



Mortality Prediction and Application Based on APC Model

Hongmin Xiao, Haifei Ma^{*}, Hongyu Zhao

Department of Mathematics, Northwest Normal University, Lanzhou, China

Email address:

xiaohm9@126.com (Hongmin Xiao), findmhf@126.com (Haifei Ma), 904160990@qq.com (Hongyu Zhao)

^{*}Corresponding author

To cite this article:

Hongmin Xiao, Haifei Ma, Hongyu Zhao. Mortality Prediction and Application Based on APC Model. *Asia-Pacific Journal of Mathematics and Statistics*. Vol. 2, No. 2, 2020, pp. 5-9.

Received: March 24, 2020; **Accepted:** September 13, 2020; **Published:** September 23, 2020

Abstract: In this paper, the APC (age period cohort) model with year of birth effect is used to fit and predict China's population mortality data. Firstly, the parameters of APC model are estimated by maximum likelihood method. Through residual analysis of the model, it is found that the model captures the effects of age, calendar year and birth year, and the data fitting effect is good. Then the ARIMA time series model is used to fit the time index and the birth year index of the mortality model. The optimal ARIMA model is selected by AICc value for prediction, and then the predicted value of mortality rate is obtained, and the prediction effect is compared according to the real value of mortality. Finally, the life expectancy of population is predicted by using the predicted mortality value, and part of the age life expectancy table is drawn. The results show that: from the perspective of fitting effect and prediction effect, APC model is suitable for predicting China's population mortality rate. At the same time, we find that China's population life expectancy tends to be gradually extended, and the aging problem is becoming increasingly serious, which will bring huge financial pressure to China's life insurance enterprises and pension institutions.

Keywords: APC Model, Residual Analysis, ARIMA Model, Life Expectancy

基于APC模型的死亡率预测及应用

肖鸿民, 马海飞^{*}, 赵弘宇

西北师范大学数学系, 兰州, 中国

邮箱

xiaohm9@126.com (肖鸿民), findmhf@126.com (马海飞), 904160990@qq.com (赵弘宇)

摘要: 本文运用带有出生年效应的APC(Age-Period-Cohort)模型对中国人口死亡率数据进行拟合与预测. 首先利用极大似然法对APC模型的参数进行估计, 通过对模型进行残差分析, 发现该模型捕捉到了年龄、日历年和出生年效应, 对数据拟合效果较好. 然后运用ARIMA时间序列模型来拟合死亡率模型的时间指数和出生年指数, 通过AICc值选出最优的ARIMA模型进行预测, 进而得到死亡率预测值, 并根据死亡率真实值对其预测效果进行比较. 最后利用计算得到的死亡率预测值对人口预期寿命进行预测, 绘制部分年龄预期寿命表. 研究结果显示: 从拟合效果和预测效果来看, APC模型适合于预测中国人口死亡率, 同时我们发现中国人口预期寿命有逐渐延长的趋势, 老龄化问题日益严重, 这将给我国的寿险企业和养老金机构带来巨大的财务压力。

关键词: APC模型, 残差分析, ARIMA模型, 预期寿命

1. 引言

随着社会经济的不断发展,我国医疗水平进一步提高,人口死亡率的降低和预期寿命的延长已经成为一种全球性趋势.国家统计局公布的数据显示,中国人口的预期寿命从解放前的34.74岁,提高到1990年的68.55岁,到2015年时已经提高到76.34岁.老龄化问题日益严重,给老年保障体系带来了巨大的财务压力.因此,选择适合我国人口数据的随机死亡率模型,提高死亡率预测的精准性,为养老金机构死亡率风险管理提供科学依据。

