

关于中国人口增长趋势的研究

【摘要】

本文从中国的实际情况和人口增长的特点出发,针对中国未来人口的老龄化、出生人口性别比以及乡村人口城镇化等,提出了 Logistic、灰色预测、动态模拟等方法进行建模预测。

首先,本文建立了 Logistic 阻滞增长模型,在最简单的假设下,依照中国人口的历史数据,运用线形最小二乘法对其进行拟合,对 2007 至 2020 年的人口数目进行了预测,得出在 2015 年时,中国人口有 13.59 亿。在此模型中,由于并没有考虑人口的年龄、出生人数男女比例等因素,只是粗略的进行了预测,所以只对中短期人口做了预测,理论上很好,实用性不强,有一定的局限性。

然后,为了减少人口的出生和死亡这些随机事件对预测的影响,本文建立了 GM(1, 1) 灰色预测模型,对 2007 至 2050 年的人口数目进行了预测,同时还用 1990 至 2005 年的人口数据对模型进行了误差检验,结果表明,此模型的精度较高,适合中长期的预测,得出 2030 年时,中国人口有 14.135 亿。与阻滞增长模型相同,本模型也没有考虑年龄一类的因素,只是做出了人口总数的预测,没有进一步深入。

为了对人口结构、男女比例、人口老龄化等作深入研究,本文利用动态模拟的方法建立模型三,并对数据作了如下处理:取平均消除异常值、对死亡率拟合、求出 2001 年市镇乡男女各年龄人口数目、城镇化水平拟合。在此基础上,预测出人口的峰值,适婚年龄的男女数量的差值,人口老龄化程度,城镇化水平,人口抚养比以及我国“人口红利”时期。在模型求解的过程中,还对政府部门提出了一些有针对性的建议。此模型可以对未来人口做出细致的预测,但是需要处理的数据量较大,并且对初始数据的准确性要求较高。接着,我们对模型三进行了改进,考虑人为因素的作用,加入控制因子,使得所预测的结果更具有实际意义。

在灵敏度分析中,首先针对死亡率发展因子 θ 进行了灵敏度分析,发现人口数量对于 θ 的灵敏度并不高,然后对男女出生比例进行灵敏度分析得出其灵敏度系数为 0.8850,最后对妇女生育率进行了灵敏度分析,发现在生育率在由低到高的变化过程中,其灵敏度在不断增大。

最后,本文对模型进行了评价,特别指出了各个模型的优缺点,同时也对模型进行了合理性分析,针对我国的人口情况给政府提出了建议。

关键字: Logistic 模型 灰色预测 动态模拟 Compertz 函数

目录

一、	问题重述	3
二、	符号定义与说明	3
三、	模型假设	3
四、	问题分析	4
1	问题背景的理解:	4
2	问题分析:	4
五、	模型建立及求解	4
1	模型一 阻滞增长模型	4
1.1	模型建立	4
1.2	模型求解	5
1.3	模型的分析与优化	6
2	模型二 GM(1, 1) 灰色预测模型	7
2.1	模型建立	7
2.2	模型求解	7
2.3	结果分析	8
3	模型三 基于计算机模拟的动态模型	8
3.1	数据处理	8
3.1.1	取平均消除异常值	8
3.1.2	对人口死亡率拟合	9
3.1.3	2001 年市镇乡男女各年龄人口数目	9
3.1.4	城镇化水平拟合	9
3.2	模型的建立及求解	10
3.2.1	模型建立	10
3.2.2	模型求解	11
3.3	模型三的改进	17
六、	模型评价	19
七、	模型灵敏度分析	19
1	人口死亡率 Compertz 函数中死亡率发展因子 θ 的灵敏度分析	19
2	模型三男女出生比例灵敏度分析	20
3	模型三妇女生育率灵敏度分析	21
八、	模型的合理性分析:	21
九、	给政府的建议	22
	【参考文献】	22
	【附件】	23

一、 问题重述

中国是世界上人口最多的国家，人口多，人均占有资源相对不足，是我国的基本国情，人口问题一直是制约中国经济发展的首要因素。因此，计划生育是我国的一项基本国策。近年来中国的人口发展出现了一些新的特点，例如，老龄化进程加速、出生人口性别比持续升高，以及乡村人口城镇化等因素，这些都影响着中国人口的发展。因此，对中国人口做出分析和预测是一个重要问题，这将为中国经济和社会发展决策提供科学依据，同时对于加速推进我国现代化建设的宏伟大业有着极为重要的现实意义。

根据附录2的数据进行建模，同时要参考中国的实际情况以及人口增长的上述特点，对中国人口增长的中短期和长期的趋势做出预测，比如未来的人口数目、性别比例、人口结构等，特别要指出模型的优点和不足。

二、 符号定义与说明

符号	符号定义	单位或备注
$x(t)$	t 年的全国人口总数	亿
$z(t)$	t 年的城镇化比例	(%)
$\mu(t)$	t 年人口死亡率	(‰)
$y(t)$	t 年老龄化比例	(%)
i	$i=1, 2, 3$ 分别表示城、镇、乡	
k	$k=0. 1. 2. \dots 90$ 分表表示人的年龄	
$q_i(t)$	t 年 i 地男女的出生比例	(%)
$w_i(t,k)$	t 年 i 地年龄为 k 的女性数量	万
$m_i(t,k)$	t 年 i 地年龄为 k 的男性数量	万
$b_i(t,k)$	t 年 i 地年龄为 k 的妇女生育率	k 只取 15. 16. 17... 49 (‰)
$b_i(t)$	t 年 i 地各自的总和生育率	(‰)

三、 模型假设

1. 不考虑我国人口向国外搬迁，同时也不考虑国外人口向国内搬迁；
2. 不考虑战争、灾害、疾病对人口数目的影响；
3. 假设在一年内，各个地区，各个年龄段的死亡率不会发生变化；

4. 假设在一年内，处于生育年龄的妇女生育率不会发生变化；
5. 由于当前男女出生比例有失调的现象，故假设在 2020 年之前男女出生比例为 116，在 2020 年之后调整到 107。

四、 问题分析

1 问题背景的理解：

新中国成立 50 多年来，我国人口发展经历了前 30 年高速增长和后 20 年低速增长两大阶段：从建国初期到上世纪 70 年代初，中国人口再生产由旧中国的高出生、高死亡率进入高出生、低死亡率的人口高速增长时期，1950-1975 年人口出生率始终保持在 30% 以上，最高达到 37%。

70 年代以后，人口过快增长的势头得到迅速扭转，人口出生率、自然增长率、妇女总和生育率有了明显下降，人口出生率由 70 年代初的 33% 大幅度下降到 80 年代的 21%，妇女总和生育率也由 6 下降到 2.3 左右。

90 年代以来，随着我国经济高速发展，人民文化和健康水平逐步提高，计划生育工作的不断深入，在 20-29 岁生育旺盛人数年均超过 1 亿的情况下，人口出生率依然呈现大幅下降的趋势，到 2000 年底人口出生率从 1990 年的 21.06% 下降到 14.03%，自然增长率由 1990 年的 14.39% 下降到 7.58%，妇女总和生育率也下降到 2 以下。

进入 90 年代末期，我国人口再生产实现了低出生、低死亡、低增长的历史性转变，我国用 20 多年时间完成了国外近 200 年的历程。到 2000 年底全国总人口为 12.6743 亿，成功实现了“九五”计划将人口控制在 13 亿的奋斗目标。

中国政府自 1980 年在全国城乡实行计划生育基本国策以来成果卓著，据国家计生委“计划生育投入与效益研究”课题组的研究成果，20 年共少生 2.5 亿个孩子。若从 70 年代算起，至今至少少生 3 亿人口，这有效地控制了人口的快速增长，为中国现代化建设、全面实现小康打下坚实的基础，这同时也是对世界人口的增长和控制做出了杰出贡献。但是由于中国人口基数大，人口增长问题依然十分严峻，1990-1999 年每年平均净增人口约 1300 万，这仍然对我国社会和经济产生巨大的压力。在我国现代化进程中，必须实现人口与经济、社会、资源、环境协调发展和可持续发展，进一步控制人口数量，提高人口质量，改善人口结构。

2 问题分析：

本题需要结合中国的实际情况和人口增长的特点来对中国人口增长的中短期和长期趋势做出预测。

首先，我们从简单模型入手，利用已有年份的人口总量数据预测将来的人口总量的变化趋势，从总体上对人口发展做出预测。

其次，把人口的增长特点考虑在内，利用动态模型并进行计算机模拟，得到符合中国实际情况的模型，包含了老龄化水平、性别比例、城镇化等更细致的结果。

最后，我们对每个模型的预测结果进行对比，评判其各自的优点及缺点，并对政府部门提出一些建设性的意见。

五、 模型建立及求解

1 模型一 阻滞增长模型

1.1 模型建立

针对未来中国的人口总数，我们建立简单的人口总数预测模型——阻滞增长模型（具

体建立方法请参考[1])。

我们得到如下等式：

$$\text{人口增长率函数: } r(x) = r(1 - \frac{x}{x_m}) \dots \dots \dots (1)$$

$$t \text{ 时刻人口数目函数: } x(t) = \frac{x_m}{1 + \left(\frac{x_m - 1}{x_0} \right) e^{-rt}} \dots \dots \dots (2)$$

$$\text{单位时间内人口增量方程: } \frac{dx}{dt} = r - st, \quad s = \frac{r}{x_m} \dots \dots \dots (3)$$

