Modeling and Simulation 建模与仿真, 2023, 12(6), 5727-5736
Published Online November 2023 in Hans. https://www.hanspub.org/journal/mos
https://doi.org/10.12677/mos.2023.126520

Hans汉斯

基于ARIMA模型的浙江省GDP预测模型研究

魏欣雨，李旭芳*

上海工程技术大学管理学院，上海

收稿日期：2023年10月8日；录用日期：2023年11月21日；发布日期：2023年11月28日

摘　要

通过选取1952年至2020年浙江省国内生产总值的相关数据，建立了时间序列分析中的自回归滑动平均求和模型ARIMA(p,d,q)，利用该模型对浙江省GDP进行短期预测，为浙江省经济的发展提供参考。建立1952年至2018年浙江省GDP数据的时间序列，利用R语言软件建立ARIMA模型，并用该模型预测的2019年和2020年浙江省GDP数据与实际数据进行比较，对建立的模型进行优化评估，最后利用优化模型对2021年至2023年浙江省GDP进行短期预测。根据建立的时间序列分析得到最优模型为ARIMA(4,2,3)，预测值与实际值的平均相对误差为2.28%，ARIMA模型能较好地反映浙江省GDP发展的趋势并进行短期预测。

关键词

GDP预测，时间序列分析，ARIMA模型

Research on the GDP Forecasting Model of Zhejiang Province Based on ARIMA Model

Xinyu Wei, Xufang Li*

School of Management, Shanghai University of Engineering Science, Shanghai

Received: Oct. 8th, 2023; accepted: Nov. 21st, 2023; published: Nov. 28th, 2023

Abstract

By selecting the relevant data of the GDP of Zhejiang Province from 1952 to 2020, the autoregressive moving average summation model ARIMA(p,d,q) in the time series analysis is established. The short-term forecast provides a reference for the economic development of Zhejiang Province. Establish a time series of Zhejiang Province's GDP data from 1952 to 2018, use R language software

*通讯作者。

文章引用: 魏欣雨, 李旭芳. 基于 ARIMA 模型的浙江省 GDP 预测模型研究[J]. 建模与仿真, 2023, 12(6): 5727-5736.
DOI: 10.12677/mos.2023.126520

to establish an ARIMA model, and use the model to compare the Zhejiang Province's GDP data in 2019 and 2020 with the actual data to optimize and evaluate the established model. Finally, the optimization model is used to make a short-term forecast of Zhejiang Province's GDP from 2021 to 2023. According to the established time series analysis, the optimal model is ARIMA(4,2,3), and the average relative error between the predicted value and the actual value is 2.28%. The ARIMA model can better reflect the trend of GDP development in Zhejiang Province and make short-term predictions.

Keywords

GDP Forecast, Time Series Analysis, ARIMA Model

1. 引言

国内生产总值(GDP)是国民经济核算中的重要参数之一，它能够综合反映一个地区或国家在经济发展方面的实际情况。准确预测未来的 GDP 对于引导宏观经济的健康发展至关重要，能够为决策机构提供有益的指导，帮助他们做出更明智的决策。浙江省是中国长江经济带的重要的省份之一，不仅是 21 世纪海上丝绸之路圈定的五省之一，也是长江经济带三极之一长江三角洲城市群的重要组成部分。浙江省 GDP 的发展在一定程度上会受到国家生产总值的影响，因此，研究浙江省的 GDP 发展趋势，有利于对其发展方案的制定提供一个科学的参考。

田美雪[1]使用山东省 1990~2020 年的 GDP 数据，采用 ARIMA 模型进行建模分析。结果显示，2021 年山东省 GDP 总量稳步提升，预测增幅为 5.7%，而 2021 年山东省 GDP 实际值增幅为 8.3%，误差为 2.6%。这充分表明该模型拟合效果好，预测精度高。杨忠裕和薛紫玥[2]使用 ARIMA 模型预测甘肃省未来两年的 GDP 数值，为政府部门制订宏观经济计划提供依据和参考。结果表明，ARIMA 模型适用于 GDP 的增长预测。查华、石舢[3]在分析江苏省 GDP 的发展状况时，采用了 ARIMA 模型进行预测分析。经过比较，他们确定了 ARIMA(0,1,1)模型对 GDP 数据的拟合效果更好。实验结果也表明两年来江苏省 GDP 呈现出稳步增长的趋势。张文韬、李瑛琪[4]对河南省 GDP 数据进行了分析，并利用残差自回归模型对建立的 ARIMA(0,1,4)模型进行验证，结果表明该模型可以用于预测河南省的 GDP，同时也提出，该模型仅适用于短期预测，无法对河南省 GDP 未来的趋势做出准确预测。

