2022 年西华师范大学数学建模校内赛

编 号 专 用 页

评阅编号（由主办方在评阅前进行编号）：

评阅记录（可供评阅时使用）：

评阅人						
评分						
备注						

参赛队员分工简要说明

队员 1 数学与信息学院 人工智能 2020 级 12 班 202011540901 陈曦	**1.** 提供问题一、二的建模思路 **2.** 使用 python 建立 BP 神经网络模型、GM(1,1)灰色模型模型 **3.** 问题一、二编程求解 **4.** 根据编程过程，对论文进行补充完善
队员 2 数学与信息学院 人工智能 2020 级 12 班 202008541219 肖智元	**1.** 收集影响总人口的变量数据，并进行数据清洗 **2.** 完成问题一、二的建模 **3.** 使用 matlab 建立 Leslie 模型 **4.** 问题三、四编程求解
队员 3 数学与信息学院 人工智能 2020 级 12 班 202008541228 朱昱霖	**1.** 分析比赛题目，选择题目；收集影响总人口的变量数据 **2.** 完成问题一、二的建模 **3.** 负责论文书写进行排版 **4.** 对四个问题的建模进行分析，并且提出相对应的改进建议

三胎政策对人口状况、经济及其影响的研究

摘 要：近年来，为改善中国人口结构、落实积极应对人口老龄化国家战略、保持中国人力资源禀赋优势，我国提出了三胎政策。本文主要根据《中国人口统计年鉴》上收集到的 1980 年到 2020 年部分数据, 在灰色预测的基础上, 引入 BP 神经网络模型, 建立了中国人口增长的 GM(1,1)和 BP 神经网络组合模型, 由此对中国人口增长的情况做出预测。并且通过 Leslie 模型和多项式回归模型预测二胎政策对人口结构和经济发展的影响。

对问题一，分别利用利用总人口、出生人口、性别比例、自然增长率、农村人口、城镇人口、GDP、结婚率共八列数据作为特征值，来影响中国每年出生人口的数量，使用灰色预测来预测 2021-2035 年的数据，将这些预测出来的变量（除去总人口，其作为预测变量）作为神经网络的输入变量，来进行网络搭建。最后使用 2021-2035 年灰色预测预测出来的总人口作为预测变量，来使用网络进行预测，得到神经网络对 2021 年到 2035 年的人口变化情况的预测值。

对问题二，为了提高预测精度，选择基于 GM (1.1) 模型与 BP 神经网络模型相结合的组合模型来进行人口预测，其次考虑到全面二胎政策带来的影响，将 2003-2020 年育龄妇女人数放到灰色预测里面进行预测，对 2003-2015 年一般三孩生育率进行拟合，得到了关于时间-生育率的一元函数。根据三胎生育意愿 12.2%以及 2021-2035 年一般三孩生育率，从而算出每年的新增人口数得到 2021-2035 年的育龄妇女人数，最终得到开放三胎后 2021-2035 年的总人口数。

对问题三，将女性年龄每一岁作为一个年龄段，共分为 91 个年龄段。结合所给数据以差分方程组的 Leslie 模型为基础，考虑不同地区、不同性别人口参数的差别等因素，按照地区和性别建立以时间和年龄为基本变量的中国人口增长模型，利用历史数据估计生育率、死亡率及人口迁移等参数，代入 Leslie 模型求解并进行预测。

对问题四，国内生产总值(GDP)是现代国民经济核算体系的核心指标，是衡量的经济重要指标。在三胎政策的影响下，GDP 总量仍在持续上升，但此时 GDP 增加主要依靠 15-64 岁劳动人口的 GDP 增加而增加，不因劳动人口的增加而增加。本文就 2010 年到 2020 年的 GDP 等相关统计数据，考虑劳动人口为主要因变量，利用多项式回归模型。利用 matlab 软件对函数拟合求出相应的参数，从而预测三胎政策对国家经济增长情况的影响。

关键词：三胎政策、BP 神经网络预测、灰色预测、Leslie 模型、人口预测

一、问题重述

2021 年 8 月 20 日，全国人大常委会会议表决通过了关于修改人口与计划生育法的决定，修改后的人口计生法规定，国家提倡适龄婚育、优生优育，一对夫妻可以生育三个子女。就积极应对人口老龄化、调整生育政策等问题，会议作出部署。将婚嫁、生育、养育、教育一体考虑，加强适婚青年婚恋观、家庭观教育引导，对婚嫁陋习、天价彩礼等不良社会风气进行治理，提高优生优育服务水平，发展普惠托育服务体系，推进教育公平与优质教育资源供给，降低家庭教育开支。

人口问题有着悠久的研究历史，也有不少经典的理论和模型。这些理论和模型都依赖生育模式、生育率、死亡率和性别比等多个因素。这些因素与政策及人的观念、社会文化习俗有着紧密的关系，后者又受社会经济发展水平的影响。国家三胎政策的开放引起了人们的热议。这一政策有人认为能够对现有经济状况和人口结构起到一定的优化作用，而另一些人则抱负面态度。

本文将针对如下四个问题进行研究：

1. 根据中国现有人口状况，预测若不考虑政策影响，中国人口从 2021 年到 2035 年的变化情况。
2. 若执行三胎政策，中国人口的变化情况又该如何？
3. 预测 2035 年时中国的人口结构情况。
4. 分析三胎政策对国家经济增长情况的影响。

二、问题分析

2.1. 问题一的分析

一个国家人口的准确预测，是制定相应宏观经济政策的重要依据，对我国经济的发展有着巨大的作用。预测是控制和规划的基础，预测的精度是控制和规划成功的前提，而选择预测的方法是提高预测精度的关键。传统的人口预测方法主要有逻辑方法，常微分方程方法和动态预测方法等。这些方法在人口预测领域起到了一定的作用，但采用这些方法时都要对数据进行模型假设。由于真实模型往往是非线性的，如果在一些简单的模型假设下就进行数据模拟，常常不能达到较好的模拟效果。

神经网络对复杂非线性系统具有曲线拟合能力，基于 BP 神经网络和 GM(1,1) 模型的组合模型进行动态预测。既利用灰色预测的需要数据资料少的优点，又吸收了 BP 神经网络容错能力，自适应能力强的优点。由于神经网络的功能之强大，型式之多样，若能将其它网络形式同灰色模型相结合，则有可能进一步提高预测精度。

2.2. 问题二的分析

将人口预测分为两部分，全面三胎政策下的人口 = 未开放三胎政策下的人口 + 后期新增三胎数。利用国家统计局给出的 1980-2020 年的人口数据，在灰色神经网络的作用下，合理预测出未开放政策下 2021-2035 年的人口总数。对于新增三胎数，在合理预测出每年育龄妇女的人数的基础上，参考三胎生育意愿，同时考虑育龄妇女三胎生育能力，即三胎生育率，从而算出政策开放后每年的新增三胎数进行累加计算，最终求得三胎政策下未来 15 年的总人口数。

2.3. 问题三的分析

中国是一个人口大国，随着经济的不断发展，生产力达到较高的水平，现在的问题已不是仅仅满足个人的需要，而是要考虑社会的需要。中国未富先老，对经济的发展产生很大的影响。当今面临的问题是如何很好的预测出未来人口的变化趋势，并根据趋势制定出合理可行的人口规划方案，以缓解人口对经济的影响。所以准确的预测出人口的发展趋势势在必行。

为了增加 Leslie 模型预测的准确度，将女性年龄每一岁作为一个年龄段，共分为 91 个年龄段。结合所给数据以差分方程组的 Leslie 模型为基础，考虑不同地区、不同性别人口参数的差别等因素，按照地区和性别建立以时间和年龄为基本变量的中国人口增长模型，利用历史数据估计生育率、死亡率及人口迁移等参数进行预测。

2.4. 问题四的分析

国内生产总值(GDP)是指在一定时期内(一个季度或一年)，一个国家或地区的经济中所生产出的全部最终产品和服务的价值，常被公认为衡量国家经济状况的最佳指标。三胎政策的推出会影响未来的人口结构，经济增长情况也必将随之出现变化。

对于问题四，已知三胎带来的人口结构变化，要求 GDP 的变化情况，显然是一个因变量与多个自变量之间地关系，即多元线性回归问题，因此建立多项式回归模型进行预测。

三、模型假设与约定

1. 在社会稳定的前提下，生育率和死亡率都较为稳定。
2. 不考虑内外的人口迁移对我国人口数量的影响。
3. 2022 年起实施三胎政策。
4. 政策全面开放满足条件家庭愿意生育。
5. 国家经济发展水平按当前态势稳步发展。

四、符号定义及说明

变量符号	含义说明
$x^l(i,k)$	k 年，i 岁(不满 i+1 岁)，性别 l(l=m 男性, l=w 女性)的人口数量
$d^l(i,k)$	死亡率(死亡人数占总人数的比例)
$s^l(i,k)$	1-$d_j^l(i,k)$～存活率
$b^l(i,k)$	生育率(k 年每位 i 岁女性平均生育婴儿数)
$a^l(k)$	婴儿性别比(男婴比例)

