

# 基于ARIMA模型的江苏省GDP预测分析

张媛媛

云南财经大学, 云南 昆明

收稿日期: 2022年3月21日; 录用日期: 2022年4月10日; 发布日期: 2022年4月20日

## 摘要

本文基于1975~2020年江苏省GDP数据, 运用R软件对1975~2015年的时间序列数据建立模型, 再通过比较AIC信息准则值以及观察自相关以及偏自相关图像, 确定最优模型: 疏系数模型。之后运用对比得到的最优模型对于江苏省2016~2020年的GDP数值进行预测分析, 并将预测值与真实值进行对比, 结果显示利用该模型进行预测的误差较小, 模型精度较高的结论。进而说明疏系数模型对于江苏省GDP预测工作的准确性。

## 关键词

ARIMA模型, 江苏省GDP预测, 时间序列, 疏系数

# GDP Forecast Analysis of Jiangsu Province Based on ARIMA Model

Yuanyuan Zhang

Yunnan University of Finance and Economics, Kunming Yunnan

Received: Mar. 21<sup>st</sup>, 2022; accepted: Apr. 10<sup>th</sup>, 2022; published: Apr. 20<sup>th</sup>, 2022

## Abstract

Based on the GDP data of Jiangsu Province from 1975 to 2020, R software was used to build a model for the time series data from 1975 to 2015, and then the optimal model was determined by comparing AIC information criterion values and observing autocorrelation and partial autocorrelation images: sparse coefficient model. After that, the optimal model obtained by comparison is used to forecast and analyze the GDP value of Jiangsu Province from 2016 to 2020, and the predicted value is compared with the real value. The results show that the prediction error of using this model is small, and the model has high accuracy. Then it shows the accuracy of the thinning coefficient model for GDP forecasting in Jiangsu Province.

## Keywords

ARIMA Model, GDP Forecast of Jiangsu Province, Time Series, Sparse Coefficient

Copyright © 2022 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

一个国家或者地区的经济发展状况一般用 GDP 来表示, 通过 GDP 可以掌握一个国家或者地区的经济发展状况, 保罗安东尼·萨缪尔森和诺德豪斯曾经将其称为“世纪最伟大的发明之一”。而江苏省的经济发展水平一直都处在国内经济发展的领先地位, 其 GDP 总值更是仅次于广东省, 居于第二位。江苏省作为经济发展大省, 其每年都为我国 GDP 的增长贡献了绝大部分, 因此分析江苏省 GDP 的发展状况, 并对其发展趋势进行预测显得尤为重要。本文从江苏省统计年鉴上查找 1975~2020 年 GDP 数据, 对其进行平稳性检验、异方差性检验以及纯随机性检验, 之后对序列拟合 ARIMA 模型, 并且利用拟合的模型预测江苏省 2016~2020 年的 GDP 数值, 并将预测值与实际值进行比较, 计算相对误差, 以此检验所构建模型的准确性。

## 2. 文献综述

随着经济的快速发展, 国内外学者对 GDP 发展规律展开了详细的研究。通过梳理文献发现预测 GDP 的方法大致可以分为四类, 分别为: 时间序列分析、回归分析、灰色预测和神经网络。

首先关于时间序列分析的研究。瞿海情和何先平(2021)建立 1978~2019 年湖北省 GDP 数据的 ARIMA(p,d,q)模型, 并利用得到的模型进行短期预测, 首先利用所建立的模型将 2018 年以及 2019 年湖北省 GDP 数值预测出来, 然后和实际值进行比较, 发现该模型对数据的拟合程度较好, 说明时间序列模型对于 GDP 的短期预测效果非常好[1]。潘典雅(2021)和瞿海情以及何先平的思路一样, 对吉林省 1993~2017 年的 GDP 数据建立 ARIMA 模型, 只是潘典雅在建立时间序列模型之前, 首先对 GDP 数据进行了一个平稳化处理以及检验, 并且对模型中出现的参数不仅进行了估计而且还进行了适用性检验[2]。陈玉霞(2021)也是利用时间序列模型对 GDP 进行预测, 但是不同于潘典雅所采用的年度 GDP 数据, 陈玉霞所采用的是 2003 年第一季度~2019 年第二季度共 66 个季度的季度 GDP 数据, 因此根据数据的特点, 陈玉霞所建立的是 SARIMA 模型, 之后运用拟合的最优模型 SARIMA 预测了下一年的 GDP 数值, 对比发现利用此模型对 GDP 进行预测准确性很高[3]。

