Statistics and Application 统计学与应用, 2022, 11(6), 1392-1400
Published Online December 2022 in Hans. http://www.hanspub.org/journal/sa
https://doi.org/10.12677/sa.2022.116145

Hans汉斯

基于ARIMA模型的我国未来人口的预测

王　瑶

云南财经大学统计与数学学院，云南　昆明

收稿日期：2022年11月14日；录用日期：2022年12月4日；发布日期：2022年12月19日

摘　要

随着中国经济的飞速发展，人们的受教育水平也日益提高，我国生育率持续下降，人口增长速度渐缓。本文通过查阅近统计年鉴近四十年来我国人口总数，建立相关ARIMA (0, 2, 1)模型，对我国未来十年人口做出短期预测，发现我国人口增长速度保持在正常范围之内，并据此提出相关建议。

关键词

人口预测，ARIMA模型，时间序列

Prediction of China's Future Population Based on ARIMA Model

Yao Wang

School of Statistics and Mathematics, Yunnan University of Finance and Economics, Kunming Yunnan

Received: Nov. 14th, 2022; accepted: Dec. 4th, 2022; published: Dec. 19th, 2022

Abstract

With the rapid development of China's economy, people's education level is also increasing, China's fertility rate continues to decline, and the population growth rate is slowing down. In this paper, by consulting the recent statistical yearbook, the total population of China in the past 40 years, establishing the relevant ARIMA (0, 2, 1) model, and making a short-term forecast of China's population in the next ten years, we find that China's population growth rate remains within the normal range, and put forward relevant suggestions accordingly.

Keywords

Population Forecast, ARIMA Model, Time Series

文章引用: 王瑶. 基于 ARIMA 模型的我国未来人口的预测[J]. 统计学与应用, 2022, 11(6): 1392-1400.
DOI: 10.12677/sa.2022.116145

1. 研究背景及意义

人口是保障社会和经济持续发展的重要载体，是整个社会的最基础的部分。庞大的人口总量不仅能为经济持续高质量发展提供充足的劳动力，同时也能在科技创新、艺术创作和体育竞技等众多人文领域发挥充足的人才优势。与之相对的，不合理的人口增长速度和人口结构也会导致社会和经济发展受到阻碍，进而影响人们的生活水平和质量以及社会的长期稳定发展。

中国人口数位居世界第一，我国人口总数的变化不仅是我国经济以及社会发展有着不可忽视的影响，同时也对于其他国家也有着不凡的意义，因此也得到了国际社会的长期关注。随着我国生育率持续降低，人口增长趋势渐缓，人口老龄化问题加剧，为了能够更好地应对未来发展中更多的问题和挑战，能够实现我国人口的均衡发展，保证我国社会和经济持续高质量发展，2015 年我国决定全面实施二胎政策。这是我国历史上人口政策方面的又一次重大的转变，这也是影响我国未来几十年的人口结构变动的另一个重要影响因素。

1.1. 研究目的

人口预测数据是国家制定人口、经济和社会发展等宏观发展战略规划中的最基础数据。随着近年来人口生育政策适度调整以及中国人口结构变动等因素的影响，中国的人口发展增速和结构各方面都变得越来越复杂，同时，人口与资源环境的关系、人口老龄化程度不断加深等问题都在不断加深，也变得更加严重，关于如何对未来人口变动趋势做出准确判断的问题不仅是人口学领域的研究重点，同时也是经济学研究的领域，人口基础数量和人口结构对我国经济发展有着重要的影响。当前，学术界已经进行了大量而深刻的关于人口各方面问题的详细研究，并且大家也普遍达成了一个共识，即专家学者们一致认为人口问题是一个发展问题。

1.2. 研究方法

首先，通过大量的文献查阅去获取与我国人口预测方向相关的研究现状，再通过统计年鉴统计我国近四十年来的人口总数，建立 ARIMA 模型进行短期预测，最后根据人口预测趋势和结果，分析其原因并提出相关建议。

