中国人口发展趋势的分析与预测

摘要:本文基于 2001-2005 年中国人口统计数据，结合中国人口特点，采用改进的 Leslie 模型进行了中短期的人口增长预测。鉴于城，镇，乡之间生育模式，平均寿命，年龄结构，出生婴儿性别比等存在显著差别，我们对三类人群分别予以考虑，采用差分方程的建模方法，通过对未来时间上总和生育率，死亡率，男女出生比例等的合理预测，建立了描述人口增长趋势与人口结构演变的数学模型。同时，结合中国具有大规模的人口流动的特点，我们引入了人口迁移率的概念，对不同年龄段的迁移规模进行了预测，进而实现了中短期的人口预测。而对于长期人口发展趋势的预测，则需要考虑人口长期变化的新特点。我们在中短期预测模型基础上引入人均寿命和出生人口性别比随时间变化的因素的影响，建立相应的模型，分析以往数据重新预测未来中国人口增长趋势。

我们利用长期预测模型的预测结果，分析并得出中国人口近期将依旧保持增长，零增长将在大约 20 年后达到，在 2024 年前后将达到峰值 14.2 亿，在 2050 年以后人口总数持续降低，在 2100 年降至 9.5 亿人左右。老龄人口在全国人口中所占百分比持续增加，2051 年的老龄人口百分比达到最大值 32.61%。

关键词：Leslie 矩阵　人口迁移率　总和生育率

1. 问题提出

中国是人口大国，改革开放以来，中国的经济取得了突飞猛进的发展，其中我国所拥有的丰富的劳动力资源为其提供了强劲的动力。然而，我们在受益于人口数量大的同时，面临着很多严峻人口方面问题，比如我国将面临人口结构老龄化严重、男女性别比例失衡、城乡人口比例失调和人口素质偏低等问题。总之，我国正面临着比 20 世纪更为复杂的人口发展形势。为此，我们要采取积极措施应对人口老龄化、男女性别比例失衡以及稳定人口数量等问题。

人口预测作为人口学的一项重要内容，其研究可给人口发展问题带来非常有意义的指导作用。一方面，我国已经完成了 5 次全国范围的人口普查，人口数据日臻完善，另一方，人口预测理论、方法和应用已经相对比较成熟。这些为我们建立人口预测模型提供了非常宝贵的资料。但是由于人口增长是一个非常复杂的问题，已有的模型或多或少的存在一些缺陷，例如预测精度低、预测时间短。如何建立一种能够根据已知的人口数据，来精确的、长期的预测具有中国特色的人口发展的数学模型是一个亟待解决的问题。在本文中我们主要针对这些问题建立相应的数学模型预测我国人口的发展。

2. 问题分析

影响人口发展趋势的因素多种多样，而中国人口结构复杂，主要又分为城，乡，镇三类人群。基于 2001 至 2005 年统计数据及经验资料进行分析，不难发现，三类人群的人口发展形势有很大的不同：对国家的人口生育政策执行力度，地区生育观念，各自生育模式都有较大的区别，年龄结构分布也不一致。因此，在模型建立中，我们将城，乡，镇人群看作 3 个模块，来进行推导和分析。

三类人群并不是独立开来，而是通过之间的人口流动联系在一起的。考虑到中国的特殊情况，三类人群间的人口流动主要有两方面：一是农村城市化，居住地迁移等导致的人口迁移，一是农村青壮年劳动力入城引起的人口流动。由此，引入迁移率的概念，并求出年龄——迁移率曲线，对上述流动情况进行定量的描述与预测。

三类人群又有自身的增长规律，我们采用 Leslie 模型对各类人群的增长趋势做出预测，建立了以城，镇，乡各龄人口数为状态变量，以总合生育率为主要控制参数的差分方程模型。在 Leslie 模型中，总和生育率，各年龄死亡率等对预测的结果都有很大的影响，基于 5 年内的统计数据，分别对未来时间段上的总和生育率，生育模式，男，女各年龄死亡率及新生婴儿性别比等做出合理的估计，建立我国人口发展趋势的中短期模型。在长期的预测模型中上述各因素对其有影响外，我们还需要针对长期预测的特点考虑人均寿命随时间变化对于各年龄段死亡率的影响，出生人口性别比在长期中的变化特点重新建立人口预测模型。

3. 模型假设

3.1 不考虑跨国跨境人口迁移因素的影响。

3.2 假定未来一段时间内没有发生自然灾害、战争等对人口数量及其分布产生的重大影响的事件。

3.3 在未来一段时间内，国家的政策（包括人口政策）没有大的变化。

3.4 人口普查和人口调查所得到的数据真实可靠，国家人口统计局发布的人口数据能够正确的反映我国人口形势。

4. 符号约定

t：表示第 2000+t 年，t=1，2，3…分别表示 2001，2002，2003..年。

j：表示不同人群，j=1、2、3 分别表示城市人群、镇人群、乡村人群。

i：表示个体的年龄，i=0、1、2…90，其中 i=0 表示婴儿，i=90 表示年龄等于或大于 90 岁。

N(t):表示 t 年人口总数。

$F^j(t)$：表示 t 年 j 人群中女性总数，$M^j(t)$：表示 t 年 j 人群中男性总数。

$F_i^j(t)$：表示 t 年 j 人群中 i 岁女性总数，$M_i^j(t)$：表示 t 年 j 人群中 i 岁男性总数

$N^j(t)$:表示 t 年 j 人群的人口总数，即$N^j(t) = F^j(t) + M^j(t)$。

$DM_i^j(t)$、$DF_i^j(t)$：分别表示 t 年 j 人群中 i 岁男性、女性的死亡率。

$B_i^j(t)$：表示 t 年 j 人群中 i 岁妇女的生育率。

$\beta^j(t)$：表示 t 年 j 人群的总和生育率，即。$\beta^j(t) = \sum_{i=1}^{90} b_i^j(t)$

h_i^j：表示 i 岁妇女的生育率占总和生育率的比例，即生育模式。

$H^j(x)$：表示连续情况下 j 人群的生育模式分布函数，

$C^j(t)$：表示 t 年 j 人群出生婴儿男女性别比（女性以 100 为基数）。

$RM_i^{(j_1,j_2)}(t)$：表示 t 年由j_1人群迁入j_2人群的 i 岁男性人口数目。

$RF_i^{(j_1,j_2)}(t)$：表示 t 年由j_1人群迁入j_2人群的 i 岁女性人口数目。

$P_i^{(j_1,j_2)}(t)$: 表示 t 年由j_1人群迁入j_2人群的 i 岁人数占j_1人群 i 岁人口总数的比例。

5. 模型建立

5.1 模型准备

1) 死亡率的分析与预测

三个人群各年龄的男，女死亡率可由人口调查数据获得。考虑到 0 岁人口死亡率（婴儿死亡率）与 1 岁以后的死亡率有很大的区别，故分别对其进行研究。

i. 婴儿死亡率：

从人口调查数据中可以看出，城、镇、乡的婴儿死亡率有较大的差别，这是因为三类人群所处环境的医疗条件，环境设施等有所不同。城、镇男女婴的死亡率，在 2001-2005 年内在一个范围内有较大波动，但无显著变化趋势，为了排除随机因素和统计误差所造成的影响，对城镇男女婴分别取 2001-2005 年间的均值作为 06 年后的男女婴死亡率的预测值（如表格 1 所示）。