五、模型的建立与求解

5.1. 问题一的建模与求解

5.1.1. BP 人工神经网络模型简介

神经网络的基本组成单元是神经元。神经元的通用模型如图 1 所示，其中常用的激活函数有阈值函数、sigmoid 函数和双曲正切函数。

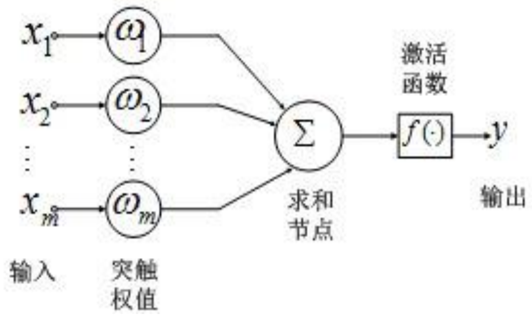

图 1 神经元模型

上面神经元的输出为:

$$y = f(\sum_{i=1}^{m} w_i x_i)$$

神经网络是将多个神经元按一定规则联结在一起而形成的网络，如图 2 所示。

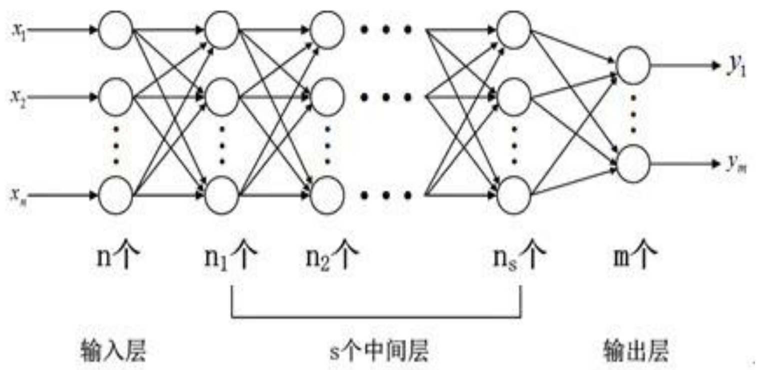

图 2 神经网络

BP 神经网络则是一种多层的前馈神经网络，其主要的特点是：信号是前向传播的，而误差是反向传播的。具体来说，对于如图 1 只含一个隐层的三层神经网络模型。

BP 神经网络的过程主要分为两个阶段，第一阶段是信号的前向传播，从输入层经过隐含层，最后到达输出层；第二阶段是误差的反向传播，从输出层到隐含层，最后到输入层，依次调节隐含层到输出层的权重和偏置，输入层到隐含层的权重和偏置。

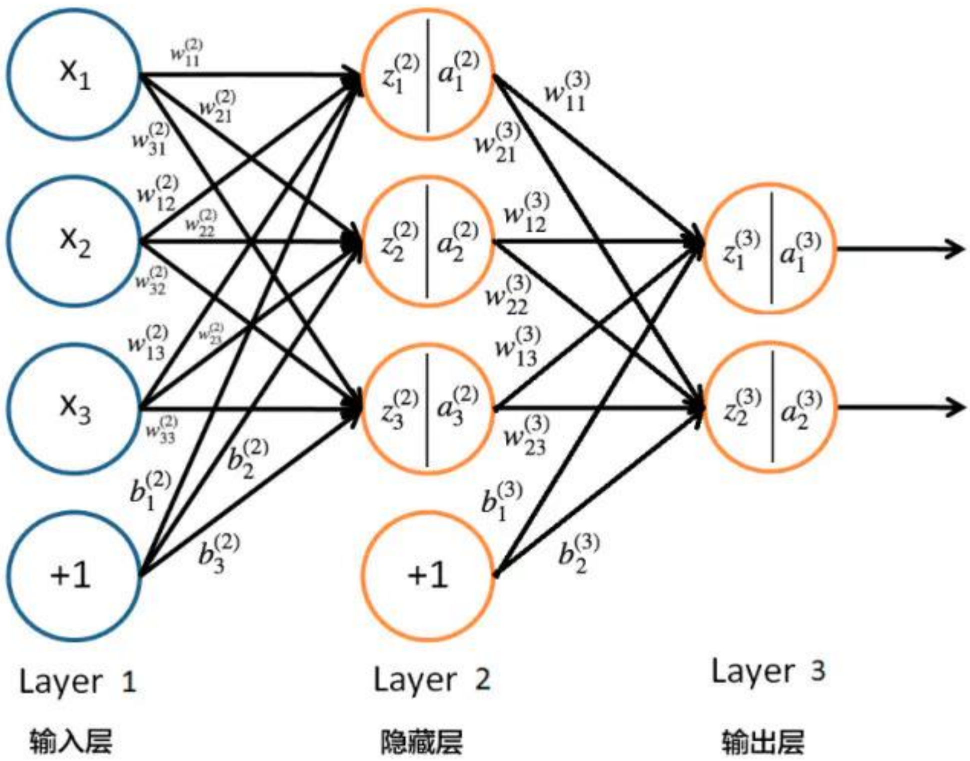

图 3 三层 BP 神经网络

5.1.2. GM (1.1)模型模型简介

GM 代表 grey model（灰色模型），GM(1,1)是一阶微分方程模型。GM(1,1)模型通过鉴别系统因素之间发展趋势的相异程度，即进行关联分析，并对原始数据进行生成处理来寻找系统变动的规律，生成有较强规律性的数据序列，然后建立相应的微分方程模型，从而预测事物未来发展趋势的状况。

GM(1,1)模型是灰色预测的核心, 它是一个单个变量预测的一阶微分方程模型, 其离散时间响应函数近似呈指数规律。

5.1.3. 模型建立

首先，通过知网上对于总人口数的影响因素进行筛选后，得到对总人口影响相对大的变量。

其次，应用 GM(1,1)灰色预测模型，将这些变量在 2021-2035 年的数据进行预测。

然后，将去掉总人口的预测数据，输入到 BP 神经网络里面进行训练。

最后，将 2021-2035 年 GM(1,1)预测的总人口作为预测变量，放到 BP 神经网络进行预测，得到最终未来 15 年的总人口数。

下面是 GM(1,1)灰色预测的模型、BP 神经网络的模型建立过程：

1. GM(1,1)灰色预测的模型建立

设$X^{(0)}=\{X^{(0)}(1),X^{(0)}(2),\dots,X^{(0)}(n)\}$为原始非负时间序列, $X^{(1)}(t)$为累加生成序列, 即

$$X^{(1)}(t)=\sum_{m=1}^{i}X^{(0)}(m),t=1,2,\dots,n$$

GM(1,1)模型的白化微分方程为：

$$\frac{dX^{(1)}}{dt}+aX^{(1)}=\mathrm{u}$$

式(2) 中, a 为待辨识参数, 亦称发展系数；u 为待辨识内生变量,亦称灰作用量. 设待辨识向量$\hat{a}=\begin{pmatrix}a\\u\end{pmatrix}$, 按最小二乘法求得$\hat{a}=(B^TB)^{-1}B^Ty$式中

$$\mathrm{B}=\begin{vmatrix}-\frac{1}{2}X^{(1)}(1)+X^{(1)}(2) & 1\\ -\frac{1}{2}X^{(1)}(2)+X^{(1)}(3) & 1\\ \cdots\cdots & \\ -\frac{1}{2}X^{(1)}(n-1)+X^{(1)}(n) & 1\end{vmatrix}$$

$$\mathrm{y}=\begin{vmatrix}X^{(0)}(2)\\X^{(0)}(3)\\\dots\\X^{(0)}(n)\end{vmatrix}$$

为所得的累加的预测值，将预测值还原即为：

$$X^{(1)}(t+1)=\left(x^{(0)}(1)-\frac{u}{a}\right)e^{-at}+\frac{u}{a}$$

$X^{(1)}\left(t+1\right)$为所得的累加的预测值，将预测值还原即为：

$$\hat{X}^{(0)}\left(t+1\right)=\hat{X}^{(1)}\left(t+1\right)-\hat{X}^{(1)}\left(t\right),\left(t=1,2,3\cdots n\right)$$

