

对共同年龄效应模型的研究及中国应用

肖鸿民*, 李芳芳, 赵苗苗

西北师范大学数学与统计学院, 甘肃 兰州

收稿日期: 2021年10月8日; 录用日期: 2021年10月29日; 发布日期: 2021年11月10日

摘要

共同年龄效应模型是一种在多个人口数据中提取所有数据对象所共有的共同年龄效应的多人口死亡率模型。我们首次运用共同年龄效应模型研究中国男女性死亡率。基于中国男女性死亡率数据, 分别使用共同年龄效应模型的极大似然估计和Lee-Carter模型的最小二乘估计对数据进行了处理, 结果表明共同年龄效应模型在处理中国男女性死亡率数据时, 拟合和预测效果都很好。

关键词

共同年龄效应模型, Lee-Carter模型, 中国, 多人口, 死亡率

Research on the Common Age Effect Model and Its Application in China

Hongmin Xiao*, Fangfang Li, Miaomiao Zhao

College of Mathematics and Statistics, Northwest Normal University, Lanzhou Gansu

Received: Oct. 8th, 2021; accepted: Oct. 29th, 2021; published: Nov. 10th, 2021

Abstract

The common age effect model is a multi-population mortality model that extracts the common age effect shared by all data objects from multiple population data. For the first time, we use the common age effect model to study the mortality of men and women in China. Based on the Chinese male and female mortality data, the maximum likelihood estimation of the Common Age Effect model and the least square estimation of the Lee-Carter model were used to process the data. The results show that while the common age effect model processes the Chinese male and female mortality data, the effects of fitting and prediction are pretty well.

*通讯作者。

Keywords

Common Age Effect Model, Lee-Carter Model, China, Multi-Population, Mortality

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着社会的发展,人们认知水平和医疗水平的提高,我国人口出生率下降,死亡率改善,自然增长率下降。人民平均寿命每年都在提升,老年人口占总人口的比重也是逐年攀升。第七次人口普查结果显示,我国人口 10 年来继续保持低速增长态势,人口老龄化程度进一步加深,我国 60 岁及以上人口有 2.6 亿人,比重达到 18.7%,其中 65 岁及以上人口有 1.9 亿,占比 13.5% [1]。这意味着我国养老压力会越来越大。到 2050 年,大批现在的中青年老去,要依靠目前不断减少的新生儿抚养,会出现少数人抚养多数人的“倒金字塔”人口结构,也会造成养老金危机。由于人口老龄化带来的长寿风险对于个人养老、保险精算、社会保障乃至国家的安定团结都是至关重要的,因此我们有必要使用死亡率模型等方法,准确预测未来人口死亡率,度量和管理工作长寿风险。

对于死亡率模型的研究首先是随年龄变化的静态死亡率模型,逐渐发展到了包含年龄、时间、队列等的动态死亡率模型。根据人口对死亡率模型分类,可以分为单人口死亡率模型和多人口死亡率模型。在单人口死亡率模型中,最早的动态死亡率模型是 Li 和 Lee (1992)提出的 Lee-Carter 模型[2],其将死亡率分解为年龄因子和时间因子。之后出现了 Age-Period-Cohort 模型(2006) [3] (简称 APC 模型)和由 Cairns、Blake 和 Dowd 提出的 CBD 模型(2006) [4]等。其中 Lee-Carter 模型和 CBD 模型是最基础的两个模型。多人口死亡率模型,通过同时模拟多个人口的死亡率数据,来解决单人口死亡率模型预测中所出现的不合理交叉和偏离现象。主要模型有 Joint-k 模型(1992) [5], Common-Age-Effect 模型(2015) [6] (简称 CAE 模型), Augmented-Common-Factor (2018) [7]模型(简称 ACF 模型)。ACF 模型族可以提取共同年龄因子和共同时间因子。Joint-k 模型和共同年龄效应模型相对应,前者提取共同时间因子,后者提取共同年龄因子。在国内,李志生等用 Lee-Carter 模型分析了中国数据[8];赵明、王晓军等研究了多人口 Lee-Carter 随机死亡率模型及中国应用[9],以及理论方法和进展综述[10];曹园研究了基于 Lee-Carter 模型的我国死亡率预测[11];肖鸿民、马海飞等基于中国数据比较了两种死亡率预测方法[12]。

早在 2015 年, Kleinow (2015)基于 Lee-Carter 模型和 10 个国家的死亡率数据,提出了一个基于 P 阶年龄和时期效应的 Lee-Carter 模型的扩展模型,即共同年龄效应(CAE)模型[6]。该模型可以扩展到一个随机死亡率模型的多个群体。这允许同时产生为所有人群考虑的死亡率情景。而后, Kleinow 等人在 2016 年,分别使用 CAE 模型与 ACF(0)模型(2005)、ACF(1)模型(2017)、ACF(2)模型(2017)对多个国家的男性和女性死亡率数据作了拟合、预测和比较[13]。基于比较准则的 CAE 模型表现得更加令人满意。后期 Kleinow 等人还研究了基于聚类的共同年龄效应多人口死亡率模型的扩展[14]。

国内外学者使用各种模型研究中国人口死亡率,取得了很好的效果。却还没有人使用共同年龄效应模型(2016)研究中国人口死亡率。由于 Kleinow 等人使用共同年龄效应模型处理多国数据,准确提取了共同年龄因子,对各国人口死亡率的预测效果良好。因此本文将进一步测试该模型,并将该模型引入中国人口死亡率。本文将共同年龄效应模型(2015)扩展到其在欧洲五国和中国男女性死亡率的适用性上,该模

型在中国男女性死亡率上的应用是我们研究的重点。本文第二部分详细介绍了 Lee-Carter 模型和共同年龄效应模型的定义及参数估计方法；第三部分分别基于欧洲五国死亡率数据和中国男女性死亡率数据做了两个实验。第一个实验使用欧洲五国的数据来测试模型效果。第二个实验将该模型应用到中国死亡率数据，运用共同年龄效应模型(2015)处理中国的男性和女性死亡率数据，提取中国男女性死亡率数据的共同年龄因子。将结果同 Lee-Carter 模型分别处理中国男女性死亡率数据的结果作比较，由绝对拟合误差 (AFE)值，死亡率对比图和标准残差图，可知共同年龄效应模型对中国死亡率数据的拟合效果很好；由死亡率预测图和残差图，可知共同年龄效应模型的预测效果也很好；第四部分对本文的内容做了一个简单的总结，还提出了深入研究和应用共同年龄效应模型的建议。

2. 模型介绍和参数估计

2.1. Lee-Carter 模型

1) 模型介绍

Lee-Carter 模型是 Lee 和 Carter (1992)提出的出现最早的单人口动态死亡率模型。其将死亡率的变化分解为随时间而变化的时间因子，和不随时间而变化的年龄因子。该模型表达形式为：

$$\ln(m_{x,t}) = \alpha_x + \beta_x k_t + \xi_{x,t} \quad (1)$$

其中 $m_{x,t}$ 代表 t 年 ($t = t_0, t_0 + 1, \dots, t_0 + T - 1$) x 岁时的中心死亡率， K_t 是时间因子，反映了死亡率随日历年变动的趋势； $\xi_{x,t}$ 表示 t 年 x 岁时中心死亡率的残差，并且 $\xi_{x,t}$ 是一个均值为 0，方差为 σ_ξ^2 的白噪声过程。

