

# 基于干预SARIMA模型对疫情后民航客运量的预测

廖俊林

华南师范大学数学科学学院, 广东 广州

收稿日期: 2023年7月19日; 录用日期: 2023年8月9日; 发布日期: 2023年8月22日

## 摘要

民航客运量不仅是交通运输部门确定合理交通设施规模的基础, 同时也是保障机场设施高效率利用的前提。因此, 对疫情后民航客运量展开科学预测显得尤为重要。针对现阶段对客运量预测中未能定量考虑到新冠疫情对客运量影响的研究缺口, 本文在建立SARIMA(1,1,1) × (0,1,1)<sub>12</sub>模型对未发生疫情下我国民航客运量展开预测的基础上, 运用干预分析方法定量衡量新冠疫情对民航客运量的影响, 进而对疫情后民航客运量展开预测。结果表明, 相比于单一SARIMA模型, 干预SARIMA模型对疫情后民航客运量短期预测效果表现良好, 后续可采用类似干预分析方法将经济状况、政策变化、航空公司策略等事件考虑进来, 以更全面分析和预测民航客运量的变化。

## 关键词

民航客运量预测, 疫情干预分析, SARIMA模型

# Prediction of Civil Aviation Passenger Traffic after the Epidemic Based on the Intervention SARIMA Model

Junlin Liao

School of Mathematical Sciences, South China Normal University, Guangzhou Guangdong

Received: Jul. 19<sup>th</sup>, 2023; accepted: Aug. 9<sup>th</sup>, 2023; published: Aug. 22<sup>nd</sup>, 2023

## Abstract

Civil aviation passenger traffic is not only the basis for the transportation sector to determine the

reasonable scale of transportation facilities, but also a prerequisite to ensure the efficient use of airport facilities. Therefore, it is particularly important to make scientific forecasts of civil aviation passenger traffic after the epidemic. Based on the SARIMA(1,1,1) × (0,1,1)<sub>12</sub> model, this paper uses intervention analysis to quantitatively measure the impact of the new epidemic on China's civil aviation passenger traffic, and then forecast the post-epidemic civil aviation passenger traffic, in order to address the gap that the current passenger traffic forecast does not quantitatively take into account the impact of the new epidemic. The results show that the intervention SARIMA model is effective in predicting passenger traffic in China after the epidemic. The results show that compared with the single SARIMA model, the intervention SARIMA model performs well in the short-term prediction of civil aviation passenger traffic after the epidemic, and the intervention analysis method can be similarly used to take into account the economic situation, policy changes, airline strategies and other events in order to analyze and predict the changes in civil aviation passenger traffic more comprehensively.

## Keywords

Civil Aviation Passenger Traffic Forecasting, Epidemic Intervention Analysis, SARIMA 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

客运量是指在一定时期内, 各种运输工具实际运送的旅客数量, 是反映航运企业一定时期内运送旅客人数的指标。近年来, 随着经济水平的不断提高以及科学技术的不断发展, 民航业取得长足进步, 正逐步走近人们的生活中。截止到 2019 年底, 民航业完成旅客运输量 65993.42 万人次, 比 2018 年增长 7.9%, 继续保持近年来高速发展水平。但是, 自疫情爆发以来, 我国民航业受到巨大冲击, 期间实施的减少民众外出等一系列政策措施在一定程度上影响民航业的发展。民航客运量作为衡量民航业发展的重要指标, 不仅在一定程度上反映民航系统的运转情况, 而且能够帮助民航部门合理配置资源、制定运输计划。因此, 在当前这个关键时间点上, 对我国民航客运量的预测与研究显得尤为重要, 不仅能了解疫情对民航业的具体影响, 而且可以为民航部门未来一段时间的运输计划及资源配置提供参考和建议。

自二十世纪以来, 客运量预测问题成为诸多学者关注的热点, 取得了较多的研究成果。目前, 对客运量预测的研究方法主要有传统时间序列方法和机器学习方法两种方法。传统时间序列方法指通过对历史数据的分析揭示时间序列蕴含的发展规律, 进而对未来短期的数据展开预测。该方法下构造出的 ARIMA 等一系列模型简单高效, 在客运量的预测中得到广泛地应用。陈文华(2006) [1]在运用一阶差分提取公路客运量年度数据所含有趋势效应的基础上运用 ARIMA 模型对未来几年浙江的公路客运量展开预测。虽然该模型在年度客运量预测表现较好, 但是由于年与年之间涉及时间间隔太大, 不能快速、充分地反映客运量中所蕴含的市场变化和有效信息, 因此需对序列时间间隔进一步缩小。由于相比于年度客运量数据, 月度客运量数据还会表现出季节效应或周期效应, 郝军章(2015) [2]在运用时序图对月度数据进行基本分析的基础上, 经过对原序列逐期差分和季节差分的处理, 建立 ARIMA 乘法模型对未来 3 个月的数据展开预测, 虽然该模型能够充分提取月度客运量时间序列的趋势效应、季节效应以及随机效应, 从而对未来客运量的数据展开科学预测, 但是其残差序列可能会具有异方差性, 从而导致其模型残差的

方差估计不准确的现象。针对此问题,赵芳卉(2021) [3]加入 GARCH 模型消除 ARIMA 乘法模型可能存在回归残差的异方差问题,进而使客运量的预测精度得到更进一步的提升。

随着人工智能技术的日益发展,机器学习方法逐渐出现在客运量预测的研究中,如李思如[4] (2021)在对序列数据进行归一化处理以及对训练集和测试集进行划分的基础上,使用 LSTM 模型对未来几个月的民航客运量展开预测。相比于 ARIMA 模型,机器学习模型凭借其对已有样本特征的充分学习和训练,能够充分提取序列中 ARIMA 模型难以提取的非线性特征,进而得到更好的预测效果,但是其模型解释性较差,且容易出现过拟合的问题。针对传统时间序列模型和机器学习模型存在的优点和不足,不少学者从组合预测的角度切入展开深入研究,其组合思路[5]主要分为两种:一种为在运用传统时间序列模型和机器学习模型得出客运量预测值的基础上,通过等权重法、简单加权、方差倒数等方法对各个预测值加权组合的并联式组合模型方法[6] [7],另一种为在运用 ARIMA 系列模型对客运量预测的基础上,运用机器学习方法对其残差序列展开拟合和预测,进而通过两者的线性求和得到预测精度更高的客运量预测结果的串联式组合模型方法。这两者组合方法相比于单一模型,能够充分利用传统时间序列模型和机器学习模型的优势,提高预测的准确性、鲁棒性和全面性。

