Forecasting Mortality in the Presence of Missing Data: An Application to Chinese Population

Joint work with
Ken Seng Tan and Johnny Siu-Hang Li,
University of Waterloo

Ping An
People's Insurance Company of China Property & Casualty

Ninth International Longevity Risk and Capital Markets Solutions Conference
Sep.6th-7th, 2013

Outline

- Introduction
- Multiple Imputation
- Graduation of Mortality
- The Lee-Carter Model
- Application
- Conclusions

Background

- Longevity risk in China

  In China, life expectancy at birth has increased from 46 to 71 from year 1950 to 2000.

- Data requirements for stochastic mortality model

- Missing data problem in China

1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001…2008
○	?	?	○	?	?	?	?	○	○	○	○	○	○	?	○

? denotes data is missing

○ denotes data is available

Objective

- Investigate how we can apply a stochastic mortality model in the presence of missing data.

- Multiple Imputation

- Graduation of Mortality

- Lee-Carter model

Existing Methods

- Zhu and Chen (2007)

  Complete-Case Analysis

- Li, Lee & Tuljapurkar (2004)

  Assumption: k(t) is always appropriately modeled by a random walk with drift.

$$k(t) = k(t-1) + c + e(t)$$

Sources of Data

- **Chinese mortality data**
- China Population Statistics Yearbook, 1987-2009.
  Age 15-84
  For year 1995, 2005,   1% Population Sample Survey
  For the rest of years,   Population Changes Sample Survey

  China Insurance Life Table (2000-2003)

- **Other Asian countries or regions mortality data**
- Japan, Taiwan   Human Mortality Database
- Hong Kong   Census and Statistics Department of the government of the Hong Kong Special Administrative Region of the P.R.C

Multiple Imputation

- Missing Pattern:

  Missing Completely at Random (MCAR)

  Let us denote the complete data as Y_{com} and partition it as $Y_{com} = (Y_{obs}, Y_{mis})$, where Y_{obs} and Y_{mis} are the observed and missing parts, respectively.

- Rubin (1976) defined missing data to be MCAR if the distribution of missingness R does not depend on Y_{com}, $P(R \mid Y_{com}) = P(R)$.

  where R is referred as an indicator variable indicating whether Y is observed (R=1) or missing (R=0).

Multiple Imputation

- Missing data procedures dealing with MCAR

- Complete-Case Analysis
- Single Imputation
- Multiple Imputation
- Maximum Likelihood

Multiple Imputation

- Regression Method

- We denote vectors Y_1 , Y_2, ..., Y_{70}, the female/male log central mortality rates at non-missing years over the period 1986-2008 for age 15 to 84, respectively. And X is a vector of non-missing calendar years from 1986 to 2008. Make the linear regressions of Y_1 , Y_2, ..., Y_{70} against X individually,

$$Y_i X \quad {}_{01} \qquad , \qquad i \quad 1,2,\ldots 70 \qquad .$$

T-statistics results of linear regressions for Mainland China & Taiwan female log central mortality rate against calendar year

age	Mainland China t-stats	Taiwan t-stats	age	Mainland China t-stats	Taiwan t-stats
15	-2.8995	-4.3030	60	-5.8259	-27.3087
16	-3.3134	-6.4304	61	-4.6576	-11.8406
…	…	…	62	-4.9951	-20.5259
34	-3.5943	-5.5658	…	…	…
35	-2.4669	-8.5851	82	-4.7460	-16.8589
36	-2.9518	-6.7239	83	-2.8922	-16.5319
…	…	…	84	-3.4278	-14.1860

Critical values: $t_{0.05}(14)=\pm2.1448$; $t_{0.1}(14)=\pm1.7613$

Multiple Imputation

- Algorithm
  - Firstly, fit the model $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, $i = 1,2,\ldots 70$ with observed data. Get the regression parameter estimates $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)$ and the associated covariance matrix $\hat{\sigma}^2 (X'X)^{-1}$.
  - Secondly, new parameters $\hat{\beta}_*$, σ^2_* are drawn from the posterior predictive distribution of the parameters where $\beta_* = \hat{\beta} + \sigma_* V_h' Z$, $\sigma^2_* = \hat{\sigma}^2 (n-2)/g$

    V_h' is the upper triangular matrix in the Cholesky decomposition.

    Z is a vector of two independent random standard variates.

    n is the number of nonmissing observations for Y and

    g is a χ^2_{n-2} random variate.

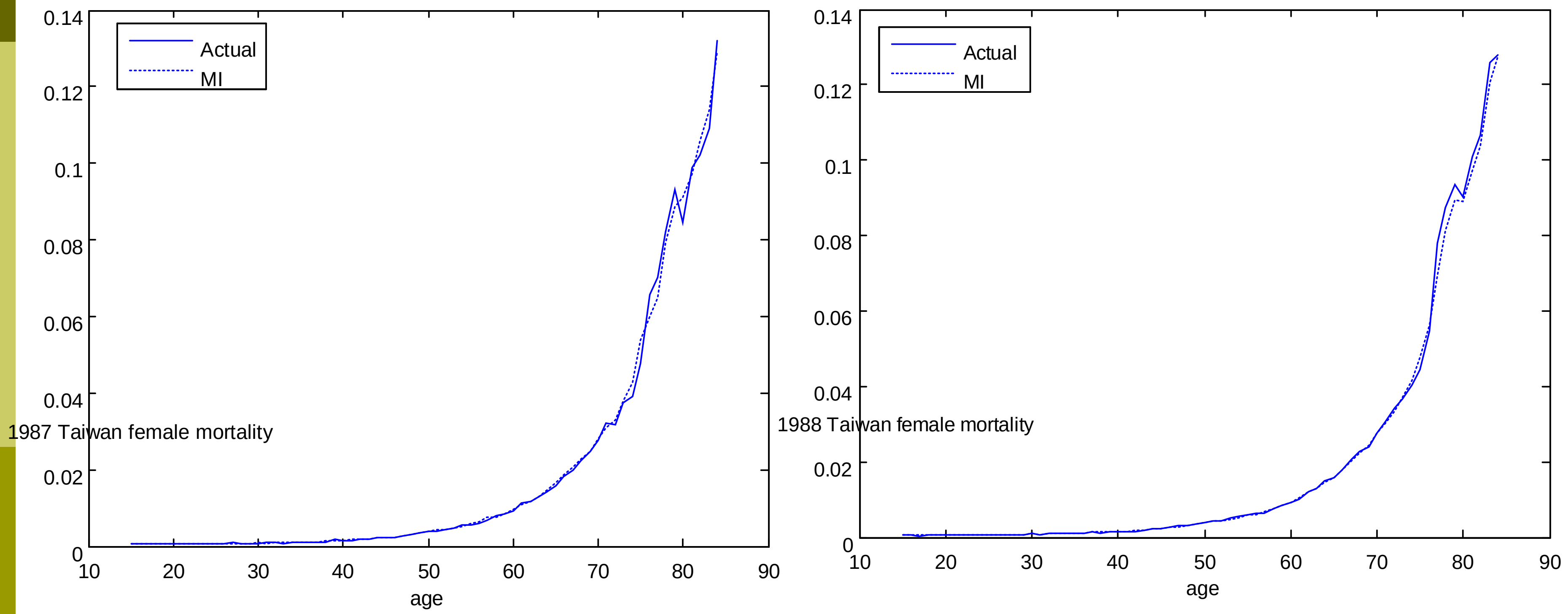
Multiple Imputation

- Thirdly, the multiple imputed values are then calculated by $\;^{*0*1*} \quad Xz$

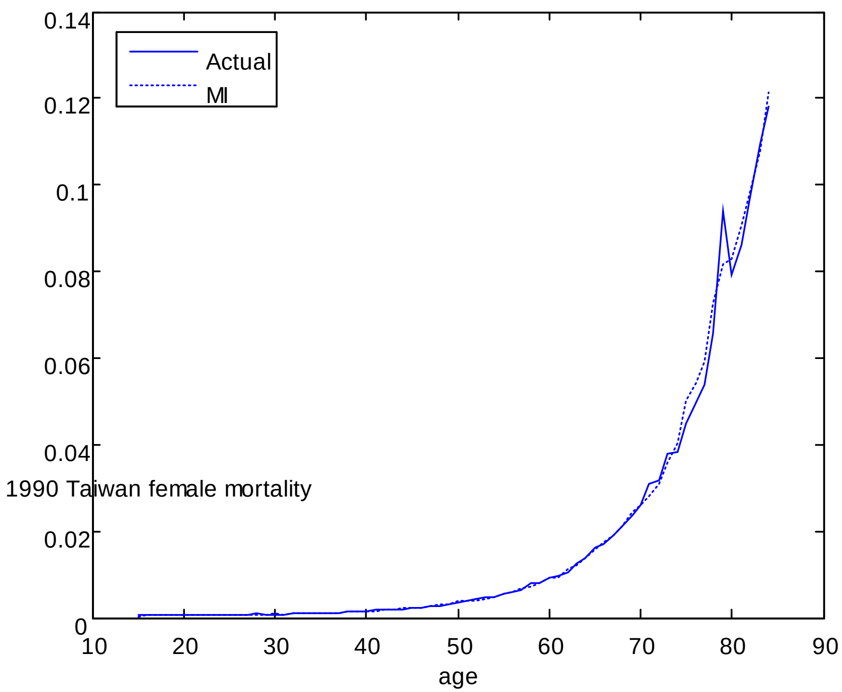
  where z is a standard normal random variable.

