Statistical and Application 统计学与应用, 2015, 4(3), 155-161
Published Online September 2015 in Hans. http://www.hanspub.org/journal/sa
http://dx.doi.org/10.12677/sa.2015.43017

Hans汉斯

Application of Lee-Carter Model in Predicting Mortality

Qiuyun Zhang

School of Mathematics and Information Sciences, Guangzhou University, Guangzhou Guangdong
Email: 18819484749@163.com

Received: Aug. 17th, 2015; accepted: Sep. 5th, 2015; published: Sep. 8th, 2015

Abstract

Changes in mortality rates have an impact on the social security system, which is a time-varying dynamic random variable. Basic pension insurance payment shall comply with the dynamic changes of population mortality. In this paper, using Lee-Carter model [1] forecasts the future mortality trends of population, the prediction of time factor κ_t in the future, the paper uses generalized difference model for prediction. And compared with the ARIMA model, the result of generalized difference model was better.

Keywords

Mortality Prediction, Lee-Carter Model, ARIMA Model, The Generalized Difference Model

Lee-Carter模型在死亡率预测中的应用

张秋芸

广州大学数学与信息科学学院，广东 广州
Email: 18819484749@163.com

收稿日期：2015年8月17日；录用日期：2015年9月5日；发布日期：2015年9月8日

摘　要

人口死亡率变化对社会保障系统有一定影响，它是时变的动态随机变量。基本养老保险的给付须遵从人

文章引用: 张秋芸. Lee-Carter 模型在死亡率预测中的应用[J]. 统计学与应用, 2015, 4(3): 155-161.
http://dx.doi.org/10.12677/sa.2015.43017

口动态死亡率变化规律。本文利用Lee-Carter [1]模型对未来人口死亡率变动趋势进行预测。对未来时间因子κ_t的预测时，用广义差分模型预测。并与ARIMA模型比较，结果广义差分模型预测效果较好。

关键词

死亡率预测，Lee-Carter模型，ARIMA模型，广义差分模型

1. 引言

在人口统计学以及精算科学中，人口死亡率预测模型一直以来都是热点问题。随时间的延续，死亡率改善导致的人口预期寿命的增加已经成为全球性趋势。死亡率预测作为养老金计划财务规划的基础，对政府养老金制度、雇主企业年金、保险公司的团体和个人养老金业务等都有重要的影响，关系到各类养老计划的财务安全和可持续发展。根据是否考虑未来死亡率变动的不确定性，把死亡率预测模型分为确定型死亡率模型(静态死亡率模型)和随机型死亡率模型(动态死亡率模型)两大类。其中，确定型死亡率模型包括 Gompertz 模型、Makeham 模型、Weibull 模型、Kannisto 模型和 T.N.Thile 模型等，但是这些模型没有考虑死亡率的随机变动，因而不适用于死亡率的预测，只适合做模型的拟合；而随机型死亡率模型包括 Lee-Carter 模型、多因素年龄-时期模型、Renshaw-Haberman 队列效应模型和 Cairns-Blake-Dowd 模型等。在随机型死亡率模型中，被公认为是随机预测方法中最典型的一个即美国人口学家 Lee Ronald D 和 Carter Lawrence R。于 1992 年提出的一种预测美国未来人口死亡率变化的概率模型，该模型的优点在于人口统计模型与统计时间序列方法相结合，不需要引入能对死亡率产生影响的医疗、环境以及社会等因素，从而有效地减少了主观判断因素对于预测结果的影响，且只有时间因子一个参数需要预测，并在时间因子趋势上没有其他假设。Lee-Carter 模型假设未来各个年龄组的死亡率将依据历史数据的变动情况，预测其未来趋势。此方法属于外推模型，对数表达形式和 ARMA 构成了 Lee-Carter 方法的主要特征。

在 Lee-Carter 模型中，对未来时间因子的预测时，用 ARIMA 模型，但是 ARIMA 模型只适合做短期预测，做中长期预测会导致误差很大，所以为了降低预测误差，可以用其它的模型来预测，本文选用广义差分模型，发现其预测误差更小。

2. 数据说明

本文采用芬兰 1971~2012 年男性死亡率数据。其中死亡率数据分为 22 组，0~5 岁的分为两组 0~1，1~4，其他 5 岁为一个组别，末组为 100~104。数据来源于人口死亡率数据库(The “Human Mortality Database” at www.mortality.org)。

