文章编号:1006-9348(2011)05-0358-05

差分方程在人口增长预测中的应用研究

曾 维
(成都理工大学信息科学与技术学院,四川 成都 610059)

摘要:在人口增长预测的研究中,关于人口总数、性别、年龄结构等预测,一般预测模型准确性较低,预测结果缺乏指导意义;而双曲模型和微分方程模型计算复杂,很难求解,不便于科技水平较低地区使用。为解决上述问题,基于灰色预测模型,提出差分方程构建人口增长预测模型。模型能仿真出不同年龄、不同类型人口变化情况,能准确预测结构变化、性别比例变化、城镇化情况等。应用模型对中国人口进行仿真预测,得到结果与国家人口发展战略研究人口发展预测相符合,为人口增长提供了较好的预测模型。

关键词:差分方程; 人口增长; 残差; 预测
中图分类号:O411.3 文献标识码:A

Study of Differential Equation Application in the Forecast of Population Growth

ZENG Wei
(College of Information Science and Technology, Chengdu University of Technology, Chengdu Sichuan 610059, China)

ABSTRACT: In the study of population growth forecast, the accuracy of common prediction model is somewhat low and the results lack of guidance. However, the hyperbolic model and the computational complexity of differential equation model are both complex in calculating and also difficult to solve. It is not convenient for use in the area with low technology. To solve these problems, based on the gray prediction model, a differential equation model of population growth has been built. The model simulates different ages, different types of demographic changes. It can accurately predict the structural changes, changes in sex ratio, urbanization, and so on. The model was applied to predict China's population. The result is consistent of the population development prediction in the study of national population development strategy. The calculation of residuals is less than 0.2 and the model predicts good results.

KEYWORDS: Differential equation; Population growth; Residual; Forecast

1 引言

人作为社会的主体,承担社会建设的全部工作。科学的预测地区人口总数、性别比例、年龄结构,有利于科学地配置人力资源,帮助政府和企业有效的解决劳动力问题。因此对人口的预测是一个由来已久的研究课题,但传统方法往往是使用单一的模型(如灰色系统模型、自回归模型和指数模型等)进行计算,考虑的影响因素不够全面。事实上,人口的增长的影响因素是多方面的,如出生率、死亡率、城市化进程、国家生育政策、经济发达程度、生活方式、传统习惯和自然灾害等。结合我国的发展实际(中国是一个经济高速发展的发展中国家,人口多、区域经济发展不平衡、人口老龄化加剧和相当长一段时间里实行计划生育的国策等),考虑其中主要的影响因素(如人口结构、人口年龄分布以及老龄化等)建立了模型。

基于差分方程的思想建立了人口预测模型,以中国2000-2005人口数据为基础数据。在对数据进行评估和修订的基础上,建立中长期人口预测函数,仿真不同年龄不同类型人口的变化情况,同时通过模型可以清楚看到各因素对人口的影响。

2 差分方程分析方法的原理依据

人口增长模型是由生育、死亡、疾病、灾害、环境、社会、经济等诸多因素影响和制约的共同结果,如此众多的因素不可能通过几个指标就能表达清楚,它们对人口增长的潜在而复杂的影响更是无法精确计算。这反映出人口系统具有明显的灰色性,适宜采用灰色模型去发掘和认识原始时间序列综合灰色量所包含的内在规律。

灰色预测模型属于全因素的非线性拟合外推类法,其特

收稿日期:2010-12-20

点是单数列预测，在形式上只用被预测对象的自身序列建立模型，根据其自身数列本身的特性进行建模、预测，与其相关的因素并没有直接参与，而是将众多直接的明显的和间接的隐藏着的、已知的、未知的因素包含在其中，看成是灰色信息即灰色量，对灰色量进行预测，不必拼凑数据不准、关系不清、变化不明的参数，而是从自身的序列中寻找信息建立模型，发现和认识内在规律进行预测。

通过对人口数据的分析，不难看出影响人口发展的主要因素有如下几个：

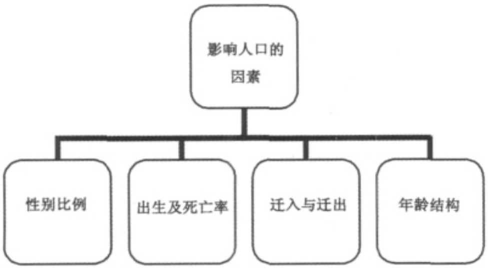

图 1　影响人口的因素

人口增长的过程中，某年某年龄的人口到了第二年就会转变为下一年龄的人口。同时，该转变又被上述因素影响。这样，人口增长数据自然构成一个数列，可以从序列角度剖析微分方程，了解其构成的主要条件，然后对近似满足这些条件的序列建立近似的微分方程模型。而对序列而言(一般指有限序列)只能获得有限差异信息，因此，用序列建立微分方程模型，实质上是用有限差异信息建立一个无限差异信息模型。

