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

Review of the Lee-Carter model and Fuzzy Set Theory

The basic Lee-Carter (LC) model

$$\ln m_{xt} = a_x + b_x k_t + \varepsilon_{xt}, \quad x = x_1, \ldots, x_N \text{ and } t = t_1, t_1 + 1, \ldots, t_1 + T - 1 \qquad (2.1)$$

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

$$\log m_{xt} = a_x + b_x^{(1)} k_t^{(1)} + b_x^{(2)} k_t^{(2)} + \cdots + b_x^{(r)} k_t^{(r)} + \varepsilon_{xt} \qquad (2.2)$$

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

$$\mu_{\tilde{A}}(x)=\begin{cases}\dfrac{x-a+l_a}{l_a}, a-l_a<x\le a\\ \dfrac{a+r_a-x}{r_a}, a<x\le a+r_a\\ 0, \qquad otherwise\end{cases}$$

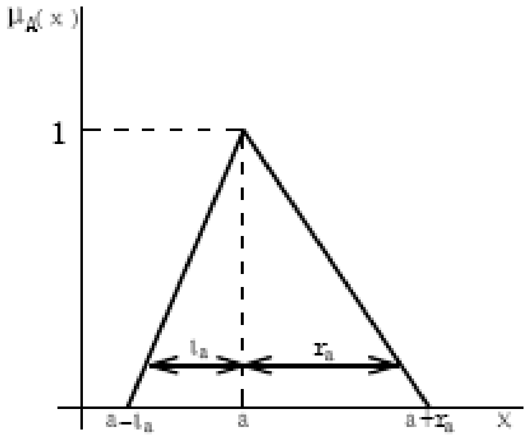


Figure 2.1 Triangular Fuzzy Number: $\tilde{A}=(a,l_a,r_a)$

- 3.Fuzzy formulation of the LC model with age-specific enhancement
- A fuzzy formulation of the Lee-Carter model is

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

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

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

$$\tilde{A}_x \otimes \tilde{B}_x^{(i)} = (a_x b_x^{(i)}, \max(\alpha_x \left|b_x^{(i)}\right|, \beta_x^{(i)} \left|a_x\right|))$$

Consequently (3.1) becomes

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

$$= (a_x + b_x^{(1)} k_t^{(1)} + b_x^{(2)} k_t^{(2)}, \max(\alpha_x, \left|b_x^{(1)}\right| \delta_t^{(1)}, \beta_x^{(1)} \left|k_t^{(1)}\right|, \left|b_x^{(2)}\right| \delta_t^{(2)}, \beta_x^{(2)} \left|k_t^{(2)}\right|)). \qquad (3.3)$$

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

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

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

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

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

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

$$= (a_x + b_x^{(1)}k_t^{(1)} + b_x^{(2)}k_t^{(2)} - y_{x,t})^2$$

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

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

Subject to:

$$\begin{cases} a_x = (1/T)\sum_t y_{x,t} \\ \sum_x b_x^{(i)} = 1, \sum_{t=t_1}^{t_n} k_t^{(i)} = 0 \quad i = 1, 2 \end{cases}$$

Fuzzy formulation of the LC model with age-specific enhancement

MATLAB can be used to obtain the optimal parameters

α_x , $b_x^{(1)}$, $\beta_x^{(1)}$, $k_t^{(1)}$, $\delta_t^{(1)}$, $b_x^{(2)}$, $\beta_x^{(2)}$, $k_t^{(2)}$ and $\delta_t^{(2)}$.

Once the optimal parameters: $b_x^{(1)}$, $k_t^{(1)}$, $b_x^{(2)}$, and $k_t^{(2)}$ are estimated

$k_t^{(2)}$ is further adjusted $k_t^{(2)\prime}$

So that the actual total deaths and the total expected deaths for each *t* matches

Fuzzy formulation of the LC model with age-specific enhancement

In other words, the parameter estimates satisfy

$$\sum_{x=x_1}^{x_k} d_{xt} = \sum_{x=x_1}^{x_k} n_{xt} \exp(\hat{a}_x + \sum_{i=1}^{2} \hat{b}_x^i \hat{k}_t^i)$$

Using ARIMA and get: $k_t^{(1)}, k_t^{(2)} \quad t = t_0 + T, t_0 + T + 1, \cdots, t_0 + T + s$

Lastly, forecast the mortality through:

$$\hat{y}_{xt} = a_x + b_x^{(1)} k_t^{(1)} + b_x^{(2)} k_t^{(2)}, \quad t = t_0 + T, t_0 + T + 1, \cdots, t_0 + T + s$$

