

# 基于RF-GM(1,1)-BP模型预测福州市财政收入

刘 威, 张 巧, 王文博, 许 可, 张慧妍, 金秀玲\*

闽江学院数学与数据科学学院(软件学院), 福建 福州

收稿日期: 2023年6月18日; 录用日期: 2023年7月13日; 发布日期: 2023年7月20日

## 摘 要

掌握市场趋势和规划收支费用对于财政部门而言具有极为重要的意义。本文选取1994~2021年福州市年财政收入相关数据, 采用随机森林(RF)模型识别出财政收入的关键影响特征, 随后建立GM(1,1)-BP组合模型, 对2022~2025年福州市的年财政收入进行预测。预测结果表明RF-GM(1,1)-BP组合模型十分适合用于预测福州市财政收入; 同时, 福州市财政收入将在2021年之后稳步增长, 并在2025年到达7,999,256万元。该结论能为相关部门实施的决策提供一定的理论参考。

## 关键词

财政收入, 随机森林, GM(1,1), BP, 组合模型

# Predicting Fuzhou City's Fiscal Revenue Based on RF-GM(1,1)-BP Model

Wei Liu, Qiao Zhang, Wenbo Wang, Ke Xu, Huiyan Zhang, Xiuling Jin\*

School of Mathematics and Data Science (School of Software), Minjiang University, Fuzhou Fujian

Received: Jun. 18<sup>th</sup>, 2023; accepted: Jul. 13<sup>th</sup>, 2023; published: Jul. 20<sup>th</sup>, 2023

## Abstract

Mastering market trends and planning revenue and expenditure expenses is of great significance for the financial department. This paper selects the relevant data of Fuzhou's annual financial revenue from 1994 to 2021, uses the random forest (RF) model to identify the key impact characteristics of financial revenue, and then establishes a GM(1,1)-BP combination model to predict the annual financial revenue of Fuzhou from 2022 to 2025. The prediction results indicate that the RF-GM(1,1)-BP combination model is very suitable for predicting the fiscal revenue of Fuzhou City;

\*通讯作者。

Meanwhile, the fiscal revenue of Fuzhou City will steadily increase after 2021 and reach 79992.56 million yuan by 2025. This conclusion can provide a certain theoretical reference for decision-making implemented by relevant departments.

## Keywords

Fiscal Revenue, Random Forest, GM(1,1), BP, Combining Model

Copyright © 2023 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

地方的财政收入作为国家财政收入的重要组成部分，它能够综合反映出一个地方的民众的经济生活水平。当政府对各项财政收支进行宏观调控时，地方政府的财政收入是其主要参考指标。因此，对于财政部门而言，对财政收入进行合理的预测能够充分有效地了解到地方财政、经济的发展状况，同时对掌握市场趋势和规划收支费用具有极为重要的意义。

国外对财政收入预测模型的研究，主要集中在对单一模型的选择和改进方面。Sexton (1987) [1]使用经济计量模型、ARIMA 模型、趋势模型三种方法对地方政府财政收入进行预测，认为前两种方法在预测效果上优于第三种。Mario 等(2019) [2]运用四个主要的欧元区经济体编制财政政策变量的季度数据集，建立四个主要欧元区国家的向量自回归模型，以此来预测各个国家的财政收入。

对于国内的相关研究，李伟(2011) [3]运用神经网络与多元线性回归组合模型对国家财政收入进行了预测，得到了较为满意的结果。方博等(2015) [4]采用 ARMA-BP 神经网络组合模型对国家财政收入进行预测，得到组合模型相较于单一模型有着更好预测效果的结论。赵海华(2016) [5]则基于回归分析思想和相关算法，建立了一个多因素财政收入预测模型，该模型使用灰色系统理论结合 RBF 神经网络来实现数据预测。我国对单一省和地方性的财政收入预测研究相对较少。因此，本文针对福州市的财政收入进行预测研究。

综上所述，国内外学者较为常用的财政收入预测方法有神经网络、灰色理论和时间序列分析等。同时他们采用的单一预测模型得到的结果具有精度不高、泛化能力较差等局限性，组合模型未考虑到特征变量对财政收入的影响。由于随机森林(RF)在选择变量时具有较好的稳定性；GM(1,1)模型能够精准模拟指数序列且预测结果比较稳定；BP 神经网络模型有着较好的容错性、很好的逼近特性和不俗的泛化能力；因此最终采用 RF-GM(1,1)-BP 组合模型对福州市财政收入进行科学的定量分析以及准确预测。