式子(1)中的 α_x, β_x, K_t 都是需要估计的参数。为了保证估计结果的唯一性，可添加附加条件： $\sum_t K_t = 0, \sum_x \beta_x = 1$ 。故 α_x 表示 x 年龄(组)的中心死亡率取对数后的均值， K_t 表示式子(1)中的 α_x, β_x, K_t 都是需要估计的参数。为了保证估计结果的唯一性，可添加附加条件： $\sum_t K_t = 0, \sum_x \beta_x = 1$ 。故 α_x 表示 x 年龄(组)的中心死亡率取对数后的均值， K_t 表示年份 t 时死亡率的强度，也称死亡指数。

2) 参数估计

Lee-Carter 模型中参数估计的方法主要包括奇异值分解法(SVD)，加权最小二乘法(WLS)和极大似然估计法(MLE)。由于加权最小二乘法(WLS)对中国数据的拟合和预测效果最好，所以本文将使用加权最小二乘法。

在使用加权最小二乘法估计 Lee-Carter 模型的参数时，需要最小化下式：

$$\sum_{x=1}^X \sum_{t=1}^T \omega_{x,t} (f_{x,t} - \alpha_x - \beta_x K_t)^2 \quad (2)$$

其中， $\omega_{x,t} = d_t, \sum_t K_t = 0, \sum_x \beta_x = 1$ 。

(2)式中的权重为死亡人数的观测值，为了最小化(2)式，需要分别求它关于 α_x, β_x 和 K_t 的一阶导数，然后令其分别等于 0。可得如下 α_x, β_x 和 K_t 的估计值的计算公式：

$$\hat{\alpha}_x = \frac{\sum_t \omega_{x,t} (f_{x,t} - \hat{\beta}_x \hat{K}_t)}{\sum_t \omega_{x,t}} \quad (3)$$

$$\hat{\beta}_x = \frac{\sum_t \omega_{x,t} \hat{K}_t (f_{x,t} - \hat{\alpha}_x)}{\sum_t \omega_{x,t} \hat{K}_t^2} \quad (4)$$

$$\hat{K}_t = \frac{\sum_x w_{x,t} \hat{\beta}_x (f_{x,t} - \hat{\alpha}_x)}{\sum_x \omega_{x,t} \hat{\beta}_x^2} \quad (5)$$

由(2)式,通过迭代,当对数死亡率的实际值和估计值的差的绝对值小于指定的偏差时, $\hat{\alpha}_x, \hat{\beta}_x$ 和 \hat{K}_t 就是 α_x, β_x 和 K_t 的估计值。

2.2. 共同年龄效应模型(CAE 模型)

1) 模型介绍

Kleinow (2015)提出了一个共同年龄效应模型(CAE 模型),发现一般的年龄效应模型可以推广到多种群的随机死亡模型之中,其表达式为:

$$\ln(m_{x,t,i}) = \alpha_{x,i} + B_x^{(1)} k_{t,i}^{(1)} + B_x^{(2)} k_{t,i}^{(2)} \quad (6)$$

其中, $x \in \{0, 1, 2, \dots, \omega\}, t \in \{1, 2, \dots, T\}, i \in \{1, 2, \dots, N\}$ 。 $m_{x,t,i}$ 表示第 i 人群中 x 岁的人在年份 t 的中心死亡率。 $\alpha_{x,i}$ 表示第 i 个人群的附加年龄因子,表示不同人群对数死亡率的均值。 $k_{t,i}^{(1)}$ 和 $k_{t,i}^{(2)}$ 是附加时间因子,反映了第 i 个人群对数死亡率随时间的变化趋势, $B_x^{(1)}$ 和 $B_x^{(2)}$ 是从所有数据群组中提取出来的共同年龄因子,代表年龄对附加时间因子 $k_{t,i}^{(1)}$ 和 $k_{t,i}^{(2)}$ 的敏感程度。 $\varepsilon_{x,t,i}$ 是随机误差项。

为了满足参数估计结果的唯一性,需要满足以下约束条件:

$$\sum_x B_x^{(1)} = 1, \sum_x B_x^{(2)} = 1, \sum_t k_{t,i}^{(1)} = 0, \sum_t k_{t,i}^{(2)} = 0.$$

2) 参数估计

共同年龄效应(CAE)模型中, $\alpha_{x,i}, B_x^{(1)}, B_x^{(2)}, k_{t,i}^{(1)}, k_{t,i}^{(2)}$ 都是需要估计的参数。本文采用极大似然方法(MLE)对共同年龄效应(CAE)模型进行参数估计,该方法假定死亡人数 $D_{x,t,i}$ 服从参数为 $\lambda_{x,t,i}$ 的泊松分布, $\lambda_{x,t,i} = m_{x,t,i} E_{x,t,i}$, 即:

$$D_{x,t,i} \sim \text{Poisson}(m_{x,t,i} E_{x,t,i}) \quad (7)$$

其中, $m_{x,t,i} = \exp(\alpha_{x,i} + B_x^{(1)} k_{t,i}^{(1)} + B_x^{(2)} k_{t,i}^{(2)})$, $D_{x,t,i}$ 表示第 i 个人群年龄为 x 年份为 t 的死亡人数, $E_{x,t,i}$ 表示第 i 个人群年龄为 x 年份为 t 的平均暴露人数。

假设每一年龄的死亡率相互独立,则 CAE 模型的似然函数可以表示为:

$$l(\alpha_{x,i}, B_x^{(1)}, k_{t,i}^{(1)}, B_x^{(2)}, k_{t,i}^{(2)}) = \prod_{x,t,i} \frac{(E_{x,t,i} m_{x,t,i})^{D_{x,t,i}}}{D_{x,t,i}!} \exp(-E_{x,t,i} m_{x,t,i}) \quad (8)$$

对上式取对数,可以得到 CAE 模型的对数似然函数:

$$\begin{aligned} L &= L(\alpha_{x,i}, B_x^{(1)}, k_{t,i}^{(1)}, B_x^{(2)}, k_{t,i}^{(2)}) \\ &= \sum_{x,t,i} \left[D_{x,t,i} (\alpha_{x,i} + B_x^{(1)} k_{t,i}^{(1)} + B_x^{(2)} k_{t,i}^{(2)}) - E_{x,t,i} \exp(\alpha_{x,i} + B_x^{(1)} k_{t,i}^{(1)} + B_x^{(2)} k_{t,i}^{(2)}) \right] + c \end{aligned} \quad (9)$$

其中, $c = \sum_{x,t,i} [D_{x,t,i} \ln(E_{x,t,i}) - \ln(D_{x,t,i}!)]$ 为常数。

通过使得对数似然函数达到极大,可以求得参数的估计值,由于某些交叉项的存在,运算复杂,所以本文采用牛顿迭代法进行,公式如下:

$$\hat{\theta}^{(n+1)} = \hat{\theta}^{(n)} - \frac{\partial L^{(n)} / \partial \theta}{\partial^2 L^{(n)} / \partial \theta^2} \quad (10)$$

其中, $\hat{\theta}^{(n)}$ 是第 n 次迭代参数, $L^{(n)} = L[\hat{\theta}^{(n)}]$ 。

设待估参数的初值分别为:

$$\hat{\alpha}_{x,i} = \frac{1}{T} \sum_t \ln \left(\frac{D_{x,t,i}}{E_{x,t,i}} \right), \hat{B}_x^{(1)} = \hat{B}_x^{(2)} = 1, \hat{k}_{t,i}^{(1)} = \hat{k}_{t,i}^{(2)} = 0 \quad (11)$$

