

新冠疫情对海南旅游业影响的统计测度研究

海南调查总队 郭涛 王宏超 王伟苏

摘要

今年年初爆发的新冠疫情对海南乃至全国、全世界的旅游业都造成了巨大的影响。本文利用乘季节自回归滑动平均模型 SARIMA, 拟在无疫情影响假设条件下对在海南过夜国内游客人数进行短期预测, 将其与疫情影响下的实际过夜人数进行对比, 探究旅游业在突发事件及政府相应管控措施下的影响程度。通过构建回归模型, 利用酒店、旅行社业务活动预期指数和商务活动指数对国内过夜游客人数进行回归拟合, 结合海南实际情况, 对海南旅游业未来恢复情况进行预测。

关 键 词: 新冠疫情 旅游业 SARIMA 模型 回归预测

目录

一、国内外研究综述.....	5
二、新冠疫情对海南省旅游业影响的预测模型.....	6
(一) 构建的基本思路.....	6
(二) 指标选择.....	6
(三) 数据时期选择.....	7
(四) 数据的预处理.....	7
(五) 平稳性检验和纯随机性检验.....	8
(六) 模型参数识别.....	10
(七) 参数估计.....	11
(八) 残差检验.....	12
(九) 样本外预测.....	14
(十) 模型的改进.....	15
三、酒店、旅行社业务活动预期指数与国内过夜游客人数的回归分析.....	20
(一) 相关性检验.....	21
(二) 平稳性检验.....	22
(三) 回归分析.....	23
四、主要结论和存在问题.....	26
(一) 主要结论.....	26
(二) 研究中存在的问题.....	27

表目录

表格 2-1: 四个初选模型的参数估计和模型检验结果	11
表格 2-2: 模型一参数估计结果	12
表格 2-3: 模型二参数估计结果	12
表格 2-4: 使用模型二的预测结果	14
表格 2-5: 改进模型参数估计结果	17
表格 3-1: 六次相关性检验结果	22
表格 3-2: 时间序列 $\ln Y$ 、 $\ln X_1$ 、 $\ln X_2$ 的 ADF 检验结果表	23
表格 3-3: 时间序列 $\ln X_3$ 、 $\ln X_4$ 的 ADF 检验结果表	24
表格 3-4: 向前回归法模型拟合及诊断对比	24
表格 3-5: 单变量回归模型拟合及诊断对比	25
表格 3-6: 回归预测结果对比	27

图目录

图 2-1 2014 年至 2019 年海南省接待国内过夜游客人数时序图	8
图 2-2 差分化处理后的海南省接待国内过夜游客时序图	9
图 2-3 差分化处理后的自相关和偏自相关图	10
图 2-4 差分化处理后的单位根检验结果	10
图 2-5 模型一残差检验结果	13
图 2-6 模型二残差检验结果	13
图 2-7 模型二拟合原序列情况	14
图 2-8 提取季节指数信息后的新序列 Z 时序图	16
图 2-9 差分化处理后的 Z 序列时序图	16
图 2-10 差分化处理后的 Z 序列自相关和偏自相关图	17
图 2-11 海南省接待国内过夜游客人数实际值和模型拟合值	18
图 2-12 最终的季节因子影响	19
图 2-13 最终的循环因子影响 2-5 模型一残差检验结果	20
图 2-14 最终的不规则因子影响	20
图 2-15 最终的季节调整后序列	21
图 3-1 酒店、旅行社业务活动预期指数和商务活动指数与国内过夜游客人数 变化图	22
图 3-2 模型 2 拟合残差图	25
图 3-3 模型 4 拟合残差图	26
图 3-4 模型 4 拟合图	27

2020 年 1 月,在我国武汉出现新型冠状病毒肺炎(COVID-19)病例,到 2 月份全国范围大部分省份城市均有突发新冠肺炎疫情,乃至全世界也有一百多个国家和地区被世界卫生组织报道有新冠肺炎疫情,全球旅游业随之遭受到巨大的冲击。海南虽然受疫情影响不大,但是旅游业及其相关服务业作为海南的支柱产业,难免受到很大影响,比较直观的就是赴海南的国内游客人数骤减。而精准的游客人数预测对旅游业的发展十分重要,因为它们可以作为海南地方政府对旅游业实施决策的工具,旨在促进经济发展、福利和就业。本研究拟利用乘季节自回归滑动平均模型(SARIMA)来研究新冠肺炎疫情对海南旅游业的影响,拟在无疫情影响假设条件下对在海南过夜国内游客人数进行短期预测,将其与疫情影响下的实际过夜人数进行对比,结合上年同期人均游客花费来计算新冠疫情给海南旅游业带来的短期损失;并以酒店和旅行社等跟旅游业密切相关企业家的商务活动指数及业务活动预期指数等先行指标为变量,构建模型来对新冠疫情给海南旅游业的影响情况进行中远期预测。

