

基于新陈代谢GM-ARIMA组合模型的广东省人口老龄化预测研究

张集锦

华南师范大学数学科学学院, 广东 广州

收稿日期: 2024年3月19日; 录用日期: 2024年4月9日; 发布日期: 2024年4月18日

摘要

本文基于2000~2022年广东省人口结构数据, 利用组合权重系数法构建新陈代谢GM(1, 1)-ARIMA模型对其未来人口结构变化情况进行预测研究。为了弥补ARIMA模型样本需求量大, 拟合更多反映线性趋势的劣势, 克服传统灰色预测在中长期预测的不可操作性和指数爆炸增长导致的预测偏离问题, 首先利用最小二乘、MAPE和组合权重系数法构建基于新陈代谢GM(1, 1)-ARIMA的组合预测模型, 接着引入TIC, MAPE和RMSE三个评价指标评估不同组合模型的精度, 最终选取利用组合权重系数法构建的组合模型进行拟合预测。预测结果表明: 该组合模型比单项模型预测精度提高0.32%, 具有参考价值。同时表明广东人口年龄结构较为年轻, 但进入初始少子化社会, 未来将不可避免地以大规模、高速度进入深度老龄化和超老龄化社会, 需引起高度重视。

关键词

人口老龄化, 新陈代谢GM(1, 1), ARIMA模型, 组合预测模型

Prediction of Population Ageing in Guangdong Province Based on Metabolic GM-ARIMA Combined Model

Jijin Zhang

School of Mathematical Sciences, South China Normal University, Guangzhou Guangdong

Received: Mar. 19th, 2024; accepted: Apr. 9th, 2024; published: Apr. 18th, 2024

Abstract

Based on the demographic data of Guangdong Province from 2000 to 2022, the article uses the

combined weight coefficient method to construct a metabolic GM(1, 1)-ARIMA model to forecast its future demographic changes. In order to make up for the disadvantages of the ARIMA model with high sample requirement and fitting more reflective of a linear trend, and to overcome the inoperability of traditional grey prediction in medium and long-term prediction and the problem of prediction deviation caused by exponential explosive growth, firstly, the combined prediction model based on metabolism GM(1, 1)-ARIMA was constructed by using the method of least squares, MAPE and combined weight coefficients, and then, the three models of TIC, MAPE and RMSE were introduced to assess the accuracy of different combinatorial models, and finally the combinatorial model constructed using the combined weight coefficient method was selected for fitting prediction. The prediction results show that the combined model is 0.32% more accurate than the single model, which is of reference value. It also shows that the age structure of Guangdong's population is relatively young, but it has entered the initial oligocephalic society, and will inevitably enter the deep aging and super-aging society in the future on a large scale and at a high speed, which needs to be paid great attention to.

Keywords

Population Ageing, Metabolism GM(1, 1), ARIMA Model, Combined Forecasting Model

Copyright © 2024 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

对于任何一个国家或地区，可持续发展的核心因素是人口的发展。随着时间的推移及经济的发展，我国社会人口的年龄结构发生了新的变化：生育率持续下降，老年人口持续增长，人口老龄化程度不断加深等等。为了进一步解决养老问题，探讨出一种准确预测未来人口发展趋势的方法具有重大的现实意义。在数据预测领域的模型选用中，BP 神经网络、灰度预测、时间序列分析、回归分析等单项模型曾被广泛应用，张栗粽等[1]探究时间序列模型在金融数据领域的预测应用，曾波等[2]利用灰度建模预测应急物资的需求情况，尽管利用单项模型进行数据预测的方法已经十分成熟，但因构建模型的思路有所差异，难免会有所欠缺，如 GM(1, 1)模型在计算时需要进行多次累加，且难以体现出数据中具有概率分布特征的规律；常规的单一时间序列分析模型难以捕捉到数据序列中的非线性部分；神经网络模型则因非线性拟合能力过强而导致数据预测的泛化能力较差。因此，使用多个模型组合后进行权重的重分配是当前主流的预测思路。伍建军等[3]利用组合权重探索了机器人的可靠性分配方法；卓毅鑫等[4]采用优化权重结构的方法对多模式融合风机气温进行了预测。陈昱杉[5]采用基于 ARIMA-GM 耦合模型对中国农村居民人均可支配收入进行预测研究。考虑到传统灰色预测模型 GM(1, 1)往往会忽略在系统发展中新干扰因素，无法进行较准确合理的中长期预测，王宁等人[6]采用新陈代谢 GM(1, 1)模型对重庆市人口老龄化进行预测研究，并表明新陈代谢 GM 模型相比于传统灰色 GM 模型效果更佳。

因此，为了弥补时间序列模型样本需求量高，拟合更多反映线性趋势的劣势，同时克服传统灰色预测在中长期预测的不可操作性和指数爆炸增长导致的预测偏离问题，本文以广东省人口为研究对象，借助广东省统计年鉴搜集 2000~2022 年广东省人口结构基础数据，运用新陈代谢 GM(1, 1)-ARIMA 组合模型预测广东省 2023~2030 年的人口结构发展趋势。