本文通过 ARIMA 模型来进行建立模型，选用浙江省 1952~2018 年的 GDP 数据，对浙江省未来三年 GDP 进行预测，以期为政府制定经济政策提供参考和依据。

2. 数据来源及 ARIMA 模型

2.1. 数据来源

GDP 国民经济发展的重要指标，它代表着一个国家或者地区的发展状况和人民的生活状况。GDP 从某一方面说明着生活的发展、经济的增长、价格的变化和地区发展速度等，所以 GDP 指标对于一个地区制定其相应的政策和明确经济发展方向有着至关重要的作用。为了更好地探究浙江省未来几年的经济发

展趋势，同时也为了保证数据的真实性与准确性，本文使用国泰安数据库的“区域经济”子数据库中的相关文件中1952~2018年GDP历史数据如表1，建立浙江省GDP收入的ARIMA模型，分析其发展趋势，预测浙江省未来三年GDP收入(单位：亿元)。

Table 1. GDP data of Zhejiang Province from 1952 to 2018
表1. 1952~2018 浙江省历年 GDP 数据

年份	GDP	年份	GDP	年份	GDP	年份	GDP
1952	24.53	1969	62.7	1986	502.47	2003	9705.02
1953	27.24	1970	69.17	1987	606.99	2004	11648.7
1954	28.65	1971	70.43	1988	770.25	2005	13417.68
1955	30.53	1972	84.39	1989	849.44	2006	15718.47
1956	33.54	1973	86.99	1990	904.69	2007	18753.73
1957	37.27	1974	86.57	1991	1089.33	2008	21462.69
1958	43.75	1975	84.23	1992	1365.06	2009	22990.35
1959	47.35	1976	87.27	1993	1909.49	2010	27722.31
1960	47.25	1977	100	1994	2666.86	2011	32318.85
1961	40.56	1978	123.72	1995	3524.79	2012	34665.33
1962	43.35	1979	157.75	1996	4146.06	2013	37756.58
1963	46.6	1980	179.92	1997	4638.24	2014	40173.03
1964	51.75	1981	204.86	1998	4987.5	2015	42886.49
1965	55.67	1982	234.01	1999	5364.89	2016	47251.36
1966	58.28	1983	257.09	2000	6036.34	2017	51768.26
1967	56.35	1984	323.25	2001	6898.339	2018	56197.15
1968	54.94	1985	429.16	2002	8003.67	-	-

2.2. ARIMA 模型简介

ARIMA模型，即差分整合移动平均自回归模型(Autoregressive Integrated Moving Average Model)，是Box-Jenkins在20世纪70年代提出的一种著名时间序列预测方法。其基本思想是将随时间变化而形成的数据序列视为一个随机序列，并使用数学模型来描述和近似该序列。一旦模型被识别出来，就可以利用时间序列的过去值和现在值来预测未来值[5]。ARIMA模型的一般形式为ARIMA(p,d,q)，其中p、d、q分别指自回归阶数、差分阶数和移动平均数。主要应用于分析非平稳且不带有明显季节性变化趋势的时间序列。ARIMA模型的结构如下所示：ARIMA(p,d,q)。其中，p表示自回归系数，q表示移动平均阶数，d表示差分次数[6]。

$$\Phi(B)\nabla^d X_t = \Theta(B)\varepsilon_t$$

$$E(\varepsilon_t)=0, Var(\varepsilon_t)=\sigma^2\varepsilon, E(\varepsilon_t\varepsilon_s)=0, s\neq t, E(X_s\varepsilon_t)=0, \forall s<t$$