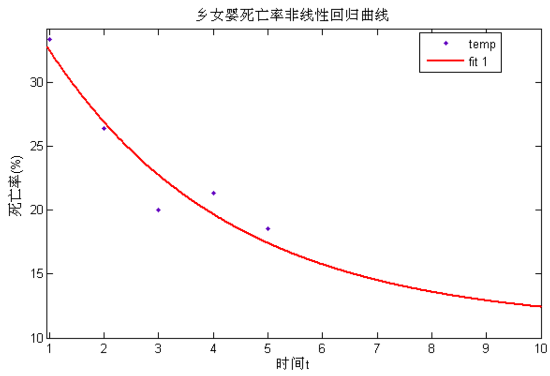
表格 1

性别	城市	镇
男性婴儿	6.116	6.4
女性婴儿	7.78	11

而在乡村，男女婴的死亡率在 2001-2005 年间呈现出递减的趋势，体现了乡村医疗条件的逐步改善。但是，考虑到死亡率不可能减至负值，也不大可能降至低于城，镇的平均死亡率，采用负指数函数的对数据进行回归分析，并对 2006 年以后的婴儿死亡率进行预测。

考虑乡村女婴死亡率的回归方程为 $DF_0^3(t) = \alpha \times e^{-\beta t} + c$，取c=11，为镇女婴的平均死亡率预测值。使用 matlab 软件进行拟合得 α =29.08(17.11, 41.06) β =0.3019 (0.1198, 0.4841) ，决定系数R^2=0.9144 均方根误差 2.041

同理乡村男婴死亡率的回归方程 $DM_0^3(t) = \alpha \times e^{-\beta t} + 6.4$ ，决定系数 R^2 =0.8677, 均方根误差 1.574 α =20.69 (13.53, 27.86) β =0.1657 (0.03791, 0.2934) 。



图表 1

ii. 其余各年龄段的死亡率：

由于自然原因，各年龄段人口的死亡率有差别。在离散的情况下 ，各年龄段对应了一个死亡率，因此可得到年龄区间为[1,90]上的一个死亡率分布函数，通过死亡率分布函数对未来时间段上的死亡率进行预测，可以避免统计数据的随机误差造成的影响。人均寿命与死亡率有着密切的关系，现从人均寿命的变化规律出发，研究城镇乡死亡率分布规律及其演变情况。

根据人口统计理论，t 年人均寿命定义为：在 t 年新生的婴儿，以 t 年的各年龄人口

死亡率作为其未来死亡率的估计值，则该新生儿的寿命期望就为该年的人均寿命，记人均寿命为 L。

人均寿命的计算公式：$L=\sum_{i=1}^{\infty} i\times P(i)$，其中$P(i)$为个体活至 i 岁的概率，$P(i)=\prod_{n=1}^{i-1}(1-D(n))\times D(i)$，D(i)为 i 岁的死亡率。考虑到 90 岁以上人口非常稀少，这里近似取$L=\sum_{i=1}^{90} i\times P(i)$作为人口寿命的近似估计。

由上述公式结合人口调查数据，可以计算出 2001-2005 年各人群男女的人均寿命，结果如表格 2 所示：

表格 2

年份	城市男性平均寿	城市女性平均寿	镇男性平均寿命	镇女性平均寿命	乡村男性平均寿	乡村女性平均寿
2001	74.2632	78.6713	72.7815	77.4768	68.8309	72.6084
2002	74.6667	79.0929	72.7821	77.1909	68.2749	72.594
2003	75.8137	79.0269	73.4754	76.8769	69.4739	73.3343
2004	76.1876	79.9876	74.1359	77.9995	69.9025	74.0472
2005	75.8986	79.9113	73.5455	76.668	70.5267	74.9918

由上表可以看出，人均寿命在 2001-2005 年间基本成递增趋势，但变化幅度很小。可以认为在中短期内人均寿命不会发生大的变化，则在中短期内死亡率的分布函数也认为不发生变化。而城，乡，镇之间，男性与女性之间却有显著的差异，需要对三类人群的男，女死亡率分别来考虑。

以城镇男性死亡率分布为例：

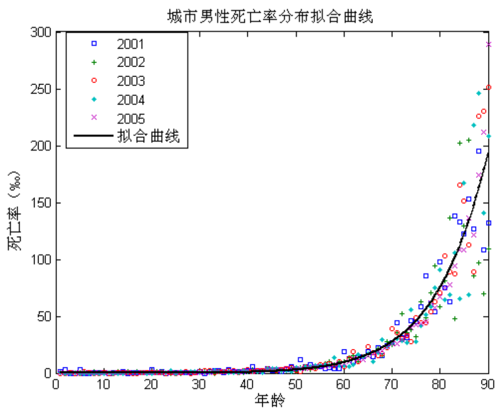
作出 2001-2005 年城镇男性死亡率$DM_i^1(t)$（t=1..5　i=1..90）散点图如图表 2 所示，死亡率分布函数为

$$f(x)=\alpha\times e^{-\beta\times t}+\gamma$$

其中常数项γ表示由于伤害，病痛等意外原因造成的人口死亡，与年龄无关；指数项随着年龄的增长而增长，表示年龄增长对死亡率的影响。

进行数据拟合，得α=0.02936，β= 0.09679 ，γ= 0.001957 ， R-square: 0.8796

同样分别对城、乡、镇，用 01-05 年的数据进行曲线拟合。拟合函数仍由f(x)决定，各参数拟合后如表格 3 所示。



图表 2

表格 3

	城市男性	城市女性	镇男性	镇女性	乡男性	乡女性
α	1.356e-005	6.402e-006	1.848e-005	4.069e-006	2.047e-005	6.097e-006
β	1.737	1.775	1.724	1.845	1.729	1.825
R^2	0.8796	0.8564	0.9022	0.8393	0.9403	0.9210

2) 妇女生育率$B_i^j(t)$和总和生育率$\beta^j(t)$的分析与预测

描述人口生育水平的指标通常有两种：育龄妇女生育率，总和生育率。较前者，总和生育率更能反映妇女的真实生育水平[1]（我们后面的分析得到了与此相同的结论），当然妇女生育率在人口预测方面也有其优势。

设某年 j 人群中 i 岁妇女数为F_i^j，i 岁妇女生育率为B_i^j，生育模式为h_i^j。则 j 人群的育龄妇女生育率定义为该年生育的妇女数与育龄妇女总数之比，即

$$B^j = \frac{\sum_{i_1}^{i_2}(F_i^j \times B_i^j)}{\sum_{i_1}^{i_2} F_i^j} \qquad (1)$$

i_2、i_1分表示育龄妇女的年龄上下限。

式(1)也可写为 $$B^j = \frac{\sum_{i_1}^{i_2}(F_i^j \times B_i^j)/F^j}{\sum_{i_1}^{i_2} F_i^j/F^j} = \frac{\sum_{i_1}^{i_2}(w_i^j \times B_i^j)}{\sum_{i_1}^{i_2} w_i^j} \qquad (2)$$

