

# Lee-Carter死亡率模型的估计与应用\*

## ——基于中国人口数据的分析

李志生 刘恒甲

**【摘要】**文章选择 Lee-Carter 死亡率模型对中国人口死亡率数据进行拟合和预测,以探讨 Lee-Carter 模型在中国的适用性和表现形式。基于1992~2007年中国人口分年龄组死亡率数据,文章对奇异值分解法(SVD)、最小二乘法(OLS)、加权最小二乘法(WLS)和极大似然法(MLE)的拟合结果和预测能力进行了比较分析。结果表明,加权最小二乘法具有最好的拟合和预测效果。文章利用最优的拟合模型,对未来中国人口平均预期寿命进行了预测,并利用Bootstrap方法进行了区间估计。

**【关键词】** Lee-Carter 模型 死亡率预测 平均预期寿命 Bootstrap 方法

**【作者】**李志生 中南财经政法大学金融学院,副教授;刘恒甲 中南财经政法大学金融学院,硕士研究生。

随着生活水平的提高和医疗卫生条件的改善,人类死亡率呈不断下降的趋势,人类寿命在不断延长。据联合国人口信息网(POPIN)的统计,1995~2005年中国、印度、日本、韩国、美国和巴西6个国家的出生人口平均预期寿命增长了20.5年,其中增幅最大是中国,增长了28.4年。人口死亡率的持续下降和寿命的不断延长,导致很多国家面临越来越高的老年依赖比<sup>①</sup>,给公共养老体系和养老金给付带来巨大的压力。

死亡率预测是人口科学的关键部分,也是人口学界和统计学界研究的重点。作为养老产品定价和养老金财务计划的基础,死亡率预测在养老金制度安排、产品供给和社会个体退休规划中都起着重要的作用。然而,大量研究结果表明,很多国家依据经验数据对人口死亡率下降和预期寿命增长趋势的预测普遍存在偏低的现象(Koissi等,2006;Stoto,1983)。这一方面使世界各国养老保障事业的发展滞后于人口老龄化进程;另一方面也给养老金的成本核算和养老产品供给者的经营带来巨大风险,严重影响各类养老计划的可持续发展。本文通过对不同死亡率模型的优劣和适用范围的比较,选择 Lee-Carter 模型对中国历史死

\* 本研究获国家自然科学基金(70801063)资助。

① 老年人口依赖比是指65岁以上的老年人口与15~64岁人口的比例。

亡率数据进行拟合,以检验 Lee-Carter 模型在中国的实用性和表现形式,通过比较不同模型估计方法的拟合效果和预测能力,得出最优的模型估计方法和模型参数,在此基础上对未来中国出生人口的平均预期寿命进行点估计和 Bootstrap 区间估计。

## 一、死亡率模型

死亡率数据最直接的来源是人口调查统计结果和保险业经验生命表。但前者反映的只是死亡率的历史状况,不能直接用来对死亡率进行预测,后者则是源于参保人的经验数据,并不具有广泛的代表性。在死亡率预测中,除了基于生物医学的方法和融合社会经济因素影响的因果模型外,最重要的一类是趋势外推模型,此类模型被广泛用于寿险精算、年金定价、长寿风险管理等方面。从模型的特征看,趋势外推模型可分为静态模型、动态模型和连续模型三类。

静态死亡率模型也称为确定型死亡率模型,主要包括:Gompertz(1825)提出的指数模型(Gompertz 模型)、Makeham(1860)在 Gompertz 模型的基础上提出的改进指数模型(Makeham 模型)、Helligman 等(1980)针对不同年龄人群提出的分年龄静态死亡率模型(HP 模型)。考虑到死亡率变动的不确定性及死亡率变动与年龄、时间的相关性,Lee 等(1992)提出了一个简洁的动态死亡率模型。在 Lee-Carter 模型的基础上,很多学者对其进行了改进,如 Renshaw 等(2006)的包含出生年效应的 Lee-Carter 模型和 Cairns 等(2006)的两因素年龄队列效应模型。第三类模型借鉴了金融经济领域广泛应用的连续模型,称为连续随机死亡率模型。Dahl(2004)对死亡力和连续利率的相似性进行了分析,提出根据死亡力的期限结构,将随机利率模型演化为死亡率模型;Biffis(2005)研究了死亡风险和信用风险的相似性,并提出了带有人口统计参数特征的死亡力仿射模型。

上述三类模型中,静态模型由于没有考虑未来死亡率变化的不确定性,不能描绘死亡率随时间变化的动态特征,一般只用于对历史数据的拟合,而很少用于预测。连续随机死亡率模型虽然在数据拟合方法上有所创新,但由于其发展较晚,模型的长期动态变化在生物学上的合理性及模型预测结果的稳健性都有待进一步研究。在动态死亡率模型中,虽然 Lee-Carter 模型最初的提出是针对美国的人口死亡率,但大量研究表明, Lee-Carter 模型对很多国家的数据都有很好的拟合(Lee 等,1994; Wilmoth, 1996; Tuljapurkar 等, 2000)。王晓军、蔡正高(2008)通过比较分析后提出中国人口死亡率数据适合用 Lee-Carter 模型进行模拟。虽然很多学者对 Lee-Carter 模型进行了改进,但是改进后的模型在使用上的复杂性并没有使其得到更为广泛的应用,加之本研究并不需要涉及人口学或生物学的更深层次,因此本文选择被学者们广泛应用的 Lee-Carter 模型对中国的死亡率数据进行拟合和预测。

