

人口拟合实验报告

李奕萱 22 PB22000161

2025 年 4 月 26 日

摘要

本报告深入探讨了通过微分方程模型分析中国人口变化趋势的实验过程和结果。随着中国人口在 2021 年首次出现下降，以及国家开放三胎政策后大城市家庭小孩数平均数仍低于 2 个的现实背景下，研究人口变化趋势对于社会经济发展和政策制定具有重要意义。我们首先系统地介绍了马尔萨斯模型和自限模型的基本原理，分析了它们的适用条件和局限性。在此基础上，针对实际情况中技术进步对资源利用效率的提升以及人口增长的时滞效应，分别引入了考虑技术进步和时间延迟的扩展模型。通过收集的历史人口数据，运用最小二乘拟合方法对模型参数进行估计，并在此基础上对未来十年的人口进行了预测。我们不仅对比了不同模型的预测结果，还详细分析了各模型的优缺点。实验结果有力地表明，该模型能够更精准地捕捉中国人口的变化趋势，为相关政策的制定和调整提供了更具前瞻性的理论支持和数据参考，有助于更好地应对未来人口变化带来的挑战和机遇。

一、前言（问题的提出）

1.1 问题背景

根据最新的第七次全国人口普查数据，2020 年中国人口达到 14.1 亿，这一庞大的人口规模在为国家发展提供丰富劳动力资源的同时，也对资源环境、社会保障等多方面带来了巨大的压力和挑战。然而，2021 年，中国人口自 1949 年以来首次出现下降，跌破 14 亿大关，这一转折性变化引起了社会各界的高度关注和深刻反思。在人口增长放缓的背景下，国家积极调整人口政策，开放三胎政策旨在鼓励家庭增加生育，以期缓解人口老龄化、劳动力减少等问题，但大城市家庭的小孩数平均数仍低于 2 个，这一现象反映出在现代化、城市化进程中，家庭生育意愿受到多重因素的制约，如养育成本高、工作与生活平衡难度大、教育和住房压力等，使得政策的实施效果尚未达到预期。这一系列复杂的人口变化情况，凸显了深入研究和分析中国人口变化趋势的紧迫性和必要性。

1.2 研究目的

本实验旨在通过微分方程模型对中国人口的变化趋势进行深入分析，并在此基础上预测未来十年的人口动态变化。通过构建和求解马尔萨斯模型、自限模型以及更复杂的扩展模型，我们将揭示不同因素如何相互作用影响人口增长，并评估各种政策干预措施的潜在效果。实验结果将为人口政策的优化、资源分配的合理规划以及社会经济的可持续发展提供科学依据和决策支持，助力应对人口变化带来的挑战 and 机遇。

二、相关工作

2.3 马尔萨斯模型

英国经济学家托马斯·罗伯特·马尔萨斯在其于 1798 年发表的《人口原理》中提出了第一个人口模型——**指数增长模型**（也称 **马尔萨斯模型**）。该模型基于一个关键假设：人类的相对出生率（即单位时间内平均每一人新生的婴儿数） b 和相对死亡率 d 均为常数，因此人口的自然增长率 $r = b - d$ 也为常数。

设时刻 t 的人口数为 $N(t)$ ，初始时刻 t_0 的人口数为 N_0 ，马尔萨斯认为 $N(t)$ 满足微分方程

$$\begin{aligned}\frac{dN}{dt} &= rN \\ N(t_0) &= N_0\end{aligned}$$

通过求解上述微分方程，可以得到人口随时间变化的解：

$$N(t) = N_0 e^{r(t-t_0)}$$

马尔萨斯模型表明，在自然增长率 r 保持不变的情况下，人口将呈指数级增长。然而，这种指数增长模式在实际中是不可持续的，因为它忽略了资源限制、环境承载力等因素对人口增长的制约作用。