一、国内外研究综述

国内外关于突发事件对旅游业影响的研究主要分为以下几类:一是时间序列预测方法,Carey 等[1]通过对 10 种不同的时间序列模型分析了香港旅游行业受金融危机的影响情况,通过验证发现 SARIMA 和 MARIMA 模型更为精准;朱迎波等[2]利用双变量 ARIMA 模型研究不同恢复期情况下 SARS 对中国入境旅游人数的影响。二是向量自回归和脉冲响应模型,Zheng Cao 等[3]利用全球向量自回归模型研究了 24 个主要国家旅游需求的协同效应,并借助模拟脉冲响应函数评估了对中国经济的冲击,结果与旅游政策制定者和监测宏观经济环境的跨国公司高度相关。三是采用灰色预测和灰色关联方法,王慧蕾[4]将灰色模型预测方法 GM(1,1) 和灰色残差修正相结合,对广东旅游业收入进行预测,得出的结论比灰色预测更加准确。四是对多种旅游业影响预测模型的效果进行评估,Hosseini Hassani[5]等评估了使用几个参数和非参数模型预测部分欧洲国家旅游需求的效果,发现没有哪个单一模型能够为任何一个国家同时提供短期、中期和长期的最优预测。五是利用网上抓取数据构建综合指数形式对旅游需求进行研究,Xin Li[6]等研究人员改进了利用搜索引擎查询到的游客数量和客房开房率等在线数据对旅游目的地需求进行预测的方法,构建了一个能准确预测中国旅游需求的

框架,并采用广义动态因子模型建立一个综合的搜索指数,最后验证了其有效性。

此外,还有一些学者采用因果关系模型、空间模型和德尔斐预测法等对传统旅游指标数据进行实证研究。总体来看,由于旅游产业影响因素众多,其发展也是复杂多变,很难从宏观大视角去精准研究突发事件对旅游产业发展的影响,大多数学者都采用个案研究的方式,对某一地区的旅游事件进行具体分析,再提出对策建议。而针对海南旅游产业受突发事件影响方面的研究大多集中在定性研究,定量的研究不多。

二、新冠疫情对海南省旅游业影响的预测模型

(一) 构建的基本思路

2020 年新冠疫情在全球迅速蔓延,对各行各业尤其是旅游业产生了巨大冲击。一方面航班客运数量减少、交通中断、出团禁令等措施导致旅游出行受阻,严格的防控措施也给旅游产品提供者的复工复产设下门槛;另一方面政府部门针对新冠疫情发出的旅行警告,必然会对游客心理产生影响,从而取消或推迟外出计划。

旅游业是海南省的主要支柱产业,受此次疫情的负面影响冲击更是成倍放大,为了定量分析新冠疫情对海南省旅游业产生的影响,本部分拟通过 2019 年以前的历史数据建立随机时间序列模型,根据模型结果对 2020 年旅游业情况进行预测,并通过与 2020 年旅游业实际情况进行对比,从而达到定量分析测算影响程度的目的。

本部分时间序列建模部分使用 EViews6 软件实现。

(二) 指标选择

考虑调查指标的科学性和可获取性,本文初步选取了接待过夜游客人数、接待国内过夜游客人数、旅游总收入、国内旅游收入等指标的月度数据来度量旅游业生产经营状况。最终选择海南省接待国内过夜游客人数指标作为旅游业的度量指标主要基于以下考虑:(1)接待游客总人数包括接待国内过夜游客人数和接待入境过夜游客人数,2019 年全年海南省接待国外过夜游客人数仅占总接待过夜游客人数的 1.73%,占比较小,此外截至 2020 年 8 月,国内疫情已基本得到控制,国外疫情形势仍然严峻,后期存在较大不确定性,可以认为国内旅游市场已在逐步恢复,而境外来华旅游暂无明显改善,考虑到随机时间序列模型对短期预

测有效的特点，本文优先选择以接待国内游客人数作为旅游业的度量指标。(2) 旅游人数指标和旅游收入指标都能较好地作为旅游业的度量指标，在旅游过程中游客支付的所有旅游费用都被统计为旅游收入，相对于旅游人数统计更容易存在调查误差，在分析旅游收入变动时价格水平变动的影响也不可忽视，因此本文优先选择统计误差较小、影响因素相对较少的旅游人数作为旅游业的度量指标。