其次关于回归分析的研究。除了运用时间序列模型对 GDP 进行预测之外, 还有很多学者借助回归分析法进行 GDP 预测, 并且不同学者所利用的回归方法也不相同。刘浏(2017)在甘肃省统计年鉴中搜寻了其 1978~2015 年的人均 GDP 数据, 对其建立了一元线性回归模型, 并利用所建立的模型预测了“十三五”期间甘肃省 GDP 以及人均 GDP 指数, 研究发现一元回归模型对于 GDP 预测效果较好[4]。魏乐航(2020)基于 1996~2019 年孝感市 GDP 数据, 对其建立自回归时间序列模型, 并预测未来 5 年孝感市 GDP 数值, 发现孝感市未来 5 年的 GDP 会呈现稳步上升的态势[5]。魏乐航所使用的自回归时间序列模型其实是将回归方法引入时间序列模型的一种方式, 而李永娣(2021)不仅使用了 ARIMA 模型, 还使用了主成分回归模型对河南省 GDP 总量以及增速进行预测, 结果显示时间序列模型与回归模型对 GDP 的预测有不同的作用, 时间序列模型更加侧重于对总量进行预测, 而回归模型则更加侧重于对引起 GDP 增加的因素进行分析[6]。

然后关于灰色预测的研究。石萍和唐俊(2015)基于 2004~2012 年包头市 GDP 数据, 运用灰色预测法

预测包头市 GDP 数值, 并通过检验发现此模型的有效性以及合理性[7]。之后王传会等(2015)在普通灰色预测模型的基础上, 为了在模型中引入经济系统中国的突变现象, 将泛函理论引入到灰色预测模型中, 之后又运用贝叶斯网络推理技术, 建立了灰色泛函预测 GFAM(1,1)模型[8]。而田梓辰和刘淼(2018)不同于王传会引入泛函理论, 而是在灰色 GM(1,1)模型的基础上, 引入改进的朗格朗日插值函数得到灰色 GM(1,1)模型, 并利用此模型预测新疆近十年的 GDP 数值, 研究表明利用此模型预测 GDP 的准确性[9]。

最后关于神经网络的研究。李南(2017)建立江西省 2000~2016 年 GDP 的 BP 神经网络模型, 并对 2018 年 GDP 数值进行预测[10]。然后晏荣堂(2018)在李南的基础上, 利用遗传算法优化的神经网络对 GDP 进行预测, 得到了一个更加准确的 GDP 预测模型[11]。朱青和周石鹏(2021)考虑到传统的时间序列方法预测 GDP 时, 对数据的要去较高, 只有当 GDP 数据为序列平稳时才可以使用, 为了解决这个难题, 朱青和周石鹏利用机器学习算法 Random Forest 选取影响 GDP 增长的重要变量, 之后利用深度学习当中的 LSTM 神经网络预测 GDP 的增长, 得到一个预测精度更高的预测模型[12]。

通过文献梳理可以发现学者们在对 GDP 的分析预测上虽然方法各异, 但是都取得了非常理想的效果。很多学者通过灰色预测理论以及神经网络分析理论对 GDP 进行预测, 但这两个方法不论是理解上还是建模难度上都比较大, 而时间序列模型不仅理解容易, 建模轻松, 而且在对 GDP 预测效果上并不比前两种方法差, 因此本文考虑选用时间序列分析模型对江苏省 GDP 数据进行预测分析。

### 3. ARIMA 模型的建立与预测

#### 3.1. 数据来源

本文所选取的数据是 1975~2020 年江苏省 GDP 数据, 是一个年度 GDP 时间序列数据, 所有的数据均可以从江苏省统计局官方获得, 本文在进行建模时, 江苏省统计局官网公布的最新的统计年鉴为《2021 年江苏省统计年鉴》, 因此 2020 年为目前所获得的最新数据。