可得如下具体的迭代表达式:

$$\hat{\alpha}_{x,i} = \hat{\alpha}_{x,i} + \frac{\sum_t (D_{x,t,i} - \hat{D}_{x,t,i})}{\sum_t \hat{D}_{x,t,i}} \quad (12)$$

$$\hat{k}_{t,i}^{(1)} = \hat{k}_{t,i}^{(1)} + \frac{\sum_x (D_{x,t,i} - \hat{D}_{x,t,i}) \hat{B}_x^{(1)}}{\sum_x \hat{D}_{x,t,i} \hat{B}_x^{(1)2}} \quad (13)$$

$$\hat{k}_{t,i}^{(2)} = \hat{k}_{t,i}^{(2)} + \frac{\sum_x (D_{x,t,i} - \hat{D}_{x,t,i}) \hat{B}_x^{(2)}}{\sum_x \hat{D}_{x,t,i} \hat{B}_x^{(2)2}} \quad (14)$$

$$\hat{B}_x^{(1)} = \hat{B}_x^{(1)} + \frac{\sum_{t,i} (D_{x,t,i} - \hat{D}_{x,t,i}) \hat{k}_{t,i}^{(1)}}{\sum_{t,i} \hat{D}_{x,t,i} \hat{k}_{t,i}^{(1)2}} \quad (15)$$

$$\hat{B}_x^{(2)} = \hat{B}_x^{(2)} + \frac{\sum_{t,i} (D_{x,t,i} - \hat{D}_{x,t,i}) \hat{k}_{t,i}^{(2)}}{\sum_{t,i} \hat{D}_{x,t,i} \hat{k}_{t,i}^{(2)2}} \quad (16)$$

其中, $\hat{D}_{x,t,i} = \hat{E}_{x,t,i} \exp(\hat{\alpha}_{x,i} + \hat{B}_x^{(1)} \hat{k}_{t,i}^{(1)} + \hat{B}_x^{(2)} \hat{k}_{t,i}^{(2)})$ 。当对数死亡率的实际值和估计值的差的绝对值小于指定的偏差时, 就得到该模型的参数估计值。

3. 基于 6 国数据的比较分析

3.1. 基于欧洲五国数据训练模型

由于 Kleinow 等人的原始数据里有大量的欧洲国家死亡率数据, 且欧洲五国具有相近的地理位置, 会有更多的共性, 所以本文基于欧洲五国死亡率数据, 测试模型的效果。本文采用的原始数据来源于人类死亡率数据库, 选取 Austria, Denmark, France, UK, Sweden 的 1948 到 2018 年的, 18 到 98 岁的整体死亡率数据。其中, 1948~2007 年的数据用于训练模型, 2008~2018 年的数据用于测试模型。其中很关键的一步就是提取它们的共同年龄效应。

首先, 由极大似然估计法, 即式(12)~(16), 拟合模型和估计参数, 参数估计结果如图 1 所示。

图 1 中, Austria, Denmark, France, UK, Sweden 分别对应的线的颜色是: 黑色, 绿色, 红色, 紫色和蓝色。 $\alpha_{x,i}$ 表示不同人群对数死亡率的平均水平, 所以 5 个国家的 $\alpha_{x,i}$ 曲线都是随着年龄增长而增大到一定的值。 $k_{t,i}^{(1)}$ 和 $k_{t,i}^{(2)}$ 是附加时间因子, 反映了不同人群对数死亡率随时间的变化趋势, 其中前者持续

下降，反应了死亡率随时间改善的趋势，后者普遍都是先升后降。 $B_x^{(1)}$ 和 $B_x^{(2)}$ 是提取的所以国家的共同年龄因子。

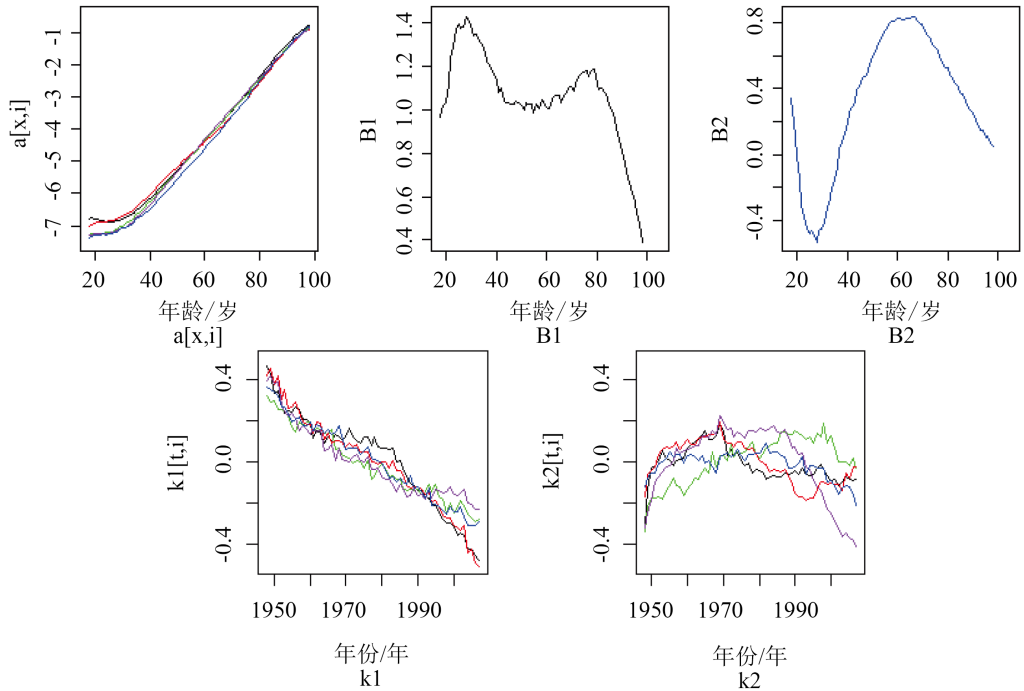


Figure 1. Parameter estimation of CAE model based on data from the 5 countries
图 1. CAE 模型基于 5 国数据的参数估计

CAE 模型中死亡率被分解为时间效应和年龄效应。对于时间项 $k_{t,i}^{(1)}$ 和 $k_{t,i}^{(2)}$ ，可以采取带漂移项的随机游走模型对数据进行拟合，预测 $k_{t,i}^{(1)}$ 和 $k_{t,i}^{(2)}$ 在将来 11 年的预测值。结果如图 2。

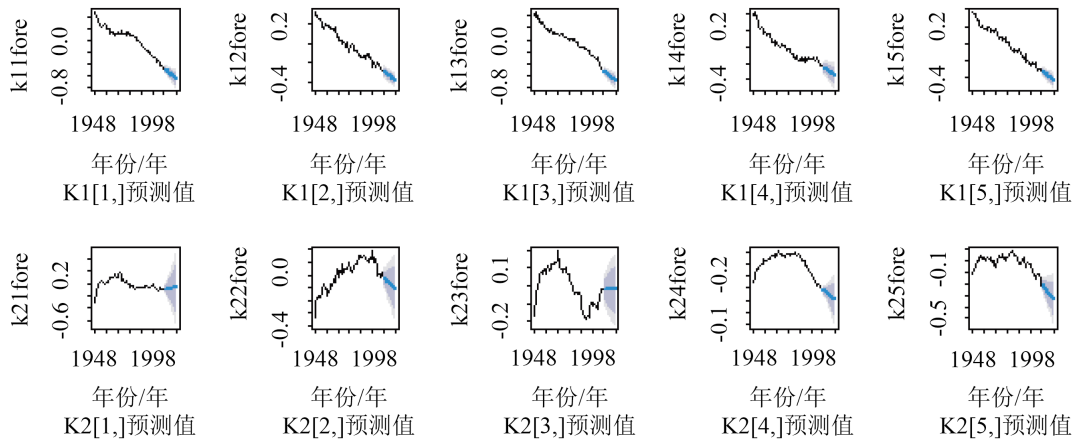


Figure 2. Estimated value of time term in the next 10 years
图 2. 时间项未来 10 年的估计值

在时间项已知的情况下，由公式

$$m_{x,s_0+t,i} \approx \exp\left(\hat{\alpha}_{x,i} + \hat{B}_x^{(1)}\hat{k}_{s_0+t,i}^{(1)} + \hat{B}_x^{(2)}\hat{k}_{s_0+t,i}^{(2)}\right) \quad (17)$$