2. 文献综述

随着经济高速发展，我国人口数量不断增加，近年来，越来越多的学者致力于研究我国人口发展趋势和未来人口增长趋势的，主要研究内容进行分类整理如下所示。

2.1. 人口发展趋势

人口数据是社会生活和经济发展中各个方面的预测和分析中最基础的数据。陈卫文(2006)基于 2000 年全国第五次人口普查数据以及 2004 年国家统计局统计公布的全国总人口数，对我国 2005~2050 年全国人口发展趋势进行预测分析，同时发现我国生育率在快速下降，我国将长期处于低生育水平[1]。而与生育率水平相关的，即是受教育人口、劳动力水平和人口老龄化。徐警武(2007)发现我国的人口再生产面临重大转型，认为我国 2007~2027 这 20 年间，我国高等教育将由大众化进入普及化，高等教育规模将进一

步扩大，这会导致大量学生入学，造成较大的入学压力[2]。张瑾和黄志龙(2014)研究发现我国未来劳动适龄人口比例与 16~64 岁劳动参与人口比例将持续下降[3]。2015 年实行二胎政策后，王亚楠和钟甫宁(2017)利用 1980~2010 年进入生育期的妇女的初育年龄对其终身生育率进行预测，研究发现，为促进未来人口出生数量能够稳步增长，以此保持人口长期处于均衡发展状态，我国生育政策的目标仍须做出适当的提高[4]。周文(2018)通过假设全面二孩政策实施后我国总和生育率分别以 2.0、1.8、1.5 和 1.18 的水平发展研究发现，只有当生育率随着全面二孩政策的实施提高至更替水平 2 左右时，我国人口结构才能得到改善，否则，我国人口规模将快速微缩，人口年龄结构进一步老化[5]。但二胎政策放开后，我国生育率水平并没有达到 2，赵玉峰和杨宜勇(2019)综合比对多种人口预测数据发现，出生人口规模将波动下降，而我国人口总量下降拐点将在未来 10 年内出现[6]。陈友华(2019)研究发现，我国正处在超低生育率早已形成、人口增长接近尾声、老龄化程度将不断加深[7]。王金营和李天然(2020)以 2002~2014 年我国老年人健康长寿影响因素跟踪调查(CLHLS)数据为基础，研究发现，未来失能老年人口规模将不断扩大，预测 2050 年失能人口约占老年人口总数的 13.68% [8]。

2.2. 人口预测方法

近年来，学者们运用多种方法对我国人口进行过预测。

孟令国、李超令和胡广(2014)采用人口－发展－环境模型(PDE)，预测了我国 2015~2050 年人口结构变化走势，认为实施二胎生育政策比较理想[9]。陈霞和肖岚(2020)在 Logistic 种群增长模型中引入适当形式的收获模型[10]。赵子铭(2019)利用时间序列方法及不同检验、最优化方法建立 ARIMA 模型，预测结果与实际结果相差不大[11]。郭雪峰、黄健元和王欢(2018)基于自适应滤波法对传统灰色模型进行残差修正后，比较了传统灰色模型和改进后模型的预测结果，发现，改进后的模型预测精度更高，适用性与可行性更强[12]。为了使预测结果更加准备，也有学者将两种模型相结合。徐翔燕和侯瑞环(2020)将灰色预测模型和支持向量机模型进行组合，选取一师阿拉尔市 19997~2017 年的人口数据进行分析，对 2018~2022 年人口进行相关预测，发现与单一模型相比，组合模型预测精度会更高，其相对误差也会更低，若以与单一模型相比，组合模型会更合适[13]。唐贤芳、崔岩和张淑丽(2020)，选择 2003~2015 年数据建立静态灰色模型，而后构建等维递补动态模型与静态灰色模型，研究发现动态预测模型在预测精度方面明显好于静态预测模型，其数据预测的可靠性也会更强[14]。对于模型的运用，盛亦男和顾大男(2020)对未来人口预测的方法做出了详细总结，发现目前的预测方法都存在一些缺陷，而基础数据的准确性反而更关键[15]。

2.3. 文献评述

我国学者对于我国人口未来发展趋势持较统一观点，普遍认为我国生育率将快速下降，即使在 2015 年施行二胎政策后，我国人口老龄化的趋势也因此得到减缓，人口老龄化加重仍是当今社会我们需要面对的重要问题，与此同时，我国未来人口总量增长趋势也会随之减缓，这也是我国当前不可避免的问题。在人口预测的模型方面，学者们大多运用灰色预测模型和 ARIMA 模型，因此本文通过查阅统计年鉴相关数据，选取了我国 1980 年至 2019 年的每年人口总数数据，利用 R 语言建立了拟合优度高的 ARIMA 模型，预测了未来十年我国人口总量以及其增长趋势。