2.4 自限模型

观察马尔萨斯模型可以发现，其预测的人口呈指数级增长，这显然与常识不符，因为人口不可能无限制地增长。为了解决这一问题，自限模型（也称 **Logistic 阻滞增长模型**）在马尔萨斯模型的基础上引入了环境容纳量的概念，从而对人口增长进行了限制。

假设在一定的环境下，资源所能供养的最多人数为 K （称为极限人口或环境容纳量）。当人口较少时，其自然增长率为 r ；但随着人口的增加，资源的限制作用逐渐显现，人口增长率会逐渐降低。具体地，当前人口数为 N 时，将人口增长率取为 $r(1 - \frac{N}{K})$ ，这是一个相对合理的选择。在此假设下，人口总数满足以下微分方程：

$$\frac{dN}{dt} = r \left(1 - \frac{N}{K} \right) N$$
$$N(t_0) = N_0$$

解得

$$N(t) = \frac{K}{1 + \left(\frac{K}{N_0} - 1 \right) e^{-r(t-t_0)}}$$

根据自限模型，随着时间的增加，人口总数将逐渐趋于极限人口 K ；当人口数达到 $K/2$ 时，人口增速最快；根据此模型预测出的人口随时间的变化为一条 S 形曲线。自限模型考虑了资源限制对人口增长的影响，使得人口增长更加符合实际情况。

2.5 考虑技术进步的模型

技术进步会提高资源利用效率，从而影响人口承载能力。为了更准确地反映技术进步对人口增长的影响，可以引入技术进步参数，建立动态环境容纳量的自限模型。

假设技术进步会动态提高环境容纳量 K 。一种可能的表达式为：

$$K(t) = K_0 \times e^{at}$$

其中， K_0 是初始环境容纳量， a 是技术进步速率。这表明环境容纳量随时间呈指数增长，反映了技术进步对资源利用效率的显著提升作用。此外，考虑到技术进步导致环境容纳量提高不一定为指数级增长，还可以引入对数增长模型：

$$K(t) = K_0 \times a \log t$$

其中， a 是对数增长系数。对数增长模型表明环境容纳量随时间增长的速度逐渐减缓，这可能更符合技术进步在不同阶段对人口承载能力的影响。

通过引入技术进步参数，模型能够更好地反映技术发展对人口增长的促进作用，同时也为预测未来人口变化提供了更丰富的分析视角。

2.6 考虑时间延迟的模型

人口增长可能受到过去人口状态的影响，例如生育高峰的滞后效应。为了捕捉这种时滞现象，可以引入时间延迟项，建立延迟微分方程模型。

建立延迟微分方程：

$$\frac{dN}{dt} = rN(t - \tau)$$

其中, τ 是延迟时间, 反映了人口增长受到过去人口状态影响的时间滞后。

这种模型能够更好地描述人口的周期性波动, 例如由于生育高峰和低谷交替出现导致的人口数量的起伏变化。

延迟微分方程模型通过引入时间延迟项, 使人口增长模型更加贴近实际情况, 有助于更准确地预测人口变化的动态过程。

三、问题分析

接下来, 我们对模型进行进一步的分析与优化, 以提高其对人口变化趋势预测的准确性。

在本次实验中, 我们使用 Python 语言实现了马尔萨斯模型、自限模型、考虑技术进步的模型和考虑时间延迟的模型。将数据集划分为训练集和测试集, 使用训练集拟合模型参数。具体地, 我们采用最小二乘拟合法, 通过最小化模型预测值与历史观测值之间的误差平方和, 来估计模型中的未知参数。在此基础上, 我们在测试集上验证模型的预测准确性, 通过计算相对均方误差 (MSE) 和相对平均绝对误差 (MAE) 等指标, 对模型的性能进行全面评估。

最后, 我们对未来十年的人口进行了预测, 并对比了不同模型的性能。

四、符号说明

表 1: 符号说明

符号	说明	单位
$N(t)$	时刻 t 的人口数	人
t	时间	年
r	人口自然增长率	无量纲
K	环境容纳量 (极限人口)	人
t_0	初始时间	年
N_0	初始时刻的人口数	人
a	技术进步速率	无量纲
τ	时间延迟	年