接待国内过夜游客人数，指接待的以旅游为目的，连续停留时间不超过一年的国内过夜游客数量，休闲度假、购物观光、探亲参会、或从事文体宗教活动都涵盖在统计范围内。

(三) 数据时期选择

本文拟选择 2014 年 1 月至 2019 年 12 月期间各月份海南省接待国内过夜游客人数指标作为旅游业的度量标准，主要基于以下考虑：(1) 2010 年《国务院关于推进海南国际旅游岛建设发展的若干意见》正式发布，海南正式拉开以旅游业为建设主导的帷幕，经过三年的政策支持和引导，来琼旅游游客人数由爆发式增长转为稳定增长，以 2014 年 1 月为分析起点更利于去繁从简，提高模型拟合和解读便利，也留有较充足的数据量以开展时间序列分析。(2) 考虑到接待国内过夜游客人数指标是由海南省旅游和文化广电体育厅发布的部门数据，在不同年份之间指标统计口径、统计方法、涵盖范围等可能存在不一致，一定程度上影响数据可比性，故不宜使用更长调查期的数据进行分析。

(四) 数据的预处理

考虑到接待国内过夜游客人数和假期存在较大相关性，易受到“五一”小长假、国庆“黄金周”、春节假期等确定性异常值影响，在数据使用前应予优先考虑。与“五一”小长假、国庆“黄金周”日期相对固定，可通过季节差分消除影响不同，春节假期由于时间不确定，无法通过季节差分直接消除，故首先分析春节因素是否对接待国内过夜游客人数产生了显著影响。通过建立简单回归模型，将春节影响因子作为自变量，海南省接待国内过夜游客人数作为因变量。若回归模型显著成立，则说明春节因素对序列有显著稳定的影响，可使用回归方程提取这类特殊影响并调整接待国内过夜游客人数序列；若回归模型不显著，则不需要对数据进行特殊预处理。

构建春节影响因子 SP，假设每年春节前 10 天和后 10 天为春节假期效应的

有效影响期，通过计算这段有效影响期落入每个月份的天数，再除以该月份的总天数，得到春节影响因子 SP 序列。例如：

2019 年	1 月	2 月	3 月至 12 月
SP	6/31	14/28	0

通过与接待国内过夜游客人数序列 Y 建立简单回归模型，得到：

$$X=437.49+38.25SP$$

选择显著性水平为 0.05，模型拟合结果显示参数显著性检验 P 值等于 0.634，远远大于 0.05 的显著性水平，该方程不能显著成立。故以春节前后 10 天为限构建的春节影响因子 SP 不是接待国内过夜游客人数序列的显著性影响因素，不再专门进行数据处理。

（五）平稳性检验和纯随机性检验

通过绘制海南省接待国内过夜游客人数时间序列图，可以看到该序列是一个非平稳时间序列，表现为序列有长期上升趋势，且呈现出每年 10 月至次年 2 月为峰值的季节性周期变动，不满足建立随机时间序列模型的基本条件。