3. ARIMA (p, d, q)模型理论分析

3.1. ARIMA 模型原理

ARIMA 模型全称为求和自回归移动平均模型，简记为 ARIMA (p, d, q)模型，是一种时间序列预测方

法，其中 AR (p)为自回归部分，MA (q)为移动平均部分。该模型旨在通过差分运算将非平稳时间序列转化为差分平稳序列，然后利用因变量的滞后值和随机误差项建立模型，以达到对未来值预测的目的，具体的数学表达式如下：

$$\begin{cases}\Phi(B)\Delta^d x_t=\theta(B)\varepsilon_t\\ \mathrm{E}(\varepsilon_t)=0,\ \ Var(\varepsilon_t)=\sigma_\varepsilon^2,\ \ \mathrm{E}(\varepsilon_t,\varepsilon_s)=0,\ \ s\neq t\\ \mathrm{E}(x_s,\varepsilon_t)=0,\ \ \forall s<t\end{cases}$$

3.2. 平稳性检验

图示法是粗略判别 AR (p)模型平稳性的一种方法，特征根判别是精确判别平稳性的方法。AR(p)模型可以简写为：

$$\phi(B)x_t=\varepsilon_t$$

假设 $\lambda_1,\lambda_2,\cdots,\lambda_p$ 是平稳序列 $\{x_t\}$ 线性差分方程的 p 个特征根，任取 $\lambda_i\left(i\in(1,2,\cdots,p)\right)$，带入特征方程，有 $\lambda_i^p-\phi_1\lambda_i^{p-1}-\phi_2\lambda_i^{p-2}-\cdots-\phi_p=0$。假如该方程的所有特征根都在单位圆内即 $|\lambda_i|<1,\ i=1,2,\cdots,p$，则该序列为平稳序列。

3.3. 纯随机性检验

Ljung 和 Box 证明 *LB* 统计量近似服从自由度为 *m* 的卡方分布，数学表达式为：

$$LB=n(n+2)\sum_{k=1}^{m}\left(\frac{\hat{\rho}_k^2}{n-k}\right)\sim\chi^2(m)$$

其中，*n* 为观测期数，*m* 为延迟期数。若 *LB* 统计量小于临界水平，则拒绝原假设，认为序列为非白噪声序列，可以继续拟合模型。

3.4. 模型选择

在建立模型时，通常会有几个模型通过上述检验，此时就需采用下表 1 所示原则来选择相关模型，并采用信息量准备来确定模型的最优阶数。

Table 1. Basic principles of model order determination
表 1. 模型定阶的基本原则

$\dot{\rho}_k$	$\dot{\varnothing}_{kk}$	模型定阶
拖尾	P 阶截尾	AR (p)模型
q 阶截尾	拖尾	MA (q)模型
拖尾	拖尾	ARMA (p, q)模型

3.5. 最小二乘估计

在 ARMA (p, q)模型场合，记

$$\tilde{\beta}=(\varnothing_1,\cdots,\varnothing_p,\theta_1,\cdots,\theta_q)'$$

$$F_t(\tilde{\beta})=\varnothing_1x_{t-1}+\cdots+\varnothing_px_{t-p}-\theta_1\varepsilon_{t-1}-\cdots-\theta_q\varepsilon_{t-q}$$

残差项为：$\varepsilon_t = x_t - F_t\left(\tilde{\beta}\right)$

残差平方和为：$Q\left(\tilde{\beta}\right) = \sum_{t=1}^{n}\left(x_t - \varnothing_1 x_{t-1} - \cdots - \varnothing_p x_{t-p} + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q}\right)^2$，使残差平方和达到最小的参数值为 $\tilde{\beta}$ 的最小二乘估计值。

3.6. 模型预测

用 $e_t(l)$ 衡量预测误差，$e_t(l) = x_{t+l} - \hat{x}_t(l)$，显然预测误差越小预测精度越高。现最常用的预测原则就是预测方差最小原则，即 $Var\left[e_t(l)\right] = \min\left\{Var\left[e_t(l)\right]\right\}$。

