**Pure Mathematics** 理论数学, 2022, 12(7), 1242-1261
Published Online July 2022 in Hans. http://www.hanspub.org/journal/pm
https://doi.org/10.12677/pm.2022.127136

Hans汉斯

# 基于模糊聚类的多人口死亡率模型研究

**李芳芳[1]，肖鸿民[1*]，李　祥[1]，康春明[2]**

[1]西北师范大学数学与统计学院，甘肃　兰州
[2]西北师范大学计算机科学与工程学院，甘肃　兰州

收稿日期：2022年6月19日；录用日期：2022年7月21日；发布日期：2022年7月28日

## 摘　要

**面对严峻的全球养老局势，单独研究某个国家的死亡率效果往往欠佳，我们需要建立多人口死亡率模型来拟合和估计死亡率。研究多人口共同因子死亡率模型，可以更好地抓住被研究人口的共性和差异。由于大部分发达国家比我国的老龄化程度更深，本文由死亡率相互独立的Individual Lee-Carter模型提取14个国家和地区的年龄和时期效应，由三种聚类方法分别可视化。将共同年龄效应模型和联合K模型与模糊k-Means聚类和队列效应结合，构建四个新模型：模糊共同年龄效应模型、模糊联合K模型、r共同年龄效应模型和r联合K模型。数值结果表明与模糊k-Means聚类结合的模型，拟合和预测效果都很好。添加了队列效应的模型，受模型和出生年份限制，效果欠佳。所以共同因子死亡率模型可以提高预测精度和提供建设性建议。本文是利用机器学习研究多人口死亡率的有益尝试。**

## 关键词

**多人口，共同年龄效应模型，联合K模型，模糊聚类，队列效应**

# Multi-Population Mortality Model Based on Fuzzy Clustering

**Fangfang Li[1], Hongmin Xiao[1*], Xiang Li[1], Chunming Kang[2]**

[1]College of Mathematics and Statistics, Northwest Normal University, Lanzhou Gansu
[2]College of Computer Science and Engineering, Northwest Normal University, Lanzhou Gansu

Received: Jun. 19th, 2022; accepted: Jul. 21st, 2022; published: Jul. 28th, 2022

## Abstract

**In the face of severe global pension situation, the effect of individual study on mortality of a coun-**

*通讯作者。

文章引用: 李芳芳, 肖鸿民, 李祥, 康春明. 基于模糊聚类的多人口死亡率模型研究[J]. 理论数学, 2022, 12(7): 1242-1261. DOI: 10.12677/pm.2022.127136

**try is often not very well, we need to build multi-population mortality model to fit and estimate mortality. The study of multi-population common factor mortality model can better capture the commonness and difference of the studied populations. Most developed countries have a deeper degree of aging than China. In this paper, the Individual Lee-Carter model with independent mortality is used to extract the age and period effects of 14 countries and regions, three clustering methods are used to visualize them respectively. The common age effect model and Joint-k model were combined with fuzzy k-Means clustering and queue effect to construct four new models: fuzzy Common Age Effect model, fuzzy Joint-k model, r-Common Age Effect model and r Joint-k model. The numerical results show that the model combined with fuzzy k-Means clustering has better fitting and prediction effect. The model with cohort effect was not very effective due to the limitation of model and birth year. Therefore, the common factor mortality model can improve the prediction accuracy and provide constructive suggestions for countries and the insurance company. This paper is a useful attempt to study multi-population mortality using machine learning.**

## Keywords

**Multi-Population, Common Age Effect Model, Joint-K Model, Fuzzy Clustering, Effect of the Queue**



## 1. 引言

第七次人口普查结果显示，我国人口保持低速增长态势，人口老龄化程度进一步加深，60 岁及以上人口达到 2.6 亿人，比重达到 18.70% [1]，其中 65 岁及以上人口占比 13.5% [1]。我国也积极出台相关服务养老的政策，为老有所养做准备。目前大部分死亡率模型对死亡率的估计偏高，给人寿年金计算和长寿风险管理带来困难，导致更大的养老金缺口。分析多个国家数据，建立多人口死亡率模型，可解决单人口模型预测中的不合理交叉或偏离现象，实现更有效的长期预测，降低政府、保险公司等要遭受的长寿风险。我们研究多个国家的共同因子死亡率模型，有利于掌握死亡率之间的共性和联系，提高预测精度,更好得管理长寿风险。

在国外，最早的动态死亡率模型是 Li 和 Lee (1992) [2]提出的 Lee-Carter (LC)模型，其将死亡率分解为年龄和时期因子。基于 LC 模型，为处理不同种群的数据，假设各个时间序列相互独立，产生了独立的 Lee-Carter (ILC)模型；Carter 和 Lee (1992) [3]首次提出多人口建模的思想，假设多个群体的死亡率具有共同时间因子，建立多人口的联合 K (Jiont-k)模型；Li 和 Lee (2005) [4]改进 Jiont-k 模型，提出增强共同因子(ACF)模型；Renshaw 和 Haberman (2006) [5]发现死亡率与人口出生年份有关，提出了考虑出生年份效应的 Renshaw-Haberman (RH)模型；Currie 等(2006) [6]对美国数据拟合时为避免 RH 模型中由参数过多出现的稳定性问题，提出简化形式即年龄－时期－队列(APC)模型；Kleinow (2015) [7]考虑了所有人口共享年龄参数的 ILC 模型的特殊情形，提出了共同年龄效应(CAE, Common Age Effect model)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中；Hatzopoulos 和 Haberman (2013) [8]应用模糊聚类技术，分析了 35 个国家时期效应的相似性，构建一个公共的年龄－时期效应模型(Common Age-Period Effect model)；Danesi，Haberman (2015) [9]研究了聚类分析在死亡率建模中的进一步运用，描述了总体死亡率随时间发展的时期效应；Enchev 等(2017) [10]用 CAE 模型、ACF 模型及其两个参数缩减模型进行了联合建模和对比，发现 CAE 模型更适合欧洲几个国家的死亡率数据；Wen 等(2021) [11]用 12 个多人

口死亡率模型拟合英格兰 10 个不同的社会－经济群体，发现死亡率同经济水平有关；Schnürch 和 Kleinow (2021) [12]等基于多种共同年龄效应的假设，给出 CAE 模型基于聚类的四种变体，发现 fuzzyCAE 模型比标准聚类方法更灵活且更易聚类。

在国内，赵明等(2020) [13]基于中国数据对 Jiont-k 模型、ACF 模型及其两个扩展模型进行对比，检验多人口模型在中国的适用情况，发现添加时期效应的多人口随机死亡率模型能得到一致的死亡率预测值，结果更稳健；王晓军、路倩(2020) [14]对死亡率预测模型进行了回顾和梳理，包括死亡率改善模型、多人口死亡率的建模和预测，发现拟合和预测准确度不断提升，提倡运用新兴统计学方法加强相关研究；赵明、王晓军(2020) [15]梳理了国内外多人口随机死亡率模型，总结了多人口建模方法在我国人口死亡率模型发展中的意义，发现该模型可为各种国家和地区提供可行的人口死亡率预测方案；赵明(2022) [16]采用 Lee-Carter 模型和贝叶斯分层模型动态预测男性人口死亡率，结果表明 Lee-Carter 适合短期预测；贝叶斯模型适合长期预测；肖鸿民等(2021) [17]基于中国男女性死亡率数据，分别使用 CAE 模型 Individual Lee-Carter 模型对我国男女性及欧洲五国死亡率数据进行了处理，结果表明 CAE 模型效果更好。

本文基于 ILC 模型，共同年龄效应模型(CAE)和联合 K 模型(Jiont-k)，构建四个新模型并进行了相应的数值分析，得到了一些启发性结论。第二部分详细介绍了这些多人口模型的定义及参数估计方法；第三部分介绍基于聚类的多人口共同因子死亡率模型的数值模拟过程，包括 fuzzy CAE 模型、fuzzy Jiont-k 模型。结果表明，基于聚类的模型的拟合和预测效果显著优于原模型，fuzzy CAE 显著优于 Kleinow (2015) [7]提出的 CAE 模型；第四部分介绍了部分年份加入队列效应的多人口共同因子死亡率模型的数值模拟过程，包括 rCAE 模型和 rJiont-k 模型数值比对结果显示，加入部分队列效应的模型的效果没有原模型好，需继续改进；第五部分，简要总结我们的工作，提供改善模型的思路。

## 2. 模型介绍

### 2.1. Individual Lee-Carter 模型

基于经典 Lee-Carter 模型(式(1))，产生了一种各种群死亡率相互独立的 Lee-Carter 模型(ILC 模型，Individual Lee-Carter)。ILC 模型是一种多种群死亡率建模方法，分别用 Lee-Carter 模型描述各个种群的死亡率，假设当 $i=1,2,\cdots,I$ 时，$k_{i,t}$ 时间序列是独立的，没有共享参数，模型表达式如式(2)。用 ILC 模型提取各个种群的 $\alpha_{x,i},\beta_{x,i},k_{i,t}$ 值，由加权最小二乘法估计参数。

$$\ln\left(m_{x,t}\right)=\alpha_x+\beta_x k_t+\varepsilon_{x,t} \tag{1}$$

$$\ln\left(m_{x,t,i}\right)=\alpha_{x,i}+\beta_{x,i}k_{i,t}+\varepsilon_{x,t,i} \tag{2}$$

首先最小化下式：

$$\sum_{x=1}^{X}\sum_{t=1}^{T1}\sum_{i=1}^{14}\omega_{x,t}\left(f_{x,t,i}-\hat{\alpha}_{x,i}-\hat{\beta}_{x,i}\hat{k}_{i,t}\right)^2 \tag{3}$$

其中，$f_{x,t,i}=\ln\left(m_{x,t,i}\right)=\alpha_{x,i}+\beta_{x,i}k_{i,t}+\varepsilon_{x,t,i}$，$\omega_{x,t,i}=d_{x,t,i}$ 为在 $t$ 年 $x$ 岁的人的死亡数，年龄数 $x\in\{0,1,2,\cdots,X\}$，年份数 $t\in\{1,2,\cdots,T1,\cdots,T\}$，国家或地区数 $i\in\{1,2,\cdots,I\}$，聚类的类别数 $l\in\{1,2,\cdots,k\}$，$\alpha_{x,i}$ 是第 $i$ 个人群的附加年龄因子，表示第 $i$ 个人群对数死亡率的均值；$k_{i,t}$ 反映第 $i$ 个人群对数死亡率随时期的变化趋势；$\beta_{x,i}$ 是年龄因子，代表年龄因素对 $k_{i,t}$ 的敏感程度，满足约束条件：$\sum_x\beta_{x,i}=1$，$\sum_t k_{i,t}=0$，$\sum_i k_{i,t}=0$。最小化式(3)，得式(4)的参数估计值，继续迭代直至满足终止条件，对得到的参数标准化：

$$\hat{\alpha}_{x,i}=\frac{\sum_t \omega_{x,t,i}\left(f_{x,t,i}-\hat{\beta}_{x,i}\hat{k}_{i,t}\right)}{\sum_t \omega_{x,t,i}}$$
$$\hat{\beta}_{x,i}=\frac{\sum_t \omega_{x,t,i}\left(f_{x,t,i}-\hat{\alpha}_{x,i}\right)\hat{k}_{i,t}}{\sum_t \omega_{x,t,i}\hat{k}_{i,t}^2} \tag{4}$$
$$\hat{k}_{t,i}=\frac{\sum_t \omega_{x,t,i}\left(f_{x,t,i}-\hat{\alpha}_{x,i}\right)\hat{\beta}_{x,i}}{\sum_t \omega_{x,t,i}\hat{\beta}_{x,i}^2}$$