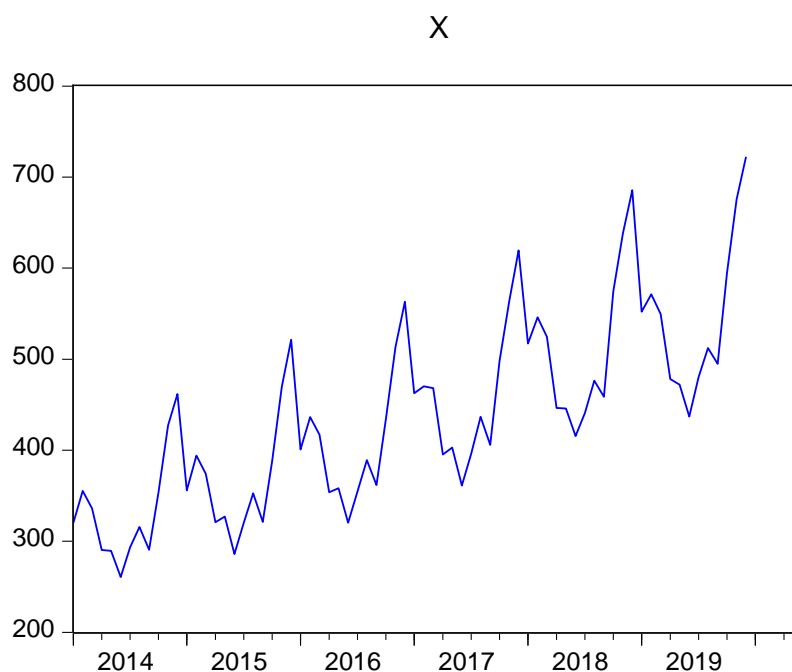


图 2-1 2014 年至 2019 年海南省接待国内过夜游客人数时序图

为消除序列的增长趋势和季节波动，对原序列 X 进行步长为 12 的一阶季节差分处理，形成新的序列 X_2 。时序图 2-2 显示， X 序列在经过差分处理后，长期趋势和周期性波动已基本消除，进一步分析 X_2 序列的自相关图，发现在时滞 $k=3$ 后， X_2 序列的自相关系数均落于 2 倍标准差范围以内并在零值附近波动，同时根据图 2-4 单位根检验结果，给出的 DF 值为 -3.708 小于显著性水平是 1% 的临界值 -3.546 ，因此我们拒绝零假设，认为 X_2 序列通过了平稳性检验。如图 2-3 所示，差分处理后的序列 X_2 时滞 $k=12$ 期的 Q 统计量 P 值小于 0.05，拒绝纯随机序列的原假设，可以认为 X_2 序列通过了纯随机性检验。综上所述， X_2 序列满足使用随机时间序列模型 $ARIMA(p, d, q)$ (P, D, Q)^s 进行建模分析的基本条件。

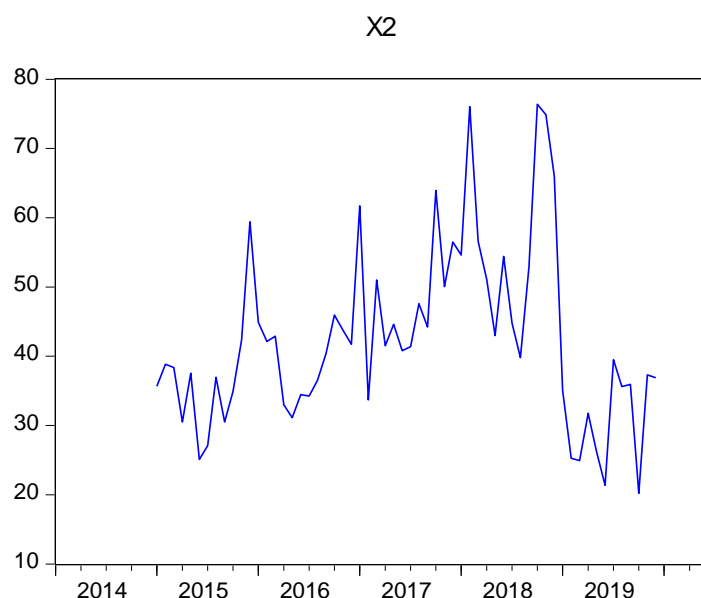


图 2-2 差分处理后的海南省接待国内过夜游客时序图

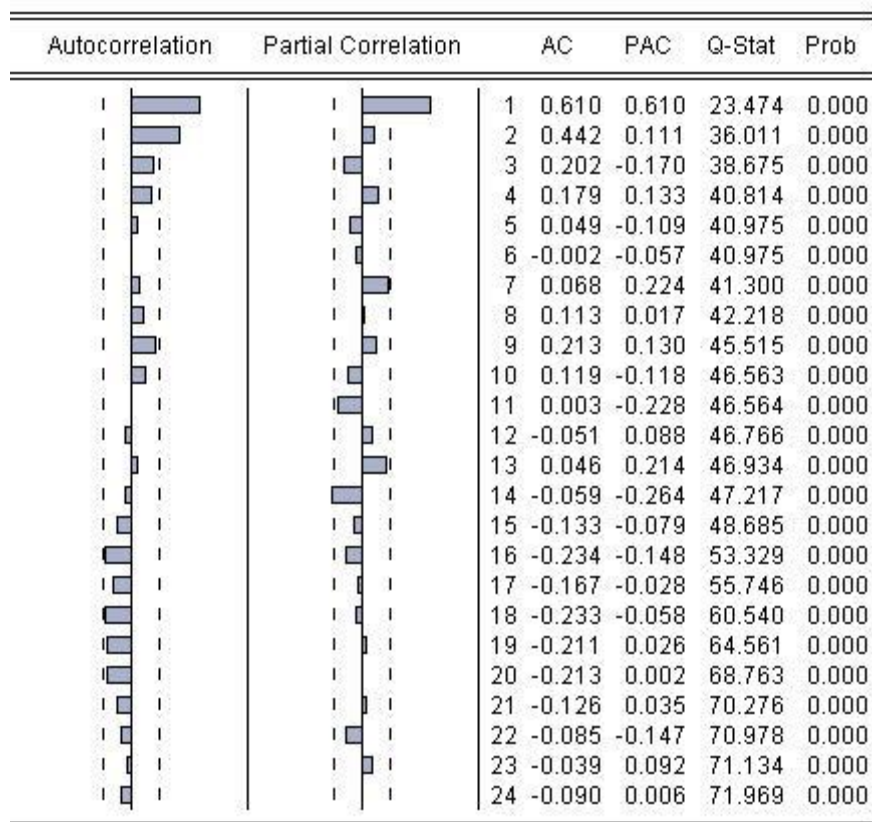


图 2-3 差分化处理后的自相关和偏自相关图

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-3.708207	0.0064
Test critical values:		
1% level	-3.546099	
5% level	-2.911730	
10% level	-2.593551	

*MacKinnon (1996) one-sided p-values.

