

GDP 组合预测模型分析与研究

张春月* 陈 威 李 诗 罗璐璐

(鞍山师范学院数学与信息科学学院, 辽宁 鞍山 114007)

摘 要 统计预测是一种运用统计方法对事物未来发展变化进行预测的方法。单项模型预测时容易受到一些偶然因素的影响并存在系统误差。基于此, 本文借助 R 语言软件, 对某省 2020-2022 年 GDP 进行预测, 结果进一步验证了组合预测模型在考虑问题时更加系统全面, 吸收了各单项模型预测的优点, 集中了更多的统计信息与预测技巧^[1], 预测精度更高, 能够大大提高预测工作的科学性和系统性, 可以为高层决策者提供一定的决策参考。

关键词 GDP 组合预测 预测精度

中图分类号: F201; F224

文献标识码: A

文章编号: 1007-0745(2022)05-0061-03

1 问题提出

国内生产总值 (GDP) 通常用于衡量一个国家或地区的宏观经济发展状况与水平^[2], 是国民经济核算体系的核心指标, 为国家制定正确的经济发展战略和政策提供有力的理论支撑。

GDP 预测是一个复杂的系统工程, 通常会受到很多因素的影响, 同时各因素间还存在着错综复杂的影响关系, 这为 GDP 的预测增加了难度, 无法保证预测的精度^[3]。

2 基本原理

各种不同的定性或定量预测模型之间的关系并非有好坏之分, 而是着手的角度不同, 若加以结合, 那么看待问题便会更加全面, 取长补短, 从而挖掘到更加具有代表性的信息, 相应地最后的预测结果也会更加有信度和效度。换句话说, 如果在预测时, 发现某个单项预测方法得到误差较大的结果, 便弃之不用, 就可能使有用的信息从手中丢失, 那当然会使预测精度受到影响。

在上面思想的推动之下, 组合模型诞生了, 它即是多个模型的结果进行组合, 以便充分利用各个模型的信息, 从而有效提高预测精度。设有 k 种方法, y_{it} 表示第 i 种方法在第 t 期的预测值, 相关资料显示, 存在以下两种计算第 t 期组合预测值的方法:

$$(1) Y_t = \sum_{i=1}^k y_{it} \omega_i, i=1, 2, \dots, k.$$

$$(2) Y_t = \frac{1}{\sum_{i=1}^k \frac{\omega_i}{y_{it}}}, i=1, 2, \dots, k.$$

对于其中 ω_i 的确定方法, 常见的权重确定方法有六种, 包括: 等权重法、优势矩阵法、方差倒数法、残差倒数法、简单加权法、标准差法等^[4]。

1. 等权重法: 即对每一种预测方法获得的权重是相同的, 这种方法虽然简单易操作, 可是它的实用性却大打折扣, 缺点在于它是对全部模型不分主次, 平均对待, 预测效果不理想。

2. 优势矩阵法: 假设两个模型对应的权重分别为 ω_1 和 ω_2 , 次数 n_1 和 n_2 表示这两个模型比真实值有更好的预测效果。那么, 这两个模型分别对应的权重不难算出。

$$\text{公式为: } \omega_1 = \frac{n_1}{n_1 + n_2}; \omega_2 = \frac{n_2}{n_1 + n_2}.$$

3. 方差倒数法: 通过误差平方和的大小来确定权重, 即将高权重赋予误差平方和小的模型, 另一种说法为预测误差平方和倒数法。预测精度经常用预测误差平方和进行度量, 受到多种因素影响, 通常来说单项预测模型的预测精度不尽相同。预测误差平方和越大导致该项预测模型的预测精度越低, 毫无疑问, 它在组合预测中的重要性就随之降低。这时, 为了保证预测效果, 就要在组合预测模型的组合预测中赋予较大的加权系数。

公式为: $\omega_i = \frac{Q_i^{-1}}{\sum_{i=1}^m Q_i^{-1}}$, 其中 Q_i 即预测值与真实值之间差的平方和。

4. 残差倒数法: 与上述方法不同, 当某单项预测模型的误差平方和越大的时候, 相应地, 它在组合预

★基金项目: 鞍山师范学院校级科研项目 (项目编号: 20kyxm71)。

*本文通讯作者, E-mail: 2008zhangchunyu@163.com。

表 1 几种权重确定法

	等权重法	优势矩阵法	方差倒数法	残差倒数法	简单加权法	标准差法
ω_1	0.25	5/39=0.14	0.046	0.16	2/20=0.10	0.21
ω_2	0.25	15/39=0.38	0.499	0.39	8/20=0.40	0.29
ω_3	0.25	15/39=0.38	0.401	0.33	6/20=0.30	0.29
ω_4	0.25	4/39=0.10	0.054	0.12	4/20=0.20	0.22

表 2 几种模型误差比较

模型误差评价指标		MAD	MAPE	SDE
单项预测	指数模型	857.86	12.45%	2128.85
	指数平滑模型	364.18	6.00%	664.37
	线性移动平均模型	471.96	7.05%	774.47
	灰色预测 GM (1,1) 模型	918.00	22.77%	1959.27
组合预测	最优组合模型	344.04	4.26%	598.60

测中的加权系数就越小。

公式为: $\omega_i = \frac{S_i^{-1}}{\sum_{i=1}^m S_i^{-1}}$, 其中, s_i 是对平方和 Q_i 进

行开方。特别注意的是, 当数据的差均为正数时等同于方差倒数法; 两种方法只有当数据差有正负之分时才会区别对待。

5. 简单加权法: 它是一种非等权平均方法之一。降序排序的对象是各单项预测模型预测误差的平方和, 再根据各单项预测模型预测误差的平方和与权重系数成反比的基本原理来安排权重, 在组合预测中的加权系数越小意味着单项预测模型排序越靠前。