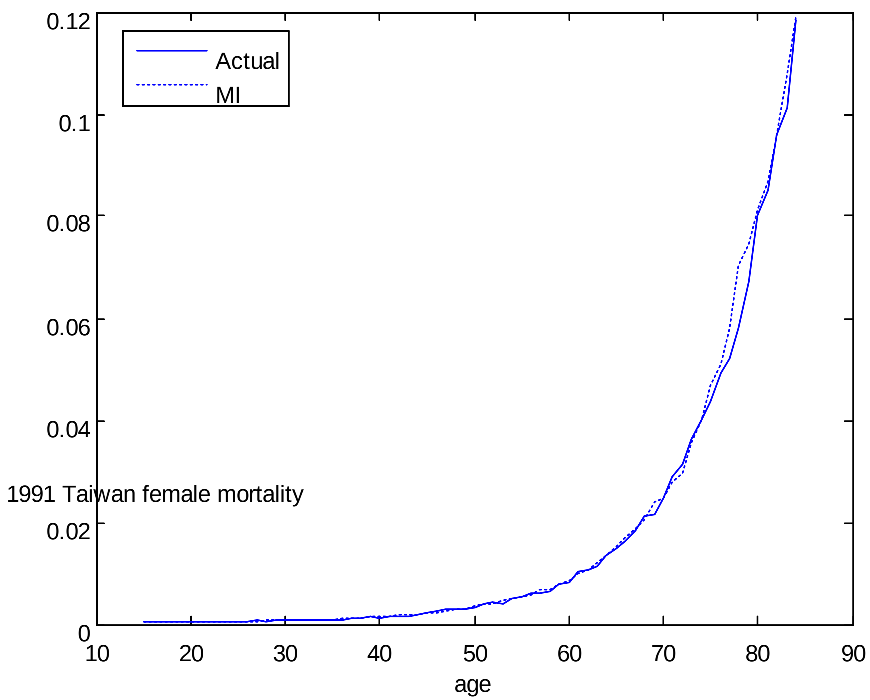
- Finally, replace the missing value with the average of the multiple imputed values.

Comparisons of actual and MI Taiwan female mortality rate for year 1987, 1988

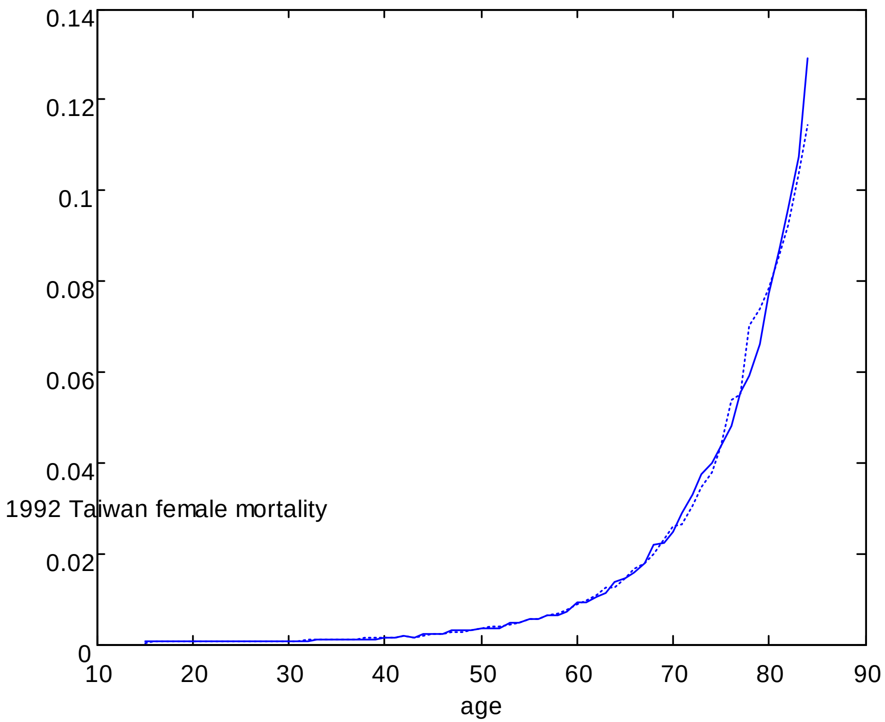


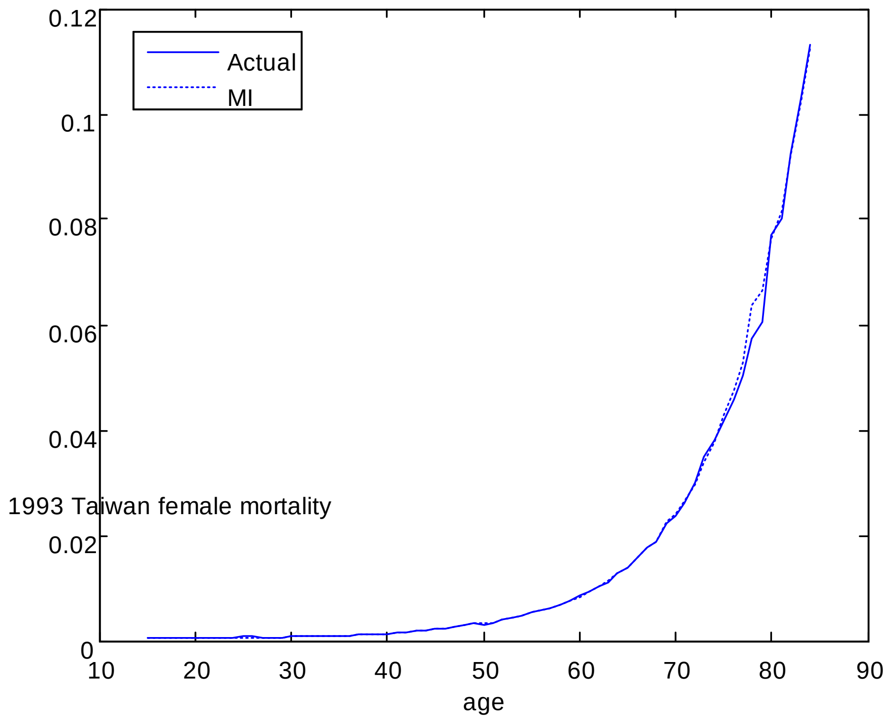
Comparisons of actual and MI Taiwan female mortality rate for year 1990, 1991



