model with age-specific enhancement

Here we adopt Diamond (1988) distance measure

For two fuzzy sets $\tilde{A}_1 = (a_1, \alpha_1)$ and $\tilde{A}_2 = (a_2, \alpha_2)$

The distance is measured by

$$D_{LR}(\tilde{A}_1, \tilde{A}_2)^2 = (a_1 - a_2)^2 + [(a_1 - \alpha_1) - (a_2 - \alpha_2)]^2 + [(a_1 + \alpha_1) - (a_2 + \alpha_2)]^2$$

That is the fuzzy-LC parameters boils down to solving the following minimization problem:

Minimize: $$\sum_x \sum_t D_{LR}[\tilde{A}_x \oplus \sum_{i=1}^2 (\tilde{B}_x^{(i)} \otimes \tilde{K}_t^{(i)}), \tilde{Y}_{x,t}]^2$$
Fuzzy formulation of the LC model with age-specific enhancement

Where

\[ F = D_{LR} \left[ \tilde{A}_x \oplus \sum_{i=1}^{2} (\tilde{B}_x^{(i)} \otimes \tilde{K}_t^{(i)}), \tilde{Y}_{x,t} \right]^2 \]

\[ = D_{LR} \left[ (a_x + b_x^{(1)} k_t^{(1)} + b_x^{(2)} k_t^{(2)}), \max(\alpha_x, b_x^{(1)} k_t^{(1)}), \beta_x^{(1)} k_t^{(1)}), b_x^{(2)} \delta_t^{(2)} \beta_x^{(2)} k_t^{(2)} \right], (y_{x,t}, e_{x,t}) \right]^2 \]

\[ = (a_x + b_x^{(1)} k_t^{(1)} + b_x^{(2)} k_t^{(2)})^2 \]

\[ + [(a_x + b_x^{(1)} k_t^{(1)} + b_x^{(2)} k_t^{(2)})^2 - \max(\alpha_x, b_x^{(1)} k_t^{(1)}), \beta_x^{(1)} k_t^{(1)}), b_x^{(2)} \delta_t^{(2)} \beta_x^{(2)} k_t^{(2)}]) - (y_{x,t} - e_{x,t})]^2 \]

\[ + [(a_x + b_x^{(1)} k_t^{(1)} + b_x^{(2)} k_t^{(2)})^2 + \max(\alpha_x, b_x^{(1)} k_t^{(1)}), \beta_x^{(1)} k_t^{(1)}), b_x^{(2)} \delta_t^{(2)} \beta_x^{(2)} k_t^{(2)}]) - (y_{x,t} + e_{x,t})]^2 \]

Subject to:

\[ a_x = (1/T) \sum_{t} y_{x,t} \]

\[ \sum_{x} b_x^{(i)} = 1, \sum_{t=t_i}^{t_n} k_t^{(i)} = 0 \quad i = 1, 2 \]
Fuzzy formulation of the LC model with age-specific enhancement

MATLAB can be used to obtain the optimal parameters

\[ \alpha_x, b_x^{(1)}, \beta_x^{(1)}, k_t^{(1)}, \delta_t^{(1)}, b_x^{(2)}, \beta_x^{(2)}, k_t^{(2)} \text{ and } \delta_t^{(2)}. \]

Once the optimal parameters: \( b_x^{(1)}, k_t^{(1)}, b_x^{(2)}, \text{ and } k_t^{(2)} \) are estimated \( k_t^{(2)} \) is further adjusted \( k_t^{(2)'} \)

So that the actual total deaths and the total expected deaths for each \( t \) matches

---

\[ 2 \]
Fuzzy formulation of the LC model with age-specific enhancement

In other words, the parameter estimates satisfy

\[
\sum_{x=x_1}^{x_k} d_{xt} = \sum_{x=x_1}^{x_k} n_{xt} \exp(a_x + \sum_{i=1}^{2} b_i \hat{k}_i)
\]

Using ARIMA and get:

\[
k^{(1)}_t, k^{(2)}_t \quad t = t_0 + T, t_0 + T + 1, \ldots, t_0 + T + s
\]

