

# 老龄多人口死亡率联合建模与 一致性预测

王晓军<sup>1,2</sup>, 赵晓月<sup>2</sup>, 陈惠民<sup>2</sup>

(1. 中国人民大学应用统计科学研究中心, 北京 100872;

2. 中国人民大学统计学院, 北京 100872)

**摘要:** 退休年龄以上的老年人群死亡率预测是养老金精算和长寿风险度量的基础。针对我国大陆地区退休年龄以上人群死亡率数据量较小且波动较大的问题, 借助多人口联合建模思想, 基于单人口 CBD 模型, 提出了一个适用于老龄死亡率建模的 Logistic 多人口模型。通过加入更多相关人口数据信息来预测我国老年人口死亡率, 选取我国台湾地区分性别死亡率相关数据, 与我国大陆地区分性别死亡率数据进行联合建模。研究发现, Logistic 多人口死亡率模型比单人口 CBD 模型表现出更好的拟合效果和长期预测一致性效果。

**关键词:** 老年人口; 死亡率; Logistic 模型; 多人口模型

**中图分类号:** C921 **文献标识码:** A **文章编号:** 1000-4149 (2021) 02-0045-12

**DOI:** 10.3969/j.issn.1000-4149.2021.00.011

## Modeling and Coherent Forecasting of the Elderly Mortality with Multiple Populations

WANG Xiaojun<sup>1,2</sup>, ZHAO Xiaoyue<sup>2</sup>, CHEN Huimin<sup>2</sup>

(1. Center for Applied Statistics, Renmin University of China, Beijing 100872,

China; 2. School of Statistics, Renmin University of China, Beijing 100872, China)

**Abstract:** The mortality prediction of the elderly above retirement age is the basis of

收稿日期: 2020-08-28; 修订日期: 2021-01-25

基金项目: 教育部哲学社会科学研究重大课题攻关项目“健康中国 2030 背景下的健康老龄化体系优化研究”(20JZD023)。

作者简介: 王晓军, 中国人民大学应用统计科学研究中心研究员, 统计学院教授; 赵晓月, 中国人民大学统计学院博士研究生; 陈惠民, 中国人民大学统计学院博士研究生。

pension actuarial valuation and longevity risk measurement. In order to solve the problem of small amount and large fluctuation of mortality data of people above retirement age in mainland China, this paper proposes a Logistic multi-population model for the elderly mortality model based on a single population Cairns-Blake-Dowd (CBD) model. By adding more relevant population data information to predict the mortality rate of the elderly population in China, this paper selects the gender specific mortality data in Taiwan and carries out joint modeling with the gender specific mortality rate in mainland China. It is found that compared with single population CBD model, the Logistic multi-population mortality model has a better fitting and long-term coherent prediction effect.

**Keywords:** aging population; mortality; Logistic model; multiple population model

## 一、引言

退休年龄以上的老年人群死亡率预测是养老金精算和长寿风险度量的基础。依据国家统计局公布的数据,截至2019年末,我国60岁及以上老年人口数达到2.54亿人,占总人口的比例超过18%。人口预期寿命逐年增加,2015年我国男、女性出生预期寿命分别达到73.64岁和79.43岁。另外,依据联合国《世界人口展望2019年》的预测结果,未来人口预期寿命仍然保持约每5年增加一岁的趋势,人口预期寿命增长的不确定性给养老金风险管理带来挑战。

在我国,分年龄的死亡人数和生存人数来源于人口普查和人口抽样调查,除了每10年的人口普查和每5年的1%人口抽样调查,其他年份均是千分之一左右的人口变动抽样调查,对应的分年龄性别人数较少,分年龄性别的死亡率波动较大。用这些数据直接建模,很难捕捉死亡率随年龄和时间的变动规律。

考虑到人类死亡率变动具有共同规律性的特点,近年来,研究者开始关注对多人口死亡率的联合建模。这里的多人口指研究人群范围为两个或两个以上的具有相似特征的不同人口群体,可以是具有相似经济文化背景或地理位置相近的国家或地区,也可以是同一地区的不同种族或性别。有学者指出,密切相关人群的死亡率模式和轨迹可能类似,可以通过考虑多人口的联合建模方式来提高个别国家的死亡率预测精度<sup>[1]</sup>。

多人口死亡率模型一般包括公共因素项和附加的特定因素项,公共因素描述整个人口死亡率随时间变化的主要长期趋势,附加因素考虑不同人口的短期差异。多人口死亡率模型可以分为三大类:一是对男女性联合建模,能够有效避免男女死亡率在中长期预测中出现死亡率交叉或分离的不合理问题;二是对不同国家或地区的多个人口进行死亡率联合建模,以达到长期的一致性预测;三是对不同国家和地区的不同性别人口死亡率进行联合建模,既考虑不同人口间的协整关系,又保证同一人口不同性别间的相依关系。

我国大陆地区高龄人口死亡数据的时间序列短,分年龄性别的风暴暴露人数少,死亡率数据的波动性较大,因此在现有大陆地区人口死亡率数据的基础上,加入更多相关人口数据信息,可以实现对我国高龄人口死亡率更好地建模和预测。本文基于CBD模型和

Logistic 两人口死亡率模型, 提出一种包括分地区和分性别的老龄多人口死亡率模型, 在性别因子的基础上加入国别和地区别因子, 假设死亡人数服从泊松分布, 利用极大似然估计进行参数估计。选择人种相同、地理位置相近的中国大陆、中国台湾分男女共四类人口的数据进行联合建模, 通过捕捉死亡率变动的共同规律, 拟对中国大陆人口死亡率做出更合理预测。

## 二、死亡率模型的研究综述

早期的死亡率模型只是能描述死亡率随年龄变动的静态模型, 之后死亡率模型中加入时间项, 发展为单个人口动态模型, 动态死亡率模型的典型代表是 1992 年提出的 Lee-Carter 模型<sup>[2]</sup>, 该模型假设对数死亡率由年龄效应、时间效应以及与年龄相关且与时间有交互效应的项组成, 因其形式简单且参数的可解释性强等特点被广泛应用。之后, 有许多研究在 Lee-Carter 模型基础上进行了扩展。伦肖 (Renshaw) 等在 Lee-Carter 模型的基础上加入了队列效应, 对于过去观察到队列效应的国家, 该模型可以更好地拟合历史数据<sup>[3]</sup>。柯里 (Currie) 提出了含有年龄、时期、队列的 Age-Period-Cohort (APC) 模型, 该模型可以视作伦肖所提模型的简化形式, 模型不需要平滑性的条件, 消除了可能存在的稳健性问题<sup>[4]</sup>。凯恩斯 (Cairns) 等又提出了适用于高龄死亡率的 CBD (Cairns-Black-Dowd) 模型, CBD 模型引入了两个时间效应因子, 区分了高年龄人群的死亡率随时间的改善效应, 使模型对高年龄段的拟合和预测效果更好<sup>[5]</sup>。凯恩斯等人在 CBD 模型的基础上, 考虑二次项和队列效应项, 给出了一些 CBD 扩展模型, 并指出加入二次项和队列项的模型拟合效果更好<sup>[6]</sup>。

