

# Forecast of Total Retail Sales of Consumer Goods in Guangxi Based on Arima and Gm(1,1) Model

Yan Liu

School of Education, Nanning University, Nanning, Guangxi, 530200, China

## Abstract

The total retail sales of consumer goods is one of the macroeconomic indicators reflecting the domestic consumption situation, which has a strong correlation with time. Its historical data has a certain impact on the future development, so the analysis and prediction of the total retail sales of consumer goods is of great significance to grasp the economic growth trend. Based on the annual data of the total retail sales of consumer goods in Guangxi from 1978 to 2020, this paper constructs a time series ARIMA model and a grey prediction GM(1,1) model under the statistical software R Studio, and respectively uses the models to forecast the total retail sales in the next three years. The results show that the relative error of the prediction of the ARIMA model is small. The prediction accuracy of GM(1,1) model is high, and short-term prediction can provide reference for the government.

## Keywords

total retail sales of consumer goods; ARIMA; GM(1,1); forecast

# 基于 ARIMA 与 GM(1,1) 模型的广西社会消费品零售总额预测

刘雁

南宁学院教育学院, 中国·广西南宁 530200

## 摘要

社会消费品零售总额是反映国内消费情况的宏观经济指标之一, 随时间变化存在很强的相关性, 其历史数据对未来的发展存在一定的影响, 因此分析预测社会消费品零售总额对把握经济增长趋势具有重要意义。论文基于广西1978—2020年的社会消费品零售总额年度数据, 在R Studio统计软件下构建时间序列ARIMA模型与灰色预测GM(1,1)模型, 并分别运用模型对未来三年的零售总额进行预测, 结果表明ARIMA模型预测的相对误差较小, GM(1,1)模型的预测精度较高, 短期预测可以为政府提供参考。

## 关键词

社会消费品零售总额; ARIMA; GM(1,1); 预测

## 1 引言

中共二十大以来, 中国消费品市场保持稳中有增, 市场规模持续扩大, 消费对经济增长的贡献率稳步提升, 成为经济增长的主要驱动力。在反映消费情况的统计指标中, 社会消费品零售总额是最常用的宏观经济指标之一。目前, 已有不少学者对社会消费品零售总额的预测进行了大量研究。马强等<sup>[1]</sup>利用乘积季节模型对2001—2020年的社会消费品零售总额数据进行时间序列分析。蒋翠清等<sup>[2]</sup>引入误差校

正方法, 构建基于VMD-DE-LSSVM误差校正组合预测模型。有关社会消费品零售总额预测研究的还有樊亮<sup>[3]</sup>、尤游和刘苏兵<sup>[4]</sup>、王爽和汪海飞<sup>[5]</sup>、王志坚和王斌会<sup>[6]</sup>等。论文以广西为例, 根据统计年鉴现有数据资料, 选取1978年至2022年的社会消费品零售总额建立ARIMA与GM(1,1)模型, 进行分析预测。

## 2 模型介绍

### 2.1 ARIMA 模型简介

ARIMA模型<sup>[7]</sup>主要用于拟合具有平稳性(或可以被转换为平稳)的时间序列, 在一个平稳的时序中, 序列的统计性质并不会随着时间的推移而改变。在一个 $p$ 阶自回归模型中, 序列的每一个值都可以用它之前 $p$ 个值的线性组合来表示, 即 $AR(p): Y_t = \mu + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_p Y_{t-p} + \epsilon_t$ 。在一个 $q$ 阶移动平均模型中, 序列的每一个值都可以用它之前 $q$ 个残差的

【基金项目】南宁学院校级科研项目(项目编号: 2023XJ24)。

【作者简介】刘雁(1995-), 女, 中国广西灵山人, 硕士, 助教, 从事经济社会统计研究。

线性组合来表示,即:

$$MA(p): Y_t = \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} + \varepsilon_t$$

将AR(p)与MA(p)结合起来成为ARMA(p, q)模型,其表达式为:

$$Y_t = \mu + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_p Y_{t-p} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} + \varepsilon_t$$

此时,序列中的每个观测值都用过去的p个观测值和q个残差的线性组合表示。ARIMA(p, d, q)模型意味着序列被差分了d次,且序列中的每个观测值都是用过去的p个观测值和q个残差值的线性组合表示的。