综合上述分析,目前学者对客运量预测的研究逐步形成一套较为完备的研究方法体系,能够很好地对未来的客运量科学而精确地预测。但是,现有文献大多数都是使用疫情前的客运量数据,鲜有文献在对客运量预测中定量考虑到新冠疫情对客运量的影响。针对这一研究缺口,本文从干预分析的角度去切入,在运用 SARIMA 模型对 2020 年未发生疫情情况下民航客运量展开预测的基础上,运用干预模型对疫情发生后带来的民航客运量趋势变动进行拟合和分析,进而对疫情发生后未来的民航客运量展开预测。

## 2. 数据来源及描述性分析

### 2.1. 数据来源

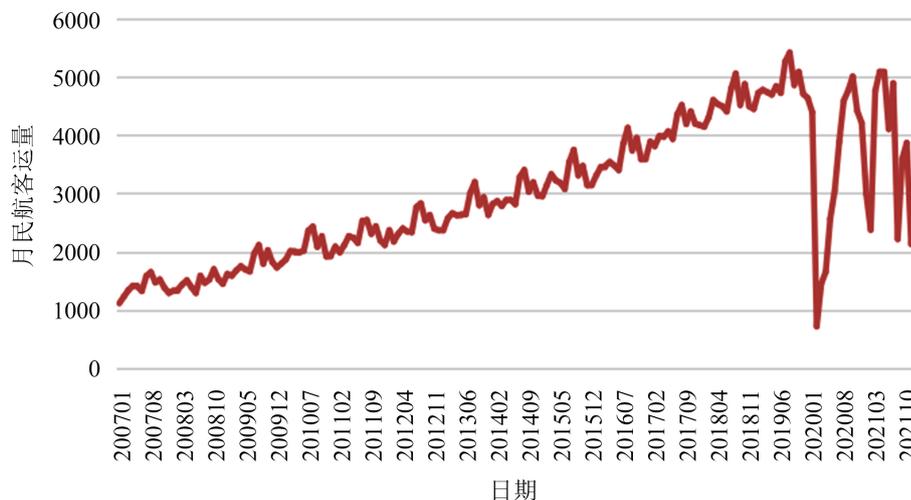
考虑到足够的样本数量是保证模型预测精确的基础保障,本文选取了 2007 年 1 月到 2021 年 12 月月度民航客运量时间序列数据。由于在 2020 年 1 月 30 日,世界卫生组织宣布新型冠状病毒肺炎疫情[8]被列为“国际关注的公共卫生紧急事件”,本文选取 2020 年 1 月 30 日作为干预事件的发生时间点,因此可将数据分为两个时期:第一个时期为 2007 年 1 月到 2020 年 1 月(新冠疫情发生之前),第二个时期为 2020 年 2 月到 2021 年 12 月(新冠疫情发生之后)。以上数据均来自中国经济社会大数据平台的国家统计局进度数据库。

### 2.2. 描述性分析

运用 excel 中“图表”功能绘制民航客运量的时序图,其结果如图 1 所示:

通过图 1 可以看出,民航客运量在 2007 年 1 月~2021 年 12 月整体上呈现出先上升后下降的趋势。得益于全球经济的发展和航空业的扩张,在疫情爆发前民航客运量一直呈现稳定增长的态势,但是,自 2020 年初以来,随着疫情的爆发和世界各地的旅行限制、封锁措施以及人们对旅行的担忧,航空公司遭遇了前所未有的挑战,航班的取消、航线的减少以及旅客需求的大幅下降导致了客运量的迅速萎缩。

尽管疫苗的研发和推广取得了一定的进展,使得民航客运量在疫情爆发后略有了一定回调趋势,但新冠病毒的变异和全球疫苗接种进度的不均衡仍然对航空业造成了持续的影响。旅行限制、健康检查措施和乘客需求的不确定性仍然存在,这使得恢复到疫情前的客运量水平变得困难。因此,若想对疫情后客运量的数据展开科学准确地预测,定量描述疫情对客运量的影响关系是必不可少的一部分。



**Figure 1.** Time series of monthly data of civil aviation passenger traffic from January 2007 to December 2021

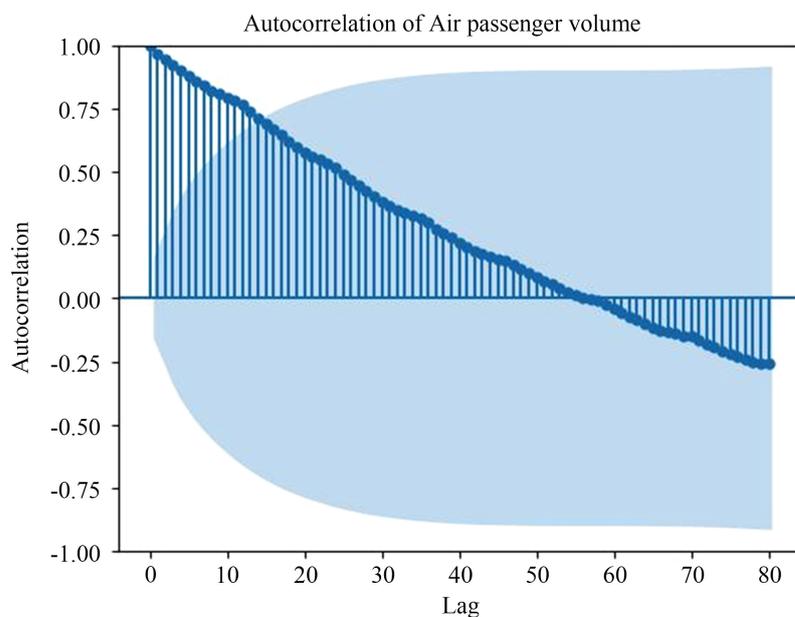
**图 1.** 2007 年 1 月~2021 年 12 月民航客运量月度数据时序图