## 2. 算法模型

### (一) 算法原理

随机森林(RF)算法是 Breiman 于 2001 年提出，是一种基于多棵树进行分类和预测的组合算法[6]。该算法通过自助抽样方法产生多个训练集，并使用这些训练集来构建多个分类回归树(CART)，作为基本的组件单元。在此过程中，Svetnik 等[7]学者利用 5 折交叉验证计算训练集中变量的重要性得分，并对变量进行排序，重复 20 次后选取排名靠前的一半变量来构建 RF 模型，记录相应测试集上的错误率，并以最小平均误差来确定最佳变量的数目。随机森林模型选择变量的指标分为置换重要性和节点纯度增加的重

要性, 置换重要性[8]就是随机打乱变量 ( $v=1,2,\dots,p$ ) 在  $OOB$  中的值, 如果变量重要, 则打乱后增加, 说明变量越重要。节点纯度增加的重要性则根据随机森林树模型中节点分裂前后的不纯度下降作为子节点纯度的增加, 它们的公式如下:

$$\begin{aligned} VIM_v^{mse} &= \frac{1}{ntree} \sum_{i=1}^{ntree} \{MSE(OOB'_i) - MSE(OOB_i)\} \\ VIM_v^{purity} &= \frac{1}{|T|} \sum_{X_v \in T} \{i_v(t) - i_v(t_L) - i_v(t_R)\} \end{aligned} \quad (1)$$

其中  $OOB'_i$  表示将第  $i$  棵树的第  $v$  个的变量  $OOB$  数据代入第  $i$  棵树,  $MSE$  表示平均平方误,  $OOB'_i$  的上标表示将  $OOB$  数据打乱,  $i_v(t)$  代表父节点的不纯度函数,  $t_L$  和  $t_R$  分别代表父节点的左右子节点,  $|T|$  表示  $X_v$  分裂变量的那些树的个数。建立随机森林回归模型选择变量, 使用  $IncNodePurity$  作为重要性的判断指标, 含义与节点纯度增加的重要性一致,  $IncNodePurity$  值越大, 说明该变量越重要。

灰色系统理论[9]是我国学者邓聚龙教授于 1982 年提出的, 适用于数据量有限、信息不完整时对不确定性问题进行研究。而灰色预测 GM(1,1)模型是灰色系统理论的核心内容, 其建模思想主要是通过已知的数据对系统信息的抽象和量化, 推导出适合的模型, 对未知数据进行预测。在该模型中, 仅需少量样本数据即可建立数据模型, 实现高精度的预测效果。灰色预测 GM(1,1)模型是通过对原始数据进行一次累加后建立一阶线性微分方程, 将得到的方程进行求解就可以得出预测模型:

$$\hat{X}^{(1)}(k+1) = \left[ \hat{X}^{(1)}(0) - \frac{\hat{u}}{\hat{a}} \right] e^{-\hat{a}k} + \frac{\hat{u}}{\hat{a}} \quad (2)$$

由于在建立微分方程时, 采用的是一次累加后的数据, 所以得到预测模型后, 还需要进行累减还原得到原始序列的灰色预测 GM(1,1)模型:

$$\hat{X}^{(0)}(k+1) = (e^{-\hat{a}} - 1) \left[ \hat{X}^{(0)}(n) - \frac{\hat{u}}{\hat{a}} \right] e^{-\hat{a}k} \quad (3)$$

BP 神经网络是一种具有连续传递函数的前馈神经网络。BP 神经网络的训练过程中使用误差反向传播算法来实现, 最为常见的为梯度下降法[10]。通过最小化均方误差, 不断地调整网络的权重和阈值以逐渐提高拟合数据的精度。这一过程可以帮助 BP 神经网络学习并优化输入数据之间的关系, 从而更好地预测未知数据的结果。BP 神经网络模型主要通过各层的权重、激活函数以及偏置来得到最终预测值, 它将输入层的输出数据乘上一定的权重放入隐藏层中, 作为隐藏层的起始值; 在隐藏层中利用模型在该层中的偏置项和激活函数作用, 得到新的值, 然后将得到的新值乘上一定的权重放入输出层中, 作为输出层的起始值; 在输出层中同样利用模型在该层中的偏置项和激活函数作用, 得到第一次迭代值。

隐藏层公式:

$$\begin{aligned} q_j &= \sum_{i=1}^5 x_i * w_{ij} - a_j \\ h_j &= f(q_j) \end{aligned} \quad (4)$$

输出层公式:

$$\begin{aligned} u &= \sum_{j=1}^7 \sum_{i=1}^5 h_i * v_j - \beta \\ \hat{y} &= f(u) \end{aligned} \quad (5)$$