## 2.2. 共同年龄效应模型类(CAE 模型类)

### 2.2.1. 共同年龄效应模型(CAE 模型)

Kleinow (2015)提出了共同年龄效应模型，发现一般的年龄效应可以推广到多种群随机死亡率模型中，如式(5)：

$$\ln\left(m_{x,t,i}\right)=\alpha_{x,i}+B_x^{(1)}k_{i,t}^{(1)}+B_x^{(2)}k_{i,t}^{(2)}+\varepsilon_{x,t,i} \tag{5}$$

年龄数 $x\in\{0,1,2,\cdots,X\}$，年份数 $t\in\{1,2,\cdots,T1,\cdots,T\}$，国家或地区数 $i\in\{1,2,\cdots,I\}$，聚类的类别 $l\in\{1,2,\cdots,k\}$，$\alpha_{x,i}$ 是第 $i$ 个人群的附加年龄因子，表示第 $i$ 个人群对数死亡率的均值；$k_{i,t}^{(1)}$ 和 $k_{i,t}^{(2)}$ 反映第 $i$ 个人群对数死亡率随时期的变化趋势；$B_x^{(1)}$ 和 $B_x^{(2)}$ 是共同年龄因子，代表共同年龄因素对 $k_{i,t}^{(1)}$ 和 $k_{i,t}^{(2)}$ 的敏感程度，也代表所有数据的共同年龄效应。为满足参数估计结果的一致性，需满足以下约束条件：$\sum_x B_x^{(1)}=1$，$\sum_x B_x^{(2)}=1$，$\sum_x B_x^{(2)}=1$，$\sum_t k_{i,t}^{(2)}=0$，$\sum_i k_{i,t}^{(2)}=0$。

也可尝试只含有 1 个 $B_x$ 和 $k_{i,t}$ 的 CAE 模型，如式(6)：

$$\ln\left(m_{x,t,i}\right)=\alpha_{x,i}+B_x k_{i,t}+\varepsilon_{x,t,i} \tag{6}$$

### 2.2.2. 模糊共同年龄效应模型(Fuzzy CAE)

由 ILC 模型可以得到每个种群的 $\alpha_{x,i},\beta_{x,i},k_{i,t}$ 估计值。基于 Simon Schnurch，TorstenKleinow 等(2021) [12]提出的模糊极大似然聚类，由模糊 k-Means 聚类算法将年龄效应 $\beta_{x,i}$ 聚为 $k$ 类，可得隶属度矩阵 $\boldsymbol{U}$，$\boldsymbol{U}$ 表示的是人口 $i$ 隶属于簇 l 的概率。将 $\boldsymbol{U}$ 和只含有一个 $B_x$ 和 $k_{i,t}$ 的 CAE 模型结合，新模型表达式如式(7)：

$$\ln\left(m_{x,t,i}\right)=\alpha_{x,i}+\left(\sum_{l=1}^{k}u_{i,l}B_{x,l}\right)k_{i,t} \tag{7}$$

需满足约束条件 $\sum_x B_{x,l}=1$，$\sum_t k_{i,t}=0$，$\sum_l u_{i,l}=1$。由极大似然法估计参数，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 $D_{x,t,i}\sim Poisson\left(m_{x,t,i}E_{x,t,i}\right)$。

$m_{x,t,i}=\exp\left(\alpha_{x,i}+\left(\sum_{l=1}^{k}u_{i,l}B_{x,l}\right)k_{i,t}\right)$，似然函数如(8)式，取对数，得对数似然函数如式(9),其中 $c$ 为常数。

$$l\left(\alpha_{x,i},B_{x,i},k_{t,i}\right)=\Pi_{x,t,i}\frac{\left(E_{x,t,i}m_{x,t,i}\right)^{D_{x,t,i}}}{D_{x,t,i}!}\exp\left(-E_{x,t,i}m_{x,t,i}\right) \tag{8}$$

$$L=L\left(\alpha_{x,i},B_{x,i},k_{i,t}\right)=\sum_{x,t,i}\left[D_{x,t,i}\left(\alpha_{x,i}+\left(\sum_{l=1}^{k}u_{i,l}B_{x,l}\right)k_{i,t}\right)-E_{x,t,i}\exp\left(\alpha_{x,i}+\left(\sum_{l=1}^{k}u_{i,l}B_{x,l}\right)k_{i,t}\right)\right]+c \tag{9}$$

使对数似然函数达到极大，结合牛顿迭代 $\hat{\theta}^{(n+1)}=\hat{\theta}^{(n)}-\dfrac{\partial L^{(n)}/\partial\theta}{\partial^2 L^{(n)}/\partial\theta^2}$，设置初始值为：$\hat{\alpha}_{x,i}=\dfrac{1}{T1}\sum_{t=1}^{T1}\ln\left(\dfrac{D_{x,t,i}}{E_{x,t,i}}\right)$，$B_{x,i}=0.1$，$k_{i,t}=0.1$。得参数估计式：

$$\begin{aligned}
\hat{\alpha}_{x,i}&=\hat{\alpha}_{x,i}-\frac{\sum_t\left(D_{x,t,i}-E_{x,t,i}\exp\left(\hat{\alpha}_{x,i}+\left(\sum_{l=1}^{k}u_{i,l}\hat{B}_{x,l}\right)\hat{k}_{i,t}\right)\right)}{\sum_t E_{x,t,i}\exp\left(\hat{\alpha}_{x,i}+\left(\sum_{l=1}^{k}u_{i,l}\hat{B}_{x,l}\right)\hat{k}_{i,t}\right)}\\
\hat{B}_{x,l}&=\hat{B}_{x,l}-\frac{\sum_t\left(D_{x,t,i}-E_{x,t,i}\exp\left(\hat{\alpha}_{x,i}+\left(\sum_{l=1}^{k}u_{i,l}\hat{B}_{x,l}\right)\hat{k}_{i,t}\right)\right)\hat{k}_{i,t}}{\sum_t E_{x,t,i}\exp\left(\hat{\alpha}_{x,i}+\left(\sum_{l=1}^{k}u_{i,l}\hat{B}_{x,l}\right)\hat{k}_{i,t}\right)\hat{k}_{i,t}^2}\\
\hat{k}_{i,t}&=\hat{k}_{i,t}-\frac{\sum_t\left(D_{x,t,i}-E_{x,t,i}\exp\left(\hat{\alpha}_{x,i}+\left(\sum_{l=1}^{k}u_{i,l}\hat{B}_{x,l}\right)\hat{k}_{i,t}\right)\right)\hat{B}_{x,l}}{\sum_t E_{x,t,i}\exp\left(\hat{\alpha}_{x,i}+\left(\sum_{l=1}^{k}u_{i,l}\hat{B}_{x,l}\right)\hat{k}_{i,t}\right)\hat{B}_{x,l}^2}
\end{aligned}\tag{10}$$

达到终止条件时，结束循环。

#### 2.2.3. 加入队列效应的 CAE 模型(rCAE)

年龄效应聚类的结果一直保持一致的国家或地区，肯定存在共性。尝试用添加了队列效应的 CAE 模型处理这些数据，得模型表达式如式(11)：

$$\ln m_{x,t,i}=\alpha_{x,i}+B_x k_{i,t}+r_{i,c}+\varepsilon_{x,t,i}\tag{11}$$

其中，$c=(t-x)$ 表示出生年份。部分出生年不在数据研究范围内，为避免空值导致计算出错，由最小二乘估计法估计参数。

$$\sum_{x=1}^{X}\sum_{t=1}^{T1}\sum_{i=1}^{I}\omega_{x,t,i}\left(f_{x,t,i}-\hat{\alpha}_{x,i}-\hat{B}_x\hat{k}_{i,t}-\hat{r}_{i,c}\right)^2\tag{12}$$

该模型满足约束条件：$\sum_t k_{i,t}=0$，$\sum_x B_x=1$，$\sum_c r_{i,c}=0$，$\sum_c c r_{i,c}=0$。最小化(12)式，得参数估计式(13)，继续迭代，直至满足终止条件。最后对参数结果标准化。

$$\begin{aligned}
\hat{\alpha}_{x,i}&=\frac{\sum_t\omega_{x,t,i}\left(f_{x,t,i}-\hat{B}_x\hat{k}_{i,t}-\hat{r}_{i,c}\right)}{\sum_t\omega_{x,t,i}},\quad \hat{B}_x=\frac{\sum_t\omega_{x,t,i}\left(f_{x,t,i}-\hat{\alpha}_{x,i}-\hat{r}_{i,c}\right)\hat{k}_{i,t}}{\sum_t\omega_{x,t,i}\hat{k}_{i,t}^2}\\
\hat{k}_{i,t}&=\frac{\sum_t\omega_{x,t,i}\left(f_{x,t,i}-\hat{\alpha}_{x,i}-\hat{r}_{i,c}\right)\hat{B}_x}{\sum_t\omega_{x,t,i}\hat{B}_x^2},\quad \hat{r}_{i,c}=\frac{\sum_c\omega_{x,t,i}\left(f_{x,t,i}-\hat{\alpha}_{x,i}-\hat{B}_x\hat{k}_{i,t}\right)}{\sum_c\omega_{x,t,i}}
\end{aligned}\tag{13}$$

### 2.3. 联合(Jiont-k)模型及其扩展模型

#### 2.3.1. 联合 K (Jiont-k)模型

Carter 和 Lee (1992)提出了多人口的的联合 K (Jiont-k)模型，假设多个群体的死亡率具有共同时期因子，具体形式如式(14)

$$\ln m_{x,t,i} = \alpha_{x,i} + \beta_{x,i} K_t + \varepsilon_{x,t,i} \tag{14}$$

年龄数 $x \in \{0,1,2,\cdots,X\}$，年份数 $t \in \{1,2,\cdots,T1,\cdots,T\}$，国家或地区数 $i \in \{1,2,\cdots,I\}$。$K_t$ 是共同时间因子，反映了第 $i$ 个人口对数死亡率随时期变化的共同趋势；为满足参数估计结果的唯一性，需满足以下约束条件：$\sum_x \beta_{x,i} = 1$，$\sum_t K_t = 0$。