## 二、Lee-Carter 模型及其拟合方法

Lee-Carter 模型的主要思路是将死亡率的变化分解为时间因子  $t$  和年龄因子  $x$ 。如果用

$m_{x,t}$  表示时间  $t$  年龄为  $x$  岁的人群的中心死亡率,那么  $m_{x,t}$  满足以下函数关系:

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

其中,  $\alpha_x$  为依赖于年龄因子的参数,反映分年龄人口死亡率自然对数的平均水平;  $k_t$  为依赖于时间因子的参数,通常称为死亡指数,反映人口死亡随时间变化的速度;  $\beta_x$  表示年龄因子对  $k_t$  的敏感度;  $\varepsilon_{x,t}$  是均值为 0、方差为  $\sigma^2$  的误差项。

为了在模型的参数估计中得到唯一解, Lee-Carter 模型同时进行以下假定:

$$\sum_t k_t = 0, \sum_x \beta_x = 1 \quad (2)$$

基于上述假设,  $\alpha_x$  的估计值  $\hat{\alpha}_x$  可以表示为  $\ln(m_{x,t})$  在时间维度上的平均值:

$$\hat{\alpha}_x = \sum_t \ln(m_{x,t}) / T = \ln[\prod_t m_{x,t}^{1/T}] \quad (3)$$

利用 Lee-Carter 模型进行死亡率预测通常需三步来实现: (1) 使用历史死亡率数据估计参数  $\alpha_x$ 、 $\beta_x$  和  $k_t$ ; (2) 通过移动平均自回归方法预测  $k_t$  的取值; (3) 将  $k_t$  的预测值和  $\alpha_x$ 、 $\beta_x$  参数代入模型, 以获得死亡率的预测值。Lee-Carter 模型参数的估计方法有两大类, 一是非似然估计方法, 主要包括矩阵奇异值分解法 (SVD)、最小二乘法 (OLS) 和加权最小二乘法 (WLS); 二是似然估计方法, 应用最多的是极大似然估计方法 (MLE)。

### (一) 矩阵奇异值分解法

为了求解模型, Lee 和 Carter 假设误差项对所有  $x$  和  $t$  都具有相同的方差, 提出了一个两阶段估计方法。在第一阶段, 通过对矩阵  $\ln(m_{x,t}) - \hat{\alpha}_x$  进行奇异值分解得到:

$$\text{SVD}[\ln(m_{x,t}) - \hat{\alpha}_x] = \sum_{i=1}^r \rho_i U_{x,i} V_{t,i} \quad (4)$$

其中,  $r$  为矩阵  $\ln(m_{x,t}) - \hat{\alpha}_x$  的秩,  $\rho_i$  为矩阵  $[\ln(m_{x,t}) - \hat{\alpha}_x]$  从大到小排列的奇异值,  $U_{x,i}$  和  $V_{t,i}$  是对应的两个奇异向量。通过对实际数据的拟合, Lee 和 Carter 发现奇异值  $\rho_1$  通常远远大于其他奇异值, 使用奇异值分解结果的第一项进行参数估计就能得到很好的结果<sup>①</sup>。如果选择这种近似方法来估计  $\beta_x$  和  $k_t$ , 那么  $[\ln(m_{x,t}) - \hat{\alpha}_x] \approx \rho_1 U_{x,1} V_{t,1}$ , 进一步分析并结合式 (2) 可得到  $\hat{\beta}_x = U_{x,1} / \sum_x U_{x,1}$  和  $\hat{k}_t = \rho_1 V_{t,1} / \sum_x U_{x,1}$ 。

矩阵奇异值分解使用纯数学近似的方法对  $\beta_x$  和  $k_t$  同时进行估计, 加之所使用的数据是死亡率的对数, 而非直接的死亡率, 这会使死亡率的拟合值与实际值产生较大偏差。为此, 在第二阶段, 还需对结果进行修正。常见的做法是对  $\hat{k}_t$  进行再估计, 再估计的原则是使实际的分年龄死亡人数  $d_{x,t}$  之和等于拟合的分年龄死亡人数之和, 即  $\sum_x d_{x,t} = \sum_x e_{x,t} e^{\hat{\alpha}_x + \hat{\beta}_x \hat{k}_t}$ , 其中  $e_{x,t}$  为  $t$  时年龄为  $x$  的人口总数, 即暴露人口数。

① 拟合方差能够解释 90% 以上离差平方和, 这一结论后来也相继被很多研究所证实 (Lee 等, 1994; Wilmoth, 1996; Tuljapourkar 等, 2000), 本文后面采用奇异值分解结果的第一项得到的拟合方程对死亡率变异的解释率也达到了 97.4%。