F^j表示 j 人群妇女总数，$F^j = \sum_1^{90} F_i^j$，w_i^j表示 i 岁育龄妇女占妇女总数的比例。

记总和生育率$\beta^j = \sum_{i_1}^{i_2} B_i^j$，生育模式$h_i^j = \frac{B_i^j}{\beta}$。生育模式描述了各年龄妇女生育率占总的妇女生育率的比例。一般情况下，生育模式在相当长的一段时期内变化缓慢，可以认为是恒定的值，因此对于第 t 年$B_i^j(t) = h_i^j \times \beta(t)$，即妇女生育率可由总和生育率和生育模式决定。其中，$\beta(t)$的值可以通过改变人口政策来进行控制，这为我们改变未来人口形势提供了一种可行的方法。

对$\beta(t)$的分析：国家人口计生委根据人口普查漏报比例以及教育卫生等部门的统计资料,于 2001 年提出了调整后的总和生育率水平为 1.8，且这个值将在未来一段时间内保持不变。按照上述公式，结合 2001-2005 年城、镇、乡妇女生育率可求出各年各人群的总和生育率，其结果如表格 4 所示：

表格 4

年份	城市总和生育率	镇总和生育率	乡村总和生育率
2001	1.00208	1.18864	1.60399
2002	0.96053	1.20339	1.65267
2003	0.9521	1.317	1.6769
2004	1.04831	1.34744	1.68698
2005	0.92648	1.27797	1.65371
均值	0.9779	1.266888	1.65485

由上表可以看出我国总和生育率已经达到一个比较稳定的状态，但是城、镇、乡之间还是存在较大的差异。然而由上表知由统计得到的数据显示我国总和生育率没有超过 1.7（比计生委提出的值要小），鉴于总和生育率对预测结果的影响很大，而且国内专家围绕总和生育率也由颇多的争议，所以在此我们完全采信由题目附录所给数据计算出来的结果。又由于我们的模型是对三个人群分开来考虑，故只需考虑上表中的均值，做为对未来$\beta(t)$的预测。

由 2001-2005 年育龄妇女生育统计数据可以看出，各人群的育龄妇女生育率均呈下降趋势，而总和生育率却保持不变，造成这一现象的原因是妇女的年龄分布发生了变化。由生育模式的分布可以看出，24-26 岁是生育的高峰期，20-29 岁是生育的旺盛期。在 2001 年时，28-32 岁妇女所占比重较大，即生育旺盛期妇女所占比重较大，随着时间的推移，这部分妇女逐渐退出生育旺盛期，而进入生育旺盛期的年轻妇女所占比重又较小小，导致了育龄妇女生育率的显著下降(见(3)式)。

以上分析可以看出，由于受到妇女年龄分布的影响，育龄妇女生育率并不能完全的反映各年龄段人口的变化趋势，事实上也没有一个稳定的变化趋势，而总和生育率却能在相当长的一段时间内保持不变，表示一个妇女在其一生中可能生育的子女数，由此更能放映人口变化规律的本质。因此，我们采用总和生育率及人口生育模式两个指标对未来人口发展进行预测。

假设在相当长一段时间内，生育模式不会发生改变。为排除统计数据的随机误差的影响，使用 2001—2005 年数据求出连续形式的生育模式曲线$H^j(x)$。

对 j=1,2,3；t=1..5 生育模式可由下式给出。

$$h_i^j(t) = \frac{B_i^j(t)}{\sum_{15}^{49} B_i^j(t)}$$

即为生育率的归一化值，分别对城，乡，镇，使用 5 年的数据对$h_i^j(t)$进行拟合。在人口统计学中，最常用的一种表示生育模式的函数为[2]

$$H(i) = \frac{(i - i_1)^{\alpha-1} e^{-\frac{i-i_1}{\theta}}}{\theta^{\alpha}\Gamma(\alpha)}$$

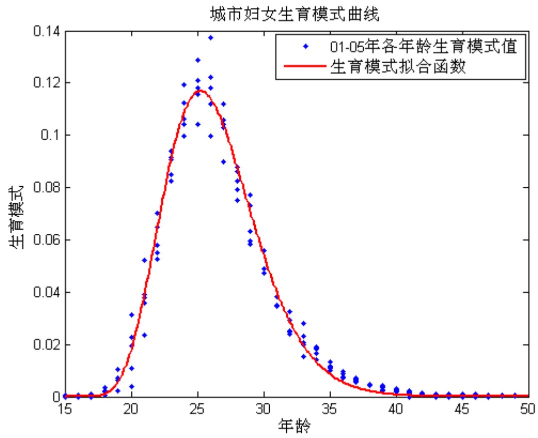
这里借用了概率论中的Γ分布函数。对于各人群进行函数拟合。

城镇妇女生育模式分布函数如图所示:

对其余数据进行拟和，参数如表格 5 所示：

表格 5

	城市妇女	镇总妇女	乡村妇女
α	10.16	8.151	6.519
θ	1.117	1.277	1.593
R-square	0.9828	0.9652	0.9585



图表 3

3) 出生婴儿男女性别比$C^j(t)$的预测

中国的出生婴儿男女性别比超出正常水平，而且城、镇、乡的情况有所不同，但是总的来说，城、镇在 1994—2005 时段内没有明显的变化趋势，仅有小幅的上下浮动，而乡村出生婴儿男女性别比$C^3(t)$在近几年呈逐年递增的趋势。考虑到国家已经采取了一定的措施来遏制这种现象，且$C^3(t)$也不可能持续地增高。所以对于$C^1(t)$、$C^2(t)$均可取 2001-2005 年间的均值，对于$C^3(t)$可采取二次拟合来预测 2006 年后的值。拟合函数为$C^3(t) = ax^2 + bx + c$，拟合参数为 a=-0.03497 (-0.1232, 0.05326)，b=0.8902(-0.2879, 2.068)，c=115.8 (112.4, 119.1)。预计到 t=10 时，即在 2010 年，性别比为 120.8，与 2005 年相比已有所下降，性别比持续增高势头得到抑制，2020 年降至 114.3，虽然与 2020 年性别比恢复正常的战略目标有所差距，但仍符合一定的实际情况。$C^1(t)$= 111.8291667，$C^2(t)$=117.9316667。

4) 人口迁移率的分析与确定

自改革开放以来，我国有大规模的人口流动。下表是 2001-2005 年我国城、镇、乡人口统计数据：

	全国总人口	城市人口	城人口所占比	镇人口	镇人口所占比	乡人口	乡人口所占比
2001	127627	30886	0.242	16553	0.1297	80201	0.6284
2002	128453	33603	0.2616	16108	0.1254	78729	0.6129
2003	129227	33612	0.2601	19668	0.1522	75934	0.5876
2004	129988	33563	0.2582	19966	0.1536	76459	0.5882
2005	130756	36246	0.2772	22385	0.1712	72125	0.5516

从表中可以看出全国人口总数缓慢增加，而城，镇人口总数增加较快，所占比例也逐年增加，乡村人口逐年递减，所占比例逐年减少，且所占比例不断减小，排除自然增长率和死亡率的影响外，导致这一现象的主要因素是城、镇、乡之间的机械迁徙与人口流动。