### 2.3.2. 模糊联合 K (Fuzzy Jiont-k)模型

同模糊 CAE 模型，使用模糊 k-Means 聚类算法将时期效应 $k_{i,t}$ 聚为 $k$ 类，可得隶属度矩阵 $\boldsymbol{U}$，$\boldsymbol{U}$ 表示的是人口 $i$ 隶属于簇 l 的概率。将 $\boldsymbol{U}$ 和只含有一个 $\beta_{x,i}$ 和 $K_t$ 的 Jiont-k 模型结合，产生表达式如式(15)，需满足约束条件：$\sum_x \beta_{x,i} = 1$，$\sum_t K_{l,t} = 0$，$\sum_l u_{i,l} = 1$。

$$\ln m_{x,t,i} = \alpha_{x,i} + \sum_{l=1}^{k} \left( u_{i,l} \beta_{x,i} K_{l,t} \right) + \varepsilon_{x,t,i} \tag{15}$$

由极大似然法估计参数。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 $D_{x,t,i} \sim Poisson\left(m_{x,t,i} E_{x,t,i}\right)$。假设每一年龄死亡率都相互独立，似然函数是：

$$l\left(\alpha_{x,i}, B_{x,i}, k_{t,i}\right) = \Pi_{x,t,i} \frac{\left(E_{x,t,i} m_{x,t,i}\right)^{D_{x,t,i}}}{D_{x,t,i}!} \exp\left(-E_{x,t,i} m_{x,t,i}\right)$$

取对数，得如式(16)的极大似然函数，其中 $c$ 为常数。

$$L = L\left(\alpha_{x,i}, \beta_{x,i}, K_{l,t}\right) = \sum_{x,t,i} \left[ D_{x,t,i} \left( \alpha_{x,i} + \sum_{l=1}^{k} \left( u_{i,l} \beta_{x,i} K_{l,t} \right) \right) - E_{x,t,i} \exp\left( \alpha_{x,i} + \sum_{l=1}^{k} \left( u_{i,l} \beta_{x,i} K_{l,t} \right) \right) \right] + c \tag{16}$$

使对数似然函数达到极大，结合牛顿迭代公式 $\hat{\theta}^{(n+1)} = \hat{\theta}^{(n)} - \frac{\partial L^{(n)} / \partial \theta}{\partial^2 L^{(n)} / \partial \theta^2}$，设置初始值为：

$\hat{\alpha}_{x,i} = \frac{1}{T1} \sum_{t=1}^{T1} \ln\left( \frac{D_{x,t,i}}{E_{x,t,i}} \right)$，$b_{x,i} = 0.1$，$K_{l,t} = 0.1$。可得参数估计式(17)：

$$\begin{aligned}
\hat{\alpha}_{x,i} &= \hat{\alpha}_{x,i} - \frac{\sum_t \left( D_{x,t,i} - E_{x,t,i} \exp\left( \hat{\alpha}_{x,i} + \sum_{l=1}^{k} \left( u_{i,l} \hat{\beta}_{x,i} \hat{K}_{l,t} \right) \right) \right)}{\sum_t E_{x,t,i} \exp\left( \hat{\alpha}_{x,i} + \sum_{l=1}^{k} \left( u_{i,l} \hat{\beta}_{x,i} \hat{K}_{l,t} \right) \right)} \\
\hat{\beta}_{x,i} &= \hat{\beta}_{x,i} - \frac{\sum_t \left( D_{x,t,i} - E_{x,t,i} \exp\left( \hat{\alpha}_{x,i} + \sum_{l=1}^{k} \left( u_{i,l} \hat{\beta}_{x,i} \hat{K}_{l,t} \right) \right) \right) \hat{K}_{l,t}}{\sum_t E_{x,t,i} \exp\left( \hat{\alpha}_{x,i} + \sum_{l=1}^{k} \left( u_{i,l} \hat{\beta}_{x,i} \hat{K}_{l,t} \right) \right) \hat{K}_{l,t}^2} \\
\hat{K}_{l,t} &= \hat{K}_{l,t} - \frac{\sum_t \left( D_{x,t,i} - E_{x,t,i} \exp\left( \hat{\alpha}_{x,i} + \sum_{l=1}^{k} \left( u_{i,l} \hat{\beta}_{x,i} \hat{K}_{l,t} \right) \right) \right) \hat{\beta}_{x,i}}{\sum_t E_{x,t,i} \exp\left( \hat{\alpha}_{x,i} + \sum_{l=1}^{k} \left( u_{i,l} \hat{\beta}_{x,i} \hat{K}_{l,t} \right) \right) \hat{\beta}_{x,i}^2}
\end{aligned} \tag{17}$$

满足终止条件时，停止迭代。

### 2.3.3. 加入队列效应的 Jiont-k 模型(r Jiont-k)

对时期效应的聚类结果一直保持一致的国家或地区，用添加了队列效应的 Jiont-k 模型处理，$c = (t - x)$ 表示出生年，模型表达式如式(18)：

$$\ln m_{x,t,i} = \alpha_{x,i} + \beta_{x,i} K_t + r_{i,c} + \varepsilon_{x,t,i} \tag{18}$$

部分出生年不在数据研究的范围内，为避免出现空值，由最小二乘法估计参数。最小化式(19)：

$$\sum_{x=1}^{X}\sum_{t=1}^{T1}\sum_{i=1}^{I} \omega_{x,t,i}\left(f_{x,t,i} - \hat{\alpha}_{x,i} - \hat{\beta}_{x,i}\hat{K}_t - \hat{r}_{i,c}\right)^2 \tag{19}$$

其中，$f_{x,t,i} = \ln\left(m_{x,t,i}\right) = \alpha_{x,i} + \beta_{x,i}K_t + r_{i,c} + \varepsilon_{x,t,i}$，$\omega_{x,t,i} = d_{x,t,i}$ 为在 $t$ 年 $x$ 岁的观测死亡数。满足约束条件：$\sum_x \beta_{x,i} = 1$，$\sum_t K_t = 0$，$\sum_c r_{i,c} = 0$，$\sum_c c r_{i,c} = 0$。最小化式(19)，得如式(20)的参数估计值，继续迭代，直至满足约束条件。最后对得到的参数结果标准化。

$$\begin{aligned}
\hat{\alpha}_{x,i} &= \frac{\sum_t \omega_{x,t,i}\left(f_{x,t,i} - \hat{\beta}_{x,i}\hat{K}_t - \hat{r}_{i,c}\right)}{\sum_t \omega_{x,t,i}}, \quad \hat{\beta}_{x,i} = \frac{\sum_t \omega_{x,t,i}\left(f_{x,t,i} - \hat{\alpha}_{x,i} - \hat{r}_{i,c}\right)\hat{K}_t}{\sum_t \omega_{x,t,i}\hat{K}_t^2} \\
\hat{K}_t &= \frac{\sum_t \omega_{x,t,i}\left(f_{x,t,i} - \hat{\alpha}_{x,i} - \hat{r}_{i,c}\right)\hat{\beta}_{x,i}}{\sum_t \omega_{x,t,i}\hat{\beta}_{x,i}^2}, \quad \hat{r}_{i,c} = \frac{\sum_c \omega_{x,t,i}\left(f_{x,t,i} - \hat{\alpha}_{x,i} - \hat{b}_{x,i}\hat{K}_t\right)}{\sum_c \omega_{x,t,i}}
\end{aligned} \tag{20}$$

## 3. 基于聚类的多人口共同因子死亡率模型

将 ILC 模型同 k 均值聚类，层次聚类和模糊 k-Means 聚类结合，对各个国家的年龄和时期效应聚类，将聚类结果与 CAE 模型，Jiont-k 模型结合来改善模型效果。

表 1 的 14 个国家大都是欧洲和亚洲国家，发达国家或发达经济体，由 Wen (2021) [11]等知不同社会经济群体的死亡率会有所不同，故存在公共因子；死亡率数据丰富全面，拟合模型的效果更好。因此研究这 14 个国家和地区的 60~89 岁，1995~2018 年的整体死亡率数据。首先采用 ILC 模型分别提取年龄效应和时期效应并对其聚类。由于 CAE 模型提取共同年龄效应，而 Jiont-k 模型提取共同时期效应，提出基于模糊 k-Means 聚类的 fuzzy CAE 和 fuzzy Jiont-k 模型。

**Table 1.** Researched countries and stats
**表 1.** 研究的国家及所属州

| 序号 | 中文名 | 英文名 | 所属洲 |
|---|---|---|---|
| 1 | 奥地利 | Austria | 欧洲中部 |
| 2 | 加拿大 | Canada | 北美洲北部 |
| 3 | 中国 | China | 亚洲东部 |
| 4 | 丹麦 | Denmark | 北欧 |
| 5 | 法国 | France | 西欧 |
| 6 | 希腊 | Greece | 欧洲 |
| 7 | 中国香港 | HongKong | 东亚，中国南部 |
| 8 | 日本 | Japan | 东亚 |
| 9 | 葡萄牙 | Portugal | 欧洲西南部 |
| 10 | 西班牙 | Spain | 欧洲 |
| 11 | 瑞典 | Sweden | 北欧 |
| 12 | 瑞士 | Switzerland | 中欧 |
| 13 | 中国台湾 | Taiwan | 东亚，中国大陆东南海域 |
| 14 | 英国 | UK | 欧洲西部 |

### 3.1. 基于 CAE 模型的模糊极大似然聚类

基于 CAE 模型的模糊极大似然估计是将 ILC 模型、隶属度矩阵 $\boldsymbol{U}$ 同 CAE 模型结合得到的改进模型。由 ILC 模型提取年龄效应，加权最小二乘法估计参数，图 1 是 ILC 模型的参数估计值。香港年龄效应在 82 岁的时候骤然变大，后逐渐恢复正常。中国的年龄效应和时期效应变动较多，也许与数据采样方式的多样性社会快速发展有关。随着年龄增长，死亡率上升，随着时代发展，死亡率下降，年龄对时期效应的敏感程度都是先增长后下降，符合自然规律。

