

# 中国“大跃进”时期的 非正常死亡人口研究

## ——基于改进的 Lee-Carter 分年龄死亡率预测模型\*

米红 贾宁

**【内容摘要】**以往文献对中国 1958~1961 年“大跃进”时期非正常死亡人口的估计缺乏深入、微观的研究。文章基于改进的 Lee-Carter 死亡率预测模型,以官方统计的分年龄别死亡率数据为基础进行回推,重构“大跃进”时期我国历年单岁组死亡率,并逆向预测估算在假定没有饥荒的情况下,“大跃进”时期正常状态死亡人口数量。继而结合国内外学者对当时总死亡人口的估计,对因饥荒而导致的“大跃进”时期的非正常死亡人口进行比较客观准确的评估性研究。结果表明,“大跃进”时期人口非正常死亡现象主要发生在 1958~1961 年间,如果不发生饥荒,则 4 年正常死亡人口介于 4339~4385 万人。由于饥荒的发生,非正常死亡人口介于 1624~2337 万人。评估分析表明,“大跃进”时期非正常死亡人口估算中值不会高于 1980 万人。

**【关键词】**大跃进; 逆预测; Lee-Carter 模型; 死亡人口

**【作者简介】**米红,浙江大学公共管理学院教授、博导,浙江大学人口与发展研究所执行所长; 贾宁,浙江大学公共管理学院博士研究生。浙江 杭州:310058

## An Assessment of Abnormal Deaths during the Great Leap Forward Using a Modified Lee-Carter Model

Mi Hong Jia Ning

**Abstract:** There is a lack of official report by the Chinese government regarding the exact number of abnormal deaths during the Great Leap Forward. Using a modified Lee-Carter mortality prediction model, this study conducts reverse prediction based on China's official age-specific mortality data to reconstruct and calculate the single-year-age mortality rate and the normal deaths assuming that there was no great famine. Then we estimate the abnormal deaths during the Great Leap Forward occurring between 1958 and 1961. If there had been no famine, there would have been 43.39 to 43.85 million normal deaths. With famine, abnormal deaths are estimated to be between 16.24 and 23.37 million. Our results suggest that the medium estimation of the abnormal deaths during the Great Leap Forward stands at no more than 19.8 million.

**Keywords:** The Great Leap Forward, Reverse Prediction, Lee-Carter Model, Death

**Authors:** Mi Hong is Professor, School of Public Affairs, Zhejiang University, Executive Director of Institute of Population & Development Studies. E-mail: spsswork@163.com. Jia Ning is PhD Candidate, Institute of Population and Development Studies, School of Public Affairs, Zhejiang University. E-mail: 11222043@zju.edu.cn.

\* 本文得到国家自然科学基金重大课题(71490732)、国家自然科学基金(71303212)、国家社科重大项目(12& ZD099)和浙江大学劳动保障与社会政策研究中心(LEPP)的支持。

1958~1961年间,由于“大跃进”运动以及牺牲农业发展的工业政策所导致的全国性粮食短缺和饥荒,被官方称为三年自然灾害,后改称为三年困难时期(Dikötter 2010)。

中外学者曾基于1953、1964和1982年的人口普查数据,参照西方模型生命表,或者辅以其他(如卫生部门的统计数据)数据进行“大跃进”时期非正常死亡人口研究,但不同研究方法所得结论差异较大(曹树基 2005)。对于“大跃进”引起的人口变动的具体情况及非正常死亡人口数量是国内外历史、人口和统计学界所关注并引起争论的焦点。因此,开展“大跃进”时期人口变动情况及非正常死亡人数研究对于认清我国建国以来人口、社会 and 经济发展曲折历程无疑具有重要意义。

## 1 文献回顾与评述

世界范围内曾有不少学者对我国“大跃进”时期非正常死亡人口进行过估算(Chen 2014)。美国国务院东亚和太平洋事务局曾在*Background Notes: China, October 1997*中将1958~1961年死于饥荒的人数估计为“数以百万计”。《剑桥中国历史》的估算为1600~2700万人,仅1960年就有超过1000万人死于饥荒(麦克法夸尔、费正清等,1990)。根据美国普查局中国科科长J. Bannister修订的数据计算结果,非正常死亡人数为2987.1万人,少出生人数为3119.5万人,人口减少总数为6106.6万人(J. Bannister, 1985)。美国人口理事委员会(PAA)前主席、普林斯顿大学教授Ansley Coale利用三次人口普查资料及1982年1‰人口生育率抽样调查资料,估算出两次人口普查间隔的历年出生人口和历年死亡人口,再按照线性趋势求得各年线性死亡人数,经与估算的实际死亡人数相比,得出纯线性死亡人数。其修订的数据计算结果为:1959~1961年非正常死亡人数为2481万人,少出生人数为3068.3万人。此外,人口总损失数量为5549.3万人。1958~1963年非正常死亡人数为2680万人。法国国立人口研究所前所长G. Calot修订的数据计算结果为:1958~1962年非正常死亡人数2850.9万人,4年少出生3197.85万人,人口总损失6048.8万人(G. Calot, 1985)。在《1959~1961年中国的“大跃进”》一书中提到,Eder认为1960~1961年非正常死亡人数为2300万人,Mosle估计1960年非正常死亡人数居于1100~3000万人。Hill则估计1958~1962年非正常死亡人数为3000万人,同时有3300万婴儿未出生或延迟出生(Kane, 1988)。宾夕法尼亚大学Luo Sheng(1988)利用Brass-Logit体系与模型生命表方法,创建一套“误差自减少”迭代算法,并设置了3种不同的人口分布与出生模式,重构3类1953~1982年的分年龄生存概率表,并与Coale、Bannister等人的计算结果进行对比,继而利用3次人口普查所得人口分布,假设1953~1964年出生性别比为105,1964~1982年的出生性别比为106,重构各年生命表和人口分年龄性别分布,估算得出1959~1963年非正常死亡人数为2889万人(Luo, 1988)。

我国一些学者对于“大跃进”时期的非正常死亡人数也有研究。西安交通大学人口研究所前所长蒋正华等根据1982年人口普查死亡资料编制了1981年人口生命表,建立动态参数估计模型,利用1953年、1964年和1982年人口普查获得的全国分年龄、性别人口数据,求解方程组后得出“一普”到“三普”间我国历年人口出生率和死亡率,并据此估算历年死亡人数,最后根据预期寿命变化估算正常死亡人数,用历年死亡人数减去正常死亡人数,求得非正常死亡人数,结果显示:1958~1963年我国非正常死亡人数约为1700万人(蒋正华、李南a, 1986; 蒋正华、李南b, 1986)。上海交通大学历史系曹树基教授运用人口学和历史地理学方法,重建1959~1961年我国各地非正常死亡人数,结果显示:1958~1961年我国非正常死亡人数约为3245.8(曹树基, 2005)。原国家统计局局长李成瑞先生经过修正Coale的研究缺陷后,测算获得我国“大跃进”时期非正常死亡人数约为2200万人。