We can get Male and Female $c_{0x}, c_{1x}, s_{0x}, s_{1x}$

with age-specific enhancement

Mortality forecasting: China data

- Mortality forecasting: China data

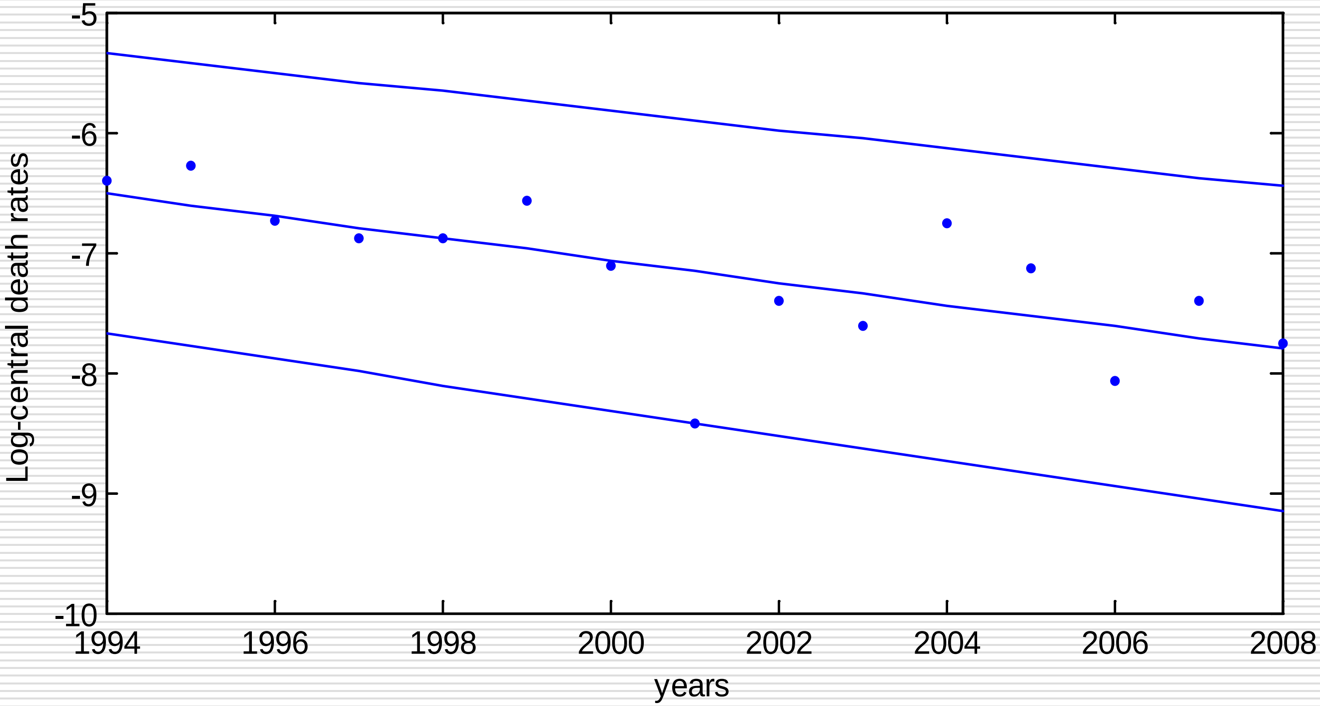


Fig 3.1 Fuzzy least-squares regression of log-central death rates (Male, X=10, age group 27-29) **with age-specific enhancemen**

