

# 我国人口自然增长率的影响因素分析及预测

周小燕

重庆电力高等专科学校人文素质学院数学教研室, 重庆

收稿日期: 2025年8月15日; 录用日期: 2025年9月5日; 发布日期: 2025年9月17日

## 摘要

近年来, 我国总和生育率屡创新低, 人口自然增长率随之持续下滑并于2022年首次出现负值。若这一趋势延续, 人口年龄结构将加速老化, 劳动年龄人口规模与占比不断收缩, 对中长期劳动力供给及整体经济社会发展有着深远影响。为此, 本文基于2000~2023年我国经济、社会、人口与政策四维度的16项原始指标, 运用因子分析将高维变量降为“社会经济繁荣-老龄化因子”(F<sub>1</sub>)与“劳动年龄人口-教育因子”(F<sub>2</sub>)两个公共因子, 累积解释方差达到95.93%。在此基础上, 将因子得分作为解释变量, 建立多元线性回归模型, 结果显示: F<sub>1</sub>每提升1个单位, 人口自然增长率平均下降2.093个千分点; F<sub>2</sub>每提升1个单位, 人口自然增长率平均上升0.903个千分点, 两者均在0.01水平显著。这意味着, 社会经济繁荣-老龄化因子已成为抑制人口自然增长的“主刹车”, 而劳动年龄人口-教育因子仅为“小油门”。进一步利用ARIMA(0,1,1)模型预测表明, 2025~2029年我国人口自然增长率将继续下探, 年均降幅约0.48个千分点, 到2029年累计缺口较2024年再扩大2.4个千分点, 人口总量收缩压力进一步加剧。因此, 提出把“松刹车”政策置于优先位置, 以快速削弱负增长主因; 同时激活“小油门”, 通过劳动力优势形成持续正向推力, 双轮驱动、同步发力, 扭转人口自然增长率向下的局面。

## 关键词

人口自然增长率, 因子分析, 多元线性回归模型, ARIMA模型

## Analysis and Forecasting of Factors Influencing China's Natural Population Growth Rate

Xiaoyan Zhou

Mathematics Department, School of Humanities and Social Sciences, Chongqing Electric Power College, Chongqing

Received: Aug. 15<sup>th</sup>, 2025; accepted: Sep. 5<sup>th</sup>, 2025; published: Sep. 17<sup>th</sup>, 2025

## Abstract

In recent years, China's total fertility rate has repeatedly hit record lows, driving the natural population growth rate down steadily and into negative territory for the first time in 2022. If this trend persists, the age structure will age rapidly, and both the size and share of the working-age population will shrink, with profound implications for medium- and long-term labour supply and overall socio-economic development. Drawing on 16 raw indicators spanning economic, social, demographic and policy dimensions from 2000 to 2023, we first use factor analysis to reduce the high-dimensional data to two common factors—"socio-economic prosperity & ageing" ( $F_1$ ) and "working-age population & education" ( $F_2$ )—which together explain 95.93 % of the total variance. Treating the factor scores as regressors, we then estimate a multiple linear regression model. The results show that a one-unit increase in  $F_1$  is associated with an average decline of 2.093 per mille in the natural population growth rate, whereas a one-unit increase in  $F_2$  is associated with an average increase of 0.903 per mille; both effects are significant at the 0.01 level. Thus, prosperity-aging is the main brake on growth, while workforce-education is only a modest accelerator. An ARIMA(0,1,1) forecast projects the natural growth rate will keep falling by about 0.48‰ annually from 2025 to 2029, enlarging the cumulative deficit by 2.4‰ relative to 2024. Policy should therefore prioritize "releasing the brake" by mitigating aging-driven drag while simultaneously "stepping on the accelerator" by leveraging labor advantages to reverse the downward trend.