2. BP 神经网络的模型建立过程，如图 4。

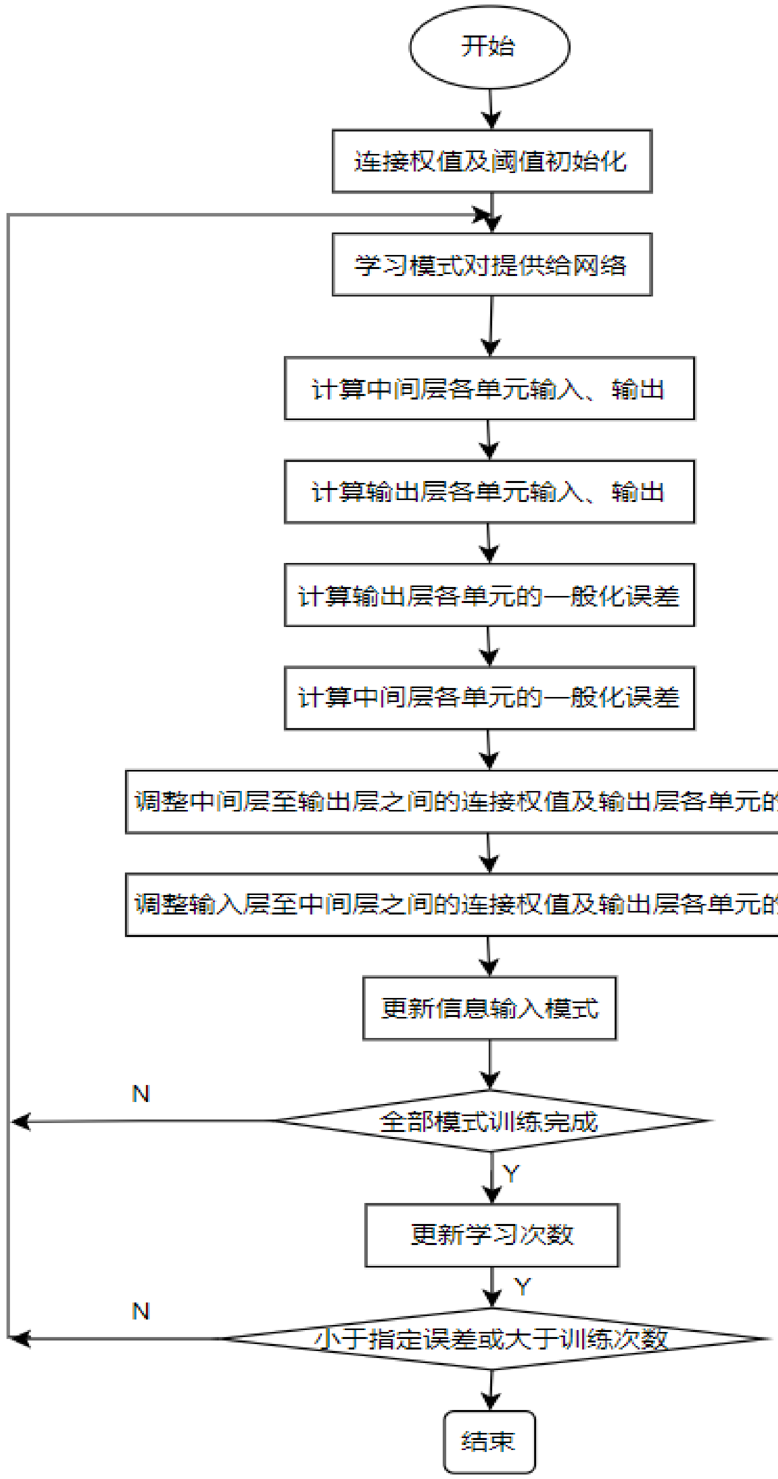

图 4 BP 神经网络建立流程图

5.1.4. 模型求解

问题一在不考虑三胎政策的情况下，2021-2035 的我国总人口数预测使用了两种 GM(1,1)模型与 BP 神经网络结合方式进行预测。

方法一：

1. 通过国家统计局的数据，获知我国 1980—2020 年的总人口、出生人口、性别比例、自然增长率、农村人口、城镇人口、GDP、结婚率数据。
2. 将以上数据，分别应用 GM(1,1)模型，预测出在 2021-2035 年的数据。
3. 另外，由于农村人口的灰色预测结果很不理想，因此之后将不再考虑此变量。
4. 将后面 6 种在 2021-2035 的预测数据作为 BP 神经网络的输入变量。
5. 然后再将 2021-2035 年的总人口预测数据作为 BP 神经网络的预测变量。
6. 最后进行 BP 神经网络的训练、预测，得到 2021-2035 年的预测总人口数。预测数据变化情况如图 5。

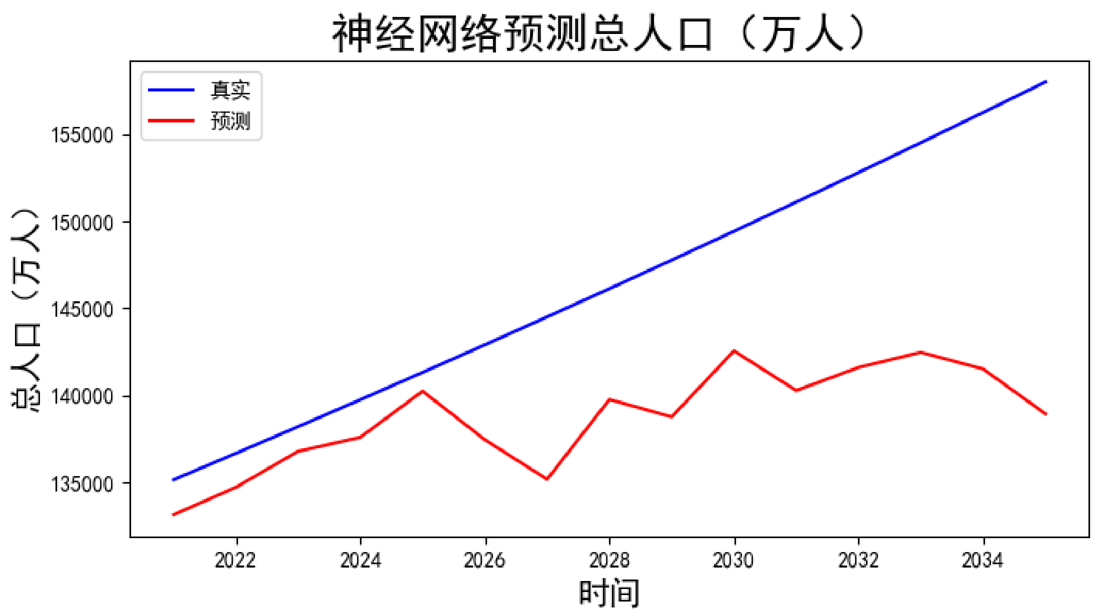

图 5 多变量下的预测总人口数

方法二：

1. 首先输入 1980—2020 年原始人口 Ti 进入 GM(1,1)灰色预测模型中，并输入预测年份;
2. 获得时间序列下 1980—2020 年总人口预测值 Pi ；
3. 将预测值 Pi 作为输入量，实际人口数据 Ti 作为输出量，对 BP 神经网络进行训练;
4. 将 2021—2035 的灰色预测值作为输入，代入 BP 神经网络，从而获得对应的未来 15 年的总人口预测数据。预测数据变化情况如图 6。

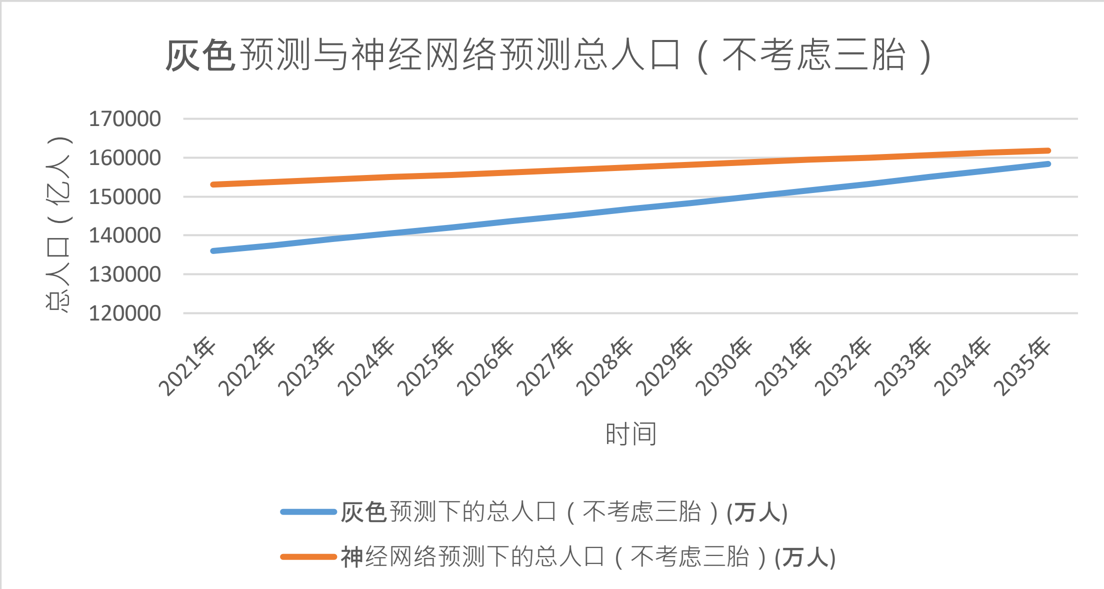

图 6 单变量下的预测总人口数

5.2. 问题二的建模与求解

5.2.1. 模型建立

将人口预测分为两部分，全面三孩政策下的人口 = 未开放三孩政策下的人口 + 后期新增三孩数。而未开放三孩政策下的人口就是问题一中的预测人口数。因此，问题二需要预测出后期新增的三孩数。

对于新增三孩数，在合理预测出每年育龄妇女的人数的基础上，参考三胎生育意愿，同时考虑育龄妇女三胎生育能力，即一般三胎生育率，从而算出政策开放后每年的新增三胎数进行累加计算。最终求得三胎政策下 2021-2035 年人口总数。

5.2.2. 模型求解

1. 将 2003-2020 年育龄妇女人数放到灰色预测里面进行预测，得到 2021-2035 年的育龄妇女人数。
2. 将 2003-2015 年一般三孩生育率进行拟合，得到了关于时间-生育率的一元函数。接着利用此函数将 2021-2035 的一般三孩生育率计算出来，拟合效果如图 7。
3. 根据三胎生育意愿 12.2%以及 2021-2035 年一般三孩生育率，从而算出每年的新增人口数。具体数据变化情况如图 8。
4. 将新增人口数与没有开放三胎政策之前的人口数进行相加，最终得到开放三