综上,尽管上述学者均对我国“大跃进”时期非正常死亡人数进行了研究,但其研究结果差异较大。基于此,本研究拟运用改进的Lee-Carter模型对大饥荒时期非正常死亡人口数量进行

深入研究。

## 2 研究设计

### 2.1 数据来源及相关问题处理

本研究数据源自 1994 ~ 2006 年《中国人口统计年鉴》,2007 ~ 2010 年《中国人口与就业统计年鉴》,2000 年数据源自中国统计局官网;总人口结构和其他相关数据源自 1953、1964 及 1982 年的人口普查及 1987、1995 和 2005 年人口抽样调查。该数据集有以下需要处理的问题(祝伟、陈秉正 2009):

#### (1) 数据的样本问题

有些年采用的是变动抽样,有些采用的是千分之一抽样,有些采用普查数据。

#### (2) 分年龄死亡率数据的缺失与年龄分组问题

如 1995 年的数据缺失;有些年份的死亡数据按每一岁进行年龄分组直至 100 岁,有些按每 5 岁分组,大部分年份数据末组为 90+,1996 年末组为 85+,2005 年的为 100+。

#### (3) 本文采用单岁组进行模型拟合

相比 5 岁组数据,单岁组数据中(尤其是抽样调查数据)会出现某些高龄组死亡率反而比低龄组死亡率低很多的情况。

因此,基于上述问题,本文进行如下处理:

(1) 本文假设人口变动抽样和千分之一抽样具有良好的随机抽样特征;

(2) 本文利用人口死亡函数对高龄死亡率数据进行补全;

(3) 对单岁组死亡率进行平滑处理;

首先采用中值滤波,确保死亡率曲线中无过多的噪点,并在此基础上利用人口死亡函数进行平滑(Eilers, Delwarde et al. 2007);

(4) 对于个别年份的缺失数据,本文通过插值法进行拟合。

### 2.2 研究方法

本研究对 Lee-Carter 模型进行了改进,使其在拟合我国死亡率数据时也有较好表现,也使其能对以往死亡率数据进行回推测算。

#### 2.2.1 经典 Lee-Carter 模型

Lee-Carter 属于外推模型,其核心是将死亡率对数化处理和运用 ARIMA 时间序列进行预测(Lee and Carter, 1992; Carter and Prskawetz, 2001)。该方法是基于分年龄死亡率,将死亡率取对数后表示为一个由 3 个参数组成的线性组合,然后运用随机时间序列方法来预测未来死亡率变化趋势的模型,其模型结构如下:

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

其中  $m_{x,t}$  表示  $x$  岁在  $t$  时间的粗死亡率,  $\alpha_x$  表示各年龄别死亡率按时间计算获得的基数,是年龄组死亡率取自自然对数后的平均值。  $\beta_x$  表示  $k_t$  对各年龄组死亡率的影响,反映各年龄组死亡率对数受  $k_t$  变化影响的大小。  $k_t$  表示  $t$  时间死亡率水平的变化,为死亡率水平指数,反映历年死亡率相对强度,测度死亡率整体水平。  $\varepsilon_{x,t}$  表示残差,其均值为 0、方差为  $\delta_{\varepsilon}^2$ 。

为了能得到唯一解,对该模型设置了两个标准化限制条件:

$$\begin{cases} \sum_{all\ x} \beta_x = 1 \\ \sum_{t=i}^j k_t = 0 \end{cases} \quad (2)$$

其中  $i, j$  分别表示死亡率观察的初始年和结束年。

## 2.2.2 经典 Lee-Carter 模型的不足

### (1) 收敛性问题

经验数据对未来死亡率所做预测大多低于实际结果, Lee-Carter 模型由于其自身队列效应而存在“天生”误差。下面以具体某年龄为例,对该问题进行诠释。

Lee-Carter 死亡率预测模型就某一岁而言,即在公式(1)中,当  $x$  固定时,  $m_{x,t}$  将随预测时间推移而逐渐减小,并收敛为零,这是由于:

将公式(1)变形获得

$$m_{x,t} = \exp(a_x + \beta_x k_t) \quad (3)$$

对于固定的某一岁而言,比如 0 岁组,即  $x=0$ ,可得:

$$m_{0,t} = \exp(a_0 + \beta_0 k_t) \quad (4)$$

死亡率  $m_{0,t}$  是关于  $k_t$  随时间  $t$  变化的函数,而  $a_0$  和  $\beta_0$  相对时间  $t$  是固定的,在做正向预测时,由于拟合出的与死亡率变化成正相关的参数  $k_t$  的变化是逐渐减小的。随着社会进步,死亡率会逐步下降,但随着  $k_t$  的单调递减,因为  $\lim_{x \rightarrow -\infty} \exp(x) \rightarrow 0$ ,  $m_{0,t}$  的值会很快趋向于 0,即“过快速减”。

基于 Lee-Carter 模型该特点,其只适用于拥有较充分历史数据的国家(如美国,日本等)对未来较少年份的预测,但不适用于基于少量数据预测多年数据。目前我国能用于拟合 Lee-Carter 模型的死亡率数据只有不到 20 年,所以,模型的收敛性问题不容忽视。

### (2) 逆预测问题

一般认为 Lee-Carter 死亡率预测模型能基于现有统计数据预测未来分年龄死亡率,并且预测值呈现逐渐降低趋势,然而实践证明 Lee-Carter 死亡率预测模型不能用于对历史死亡率的逆预测,原因在于,在利用现在数据倒推历史数据时,  $k_t$  将呈上升趋势,这与正向预测时  $\lim_{t \rightarrow -\infty} m_{x,t} = \lim_{x \rightarrow -\infty} \exp(a_x + \beta_x k_t) \rightarrow 0$  不同;而变成  $\lim_{t \rightarrow +\infty} m_{x,t} = \lim_{x \rightarrow +\infty} \exp(a_x + \beta_x k_t) \rightarrow +\infty$ ,实际上,只要  $\lim_{t \rightarrow +\infty} m_{x,t} = \lim_{x \rightarrow +\infty} \exp(a_x + \beta_x k_t) > 1$  即为错误,因为死亡率不可能大于 1。

## 2.2.3 改进后的模型

为了能使模型对历史死亡率进行逆预测,首先要解决的是模型的后项收敛问题, Logistic 曲线十分适用,因为即使对模型后项进行收敛,对历史数据的预测也不同于对未来的预测,未来一些年龄的死亡率可以以 0 为极限,这对于正向预测几年,几十年不会出现什么问题,然而,模型反向收敛后,传统 Logistic 曲线极限为 1,即逆预测出的死亡率将接近于 1,是不符合实际情况的。因此,还需增加一个调整参数,该参数起到以下 3 个作用:

### (1) 限定函数的收敛性

调整未来预测年份死亡率的极限。对于历史回推逆预测,其分年龄死亡率将呈现逆向上升趋势,若照搬经典 Lee-Carter 模式,则死亡率极限为 1。但而实际情况是,由于历史上存在个别年份的人口普查和人口统计资料对当时的死亡率有一个大概分析,因此可以大概了解当时的死亡率水平,估算出极限范围,其值应当小于 1,而且,也不会非常接近于 1。