2. 方法

2.1. ARIMA 模型

ARIMA 模型全称为自回归差分移动平均模型,其基本思想是试图通过数据的自相关性和差分的方式,提取出隐藏在数据背后的时间序列模式,然后用这些模式来预测未来的数据[7]。主要表达式为:

$$\begin{cases} \Phi(B)\Delta^d x_t = \Theta(B) \\ E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ E(\varepsilon_t x_s) = 0, \forall s < t \end{cases} \quad (1)$$

式中, $\Delta^d = (1-B)^d$; $\Phi(B)$ 为移动平均系数多项式, $\Theta(B)$ 为自回归系数多项式。对于非平稳时间序列, ARIMA(p, d, q)模型通过 d 阶差分运算将序列平稳化后,再利用 ARMA(p, q)模型对平稳后的序列进行模型定阶与拟合,从而实现非平稳时间序列的预测分析。

2.2. 新陈代谢 GM(1, 1)模型

新陈代谢 GM(1, 1)模型是灰色预测模型的一种优化形式。常规的灰色 GM(1, 1)模型只考虑 $\hat{x}^{(0)}(k)$ 以前的历史数据,往往会忽略在系统发展过程中不断进入系统的新干扰因素,只具有较好的短期预测效果,无法进行较准确的中长期预测。为了弥补上述缺陷,获得精度更高、时间更久远的预测,在继承常规 GM(1, 1)模型优点和充分考虑相继不断进入系统的随机扰动项和驱动因素的基础上,引入了新陈代谢 GM(1, 1)模型。预测原理是将 GM(1, 1)模型预测的最新数据 $\hat{x}^{(0)}(k+1)$ 加入到原始数据序列 $\{x^{(0)}\}$ 中,再去掉最老的数据 $\hat{x}^{(0)}(1)$,以保证数据序列的维度不变,然后用最新的数据序列 $x^{(0)} = (x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(k), x^{(0)}(k+1))$ 重复进行常规 GM(1, 1)模型的建立与检验,预测出新数据 $\hat{x}^{(0)}(k+2)$,然后再将 $\hat{x}^{(0)}(k+2)$ 加入到 $x^{(0)} = (x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(k), x^{(0)}(k+1))$,再去掉 $\hat{x}^{(0)}(2)$,形成新数列,继续 GM(1, 1)模型的预测与检验。如此反复,依次递补,直到完成预测目标,此即为新陈代谢 GM(1, 1)模型。

2.3. 组合模型构建

1969 年,由 Bates 等人提出组合预测模型方案,即先从各种单项预测模型中提取出有效的系统信息,找到一种准则或方式,使这些不同的预测模型进行合理有效地组合,并选择合适的权系数进行加权,最终得到最合理的组合预测模型。在进行某具体问题预测前,需要对该问题涉及到的数据进行采集,即 $Y = (y_1, y_2, \dots, y_n)$ 。在本题中,本文选取两个预测模型 ARIMA 和新陈代谢 GM(1, 1)模型,在 t 时刻的预测值分别记为 $(\varphi_1(t), \varphi_2(t))$,预测对应的权重为 $\lambda = (\lambda_1, \lambda_2)$,满足条件 $\lambda_1 + \lambda_2 = 1$,最终得到组合预测模型表达式为 $\varphi(t) = \lambda_1 \varphi_1(t) + \lambda_2 \varphi_2(t)$ 。

在模型构建的基础上进行数据拟合后,可根据预测值与实际值进行比对分析,即通过误差来确定权重,常用的方法有拟合误差、对数误差与相对误差。本文使用最小二乘法与平均绝对百分数误差(MAPE)共同确定出组合模型的权重系数。

2.3.1. 最小二乘法确定权数

最小二乘法,简单来说是以拟合误差平方和达到最小来确定权重系数。设 $\varphi(t) = \beta_1 \varphi_1(t) + \beta_2 \varphi_2(t)$ 为 $x(t)$ 的组合预测值;设 $e(t)$ 为组合预测在 t 时刻的预测误差,可得 $e(t) = \varphi(t) - x(t)$;设 β_1, β_2 分别为 2 种预测方法的权重系数,权重系数满足条件为: $\beta_1 + \beta_2 = 1, \beta_1 \geq 0, \beta_2 \geq 0$ 。设 Q 为组合预测的误差平方和,则:

$$Q = \sum_{i=1}^n |e_i^2| = \sum_{i=1}^n |\beta_1 e_1^2(t) + \beta_2 e_2^2(t)| \tag{2}$$

进一步转化为以误差平方和最小为准则的方程求解问题，表达式为：

$$\begin{aligned} \min Q &= \sum_{i=1}^n |\beta_1 e_1^2(t) + \beta_2 e_2^2(t)| \\ \text{st. } &\beta_1 + \beta_2 = 1, \beta_1 \geq 0, \beta_2 \geq 0 \end{aligned} \tag{3}$$