4. ARIMA 模型建模分析——以我国近四十年人口总数为例

4.1. 平稳性检验

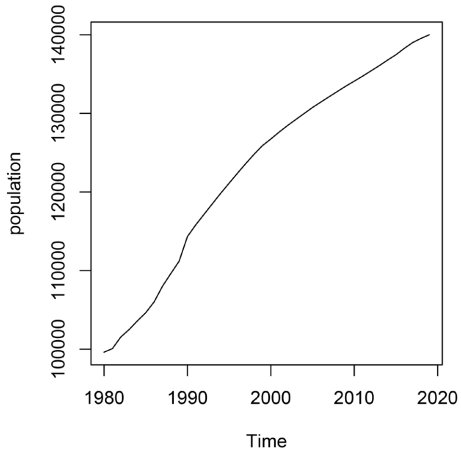

Figure 1. Time series chart of China's total population from 1980 to 2019
图 1. 1980~2019 年中国人口总数时序图

选取统计局官网公布的统计年鉴中 1980 年至 2019 年的总人口数作为观测样本，近四十年来我国人口波动时序图如图 1 所示，人口数量呈稳定增长趋势，对序列进行一阶差分后发现，ADF 检验并未通过，再进行二阶差分，二阶差分后时序图如下图 2 所示。观察二阶差分后时序图，认为时序图具有一定的平稳性。

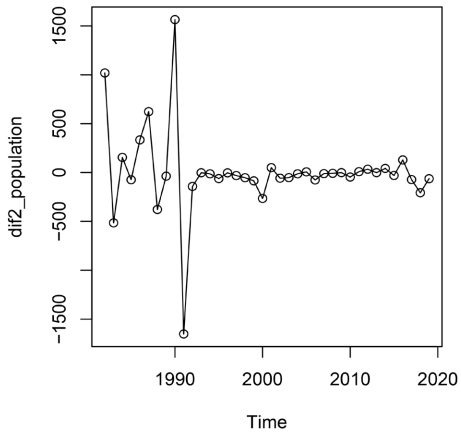

Figure 2. Second order difference sequence diagram of China's total population from 1980 to 2019
图 2. 1980~2019 年中国人口总数二阶差分后序列时序图

由表 2 可知，虽然有趋势有截距项有趋势两种类型下的滞后 4 阶的 P 值均大于 5%临界水平，但无趋势无截距和无趋势有截距项类型下的滞后 1~4 阶 P 值均小于 5%临界水平，因此我国近四十年人口总数二阶差分后序列是平稳的。

Table 2. Unit root test of China's total population from 1980 to 2019
表 2. 1980~2019 年中国人口总数单位根检验

	Type 1: no drift no trend		Type 2: with drift no trend		Type 3: with drift and trend	
lag	ADF	p.value	ADF	p.value	ADF	p.value
0	−10.21	0.01	−10.09	0.0100	−9.97	0.010
1	−7.10	0.01	−7.03	0.0100	−7.05	0.010
2	−4.36	0.01	−4.33	0.0100	−4.34	0.010
3	−3.03	0.01	−3.01	0.0459	−3.08	0.151

4.2. 纯随机性检验

如下表 3 所示，原序列在各阶数下 LB 统计量的 P 值均小于 5%临界水平，因此该序列拒绝原假设，即，我国 1980~2019 年人口总数二阶差分序列为非白噪声序列，可直接拟合模型。