## (二) 最小二乘法

最小二乘法是另一种求解 Lee-Carter 模型参数的简洁方法。根据式(2)和式(3),通过以下两个步骤即可求得 $\beta_x$ 和 $k_t$ 的估计值:第一步,将式(1)两边对年龄 $x$ 求和,得到 $\hat{k}_t = \sum_x [\ln(m_{x,t}) - \hat{\alpha}_x]$ ;第二步,利用 $\hat{k}_t$ 对 $[\ln(m_{x,t}) - \hat{\alpha}_x]$ 进行线性回归,建立回归方程 $\ln(m_{x,t}) - \hat{\alpha}_x = \beta_x \hat{k}_t + \varepsilon_{x,t}$ ,然后基于最小二乘法最小化 $\sum_{x,t} [\ln(m_{x,t}) - \hat{\alpha}_x - \beta_x \hat{k}_t]^2$ ,可以得到 $\hat{\beta}_x = \sum_{t=1}^T \hat{k}_t (\ln m_{x,t} - \hat{\alpha}_x) / \sum_{t=1}^T \hat{k}_t^2$ 。

虽然最小二乘法求解模型时使用的也是死亡率的对数,但由于 $k_t$ 和 $\beta_x$ 是通过两步先后进行估计,第二步利用最小二乘法对 $\beta_x$ 的估计能有效地调整第一步产生的拟合偏差,因此不需对 $\beta_x$ 或 $k_t$ 进行再调整。

## (三) 加权最小二乘法

奇异值分解法和最小二乘法对不同年龄人群的死亡率赋予了相同的权重,但在现实情况下,不同年龄人群对应的人口数和死亡人口数都存在较大的差异,因此这两种方法在死亡率很低的条件下使用效果较差(Koissi 等, 2006)。为此, Wilmoth (1996)提出采用加权最小二乘法来估计 Lee-Carter 模型。

加权最小二乘法对 $k_t$ 的估计和最小二乘法相同。Wilmoth 证明 $\ln(m_{x,t})$ 的方差近似等于死亡人数 $d_{x,t}$ 的倒数,因此可以将 $d_{x,t}$ 作为残差平方和的权重。最小化经加权处理后的残差平方和 $\sum_{x,t} d_{x,t} [\ln(m_{x,t}) - \hat{\alpha}_x - \beta_x \hat{k}_t]^2$ ,得到 $\beta_x = \sum_{t=1}^T d_{x,t} k_t (\ln m_{x,t} - \alpha_x) / \sum_{t=1}^T d_{x,t} k_t^2$ 。与最小二乘法一样,加权最小二乘法也不必对参数的估计结果进行再调整。

## (四) 极大似然估计

Brouhns 等(2002)去掉了估计 Lee-Carter 模型时所作的 $\varepsilon_{x,t}$ 同方差假设,提出了似然估计的方法。他们假设死亡人数服从泊松分布,即 $d_{x,t} \sim \text{Poisson}(e_{x,t} e^{\hat{\alpha}_x + \hat{\beta}_x \hat{k}_t})$ 。通过推导,最小化以下似然函数,即可实现模型的估计:

$$l(\alpha, \beta, k) = \sum_{x,t} [d_{x,t} (\alpha_x + \beta_x k_t) - e_{x,t} e^{\alpha_x + \beta_x k_t}] + \text{constant} \quad (5)$$

这一方法也称为泊松对数双线性模型法。针对双线性特征项 $\beta_x k_t$ 和对数项 $\ln(m_{x,t})$ 给模型估计带来的困难,本文利用 Goodman (1979)的参数迭代方法来实现三类参数的估计。从参数初始值, $\hat{\alpha}_x^{(0)}=0$ 、 $\hat{\beta}_x^{(0)}=1$ 和 $\hat{k}_t^{(0)}=0$ <sup>①</sup>开始,按以下步骤更新参数:

$$\begin{aligned} \hat{\alpha}_x^{(v+1)} &= \hat{\alpha}_x^{(v)} + \frac{\sum_t (d_{x,t} - \hat{d}_{x,t}^{(v)})}{\sum_t \hat{d}_{x,t}^{(v)}}, \quad \hat{\beta}_x^{(v+1)} = \hat{\beta}_x^{(v)}, \quad \hat{k}_t^{(v+1)} = \hat{k}_t^{(v)}; \\ \hat{k}_t^{(v+2)} &= \hat{k}_t^{(v+1)} + \frac{\sum_x (d_{x,t} - \hat{d}_{x,t}^{(v+1)}) \hat{\beta}_x^{(v+1)}}{\sum_x \hat{d}_{x,t}^{(v+1)} (\hat{\beta}_x^{(v+1)})^2}, \quad \hat{\alpha}_x^{(v+2)} = \hat{\alpha}_x^{(v+1)}, \quad \hat{\beta}_x^{(v+2)} = \hat{\beta}_x^{(v+1)}; \end{aligned}$$

① 也可以采用随机值作为参数初始值。

$$\hat{\beta}_x^{(v+3)} = \hat{\beta}_x^{(v+2)} + \frac{\sum_l (d_{x,l} - \hat{d}_{x,l}^{(v+2)}) \hat{k}_l^{(v+2)}}{\sum_l \hat{d}_{x,l}^{(v+2)} (\hat{k}_l^{(v+2)})^2}, \hat{\alpha}_x^{(v+3)} = \hat{\alpha}_x^{(v+2)}, \hat{k}_l^{(v+3)} = \hat{k}_l^{(v+2)}.$$

其中,第  $v$  次更新需满足条件  $\hat{d}_{x,l}^{(v)} = e_{x,l} e^{\hat{\alpha}_x^{(v)} + \hat{\beta}_x^{(v)} \hat{k}_l^{(v)}}$ 。当参数更新使式(5)的增加量很小(比如小于  $10^{-6}$ )时停止迭代,得到参数估计值。