Mortality forecasting: China data

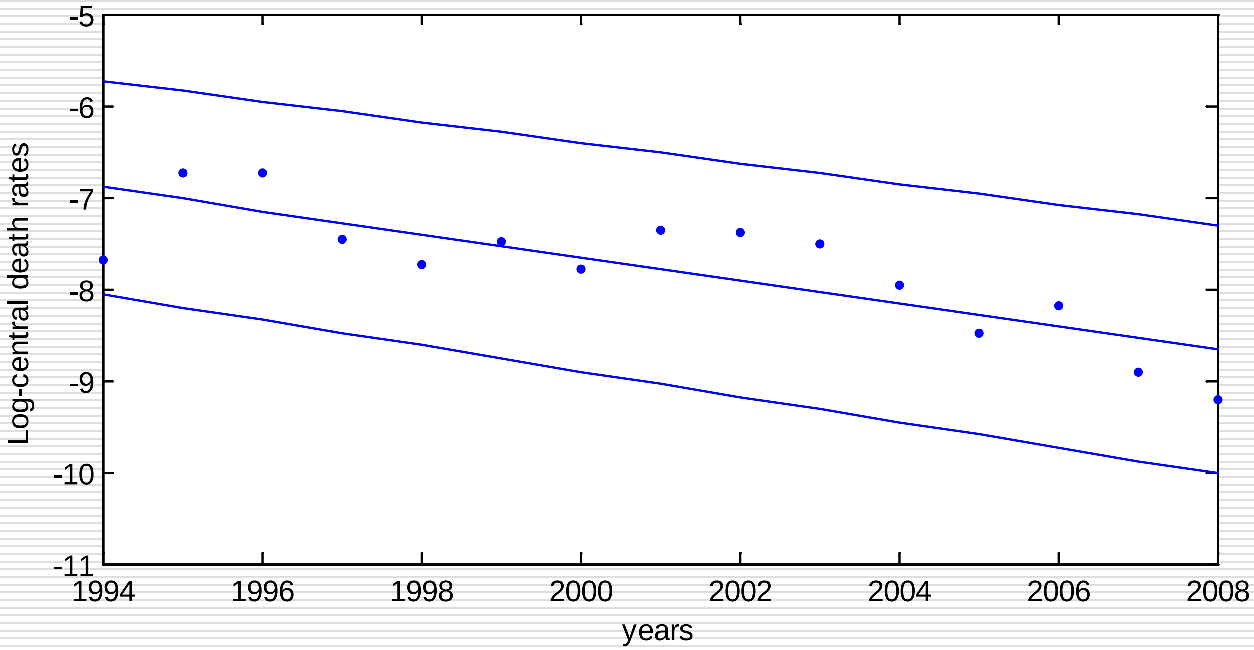


Fig 3.2 Fuzzy least-squares regression of log-central death rates (Female, X=10, age group 27-29) **with age-specific enhancement**

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], …, [87,89].
- In other words, we have t=1, 2, …,15 and x=1, 2, …30.

Mortality forecasting: China data

Estimating fuzzy-LC parameters with China death rates data

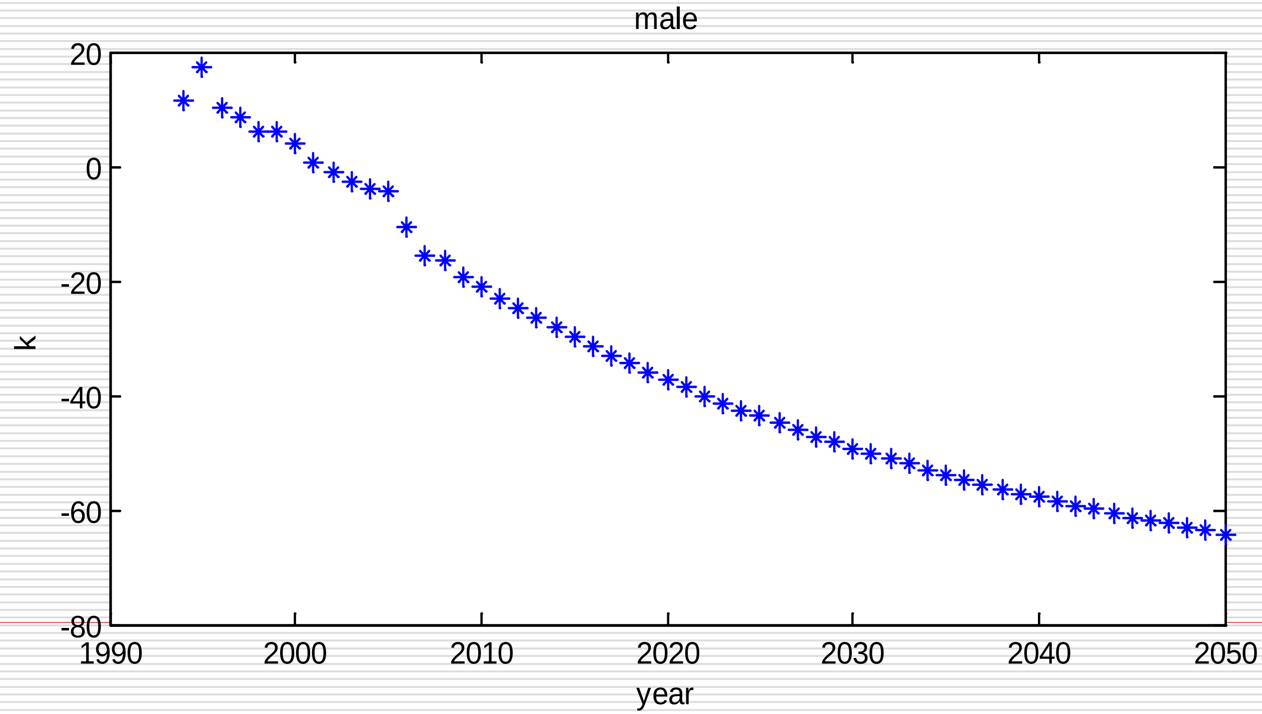
To estimate fuzzy-LC parameters using MATLAB