1.2 模型求解

首先，我们利用方程（3）以及 1990 年到 1999 的中国总人口数据（数据见附件 1）用 MATLAB 软件对方程（3）进行线形最小二乘法拟合，得到 $r = 0.05$ ， $x_m = 14753(\times 10 \text{万})$ 。则人口总量函数为：

$$x(t) = \frac{14753}{1 + \left(\frac{14753}{11433} - 1 \right) \times e^{-0.0568t}} \dots \dots \dots (4)$$

然后，我们用(4)式，得出计算结果，并与实际数据作比较，分别得到阻滞增长模型拟合中国人口数据的结果（表 2）和阻滞增长模型拟合曲线（如图 1）：

年份	1990	1991	1992	1993	1994	1995	1996	1997
实际人口（亿）	11.43	11.58	11.72	11.85	11.99	12.11	12.24	12.36
计算人口 x（亿）	11.43	11.57	11.72	11.85	11.98	12.11	12.23	12.35
年份	1998	1999	2000	2001	2002	2003	2004	2005
实际人口（亿）	12.48	12.58	12.67	12.76	12.85	12.92	13.00	13.08
计算人口 x（亿）	12.46	12.57	12.67	12.77	12.86	12.96	13.04	13.13
年份	2006							
实际人口（亿）	13.14							
计算人口 x（亿）	13.21							

表 2 阻滞增长模型拟合中国人口数据的结果

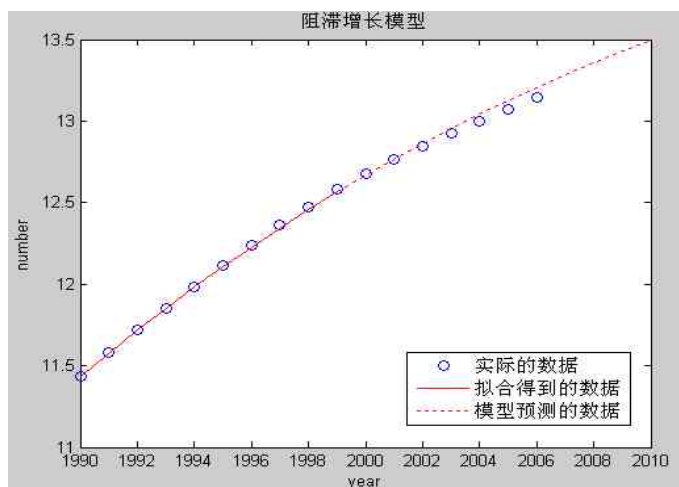


图 1 阻滞增长模型拟合曲线一

可以看出，这个模型拟合时虽然初始的一段（1990 年至 2002 年）吻合得不错，但是最后一段（2003 年至 2006 年）不太好。

1.3 模型的分析与优化

在上述模型中，我们并没有把 2000 年到 2006 年的实际数据参与函数的拟合，目的是为了用它们作模型的检验。我们用模型计算的数据和这一段的实际数据比较，来检验模型是否合适，经计算得到 2000~2006 年人口计算与实际数据的相对误差（如表 3）：

年份	2000	2001	2002	2003	2004	2005	2006
相对误差 (%)	0	0.078	0.078	0.31	0.31	0.38	0.53

表 3 2000~2006 年人口计算与实际数据的相对误差

从表中可以看出，误差均比较小，所以可以认为该模型是相当满意的。

经过上面的分析求解，我们应当把 2000 年到 2006 年的实际数据加进去重新拟合，得到新的 $x(t)$ 如下：

$$x(t) = \frac{14208}{1 + \left(\frac{14208}{11433} - 1\right) \times e^{-0.067049t}} \dots\dots\dots (5)$$

我们用（5）式对未来 14 年进行预测，得到阻滞增长模型预测 2007~2050 年的人口数据（如表 4）和阻滞增长模型拟合曲线二（如图 2）：

年份	2007	2008	2009	2010	2011	2012	2013	2014
预测人口值(亿)	13.19	13.25	13.31	13.36	13.41	13.46	13.51	13.55
年份	2015	2016	2017	2018	2019	2020	2021	2022
预测人口值(亿)	13.59	13.63	13.67	13.70	13.73	13.76	13.79	13.82
年份	2023	2024	2025	2026	2027	2028	2029	2030
预测人口值(亿)	13.84	13.86	13.89	13.91	13.93	13.94	13.96	13.98
年份	2031	2032	2033	2034	2035	2036	2037	2038
预测人口值(亿)	13.99	14.01	14.02	14.03	14.04	14.05	14.06	14.07
年份	2039	2040	2041	2042	2043	2044	2045	2046
预测人口值(亿)	14.08	14.09	14.10	14.10	14.11	14.12	14.12	14.12
年份	2047	2048	2049	2050				
预测人口值(亿)	14.13	14.14	14.14	14.15				

表 4 阻滞增长模型预测 2007~2020 年的人口数据

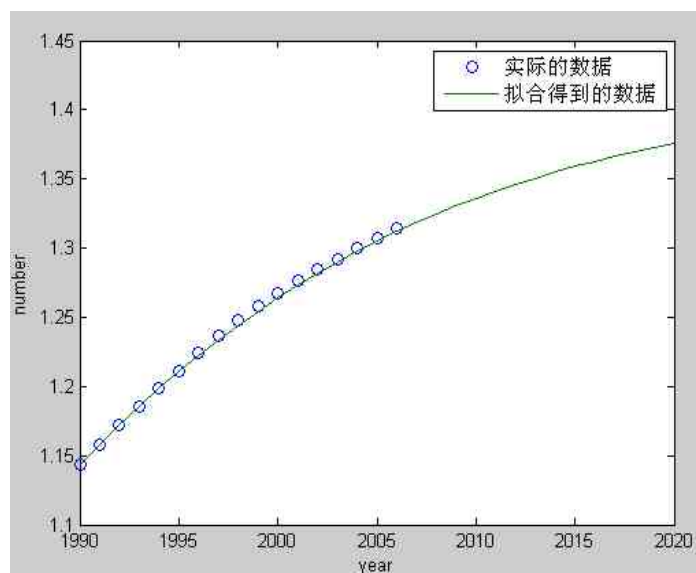


图 2 阻滞增长模型拟合曲线二

2 模型二 GM(1, 1) 灰色预测模型

2.1 模型建立

由于人的出生和死亡是随机的，因此我们利用灰色预测模型中的累加效果，尽量减小这种随机影响，在此我们采用模灰色GM(1, 1)模型（具体讲解请参考[3]）。为了使预测效果更佳，并不直接用总人口序列建模，而是首先求出各年净增人口序列，即1990～2005年各年净增人口数据（如表5），然后应用净增人口序列建模计算净增人口预测值。

年份	1990	1991	1992	1993	1994	1995	1996	1997
人口总数（亿）	11.43	11.58	11.72	11.85	11.99	12.11	12.24	12.36
净增人口数(万)	1629	1490	1348	1346	1333	1271	1268	1237
年份	1998	1999	2000	2001	2002	2003	2004	2005
人口总数（亿）	12.48	12.58	12.67	12.76	12.85	12.92	13.00	13.08
净增人口数(万)	1135	1025	957	884	826	774	761	768

表 5 1990～2005年各年净增人口数据

2.2 模型求解

原始样本序列为：

$$\mathbf{q}^{(0)} = [1629 \ 1490 \ 1348 \ 1346 \ 1333 \ 1271 \ 1268 \ 1237 \ 1135 \ 1025 \ 957 \ 884 \ 826 \ 774 \ 761 \ 768]$$

通过灰色GM(1, 1)模型，我们得到未来第t年的较1989年(1989年人口总数为11.27亿)的累积人口净增量（记1990年t=1，1991年t=2，之后以此类推）

$$\widehat{\mathbf{q}}^{(1)}(t) = 32954.4 - 31325.4 * \exp(-0.049579 * (t - 1)) \dots \dots \dots (6)$$

利用 $\widehat{\mathbf{q}}^{(0)}(t) = \widehat{\mathbf{q}}^{(1)}(t) - \widehat{\mathbf{q}}^{(1)}(t-1)$ 以及(6)式，得到人口的净增长预测值 $\widehat{\mathbf{q}}^{(0)}(t)$ 后，利用 $\mathbf{x}(t) = \widehat{\mathbf{q}}^{(0)}(t) + \mathbf{x}(t-1)$ 迭代可得出未来第t年的人口数量（记1990年t=1，1991年t=2，之后以此类推）为

$$\mathbf{x}(t) = 145660 - 31325.4 * \exp(-0.049579 * (t - 1)) \dots \dots \dots (7)$$