### 3. 基于 SARIMA 模型对民航客运量的预测

#### 3.1. 模型选择

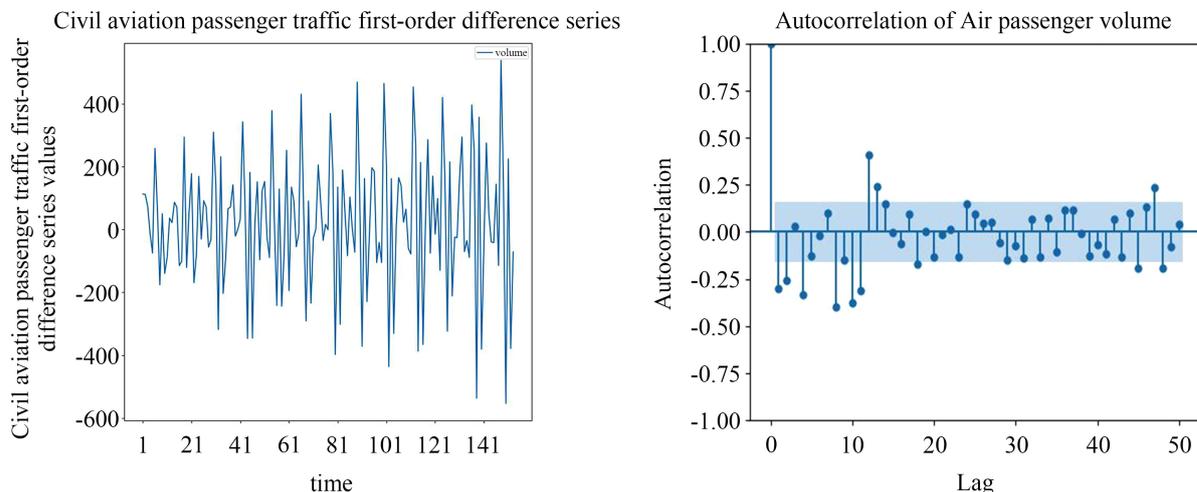
在民航客运量的预测中，常用的方法有传统时间序列分析方法和 LSTM、RF 等机器学习方法，考虑到民航客运量数据集样本量较少，且相比于机器学习算法，**ARIMA 系列模型简单高效，具有良好的可解释性**，本文选择使用 ARIMA 系列模型对民航客运量序列进行拟合并对未来民航客运量进行预测。

考虑到该时间序列呈现明显的线性趋势效应，对原序列展开一阶差分处理并依次绘制其时序图及原序列及一阶差分序列自相关图，具体结果分别如图 2，图 3 所示。



**Figure 2.** Autocorrelogram of the original civil aviation passenger volume series

**图 2.** 原民航客运量序列的自相关图



**Figure 3.** Timing diagram and autocorrelation diagram of first-order difference sequence (left is the timing diagram, right is the autocorrelation diagram)

**图 3.** 一阶差分序列的时序图和自相关图(左为时序图, 右为自相关图)

通过图 1~3 可以看出, 在疫情发生之前(2020 之前), 该时间序列呈现明显的趋势效应和季节效应, 具体表现出的时间序列特征如下:

1) **该时间序列呈现增长的趋势效应。**如图 2 可知, 原始序列的自相关图在 56 阶之前持续为正, 56 阶之后持续为负, 呈现明显倒三角特征, 结合时序图说明该序列具有增长的趋势效应。

2) **该时间序列呈现夏季偏高, 冬季偏低的季节效应特征。**通过图 3 可知, 在经过差分处理消除原序列趋势效应的基础上, 原序列的时序图和自相关图均呈现一定的季节周期特征。通过观察对比每年各个月的数据, 可以看出在 1 年中 7, 8 月的数据相对是最高的, 而 1, 2 月的数据相对是最低的。由于 7, 8 月正值学生暑假时期, 且夏天少雨, 人们倾向于选择在夏季进行长途旅行、探亲访友或度假, 从而推动了客运量的增加, 而 1, 2 月份天气寒冷, 节假日较少且春节机票价格相对较贵, 因此这段时间的客运量会相对减少。

3) **该时间序列的趋势效应和季节效应具有一定的相关性。**通过图 1 时序图可以看出, 随着趋势的递增, 每个周期波动范围也逐步扩大, 呈现喇叭形的形状, 说明其趋势效应和季节效应具有一定相关性。

基于疫情前客运量序列数据表现出的特征, 本文选择运用 **SARIMA 乘法模型**构建未发生疫情下客运量的预测模型。

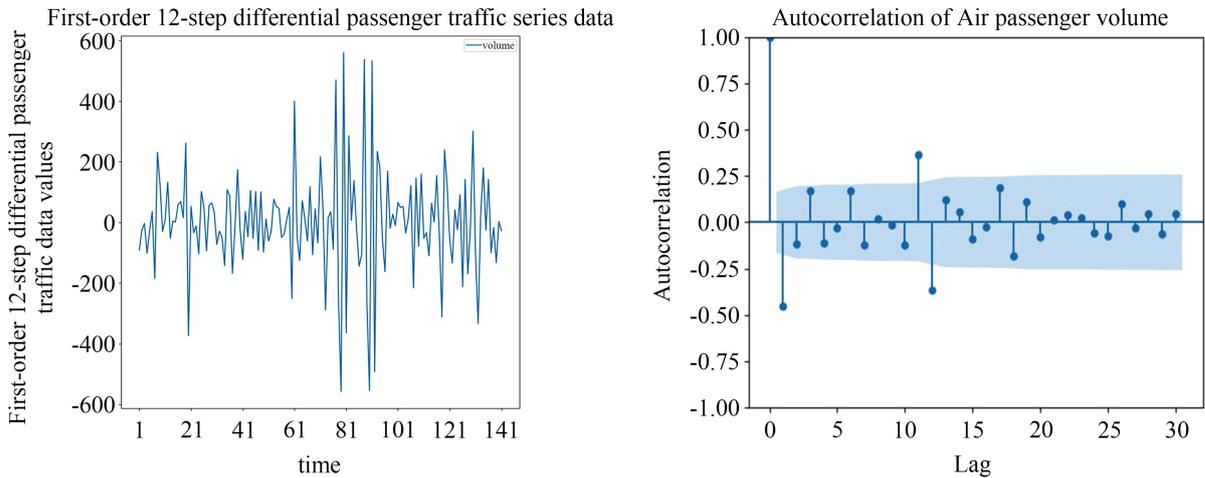
## 3.2. 模型的建立与检验

### 3.2.1. 差分平稳化

通过 python 对一阶差分序列再进行 12 步的周期差分, 力求提取原序列季节效应。差分后序列的时序图和自相关图如图 4 所示。

通过图 4 左边时序图可以看出, 经过一阶 12 步差分之后, 客运量围绕在 0 值附近, 在[-600, 600]的范围内波动, 且没有明显的周期效应和趋势效应, 结合图 4 右边自相关图显现出除 0 阶, 1 阶, 11 阶, 12 阶之外, 其他各阶自相关系数均落在两倍标准差范围内, 且没有倒三角和周期现象的特征规律, 可初步判断一阶 12 步差分变换充分提取原序列中蕴含的趋势效应和季节效应, 从而使序列平稳化。