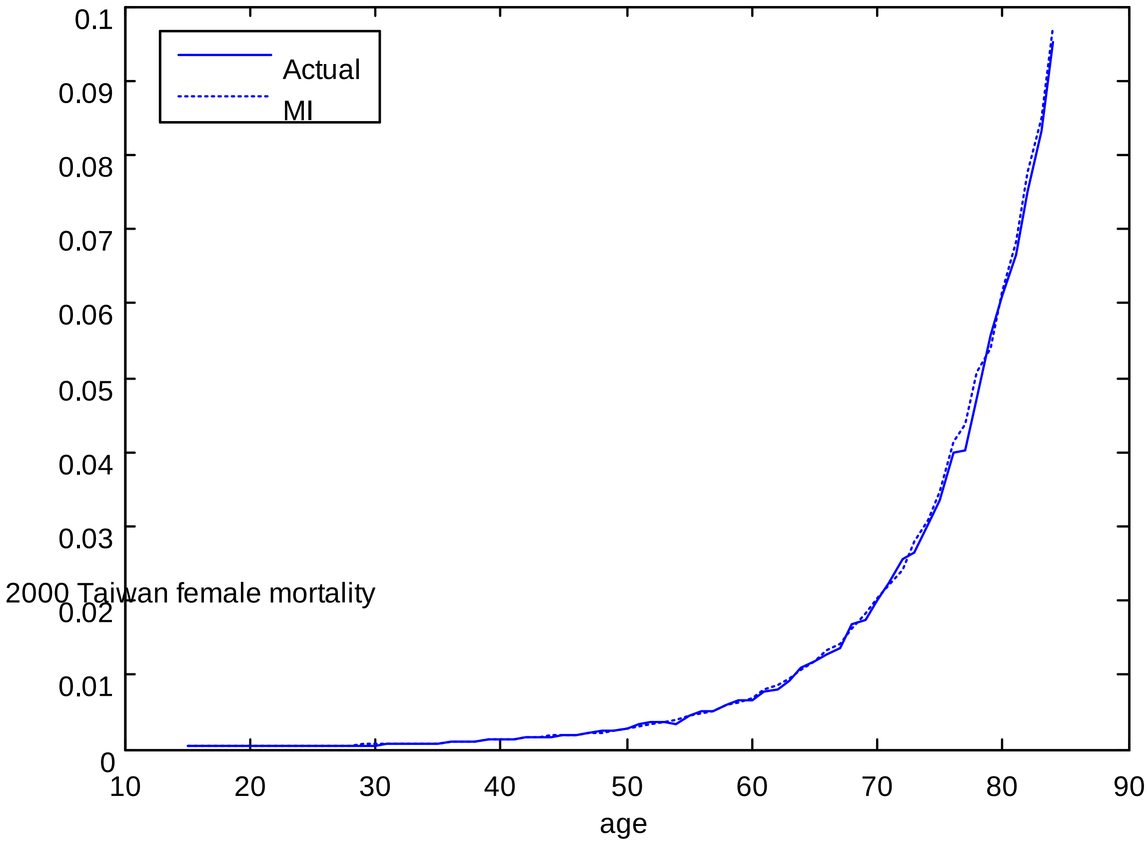


Comparisons of actual and MI Taiwan female mortality rate for year 1992, 1993





Comparisons of actual and MI Taiwan female mortality rate for year 2000



MAPE measure

- Mean absolute percentage error (MAPE)

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \hat{Y}_i}{Y_i}\right| * 100\%$$

- where Y_i and $\hat{Y}_i$ are the observed and estimated values, and n is the number of observations.

Table 3 MAPE of Taiwan mortality data derived by Multiple Imputation

Year	1987	1988	1990	1991	1992	1993	2000
MAPE	6.49%	6.52%	5.59%	5.50%	6.12%	5.34%	5.39%

Graduation of Mortality

- **Cubic smoothing spline**

  $x_1, x_2, \ldots, x_n$ are points satisfying $x_1 < x_2 < \cdots < x_n$, $y_1, y_2, \ldots, y_n$ are observations at points $x_1, x_2, \ldots, x_n$.

  $f(x)$ is a smooth and continuous twice differentiable curve that minimizes the penalized sum of squares

  $$p\sum_{j=1}^{n} w(j)\left|y(j) - f(x(j))\right|^2 + (1-p)\int \left|D^2 f(t)\right|^2 dt$$

  where p is a positive smoothing parameter.

- Choosing the smoothing parameter p

Graduation of Mortality

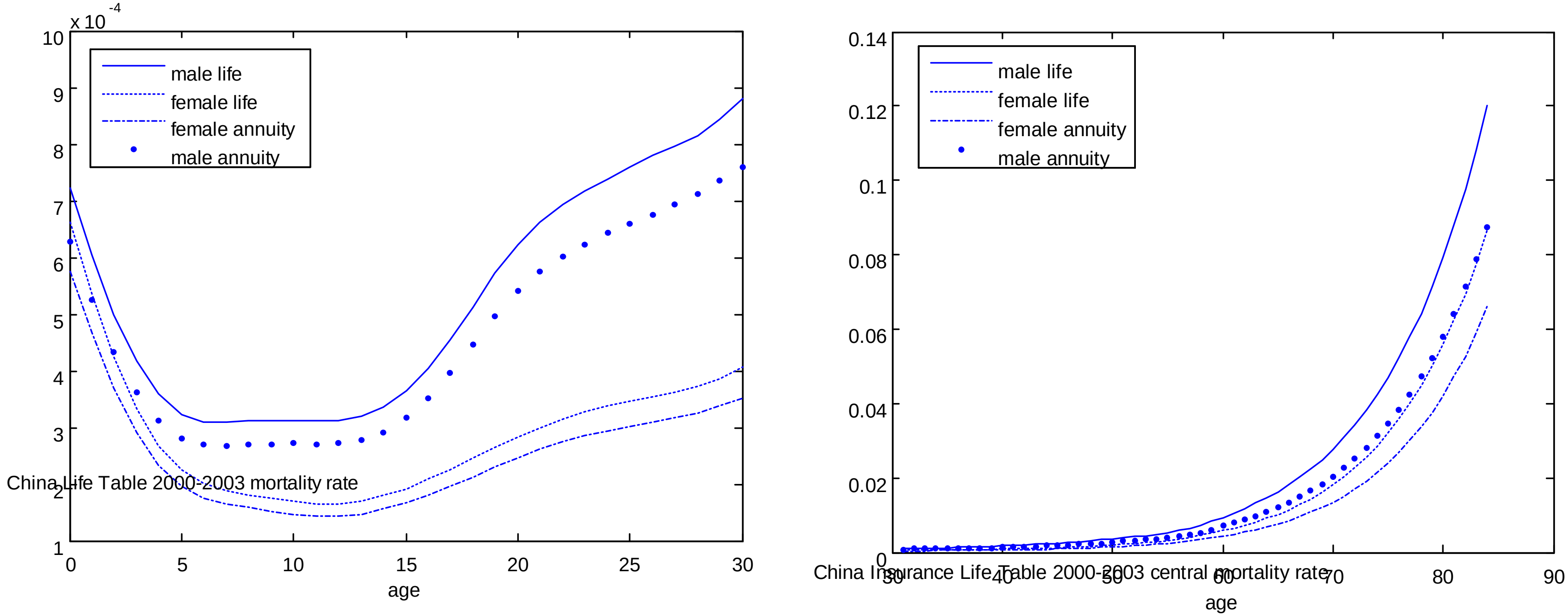


Figure 2 China Insurance Life Table (2000-2003) central mortality rate

The Lee-Carter Model

$$\ln(m(x,t))\ a(x)\ b(x)k(t)\ (x,t)$$

$m(x,t)$ is central mortality rate at age x, year t

$a(x)$ is the baseline age schedule of mortality

$b(x)$ is an age-specific component that represents how rapidly or slowly mortality at each age varies when the general level of mortality changes

$k(t)$ is the mortality improvement along with the year

(x,t) is the error term, a normal distribution with mean 0 and variance .