#### 3.2. 原时间序列的平稳性检验

首先将 1975~2015 年的江苏省 GDP 数据录入到软件 R 当中, 首先画出该序列的时序图, 可见图 1, 对其平稳性进行初步的判断。从图中可以非常明显的看出, 江苏省 GDP 数据具有非常明显的上升趋势。接着利用 ADF 检验原序列的平稳性, 结果显示  $p = 0.9805$ , 因此不能拒绝序列非平稳的原假设, 故可判定该序列是非平稳的。

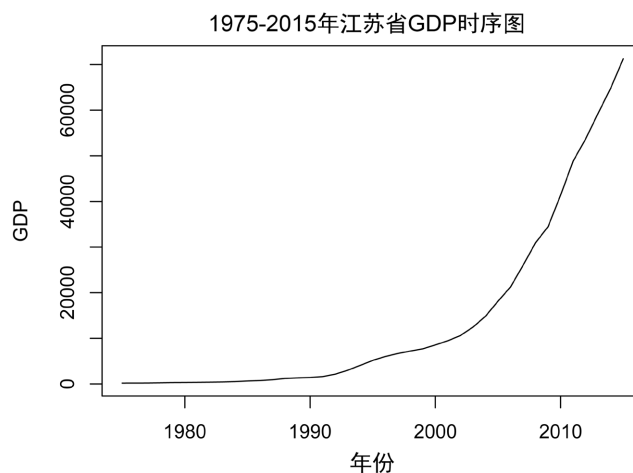


Figure 1. Sequence chart of GDP of Jiangsu Province from 1975 to 2015

图 1. 1975~2015 年江苏省 GDP 时序图

### 3.3. 数据处理

#### 3.3.1. 数据预处理

第二部分通过画时序图以及 ADF 检验已知该原始序列是非平稳的，因此下面利用 ADF 检验对一阶差分后的序列进行平稳性检验，检验结果可见表 1，从检验结果中可以看出  $p = 0.8141$ ，同样的不能拒绝一阶差分序列为非平稳的原假设，故判定此序列一阶差分后为非平稳的。

**Table 1.** ADF test results of original sequence and first order difference sequence

**表 1.** 原序列及一阶差分序列的 ADF 检验结果

	原序列	一阶差分序列
ADF 值	-0.4221	-1.3869
滞后阶数	3	3
$p$ 值	0.9805	0.8141

备择假设：序列平稳

由于原序列以及一阶差分序列都非平稳，因此对序列进行二阶差分处理，并利用 ADF 检验对两阶差分后的序列进行平稳性检验，下表 2 为单位根检验结果。从结果表中可知，若给定显著性水平  $\alpha = 0.05$ ，任一模型都有  $p < \alpha$ ，因此可以得出二阶差分后的序列是平稳的结论。

**Table 2.** ADF test results for second order difference sequences

**表 2.** 二阶差分序列的 ADF 检验结果

无截距无趋势			有截距无趋势			有截距有趋势		
Lag	ADF	p.value	Lag	ADF	p.value	Lag	ADF	p.value
0	-7.49	0.0100	0	-7.78	0.0100	0	-7.92	0.0100
1	-6.51	0.0100	1	-7.22	0.0100	1	-7.71	0.0100
2	-2.93	0.0100	2	-3.46	0.0178	2	-3.84	0.0275
3	-2.55	0.0132	3	-3.28	0.0251	3	-4.28	0.0100