但由于微分方程求解困难，在模型建立过程中笔者将微分方程转化为差分方程，应用差分方程构建预测模型。

3　模型建立

根据预测时间的不同，分别建立了中短期和长期预测模型。

3.1　人口增长中短期预测模型

在此模型中主要是对某一年的总人口数进行预测，对总人口的预测是对这几个指标的综合处理，建立了一个有关生育率，死亡率等因素的求解人口增长差分方程模型，人口的发展实际是一个动态的循环递推过程，可知第 n 年的人口数由各年龄组人数组成，可得式

$$Z_n = \sum_{k=0,1} \sum_{j=0}^{2} \sum_{i=0}^{90} X_{ijk}^n \tag{1}$$

把第 i 年龄组作为研究对象，其中又分两种情况讨论：(1) 当 $i = 0$ 时，即这一年才出生的 0 年龄人口组，是该年各年龄组育龄妇女人数(Y_j^n)在生育率为(b_j^n)的情况下所出生的新增人口，；(2) 当 $1 < i < 90$ 时，由于年龄的动态变化，今年第 $i+1$ 年龄组的人口由上一年第 i 年龄组的人口数(X_{ij}^n)转化而来，在转化过程中死亡率(D_{ij}^n)的存在及可能从其他的类型迁入第 j 类型人口或迁出第 j 类型人口数(A_{ij}^n：净迁出为负；净迁入为正)，所以要除去退出该递推过程的人口数。所以建立差分方程模型：

$$\begin{cases} X_{i+1,j,k}^{n+1} = X_{ijk}^n - D_{ijk}^n X_{ijk}^n + A_{ijk}^n \quad (0 < i < 89) \\ X_{0jk}^n = \beta_j^n Y_j^n S_{jk}^n B_j^n \\ X_{90j,k}^{n+1} = X_{89jk}^n - D_{89jk}^n X_{89jk}^n + X_{90jk}^n (1 - D_{90jk}^n) \\ Y_j^n = \sum_{i=15}^{49} X_{ijk}^n \quad (k = 1) \end{cases} \tag{2}$$

通过模型，控制参数 i,j,k 值的变化能够得到下列预测值：

1) 第 n 年总人口数为 $Z_n = \sum_{j=0}^{2} X_{ijk}^n (i = 0,1\cdots 90+, k = 0,1)$；

2) 第 n 年的男、女人数分别为 $\sum_{j=0}^{2} X_{ijk}^n$ ($k = 0$ 为男人数；$k = 1$ 为女人数)

3) 第 n 年的总出生人数为：$\sum_{k=0,1} X_{0jk}^n$

通过分析可以看到生育率(B_j^n)、死亡率(D_{ijk}^n)，出生男女性别比和迁入或迁出(A_{ijk}^n)的发展变化都有其自身的规律。因此可作为分析和预测的重要指标，对不同指标所给的数据特点选取不同的预测方法。

3.1.1　生育率的确定

充分利用了灰色系统 $GM(1,1)$ 模型在短期预测问题上准确度比较高的优点对市镇乡的生育率预测建立 $GM(1,1)$ 模型进行灰色预测。根据人口统计资料及数据处理后，构造原始数据列

$$x^{(0)}(k) = \{x^0(1), x^0(2), \cdots x^0(n)\} \quad k = 1,2,\cdots n \tag{3}$$

对原始数据进行一次累加生成，得到灰化数据列 $x^l(k) = \sum_{k=1}^{n} x^0(k)$，建立 $GM(1,1)$ 微分方程

$$\frac{dx^{(1)}}{dt} + ax = u \tag{4}$$

用最小二乘法求解参数向量 a 和 u，

$$a = [a,u]^T = (B^T B)^{-1} B^T Y \tag{5}$$

其中，

$$B = \begin{bmatrix} -\frac{1}{2}[x^{(1)}(1) + x^{(1)}(2)] & 1 \\ -\frac{1}{2}[x^2(2) + x^1(3)] & 1 \\ \vdots & \vdots \\ -\frac{1}{2}[x^{(1)}(n-1) + x^1(n)] & 1 \end{bmatrix}$$