胎后 2021-2035 年的总人口数。具体数据变化情况如图 9。

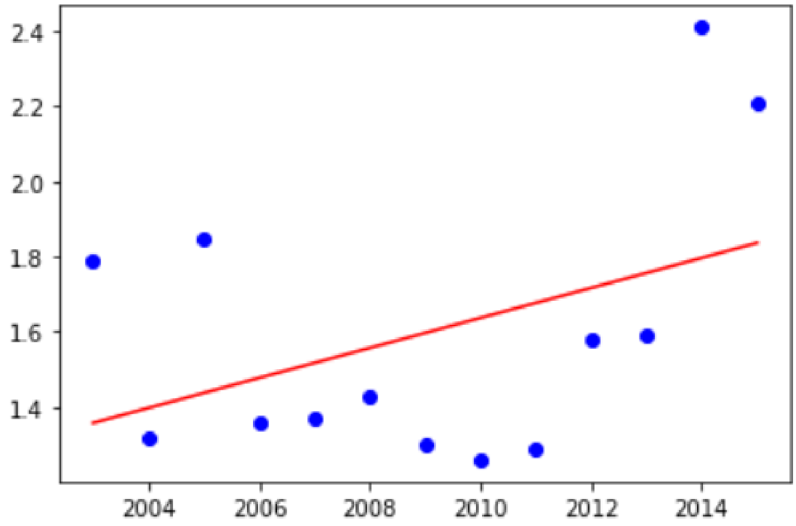

图 7 一般三孩生育率

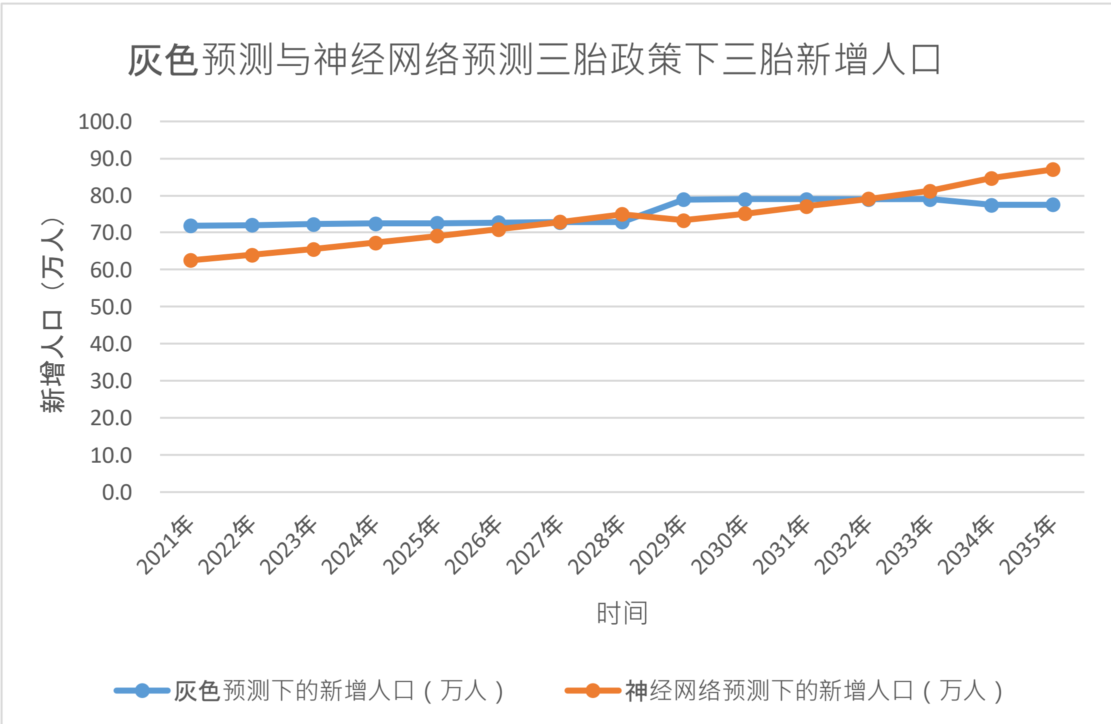

图 8 新增三孩数

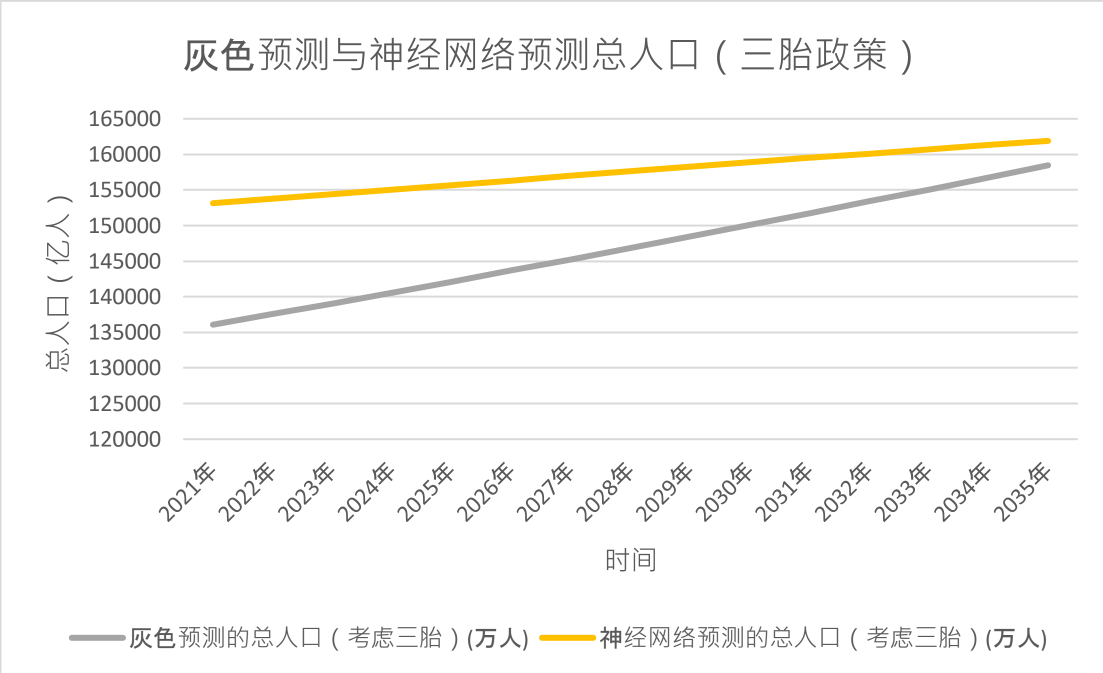

图 9 三胎政策下的总人口数

5.3. 问题三的建模与求解

5.3.1. Leslie 模型模型简介

Leslie 矩阵模型利用某一初始时刻种群的年龄分布，动态预测种群年龄分布随时间的演变过程。这是一个线性的种群数量动力学模型，通过 Leslie 矩阵可以得出，当时间充分时，种群的年龄分布趋于稳定，总数量趋于指数增长、指数衰变，或保持不变。

我们把整个社会中的人群按年龄等分成 n 组，每组中该年的人口总数为 a_i，$i=1,2,\dots,n$，每组人口的每年的普遍存活率为 c_i，$i=1,2,\dots,n-1$ (设最后一组下一年全部死亡)，每组人口的每年普遍生育率为 b_i，$i=1,2,\dots,n$ ，则下一年每组中的人口总数 a_i'，$i=1,2,\dots,n$，就满足递推关系式

$$\begin{cases} a_i' = a_{i-1}c_{i-1}, i = 2,3,\dots,n \\ a_i' = \sum_{i=1}^{n} a_i b_i \end{cases}$$

该式可写成矩阵乘向量的形式:

$$\vec{a}=\begin{pmatrix} b_1 & b_2 & \dots & b_{n-1} & b_n \\ c_1 & 0 & \dots & 0 & 0 \\ 0 & c_2 & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & c_{n-1} & 0 \end{pmatrix}$$

5.3.2. 模型建立

通过模型介绍可知，Leslie 模型的思想是将女性的一生分为多个年龄段，只讨论年龄段之间女性数量的传递关系，而忽略了在同一年龄段之间女性数量的变化，因此要使得 Leslie 模型能更准确的预测人口数量的变化，需要增加年龄段的划分，最理想的划分按照每一岁划分一次。考虑女性的生育率随时间变化的趋势情况，特别是在全面二孩政策后，势必会对女性的生育率有一个正向的冲击。所以我们对于女性生育率的预测一定要包括政策的影响。

男婴数量：$x^m(i,k+1) = s^m(0,k)a(k)\sum_{i=i_1}^{i_2} b^w(i,k)x^w(i,k) + v^m(0,k)$

女婴数量：$x^w(i,k+1) = s^w(0,k)(1-a(k))\sum_{i=i_1}^{i_2} b^w(i,k)x^w(i,k) + v^w(0,k)$

人口数量演变：$x^l(i+1,k+1) = s^l(i,k)a(k)x^l(i,k) + v^l(i,k)$

k 年每位育龄女性的生育数：$\beta^w(k) = \sum_{i=i_1}^{i_2} b^w(i,k)$

若假定所有女性在育龄期间都保持这个生育数,则"(k)～总和生育率(每位女性一生的平均生育数)。

生育模式：$h^w(i,k) = \frac{b^w(i,k)}{\beta^w(k)}$(i 岁女性生育数在育龄女性中的比例),且 $\sum_{i=i_1}^{i_2} b^w(i,k)$=1

人口分布向量：$x^l(k) = [x^l(1,k), x^l(2,k), \ldots, x^l(n,k)]^{\mathrm{T}}$

存活率矩阵：$S^w(k) =$

$$\begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ s^l(1,k) & 0 & \cdots & 0 & 0 \\ 0 & s^l(1,k) & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & s^l(1,k) & 0 \end{pmatrix}$$

生育模式矩阵：$H^w(k) =$

$$\begin{pmatrix} 0 & 0 & 0 & 0 & h^{w}(i_1,k) & \cdots & h^{w}(i_2,k) & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}$$

5.3.3. 模型求解

$$x^{l}(k+1)=f(s^{l}(k),a^{l}(k),\beta^{l}(k),H^{l}(k),v^{l}(k),x^{l}(k)),l=m、w,k=1,2,\ldots$$

$x^{l}(i,k)$全面、完整地描述人口的演变过程.

人口总数：$\mathrm{X(k)}=\sum_{i=0}^{n}[x^{m}(i,k)+x^{w}(i,k)]$

$\mathrm{L(k)}=\sum_{j=1}^{2}(\sum_{i=i_{m1}}^{i_{m2}}x^{m}(i,k)+\sum_{i=i_{w1}}^{i_{w2}}x^{w}(i,k))\quad \rho(\mathrm{k})=\frac{\mathrm{X(k)-L(k)}}{\mathrm{L(k)}}$