备择假设：原序列平稳

#### 3.3.2. 序列纯随机性检验

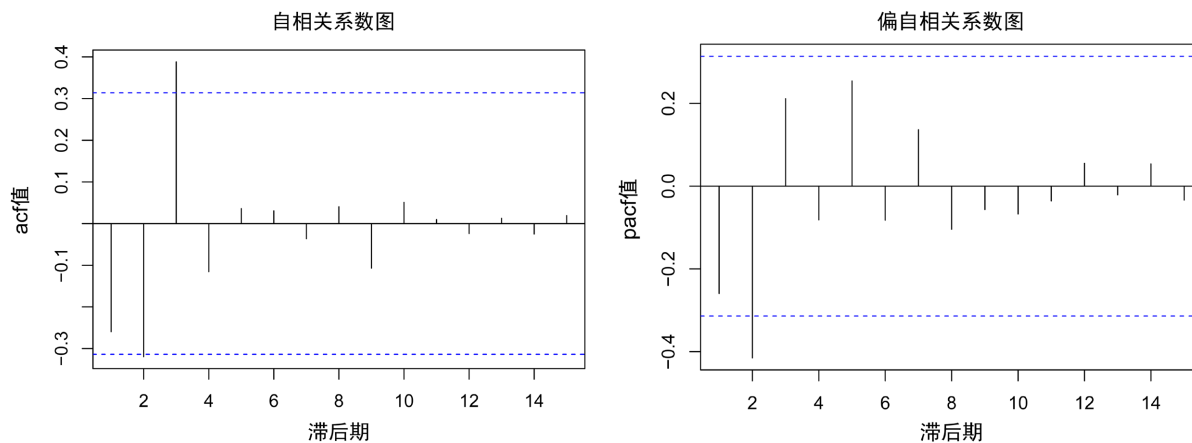
平稳性检验以及异方差检验结束后，下面需要对序列进行纯随机性检验，由于所选取的数据仅仅是 1975~2020 年共 46 年的江苏省 GDP 数据，进行二阶差分后会损失两个数值，因此检验的序列仅为 44 个，作为一个时间序列数据，相对而言比较短，因此这里的纯随机性检验仅检验到滞后 6 期，检验结果可见表 3。从检验结果表中可以看出，若给定显著性水平  $\alpha = 0.05$ ，不论是滞后 3 期还是滞后 6 期，其对应的  $p$  值都小于  $\alpha = 0.05$ ，故有充分的理由拒绝原假设，进而进行模型识别与定阶的任务。

#### 3.3.3. 模型的识别与定阶

为了建立正确的时间序列模型，首先需要知道模型的阶数，方法是通过二阶差分后的自相关图以及偏自相关图进行判断，自相关以及偏自相关的检验结果可见图 2。

**Table 3.** Pure randomness test results**表 3.** 纯随机性检验结果

滞后阶数	X-squared	p-value
3	12.491	0.005878
6	13.097	0.04152

**Figure 2.** Sequence autocorrelation and partial autocorrelation graphs**图 2.** 序列的自相关以及偏自相关图

从 acf 图以及 pacf 图可以看出，自相关图 3 阶截尾，而偏自相关图 2 阶截尾，因此可以拟合模型 ARIMA(2,2,3)、ARIMA(0,2,3)、ARIMA(2,2,0)，由于以上模型均通过了残差的白噪声检验，因此用上述任一模型对江苏省的 GDP 进行预测都是合理的，但是为了选择最优模型预测江苏省 GDP 走势，下面借助信息准则 AIC 辅助，判断三个模型的拟合结果可见表 4。

**Table 4.** Comparison of fitting results of different models**表 4.** 不同模型拟合结果比较

拟合模型	最大似然值	AIC
ARIMA(2,2,3)	-308.47	628.94
ARIMA(0,2,3)	-308.66	625.32
ARIMA(2,2,0)	-314.41	634.81