## Keywords

Natural Population Growth Rate, Factor Analysis, Multiple Linear Regression Model, ARIMA Model

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

人口问题是“国之大者”。自 20 世纪 90 年代起，我国总和生育率跌破 2.1 (更替水平)，人口再生产类型已由“高出生、低死亡、高增长”转变为“低出生、低死亡、低增长”。全国人口自然增长率于 2022 年首次转负(-0.60‰)，2023 年更下探至-1.48‰，表明我国人口问题已经到了刻不容缓的阶段。党的二十大报告提出“优化人口发展战略，建立生育支持政策体系”，如何在负增长时代保持经济社会活力，成为社会的共同关切。

既有文献从教育经费、全国人均 GDP、公共卫生与城镇化、房价等角度探讨了人口自然增长率的影响因素，但多聚焦单一变量或几个变量，难以揭示多重因素交织下的结构性机制。近年来，黄祺元[1]运用因子分析将 8 个指标压缩为 3 个公共因子，提高了变量整合度；在预测方法上，巩永丽、张德生等[2]构建非参数自回归模型，康意、王国军[3]采用二次指数平滑，均取得良好预测效果。

本文在现有研究基础上，基于 2000~2023 年我国各项数据表现，构建经济、社会、人口与政策四维度 16 项指标体系，运用因子分析降维的思想，找出影响人口自然增长率的公共因子。随后，以因子得分建立多元线性回归模型，量化公共因子对人口自然增长率的边际效应。为提高预测精度，将样本扩展至 1978~2024 年，引入 ARIMA 模型对 2025~2029 年人口自然增长率进行动态外推。研究旨在回答：(1) 多重因素交织下，人口自然增长率的核心抑制因子与驱动因子是什么；(2) 若延续现有政策框架，未来五年人口负增长态势将如何演化；(3) 如何设计有效政策扭转人口自然增长率向下的局面。研究结果将为优化

人口发展战略提供实证依据与政策参考。

2. 数据来源与指标体系

2.1. 数据来源

本文数据来自国家统计局，时间跨度 24 年，从 2000 到 2023 年，包含教育经费(万元)、人均 GDP (元)、卫生总费用(亿元)等，部分数据详情，见表 1 所示。

Table 1. National indicators by year, 2000~2023  
表 1. 2000~2023 年国家各项指标情况

| 年份    | 教育经费(万元)    | 全国人均 GDP(元) | 卫生总费用(亿元) |
|-------|-------------|-------------|-----------|
| 2000  | 38,490,806  | 8024        | 4586.63   |
| 2001  | 46,376,626  | 8818        | 5025.93   |
| 2002  | 54,800,278  | 9631        | 5790.03   |
| ..... | .....       | .....       | .....     |
| 2023  | 645,950,392 | 91,746      | 90575.81  |

2.2. 指标体系

本文根据以往研究者的结论并结合经验与猜想，确定了 16 个变量作为影响人口自然增长率(‰)的指标，具体指标，见表 2 所示。

Table 2. Indicators and their sources  
表 2. 各指标及来源

| 序号 | 指标名称                 | 指标来源               |
|----|----------------------|--------------------|
| 1  | 教育经费(万元)             | 陈丽萍(2014) [4]      |
| 2  | 全国人均 GDP(元)          | 张琼枝，陈望学等(2017) [5] |
| 3  | 卫生总费用(亿元)            |                    |
| 4  | 城镇化(%)               |                    |
| 5  | 住宅商品房平均销售价格(元/平方米)   | 靳天宇，刘东浩(2019) [6]  |
| 6  | 就业人员(万人)             | 肖和宏(2021) [7]      |
| 7  | 少儿抚养比(%)             |                    |
| 8  | 老年抚养比(%)             |                    |
| 9  | 居民人均可支配收入(元)         | 李资博(2022) [8]      |
| 10 | 年末参加生育保险人数(万人)       |                    |
| 11 | 每十万人人口平均在校生数(人)      |                    |
| 12 | 居民消费水平(元)            | 经验与猜想              |
| 13 | 居民消费价格指数(1978 = 100) |                    |
| 14 | 每万人拥有卫生技术人员数(人)      |                    |
| 15 | 人口死亡率(‰)             |                    |
| 16 | 人口出生率(‰)             |                    |