The Lee-Carter Model

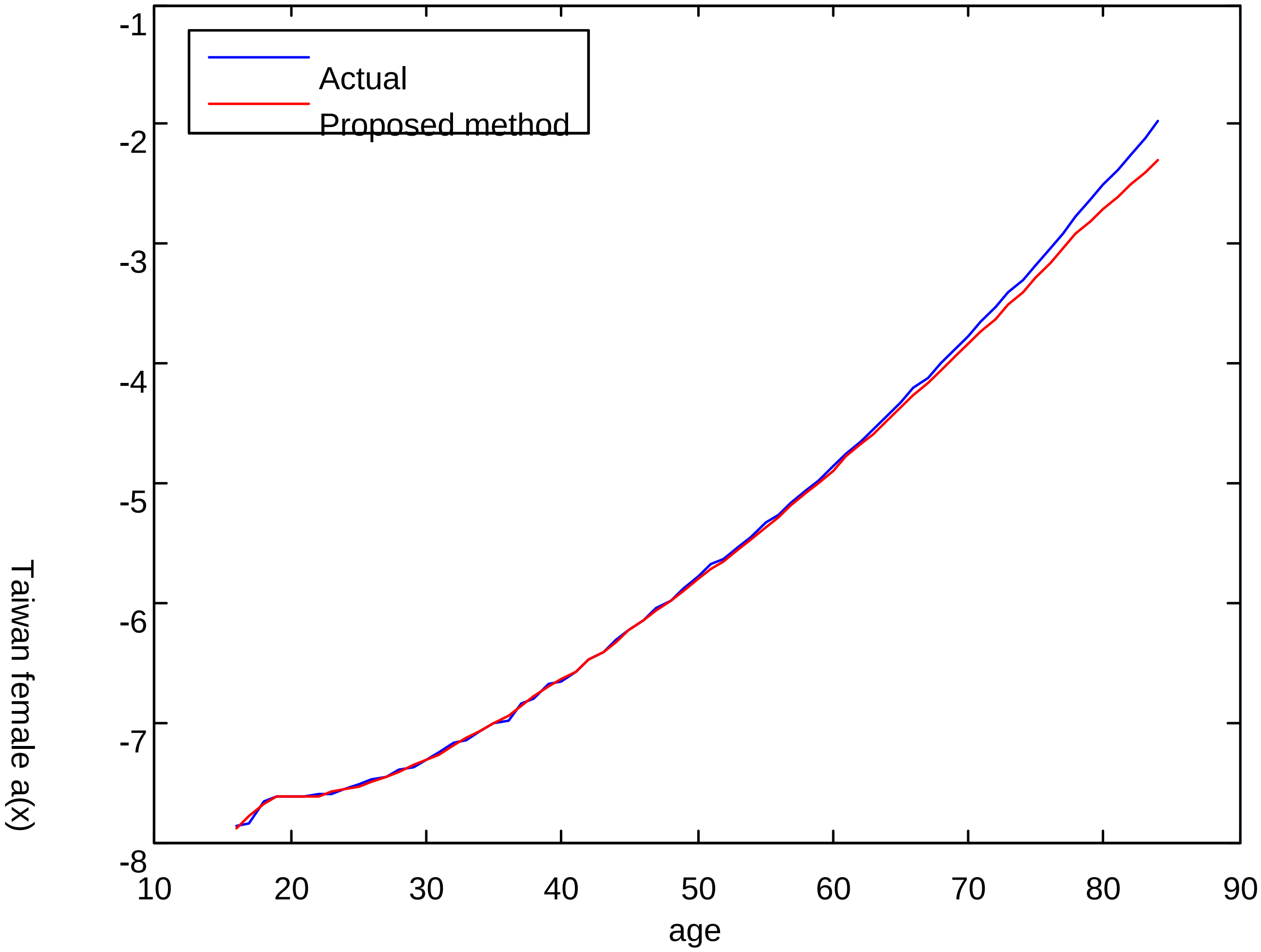
- The parameters are derived by Maximum Likelihood Estimation (MLE) assuming that death number subjects to Poisson Distribution.
- The resulting estimated $k(t)$'s are then modeled and projected as a stochastic time series using standard Box-Jenkins methods.

- Mortality rate projection:

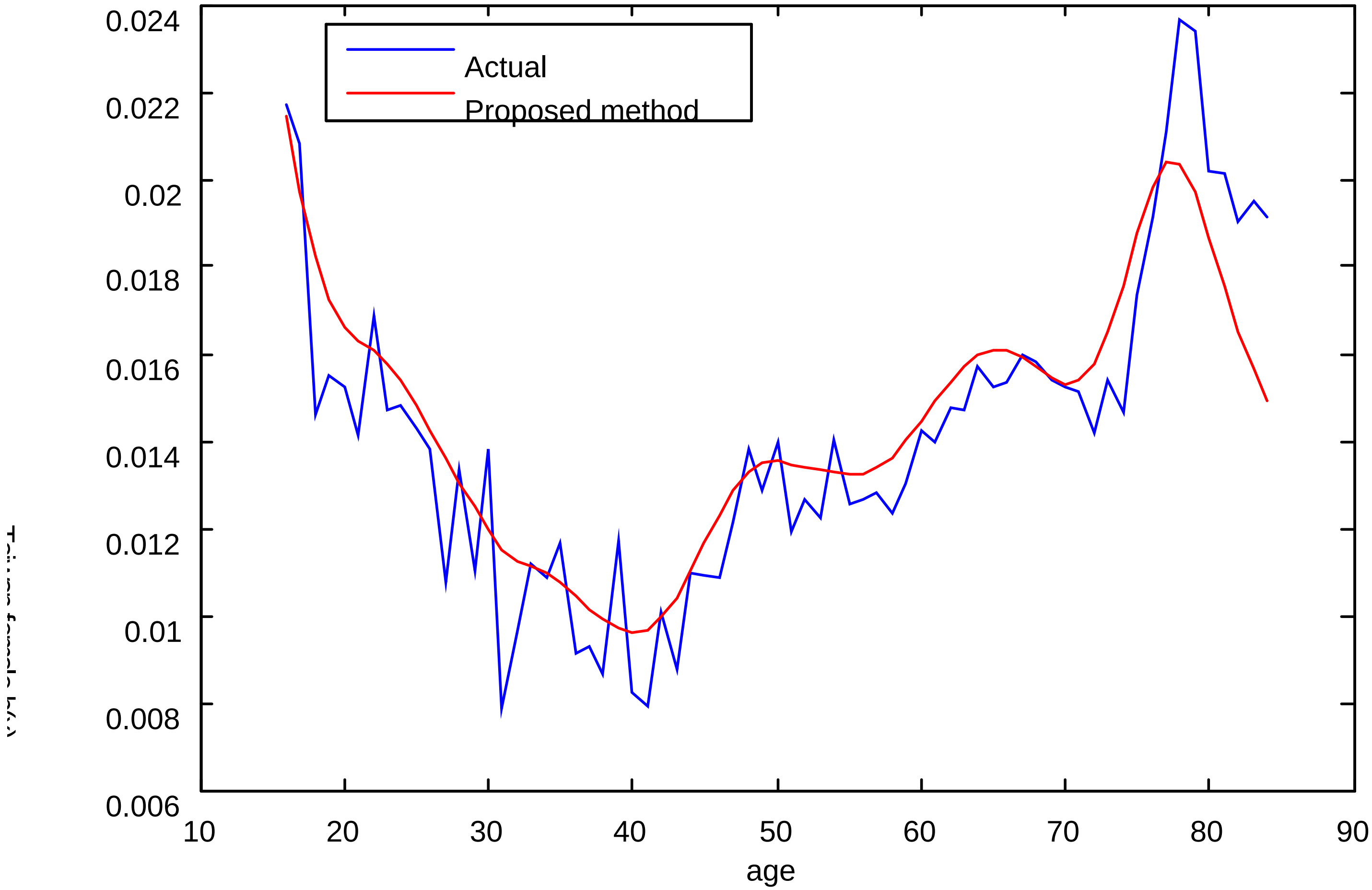
$$\log(m(x,T)) = \log(m(x,t)) + b(x)(k(T) - k(t))$$

- Life expectancy projection: $e_x = \sum_{t \geq 1} {}_tp_x$

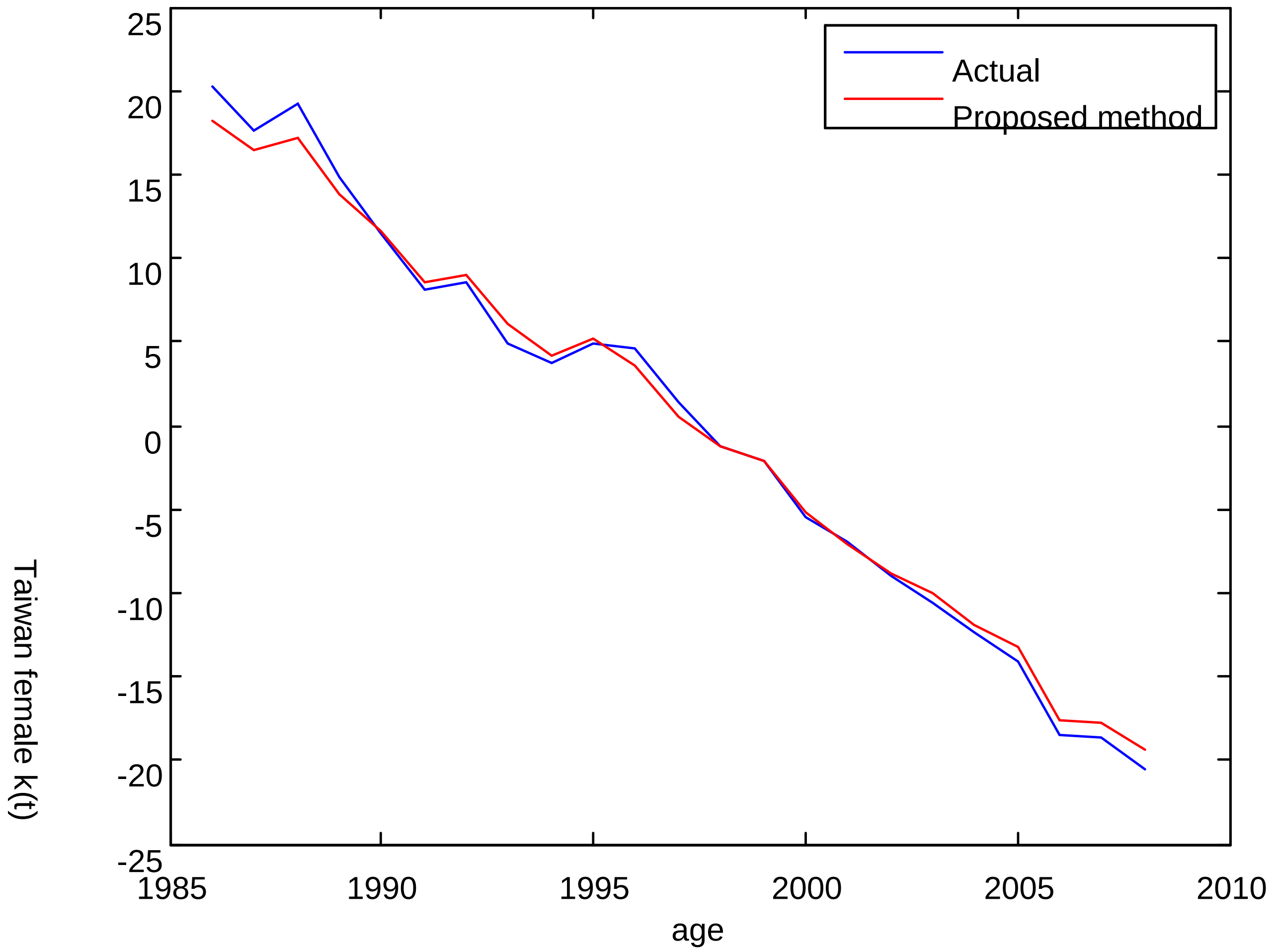
Comparisons of Lee-Carter Model parameter a(x) of Taiwan, female