We need to define the formulation as follows

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

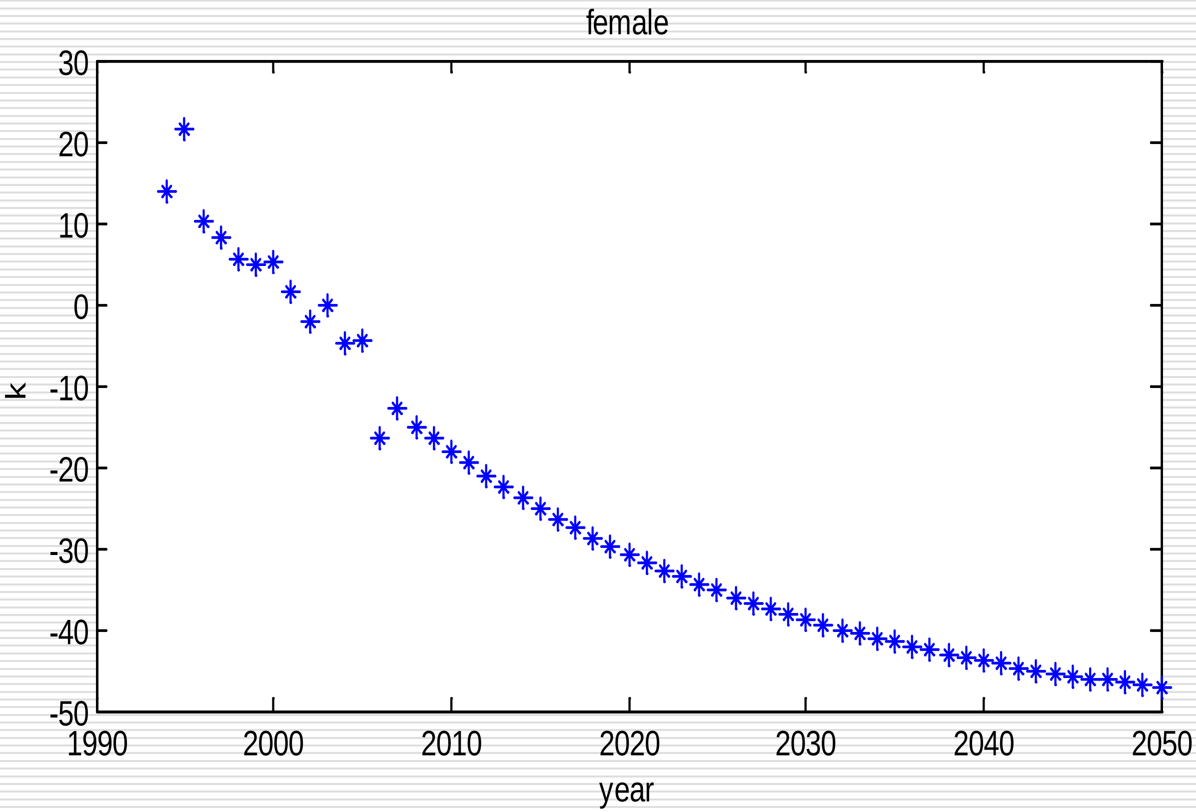
The calculation methods and results of a_x , b_x , k_t are same as the fuzzy formulation of Lee-Carter with r=1

- 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.



Male K_t with China death rates data

Mortality forecasting: China data



Female K_t with China death rates data

Mortality forecasting: China data

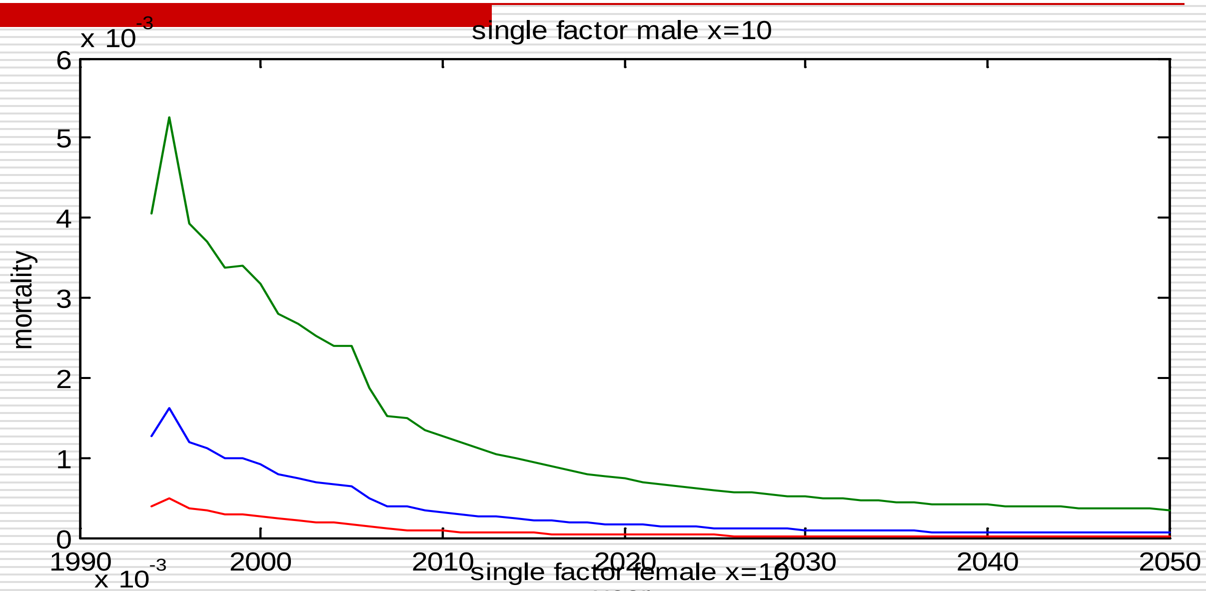
- We can calculate the spread of mortality. The calculation methods and results of the spread are same as the fuzzy formulation of Lee-Carter with r=1
- Then we can draw the mortality forecasting figure
- Fig 4.3 for Single factor male（x=10,age group 27-29）mortality and fig 4.4 for Single factor female（x=10,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