由于城、乡、镇之间生育模式，生育观念和政策等的差别，只有对城、镇、乡分开考虑，同时对相互间的人口流动进行合理的估计，才可能准确地预测我国人口的发展趋势。

考虑到中国的特殊情况，迁移的人口主要分为两部分，一是受到农村人口城市化的影响和农村人口进入到城、镇，成为固定户口的迁徙，另一部分是受到青壮年人口进城务工求学等影响，形成的大规模人口流动。由于高龄人口迁移率很小及所占人口比例不大，在模型中不考虑 80 岁以上高龄人口的迁徙。

下面通过迁移率及迁移率——年龄函数对各类人群间的流动情况进行分析。

设在 t 年，某类人群总数为 Q(t)，其中 i 岁人数为$Q_i(t)$，定义迁移率函数为：

$$s_i(t)=\frac{M_i(t)}{Q_i(t)}$$

$M_i(t)$为 t 年流出该人群的 i 年龄人口数量，由此可见，迁移率是与年龄有关的，表示 i 年龄迁移人数占当前 i 年龄总人数的百分比。若$s_i(t)>0$，则表示为迁出率，若$s_i(t)<0$，则表示为迁入率。

通过 2001-2005 年的原始数据，分别对城，镇，乡男性，女性的迁移率进行分析。

以乡村女性为例，在 t 年 i 年龄的人口数为 $F_i^3(t)$，假设不考虑迁徙，则 t+1 年人口数量仅仅由 t 年的人口数量及其死亡率决定，设为${\tilde{F}_i}^3(t)$。

则${\tilde{F}_{i+1}}^3(t+1)=F_i^3(t)\times(1-DF_i^3(t))$，$DF_i^3(t)$为死亡率。

若 t+1 年实际人口数量$F_i^3(t+1)$已知，则可算得 t 年的人口迁移率为

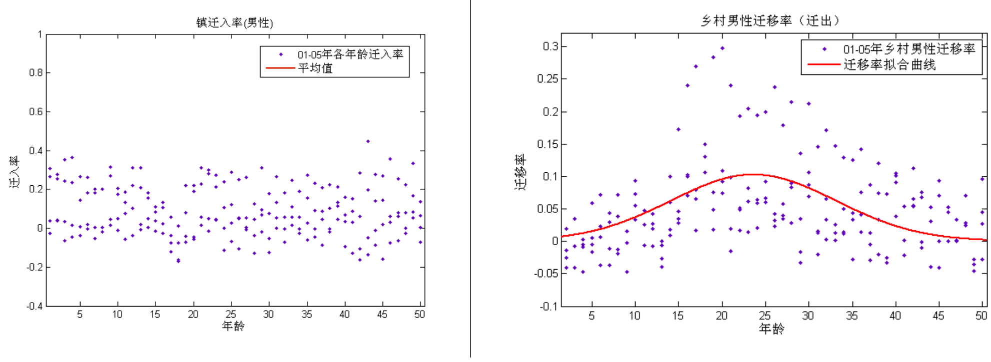
$$s_i(t)=\frac{{\tilde{F}_{i+1}}^3(t+1)-F_i^3(t)}{F_i^3(t)}$$

经资料查询得到 2001—2005 年全国总人口数估计值，设第 t 年人口数为 N(t)，由 5 年的 1%人口统计资料得到第 t 年乡村人口总数占全国总人口数的百分比$\alpha(t)$， 则 t 年年龄为 i 的妇女数量为：$F_i^3(t)=N(t)\alpha(t)\omega_i^3(t)$，其中，$\omega_i^3(t)$为 t 年 i 岁乡村妇女占总乡

村人数的百分比，可由原始资料直接得到。

因此，可由上获得乡村女性各年的迁移率，同理，可以求得城，镇，乡各年的男，女迁移率。

5 年内镇男性迁入率，乡男性的 迁出率——年龄 散点图如图表 4 所示：



图表 4

以上两图体现了镇与乡之间人口迁移情况的较大区别。

可以看出镇迁入率与年龄无关，保持在一个稳定值附近。进一步分析可知，镇迁入率主要是受农村搬迁率，固定户口迁徙等影响，而青壮年外出务工对其造成影响不大。还可看出，镇迁入率其 5 年内均值大于 0，迁入人数大于迁出人数，从 1%人口统计也可看出，镇人口所占比例不断在增长。

而在乡村男性迁出率——年龄图中，在 15-33 岁之间的迁出率出现峰值。与镇迁移的情况有很大的不同，乡村的迁移率主要受到了劳动力人口外出务工的影响，在劳动力旺盛的年龄段内迁出率明显增加，在其他年龄段渐趋平缓。

采用高斯函数对乡男性的迁移率进行拟合，得到乡男性迁移率——年龄分布函数：

函数定义为 $f(x) = ae^{-\frac{(x-b)^2}{c^2}} + \mu$

其中常数部分μ称为固定迁移率，表示与年龄无关的迁移，如户口迁徙，农村城市化等，这部分迁移人口部分迁移出镇，部分迁移入城市。μ可由城市化增长率近似确定，由 5 年 1%统计数据计算出城市化增长率为 0.0192，取μ=0.02。

$ae^{-\frac{(x-b)^2}{c^2}}$为指数部分，表示在 20-30 岁之间的峰值突起部分，主要由于民工流等青壮年劳动人口流入城市导致。模型中合理地假定这部分人口只迁移入城市，且不考虑镇居民向城市方向的迁移。

同样拟合得到乡村女性迁移率——年龄分布函数，如表格 6 所示：

表格 6

	a	b	c
乡村男性	0.09445	22.11	8.731
乡村女性	0.07418	20.34	7.685

在没有发生重大事件的情况下，迁移率函数随时间变化较小，可近似将迁移率函数在 i 点的值做为未来时间段年龄 i 人群迁移率的预测。

5.2 中短期预测模型建立

从人口统计数据可以看出，不管是在死亡率上（包括婴儿死亡率和自然死亡率）还是在人口流动百分比方面，男性和女性之间都有很大的差别，故在模型中，将男性与女

性分开来考虑。先假定不存在人口流动带来的影响，各人群中人口发展的机理完全相同，故可等同考虑，后面若没有特别说明，均由 j 代指某一特定人群。

t+1 年 j 人群婴儿总数是育龄妇女生育子女的总数，妇女生育率$B_i^j(t) = H_i^j(t) \times \beta(t)$, $\beta(t)$为总和生育率。将男性婴儿和女性婴儿分开来考虑，则有

男性婴儿数 $$M_0^j(t+1) = \frac{C^j(t)}{C^j(t)+100} \times \sum_{i_1}^{i_2} B_i^j(t) \times M_i^j(t) \quad (3)$$

女性婴儿数 $$F_0^j(t+1) = \frac{100}{C^j(t)+100} \times \sum_{i_1}^{i_2} B_i^j(t) \times F_i^j(t) \quad (4)$$

t+1 年 i+1 岁人口总数是 t 年 i 岁人口中存活下来的数量，鉴于男女死亡率的有所不同，此处仍将男女人口分开来考虑

男性人口数 $$M_{i+1}^j(t+1) = (1 - DM_i^j(t)) \times M_i^j(t+1) \quad (5)$$

女性人口数 $$F_{i+1}^j(t+1) = (1 - DF_i^j(t)) \times F_i^j(t+1) \quad (6)$$

由(3)、(4)、(5)、(6)式可以得到男女的人口转移矩阵，现将女性的人口转移矩阵$TF^j(t)$表示如下：

$$\begin{bmatrix} \frac{100}{C^j(t)+100} \times B_0^j(t) & & \cdots & \frac{100}{C^j(t)+100} \times B_{89}^j(t) & \frac{100}{C^j(t)+100} \times B_{90}^j(t) \\ 1 - DF_0^j(t) & 0 & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 \quad 0 & & \cdots & 1 - DF_{88}^j(t) \quad 0 & 0 \\ 0 \quad 0 & & & 0 \quad 1 - DF_{89}^j(t) & 0 \end{bmatrix}$$