图 2-4 差分化处理后的单位根检验结果

在建模过程中也曾尝试使用逐期一阶差分加步长为 12 的一阶季节差分对原序列进行处理，但差分化后的序列短期自相关并不明显，怀疑可能是过度差分导致信息大量丢失，两次差分后的序列成为了白噪声序列，无法继续开展后续建模，多次尝试后最终选择了仅对数据进行步长为 12 的一阶季节差分处理。

（六）模型参数识别

如前文所述，根据差分化处理结果，初步建立的随机时间序列模型 $ARIMA(p, d, q) (P, D, Q)^S$ 中 $d=0, D=1, S=12$ 。进一步分析图 2-3 中的自相关和偏自相关系数，自相关函数在 k 小于 12 时均表现为随时滞 k 增加，呈衰减的正弦波趋向于 0，具有拖尾性，偏自相关函数表现为 1 阶截尾，但在 $k=7$ 和 $k=11$ 时有

增大趋势,故优先考虑建立自回归移动平均混合模型。本模型考虑 p 选择 2 或 3, q 选择 2 或 3。此外还需补充季节自回归和季节移动平均的阶数 P 和 Q , 观察 $k=12, 24, 36$ 时自相关和偏自相关函数情况, 发现该期自相关函数和偏自相关函数均已落在两倍标准差范围内, 且逐步衰减至零附近波动, 此时可以考虑 P 选择 0, Q 考虑选择 0。

初选模型包括: $ARIMA(2, 0, 2)(0, 1, 0)^{12}$, $ARIMA(2, 0, 3)(0, 1, 0)^{12}$, $ARIMA(3, 0, 2)(0, 1, 0)^{12}$, $ARIMA(3, 0, 3)(0, 1, 0)^{12}$ 。

(七) 参数估计

采用最大似然估计, 对四个初选模型分别进行参数估计, 结果如表 2-1 所示。剔除各模型中未通过系数显著性检验的变量, 最后通过各项检验, 可作为接待国内过夜游客人数的时间序列度量模型见表 2-2 和表 2-3。虽然 AIC 准则、SC 准则和 HQC 准则三项定阶的最小信息准则都认为表 2-2 的结果更优, 但仍应考察残差情况再进一步决定。

表 2-1 四个初选模型的参数估计和模型检验结果

	变量	系数	T-统计量	P 值
$ARIMA(2, 0, 2)(0, 1, 0)^{12}$	C	43.037	17.401	0.000
	AR(1)	-0.496	-3.568	0.001
	AR(2)	-0.123	-0.948	0.348
	MA(1)	1.228	36.767	0.000
	MA(2)	0.966	47.697	0.000
$ARIMA(2, 0, 3)(0, 1, 0)^{12}$	C	44.295	10.490	0.000
	AR(1)	0.384	1.604	0.115
	AR(2)	0.437	2.986	0.004
	MA(1)	0.369	1.639	0.107
	MA(2)	-0.121	-0.456	0.651
	MA(3)	-0.766	-3.743	0.001
$ARIMA(3, 0, 2)(0, 1, 0)^{12}$	C	43.580	13.555	0.000
	AR(1)	-0.518	-3.562	0.001
	AR(2)	0.039	0.271	0.787
	AR(3)	0.238	1.879	0.066
	MA(1)	1.265	25.896	0.000
	MA(2)	0.964	26.323	0.000

ARIMA (3, 0, 3) (0, 1, 0) ¹²	C	43.229	11.861	0.000
	AR(1)	-0.324	-0.583	0.563
	AR(3)	0.273	1.976	0.054
	AR(2)	0.151	0.540	0.592
	MA(1)	1.063	1.819	0.075
	MA(2)	0.715	1.016	0.314
	MA(3)	-0.181	-0.344	0.732

表 2-2 模型一参数估计结果

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	42.35512	2.249881	18.82549	0.0000
AR(1)	-0.527752	0.106838	-4.939739	0.0000
MA(1)	1.301562	0.050773	25.63513	0.0000
MA(2)	0.994994	0.038925	25.56160	0.0000
R-squared	0.479212	Mean dependent var		42.54932
Adjusted R-squared	0.450805	S.D. dependent var		12.89114
S.E. of regression	9.553322	Akaike info criterion		7.417044
Sum squared resid	5019.627	Schwarz criterion		7.557894
Log likelihood	-214.8028	Hannan-Quinn criter.		7.472026
F-statistic	16.86972	Durbin-Watson stat		2.043648
Prob(F-statistic)	0.000000			
Inverted AR Roots	-.53			
Inverted MA Roots	-.65-.76i	-.65+.76i		