对 d 阶齐次非平稳序列 $\{X_t\}$ 而言，$\{\nabla^d X_t\}$ 是一个平稳序列，设其适合 ARIMA(p，q)模型，即 $(1-B)^d$ 为 d 阶差分。

$$\Phi(B)=1-\phi_1 B-\phi_2 B^2-\cdots\phi_P B^P$$

$$\Phi(B)=1-\theta_1 B-\theta_2 B^2-\cdots\theta_q B^P$$

2.3. ARIMA 模型建立及预测

2.3.1. 模型识别

确定合适的 p、d、q 值是识别 ARIMA 模型的首要任务。首先需要检验时间序列数据是否平稳。可以通过观察时间序列折线图初步判断，看其是否在一个固定值附近波动。若是平稳的，可以进一步使用 ADF 单位根检验方法进行验证，这种方法更准确且更具说服力。如果折线图呈增长或下降趋势，则意味着序列不平稳。此时，可以尝试取对数或进行差分操作来处理数据，再判断处理后的序列是否趋于平稳。如果仍然不平稳，需要重复这个过程，直到序列变为平稳状态。处理过程中所进行的差分次数即为 d 的值。一旦时间序列平稳，就可以使用 ARMA 模型求解问题。因此，ARIMA(p,d,q)模型的问题转变为 ARMA(p,q)模型的问题。然而，为了保证模型简洁且准确，需要避免过多差分导致的误差问题。

要得到平稳时间序列，可以通过绘制自相关图(ACF)和偏自相关图(PACF)来进行观察和分析，初步判断从而选择 ARIMA 模型中的 p 和 q 值。根据表 2 中的准则，对于不同的 p 和 q 值，所选取的 ARIMA(p,d,q)模型也会有所不同。此时，可以根据 AIC 或 BIC 准则来评估模型的优劣。当 AIC 或 BIC 值最小时，说明相应的 ARIMA(p,d,q)模型拟合效果最好。通过这种方法，可以为模型选择提供指导，并找到最佳的 ARIMA 模型。

Table 2. Selection principles for the ARMA(p,q) model
表 2. ARMA(p,q)模型的选择原则

ACF	PACF	选择模型
拖尾	p 阶截尾	ARMA(p,0)
q 阶截尾	拖尾	ARMA(0,q)
拖尾	拖尾	ARMA(p,q)

2.3.2. 模型估计

在 ARIMA 模型的估计中，通常采用极大似然估计法来估计 ARIMA(p,d,q)模型中的未知参数。极大似然估计是一种常用的参数估计方法，通过最大化观测数据的似然函数，确定模型中未知参数的值。在 ARIMA 模型中，极大似然估计法的目标是找到使观测数据的似然函数达到最大的参数值。具体而言，首先是根据所选择的(p,d,q)阶数，通过遍历不同的参数组合，建立多个 ARIMA 模型。然后，使用训练数据集对每个 ARIMA 模型进行拟合。

2.3.3. 模型检验

模型检验是用来评估时间序列模型的拟合质量的方法。如果模型的拟合效果较差，就需要进行模型的重新选择，直到得到最佳的拟合效果。模型检验包括对参数估计值和残差序列的检验。常用的检验指标包括参数估计值的显著性和残差序列是否为白噪声。当参数估计值显著，并且残差序列满足白噪声的条件时，可以认为模型的拟合效果较好，通过了模型检验。相反，如果参数估计值不显著，或者残差序

列不满足白噪声的条件，就需要重新选择模型，使其通过模型检验。在模型检验中，可以使用 Ljung-Box 统计量检验来判断残差序列是否满足白噪声的条件。Ljung-Box 统计量是基于残差序列的自相关函数(ACF)和偏自相关函数(PACF)构建的，用于检验残差序列的相关性是否显著。如果残差序列通过 Ljung-Box 统计量检验，即相关性不显著，可以认为残差序列是白噪声。

2.3.4. 模型预测

根据模型检验和比较的最终结果，在 R Studio 软件中利用构建的 ARIMA(p,d,q)模型进行预测。预测功能可以用来描绘原始时间序列图的未来变化趋势，并通过对比预测值与实际值进行误差分析，进一步验证模型的可行性。