由此 $$\vec{F}^j(t+1) = TF^j(t) \times \vec{F}^j(t)$$

此处$\vec{F}^j(t)$ 为女性人口分布向量，其各分量之和等于$F^j(t)$。

同理可以到男性人口转移矩阵$TM^j(t)$和男性人口分布向量，由于与女性的区别不大，在此就不再列出。

考虑人口流动，由上面的分析知，只有乡村人群向镇人群和城市人群有非常显著的人口流动，用$RF_i^{(3,2)}(t)$、$RF_i^{(3,1)}(t)$分别表示 t 年由乡村人群流向镇人群和城市人群中 i 岁的女性人口数，以$RF_i^{(3,1)}(t)$为研究对象。

$$RF_i^{(3,1)}(t) = PF_i^{(3,1)}(t) \times F_i^3(t) \qquad (PF_i^{(3,1)}(t)\text{是人口迁入率})$$

其向量的表示形式为$\overrightarrow{RF}^{(1,3)}(t) = \overrightarrow{PF}^{(1,3)}(t) \cdot \vec{F}^3(t)$

同理可以得到男性人口流动向量，在此不再累赘。

综和考虑人口的自然演变和人口流动，三个人群的人口变化可由下面方程表示（只考虑女性的，男性的可同理给出）：

$$
\begin{cases}
\vec{F}^1(t+1) = TF^1(t) \times \vec{F}^1(t) + \overrightarrow{RF}^{(3,1)}(t) \\
\vec{F}^2(t+1) = TF^2(t) \times \vec{F}^2(t) + \overrightarrow{RF}^{(3,2)}(t) \\
\vec{F}^3(t+1) = TF^3(t) \times \vec{F}^3(t) - \overrightarrow{RF}^{(3,1)}(t) - \overrightarrow{RF}^{(3,2)}(t)
\end{cases} \tag{7}
$$

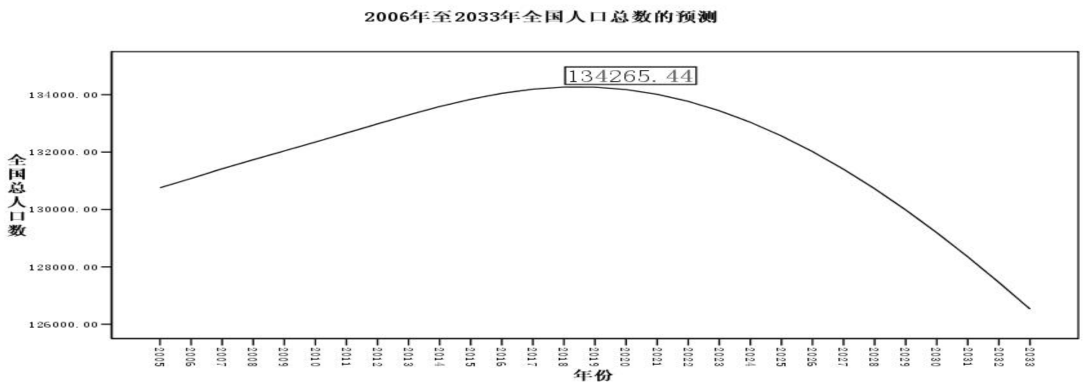
5.3 中短期预测模型求解

对以上模型使用 matlab 编程，通过数值迭代可求得结果。进一步可以求出在未来 t 时刻内城，镇，乡人口总数，各年龄人口数及所占比例。并基于这些数据，对我国未来人口的发展趋势做出预测。

5.4 中短期预测模型结果分析

在预测过程中，人口的演变与总和生育率β(t)有着密切的关系，β(t)的不同，将引起人口未来预测数据的明显变化，特别是随着时间的推移引发的长期效应更不可忽视。通过对 2001-2005 年原始数据的分析，得到 01-05 年城，镇，乡的平均β值分别等于 0.9779，1.2669，1.6549 的结论，均明显小于普遍认可的估计值 1.8。但是，β到底为多少目前仍然没有确切的定论，并且通过 β 调节值也是国家对人口发展情况的重要控制手段，在模型分析中，我们将在必要时对不同的β值分别进行分析。

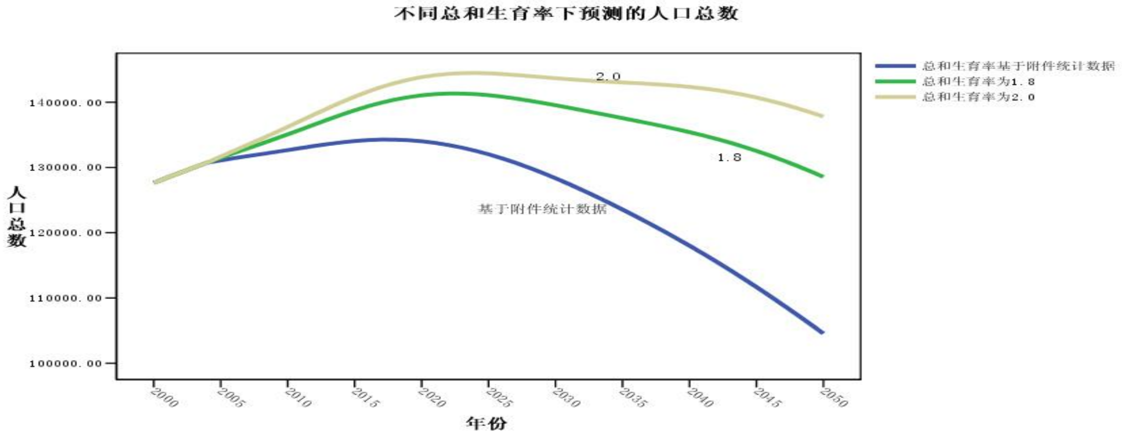
1) 人口总数的预测



图表 5

以原始数据所得的β值做为总和生育率对未来人口趋势进行预测，从图表 5 中可以看到，在未来一段时间内全国总人口仍将保持上升的势头，其原因是受 1980-1990 年之间的第三次生育高峰的影响，在 2005-2020 年生育旺盛期妇女数量将处于一个高峰，使人口数量有明显的增加。在 2018 年左右将达到人口峰值，净增加人口数为 3000 万，之后逐渐下降。