$$Y = \begin{bmatrix} x^0(2) \\ x^0(3) \\ \vdots \\ x^0(n) \end{bmatrix}$$

建立时间响应函数，求解微分方程得

$$x^{(1)}(k+1) = (x^{(1)}(0) - u/a)e^{-ak} + u/a \quad (6)$$

求导还原后得 $GM(1,1)$ 预测方程

$$x^{(1)}(k+1) = -a(x^{0}_{(xy)}(1) - \frac{u}{a})e^{-at} \quad (7)$$

把处理后的数据带入求解便可得到生育率。

3.1.2　死亡率的确定

根据提供的数据死亡率求平均值计算。

3.1.3　出生男女所占比例

鉴于男女出生比例保持一定比例的特点，这里使用 $Matlab$ 工具箱进行一元线性回归可以对第 j 类人口出生男女性性别比例(S^n_{jk}) 进行有效预测。

3.1.4　人口的迁入与迁出量的确定

但对于迁移人数问题即我国实际中人口城镇化，题中没有给出直接相关的数据，假设人口城镇化水平年平均增长为 1%，为了简化问题，设转移到城镇的人口各占 1/2，男女分配比例为 1.2:1(根据农村人口的性别比)，那么第 n 年的城市增加人口 = 第 $n-1$ 年的人数 - 死亡人数 + 转移的人数，转移人数为第 $n-1$ 年的人数的 1%，镇人口同理，而农村人口减去上一年转移到城镇的人数，这也是一个逐年递推转移的过程

$$A^n_{ij} = \frac{X^n_{ijk}}{\sum_{i=0}^{90} X^n_{ijk}} \cdot Z_n \cdot 1\% \cdot g_{jk} \quad (8)$$

其中 $g_{jk} = \begin{bmatrix} \frac{1.2}{4.4} & \frac{1.2}{4.4} & \frac{-1.2}{2.2} \\ \frac{1}{4.4} & \frac{1}{4.4} & \frac{-1}{2.2} \end{bmatrix}$，表示各类型间男女的迁移系数。

3.2　人口增长长期预测模型

对于人口的长期预测，仍然利用人口增长中短期预测模型中的差分方程。由于 $GM(1,1)$ 模型进行长期预测误差较大，所以需改变出生率的计算方式。据资料表明，在稳定条件下各年龄妇女的生育率(C^n_{ij}) 基本不变(与是时间无关) 。本模型将其处理为定值，建立模型如下:

$$\begin{cases} X^{n+1}_{i+1,j,k} = X^n_{ijk} - D^n_{ijk}X^n_{ijk} + A^n_{ijk} \quad (0 < i < 89) \\ X^n_{0jk} = \beta S^n_{ijk} \sum_{i=15}^{49} (X^n_{ij1} C^n_{ij}) \\ X^{n+1}_{90j,k} = X^n_{89jk} - D^n_{89jk}X^n_{89jk} + A^n_{89j} + X^n_{90jk}(1 - D^n_{90jk}) \end{cases} \quad (9)$$

由于是进行长期预测，以及其他因素(医疗水平的提高、人口文化以及妇女地位的提高等) 的影响，使模型中的指标出现一定的变化。根据生活经验可知主要有如下几种影响:

1) 医疗水平的提高使得死亡率下降;

2) 人口文化提高使得人口城镇化加快;

3) 妇女地位的提高使得出生男女性别比例的降低和 0 岁女婴死亡率降低。

据此对模型中的死亡率(D^n_{ij}) 、出生男女所占比例(S^n_{ijk}) 和迁移量(A^n_{ij}) 的变化率进行修正:

1) 对死亡率的修正，年龄在 75 以后为:

$$D^n_{ijk} - n\delta_1; i \geqslant 75;$$

2) 镇乡 0 岁女婴的死亡率为:

$$D^n_{ijk} - n\delta_2; i = 0, j = 1,2$$

3) 对出生男女所占比例的修正量为 δ_3，其中女的修正后为 $S^n_{ij1} + n\delta_3$，男相对为 $S^n_{ij0} - n\delta_3$;

4) 城镇化水平增加率修正后为 $0.1 + n\delta_4$。

根据实际预测情况调整上述修正量可以提高模型精确度。

4　预测模型的仿真

4.1　短期人口预测模型求解

4.1.1　生育率计算

构建模型时，首先通过灰色系统 $GM(1,1)$ 模型对生育率(b^n_j) 预测。当原始数据完整时，可直接带入数据进行计算。如原始数据有缺失，可以通过线性插值估算出缺失数据，再进行计算。