随着研究的进行, 研究者发现单人口死亡率模型在死亡率的长期预测中容易出现不合理的交叉或偏离情况, 有违人类死亡率变动的一般规律性, 从而提出多人口联合建模和预测的思路。有学者在 Lee-Carter 模型基础上进行扩展, 提出了公共因子 (Common Factor, CF) 模型及增广公因子 (Augmented Common Factor, ACF) 模型, 它包含了整个人群的一个共同因素, 以及扩展模型中的一个额外的性别特定因素<sup>[1]</sup>。公共因素描述了整个人口死亡率随时间变化的长期主要趋势, 附加因素考虑了每个性别的短期差异。使用公共因子模型避免了不同人口死亡率随时间出现的交叉或偏离问题, 并确保根据观察到的历史模式, 每个年龄段的男性死亡率与女性死亡率的比率在长期收敛到一个常数。又有学者使用泊松分布直接对死亡人数建模, 同时将模型结构广义化, 提出了泊松公共因子模型 (Poisson Common Factor Model, 简称 PCFM 模型), 该模型为数据分析、参数估计和模型选择提供了严格的统计框架, 同时模型可以扩展为多个特定因素, 可以更好地捕获残差中的其他系统特征<sup>[7]</sup>。他们利用 10 个人口死亡率数据, 进一步评估了泊松公共因子模型的拟合和预测性能, 结果表明, 该方法能确保每个年龄段的男女死亡率在长期内收敛到一个常数, 并提供更准确的男女死亡率预测<sup>[8]</sup>。此后, 又有学者在扩展公共因子模型的基础上加入了国别 (不同国家或地区) 因子, 提出了两层扩展公共因子模型, 并使用英国的英格兰和威尔士、苏格兰和北爱尔兰的数据, 对模型进行了验证, 对死亡率作出了一致的预测<sup>[9]</sup>。

学者们又对单人口 CBD 模型进行扩展, 在截距项和斜率项中分别加入了性别参数, 提

出了一个适用于高龄的 Logistic 两人口死亡率模型, 并使用模型对比利时、瑞典、瑞士、英国四个国家的高龄死亡率数据进行建模预测, 取得了良好的拟合和预测效果<sup>[10]</sup>。

### 三、Logistic 死亡率模型

#### 1. CBD 模型

凯恩斯等人提出的 CBD 模型是一种单人口 Logistic 动态模型, 适用于高龄人群死亡率的拟合和预测<sup>[5]</sup>。模型形式如下:

$$\text{logit}(q_{x,t}) = \kappa_t^{(1)} + (x - \bar{x}) \kappa_t^{(2)} + \varepsilon_{x,t} \quad (1)$$

其中,  $q_{x,t}$  是  $t$  年  $x$  岁的死亡概率,  $\kappa_t^{(1)}$ 、 $\kappa_t^{(2)}$  是两个时间项,  $\varepsilon_{x,t}$  为随机误差项。

#### 2. Logistic 两人口死亡率模型

如前所述, 采用单人口模型对男性和女性高龄人口死亡率进行单独建模和预测时往往在预测期内会出现两性死亡率交叉或分离等脱离实际的现象, 因此需要考虑对两性死亡率的联合建模。

2019 年, 学者们提出适用于高龄的 Logistic 两人口死亡率模型<sup>[10]</sup>, 模型形式如下:

$$\text{logit}q_{x,t}^{(i)} = \alpha_i + \beta_i(x - \bar{x}) + \tau_i^{(i)} + \nu_i^{(i)}(x - \bar{x}) \quad (2)$$

其中,  $q_{x,t}^{(i)}$  表示  $x$  岁的人在时间  $t$  的死亡概率,  $i$  代表性别 ( $i=1$  代表女性,  $i=2$  代表男性),  $\alpha_i$  和  $\beta_i$  是两个性别的共同参数,  $\tau_i^{(i)}$  和  $\nu_i^{(i)}$  是分性别参数,  $\alpha_i + \tau_i^{(i)}$  和  $\beta_i + \nu_i^{(i)}$  分别代表死亡率曲线对性别  $i$  在时间  $t$  的截距项和斜率项。  $\bar{x}$  表示建模年龄段的平均年龄。

在模型 (2) 的基础上加入个体曲率项可以进一步扩展为模型 (3), 具体形式如下:

$$\text{logit}q_{x,t}^{(i)} = \alpha_i + \beta_i(x - \bar{x}) + \tau_i^{(i)} + \nu_i^{(i)}(x - \bar{x}) + \gamma_i^{(i)} [(x - \bar{x})^2 - \sigma^2] \quad (3)$$

其中,  $\gamma_i^{(i)}$  是关于性别  $i$  的曲率项,  $\sigma^2$  是建模所覆盖的年龄范围下  $(x - \bar{x})^2$  的均值。如果男女之间的曲率相似, 即第三项为公共曲率项, 则模型 (3) 可简化为模型 (4):

$$\text{logit}q_{x,t}^{(i)} = \alpha_i + \beta_i(x - \bar{x}) + \tau_i^{(i)} + \nu_i^{(i)}(x - \bar{x}) + \gamma_i [(x - \bar{x})^2 - \sigma^2] \quad (4)$$

其中,  $\gamma_i$  为两性别的共同参数项。

#### 3. Logistic 多人口死亡率模型

以上两人口模型可以避免男女死亡率随时间的交叉问题, 但没有考虑不同人口之间死亡率的长期一致性。本文在 Logistic 两人口死亡率模型的基础上, 加入人口别 (不同国家或地区) 因子, 提出一种 Logistic 多人口死亡率模型, 既考虑不同人口之间的协整关系, 又考虑性别间的相依关系, 构建高维多人口死亡率模型, 捕捉人类死亡率变动的共同规律, 使对不同人口的死亡率预测具有一致性。模型形式如下:

$$\text{logit } q_{x,t,i,j} = \alpha_i + \beta_i(x - \bar{x}) + \tau_{t,i,j} + \nu_{t,i,j}(x - \bar{x}) \quad (5)$$