### (2) 提高了参数的拟合优度

### (3) 不同年龄组的变化趋势具有多样性

各年龄组的逐年变化趋势趋于单一性、同一性,实际上不同年龄组的逐年变化趋势应该呈现多样性,即增长(或减小)的加速度不同。

为了解决以上问题,本文对模型进行了改进,所提出的新模型记为: Lee-Carter-MY,其形式为:

$$m_{x,t} = \frac{1}{\gamma + e^{a_x + \beta_x k_t + \varepsilon_{x,t}}} \quad (5)$$

其中  $a_x$  表示关于时间的平均对数死亡率,  $\beta_x$  表示死亡变化率,  $k_x$  为随机时期效应, 其服从带漂移的随机游过程或 ARIMA 过程。  $\gamma$  是调整参数, 用于调整模型的收敛效果, 对于不同年龄段, 选用不同  $\gamma$  值, 既可以得到  $\beta_x$  较好的拟合效果, 又可以保证死亡率的收敛性, 对于逆预测尤为重要。

对于参数的求解, 暂不考虑残差, 且将  $\gamma$  视为常数, 故经简单变换

$$\frac{1}{m_{x,t}} - \gamma = e^{a_x + \beta_x k_t} \quad (6)$$

取对数后获得:

$$\ln\left(\frac{1}{m_{x,t}} - \gamma\right) = a_x + \beta_x k_t \quad (7)$$

以上方程式反映出新模型的假设: 死亡率倒数的对数值呈线性变化。为了能得到唯一解, 仿照经典 Lee-Carter 模型, 对该模型设定条件, 同方程式 (2)。

本模型的参数求解过程如下:

第一步: 以  $\gamma = 1$  完成  $\alpha_x$  和  $k_t$  的估计和  $\beta_x$  的初次估计;

第二步: 基于以上参数, 测算误差, 调整参数  $\gamma_x$ , 并重新估计  $\beta_x$ 。

(1)  $\alpha_x$  的估计

$\alpha_x$  和  $k_t$  的估计前提假设是  $\gamma = 1$ , 即模型 (5) 变为:

$$m_{x,t} = \frac{1}{1 + e^{a_x + \beta_x k_t}} \quad (8)$$

该模型类似于经典 Lee-Carter 模型, 只是将死亡率变化曲线由指数型变为了“S”型, 并且后向收敛性由发散限制为 1。

将方程式 (7) 两边对  $t$  求和获得:

$$\sum_{t=i}^{t=j} \left[ \ln\left(\frac{1}{m_{x,t}} - 1\right) \right] = \sum_{t=i}^j a_x + \sum_{t=i}^j \beta_x k_t \quad (9)$$

上式中  $i$  是死亡率实际观察数据的开始年,  $j$  是死亡率实际观察的结束年。在满足约束条件后, 因

为:  $\sum_{t=i}^j k_t = 0$  求得:

$$a_x = \frac{\sum_{t=i}^j \left[ \ln\left(\frac{1}{m_{x,t}} - 1\right) \right]}{n} \quad (10)$$

其中  $n$  为观察年份。

(2)  $k_t$  的估计:

对固定某年份的单岁组年龄求和得:

$$\sum_{all\ x} \left[ \ln\left(\frac{1}{m_{x,t}} - 1\right) - a_x \right] = \sum_{all\ x} \beta_x k_t \quad (11)$$

当  $t$  不变时, 也就是说对于同一年份, 不同年龄组  $k_t$  值是不变的。

又因为  $\sum_{x=0}^{100} \beta_x = 1$  所以有:

$$k_t = \sum_{x=0}^{100} \left[ \left( \frac{1}{\ln m_{x,t}} - 1 \right) - a_x \right] \quad (12)$$

(3)  $\beta_x$  的初步估计:

根据以上的估计值, 用最小二乘法求解方程:

$$\frac{[1/m_{x,t}] - 1}{\exp(a_x)} = \exp[\beta_x k_t] \quad (13)$$

利用非线性方程的拟合方法,对于某一个年龄组  $\exp(a_x)$  是一个常数项,将方程左边看成因变量,  $k_t$  看成因变量,便可估算出每个年龄组的  $\beta_x$ 。至此,基于  $\gamma = 1$  条件下的  $a_x$ ,  $k_t$ ,  $\beta_x$  拟合完成。

对于参数调整与重估计,需要选择合适的  $\gamma$  值,与  $\beta_x$  配合,实现既满足优化参数拟合的效果,又能对历史死亡率的变化规定一个合理的极限,以至保证逆向预测的准确性。

#### (1) 满足优化参数拟合的效果

通过大量实验,依靠 R-square 指标进行判断参数  $\beta_x$  的拟合优度。首先需要 R-square 的值接近于 1 的数量多,其次需要 Sum(R-square) 的值较大。

#### (2) 满足不同年龄组死亡率极限的合理设定

主要需要设定高、低龄的死亡率极限,即上下限。本文利用误差自调整估计参数  $\gamma$ 。另外基于历史数据,可以大概估算出某年高低龄的死亡率大概在何种水平,作为参考。

$\gamma$  的估计步骤:

第一步:将由  $\gamma = 1$  估计获得的  $a_x$ 、 $\beta_x$ 、 $k_t$  带入方程(5),方程左边设为拟合年份的死亡率准确值  $m_{x,t}$ ,方程右边为  $\frac{1}{\gamma + e^{a_x + \beta_x k_t}}$ ,令左边等于右边,则算得误差项  $\gamma$ ,这里它是随  $x$  和  $t$  变化的值  $\gamma_{x,t}$ 。

$$\gamma_{x,t} = \frac{1}{m_{x,t}} - \exp(a_x + \beta_x k_t) \quad (14)$$

第二步:对其进行关于  $t$  的均值化,得  $\bar{\gamma}_2 = \frac{\sum_{t=1}^j \gamma_{x,t}}{n}$ ,再进行中值滤波和平滑化处理,再将处理后的  $\bar{\gamma}_2$  带入式(8)代替  $\gamma$ ;

第三步:基于新的  $\gamma_x$ ,重新估计  $\beta_x$ ,新得到的  $\gamma_x$  为随  $x$  变化的参数,表明不同年龄组有不同的收敛性,且有不同的变化模式。

倘若参数  $\gamma$  值恒定为 1,则会导致随着  $k_t$  的急剧变化,0 岁组和高龄组的死亡率极限将逐渐逼近 1,这与经典 Lee-Carter 模型在做正向预测时 0 岁组的死亡率将趋于 0 是同一类的问题。

#### 2.2.4 逆推质量评估

对于本模型的质量评估可从以下两个方面展开(杨明旭 2011):