其中  $x_i$  为输入数据,  $w_{ij}$  为输入层权重,  $a_j$  为隐藏层偏置,  $f()$  为激活函数,  $v_j$  为隐藏层权重, 为输出

层偏置。接着，通过最小化损失函数，不断更新权重和偏置，得到最优的预测值，损失函数公式如下：

$$L(e) = \frac{1}{n} \sum_{k=1}^p e_k^2 = \frac{1}{n} \sum_{k=1}^p (\hat{y}_k - y_k)^2 \quad (6)$$

其中  $\hat{y}_k$  为经过神经网络后的输出值， $y_k$  为期望输出结果即真实值。

## (二) 评价标准

均方根误差(RMSE)是指预测值和真实值之间差距的平方和的平均值的开根号，可以用于评估 BP 模型预测结果的精度，能够很好地反映预测值与实际值之间的误差且受异常值的影响程度较小，因此使用 RMSE 来验证 BP 模型准确性，所得到的结果更加直观。它的计算公式如下：

$$X_{RMSE} = \sqrt{\frac{\sum_{i=1}^N (X_{obs,i} - X_{model,i})^2}{N}} \quad (7)$$

R 方(R-squared)是指模型能够解释因变量的变异程度，可以消除观察值尺度的影响，对 BP 模型性能的好坏有着准确的评价，R 方越大说明模型拟合的越好，预测结果越精确。它的计算公式如下：

$$R^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y})^2}{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2} \quad (8)$$

## (三) 研究框架

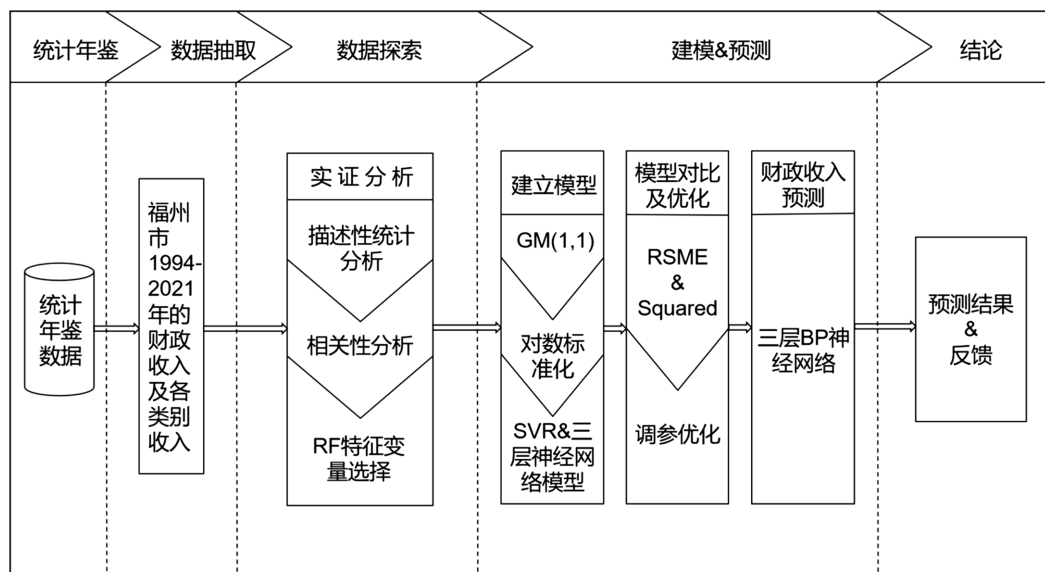


Figure 1. Flow chart of Fuzhou fiscal revenue forecast research  
图 1. 福州市财政收入预测研究流程图

福州市财政收入预测研究流程如图 1 所示，具体为：一、从《福州市统计年鉴》收集整理得到 1994~2021 年福州市财政收入的相关年数据；二、对数据进行描述性统计分析、相关性分析，再根据随机森林(RF)重要性选择变量；三、建立所选取变量的 GM(1,1)模型，再对数据进行对数标准化，最后分别建立 SVR 和 BP 神经网络模型。四、根据均方根误差和 R 方两种评价标准的结果，得到最优模型；五、利用 GM(1,1)-BP 组合模型，得到福州市在 2022~2025 年财政收入的预测值。

### 3. 实证分析

#### (一) 数据介绍

预测福州市未来 4 年财政收入的相关数据均来自于《福州市统计年鉴》(1994~2022 年), 数据真实可靠。财政收入( $y$ )与许多因素相关, 选取了 12 个对财政收入影响较大的变量来预测未来 4 年的财政收入。由于 1994 年我国的财政体制进行了重大的改革, 实施了分税制的财政政策。因此, 采用 1994 年之后的数据进行分析。