### 2.2 GM(1, 1) 模型简介

灰色系统理论是邓聚龙教授于1982年提出,用于进行预测、决策、评估、规划控制、系统分析与建模的重要方法,其主要优点在于它可以对时间序列短、统计数据少、信息不完全系统的分析与建模,具有独特的功效,因此得到了广泛的推广<sup>[8]</sup>。其中,GM(1,1)模型是灰色系统理论中应用最广泛的一种灰色动态预测模型,其预测步骤如下:

设一单调非负原始数据序列为 $X^{(0)} = \{X^{(0)}(1), X^{(0)}(2), \dots, X^{(0)}(n)\}$ ,先对 $X^{(0)}$ 进行一次累加得到一次累加序列 $X^{(1)} = \{X^{(1)}(1), X^{(1)}(2), \dots, X^{(1)}(n)\}$ ,然后对 $X^{(1)}$ 进行一阶线性微分方程:  $\frac{dx^{(1)}}{dt} + aX^{(1)} = u$ ,即GM(1,1)模型。求解微分方程,并将GM(1,1)所得数据经过累减还原为 $\hat{X}^{(0)}(k+1)$ ,即 $X^{(0)}$ 的灰色预测模型为:

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

## 3 模型构建

### 3.1 ARIMA 模型构建

#### 3.1.1 数据预处理

以广西1978—2022年社会消费品零售总额为原始数据,数据来源于广西统计年鉴。为了研究需要,只选择前面43个数据进行建模,2021和2022年数据用于预测对比,来检验模型预测的准确性。

对原始数据做出时序图,如图1所示,数据前期有缓慢上升的趋势,2000年以后上升速度加快,呈指数形式增长,数据极不平稳。对原始数据取对数以消除指数趋势,所得时序图如图2所示,取对数后的时序图具有明显的线性趋势。

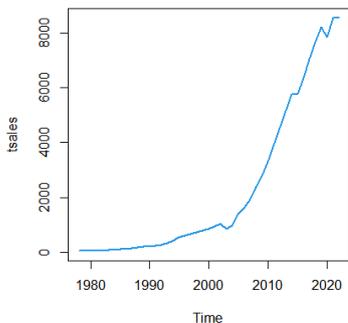


图1 社会消费品零售总额时序图

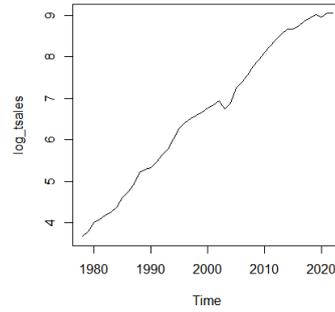


图2 取对数后社会消费品零售总额时序图

对取对数后的序列做一阶差分,记为d\_log\_tsales,以消除线性趋势,如图3所示,数据较为平稳,但是序列大部分都是在零均值上方。尝试对取对数后的序列做二阶差分,记为dd\_log\_tsales,如图4所示,数据明显消除线性趋势,并且在0处上下波动。图4序列是否平稳需要经过ADF单位根检验,检验结果显示序列此时是平稳的,可以继续下一步。

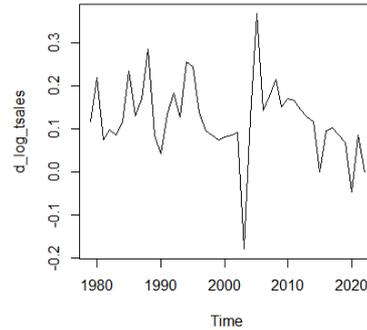


图3 序列 d\_log\_tsales 时序图

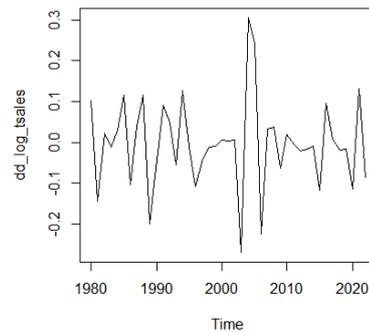


图4 序列 dd\_log\_tsales 时序图

#### 3.1.2 模型识别与参数估计

对二阶差分序列dd\_log\_tsales做自相关(ACF)图和偏自相关(PACF)图,如图5和图6所示。

自相关系数在滞后阶数k=2处显著不为0,在k>2时几乎全部落入随机区间,因此,取q=2。偏相关系数在k=2,14处显著不为0,因此p=2或14。一般建模时除了考虑模型精度外,还要尽可能使模型简化易于解释,因此在初步选择模型时,应从最低阶数考虑,为ARIMA(1,2,1),再依次考虑ARIMA(1,2,2),ARIMA(2,2,2)。由arima()函