其中,  $q_{x,t,i,j}$  表示  $x$  岁的人在时间  $t$  的死亡概率,  $i$  代表性别 ( $i=1$  代表女性,  $i=2$  代表男性),  $j=n$  表示  $n$  个不同的人口 (国家或地区) 数据,  $\alpha_i$  和  $\beta_i$  是共同参数,  $\tau_{t,i,j}$  和  $\nu_{t,i,j}$  是分性别参数。  $\alpha_i + \tau_{t,i,j}$  和  $\beta_i + \nu_{t,i,j}$  分别代表死亡率曲线对性别  $i$  在时间  $t$  的截距项和斜率项,  $\bar{x}$  表示建模年龄段的平均年龄。

在模型 (5) 的基础上加入分性别分人口曲率项可以进一步扩展为:

$$\text{logit } q_{x,t,i,j} = \alpha_t + \beta_t(x - \bar{x}) + \tau_{t,i,j} + \nu_{t,i,j}(x - \bar{x}) + \gamma_{t,i,j} [(x - \bar{x})^2 - \sigma^2] \quad (6)$$

其中,  $\gamma_{t,i,j}$  是关于性别  $i$  人口  $j$  的曲率项,  $\sigma^2$  是建模所覆盖的年龄范围下  $(x - \bar{x})^2$  的均值。

为确保死亡率预测的一致性, 假设共同的参数遵循带漂移的随机游走, 性别参数和人口别参数均遵循自回归过程 AR (1):

$$\alpha_t = \delta_1 + \alpha_{t-1} + \varepsilon_{t,1} \quad (7)$$

$$\beta_t = \delta_2 + \beta_{t-1} + \varepsilon_{t,2} \quad (8)$$

$$\gamma_t = \delta_3 + \gamma_{t-1} + \varepsilon_{t,3} \quad (9)$$

$$\tau_{t,i,j} = \varphi_{0,1}^{(i,j)} + \varphi_{1,1}^{(i,j)} \tau_{t-1,i,j} + \omega_{t,1}^{(i,j)} \quad (10)$$

$$\nu_{t,i,j} = \varphi_{0,2}^{(i,j)} + \varphi_{1,2}^{(i,j)} \nu_{t-1,i,j} + \omega_{t,2}^{(i,j)} \quad (11)$$

$$\gamma_{t,i,j} = \varphi_{0,3}^{(i,j)} + \varphi_{1,3}^{(i,j)} \gamma_{t-1,i,j} + \omega_{t,3}^{(i,j)} \quad (12)$$

其中,  $\delta_1$ ,  $\delta_2$  和  $\delta_3$  是漂移项,  $\varphi_{0,1}^{(i,j)}$ ,  $\varphi_{0,2}^{(i,j)}$  和  $\varphi_{0,3}^{(i,j)}$  是截距项,  $\varphi_{1,1}^{(i,j)}$ ,  $\varphi_{1,2}^{(i,j)}$  和  $\varphi_{1,3}^{(i,j)}$  是自回归参数,  $\varepsilon_{t,1}$ ,  $\varepsilon_{t,2}$ ,  $\varepsilon_{t,3}$ ,  $\omega_{t,1}^{(i,j)}$ ,  $\omega_{t,2}^{(i,j)}$  和  $\omega_{t,3}^{(i,j)}$  是正态误差项。它们相互关联, 跨时间独立。

#### 4. 参数估计方法

运用极大似然估计方法估计参数, 以模型 (6) 为例, 假设死亡人数服从泊松分布:

$$D_{x,t,i,j} \sim \text{Poisson}(e_{x,t,i,j} \mu_{x,t,i,j}) \quad (13)$$

其中,  $D_{x,t,i,j}$  是死亡人数,  $e_{x,t,i,j}$  是中心暴露数。令  $\text{logit } \mu_{x,t,i,j} = \eta_{x,t,i,j}$ , 则  $\mu_{x,t,i,j} = \exp(\eta_{x,t,i,j}) / [1 + \exp(\eta_{x,t,i,j})]$ 。似然函数为:

$$l = \sum_{x,t,i,j} [d_{x,t,i,j} \ln e_{x,t,i,j} \mu_{x,t,i,j} + d_{x,t,i,j} \ln \mu_{x,t,i,j} - e_{x,t,i,j} \mu_{x,t,i,j} - \ln(d_{x,t,i,j}!)] \quad (14)$$

参数估计迭代更新方程为:  $\theta^* = \theta - \frac{\partial l}{\partial \theta} / \frac{\partial^2 l}{\partial \theta^2}$ , 迭代更新步骤如下。

步骤 1: 设置参数初始值 (建议  $\alpha_t = -2$ ,  $\beta_t = 0.1$ ,  $\tau_{x,t,i,j} = 0$ ,  $\nu_{x,t,i,j} = 0$ ,  $\gamma_{x,t,i,j} = 0$ ), 并计算所有的  $\mu_{x,t,i,j}$  和  $\eta_{x,t,i,j}$  的拟合值;

步骤 2: 对所有的  $t$  更新  $\alpha_t$ , 然后重新计算所有  $\mu_{x,t,i,j}$  和  $\eta_{x,t,i,j}$  的拟合值;

步骤 3: 对所有的  $t$  更新  $\beta_t$ , 然后重新计算所有  $\mu_{x,t,i,j}$  和  $\eta_{x,t,i,j}$  的拟合值;

步骤 4: 对所有  $i$  和  $t$  更新  $\tau_{x,t,i,j}$ , 然后重新计算所有  $\mu_{x,t,i,j}$  和  $\eta_{x,t,i,j}$  的拟合值;

步骤 5: 对所有  $i$  和  $t$  更新  $\nu_{x,t,i,j}$ , 然后重新计算所有  $\mu_{x,t,i,j}$  和  $\eta_{x,t,i,j}$  的拟合值;

步骤 6: 对所有  $i$  和  $t$  更新  $\gamma_{x,t,i,j}$ , 然后重新计算所有  $\mu_{x,t,i,j}$  和  $\eta_{x,t,i,j}$  的拟合值;

步骤 7: 对所有  $i$  和  $t$ , 调整  $\alpha_t^* = \alpha_t + c_\tau$  和  $\tau_{t,i,j}^* = \tau_{t,i,j} - c_\tau$ , 其中  $c_\tau = \sum_t \tau_t^{(1)} / N_t$ ,  $N_t$  为年数;