##### (1) 与已有研究结果对比

Luo Sheng(1988) 基于几次普查数据,分析年龄分布,构建死亡函数,重构 1953~1982 各年的分年龄(5 岁组)、性别人口分布及死亡率数据,并重构这期间的各年的生命表,弥补了我国在这期间的统计空白。虽然 Coale、Bannister 对这个问题也进行过研究,但是 Coale 和 Bannister 并未得出分年龄组的数据。本文将通过改进的 Lee-Carter 模型重构 1953 年至今的单岁组死亡率,利用此结果已重构出各年的生命表,并将其与 Luo 的结果进行对比(见表 1)。

由于本文未进行分性别的人口逆预测研究,故而所得结果为男女合计的总人口分年龄死亡率。其逆预算估算结果介于 Luo 文预测的分性别实际人口估测数据之间。需要在这里特别说明的是:本文为了比较有效精确计算出“大跃进”时期的非正常死亡人口数,首先利用 Lee-Carter-MY 模型逆向预测出的人口分年龄死亡率是“假设没有发生饥荒”的人口数据,而 Luo 文的估算则为实际发生饥荒的人口数据。且该结果显示:在“大跃进”时期死亡率明显升高,且男性高龄死亡率明显高于女性。

表1 1981年分年龄死亡率估计值对比

Table 1 Comparison of the Estimated Age-Specific Mortality in 1981 ‰

年龄	本文(总)	Luo(男)	Luo(女)
0	50.4	47.4	54.4
1	4.3	5.9	5.2
5	1.3	1.3	1.2
10	1.3	1.5	1.6
15	1	0.3	0.5
20	1.2	1.1	1
25	1.82	3.3	2.1
30	1.9	0.8	2.6
35	2.3	1.9	2.7
40	3.3	2.2	2
45	4.9	2	2.1
50	7.5	7.7	8.4
55	11.8	13.3	10.4
60	22	23.1	16.1
65	33.3	37.3	25.8
70	55.1	64.2	50
75	84.7	114.1	91.1
80	140.8	191.1	164.6
85	190.3		
90	236.2		
95	270.3		
100	311.9		

资料来源: Luo S, 1988 及本文逆向预测与仿真数据。

### (2) 与“三普”研究结果的对比

1982年的第三次人口普查数据的高质量性已获得学界的一致认可。本研究将估计结果与 Luo 的研究结果进行对比。

综上,在还原基础数据效果、逆预测结果、实际统计数据对比及与先前学者的研究一致性来看, Lee-Carter-MY(改进后的模型)的逆向预测精度是高的、效果是好的。

对于死亡率的估算,利用本研究方法求得结果与蒋正华、李南(1986)和 Luo(1988)学者的结论有较高相关性和一致性;本研究对期望寿命的预测为 67.32 岁,与蒋正华和李南的估算(67.88 岁)较为接近,比 Luo 的估计值(66.47 岁)偏高。

### 3 分年龄死亡率重构结果

参数的拟合方法要建立在基础数据的有效分析的基础之上,不同时期的死亡率都有各自的变化模式。针对我国不足 20 年基础数据的特点,若仅利用 20 年不到的基础数据去预测或逆预测 30 年甚至 50 年的死亡率变化模式,显然是不科学的。本文进行 5 年一组的逆预测,并将结果重构进基础数据,并依据已有统计数据对其进行相应优化处理。以此迭代预测,可以实现不同时期的逆预测都会有各自的一套拟合参数,死亡率呈现不同变化趋势的特点,并能保证逆预测的平稳性。因此,本文采用 5 年一次的预测方法。即先用 2009~1994 年的死亡率重构出 1993~1989 的数据,再整合 2009~1989 的数据,重构出 1988~1984 的死亡率,以此类推,回推至 1950 年,其间辅以个别有实际统计的年份数据,获得逆向仿真和逆向预测结果如图 1~图 4 所示。

(1)  $a_x, k_t, \beta_x, \gamma_x$  的估计结果

$a_x$  反映了各年龄组死亡率对数变化的基数,描述死亡率年龄模式的相对时间趋势;与经典模型的正向预测时呈现“U”型趋势不同,在逆预测中,由于该参数位于分母,故呈现倒“U”型趋势(见图1)。其次,由于在新模型 Lee-Carter-MY 中  $k_t$  与死亡率呈负相关,越靠前的历史年份  $k_t$  值越低,反映了死亡率在逆向时间上呈整体上升的趋势(见图2)。

图1  $a_x$  的拟合值

Figure 1 Fitted Value of  $a_x$

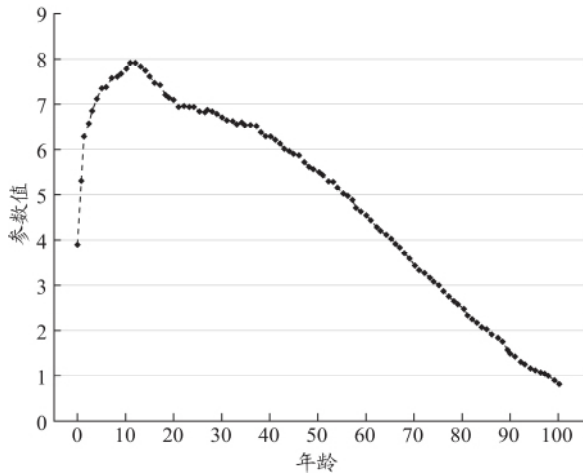
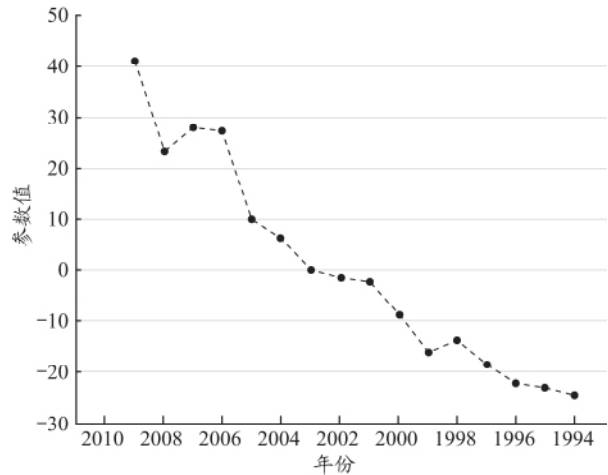


图2  $k_t$  的拟合值

Figure 2 Fitted Value of  $k_t$



资料来源: 本文仿真数据和逆向预测数据。

再次,低龄组人口的  $\beta_x$  值较高,主要是由于新生人口具有较高死亡率,对死亡率趋势的变化也最敏感,其中10岁、11岁组的  $\beta_x$  过低,主要是受基础统计数据的制约,导致对这两个年龄的拟合效果很差,85岁以上  $\beta_x$  呈下降趋势,原因在于高龄人口的死亡率随时间变化较小,实际死亡率对参数的变化不敏感(见图3)。