我们由(7)式预测从2007～2050年中国人口总数，得到灰色GM(1, 1)模型预测2007～

2050年的人口数据（如表6）：

年份	2006	2007	2008	2009	2010	2011	2012	2013
预测人口(亿)	13.149	13.217	13.283	13.345	13.404	13.460	13.513	13.564
年份	2014	2015	2016	2017	2018	2019	2020	2021
预测人口(亿)	13.613	13.659	13.703	13.744	13.784	13.822	13.858	13.892
年份	2022	2023	2024	2025	2026	2027	2028	2029
预测人口(亿)	13.925	13.956	13.985	14.013	14.040	14.066	14.090	14.113
年份	2030	2031	2032	2033	2034	2035	2036	2037
预测人口(亿)	14.135	14.156	14.175	140194	140212	14.229	14.246	14.261
年份	2038	2039	2040	2041	2042	2043	2044	2045
预测人口(亿)	14.276	14.290	14.303	14.316	14.328	14.340	14.350	14.361
年份	2046	2047	2048	2049	2050			
预测人口(亿)	14.371	14.380	14.389	14.398	14.406			

表 6 灰色GM(1, 1)模型预测2007~2050年的人口数据

2.3 结果分析

对我国2006 年实际总人口数进行检验性预测，预测值为13.149亿，实际值为13.1448亿，相对误差为0.03%。然后进行后误差检验,检验合格后的模型即可用于预测。对1990~2005年数据进行检验，得到1990~2005年人口计算与实际数据的相对误差（如表7）：

年份	1990	1991	1992	1993	1994	1995	1996	1997
相对误差 (%)	0	0.0233	0.016	0.1207	0.1001	0.0735	0.0008	0.0857
年份	1998	1999	2000	2001	2002	2003	2004	2005
相对误差 (%)	0.1371	0.1399	0.1286	0.0995	0.0568	0.0054	0.0169	0.0107

表 7 1990~2005 年人口计算与实际数据的相对误差

$S1=246.9858$, $S2=8.0308$, $C=S2/S1=0.0325<0.35$ ，表明模型预测的精度很高。根据灰色系统理论,当发展系数 $a \in (-2,2)$ 且 $a \geq -0.3$ 时, 所建GM(1, 1)模型可用于中长期预测。

但是由于中国的人口发展特点并不能保证很长时间预测的准确性，比如对中国100年后的人口数量便难以预测。

3 模型三 基于计算机模拟的动态模型

3.1 数据处理

在此模型中，我们综合考虑人口的出生死亡率、出生人口性别比例、城镇化水平等因素，把数据做如下处理：

3.1.1 取平均消除异常值

附录2中的数据是根据实际情况调查得知的，但是在某些年份，可能由于某些自然灾害、意外事故等因素，导致人口死亡率上升，出生率下降。为了排除这些意外因素对人口发展造成的异常波动，我们把2001到2005年五年的所有数据作平均化处理，即从0至90岁的各个年龄的各种比率做对应的平均；同时对于每一年的抽样调查人数也修正，首先求出每一年抽样的城镇乡男女人数各自占该年总抽样人数的比例，之后将其代入

$s(t) = a + bt$ （ $s(t)$ 为第 t 年比例，记2001年 $t=0$ ，2002年 $t=1$ ，之后以此类推）拟合，得到

其各自的拟合函数，并且经检验得到每一个的欧氏距离数量级均为 10^{-3} ，可知这个修正比较满意。

3.1.2 对人口死亡率拟合

附件2给出了我们1978年到2005年的中国人口死亡率,我们用其中1990至2005年的死亡率(其中1990、1994、2004、2005年的数据出现异常波动,是坏点,应该排除)对Compertz函数—— $\mu(t) = B \exp(\theta t)$ (其中 B 表示1991年的死亡率, θ 表示外界因素对死亡率的综合影响,称之为死亡率发展因子)进行拟合,得到 $\mu(t) = 0.0067 * \exp(-0.0042t)$ (记1991年 $t=0$, 1992年 $t=1$, 之后以此类推)..... (8)

之后对(8)式进行残差检验,得到死亡率残差表(如表8):

年份	1990	1991	1992	1993	1994	1995	1996	1997
残差(‰)	—	0	-0.0320	-0.0042	—	-0.0188	-0.0014	-0.0240
年份	1998	1999	2000	2001	2002	2003	2004	2005
残差(‰)	-0.0067	-0.0195	-0.0025	0.0044	0.0112	0.0280	—	—

表 8 死亡率残差表

我们对残差求平方和之后再开方即拟合出的数据与原始数据的欧氏距离,得到其为 5.7×10^{-5} , 此值在允许的范围内,所以拟合结果有一定的可信度。

3.1.3 2001年市镇乡男女各年龄人口数目

在附录2中,只是对人口的抽样数据,我们需要知道2001年总人口在市镇乡各个年龄的人口数量。根据大数定理可知抽样数据中市镇乡男女的分布应与总人口中的分布大致相同,由此算法如下:

Step1: 利用3.1.1的2001年抽样数据中市镇乡男女数目占抽样数据总和的比例与2001年全国的人口总数相乘,得到全国市镇乡男女的数目;

Step2: 用Step1得到的全国市镇乡男女的数目与3.1.1得到的与其对应的市镇乡男女年龄的平均比率相乘,得到该年市镇乡男女各个年龄的人口总数。

我们用所得到的数据进行预测。

3.1.4 城镇化水平拟合

首先说明城镇化进程的一般规律: 第一阶段为城镇化的初期阶段, 城镇人口增长缓慢, 当城镇人口超过10%以后, 城镇化进程逐渐加快, 当城镇化水平超过30%时进入第二阶段, 城镇化进程出现加快趋势, 这种趋势一直要持续到城镇人口超过70%以后才会趋缓, 此后为城镇化进程第三阶段, 城镇化进程停滞或略有下降趋势。由此可知, 城镇化进程会经历发生、发展、成熟三阶段, 其规律性的变化是发生阶段速度缓慢, 发展阶段速度加快, 成熟阶段速度又趋缓慢。

由此可见城镇化过程具有S型曲线特征, 所以我们用Logistic方程描述乡村人口城镇化水平的发展。

$$\text{Logistic方程为: } z(t) = \frac{z_m}{1 + \left(\frac{z_m - 1}{z_0} \right) e^{-rt}}, \text{ 我们定义 } z_m \text{ 为城镇化率的最大值, 由于}$$

当今世界发达国家的城镇化率已达90%, 并且当城镇化率达到90%后, 很难再增长, 所以这里取 $z_m = 90\%$ 。我们利用往年中国的城镇化率(城镇化率见附件3)对此方程进行

拟合，得到 $r = 0.02466$ ，拟合到城镇化拟合曲线（如图3）：

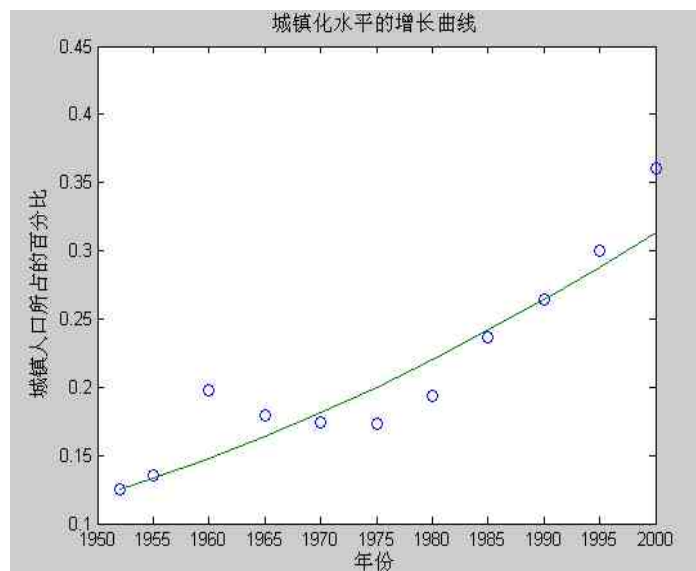


图 3 城镇化拟合曲线

从拟合的曲线看出，拟合的效果不是很好，有些点偏离曲线较严重，这可能是由于一些人为或自然的因素导致这些年份出现异常，但它对于我国将来的预测还是有一定的指导意义。

3.2 模型的建立及求解

3.2.1 模型建立

以2001年为起始年份，利用动态模拟对未来人口进行预测。

首先，我们利用 3.1.1 得到的市镇乡各个年龄的平均生育率求到 2001 年的总和生育率，发现其并没有达到实际情况的 1.8 左右，因此，我们对其进行如下修正：

$$b_i(2001, k)_{\text{修正}} = b_i(2001, k) + \frac{b_i(2001, k)}{\sum_{i=1}^3 b_i(t)} \left(5.4 - \sum_{i=1}^3 b_i(t) \right), \text{ 得到修正后的市镇乡各个年龄的}$$

