

中国人口增长预测模型的建立与分析

摘要

针对我国人口发展过程中出现的老龄化进程加快,出生人口性别比持续升高,乡村人口城镇化的新特点,我们基于 LESLIE 矩阵,着重考虑城镇与乡村间的人口迁移及女性人口比例变化对我国人口增长的影响,经过两次改进建立了便于计算机求解的差分方程模型,对我国 2005 年以后 45 年的人口增长进行了预测。随后利用时间段参数设置法,对差分方程模型又进行了一次改进。然后运用等维灰色系统预测法对该差分方程模型的中短期预测进行了检验,同时根据 2001 年人口基本数据运用此模型对 2001 年~2005 年进行了预测,并用实际数据对预测结果进行了检验。

我们将预测区间分为 2006~2020 年、2021~2035 年、2036~2050 年三个区间,以量化短期、中期与长期。通过调整模型中相关参数及输入条件,定量地分析了男女性别比例、老龄化和乡村人口城镇化对我国人口增长的影响。预测结果表明,从短期来看,我国的出生性别比变化不明显,将在短期内维持基本不变,老龄化进程在 15 年内在上升了 8 个百分点,人口扶养比持续升高,这将加重我国的人口压力,乡村人口城镇化水平进展缓慢;从中期来看,总人口性别比将保持在 1 与 1.1 之间,老龄化进程将呈线性增加趋势,乡村人口城镇化水平将持续发展;从长期来看,老龄化进程将在 2035 到 2045 年经历老龄人口高峰平台,老龄人口比重在 0.3 以上,育龄妇女人数持续下降,总人口数将在 2023 年达到峰值 14.05 亿。

关键词: LESLIE 矩阵, 人口预测, 性别比例, 城镇化, 老龄化, 灰色系统预测

一、问题的重述

人口问题是中国社会发展的重要问题,对中国人口的中长期预测有助于政府制定相应的政策保持中国的长治久安。

现需要解决的问题如下:

1. 主要根据 2001~2005 年的人口统计数据,对中国人口增长的中短期和长期趋势作出预测,特别要关注老龄化,出生人口性别比及乡村人口城镇化等因素。
2. 指出所建模型的优点和不足之处。

二、模型假设

1. 在未来 50 年人口生存的社会环境相对稳定(即没有战争及毁灭性灾难)。
2. 国际人口迁入与迁出量相等。
3. 在本世纪中叶前,我国计划生育政策稳定。
4. 题目所给抽样数据是随机的,真实地反映了整体实际情况。

三、符号说明

| | |
|--------------------------------------|---------------------------------|
| $d_i^1(t), d_i^2(t), d_i^3(t)$ | 分别表示乡村、镇、市第 t 年 i 岁人口的死亡率; |
| $x_i^1(t), x_i^2(t), x_i^3(t)$ | 分别表示乡村、镇、市第 t 年 i 岁的人口数; |
| $b_i^1(t), b_i^2(t), b_i^3(t)$ | 分别表示乡村、镇、市第 t 年 i 岁的女性生育率; |
| $k_i^1(t), k_i^2(t), k_i^3(t)$ | 分别表示乡村、镇、市第 t 年 i 岁人口的女性比; |
| $c^1(t), c^2(t), c^3(t)$ | 分别表示乡村、镇、市第 t 年的婴儿死亡率; |
| $f^1(t), f^2(t), f^3(t)$ | 分别表示乡村、镇、市第 t 年的出生人数; |
| $h_i^1(t), h_i^2(t), h_i^3(t)$ | 分别表示乡村、镇、市第 t 年 i 岁女性的生育模式; |
| $\beta^1(t), \beta^2(t), \beta^3(t)$ | 分别表示乡村、镇、市第 t 年的总和生育率; |
| $N^1(t), N^2(t), N^3(t)$ | 分别表示乡村、镇、市第 t 年的总人数; |
| $w_i^1(t), w_i^2(t), w_i^3(t)$ | 分别表示乡村、镇、市第 t 年 i 岁女性的总人数; |
| $wd_i^1(t), wd_i^2(t), wd_i^3(t)$ | 分别表示乡村、镇、市第 t 年 i 岁女性的死亡率; |
| $m_i^1(t), m_i^2(t), m_i^3(t)$ | 分别表示乡村、镇、市第 t 年 i 岁男性的总人数; |
| $md_i^1(t), md_i^2(t), md_i^3(t)$ | 分别表示乡村、镇、市第 t 年 i 岁男性的死亡率; |
| r | 表示为迁移人口中女性所占比例; |
| z^1, z^2, z^3 | 分别表示乡村、镇、市出生人口中女性所占的比例; |

四、问题的分析

人口发展过程的定量预测,需要预测出未来的人口发展趋势,包括人口总数、人口的性别、年龄和城乡构成,人口出生、死亡和自然增长率的变化以及在未来的人口构成中劳动力和抚养水平及老龄化水平等各项人口指数全部测算出来。人口增长的决定因素为出生率、死亡率和人口基数,但人口分布,人口素质,宏观政策和人口结构(如:年龄结构,性别比例等)等众多因素能够影响出生率与死亡率的波动,从而从根本上影响我国人口的增长。鉴于我国人口问题已有多方面的研究,我们针对近年来我国的人口发展出现的一些新特点,忽略国际人口流动,故可以认为我国人口为一个封闭的系统。对于封闭的系统来说,某时刻人口总量=人口基数+新生人口数-死亡人口数。为了提供更多关于市、镇、乡的人口增长分布趋势,我们对三者分别进行研究。但由于市、镇、

乡人口在我国城镇化进程中迁移流动量大，故必须把人口迁移考虑进去。在本论文中，我们希望对我国人口进行中短期及长期预测。现在给出三个区间，（2006年□2020年）、（2026年□2035年）、（2036年□2050年）分别表征从2005年算起的短期、中期与长期。建立 LESLIE 矩阵差分方程模型，采用人口普查的相关数据、运用灰色系统预测方法对模型的合理性进行检验，从而可以预测我国人口发展的中短期和长期趋势。

五、 模型建立与求解

5.1 基于 LESLIE 矩阵的模型的建立^[1]

考虑到市，镇，乡村的人口性别比例，妇女生育率以及人口的死亡率都有所差别，我们分别建立市，镇，乡村的差分方程模型。市，镇，乡村合起来即可得到全国人口增长的差分方程模型。

（1）首先，在不考虑人口迁移的情况下（以人口的户籍变动为准），以乡村为例，建立人口增长模型。

记人口最大年龄为 m 岁（由所给数据取 m 为 90）。 $x_i^1(t)$ 为乡村第 t 年 i 岁的人口数（用上标 1 表示乡村，上标 2 表示镇，上标 3 表示市）， $d_i^1(t)$ 为乡村第 t 年 i 岁人口的死亡率，由人口死亡率数据为所占该类年龄段人口的千分比（‰），我们可以得到乡村第 t 年 i 岁人口的死亡率为：

$$d_i^1(t) = \frac{x_i^1(t) - x_{i+1}^1(t+1)}{x_i^1(t)} \times 1000\% \quad i=0,1,2,\dots,m-1; \quad t=0,1,2,\dots$$