Lastly, forecast the mortality through:

\[
\hat{y}_{xt} = a_x + b^{(1)}_x k^{(1)}_t + b^{(2)}_x k^{(2)}_t, \quad t = t_0 + T, t_0 + T + 1, \ldots, t_0 + T + s
\]

We can get Male and Female with age-specific enhancement.
Mortality forecasting: China data

Fig 3.1 Fuzzy least-squares regression of log-central death rates (Male, X=10, age group 27-29) with age-specific enhancements
Fig 3.2 Fuzzy least-squares regression of log-central death rates (Female, $X=10$, age group 27-29) with age-specific enhancement
Mortality forecasting: China data

- This paper uses 15 yearly observations of age-specific death rates for both males and females in China from 1994 to 2008, covering ages 0 to 89.
- These data are provided by the China Population Statistical Yearbooks and the China Statistical Yearbooks compiled by the National Bureau of Statistics of China.
- We implement the fuzzy formulation of Lee-Carter with \( r=1 \) and \( r=2 \) by partitioning the age into 30 groups consist of \([0,2], [3,5], [6,8], \ldots, [87,89]\).
- In other words, we have \( t=1, 2, \ldots, 15 \) and \( x=1, 2, \ldots, 30 \).
Mortality forecasting: China data

Estimating fuzzy-LC parameters with China death rates data

To estimate fuzzy-LC parameters using MATLAB

We need to define the formulation as follows

Minimize: \(\sum_x \sum_t D_{LR}[\tilde{A}_x \oplus \sum_{i=1}^{2} (\tilde{B}^{(i)}_x \otimes \tilde{K}^{(i)}_t), \tilde{Y}_{x,t}]^2\)

The calculation methods and results of \(a_x, b_x, k_t\) are same as the fuzzy formulation of Lee-Carter with \(r=1\)
Then we forecast $K_t$ from 2009 to 2050 with ARIMA. We select ARIMA $(0,2,1)$ for male and ARIMA $(1,2,0)$ for female using ADF test and R test. The results are given in fig 4.1 and fig 4.2.
Mortality forecasting: China data

Female $K_t$ with China death rates data
Mortality forecasting: China data

- We can calculate the spread of mortality. The calculation methods and results of the spread are same as the fuzzy formulation of Lee-Carter with $r=1$
- Then we can draw the mortality forecasting figure
- Fig 4.3 for Single factor male ($x=10$, age group 27-29) mortality and fig 4.4 for Single factor female ($x=10$, age group 27-29) mortality here
Mortality forecasting: China data

In fig.4.3 and fig.4.4, the middle curve is the forecasting mortality; The upper and lower curves are the upper limit and lower limit of forecasting mortality.
Mortality forecasting: China data

- Then we can calculate the life expectancy and compare them with the original data.
- We summarize the results of life expectancy:
  - In 2008, the expectancy of male is 79.07, the expectancy of female is 81.56.
  - In 2050, the expectancy of male is 85.89, the expectancy of female is 85.94.
- The expectancy of male increased 6.82, the expectancy of female increased 4.08, and the difference of male and female decreased to 0.05 from 2.79.
Mortality forecasting: China data

- We estimating parameters of the fuzzy-LC with age-specific enhancement directly using MATLAB

Inset Table 4.6 for $b_x^{(1)}$, $b_x^{(2)}$ with age-specific enhancement here

Keep $k_r^{(1)}$ stable, justify $k_r^{(2)}$ based on (3.11), get $k_r^{(2)'}$

Then we forecast the $k_r^{(1)}$ and $k_r^{(2)'}$ to 2050 with ARIMA.

(1) $k_r^{(1)}$: We select ARIMA $(0,2,1)$ for male and ARIMA $(0,1,1)$ for female using ADF test and R test.
Mortality forecasting: China data

(2) $k_t^{(2)}$: We select ARIMA $(0,2,1)$ for male and ARIMA $(1,2,0)$ for female using ADF test and R test.