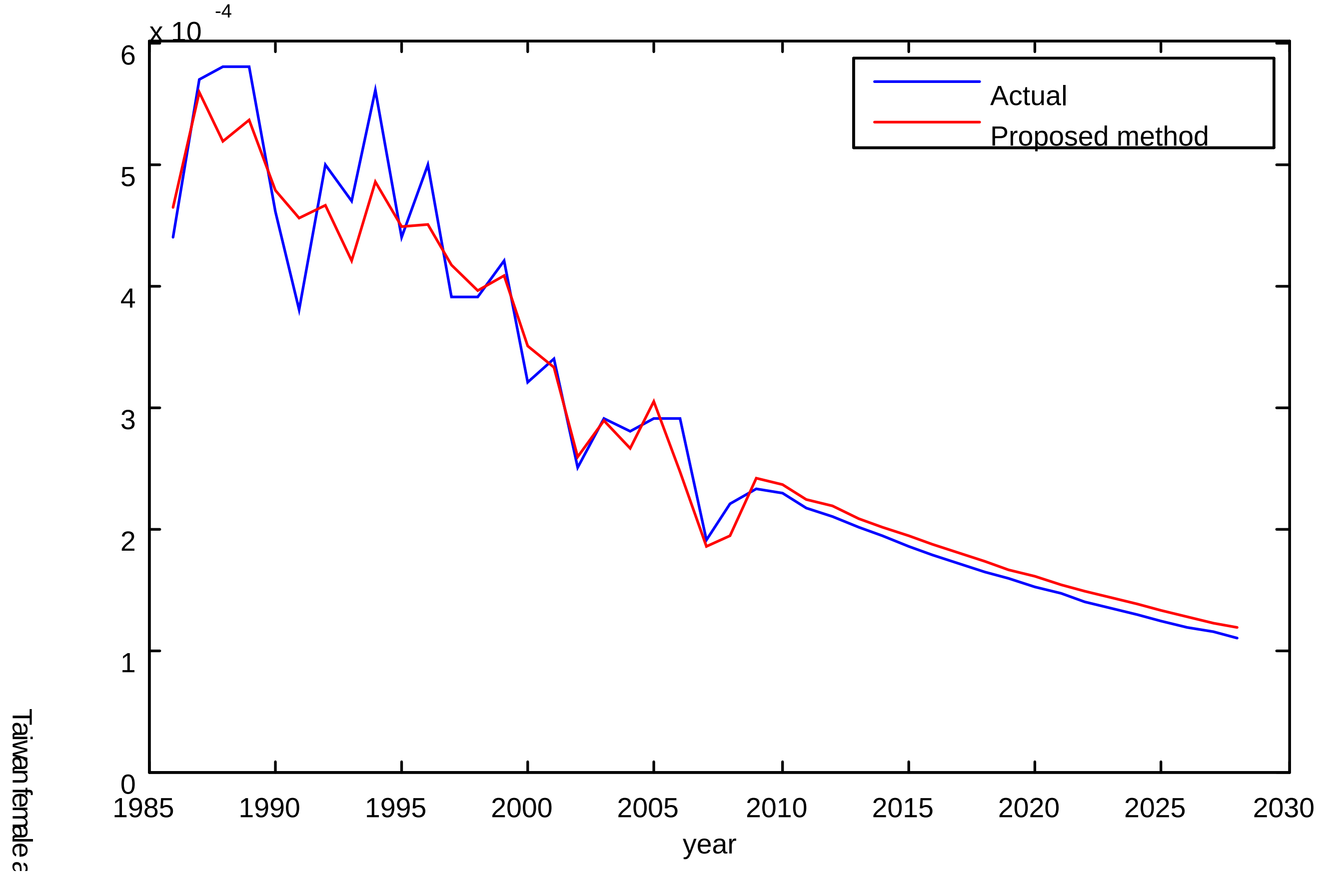
Comparisons of Lee-Carter Model parameter b(x) of Taiwan, female