步骤 8: 对所有  $i$  和  $t$ , 调整  $\beta_t^* = \beta_t + c_\nu$  和  $\nu_{t,i,j}^* = \nu_{t,i,j} - c_\nu$ , 其中  $c_\nu = \sum_t \nu_t^{(1)} / N_t$ ,  $N_t$  为年数;

步骤 9: 计算对数似然函数;

步骤 10: 重复步骤 2 到 9, 直到对数似然函数收敛。对于其他形式的模型, 可以简单修改上面的一些步骤, 实现参数估计。

#### 四、死亡率联合建模

##### 1. 数据描述

中国大陆人口死亡率数据来自历年《中国人口统计年鉴》和《中国人口和就业统计年鉴》，目前公布的数据有1994—2018年共25年连续数据。其中2000年和2010年全国人口普查数据，1995年、2005年、2015年全国1%人口抽样调查数据，对于普查年数据和1%人口抽样调查数据，分年龄分性别有0岁到100岁及以上数据。其他年份为1‰左右变动抽样数据，其中1996年数据从0岁到85岁及以上，对于85—89岁死亡率采用插值法处理，其他年份数据从0岁到90岁及以上。

中国台湾人口死亡率数据来自人类死亡率数据库HMD (Human Mortality Database)<sup>①</sup>，根据公布的分年龄分性别数据，台湾地区有1970—2014年共45年连续数据，台湾区分年龄分性别数据从0岁到110岁及以上。相比中国大陆地区而言，中国台湾地区的死亡率数据时间段较长，数据质量较高。柳向东等认为我国大陆地区和台湾地区人口死亡率存在长期均衡关系<sup>[11]</sup>。段白鸽指出我国大陆和台湾地区各年龄死亡率、平均预期寿命之间的差异明显缩小，未来死亡率呈现趋同化发展<sup>[12]</sup>。因此我们选择中国大陆和中国台湾分性别老龄数据进行联合建模，对中国大陆老龄人口死亡率数据进行修匀和预测。

##### 2. 模型拟合和模型选择

我们选择中国大陆、中国台湾共同数据区间1994—2014年男性和女性分年龄60—89岁死亡率数据进行联合建模。考虑到中国大陆人口基数较大，为了充分利用死亡率信息，以中国大陆地区的风险暴露数为基准，按1:1的比例调整中国台湾地区的风险暴露数，再以台湾地区的中心死亡率调整死亡人数，最后运用极大似然方法对死亡人数进行建模。

选择赤池信息准则 (Akaike Information Criterion, AIC) 和贝叶斯信息准则 (Bayes Information Criterion, BIC) 来评估模型的拟合效果。其中，

$$AIC = 2k - 2\ln(L) \quad (15)$$

$$BIC = k\ln(n) - 2\ln(L) \quad (16)$$

其中， $k$ 为模型参数个数， $n$ 为样本量， $L$ 为似然函数。AIC准则和BIC准则的区别在于BIC准则考虑了样本量，对模型复杂度的惩罚力度更大，当样本量较大时，可以有效防止模型拟合精度过高而造成的模型复杂度过高的现象。AIC和BIC值越小模型效果越好。

表1给出了Logistic多人口模型的2个模型拟合得到的AIC值和BIC值。模型(6)的AIC和BIC值均比模型(5)小了大约13%，因此选择模型(6)对数据进行建模。

| 指标  | 模型(5)   | 模型(6)   |
|-----|---------|---------|
| AIC | 1846620 | 1606751 |
| BIC | 1847845 | 1608466 |

我们还可以通过计算样本内平均绝对百分比误差 (Mean Absolute Percentage Error, MAPE) 来评价模型的拟合效果，MAPE表达式为：

<sup>①</sup> 人类死亡率数据库HMD (Human Mortality Database)，网址 <http://www.mortality.org>。

$$MAPE = \sum_{i=1}^n \left| \frac{observed_i - predicted_i}{observed_i} \right| * \frac{100}{n} \quad (17)$$

表2给出了模型(1)一模型(6)中中国大陆、中国台湾拟合结果的MAPE值, 其中模型(1)为单人口CBD模型, 模型(2)、(3)为仅加入性别项的两人口模型, 模型(5)、(6)为加入了性别项和人口项的多人口模型。结果显示, 无论是男性还是女性, 加入性别项的两人口模型与同时

| 模型    | 中国大陆 |       | 中国台湾 |      |
|-------|------|-------|------|------|
|       | 男性   | 女性    | 男性   | 女性   |
| CBD模型 | 9.51 | 10.83 | 5.29 | 4.27 |
| 模型(2) | 9.26 | 10.76 | 4.47 | 3.33 |
| 模型(3) | 8.92 | 10.23 | 2.85 | 2.79 |
| 模型(5) | 9.26 | 10.76 | 4.47 | 3.33 |
| 模型(6) | 8.92 | 10.23 | 2.85 | 2.79 |

加入性别和人口项的多人口模型的MAPE值相同, 均小于单人口CBD模型, 而加入曲率项的模型(3)和模型(6)的男女死亡率拟合结果均优于没有加曲率项的模型。与CBD模型相比, 在模型(6)下的拟合结果中国大陆男女MAPE值分别减小了6%和5.5%, 中国台湾男女MAPE值分别减小了46%和35.7%, 拟合效果有明显改善。

运用模型(6)对数据进行联合建模, 图1给出了模型(6)的拟合残差图, 可见, 中国大陆、中国台湾男女的残差分布均满足随机性。对于中国大陆而言, 可以看到数据质量较好的普查年份和1%抽样调查年份的残差明显较小, 说明模型对这部分的数据拟合效果较好, 这也符合我们的预期结果。对于中国台湾而言, 整体拟合残差值较小。