五、结果 (与对比)

5.7 实验过程

1. **数据准备:** 收集中国过去几十年的人口数据, 并整理成适合模型输入的格式。

2. **数据划分**: 将数据划分为训练集和测试集, 训练集用于模型参数拟合, 测试集用于验证模型的预测准确性。
3. **模型拟合**: 使用最小二乘法对马尔萨斯模型、自限模型、考虑技术进步的模型和考虑时间延迟的模型分别进行参数拟合。
4. **模型预测**: 利用拟合得到的参数, 对未来十年的人口进行预测。
5. **结果评价**: 计算模型在训练集和测试集上的均方误差 (MSE) 和平均绝对误差 (MAE), 评价模型的性能。

5.8 结果展示

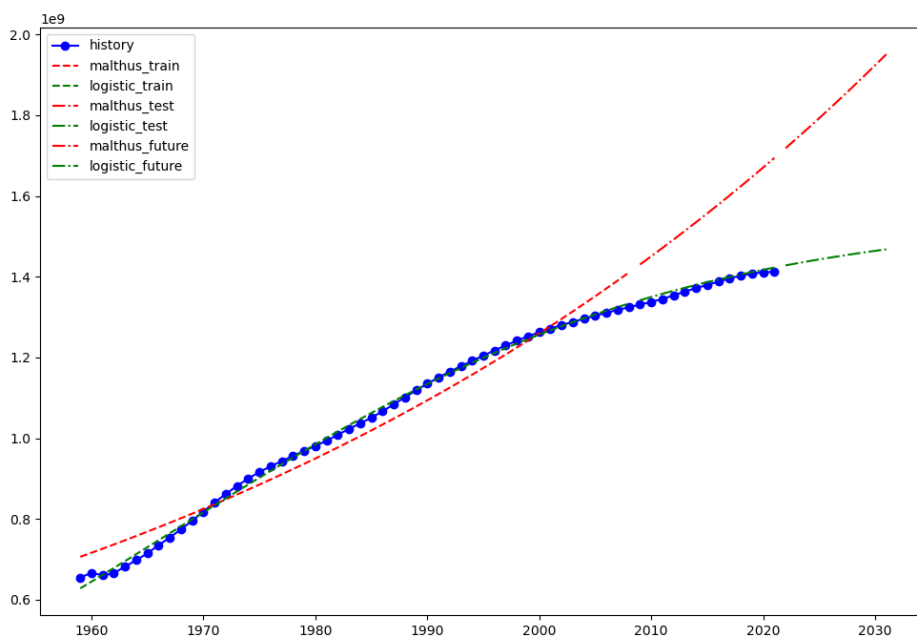


图 1: 基础模型

模型方法	MSE (训练集)	MAE (训练集)	MSE(测试集)	MAE (测试集)
马尔萨斯模型	0.0008	0.0250	0.0181	0.1289
自限模型	0.0000	0.0055	0.0000	0.0057
技术进步模型 (指数)	0.0000	0.0052	0.0001	0.0086
技术进步模型 (对数)	0.0001	0.0068	0.0041	0.0523
时间延迟模型	0.0000	0.0050	0.0004	0.0166

表 2: 误差评价 (相对误差)