死亡率的预测对人口科学研究和养老金制度安排是至关重要的,而死亡率预测首先要从死亡率建模开始.在国外,使用最为广泛的是由Lee等[1]提出的Lee-Carter模型,之后各国学者结合本国数据对死亡率进行拟合与预测.Renshaw等[2]首次提出了考虑出生年份效应的死亡率模型,即RH模型.运用英格兰和威尔士男性死亡率数据进行拟合,并将其与Lee-Carter模型进行对比后发现扩展的RH模型更优.Currie等[3]在RH模型的基础上提出了一个简化形式,即APC模型.发现其对美国历史数据有很高的拟合度,并且避免了RH模型参数估计时出现的稳定性问题. Steven等[4]运用APC模型对死亡率进行预测并计算了不同时期的预期寿命.Brouhns等[5]运用极大似然方法对Lee-Carter模型的参数进行估计. Cairns等[6]通过对六个随机死亡率模型的分析,发现要从拟合和预测两个方面来判定一个模型的合理性.Li等[7]用扩展的Lee-Carter模型对日本和美国的数据进行拟合,预测死亡率及其预期寿命.在国内的研究中,对人口死亡率并没有进行深入探讨,年金和保险机构大多是依据静态的死亡率数据进行一些简单的分析预测.李志生等[8]运用Lee-Carter模型对中国人口死亡率数据进行拟合和预测,并在此基础上对人口平均预期寿命进行了推算.之后黄顺林等[9]运用RH模型对中国男性人口死亡率进行了预测,并将其应用到预期寿命预测和年金系数的估计中.韩猛等[10]对Lee-Carter模型进行了改进,通过一个双随机过程对Lee-Carter模型中的时间项进行建模.王晓军等[11]未对缺失数据进行处理,而是筛选完整的6个间隔年份数据通过韩猛提出的一个双随机过程对有限数据下的中国人口死亡率进行预测.此外,杨泽祥等[12]运用死亡结构法结合粗死亡率推算出了平均预期寿命.苏晶晶等[13]对APC模型参数估计研究最新进展进行详细介绍,并对各种参数估计方法的优缺点和适用范围进行比较说明.曾燕等[14]采用Bootstrap方法与Lee-Carter模型拟合死亡率数据,运用带漂移项的随机游走模型进行预测.曹园[15]运用Lee-Carter模型对中国人口死亡率进行预测后,根据预测的死亡率数据对中国人口平均预期寿命进行计算.樊毅等[16]基于中国全人口死亡率数据,对八个随机死亡率模型进行了全面的比较分析,发现由Lee-Carter模型扩展而来的APC模型最适合于中国人口死亡率的拟合和预测.王晓军等[17]运用台湾高龄人口死亡率数据对贝叶斯分层模型等八种死亡率模型进行了比较分析,最后通过拟合中国大陆人口数据,发现

CBD模型对我国大陆地区高龄死亡率的拟合和预测较好。

值得注意的是,目前国内学者对中国人口死亡率的研究大多是基于Lee-Carter模型来进行的,对文[11]中提出的最适合中国人口死亡率拟合的APC模型,国内很少有人将其应用到死亡率的预测和长寿风险的度量中.考虑到Lee-Carter模型最初的提出是以美国人口死亡率数据为基础,而我国死亡率数据与外国数据差异较大,所以本文将运用APC模型,结合中国人口死亡率数据,对模型进行参数估计,并预测中国人口死亡率和平均预期寿命,为相关机构度量长寿风险提供更多可靠的依据。

2. Age-Period-Cohort模型

Renshaw and Haberman(2006)首先提出了带有出生年效应的人口死亡率模型,具体形式如下:

$$\ln m_{x,t} = \alpha_x + \beta_x^{(1)} k_t + \beta_x^{(2)} \gamma_{t-x} \quad (1)$$

其中, $m_{x,t}$ 表示年龄为 x 岁时间为 t 的中心死亡率; α_x 、 $\beta_x^{(1)}$ 和 $\beta_x^{(2)}$ 为依赖于年龄因子 x 的参数; k_t 表示死亡率随时间 t 的变化程度; γ_{t-x} 是一个随机的出生年效应,反映出出生年 $(t-x)$ 对死亡率的影响.为了在模型的参数估计中得到唯一解,需对模型的参数进行规范化约定,即同时满足约束条件:

$$\sum_x \beta_x^{(1)} = 1, \sum_t k_t = 0, \sum_{t-x} k_{t-x} = 0, \sum_x \beta_x^{(2)} = 1 \quad (2)$$