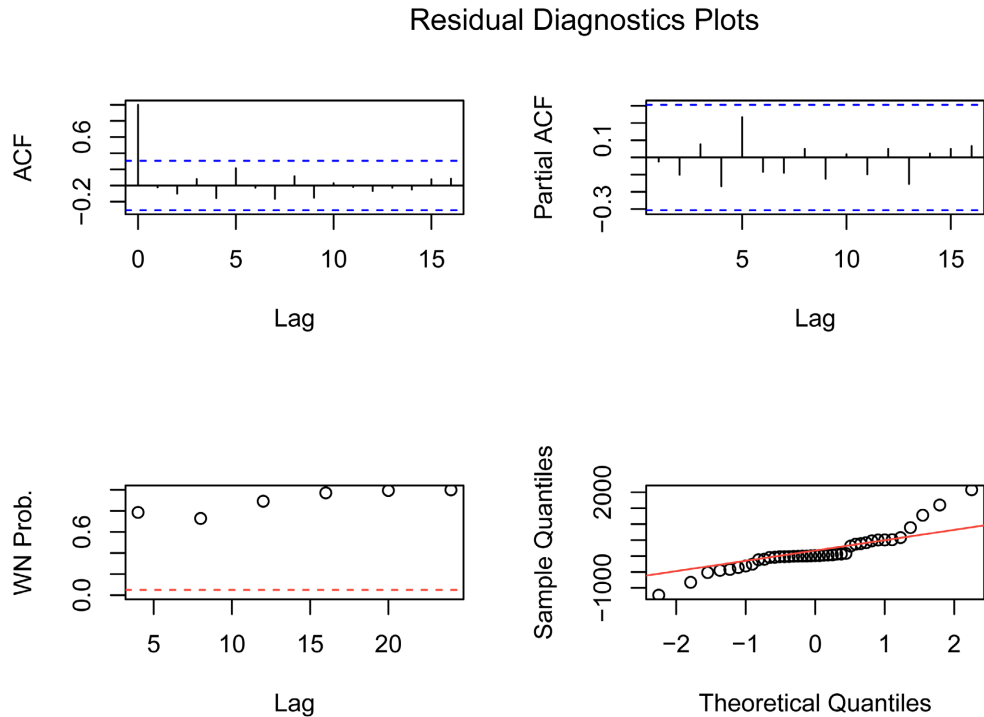
根据 AIC 指标数值最小的原则，最优的拟合模型是 ARIMA(0,2,3)，通过观察 acf 图以及 pacf 图可以看出，acf 图滞后 2 阶和滞后 3 阶显著不为 0，只有滞后 1 阶是显著为 0 的，同理，pacf 滞后 1 阶显著为 0，滞后 2 阶显著不为 0，因此可以建立疏系数模型 ARIMA((p<sub>1</sub>),2,(p<sub>2</sub>,p<sub>3</sub>))，通过 R 软件得到疏系数模型所对应的 AIC 准则值为 625.19，小于 ARIMA(0,2,3)模型的 AIC 值，因此采用 ARIMA((p<sub>1</sub>),2,(p<sub>2</sub>,p<sub>3</sub>))模型更优，所得到的最优模型为：

$$(1 + 0.1774B^2)(1 - B)^2 x_t = (1 + 0.4415B^2 - 0.71122B^3) \varepsilon_t$$

### 3.3.4. 模型的显著性检验

最后一步检验模型的显著性，检验结果可见图 3，检验结果显示残差序列是白噪声，因此可以说明

此模型拟合较优，将序列的所有信息都充分提取到。



**Figure 3.** Model significance test  
**图 3.** 模型显著性检验

### 3.3.5. ARIMA 模型的预测

下面利用上述得到的拟合模型，预测江苏省 2016~2020 年的 GDP 数据，之后用实际值与预测得到的拟合值进行比较，并且设定误差公式为： $error = \frac{test - fore}{test}$ ，其中 *test* 代表真实值，*fore* 代表预测值。

**Table 5.** GDP forecast results of Jiangsu Province from 2016 to 2020  
**表 5.** 2016~2020 年江苏省 GDP 预测结果

年份	真实值	预测值	误差
2016 年	77350.85	77681.14	-0.00427
2017 年	85869.76	84106.45	0.02053
2018 年	93207.55	90531.76	0.02871
2019 年	98656.82	96957.07	0.01722
2020 年	102719	103382.38	-0.00646

从表 5 中可以看出，将 2016 年至 2020 年这五年数值带入到预测模型中，所产生的预测误差都在 5% 之内，因此可以说明该模型用于拟合江苏省 GDP 数据的合理性。图 4 展示的是江苏省 GDP 未来的走势，从图中可以非常明显地看出，江苏省 GDP 在未来仍然将保持高速增长态势。