考虑到运用时序图和自相关图检验来判断该序列的平稳性具有一定主观性。为了使检验结果更加客观准确, 本文引入 ADF 单位根统计量进一步辅助判别。ADF 单位根统计检验结果如表 1 所示。



**Figure 4.** Timing and autocorrelation diagrams of a first-order 12-step difference sequence (left is the timing diagram, right is the autocorrelation diagram)

**图 4.** 一阶 12 步差分序列的时序图和自相关图(左为时序图, 右为自相关图)

**Table 1.** ADF test results for the 1st order 12-step difference series

**表 1.** 一阶 12 步差分序列的 ADF 检验结果

类型	延迟阶数	模型结构	统计量 $\tau$ 值	P 值
类型一	0	$x_t = \epsilon_t$	-19.274	<0.0001
	1	$x_t = \varnothing_1 x_{t-1} + \epsilon_t$	-15.388	<0.0001
	2	$x_t = \varnothing_1 x_{t-1} + \varnothing_2 x_{t-2} + \epsilon_t$	-10.581	<0.0001
类型二	0	$x_t = \varnothing_0 + \epsilon_t$	-19.205	<0.0001
	1	$x_t = \varnothing_0 + \varnothing_1 x_{t-1} + \epsilon_t$	-15.332	<0.0001
	2	$x_t = \varnothing_0 + \varnothing_1 x_{t-1} + \varnothing_2 x_{t-2} + \epsilon_t$	-10.544	<0.0001
类型三	0	$x_t = \infty + \beta t + \epsilon_t$	-19.146	<0.0001
	1	$x_t = \infty + \beta t + \varnothing_1 x_{t-1} + \epsilon_t$	-15.303	<0.0001
	2	$x_t = \infty + \beta t + \varnothing_1 x_{t-1} + \varnothing_2 x_{t-2} + \epsilon_t$	-10.540	<0.0001

如表 1 可知, 类型一和类型二中几乎所有的统计量 P 值均小于 0.05, 在显著水平为 0.05 的情况下, 拒绝原假设, 可以确定该序列经过一阶 12 步差分后使序列平稳化。

### 3.2.2. 纯随机性检验

运用 LB 统计量对原始序列展开纯随机性检验, 得出的结果如表 2 所示;

**Table 2.** Results of the pure randomness test for the LB statistic of the first-order 12-step difference series

**表 2.** 一阶 12 步差分序列的 LB 统计量纯随机性检验结果

延迟阶数	LB 统计量的值	P 值
2	42.351478	$1.567118 \times 10^{-7}$
4	88.179154	$1.110630 \times 10^{-13}$
6	103.340766	$5.389865 \times 10^{-14}$

如表 2 可知, 各阶延迟下 LB 统计量的 P 值均小于 0.05, 拒绝序列为纯随机序列的原假设, 说明该一阶 12 步差分后的序列为非白噪声序列, 具有短期相关性的特点。

### 3.2.3. 模型定阶

绘制一阶 12 步差分序列的自相关图和偏自相关图对 SARIMA 模型进行定阶, 具体结果如图 5 所示。

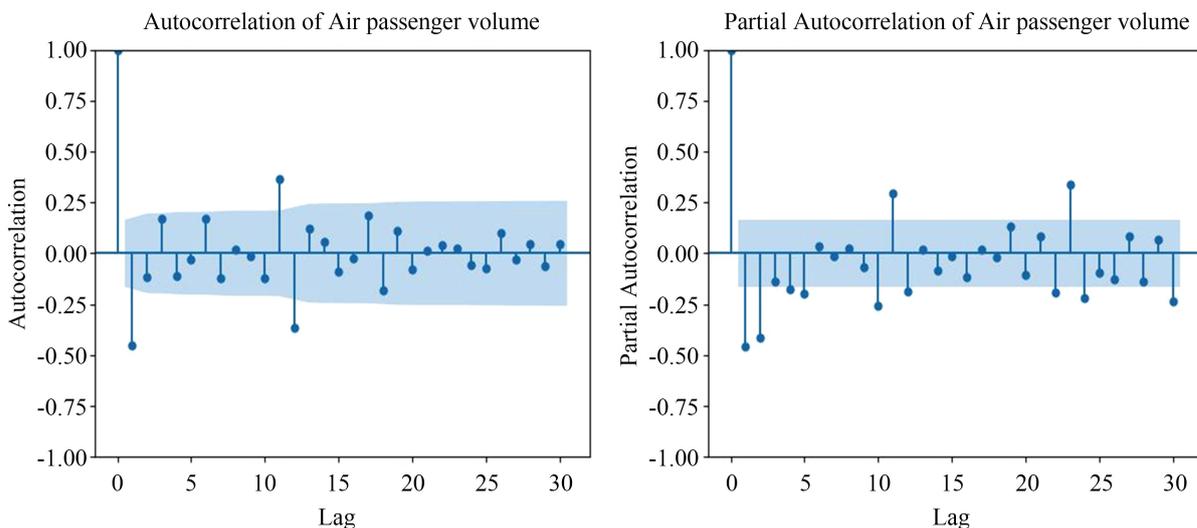


Figure 5. Autocorrelation and partial autocorrelation plots of first-order 12-step difference series (left autocorrelation plot, right partial autocorrelation plot)

图 5. 一阶 12 步差分序列的自相关图和偏自相关图(左自相关图, 右偏自相关图)

结合图 5 的自相关图和偏自相关图, 最终可使用  $ARIMA(0,1,1) \times (0,1,1)_{12}$  作为本文的预测模型, 考虑到图示定阶法具有一定的主观性, 本文设计  $ARIMA(0,1,1) \times (0,1,1)_{12}$ ,  $ARIMA(1,1,1) \times (0,1,1)_{12}$ ,  $ARIMA(1,1,1) \times (1,1,1)_{12}$ ,  $ARIMA(0,1,1) \times (1,1,1)_{12}$  四个模型, 通过 BIC 准则选取四个模型相对最优的模型作为最终的疫情前民航客运量的拟合和预测模型, 四个模型对应的 BIC 值如表 3 所示。