社会总从业人数( $x_1$ ): 单位为万人, 社会总从业人数对市财政收入有显著的影响, 从业的人数越多, 税收越多, 居民消费水平越高, 间接影响市财政收入。

在岗职工工资总额( $x_2$ ): 单位为万元, 在岗职工工资总额越高, 政府所收的税收越多, 进而影响市财政收入。

社会消费品销售总额( $x_3$ ): 单位为亿元, 社会消费品零售总额创造税收, 而财政收入主要来自税收, 所以社会消费品零售总额影响着市财政收入。

城镇居民人均可支配收入( $x_4$ ): 单位为元, 城镇居民人均可支配收入提高, 意味着居民的工作积极性提高, 促进市财政收入快速增长。

城镇居民消费性支出( $x_5$ ): 单位为元, 城镇居民消费性支出增加, 意味着居民消费水平提升, 政府可以从居民消费中获得更多税收, 市财政收入也得到增加。

年末总人口( $x_6$ ): 单位为万人, 在经济发展水平不变的情况下, 一个地区的人口总数越多, 则人均市财政收入越少。

地区生产总值( $x_7$ ): 单位为亿元, 地区生产总值代表一个地区的经济发展水平, 一个地区经济发展水平好, 就会带动这个地区的财政收入, 所以地区生产总值对市财政收入有着重大的影响。

第一产业产值( $x_8$ ): 单位为亿元, 第一产业主要是农业方面, 农业发展对市财政收入有着一定的影响。

居民消费价格指数( $x_9$ ): 将 1990 年的值视为 100, 得到的数据与之比较, 单位为%, 居民消费价格指数也叫做 CPI, 它反映了居民家庭所购买的消费品以及服务项目的价格变化的指标, 影响着市财政收入。

第三产业与第二产业产值比( $x_{10}$ ): 单位为亿元, 第三产业与第二产业产值比代表了产业结构的变动, 产业结构得到优化, 则市财政收入也会得到提升。

工业增加值( $x_{11}$ ): 单位为亿元, 工业增加值反映了企业生产过程中新创造的价值, 工业值变化, 可以判断短期地区经济情况, 与市财政收入息息相关。

建筑业增加值( $x_{12}$ ): 单位为万元, 建筑业增加值是 GDP 的重要组成部分, 反映了建筑业企业生产经营活动的最终成果, 与市财政收入息息相关。

#### (二) 福州市财政收入相关变量描述性统计分析

对福州市财政收入相关变量进行描述性统计分析, 得到福州市财政收入的最小值为 209931.00 万元, 最大值为 7498470.00 万元, 两者相差较大, 说明福州市财政收入的变化比较明显, 尤其是 2010 年之后, 福州市财政收入大幅度增加; 而通过福州市财政收入的标准差和均值, 可以知道福州市各年财政收入数据较为离散。因此, 通过建立模型来研究福州市财政未来的收入是有必要的。具体数据见表 1。

**Table 1.** General description of relevant variables of Fuzhou's financial revenue

**表 1.** 福州市财政收入相关变量概括性描述

|            | 最小值       | 最大值         | 均值         | 标准差        |
|------------|-----------|-------------|------------|------------|
| $x_1$ (万人) | 274.03    | 420.00      | 350.10     | 50.49      |
| $x_2$ (万元) | 395044.00 | 14211914.00 | 4623190.00 | 4649242.00 |
| $x_3$ (亿元) | 101.99    | 4549.41     | 1567.16    | 1470.45    |

Continued

|               |           |            |            |            |
|---------------|-----------|------------|------------|------------|
| $x_4$ (元)     | 4108.00   | 53421.00   | 22029.55   | 15555.56   |
| $x_5$ (元)     | 3338.00   | 35664.00   | 15353.28   | 10226.96   |
| $x_6$ (万人)    | 555.47    | 723.36     | 634.00     | 50.30      |
| $x_7$ (亿元)    | 349.08    | 11324.48   | 357.00     | 3290.00    |
| $x_8$ (亿元)    | 72.87     | 637.03     | 266.00     | 161.00     |
| $x_9$ (%)     | 181.7     | 376.9      | 289.00     | 53.20      |
| $x_{10}$ (亿元) | 0.86      | 1.49       | 1.11       | 15.70      |
| $x_{11}$ (亿元) | 103.37    | 2758.62    | 1100.00    | 874.00     |
| $x_{12}$ (万元) | 28.07     | 1548.49    | 399.00     | 452.00     |
| $y$ (万元)      | 209931.00 | 7498470.00 | 2808143.00 | 2537309.00 |