3. 浙江省 GDP 时间序列 ARIMA 模型的应用

本文对浙江省 1952~2020 年的 69 个 GDP 数据进行了分析，为了检验模型的说服力以及正确性，现在选取前面 67 个 GDP 数据用来建模，并用后面 2 年的数据来检验模型的拟合效果，最后再来预测 2021~2023 年的 GDP。

3.1. 数据平稳性检验与处理

根据 1952~2018 年的浙江省 GDP 数据，画出时间序列图，如图 1 所示。

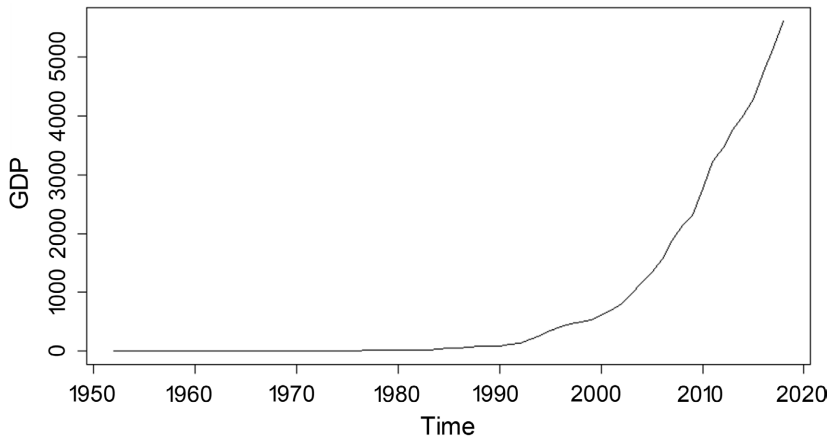

Figure 1. Time series plot of GDP of Zhejiang Province, 1952~2018
图 1. 1952~2018 年浙江省 GDP 的时间序列图

从图中可以观察到 GDP 数据呈现增长趋势，没有出现周期性和季节性波动，显然，这个时间序列初步被判定为非平稳。为了进一步验证，对该时间序列进行了 ADF 单位根检验，结果显示三个模型的统计量 P 值都为 0.99，大于显著性水平 0.05，因此可以得出这个时间序列是非平稳的结论

对于这个非平稳的序列，进行了第一次差分操作后，得到了一次差分后的时序图，如图 2 所示。在图中可以明显看到增长趋势，初步表明序列仍然不平稳。进行了单位根检验后，发现三个类型的模型的统计量的 P 值，均大于 0.05，这进一步证明了序列的非平稳性。

为了进一步处理这个序列的非平稳性，需要进行第二次差分操作。经过二次差分后得到的时序图如图 3 所示。从图中可以观察到序列在 0 值附近上下波动，但仍需进行单位根检验以得出结论。进行了单位根检验后，发现三个类型的模型的统计量的 P 值，均为 0.01，小于 0.05，说明此时的时间序列达到了平稳性要求。因此，可以确定在 ARIMA(p,d,q)模型中，取 d = 2。