图3  $\beta_x$  的拟合值

Figure 3 Fitted Value of  $\beta_x$

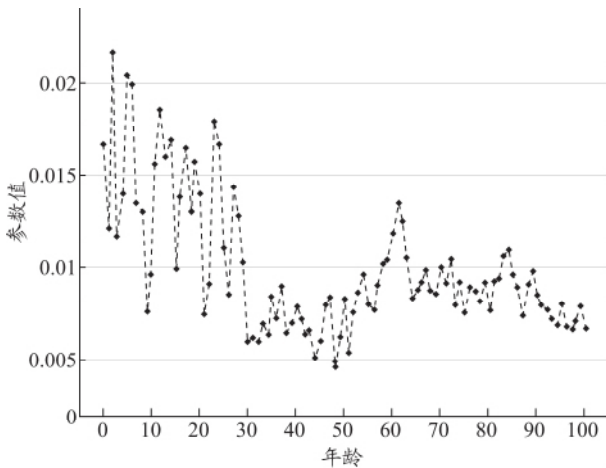
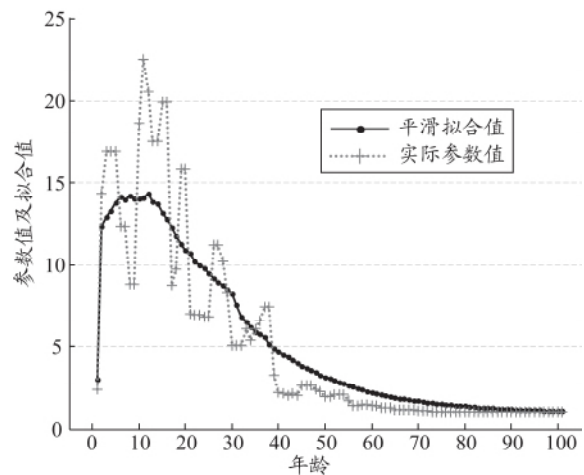


图4  $\gamma_x$  的拟合值

Figure 4 Fitted Value of  $\gamma_x$



资料来源: 本文仿真数据和逆向预测数据。

$\gamma_x$  的平滑拟合在于得到其趋势性。拟合后的值类似于  $a_x$ , 也呈倒“U”型。各相邻年龄的  $\gamma_x$  值不



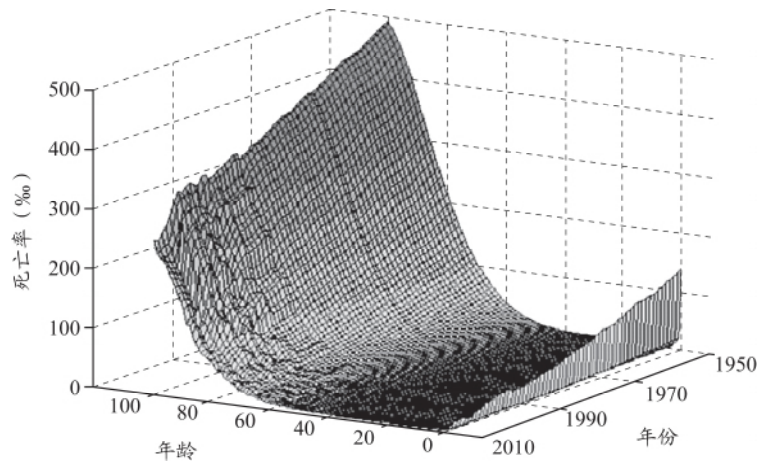
应该出现较大波动, 因为其值将决定各岁组的预测收敛极限, 而相邻岁组的死亡率预测极限差距应该不大(见图4)。

## (2) 拟合结果

$\beta_x$  拟合优度即 R-Square 大于 0.6 的点有 95 个, R-Square 大于 0.7 的点有 70 个, 拟合效果较好。最后, 应用新模型 Lee-Carter-MY 进行逆向仿真和预测的结果如图 5 和图 6 所示。

图 5 1953 ~ 1993 年单岁组死亡率

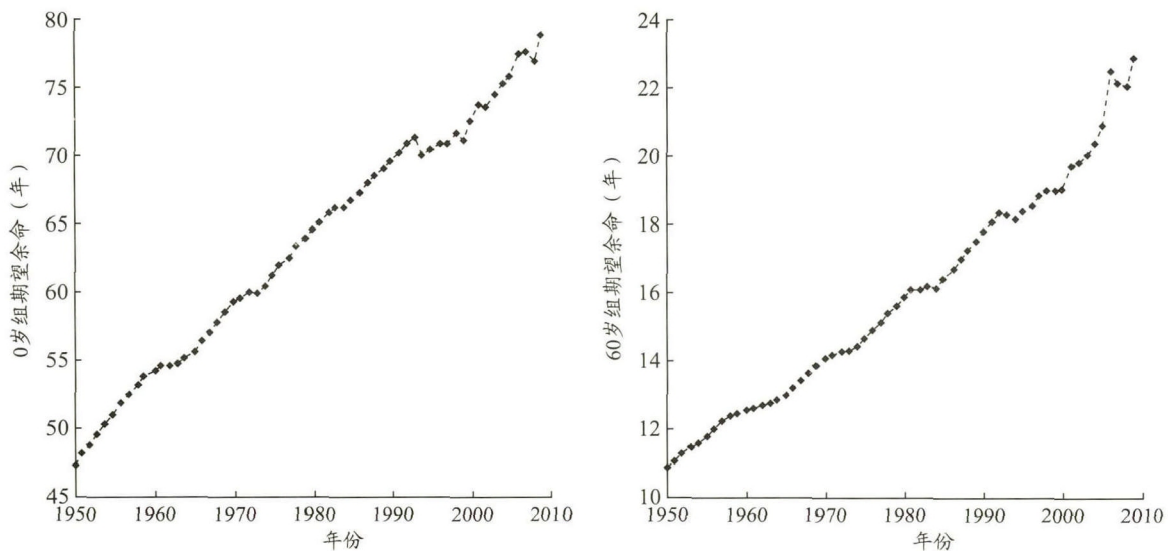
Figure 5 Mortality Rate of Single Age Group from 1953 to 1993



资料来源: 本文仿真数据和逆向预测数据计算所得。

图 6 0 岁组与 60 岁组期望寿命变化趋势

Figure 6 Trends in Life Expectancy at ages 0 and 60



资料来源: 本文仿真数据和逆向预测数据。

## 4 我国“大跃进”时期(1958 ~ 1961 年间) 非正常死亡人口分析

### 4.1 几个概念及研究边界

(1) 总死亡人口(实际死亡人口): 指这一时期全部死亡人口, 包括正常死亡人口和因饥荒而死亡的人口。

(2) 非正常死亡人口: 死于饥荒及由于饥荒引发疾病死亡的人口。即:

$$\text{非正常死亡人口} = \text{总死亡人口} - \text{正常死亡人口} \quad (15)$$

或:

$$\text{非正常死亡人口} = (\text{实际死亡率} - \text{正常死亡率}) \times \text{实际人口数} \quad (16)$$