L(k) ~劳动力人数　　$[i_{m1},i_{m2}],[i_{w1}i_{w2}]$ ~ 劳动力年龄

通过 matlab 应用上述模型作图，得到如图 10：

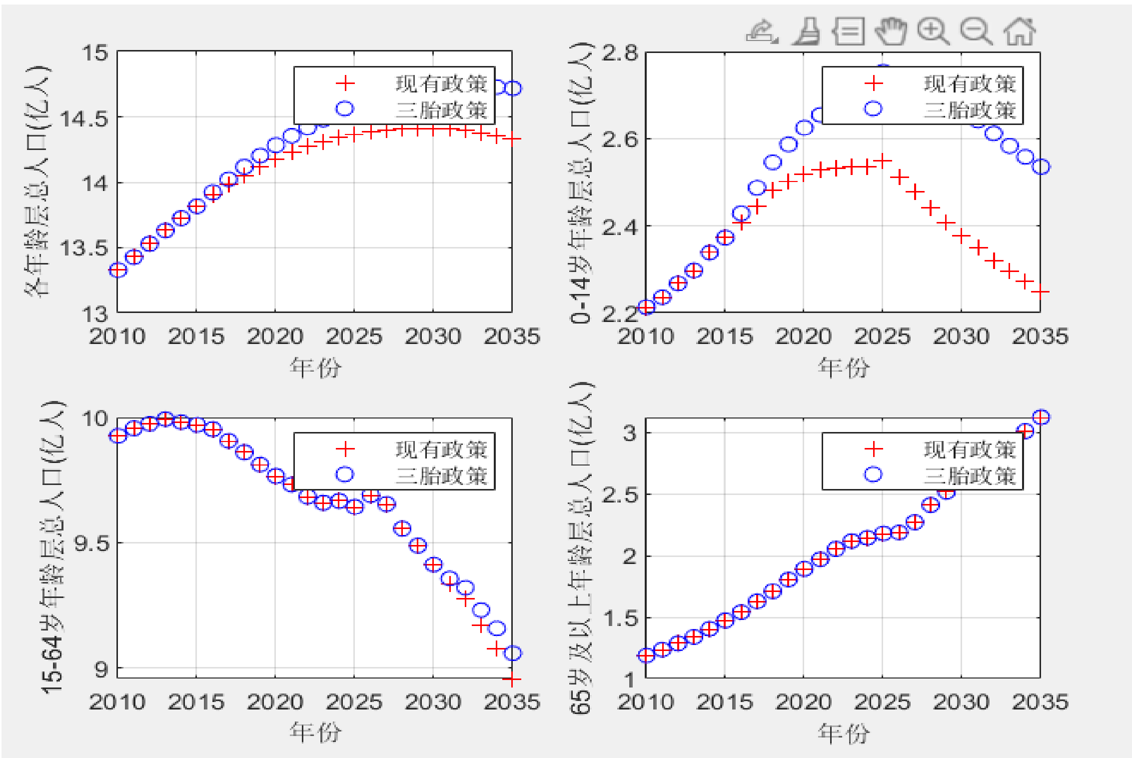

图 10

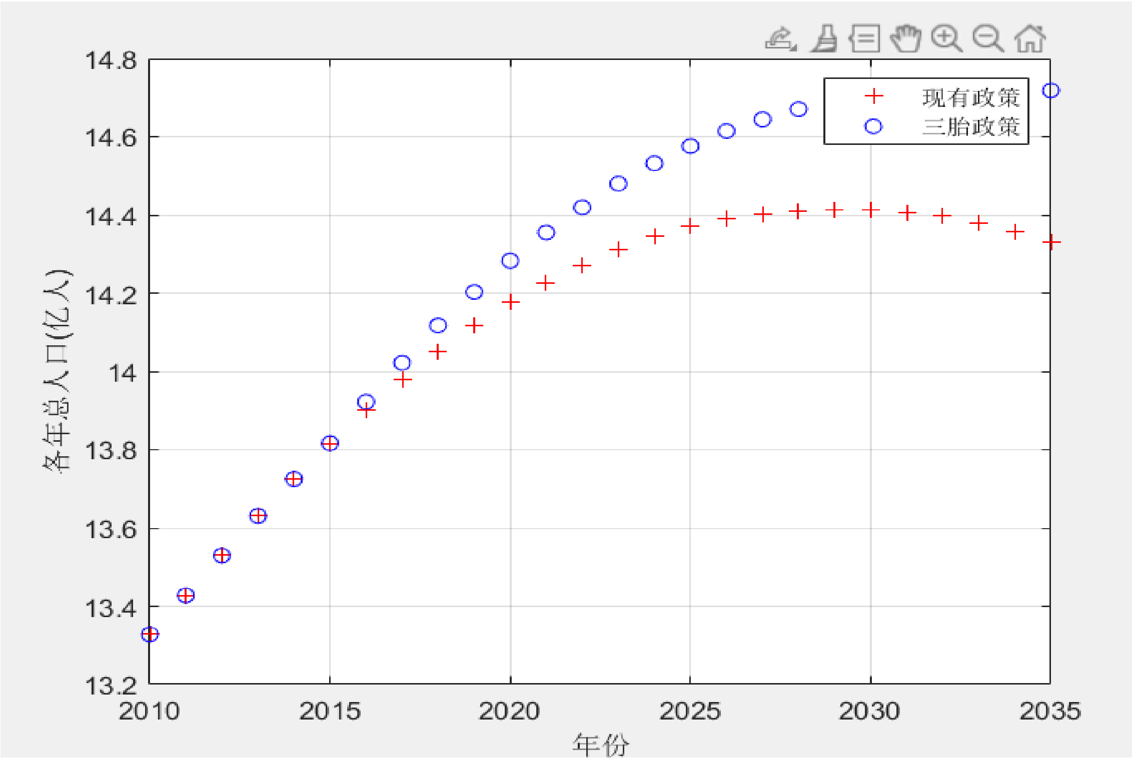

图 11

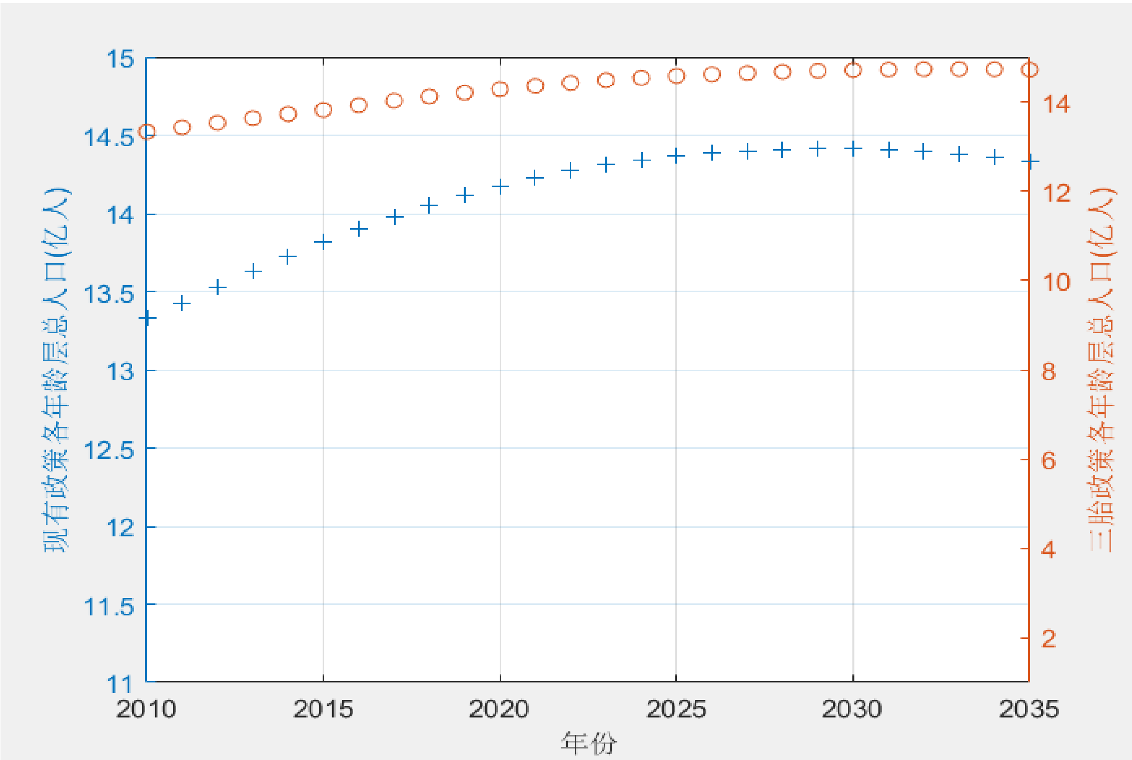

图 12

5.4. 问题四的建模与求解

5.4.1. 多项式回归模型简介

多项式回归，回归函数是回归变量多项式的回归。多项式回归模型是线性回归模型的一种，此时回归函数关于回归系数是线性的。由于任一函数都可以用多

项式逼近，因此多项式回归有着广泛应用。

一元 m 次多项式回归方程为：$\hat{y} = b_0 + b_1x + b_2x^2 + \ldots + b_mx^m$

5.4.2. 模型建立

基于多项式回归模型对中国 GDP、人均 GDP 进行分析、回归、预测，与问题三中 Leslie 模型预测的人口结构相结合，预测、分析三胎政策对中国的经济增长影响

5.4.3. 模型求解

通过 matlab 应用多项式回归模型作图，得到如图 13：

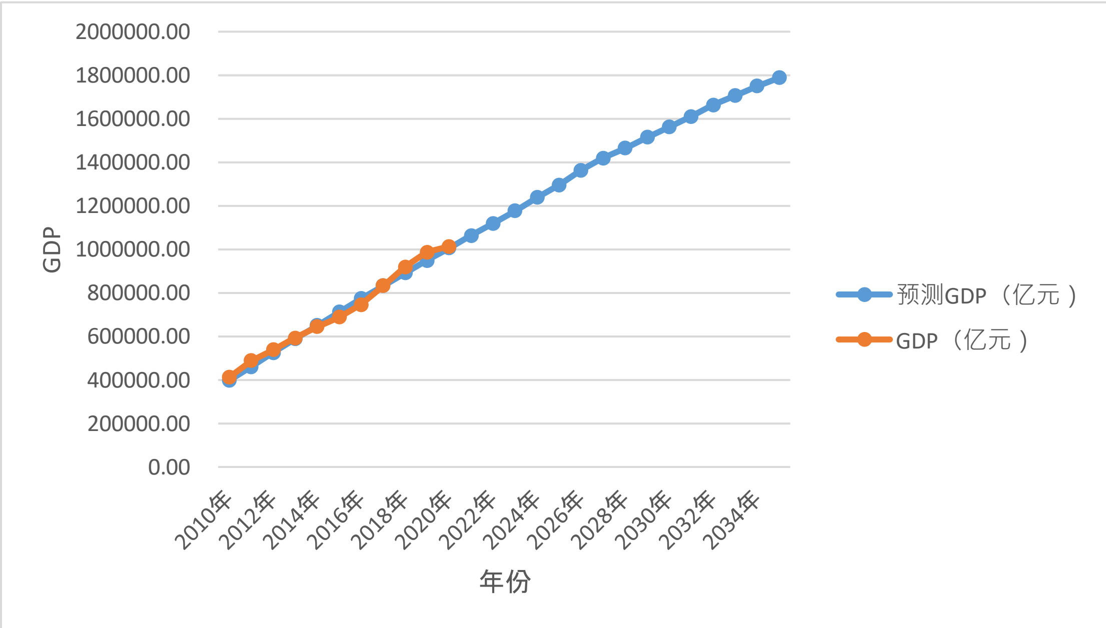

图 13

六、模型的检验与误差分析

对 GM(1,1)模型预测出的未来 15 年总人口数的分析

1) 后验差比检验：

在进行模型构建后，会得到后验差比 C 值，该值为残差方差 / 数据方差；其用于衡量模型的拟合精度情况，C 值越小越好，一般小于 0.65 即可。具体判定等级如表 1。