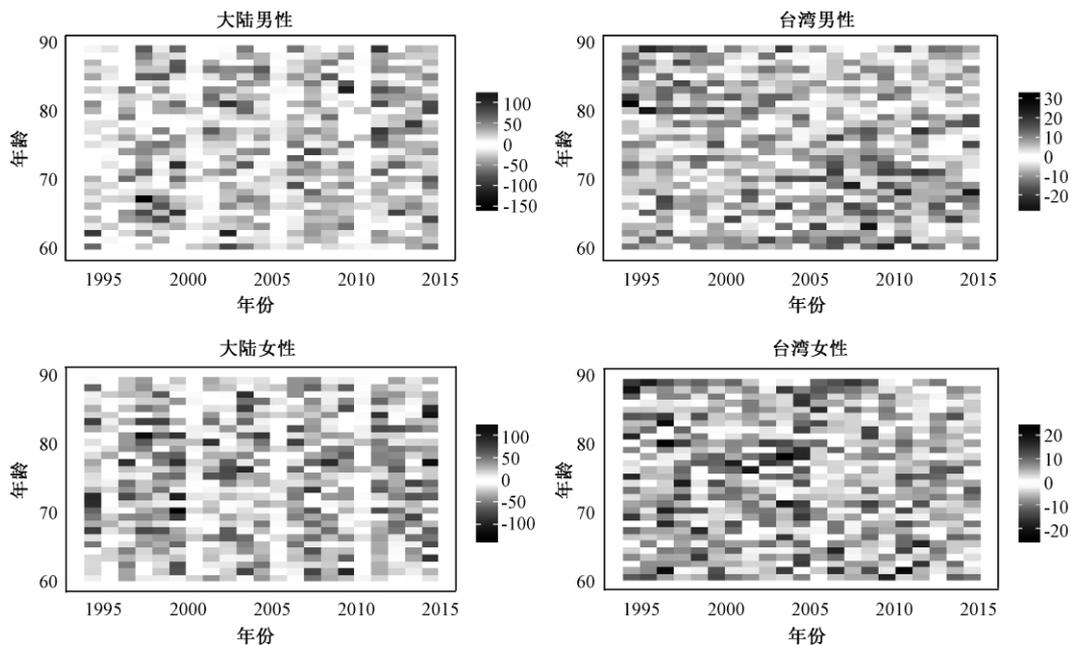


图1 模型(6)的拟合残差图

图2给出了模型(6)下中国大陆男女死亡率原始数据和修匀结果, 可以看出模型(6)修匀效果有较好的光滑性, 男女死亡率随时间变化呈递减趋势。

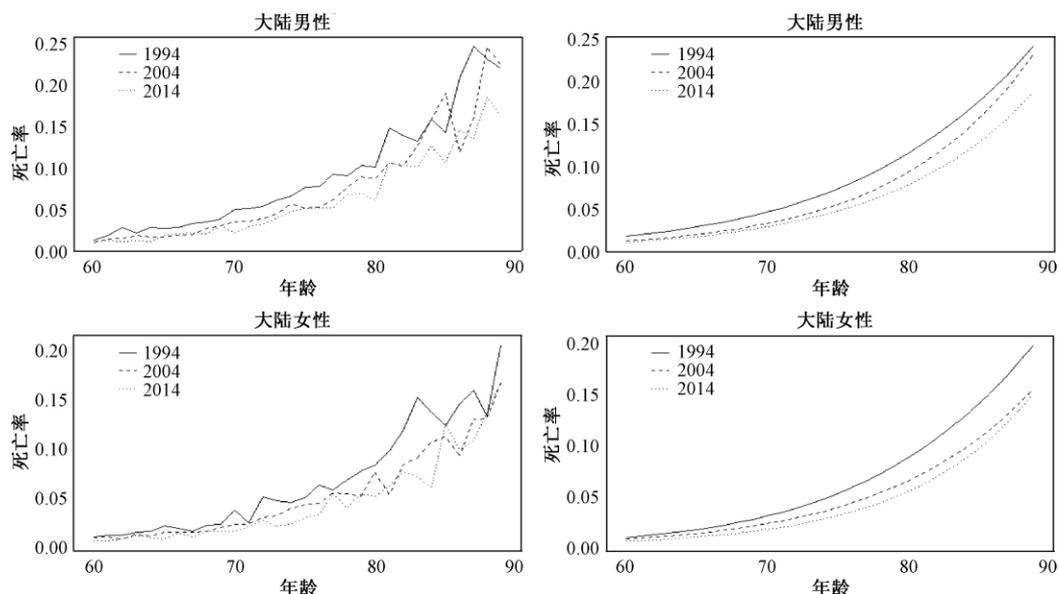


图2 中国大陆男女死亡率在模型(6)下的修匀结果

对于上述拟合结果，考虑到我国大陆地区分年龄死亡率波动性较大，高龄部分数据稀疏的情况，我们选择数据量较大，数据质量较好的普查年和1%抽样调查年的1995年、2000年、2005年、2010年四年的分性别死亡率数据，将几个模型的拟合结果做对比。表3给出了中国大陆地区这四个年份分性别拟合结果的MAPE值。结果表明，对于中国大陆地区，模型(6)对数据拟合优于CBD模型，而且对于数据质量较好的年份，男女MAPE值分别减少了29%和25.85%，拟合结果也优于单人口的CBD模型。

表3 中国大陆地区 MAPE 值 %

| 性别 | CBD 模型 | 模型 (2) | 模型 (3) | 模型 (5) | 模型 (6) |
|----|--------|--------|--------|--------|--------|
| 男性 | 4.26   | 3.19   | 3.02   | 3.19   | 3.02   |
| 女性 | 4.38   | 3.68   | 3.25   | 3.68   | 3.25   |

图3、图4分别给出了中国大陆地区男性和女性在CBD模型和模型(6)下1995年、2000年、2005年、2010年的拟合效果图。可见，无论是男性还是女性，CBD模型拟合结果均低于模型(6)，模型(6)在CBD模型的基础上提高了拟合精度。

### 3. 死亡率预测

在死亡率预测中，为确保预测一致性，假设共同的参数遵循带漂移的随机游走，性别参数和人口别参数均遵循自回归过程AR(1)。图5给出了共同参数 $\alpha_t$ 和 $\beta_t$ 的估计和预测结果，参数 $\alpha_t$ 反映了死亡率整体下降的趋势，参数 $\beta_t$ 呈缓慢下降趋势。图6给出了中国大陆分性别参数的估计和预测结果，可以看到分性别和人口别的参数在弱平稳过程下收敛到常数，同时也给出了参数的95%预测区间。