平均生育率，并且在未来的预测中，均采用这个生育率，即把生育率当作一个定值，同时假设现阶段我国总和生育率为 1.8。

然后，按照下面的算法进行计算机模拟求解：

Step1: 利用3.1.3得到的数据与修正后的市镇乡各个年龄的平均生育率对2002年的市镇乡出生的男女总数进行预测，其中男女的出生比例采用模型假设5所设定的值（此值参考了大量信息）；

Step2: 利用3.1.1得到的男女各个年龄的平均死亡率代入 $\mu(t) = B \exp(\theta t)$ （ B 为各个年龄的平均死亡率， $\theta = -0.0042$ ）求出2002年的市镇乡男女各个年龄的死亡率；

Step3: 用Step2得到的数据与3.1.3的数据求出2002年市镇乡男女各自0到88岁存活的人数；

Step4: 利用3.1.3得到的89岁的人数计算出90岁时存活的人数，然后加上90+存活的人数得到2002年市镇乡男女各自90+的总人数；

Step5: 将Step1、Step3、Step4综合便得到2002年市镇乡男女各个年龄的人口总数；

Step6: 利用Step5得到的数据便可计算出2002年全国男女比例，市镇乡各自男女的比例，人口结构，乡村人口向城镇搬迁情况；

Step7: 之后利用预测到的2002年的数据代替初始数据从Step1开始循环, 则可以得到2003年的人口情况, 之后以此类推, 便可得到未来某一年的人口情况。

算法框图如图4:

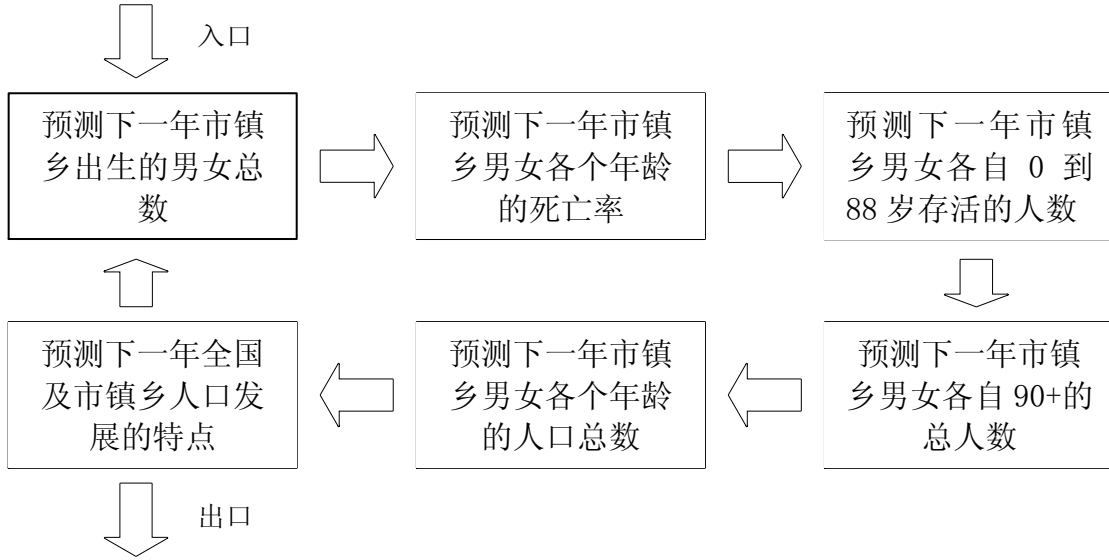


图 4 人口预测算法框图

得到的迭代方程组为:

$$\begin{cases}
 x(t) = \sum_{r=0}^{90} \sum_{i=1}^3 [w_i(t,k) + m_i(t,k)] \\
 x(t+1) = x(t) + \sum_{r=15}^{49} \sum_{i=1}^3 b_i(t,k) w_i(t,k) - \sum_{r=0}^{90} \sum_{i=1}^3 [\mu_{w_i}(t,k) w_i(t,k) + \mu_{m_i}(t,k) m_i(t,k)] \\
 w_i(t+1, k+1) = w_i(t,k) [1 - \mu_{w_i}(t,k)] \quad (k = 1 \dots 89) \\
 m_i(t+1, k+1) = m_i(t,k) [1 - \mu_{m_i}(t,k)] \quad (k = 1 \dots 89) \\
 w_i(t+1, 90) = w_i(t, 89) [1 - \mu_{w_i}(t, 89)] + w_i(t, 90) [1 - \mu_{w_i}(t, 90)] \quad (i = 1, 2, 3) \\
 m_i(t+1, 90) = m_i(t, 89) [1 - \mu_{m_i}(t, 89)] + m_i(t, 90) [1 - \mu_{m_i}(t, 90)] \quad (i = 1, 2, 3) \\
 w_i(t+1, 0) = \frac{100}{100 + q_i(t)} \sum_{r=15}^{49} b_i(t,k) w_i(t,k) \quad (i = 1, 2, 3) \\
 m_i(t+1, 0) = \frac{q_i(t)}{100 + q_i(t)} \sum_{r=15}^{49} b_i(t,k) w_i(t,k) \quad (i = 1, 2, 3)
 \end{cases} \quad \dots (9)$$

其中 $q_i(t)$ 为男女出生比例, $w_i(t,k)$ 和 $m_i(t,k)$ 分别表示 t 年时年龄为 k 的女性和

男性数量, $\mu_{w_i}(t,k)$ 和 $\mu_{m_i}(t,k)$ 表示 t 年时年龄为 k 的女性和男性死亡的比例, $b_i(t,k)$ 表示 t 年时年龄为 k 的妇女的生育率, $i=1, 2, 3$ 时分别表示来自城市乡镇和农村。

3.2.2 模型求解

结果一: 人口总数预测

由方程组 (9) 我们得到2002~2025年的人口数目 (如表9):

年份	2002	2003	2004	2005	2006	2007	2008	2009
人口数目(亿)	12.867	12.966	13.059	13.148	13.237	13.328	13.422	13.520
年份	2010	2011	2012	2013	2014	2015	2016	2017
人口数目(亿)	13.621	13.722	13.821	13.914	14.001	14.079	14.148	14.207
年份	2018	2019	2020	2021	2022	2023	2024	2025
人口数目(亿)	14.257	14.296	14.324	14.342	14.349	14.349	14.343	14.333

表 9 2002~2020年的人口数目

同时我们也对未来100年进行了预测，得到了未来100年人数曲线（如图5）：

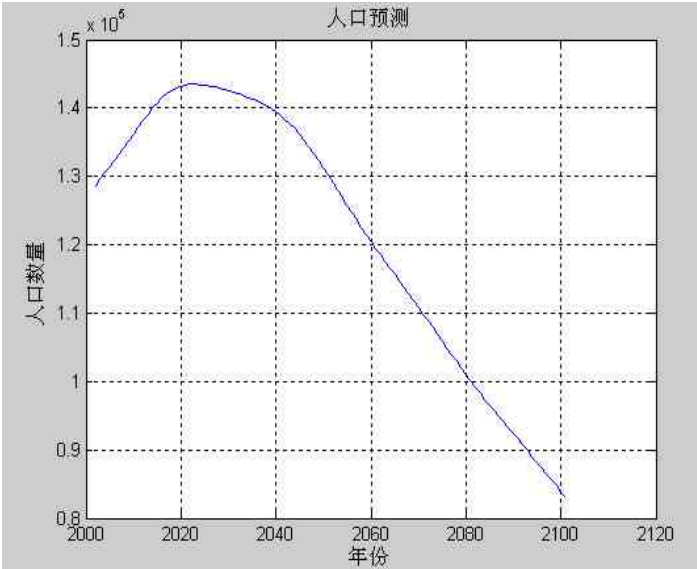


图 5 未来100年人数曲线

在预测的数据中，我们用2002~2006年的实际数据对模型进行检验，得到2002~2006相对误差表（如表10）：

年份	2002	2003	2004	2005	2006
相对误差(%)	0.13	0.33	0.46	0.55	0.70

表 10 2002~2006相对误差表

从表中我们可以看出，随着年份的推移，相对误差越来越大，但在 2006 年的相对误差仍小于 1%，所以我们可以认为模型是合理的。从图中我们可以出如果人口按这种趋势发展的话，将在 2022 年达到峰值，为 14.349 亿。