就可以预测出当前时刻 s_0 之后的中心死亡率。全年齡段死亡率预测结果如下图 3 所示。本文选取了不同城市不同年份的死亡率作图，其中实际死亡率是红色，预测死亡率是蓝色。可以看出，CAE 模型对这 5 个国家整体死亡率数据的拟合和预测效果良好。

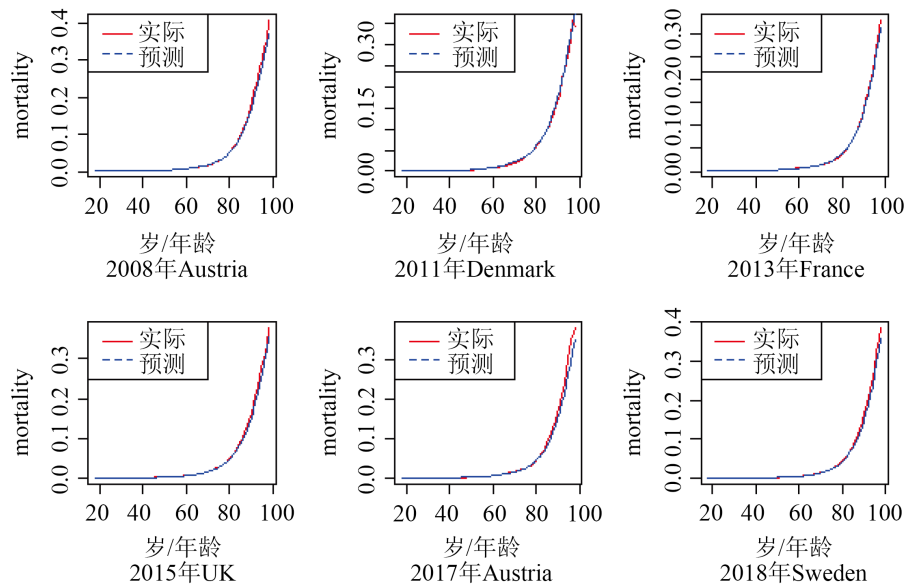


Figure 3. Comparison of actual and projected mortality rates for different countries in different years
图 3. 不同国家不同年份实际和预测死亡率对比

如下图 4，是不同国家不同年份的实际和预测的对数死亡率。可以看出在 11 年的预测区间内，对数死亡率之间的差距特别小，说明对老年数据的拟合和预测效果较好。而前期对数死亡率值有部分偏差，主要是由于前期部分死亡率太小，导致其对数变动大。

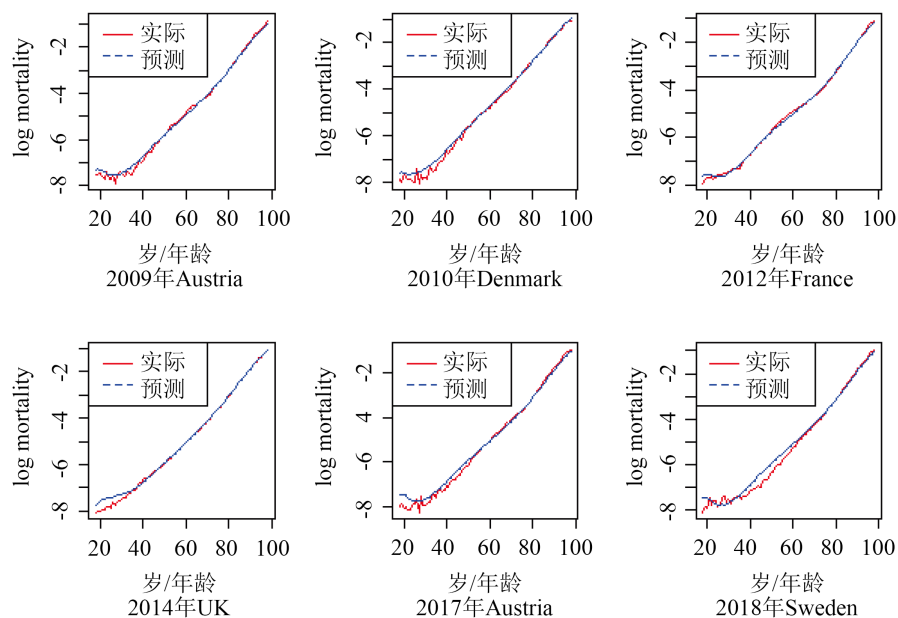


Figure 4. Actual and projected logarithmic mortality rates for different countries in different years
图 4. 不同国家不同年份的实际和预测的对数死亡率

如下图 5，是法国年龄为 75，85，95 岁的人的实际和预测死亡率。由图可知年龄越大，死亡率越高。

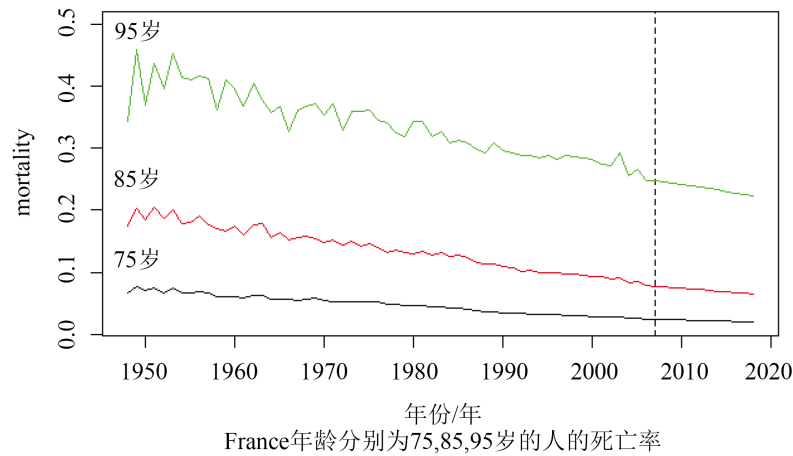


Figure 5. Mortality rates of people aged 75, 85, 95 years in France
图 5. France 年龄分别为 75、85、95 岁的人的死亡率

接下来采用 AFE (绝对拟合误差)来评估 CAE 模型对 5 个国家的拟合和预测效果。AFE 表达式为：

$$AFE_i = \sum_{x,t} |m_{x,t,i} - \hat{m}_{x,t,i}|。$$

结果如下表 1 所示，计算所得的 5 个 AFE 值如下。

Table 1. AFE values of 5 countries
表 1. 5 国的 AFE 值

国家	Austria	Denmark	France	UK	Sweden
AFE	3.103497	1.826948	1.15589	1.480876	2.370731