在对英格兰和威尔士的数据进行分析后,Renshaw and Haberman(2006)发现使用加入出生年效应的模型相对Lee-Carter模型有了明显的改善,但被证实RH模型是缺少稳定性的.为此Currie(2006)在RH模型的基础上提出了一个简化形式,即Age-Period-Cohort模型,具体形式如下:

$$\ln m_{x,t} = \alpha_x + k_t + \gamma_{t-x} \quad (3)$$

显然它是RH模型的一个特例,只需令 $\beta_x^{(1)} = \beta_x^{(2)} = 1$ 即可,其他符合与参数意义和前文相同。

对上述APC模型进行参数估计时,通常假定死亡人数 $d_{x,t}$ 服从泊松分布,即

$$d_{x,t} \sim \text{Poisson}(m_{x,t} e_{x,t}) \quad (4)$$

其中, $d_{x,t}$ 表示年龄为 x 时间为 t 的死亡人数, $e_{x,t}$ 表示年龄为 x 时间为 t 的平均暴露人数, $m_{x,t} = \exp(\alpha_x + k_t + \gamma_{t-x})$. 则APC模型的似然函数为:

$$L = \sum_{x,t} [d_{x,t}(\alpha_x + k_t + \gamma_{t-x}) - e_{x,t} \exp(\alpha_x + k_t + \gamma_{t-x})] + C \quad (5)$$

上式中 C 为常数,牛顿迭代公式为:

$$\hat{\theta}^{(n+1)} = \hat{\theta}^{(n)} - \frac{\partial L^{(n)} / \partial \theta}{\partial^2 L^{(n)} / \partial \theta^2} \quad (6)$$

其中 $\hat{\theta}^{(n)}$ 是第 n 次迭代参数, $L^{(n)} = L(\theta^{(n)})$ 。由此得到APC模型的参数更新公式为:

$$\hat{\alpha}_x^{(n+1)} = \hat{\alpha}_x^{(n)} - \frac{\sum_t w_{x,t}(d_{x,t} - \hat{d}_{x,t})}{\sum_t w_{x,t} \hat{d}_{x,t}} \quad (7)$$

$$\hat{k}_t^{(n+1)} = \hat{k}_t^{(n)} - \frac{\sum_x w_{x,t}(d_{x,t} - \hat{d}_{x,t})}{\sum_x w_{x,t} \hat{d}_{x,t}} \quad (8)$$

$$\hat{\gamma}_c^{(n+1)} = \hat{\gamma}_c^{(n)} - \frac{\sum_{t-x=c} w_{x,t}(d_{x,t} - \hat{d}_{x,t})}{\sum_{t-x=c} w_{x,t} \hat{d}_{x,t}} \quad (9)$$

其中 $\hat{d}_{x,t} = e_{x,t} \exp(\alpha_x + k_t + \gamma_{t-x})$, $w_{x,t}$ 表示权重,当数据缺失时权重为0,否则均为1。当参数更新使式(5)的增量足够小(通常为小于 10^{-10})时停止迭代,得到参数估计值。

3. 死亡率预测

3.1. 数据与拟合结果

本文采用《中国人口统计年鉴》中1995~2014年连续20年的中国人口0~90岁死亡率数据。由于大部分日历年数据的末组都为90+,所以本文的年龄上限设为90,其中1996年数据的最高年龄是85岁以上,对于数据不足90+的部分,用均值插值法进行补充处理,对1995、2000、2005和2010年数据中100+的部分进行合并处理,然后结合极大似然法得到模型参数拟合结果(如图1)。