Table 3. BIC values of the four models

表 3. 四个模型的 BIC 值

模型形式	BIC 值
$ARIMA(0,1,1) \times (0,1,1)_{12}$	1654.2714
$ARIMA(1,1,1) \times (0,1,1)_{12}$	1633.5831
$ARIMA(1,1,1) \times (1,1,1)_{12}$	1635.5814
$ARIMA(0,1,1) \times (1,1,1)_{12}$	1656.2316

如表 3 可知,  $ARIMA(1,1,1) \times (0,1,1)_{12}$  乘法模型 BIC 信息量较小, 是四个模型中相对最优的模型。因此, 下文将运用  $ARIMA(1,1,1) \times (0,1,1)_{12}$  模型展开参数估计、模型检验与预测。

### 3.2.4. 参数估计

在对模型进行定阶的基础上, 运用最小二乘估计方法对模型进行参数估计, 得出的结果如表 4。

**Table 4.** Parameter estimation and significance test results of  $ARIMA(1,1,1) \times (0,1,1)_{12}$  model  
**表 4.**  $ARIMA(1,1,1) \times (0,1,1)_{12}$  模型的参数估计及显著性检验结果

参数	参数值	t 统计量的值	P 值
$\theta_1$	0.005	-42.352	<0.001
$\varphi_1$	-0.005	42.575	<0.001
$\theta_{12}$	0.9927	4.43	0.046

通过表 4 可以看出, 使用最小二乘估计得出模型的口径为:

$$\nabla_{12}x_t = \frac{1-0.005B}{1+0.005B}(1-0.9927B^{12})\varepsilon_t, \text{Var}(\varepsilon_t) = 1908.8483 \quad (1)$$

此外, 从表 4 可以看出, 各参数统计量对应的 P 值均小于 0.05, 通过了参数的显著性检验。

### 3.2.5. 模型检验

为进一步保证上述构造的  $ARIMA(1,1,1) \times (0,1,1)_{12}$  模型显著有效, 能充分提取民航客运量的信息, 利用 LB 统计量对其残差项进行纯随机性检验。纯随机性检验结果如表 5 所示。

**Table 5.** Significance test results of  $ARIMA(1,1,1) \times (0,1,1)_{12}$  model  
**表 5.**  $ARIMA(1,1,1) \times (0,1,1)_{12}$  模型的显著性检验结果

延迟阶数	LB 统计量值	P 值
6	4.4927	0.6103
12	12.7396	0.3883
18	17.2824	0.5037

如表 5 可知, 该模型各阶延迟下 LB 统计量的 P 值均显著大于 0.05, 可认为该拟合模型的残差序列属于白噪声序列, 即拟合模型通过模型有效性检验。

### 3.3. 模型的预测与评价

基于上文构建得出的  $ARIMA(1,1,1) \times (0,1,1)_{12}$  预测模型, 对未来 24 个月(2020 年 1 月~2021 年 12 月)未发生疫情情况下我国的民航客运量展开预测。2020 年 1 月~2020 年 5 月的预测值和预测区间如表 6 所示:

**Table 6.** Forecast values and forecast intervals for January~May 2020 in the absence of an outbreak  
**表 6.** 未发生疫情情况下 2020 年 1 月~5 月的预测值和预测区间

月份	预测值	预测区间(置信水平为 95%)
2020 年 1 月	4958.21	[4698.2039, 5218.2170]
2020 年 2 月	5033.60	[4765.5411, 5301.6615]
2020 年 3 月	5091.24	[4816.8461, 5365.6454]
2020 年 4 月	5047.39	[4766.8406, 5327.9399]
2020 年 5 月	5139.11	[4852.5450, 5425.8775]

通过 python 中 plt 库，将 2007 年 1 月~2019 年 12 月的序列实际值，序列拟合值以及 2020 年之后的序列预测值及实际值联合作图，绘制得出的图像如图 6 所示：

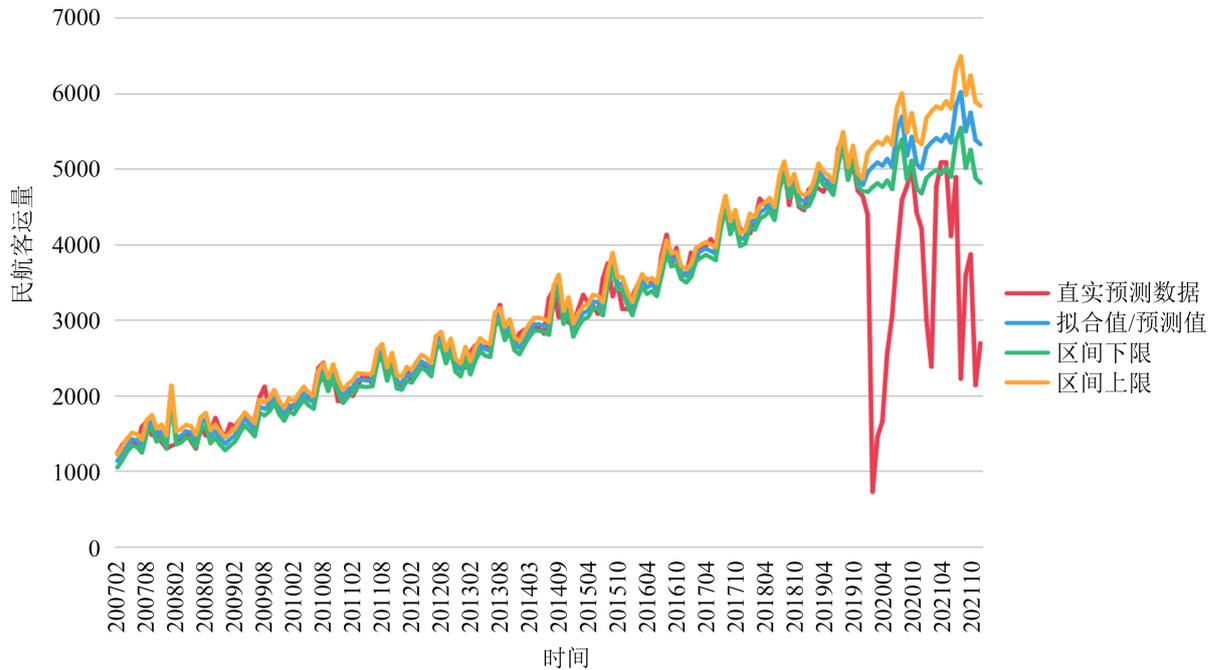


Figure 6. Fitting and forecasting effect of the series of civil aviation passenger traffic in China