现分别取β为 1.8，2.0，并对人口总数进行预测如图表 6 所示：

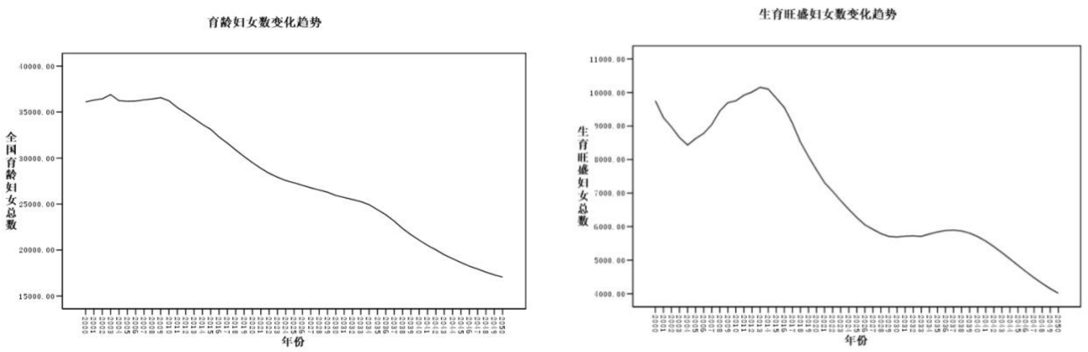


图表 6

从图表 6 中可看出，随着β值的增加，将使人口高峰到来的时间往后推移（如当β为 1.8 时，人口高峰出现在 2024 年左右，推迟了近 6 年），同时也会使最大人口总量增加。可见，β值不能过高，若β值过高，将导致我国人口在未来一段长时期内继续的增长，人口高峰期数量增大，而我国目前已是人口大国，不能承受人口的继续膨胀带来的各种危机。

2) 生育状况的预测 （β值基于原始数据统计）

绘制 2006-2050 年时间内育龄妇女总数，生育旺盛期妇女总数如图表 7 所示：

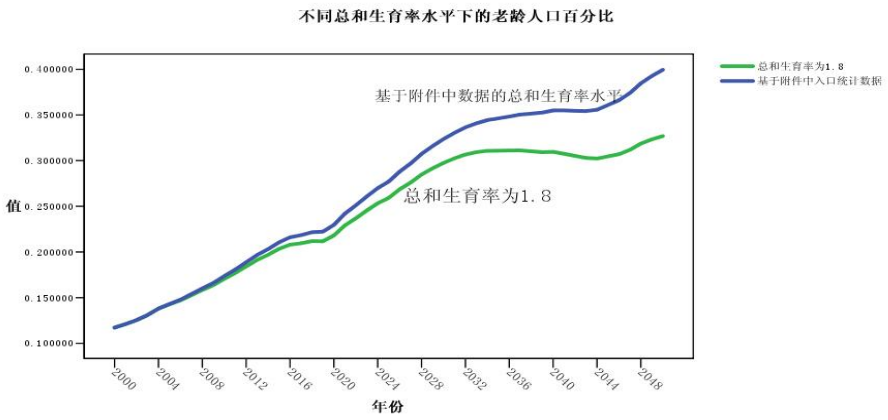


图表 7

不难看出，育龄妇女总数与生育旺盛其妇女总数成下降的趋势，但有所波动，且在 2009 年左右出现最大值。由所给资料分析可知该段妇女主要由第三次人口生育高峰造成的年龄推移产生的，随着该段妇女逐渐过渡出生育期，育龄妇女总数，生育旺盛期妇女总数将不断下降。

3) 年龄结构分布的预测（β值分别取原始统计数据，1.8）

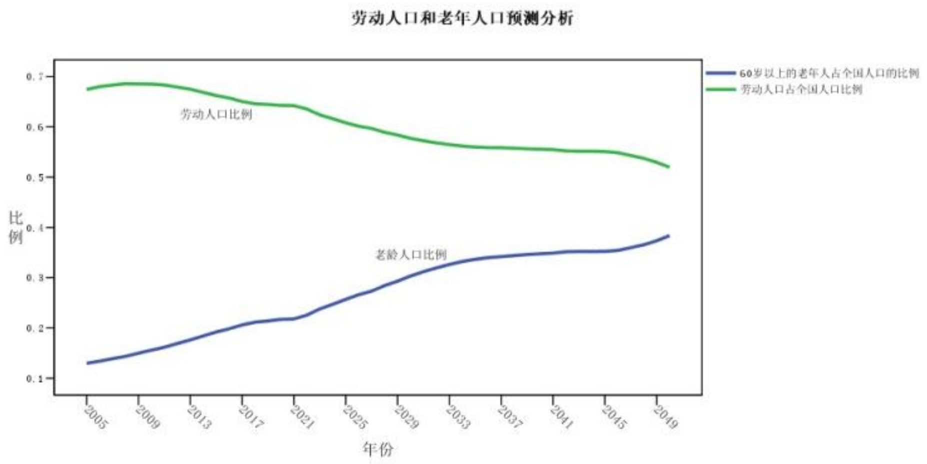
定义 60 岁以上为老年人口。对老年人在全国人口中所占的比率进行预测，由图表 8 所示，总体呈递增的趋势。总和生育率对未来特别是下世纪中叶的老龄化进程有显著的影响，基于附件中数据统计的总和生育率水平是低于 1.8 的。总和生育率低，虽然可以减缓人口总数的增长，降低峰值的人口数量，但是也会导致新生婴幼儿减少，在 2020 年以后随着大部分人口步入老龄阶段，社会老龄化现象将越加严重，使得老龄化指数升



图表 8

高。以原始数据统计的总和生育率水平计算得2050年的老龄人口百分比为39.95%；β = 1.8时计算得32.69%，相比有了显著的降低。由上分析可知，过小的β值会引起老龄化现象更加严重，不利于经济的发展。β值也不能过大，否则会造成人口增长大，高峰时期总人数过大等，也不利于稳定的发展。因此，β应稳定在一定之内比较合适（如1.8）。图表 9为劳动力人口百分比，老龄人口百分比(15-60岁)的变化趋势。老年人口所占比例将越来越大。

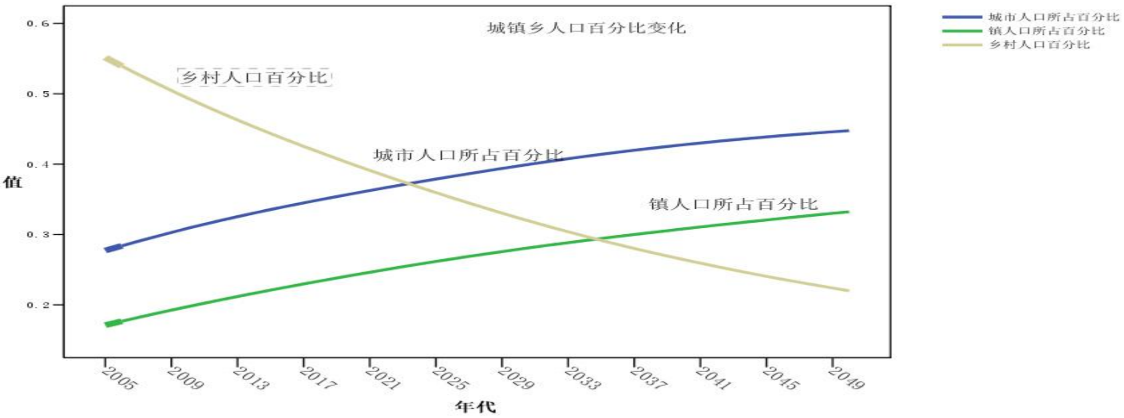
在下个世纪，随着时间的推移，社会老龄化现象将越来越严重，应引起政府及相关部门的重视。