三、数据与拟合结果

(一) 原始数据的选择及处理

本文采用的原始数据来源于 1993~2008 年的《中国人口统计年鉴》中的分年龄组死亡率数据、死亡人口和暴露人口数据,包括 0~4 岁、5~9 岁……90 岁以上共 19 个年龄分组。其中,1996 年的《中国人口统计年鉴》并未提供相关分组死亡率指标,所以我们得到的数据区间为 1992~1994 年、1996~2007 年,共 15 年。

为获得出生人口的平均预期寿命,本文基于统计年鉴中 0~4 岁组和 0 岁人口的死亡人口数据和暴露人口数据,经过计算将 0~4 岁组的死亡率拆分为 0 岁组与 1~4 岁组。另外,超高龄人口死亡率的变化不太稳定,对整个人口结构的影响很大。但是,《中国人口统计年鉴》将 90 岁以上人口合并为一组进行统计,并未像 0~89 岁那样给出更为详细的分组数据。为了能够在估计模型的过程中得到反映 90 岁及以上人口死亡率变化特征的参数值,本文采用 Coale 等(1990)提出的死亡率分组扩展方法,将 90 岁及以上分组扩展为 90~94 岁、95~99 岁和 100 岁及以上 3 个分组。最终得到并采用 15 年 22 个年龄分组的死亡率数据。

(二) 拟合结果

4 种方法对代表死亡率自然对数平均水平的参数的估计结果如表 1 所示,其中 SVD、OLS 和 WLS 求解  $\alpha_x$  的方法相同。0 岁组和 65 岁及以上年龄分组的死亡率自然对数值相对较高,而 1~4 岁组至 60~64 岁组相对较低。

表 1 4 种方法下  $\alpha_x$  的估计值

年龄组(岁)	SVD、OLS 和 WLS	MLE	年龄组(岁)	SVD、OLS 和 WLS	MLE
0	-3.7131	-3.7406	50~54	-5.2776	-5.2712
1~4	-6.6114	-6.6265	55~59	-4.7849	-4.7811
5~9	-7.5307	-7.5038	60~64	-4.2461	-4.2431
10~14	-7.7218	-7.7016	65~69	-3.7278	-3.7463
15~19	-7.2381	-7.2115	70~74	-3.1519	-3.2007
20~24	-6.8418	-6.8306	75~79	-2.7930	-2.7856
25~29	-6.7567	-6.7434	80~84	-2.3051	-2.2973
30~34	-6.5974	-6.5924	85~89	-1.9326	-1.9188
35~39	-6.3785	-6.3703	90~94	-1.2379	-1.2055
40~44	-6.0515	-6.0493	95~99	-0.4748	-0.4760
45~49	-5.6600	-5.6567	100+	0.0000	0.0000



图 1 给出了参数  $\beta_x$  和  $k_t$  的估计结果。从中可以看出,4 种方法得到的死亡率指数  $k_t$  的估计值均呈近似线性下降的趋势,表明死亡率随时间推移而减小的速度较为稳定,与历史死亡率总体趋于下降的特征一致。4 种方法得到的  $\beta_x$  估计

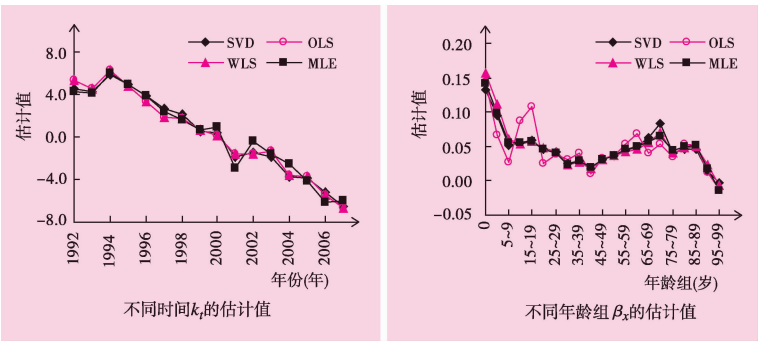


图 1 参数估计结果

计值在不同年龄组间具有相似的走势。从图 1  $\beta_x$  的估计值看,低龄组人口的  $\beta_x$  值很高,特别是 0 岁组和 1~4 岁组。这主要是由于新生人口具有较高的死亡率,对死亡率趋势的变化也最为敏感。85~89 岁及以上高龄组的  $\beta_x$  值呈下降趋势,并趋近于 0,其原因在于高龄人口的死亡率特征随时间的变化较小,实际死亡率对死亡率指数不敏感。

从模型的拟合程度来看,4 种方法都表现出良好的效果,SVD、OLS、WLS 和 MLE 得到的拟合方程对死亡率离差平方和的解释程度分别为 97.39%、97.51%、97.48% 和 98.83%。为了进一步比较不同方法的拟合优劣,我们对残差的统计特征进行分析。残差是否独立和是否具有相同的概率分布是检验模型拟合好坏的重要指标。我们通过残差分布的三维图可以有效地检验其独立性和同分布性。如果残差序列表现出随机正态分布的特征,残差图中将不会出现明显的条状区域。