(3) 人口损失: 非正常死亡人口和因饥荒而未出生的人口。

(4) 正常死亡人口数: 若不发生饥荒正常死亡的人口。

测算正常死亡人口,首先需要获得各年单岁组正常死亡率,再结合人口预测方法,重建各年的单岁组人口数。

本文的主要工作在于:假定不发生饥荒的情况下,“大跃进”时期(1958~1961年间)每年里的单岁组的正常死亡率应该是多少?另外,由于本文没有直接涉及该时期实际发生的每年单岁组的死亡率及其实际发生的死亡人口的估算,故以下所开展的对“大跃进”时期(1958~1961年间)的非正常死亡人口的估算,实际上是对上述各位中外学者已有研究结果的一种评估。即:基于他们对实际总死亡人口的估计分析结果,结合本文的逆向预测估算工作,得到对该时期的非正常死亡人口的估计值。还需要说明的是,本文所选取1958~1961年及1958~1963年这两个时间段,既考虑了灾荒从发生、逐步蔓延到情况严重和逐步减轻的全过程,也考虑了与前辈学者在时间段口径相一致的问题。

#### 4.2 逆向预测分析过程

(1) 重建1953年分年龄人口构成。

参照1953年人口普查的年龄结构,借鉴蒋正华、米红和张友干(1996)的重构方法,得出1953年的单岁组人口年龄结构。

(2) 基于当时情况,设定1953~1964年正常的婴儿出生率及出生性别比。

利用1953年和1964年两次普查所得结果的均值(李成瑞,1996),获得1954~1963年各年的出生率,本文设定的总出生率为40.5‰,出生性别比为105。

(3) 将重建的1953年人口结构、各年出生率、性别比、分年龄死亡率代入人口预测模型,计算获得1953~1964年各年的人口结构。

(4) 参照几位学者对1958~1963年及1958~1961年各年实际总人口的估计,按预测得到的人口结构进行各年龄组分摊,再结合单岁组正常死亡率,求得不同实际总人口下的正常死亡人口。

(5) 总死亡人口量参照已有研究成果,估算1958~1963年及1958~1961年非正常死亡人口数量。

#### 4.3 逆向预测结果

##### 4.3.1 总人口及总损失人口

对于“大跃进”时期我国的总人口数及总损失人口的估算,本研究主要是在对历史文献的考查的基础上重构了“大跃进”前后各年度的总人口数(见表2)。

因此,若用本文估计的每年应有总人口(假设“大跃进”未发生)减去各位学者估计的每年的实际人口,则为总损失人口,包括非正常死亡及未出生人口。

由于学者们对每年实际人口的估计差别不大,所以用本文估计的应有总人口(假设不发生饥荒)减去实际人口得出的趋势基本一致。

因此,为与中外各位学者的研究时段相一致,本文将选取两个时段评估“大跃进”时期的人口非正常死亡人口:其一是1958~1963年;其二是1958~1961年。

表2 历史文献及本文重构的“大跃进”前后各年度总人口数对照表  
Table 2 Comparison of Annual Total Population in Our Work and Previous Studies

年份	本文	Coale	Bannister	Calot	蒋正华
1953	58800				
1954	60200				60300
1955	61400	61400	60700	61400	61500
1956	62600	62600	61900	62700	62600
1957	63700	63900	63300	64500	63900
1958	64900	64500	63900	65800	64800
1959	66100	65200	65100	66300	65500
1960	67300	64900	65300	65300	65200
1961	68600	64600	64800	65000	64900
1962	69900	65500	65300	66400	66800
1963	71200	68700	67400	68000	68700

资料来源: (1) 本文仿真数据和逆向预测数据计算所得; (2) Coale ,1984; Bannister ,1985; Calot ,1985; 蒋正华、李南 ,1986。

注: (1) 本文逆预测分析和重建的总人口数据为假设未发生“大跃进”,按正常的出生率和死亡率变化而得到的“应有人口”总数如表4第2列所示); (2) 各位学者重建人口为“大跃进”发生后的实际人口(如表第3列-第6列);从1958年开始(本文预测的人口总数开始),其结果开始大于各位学者重构的人口,原因在于“大跃进”导致的人口过量非正常死亡。

此外,总损失人口从1959年起快速上升,至1961年或1962年达到顶峰,1961~1962年间变化不大,而1962年及以后由于生育的补偿效应,出生率开始高于正常年份的出生率,且此时死亡率也开始大幅下降,从而导致人口损失曲线的下降。由此明显看出“大跃进”实际引起的饥荒是从1958年就在我国的部分省份逐渐开始,而1959年、1960年及1961年则是在全国性蔓延和情况特别严重的3年。

因此,本文认为研究“大跃进”时期的非正常死亡人口的时段,选择“1958~1961”应是最为严谨的和科学的。

#### 4.3.2 各年正常死亡人口及总正常死亡人口测算

总损失人口包括未出生人口及非正常死亡人口,研究非正常死亡人口,未考虑各年龄组不同死亡率,通过对单岁组死亡人口求和得到。计算步骤为:

(1) 将几位学者得出的各年总人口,按照本文预测的各年人口结构,拆分为各单岁组人口数;

(2) 计算单岁组死亡人口,即:

$$\text{单岁组死亡人口} = \text{单岁组人口} \times \text{单岁组死亡率} \quad (17)$$

(3) 对各单岁组死亡人口按年龄求和:

$$D_t = \sum_{i=0}^{100} P_{i,t} \times q_{i,t} \quad (18)$$

其中  $D_t$  表示  $t$  年的总死亡人口,  $P_{i,t}$  表示  $t$  年年龄为  $i$  岁的人口,  $q_{i,t}$  表示  $t$  年年龄为  $i$  岁的死亡率。总死亡人口数除以各年总人口数,获得各年粗死亡率,结果如下(见表3):

表 3 1953 ~ 1964 年各年正常死亡人口与粗死亡率(假设没有发生“大跃进”)

Table 3 Annual Natural Deaths and Crude Death Rate During 1953-1964 (Assuming no Great Famine)

年份	总人口(人)	正常死亡人口(人)	粗死亡率(‰)
1953	594350000	11669052	19.63
1954	604377430	11535132	19.09
1955	614701515	11426000	18.59
1956	625567021	11338941	18.13
1957	636915594	11251785	17.67
1958	648748888	11103846	17.12
1959	660895459	11040751	16.71
1960	673276278	10962556	16.28
1961	685899625	10744194	15.66
1962	698873105	10532090	15.07
1963	712432481	10339914	14.51
1964	726569955	10091475	13.89

资料来源: 本文仿真数据和逆向预测数据计算所得。

因此,由本文逆向预测所得 1958 ~ 1961 年总正常死亡人口为 43851347 人,但是这个结果并不能等同于 Coale 与蒋正华等学者得出的实际死亡人口(见表 4)。原因在于,本文这个结果的前提假设是并未发生“大跃进”。

表 4 几位学者估计的各年份粗死亡率

Table 4 Estimated Annual Crude Death Rate by Previous Studies

年份	Bannister	Coale	Calot	蒋正华
1954	24.2	29.1	19.96	
1955	22.33	22.4	22.31	17.23
1956	20.11	20.8	16.85	16.77
1957	18.12	19	13.24	16.55
1958	20.65	20.4	15.98	17.25
1959	22.06	23.3	19.2	18.96
1960	44.6	38.8	40.76	31.25
1961	23.01	20.5	27.03	24.57
1962	14.02	13.7	18.28	18.08
1963	13.81	13	21.22	16.72
1964	12.45	13.5	20.82	13.03