后验差比值 C	小误差概率 P	预测精度等级
<0.35	>0.95	好
<0.50	>0.80	合格
<0.65	>0.70	勉强
>=0.65	<=0.65	不合格

表 1

2) 模型残差检验：

模型残差检验为事后检验法。主要查看相对误差值和级比偏差值。相对误差值=预测拟合值与残差值的差值绝对值 / 原始值。相对误差值越小越好，一般情况下小于 20%即说明拟合良好。级比偏差值也用于衡量拟合情况和实际情况的偏差，一般该值小于 0.2 即可。

下面是育龄妇女人数的检验分析结果：

```
1 print(result) # 2021-2035的相关数据
```

```
{'a': {'value': 0.014361542194514944, 'desc': '发展系数'}, 'b': {'value': 370754.19840219116, 'desc': '灰色作用量'}, 'predict': {'value': array([242879.88254561, 239416.68084759]), 'desc': '往后预测2个的序列'}, 'C': {'value': 0.3511302856184921З, 'desc': '后验差比<=0.5, 模型精度等级为合格'}}
```

对 Leslie 模型的误差分析

(1) 模型误差：

Leslie 模型针对各年龄段生育率与死亡率，而并没有对迁入迁出率、城镇化指数等影响因素分析。只进行了三胎政策下人口结构的近似模拟。

(2) 观测误差：

Leslie 模型所使用各个年龄段的死亡率会随着科技的进步等因素随着年份进行改变，从而在长期上影响模型的结果。

七、模型的评价与推广

7.1.1. 灰色 GM(1,1)模型评价推广

1. 灰色 GM(1,1)模型优点：

灰色 GM(1,1)预测模型在计算过程中主要以矩阵为主，它与 MATLAB 的结合解决了它在计算中的问题. 由 MATLAB 编制的灰色预测程序简单实用，容易操作，预测精度较高。

2. 灰色 GM(1,1)模型的缺点：

该模型是指运用曲线拟合和灰色系统理论对我国人口发展进行预测的方法，因此它对历史数据有很强的依赖性，而且 GM(1,1)的模型没有考虑各个因素之间的联系. 因此，误差偏大，尤其是对中长期预测，例如对中国人口总数变化情况做长期预测时，误差偏大，脱离实际. 下面我们来讨论 GM(1,1)模型的适用范围.

7.1.2. BP 神经网络模型评价推广

1. 运用 BP 神经网络进行人口预测的可行性：

一个好的人口预测模型首先应符合人口基本理论和数学建模要求，这是选择模型的关键. 其次要保证模型数据可得，一致和可比性，在数据预测检验阶段应充分拟合原始数据,特别是有波动的数据，因为波动性数据往往蕴藏了系统重要信息. 用神经网络对人口历史数据进行分析拟合，是人口预测的有效方法. 与传统的人口预测方法相比，将基于 BP 神经网络的时间序列预测方法用于人口预测，避免了繁琐的常规建模过程. 神经网络模型良好的适应和自学习能力，使预

测系统计算简单，灵活，运用计算机强大的组合能力，可以更好地实现人口分类预测和管理，大大提高人口预测效率和预测精度。

2. BP 神经网络模型的不足：

BP 神经网络需要大量的样本数据用来训练和测试，当样本数量不够时，预测的误偏差很大。

7.1.3. Leslie 模型评价推广

1. Leslie 模型的优点：

采用 Leslie 模型，建立年龄结构的离散模型，并通过合理假设，在时间跨度不大的前提下，对人口结构仅此进行了预测，在预测中使用了第六次人口普查的数据，使模型计算结果更加准确。

参考文献

[1]张之政,林雨泽,熊伟宇,张贺。全面二胎政策前后人口结构预测模型[J]。中国高新区,2018(03):236。

[2]朱海,李卓,杨晓芳。基于 Leslie 模型预测二胎政策对经济发展的影响[J]。黑龙江科学,2020,11(04):162-164。

[3]吕俊兴,徐天琛,王辉,唐煜。新二胎政策下基于 Leslie 矩阵等数学模型的山东省人口预测[J]。青岛大学学报(自然科学版),2017,30(01):14-20。

[4]吕静妍,孙培成,孙柏芳。全面二胎政策对我国人口结构的影响[C]//。荆楚学术 2017 年第 8-9 期。[出版者不详],2017:151-159。

[5]计会凤,金赛,许颖,高更和。基于神经网络模型的中国乡村人口总量分析与预测[J]。农村经济与科技,2022,33(04):7-10。

[6]刘中侠,蒋诗泉。基于灰色 GM（1,1）和 BP 神经网络组合预测模型及应用[J]。铜陵学院学报,2016,15(3):102-104。DOI:10。3969/j。issn。1672-0547。2016。03。027。

附录

附录 1 问题一所需的八个特征值（部分数据）

时间	年末总人口(万人)	城镇人口(万人)	乡村人口(万人)	出生人口	中国性别比例(按照女生=100)	总和生育率	自然增长率	结婚率	国民总收入(亿元)
1980	98705	19140	79565	1776	105.98	2.61	11.87	14.6	4586.1
1981	100072	20171	79901	2064	106.11	2.55	14.55	20.8	4933.7
1982	101654	21480	80174	2230	106.19	2.54	15.68	16.5	5380.5
1983	103008	22274	80734	2052	106.61	2.56	13.29	14.9	6043.8
1984	104357	24017	80340	2050	106.61	2.61	13.08	15	7314.2
1985	105851	25094	80757	2196	107.04	2.65	14.26	15.7	9123.6
1986	107507	26366	81141	2374	107.04	2.67	15.57	16.4	10375.4
1987	109300	27674	81626	2508	106.19	2.64	16.61	17.2	12166.6
1988	111026	28661	82365	2445	106.27	2.58	15.73	16.6	15174.4
1989	112704	29540	83164	2396	106.4	2.46	15.04	16.8	17188.4
1990	114333	30195	84138	2374	106.27	2.31	14.39	16.4	18923.3
1991	115823	31203	84620	2250	105.52	2.14	12.98	16.5	22050.3
1992	117171	32175	84996	2113	104.27	1.98	11.6	16.5	27208.2
1993	118517	33173	85344	2120	104.18	1.84	11.45	15.5	35599.2

1994	119850	34169	85681	2098	104.51	1.73	11.21	15.6	48548.2
1995	121121	35174	85947	2052	104.21	1.66	10.55	15.4	60356.6
1996	122389	37304	85085	2057	103.34	1.62	10.42	15.3	70779.6
1997	123626	39449	84177	2028	104.36	1.6	10.06	14.7	78802.9

附录 2 预测人口

时间	城镇人口(万人)	乡村人口(万人)	出生人口	中国性别比例(按照女生=100)	总和生育率	自然增长率	结婚率	国民总收入(亿元)	年末总人口(万人)
2022年	60976	79921	1680	105.3	1.3	6.6	12.6	325820.6	136654
2023年	63572	79800	1657	105.3	1.3	6.3	12.4	371886.5	138187
2024年	66279	79678	1635	105.2	1.3	6.1	12.2	424465.5	139737
2025年	69101	79557	1612	105.2	1.2	5.9	12.1	484478.4	141304
2026年	72043	79437	1591	105.2	1.2	5.7	11.9	552976.1	142889
2027年	75110	79316	1569	105.1	1.2	5.5	11.8	631158.3	144492
2028年	78308	79195	1548	105.1	1.1	5.3	11.6	720394.3	146113
2029年	81642	79075	1527	105.1	1.1	5.1	11.4	822246.9	147752
2030年	85119	78955	1506	105.0	1.1	4.9	11.3	938499.8	149409
2031年	88743	78835	1486	105.0	1.0	4.7	11.1	1071189.1	151085
2032年	92521	78715	1465	105.0	1.0	4.6	11.0	1222638.6	152780
2033年	96460	78596	1446	104.9	1.0	4.4	10.8	1395500.7	154494
2034年	100567	78476	1426	104.9	1.0	4.3	10.7	1592802.8	156227
2035年	104849	78357	1407	104.9	0.9	4.1	10.6	1818000.4	157980

附录 3 三胎问题下的总人数

时间	平均育龄妇女人数(万人)	灰色预测育龄妇女	神经网络预测育龄妇女
2003	35068	35067	34511
2004	35123	36824	35156
2005	46766	36243	46727
2006	32892	35671	32889
2007	32481	35109	33017
2008	31951	34555	33017
2009	31736	34010	31522
2010	30820	33474	32399
2011	32550	32946	32538
2012	31705	32447	32066
2013	30887	31915	30633
2014	30473	31412	30598
2015	36567	30917	36548
2016	36749	30429	24499
2017	32749	29949	24407
2018	34600	29477	24358
2019	34100	29012	24352
2020	32229	28555	24387

附录 4 全国各年龄、性别的死亡人口状况(2009.11.1-2010.10.31)(部分数据)

年 龄	平均人口			死亡人口			死亡率		
	合计	男	女	合计	男	女	合计	男	女
总 计	1330613984	681275604	649338380	7421990	4293783	3128207	5.58	6.30	4.82
0-4 岁	77222937	41973542	35249395	99808	55145	44663	1.29	1.31	1.27
0	15781004	8586163	7194841	60217	32026	28191	3.82	3.73	3.92
1	15609741	8529096	7080645	17367	9892	7475	1.11	1.16	1.06
2	15620703	8481993	7138710	9799	5711	4088	0.63	0.67	0.57
3	15308734	8301050	7007684	6914	4160	2754	0.45	0.50	0.39
4	14902755	8075240	6827515	5511	3356	2155	0.37	0.42	0.32
5-9 岁	70449638	3823699	3221264	21183	13621	7562	0.30	0.36	0.23