### (三) 福州市财政收入各变量相关性分析

通过相关性分析得到各自变量与因变量间的相关关系, 观察各自变量与财政收入的相关程度, 对筛选变量做出初步判断。福州市财政收入各变量相关系数见表 2, 从中可以得到在岗职工工资总额( $x_2$ )、社会消费品销售总额( $x_3$ )、城镇居民人均可支配收入( $x_4$ )、城镇居民消费性支出( $x_5$ )、第一产业产值( $x_8$ )、工业增加值( $x_{11}$ )变量与财政收入( $y$ )间均具有较高的相关关系且相关性最高的达到了 0.99, 第三产业与第二产业产值比( $x_{10}$ )变量与财政收入( $y$ )间具有最低的相关关系达到了 0.70。另外, 其他自变量间的相关程度也比较高, 说明各变量间存在多重共线性。

**Table 2.** Correlation coefficient of various variables of Fuzhou's financial revenue

**表 2.** 福州市财政收入各变量相关系数表

|          | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $x_8$ | $x_9$ | $x_{10}$ | $x_{11}$ | $x_{12}$ | $y$ |
|----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|----------|----------|-----|
| $x_1$    |       |       |       |       |       |       |       |       |       |          |          |          |     |
| $x_2$    | 0.92  |       |       |       |       |       |       |       |       |          |          |          |     |
| $x_3$    | 0.93  | 1.00  |       |       |       |       |       |       |       |          |          |          |     |
| $x_4$    | 0.96  | 0.99  | 0.99  |       |       |       |       |       |       |          |          |          |     |
| $x_5$    | 0.96  | 0.98  | 0.99  | 1.00  |       |       |       |       |       |          |          |          |     |
| $x_6$    | 0.98  | 0.95  | 0.96  | 0.98  | 0.98  |       |       |       |       |          |          |          |     |
| $x_7$    | 0.91  | 0.99  | 0.99  | 0.99  | 0.98  | 0.95  |       |       |       |          |          |          |     |
| $x_8$    | 0.93  | 0.99  | 0.99  | 0.99  | 0.99  | 0.97  | 1.00  |       |       |          |          |          |     |
| $x_9$    | 0.98  | 0.94  | 0.96  | 0.98  | 0.98  | 0.99  | 0.94  | 0.97  |       |          |          |          |     |
| $x_{10}$ | 0.63  | 0.76  | 0.75  | 0.73  | 0.72  | 0.65  | 0.79  | 0.75  | 0.64  |          |          |          |     |
| $x_{11}$ | 0.96  | 0.99  | 0.99  | 1.00  | 1.00  | 0.98  | 0.98  | 0.99  | 0.97  | 0.69     |          |          |     |
| $x_{12}$ | 0.86  | 0.99  | 0.98  | 0.97  | 0.96  | 0.92  | 0.99  | 0.98  | 0.91  | 0.82     | 0.96     |          |     |
| $y$      | 0.95  | 0.99  | 0.99  | 0.99  | 0.99  | 0.97  | 0.98  | 0.99  | 0.96  | 0.70     | 0.99     | 0.96     |     |

## 4. RF-GM(1,1)-BP 模型的建立

### (一) RF 选择变量

由于各变量间具有多重共线性, 因此采用随机森林(RF)模型, 筛选出对财政收入来说最重要的几个变量, 确保在保留数据特征的前提下, 减少数据的维度, 使模型预测结果更加精确。通过 5 次重复十折



交叉验证得到：当自变量保留 4~7 个时，所建立的模型最优，通过分析 5 次重复十折交叉验证法的结果并结合所学知识，决定保留对预测财政收入最重要的 5 个变量，将自变量按照下标顺序排序输出，具体结果见表 3。