数建立相应的模型,并通过AIC准则判断这三个模型的优劣。由此得到ARIMA(1,2,1)的AIC值为-75.08,ARIMA(1,2,2)的AIC值为-77.93,ARIMA(2,2,2)的AIC值为-76.19。根据AIC值越小模型越好的原则,初步判定ARIMA(1,2,2)为优选模型。

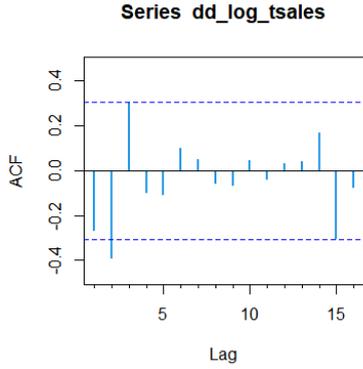


图5 序列 dd\_log\_tsales 自相关图

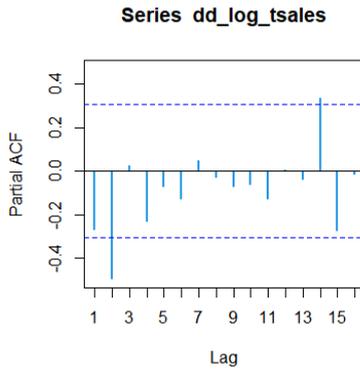


图6 序列 dd\_log\_tsales 偏自相关图

### 3.1.3 模型检验

一般来说,模型的显著性检验为残差序列的白噪声检验,若残差序列为白噪声,则意味满足均值为0的正态分布,模型前后信息提取充分,从而模型合适。如图7所示,残差图围绕零均值上下波动;若残差序列不是白噪声,则说明模型有改进的必要。下面通过qqnorm()和qqline()函数检验残差是否满足正态分布,如图8所示,数据中的点大致落在直线上,说明结果可行。

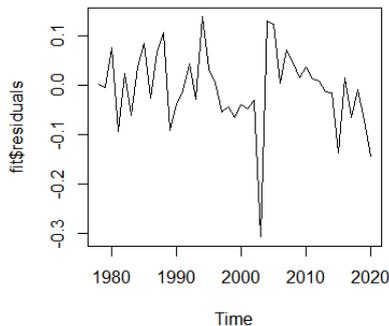


图7 残差图

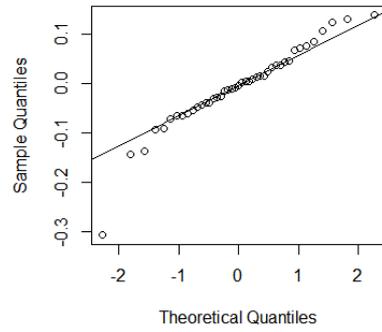


图8 残差序列正态 Q-Q 图

下面通过Box.test()函数检验残差的自相关系数是否都为0,结果显示,X-squared=17.984,df=18,p-value=0.4567,残差没有通过显著性检验,即可以认为残差的自相关系数为零,ARIMA(1,2,2)能较好地拟合数据。

### 3.1.4 模型预测

下面对2021至2022年广西社会消费品零售总额进行预测,利用2021和2022年的实际值与拟合值计算相对误差,结果如表1所示。

表1 2021—2023年实际值与预测值对比

年份	实际值(亿元)	预测值(亿元)	相对误差
2021	8538.5	8441.497	-1.14%
2022	8539.1	9701.722	13.62%
2023	—	10926.449	—

由表1可知,2021年的相对误差为-1.14%,2022年的相对误差为13.62%,误差均小于13.65%,说明模型拟合良好。图9表明未来三年的预测值保持上升的趋势。