Comparisons of Lee-Carter Model parameter k(t) of Taiwan, female

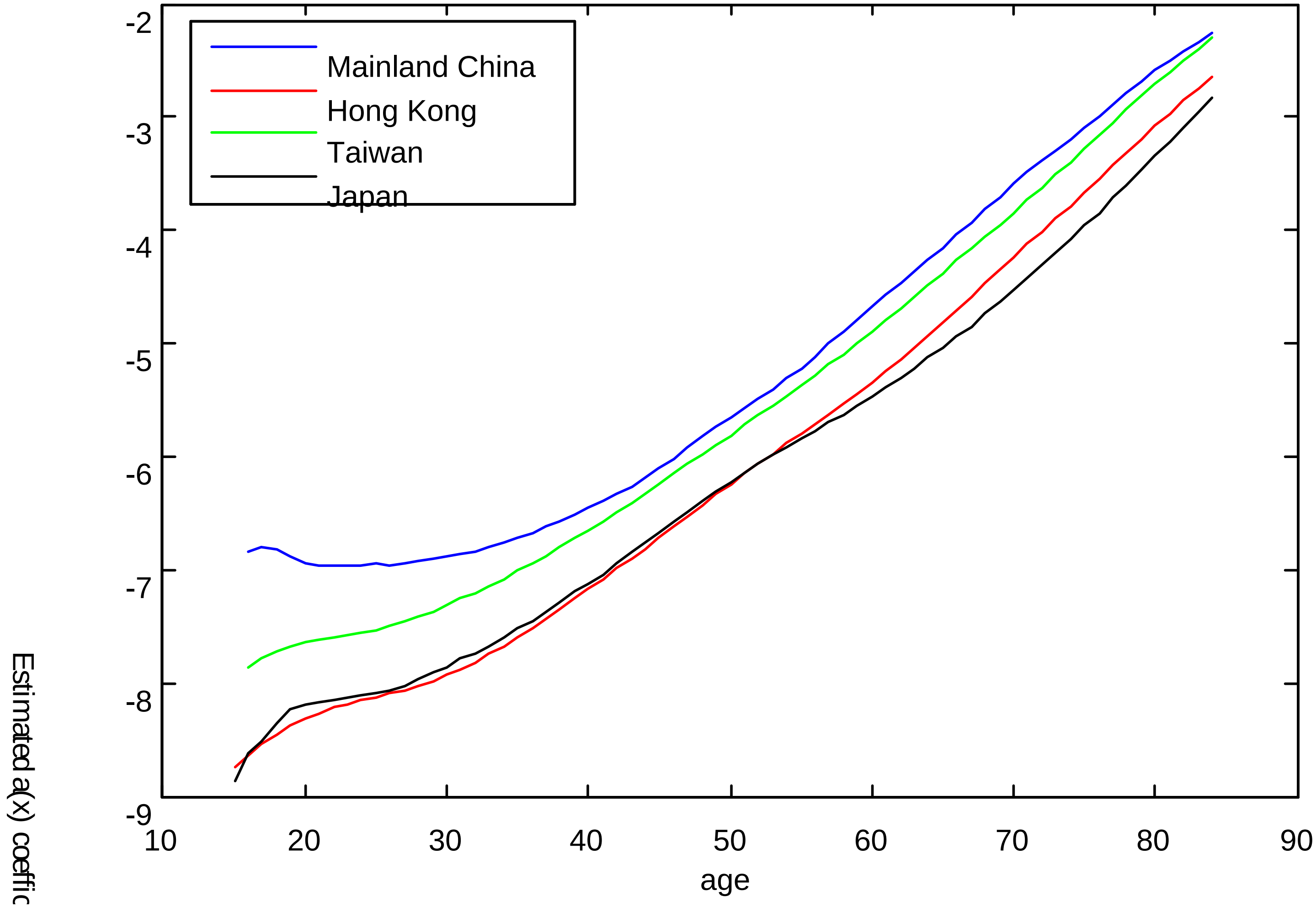


Mortality Projection of Taiwan, female, age 16

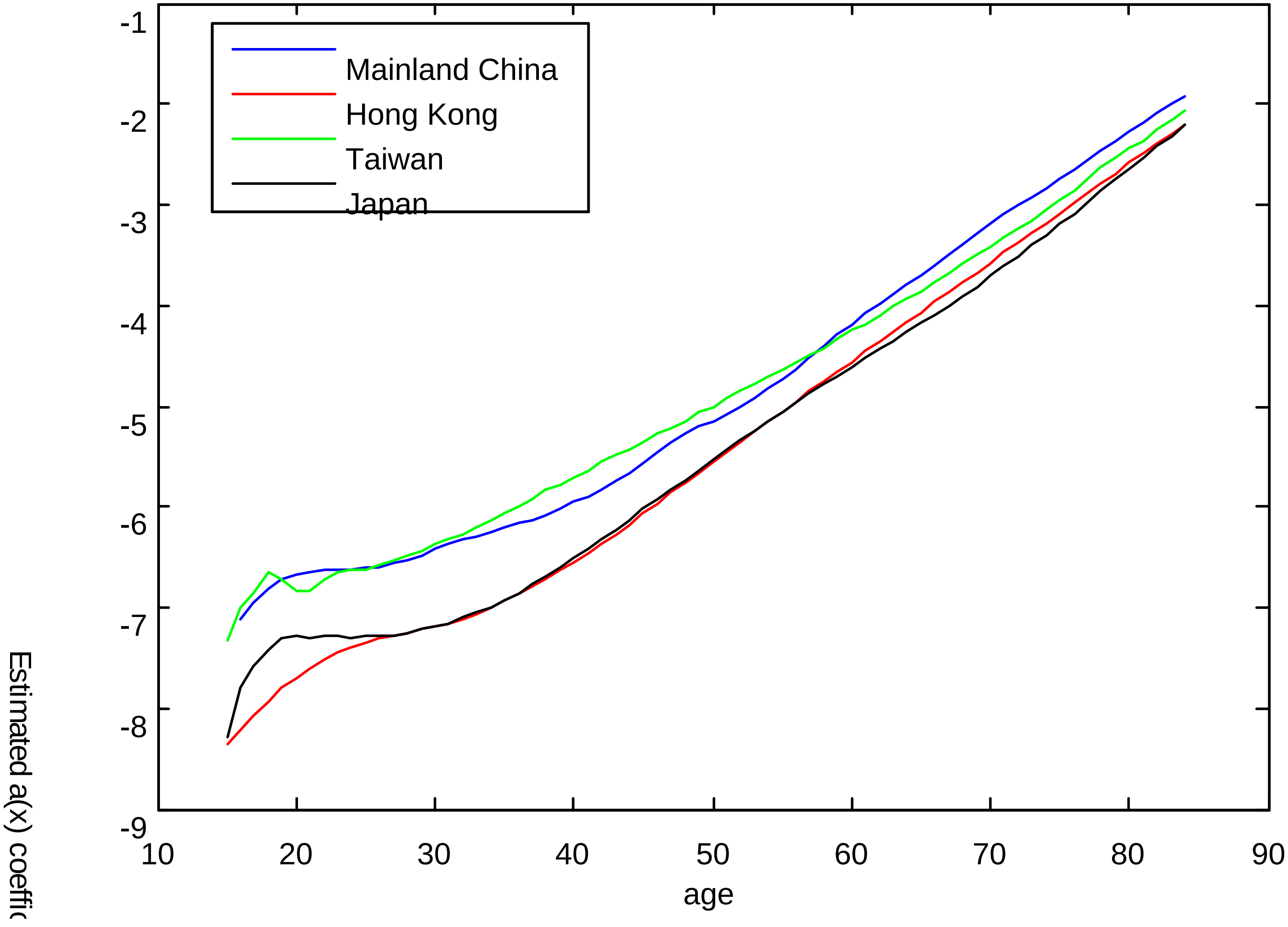


MAPE of mortality forecast for 2009-2028
at age 16,female is 5.27%.

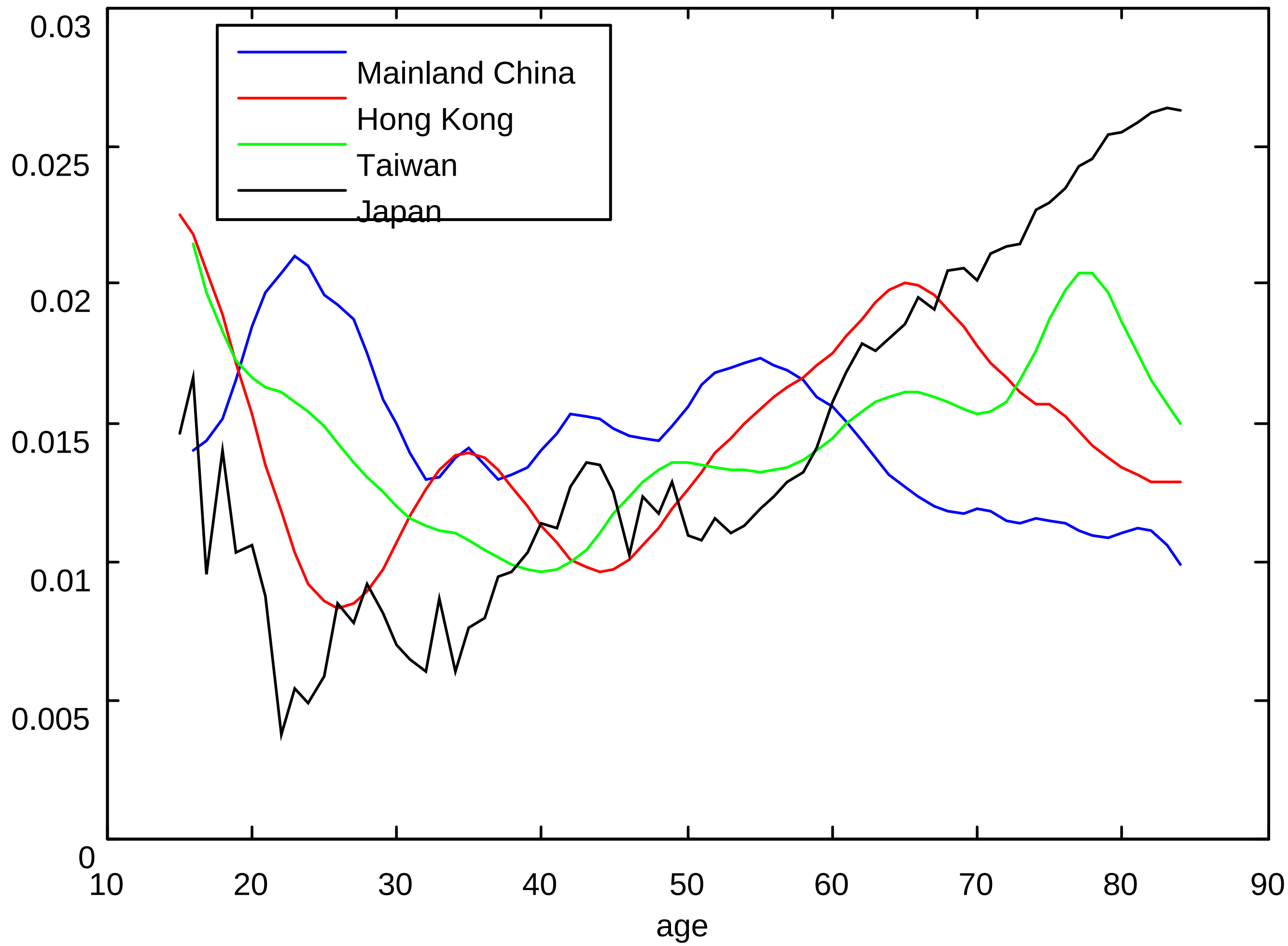
Comparisons of Lee-Carter Model parameter a(x) of Mainland China with other Asian populations, female