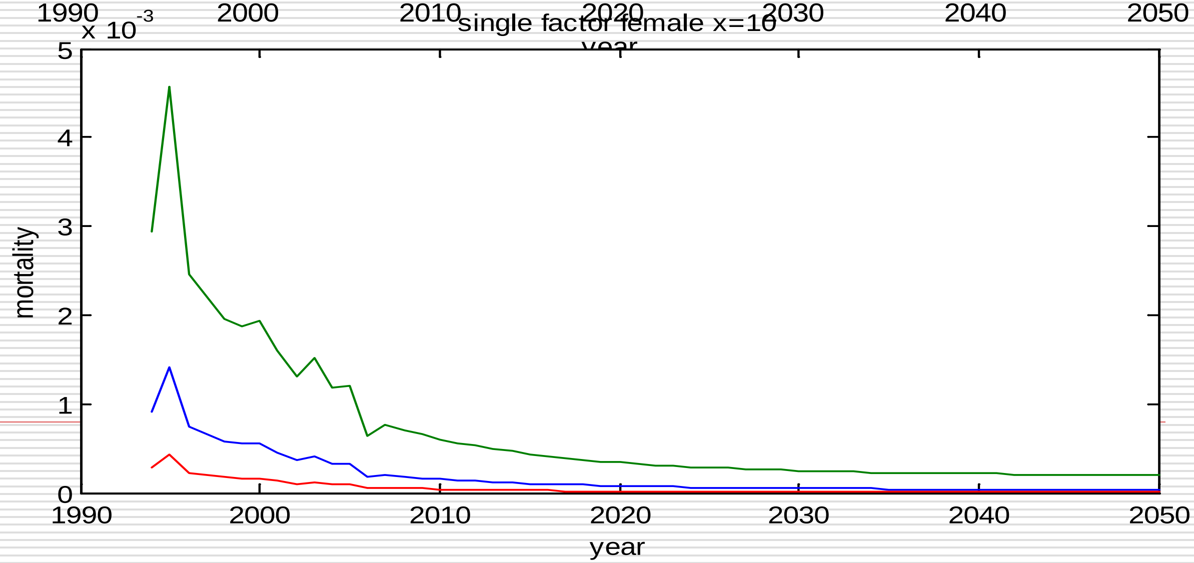
  mortality here

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

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

(1) $k_t^{(1)}$: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 $k_t^{(2)'}$ with age-specific enhancement

T	Male			Female		
	$k_t^{(1)}$	$k_t^{(2)}$	$k_t^{(2)'}$	$k_t^{(1)}$	$k_t^{(2)}$	$k_t^{(2)'}$
1	10.4040	10.2330	11.8243	12.9460	12.1762	13.5574
2	17.5866	16.8844	16.9228	20.2478	20.1532	22.0503
3	9.1799	8.9434	9.4664	9.8251	8.5309	10.0529
4	8.9691	8.7685	8.4060	6.3601	6.7724	7.7928
5	6.1871	6.0817	5.4046	3.8698	4.2526	5.4819
6	5.5871	5.5057	6.0159	3.6555	3.3305	4.5185
7	2.5932	3.0934	3.2279	4.7060	4.0327	4.6582
8	-0.7181	-0.4087	0.6446	1.0441	1.3139	1.0722
9	-1.0705	-1.7893	-1.0068	-4.2277	-3.2679	-1.3828
10	-3.3719	-3.3982	-1.6451	-2.0950	-0.4707	-0.7302
11	-4.3293	-5.2782	-4.4506	-5.4636	-6.2049	-5.1482
12	-4.7509	-5.1788	-4.0797	-4.9691	-4.5718	-3.6425
13	-11.7264	-11.2937	-10.7726	-17.2602	-16.8087	-17.2865
14	-17.0810	-15.6848	-15.8321	-13.2665	-14.1426	-12.3293
15	-17.4589	-16.4782	-17.1257	-15.3723	-15.0958	-16.1282