图表 9

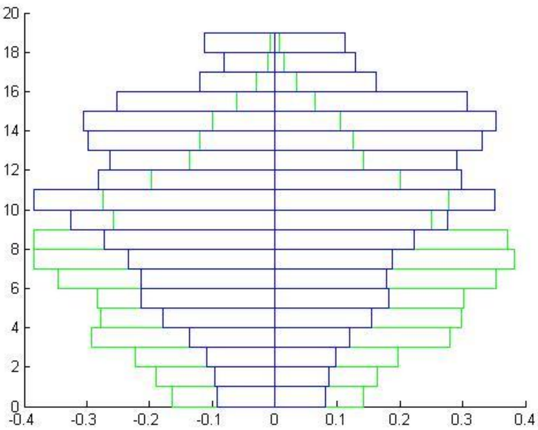
4) 城乡人口分布的预测

分析城镇乡占总人口百分比的变化趋势（图表 10）我们可以发现农村人人口百分比逐年下降，且其处于劳动年龄的人口数占其自身的比重也是逐年下降的，其原因为乡村人口迁往城镇，乡村年轻人大量涌入城市务工等。城市化水平的不断提高，也是国家经济发展的一种体现。在2050年，我国人口的城镇化率预计将达到62%。但是，随着年轻劳动力的不断流失，农村人口的老龄化现象也将越加严重。



图表 10

5) 人口结构的分析



图表 11

分别绘制出 2005 与 2040 年的人口结构的金字塔模型(左边为男性，右边为女性)，从图表 11 中明显的看到了变化的趋势。在 2005 年，人口的平均寿命较小，年轻人口占了更多的比率。而在 2040 年，老龄人口形成了一个庞大的集体，社会老龄化现象十分严重。

5.5 长期预测模型

在中短期的人口预测模型中，我们假设 2005 年之后城，镇，乡的男女死亡率分布和女性生育率与时间无关，在中短期预测中，时间跨度不大，这种假设是合理的。但从长期来看，总和生育率基本保持稳定。而人均寿命逐年增加，因此死亡率应成下降的趋势。

对于新生婴儿性别比，在短期预测模型中，我们认为城，镇的出生性别比保持不变，而乡村的出生性别比采用二次函数拟合。在不长的时间内这样的假设是合理的，而在长期预测模型中则需要对其进行另外的推导。

在考虑迁移率时，由于我国的城市化是一直在进行的，流动人口依旧是由农村迁往城镇，具体流动方式和趋势长时间内并不会发生改变。假设在长期预测模型中迁移率分布函数形式保持不变。

由此我们在长期人口预测中着重考虑死亡率随时间按变化趋势，以及婴幼儿出生性别比的变化。

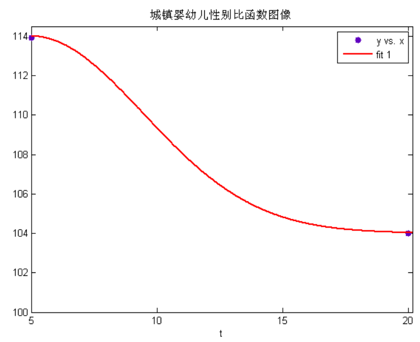
1) 婴幼儿出生性别比的预测

由原始数据，现阶段城，镇，乡性别比均处于偏高的水平，但在长时间内，都不可能保持不变，主要是受到国家调控政策的影响。考虑在国家的战略目标是在 2020 年使性别比达到正常水平，即 102 到 107 之间。

分别以城，乡，镇 2005 年性别比为初始数据，采用高斯型函数描述性别比随时间的变化规律，设 2005 年性别比为 a，2020 年性别比为 104，则函数确定为

$$f(x) = (a-104) * e^{-\frac{(x-5)^2}{40}} + 104$$

以城镇出生性别比为例，其函数图像如图表 12 所示。



图表 12

2) 在长期预测模型中的死亡率的分析与预测

考虑人均寿命随时间是逐年增加的，因此我们在长期预测中不能忽略人均寿命增加对各年龄死亡率的影响，为此我们定义修正因子$\alpha_t(\alpha_t < 1)$对 t 时刻的死亡率分布函数进行修正。这样我们针对长期预测型求出各时刻点的死亡率分布函数。

设人均寿命每年增加量相同为γ（岁/年）。查阅相关资料[3]$\gamma = 0.1$，假定初始时刻t_0的死亡率随年龄的分布函数已知为 f（x, t_0）,其人均寿命为 L（t_0），则$L(t_0)$可由 f（x, t_0）求出。设下一时刻t_0+1的死亡率随年龄的分布函数为 f(x, t_0+1)=$\alpha_{t_0+1}\cdot$ f（x, t_0），人均寿命为 L（t_0+1)，其中α_{t_0+1} 的取值满足$L(t_0+1)-L(t_0)=\gamma$;

由于γ已知，给定初始时刻分布函数 f（x, t_0），即可求得α_{t_0+1}的值，进而确定 f(x, t_0+1)，以及未来任一时刻的分布函数 f(x, t_0+n) n=1,2...

对于城，镇，乡的男性和女性，分别采用 5 年统计数据中各年龄点死亡率进行初始时刻分布函数的拟合。

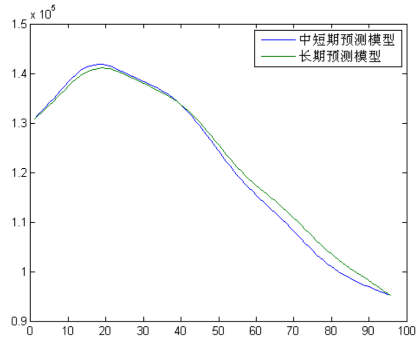
求得第 t 年的死亡率分布函数后，对其离散化，做为第 t 年各年龄死亡率的估计值。

5.6 长期预测模型求解：

采用以上方法对未来各项因素的进行估计，代入模型，便可求得未来城，镇，乡各年龄男女的数量。

5.7 长期预测模型结果分析

使用中短期模型和长期模型对未来 100 年内的预测进行比较（$\beta = 1.8$），考虑，见图表 13 可以看出，在中短期（30 年）内，两个模型预测值差别不大。这是因为在长期模型中考虑到的死亡率的变化，婴幼儿死亡率的减小，新生婴儿性别比的改善等在短期



图表 13

内其效应并不能得到充分的体现。同时，中短期模型部分参数的简化也是建立在合理的推断上的，有着较好的预测能力。

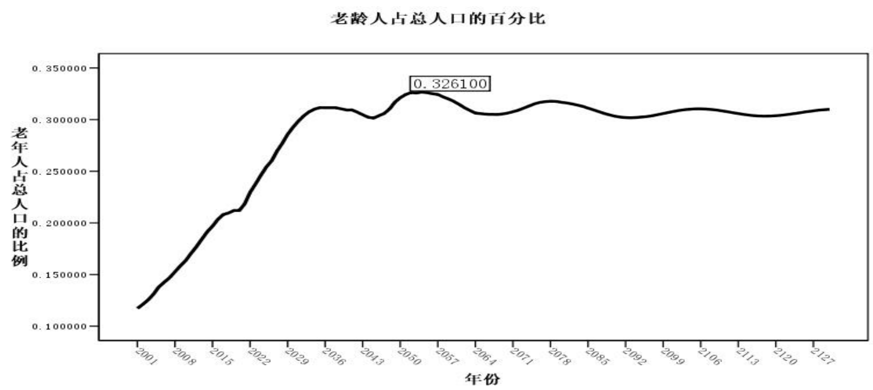
随着时间的推移，长期模型中考虑的因素其累计效应将得到充分的体现，与中段期模型有了明显的差别，由上图看出，在 21 实际后半叶，长期模型的预测值比短期模型偏高，究其原因是人均寿命的增加，婴幼儿死亡率的降低等因素随着时间的推移产生了累计效应，对预测值产生了较为明显的影响。