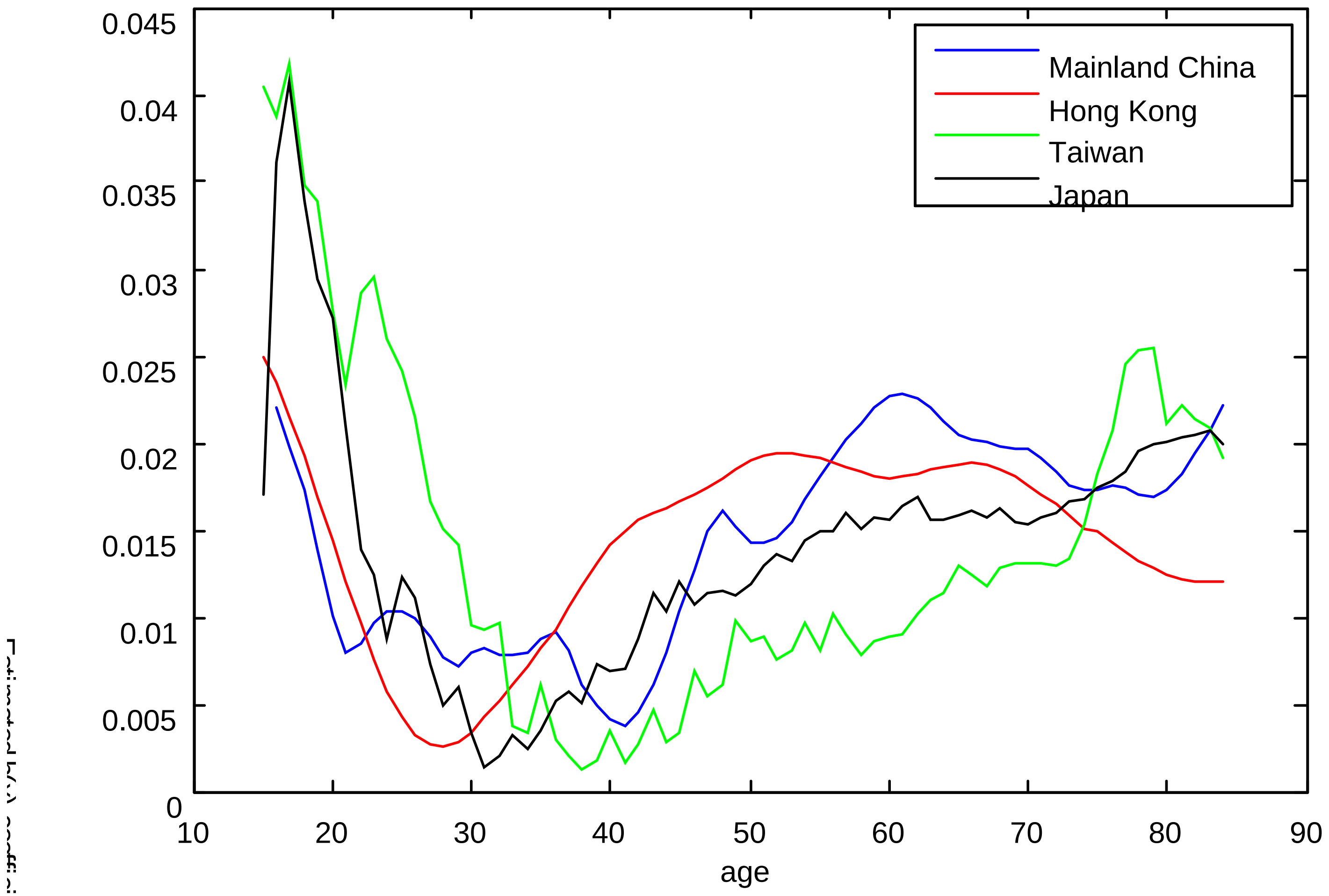
Comparisons of Lee-Carter Model parameter a(x) of Mainland China with other Asian populations, male



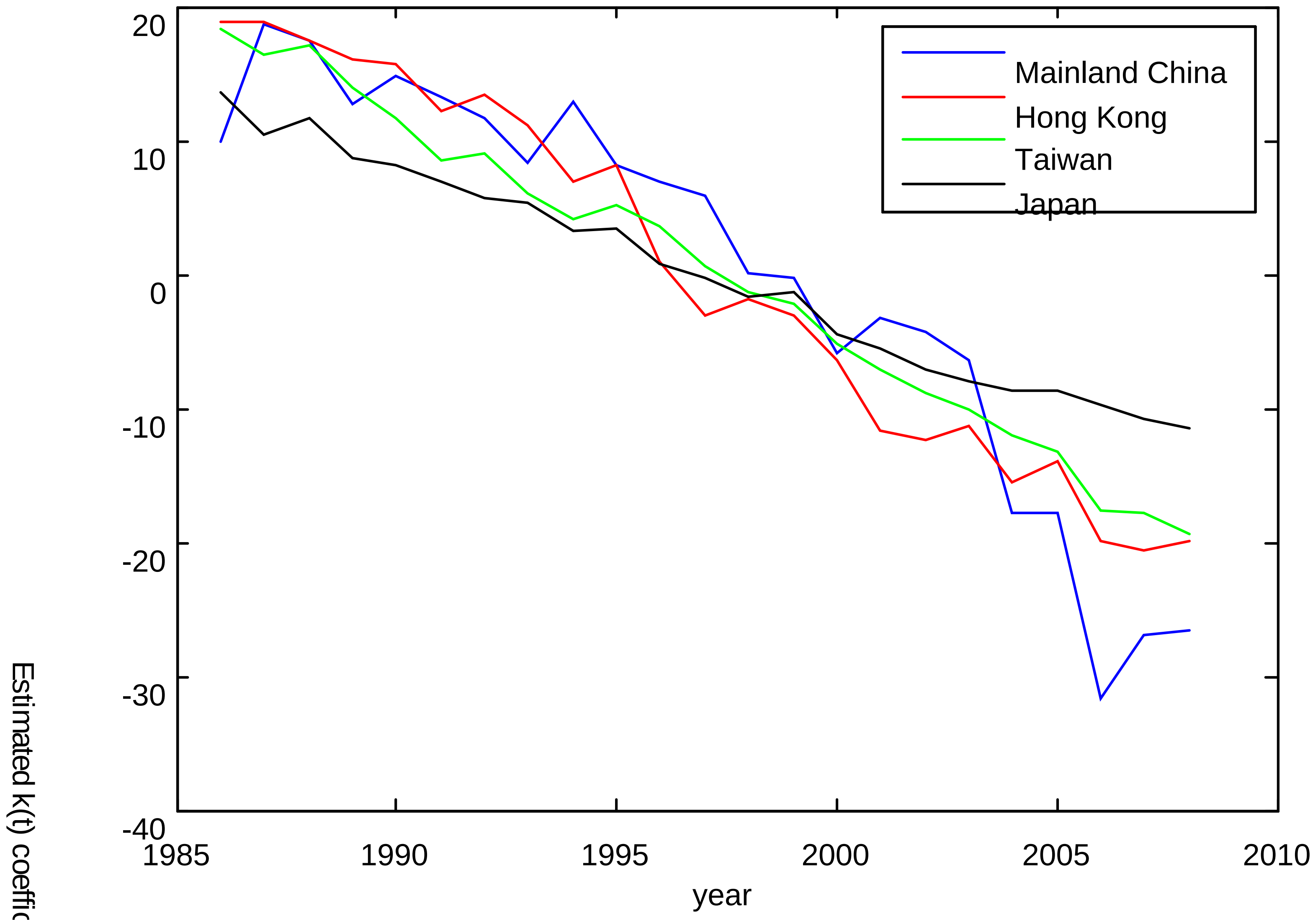
Comparisons of Lee-Carter Model parameter b(x) of Mainland China with other Asian populations, female