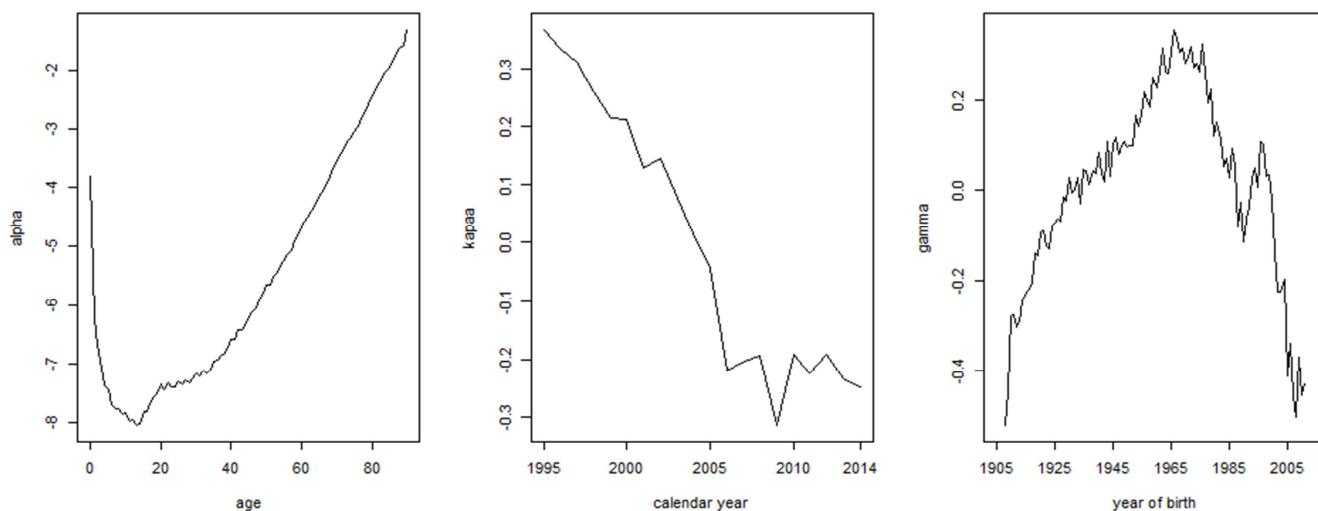


图1 APC模型参数估计图。

参数 α_x 反映了死亡率随年龄变化的情况,从图1可以看出,随着年龄的增长 α_x 呈现出先下降后上升的趋势; k_t 的估计值随着时间的推移而逐渐下降,反映出死亡率随时间改善的趋势;参数 γ_{t-x} 表示的是出生年对死亡率的影响,图1中随着时间的变化, γ_{t-x} 呈现出先上升后下降的趋势(个别波动除外)。

图2是死亡率分别关于年龄、日历年和出生年的残差图,从图中可以看出APC模型的拟合散点图在年龄、时间和出生年效应上的分布均匀且接近零轴,偏移程度较小,说明APC模型的拟合效果较好。