**Table 3.** Importance principle selection variables table

**表 3.** 重要性原则选择变量表

|          | $\%IncMSE$             | $IncNodePurity$        |
|----------|------------------------|------------------------|
| $x_1$    | $5.283 \times 10^{11}$ | $1.196 \times 10^{13}$ |
| $x_2$    | $6.627 \times 10^{11}$ | $1.214 \times 10^{13}$ |
| $x_3$    | $6.648 \times 10^{11}$ | $1.672 \times 10^{13}$ |
| $x_4$    | $6.028 \times 10^{11}$ | $1.565 \times 10^{13}$ |
| $x_5$    | $6.113 \times 10^{11}$ | $1.325 \times 10^{13}$ |
| $x_6$    | $9.115 \times 10^{11}$ | $1.833 \times 10^{13}$ |
| $x_7$    | $7.749 \times 10^{11}$ | $1.777 \times 10^{13}$ |
| $x_8$    | $6.535 \times 10^{11}$ | $1.315 \times 10^{13}$ |
| $x_9$    | $4.643 \times 10^{11}$ | $1.086 \times 10^{13}$ |
| $x_{10}$ | $7.632 \times 10^{11}$ | $1.529 \times 10^{13}$ |
| $x_{11}$ | $8.666 \times 10^{11}$ | $1.884 \times 10^{13}$ |
| $x_{12}$ | $8.201 \times 10^{11}$ | $1.627 \times 10^{13}$ |

表 3 中， $\%IncMSE$  就是 *increase in MSE*，指使用随机值代替各个变量，变量替换后模型的误差，而  $IncNodePurity$  就是 *increase in node purity*，指每个变量对回归树的每个节点观测值的影响，两者值越大说明该变量越重要，都可以作为重要性的判断指标，采用  $IncNodePurity$  来选择变量。在经济意义方面，社会总从业人数( $x_1$ )、在岗职工工资总额( $x_2$ )、社会消费品销售总额( $x_3$ )、城镇居民人均可支配收入( $x_4$ )、城镇居民消费性支出( $x_5$ )这 5 个变量对财政收入的影响均反映在税收上，因此，只需保留一个变量；由于 2006 年国家取消了农业税，导致第一产业产值( $x_8$ )对财政收入影响较小，因此剔除该变量；居民消费价格指数( $x_9$ )影响了居民的消费额，对财政收入的影响也反映在税收上，所以剔除该变量；第三产业与第二产业产值比( $x_{10}$ )代表了产业结构的变动，对财政收入有一定影响，但相较于其余变量权重较小，因此剔除该变量。

综上所述，结合随机森林重要程度识别出 5 个影响福州市财政收入的变量分别为：社会消费品销售总额( $x_3$ )、年末总人口( $x_6$ )、地区生产总值( $x_7$ )、工业增加值( $x_{11}$ )、建筑业增加值( $x_{12}$ )。

## (二) GM(1,1)-BP 模型的建立

未来的财政收入情况关系到政府工作规划，需要十分精确的结果。本文所得到的数据单位和大小都存在一定的差距，直接进行建模预测，得到的结果误差极大，不具备可信度，而对于经济数据来说，最适合的标准化方法为对数标准化[11]。为了得到精确的预测结果，对数据进行对数标准化，消除数据间大小和单位的影响。再建立 BP 神经网络模型。

第一步，将数据划分为训练集和测试集，利用测试集检验模型优劣，防止出现过拟合现象，提高模型的泛化能力。由于数据为时间序列，具有连续性，数据的划分不能采用随机形式，选取 1994~2012 年福州市财政收入相关数据作为训练集，2013~2021 年福州市财政收入相关数据作为测试集。

第二步，根据得到的训练集建立 BP 神经网络模型，神经网络具有多种不同的模型，采用三层神经网络，该模型共分为三层，分别为输入层、隐藏层和输出层。再将 BP 神经网络中最大迭代次数设置为 10,000，误差精度设置为 0.00001；同时，由于适合的隐藏层节点数会大大增加模型的精确度。因此，通

过调参优化的方式,将隐藏层神经元个数设置为7,建立最优的BP神经网络模型。

第三步,将建立的支持向量机回归模型与BP神经网络模型相比较,再构建对比模型SVR。

### (三) BP模型检验及优化

将划分好的测试集分别代入SVR和BP神经网络中,利用R中的函数计算两个模型的RMSE(均方根误差)和R-squared(R方),根据得到的结果判断模型的优劣,选择出最适合的模型,检验结果见表4。

**Table 4.** Model inspection and comparison table

**表 4.** 模型检验及对比表

|           | RF-SVR | RF-BP  |
|-----------|--------|--------|
| RMSE      | 0.4903 | 0.0354 |
| R-squared | 0.9022 | 0.9497 |
| MAE       | 0.4544 | 0.0312 |

从表4得到的结果来看:RF-BP模型的RMSE(均方根误差)为0.0354,R-squared(R方)为0.9497;RF-SVR模型的RMSE(均方根误差)为0.4903,R-squared(R方)为0.9022。均方根误差(RMSE)是指预测值和真实值之间差距的平方和的平均值的开根号,用来评价模型的准确性;R方(R-squared)是指模型能够解释因变量的变异的程度,用来评价模型的拟合程度,均方根误差越小、R方越大说明模型拟合的越好,预测结果越精确。因此,使用RF-BP模型对福州市财政收入进行预测,会得到精确的结果。