### 3. 人口自然增长率的影响因素分析

#### 3.1. 因子分析简介

因子分析的基本思想是根据相关性大小把原始变量分组，使得同组内的变量之间相关性较高，而不同组的变量间的相关性则较低。把  $p$  个原始变量及标准化后的变量向量均用  $X$  表示，用  $F_1, F_2, \dots, F_m (m < p)$  表示标准化的公共因子。则模型：

$$\begin{cases} X_1 = a_{11}F_1 + a_{12}F_2 + \dots + a_{1m}F_m + \varepsilon_1 \\ X_2 = a_{21}F_1 + a_{22}F_2 + \dots + a_{2m}F_m + \varepsilon_2 \\ \dots\dots \end{cases}$$

称为因子模型。其中， $\varepsilon = \varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$  与  $F$  相互独立，且  $E(\varepsilon) = 0$ ，协方差矩阵是对角矩阵[9]。

#### 3.2. 建模与求解

由于各指标之间的度量单位不同，且取值范围也存在较大的差异。为了消除量纲的影响，先对原始指标做数据标准化处理，使之变量的均值为 0，方差为 1，便于后续分析和建模。

通过“SPSS - 分析 - 降维 - 因子”对数据进行因子分析。首先根据各变量的相关阵判断，各变量之间有较强的相关性，如表 3 所示。

Table 3. Correlation matrix of variables

表 3. 各变量之间的相关矩阵

|           | 教育经费(万元) | 人均 GDP (元) | 卫生总费用(亿元) | ..... | 人口出生率(‰) |
|-----------|----------|------------|-----------|-------|----------|
| 教育经费(万元)  | 1.000    | 0.999      | 0.992     | ..... | 0.814    |
| 人均 GDP(元) | 0.999    | 1.000      | 0.992     | ..... | 0.825    |
| 卫生总费用(亿元) | 0.992    | 0.992      | 1.000     | ..... | 0.778    |
| .....     | .....    | .....      | .....     | ..... | .....    |
| 人口出生率(‰)  | 0.814    | 0.825      | 0.778     | ..... | 1.000    |

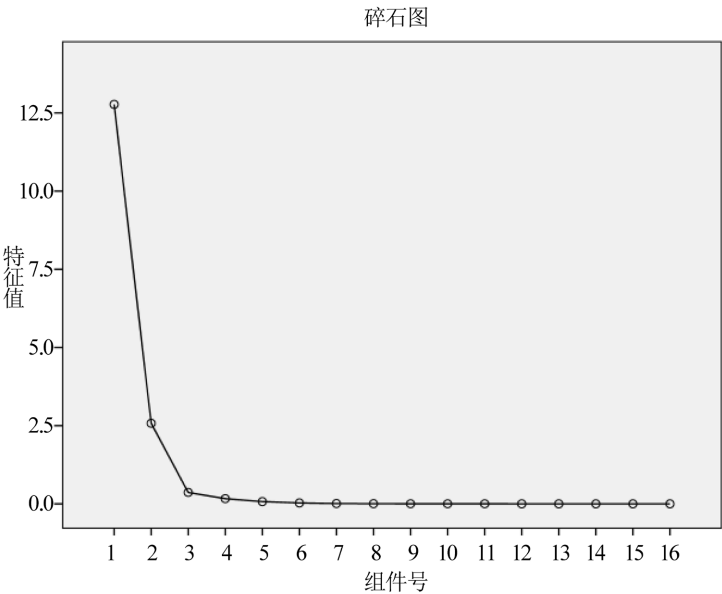


Figure 1. Scree plot  
图 1. 碎石图

然后，求解初始公共因子及因子载荷矩阵。按照特征根大于 1 的原则，选入 2 个公共因子，如图 1 所示。其累计方差贡献率为 95.93%，表明这两个因子共同解释了原始变量 95.93% 的方差，如表 4 所示。