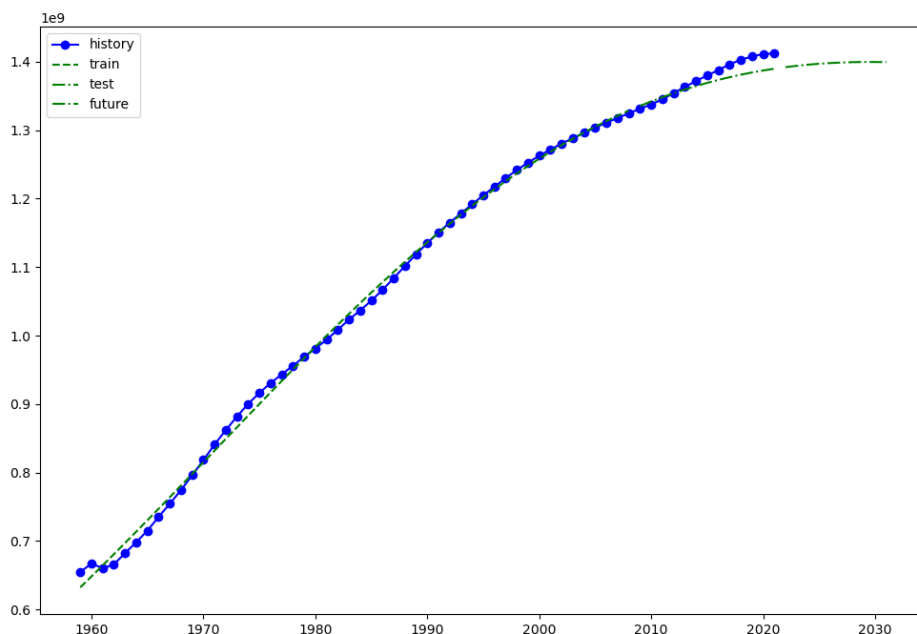


图 2: 技术进步模型 (指数)

5.9 结果对比

从误差评价表中可以看出, 自限模型在训练集和测试集上的 MSE 和 MAE 均为最低, 表明其拟合效果和预测准确性最好。该模型在平衡数据拟合和泛化能力方面表现出色, 能够较为精准地捕捉人口变化的整体趋势, 同时避免了对数据中随机噪声的过度拟合, 具有较高的鲁棒性和可靠性。

考虑技术进步和时间延迟的扩展模型在训练集上的误差与自限模型相当, 但在测试集上的误差略有增加, 这可能是由于模型复杂度增加导致的过拟合现象。这些扩展模型引入了更多参数和复杂结构以捕捉人口变化中更细致的特征, 但也更容易受到数据中噪声和异常值的影响, 从而在未见过的数据上表现不如自限模型稳定。尽管如此, 它们能够更好地捕捉人口变化的细节特征, 如技术进步对环境容纳量的动态影响以及时间延迟对人口周期性波动的作用, 为人口变化的深入分析提供了更丰富的视角和更全面的理解。

与基础模型相比, 扩展模型的优势在于其能够考虑更多现实因素的影响, 从而在理论上更贴近实际人口变化的复杂性。例如, 考虑技术进步的模型能够反映出技术发展对资源利用效率提升进而促进人口增长的作用; 考虑时间延迟的模型则能够模拟生育行为的时滞效应, 这对于预测人口结构变化和制定相应政策具有重要意义。然而, 模型复杂度的增加也带来了更高的数据需求和参数估计难度, 需要更丰富、更高质量的数据支持以及更精细的模型调优策略来充分发挥其预测潜力。

总体而言, 自限模型以其简洁性和高效性在本次实验中展现了最佳的预测性能, 适合用于对人口变化的大趋势进行快速、准确的预测。而扩展模型则在探索人口变化的细微机制和复杂影响因素方面具有独特价值, 通过对它们的深入研究和改进, 可以进一步提升人口预测的科学性和实用性, 为政策制定者提供更全面、更深入的决策依据。