## 5. 财政收入预测

### (一) GM(1,1)预测自变量的值

GM(1,1)模型有着几个优点:1)具有时效性。对于变化较快的数据,可以更好地满足预测需求;2)适用性强。对于不规则时间序列的预测能力强,可以更准确地反映数据的发展规律;3)对异常值鲁棒性强。即使数据中存在异常情况,也可以保证预测的准确性;4)中短期预测准确。对于中短期的预测更加精确。而福州市财政收入各变量的数据没有明显的规律,预测未来4年的值也属于短期预测。因此,与其他模型相比,在预测未来4年福州市社会消费品零售总额、年末总人口、地区生产总值、工业增加值以及建筑业增加值的情况下,采用GM(1,1)模型具有更大的优势。

为预测未来4年福州市财政收入总值,利用GM(1,1)模型对通过随机森林法选出的5个关键变量进行预测,得到影响福州市财政收入各变量未来4年的值。

首先,将社会消费品销售总额( $x_3$ )、年末总人口( $x_6$ )、地区生产总值( $x_7$ )、工业增加值( $x_{11}$ )、建筑业增加值( $x_{12}$ )这5个特征变量分别进行一次累加生成一次累加序列,建立起5个灰色微分方程进行求解,运用后减运算还原得模型输入序列预测序列,具体结果见表5。

**Table 5.** Prediction results and testing of related factors

**表 5.** 相关因素预测结果及检验

|       | $x_3$ (亿元) | $x_6$ (万人) | $x_7$ (亿元) | $x_{11}$ (亿元) | $x_{12}$ (万元) |
|-------|------------|------------|------------|---------------|---------------|
| 2022年 | 7666.86    | 727.4      | 14654.55   | 3962.96       | 2268.40       |
| 2023年 | 8618.83    | 734.45     | 16517.53   | 4358.08       | 2646.31       |
| 2024年 | 9689.00    | 741.57     | 18617.35   | 4792.58       | 3087.17       |
| 2025年 | 10892.05   | 748.75     | 20984.12   | 5270.40       | 3601.48       |
| 后验差比值 | 0.34       | 0.03       | 0.12       | 0.22          | 0.21          |



通过 GM(1,1)模型得到 2025 年社会消费品销售总额( $x_3$ )的预测值为 10892.05 亿元, 年末总人口( $x_6$ )的预测值为 748.75 万人, 地区生产总值( $x_7$ )的预测值为 20984.12 亿元, 工业增加值( $x_{11}$ )的预测值为 5270.40 亿元, 建筑业增加值( $x_{12}$ )的预测值为 3601.48 万元。接着, 比较后验差比值(C)以及预测图来检验模型预测精度。从表 5 中可以得到社会消费品销售总额、年末总人口、地区生产总值、工业增加值、建筑业增加值的后验差比值(C)均小于 0.35, 说明模型精度等级为好, 模型预测结果可靠, 可用于预测福州市财政收入。

## (二) RF-GM(1,1)-BP 模型预测福州市未来 4 年财政收入

将自变量在福州市未来 4 年的预测值代入建立隐藏层节点为 7、最大迭代次数为 10,000、误差精度为 0.00001 的 BP 神经网络模型中, 得到福州市未来 4 年财政收入的预测值, 具体结果见表 6。

**Table 6.** Forecast results of financial revenue of Fuzhou in the next four years

**表 6.** 福州市财政收入未来 4 年预测结果

|        | 福州市财政收入预测值(万元) |
|--------|----------------|
| 2022 年 | 7,959,923      |
| 2023 年 | 8,015,241      |
| 2024 年 | 8,025,953      |
| 2025 年 | 7,999,256      |

从表 6 中, 可以得到 2022 年福州市财政收入预测值为 7,959,923 万元, 2023 年福州市财政收入预测值为 8,015,241 万元, 2024 年福州市财政收入预测值为 8,025,953 万元, 2025 年福州市财政收入预测值为 7,999,256 万元, 可以看到福州市财政收入将在 2021 年之后稳步增长, 表明福州市经济发展状态良好。

## 6. 结论与建议

### (一) 结论

1) 福州市财政收入将在 2021 年之后稳步增长, 表明福州市针对经济发展的政策十分优秀。

2) 工业增加值对福州市财政收入而言最为重要, 年末总人口数次之, 然后是地区生产总额, 社会消费品销售总额, 建筑业增加值。政府的财政收入大多来自税收, 工业增加值代表着企业创造的价值, 说明企业创造的价值在税收收入中所占的比例较大; 年末总人口数对财政收入起着积极作用, 说明福州市人民的收入和人民生活水平较高。