使用上述参数预测值，可以对死亡率做出预测，将死亡率外推到2050年，并与单人口

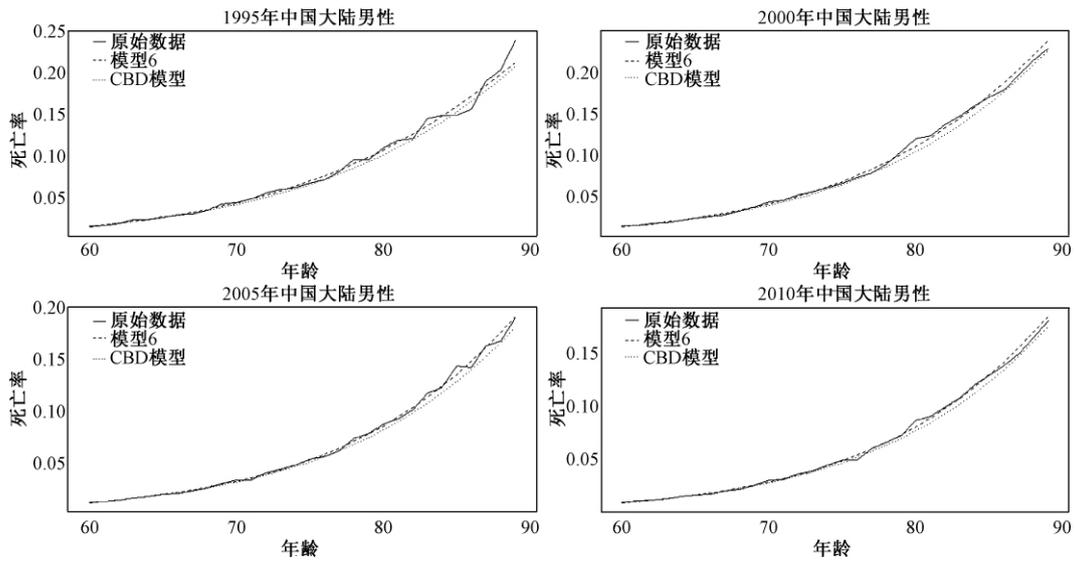


图3 中国大陆男性在 CBD 模型和模型 (6) 下的拟合结果

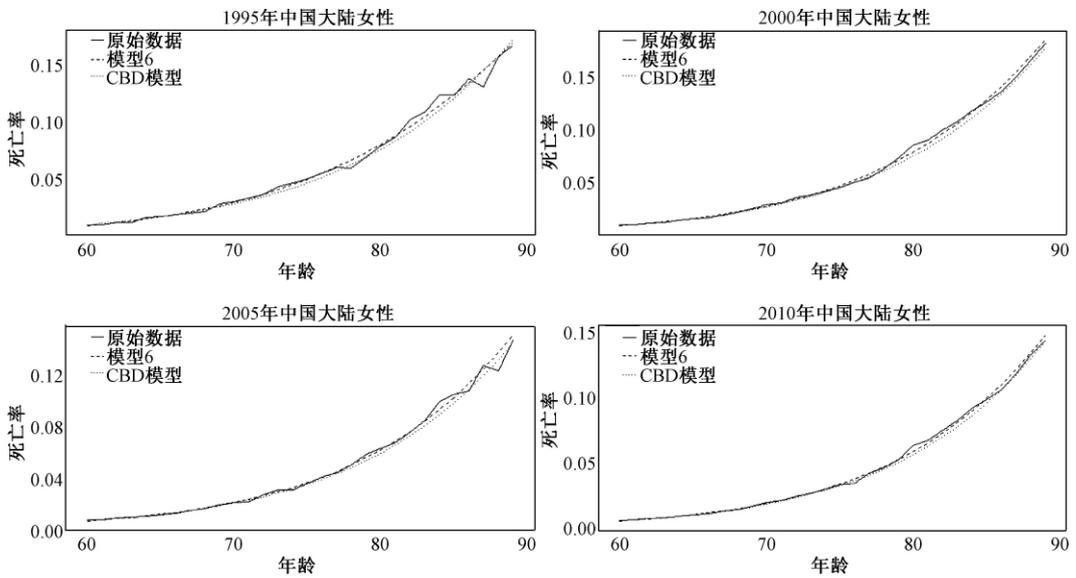


图4 中国大陆女性在 CBD 模型和模型 (6) 下的拟合结果

死亡率预测结果进行对比。图7给出了中国大陆地区、中国台湾地区在 CBD 模型、模型(3)、模型(6)下2050年的分性别分地区的预测结果对比图。图中 CBD 模型死亡率曲线从上到下依次为中国大陆男性、中国台湾男性、中国大陆女性、中国台湾女性, 可以看到在长期预测中, CBD 模型男女死亡率产生偏离现象; 模型(3)结果显示男女死亡率呈现一致性预测, 男女死亡率曲线趋于平行, 但不同人口之间死亡率差异性较明显; 模型(6)的结果显示, 男女死亡率呈现一致性预测, 男女死亡率曲线趋于平行, 且不同人口之间死亡率也呈现一致性预测。