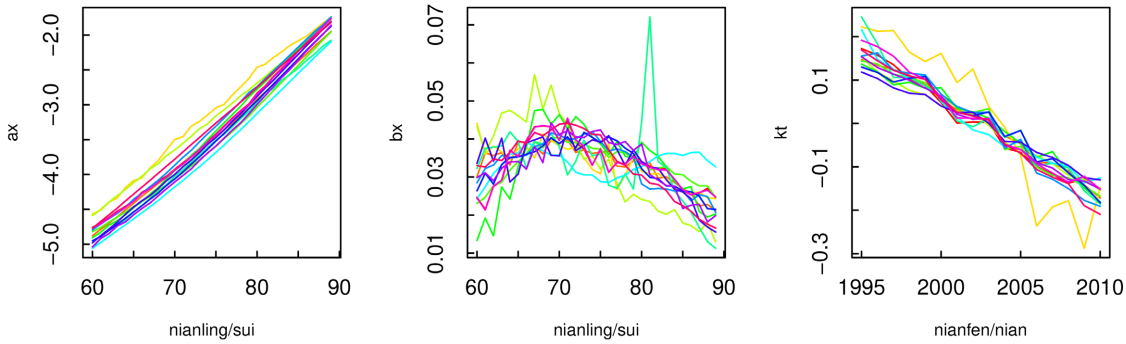


**Figure 1.** Results of parameter estimation obtained by ILC model
**图 1.** ILC 模型得到的参数估计结果

#### 3.1.1. Cmeans 聚类

如下图 2 (左)是基于簇内平方和绘制的簇数选择图，效果较好的簇数是两类，三类或四类。基于 Hartigan-Wong 算法对年龄效应聚类。簇内误差平方和越小，簇间平方和越大，效果越好，between-ss/Total-ss 分别为 25.7%，47.6%，66.6%。聚类地区有限，选取簇数为 3 类。如下图 2 (右)为基于欧式距离聚为四类的可视化展示。

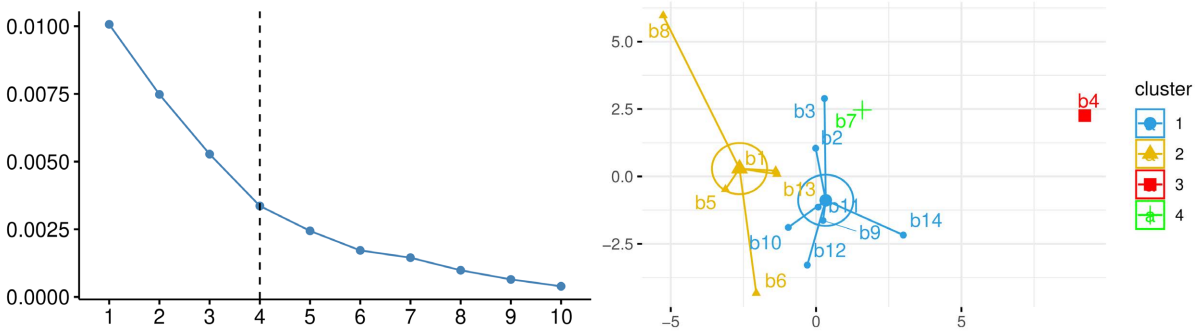


**Figure 2.** Results of C means algorithm clustering into 4 classes
**图 2.** C means 算法聚为 4 类的结果

#### 3.1.2. 层次聚类

由欧式距离，计算 14 个 b[x,i]的相似性，使用 Ward.D2 算法，由层次聚类对年龄效应聚类，切分聚类树，结果如图 3 所示，分别为聚为三类和四类的结果。当聚为三类时，b4 和 b7 分别为组，b14，b12，b9，b10，b3，b2，b11，b8，b6，b5，b1，b13 为一组。对第三类内部分割，b14，b12，b9，b10 为一组，b3，b2，b11 为一组，b8，b6，b5，b1，b13 为一组。由于丹麦死亡率在考虑时间段内停滞不前，将丹麦年龄效应聚为一类；第二簇的中国香港在二战后死亡率下降到了一个非常低的水平，且是世界人均寿命

最高的地区；其他的所有国家为第三簇。

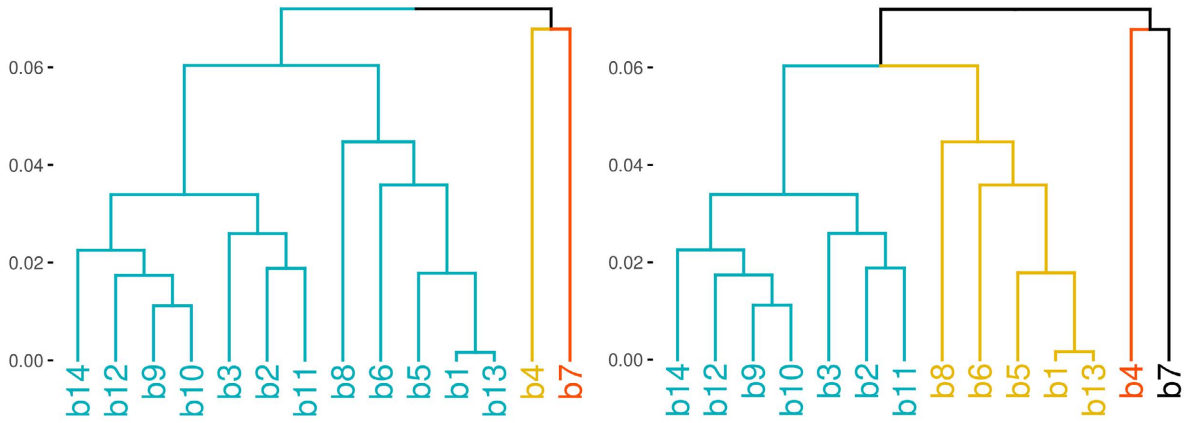


**Figure 3.** The result of grouping the age effect into three and four classes by Hierarchical clustering
**图 3.** 层次聚类将年龄效应聚为三类和四类的结果

表 2 是基于 k 均值聚类，层次聚类和模糊 k-Means 聚类将这 14 个地区的 b[x,i]聚为两类,三类和四类的结果。聚为两类和三类时前两种聚类方法结果一致，模糊 k-Means 聚类结果和前两个聚类方法差别很大。由前两种聚类方法，丹麦被单独分为一组，香港也自成一组，其他所有国家为一组。聚类没有准确得将欧洲，亚洲，美洲的国家分开，知聚类结果不仅同地理位置有关，还和经济发展水平，政策等其他因素相关。

**Table 2.** The clustering results of age effect by three clustering algorithms
**表 2.** 三种聚类算法对年龄效应聚类的结果

| 各地区年龄效应 | b1 | b2 | b3 | b4 | b5 | b6 | b7 | b8 | b9 | b10 | b11 | b12 | b13 | b14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| cmeans 两类 | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| cmeans 三类 | 1 | 1 | 1 | 2 | 1 | 1 | 3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| cmeans 四类 | 1 | 2 | 2 | 3 | 1 | 1 | 4 | 1 | 2 | 2 | 2 | 2 | 1 | 2 |
| Hclust 两类 | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Hclust 三类 | 1 | 1 | 1 | 2 | 1 | 1 | 3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Hclust 四类 | 1 | 2 | 2 | 3 | 1 | 1 | 4 | 1 | 2 | 2 | 2 | 2 | 1 | 2 |
| Fcm 两类 | 1 | 2 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 1 | 2 | 2 | 1 | 2 |
| Fcm 三类 | 1 | 2 | 2 | 2 | 1 | 1 | 2 | 1 | 3 | 3 | 3 | 3 | 1 | 3 |
| Fcm 四类 | 1 | 2 | 2 | 3 | 1 | 1 | 2 | 1 | 4 | 4 | 2 | 4 | 1 | 4 |

### 3.1.3. 模糊共同年龄效应模型(Fuzzy CAE)

Simon Schnurch，TorstenKleinow 等(2021) [12]提出的模糊极大似然聚类，是将模糊 k-Means 聚类算法同极大似然估计结合，估计参数的方法。称为模糊聚类，是因为样本不刚性得属于某一类，而是有属于每个类的概率。每个个体属于所有类的概率和为 1。首先运用 inaparc 包的 kmpp 和 imembrand 函数分析年龄效应，得初始隶属度矩阵 $\boldsymbol{U0}$ 和初始聚类中心 $\boldsymbol{V0}$。基于模糊 k-Means 聚类算法(fcm, fuzzy-k-Means-clustering algrithom)对年龄效应聚类，得到最终隶属度矩阵 $\boldsymbol{U}$ 和聚类中心 $\boldsymbol{V}$。

表 2 是运用模糊 k-Means 聚类算法将年龄效应聚为两类，三类和四类的结果，结果与前两种聚类方法有差异。聚为 3 类时，分别将奥地利，法国，希腊，日本和中国台湾聚为一簇，加拿大，中国，丹麦，中国香港聚为一簇，葡萄牙，西班牙，瑞典，瑞士和英国聚为一簇。第一簇都是属于发达地区和国家；第二簇特殊，丹麦和中国香港有独特的年龄效应，中国大陆人口抽样方式多样，加拿大地理位置特殊，人口基数小；第三簇全为欧洲发达国家，聚类效果较好。

选择聚为三类。图 4 为用模糊聚类将年龄效应聚为 3 类，并将结果转化为 k-Means 聚类结果的展示。

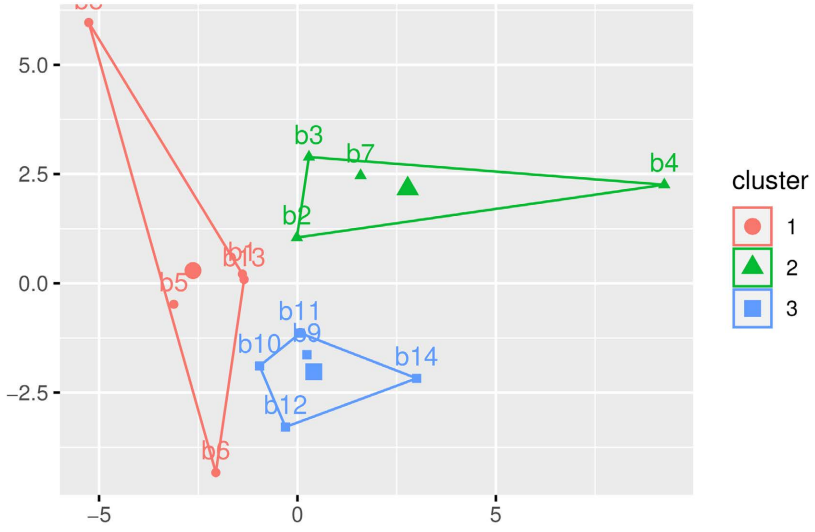


**Figure 4.** Visual display of fuzzy clustering results
**图 4.** 模糊聚类结果的可视化展示

由模糊聚类得到的隶属度矩阵 $\boldsymbol{U}$ 如表 3 所示，把这 14 个地区分为 3 个簇，三个簇内的个数比为 5:4:5，每一列的数字代表年龄效应隶属于每个簇的概率。