		3	5							
5	14909878	8084689	6825189	4968	2996	1972	0.33	0.37	0.29
6	13682248	7439647	6242601	4357	2762	1595	0.32	0.37	0.26
7	13813184	7497063	6316121	3936	2598	1338	0.28	0.35	0.21
8	13750209	7465949	6284260	3892	2559	1333	0.28	0.34	0.21
9	14294119	7749645	6544474	4030	2706	1324	0.28	0.35	0.20
10-14 岁	77144787	4136844 3	3577634 4	23088	15243	7845	0.30	0.37	0.22
10	14426198	7800080	6626118	4396	2875	1521	0.30	0.37	0.23
11	14826750	7993164	6833586	4338	2805	1533	0.29	0.35	0.22

附录 5 全国各年龄、性别的人口（部分数据）

年　龄	人　口　数			占总人口比重			性别比
	合计	男	女	合计	男	女	（女=100）
总　计	1332810869	682329104	650481765	100.00	51.19	48.81	104.90
0-4 岁	75532610	41062566	34470044	5.67	3.08	2.59	119.13
0	13786434	7461199	6325235	1.03	0.56	0.47	117.96
1	15657955	8574973	7082982	1.17	0.64	0.53	121.06
2	15617375	8507697	7109678	1.17	0.64	0.53	119.66
3	15250805	8272491	6978314	1.14	0.62	0.52	118.55
4	15220041	8246206	6973835	1.14	0.62	0.52	118.24
5-9 岁	70881549	38464665	32416884	5.32	2.89	2.43	118.66
5	14732137	7988151	6743986	1.11	0.60	0.51	118.45
6	14804470	8034452	6770018	1.11	0.60	0.51	118.68
7	13429161	7292300	6136861	1.01	0.55	0.46	118.83
8	13666956	7423559	6243397	1.03	0.56	0.47	118.90
9	14248825	7726203	6522622	1.07	0.58	0.49	118.45
10-14 岁	74908462	40267277	34641185	5.62	3.02	2.60	116.24
10	14454357	7830808	6623549	1.08	0.59	0.50	118.23
11	13935714	7522558	6413156	1.05	0.56	0.48	117.30
12	15399559	8288987	7110572	1.16	0.62	0.53	116.57
13	15225032	8161000	7064032	1.14	0.61	0.53	115.53
14	15893800	8463924	7429876	1.19	0.64	0.56	113.92
15-19 岁	99889114	51904830	47984284	7.49	3.89	3.60	108.17

灰色预测育龄妇女人口数

```
from sklearn.metrics import mean_absolute_error
import pandas as pd
```

```
import numpy as np
import matplotlib.pyplot as plt

#读取数据
data = pd.read_excel('./育龄妇女数.xls')
# 先训练总人口
train = data['灰色预测'].values[:18] #训练数据
test = data['灰色预测'].values[18:31] #测试数据
def GM11(x,n):
    '''
    灰色预测
    x：序列，numpy 对象
    n:需要往后预测的个数
    '''
    x1 = x.cumsum()#一次累加
    z1 = (x1[:len(x1) - 1] + x1[1:])/2.0#紧邻均值
    z1 = z1.reshape((len(z1),1))
    B = np.append(-z1,np.ones_like(z1),axis=1)
    Y = x[1:].reshape((len(x) - 1,1))

    #a 为发展系数 b 为灰色作用量
    [[a],[b]] = np.dot(np.dot(np.linalg.inv(np.dot(B.T, B)), B.T), Y)#
计算参数
    result = (x[0]-b/a)*np.exp(-a*(n-1))-(x[0]-b/a)*np.exp(-a*(n-2))
    S1_2 = x.var()#原序列方差
    e = list()#残差序列
    for index in range(1,x.shape[0]+1):
        predict                                                    =
(x[0]-b/a)*np.exp(-a*(index-1))-(x[0]-b/a)*np.exp(-a*(index-2))
        e.append(x[index-1]-predict)
    S2_2 = np.array(e).var()#残差方差
    C = S2_2/S1_2#后验差比
    if C<=0.35:
        assess = '后验差比<=0.35，模型精度等级为好'
    elif C<=0.5:
        assess = '后验差比<=0.5，模型精度等级为合格'
    elif C<=0.65:
        assess = '后验差比<=0.65，模型精度等级为勉强'
    else:
        assess = '后验差比>0.65，模型精度等级为不合格'

    #预测数据
    predict = list()
    for index in range(x.shape[0]+1,x.shape[0]+n+1):
```

```
predict.append((x[0]-b/a)*np.exp(-a*(index-1))-(x[0]-b/a)*np.exp(-a*(
index-2)))
    predict = np.array(predict)
    return {
            'a':{'value':a,'desc':'发展系数'},
            'b':{'value':b,'desc':'灰色作用量'},
            'predict':{'value':result,'desc':'第%d 个预测值'%n},
            'C':{'value':C,'desc':assess},
            'predict':{'value':predict,'desc':'往后预测%d 个的序列
'%(n)},
            }
#GM11 动态建模
yPre = []
for i in range(test.shape[0]):
    #只预测 1 个数
    result = GM11(train,1)
    yPre.append(result['predict']['value'][0])
    #更新训练集
    train = train.tolist()[:-1]
    train.append(test[i])
    train = np.array(train).reshape(-1)
#计算 MAE
MAE = mean_absolute_error(test,yPre)
#打印模型
print(result['C']['desc'])
print(result['a']['desc'],np.round(result['a']['value'],2))
print(result['b']['desc'],np.round(result['b']['value'],2))
```

一般三孩生育率

```
from sklearn.linear_model import LinearRegression
import matplotlib.pyplot as plt
import numpy as np

if __name__ == '__main__':
    X
=[2015,2014,2013,2012,2011,2010,2009,2008,2007,2006,2005,2004,2003]
    Y
=[2.21,2.41,1.59,1.58,1.29,1.26,1.3,1.43,1.37,1.36,1.85,1.32,1.79]

    #转换成 numpy 的 ndarray 数据格式，n 行 1 列,LinearRegression 需要列格
式数据，如下：
    X_train = np.array(X).reshape((len(X), 1))
```

```
    Y_train = np.array(Y).reshape((len(Y), 1))
    # 转换后数据格式如下
    # X_train = [[12.46], [0.25], [5.22], [11.3], [6.81], [4.59], [0.66],
[14.53], [15.49], [14.43], [2.19], [1.35],
    #              [10.02], [12.93], [5.93], [2.92], [12.81], [4.88],
[13.11], [5.8]]
    # Y_train = [[29.01], [4.7], [22.33], [24.99], [18.85], [14.89],
[10.58], [36.84], [42.36], [39.73], [11.92], [7.45],
    #             [22.9], [36.62], [16.04], [16.56], [31.55], [20.04],
[35.26], [23.59]]


    #新建一个线性回归模型，并把数据放进去对模型进行训练
    lineModel = LinearRegression()
    lineModel.fit(X_train, Y_train)

    #用训练后的模型，进行预测
    Y_predict = lineModel.predict(X_train)

    #coef_是系数，intercept_是截距
    a1 = lineModel.coef_[0][0]
    b = lineModel.intercept_[0]
    print("y=%.4f*x+%.4f" % (a1,b))

    #对回归模型进行评分，这里简单使用训练集进行评分，实际很多时候用其
他的测试集进行评分
    print("得分", lineModel.score(X_train, Y_train))

    #简单画图显示
    plt.scatter(X, Y, c="blue")
    plt.plot(X_train,Y_predict, c="red")
    plt.show()
```

问题二 灰色预测总人口数

```
def GM11(x,n):
    '''
    灰色预测
    x：序列，numpy 对象
    n:需要往后预测的个数
    '''
    x1 = x.cumsum()#一次累加
    z1 = (x1[:len(x1) - 1] + x1[1:])/2.0#紧邻均值
    z1 = z1.reshape((len(z1),1))
    B = np.append(-z1,np.ones_like(z1),axis=1)
    Y = x[1:].reshape((len(x) - 1,1))
```

```
    #a为发展系数 b为灰色作用量
    [[a],[b]] = np.dot(np.dot(np.linalg.inv(np.dot(B.T, B)), B.T), Y)#
计算参数
    result = (x[0]-b/a)*np.exp(-a*(n-1))-(x[0]-b/a)*np.exp(-a*(n-2))
    S1_2 = x.var()#原序列方差
    e = list()#残差序列
    for index in range(1,x.shape[0]+1):
        predict                                                          =
(x[0]-b/a)*np.exp(-a*(index-1))-(x[0]-b/a)*np.exp(-a*(index-2))
        e.append(x[index-1]-predict)
    S2_2 = np.array(e).var()#残差方差
    C = S2_2/S1_2#后验差比
    if C<=0.35:
        assess = '后验差比<=0.35，模型精度等级为好'
    elif C<=0.5:
        assess = '后验差比<=0.5，模型精度等级为合格'
    elif C<=0.65:
        assess = '后验差比<=0.65，模型精度等级为勉强'
    else:
        assess = '后验差比>0.65，模型精度等级为不合格'

    #预测数据
    predict = list()
    for index in range(x.shape[0]+1,x.shape[0]+n+1):

predict.append((x[0]-b/a)*np.exp(-a*(index-1))-(x[0]-b/a)*np.exp(-a*(
index-2)))
    predict = np.array(predict)
    return {
            'a':{'value':a,'desc':'发展系数'},
            'b':{'value':b,'desc':'灰色作用量'},
            'predict':{'value':result,'desc':'第%d个预测值'%n},
            'C':{'value':C,'desc':assess},
            'predict':{'value':predict,'desc':'往后预测%d 个的序列
'%(n)},
            }
#GM11动态建模
yPre = []
for i in range(test.shape[0]):
    #只预测1个数
    result = GM11(train,1)
    yPre.append(result['predict']['value'][0])
    #更新训练集
```

```
    train = train.tolist()[:-1]
    train.append(test[i])
    train = np.array(train).reshape(-1)
#计算MAE
MAE = mean_absolute_error(test,yPre)
#打印模型
print(result['C']['desc'])
print(result['a']['desc'],np.round(result['a']['value'],2))
print(result['b']['desc'],np.round(result['b']['value'],2))
```