Conclusions: China data

- Lastly, forecast the mortality from 2009 to 2050 with

$$\hat{y}_{xt} = a_x + b_x^{(1)} k_t^{(1)} + b_x^{(2)} k_t^{(2)}, \quad t = t_0 + T, t_0 + T + 1, \cdots, t_0 + T + s$$

- 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

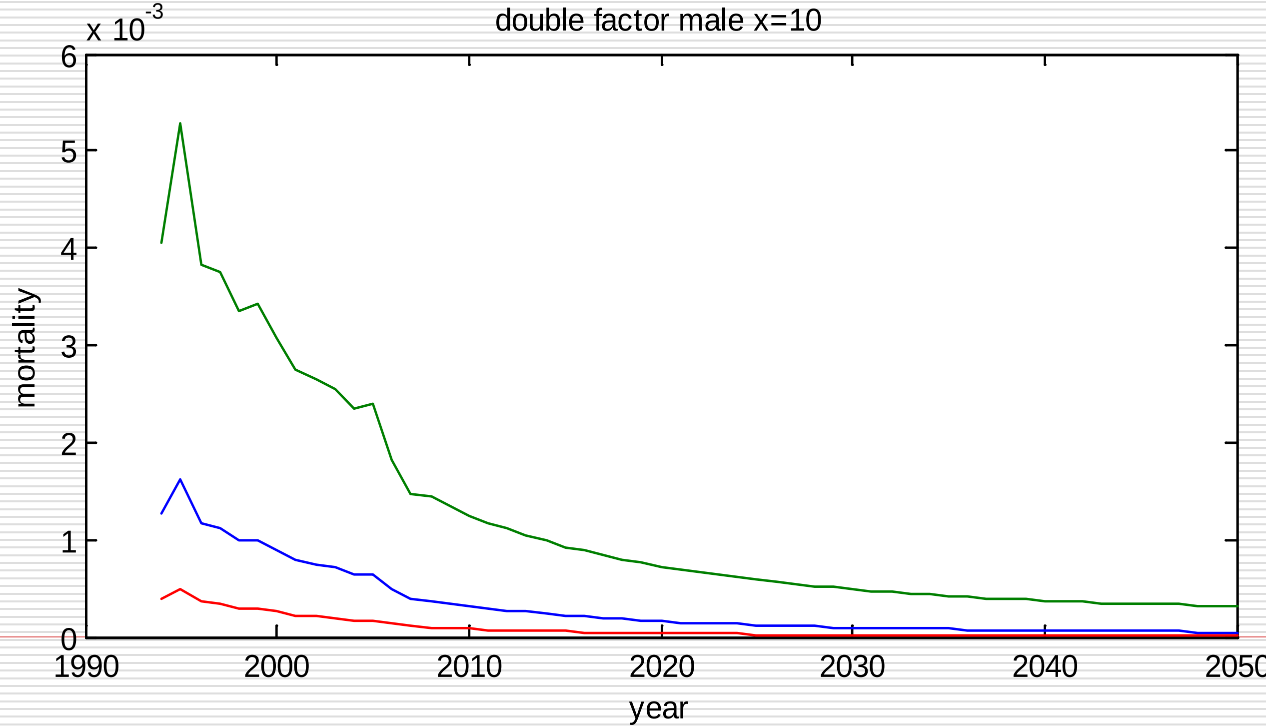
- We compared the fuzzy-LC with age-specific enhancement with single factor fuzzy-LC and found
  - (1) The fuzzy-LC's sum error square with age-specific enhancement is smaller than the single factor fuzzy-LC's.
  - (2) In both the fuzzy-LC's and fuzzy-LC's with age-specific enhancement, the expectancy 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:

$$e_{x,t} = s_{0x} + s_{1x} \times t \quad t=1,2,3,......57$$

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

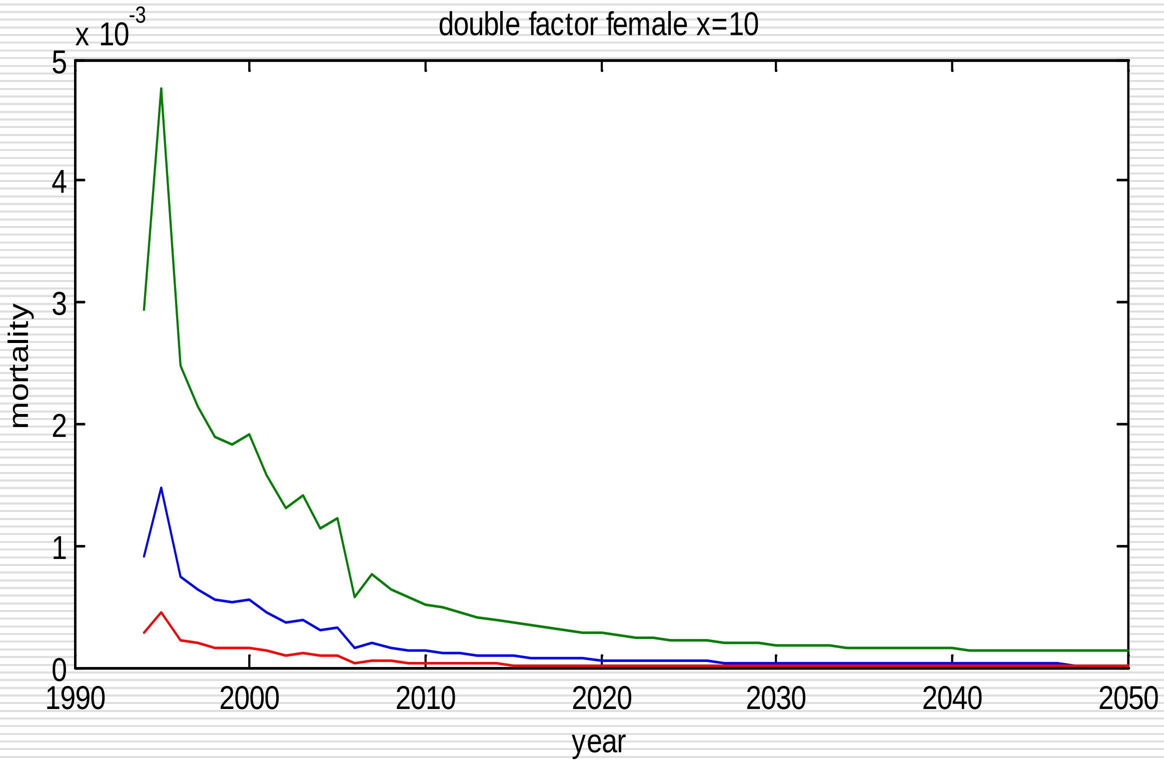


Fig.4.12 Female life expectancy with age-specific enhancement