图 2 中,如果残差在同一年份的不同年龄组之间趋于一致,表明同方差性假设不成立;而残差值在相同年龄组的不同年份上趋于一致,则表明违反了残差独立性假设。我们发现,SVD 和 OLS 方法的拟合残差在低龄人口分组中表现出明显的异方差特征,WLS 和 MLE 方法的拟合残差则呈现出更好的随机分

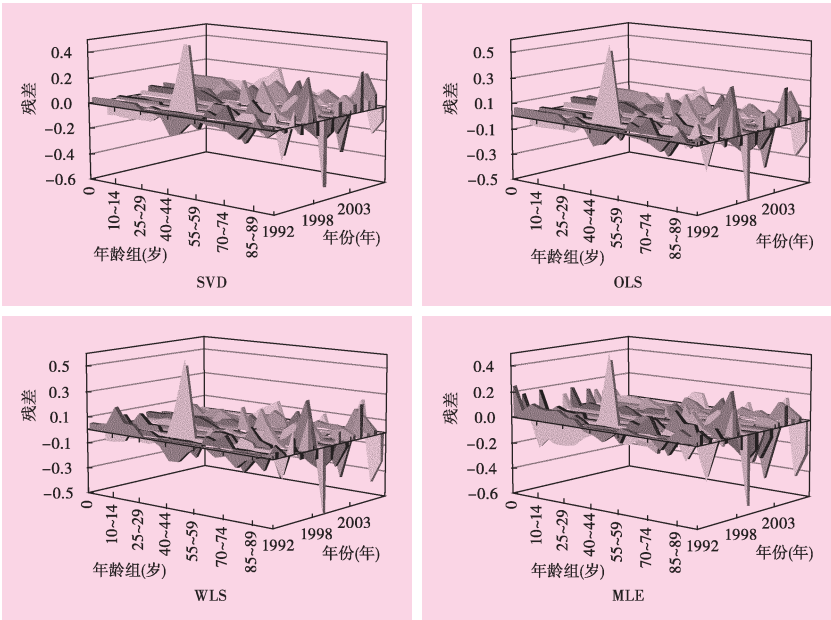


图 2 模型拟合残差三维图

布特征。这主要是因为不同年龄人群的实际人口数和死亡人口数存在较大差异,SVD 和 OLS 方法对不同年龄人口的死亡率赋予相同的权重,并不能满足误差项同方差的假设。此外,WLS 和 MLE 方法拟合残差的绝对值也相对较小,并且基本对称分布于 0 附近,说明对所有年龄分组死亡率的拟合效果趋于一致。综上所述,我们认为 WLS、MLE 方法克服了SVD、OLS 方法不适用于低龄人口死亡率估计和同方差假设的不足,具有更好的拟合效果。

四、模型的预测效果

下面检验使用不同估计方法得到的拟合模型的预测效果。我们以 1992~2002 年的死亡率数据对 2003~2007 年的死亡率进行预测,并将预测结果与 2003~2007 年的实际死亡率进行比较。我们利用 ARIMA 模型实现时变参数的预测,建立以下模型:

$$(1-\sum_{i=1}^p \phi_i L^i)(1-L)^d k_t = \lambda + (1 + \sum_{i=1}^q \theta_i L^i) \varepsilon_{t-1} \tag{6}$$

其中, $L$  为滞后算子, $p$  为自回归项数, $q$  为移动平均项数, $d$  为使  $k_t$  成为平稳序列所做的差分次数, $\lambda$  为常数项, $\varepsilon$  为误差项。

根据 Box-Jenkins 方法,利用前文中得到的  $k_t$  值对式(6)进行拟合,发现最优的模型选择为 ARIMA(0,1,0),即  $k_t = k_{t-1} + \lambda + \varepsilon_t$ 。表 2 的数据显示,ARIMA(0,1,0)模型对 4 种估计方法得到的指数均有较好的拟合,由于 OLS 和 WLS 求解  $k_t$  的方法相同,因此得到相同的模型参数。

表 2 ARIMA(0,1,0)模型的估计参数

估计方法	模型参数			
	$\lambda$	R-squared	$t_\lambda$	Prob.
SVD	-0.7972	0.9192	-2.6519	0.0021
OLS	-0.8535	0.9030	-2.6113	0.0227
WLS	-0.8535	0.9030	-2.6113	0.0227
MLE	-0.7188	0.8535	-1.5893	0.1380

将  $k_{t+1} \cdots k_{t+n}$  代入 Lee-Carter 模型即可得到时间  $t+1$  到  $t+n$  的死亡率预测值。根据 1992~2002 年的死亡率数据,分别对 0 岁、20~24 岁、60~64 岁、80~84 岁 4 个年龄组在 2003~2007 年的死亡率进行预测,表 3 列出了 4 种方法在不同年份针对不同年龄组的预测

偏差的平均水平。首先,平均预测偏差在不同年龄组上存在差异,20~24 岁和 60~64 岁两组的预测偏差较小,而 0 岁组和 80~84 岁两组的预测偏差较大。产生上述结果的主要原因是在所有年龄分组中,婴幼儿及高龄人口死亡率的变化本身具有较大的不稳定性,这给死亡率预测带来一定难度。其次,平均预测偏差在不同年份也存在差异,随着时间的推移,预测偏差有逐步增大的趋势。可见,从总体上看,短期预测效果好于长期预测效果,这主要是由于预测偏差随着时间的推移存在累积的现象。