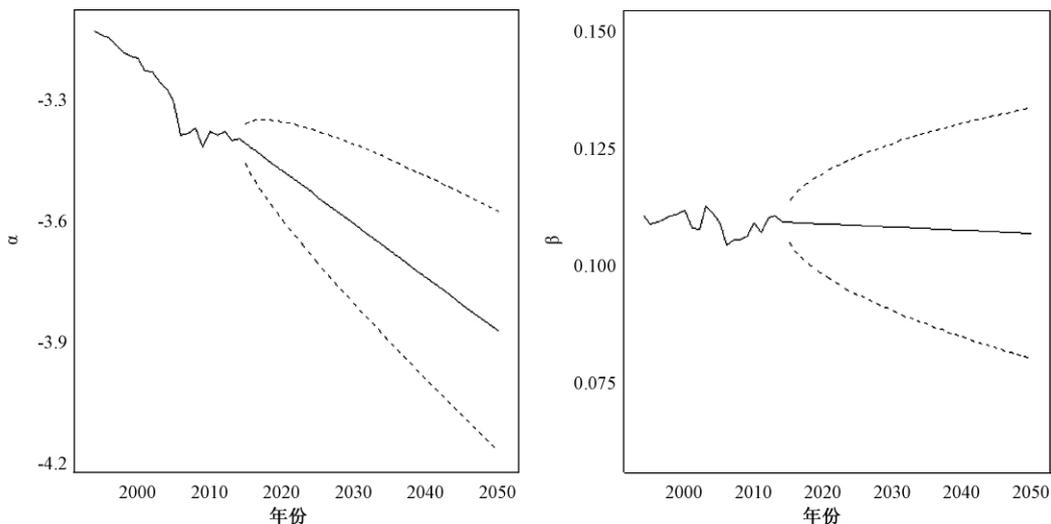


图5 共同参数的估计和预测值（实线）和95%预测区间（虚线）

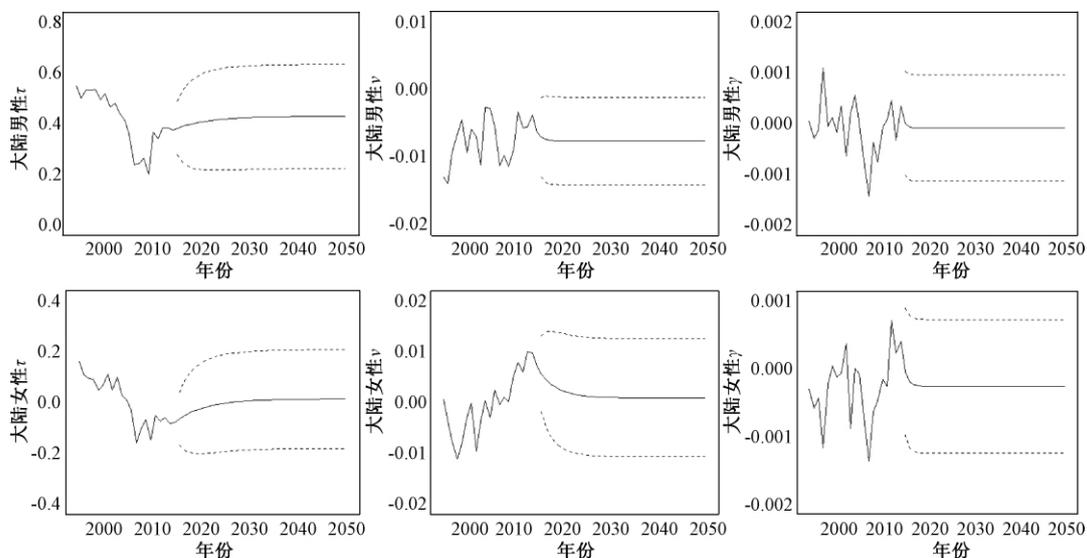


图6 中国大陆分性别参数的估计和预测值（实线）和95%预测区间（虚线）

在模型（6）参数估计的结果下，可以计算得出中国大陆地区男女死亡率预测值。表4给出了2020年、2030年、2040年、2050年的60—89岁死亡率预测值。可见，随着时间推移，男女死亡率均呈下降趋势，60岁男性死亡率从2020年的0.0105降到了2050年的0.0075，大约降低28.6%；60岁女性死亡率从2020年的0.0058降到2050年的0.0043，约降低25.9%。同一时期女性死亡率明显低于男性，随着时间推移男女死亡率差距逐渐减小。