**Table 4.** Total variance explained

**表 4.** 总方差解释

| 成分 | 提取载荷平方和 |        |        |
|----|---------|--------|--------|
|    | 总计      | 方差百分比  | 累积%    |
| 1  | 12.772  | 79.827 | 79.827 |
| 2  | 2.576   | 16.102 | 95.93  |

接着，为了使得到的公共因子的实际意义更容易解释，对公共因子进行方差最大化正交旋转并按照系数大小排序得到结果如表 5。公共因子  $F_1$  在卫生总费用、可支配收入、人均 GDP、消费水平、教育经费、每万人拥有卫生技术人员、房价、老年抚养比、参加生育保险人数、城镇化、出生率、死亡率上面的载荷值都很大。卫生总费用、每万人拥有卫生技术人员是反映社会医疗保障间接反映的就是社会的经济水平，可支配收入、人均 GDP、消费水平、房价、城镇化是反映社会经济水平，教育经费、参加生育保险人数反映的是社会的教育、生育福利水平也是间接反映了社会的经济水平，老年抚养比是反映人口结构老年化占比，出生率、死亡率则直接反映了人口自然增长率的情况。因而， $F_1$  为反映我国经济发展水平及老龄化的公共因子，在这个因子上得分越高，经济发展水平越高，老龄化越突出，因而  $F_1$  可以归纳为社会经济繁荣 - 老龄化因子。公共因子  $F_2$  上面载荷值较大的仅有三个原始指标，分别是就业人员、少儿抚养比、每十万人口平均在校生数。反映劳动年龄人口占比大、学龄人口相对较少的“劳动力结构”特征，劳动年龄人口越充裕，教育负担相对轻松，因而  $F_2$  可以归纳为劳动年龄人口 - 教育因子。

**Table 5.** Rotated component matrix

**表 5.** 旋转后的成分矩阵

|                      | 成分 1   | 成分 2   | 主要归属      |
|----------------------|--------|--------|-----------|
| 卫生总费用(亿元)            | 0.996  | -0.035 | 因子 1      |
| 居民人均可支配收入(元)         | 0.994  | 0.066  | 因子 1      |
| 人均 GDP(元)            | 0.994  | 0.086  | 因子 1      |
| 居民消费水平(元)            | 0.994  | 0.047  | 因子 1      |
| 教育经费(万元)             | 0.992  | 0.086  | 因子 1      |
| 每万人拥有卫生技术人员数(人)      | 0.992  | -0.016 | 因子 1      |
| 住宅商品房平均销售价格(元/平方米)   | 0.991  | 0.073  | 因子 1      |
| 老年抚养比(%)             | 0.989  | -0.137 | 因子 1      |
| 居民消费价格指数(1978 = 100) | 0.969  | 0.227  | 因子 1      |
| 年末参加生育保险人数(万人)       | 0.965  | 0.246  | 因子 1      |
| 城镇化率(%)              | 0.963  | 0.256  | 因子 1      |
| 人口出生率(‰)             | 0.818  | 0.356  | 因子 1      |
| 人口死亡率(‰)             | -0.803 | 0.46   | 因子 1，且为负向 |
| 就业人员(万人)             | 0.194  | 0.94   | 因子 2      |
| 少儿抚养比(%)             | -0.571 | -0.792 | 因子 2，且为负向 |
| 每十万人口平均在校生数(人)       | 0.561  | -0.775 | 因子 2，且为负向 |

最后，为了量化人口自然增长率与公共因子之间的影响关系，计算公共因子得分。

3.3. 人口自然增长率与公共因子之间的关系

3.3.1. 多元线性回归模型

设因变量为  $y$ ,  $k$  个自变量分别为  $x_1, x_2, \dots, x_k$ , 描述因变量  $y$  如何依赖于自变量  $x_1, x_2, \dots, x_k$  和误差项  $\varepsilon$  的方程称为多元线性回归模型。其一般形式可表示为:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$$