图 3 绘出了不同方法在不同时间、针对不同年龄组的预测偏差。显然,除了低估 0 岁组的死亡率外,MLE 方法均高估了其他 3 个年龄组的死亡率,预测值与实际值存在较大偏差,其中对 0 岁组的平均预测偏差达 -0.0186,对其余 3 组的平均预测偏差为 0.0952。SVD 和 OLS 两种方法在大多数情况下也存在一定的预测偏差,平均偏差分别为 0.0033 和 0.0045。

表 3 4 种方法预测偏差的平均值

年龄组(岁)	2003 年	2004 年	2005 年	2006 年	2007 年	年平均
0	-0.0077	-0.0298	-0.0022	-0.0019	0.0031	-0.0077
20~24	0.0000	0.0003	0.0001	0.0001	0.0000	0.0001
60~64	0.0032	0.0021	0.0033	0.0049	0.0056	0.0038
80~84	-0.0062	0.0016	0.0048	0.0194	0.0163	0.0072
平均值	-0.0027	-0.0064	0.0015	0.0056	0.0062	0.0009

比较而言,WLS 方法的预测表现最好,从平均水平看存在一定的低估(平均预测偏差为 -0.0008),但在不同时间、针对不同年龄组的预测偏差在方向和大小上都存在差异。

五、未来人口死亡率和人口平均预期寿命的估计

根据上述分析,无论是从模型对历史数据的拟合效果,还是从模型基于历史数据的预测效果来看,WLS 方法估计得到的死亡率模型均有较好的表现。下面以 WLS 方法得到的死亡率模型为基础,对不同年份、不同年龄组人口未来死亡率和平均预期寿命进行估计,并利用 Bootstrap 方法对平均预期寿命进行区间估计。

(一) 未来人口死亡率

基于中国 1992~2007 年的死亡率数据,本文利用 WLS 方法求解 Lee-Carter 模型,并对 2008~2017 年分年龄组人口死亡率进行预测,结果如表 4 所示。表 4 数据显示,90~94 岁和 95~99 岁两个高龄组的死亡率随时间变化而下降的趋势不明显,这主要是由于其本身具有较高的死亡率,与前文得出的高龄人口  $\beta_x$  值较低的结论是一致的。

图 4 绘出了 2008、2011、2014 和 2017 年人口死亡年龄的分布情况。可以看到,死亡人口的年龄分布表现出一定的左偏特征,而且随着时间的推移,左偏的特征越来越明显。这表示越来越多的人将在更高的年龄才会死亡,人口平均预期寿命呈现不断增加的趋势。