于是第 $(t+1)$ 年 $(i+1)$ 岁人口数为：

$$x_{i+1}^1(t+1) = (1 - d_i^1(t))x_i^1(t) \quad (1)$$

记 $b_i^1(t)$ 为乡村第 t 年 i 岁女性生育率，该生育率表示所有 i 岁女性平均生育婴儿数，生育率为该年龄妇女生育子女与该类年龄妇女的千分比， $[i_1, i_2]$ 为育龄区间（根据题目附录二数据取 15 到 49 岁为育龄区间）， $k_i^1(t)$ 为乡村第 t 年 i 岁人口的女性比率，则乡村第 t 年的出生人口数 $f^1(t)$ 为

$$f^1(t) = \sum_{i=i_1}^{i_2} b_i^1(t)k_i^1(t)x_i^1(t) \quad (2)$$

考虑到婴儿并不是全部都能活到 t 年统计时刻（如 2005 年统计时刻为 11 月 1 日零时零分）其中有些婴儿由于疾病等原因死亡，能够活到 t 年统计时刻的婴儿数是 $x_0(t)$ ，因此， $f^1(t) - x_0(t)$ 就是 $t-1$ 年到 t 年的婴儿死亡数^{[2][P54-59]}，记婴儿死亡率为 $c^1(t)$ ，

$$\text{则：} \quad c^1(t) = \frac{f^1(t) - x_0(t)}{f^1(t)} \times 1000\%$$

$$\text{于是第 } t \text{ 年的婴儿数} \quad x_0^1(t) = (1 - c^1(t))f^1(t) \quad (3)$$

由（1）、（2）、（3）式可得在 $(t+1)$ 年的一岁人口数：

$$x_1^1(t+1) = (1 - c^1(t))(1 - d_0^1(t)) \sum_{i=i_1}^{i_2} (b_i^1(t)k_i^1(t)x_i^1(t)) \quad (4)$$

将 $b_i^1(t)$ 分解为

$$b_i^1(t) = \beta^1(t) h_i^1(t) \quad (5)$$

其中： $h_i^1(t)$ 是生育模式，用以调整育龄妇女在不同年龄时的生育率高低，且：

$$\sum_{i=i_1}^{i_2} h_i^1(t) = 1;$$

$\beta^1(t)$ 是乡村妇女的总和生育率，

$$\beta^1(t) = \sum_{i=i_1}^{i_2} b_i^1(t)。$$

记 $b_i^1(t)^* = (1 - c^1(t))(1 - d_0^1(t))h_i^1(t)k_i^1(t)$ 则由 (4)、(5) 式可得

$$x_i^1(t+1) = \beta^1(t) \sum_{i=i_1}^{i_2} b_i^1(t)^* x_i^1(t) \quad (6)$$

于是我们可以得到乡村人口差分方程模型为：

$$\begin{cases} x_0^1(t) = (1 - c^1(t))f^1(t) \\ x_i^1(t+1) = \beta^1(t) \sum_{i=i_1}^{i_2} b_i^1(t)^* x_i^1(t) \\ x_{i+1}^1(t+1) = (1 - d_i^1(t))x_i^1(t) \end{cases} \quad (i = 1, 2, \dots, m-1)$$

将其表示为矩阵形式，记

$$A^1(t) = \begin{pmatrix} 0 & 0 & \cdots & \beta^1(t)b_{i_1}^1(t)^* & \cdots & \beta^1(t)b_{i_2}^1(t)^* & \cdots & 0 & 0 \\ 1-d_1^1(t) & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & 0 \\ 0 & 1-d_2^1(t) & \cdots & 0 & \cdots & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 1-d_{m-1}^1(t) & 0 \end{pmatrix}_{m \times m}$$

此处， $A^1(t)$ 即为 LESLIE 矩阵；

同时，记

$$X^1(t) = (x_1^1(t) \ x_2^1(t) \ x_3^1(t) \ \cdots \ x_m^1(t))$$

则由上述方程得：

$$X^1(t+1) = X^1(t)A^1(t)$$

运用同样方法可分别得出镇与城的人口差分方程模型为：

$$X^2(t+1) = X^2(t)A^2(t)$$

$$X^3(t+1) = X^3(t)A^3(t)$$

其中 $X^2(t)$ ， $X^3(t)$ ， $A^2(t)$ ， $A^3(t)$ 与 $X^1(t)$ ， $A^1(t)$ 的推导方法相同。

(2) 由于市，镇，乡村之间并不是相互独立的，他们之间有着频繁的人口流动，在实际问题中不能被忽视，下面我们在考虑人口迁移的情况下对上述模型进行改进。

考虑到在实际发生的人口迁移中多数由于贫富差距引起^[3]，我们在对模型进行改进时仅考虑乡村、镇、市的人口净迁移人口量，可以看到镇、市人口净迁移量都为正。我们假设每年乡村到城镇的人口迁移数为上年人口总数的 α 倍，注意到一个地区人口数量与经济发达程度有很大联系^[3]，我们以市，镇总人口的比例来分配乡村到其人口迁移的

数量。由于我们以每年的各个年龄段为预测变量，必须考虑各个年龄段的迁移数量，为简化起见我们以 2001 至 2005 年各个年龄段五年人口数量之和占五年总人口数量的比例分配各个年龄段的迁移数量，并记 s_i 为 i 年龄所占比例。

记 $N^1(t)$, $N^2(t)$, $N^3(t)$ 分别为乡村，镇，市在第 t 年的总人口数，显然满足：

$$N^1(t) = \sum_{i=0}^m x_i^1(t), \quad N^2(t) = \sum_{i=0}^m x_i^2(t), \quad N^3(t) = \sum_{i=0}^m x_i^3(t)$$

由以上分析，第 $t+1$ 年乡村 1 岁人口数量 $x_1^1(t+1)$ 为：

$$x_1^1(t+1) = \beta^1(t) \sum_{i=i_1}^{i_2} b_i^1(t) * x_i^1(t) - s_0 \alpha N^1(t)$$

$(i+1)$ 岁人口的数量 $x_{i+1}^1(t+1)$ 应为：

$$x_{i+1}^1(t+1) = (1 - d_i^1(t)) x_i^1(t) - s_i \alpha N^1(t) \quad (i = 1, 2, 3, \dots, m-1)$$

第 $(t+1)$ 年镇的 1 岁人口数量 $x_1^2(t+1)$ 为：

$$x_1^2(t+1) = \beta^2(t) \sum_{i=i_1}^{i_2} b_i^2(t) * x_i^2(t) + \left(\frac{N^2(t)}{N^3(t) + N^2(t)} \right) s_0 \alpha N^1(t)$$

$(i+1)$ 岁人口的数量 $x_{i+1}^2(t+1)$ 应为：

$$x_{i+1}^2(t+1) = (1 - d_i^2(t)) x_i^2(t) + \left(\frac{N^2(t)}{N^3(t) + N^2(t)} \right) s_i \alpha N^1(t) \quad (i = 1, 2, 3, \dots, m-1)$$

第 $(t+1)$ 年市的 1 岁人口数量 $x_1^3(t+1)$ 应为：

$$x_1^3(t+1) = \beta^3(t) \sum_{i=i_1}^{i_2} b_i^3(t) * x_i^3(t) + \left(\frac{N^3(t)}{N^3(t) + N^2(t)} \right) s_0 \alpha N^1(t)$$