最终的组合预测模型表达式为：

$$\varphi^{(1)}(t) = \beta_1 \varphi_1(t) + \beta_2 \varphi_2(t) \tag{4}$$

2.3.2. MAPE 法确定权数

平均绝对百分数误差(MAPE)常作为衡量模型精度的指标之一，其误差大小反映了模型优劣的程度，具体的数学表达式为：

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\varphi(t) - x(t)}{x(t)} * 100\% \right|, t = 1, 2, \dots, n \tag{5}$$

在组合模型中利用 MAPE 法进行权重系数分配时所遵循的原则，即 MAPE 数大的模型在组合模型中分配到的权重系数小。在本题中，应将 2 个单向模型的 MAPE 按照递增顺序排列，依次得到 d_1, d_2 ，则第 i 个模型的权重系数表达式为：

$$\alpha_i = \frac{d_i}{d_1 + d_2}, i = 1, 2 \tag{6}$$

由上述方法计算后得到的组合模型表达式为：

$$\varphi^{(2)}(t) = \alpha_1 \varphi_1(t) + \alpha_2 \varphi_2(t) \tag{7}$$

2.3.3. 组合权重系数

经过上述两种权重计算方法的论述，综合比较分析后得出了一种组合使用两种上述方法的权重表达式，假设使用 $\lambda_i, i = 1, 2$ 表示第 i 个模型的最终权重系数如下：

$$\lambda_i = \frac{\alpha_i \beta_i}{\alpha_1 \beta_1 + \alpha_2 \beta_2}, i = 1, 2 \tag{8}$$

式中， α_i 为第 i 个利用 MAPE 为准则确定的权重系数； β_i 为第 i 个利用最小二乘法为准则确定的权重系数。这种方法将两种单项权重系数进行结合，最终模型给出的预测结果受到两方面的制约。具体表达式为：

$$\varphi^{(3)}(t) = \lambda_1 \varphi_1(t) + \lambda_2 \varphi_2(t) \tag{9}$$

综上所述，得到组合模型权重分配表 1 如下。

Table 1. Combined model weight allocation table
表 1. 组合模型权重分配表

权重系数	ARIMA 模型	新陈代谢 GM 模型
最小二乘法权重	β_1	β_2
MAPE 权重	α_1	α_2
组合权重系数	$\frac{\alpha_1 \beta_1}{\alpha_1 \beta_1 + \alpha_2 \beta_2}$	$\frac{\alpha_2 \beta_2}{\alpha_1 \beta_1 + \alpha_2 \beta_2}$

2.4. 模型评价函数

在预测模型精度的评估中,常常使用 5 种基于预测误差的评价指标:均方根误差(RMSE)、R-平方(R^2)、平均绝对百分误差(MAPE)、平均绝对误差(MAE)和希尔不等系数(TIC)。本文采用 TIC、RMSE 和 MAPE 对不同预测模型进行精度评价。具体数学表达式如下:

$$\begin{aligned} \text{TIC} &= \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (\varphi(t) - x(t))^2}}{\sqrt{\frac{1}{n} \sum_{i=1}^n \varphi(t)^2 + \frac{1}{n} \sum_{i=1}^n x(t)^2}} \\ \text{RMSE} &= \sqrt{\frac{1}{n} \sum_{i=1}^n (\varphi(t) - x(t))^2} \\ \text{MAPE} &= \frac{1}{n} \sum_{i=1}^n \left| \frac{\varphi(t) - x(t)}{x(t)} * 100\% \right| \end{aligned} \quad (10)$$

3. 广东省人口老龄化现状

本文所用数据来源于《广东省统计年鉴》，从中搜集整理出 2000~2022 年广东省的人口结构的基础数据对其特征进行分析。

3.1. 65 岁及以上老年人口创历史新高

根据国际通行判断标准,广东省早在 2013 年 65 岁及以上老年人口比重首次突破 8%,进入老龄化社会。从图 1 数据来看,广东省 65 岁及以上老年人口比重从 2010~2022 年不断增加,整体保持持续增长趋势,并于 2022 年达到顶峰。常住人口中 65 岁及以上人口规模为 1214.4 万人,达到历史新高,比 2021 年增加近 57.4 万人,增幅达到 4.9%。65 岁及以上人口比重从 2021 年的 9.12% 上升 9.59%,人口老龄化态势持续加快、程度持续加深。