式中,  $\beta_0, \beta_1, \dots, \beta_k$  是模型的参数;  $\varepsilon$  是误差项。

3.3.2. 模型的建立与求解

绘制公共因子与人口自然增长率之间的散点图, 如图 2 所示, 人口自然增长率与两个公共因子之间存在一定的线性相关关系。

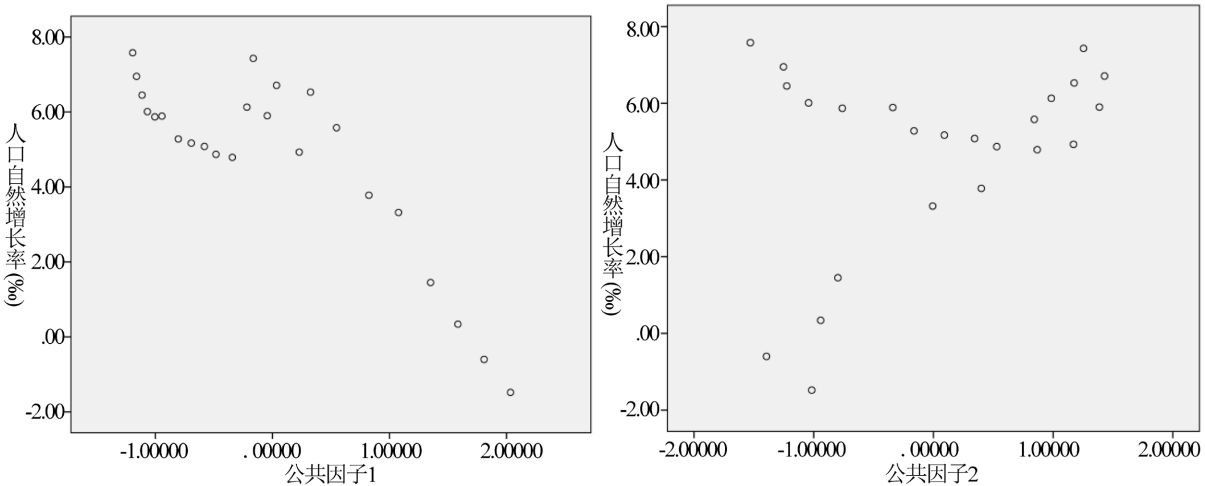


Figure 2. Scatter plot  
图 2. 散点图

Table 6. Model summary  
表 6. 模型摘要

| R     | R 方   | 调整后 R 方 | 显著性 F 变化量 |
|-------|-------|---------|-----------|
| 0.926 | 0.858 | 0.844   | 0         |

Table 7. Model coefficients  
表 7. 模型系数

| 模型     | 未标准化系数 |       | 标准化系数 | t       | 显著性 |
|--------|--------|-------|-------|---------|-----|
|        | B      | 标准误差  | Beta  |         |     |
| (常量)   | 4.748  | 0.198 |       | 23.938  | 0   |
| 公共因子 1 | -2.093 | 0.203 | -0.85 | -10.331 | 0   |
| 公共因子 2 | 0.903  | 0.203 | 0.367 | 4.455   | 0   |

通过“SPSS - 分 - 回归 - 线性”对数据进行多元线性回归分析。根据表 6 和表 7 可知, 调整后的 R 方为 0.844, 表明模型的拟合效果非常不错。故最终线性回归模型:

$$y = 4.748 - 2.093F_1 + 0.903F_2$$

### 3.4. 结果分析

两个公共因子对人口自然增长率(‰)的作用方向相反,且都极其显著( $p < 0.01$ )。  $F_1$  (社会经济繁荣 - 老龄化因子)每提高 1 个单位,人口自然增长率平均下降 2.093 个千分点;  $F_2$  (劳动年龄人口 - 教育因子)每提高 1 个单位,人口自然增长率平均上升 0.903 个千分点。