表 2-3 模型二参数估计结果

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	43.15812	3.292141	13.10944	0.0000
AR(1)	0.725377	0.097780	7.418454	0.0000
MA(3)	-0.330509	0.136076	-2.428861	0.0184
R-squared	0.411361	Mean dependent var		42.54932
Adjusted R-squared	0.390338	S.D. dependent var		12.89114
S.E. of regression	10.06551	Akaike info criterion		7.505616
Sum squared resid	5673.613	Schwarz criterion		7.611253
Log likelihood	-218.4157	Hannan-Quinn criter.		7.546853
F-statistic	19.56732	Durbin-Watson stat		2.221160
Prob(F-statistic)	0.000000			
Inverted AR Roots	.73			
Inverted MA Roots	.69	-.35+.60i	-.35-.60i	

（八）残差检验

对所选两个模型进行残差序列的 X^2 检验，结果如图 2-5 和图 2-6 所示，模型一有部分 P 值小于 0.05 的显著性水平，可以认为残差不是一个白噪声序列，

模型尚未能完全提取序列中所含信息。模型二右侧一列 P 值均大于 0.05，说明所有 Q 值都小于检验水平为 0.05 的 χ^2 分布临界值（位于临界值左侧），即可以认为模型二的随机误差项是白噪声序列，是一个更合理的、可用于最终预测的模型。使用模型二拟合原序列情况见图 2-7。



Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 -0.024	-0.024	0.0345	
		2 0.084	0.083	0.4753	
		3 0.286	0.292	5.7256	
		4 -0.072	-0.066	6.0682	0.014
		5 0.069	0.015	6.3831	0.041
		6 -0.029	-0.107	6.4395	0.092
		7 0.002	0.035	6.4399	0.169
		8 0.083	0.074	6.9281	0.226
		9 0.043	0.099	7.0606	0.315
		10 0.137	0.122	8.4421	0.295

图 2-5 模型一残差检验结果

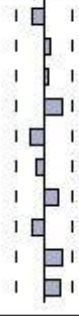
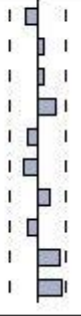
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 -0.112	-0.112	0.7779	
		2 0.063	0.051	1.0306	
		3 0.037	0.050	1.1192	0.290
		4 0.160	0.169	2.7900	0.248
		5 -0.117	-0.089	3.6995	0.296
		6 -0.075	-0.125	4.0854	0.395
		7 0.125	0.106	5.1698	0.396
		8 -0.111	-0.094	6.0351	0.419
		9 0.171	0.197	8.1460	0.320
		10 0.140	0.216	9.5930	0.295