3) 社会消费品零售总额在 2022、2023、2024、2025 年预测值为 7666.86 亿元、8618.83 亿元、9689.00 亿元、10892.05 亿元; 年末总人口数在 2022、2023、2024、2025 年预测值为 727.40 万人、734.45 万人、741.57 万人、748.75 万人; 地区生产总额在 2022、2023、2024、2025 年预测值为 14654.55 亿元、16517.53 亿元、18617.35 亿元、20984.12 亿元; 工业增加值在 2022、2023、2024、2025 年预测值为 3962.96 亿元、4358.08 亿元、4792.58 亿元、5270.40 亿元; 建筑业增加值在 2022、2023、2024、2025 年预测值为 2268.40 万元、2646.31 万元、3087.17 万元、3601.48 万元。

4) 福州市财政收入总值在 2022、2023、2024、2025 年的预测值为 7,959,923 万元、8,015,241 万元、8,025,953 万元、7,999,256 万元, 表明福州市财政收入将在 2021 年后不断增长。在疫情结束后, 各个地区都在加快恢复经济建设, 福州市在未来的财政收入有极大的增长, 福州市政府的治理能力和管理水平较高。

### (二) 建议

1) 福州市针对经济发展的政策还有可以提升的地方, 如改善财政支出结构等。

2) 企业创造的价值在税收收入中所占的比例较大, 政府应该加强对地方产业政策的制定和实施, 引导产业结构调整 and 转型升级, 提高地方财政收入; 提高政策影响, 提高福州市人民生活水平, 保持年末总人口数对财政收入增长的积极作用。

3) 政府可以加大社会消费品零售、工业和建筑业方面的投入, 并通过政策鼓励, 增加福州市人口总数, 提高福州市地区生产总值, 来增加地区的财政收入。

4) 地区财政收入增长较多, 并不一定代表地区真正的繁荣, 相关部门应该更加注意, 避免出现资源的过度消耗以及税收的过度增加。

## 基金项目

项目来源: 福建省科技厅。

项目名称: 基于 ATOT 技术的智能养老系统设计与开发。

项目编号: 2023350104000282。

## 参考文献

- [1] Sexton, T.A. (1987) Forecasting Property Taxes: A Comparison and Evaluation of Methods. *National Tax Journal*, **40**, 47-59. <https://doi.org/10.1086/NTJ41789674>
- [2] Alloza, M., Burriel, P. and Pérez, J.J. (2019) Fiscal Policies in the Euro Area: Revisiting the Size of Spillovers. *Journal of Macroeconomics*, **61**, Article ID: 103132. <https://doi.org/10.1016/j.jmacro.2019.103132>
- [3] 李伟. 神经网络在财政数据中的应用[D]: [硕士学位论文]. 长春: 吉林大学, 2011.
- [4] 方博, 何朗. 关于 ARMA-BP 神经网络组合模型的财政收入预测[J]. *基础科学*, 2014(11): 709-713.
- [5] 赵海华. 基于灰色 RBF 神经网络的多因素财政收入预测模型[J]. *经济与管理科学*, 2016(7): 79-81.
- [6] Breiman, L. (2001) Random Forests. *Machine Learning*, **45**, 5-32. <https://doi.org/10.1023/A:1010933404324>
- [7] Svetnik, V., Liaw, A., Tong, C., et al. (2004) Application of Breiman's Random Forest to Modeling Structure-Activity Relationships of Pharmaceutical Molecules. *Multiple Classifier Systems*, **3077**, 334-343. [https://doi.org/10.1007/978-3-540-25966-4\\_33](https://doi.org/10.1007/978-3-540-25966-4_33)
- [8] 曹桃云. 基于随机森林的变量重要性研究[J]. *统计与决策*, 2022(4): 60-63.
- [9] 杨华龙, 刘金霞, 郑斌. 灰色预测 GM(1,1)模型的改进及应用[J]. *数学的实践与认识*, 2011(23): 39-46.
- [10] 汪镭, 周国兴, 吴启迪. 人工神经网络理论在控制领域中的应用综述[J]. *同济大学学报(自然科学版)*, 2001, 29(3): 357-361.
- [11] 孙红卫, 吕春燕, 祁爱琴, 等. 综合评价中数据标准化的原理研究[J]. *中国卫生统计*, 2015(2): 342-349.