由此可见,社会经济繁荣 - 老龄化因子是抑制人口自然增长的“主刹车”,劳动年龄人口 - 教育因子只是“小油门”;要想提高人口自然增长率,就必须在“松刹车”和“踩油门”两方面同时发力。一方面,在社会经济繁荣发展的同时,需要完善医疗保障、控制教育成本、降低生育成本等,削弱社会经济繁荣对生育意愿的抑制效应,减缓老龄化对人口自然增长的负面拉力;另一方面,充分挖掘劳动力结构优势,稳定就业规模提升劳动参与率,保障收入稳定,释放劳动年龄人口对生育的潜在促进作用,从而放大“小油门”的正向效应。只有这样,才能在人口结构转型期,实现人口自然增长率的稳步回升。

## 4. 人口自然增长率的预测

### 4.1. ARIMA 模型简介

在时间序列分析中,研究的对象都是在平稳序列的基础上,而生活中的时间序列大部分都是非平稳序列,那非平稳序列如何转化为平稳序列,且尽可能多的提取它的随机信息。Cramer 分解定理中表明,序列中的确定性信息能够通过差分方法进行充分获取。 $d$  阶差分表示:

$$\nabla^d x_t = \sum_{i=0}^d (-1)^{i+1} C_d^i x_{t-i}$$

求和自回归移动平均模型的结构如下所示,简记为  $ARIMA(p, d, q)$ :

$$\nabla^d x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

模型中有三个限制条件:

- (1)  $\phi_p, \theta_q$  不能同时为零。
- (2)  $E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_s \varepsilon_t) = 0, (s \neq t)$ , 要求随机干扰项为零均值的纯随机序列。
- (3)  $E(\varepsilon_s \varepsilon_t) = 0, \forall s < t$ , 过去的序列值与当期的干扰项无相关性。

其实,ARIMA 模型就是进行差分后的 ARMA 模型,模型的条件和性质与 ARMA 保持一致。所以在实际应用中遇到非平稳的时间序列,通常是将非平稳的时间序列通过差分运算转化为平稳的时间序列,继而对平稳时间序列进行建模分析[10]。

### 4.2. ARIMA 模型的建立与求解

为了保证时间序列分析的效果,绘制 1978~2024 年的人口自然增长率的时序图,由图 3 观测,时序图为非平稳序列,同时未见明显的周期特征,因而需对序列进行简单的差分运算,将非平稳序列转化为平稳的时间序列。

通过一阶差分,将非平稳序列转化为平稳序列,为后续建立 ARIMA 模型做好准备,由图 4 可知,表明在一阶差分之后,序列趋于平稳。

为了确认 ARIMA 模型中的参数  $p, q$ ,通过 Python 中提供的一个“自动定阶”工具函数,使用 AIC, BIC 最小准则来确认最优的参数  $p, q$ ,由表 8 所示,  $p = 0, q = 1$ 。结合差分运算,初步确定模型为  $ARIMA(0,1,1)$ 。

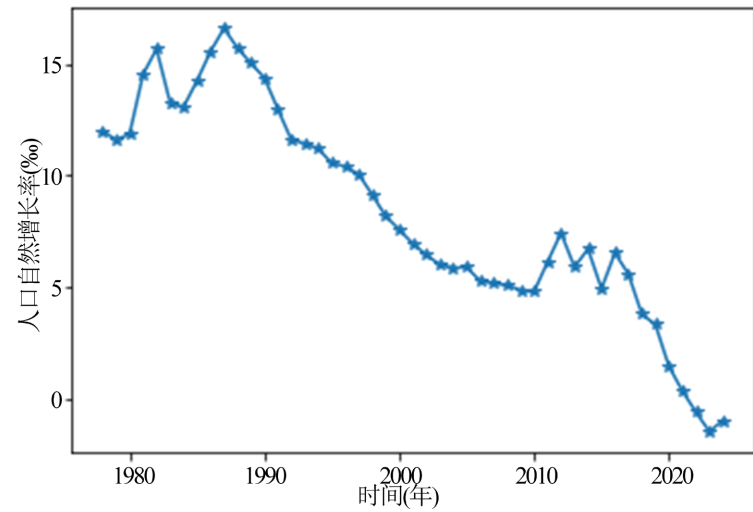


Figure 3. Trends in the natural population growth rate, 1978~2024  
图 3. 1978~2024 年人口自然增长率变化情况