图 2-6 模型二残差检验结果

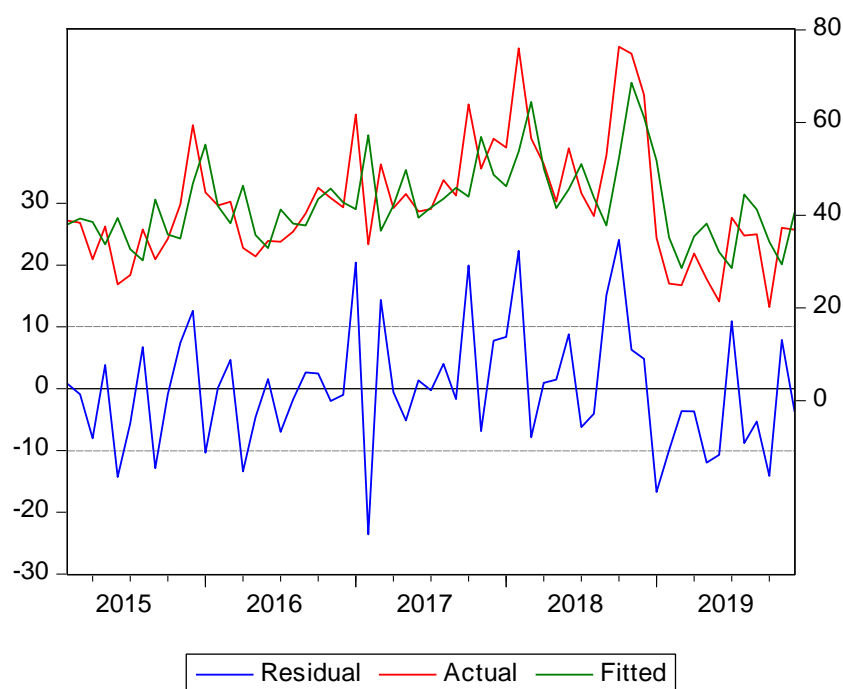


图 2-7 模型二拟合原序列情况

（九）样本外预测

选择模型二对 2020 年 1 月至 6 月海南省接待国内过夜游客人数进行简单预测，估计结果分别为 595.56 万人、611.97 万人、592.16 万人、520.89 万人、514.88 万人和 479.80 万人。通过与实际统计数据对比，可以分析出海南省接待国内过夜游客人数受疫情影响的程度，结果表明，新冠疫情从 1 月开始对海南省旅游业产生影响，但由于疫情是 1 月下旬开始爆发并受到重视，故对该月数据产生的影响有限；2 月、3 月正是加强疫情防控、限制人员流动的关键月份，旅游业受到的直接冲击最为严重，接待国内过夜游客数量不足预测值的三成；随着疫情缓解以及现存患者清零，疫情对旅游业的影响开始消退，海南省旅游业从 4 月开始进入恢复期，五一长假、跨省旅游地逐步放开等因素都对旅游人数有正向促进，截至 6 月海南省接待国内过夜游客人数已恢复至应有水平的七成左右。

表 2-4 使用模型二的预测结果

	预测值	实际值	差值	实际值/预测值
2020 年 1 月	595.56	355.12	240.44	59.63%
2020 年 2 月	611.97	53.48	558.49	8.74%
2020 年 3 月	592.16	146.99	445.17	24.82%
2020 年 4 月	520.89	226.15	294.74	43.42%
2020 年 5 月	514.88	314.02	200.86	60.99%
2020 年 6 月	479.80	340.49	139.31	70.97%

(十) 模型的改进

考虑到直接进行步长为 12 的一阶季节差分处理可能导致较多信息损失,并减少了可用于推断参数的样本量,易导致后期结论产生偏差,现考虑首先将 X 序列中的季节影响因素用非差分方法提取,再使用时间序列分析方法对新序列进行分析。

1. 数据预处理

通过对图 2-1 的分析可知,海南省接待国内过夜游客人数在 11 月、12 月位于全年峰值且较为明显,在 1 月至 3 月也处于较高水平,其他月份的水平均较低,呈现出与旅游淡旺季高度相关、逐年同月有相同变化方向的特点,满足使用季节指数提取季节趋势信息的条件。通过计算 2014 年至 2019 年各月份接待国内过夜游客人数均值与所有月份接待国内过夜游客人数均值的比值,得到各月份季节指数 M_i ($i=1, 2, \dots, 12$) 如下:

	1 月	2 月	3 月	4 月	5 月	6 月
M_i	0.993	1.056	1.016	0.870	0.874	0.792
	7 月	8 月	9 月	10 月	11 月	12 月
M_i	0.870	0.945	0.888	1.082	1.252	1.360

根据计算结果,对原有序列 X 进行剔除季节性趋势处理,将 2014 年至 2019 年各月份国内接待过夜游客人数除以对应的季节指数 M_i 得到新的序列 Z。图 2-8 显示,新序列 Z 季节性趋势已大部分消除,长期趋势仍较为明显。

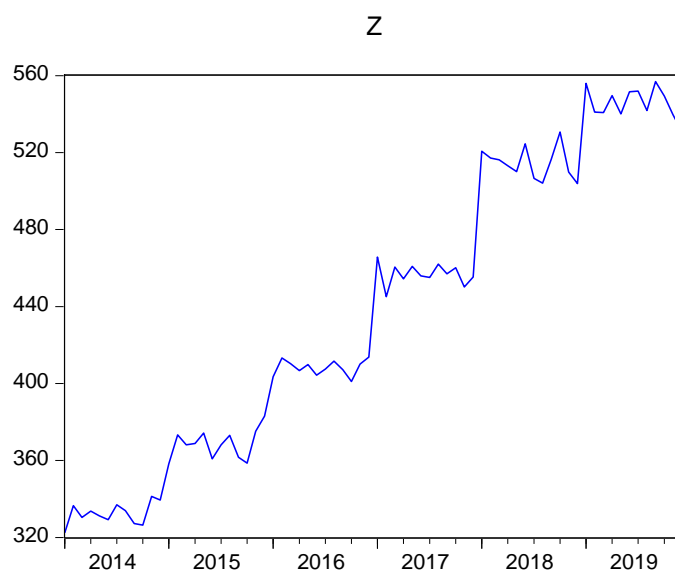


图 2-8 提取季节指数信息后的新序列 Z 时序图

2. 平稳性检验和纯随机性检验

对序列 Z 进行一阶逐期差分后形成新序列 Z_1 , 检验 Z_1 序列的平稳性和纯随机性。时序图 2-9 显示, 差分处理后 Z_1 序列原有趋势已基本消除, 在时滞 $k=1$ 后, 序列的自相关系数迅速趋于 0, 且都落于 2 倍标准差范围内并在零值附近波动, 可以认为 Z_1 序列通过平稳性检验。如图 2-10 所示, 差分处理后的序列 Z_1 时滞 $k=12$ 期的 Q 统计量 P 值小于 0.05, 拒绝纯随机序列的原假设, 可以认为该序列通过了纯随机性检验。