3. Lee-Carter 模型

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 x 岁的人群在第 t 年的中心死亡率，那么 $m_{x,t}$ 满足以下函数关系式。

$$\ln\left(m_{x,t}\right)=\alpha_x+\beta_x\kappa_t+\xi_{xt} \tag{1}$$

其中，α_x 反应 x 岁年龄组别的对数中心死亡率的平均水平，κ_t 为人口死亡率随时间变化的速度，β_x 为年龄因子对 κ_t 的敏感度，即 x 岁年龄组别的死亡率随着时间变化的大小，ξ_{xt} 为随机误差项，假设其服从正态分布 $N\left(0,\delta^2\right)$。为了得到唯一确定的参数估计值，加入约束条件 $\sum_x\beta_x=1$ 以及 $\sum_t^T\kappa_t=0$。$\sum_t^T\kappa_t=0$ 是为了保证参数 α_x 的平均值的含义，即 $\alpha_x=\frac{1}{T}\sum_t^T\ln\left(m_{x,t}\right)$。

3.1. 模型的求解

Lee-Carter模型参数的估计方法主要包括矩阵奇异值分解法(SVD)、最小二乘(OLS)和加权最小二乘法(WLS)；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Koissi [2]证明了，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。为此，本文采用加权最小二乘法来估计Lee-Carter模型。

加权最小二乘法通过以下两个步骤求得 $\hat{\kappa}_t$， $\hat{\beta}_x$

第一步，将式(1)两边对年龄x 求和，得到 $\hat{\kappa}_t = \sum_x \left[\ln\left(m_{x,t}\right) - \hat{\alpha}_x\right]$。

第二步，Wilmoth [3]证明 $\ln\left(m_{xt}\right)$ 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 $\sum_{x,t} d_{x,t}\left[\ln\left(m_{x,t}\right) - \hat{\alpha}_x - \hat{\beta}_x\kappa_t\right]^2$，即 $\min \sum_{x,t} d_{x,t}\left[\ln\left(m_{x,t}\right) - \hat{\alpha}_x - \hat{\beta}_x\kappa_t\right]^2$ 得到 $\hat{\beta}_x = \sum_{t=1}^{T} d_{x,t}\kappa_t\left[\ln\left(m_{xt}\right) - \hat{\alpha}_x\right] \Big/ \sum_{t=1}^{T} d_{x,t}\kappa_t^2$ [4]。

3.2. 拟合的结果

用加权最小二乘法拟合数据，估计参数的结果如下(表1，图1)。

Table 1. Estimation results of corresponding parameters of α_x, β_x
表 1. 相应的 α_x， β_x 参数的估计结果

年龄组 x	α_x	β_x	年龄组 x	α_x	β_x
0~1	−5.19622	0.083816	50~54	−4.78424	0.045025
1~4	−8.08868	0.079879	55~59	−4.35742	0.047078
5~9	−8.45755	0.092181	60~64	−3.92364	0.049946
10~14	−8.436	0.071453	65~59	−3.49202	0.051193
15~19	−7.12615	0.041819	70~74	−3.039	0.048626
20~24	−6.71815	0.026724	75~79	−2.56664	0.043405
25~29	−6.61755	0.030425	80~84	−2.08531	0.03624
30~34	−6.40915	0.033821	85~89	−1.61177	0.027287
35~39	−6.07316	0.038381	90~94	−1.12668	0.0176
40~44	−5.6565	0.042175	95~99	−0.6368	0.014908
45~49	−5.21892	0.044414	100~104	−0.1517	0.012784

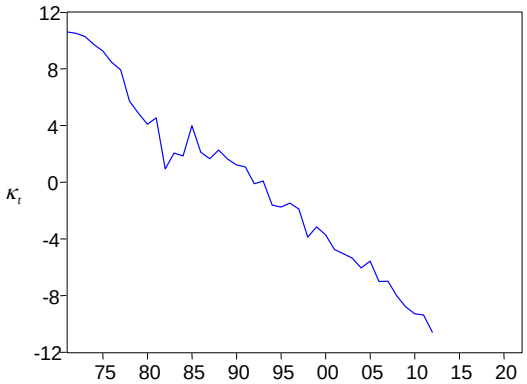


Figure 1. The charge of estimates of κ_t over time
图 1. κ_t 的估计值随时间的变化

从表 1 中 α_x 的估计值可以看出 0~1 岁年龄组至 70~74 岁年龄组，α_x 的估计值相对较高，75 岁以上年龄组 α_x 的估计值相对较低。

表 1 β_x 的估计值显示，低龄组人口(0~1 岁至 10~14 岁)的 β_x 的估计值较高，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，高年龄组(90~94 岁及以上年龄组)的估计值较低，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。