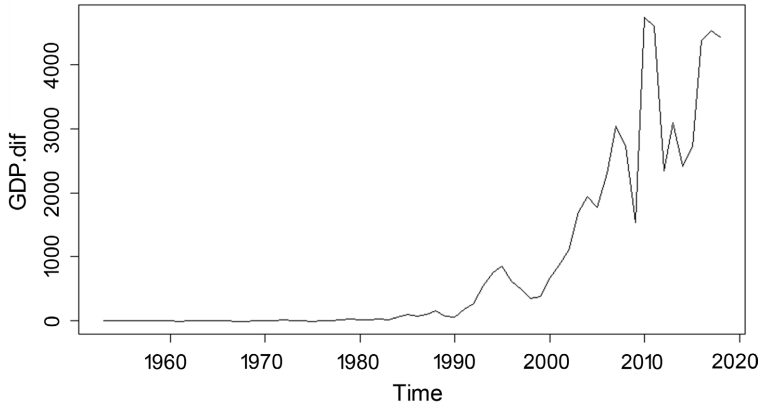

Figure 2. First-order differential GDP time series plot
图 2. 一阶差分 GDP 时间序列图

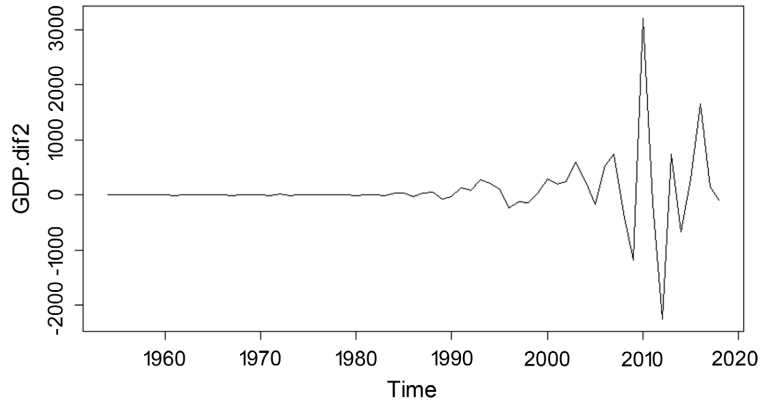

Figure 3. Second-order differential GDP time series plot
图 3. 二阶差分 GDP 时间序列图

3.2. 确定 ARIMA 模型的阶数

从上述分析得到，二次差分后的浙江省 GDP 序列是平稳的，利用 R Studio 软件，画出二次差分后序列的自相关 ACF 图和偏自相关 PACF 图，如图 4 和图 5 所示。

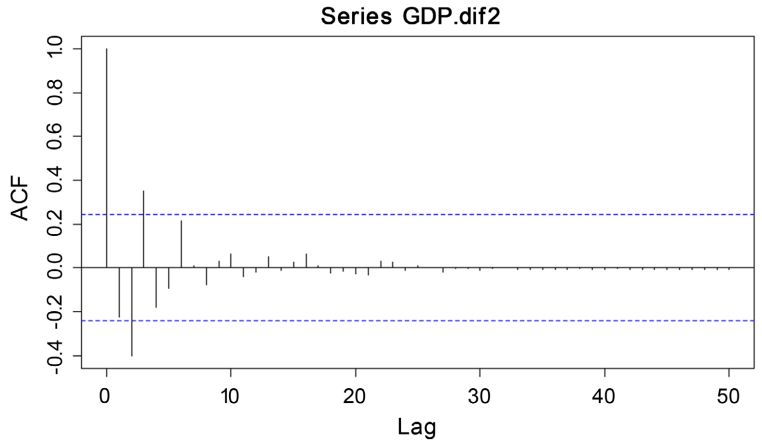

Figure 4. Autocorrelation plot after second-order difference
图 4. 二阶差分后的自相关图

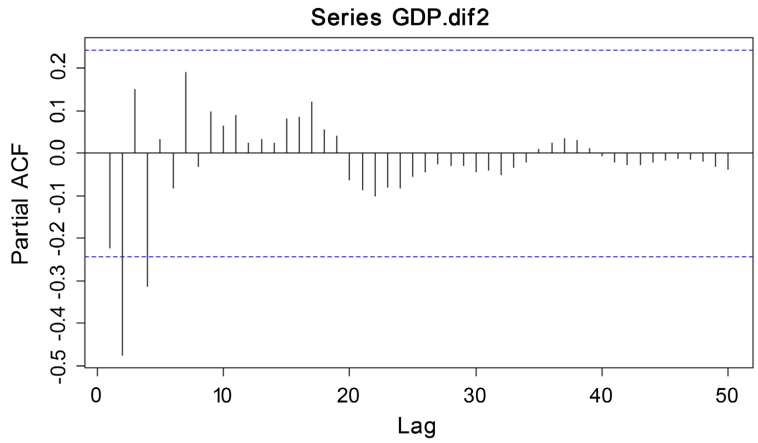

Figure 5. Partial autocorrelation plot after second-order difference
图 5. 二阶差分后的偏自相关图