图 6. 我国民航客运量序列的拟合和预测效果图

通过图 6 可以看出：

1) 在拟合区域中，每个序列时间点的拟合值与真实值的差异较小，且拟合曲线均落在 95% 置信区间的范围之内，说明该乘法模型对未发生疫情前民航客运量序列数据的拟合效果较好。

2) 在预测区域中，该拟合模型对未发生疫情时期客运量的预测精度较高，但从 2020 年 2 月疫情爆发之后，预测时间点真实值远远比实际值预测值小，其实际曲线均落在 95% 预测区间范围之外，间接说明新冠疫情的爆发导致民航客运量的大幅减少，使得民航业发展受到一定的困扰。

#### 4. 基于干预模型衡量疫情对民航客运量的影响程度

##### 4.1. 干预问题分析与有效性检验

近年来，新冠肺炎疫情事件对人们生活和社会经济造成不同程度的冲击，阶段性地影响着人们日常生活和社会发展。飞机作为我国重要交通工具，其疫情后客运量变化能够较为直观地体现新冠疫情事件对人民生活的影响，研究分析疫情对客运量的冲击程度和发展趋势有利于更好预防和应对将来疫情对客运量可能造成的影响。

突发事件对客运量的影响实际上是一个突发事件的干预问题[9]，这种干预可以运用干预分析模型[10]，从定量角度评估新冠疫情干预事件对民航客运量的影响。

干预分析模型在客运量上的预测过程[8]如下：

假设  $Y_t$  为疫情后民航客运量的某个指标值， $X_t$  表示无疫情干预影响时的民航客运量值， $Z_t$  为疫情干预影响值，则可以构造(2)式：

$$Y_t = X_t + Z_t \tag{2}$$

式中,  $Y_t$  为疫情前后民航客运量的真实值, 由于上文构建  $ARIMA(1,1,1) \times (0,1,1)_{12}$  对疫情前民航客运量数据的拟合效果和预测效果表现良好, 因此将  $X_t$  近似为 SARIMA 模型得出的民航客运量预测值。

通过式(2), 可以提取出 2020 年 2 月以后疫情干预的影响趋势, 为进一步检验疫情对民航客运量的影响是确切存在的, 本文首先使用 LB 统计量对干预影响趋势进行纯随机性检验。纯随机性检验的结果如表 7 所示。

**Table 7.** Results of the pure randomness test for the LB statistic of the intervention impact series  
**表 7.** 干预影响序列的 LB 统计量纯随机性检验结果

延迟阶数	LB 统计量的值	P 值
1	4.385508	0.036246
2	5.143568	0.076399
3	5.409479	0.144154

如表 7 可知, 干预影响序列在延迟阶数为 1 时其 LB 统计量的 P 值小于 0.05, 拒绝序列为纯随机序列的原假设, 可认为该序列为非白噪声序列, 进而从统计检验的角度说明疫情对民航客运量造成的影响是真实存在的。

但值得注意的是, 当延迟阶数大于 2 时, 其 LB 统计量的 P 值均大于 0.05, 说明干预序列仅仅只有短期自相关性, 因此下文在建立干预模型时延迟阶数不宜设得过大。

## 4.2. 干预影响分析方程的建立与求解

### 4.2.1. 基本原理

干预模型[10]是指在单变量序列的 ARIMA 系列模型中加进各种干预变量影响的模型。一般来说, 干预变量主要分为两种: 第一种为持续性的干预变量(用  $S_T$  表示), 一种为短暂性的干预变量(用  $P_T$  表示), 两种干预变量的表达形式如式(3)和式(4)所示。

$$S_T = \begin{cases} 0, & \text{干预变量发生之前}(t < T) \\ 1, & \text{干预变量发生之后}(t > T) \end{cases} \quad (3)$$

$$P_T = \begin{cases} 0, & \text{干预变量发生之前}(t = T) \\ 1, & \text{干预变量发生之后}(t \neq T) \end{cases} \quad (4)$$

其中,  $t$  是时间变量,  $T$  表示干预变量发生年份。

综合现有文献[10] [11] [12], 干预事件的影响形式虽然千姿百态, 但按其影响的形式, 可以归结为以下 4 种基本类型或其组合形式。

- 1) 干预事件的影响突然开始, 且其影响是长期的。这种干预事件的影响可以表示为:

$$Z_t = wS_T, \quad (5)$$

其中,  $w$  为干预影响的强度。

- 2) 干预事件的影响逐渐开始, 且其影响是长期的。这种干预事件的影响可表示为:

$$Z_t = \frac{wB^b}{1 - \delta_1 B - \dots - \delta_k B^k} S_T, \quad (6)$$

其中,  $w$  为干预影响的强度,  $B$  为后移算子,  $b$  为延迟期数。

3) 干预事件的影响突然开始, 且其影响是短期的。这种干预事件的影响可表示为:

$$Z_t = wP_T, \tag{7}$$

其中,  $w$  为干预影响的强度。

4) 干预事件的影响逐渐开始, 且其影响是短期的。这种干预事件的影响可表示为:

$$Z_t = \frac{wB^b}{1 - \delta_1 B - \dots - \delta_k B^k} P_T, \tag{8}$$

其中,  $w$  为干预影响的强度,  $B$  为后移算子,  $b$  为延迟期数。

通过 excel 软件绘制干预序列  $\{Z_t\}$  的时序图, 其时序图如图 7 所示。

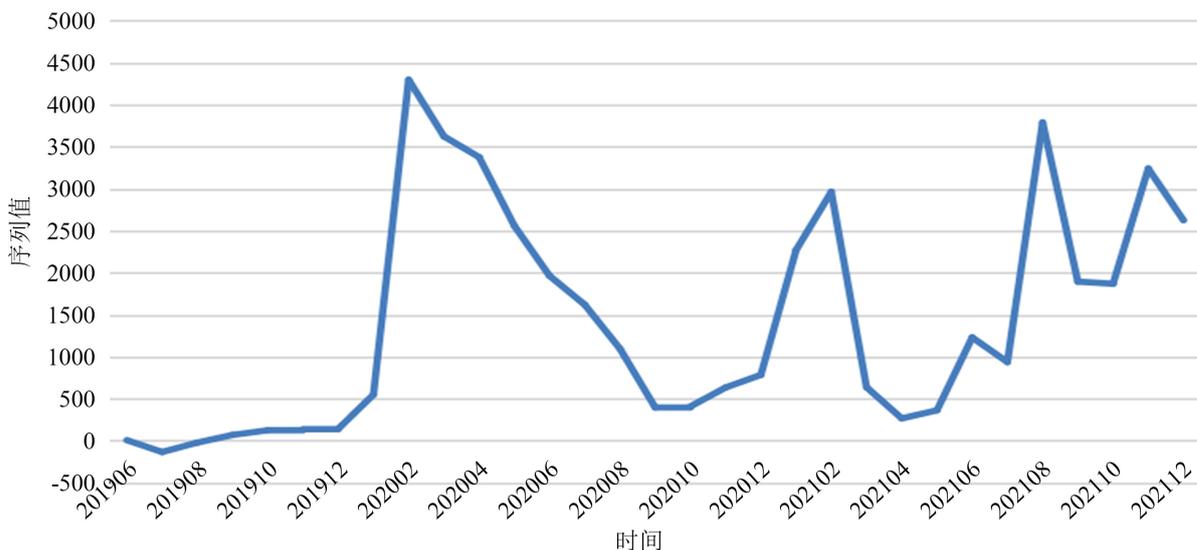


Figure 7. Trends in the impact of new crown epidemic interventions

图 7. 新冠疫情干预的影响趋势

结合图 7 新冠疫情干预序列时序图和实际情况可得出, 新冠疫情对我国民航客运量的减少造成一定的影响, 具体影响可分为以下两个部分:

1) **新冠疫情对民航客运量造成的短期影响:** 在 2020 年 2 月, 新冠疫情对民航客运量的干预影响突然大幅度上升。一方面, 随着新冠疫情在中国迅速蔓延, 政府采取了严格的紧急旅行限制措施, 包括封锁城市和关闭国际航班等, 导致航班数量急剧减少。另一方面, 由于疫情病毒的突然爆发使人们陷入了一丝恐慌, 减少人们的外出风险, 这均大致民航客运量出现了突然的暴减趋势。

2) **新冠疫情对民航客运量造成的长期影响:** 自 2020 年 3 月后, 随着国家一系列疫情防控措施的采取以及人们的积极配合, 新冠疫情对民航客运量逐渐变平缓。但值得注意的是, 由于新冠疫情的复杂性、全球性和长期性挑战, 抗争疫情是一个长期的过程, 需要持续的努力、合作和适应, 直到全球范围内实现疫情的控制和结束。

综合来说, 新冠疫情对铁路客运量的减少既体现在紧急政策的突变短期影响, 也有疫情本身的缓慢长期影响上。因此, 本文建立以下(9)式干预影响过程组合模型对干预序列进行拟合与预测。

$$Z_t = Z_{t1} + Z_{t2} = w_1 P_T + \frac{w_2 B^b}{1 - \delta_1 B} S_T, \tag{9}$$

其中,

$$S_T = \begin{cases} 0, & (t < 2020\text{年}3\text{月}) \\ 1, & (t > 2020\text{年}3\text{月}) \end{cases}, P_T = \begin{cases} 0, & (t = 2020\text{年}2\text{月}) \\ 1, & (t \neq 2020\text{年}2\text{月}) \end{cases}$$

#### 4.2.2. 干预方程的拟合

对于干预序列  $\{Z_{t1}\}$ , 根据其在 2020 年 2 月序列值可计算得出  $w_1 = 4304.01$ 。

运用最小二乘方法对干预序列  $\{Z_{t2}\}$  展开参数估计与检验, 其参数估计及检验结果如表 8 所示:

**Table 8.** Parameter estimation and significance test results of the intervention model

**表 8.** 干预模型的参数估计及显著性检验结果

参数	参数值	t 统计量的值	P 值
$w_2$	-0.9994	-2.796	0.005
$\delta_1$	0.6040	2.328	0.02

如表 8 可以看出, 干预序列  $\{Z_{t2}\}$  各参数 T 统计量对应的 P 值均小于 0.05, 通过参数的显著性检验。对拟合模型的残差进行纯随机性检验, 得出显著性检验结果如表 9 所示:

**Table 9.** Significance test results of the intervention model

**表 9.** 干预模型的显著性检验结果

延迟阶数	LB 统计量值	P 值
1	0.2004	0.6544
2	0.2284	0.8920
3	0.9308	0.8179

如表 9 可知, 该模型各阶延迟下 LB 统计量的 P 值均显著大于 0.05, 可认为该拟合模型的残差序列属于白噪声序列, 即拟合模型通过模型有效性检验。

综上所述, 该干预影响过程模型有效衡量新冠疫情对民航客运量的影响, 其一般方程为:

$$Z_t = Z_{t1} + Z_{t2} = 4304.1P_T + \frac{-0.9904B^b}{1-0.6040B} S_T, \quad (10)$$

#### 4.2.3. 干预影响趋势的预测

运用干预影响过程模型对 2021 年 11 月和 2021 年 12 月的数据展开预测, 得出的预测结果和 95% 预测区间如表 10 所示:

**Table 10.** Predicted values and prediction intervals for the impact trend of the intervention impact process model for Nov~Dec 2021

**表 10.** 干预影响过程模型对 2021 年 11 月~12 月的影响趋势的预测值和预测区间

月份	预测趋势值	真实趋势值	预测区间(置信水平为 95%)
2021 年 11 月	1833.71	3247.24	[-108.117, 3775.5485]
2021 年 12 月	1807.67	2632.50	[-499.8271, 4115.1817]

如表 10 可以得出, 2021 年 11 月与 12 月实际数据均在干预模型的预测区间之内, 说明干预模型可以有效预测未来短期新冠疫情对民航客运量的影响。值得注意的是, 由于 2021 年 11 月新冠病毒出现突变株, 导致部分地区新冠疫情再度二次爆发, 因此干预模型得出我国民航客运量的预测值与真实值仍存在一定的偏差。若想进一步提高预测精度, 需进一步将疫情突变株出现对民航客运量的影响考虑在内。

#### 4.2.4. 干预模型的预测效果评价

基于干预模型的一般方程形式:  $Y_t = X_t + Z_t$ , 根据上文构建的 ARIMA 乘法模型得出的 2021 年 11 月和 2021 年 12 月未发生疫情情况下的预测值  $Y_{21,11}$ ,  $Y_{21,12}$  以及干预影响过程模型得出相同时间段新冠过程影响程度预测值  $Z_{21,11}$ ,  $Z_{21,12}$ , 得出干预模型对新冠疫情情况下 2021 年 11 月和 12 月的预测值如表 11 所示。