可以发现 AFE 的值都不是很大，说明拟合效果整体比较好。且对 UK 的拟合最好，对 Austria 的拟合误差最大，为 3.103497。

3.2. 共同年龄效应模型在中国数据上的应用

共同年龄效应(CAE)模型是基于国外人口死亡率数据提出的，本文前面使用欧洲五国数据测试模型效果，效果良好。现在希望把共同年龄效应模型应用到国内，观察效果是否良好。良好的话就可以用它来拟合和预测我国死亡率。本文采用中国的男性和女性的 1995 到 2019 年的，0 到 89 岁的整体死亡率数据。其中，1995~2010 年的数据用于训练模型，2011~2019 年的数据用于测试模型。接下来将会分别使用 Lee-Carter 模型和 CAE 模型来处理中国男女性死亡率数据。然后比较两个模型的拟合和预测效果。

3.2.1. Lee-Carter 模型处理中国男女性死亡率数据

由于使用 Lee-Carter 模型处理中国数据时，WLS 进行参数估计的效果最好，所以本文使用 WLS 来进行参数估计。

如下图 6 是使用 WLS 进行参数估计的结果。画出了 $a_{wls}, b_{wls}, k_{wls}$ 的值，图中黑色线是男性死亡率，红色线是女性死亡率。由 a_{wls} 图像可知男性死亡率普遍高于女性死亡率，结果符合我国实际。且男女性对数死亡率都是先减小后增大的。 k_{wls} 也是随着年份在降低，说明死亡率在不断改善，不过改善的趋势比较缓慢。

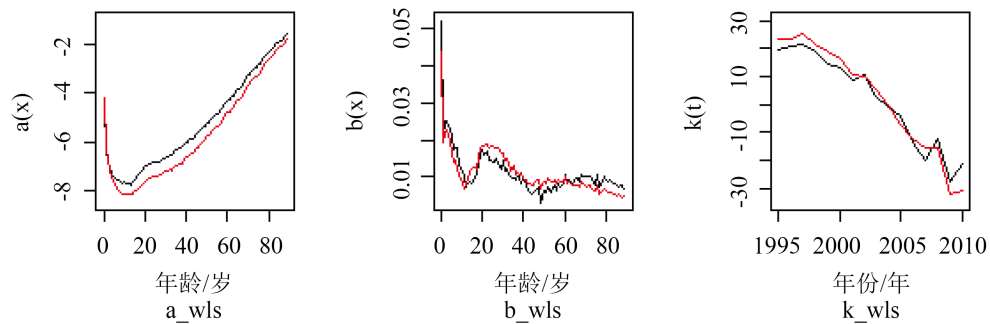


Figure 6. Parameter estimates of the Lee-Carter model for male and female mortality in China
图 6. Lee-Carter 模型对中国男女性死亡率的参数估计值

由于 Lee-Carter 模型将死亡率的变化分解为随时间而变化的时间因子和不随时间而变化的年龄因子。因此采用带漂移项的随机游走模型对数据进行拟合，并预测时间项。如下图 7 是对时间项进行预测的结果。

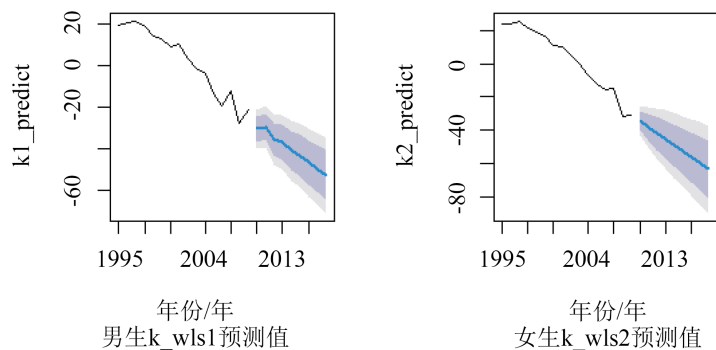


Figure 7. Predicted value of time term k
图 7. 时间项 k 的预测值

当得到预测的时间项时，就可以用 $m_{x,s_0+t,i} \approx \exp(\hat{\alpha}_{x,i} + \hat{B}_x^{(1)} \hat{k}_{s_0+t,i}^{(1)} + \hat{B}_x^{(2)} \hat{k}_{s_0+t,i}^{(2)})$ 来算出未来一段时间的预测死亡率。从而得到全年龄段死亡率。如下图 8，画出 2015 年的男性和 2013 年的女性的实际死亡率和用 Lee-Carter 模型预测的死亡率。其中黑色线表示男性，红色线表示女性。

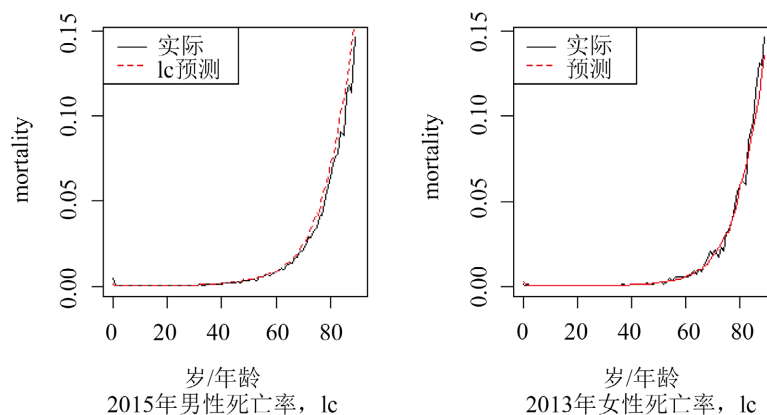


Figure 8. Actual and projected mortality rates for men and women
图 8. 男性和女性的实际和预测死亡率

3.2.2. 共同年龄效应模型处理中国男女性数据

使用共同年龄效应(CAE)模型拟合中国男女性数据,提取中国男性和女性死亡率共同年龄效应 $B_x^{(1)}$ 和 $B_x^{(2)}$ 。首先,根据极大似然估计法,即式(12)~(16),拟合模型和估计参数,参数估计结果如下图 9 所示。黑色线代表男性,红色线代表女性。由 $a[x,i]$,可知男性死亡率高于女性死亡率。由于 $k1[t,i]$, $k2[t,i]$ 随着年份有很强的下降的趋势,说明死亡率有所改善。 $B1$ 和 $B2$ 则是提取的中国男女性死亡率共同趋势。

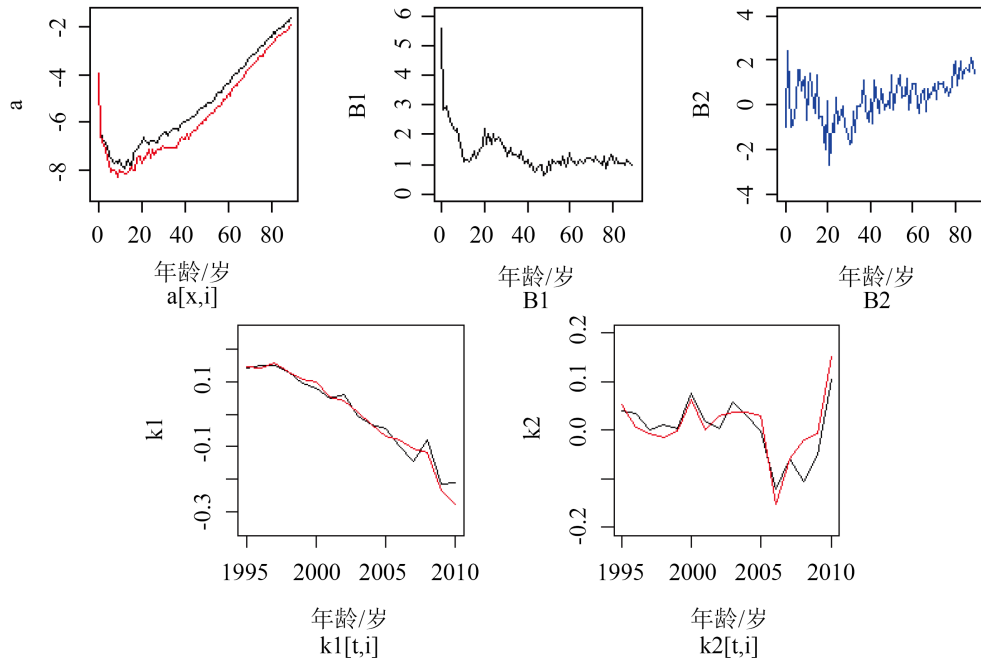


Figure 9. Estimation results of male and female data parameters processed by CAE model in China
图 9. CAE 模型处理中国男女性数据参数估计结果