$(i+1)$ 岁的人口数量应为：

$$x_{i+1}^3(t+1) = (1 - d_i^3(t)) x_i^3(t) + \left(\frac{N^3(t)}{N^3(t) + N^2(t)} \right) s_i \alpha N^1(t) \quad (i = 1, 2, 3, \dots, m-1)$$

于是我们可以得到在考虑人口迁移的情况下市，镇，乡村的差分方程模型为：

$$\begin{cases}
x_0^1(t) = (1 - d_b^1(t))f_1(t) \\
x_0^2(t) = (1 - d_b^2(t))f_2(t) \\
x_0^3(t) = (1 - d_b^3(t))f_3(t) \\
x_1^1(t+1) = \beta^1(t) \sum_{i=i_1}^{i_2} b_i^1(t) * x_i^1(t) - s_0 \alpha N^1(t) \\
x_1^2(t+1) = \beta^2(t) \sum_{i=i_1}^{i_2} b_i^2(t) * x_i^2(t) + \left(\frac{N^2(t)}{N^3(t) + N^2(t)} \right) s_0 \alpha N^1(t) \\
x_1^3(t+1) = \beta^3(t) \sum_{i=i_1}^{i_2} b_i^3(t) * x_i^3(t) + \left(\frac{N^3(t)}{N^3(t) + N^2(t)} \right) s_0 \alpha N^1(t) \\
x_{i+1}^1(t+1) = (1 - d_i^1(t))x_i^1(t) - s_i \alpha N^1(t) \\
x_{i+1}^2(t+1) = (1 - d_i^2(t))x_i^2(t) + \left(\frac{N^2(t)}{N^3(t) + N^2(t)} \right) s_i \alpha N^1(t) \\
x_{i+1}^3(t+1) = (1 - d_i^3(t))x_i^3(t) + \left(\frac{N^3(t)}{N^3(t) + N^2(t)} \right) s_i \alpha N^1(t)
\end{cases} \quad (i=1,2,3,\dots,m-1)$$

(3) 从理论上, 男女出生比例应该为 1: 1, 但是根据题目附录二材料, 我国男女出生比例的实际情况是有差异的, 生男孩比例普遍比生女孩高。我们应该对每一个孕龄妇女所生子女的性别加以区分, 而不能笼统地概括为 1: 1, 这样可以更加精确地得到每个年龄男女比例的预测值。同时, 由于城镇化进程中, 乡村人口的迁移带来了城镇人口男女比例的变化。在这个过程中, 我们也应该着重对性别因素进行考虑。以下, 是对前面模型的修正, 以期使模型对性别比例变化产生足够的灵敏度。

出生人口中女性所占比例, 对人口的增长起着至关重要的作用^[6], 有必要对其数量进行单独预测。为此, 我们在 (2) 模型的基础上分别设乡村第 t 年 i 龄女性人口数量为 $w_i^1(t)$, 男性人口数量为 $m_i^1(t)$, 镇第 t 年 i 龄女性人口数量为 $w_i^2(t)$, 男性人口数量为 $m_i^2(t)$, 市第 t 年 i 龄女性人口数量为 $w_i^3(t)$, 男性人口数量为 $m_i^3(t)$, 在迁移人口数量中女性所占比例为 r 。

根据所给乡村, 镇, 市男女出生比例数据, 进行数据的预处理, 得到乡村, 镇, 市出生人口中女性所占比例, 记为 z^1, z^2, z^3 。

那么乡村, 镇, 市女性第 t 年 0 岁人口数量为:

$$\begin{aligned}
w_0^1(t) &= z^1(1 - c^1(t))f^1(t) \\
w_0^2(t) &= z^2(1 - c^2(t))f^2(t) \\
w_0^3(t) &= z^3(1 - c^3(t))f^3(t)
\end{aligned}$$

相应的男性第 t 年 0 岁人口数量为:

$$\begin{aligned}
m_0^1(t) &= (1 - z^1)(1 - c^1(t))f^1(t) \\
m_0^2(t) &= (1 - z^2)(1 - c^2(t))f^2(t)
\end{aligned}$$

$$m_0^3(t) = (1 - z^3)(1 - c^3(t))f^3(t)$$

第 $t+1$ 年 1 到 m 龄女性以及男性人口的预测只须分别用 $wd_i^k(t)$, $md_i^k(t)$ 替代 (2) 模型中的 $d_i^k(t)$ ($k=1,2,3$), 并在迁入或迁出的人数前分别乘以迁移率 r 或 $1-r$, 以乡村女性第 $t+1$ 年 $i+1$ 龄人口数预测为例, 有:

$$w_{i+1}^1(t+1) = (1 - wd_i^1(t))x_i^1(t) - rs_i\alpha N^1(t)$$

其他预测以此类推。

5.2 模型的分析与求解

5.2.1 模型参数的设定

我们根据所建立模型, 需要以下几个输入量:

总和生育率 $\beta^k(t)$, 婴儿死亡率 $c^k(t)$, 生育模型 $h_i^k(t)$, 每年乡村到城镇的人口迁移数为上年人口总数的 α 倍数, 乡村到城镇的人口迁移中女性所占比例 r , 各年龄人口的死亡率, 出生人口中女性所占比例 $z^k(t)$, 总人口数 $N^k(t)$;

根据《中国统计年鉴 2006》^[4], 我们得到 1987 年到 2005 年中每年的总人口数。可以作为模型预测的初值。

在题目附录二中, 给出了市、镇、乡男女出生比例。通过对其时间序列分析, 得出在很短时间内, $z^k(t)$ 都稳定到某一值:

| $z^1(t)$ | $z^2(t)$ | $z^3(t)$ |
|-------------|-------------|-------------|
| 0.455273202 | 0.458859428 | 0.472078522 |

本论文只关注从 2005 起, 未来 45 年我国人口的发展趋势。由于社会环境稳定, 可以简单地认为未来 45 年中, 各年龄人口的死亡率保持一定。

国家政策可以影响^[5]乡村到城镇的人口迁移中女性所占比例 r 及迁移总量的倍率 α 。在以下求解中, 我们可以分析不同 r 、 β 值下人口增长曲线的变化。

同时, 国家计划生育政策也将影响总和生育率 $\beta^k(t)$, 从题目附录一中可知, 为了保持社会经济的稳定发展, 应该把总和生育率控制在 1.8 左右。由于乡村与城镇总和生育率的不同, 我们将各生育率定为如下:

| $\beta^1(t)$ | $\beta^2(t)$ | $\beta^3(t)$ |
|--------------|--------------|--------------|
| 2.0 | 1.8 | 1.6 |

根据《2006 中国卫生统计年鉴》上 1991 年至 2004 年婴儿死亡率 $c^k(t)$ 的变化^[7], 即《中国儿童发展纲要》(2001—2010 年) 提出的每 10 年降低 20 个百分点^[8]。经过数据处理, 得出, 2050 年婴儿死亡率

| $c^1(2050)$ | $c^2(2050)$ | $c^3(2050)$ |
|-------------|-------------|-------------|
| 10.4 | 8.8 | 7.5 |

则可以得出, 在未来 45 年里, 乡、镇、市总和生育率为 17.4, 12.9, 8.8。

我们首先对模型（3）运用 Matlab 编程对其参数进行分析，由上所定参数，并且 r 取 0.5。此时仅有 α 未知，为分析其对总人口数变化规律的影响，分别令其取 0.01, 0.001, 0.0001。以 2005 年数据为预测起始数据得到图 1：