为更好地对干预模型的预测效果进行比较, 本文也尝试直接运用 ARIMA 乘法模型对 2007 年 1 月~2021 年 10 月进行拟合, 并对 2021 年 11 月和 12 月的数据进行预测, 其预测值如表 11 所示。

**Table 11.** Comparison of predictive values of ARIMA multiplicative models before and after the intervention

**表 11.** 干预前后 ARIMA 乘法模型的预测值比较

	2021 年 11 月	2021 年 12 月
ARIMA 乘法模型预测值	4755.5614	5236.7814
干预 SARIMA 模型预测值	3055.437873	3522.8374
真实值	2141.9	2698

如表 11 可知, 相比于直接采用 ARIMA 乘法模型, 干预 SARIMA 模型预测值更接近于真实值, 进一步说明了干预 SARIMA 模型对疫情后民航客运量的短期预测效果表现良好。

## 5. 总结

### 5.1. 结论

本文在运用 SARIMA 模型对疫情未发生情况下未来理论民航客运量展开预测的基础上, 通过分析新冠疫情的影响形式, 构造干预影响模型对干预趋势进行拟合, 进而结合 SARIMA 模型和干预影响模型对疫情后未来民航客运量的值展开预测。通过本文研究, 可以得出以下几个结论:

1) ARIMA(1,1,1) × (0,1,1)<sub>12</sub> 乘法模型对未发生疫情情况下民航客运量的拟合和预测效果表现良好。通过该模型, 可得出在未来不发生疫情的情况下, 民航客运量仍会呈现逐年持续上涨的趋势, 且具有夏季偏多, 冬季偏少的季节效应;

2) 干预模型精确、科学地通过定量方式衡量新冠疫情对民航客运量的影响机制。通过干预模型, 可以定量得出新冠疫情对民航客运量的影响程度, 进而结合民航客运量在未发生疫情下的理论值对疫情后未来的民航客运量展开预测;

3) 相比于单独使用 SARIMA 模型, 干预 SARIMA 模型对疫情后短期民航客运量的预测表现更加精确。除了新冠疫情这一干预因素, 民航客运量可能还受到经济状况、政策变化、航空公司策略等其他外部因素的影响。后续可以将这些因素考虑进预测模型中, 以更全面地分析和预测民航客运量的变化。

### 5.2. 不足及改进方向

尽管干预 SARIMA 模型在对疫情后民航客运量的预测表现良好, 但仍有很多不足之处, 具体有以下

方面:

1) **干预 SARIMA 模型对非线性趋势的提取程度较差**, 导致其得出的预测值仍存在一定的偏差。由干预影响方程和 SARIMA 模型的方程形式可以看出, 干预 SARIMA 模型主要提取序列的线性趋势。若序列存在非线性特征, 则干预 SARIMA 模型的预测结果可能会有一定的偏差。针对此问题, 考虑到 LSTM、RF 等机器学习模型可以有效提取序列的非线性特征, 可以运用串联组合预测方法, 在运用干预 SARIMA 模型得出疫情发生后各时序点预测值的基础上, 通过机器学习模型对其残差进行拟合和预测, 进而在不失去模型可解释度的同时进一步提高干预模型的预测精度。

2) **该干预 SARIMA 模型忽视了干预因素与 SARIMA 模型的内部关系**。该类模型将民航客运量分解为 SARIMA 模型拟合值与干预因素, 一定程度上忽视了干预因素和 SARIMA 模型内部可能存在的关系, 进而使模型预测值出现一定的偏差。针对此问题, 可以从干预 SARIMA 模型方程引入交互项的角度对模型进行改进。通过引入交互项, 分析干预因子与 SARIMA 模型结果的相互影响程度, 以捕捉干预因素与 SARIMA 模型的动态联动效应, 进而使模型预测变得更加精确。

## 参考文献

- [1] 陈文华, 边浩毅, 王怡民, 等. 基于 ARIMA 模型的公路客运量预测[J]. 重庆工商大学学报(西部论坛), 2006(5): 95-97.
- [2] 郝军章, 崔玉杰, 韩江雪. 基于 SARIMA 模型在我国铁路客运量中的预测[J]. 数学的实践与认识, 2015, 45(18): 95-104.
- [3] 赵芳卉, 陈琳, 李冬青. 基于 SARIMA-GARCH 模型的民航客运量研究[J]. 统计与管理, 2021, 36(2): 17-24.  
<https://doi.org/10.16722/j.issn.1674-537x.2021.02.003>
- [4] 李思如. SARIMA 模型与 LSTM 神经网络的航空客运量的预测比较[J]. 河南科技, 2021, 40(36): 18-21.
- [5] 郭猛. 基于组合模型的北京民航客运量预测研究[D]: [硕士学位论文]. 石家庄: 河北经贸大学, 2020.  
<https://doi.org/10.27106/d.cnki.ghbj.2020.000510>
- [6] 陈宗楠. 基于组合预测研究新冠疫情对民航客运量的影响[J]. 电脑知识与技术, 2023, 19(3): 16-18.  
<https://doi.org/10.14004/j.cnki.ckt.2023.0148>
- [7] 刘昶, 吴君华, 张凤娇, 等. 基于 ARIMA-BP 组合模型的城市公交客运量预测[J]. 智能计算机与应用, 2022, 12(7): 80-83+89.
- [8] 何桂焯. 新冠疫情对黄金期货价格的干预影响研究[J]. 中国市场, 2021(34): 44-46.  
<https://doi.org/10.13939/j.cnki.zgsc.2021.34.044>
- [9] 汪志红. 突发事件对我国铁路客运量的干预特征研究[J]. 铁道运输与经济, 2022, 44(1): 44-51.  
<https://doi.org/10.16668/j.cnki.issn.1003-1421.2022.01.07>
- [10] 王鑫, 肖枝洪. 基于干预模型与 BP 神经网络集成的 GDP 预测[J]. 统计与决策, 2012(20): 141-144.  
<https://doi.org/10.13546/j.cnki.tjyj.2012.20.039>
- [11] 冯文权, 蔡基栋. 干预分析模型及其应用[J]. 武汉大学学报(自然科学版), 1994(6): 29-35.
- [12] 高旭, 李兴东. 干预分析模型在时间序列中的应用[J]. 科学技术创新, 2021(29): 26-28.