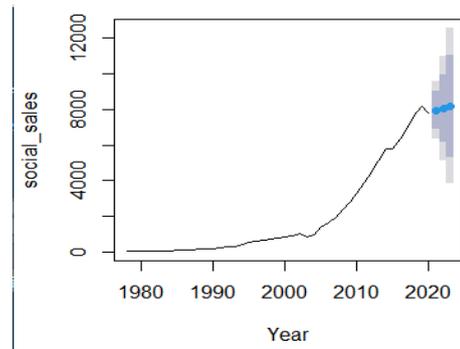


图9 社会消费品零售总额 ARIMA(1,2,2) 预测图

### 3.2 GM(1,1) 模型构建

建立灰色预测模型前需要进行可行性分析,只有数据通过级比检验,级比全部落在可容覆盖区间,才能够使用GM(1,1)模型。在对广西1987—2020年共43条数据进行计算时,发现大部分级比没有落入可容区间,无法通过可行性检验。GM(1,1)预测模型选用数据的基本原则是“重近轻远”,数据个数不低于4个时即可建立模型,但数据的波动会影响模型的预测精度。已知逐渐减少以前的数据可以令级比全部落入可容区间,经过反复测试直到数据选取到

1991—2020年，才全部通过可行性检验，故将1991—2020年数据作为原始序列。

GM(1,1)估计结果为发展系数  $-a=0.1100451$ ，灰色作用量  $u=495.584$ ，平均相对误差  $=46.36395\%$ 。对于模型精度检验，论文采用后验差比值检验法， $C$ 值  $=0.3210236 < 0.35$ ，说明GM(1,1)预测精度等级好。因此，在运用GM(1,1)模型1991—2020年数据的基础上对未来三年社会消费品零售总额进行预测，结果如图10所示。在没有突发事件的前提下，未来社会消费品零售总额有稳定增长趋势，预测精度较高，也证明运用该模型对社会消费品零售总额未来发展趋势的研究是合理的。

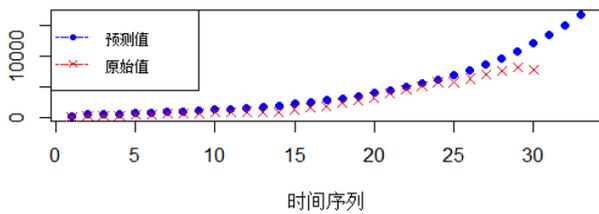


图10 社会消费品零售总额 GM(1,1) 预测图

#### 4 结语

论文通过对广西近40年来社会消费品零售总额进行ARIMA与GM(1,1)建模并对2021年和2022年以及未来一年的数据进行预测。ARIMA模型对2021和2022年预测的

相对误差均在14%左右，说明误差相对较小。GM(1,1)模型预测精度等级好，但相对误差较大，而且随着时间跨度的增加，模型的误差值也在增大。尽管如此，两种模型预测值仍有上升的趋势，预测效果不错，短期预测可以为政府决策提供参考。

#### 参考文献

- [1] 马强,张琛婕,陈雪平等.基于乘积季节模型的社会消费品零售总额分析[J].数学的实践与认识,2021,51(6):87-94.
- [2] 蒋翠清,李元申,丁勇等.基于VMD-DE-LSSVM误差校正的社会消费品零售总额组合预测模型[J].运筹与管理,2023,32(6):126-131+144.
- [3] 樊亮.基于ARIMA模型对社会消费品零售总额的预测分析——以甘肃省陇南市各县为例[J].通化师范学院学报,2020,41(8):29-34.
- [4] 尤游,刘苏兵.岭回归和主成分回归下的芜湖市社会消费品零售总额实证研究[J].信阳农林学院学报,2020,30(3):28-32.
- [5] 王爽,汪海飞.中国社会消费品零售总额的成分分解及预测——基于X-12-ARIMA模型的应用[J].经济研究导刊,2020(17):75-77.
- [6] 王志坚,王斌会.基于ARMA模型的社会消费品零售总额预测[J].统计与决策,2014(11):77-79.
- [7] 王黎明,王连,杨楠.应用时间序列分析[M].上海:复旦大学出版社,2009.
- [8] 郭改文,王诗涵.基于灰色理论与ARIMA模型的股票价格预测[J].河南教育学院学报(自然科学版),2023,32(2):22-27