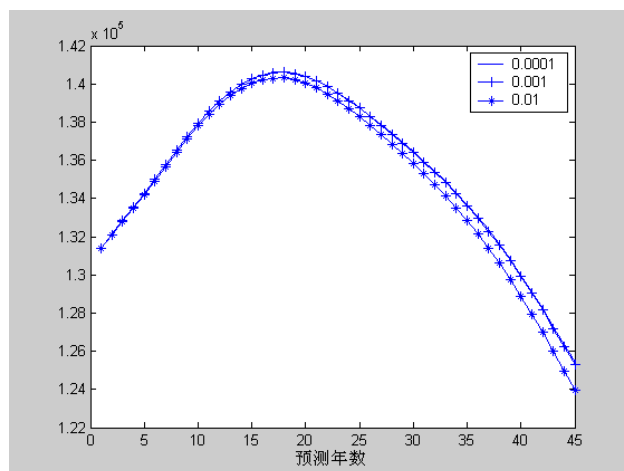


图 1 未来 45 年总人口预测

从图中可以看出 α 对总人口数的影响很小，在前期 1 到 15 年几乎没有影响。那么可以认为 $\alpha = 0.01$ 。

对于 r ，在 α 确定的情况下分别令其取 0.2, 0.5, 0.8 得到图 2：

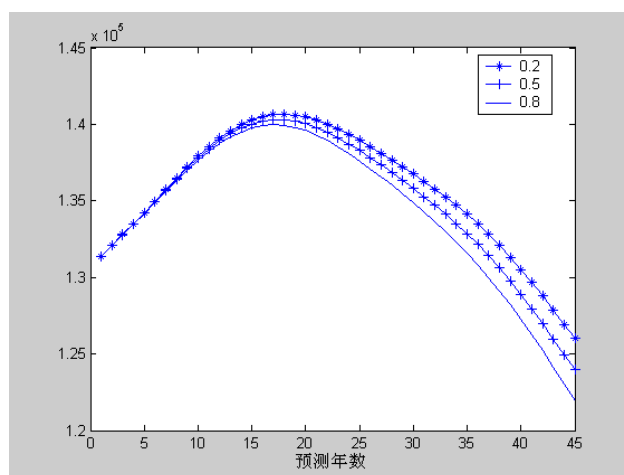


图 2 未来 45 年总人口预测

从图中可以看出 r 对人口高峰的到来时间有一定影响，但是影响不够显著。所以可以简化认为 $r = 0.5$ 。

以上述给定参数，对我国总人口数从 2005 年预测到 2050 年的变化图 3（其中，）：

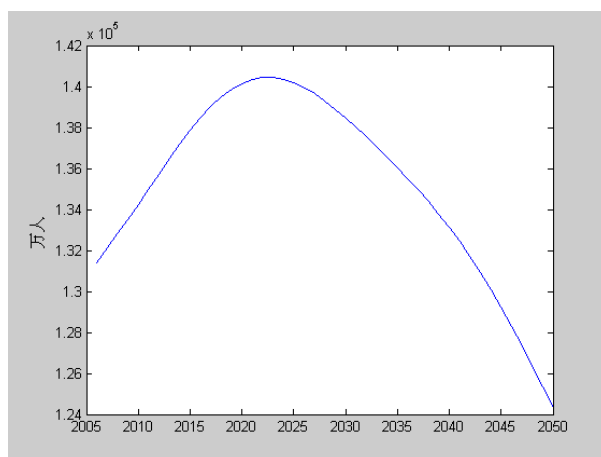


图 3 未来 45 年总人口预测

从预测数据可以得出总人口在 2023 年达到峰值 14.045 亿。

市，镇，乡村以及总人口的男女比例从 2005 年到 2050 年的变化图 4：

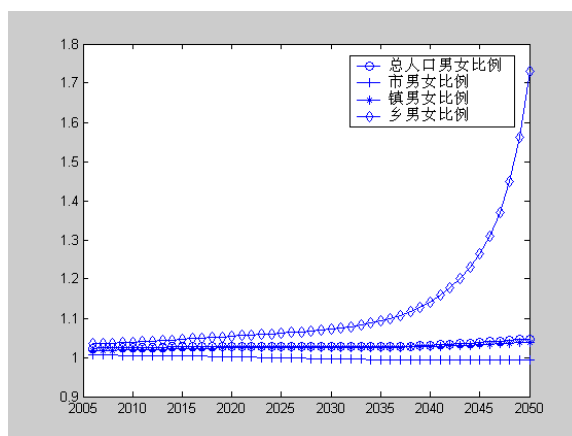


图 4 市，镇，乡村以及总人口的男女比例

从图中可以看出总人口，市，镇男女比例一直保持在 1 附近，只有乡村男女比例在 2035 年后呈指数形增长，但在 2050 年仍小于 1.8。

总人口的抚养比从 2005 年到 2050 年的变化图 5：

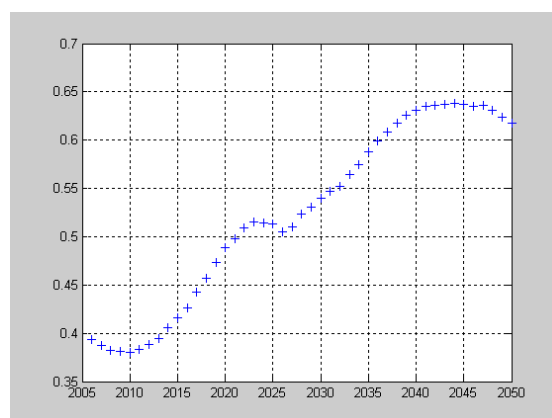


图 5 总人口的抚养比

从图中可以看出 2005 至 2010 总人口的抚养比一直减小，2010 年达最小值：0.38. 2010 年以后抚养比一直增加，2043 年左右达最大值：0.64。

人口城镇化比例从 2005 到 2050 年的变化图如 6：

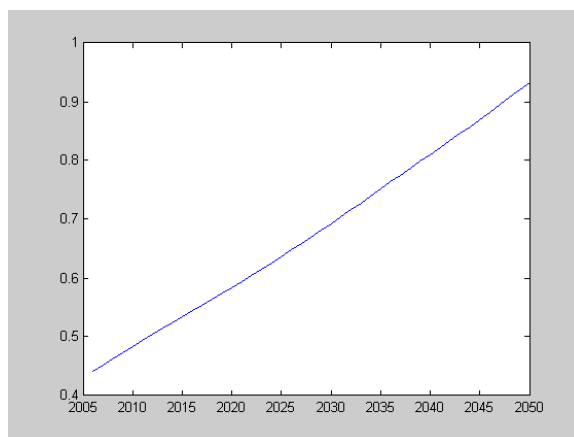


图 6 人口城镇化比例

总人口考虑 60 岁以上人口的老龄化进程从 2005 年到 2050 如图 7:

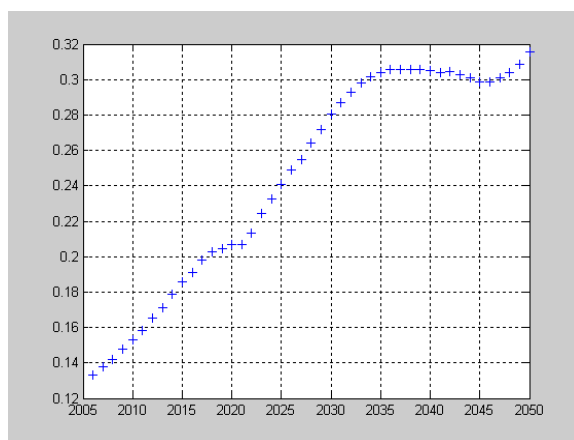


图 7 60 岁以上人口的老龄化

从图中可以看出我国在 2035 年到 2045 年形成老龄人口高峰平台，60 岁以上老年人口比重在 0.3 以上。

总人口中育龄妇女人数从 2005 年到 2050 年变化如图 8:

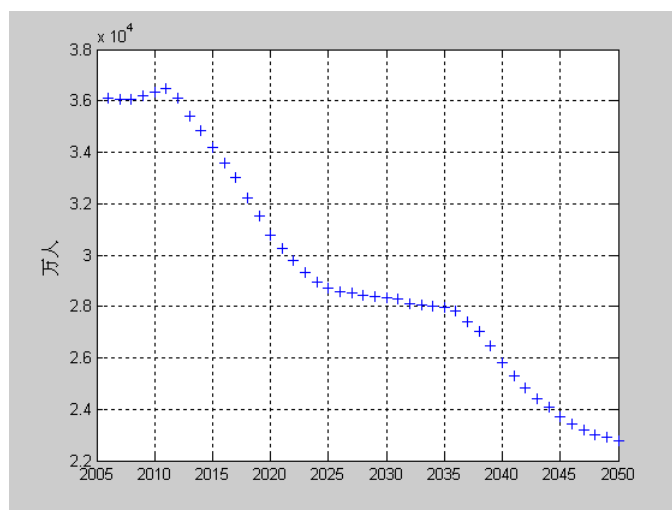


图 8 总人口中育龄妇女人数

从图中可以看出，在 2010 年后育龄妇女人数一直下降，2045 年降到 2.4 亿以下。

5.3 对差分方程的修正

影响人口增长因素很多，如农村人口向城镇的迁移、婴儿死亡率、总和生育率等。在对长时间的预测中，我们只能假定一些因素是不变的。但是，我们应该清楚地看到，随着时间的推移、大环境的变化必然带来因素的改变。为了解决这个问题，我们需要不断地对模型中一些参数进行修正。在这里，我们采用时间段参数设置法^[9]，将时间区间设置等间距的 15 年。在区间内，我们认为参数是不变的。在一个区间数据预测完成后，我们在上一区间预测数据的基础上修正因素，进行下一区间预测，以达到最大限度补充信息量，保证预测效果的显著。

下面仅在 2020 年及 2035 年修正婴儿死亡率，得到以下预测见表 1（人口单位/10 亿）

表 1 差分方程修正

| 年份 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 |
|------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 一般预测 | 1.3137 | 1.3209 | 1.3279 | 1.3348 | 1.3419 | 1.3491 | 1.3566 | 1.3641 | 1.3713 | 1.378 |
| 修正预测 | 1.3137 | 1.3209 | 1.3279 | 1.3348 | 1.3419 | 1.3491 | 1.3566 | 1.3641 | 1.3713 | 1.378 |
| 年份 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 | 2024 | 2025 |
| 一般预测 | 1.384 | 1.3894 | 1.3938 | 1.3974 | 1.4003 | 1.402 | 1.4031 | 1.4031 | 1.4021 | 1.4004 |
| 修正预测 | 1.384 | 1.3894 | 1.3938 | 1.3974 | 1.4003 | 1.4022 | 1.4033 | 1.4034 | 1.4025 | 1.4008 |
| 年份 | 2026 | 2027 | 2028 | 2029 | 2030 | 2031 | 2032 | 2033 | 2034 | 2035 |
| 一般预测 | 1.3978 | 1.3947 | 1.3911 | 1.3871 | 1.3828 | 1.3781 | 1.3733 | 1.3684 | 1.3634 | 1.3582 |
| 修正预测 | 1.3983 | 1.3953 | 1.3917 | 1.3877 | 1.3835 | 1.3788 | 1.3741 | 1.3692 | 1.3642 | 1.3591 |
| 年份 | 2036 | 2037 | 2038 | 2039 | 2040 | 2041 | 2042 | 2043 | 2044 | 2045 |
| 一般预测 | 1.3528 | 1.3472 | 1.3413 | 1.3352 | 1.3285 | 1.3215 | 1.3142 | 1.3061 | 1.2977 | 1.2888 |
| 修正预测 | 1.3538 | 1.3484 | 1.3426 | 1.3366 | 1.3301 | 1.3232 | 1.316 | 1.3081 | 1.2999 | 1.2911 |
| 年份 | 2046 | 2047 | 2048 | 2049 | 2050 | | | | | |
| 一般预测 | 1.2795 | 1.27 | 1.2599 | 1.2498 | 1.2399 | | | | | |
| 修正预测 | 1.2819 | 1.2726 | 1.2626 | 1.2526 | 1.2428 | | | | | |

从表中可以看出，参数设置法对人口增长产生了一定影响。如果把各个影响人口增长的因素加到对分段参数设置法修正中，会得到更普配实际的预测。但由于本论文成文仓促，本修正方法没有做完，是此论文的一个不足之处。

六、模型的比较与检验

为了对差分方程模型进行检验与比较，我们将模型的时间节点朝前移五年，用 2001 年数据为输入数据，得出 2002 年到 2005 年的预测数据，经过同实际统计数据的对比，我们可以看到差分方程模型的优点与不足。

另外，对模型进行检验，我们可以比照《国家人口发展战略研究报告》(2007)的相关数据指标。但是为了更加独立地校验模型，我们另建立了灰色预测模型，以期看到差分方程模型与灰色预测结果的匹配程度。

6.1 以普查统计数据检验差分方程模型

带入 2001 年数据，由差分方程模型（3）得出预测结果见表 2。

表 2

| | 2002 年 | 2003 年 | 2004 年 | 2005 年 |
|----------|----------|----------|----------|----------|
| 实际值（万人） | 128453 | 129227 | 129988 | 130756 |
| 预测值（万人） | 128000 | 128980 | 129890 | 130760 |
| 绝对误差（万人） | 453 | 247 | 98 | -4 |
| 相对误差 | 0.003527 | 0.001911 | 0.000754 | 0.000031 |

通过实际人口量与预测人口数量的比较，我们可以看到，差分方程模型在总人口的预测上与实际情况符合得很好。

6.2 灰色预测模型的建立与求解

6.2.1 灰色数列预测原理

灰色数列预测最大的特点是单数据序列的预测。在形式上，只运用预测对象自身的时间序列，而与预测对象相关联的其他因素没有参与运算和建模，灰色系统理论把影响客观系统的诸多因素及它们之间的关系定为灰色量。对这样的灰色量进行预测，就可从自身的时间序列中寻找有用信息，发现内在规律，建立模型进行预测。