图 1 给出了 κ_t 的估计值随时间变化的图像，从中可以看出 κ_t 的估计值随着时间的变化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. 模型的预测

对于未来时间因子 κ_t 的预测，本文用 ARIMA 模型和广义差分模型。并将两者比较，选出较优的模型。

4.1. ARIMA 模型

ARIMA 模型是 Box 和 Jenkins，1970 年[5]提出的以随机理论为基础的时间序列分析方法，又称为“Box-Jenkins 模型”，这一模型在经济领域的预测分析中得到广泛的应用。时间序列是依赖时间 t 的一组随机变量，构成该时序的单个序列值虽然具有不确定性，但对整个时间序列来说，它的变化却具有一定的规律性，可以用相应的数学模型来近似描述。一个 ARIMA(p, d, q)模型，由三个过程组成：自回归模型(AR(p))，移动平均模型 MA(q)，单整(I(d))。

1) 一般的 p 阶自回归过程 AR(p)是

$$x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \cdots + \varphi_p x_{t-p} + \varepsilon_t \tag{2}$$

2) 一般的 q 阶的移动平均过程 MA(q)可以表示为

$$x_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \tag{3}$$

将纯 AR(p)与纯 MA(q)结合得到一个一般的自回归平均过程 ARMA(p, q)

$$x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \cdots + \varphi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \tag{4}$$

3) 单整自回归移动平均过程 ARIMA(p, d, q)

假设一个随机过程含有 d 个单位根，其经过 d 次差分之后，可以变换为一个平稳的自回归移动平均过程。则该随机过程称为单整自回归移动平均过程，记为 ARIMA(p, d, q)。

因为 κ_t 是随着时间变化，所以可以用 ARIMA 模型来预测未来的时间因子 κ_t，通过比较后发现 ARIMA(0, 1, 1)较合适，其预测模型为

$$\kappa_t = -0.518369 + \kappa_{t-1} + \varepsilon_t - 0.442787\varepsilon_{t-1} \tag{5}$$

4.2. 广义差分模型

在经济计量研究中，随机误差项自相关是一种十分普遍的现象。在这种情况下，仍应用普通最小二乘法(即OLS方法)建立的预测模型，尽管所得的结果为无偏的，但估计量不具有最小方差性，t、F检验失效，并且会使得预测置信区间增大，降低了预测精度。当然也就失去了预测的意义。如果在模型 $y_t = b_0 + b_1 x_t + u_t$ 中作广义差分变换[6]，得到广义差分模型即可消除自相关性从而提高预测的精度。

设线性回归模型 $y_t = b_0 + b_1 x_t + u_t$ 存在一阶自相关性：$u_t = \rho u_{t-1} + v_t$，其中 v_t 为满足古典回归模型基本假定的随机误差项。将模型滞后一期，得

$$y_{t-1} = b_0 + b_1 x_{t-1} + u_{t-1}$$

在方程两边同乘以 ρ，并与原模型相减得

$$y_t-\rho y_{t-1}=b_0(1-\rho)+b_1(x_t-\rho x_{t-1})+(u_t-\rho u_{t-1}) \tag{6}$$

定义变量变换

$$\begin{cases} y_t^*=y_t-\rho y_{t-1} \\ x_t^*=x_t-\rho x_{t-1} \end{cases} \tag{7}$$

称式(7)为广义差分变换，模型(6)可以表示成如下形式

$$y_t^*=A+b_1x_t^*+v_t \tag{8}$$

其中， $A=b_0(1-\rho)$。式(8)是经过广义差分变换得到的模型，称为广义差分模型。

4.2.1. 建立线性回归方程

从图 1 可以看出 κ_t 与时间 t 近似成线性关系。所以设方程为 $\kappa_t=b_0+b_1t$，为了计算简便，把 1971 年作为第一年 $t=1$，其他的年份以此类推。用 EVIEWS 软件求出 $b_0=10.45415$，$b_1=-0.486236$。相应的 p 值都小于 0.05，说明 κ_t 与 t 的线性关系显著。