我们利用此模型对未来 500 年进行预测，观察其更远年份的发展趋势，得到未来 500 年人数曲线（如图 6）

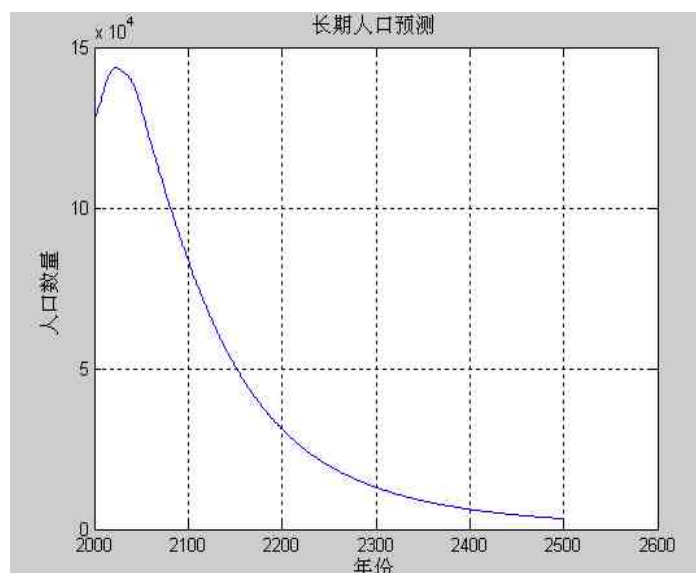


图 6 未来 500 年人数曲线

从曲线中我们可以看出，此模型对于长期预测数据不准确，年份越久，人口数目越少，到 2100 年人口低于 10 亿，这显然是不合理的。出现这种情况的原因可能是由于本模型并没有考虑政府的调控因素，比如人口出生率低的时候，政府就会鼓励生育。

结果二：性别比例预测

我们利用方程 (9) 也可以预测未来人口的性别比例，得到未来 100 年适婚男女人数之差（如图 7）：

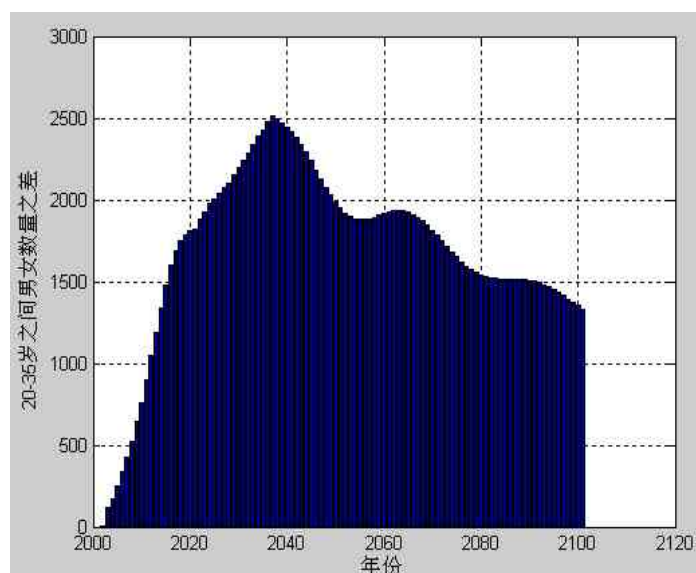


图 7 未来 100 年适婚男女人数之差

从图中我们可以看出，在 2037 年时，20-35 岁之间男女人数之差达到最大，为 2514.2 万人，并且在 2000 年到 2037 年两者之差呈增长趋势，如果按照这种情况发展的话，我国的男女比例将会严重失调，将对人口结构产生重大负面影响，对未来社会的良性稳定运行、社会伦理道德体系也会造成一定冲击。同时男女失调也将带来一系列的社会问题，如婚配失当、人口拐卖、性行为错乱等，影响社会尤其是婚姻的稳定。

为使社会呈现良性发展，政府应该进一步提高我国计生工作的质量，这是解决性别比失调的关键。而在计生工作上，人的观念是决定因素，政府部门应多进行宣传教育，一方面对独女户进行扶持，解决其后顾之忧，另一方面对违反法规因歧视性性别偏好而

终止妊娠的人进行惩罚，把偏高的出生性别比降下来，使人口控制和经济发展、社会进步走上健康、和谐的轨道。

结果三：人口结构及老龄化预测

同样，我们根据模型三的结果也可以预测我国未来的人口结构，我们得到 2010 年、2020 年、2030 年、2040 年的人口结构饼状图（如图 8）：

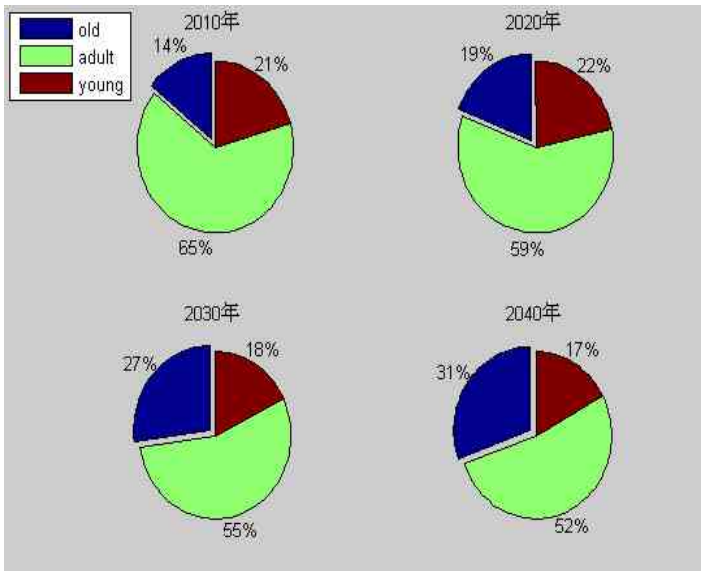


图 8 2010 年、2020 年、2030 年、2040 年的人口结构饼状图

从图中我们可以很清晰的看到，老年人的比例逐渐增大。根据这种趋势，我们可以预测未来成年人仍然在总人口中占有相当大的比重，而青年人的比重会下降，老年人的比重会上升，长期发展下去，老龄化会越来越严重，社会将会呈现出衰退型。

目前国际通行以老年人口占总人口的比重作为衡量老年年龄化的指标，60 岁以上老年人口占总人口的比重达到 10% 以上，或者 65 岁以上的老年人口占总人口的比重达 7% 以上，标志着这个国家或地区的人口进入了老年型，人口开始老化。在这里我们以 60 岁为衡量的界限，利用模型二的数据进行预测，得到未来 100 年的人口老龄化曲线（如图 9）：

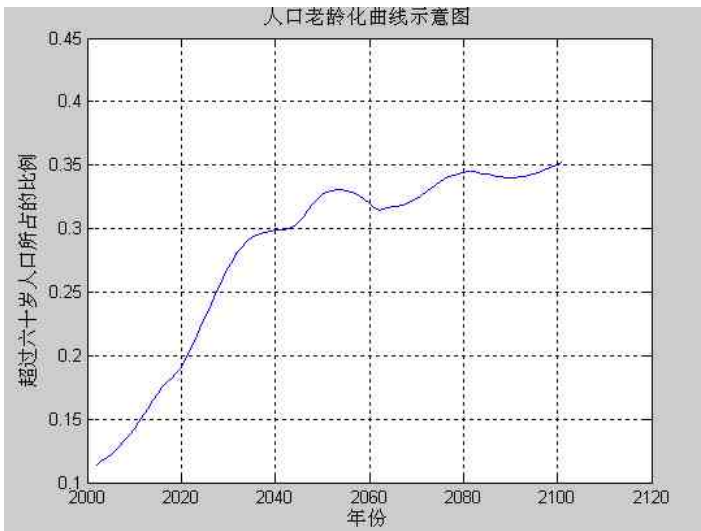


图 9 未来 100 年的人口老龄化曲线

从曲线上我们可以看到，人口老龄化呈增长趋势，2040 年之前几乎呈线性增长，之后波动上扬。我们预测到 2010 年的老龄化比例为 14.25%，2020 年为 19.10%，2030

年为 26.90%，2040 年为 30.00%，这些数据显示我国的人口老龄化将会越来越严重。但这是没有任何外界干扰的情况下的发展状态，因此预测有一定的局限性，不过这也给政府部门敲响了警钟。

同时，我们也用现今一些年份的人口出生率与 60 年后年龄 60 岁的人口数目进行了对比（如图 10）：

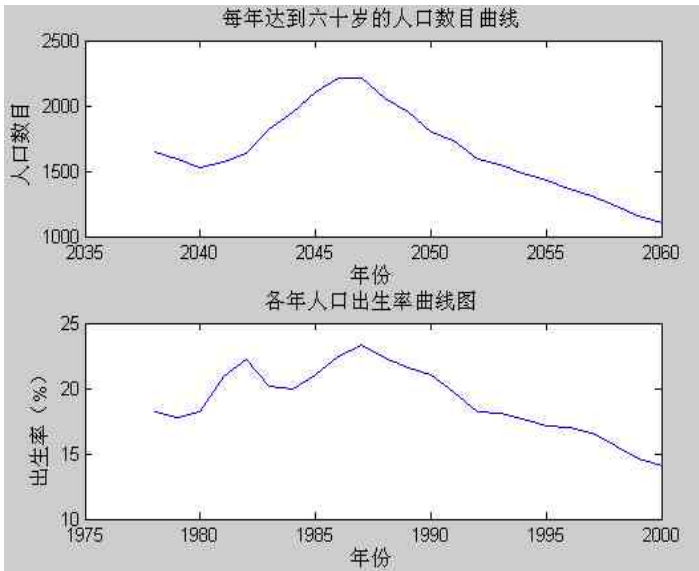


图 10 1978~2000 出生率与 2038~2060 年 60 岁人口数目对比图

从对比图中我们可以看出两者的变化趋势基本相同，可得知人口老龄化是有一定的滞后性的，即每年的生育率（出生率）都会影响到后六十年步入六十岁的人口数目，也就是说现在所执行的人口政策直接关系到我们子孙后代。