笔者根据中国人口数据求解出各类育龄妇女生育率的预测方程:

$$x^{(1)}(k+1) = -a(x^{0}_{(xy)}(1) - \frac{u}{a})e^{-at} \quad (10)$$

根据对数据的处理计算得到了城镇乡分别的预测系数 a，u(见表 2) 。

表 2　预测系数

	城	镇	乡
a	0.0796	0.0322	0.0393
u	51.9276	58.1562	60.9107

将系数带入方程，则可对某一年各类型人口的生育率进行计算预测，如对 2010 ~ 2015 预测如表 3。

表 3　2010 ~ 2015 生育力预测

	城	镇	乡
2010	15.4734	28.2938	33.3502
2011	14.2773	27.3971	32.0547
2012	13.1737	26.5289	30.8096
2013	12.1555	25.6881	29.6128
2014	11.2159	24.874	28.4625
2015	10.3489	24.0856	27.3569

4.1.2　性别比例计算

对于男女性别比例，用一元线性回归(Matlab) 计算出的预测方程如表 4。

表 4　男女性别比例预测方程

	y
城	$0.11x + 110$
镇	$0.084x + 120$
乡	$0.44x + 120$

由此算式预测得到 2010 ~ 2015 男女性别比例年数据如表 5。

表 5　男女性别比例数据

	城	镇	乡
2010	111.76	121.344	127.04
2011	111.87	121.428	127.48
2012	111.98	121.512	127.92
2013	112.09	121.596	128.36
2014	112.2	121.68	128.8
2015	112.31	121.764	129.24

4.1.3　中短期总人口预测

对上面各指标进行预测分析后，把人口迁移的计算式(8)，在环境较为稳定的情况下死亡率 D^n_{ijk} 可视为定值。该值可通过数据取平均算出，将 A^n_{ij} 和 D^n_{ijk} 带入式(2) 解出 $X^{n+1}_{i+1,j,k}$，再带入式(1) 求出第 n 年总人口数 Z_n。为与实际数据经行比较，笔者通过模型对 2006 ~ 2015 经行了预测运算，预测数据如表 6。

表 6　总人口预测

	总人口(百万)
2006	1319.559
2007	1330.698
2008	1340.558
2009	1349.236
2010	1356.751
2011	1363.155
2012	1368.262
2013	1371.984
2014	1374.452
2015	1375.691

利用 Matlab 画出短期总人口预测图如图 2 所示。

由于模型的一般性，还可根据式 $X^{n+1}_{i+1,j,k} = X^n_{ijk} - D^n_{ijk}X^n_{ijk} + A^n_{ij}$，求出第 n 年城镇乡男女分别的人数，以便在此基础上做更精确的预测。

4.2　长期人口预测模型求解

因为在模型建立中所讨论的一些因素的影响，对模型

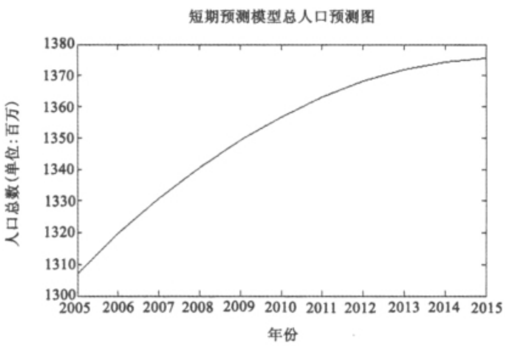

图 2　短期人口预测图

(2) 中的一些指标进行修正，而修正的量视情况而定，修正标准因人而异。对数据修正处理后运用差分方程(2)，求解后长期的总人口数 $\sum_{k=0,1}\sum_{j=0}^{2}\sum_{i=0}^{90}X^n_{ijk}$，得到长期模型预测数据(单位:百万人) 及人口变化如图 3。

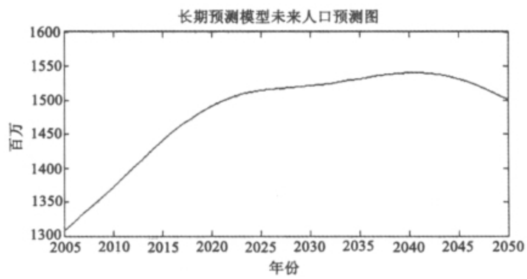

图 3　长期模型预测图

进而可以根据改变 i,j 的值得到了 2005 ~ 2050 年的各类型男女分别的人口数以及每年的男女比例(见图 4)。

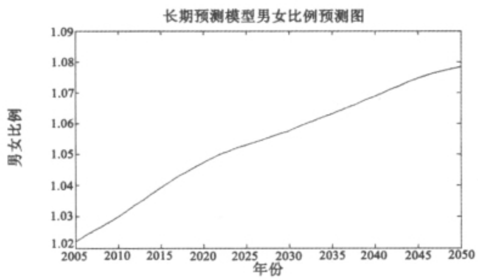

图 4　长期预测模型男女比例预测图

另外也可以改变 i 的值来计算各年龄的人数，所以定 $i >$ 65 或 60 求每年老年人口的数量，从而对我国的人口老龄化问题进行长期的预测和分析(见图 5)。