资料来源: Coale 1984; J. Banister 1985; G. Calot 1985; 蒋正华, 李南 b, 1986; 袁永熙, 1991。

实际的死亡人口应以饥荒年间各年的实际总人口为基数,再以各年的分年龄死亡率为测算依据得出的数据。

利用几位学者研究获得的总人口数据,对单岁组死亡人口求和得到实际总死亡人口数(见表 5):

表 5 实际死亡人口数对照表  
Table 5 Comparison of Total Actual Deaths

年份	Coale	Bannister	Calot	蒋正华
1958	11237012	11126074	11460631	11290479
1959	11127944	11095876	11315064	11168540
1960	10716662	10781932	10797299	10780114
1961	10307882	10340929	10375891	10365035
1962	10133692	10106619	10270911	10327841
1963	10102460	9919865	10010648	10115408
总计(1958~1961)	43389500	43344811	43948885	43604168
总计(1958~1963)	63625652	63371294	64230444	64047418

资料来源: 本文仿真数据和逆向预测数据计算所得。

通过对单岁组死亡人口求和获得实际总死亡人口后,再结合几位学者已有的研究数据,计算出1958~1963年各年总死亡人数(见表6):

表 6 1958~1963年各年总死亡人口①  
Table 6 Annual Total Deaths in 1958-1963

年份	Coale	Bannister	Calot	蒋正华
1958	13162488	13192253	10515799	11101065
1959	15200454	14350030	12736320	12323810
1960	25163740	29101500	26633807	20470000
1961	13236440	14904728	17567878	15953792
1962	8974185	9159266	12136458	11840411
1963	8925930	9310702	14437451	11303222
总计(1958~1961)	66763112	71548511	67453804	59848667
总计(1958~1963)	84663237	90018478	94027713	82992301

资料来源: 本文仿真数据和逆向预测数据计算所得。

#### 4.3.3 非正常死亡人口测算

利用本文逆向预测所获得的正常死亡人数以及按李成瑞、Coale及蒋正华和李南学者对各年总人口的估算,通过单岁组死亡率重新计算得出总死亡人数,若以1958~1963为时间段,对各学者所得非正常死亡人口数进行修正,估算结果如表7的第二行和第三行所示。

表 7 本文对各位学者研究结果的修正(1958~1963年)

Table 7 Correction of Previous Calculations (1958-1963)

死亡人数	Coale	李成瑞	蒋正华
总死亡人数(修正)	8466	8466	8299
假定无“大跃进”发生的正常死亡人数(本文计算)	6363	6462	6405
“大跃进”时期的非正常死亡人数	2103	2004	1894

资料来源: (1) 本文仿真数据和逆向预测数据计算所得。(2) Coale,1984; 李成瑞,1998; 蒋正华,李南 a,1986; 袁永熙,1991。

① 1958~1963年各年总死亡人数计算模型为: 各年总死亡人口 = 各年实际总人口 × 各年实际死亡率。

若以 1958 ~ 1961 年为时间段,由于李成瑞的研究缺乏分年的总死亡人口数据,故本研究只对蒋正华与 Coale 的研究进行修正。

对 Coale 有关 1958 ~ 1961 年非正常死亡人口的修正结果高于 1958 ~ 1963 年的非正常死亡人口,主要由于在 Coale 的研究中,1962 年和 1963 年的出生率相对上升较快,从 1961 年的 22‰ 升到 1963 年的 47‰,而实际死亡率相对下降也较快(从 1961 年的 20.5‰ 下降到 1963 年的 13‰),实际死亡率甚至低于了本研究估算的正常死亡率,故从 Coale 的研究可以看出:若将 1962 年和 1963 年归入“大跃进”时期,则会造成 1962 年和 1963 年补偿的人口抵消因饥荒而非正常死亡的人口。这也再次说明了“大跃进”时期的人口非正常死亡人口的估算的科学时间段应该是 1958 ~ 1961 年。Coale 在其结论中定义大饥荒的时间段为 1958 ~ 1963 年,但是按照对其出生率和死亡率的分析可以看出,1962 年和 1963 年实际已经不属于饥荒期了。因此,本文认为 Coale 关于非正常死亡人口的最终估算数偏大,且结论缺乏严谨性。

综上所述,本文应用改进的 Lee-Carter 模型,在假定无“大跃进”发生时,逆向预测出了 1958 ~ 1961 年属于正常状态下的死亡人数,并结合 Coale 和蒋正华两位学者针对“大跃进”时期(1958 ~ 1961 年间)的研究结果,进一步计算出其非正常死亡人口数。

因此,大跃进时期(1958 ~ 1961 年间)如果不发生饥荒,则 4 年正常死亡人口介于 4339 ~ 4385 万人之间。由于饥荒的发生,非正常死亡人口介于 1624 ~ 2337 万人之间。

## 5 结论

(1) 本文首先基于中国 1953 年以来的人口普查历史数据特点,对经典的 Lee-Carter 分年龄死亡率预测模型进行了改进和创新,形成了分年龄死亡率逆向预测仿真模型。仿真结果表明:Lee-Carter-MY 对拟合我国分年龄死亡率数据精度有所提高,并且能够运用到分年龄死亡率的逆向(对历史年份死亡率的回推)预测与仿真分析之中。

同时,本文通过 Lee-Carter-MY 模型还原了部分历史数据,并结合逆预测结果与实际统计数据的对比及与以往文献的研究等进行了综合评估,看出 Lee-Carter-MY 模型的逆向预测精度是高的、仿真效果是好的。

(2) 基于 Lee-Carter-MY 分年龄死亡率逆向预测仿真模型,本文重构了 1953 年以来各单岁组的“正常状态下”的分年龄死亡率(假设没有发生“大跃进”时期的饥荒),结果表明,中国人均期望寿命 1953 年为 48.43 岁,1959 年应为 52.23 岁,到 1964 年应为 55.24 岁。通过对已有数据及人口回推数据的分析,本文认为从 1962 ~ 1963 年,随着总和生育率大幅度上升至 7.3 及死亡率的大幅的下降,非正常死亡人口已经非常少,且人口已经出现补偿趋势。

另外,本文有关“大跃进”时期人口变动尤其是对非正常死亡人口测算的时间段口径选择为 1958 ~ 1961 年,是严谨的和科学的。主要基于以下考量:实际的饥荒是从 1958 年开始发生,在 1959 年开始蔓延并在 1960 年及 1961 年情况严重,1962 年开始并在 1963 年大大减轻并转为补偿生育。之前的学者主要是针对 1958 ~ 1963 年间研究,本文的修正结果表明:非正常死亡人数为 1624 ~ 2337 万人,属于 1958 ~ 1961 年的时间段;而 1958 ~ 1963 年间的非正常死亡人数为 1894 ~ 2103 万人。之所以后者的非正常死亡人数并未随年份增加,反而有所减少,其主要原因是由于 1962 年和 1963 年生育补偿的出现。