The results are as following:
Table 4.7 $k_t^{(1)}$, $k_t^{(2)}$ and justified $k_t^{(2)\prime}$ with age-specific enhancement

<table>
<thead>
<tr>
<th>T</th>
<th>Male</th>
<th></th>
<th></th>
<th>Female</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$k_t^{(1)}$</td>
<td>$k_t^{(2)}$</td>
<td>$k_t^{(2)\prime}$</td>
<td>$k_t^{(1)}$</td>
<td>$k_t^{(2)}$</td>
</tr>
<tr>
<td>5</td>
<td>6.1871</td>
<td>6.0817</td>
<td>5.4046</td>
<td>3.8698</td>
<td>4.2526</td>
</tr>
<tr>
<td>6</td>
<td>5.5871</td>
<td>5.5057</td>
<td>6.0159</td>
<td>3.6555</td>
<td>3.3305</td>
</tr>
<tr>
<td>7</td>
<td>2.5932</td>
<td>3.0934</td>
<td>3.2279</td>
<td>4.7060</td>
<td>4.0327</td>
</tr>
<tr>
<td>8</td>
<td>-0.7181</td>
<td>-0.4087</td>
<td>0.6446</td>
<td>1.0441</td>
<td>1.3139</td>
</tr>
<tr>
<td>9</td>
<td>-1.0705</td>
<td>-1.7893</td>
<td>-1.0068</td>
<td>-4.2277</td>
<td>-3.2679</td>
</tr>
<tr>
<td>10</td>
<td>-3.3719</td>
<td>-3.3982</td>
<td>-1.6451</td>
<td>-2.0950</td>
<td>-0.4707</td>
</tr>
<tr>
<td>11</td>
<td>-4.3293</td>
<td>-5.2782</td>
<td>-4.4506</td>
<td>-5.4636</td>
<td>-6.2049</td>
</tr>
<tr>
<td>12</td>
<td>-4.7509</td>
<td>-5.1788</td>
<td>-4.0797</td>
<td>-4.9691</td>
<td>-4.5718</td>
</tr>
<tr>
<td>13</td>
<td>-11.7264</td>
<td>-11.2937</td>
<td>-10.7726</td>
<td>-17.2602</td>
<td>-16.8087</td>
</tr>
</tbody>
</table>
Conclusions: China data

- Lastly, forecast the mortality from 2009 to 2050 with
  \[ \hat{y}_{xt} = a_x + b_x^{(1)}k_{t}^{(1)} + b_x^{(2)}k_{t}^{(2)}, \quad t = t_0 + T, t_0 + T + 1, \ldots, t_0 + T + s \]

- Then we can calculate the life expectancy and compare them with the original data. We summarize the results of life expectancy:
  - (1) In 2008, the expectancy of male is 79.36, the expectancy of female is 82.10.
  - (2) In 2050, the expectancy of male is 85.92, the expectancy of female is 86.86.
  - (3) The expectancy of male increased 6.56, the expectancy of female increased 4.76, and the difference of male and female decreased to 0.94 from 2.74.
Conclusions: China data

- We compared the fuzzy-LC with age-specific enhancement with single factor fuzzy-LC and found
  - (1) The fuzzy-LC’s sum error square with age-specific enhancement is smaller than the single factor fuzzy-LC’s.
  - (2) In both the fuzzy-LC’s and fuzzy-LC’s with age-specific enhancement, the expectancy of male and female increased from 2008 to 2050, and the difference of expectancy of male and female decreased from 2008 to 2050.
Conclusions: China data

- We can calculate the spread of mortality based on the formula of spread:
  
  \[ e_{x,t} = s_{0x} + s_{1x} \times t \quad t=1,2,3,\ldots,57 \]

Fig. 4.13 Male (x=10, age group 27-29) mortality curve with age-specific enhancement.
Conclusions: China data

Fig.4.14 Female (x=10, age group 27-29) mortality curve with age-specific enhancement
Fig. 4.12
Female life expectancy with age-specific enhancement
Fig. 4.11 Male life expectancy with age-specific enhancement
Questions ?

Thanks !