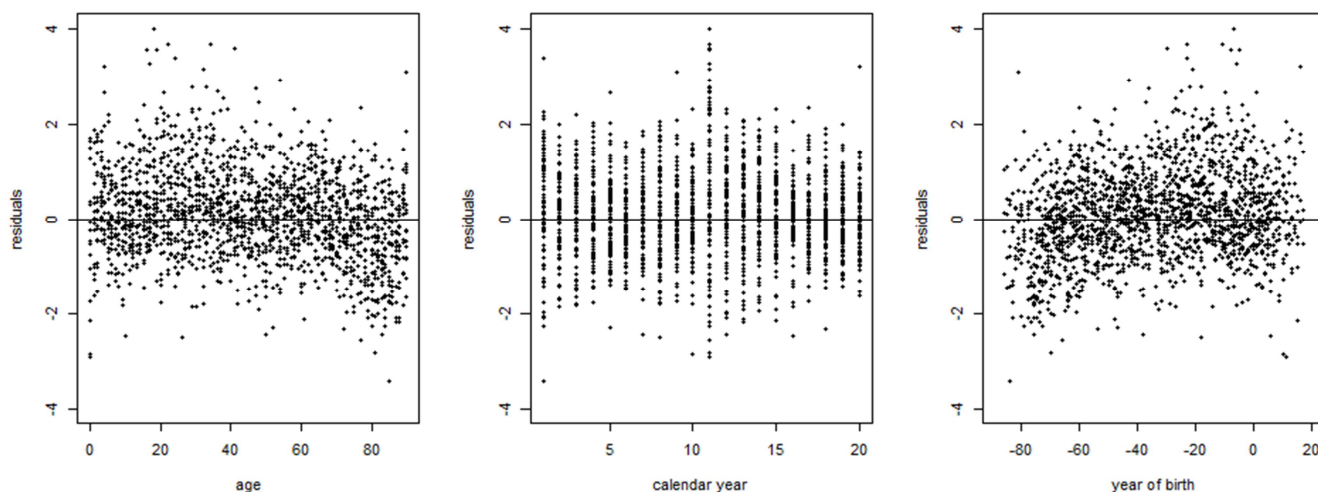


图2 APC模型拟合残差图。

3.2. 死亡率预测

APC模型中死亡率的动态变化受时间指数和出生年指数的影响,因此本文对于参数估计值 \hat{k}_t 、 $\hat{\gamma}_{t-x}$ 序列,分别运用ARIMA(p,d,q)模型去拟合,通过每个拟合过程得到的AICc值来判断模型拟合结果的好坏,模型选择的结果如表1所示。

表1 ARIMA模型的选择。

$\hat{\gamma}_{t-x}$ 序列ARIMA模型选择	AICc	\hat{k}_t 序列ARIMA模型选择	AICc
ARIMA(2,2,2)	-304.68	ARIMA(2,1,2) with drift	1e+20
ARIMA(0,2,0)	-202.65	ARIMA(0,1,0) with drift	-47.74
ARIMA(1,2,0)	-252.81	ARIMA(1,1,0) with drift	-47.02
ARIMA(0,2,1)	-290.09	ARIMA(0,1,1) with drift	-46.45
ARIMA(1,2,2)	-296.97	ARIMA(0,1,0)	-45.56
ARIMA(2,2,1)	-298.35	ARIMA(1,1,1) with drift	-43.81
ARIMA(3,2,2)	-304.42		
ARIMA(2,2,3)	-311.40		
ARIMA(1,2,3)	-298.90		
ARIMA(3,2,3)	-302.21		

从表1可以看出, \hat{k}_t 和 $\hat{\gamma}_{t-x}$ 序列的AICc最小值为-311.40和 -47.74, 所以分别用ARIMA(2,2,3)和带漂移项的ARIMA(0,1,0)模型对 \hat{k}_t 和 $\hat{\gamma}_{t-x}$ 序列进行预测,得到 \hat{k}_t 和 $\hat{\gamma}_{t-x}$ 的预测值 \hat{k}_t 、 $\hat{\gamma}_{t-x}$,再根据 $\ln \hat{m}_{x,t} = \alpha_x + \hat{k}_t + \hat{\gamma}_{t-x}$,可以得到死亡率的预测值 $\hat{m}_{x,t}$ 。

图3是2015~2017年的死亡率实际数据与预测数据对比图,图中除去65岁以上死亡率预测值与真实值稍有差异,其他数据与真实值均非常接近,拟合较好.说明用APC模型来拟合和预测中国人口死亡率是比较可信的。