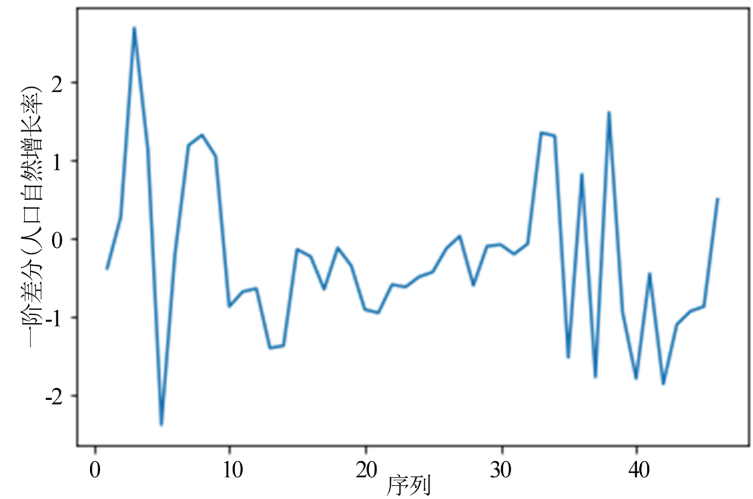


Figure 4. Time-series plot after first-order differencing  
图 4. 一阶差分后时序图

Table 8. Optimal parameter determination  
表 8. 最优参数确定

| 准则  | 最优 $p, q$ |
|-----|-----------|
| AIC | (0, 1)    |
| BIC | (0, 1)    |

Table 9. Residual test results  
表 9. 残差检验结果

|                    | 数值   | p 值  | 含义            |
|--------------------|------|------|---------------|
| Ljung-Box (L1) (Q) | 0.27 | 0.61 | >0.05, 无显著自相关 |
| Jarque-Bera (JB)   | 0.64 | 0.73 | >0.05, 正态性通过  |
| 异方差                | 0.87 | 0.78 | >0.05, 同方差    |

为了确认模型是否可用，对拟合模型的残差序列进行白噪声检验，由表 9 所示，残差既无自相关，也无异方差，且服从正态分布，表明模型检验通过，该模型可用。因此，最终确认模型为： $ARIMA(0,1,1)$ 。

### 4.3. 预测未来 5 年我国人口自然增长率

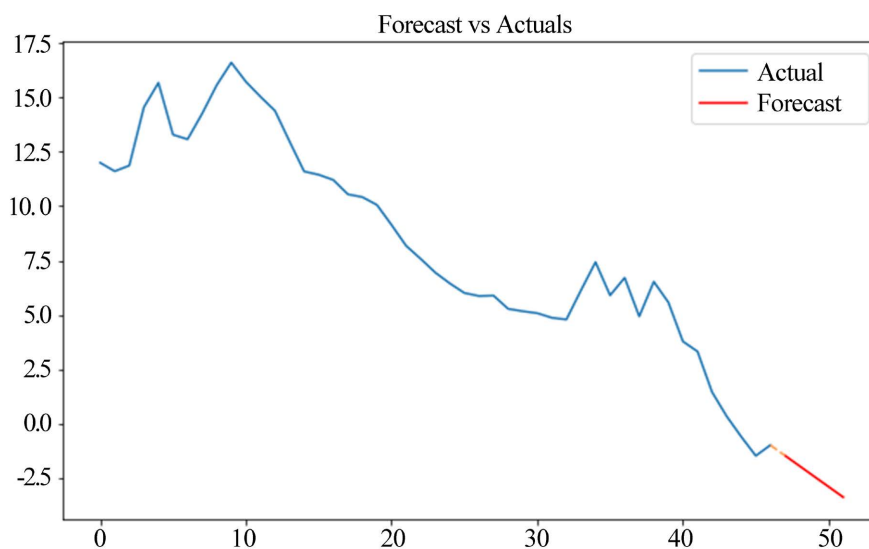
利用最终确认的  $ARIMA(0,1,1)$  模型，预测未来 5 年我国人口自然增长率，预测结果，见表 10 所示。