由于 CAE 模型将死亡率的变化分解为时间因子和年龄因子,共同年龄效应。因此采用带漂移项的随机游走模型对数据进行拟合,并预测时间项。如下图 10 是对时间项进行预测的结果。

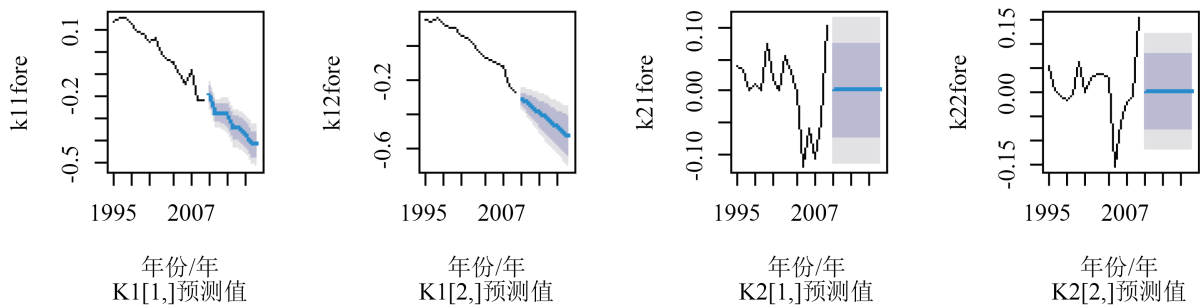


Figure 10. Predicted value of time term in the next 10 years
图 10. 时间项在未来 10 年的预测值

当得到预测的时间项时,就可以用 $m_{x,s_0+t,i} \approx \exp(\hat{a}_{x,i} + \hat{B}_x^{(1)} \hat{k}_{s_0+t,i}^{(1)} + \hat{B}_x^{(2)} \hat{k}_{s_0+t,i}^{(2)})$ 来算出未来一段时间的预测死亡率。从而得到全年龄段死亡率。如下图 11,画出用 CAE 模型的 2015 年的男性和女性的实际和预

测死亡率。其中红色线表示实际死亡率值，蓝色线代表预测的死亡率值。

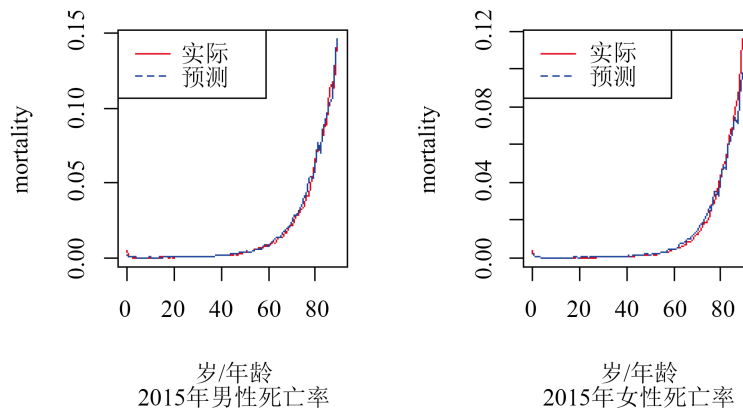
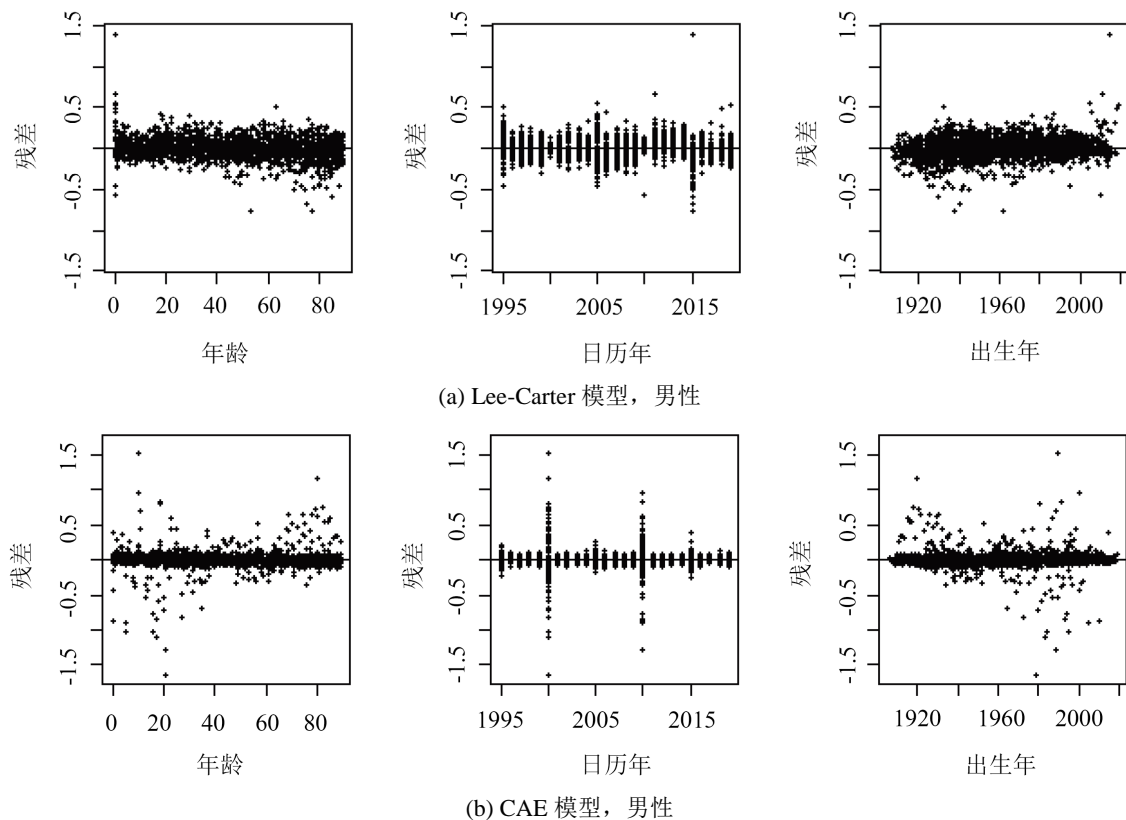


Figure 11. Actual and predicted mortality rates for males and females in a CAE model

图 11. CAE 模型下男性和女性的实际和预测死亡率

3.3. 模型对比

为了比较模型的拟合效果，画出它的标准残差图。如下图 12 所示。由图可知 Lee-Carter 模型和共同年龄效应(CAE)模型都较好得捕捉到了年龄效应和时间效应。Lee-Carter 模型的残差图整体比较稳定，CAE 模型对年龄效应和时间效应的提取更集中一点，虽然存在部分残差比较大的点，但在具有共同年龄效应的数据对象增多的条件下，它的优势才能更好得显现出来。故 CAE 模型的拟合效果是比较好的。