**Table 3.** The ultimate membership matrix by fuzzy clustering
**表 3.** 模糊聚类得到的最终隶属度矩阵

| 序号 | Cluster 1 | Cluster 2 | Cluster 3 | 序号 | Cluster 1 | Cluster 2 | Cluster 3 |
|---|---|---|---|---|---|---|---|
| b1 | 0.8931 | 0.0462 | 0.0608 | b8 | 0.4496 | 0.2901 | 0.2602 |
| b2 | 0.1018 | 0.5733 | 0.3249 | b9 | 0.1421 | 0.1896 | 0.6684 |
| b3 | 0.1756 | 0.5492 | 0.2752 | b10 | 0.1856 | 0.1636 | 0.6507 |
| b4 | 0.2477 | 0.4159 | 0.3364 | b11 | 0.1527 | 0.4079 | 0.4394 |
| b5 | 0.6687 | 0.1235 | 0.2079 | b12 | 0.1874 | 0.2123 | 0.6003 |
| b6 | 0.4836 | 0.2154 | 0.3010 | b13 | 0.8870 | 0.0482 | 0.0648 |
| b7 | 0.2777 | 0.4064 | 0.3159 | b14 | 0.1436 | 0.32694 | 0.5295 |

将 $\boldsymbol{U}$ 同只含有一个 $\boldsymbol{B}$ 和 $\boldsymbol{k}$ 的 CAE 模型结合，产生 fuzzy CAE 模型，表达式如(21)式，其中 $u_{i,l}$ 是地区 $i$ 的年龄效应隶属于簇 $l$ 的隶属度，由极大似然法估计参数。1995~2010 年的数据用于拟合，2011~2018 年的数据用于检验。考虑年龄数为 30，拟合年数为 16，用于预测的年数为 8，地区数 $i$ 为 14，簇数 $k$ 为 3。

参数估计结果如图 5，聚类后各参数变化趋势同 ILC 模型结果，群体年龄效应有更多相似点。中国

大陆年龄效应变动较多，聚类后中国台湾有了更大的时期效应，说明所处时期对中国台湾的死亡率及其所属簇的影响增大。

$$\ln\left(m_{x,t,i}\right)=a_{x,i}+\left(\sum_{l=1}^{k}u_{i,l}B_{x,l}\right)k_{i,t} \tag{21}$$

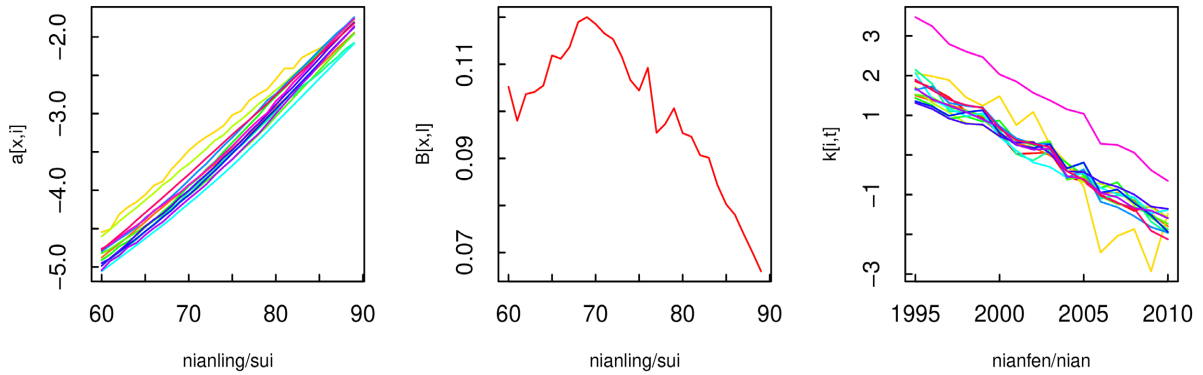


**Figure 5.** Results of parameter estimation by fuzzyCAE model
**图 5.** FuzzyCAE 模型的参数估计结果

图 6 是各个簇的年龄效应，相同簇的年龄效应形状及变化趋势相似，提取共同年龄效应的效果好。

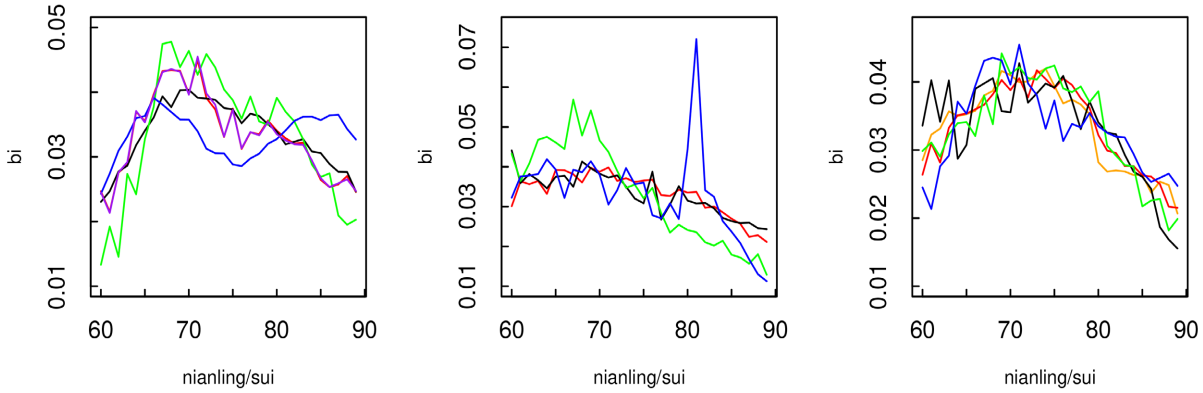


**Figure 6.** Age effects of various clusters in the fuzzyCAE model
**图 6.** FuzzyCAE 模型的各个簇的年龄效应

预测未来 8 年各地区死亡率，采用 zoo，forecast 和 stats 包，由带漂移项的随机游走模型拟合和外推数据，得 $k_{i,t}$ 的预测值。表 4 是最优 ARIMA 序列和 AIC，BIC，$\sigma^2$ 的值，每个序列都带漂移项。AIC 和 BIC 的绝对值都小于 100，拟合效果和模型复杂度较好，$\sigma^2$ 大部分都达到了 10^(−4)，模型较稳定。

**Table 4.** Optimal value of ARIMA sequence, AIC, BIC and sigma^2
**表 4.** 最优 ARIMA 序列，AIC，BIC，sigma^2 值

| | 最优 arima | AIC | BIC | sigma^2 |
|---|---|---|---|---|
| k1arima | (0,1,0) | −73.47 | −72.06 | 0.0004 |
| k2arima | (0,1,0) | −97.07 | −95.66 | 0 |
| k3arima | (0,1,0) | −38.83 | −38.12 | 0.0039 |

**Continued**

| | | | | |
|---|---|---|---|---|
| k4arima | (0,1,0) | −91.78 | −90.36 | 0.0001 |
| k5arima | (1,1,0) | −72.18 | −70.05 | 0.0004 |
| k6arima | (0,1,0) | −77.23 | −76.23 | 0.0003 |
| k7arima | (1,1,0) | −65.29 | −63.17 | 0.0006 |
| k8arima | (0,1,0) | −73.18 | −71.76 | 0.0004 |
| k9arima | (1,1,0) | −75.92 | −73.8 | 0.0003 |
| k10arima | (1,1,0) | −73.39 | −71.26 | 0.0003 |
| k11arima | (0,1,0) | −93.12 | −91.71 | 0 |
| k12arima | (1,1,0) | −87.34 | −85.22 | 0.0001 |
| k13arima | (2,1,0) | −80.96 | −76.96 | 0.0002 |
| k14arima | (0,1,1) | −86.54 | −84.41 | 0.0001 |

图 7 是属于第一簇的奥地利，法国，希腊，日本，中国台湾和属于第二簇的丹麦时期效应 $k_{i,t}$ 在未来 8 年的预测值。计算公式为 $m_{x,s_0+t,i} \approx \exp\left(\hat{\alpha}_{x,i} + \hat{B}_x \hat{k}_{s_0+t,i} + \varepsilon_{x,t,i}\right)$。图 8 分别作出这六个国家和地区 2015 年的真实死亡率，fuzzy CAE 模型预测死亡率，CAE 模型预测死亡率(红色，蓝色和橘色)，知 fuzzy CAE 模型比原模型效果好很多。Fuzzy CAE 模型的预测值和真实值很接近，除过 90 岁前后，死亡率变动大，预测值稍微高于真实值，这受高龄数据影响。