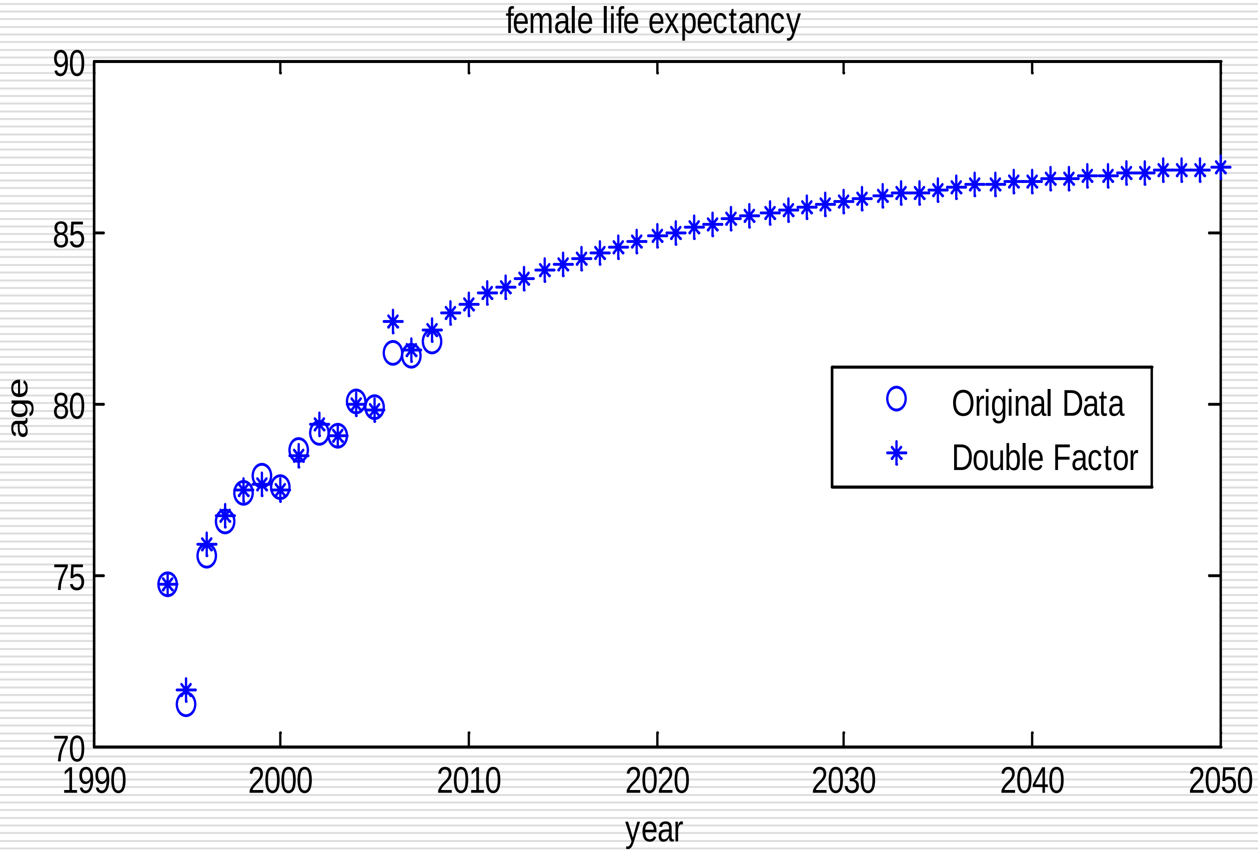
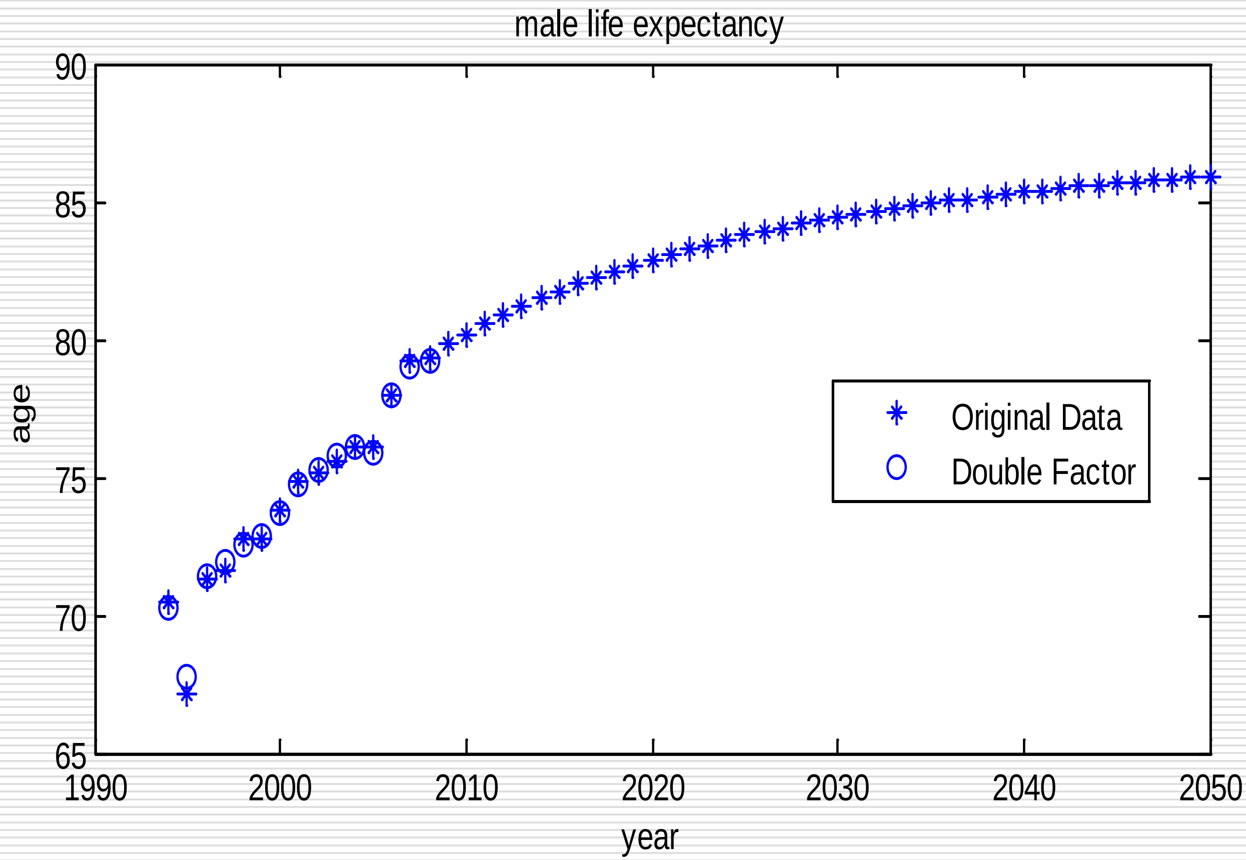


Fig.4.11 Male life expectancy with age-specific enhancement



Questions？

Thanks！