4.2.2. 检验自相关系

用 EVIEWS 将模型的残差进行拉格朗日 LM 乘数检验。选择滞后期为一阶。显示 $\text{LM}=13.8307>\chi^2_{0.95}(1)=3.8415$，对应的$P=0.0005$小于0.05，说明随机误差项存在一阶自相关。

4.2.3. 自相关修正

对线性回归模型进行广义差分变换，得出 $b_0=10.35274$，$b_1=-0.484094$，$\rho=0.54039$，再进行LM检验，$\text{LM}=2.929275<\chi^2_{0.95}(1)=3.8415$，相应的$P$值为0.2312，说明随机误差项不存在自相关性。最后预测模型为

$$\kappa_t=10.35274-0.484094t+\left[\text{AR}(1)=0.45039\right] \tag{9}$$

5. ARIMA 模型与广义差分模型比较

通过计算模型拟合的平均绝对误差、平均相对误差、均方根误差、Theil不等系数、偏差比率、方差比率和协方差比率(表2)，可以明显看出，广义差分模型的拟合精度优于ARIMA模型。所以对于 κ_t 的预测，选择广义差分模型。

6. 模型的预测结果检验

为评价模型的有效性，运用 Lee-Carter 模型预测未来 3 年 2013，2014，2015，各年年龄段的死亡率，再求出 2011~2015 年 5 年的平均死亡率与联合国世界人口组织分布 2011~2015 年的死亡率作比较，如下图 2。

结果显示 0~60 岁误差较小，60~80 岁误差相对较大，而且略大于联合国世界人口组织公布的死亡率，这是因为 Lee-Carter 模型是用以前的数据来预测未来的，不能够准确估算出未来随着时间，医学的进步和公共卫生新技术的应用以及文化教育水平提高对死亡率的影响大小。其次高年龄组误差较大，主要是因为高年龄组死亡率不确定性较大。这些都是避免不了的。

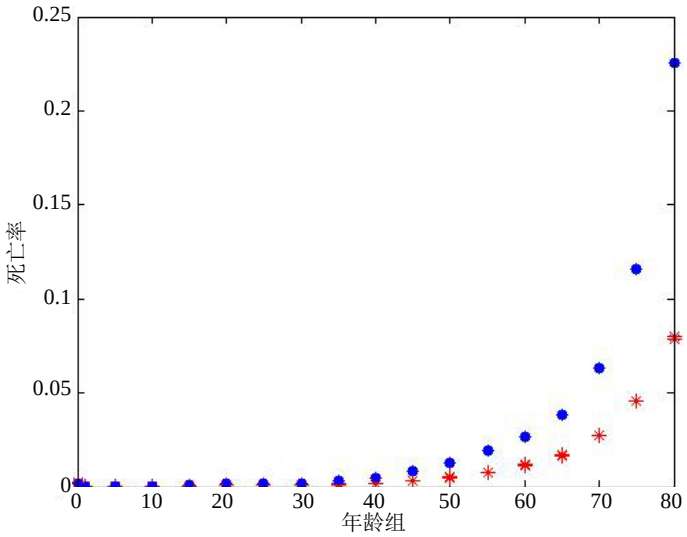
总的来说该模型在预测未来死亡率上还是可行的。

7. 死亡率降低对养老保险金的影响

随着死亡率的降低，人口寿命的延长，越来越多的国家将会面临人口老龄化所带来的压力，芬兰也不例外。用以上 Lee-Carter 模型，预测 2013~2022 年芬兰男性婴儿的预期寿命如下图 3 所示。

Table 2. Comprehensive evaluation index
表 2. 综合评价指标

	广义差分模型	ARIMA(0, 1, 1)模型
均方根误差(RMSE)	0.975089	1.077868
平均绝对误差(MAE)	0.727842	0.772532
平均相对误差(MPE)	62.72091	61.00301
Theil不等系数(U)	0.084163	0.090164
偏差比率	0.000355	0.003708
方差比率	0.002758	0.088758
协方差比率	0.996887	0.907534