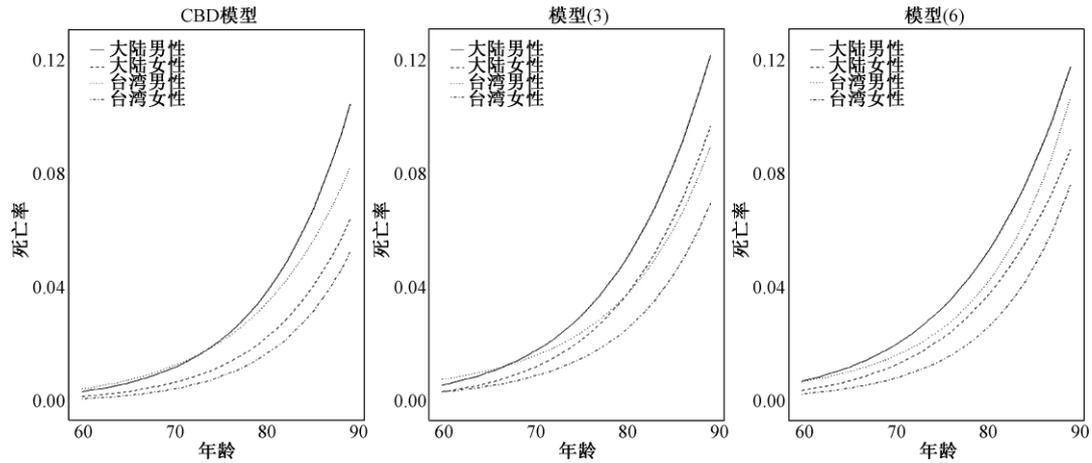


图7 2050年中国大陆和中国台湾分男女死亡率预测结果

表4 中国大陆地区男女死亡率预测值

| 年龄 | 2020年  |        | 2030年  |        | 2040年  |        | 2050年  |        |
|----|--------|--------|--------|--------|--------|--------|--------|--------|
|    | 男性     | 女性     | 男性     | 女性     | 男性     | 女性     | 男性     | 女性     |
| 60 | 0.0105 | 0.0058 | 0.0095 | 0.0054 | 0.0084 | 0.0048 | 0.0075 | 0.0043 |
| 61 | 0.0117 | 0.0066 | 0.0105 | 0.0061 | 0.0093 | 0.0054 | 0.0082 | 0.0048 |
| 62 | 0.0129 | 0.0074 | 0.0116 | 0.0068 | 0.0103 | 0.0061 | 0.0091 | 0.0054 |
| 63 | 0.0143 | 0.0083 | 0.0128 | 0.0076 | 0.0114 | 0.0068 | 0.0101 | 0.0060 |
| 64 | 0.0158 | 0.0093 | 0.0142 | 0.0086 | 0.0126 | 0.0076 | 0.0111 | 0.0067 |
| 65 | 0.0175 | 0.0104 | 0.0157 | 0.0096 | 0.0139 | 0.0085 | 0.0123 | 0.0075 |
| 66 | 0.0193 | 0.0117 | 0.0174 | 0.0107 | 0.0154 | 0.0095 | 0.0136 | 0.0084 |
| 67 | 0.0214 | 0.0131 | 0.0192 | 0.0120 | 0.0170 | 0.0106 | 0.0150 | 0.0094 |
| 68 | 0.0236 | 0.0147 | 0.0212 | 0.0134 | 0.0188 | 0.0119 | 0.0166 | 0.0105 |
| 69 | 0.0261 | 0.0164 | 0.0234 | 0.0150 | 0.0207 | 0.0133 | 0.0183 | 0.0117 |
| 70 | 0.0288 | 0.0184 | 0.0259 | 0.0167 | 0.0229 | 0.0148 | 0.0202 | 0.0130 |
| 71 | 0.0318 | 0.0205 | 0.0285 | 0.0186 | 0.0252 | 0.0165 | 0.0222 | 0.0145 |
| 72 | 0.0351 | 0.0229 | 0.0315 | 0.0208 | 0.0278 | 0.0184 | 0.0245 | 0.0162 |
| 73 | 0.0387 | 0.0256 | 0.0347 | 0.0231 | 0.0307 | 0.0205 | 0.0270 | 0.0180 |
| 74 | 0.0427 | 0.0285 | 0.0382 | 0.0258 | 0.0338 | 0.0228 | 0.0297 | 0.0200 |
| 75 | 0.0470 | 0.0317 | 0.0421 | 0.0286 | 0.0372 | 0.0253 | 0.0327 | 0.0222 |
| 76 | 0.0517 | 0.0353 | 0.0463 | 0.0318 | 0.0409 | 0.0281 | 0.0360 | 0.0247 |
| 77 | 0.0569 | 0.0392 | 0.0510 | 0.0353 | 0.0450 | 0.0312 | 0.0396 | 0.0274 |
| 78 | 0.0626 | 0.0436 | 0.0560 | 0.0392 | 0.0495 | 0.0346 | 0.0435 | 0.0304 |
| 79 | 0.0687 | 0.0483 | 0.0615 | 0.0434 | 0.0544 | 0.0383 | 0.0478 | 0.0336 |
| 80 | 0.0755 | 0.0536 | 0.0676 | 0.0480 | 0.0597 | 0.0424 | 0.0525 | 0.0372 |
| 81 | 0.0827 | 0.0593 | 0.0741 | 0.0531 | 0.0655 | 0.0469 | 0.0576 | 0.0412 |
| 82 | 0.0907 | 0.0656 | 0.0812 | 0.0587 | 0.0718 | 0.0518 | 0.0632 | 0.0455 |
| 83 | 0.0992 | 0.0724 | 0.0889 | 0.0647 | 0.0786 | 0.0571 | 0.0692 | 0.0502 |
| 84 | 0.1085 | 0.0799 | 0.0973 | 0.0714 | 0.0861 | 0.0630 | 0.0758 | 0.0553 |
| 85 | 0.1185 | 0.0880 | 0.1063 | 0.0785 | 0.0941 | 0.0693 | 0.0829 | 0.0609 |
| 86 | 0.1293 | 0.0968 | 0.1161 | 0.0864 | 0.1028 | 0.0762 | 0.0906 | 0.0670 |
| 87 | 0.1409 | 0.1063 | 0.1266 | 0.0948 | 0.1122 | 0.0837 | 0.0990 | 0.0736 |
| 88 | 0.1533 | 0.1166 | 0.1379 | 0.1040 | 0.1223 | 0.0919 | 0.1080 | 0.0808 |
| 89 | 0.1666 | 0.1277 | 0.1500 | 0.1138 | 0.1332 | 0.1006 | 0.1177 | 0.0885 |

## 五、结论

本文在 CBD 模型和 Logistic 两人口死亡率模型基础上对模型进行扩展,在分性别因子基础上加入国家和地区别因子,提出了 Logistic 多人口模型,用于对多个人口分性别的联合建模,以捕捉不同人口死亡率的共同特性,更好地预测死亡率变动趋势。模型打破了单人口模型与两性别或多人口联合建模的局限,将性别与人口别同时联合建模,将模型推向高维。

针对中国大陆地区高龄死亡率数据波动大,数据质量差的问题,我们选择中国大陆地区、台湾地区分男女死亡率相关数据,以中国大陆数据为基础,并按 1:1 数据量调整台湾数据,对其进行联合建模和预测,来拟合和预测中国大陆未来死亡率变化模式;与单人口 CBD 模型相比,从拟合结果 MAPE 值来看,中国大陆、中国台湾男女死亡率拟合效果均有所改善。在长期预测中,Logistic 多人口模型避免了单人口 CBD 模型可能出现的死亡率交叉或偏离问题,能够给出死亡率的一致预测。

在进一步的研究中,Logistic 多人口模型还可以尝试加入队列效应对模型进行扩展,以适用于能够捕捉到队列效应国家或地区的数据;也可以采用贝叶斯技术,该技术可以提供合理的框架来同时处理死亡率结构和时间序列过程。

## 参考文献:

- [1] LI N, LEE R. Coherent mortality forecasts for a group of populations: an extension of the Lee-carter method [J]. Demography, 2005, 42 (3): 575-594.
- [2] LEE R D, CARTER L R. Modeling and forecasting U. S. mortality [J]. Journal of the American Statistical Association, 1992, 87 (419): 659-671.
- [3] RENSHAW A E, HABERMAN S. A cohort-based extension to the Lee-carter model for mortality reduction factors [J]. Insurance: Mathematics and Economics, 2006, 38 (3): 556-570.
- [4] CURRIE I D. Smoothing and forecasting mortality rates with p-splines [R]. Institute of Actuaries, 2006.
- [5] CAIRNS A J G, BLAKE D, DOWD K. A two-factor model for stochastic mortality with parameter uncertainty: theory and calibration [J]. Journal of Risk and Insurance, 2006, 73 (4): 687-718.
- [6] CAIRNS A J G, BLAKE D, DOWD K et. al. A quantitative comparison of stochastic mortality models using data from England & Wales and the United States [J]. North American Actuarial Journal, 2009, 13 (1): 1-35.
- [7] LI J. A poisson common factor model for projecting mortality and life expectancy jointly for females and males [J]. Population Studies, 2013, 67 (1): 111-126.
- [8] LI J, TICKLE L, PARR N. A multi-population evaluation of the poisson common factor model for projecting mortality jointly for both sexes [J]. Journal of Population Research, 2016, 33 (4): 333-360.
- [9] CHEN R Y, MILLOSSOVICH P. Sex-specific mortality forecasting for UK countries: a coherent approach [J]. Eur. Actuar. J, 2018, 8 (1): 69-95.
- [10] LI J, LIU J. A Logistic two-population mortality projection model for modelling mortality at advanced ages for both sexes [J]. Scandinavian Actuarial Journal, 2019 (2): 97-112.
- [11] 柳向东, 范洋洋. 基于均衡关系的中国人口死亡率预测模型 [J]. 统计与信息论坛, 2016 (10): 3-9.
- [12] 段白鸽. 我国全年龄段人口平均预期寿命的动态演变 [J]. 人口与经济, 2015 (1): 49-63.

[责任编辑 刘爱华]