因此,本文在得出“大跃进”时期历年正常状态下单岁组死亡率及总死亡人口数(假设没有发生饥荒的状态)之后,结合改进后的模型 Lee-Carter-MY 进行逆向仿真和逆向预测估算出 1958 ~ 1961 年间的正常状态的死亡人口数介于 4339 万 ~ 4385 万人,其结果与前辈学者李成瑞、蒋正华的估算相近。

(3) 本文对学者 Coale、蒋正华关于对非正常死亡人口的研究做出评估。在 1958 ~ 1961 年间,若

按蒋正华对总死亡人口估算的数据来计算,得出非正常死亡人口应为 1624 万人,若按 Coale 的数据计算则为 2337 万人。两者相差约 700 万人,主要是由于两位学者对实际总死亡人口数的估算差别较大。即“大跃进”时期(1958~1961 年间)的非正常死亡人口数的中值为 1980 万人,其变化值处在区间[1624 万人,2337 万人]之中。另外需要说明的是,Coale 在其另一个时段 1958~1963 年的估算结论反而比 1958~1961 年时段要小得多。通过对其出生率和死亡率的分析可以看出:其原因在于 1962 和 1963 实际已经不属于饥荒期了,而是处于生育补偿期了。其中,1962 年的总和生育率为 6.02,1963 年的总和出生率更是已经达到了 7.5,这无疑也是迄今为止新中国以来的最高总和生育率!因此,本文认为 Coale 关于“大跃进”时期(1958~1961 年间)实际死亡人口的最终估算数在整体上是偏大的,从而,也影响到其非正常死亡人口数的估算值也是偏高的。

综上所述,就估算中值来说,本文认为关于“大跃进”时期(1958~1961 年间)的非正常死亡人口数不应高于 1980 万人。

#### (4) 局限性讨论

本文尚存在局限性及有待进一步深入的地方。首先,本文是基于分年龄的死亡率展开分析,未对死亡率进行分性别研究;其次,本研究在分析“大跃进”时期人口变动情况时,限于篇幅,并未对“大跃进”时期的实际总死亡人口进行测算,而是直接借鉴和引用了先前学者们的研究成果,这些将会在未来的研究中有待加强。

#### 参考文献/References:

- 1 Carter, L. R. and A. Prskawetz. 2001. Examining Structural Shifts in Mortality Using the Lee-Carter Method. *Mpidr Working Papers* 7: 57-68.
- 2 Chen, Y. 2014. The Great Famine in China, 1958-1962: A Documentary History. *Historian* 76: 157-158.
- 3 AJ Coale J. 1984. Rapid Population Change in China, 1952-1982. Washington, D. C.: National Academy Press, 13: 89.
- 4 Dikötter, F. 2010. Mao's Great Famine: the History of China's most Devastating Catastrophe, 1958-1962. *Library Journal* 18: 54-56.
- 5 Eilers, P., H., C. A., Delwarde, and M., Denuit. 2007. Smoothing the Lee-Carter and Poisson Log-bilinear Models for Mortality Forecasting: A Penalized Likelihood Approach. *Statistical Modelling* 1: 29-48.
- 6 Kane, P. 1988. Famine in China 1959-61. Demographic and Social Implications. Macmillan Press.
- 7 Lee, R. D. and L., R., Carter, 1992. Modeling and Forecasting U. S. Mortality. *Journal of the American Statistical Association* 419: 659-671.
- 8 Luo S. 1988. Reconstruction of Life Tables and Age Distributions for the Population of China, by Year, from 1953 to 1982; Phd Dissertation, University of Pennsylvania.
- 9 (美)班久帝(J. Banister). 对中国人口普查结果的惊奇与确认. 国际人口科学研究联合会(IUSSP) 1985 年佛罗伦萨会议文集, 1985  
J. Banister. 1985. Surprising and Confirmation of the Results of China's Census. *Collected Papers of IUSSP 1985 Florence Conference*.
- 10 曹树基. 1959-1961 年中国的人口死亡及其成因. *中国人口科学*, 2005; 1: 14-28  
Cao Shuji. 2005. China's Population Death During 1959-1961 and Its Causes. *Chinese Journal of Population Science* 1: 14-28.
- 11 (法)G. 卡洛(G. Calot). 为便于分析中国人口统计资料的建议. 国际人口科学研究联合会(IUSSP) 1985 年佛罗伦萨会议文集, 1985  
(France) G. Calot. 1985. Suggestions on China's Demographic Data Analysis. *Collected Papers of IUSSP 1985 Florence*

- Conference.
- 12 蒋正华,李南 a. 中国人口动态参数的校正. 西安交通大学学报,1986; 3:46-64  
Jiang Zhenghua and Li Nan. 1986. Parameter Correction of China's Population Dynamics. Journal of Xi'an Jiao Tong University 3:46-64.
  - 13 蒋正华,李南 b. 中国人口动态估计的方法和结果. 中国人口年鉴. 经济管理出版社. 1986  
Jiang Zhenghua and Li Nan. 1986. Methodology and Results of China's Population Dynamics Estimation. Demographic Yearbook of China. Economic Management Press.
  - 14 蒋正华,米红,张友干. 1946~1949年中国大陆人口向台湾及海外迁移估计. 中国人口科学,1996;4:1-12  
Jiang Zhenghua, Mi Hong and Zhang Yougan. 1996. Estimation of Population Migration from Chinese Mainland to Taiwan and Overseas During 1946 and 1949. Population Science of China 4:1-12.
  - 15 李成瑞. “大跃进”引起的人口变动. 人口研究,1998;1:3-12  
Li Chengrui. 1998. Population Change Led by the Big Leap. Population Research 1:3-12.
  - 16 (美)麦克法夸尔,(美)费正清,谢亮生. 剑桥中华人民共和国史:(上卷)革命的中国的兴起 1949~1965年. 中国社会科学出版社,1990  
Roderick Macfarquhar, John King Fairbank, Xie Langsheng. 1990. Cambridge Chinese History: Volume 1 Rise of Revolutionary China 1949-1965.
  - 17 杨明旭. 人口死亡率逆预测方法创新及实证研究. 硕士论文[10384],厦门大学. 2011  
Yang Mingxu. 2011. Method Innovation and Empirical Study of Mortality Reverse Prediction. Master Thesis [10384], Xiamen University.
  - 18 袁永熙. 中国人口总论. 中国财政经济出版社,1991  
Yuan Yongxi. 1991. On China's Population. Finance and Economic Publishing House of China.
  - 19 祝伟,陈秉正. 中国城市人口死亡率的预测. 数理统计与管理,2009;4:736-743  
Zhu Wei and Chen Bingzheng. 2009. Prediction of China's Urban Population Mortality. Application of Statistics and Management 4:736-743.

(责任编辑:沈 铭 收稿时间:2015-08)