面对人口年龄结构老化的挑战，我们应以科学的态度对待，以积极的态度迎接人口年龄结构老化的挑战，积极采取对策，充分利用人口老化前期总抚养比较低的有利形势，大力发展社会经济，建立起完善的社会保障制度。同时，我们又要认真研究人口年龄结构老化的消极后果，尽量避免或减缓人口年龄结构老化对中国社会经济的消极影响。

因此，从战略的高度认识中国的人口年龄结构老化问题，从宏观上看清人口年龄结构老化的趋势，及早地制定人口发展和经济发展相协调的总体战略规划，统筹兼顾，综合治理十分必要。

结果四：乡村人口城镇化预测

利用模型三得到的数据，我们还可以对乡村人口城镇化进行预测，得到未来 100 年城镇人口占总人口比重的增长曲线（如图 10）：

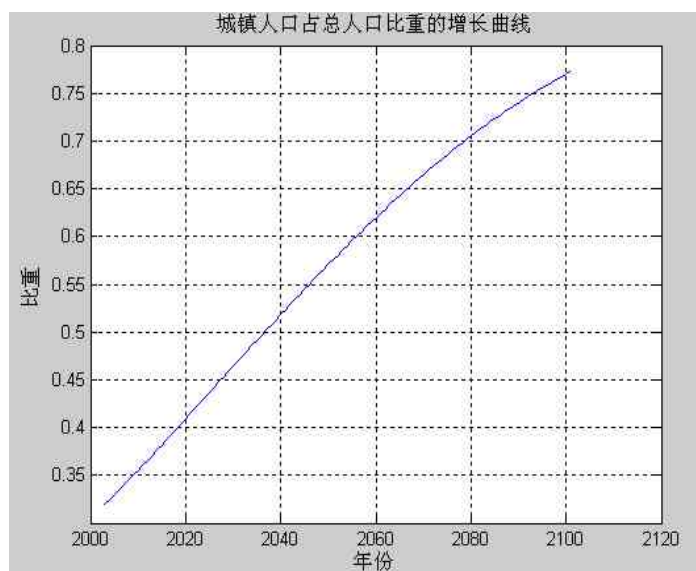


图 11 未来 100 年城镇人口占总人口比重的增长曲线

从曲线上可以看出城镇人口的比重在不断增加，由预测可见我国城镇化水平将不断提高。

同时，我们也得到城、镇、乡人口未来 50 年内的变化曲线（如图 11）：

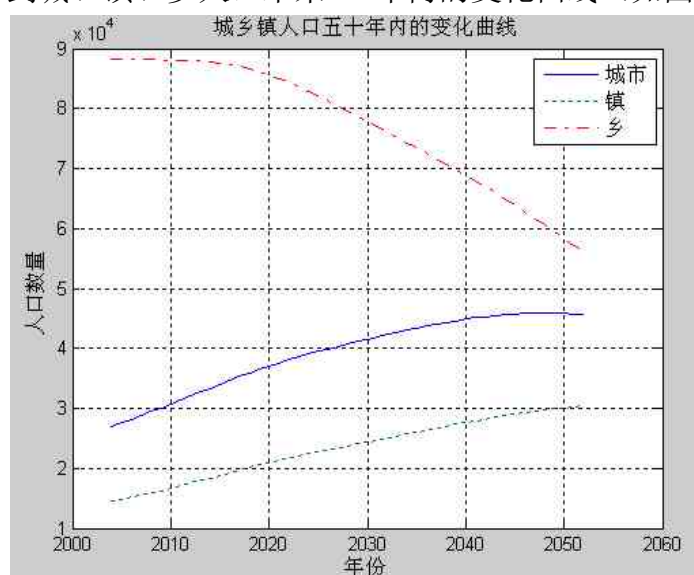


图 12 城、镇、乡人口未来 50 年内的变化曲线

从图中，我们可以看出乡村人口总量将越来越小，下降趋势比较快，城市和镇人口总量比例越来越大，我们列出 2010、2020、2030、2040 年城、镇、乡人口总量表（如表 11）：

年份	2010	2020	2030	2040
城市人口（亿）	3.0725	3.7030	4.1536	4.4860
镇人口（亿）	1.6702	2.0948	2.4413	2.7623
乡村人口（亿）	8.8079	8.5686	7.7803	6.8908

表 11 2010、2020、2030、2040 年城、镇、乡人口数目表

表中的数据也反映了城镇化水平在不断提高。

我们给出 2000 年到 2100 年，每年乡村向城镇转移的人口数目(如图 13)：

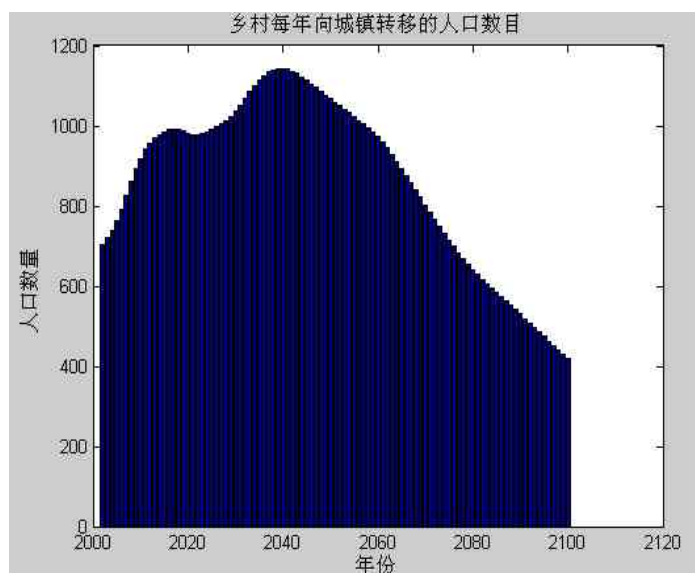


图 13 2000~2100 年乡村向城镇转移人口数目

图中表明在 2040 年之前，转移人口在逐年增加，并在 2041 达到最大值，1142.1 万。这就要求城镇要做好相应的准备，应对即将到来的人口转移。

结果五：人口抚养比预测

利用预测的各个年龄人口数目，我们对人口抚养比进行了预测，其中人口抚养比是指人口总体中非劳动年龄人口数与劳动年龄人口数之比。在这里我们定义 15 至 64 岁人口为劳动年龄人口，得到 2000 至 2100 年的人口抚养比变化曲线图（如图 14）：

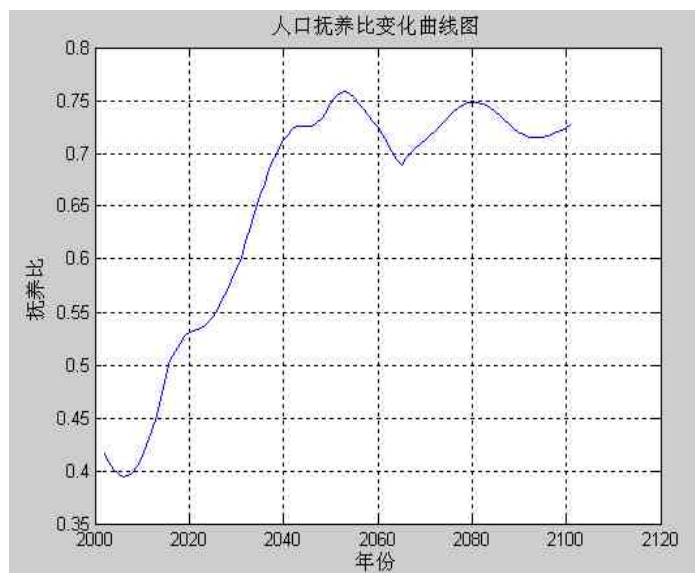


图 14 2000 至 2100 年的人口抚养比变化曲线图

从曲线图中我们可以看出，2015 年之前人口抚养比低于 50%，为人口红利时期（人口抚养比低于 50% 的年份，作为在人口年龄结构上最有利于经济发展，有可能获取生育率下降所带来的“人口红利”的时期）；2053 年人口抚养比达到最大，为 75.8%。

根据预测的情况我们看出，我国未来只有不到 10 年时间处于经济发展期了，因此在这十年内，我国应加紧利用这一时期内丰富的劳动资源，必须长时期保证经济又好又快增长，同时必须尽快建立起覆盖全体居民的社会保障体系，以应对到来的人口负债时期。