观察分析两个图可以得到，自相关系数在 3 阶之后是逐渐趋于零的，偏自相关系数在 4 阶之后是也逐渐趋于零的，因此，P 值取 4，q 值取 3。但是，需要注意的是，这样选择模型的过程具有一定的主观性。为了消除误差，建立了多个 ARIMA 模型，并分别选择了 ARIMA(0,2,2)，ARIMA(0,2,3)，ARIMA(2,2,2)，ARIMA(4,2,2)，ARIMA(4,2,3)等。针对每个模型，计算了相应的 AIC 值，具体数据结果可见于表 3。从表中观察可以得出，ARIMA(4,2,3)模型的 AIC 值最小，所以可以认为该模型是最佳选择。

Table 3. Comparison of ARIMA models
表 3. ARIMA 模型的比较

	ARIMA(0,2,2)	ARIMA(0,2,3)	ARIMA(2,2,2)	ARIMA(2,2,3)	ARIMA(4,2,2)	ARIMA(4,2,3)
AIC	1002.57	997.01	997.98	999.76	1001.52	996.78

3.3. 模型的检验

对 ARIMA(4,2,3)模型的残差进行白噪声检验。通过 Ljung-Box 统计量检验，在给定的显著性水平下，观察残差的 P 值。具体检验结果及图示可参考表 4。根据检验结果，得到残差序列的 P 值分别为 0.9956 和 0.9999，均大于显著性水平 0.05。因此，可以认为该模型通过了白噪声检验，表明残差序列符合白噪声性质。基于这一结论，可以使用该模型进行预测。

Table 4. Residual white noise test results
表 4. 残差白噪声检验结果

滞后阶数	χ^2 检验统计量	P 值
6	0.6466	0.9956
12	1.359	0.9999

3.4. 模型的预测

利用通过检验的 ARIMA(4,2,3)模型，预测 2019~2023 年的浙江省 GDP，预测结果如表 5 和图 6 所示。同时，将 2019 年和 2020 年的预测结果与实际结果进行对比，并计算相对误差和实际误差，结果如表 6 所示。

Table 5. Predictions of the ARIMA(4,2,3) model
表 5. ARIMA(4,2,3)模型的预测结果

年份	预测值	预测最小值	预测最大值
2019	60,277	59,407	61,147
2020	63,817	62,061	65,574
2021	67,498	65,089	69,906
2022	71,468	68,256	74,681
2023	75,757	71,640	79,874

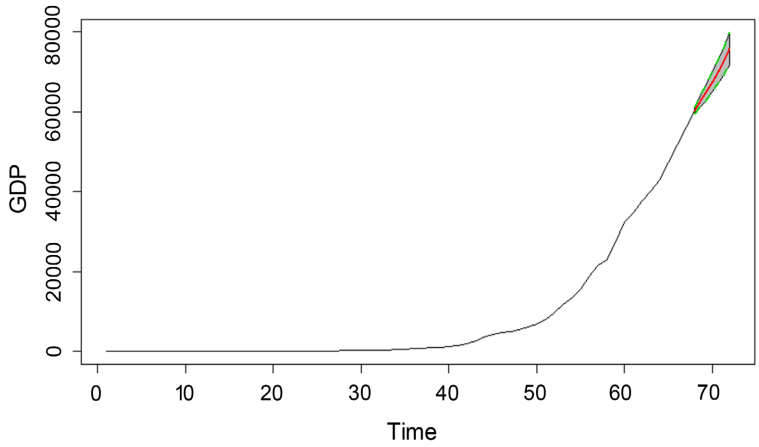

Figure 6. Trend plot of prediction results of ARIMA(4,2,3) model
图 6. ARIMA(4,2,3)模型的预测结果趋势图

Table 6. ARIMA(4,2,3) comparison of predicted and actual values
表 6. ARIMA(4,2,3)预测值与实际值比较

年份	实际值	预测值	相对误差(%)	平均误差(%)
2019	62351.74	60,277	3.33%	2.28%
2020	64613.34	63,817	1.23%	

根据表中的结果分析，平均相对误差为 2.28%。从数据上来看，该模型在短时间内对 GDP 的预测相对准确。浙江省在 1952 年至 2000 年左右的 GDP 增长相对平稳，但自 2001 年至 2020 年，增长非常迅速。这可能与阿里巴巴公司在杭州成立以及电子商务时代的到来有关。据此，我们可以推断，未来浙江省的 GDP 将持续增长并继续突破。