Table 3. White noise test
表 3. 白噪声检验

指标名称	延迟 1 阶	延迟 2 阶
X-squared	6.4627	6.788
p-value	0.01102	0.03357

4.3. 模型识别与定阶

1) 模型识别

我国 1980~2019 年人口总数二阶差分后序列自相关和偏自相关图如下所示。

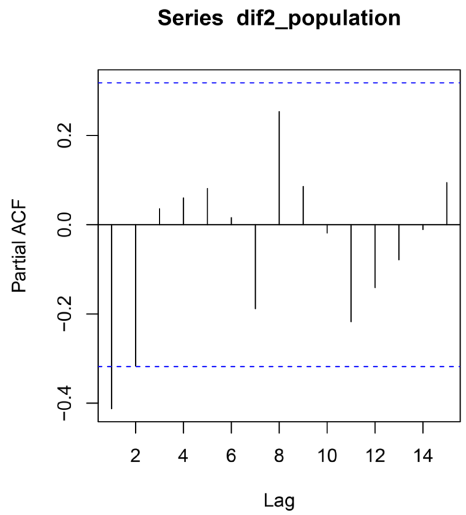

Figure 3. Autocorrelation graph of primitive sequence
图 3. 原序列的自相关图

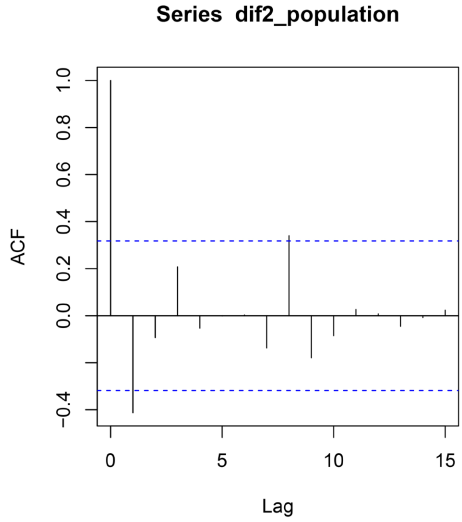

Figure 4. Partial autocorrelation graph of primitive sequence
图 4. 原序列的偏自相关图

由图 3 和图 4 可知，自相关图无法具体确定是几阶截尾或拖尾，因此可以初步拟合 ARIMA (0, 2, 1)、ARIMA (0, 2, 2)、ARIMA (1, 2, 1)和 ARIMA (1, 2, 2)模型，观察其 AIC 值，再确定最终拟合模型，R 在序列自动定阶中给出的合理模型为 ARIMA (0, 2, 1)，因此也加入模型的 AIC 比较中。

2) 模型定阶

Table 4. ARIMA model fitting effect
表 4. ARIMA 模型拟合效果

拟合模型	AIC 信息准则
ARIMA (0, 2, 1)	563.58
ARIMA (0, 2, 2)	564.73
ARIMA (1, 2, 1)	565.17
ARIMA (1, 2, 2)	567.58

如上表 4 所示，ARIMA (0, 2, 1)的 AIC 值最小，同时其模型检验见图 5，根据左下图显示各阶延迟下白噪声检验统计量的 P 值都显著大于 5%临界水平，表示拟合模型 ARIMA (0, 2, 1)的残差序列为白噪声序列，即该拟合模型均显著。

4.4. 模型预测

根据图 6 的时序图可以看出未来十年的人口总数增长趋势与历史波动趋势大致吻合，具体的人口总数预测值见表 5。

5. 结论与建议

5.1. 结论

本文针对我国总人口的预测问题建立了相关 ARIMA 模型，人口总数二阶序列呈现趋势平稳，也通过了纯随机性检验，根据偏自相关和自相关图相结合建立的模型 ARIMA (0, 2, 1)的 AIC 值最小，同时，

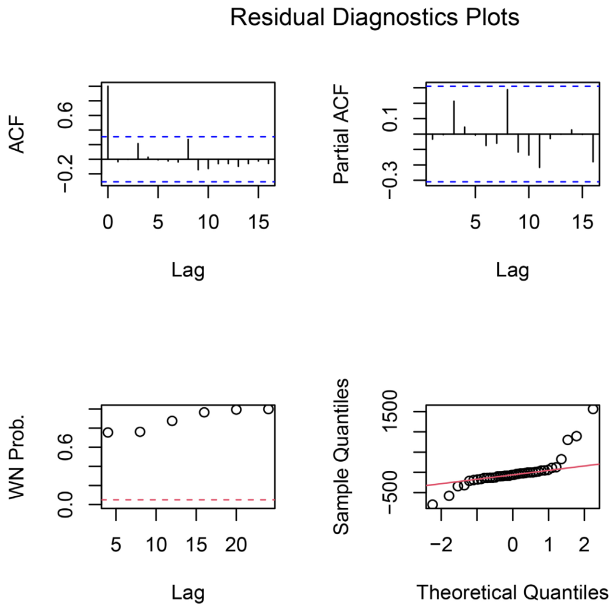

Figure 5. Significance test of ARIMA (0, 2, 1) model
图 5. ARIMA (0, 2, 1)模型的显著性检验