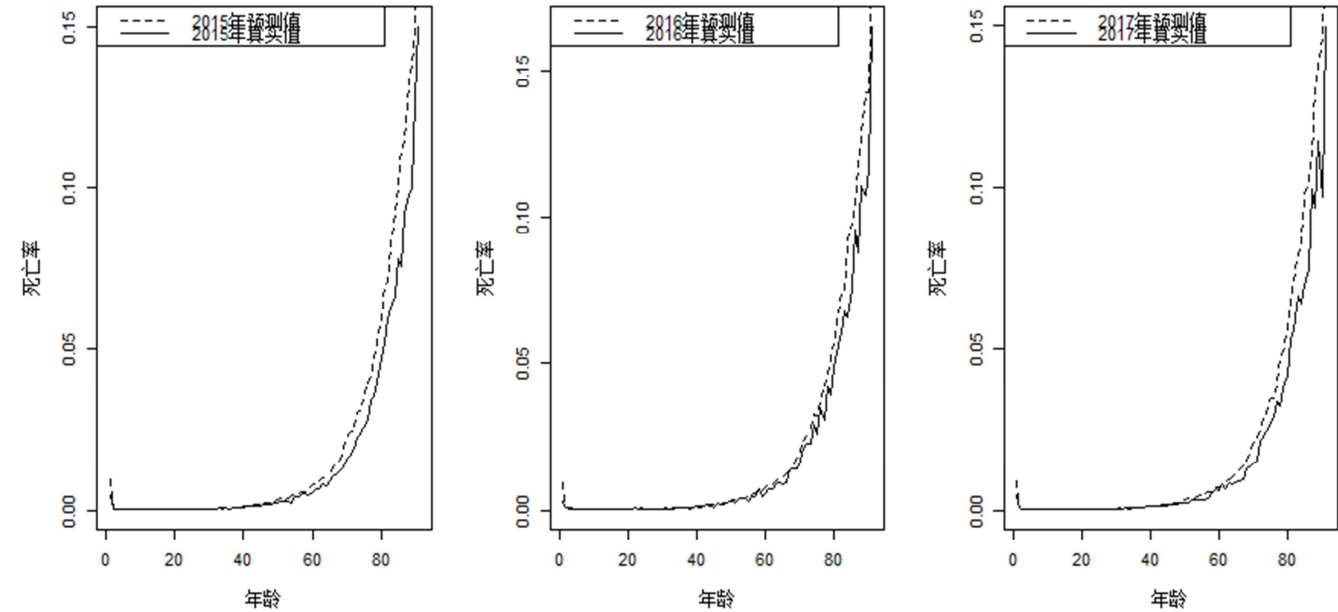


图3 2015~2017年死亡率数值对比。

4. 预期寿命预测

人口死亡率的降低和寿命的延长已经成为一种全球性趋势,为了更直观的反映这一情况,本节我们将用人口粗死亡率来计算人口预期寿命,具体步骤如下:

- (1) 普查年龄别平均人数 E_x 和死亡人数 D_x ,计算年龄别死亡率 m_x :

$$m_x = \frac{D_x}{E_x}, x = 0,1,2, \cdots, \omega - 1 \tag{10}$$

- (2) 用指数法死亡率概率公式计算死亡概率 q_x :

$$q_x = 1 - \exp(-m_x) \tag{11}$$

- (3) 确定死亡人口基数,一般取 $l_0 = 100000$;
- (4) 计算死亡人数 d_x :

$$d_x = l_0 * q_x \tag{12}$$

- (5) 计算尚存人数 l_x :

$$l_{x+1} = l_x - d_x, x = 0,1,2, \cdots, \omega - 1 \tag{13}$$

- (6) 计算平均生存人年数 L_x :

$$L_x = \begin{cases} \frac{1}{4}l_0 + \frac{3}{4}l_1, x = 0 \\ \frac{1}{2}(l_x + l_{x+1}) + \frac{1}{24}(d_x - d_{x-1}), x = 1, 2, 3, 4 \\ \frac{1}{2}(l_x + l_{x+1}), 5 \leq x \leq \omega - 2 \\ \frac{1}{2}(l_x), x = \omega - 1 \end{cases} \quad (14)$$

(7) 计算平均生存总人年数 T_x (向下累积):

$$T_x = \sum_{i=x}^{\omega-1} L_i, x = 0, 1, 2, \dots, \omega - 1 \quad (15)$$

(8) 计算平均预期寿命 e_x :

$$e_x = \frac{T_x}{l_x}, x = 0, 1, 2, \dots, \omega - 1 \quad (16)$$

根据上述步骤可以计算中国人口平均预期寿命,表2给出了部分年份中国人口不同年龄的预期寿命表.从表中不难看出,各年龄中国人口预期寿命出现了逐渐增长的趋势,其中新生儿的预期寿命从2015年的77.10岁增加到了2025年的79.60岁,同样其他年龄的预期余寿也明显增加.这将给养老金的管理带来巨大的困难,也是长寿风险的根本来源.本文基于APC模型对中国人口死亡率及预期寿命进行预测,为长寿风险的管理提供依据.