根据所得模型，预测浙江省 2021 年至 2023 年的 GDP 分别为 67498 亿元、71468 亿元和 75757 亿元。然而，由于自 2020 年起新冠肺炎疫情及国际疫情变异的影响，预测未来三年的 GDP 可能会受到影响并增加误差。因此，所预测的数据仅作为参考。

4. 结论

时间序列是一种广泛应用于预测未来数据的方法，它利用过去和现在的观测数据，其模型相对简单，对数据要求不高。

在建立模型、分析和预测过程中，首先要确保建模数据满足平稳性的条件。如果不满足平稳性，需要进行相应的处理，例如差分或取对数。然后，根据绘制的自相关图(ACF)和偏自相关图(PACF)来确定模型，本文中为 ARIMA(4,2,3)。接着，通过参数的显著性检验和残差的白噪声检验来增加模型的可信度和说服力。同时，比较预测结果与实际值之间的误差，也可以证明模型的可行性。

在实际应用中，我们需要注意，ARIMA 模型虽然在短期预测方面表现良好，但随着预测时间的延长，模型的误差会逐渐增大。这是因为 ARIMA 模型无法考虑到一些特殊的经济因素，比如国内外新冠肺炎疫情的影响、中国进口博览会等大型活动的举办以及国家和地区政策的推行等。这些因素都可能对 GDP 增长产生重要影响，但 ARIMA 模型无法直接体现。

因此，在使用 ARIMA 模型进行预测时，需要将其仅作为参考，结合其他经济模型或考虑额外的经济因素来提高预测的准确性和可靠性。

基金项目

上海市科委软科学重点项目“数字化驱动上海制造业绿色创新的机制及路径研究”(22692105100)。

参考文献

[1] 田美雪. ARIMA 模型在山东省 GDP 预测中的应用[J]. 中国管理信息化, 2023, 26(1): 139-141.

[2] 杨忠裕, 薛紫玥. 基于 ARIMA 模型的甘肃省 GDP 的分析与预测[J]. 中国市场, 2023(6): 1-4.

[3] 查华, 石舢. 基于 ARIMA 模型对江苏省 GDP 的预测[J]. 兰州文理学院学报(自然科学版), 2022, 36(3): 33-36+54.

[4] 张文韬, 李瑛琪. 基于 ARIMA 模型的河南省 GDP 指数分析[J]. 洛阳师范学院学报, 2019, 38(2): 11-14.

[5] 符文智. 基于 ARIMA-RF 组合模型的股价预测研究[J]. 科学技术创新, 2023(8): 40-43.

[6] 王燕. 应用时间学列分析[M]. 第 4 版. 北京: 人民大学出版社, 2015: 1-127.

附　录

```
#读入数据
getwd()
setwd(dir = "C:/Users/10475/Desktop/时间序列")
library(readxl)
ZJGDP <- read_excel("C:/Users/10475/Desktop/时间序列/ZJGDP.xlsx")
View(ZJGDP)  #导入 Excel 数据
#创建时间序列
G=ZJGDP$gdp
GDP=ts(G,start = c(1952),frequency = 1)
#数据平稳性检验与处理
plot(GDP)
adf.test(GDP,nlag = 2)  #单位根检验 GDP
GDP.dif<-diff(GDP)
plot(GDP.dif)  #一阶差分及时序图
adf.test(GDP.dif,nlag = 2)  #单位根检验 GDP.dif
GDP.dif2<-diff(GDP,1,2)
plot(GDP.dif2)  #二阶差分及时序图
adf.test(GDP.dif2,nlag = 2)  #单位根检验 GDP.dif2
acf(GDP.dif2,50)
pacf(GDP.dif2,50)  #自相关和偏自相关
#模型建立
GDP.fit<-arima(GDP,order = c(4,2,3))
GDP.fit  #拟合模型选取 AIC 值最小的模型
for(i in 1:2) print(Box.test(GDP.fit5$residuals,lag = 6*i))  #残差白噪声检验
GDP.fore<-forecast(GDP.fit5,5)  #预测未来五年数据
```