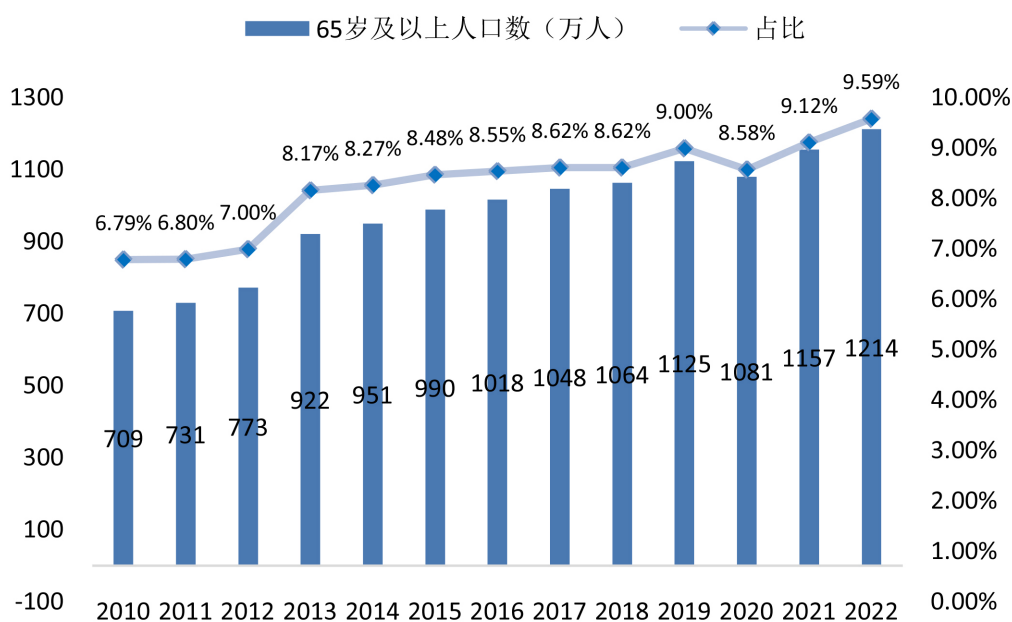


Figure 1. Changes in the population aged 65 and over in Guangdong Province, 2010~2022

图 1. 2010~2022 年广东省 65 岁及以上人口变化情况

3.2. 不同城市人口老龄化情况存在差异

从图2数据看,由于外省流入人口的年龄结构相对年轻,广东常住人口老龄化进程较全国缓慢,2020年60岁及以上人口比重达12.35%,低于全国平均水平18.70%,但不同城市的人口老龄化情况存在明显差异。有14个直辖城市60岁及以上人口比重高于全省平均水平,其中云浮、潮州、梅州、韶关和江门5个城市60岁及以上人口比重高于全国平均水平,老龄化程度较高。而东莞市和深圳市60岁人上人口占比低于6%,老龄化程度较低。作为全国第一的人口大省,广东省外省流入人口的年龄结构相对年轻,整体老龄化程度与全国相比较缓慢,但同时也应该注意到,有66.7%的城市老龄化程度高于全省平均水平,说明未来广东省将不可避免地大规模加深老龄化程度。