表2 不同年份中国人口不同年龄的预期寿命表。

年龄 (岁)	2015年	2017年	2019年	2021年	2023年	2025年
0	77.10	77.61	78.11	78.62	79.13	79.60
15	63.15	63.58	64.01	64.42	64.83	65.23
25	53.40	53.81	54.21	54.62	54.99	55.36
35	43.67	44.06	44.45	44.83	45.20	45.57
45	34.16	34.51	34.86	35.20	35.54	35.87
55	25.13	25.43	25.74	26.02	26.30	26.58
65	16.77	17.05	17.31	17.56	17.80	18.02
75	9.58	9.78	9.96	10.11	10.29	10.45
85	3.49	3.54	3.60	3.62	3.67	3.70

5. 结论

本文利用中国1995~2014年的死亡率历史数据对Lee-Carter扩展模型(APC模型)进行拟合,作出模型标准残差图.通过选择最优时间序列模型对死亡率进行预测,并与死亡率真实数据作对比,对中国人口平均预期寿命进行预测,得出以下结果:

拟合方面,从残差图来看,APC模型很好的捕捉到了年龄、日历年和出生年效应,并且均匀的分布在零轴两侧,相对稳定,说明APC模型适合于中国人口死亡率数据的拟合。

预测方面,根据死亡率对比图,发现2015~2017年死亡率真实值与模型预测值非常接近,其预测结果在一定程度上是比较可靠的.从基于粗死亡率推算出的中国人口平均预期寿命表可以看出,中国人口预期寿命有逐渐延长的趋势,这对社会保障、企业年金和保险公司带来了巨大的财务压力。

基金项目

国家自然科学基金(12061066)。

参考文献

- [1] Lee R D, Carter L R. Modeling and forecasting US mortality [J]. Journal of American Statistical Association, 1992,87(14):659-675.
- [2] 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.
- [3] Currie I D, Durban M, Eilers P H C. Generalized linear array models with applications to multidimensional smoothing[J]. Journal of the Royal Statistical Society, 2006,68(2):259-280.
- [4] Steven Haberman, Arthur Renshaw. On age-period-cohort parametric mortality rate projections[J]. Insurance Mathematics and Economics, 2009,45(2):255-270.
- [5] Brouhns N, Denuit M, Vermunt J. A Poisson log-linear regression approach to the construction of projected lifetables[J]. Insurance: Mathematics and Economics, 2002,31(3):373-393.
- [6] Cairns A J G, Blake D, Dowd K, et al. Mortality density forecasts: An analysis of six stochastic mortality models [J]. Social Science Electronic Publishing, 2011, 48(3): 355-367.
- [7] Li N, Lee R, Gerland P. Extending the Lee-Carter Method to Model the Rotation of Age Patterns of Mortality Decline for Long-Term Projections[J]. Demography, 2013,50(6):2037-2051.
- [8] 李志生,刘恒甲.Lee-Carter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):47-56.
- [9] 黄顺林,王晓军.加入出生年效应的死亡率预测及其在年金系数估计中的应用[J].统计与信息论坛,2010(05):81-86
- [10] 韩猛,王晓军.Lee-Carter模型在中国城市人口死亡率预测中的应用与改进[J].保险研究,2010,(10): 3-9。
- [11] 王晓军,任文东.有限数据下Lee-Carter模型在人口死亡率预测中的应用[J].统计研究,2012,(6):87-94。
- [12] 杨泽祥,赵萍,王欣.人口生命表编制的一种新方法:死亡结构法[J].统计与决策,2014(01):4-7
- [13] 苏晶晶,彭非.年龄-时期-队列模型参数估计方法最新研究进展[J].统计与决策,2014(23):21-26。
- [14] 曾燕,陈曦,邓颖璐.创新的动态人口死亡率预测及其应用[J].系统工程理论与实践,2016,36(7):1710-1718。
- [15] 曹园.基于Lee-Carter模型的我国死亡率预测[J].统计与决策,2018(9):32-36。
- [16] 樊毅,张宁.基于全人口死亡率数据的随机死亡率模型拟合效果比较[J].统计与决策,2018(23):33-37。
- [17] 王晓军,路倩.高龄人口死亡率预测模型的比较与选择[J].保险研究,2019(3):82-102。