3.3 模型三的改进

经过模型三的求解，我们可以看出我国未来的人口规模在不断加大，老龄化现象越

来越严重,因此应该对人口出生率进行控制,从而达到对人口规模和老龄化比例的控制。

在理想的发展状态下,我国人口规模的理想数值为 8 亿,老龄化比例的理想值为 10 %,所以模型改进中把其当作标准,对每一年的人口规模和老年人比例做归一化处理:

$$p(t) = \frac{x(t)}{8}, \quad o(t) = \frac{y(t)}{10\%} \dots\dots\dots (10)$$

之后判断 $p(t) - o(t)$ 的正负,用其对下一年的出生率做出调整,调整后的等式为:

$$b_i(t+1,k) = b_i(t,k) + \frac{p(t) - o(t)}{\sum_{i=1}^3 b_i(t)} (c * b_i(t,k)) \dots\dots\dots (11)$$

(11) 式中 c 代表人为的控制因子,是一个未知数,需要对其进行筛选,选择最优的值使我国人口良性发展。模型改进的其他部分与模型三一致。

经过改进我们得到了不同控制因子下的人口发展曲线图(如图 12)。

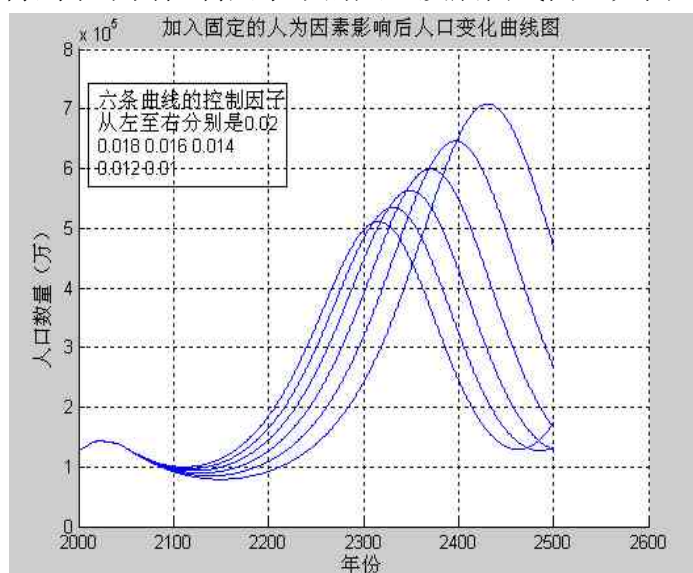


图 15 不同控制因子下的人口发展曲线图

在图中,我们对未来 500 年进行预测,目的是为了看出在不同的控制强度 c 下,人口发展的变化趋势,但并不具有实际意义。曲线表明在预测年份不长时,不同控制因子下人口数量并没有太大的差别;当年份较长时,控制因子大的人口数量将提前达到最大值,并且最大值随着控制因子的减小而增大。在达到最大值后,人数便会下降。

我们可以用牛顿第二定律去证明这种变化情况,假设路程 S 代表人口数量,控制因子 c 越大时, a 越大,由 $S = \frac{v^2}{2a}$ 可得到在人口增长速度 v 一定的情况下,控制因子越大,人口数量的峰值越小;由 $t = \frac{v}{a}$ 可得在人口增长速度 v 一定时,控制因子越大,人口数量到达峰值的时间就越短。这只是用类比的方法对这种现象做粗略的解释,由于时间关系,我们就不做深入研究了。

我们针对我国当前的情况,在现阶段我国应该采取较高的控制因子,使我国人口降到一定数值后,再减小控制因子,使人口数目保持在一定值。我们因此建议政府:现在中国人口处于高峰时刻,应该坚持计划生育这项基本国策,而非某些人所说的为了避免人口老龄化而增加人口生育率,选择非计划生育,否则带来的人口反弹将使我国面临更

加严重的人口问题。

在预测为来人口发展情况时，我们把模型三看作不变的生育率做了预测，这样使我们的预测更准确，更具有指导意义。

六、 模型评价

模型一采用 Logistic 阻滞增长模型对人口增长率做出了合理、简化的假设，即将人口增长率看作是人口数量的函数。同时它的中期预报结果比较准确，可以对人口数量进行粗略的预测。但是，在此模型中，并没有考虑年龄因素，而且它只能对人口数量作出中短期预测，理论上很好，实用性不强，由一定的局限性。

模型二基于灰色预测理论,研究了基本灰色 GM(1, 1) 预测模型，灰色模型预测相对误差小，精度高，适用于人口数量的中长期预测。但是同模型一类似，本模型也没有考虑年龄因素，只是利用现有人口数据对未来总数的预测，没有结合中国国情与人口增长的特点。

模型三把市镇乡人口比例、各年龄死亡率，生育率等各种因素都考虑在内，利用了计算机模拟的方法进行了预测。在预测前，我们对所给的数据进行了合理的处理：一方面对数据取平均排除了意外因素对人口发展造成的异常波动，同时对抽样人口中市镇乡男女所占总抽样人口的比例作了修正，得到比较准确的比例，为求解 2001 年的市镇乡男女数目服务；另一方面对人口死亡率和城镇化水平进行了拟合，求出了各自的变化趋势。在模型结果分析中，我们进行了误差分析得出模型可行。本模型能够对未来的人口总数、性别比例、人口结构、老龄化水平、城镇化水平、人口抚养比等进行预测，使结果更为细化，但不足之处是所要处理的数据量比较大，可能会由于原始调查数据的误差降低预测的准确性。

同时我们还对模型三进行了改进，加上了人为的调控作用。当人口规模过大或老龄化水平过高时，对下一年的生育率进行调整，使这些不好的情况得到改善。改进中得到的控制因子只是一种平均的水平，而在实际情况中，应根据当年的情况来确定下一年控制因子的大小。

为了对三个模型的优缺点有更直观化的了解，我们列出模型优缺点对比表（如 12）；

模型	优点	缺点
Logistic 阻滞模型	所需的数据少，计算简单，所耗时间少，中短期预测较准确	只能研究单调的情形，只是对数据进行单纯处理，没有考虑事物内部因素，不能进行长期预测
灰色预测模型	所需数据少，计算简单，所耗时间少，中长期预测效果好	只能研究单调的情形，只是对数据进行单纯处理，没有考虑事物内部因素，不能进行长期预测
动态模拟	模型准确，充分考虑到个体的情况，能够进行长期预测	初始数据要求准确度高，计算复杂，所耗时间长

表 12 模型优缺点对比

七、 模型灵敏度分析

1 人口死亡率 Compertz 函数中死亡率发展因子 θ 的灵敏度分析

在人口死亡率的拟合中，我们得到了死亡率发展因子 $\theta = -0.0042$ ，考虑到所选取的

死亡率值可能会出现一些误差，因此对其进行灵敏度分析，得到不同 θ 下人口总数预测曲线（如图 13）

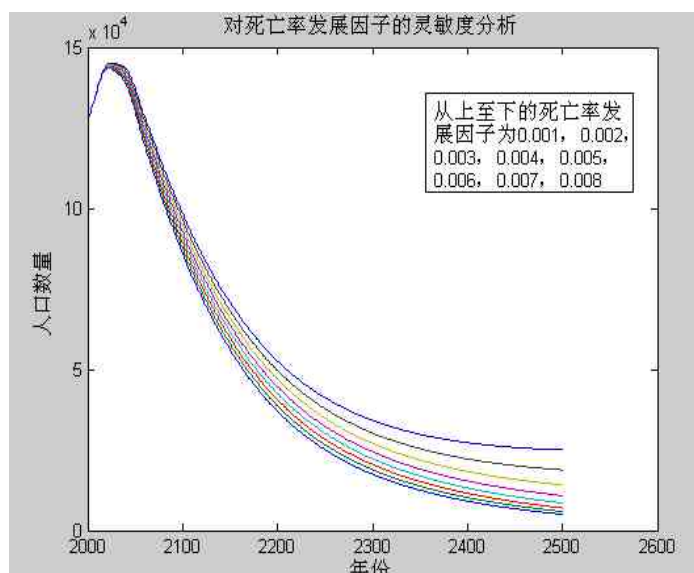


图 16 不同 θ 下人口总数预测曲线

从图中我们可以看出，人口总数对 θ 的灵敏度并不高，因此我们可以认为用拟合得出的 θ 进行人口总数的预测是可行的。

2 模型三男女出生比例灵敏度分析

在模型三中，考虑出生的男女比例时，我们假设在 2020 年之前男女出生比例为 116，这个值是从现今中国的实际生育情况得到的；在 2020 年之后男女出生比例是 107，这个值是由生物学规律决定的（正常情况下，应保持在 103 到 107 之间）。在预测时，这个比例的确定性不高，因此需要对其进行灵敏度分析。