公式为: $\omega_i = \frac{i}{\sum_{i=1}^k i} = \frac{2i}{k(k+1)}$, $i=1,2,\dots,k$, 式中

按误差平方和 Q_i 降序后排列 k 个模型, 将高权重赋予误差平方和小的模型。

6. 标准差法。

公式为: $\omega_i = \frac{1}{k-1} (1 - \frac{S_i}{\sum_{i=1}^k S_i})$, $i=1,2,\dots,k$, 其中

S_i 表示第 i 个模型的标准差。

3 实证分析

用以上几种权重确定方法和模型组合方法, 对指数曲线预测模型、二次曲线指数平滑法、线性二次移

动平均模型、灰色 GM (1, 1) 四个单项预测模型进行两种方法的组合预测, 比较 12 种组合模型, 得出相对最优组合模型。

3.1 数据来源

数据来自国家统计局 2020 年统计年鉴^[5], 以 1978—2016 年为训练集, 2017—2019 年为测试集, 旨在选择一种合适的模型预测 2020—2022 年 GDP 指标。

3.2 模型判断

3.2.1 图形识别法

借助 R 检验 GDP 序列平稳性与纯随机性, 代码如下:

Box-Pierce test

data: GDP

X-squared = 116.52, df = 6, p-value < 2.2e-16

> adf.test (GDP)

Augmented Dickey-Fuller Test

data: GDP

Dickey-Fuller = -2.0431, Lag order = 3, p-value = 0.5565

alternative hypothesis: stationary

由 GDP 时序图与相关图可得: GDP 序列的时序图有明显的递增趋势, 自相关系数递减到零的速度相当缓慢, 因此该序列是非平稳的。纯随机检验的结果显示,

表 3 测试集预测精度比较

年份	指数模型		指数平滑模型		线性移动平均		灰色预测模型		最优组合预测模型	
	预测值	精度	预测值	精度	预测值	精度	预测值	精度	预测值	精度
2017	34910.39	39.07%	20107.77	92.69%	21334.33	98.35%	34093.56	42.84%	21498.99	99.11%
2018	39678.07	31.23%	23241.585	98.86%	21793.58	92.70%	39799.80	30.71%	23592.60	99.65%
2019	45096.87	18.56%	26389.44	93.83%	23764.61	95.61%	46567.83	12.64%	26341.76	94.02%

在各阶延迟下 LB 检验统计量的 P 值都非常小 (<0.05)，因此可判定该序列为非白噪声序列。

3.2.2 差分法

对 GDP 序列做二阶差分，对差分后的序列 (GDP_diff) 再次进行平稳性和纯随机性检验，代码如下：

Box-Pierce test

data: GDP_diff

X-squared = 6.8371, df = 6, p-value = 0.3362

Box-Pierce test

data: GDP_diff

X-squared = 8.3332, df = 12, p-value = 0.7586

由差分后 GDP 时序图可得出，二阶差分后，GDP 序列虽趋于平稳，但却为白噪声序列，即无法再提取有效信息，属随机游走序列，因此无法用 ARIMA 模型对其进行建模。观察 GDP 序列时序图，可观察出该序列不具有明显的季节成分，具有非线性趋势。

3.3 确定

采用指数曲线预测模型、二次曲线指数平滑法、线性二次移动平均模型、灰色 GM (1, 1) 四种单一预测方法和组合预测模型，得到结果如下：

通过计算各种组合预测模型的误差指标值 SDE (预测误差平方)，MDE (绝对误差)，MAPE (相对误差)，比较得出最优组合预测模型，即权数为： $\omega_1=0.046$ ， $\omega_2=0.499$ ， $\omega_3=0.401$ ， $\omega_4=0.054$ ，运用 $Y_i = \frac{1}{\sum_{i=1}^k \frac{\omega_i}{y_i}}$ ， $i=1, 2, \dots, k$ 的组合预测模型。

3.4 精度比较

通过以上模型各误差评价指标以及序列测试集精度的比较可知：组合预测模型的 MAD、MAPE、SDE 均相对最小，且在 2017-2019 年测试集上预测精度均在 90% 以上，效果良好。因此可得出：单项模型的预测效果不及组合模型的预测效果，应采用组合预测模

型进行预测。

3.5 预测

通过以上各模型比较，采用组合预测模型对 2020-2022 年 GDP 数据进行预测，结果为：2020 年为 28198.70 亿元，2021 年为 30514.33 亿元，2022 年为 32976.53 亿元。

4 总结

当下，国内经济形势发展已步入新常态阶段，GDP 增速整体是呈下行的发展趋势。那么如何精确分析 GDP 增长呈现何种形式，如何准确判断十四五期间乃至未来十几年，经济走势是下行或是步入新的增长周期显得尤为重要。因此，本文提出 GDP 新组合预测模型，旨在提升 GDP 的预测精度，为准确判断经济走势，为相关部门制定正确的经济发展战略和政策都有重要的现实意义。

参考文献：

- [1] 徐国祥. 统计预测和决策 (第五版) [M]. 上海: 上海财经大学出版社, 2018.
- [2] 刘宁. 几种浙江省 GDP 的预测方法及其比较 [D]. 济南: 山东师范大学硕士学位论文, 2019.
- [3] 陈齐海. GDP 组合预测模型的构建及应用研究 [D]. 南昌: 南昌大学硕士学位论文, 2018.
- [4] 雷振兴. 煤矿安全事故的分类和预测研究 [D]. 西安: 西安理工大学硕士学位论文, 2010.
- [5] 国家统计局 2020 年统计年鉴 [EB/OL]. <https://data.stats.gov.cn/easyquery.htm?cn=E0103> 2021.