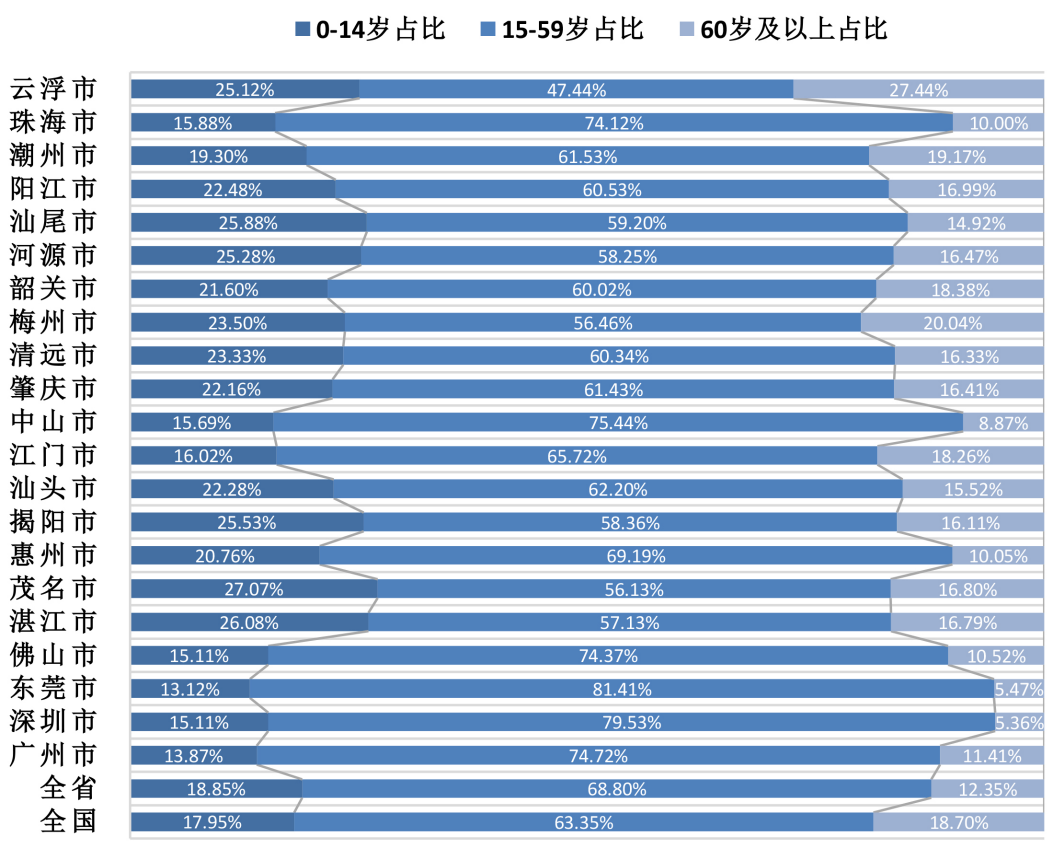


Figure 2. Population structure of Guangdong Province in 2020 (Seventh Population Census)
图2. 广东省2020年人口结构(第七次人口普查)

3.3. 老年人口抚养比持续攀升

根据2010~2022年广东省人口结构数据(表2)可知,广东省15~64岁人口在2019年达到峰值(9332万人),增长趋势变缓。短期内受人口惯性影响,劳动年龄人口绝对规模虽然仍在增加,人口红利仍在,但下降趋势难以改变。老年人口规模却在持续上升,2022年达到1214万人,占比9.59%,老年抚养比贡献逐渐增大,2022年达到1334%。在总人口增速放缓、老年人口快速增加、人口规模较大的叠加影响下,广东将不可避免地以大规模、高速度进入深度老龄化和超老龄化社会。预计未来,广东劳动年龄人口规模持续下降,老年人口规模加速扩张,出生人口规模趋向稳定;少儿抚养比渐趋稳定,而老年抚养比和总抚养比将高速攀升,劳动年龄人口整体老化,人口红利进入持续下行通道。

Table 2. Population structure of Guangdong Province, 2010~2022
表 2. 2010~2022 年广东省人口结构数据

年份	65 岁及以上 (万人)	占比	15~64 岁 (万人)	占比	老年人口抚养比
2010	709	6.79%	7963	76.33%	8.90%
2011	731	6.80%	8207	76.30%	8.91%
2012	773	7.00%	8502	77.00%	9.09%
...	...	...	...	...	...
2020	1081	8.58%	9145	72.57%	11.82%
2021	1157	9.12%	9151	72.15%	12.64%

4. 组合预测模型应用研究

4.1. ARIMA 模型人口老龄化预测

利用广东省 2000~2022 年 65 岁及以上数据绘制人口时序图(见图 3)可知，原始数据具有明显的增长趋势，为非平稳时间序列。为此，对原始序列进行一阶差分处理，绘制一阶差分时序图。由一阶差分时序图(图 3)可知，该序列基本围绕在 0 值附近上下波动，对其进行 ADF 检验，检验后结果如表 3 所示，P 值小于 0.5，说明该序列是零均值平稳序列。进一步进行白噪声检验可得该序列是平稳非白噪声序列，可进一步通过 ARMA 模型进行拟合分析。

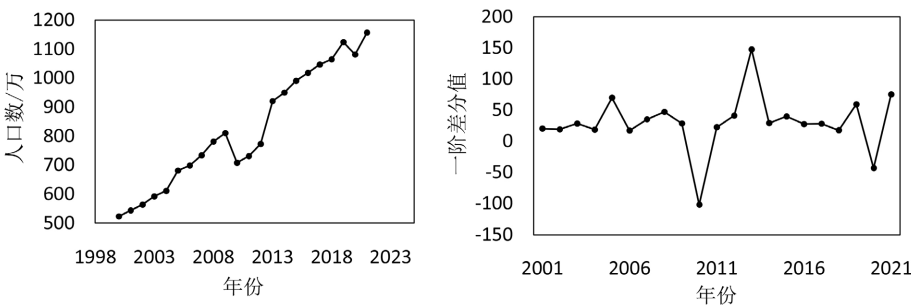


Figure 3. Raw data (left) and first order difference (right) timing diagrams
图 3. 原始数据(左)和一阶差分(右)时序图

Table 3. ADF inspection structure
表 3. ADF 检验结果

类型	滞后	Rho	Pr < Rho
Zero Mean	0	-15.4262	0.0020
	1	-8.4780	0.0310
	2	-11.9712	0.0077
Single Mean	0	-22.1542	0.0005
	1	-22.5332	0.0003
	2	1293.665	0.9999
Trend	0	-22.1346	0.0050
	1	-22.7371	0.0031
	2	702.4104	0.9999