6.2.2 建立 $GM(1,1)$ 模型

(1) 数据的检验与处理

对于给定的原始时间1996~2005年数据列：

$$\begin{aligned}
 x^{(0)} &= (x^{(0)}(1), x^{(0)}(2), x^{(0)}(3), x^{(0)}(4), x^{(0)}(5), x^{(0)}(6), x^{(0)}(7), x^{(0)}(8), x^{(0)}(9), x^{(0)}(10)) \\
 &= (122389, 123626, 124761, 125786, 126743, 127627, 1284 \\
 &\text{数列的级比}
 \end{aligned}$$

$$\begin{aligned}
 \lambda(k) &= \frac{x^{(0)}(k-1)}{x^{(0)}(k)} \cdots \cdots (k=2, 3, \cdots, n) \\
 &= (0.9900, 0.9909, 0.9919, 0.9924, 0.9931, 0.9936, 0.9940, 0.9941, 0.9941)
 \end{aligned}$$

如果所有的级比 $\lambda(k)$ 都落在可容覆盖 $X = \left(e^{-\frac{2}{n+1}}, e^{\frac{2}{n+1}} \right)$ 内，则数列 $x^{(0)}$ 可以作为

模型 $GM(1,1)$ 和进行数据灰色预测。

经过计算， $X = (0.8338, 1.1994)$ ，所以 $\lambda(k)$ 全部落入 X 中。数列 $x^{(0)}$ 可以作为模型 $GM(1,1)$ 和进行数据灰色预测。

(2) 确定数据序列

对 $x^{(0)}$ 做AGO生成，有 $x^{(1)} = \text{AGO } x^{(0)}$ ，其中 $x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i) \quad (k=1, 2, \cdots, n)$ ，

$$\begin{aligned}
 \text{则} \quad x^{(1)} &= (x^{(1)}(1), x^{(1)}(2), x^{(1)}(3), \cdots, x^{(1)}(n)) \\
 &= (x^{(0)}(1), x^{(0)}(1) + x^{(0)}(2), \cdots, x^{(0)}(n-1) + x^{(0)}(n)) \\
 &= (122389, 246015, 370776, 496562, 623305, 750932, 87938, \\
 &\quad 1008612, 1138600, 1269356)
 \end{aligned}$$

求均值数列

$$\begin{aligned} z^{(1)}(k) &= 0.5 \times x^{(1)}(k) + 0.5 \times x^{(1)}(k-1) \quad (k=2, \dots, n) \\ &= (184202, 308395.5, 433669, 559933.5, 687118.5, 815158.5, 943998.5, \\ &\quad 1073606, 1203978) \end{aligned}$$

于是建立 $GM(1,1)$ 模型,

灰微分方程为^[10]:

$$x^{(0)}(k) + ax^{(1)}(k) = b \quad (k=2, 3, \dots, n)$$

相应的白化微分方程为:

$$\frac{dx^{(1)}(t)}{dt} + a \times x^{(1)}(t) = b, \quad (1)$$

$$\text{记 } u = (a, b)^T, \quad Y_1 = (x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n))^T, \quad B = \begin{bmatrix} -z^{-1}(2) & 1 \\ -z^{-1}(3) & 1 \\ \vdots & \vdots \\ -z^{-1}(n) & 1 \end{bmatrix},$$

则由最小二乘法, 求使得

$$J(\hat{u}) = (Y_1 - B\hat{u})^T (Y_1 - B\hat{u})$$

达到最小值的 $\hat{u} = \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix}^T = (B^T B)^{-1} B^T Y_1$ 。

于是求解方程得

$$x^{(1)}(k+1) = \left(x^{(0)}(1) - \frac{b}{a} \right) e^{-ak} + \frac{b}{a} \quad (k=2, \dots, n-1)$$

$$a = -0.0069$$

$$b = 122680$$

由上可以得到白化方程的时间相应式

$$\hat{x}^{(1)}(k+1) = \left(x^{(0)}(1) - \frac{b}{a} \right) e^{-ak} + \frac{b}{a} = 17912000e^{0.0069k} - 17790000$$

$$\text{而且} \quad \hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k)$$

$$\text{得还原方程: } \hat{x}^{(0)}(t+1) = 1.2402 \times 10^7 e^{0.0069k} \quad (k=2, \dots, n-1)$$

(3)对灰色预测模型的改进

一般情况下, $GM(1,1)$ 模型通过对数据序列长度的不同取舍, 得到不同的预测结果,

而组成一个预测区间，可供决策选用。但是未来时刻越远， $GM(1,1)$ 模型预测的灰区间越大。可见预测精度的提高在于不断补充新的信息，使灰度逐步白化。所以，用 $GM(1,1)$ 模型预测时不是建立一个模型一直预测下去，而只用已知数列建立的 $GM(1,1)$ 模型预测的第一个预测值，补充在已知数列之后，同时为不增加数据序列的长度（维度），去掉其第一个已知数据，保持数据序列的等长度，再建立 $GM(1,1)$ 模型，预测下一个值。将预测值再补充到数据序列之后，再去掉该数据序列的第一个数据。这样新陈代谢，逐个预测，依次递补，直到完成预测目的或达到一定的精度要求为止。这种预测方法称“等维灰数递补动态预测”^[11]法，即新陈代谢的灰色预测方法。它可以达到两个目的：①及时补充和利用新的信息，提高了灰色区间的白化度。当然随着递补次数的增加，灰度也在增加，信息量减少，因此也不应是无止境的。②每预测一步灰参数做一次修正，模型得到改进。这样灰参数不断修正，模型逐步改进，因而预测值都产生在动态之中。但是，需要注意的是，每次进行新陈代谢后，都需要对模型进行必要的检验，以保证建模的可行性。

6.3 灰色预测的检验

6.3.1 灰色预测数列长度的选择

在应用“等维灰数递补动态预测”方法，我们知道选择不同的数列长度（维度）对预测模型的影响较大。我们在《中国统计年鉴——2006》中查询到1978年至2005年我国人口数及构成^[4]，运用matlab软件编写了“等维灰数递补动态预测”程序（程序代码见附录二）对未来45年的总人口预测。在对总人口数量的预测中，我们采用了三种数列长度。图形表示如图9，（具体数据见附录一）。