•表示预测的死亡率，*表示联合国世界人口组织公布的死亡率

Figure 2. Comparison of mortality average predicted value and the actual value for five years among 2011-2015
图 2. 2011~2015 年五年死亡率平均预测值与实际值的比较

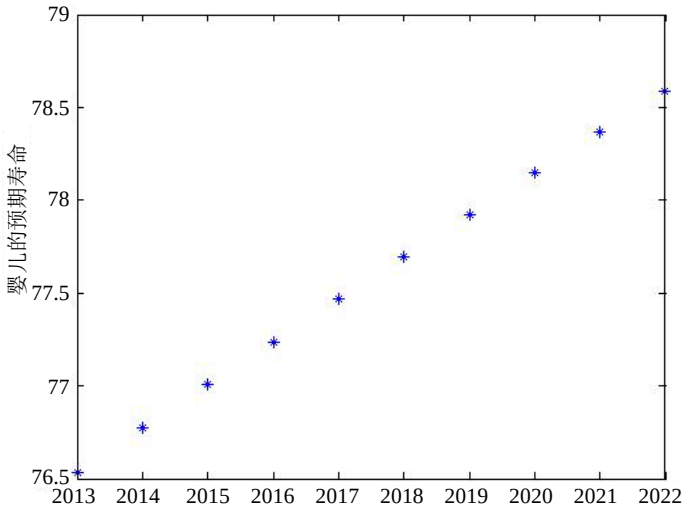


Figure 3. The change of life expectancy of male infants over time
图 3. 男性婴儿预期寿命随着时间的变化

从图 3 中可以看出随着时间的变化，芬兰男性人口预期寿命不断地增加，到了 2022 年预期寿命达到了 78.5 岁，这就意味着，领取养老金的人会越来越多，这将会对芬兰的财政带来严峻的挑战。

不过，面对不断加剧的老龄化态势，芬兰采取多轮延迟退休，如上世纪 70 年代以前，芬兰人的退休年龄是 60 岁，1978 年降至 58 岁，1980 年降到 55 岁，1987 年上调至 60 岁。2005 年芬兰将退休年龄由 60 岁提高到 63 岁，身体健康且愿意继续工作的可延长至 68 岁。以此来缓解财政压力，并取得了一定效果。这对我国有一定的启示的作用。

因为我国同样面临预期寿命的延长带来的财政的压力，建国初的时候，我国人均预期寿命只有 40 岁左右，现在人均的预期寿命，“第六次全国人口普查”的数据是 74.8 岁。也就是说，建国 60 多年来，我国经济社会发展、人口数量、人口结构、人口预期寿命，都发生了巨大的变化。在建国初期制定的退休年龄政策，即 1953 年《劳动保险条例》规定的：女工人退休年龄是 50 岁，女干部 55 岁，男职工 60 岁，很显然和当前经济社会的发展不相适应，所以我国有必要像芬兰一样对退休年龄作出调整。

8. 总结

本文是用 Lee-Carter 模型的对芬兰男性人口死亡率进行预测。对于未来时间 κ_t 的预测，本文应用了广义差分模型。较于已有方法 ARIMA 模型，广义差分模型预测误差更小。最后分析了死亡率降低，对芬兰国家的财政影响，以及对我国的一些启示。

参考文献 (References)

[1] Lee, R.D. and Carter, L.R. (1992) Modeling and forecasting US mortality. *Journal of the American Statistical Association*, **87**, 659-671.

[2] Koissi, M.C., Shapiro, A.F. and Hognas, G. (2006) Evaluating and extending the lee-carter model for mortality forecasting: Bootstrap confidence interval. *Insurance Mathematics and Economics*, **38**, 1-20. http://dx.doi.org/10.1016/j.insmatheco.2005.06.008

[3] Wilmoth, J.R. (1996) Mortality projections for Japan: A comparison of four methods. Health and mortality among elderly population. Oxford University Press, New York.

[4] 李志生, 刘恒甲 (2010) Lee-Carter 死亡率模型的估计与应用——基于中国人口数据的分析. *中国人口科学*, **3**, 47-56.

[5] Box, G.E.P. and Jenkins, G.M. (1970) Time series analysis: Forecasting and control. Holden-Day, San Francisco.

[6] 王新军 (1993) 广义差分模型及预测应用. *山东经济*, **1**, 57-60.