图 9 是奥地利 75 岁和 85 岁人口的实际和预测死亡率(黑色，红色)。实际死亡率波动很大，预测死亡率波动小，较为平缓且稍微高于实际值，符合实践。

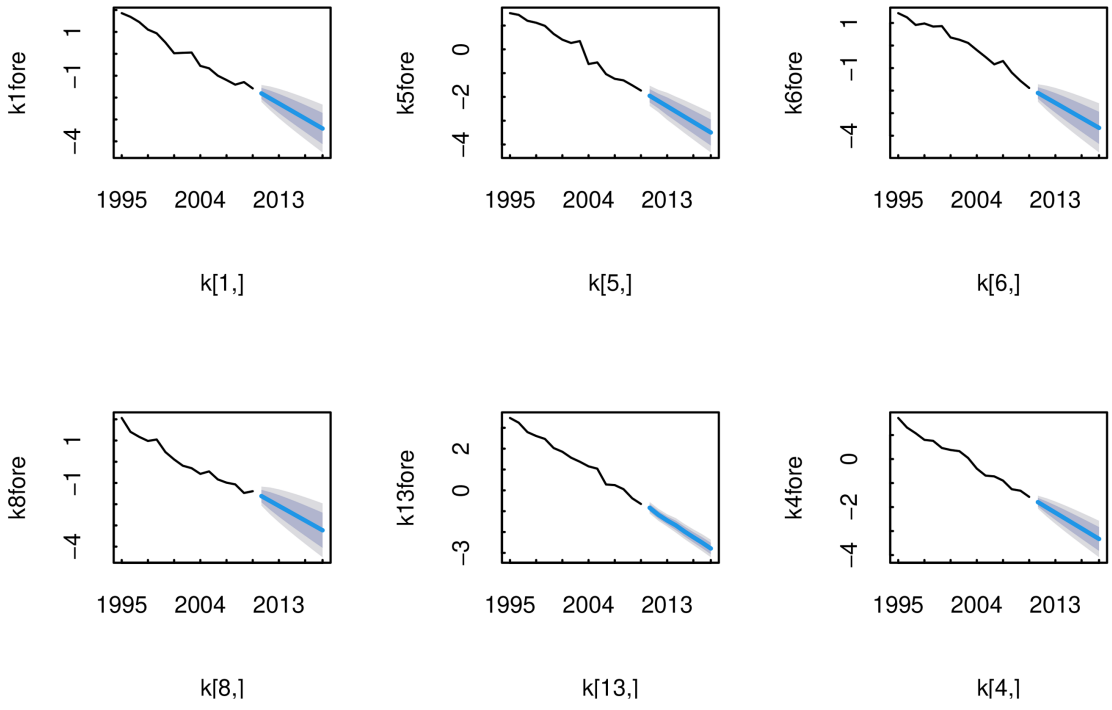


**Figure 7.** Predicted values over the next eight years for six countries and regions
**图 7.** 六个国家和地区在未来八年的时间项的预测值

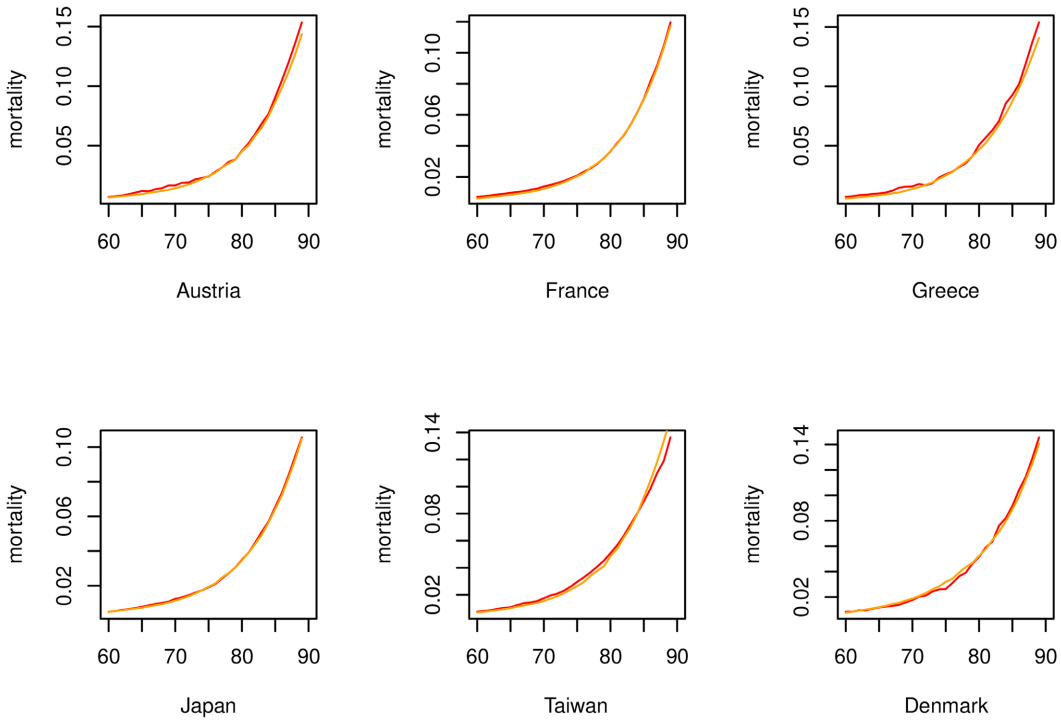


**Figure 8.** Actual and predicted mortality rates for six countries and regions in 2015
**图 8.** 2015 年六个国家和地区实际和预测死亡率

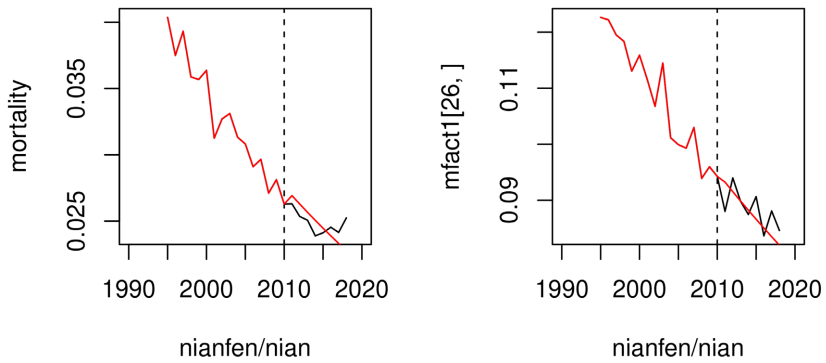


**Figure 9.** Actual and predicted mortality rates for people aged 75 and 85 in Austria
**图 9.** 奥地利 75 岁和 85 岁人口实际和预测死亡率

由绝对拟合误差(AFE)评估模型拟合和预测效果。AFE 表达式为：$\mathrm{AFE}_i = \sum_{x,t}\left|m_{x,t,i} - \hat{m}_{x,t,i}\right|$。表 5 是 14 个国家和地区，fuzzy CAE 模型的绝对拟合误差。AFE 值大都在 0.02 到 0.05 之间，拟合和预测效果良好。

**Table 5.** AFE values of 14 countries and regions
**表 5.** 14 个国家和地区的 AFE 值

| 地区 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| AFE | 0.0319 | 0.0415 | 0.07482 | 0.0408 | 0.0331 | 0.0367 | 0.0546 |
| 地区 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| AFE | 0.0280 | 0.0503 | 0.0378 | 0.0294 | 0.0388 | 0.0391 | 0.0367 |

### 3.2. 基于 Jiont-k 模型的模糊极大似然聚类

基于 Jiont-k 模型的模糊极大似然聚类是将 ILC 模型、隶属度矩阵 $\boldsymbol{U}$ 同 Jiont-k 模型结合的方法。在 ILC 模型里用加权最小二乘估计(WLS)提取各个国家的时期效应，由三类聚类方法处理时期效应 k[i,t]。

#### 3.2.1. Cmeans 聚类

基于 Hartigan-Wong 算法将 $k_{i,t}$ 聚为三类，四类和五类，between-ss/total-ss 分别为 70.3%，80.4%，84.7%。图 10 是将 14 个时期效应聚为四类和五类的结果。前三类完全一样，第四类被再分割。中国大陆地区；被单独分为一类葡萄牙和英国被分为一类；奥地利，香港，日本和中国台湾被分为一类，都处于发达地区，后三者都处于东亚，时期效应对它们的影响相似；剩余国家是第四类，大都是欧洲发达国家。第四类细分的两小组时期效应类似，在高效和简洁间取舍，决定聚为四类。

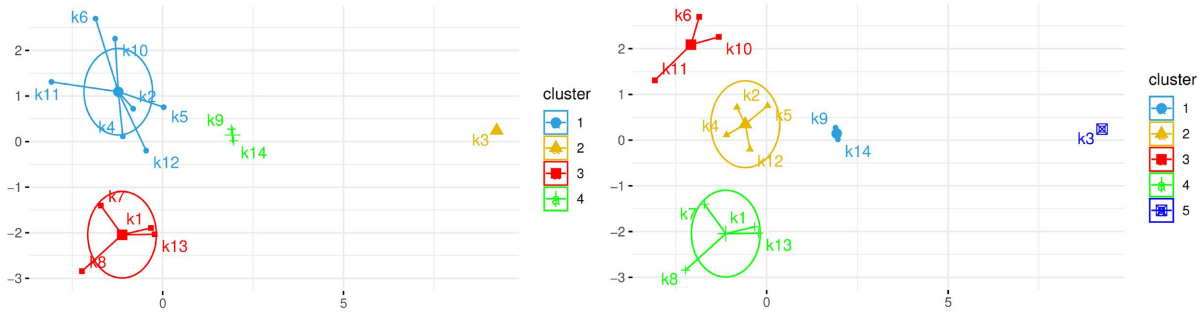


**Figure 10.** The 14 period effects were grouped into 4 and 5 classes respectively
**图 10.** 将 14 个时期效应分别聚为 4 类和 5 类

#### 3.2.2. 层次聚类

由欧式距离，计算 14 个时期效应 $k_{i,t}$ 的相似性，由 Ward.D2 算法层次聚类并切分聚类树。图 11 将聚类树切分为三类和四类。切分为三类时，中国单独为一类，希腊，西班牙，瑞典为一类，其他的为一类，同 cmeans 聚类结果相同。聚为四类时，将中国香港和日本聚为一类，都属东亚发达地区，认为聚为四类更合理。

#### 3.2.3. 模糊联合 K 模型(Fuzzy Jiont-k)

和 fuzzy CAE 模型同理，将 $\boldsymbol{U}$ 同 Jiont-k 模型结合得到改进模型。基于前两种聚类结果，将时期效应 $k_{i,t}$ 聚为 4 类。基于 $k_{i,t}$ 得 $\boldsymbol{U0}$ 和 $\boldsymbol{V0}$，由 fcm 函数进行模糊 k-Means 聚类，得隶属度矩阵 $\boldsymbol{U}$ 和聚类中心 $\boldsymbol{V}$，将 $\boldsymbol{U}$ 代入 Jiont-k 模型得 fuzzy Jiont-k 模型，表达式如下：

$$\ln\left(m_{x,t,i}\right)=a_{x,i}+\sum_{l=1}^{k}\left(u_{i,l}b_{x,i}K_{l,t}\right) \tag{22}$$

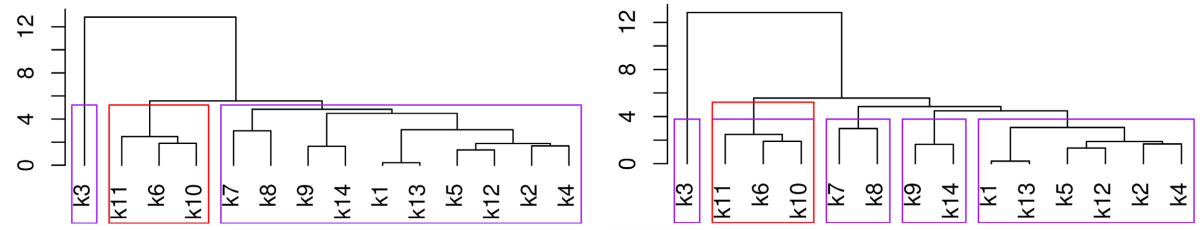


**Figure 11.** Hierarchical clustering tree is divided into 3 and 4 classes
**图 11.** 将层次聚类树分为 3 类和 4 类

图 12 是聚类为 4 类的可视化表示，前两类结果同 k-Means 聚类和层次聚类，中国单独为一类，后两类与前两种聚类结果相似，第三类是奥地利，中国香港，日本，瑞士，中国台湾；第四类将丹麦，法国，希腊，西班牙，瑞典分为一类，这五个国家都是欧洲国家。