我们在模型三的基础上，在其他因素不变时，取不同的比例进行预测，得到不同出生比例下性别比预测图象（图中的值对比例进行了化简）（如图 12）：

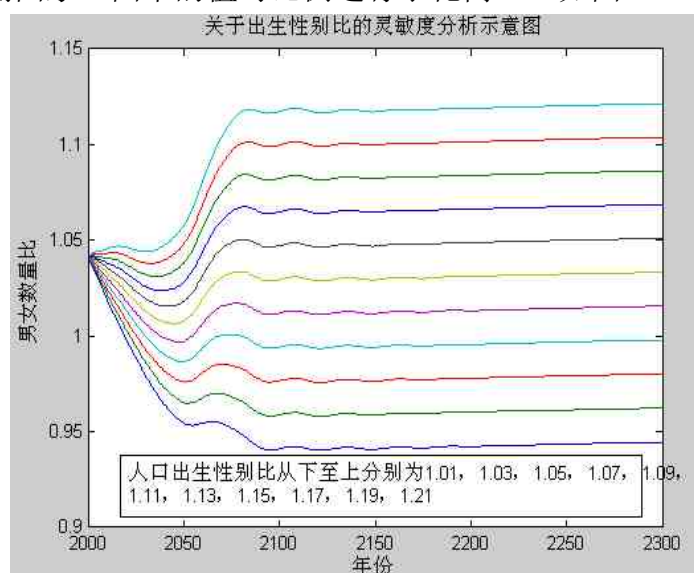


图 17 不同出生比例下性别比预测图象

从图上可以看到，在 2000 至 2100 年的数据不稳定，波动范围大，所以我们取之后 200 年的数据进行处理。我们以男女人数比 100 为标准进行衡量，对每一个出生比例下后 200 年的男女比例值求平均，之后用平均值减去 100，得出各出生比例下男女比例偏

离标准的值（如表 13）：

男女出生比例	101	103	105	107	109	111
偏离值	-5.77	-3.99	-2.20	-0.42	1.36	3.14
男女出生比例	113	115	117	119	121	
偏离值	4.91	6.68	8.45	10.21	11.97	

表 13 各男女出生比例下男女比例偏离标准值

从表中我们可以看出当男女出生比例为 107 时，男女人数比例与标准最接近，即男女人数平衡。之后计算两两相邻的偏离值差的平均值，得到灵敏度系数为 0.8850。

3 模型三妇女生育率灵敏度分析

在模型三中，我们在对人口数目进行预测时，假设总和生育率不变，但是在人口生育率比较大时，人口的总数波动可能会比较大，因此我们对其进行灵敏度分析。

我们得出在不同生育率下，人口总数的预测曲线（如图 18）：

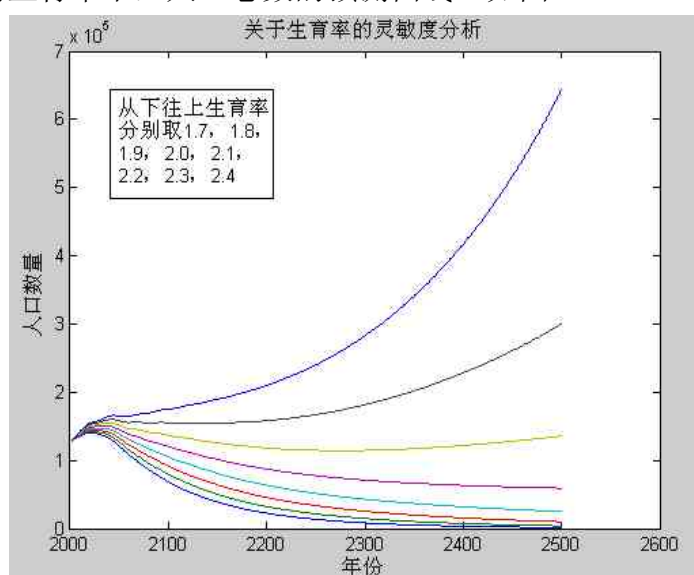


图 18 不同生育率下人口总数的预测曲线

从图中我们可以看到，生育率越低，曲线波动越小，灵敏度越低；生育率越高，曲线波动越大，灵敏度越大，并且在生育率取某个值时，人口会趋于稳定，这个值在 2.1 与 2.2 之间，具体什么值我们就不再分析，因为模型本身具有一定的误差。

八、模型的合理性分析：

本文首先从最简单的情况入手，不考虑人口的年龄层次和男女出生比例，采用 Logistic 模型和灰色预测模型对人口的总数进行预测。由于这两个模型对于种群数量是单调改变的情况预测准确度很高，而人口总数的变化并不总是单调的，所以只能对人口进行短期或中短期预测，长期预测结果只可作为参考。

然后，把年龄层次、男女出生比例、妇女生育率等因素考虑进去，建立了基于计算机模拟的动态模型。我们首先对数据做了合理的处理，使其可信度和准确度更高，之后在模型的建立求解过程中，用各个地区各个年龄的人口数、死亡率、妇女的生育率、男女出生比例对我国未来的人口发展作出预测，包含人口总数、性别比例、人口结构、老龄化水平、城镇化水平、人口抚养比例等在内的多项预测，使预测结果更具体。

同时，我们还加入了对人为因素的考虑。在模型三的改进中，加入了控制因子，通

通过对妇女生育率的调控来达到控制人口规模以及老龄化比例，使人口的预测更具有实际意义。

可以看出，所建立的三个模型由浅入深，由简单到复杂，由粗略到具体，对问题进行了预测求解。各个模型均在合理的假设下，对人口发展得到了合理的结果。

九、 给政府的建议

1. 控制好男女比例，使其处于正常水平——107 左右，消除性别歧视；
2. 多创造一些就业机会，有效解决乡村人口向城镇转移带来的巨大压力；
3. 积极推进和完善社会养老保险制度，以应对人口老龄化高峰的到来；
4. 坚持计划生育这一项基本国策，并在适当的时候调整改革生育政策，以解决未来可能出现的人口问题如劳动人口减少和老龄化程度加重；
5. 抓住人口红利时期，大力发展经济，完善社会的保障体系，为应对人口负债时期的到来做好准备。

【参考文献】

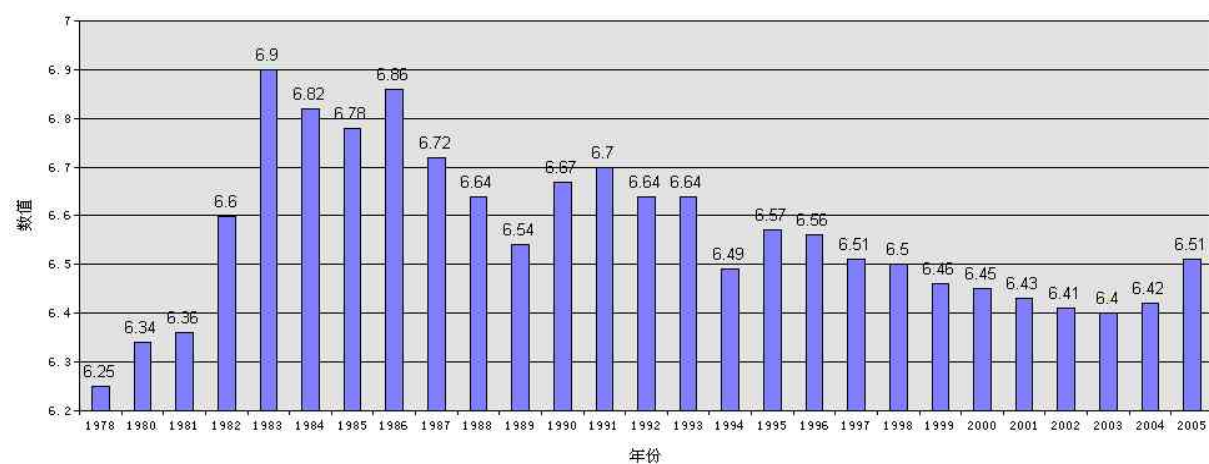
- [1] 姜启源，谢金星，叶俊。《数学模型（第三版）》[M] 北京，高等教育出版社，2004 年4月；
- [2] 苏金明，阮沈勇。《MATLAB 实用教程》[M] 北京，电子工业出版社，2005 年7月；
- [3] 邓聚龙。《灰色预测与决策》[M] 武汉，华中工学院出版社，1986：30~95；
- [4] 国家统计局。《中国统计年鉴》[M] 北京：中国统计出版社, 2006 年；
- [5] Mark M. Meerschaert。《数学建模方法与分析》[M] 北京，机械工业出版社，2005 年6月；

【附件】

1. 1990 至 2006 中国总人口数据表

年份	1990	1991	1992	1993	1994	1995	1996	1997
实际人口（亿）	11.43	11.58	11.72	11.85	11.99	12.11	12.24	12.36
年份	1998	1999	2000	2001	2002	2003	2004	2005
实际人口（亿）	12.48	12.58	12.67	12.76	12.85	12.92	13.00	13.08
年份	2006							
实际人口（亿）	13.14							

2. 中国人口死亡率（‰）



3 中国人口城镇化率

年份	1952	1955	1960	1965	1970	1975	1980	1985	1990	1995	2000
城镇化率（%）	12.46	13.48	19.76	17.98	17.38	17.34	19.39	23.71	26.41	30.00	36.09