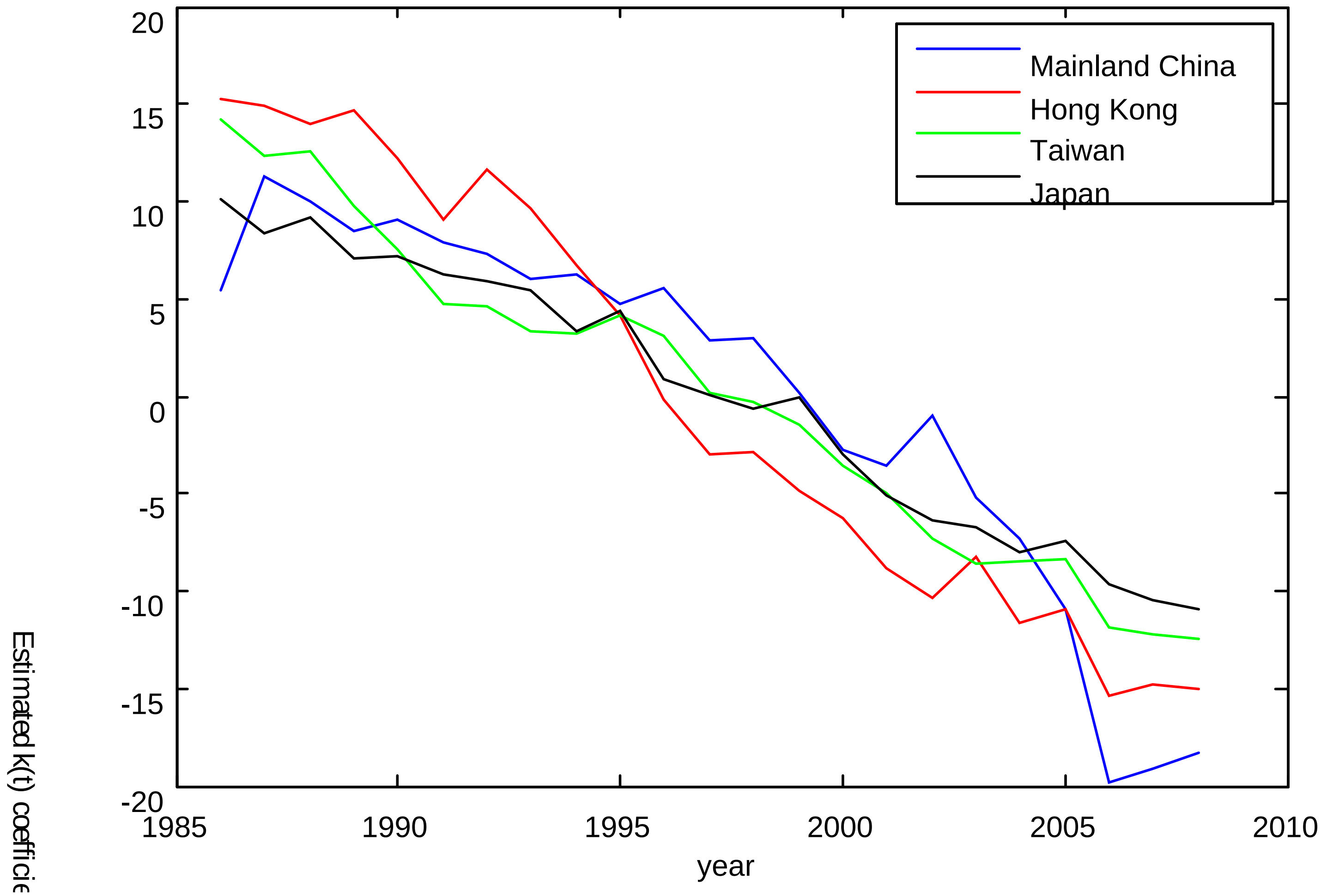
Comparisons of Lee-Carter Model parameter b(x) of Mainland China with other Asian populations, male



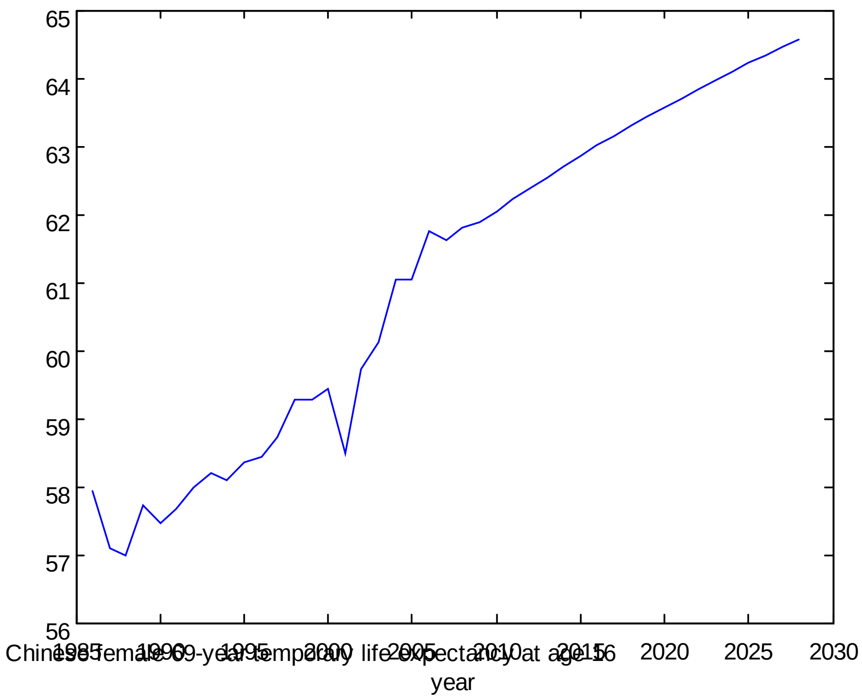
Comparisons of Lee-Carter Model parameter k(t) of Mainland China with other Asian populations, female

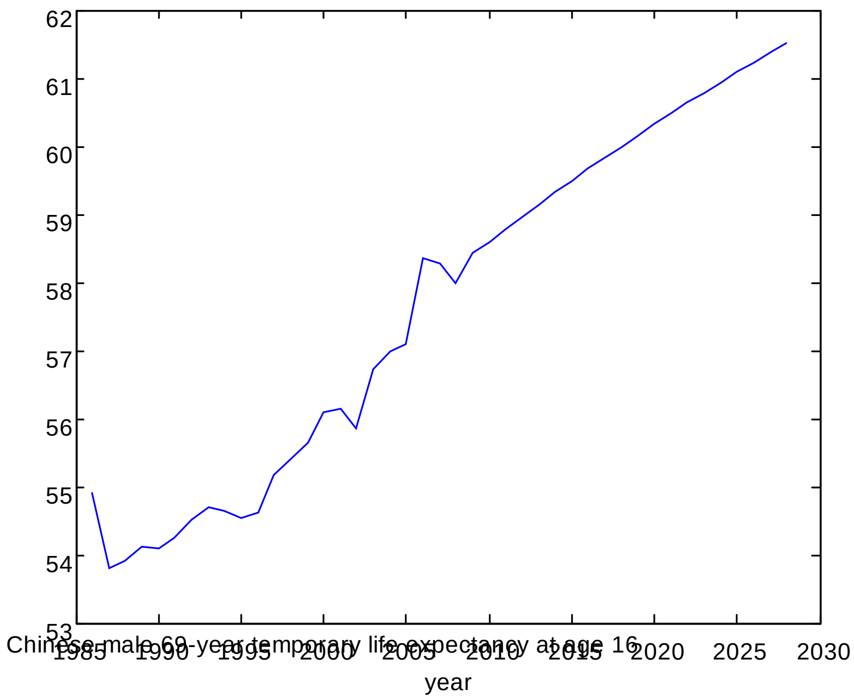


Comparisons of Lee-Carter Model parameter k(t) of Mainland China with other Asian populations, male



Life expectancy projection





Application

- We consider a simple portfolio of a single premium 20-year term annuity-due with unit annual benefit sold to a life-aged-60 in 2008. Valuations are based on the following assumptions:
  - Mortality follows the Chinese female population experience;
  - The interest rate is fixed at 5% per annum effective;
  - Net single premium payable at age 60;
  - No allowance is made for profit and expense.

Application

- Method 1 Using 2008 Chinese female mortality rate
- Method 2 Using LC model fitted by the proposed method data

Table 4 Comparison of single premium based on two methods

Method 1	Method 2	Absolute Error	Absolute Percentage Error
11.8001	12.3143	0.5142	4.36%

Conclusions

- Multiple Imputation works quite well dealing with missing data problem. Our method gives a way to forecast mortality for populations with missing data.
- Compared to the results of other countries, Lee-Carter parameters estimated by our proposed methods are reasonable.
- Based on the projection of our proposed method, taking account of the longevity risk leads an unignorable difference in annuity premium in China.