使用长期预测模型，主要对 2050 年后的各项数据预测分析，主要结论如下：

a) 城镇化率的增加速率逐渐变缓，在 2100 年城镇化率达 90%以上。
b) 在 2024 年左右人口总数达到峰值 14.2 亿，在 2050 年以后人口总数持续降低，在 2100 年降至越 9.5 亿人左右。具体人口数据由下表所示：

时间	2005	2010	2015	2020	2025	2030	2035	2040	2045
人口总数（万）	130756	134384.1697	138117.5629	140631.3706	140894.5617	139443.3736	137575.9568	135659.3826	133038.0366
时间	2055	2060	2065	2070	2075	2080	2085	2090	2095
人口总数（万）	124578.2539	120224.5756	116720.6286	113568.7782	110083.3515	106304.4571	102893.4493	100129.8518	97623.193

c) 在 2100 年以后，男女比例接近与 1:1，恢复至正常水平。
d) 人口老龄化问题依然十分严重：



图表 14

图表 14 为在 2005-2120 时间范围，对老龄人口（60～90）百分比进行预测的结果。可见，从 2001 起至 2045 年左右，老年人人口及其所占百分比都将保持很高的增长速度，在 2050 年左右达到一个稳定值，在 21 世纪下半叶，将一直稳定在 30%左右。

6. 模型优缺点分析

6.1. 模型优点：

1) 本模型通过对人口统计数据处理，分析各因素对未来人口变化影响的机理，结合统计分析与机理分析的方式建立相应的差分方程模型，对人口演变进行相应刻画与描述，

它能够较好地反映人口各项统计数据的变化，可信度较高。
2) 模型中尽可能多的考虑到影响未来人口变化的因素，针对中短期和长期预测的不同分别确定影响未来人口变化的因素的变化规律，分别找出适宜不同预测方式的参数模型，由此能够较好地进行中短期和长期模拟人口的演变规律，为未来人口增长的预测提供了令人信服的结果。
3) 由于所给数据有限且波动大，为此在在构造 Leslie 矩阵时我们综合已有数据合理的处理和简化各因素对 Leslie 矩阵作用方式，达到更加符合实际的目的。使得模型预测的结果有较好的精度前体下不失简单性。
4) 多因素考虑人口预测对于惯性比较大的这一社会系统来说，其实际上是动态的模拟该复杂人口系统的演变过程，并结合已有数据显现出来的因素变化趋势最终实现人口预测。

6.2. 模型缺点：

1) 模型在分析各因素的影响时仅依赖 2001-2005 年的人口统计的抽样数据，不考虑数据来源的真实性以及统计的准确性等其他因素影响，当预测的时间跨度较长时，仍可能因为情况变化较大而导致预测结果不准确。
2) 为了简化模型，在分析城镇乡三者之间的人口迁移时我们没有考虑镇向城市之间，镇向乡方向，城向镇，城向乡方向的人口迁徙。
3) 在考虑生育率与死亡率按年龄分布的时间序列时我们仅仅是以长度为 4 的时间序列来预测未来时间的生育率与死亡率年龄分布，且其与时间值无关。这对于长期人口预测来说具有它的不足之处的。
4) 人口出生性别比，总和生育率受国家政策影响程度明显，而且短期内国家不可能放任其按目前趋势发展下去，因此在预测未来人口变化时无法确确的考虑这一因素的影响。
5) 在长期预测中我们简单假定人口平均寿命年递增值保持恒定，这与实际人口寿命的年变化有所不同。
6) 长期人口预测采用的模型相比中短期而言更加贴近实际情况，为此建立模型时我们不妨仅仅考虑长期的人口预测。

7. 模型拓展

生育率和死亡率预测是人口预测的中心问题。我们可以通过将生育率和死亡率作为概率处理来提供对结果的随机性描述，且考虑其随机性与时间的关系，为此我们将生育率和死亡率理解为随机性时变序列。 在查阅相关资料[4]后我们发现事实上 Lee,R.D.和 L.Carter[3]根据在生育和死亡率的变化中水平变化远比模式变化显著的特点,利用矩阵奇异值分解(SVD)将时变的按年龄生育率和死亡率即时间序列向量问题转化成了时间序列变量问题,进而利用标准的时间序列分析方法进行了建模及预测研究。这样我们可以利用已有的数据作出对生育率和死亡率时间序列的随机模型预测。将预测出的生育率和死亡率时间序列向量再利用我们已建立的模型实现预测未来人口变化。只要对生育率和死亡率的预侧随机时间序列这一变化方式能够更加准确的反应实际情况，那么利用这些预测数据得到的最终预测人口数量值更加接近实际情况。

预测结果表明，以独生子女为主要特征的现行生育政策现已基本完成它的历史使命，在经过了人为的加速人口转变之后，在未来的不久中国将有很长一段时间面临劳动力供给萎缩以及老龄人口迅速膨胀，这种趋势难以在短期内扭转。中国各地区未来面临的许多人口挑战也会大相径庭，中国老龄化进程在城乡间有很大的差异，由于过去几十年各地的计划生育执行力度大不相同，以及乡村年轻人涌入城市，这给国家的养老保障，经济发展长生深远影响。人口问题对于中国来说依旧非常严峻。

人口问题说到底是发展问题,提高人口质量与控制人口数量是相辅相成的。人口素质的提高是经济增长的一个重要源泉。世界银行研究表明,劳动力受教育平均时间每增加一年,国内生产总值就会增加 9%。随着我国现代化的不断推进和市场经济的飞速发展,迁移流动人口大量增加,农村剩余劳动力纷纷涌入城市,这给计划生育工作带来新的课题。今后,我国在逐步解决人口数量问题的同时,人口素质、人口结构、人口分布等问题日趋突出。我们必须摆脱惟人口数量观念,进一步稳定低生育水平,严格控制人口增长,全面提高人口素质,建立完善的优生优育体系和社会保障体系,逐步实现人口、资源、经济和社会长期协调发展。21 世纪,我国人口工作将任重而道远。

[1] 严浩，人口生育指标和预测参数的选择研究，宏观经济研究：2007 年第 7 期，38-44，社会发展

[2] 姜启源，数学模型（第三版），北京；高教出版社，2003 年

[3] 曾毅 李玲，21 世纪中国人口与经济发展　社会科学文献出版社 2006

[4] Lee R.D.and Carter L.Modling and Forcasting US Mortality:report of the 1990 meetings of the Population Assciation of America.Toronto,1990