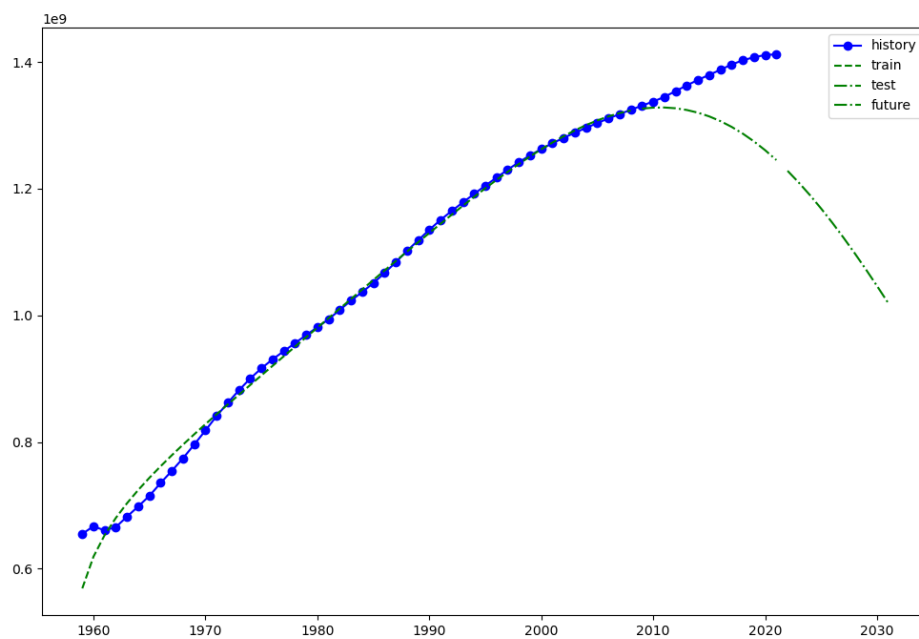


图 3: 技术进步模型 (对数)