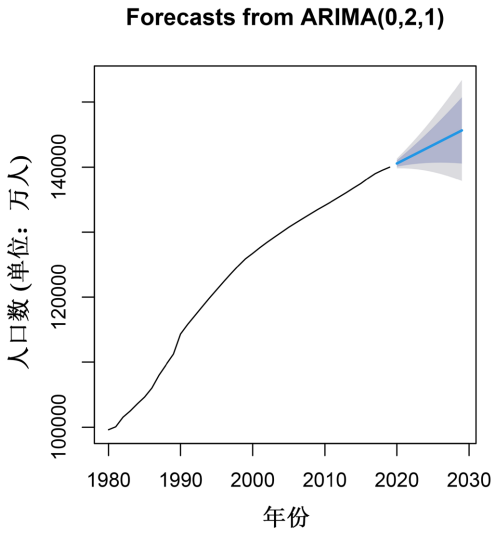

Figure 6. Forecast of China's total population in the future from 2020 to 2029
图 6. 2020 年~2029 年我国未来人口总数预测图

Table 5. Forecast of China's total population in the future from 2020 to 2029
表 5. 我国 2020 年~2029 年未来人口总数预测

Month	Forecast
2020	140569.1
2021	141133.1
2022	141697.2
2023	142261.2

Continued

2024	142825.3
2025	143389.4
2026	143953.4
2027	144517.5
2028	145081.5
2029	145645.6

ARIMA (0, 2, 1)模型显著，于是选择 ARIMA (0, 2, 1)模型对我国未来人口做出预测，其预测结果与历史波动趋势相吻合，而在实施二胎政策后，我国人口增长趋势并没有发生显著性变化，这说明我国人口并没有因为二胎政策而有显著变化。

5.2. 建议

根据我国人口总数预测，提出以下几点建议：

1) 生育政策的调整目标须适当提高。二胎政策实行以来，我国人口数并没有得到显著的增加，说明生育目标仍须适当提高，以此来增加我国新生人口数。

2) 重点解决我国年轻一代就业和住房压力。生育压力主要来自于就业和住房压力，在就业压力和住房压力得到适当缓解后，我国生育水平得到提升，人口增长率也会随着提升，同时，我国经济发展也将得到提升。

参考文献

[1] 陈卫. 中国未来人口发展趋势:2005~2050 年[J]. 人口研究, 2006(4): 93-95.

[2] 徐警武. 人口发展趋势与教育战略规划研究[J]. 内蒙古师范大学学报(教育科学版), 2007(1): 18-21.

[3] 张瑾, 黄志龙. 我国人口发展趋势及对策研究[J]. 宏观经济管理, 2014(1): 59-61+72.

[4] 王亚楠, 钟甫宁. 1990 年以来中国人口出生水平变动及预测[J]. 人口与经济, 2017(1): 1-12.

[5] 周文. 全面二孩政策下中国未来 30 年人口趋势预测[J]. 统计与决策, 2018, 34(21): 109-112.

[6] 赵玉峰, 杨宜勇. 我国中长期人口发展趋势及潜在风险[J]. 宏观经济管理, 2019(8): 11-17+24.

[7] 陈友华. 中国人口发展: 现状、趋势与思考[J]. 人口与社会, 2019, 35(4): 2-17.

[8] 王金营, 李天然. 中国老年失能年龄模式及未来失能人口预测[J]. 人口学刊, 2020, 42(5): 57-72.

[9] 孟令国, 李超令, 胡广. 基于 PDE 模型的中国人口结构预测研究[J]. 中国人口·资源与环境, 2014, 24(2): 132-141.

[10] 陈霞, 肖岚. Logistic 模型的改进与中国人口预测[J]. 成都信息工程大学学报, 2020, 35(2): 239-243.

[11] 赵子铭. 基于 ARIMA 模型的中国人口预测[J]. 赤峰学院学报(自然科学版), 2019, 35(9): 10-12.

[12] 郭雪峰, 黄健元, 王欢. 改进的灰色模型在流动人口预测中的应用[J]. 统计与决策, 2018, 34(8): 76-79.

[13] 徐翔燕, 侯瑞环. 基于 GM(1,1)-SVM 组合模型的中长期人口预测研究[J]. 计算机科学, 2020, 47(S1): 485-487+493.

[14] 唐贤芳, 崔岩, 张淑丽. 基于等维递补灰色模型的人口预测分析[J]. 洛阳师范学院学报, 2020, 39(2): 5-8.

[15] 盛亦男, 顾大男. 概率人口预测方法及其应用——《世界人口展望》概率人口预测方法简介[J]. 人口学刊, 2020, 42(5): 31-46.