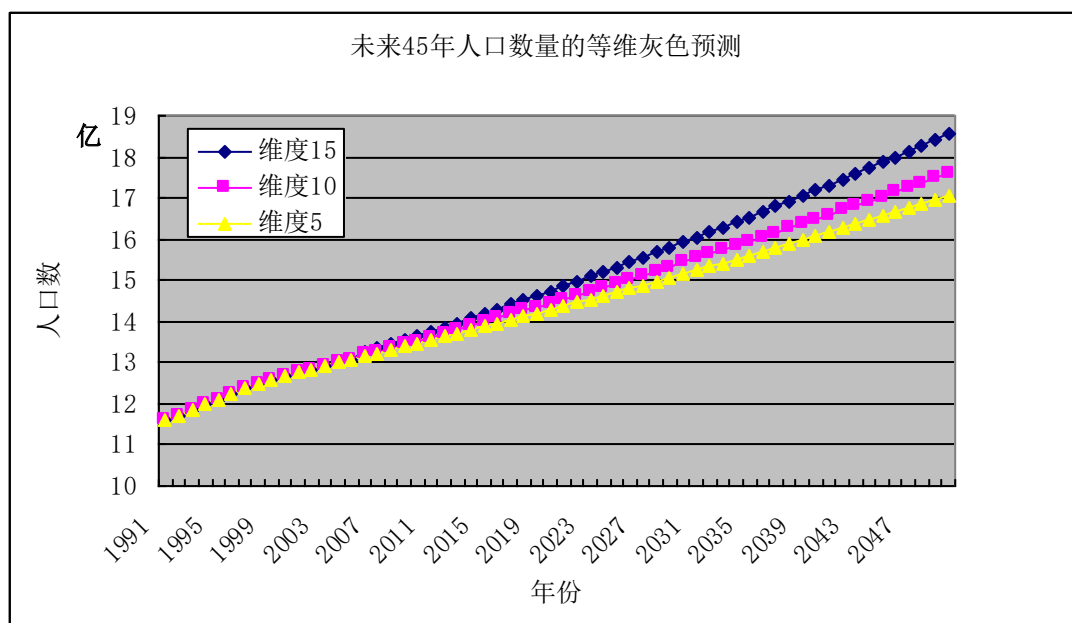


图9 不同维度的灰色预测

运用维度为十的“等维灰数递补动态预测”方法对未来十五年的总人口预测，得到预测结果如下表3：

表3

| 年份 | 2006 | 2007 | 2008 | 2009 | 2010 |
|---------|--------|--------|--------|--------|--------|
| 人口数(十亿) | 1.3153 | 1.3231 | 1.3309 | 1.3388 | 1.3467 |
| 年份 | 2011 | 2012 | 2013 | 2014 | 2015 |
| 人口数(十亿) | 1.3547 | 1.3627 | 1.3708 | 1.3789 | 1.387 |
| 年份 | 2016 | 2017 | 2018 | 2019 | 2020 |
| 人口数(十亿) | 1.3952 | 1.4035 | 1.4118 | 1.4201 | 1.4285 |

从上图及具体的预测数据中，我们看到，在未来15年内数据预测与《国家人口发展战略研究报告》（见题目附录二）的吻合较好。说明等维灰色预测在本论文的人口预测中，比较适合短期预测。

6.3.2 灰色预测与差分方程模型结果的比较

我们用灰色预测数据与前面建立的差分方程模型得出的数据进行比较。走势图如图10。

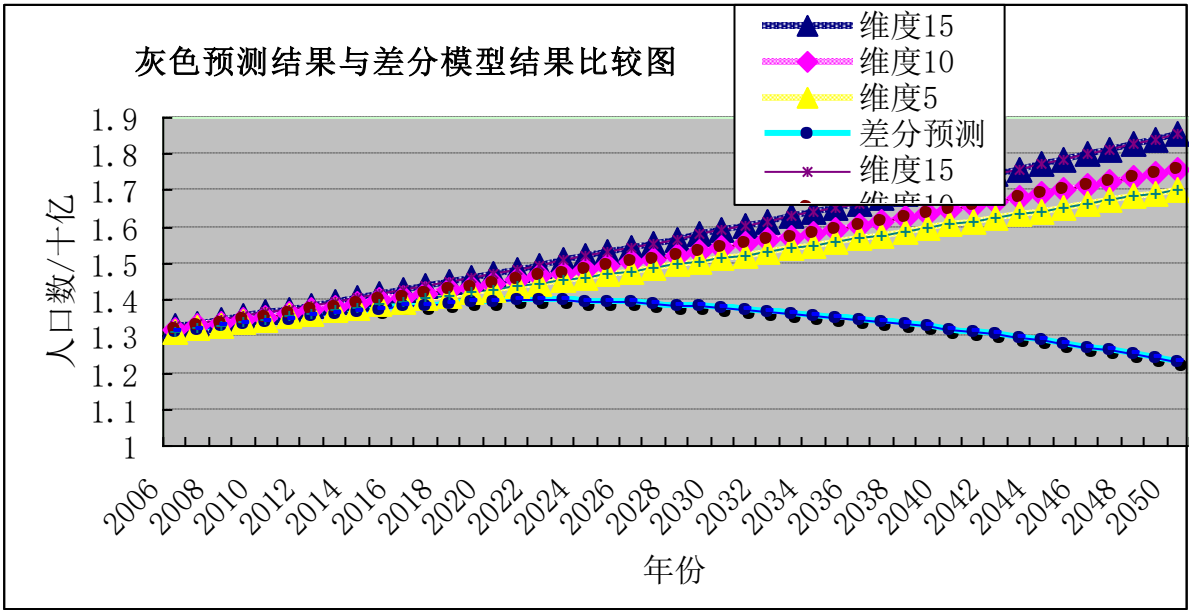


图10 灰色预测结果与差分模型结果比较

从上图，我们看到人口差分模型结论在短期内是比较合理的。
维度为十的灰色预测结果与差分模型结果的比较如表4（人数单位为十亿）：

表4

| 年份 | 2006 | 2007 | 2008 | 2009 | 2010 |
|------|--------|--------|--------|--------|--------|
| 灰色预测 | 1.3153 | 1.3231 | 1.3309 | 1.3388 | 1.3467 |
| 差分模型 | 1.3135 | 1.3206 | 1.3274 | 1.3341 | 1.341 |
| 年份 | 2011 | 2012 | 2013 | 2014 | 2015 |
| 灰色预测 | 1.3547 | 1.3627 | 1.3708 | 1.3789 | 1.387 |

| | | | | | |
|------|--------|--------|--------|--------|--------|
| 差分模型 | 1.348 | 1.3552 | 1.3624 | 1.3694 | 1.3758 |
| 年份 | 2016 | 2017 | 2018 | 2019 | 2020 |
| 灰色预测 | 1.3952 | 1.4035 | 1.4118 | 1.4201 | 1.4285 |
| 差分模型 | 1.3815 | 1.3866 | 1.3908 | 1.3941 | 1.3967 |

七、 模型的评价与改进

本论文通过对差分方程模型的三次改进,着重考虑城镇与乡村间的人口迁移及女性人口比例变化对我国人口增长的影响,并在中长期预测中参考了时间段参数设置法,为减小长期预测的误差带来了一种新的方法。所以本模型较多地从我国实际情况出发,深刻地描述了人口在新时期的发展特点,能较好地反映我国的实际情况。通过以上对模型的比较与检验可以看出,模型在中短期预测中有较高的准确率。在参数设置中,对某些参数进行了灵敏度分析,得出对人口增长的不同影响程度。

但是我们在整个建模过程中没有考虑到文化背景和人口素质的差异会对人口增长以及人口的迁移产生影响,如在乡村由于重男轻女、多子多福等落后思想的存在会导致人口性别比的失调和人口增长快等。没有考虑到可能会由于欧洲等国家老龄化现象普遍,我国生产力水平提高致使劳动力富余而大量向国外输出劳动力而引起的人口增长变化等因素,所以预测结果精度还有待进一步提高。对于这些不足,我们可以采用设置参数来反映这些因素的作用来加以改进。

参考文献:

- [1] 王彦 马伯强,《20世纪80年代以来我国人口发展的数学模型和展望》,北京大学学报(自然科学版),第39 卷 增刊; 29—30页, 2003。
- [2] 宋健等,《人口预测和人口控制》,北京:人民出版社,1981。
- [3] 卢向虎、王永刚,中国“乡—城”人口迁移规模的测算与分析(1979- 2003),《西北人口》,第1 期 No.1(107)2006: 14-16页, 2006。
- [4] 中华人民共和国国家统计局,《中国统计年鉴——2006》,北京:中国统计出版社,2007。
- [5] 中国城市发展报告编辑委员会,《2006 中国城市发展报告》,北京:年鉴社,2007。
- [6] Fred Brauer,Carlos Castillo-Chavez,《mathematical models in population biology and epidemiology》,Springer, New York : 2001。
- [7] 国家统计局编辑部,《2006中国卫生统计年鉴》,北京:国家统计局出版社,2007。
- [8] 国务院新闻办公室,《中国儿童发展纲要(2001—2010年)》,北京:中国法制出版社,2001。
- [9] 李永胜,人口预测中的模型选择与参数认定,《财经科学》,总 203 期: 68-72 页,2004。
- [10] 韩中庚,《数学建模方法及应用》,北京:高等教育出版社,2006。
- [11]王海应 张昌维 , 利用灰色理论预测油田开发指标,《新疆石油学院学报》,第 16 卷第 3 期: 27~28 页, 2004。