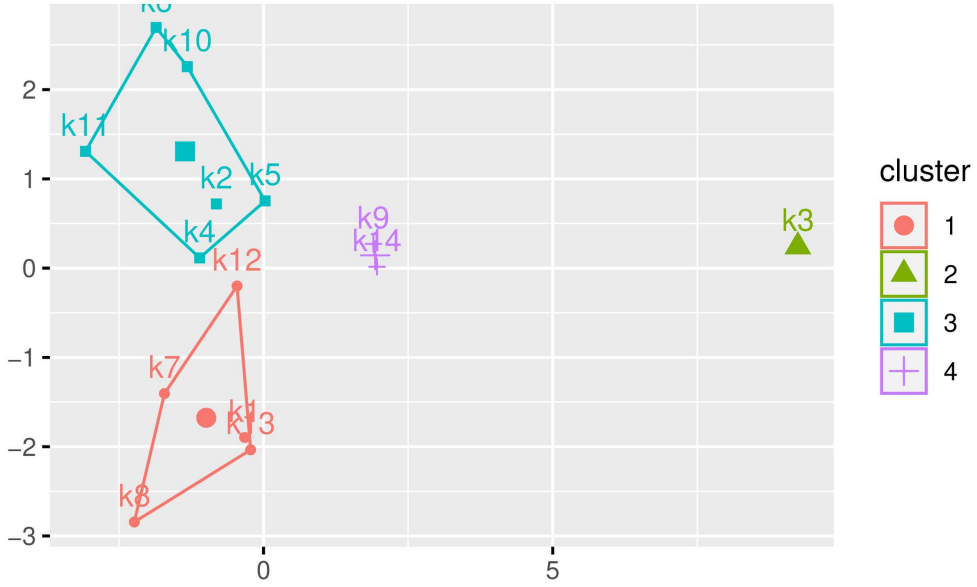


**Figure 12.** Fuzzy Jiont-K model clustered the period effects into four classes
**图 12.** Fuzzy Jiont-k 模型将时期效应聚为 4 类

由极大似然法估计参数，1995~2010 年的用于拟合模型，2011~2018 年的用于预测和检验。

参数估计结果如图 13，聚类后参数变化趋势同 ILC 模型基本相同，每个簇年龄效应有更多相似点。中国的时期效应依然变动较多，聚类后中国台湾地区有了比 ILC 模型更大的时期效应，说明时期效应对台湾的死亡率及其所属簇的影响增大，奥地利的时期效应也增强。

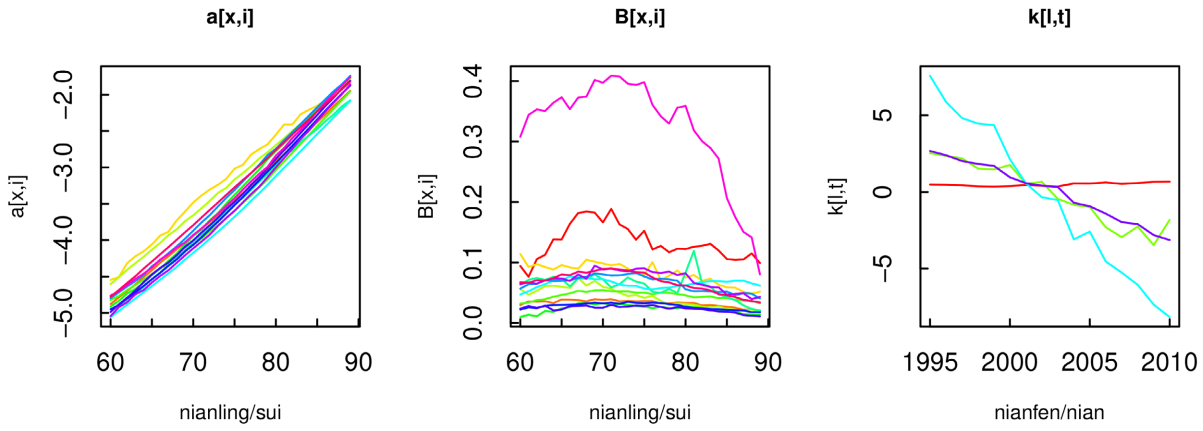


**Figure 13.** Parameter estimation results of Fuzzy Jiont-k model
**图 13.** Fuzzy Jiont-k 模型的参数估计结果

图 14 是各个簇的时期效应，每个簇的个体时期效应变化较一致，聚类效果较好。

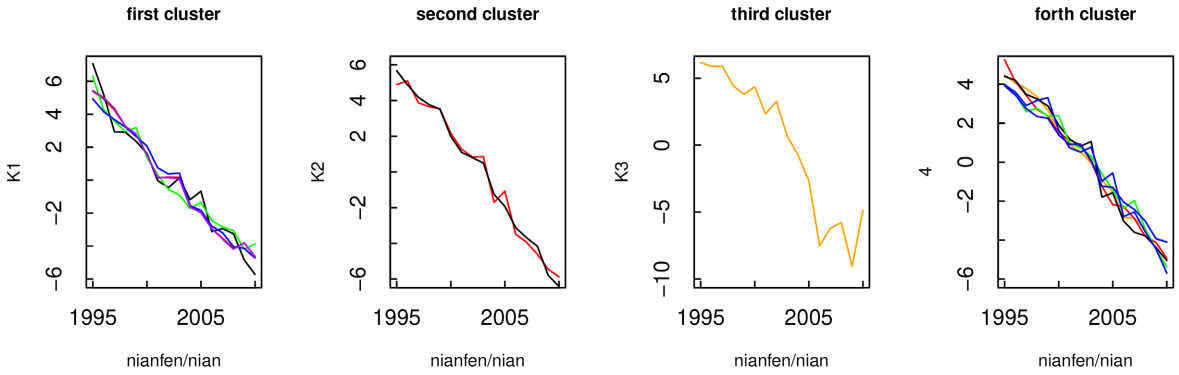


**Figure 14.** Period effects of each cluster in fuzzy Jiont-k models
**图 14.** Fuzzy Jiont-k 模型各个簇的时期效应

预测未来 8 年各地区死亡率。采用 zoo，forecast 和 stats 包，由带漂移项的随机游走模型拟合数据，得 $k_{l,t}$ 的预测值。将预测值代入 $m_{x,s_0+t,i} \approx \exp\left(\hat{\alpha}_{x,i} + \hat{b}_{x,i}\hat{K}_{s_0+t} + \varepsilon_{x,t,i}\right)$，得预测的死亡率。图 15 是 2015 年瑞士，丹麦和法国的由 fuzzy Jiont-k，Jiont-k 预测的死亡率和实际死亡率，(蓝色，橙色和红色)，知原模型和改进模型的拟合和预测效果都很好，但 fuzzy Jiont-k 模型效果更好。丹麦实际死亡率变动较大，预测的死亡率更平滑，改进了这个缺点。

图 16，尝试了法国人口不同年份 75 岁和 85 岁的死亡率，(红色是实际，蓝色是 fuzzy Jiont-k 模型，橘色是 Jiont-k 模型)，fuzzy Jiont-k 模型的预测效果好。若只观察 fuzzy Jiont-k 模型，拟合效果整体较好，预测死亡率低于实际死亡率，符合经验。由 AFE 评估模型效果。表 6 是 AFE 值，都小于 0.029，拟合和预测效果较好。Fuzzy CAE 模型的 AFE 值大于 fuzzy Jiont-k，fuzzy Jiont-k 模型效果更好，但基于共有年龄效应的模型很重要，可继续改进或者将二者结合。Kleinow (2015) [7]，赵明(2020) [13]等提出给模型添加队列效应改进效果。后文分别给 CAE 模型和 Jiont-k 模型添加队列效应。

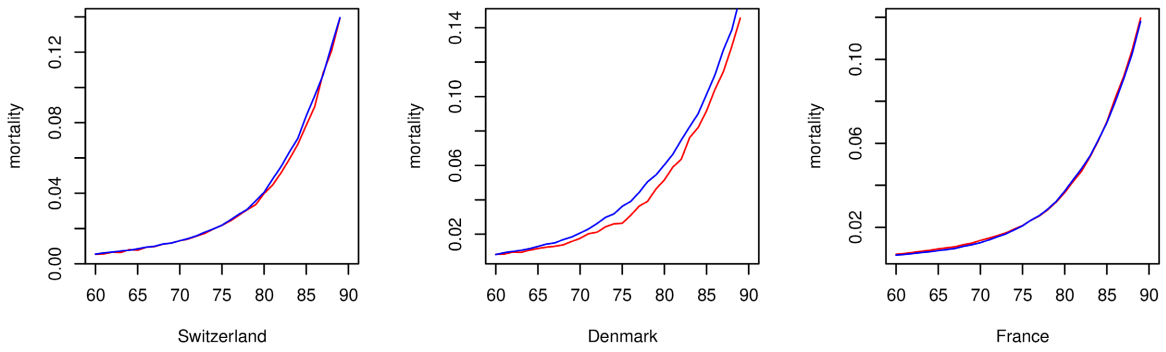


**Figure 15.** Actual and predicted death rates for Switzerland, Denmark and France in 2015
**图 15.** 瑞士，丹麦，法国 2015 年的实际和预测死亡率

**Table 6.** AFE values of 14 countries and regions
**表 6.** 十四个国家和地区的 AFE 的值

| 序号 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| AFE | 0.0148 | 0.0010 | 0.0289 | 0.0061 | 0.0057 | 0.0047 | 0.0263 |
| 序号 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| AFE | 0.0126 | 0.0084 | 0.0063 | 0.0032 | 0.0030 | 0.0216 | 0.0138 |

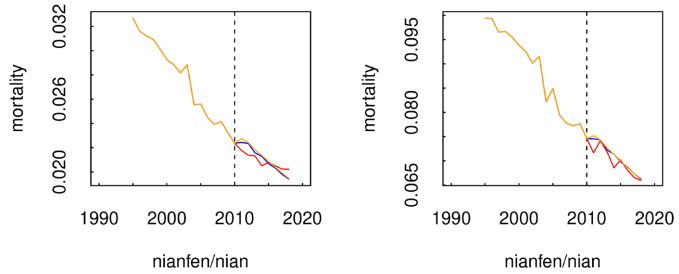


**Figure 16.** Actual and predicted mortality rates in French people aged 75, 85
**图 16.** 法国 75，85 的实际和预测死亡率

## 4. 加入队列效应的多人口共同因子死亡率模型

### 4.1. 添加了队列效应的共同年龄效应模型(rCAE)

给 CAE 模型添加队列效应，产生 rCAE 模型。对年龄效应聚类，奥地利，法国，希腊，日本，中国台湾地区总被分为一类，后两个都属于东亚，有相似地理结构，选取中国台湾和日本的 53~89 岁，1970~2019 年的整体死亡率数据，1970~2010 年用来拟合，2011~2019 年检验效果。队列效应用 $r_{t-x}$ 表示，$t-x$ 表示出生年份。由于数据有限，找不到某些年纪较大人的出生年份，比如在 2019 年 89 岁的人，出生年是 1930 年，1930 年不在研究的数据范围内。当出生年份在研究年份内，认为有队列效应，否则默认没有队列效应。即当 $t \geq (x+1)$ 时，认为有队列效应，如(11)式；否则，认为没有队列效应，如(5)或(6)式。当出生年不在研究年份内，极大似然估计(MLE)会出现空值且易报错，故选择加权最小二乘法估计参数。图 17 是参数估计图(红色是台湾，绿色是日本)，各参数变化趋势同 CAE 模型，台湾地区的对数中心死亡率均值普遍大于日本，后缓慢靠近；台湾地区时期效应 $k_{i,t}$ 远大于日本，而日本的队列效应大于中国台湾地区。