(二) 未来人口平均预期寿命

估计分年龄人口预期寿命的最直接的方法是通过目标时间、目标人群中具有相同年

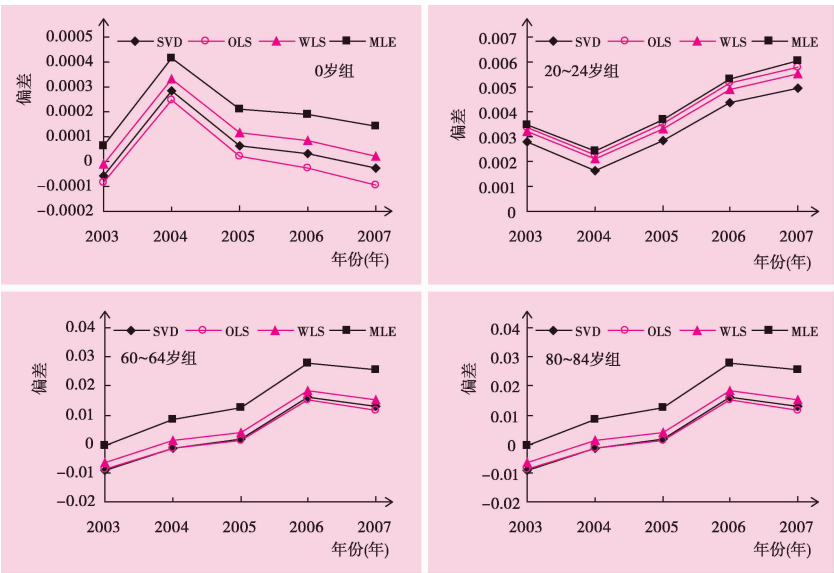


图 3 2003~2007 年死亡率预测值与实际值之差



表 4 未来人口死亡率的预测结果 ‰

年龄组(岁)	2008 年	2009 年	2010 年	2011 年	2012 年	2013 年	2014 年	2015 年	2016 年	2017 年
0	8.85	7.88	7.02	6.25	5.57	4.96	4.41	3.93	3.50	3.12
1~4	0.65	0.60	0.55	0.51	0.47	0.43	0.40	0.37	0.34	0.31
5~9	0.34	0.32	0.31	0.29	0.28	0.26	0.25	0.24	0.23	0.21
10~14	0.30	0.28	0.27	0.26	0.25	0.24	0.23	0.22	0.21	0.20
15~19	0.47	0.44	0.42	0.40	0.38	0.36	0.35	0.33	0.31	0.30
20~24	0.75	0.72	0.69	0.67	0.64	0.61	0.59	0.57	0.54	0.52
25~29	0.85	0.82	0.79	0.76	0.74	0.71	0.69	0.66	0.64	0.62
30~34	1.15	1.13	1.11	1.09	1.07	1.05	1.03	1.01	0.99	0.97
35~39	1.39	1.36	1.33	1.30	1.27	1.25	1.22	1.19	1.16	1.14
40~44	2.06	2.03	2.00	1.96	1.93	1.91	1.88	1.85	1.82	1.79
45~49	2.75	2.68	2.61	2.54	2.47	2.41	2.34	2.28	2.22	2.16
50~54	3.88	3.76	3.65	3.54	3.43	3.32	3.22	3.12	3.02	2.93
55~59	6.07	5.85	5.64	5.44	5.25	5.06	4.88	4.70	4.53	4.37
60~64	10.13	9.74	9.36	8.99	8.65	8.31	7.99	7.68	7.38	7.09
65~69	15.87	15.14	14.43	13.76	13.13	12.52	11.94	11.38	10.85	10.35
70~74	25.42	23.95	22.57	21.27	20.04	18.88	17.79	16.76	15.80	14.88
75~79	45.30	43.76	42.28	40.85	39.47	38.13	36.84	35.59	34.39	33.22
80~84	69.98	67.20	64.53	61.97	59.51	57.15	54.88	52.70	50.61	48.60
85~89	100.46	96.35	92.41	88.63	85.01	81.53	78.20	75.00	71.93	68.99
90~94	242.54	237.64	232.83	228.13	223.51	219.00	214.57	210.23	205.98	201.82
95~99	652.13	655.67	659.22	662.80	666.40	670.01	673.64	677.30	680.97	684.67

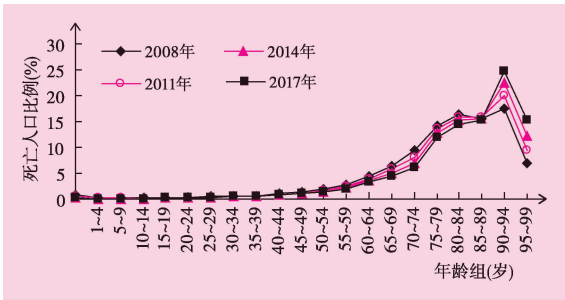


图 4 死亡人口年龄分布

龄的一个随机抽样进行跟踪调查。但是，跟踪调查一批人的整个生命过程在现实中是难以实现的，而且这种方法得到的结果也是多年以前的历史寿命。因此，实际计算中通常采用生命表方法，即以目标时间各年龄组人口的死亡率水平来代替分年龄人口在不同年龄的死亡率水平，并计算各年龄人口的平均生存人数，由此推算

出分年龄人口在目标时间的预期寿命。基于 WLS 方法得到的死亡率模型，对 2008~2017 年 0 岁、65 岁和 85 岁人口的平均预期寿命进行估计，结果如表 5 所示。

接下来用 Bootstrap 方法对 2008~2017 年中国出生人口的平均预期寿命进行区间估计。Bootstrap 方法有两种实现方式，一是根据原有数据的具体分布进行再抽样来得到 Bootstrap 数据，也叫参数 Bootstrap 方法；二是对模型拟合的残差进行再抽样得到 Bootstrap 数据，通常称为残差 Bootstrap。残差 Bootstrap 不需要知道原有数据的具体分布特征，是一种非参数方法。

表 5 部分年龄人口平均预期寿命的预测结果

年龄(岁)	2008 年	2009 年	2010 年	2011 年	2012 年	2013 年	2014 年	2015 年	2016 年	2017 年
0	73.82	74.23	74.63	75.01	75.38	75.74	76.09	76.42	76.75	77.07
65	14.20	14.43	14.65	14.87	15.08	15.29	15.49	15.69	15.89	16.08
85	1.96	1.98	2.00	2.02	2.04	2.06	2.08	2.09	2.11	2.12

本文采用第二种方法，对模型的拟合残差进行 Bootstrap 并通过残差方程<sup>①</sup>求解出死亡率的估计值，进而获得未来死亡率的区间估计。图 5 为 2 000 次残差 Bootstrap 的预测结果。

基于生命表的人口平均预期寿命求解方法，假设人口结构和死亡率在未来不发生改变，即未来不同年龄组人口的死亡率和现在是一致的。但是，现实中不同年龄组人口的死亡率是随时间的推移而改变的，一般表现为下降趋势。因此基于生命表的平均预期寿命估计方法存在低估未来人口寿命的风险，这是长寿风险的主要来源之一。Lee-Carter 模型通过分析人口死亡率变化的动态特征，首先对未来人口死亡率进行合理的预测，然后基于死亡率的预测值对人口平均预期寿命进行估计，这在一定程度上克服了由于假定未来人口死亡率不变而带来的平均预期寿命低估的缺陷。