附录

附录一：灰色预测 45 年的数据

| | 维度 15 | 维度 10 | 维度 5 | 维度 15 | 维度 10 | 维度 5 |
|------|--------|--------|--------|-------|--------|--------|
| 2006 | 1.3269 | 1.3189 | 1.3153 | 2029 | 1.5795 | 1.5326 |
| 2007 | 1.3365 | 1.327 | 1.3231 | 2030 | 1.5916 | 1.5427 |
| 2008 | 1.346 | 1.3353 | 1.3309 | 2031 | 1.6038 | 1.5528 |
| 2009 | 1.3556 | 1.3439 | 1.3388 | 2032 | 1.6162 | 1.5631 |
| 2010 | 1.3653 | 1.3526 | 1.3467 | 2033 | 1.6288 | 1.5733 |
| 2011 | 1.3753 | 1.3617 | 1.3547 | 2034 | 1.6414 | 1.5837 |
| 2012 | 1.3856 | 1.3708 | 1.3627 | 2035 | 1.654 | 1.5941 |
| 2013 | 1.3962 | 1.38 | 1.3708 | 2036 | 1.6667 | 1.6046 |
| 2014 | 1.4072 | 1.3891 | 1.3789 | 2037 | 1.6795 | 1.6152 |
| 2015 | 1.4184 | 1.3979 | 1.387 | 2038 | 1.6924 | 1.6258 |
| 2016 | 1.4298 | 1.4072 | 1.3952 | 2039 | 1.7054 | 1.6365 |
| 2017 | 1.4412 | 1.4165 | 1.4035 | 2040 | 1.7186 | 1.6472 |
| 2018 | 1.4524 | 1.4259 | 1.4118 | 2041 | 1.7318 | 1.6581 |
| 2019 | 1.4633 | 1.4353 | 1.4201 | 2042 | 1.7452 | 1.669 |
| 2020 | 1.4736 | 1.4447 | 1.4285 | 2043 | 1.7586 | 1.6799 |
| 2021 | 1.485 | 1.4542 | 1.4369 | 2044 | 1.7722 | 1.691 |
| 2022 | 1.4966 | 1.4638 | 1.4454 | 2045 | 1.7858 | 1.7021 |
| 2023 | 1.5084 | 1.4734 | 1.4539 | 2046 | 1.7995 | 1.7133 |
| 2024 | 1.5201 | 1.4832 | 1.4625 | 2047 | 1.8134 | 1.7246 |
| 2025 | 1.5319 | 1.4929 | 1.4711 | 2048 | 1.8273 | 1.7359 |
| 2026 | 1.5438 | 1.5027 | 1.4798 | 2049 | 1.8414 | 1.7473 |
| 2027 | 1.5556 | 1.5126 | 1.4885 | 2050 | 1.8556 | 1.7588 |
| 2028 | 1.5675 | 1.5226 | 1.4973 | | | |

附录二：灰色预测的 MATLAB 程序

```
total=[34743.9554 21468.0445 74544.0000];
load dmat M;load bmat b;z=[1.8,2.1,2.5];r=0.4;s=[0.47 0.45 0.45];yer=[10.1,17.3,24.5];% 婴儿死亡率
yer=[10.1,17.3,24.5];
bili=[0.010519802
.....
0.001443894
];
x1=zeros(3,45,90);women=zeros(3,45,91);men=zeros(3,45,91);%j=1;k=3;
for j=1:45
    q=sum(total)/100;
for k=3:-1:1
    if j==1
    for i=1:90
        if k==3
            women(k,j,i+1)=(M(i,k+3*(k-1)+2)/100)*total(k)*(1-(M(i,k+3*(k-1)+3)/1000))-r*q*bili(i);
```

```

    men(k,j,i+1)=(M(i,k+3*(k-1))/100)*total(k)*(1-(M(i,k+3*(k-1)+1)/1000))-(1-r)*q*bili(i);%q*f1*bili(1);
elseif k==1
women(k,j,i+1)=(M(i,k+3*(k-1)+2)/100)*total(k)*(1-(M(i,k+3*(k-1)+3)/1000))+r*(total(1)/(total(1)+total(
2)))*q*bili(i);%q*f1*bili(1);
men(k,j,i+1)=(M(i,k+3*(k-1))/100)*total(k)*(1-(M(i,k+3*(k-1)+1)/1000))+(1-r)*(total(1)/(total(1)+total(2
)))*q*bili(i);%q*f1*bili(1)
else
women(k,j,i+1)=(M(i,k+3*(k-1)+2)/100)*total(k)*(1-(M(i,k+3*(k-1)+3)/1000))+r*(total(2)/(total(1)+total(
2)))*q*bili(i);%q*f1*bili(1);
men(k,j,i+1)=(M(i,k+3*(k-1))/100)*total(k)*(1-(M(i,k+3*(k-1)+1)/1000))+(1-r)*(total(2)/(total(1)+total(2
)))*q*bili(i);
end
x1(k,j,i)=women(k,j,i+1)+men(k,j,i+1);
end
else
for i=1:90
if k==3
women(k,j,i+1)=(1-M(i,3+k+3*(k-1))/1000)*women(k,j-1,i)-r*q*bili(i);
men(k,j,i+1)=(1-M(i,1+k+3*(k-1))/1000)*men(k,j-1,i)-(1-r)*q*bili(i);
elseif k==1
    women(k,j,i+1)=(1-M(i,3+k+3*(k-1))/1000)*women(k,j-1,i)+(total(1)/(total(1)+total(2)))*r*q*bili(i);
    men(k,j,i+1)=(1-M(i,1+k+3*(k-1))/1000)*men(k,j-1,i)+(total(1)/(total(1)+total(2)))*(1-r)*q*bili(i);
else
    women(k,j,i+1)=(1-M(i,3+k+3*(k-1))/1000)*women(k,j-1,i)+(total(2)/(total(1)+total(2)))*r*q*bili(i);
    men(k,j,i+1)=(1-M(i,1+k+3*(k-1))/1000)*men(k,j-1,i)+(total(2)/(total(1)+total(2)))*(1-r)*q*bili(i);
end
x1(k,j,i)=women(k,j,i+1)+men(k,j,i+1);
end
end
hi=b(:,k)/sum(b(:,k));
f=0;
for i=1:35
f=f+(1-yer(k)/1000)*z(k)*hi(i)*women(k,j,i+15);%(women(k,j,i+15)/(men(k,j,i+15)+women(k,j,i+15)))*x
1(k,j,i+14);
end
women(k,j,1)=s(k)*f;
men(k,j,1)=(1-s(k))*f;
total(k)=sum(x1(k,j,:))+f;
end
j
total
sum(total)
end

```