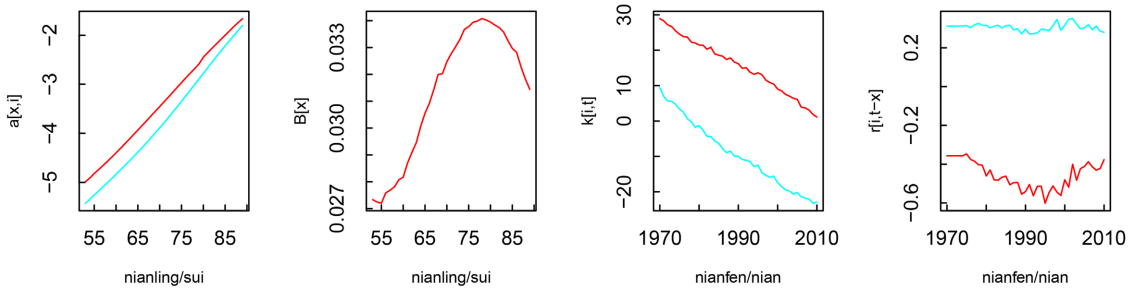


**Figure 17.** Results of parameter estimates for model rCAE
**图 17.** rCAE 模型的参数估计值

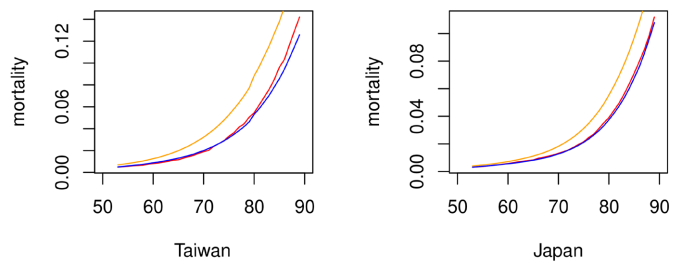


**Figure 18.** Actual and predicted mortality rates for Taiwan, China and Japan in 2015
**图 18.** 2015 年中国台湾地区和日本的实际和预测死亡率

图 18 是 2015 年中国台湾和日本地区的死亡率对照图(红色是实际，蓝色是 CAE 模型，橙色是 rCAE 模型)。rCAE 模型的预测效果没有 CAE 模型好，这是由于采用的不是队列数据，年份数据不够导致抽取的时期效应不完全。由估计的未来 9 年的时期效应，说明就 rCAE 模型死亡率下降趋势基本一致，但台湾的时期效应下降速度比日本稍微快一点，随着时间的推移，下降趋势越缓慢，时期效应对死亡率的影响也缓慢减小。由 AFE 评估模型对 2 个国家的拟合和预测效果，AFE 值分别为 24.65%，6.26%，模型拟合和预测效果良好。

## 4.2. 添加了队列效应的 Jiont-k 模型(rJiont-k)

给 Jiont-k 模型添加队列效应产生 rJiont-k 模型。对时期效应的聚类大多会将奥地利，中国香港，日本和中国台湾聚类到一起，选取中国台湾地区和日本 53~89 岁，1970~2019 年的整体死亡率数据：1970~2010 年的用来拟合，2011~2019 年的用于检验。出生年份在研究年份以内时，认为有队列效应，否则默认没有队列效应。即当 t >= (x + 1)时，认为有队列效应，如(18)式；否则，认为没有队列效应，如(14)式。对出生年不在研究年份内的情况，选择加权最小二乘法(WLS)估计参数。图 19 是参数估计图(红色是台湾，绿色是日本)，日本的年龄效应基本都大于中国台湾地区，提取的公共时期效应刚开始快速下降，后期下降速度缓慢降低，前期中国台湾地区的队列效应高于日本的，后期日本的队列效应高于中国台湾。

如图 20，比较了实际死亡率，Jiont-k 模型和 rJiont-k 模型预测的死亡率(红色，蓝色和橘色)，Jiont-k 模型的预测效果比 rJiont-k 模型好。rJiont-k 模型在未来 9 年的时期效应，下降速度缓慢，所处时期对死亡率的影响缓慢减小。由 AFE 评估模型的拟合和预测效果，两国家和地区的 AFE 值分别是：24.01%，6.36%，认为拟合和预测效果良好。

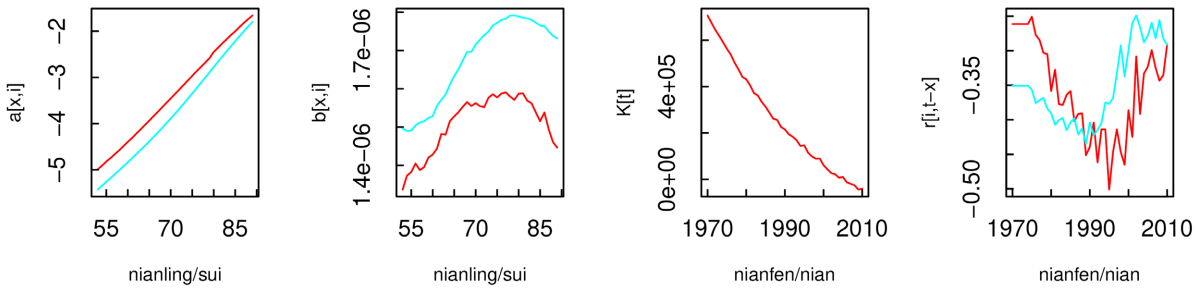


**Figure 19.** Result of parameter estimation of rJiont-k model
**图 19.** rJiont-k 模型的参数估计值

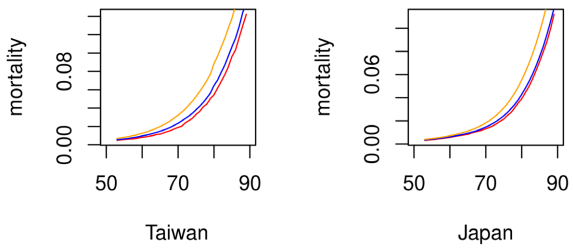


**Figure 20.** Actual and predicted mortality rates for Taiwan, China and Japan in 2015
**图 20.** 2015 年中国台湾地区和日本的实际和预测死亡率

## 5. 结论

本文的研究数据主要来自于人类死亡率数据库和中国人口就业统计年鉴的 14 个国家和地区的 18~89

岁1970~2019年的整体死亡率数据。因为它们大都是欧亚国家，发达国家或发达经济体，由Wen (2021) [11]等知不同社会经济群体的死亡率会有所不同，故存在公共因子；死亡率数据丰富全面，拟合多人口共同因子死亡率模型的效果更好。编程软件是R软件，使用的R包主要有ggplot2，factoextra，cluster，e1071，ppclust，inaparc，zoo，forecast，stats等。

本文的创新点是提出四个新模型并作比较分析，数值结果表明基于聚类的模型的拟合和预测效果比原模型好。(fuzzy CAE模型优于CAE模型，fuzzy Jiont-k模型优于Jiont-k模型，提取部分队列效应模型的效果没有原模型好，提取时期效应模型的收敛速度普遍快于提取年龄效应的模型。)给CAE模型和Jiont-k模型加入队列效应，由于许多人的出生年不在研究范围内，对效果产生了不好的影响。可以尝试由队列数据代替一般数据，提高代码效率来改进模型。由于基于聚类的模型效果比原模型好，在研究数据公共因子时，可先对元素聚类，再对不同的簇分别处理。我们还可研究单一的队列数据对死亡率的影响。考虑其他死亡率模型研究多人口死亡率，如ACF模型，APC模型，EAPC模型等。

本文提出的模型为多人口死亡率研究做了新的尝试，提供了新的方向。目前拟合我国死亡率效果最好的模型是CBD模型和APC模型，我们还可使用CAE，fuzzyCAE，fuzzyJiont-k等多人口模型研究我国不同省份，不同经济水平地区的死亡率。还可将神经网络等机器学习方法引进到多人口死亡率研究中。多人口死亡率模型可以在一个更全面的体系内考虑多种群死亡率，有很大的潜力亟待开发，随着社会及相关学科的发展，多人口死亡率模型会发挥更大的作用。

## 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[Z/OL]. 国家统计局. http://www.stats.gov.cn/ztJc/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] 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

[4] Li, N. and Lee, R. (2005) Coherent Mortality Forecasts for a Group of Populations: An Extension of the Lee-Carter Method. *Demography*, **42**, 575-594. https://doi.org/10.1353/dem.2005.0021

[5] Renshaw, A.E. and Haberman, S. (2006) A Cohort-Based Extension to the Lee-Carter Model for Mortality Reduction Factors. *Insurance*: *Mathematics and Economics*, **38**, 556-570. https://doi.org/10.1016/j.insmatheco.2005.12.001

[6] 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

[7] Kleinow, T. (2015) A Common Age Effect Model for the Mortality of Multiple Populations. *Insurance Mathematics and Economics*, **63**, 147-152. https://doi.org/10.1016/j.insmatheco.2015.03.023

[8] Hatzopoulos, P. and Haberman, S. (2013) Common Mortality Modeling and Coherent Forecasts. An Empirical Analysis of Worldwide Mortality Data. *Insurance Mathematics and Economics*, **52**, 320-337. https://doi.org/10.1016/j.insmatheco.2012.12.009

[9] Danesi, I.L., Haberman, S. and Millossovich, P. (2015) Forecasting Mortality in Subpopulations Using Lee-Carter Type Models: A Comparison. *Insurance Mathematics and Economics*, **62**, 151-161. https://doi.org/10.1016/j.insmatheco.2015.03.010

[10] Enchev, V., Kleinow, T. and Cairns, A.J.G. (2017) Multi-Population Mortality Models: Fitting, Forecasting and Comparisons. *Scandinavian Actuarial Journal*, **2017**, 319-342. https://doi.org/10.1080/03461238.2015.1133450

[11] Wen, J., Cairns, A. and Kleinow, T. (2021) Fitting Multi-Population Mortality Models to Socio-Economic Groups.

*Annals of Actuarial Science*, **15**, 144-172. https://doi.org/10.1017/S1748499520000184

[12] 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

[13] 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.

[14] 王晓军, 路倩. 动态死亡率模型的研究进展[J]. 应用概率统计, 2020, 36(4): 415-440.

[15] 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.

[16] 赵明. 中国男性人口死亡率动态预测的方法比较——基于 Lee-Carter 模型与贝叶斯分层模型的研究[J]. 人口与发展, 2022, 28(1): 40-49.

[17] 肖鸿民, 李芳芳, 赵苗苗. 对共同年龄效应模型的研究及中国应用[J]. 应用数学进展, 2021, 10(11): 3743-3757.