分析图可知在 2028 左右以后，老龄化程度有较快的增长，到达 2040 年后开始进入一个平稳状态，保持在 0.3 的水平。该图形跟国家人口发展战略研究人口发展预测相符合，说该模型有较强的可靠性。

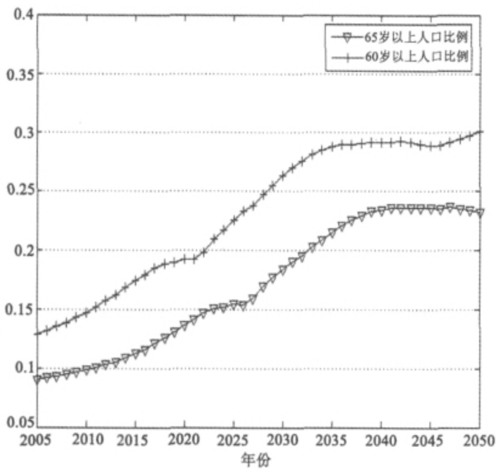

图5　人口老龄化长期预测图

4.3　可靠性分析

对于预测模的可靠性，一般是通过残差检验。令残差为 $\varepsilon(k)$，计算

$$\varepsilon(k) = \frac{x^{(0)}(k) - \hat{x}^{(0)}(k)}{x^{(0)}(k)} \quad (k = 1,2,\ldots n) \qquad (10)$$

如果 $\varepsilon(k) < 0.2$，则可认为达到一般要求；如果，$\varepsilon(k) < 0.1$ 则认为达到较高的要求。

计算 $GM(1,1)$ 模型中各预测数据如下：

表7　各级生育率预测的残差分析

	市生育率	镇生育率	乡生育率
1996	0.1686	0.0324	0.0035
1997	-0.1507	-0.0569	0.0069
1998	-0.1081	-0.0072	0.0067
1999	-0.056	0.0399	-0.0062
2000	-0.0352	0.0129	-0.0132
2001	-0.0187	-0.0188	-0.0226
2002	-0.0942	-0.0499	-0.0077
2003	-0.0097	0.0126	0.0121
2004	0.1453	0.0441	0.034
2005	0.1258	-0.0192	-0.0134

上述数据中所有残差都小于 0.2 表明 GM(1,1) 模型预测效果很好。

5　结束语

模型在建立的过程中，考虑了多个因素(出生率、死亡率、各个年龄男女比等)比以往的人口预测模型如指数增长模型、阻滞增长模型(logistic 模型)有了很大的改进。由于模型是根据人口密度推总人口分布的，人口与时间和年龄都有联系，所以可以推出将来的人口的年龄结构。

此外，模型还涉及到女性性别比、总和生育率，这对我国的计划生育政策提供依据和手段。对今后一段时间内的预测很好的吻合了我国最新的国家人口战略发展研究报告。

同时，根据模型的推演方式，也可代入其它数据从而得到特定地区的预测模型，具有较好地可移植性。可以方便地应用于对地区人口的预测，帮助改善区域人口计划。

参考文献：

[1] 胡仕明，黄国石. Logistic 模型的反馈控制分析[C]. Systems Engineering, Systems Science and Complexity Research - Proceeding of 11th Annual Conference of Systems Engineering Society of China, 2000.

[2] 王学萌，张继忠，王荣. 灰色系统分析及实用计算机程序[M]. 武汉：华忠科技大学出版社，2001.

[3] 吴自库. Logistic 人口模型参数伴随同化识别[J]. 科学技术与工程，2008,(16).

[4] 谭东风，王斌. 基于战斗方程的婚姻人口模型[J]. 计算机仿真，2008,(2).

[5] 罗建梅，原三领，中素慧. 一类非线性年龄结构人口模型[J]. 上海理工大学学报，2008,(1).

[6] 黄荣清. 关于人口预测问题的思考[J]. 人口研究，2004,28(1).

[作者简介]

曾　维(1979-)，男(汉族)，四川省绵阳市人，讲师，博士研究生，主要研究领域为电子通讯，矿物电子材料。