Forecasts form ARIMA(2,2,3) with non-zero mean

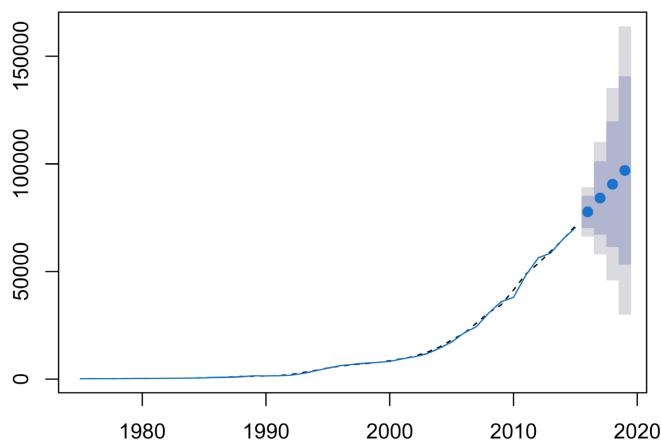


Figure 4. GDP forecast chart of Jiangsu Province

图 4. 江苏省 GDP 预测图

## 4. 结论与展望

### 4.1. 结论

本文建立了江苏省 1975~2020 年的 ARIMA 疏系数模型, 通过时序图发现 GDP 原序列非平稳, 故将其进行一次普通差分, 通过时序图发现差分一次的 GDP 数据仍然不平稳, 故对其进行二阶差分处理, 通过图示法以及 ADF 检验得出二阶差分后的数据平稳的结论。其后观察两次差分序列的 acf 以及 pacf 图特点, 对模型进行定阶处理, 比较各个可能的拟合模型, 基于 AIC 准则最小化原理选出最优拟合模型, 即疏系数模型  $ARIMA((p_1), 2, (p_2, p_3))$ , 并利用拟合模型预测江苏省 2016~2020 年 GDP 数值, 并将实际值和预测值进行比较, 并计算二者之间的误差, 结果这五年的预测值与真实值之间的误差均较小, 说明模型的拟合效果较好。最后通过图示法展现江苏省 GDP 增长趋势, 发现江苏省近几年的 GDP 一直保持稳定的速度持续上涨, 说明江苏省的发展态势较好。

### 4.2. 展望

本文在研究江苏省 GDP 发展特征时, 采用的 1975~2020 年的年度数据, 是从一个比较宏观和笼统的方面对江苏省 GDP 数据进行研究, 其实还可以查阅江苏省 GDP 月度数据, 更加细微地研究其增长变化情况。此外本文采用的是时间序列模型对 GDP 进行预测, 而关于预测的模型还有很多, 比如灰色预测模型, 神经网络模型等, 后续工作还可以利用这些模型对江苏省 GDP 数据进行预测分析, 并且比较选择最优预测模型。

## 参考文献

- [1] 瞿海情, 何先平. 基于时间序列分析的湖北省 GDP 预测模型研究[J]. 湖北经济学院学报(人文社会科学版), 2021, 18(9): 37-39.
- [2] 潘典雅. 基于 ARIMA 模型的吉林省 GDP 分析及预测[J]. 中国集体经济, 2021(27): 15-16.
- [3] 陈玉霞. 基于 SARIMA 模型的贵州省季度 GDP 预测[J]. 经营与管理, 2021(8): 170-175.
- [4] 刘浏. 基于线性回归的甘肃省“十三五”时期 GDP 预测[J]. 淮海工学院学报(人文社会科学版), 2017, 15(3): 90-92.
- [5] 魏乐航. 基于自回归时间序列的孝感市 GDP 分析与预测[J]. 住宅与房地产, 2020(27): 37-38.
- [6] 李永娣. 基于 ARIMA、平减指数法和主成分回归模型的 GDP 总量及增速预测——以河南省为例[J]. 河南科技, 2021, 40(13): 149-155.

- [7] 石萍, 唐俊. 灰色 GM(1,n)模型在经济预测中的应用——以包头市经济发展为例[J]. 内蒙古大学学报(自然科学版), 2015, 46(1): 43-47.
- [8] 王传会, 公维凤, 方志耕, 郭玉强. 灰色泛函预测 GFAM(1,1)模型及其应用[J]. 系统工程, 2015, 33(7): 115-120.
- [9] 田梓辰, 刘淼. 基于改进灰色 GM(1,1)模型的 GDP 预测实证[J]. 统计与决策, 2018, 34(11): 83-85.
- [10] 李南. 基于 BP 神经网络的江西省 GDP 预测[J]. 科技广场, 2017(10): 47-49.
- [11] 晏荣堂. 基于遗传算法优化的神经网络预测 GDP [J]. 科技视界, 2018(16): 4-5+31.
- [12] 朱青, 周石鹏. 基于 LSTM 模型的国民经济 GDP 增长预测建模研究[J]. 经济研究导刊, 2021(19): 5-9+37.