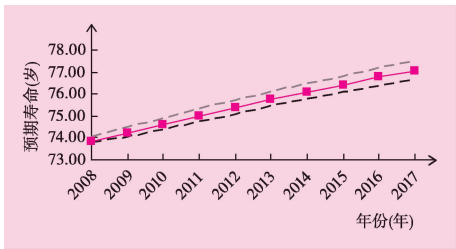


图 5 2008~2017 年出生人口平均预期寿命在 95%置信区间下的区间估计

## 六、结 论

本文利用中国 1992~2007 年的历史死亡率数据对 Lee-Carter 死亡率模型进行了量化分析,结果表明, Lee-Carter 模型对中国死亡率数据具有很好的拟合,但不同参数估计方法得到的结果存在差异,从拟合模型的解释率和拟合残差的统计特征看,加权最小二乘法得到的估计模型具有最好的拟合效果。为了比较不同拟合模型的预测效果,本文利用 1992~2002 年的死亡率数据对 2003~2007 年的死亡率进行了预测,并将预测结果和实际死亡率数据进行比较,结果显示,加权最小二乘法得到的死亡率模型具有最小的预测偏差。利用加权最小二乘法得到的最优拟合模型,本文对未来中国分年龄组人口死亡率和预期寿命进行了预测,并通过 Bootstrap 方法给出了相应的区间估计。

人口死亡率特征及其动态发展趋势不仅关系到政府政策的制定和社会资源的再分配,也涉及社会个体的养老、医疗和保险等诸多方面。人口死亡率的持续下降和平均预期寿命的不断延长是近半个世纪以来世界各国人口发展的基本态势。寿命的延长一方面提高了退休生活的成本;另一方面也加大了退休计划中的风险因素——长寿风险。目前国内外学者

① 残差方程为  $\xi_{it} = \text{sign}(d_{x,t} - \hat{d}_{x,t}) \times [d_{x,t} \ln(d_{x,t} / \hat{d}_{x,t}) - (d_{x,t} - \hat{d}_{x,t})]^{1/2}$ ; 死亡率估计值  $\hat{m}_{x,t} = \hat{d}_{x,t} / e_{x,t}$ 。

关于长寿风险的讨论还局限在定性的层面上,在定量计算方面还没有形成系统的研究成果。本文关于 Lee-Carter 死亡率模型在中国的适用性和应用研究是长寿风险管理等相关问题的基础,将为长寿风险的识别和量化、年金产品的定价、养老产品的设计、个体退休计划的制订、养老保障体系的完善等提供重要参考。结合本文得到的最优死亡率模型,利用不同的下方风险计量方法构建长寿风险评估体系和定量计算矩阵,是未来研究的重点与方向。

#### 参考文献:

1. 王晓军、蔡正高(2008):《死亡率预测模型的新进展》,《统计研究》,第9期。
2. A.E.Renshaw, S. Haberman(2006), A Cohort Based Extension to the Lee-Carter Model for Mortality Reduction Factors. *Insurance: Mathematics and Economics*. Vol. 38, pp. 556-570.
3. A.J.Coale, E.E. Kisker (1990), Defects in Data on Old-age Mortality in the United States; New Procedures for Calculating Schedules and Life Tables at the Higher Ages. *Asian and Pacific Population Forum*. Vol.4, pp.1-31.
4. B.Gompertz(1825), On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies. *Philosophical Transactions of the Royal Society of London*. Vol.115, pp. 513-583.
5. E.Biffis(2005), Affine Processes for Dynamic Mortality and Actuarial Valuations. *Insurance: Mathematics and Economics*. Vol.37, pp. 443-468.
6. J.R.Wilmoth(1996), Mortality Projections for Japan: A Comparison of Four Methods. *Health and Mortality Among Elderly Population*. Oxford University Press. New York.
7. L.A.Goodman (1979), Simple Models for the Analysis of Association in Cross-classifications Having Ordered Categories. *Journal of the American Statistical Association*. Vol. 74, pp. 537-552.
8. L.Helligman, J.Pollard(1980), The Age Pattern of Mortality. *Journal of the Institute of Actuaries*. Vol.107, pp. 49-75.
9. M.A.Stoto(1983), The Accuracy of Population Projections. *Journal of the American Statistical Association*. Vol. 78, pp. 13-20.
10. M.C.Koissi, A.F.Shapiro, G.Hognas (2006), Evaluating and Extending the Lee-Carter Model for Mortality Forecasting: Bootstrap Confidence Interval. *Insurance: Mathematics and Economics*. Vol. 38, pp.1-20.
11. M.Dahl(2004), Stochastic Mortality in Life Insurance: Market Reserves and Mortality-linked Insurance Contracts. *Insurance: Mathematics and Economics*. Vol. 35, pp. 113-136.
12. N.Brouhns, M.Denuit, J.K.Vermunt(2002), A Poisson Log-bilinear Regression Approach to the Construction of Projected Lifetables. *Insurance: Mathematics and Economics*. Vol.31, pp. 373-393.
13. R.D.Lee, L.R.Carter(1992), Modeling and Forecasting U.S. Mortality. *Journal of the American Statistical Association*. Vol. 419, pp. 659-675.
14. R.D.Lee, R.Rofman(1994), Modeling and Projecting Mortality in Chile. *Notas de Población*. Vol.22, pp.182-213.
15. S.Tuljapurkar, L.Nan, C.Boe (2000), A Universal Pattern of Mortality Decline in the G7 Countries. *Nature*. Vol. 405, pp.789-792.
16. W.M.Makeham(1859), On the Law of Mortality, and the Construction of Annuity Tables. *Journal of the Institute of Actuaries*. Vol. 8, pp.301-310.

(责任编辑:朱 萍)