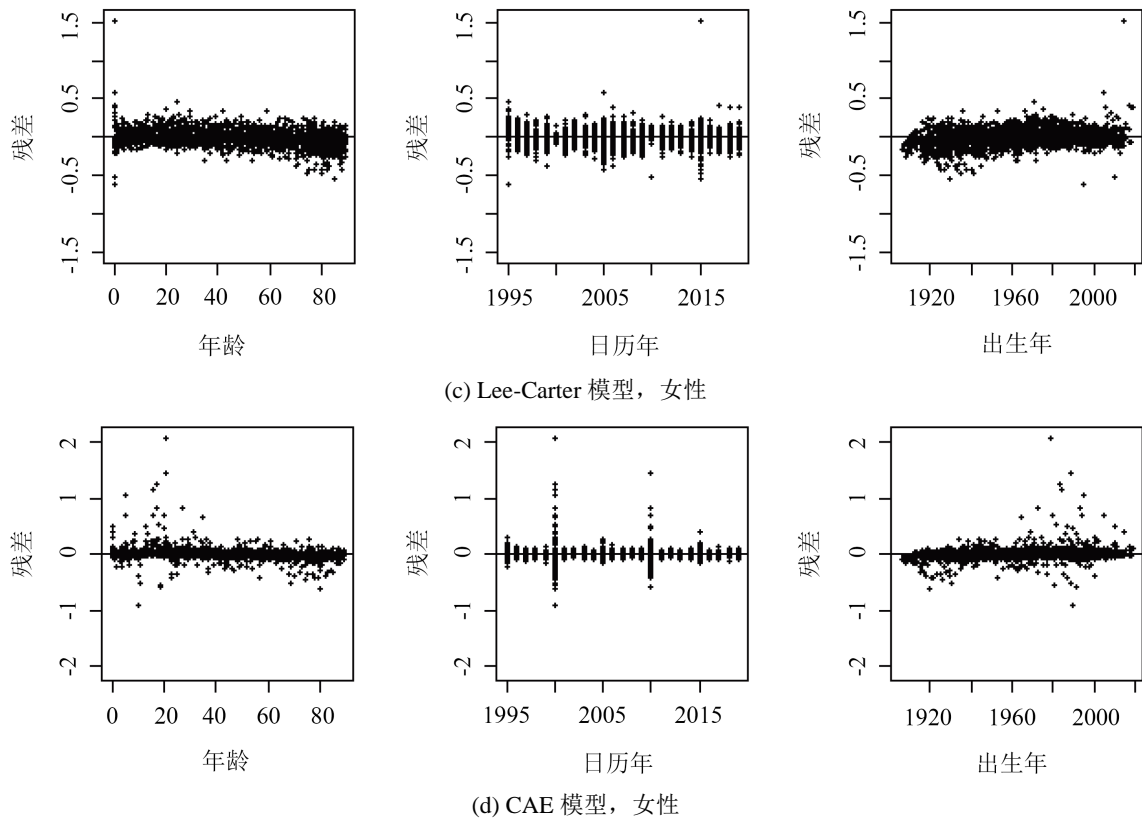


Figure 12. Lee-Carter model, standard residual chart of CAE model for male and female mortality in China
图 12. Lee-Carter 模型, CAE 模型对中国男女性死亡率的标准残差图

如下图 13 分别画出实际的, Lee-Carter 模型预测的, CAE 模型预测的男性和女性的死亡率。红色是男性, 蓝色是女性。由图可知男性死亡率普遍高于女性死亡率。

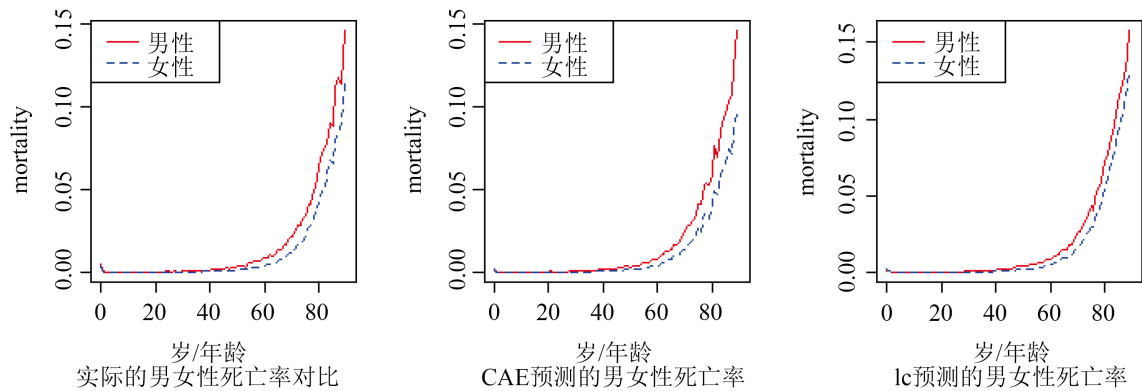


Figure 13. Actual, Lee-Carter, CAE model mortality values
图 13. 实际, Lee-Carter, CAE 模型下死亡率值

继续使用 AFE (绝对拟合误差)来比较这 2 个模型对数据的拟合效果, AFE 的计算公式为:

$$AFE_i = \sum_{x,t} |m_{x,t,i} - \hat{m}_{x,t,i}|$$
。由公式得到如下表 2, 由表 2, CAE 模型对中国男性和女性的绝对拟合误差, 都小于 LC 模型对中国男性和女性的绝对拟合误差。所以, CAE 模型对中国男女性数据的拟合效果良好。

Table 2. AFE values of Chinese male and female data processed by the two models
表 2. 两模型处理中国男女性数据的 AFE 值

数据对象	lc, 男性	lc, 女性	cae, 男性	cae, 女性
AFE	7.503822	5.671689	2.722456	2.56403

再研究两模型对中国男女性死亡率数据的预测效果。分性别画出 2015 年男女性实际, CAE, Lee-Carter 预测的男女性死亡率。如下图 14 所示。可知对男女性死亡率的预测效果都比较好, 其中 LC 模型对男女性死亡率预测有偏大的倾向, 相比之下, 除了某些超高年龄的死亡率值外, CAE 模型的预测值和实际值比较接近。可知 CAE 模型对中国男女性数据的拟合和预测效果良好。

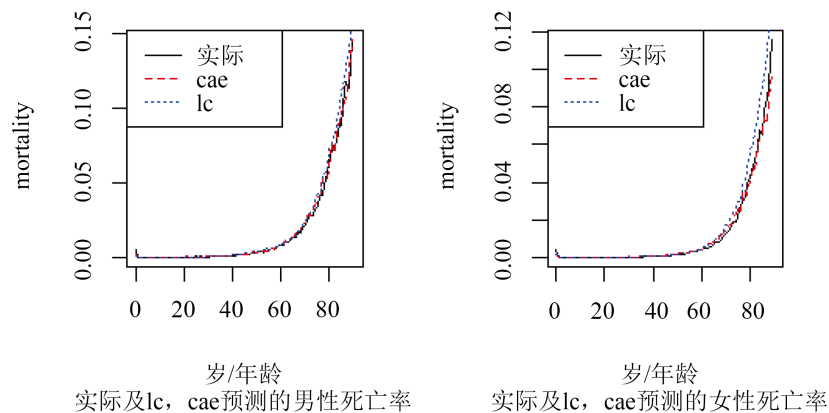
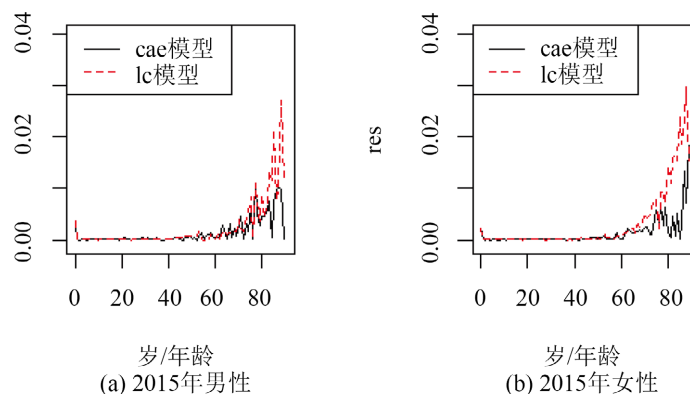