问题二 神经网络模型

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import BP
from sklearn import metrics
from sklearn.metrics import mean_absolute_error
from sklearn.metrics import mean_squared_error
from sklearn import preprocessing
# 导入必要的库

df1 = pd.read_excel('三胎问题下的总人数.xls', 0)
df1 = df1.iloc[:, 1:3]

# 进行数据归一化
min_max_scaler = preprocessing.MinMaxScaler()
df0 = min_max_scaler.fit_transform(df1)
df = pd.DataFrame(df0, columns=df1.columns)
x = df.iloc[:, -1]       # 灰度预测人口数（2003-2020，2021-2035）
y = df.iloc[:18, :-1]    # 2003-2020 真实人口数

# 划分训练集测试集
cut = 13  # 目前是 20 ，一共 33
x_train, x_test = x.iloc[:13], x.iloc[13:]
y_train, y_test = y, x.iloc[13:]
x_train, x_test = x_train.values, x_test.values
y_train, y_test = y_train.values, y_test.values

'''
x_train: 2003-2020 灰度预测的人口数
Y_train: 2003-2020 真实人口数
x_test:  2021-2035 灰度预测的人口数
Y_test:  2021-2035 灰度预测的人口数(只是用来计算误差)
'''
```

```
# 神经网络搭建
bp1 = BP.BPNNRegression([1, 16, 1])
train_data = [[sx.reshape(1, 1), sy.reshape(1, 1)] for sx, sy in zip(x_train, y_train)]
test_data = [np.reshape(sx, (1, 1)) for sx in x_test]

# 神经网络训练
bp1.MSGD(train_data, 60000, len(train_data)-4, 0.1)

# 神经网络预测
y_predict = bp1.predict(test_data)
y_pre = np.array(y_predict)  # 列表转数组
y_pre = y_pre.reshape(20, 1)
y_pre = y_pre[:, 0]

# 反归一化
df0[0:20, -1] = y_pre

a = min_max_scaler.inverse_transform(df0)

# 可视化，展示在测试集上的表现
plt.rcParams['font.sans-serif'] = ['SimHei']  # 使标题乱码变为中文
plt.figure(figsize=(8, 4))
plt.title('时间与总人口',  # 标题字体比坐标轴更大些
         fontsize = 20)
plt.xlabel('时间',fontsize=15)
plt.ylabel('总人口',fontsize=15)
plt.plot([i+2007 for i in list(range(15))], a[0:15, -1], color='blue')
plt.plot([i+2007 for i in list(range(15))], a[-15:, -1], color='red')
plt.legend(labels=['预测', '真实'])
plt.show()

predict_population = a[0:20, -1]
print(len(predict_population))
print(predict_population)

# 输出精度指标
print('测试集上的MAE/MSE/RMSE：')
print('平均绝对误差:', mean_absolute_error(y_pre, y_test))
print('均方误差(损失函数):', mean_squared_error(y_pre, y_test))
print('均方根误差:', mean_squared_error(y_pre, y_test)**0.5)
mape = np.mean(np.abs((y_pre-y_test)/(y_test))*100)
print('MAPE = ', mape, '%')
```

```
# 画出真实数据和预测数据的对比曲线图
print("R2 = ", metrics.r2_score(y_test, y_pre))
```

问题三 所建模型 matlab 求解代码

```
clear;
clc;
warning off;
close all;
%% 从 Excel 表格中读取数据
female = xlsread('人口数量与比例.xls',1,'D7:D126'); %2010 年各年龄段女
性人口数量
rsex = xlsread('人口数量与比例.xls',1,'H7:H126'); %2010 年各年龄段性别
比，女=100 的男性数量
death = xlsread('存活率.xls',1,'J8:J127'); %2010 年各年龄段死亡率，千
分之单位
%% 从数据中读取各年龄女性人口数、女性人口比例和存活率
j = 1;
for i=1:1:length(female)
    if mod(i-1,6) ~= 0
        x(j,1) = female(i)/1e8;
        w(j,1) = 100/(rsex(i)+100);
        s(j,1) = 1-death(i)/1000;
        j = j+1;
    end
end
%% 总和生育率
    b1(1:15,1) = 0; %0-14 岁
b1(16:50,1) = 1.6/35; %15-49 岁，现有政策
b1(51:100,1) = 0; %50-100 岁

b2(1:15,1) = 0; %0-14 岁
b2(16:50,1) = 1.8/35; %15-49 岁,三胎政策
b2(51:100,1) = 0; %50-100 岁
%% 计算 Leslie 人口预测模型中 Leslie 矩阵
L1 = zeros(100); %现有政策
for i=1:1:length(b1)
    L1(1,i) = w(i)*b1(i);
    if i ~= length(b1)
        L1(i+1,i) = s(i);
    end
end

L2 = zeros(100); %三胎政策
for i=1:1:length(b2)
```

```
    L2(1,i) = w(i)*b2(i);
    if i ~= length(b2)
        L2(i+1,i) = s(i);
    end
end
%% 预测每年的女性人口数量
t(1,1) = 2010; %数据起始年
period = 25; %预测年数
x1 = x;
x2 = x;
for i=2:1:period+1
    t(i,1) = 2010+i-1;
    x1(:,i) = L1*x1(:,i-1); %现有政策
    if t<2016
        x2(:,i) = L1*x2(:,i-1);
    else
        x2(:,i) = L2*x2(:,i-1); %三胎政策
    end
end
%% 转换成总人口数量
X1 = x1./w; %现有政策
X2 = x2./w; %三胎政策
%% 将结果写入 Excel 文件
filename = ['2010 年人口普查后',num2str(period),'年人口预测.xlsx'];
name = {'年龄/年份'};
% 现有政策 sheet 数据
sheet = '现有政策';
xlRange = 'A1';
xlswrite(filename,name,sheet,xlRange);
rowname = [0:1:99]';
xlRange = 'A2';
xlswrite(filename,rowname,sheet,xlRange);
colname = 2010:1:period+2010;
xlRange = 'B1';
xlswrite(filename,colname,sheet,xlRange);
xlRange = 'B2';
xlswrite(filename,X1,sheet,xlRange);
% 二胎政策 sheet 数据
sheet = '三胎政策';
xlRange = 'A1';
xlswrite(filename,name,sheet,xlRange);
rowname = [0:1:99]';
xlRange = 'A2';
xlswrite(filename,rowname,sheet,xlRange);
```

```
colname = 2010:1:period+2010;
xlRange = 'B1';
xlswrite(filename,colname,sheet,xlRange);
xlRange = 'B2';
xlswrite(filename,X2,sheet,xlRange);
%% 画图
figure(1)
hold on;
box on;
grid on;
plot(t,sum(X1,1),'r+');
plot(t,sum(X2,1),'bo');
xlabel('年份');
ylabel('各年总人口(亿人)');
legend({'现有政策','三胎政策'});

figure(2)
hold on;
box on;
grid on;

yyaxis left;
plot(t,sum(X1,1),'+');
ylim([11 15]);
ylabel('现有政策各年龄层总人口(亿人)');

yyaxis right
plot(t,sum(X2,1),'o');
ylim([1 15]);
ylabel('三胎政策各年龄层总人口(亿人)');
xlabel('年份');

figure(3)
subplot(2,2,1)
hold on;
box on;
grid on;
plot(t,sum(X1,1),'r+');
plot(t,sum(X2,1),'bo');
xlim([2010 2035]);
xlabel('年份');
ylabel('各年龄层总人口(亿人)');
legend({'现有政策','三胎政策'});
```

```
subplot(2,2,2)
hold on;
box on;
grid on;
plot(t,sum(X1(1:15,:),1),'r+');
plot(t,sum(X2(1:15,:),1),'bo');
xlim([2010 2035]);
xlabel('年份');
ylabel('0-14 岁年龄层总人口(亿人)');
legend({'现有政策','三胎政策'});

subplot(2,2,3)
hold on;
box on;
grid on;
plot(t,sum(X1(16:65,:),1),'r+');
plot(t,sum(X2(16:65,:),1),'bo');
xlim([2010 2035]);
xlabel('年份');
ylabel('15-64 岁年龄层总人口(亿人)');
legend({'现有政策','三胎政策'});

subplot(2,2,4)
hold on;
box on;
grid on;
plot(t,sum(X1(66:end,:),1),'r+');
plot(t,sum(X2(66:end,:),1),'bo');
xlim([2010 2035]);
xlabel('年份');
ylabel('65 岁及以上年龄层总人口(亿人)');
legend({'现有政策','三胎政策'});
```

问题四 模型求解代码

```
import pandas as pd
import numpy as np
from sklearn.linear_model import LinearRegression
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.preprocessing import PolynomialFeatures

datarj = pd.read_excel('人均 GDP 以及预测.xls')

from sklearn.preprocessing import PolynomialFeatures
```

```
x = datarj[['年']][:11]
y1 = datarj[['人均GDP']][:11]




p = PolynomialFeatures(degree = 1)
x_1 = p.fit_transform(x)
GDPY = LinearRegression()
GDPY.fit(x_1,y1)

yca = GDPY.predict(x_1)

xyc = datarj[['年']]
xyc_1 = p.fit_transform(xyc)
ycayc = GDPY.predict(xyc_1)


import matplotlib.pyplot as plt
plt.rcParams['font.sans-serif'] = ['SimHei']#使标题乱码变为中文
plt.figure(figsize = (8,4))
plt.title('时间与GDP',#标题字体比坐标轴更大些
         fontsize = 20)
plt.xlabel('时间',fontsize = 15)
plt.ylabel('GDP',fontsize = 15)
plt.plot(x,y1,color = 'blue')
plt.plot(x,yca,color = 'red')
plt.plot(xyc,ycayc,color = 'green')
plt.show()
```