**Table 10.** Projected natural population growth rate in China for the next five years

**表 10.** 未来 5 年我国人口自然增长率

| 年份   | 预测值(‰)   |
|------|----------|
| 2025 | -1.46856 |
| 2026 | -1.94712 |
| 2027 | -2.42568 |
| 2028 | -2.90425 |
| 2029 | -3.38281 |

预测图见图 5 所示，蓝色曲线展示了 1978~2024 年我国人口自然增长率的实际趋势，由 1978 年的 12.00‰ 高位波动下降至 2024 年的 -0.99‰，已呈现“高台跳水”式下滑；红色线段则给出了 2025~2029 年的预测值，分别为 -1.47‰、-1.95‰、-2.43‰、-2.90‰ 和 -3.38‰。对比可见，未来 5 年我国人口自然增长率负增长态势将延续并加速：年均降幅约 0.48 个百分点，到 2029 年缺口比 2024 年再扩大 2.4 个百分点，人口总量缩减压力进一步加剧，急需通过“松刹车”与“踩油门”并重的组合拳积极应对。



**Figure 5.** Projected natural population growth rate in China for the next five years (‰)

**图 5.** 未来 5 年我国人口自然增长率(‰)

## 5. 结论

1、人口自然增长率抑制因子为社会经济繁荣 - 老龄化因子(“主刹车”)。经济越发达、养育成本越高，生育意愿越低，老龄化越突出，对人口自然增长形成强力下拉；驱动因子为劳动年龄人口 - 教育因子(“小油门”)。劳动年龄人口占比高、教育负担相对轻，可为生育提供潜在正向拉力，但力度有限。

2、人口自然增长率在未来5年将继续“高台跳水”，由2024年的-0.99‰降至2029年的-3.38‰。年均降幅约0.48个百分点，五年累计再扩大负增长缺口2.4个百分点，人口总量缩减压力显著加剧。

3、为了有效改变当前局面，必须双轮驱动、同步发力。把“松刹车”政策置于优先位置，以快速削弱负增长主因；同时激活“小油门”，通过劳动力优势形成持续正向推力。一方面：在社会经济繁荣发展的同时，需要完善医疗、教育、生育等全生命周期生育支持政策。另一方面，提升劳动参与率，保障收入稳定，利用充裕劳动年龄人口，创造生育友好型的就业环境。

## 参考文献

- [1] 黄祺元. 两广人口自然增长率的影响因素分析及预测[D]: [硕士学位论文]. 桂林: 广西师范大学, 2024.
- [2] 巩永丽, 张德生, 武新乾. 人口增长率的非参数自回归预测模型[J]. 数理统计与管理, 2007(5): 759-764.
- [3] 康意, 王国军. “十二五”期间我国人口自然增长率的预测及政策建议[J]. 北京联合大学学报, 2014, 28(1): 64-69.
- [4] 陈丽萍. 浅谈人口与教育[J]. 科技创新导报, 2014(1): 235.
- [5] 张琼枝, 陈望学, 周慧婷, 等. 影响人口增长率的回归分析[J]. 经济研究导刊, 2017(36): 178-180.
- [6] 靳天宇, 刘东浩. 房价对城市人口出生率的影响——基于中国省级面板数据的分析[J]. 山东社会科学, 2019(1): 176-181.
- [7] 肖和宏. 基于计量经济学分析影响人口自然增长的因素[J]. 西部皮革, 2021, 43(19): 136-137.
- [8] 李资博. 我国人口自然增长率的影响因素探寻[J]. 商展经济, 2022(13): 116-118.
- [9] 任雪松, 于秀林. 多元统计分析[M]. 北京: 中国统计出版社, 2010: 393.
- [10] 王燕. 时间序列分析: 基于R[M]. 北京: 中国人民大学出版社, 2015.