Figure 14. Actual male and female mortality rates, Lee-Carter, CAE, 2015

图 14. 2015 年男女性实际, Lee-Carter, CAE 预测的男女性死亡率

接下来分别计算了男性, 女性死亡率在两个模型的死亡率预测残差的绝对值, 并分别作出了男女性在 2014 年和 2019 年的死亡率残差绝对值图, 如下图 14 所示。黑色线是 CAE 模型, 红色线是 lc 模型。由图 15 可知, 在预测时间区间的两端, lc 模型的残差较小, 在预测区间的中间, CAE 模型的残差较小。这个可能是由于处理的对象比较少, 提取的共同年龄因子也比较小的缘故, 当处理对象增多时, 效果应该会得到改善。而且 LC 模型对较高年龄段的预测有待改善。从图中可以看出, 两个模型预测的残差值在 60 岁之前普遍比较稳定, 在 60 岁之后出现波动, 到了高龄波动更大。CAE 模型的残差在 60 岁之后普遍落到 lc 模型之下, 距离 0 值更近一些。即预测残差相对较低, 所以 CAE 模型对中国男女性死亡率数据的预测效果是比较好的。



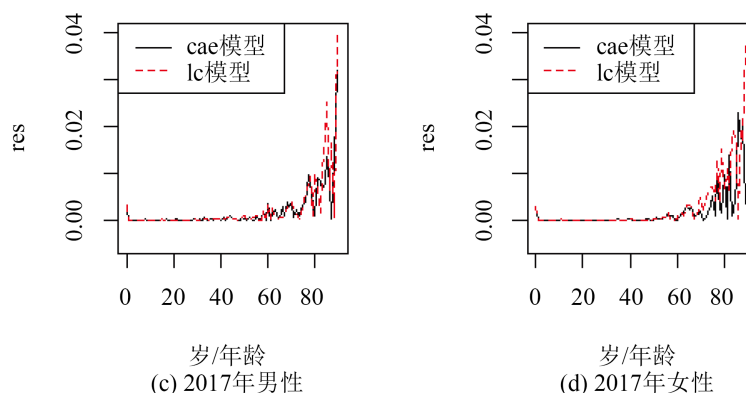


Figure 15. Prediction residuals based on male and female mortality in CAE model and LC model

图 15. CAE 模型和 lc 模型基于男女性死亡率的预测残差

4. 总结

本文首先采用欧洲的 Austria、Denmark、France、UK、Sweden 五个国家的死亡率数据，检验共同年龄效应模型。发现其对欧洲五国的数据拟合和预测效果良好。然后，首次将该模型运用到中国男女性死亡率数据，提取中国男女性死亡率数据的共同因子。分别用 Lee-Carter 模型的 WLS (最小二乘估计) 和 CAE 模型的 MLE (极大似然估计) 以及牛顿迭代法进行了参数估计，并预测死亡率。由死亡率对照图、标准残差图、绝对拟合误差值、残差图，可以发现 CAE 模型对中国的死亡率数据处理更加集中，效果更好。随着具有共同年龄效应的数据对象的增加，CAE 模型的效果也会越来越好。还可以尝试给 CAE 模型添加队列效应，或者研究省份、镇、县之间的死亡率。死亡率下降带来的长寿风险给我国的基本养老体系和保险公司等带来了很大的损失，因此有必要提高死亡率的预测精度，帮助保险公司、政府、企业等减少长寿风险带来的损失。

基金项目

国家自然科学基金资助项目：随机动态死亡率模型的统计性质及应用研究(12061066)，甘肃省自然科学基金资助项目：高龄动态随机死亡率模型的构建及其应用研究(20JR5RA528)。

参考文献

- [1] 国务院第七次全国人口普查领导小组办公室，第七次全国人口普查公报(第五号)——人口年龄构成情况 [EB/OL]. http://www.stats.gov.cn/zjtj/zdtjgz/zgrkpc/dqcrkpc/ggl/202105/t20210519_1817698.html, 2021-05-11.
- [2] Lee, R.D. and Carter, L.R. (1992) Modeling and Forecasting US Mortality. *Journal of American Statistical Association*, **87**, 659-675. <https://doi.org/10.2307/2290201>
- [3] Currie, I.D., Durban, M. and Eilers, P.H.C. (2006) Generalized Linear Array Models with Applications to Multidimensional Smoothing. *Journal of the Royal Statistical Society*, **68**, 259-280. <https://doi.org/10.1111/j.1467-9868.2006.00543.x>
- [4] Cairns, A.J.G., Blake, D. and Dowd, K. (2006) A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration. *Journal of Risk and Insurance*, **73**, 687-718. <https://doi.org/10.1111/j.1539-6975.2006.00195.x>
- [5] Carter, L.R. and Lee, R.D. (1992) Modeling and Forecasting US Sex Differentials in Mortality. *International Journal of Forecasting*, **8**, 393-411. [https://doi.org/10.1016/0169-2070\(92\)90055-E](https://doi.org/10.1016/0169-2070(92)90055-E)
- [6] Kleinow, T. (2015) A Common Age Effect Model for the Mortality of Multiple Populations. *Mathematics and Economics*, **63**, 147-152. <https://doi.org/10.1016/j.insmatheco.2015.03.023>
- [7] Kang, M., Liu, Y., Li, J.S.-H. and Chan, W.-S. (2018) Mortality Forecasting for Multiple Populations: An Augmented

-
- Common Factor Model with a Penalized Log-Likelihood. *Communications in Statistics: Case Studies, Data Analysis and Applications*, **4**, 118-141. <https://doi.org/10.1080/23737484.2018.1529546>
- [8] 李志生, 刘恒甲. Lee-Carter 死亡率模型的估计与应用——基于中国人口数据的分析[J]. 中国人口科学, 2010(3): 46-56+111.
- [9] 赵明, 王晓军. 多人口 Lee-Carter 随机死亡率模型比较与中国应用[J]. 中国人口科学, 2020(2): 81-96+128.
- [10] 赵明, 王晓军. 多人口随机死亡率模型研究: 理论方法与进展综述[J]. 统计研究, 2020, 37(7): 30-41.
- [11] 曹园. 基于 Lee-Cater 模型的我国死亡率预测[J]. 统计与决策, 2018, 34(9): 32-36.
- [12] 肖鸿民, 马海飞, 康彦玲. 两种死亡率预测方法的比较[J]. 统计与决策, 2020, 36(23): 5-8.
- [13] Enchev, V., Kleinow, T. and Cairns, A.J.G. (2017) Multi-Population Mortality Models: Fitting, Forecasting and Comparisons. *Scandinavian Actuarial Journal*, **4**, 319-342. <https://doi.org/10.1080/03461238.2015.1133450>
- [14] Simon, S., Torsten, K. and Ralf, K. (2021) Clustering-Based Extensions of the Common Age Effect Multi-Population Mortality Model. *Risks*, **9**, 45. <https://doi.org/10.3390/risks9030045>