对观察值序列进行平稳化后，根据序列的自相关图和偏自相关图初步确定 q 在 $(0, 2)$ 范围， p 在 $(0, 1)$ 范围。利用赤池信息准则(AIC)和贝叶斯信息准则(BIC)确定最后的阶数。由最小信息准则可知(见表 4)最终选取 $p = 0, q = 1$ 的 ARIMA $(0, 1, 1)$ 模型对该序列进行拟合。利用条件最小二乘法对模型进行参数估计。由检验结果可知(见表 5)，该模型各参数 p 值均小于 0.05 则参数显著，最终确立 ARIMA $(0, 1, 1)$ 模型为：

Table 4. Minimum information standard (MIS)

表 4. 最小信息准则

模型	AIC	BIC
ARIMA (1, 1, 0)	223.1781	225.2671
ARIMA (0, 1, 1)	223.134	225.223
ARIMA (1, 1, 1)	223.3796	226.5132
ARIMA (0, 1, 2)	223.9027	226.0362
ARIMA (1, 1, 2)	224.8555	229.0336

Table 5. Parameter estimation

表 5. 参数估计

条件最小二乘法参数估计					
参数	估计	标准误差	T 值	近似 $\text{Pr} > t$	滞后
MU	29.98539	8.90756	3.37	0.0032	0
MA1, 1	0.13688	0.22942	1.60	0.0378	1

$$(1 - B)\varphi(t) = 29.98539 + (1 - 0.13688B)\varepsilon_t \tag{11}$$

最后，对于模型残差进行检验与拟合判断模型有效性，对 ARIMA $(0, 1, 1)$ 模型残差进行纯随机性检验，检测结果 P 值均大于 0.05，说明残差为随机性残差，模型信息提取充分，利用该模型预测结果见表 6。

Table 6. Individual model predicted fitted values

表 6. 单项模型预测拟合值

年份	观测值	ARIMA 拟合值	新陈代谢 GM 拟合值
2001	544.21	553.6353928	564.4606
2002	563.61	575.4855398	585.727
2003	592.23	595.2209157	607.7946
...	...	...	...
2020	1081.3	1150.171925	1139.7348
2021	1157.3	1120.712577	1182.675
2022	1214.4	1182.018373	1227.2329

4.2. 新陈代谢 GM(1, 1)模型人口老龄化预测

利用新陈代谢 GM(1, 1)模型得到对应的数据预测值, 通过对预测结果的级比偏差与相对误差检验, 将原数据序列各项级比值落入区间(0.9707, 1.0021)内, 平均相对残差为 0.04708, 平均级比偏差为 0.031212, 模型精度较高。相关预测数据如下。

4.3. 组合模型人口老龄化预测

4.3.1. 权重构建

本文利用三种不同的评价指标, 算的各单项模型的精度如图 4。由 TIC 和 RMSE 可知, 新陈代谢 GM(1, 1)具有更高的精度, 而从 MAPE 来看, ARIMA(0, 1, 1)模型精度比新陈代谢 GM(1, 1)略胜一筹。考虑到不同的单项模型具有不同的优点和拟合精度, 本文将两组模型的预测数据进行合理的权重分配, 进一步提高预测精度。

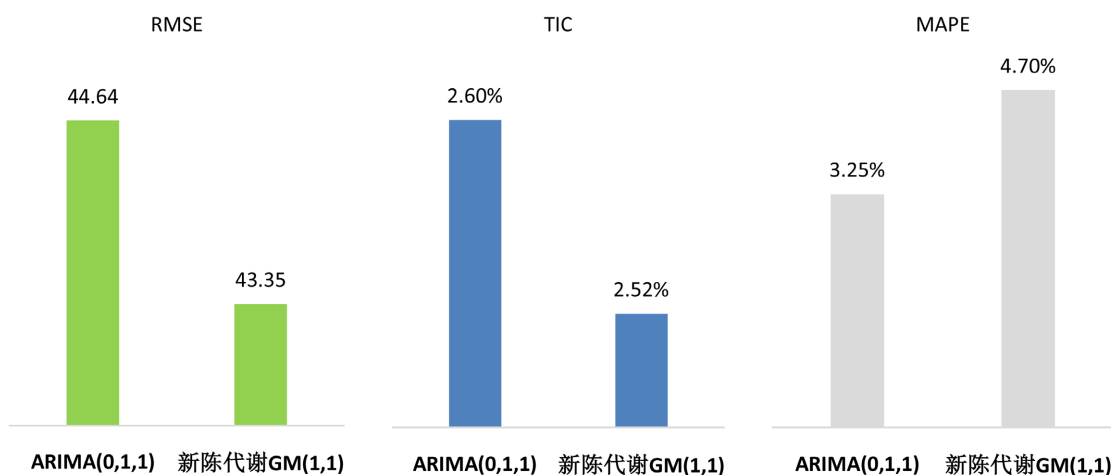


Figure 4. Comparison of individual model accuracy

图 4. 单项模型精度对比

利用最小二乘法将 ARIMA 模型的预测值作为 $\varphi_1(t)$, 新陈代谢 GM(1, 1)模型预测值作为 $\varphi_2(t)$, 根据公式(4), 利用 spss 进行最小二乘回归, 得到各模型的权重系数 $\beta_1 = 0.466$, $\beta_2 = 0.543$, 则基于最小二乘法的组合模型为:

$$\varphi^{(1)}(t) = 0.466\varphi_1(t) + 0.543\varphi_2(t) \quad (12)$$

利用平均绝对百分误差 MAPE 可知, $d_1 = 4.70\%$, $d_2 = 3.25\%$, 根据公式(7)可得 $\alpha_1 = 0.5912$, $\alpha_2 = 0.4088$, 则基于 MAPE 法构建的组合模型为:

$$\varphi^{(2)}(t) = 0.5912\varphi_1(t) + 0.4088\varphi_2(t) \quad (13)$$

根据前文构建的组合模型权重系数表可知, 结合前文算出的组合权数, 利用公式(9)可得 $\lambda_1 = 0.5538$, $\lambda_2 = 0.4462$, 则基于组合权重系数构建的组合模型为:

$$\varphi^{(3)}(t) = 0.5538\varphi_1(t) + 0.4462\varphi_2(t) \quad (14)$$

4.3.2. 不同组合模型结果对比

为提高预测值的预测精度, 减少权重计算重分配时产生的误差, 将上述两种预测方法结合为新陈代谢

谢 GM-ARIMA 组合模型对数据进行重新预测。重分配两种模型所占权重后进行数据加和，得到组合预测模型的最终预测值，结果见表 7。

Table 7. Combined model predictions
表 7. 各组合模型预测结果

年份	实际值	最小二乘法组合模型		MAPE 组合模型		组合权重的组合模型	
		拟合值	模型误差	拟合值	模型误差	拟合值	模型误差
2001	544	564.50	-20.29	558.06	-13.85	558.47	-14.26
2002	564	586.23	-22.62	579.67	-16.06	580.05	-16.44
2003	592	607.41	-15.18	600.36	-8.13	600.83	-8.60
...	...	...	...	...	...	...	...
2020	1081	1154.86	-73.56	1145.91	-64.61	1145.52	-64.22
2021	1157	1164.44	-7.44	1146.04	10.96	1148.36	8.64
2022	1214	1203.41	10.59	1200.51	13.49	1217.21	-3.21

通过对表 7 中各组合模型与实际值间产生的绝对误差进行对比分析：三个组合模型中，最小二乘法模型对应数据的变化幅度最大，预测值与实际值偏差最大；组合权重系数模型与 MAPE 预测模型的数据起伏变化相似，但组合权重系数表现更佳，特别预测误差在个别年份明显比前两个模型的预测误差小。综合来看，组合权重的组合预测模型在误差稳定性与预测偏离度方面表现最佳，与实际值间产生的误差变化幅度及综合包络面积最小(见图 5)，充分验证了利用组合模型进行数据预测后，采取权重重分配的方法可以有效提高模型的预测精度的结论。

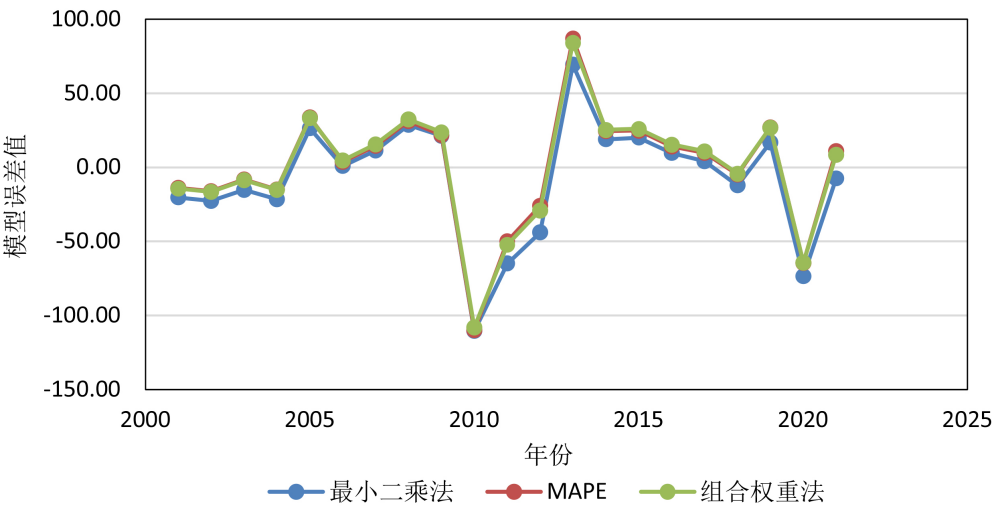


Figure 5. Comparison of errors of different combination models
图 5. 不同组合模型误差对比

利用三种评价函数得到模型评价优劣的实际数据(见表 8)：组合权重所确定的组合预测模型均优于单一权重组合模型，模型预测精度具有显著性的提高。结果表明：组合预测模型的预测精度整体优于其它两种单一预测模型，预测精度提高约 0.32%。该预测数据可为广东省人口老龄化发展预测提供参考。