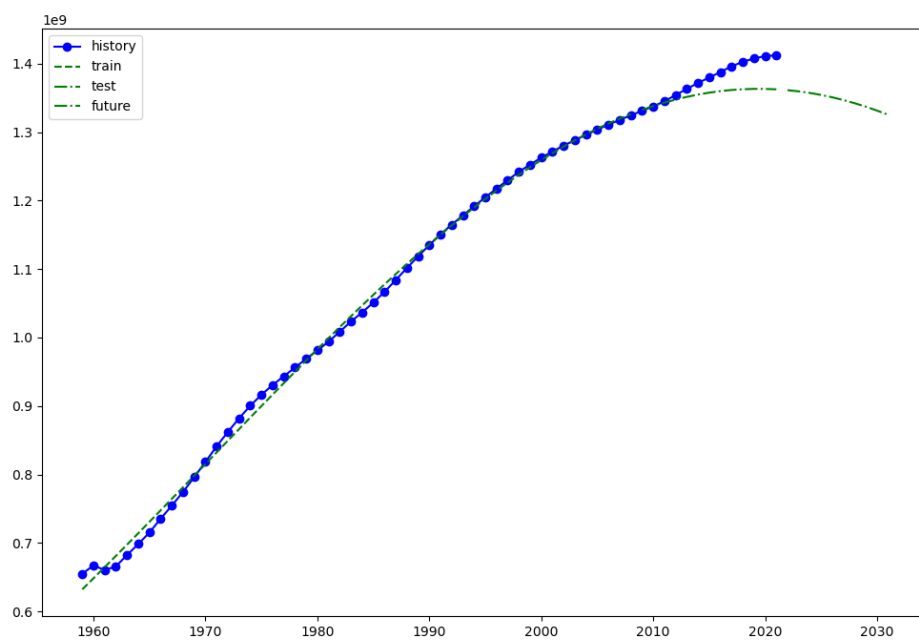


图 4: 时间延迟模型

六、结论

通过本次实验，我们利用微分方程模型对中国人口的变化趋势进行了深入分析和预测。实验结果表明：

- 1. 自限模型表现最佳：**自限模型在训练集和测试集上的误差均为最低，能够最准确地拟合历史数据和预测未来人口变化。这表明在考虑资源限制和环境承载力的情况下，自限模型能更真实地反映人口增长的规律。其稳定的性能和较高的预测精度使其成为本次实验中最为可靠的模型选择。
- 2. 扩展模型的补充价值：**考虑技术进步和时间延迟的扩展模型虽然在预测准确性上略逊于自限模型，但它们能够更好地捕捉人口变化中的细微特征。例如，考虑技术进步的模型能够动态反映技术发展对环境容纳量的提升作用，而时间延迟模型则能体现生育行为的滞后效应。这些模型为政策制定者提供了更全面的视角，有助于深入理解人口变化的复杂机制，并制定更具针对性和前瞻性的政策。
- 3. 模型的局限性：**所有模型均基于历史数据进行拟合和预测，存在一定的局限性。它们未能充分考虑未来可能发生的重大政策变化、社会经济转型、突发公共卫生事件等不确定因素对人口的潜在影响。此外，模型对数据质量和数量有较高要求，数据的不准确或不完整可能导致预测结果的偏差。在实际应用中，需结合专家判断和情景分析等方法，对模型预测结果进行综合评估和调整。

综上所述，本次实验所采用的微分方程模型在分析和预测中国人口变化趋势方面具有一定的有效性和实用性。自限模型以其简洁性和高效性在本次实验中展现出最佳的预测性能，而扩展模型则在探索人口变化的复杂性方面具有独特价值。未来研究可通过进一步优化模型结构、引入更多相关变量、提高数据质量等途径，不断提升人口预测的准确性和可靠性，为应对人口变化挑战提供更有力的决策支持。

七、问题与改进方向

在本次实验中，我们发现以下问题和改进方向：

- 1. 数据质量与完整性：**人口数据的准确性和完整性对模型性能有重要影响。现有数据可能存在缺失值、异常值或测量误差等问题，这些问题会直接影响模型的拟合效果和预测准确性。未来应进一步收集和整理更详细、更准确的人口数据，包括年龄结构、性别比例、地域分布等多维度信息。此外，还可以探索数据预处理方法，如数据清洗、缺失值填充和异常值检测等，以提高数据质量。
- 2. 模型的动态适应性：**人口系统受到多种动态因素的影响，如政策调整、经济发展、科技进步等。这些因素的变化可能导致人口增长模式发生显著变化，而静态模型可能难以适应这些变化。未来可以探索开发动态模型，实时更新模型参数，以适应人口变化的新趋势。例如，可以采用时间序列分析方法或在线学习算法，使模型能够根据新数据及时调整参数，从而更好地捕捉人口变化的动态特征。
- 3. 多模型融合：**不同模型具有各自的优势和局限性。自限模型在整体趋势预测上表现最佳，而扩展模型则能够捕捉到更细致的特征。未来可以研究多模型融合方法，综合各模型的预测结果，提高预测的准确性和可靠性。例如，可以采用加权平均、Stacking 或 Bagging 等融合策略，充分发挥各模型的优势，弥补单一模型的不足。

4. **模型的解释性与可视化**：复杂模型虽然能够提供更精准的预测，但往往缺乏直观的解释性和可视化效果。未来可以开发更具解释性的模型或可视化工具，帮助政策制定者和公众更好地理解模型的预测结果和内在机制。例如，可以采用决策树、规则模型或 SHAP 值分析等方法，增强模型的可解释性；同时，利用数据可视化技术，直观展示人口变化的趋势、结构和影响因素。
5. **跨学科研究与合作**：人口问题涉及到多个学科领域，如经济学、社会学、生态学等。未来可以加强跨学科研究与合作，整合不同领域的理论和方法，构建更全面、更综合的人口模型。例如，可以结合经济学模型分析人口变化对经济发展的反馈作用，或利用社会学理论探讨文化、教育等因素对生育行为的影响。

通过针对上述问题采取相应的改进措施，我们有望在未来的研究中进一步提升人口预测模型的性能和实用性，为应对复杂多变的人口形势提供更有力的支持。