Table 8. Combined model evaluation comparison
表 8. 组合模型评价对比

评价函数	最小二乘组合模型	MAPE 组合模型	组合权重组合模型
TIC	2.3210%	2.2911%	2.2890%
MAPE	3.7600%	3.6600%	3.6100%
RMSE	40.03482097	39.62584825	39.46566296

4.3.3. 组合权重组合模型预测分析

由前文可知，利用组合权重法的组合模型在广东省人口老龄化预测问题中模型精度较高。本文利用此模型，基于 2000~2022 年广东省人口结构数据预测未来 8 年的人口结构变化情况，得到预测结果如下(表 9)。

Table 9. Projections of the population structure of Guangdong Province in the next eight years
表 9. 未来 8 年广东省人口结构预测情况

单位：万人		0~14 岁		15~64 岁		65 岁及以上		老年人口抚养比	总抚养比
年份	人口	人数	比重	人数	比重	人数	比重		
2023	12716.6	2368.7	18.63%	9076.7	71.38%	1271.2	10.00%	14.01%	40.10%
2024	12755.5	2380.4	18.66%	9059.1	71.02%	1315.9	10.32%	14.53%	40.80%
2025	12838.3	2427.8	18.91%	9048.2	70.48%	1362.2	10.61%	15.06%	41.89%
2026	12929.1	2478.4	19.17%	9041.5	69.93%	1409.1	10.90%	15.59%	43.00%
2027	13012.5	2512.9	19.31%	9037.4	69.45%	1462.1	11.24%	16.18%	43.98%
2028	13103.4	2551.6	19.47%	9034.8	68.95%	1516.8	11.58%	16.79%	45.03%
2029	13229.1	2620.7	19.81%	9033.2	68.28%	1575.1	11.91%	17.44%	46.45%
2030	13361.6	2690.5	20.14%	9032.2	67.60%	1638.7	12.26%	18.14%	47.93%

根据预测结果可知，未来几年，广东省常住人口数量将趋于增长速度逐渐减小的稳定增长。考虑到年轻人婚育观念、婚育意愿和婚育行为发生转变，生育、养育、教育成本持续攀升，低生育率形势难以即刻扭转，出生人口将处于下滑期。劳动年龄人口规模始终超过 9000 万人，占比 67%~72%，位于较高水平，但也应注意到短期内受人口惯性影响，劳动年龄人口绝对规模虽然仍在增加，人口红利仍在，但下降趋势难以改变。与此同时，65 岁及以上老年人口将持续攀升，到 2030 年预计达到 1638.7 万人，占比 12.26%，人口老龄化态势持续加快，程度持续加深。总体而言，广东人口年龄结构较为年轻，但进入初始少子化社会，未来将不可避免地以大规模、高速度进入深度老龄化和超老龄化社会，老年抚养比贡献逐渐增大。预计到 2030 年，社会总抚养比达到 47.93%，意味着将近一半人口需要抚养，老龄化情况严峻。

5. 结论

本文采用组合权重系数的计算方法将新陈代谢 GM(1, 1)模型和 ARIMA 模型进行组合对广东省未来 8 年的人口结构情况进行预测，并对预测结果进行综合分析。该组合模型不仅计算简便，而且能够充分吸收单个模型的优点从而提高预测精度，可操作性强。根据预测结果可知，广东人口年龄结构较为年轻，

但进入初始少子化社会，未来将不可避免地以大规模、高速度进入深度老龄化和超老龄化社会，需引起高度重视。

致 谢

感谢老师和同学一路上的指导与帮助。

参考文献

- [1] 张栗棕, 王谨平, 刘贵松, 等. 面向金融数据的神经网络时间序列预测模型[J]. 计算机应用研究, 2018, 35(9): 2632-2637.
- [2] 曾波, 孟伟, 刘思峰, 等. 面向灾害应急物资需求的灰色异构数据预测建模方法[J]. 中国管理科学, 2015, 23(8): 84-91.
- [3] 伍建军, 杨钥姣, 王振飞. 基于组合权重的微型机器人可靠性分配方法研究[J]. 制造业自动化, 2022, 44(2): 93-97.
- [4] 卓毅鑫, 秦意茗, 胡甲秋, 等. 基于组合权重的多模式融合风机气温预测方法[J]. 南方电网技术, 2023, 17(2): 111-117.
- [5] 陈昱杉. 基于 ARIMA-GM 耦合模型的中国农村居民人均可支配收入预测研究[J]. 农村经济与科技, 2023, 34(19): 221-224.
- [6] 王宁, 张爽, 曾庆均. 基于新陈代谢 GM(1, 1)模型的重庆市人口老龄化预测研究[J]. 西北人口, 2017, 38(1): 66-67.
- [7] 王燕. 应用时间序列分析[M]. 第 6 版. 北京: 中国人民大学出版社, 2022.