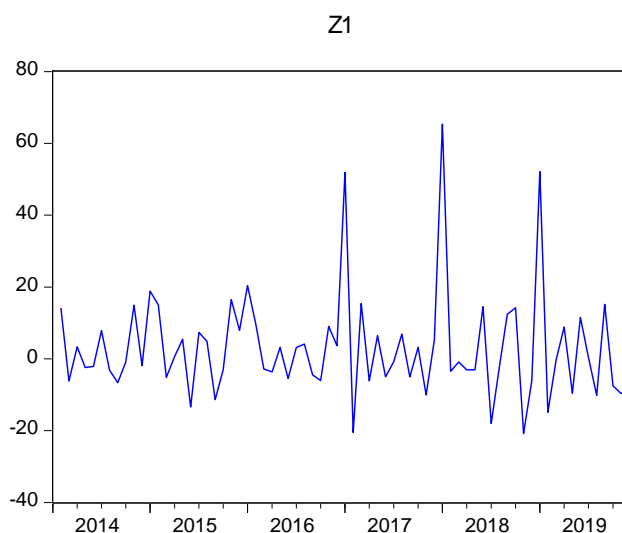


图 2-9 差分处理后的 Z 序列时序图

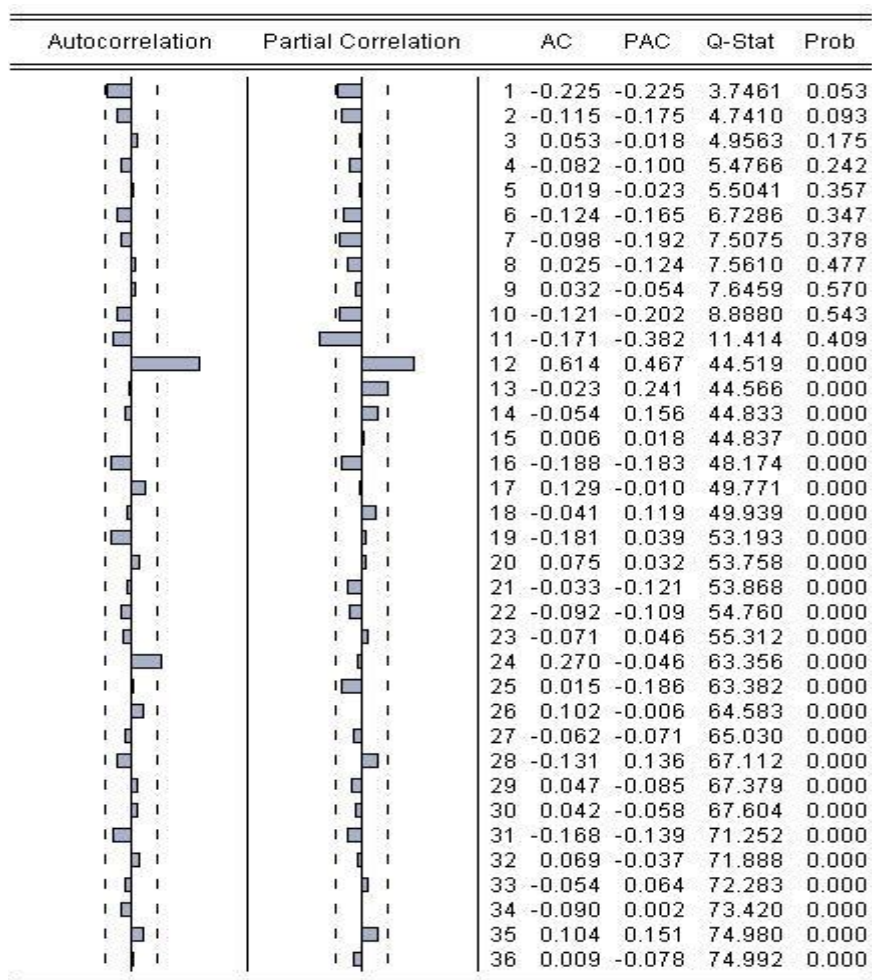


图 2-10 差分化处理后的 Z 序列自相关和偏自相关图

3. 模型参数识别和参数估计

观察 Z1 序列的自相关图和偏自相关图，自相关函数在 k 小于 12 时均呈衰减的正弦波趋向于 0，具有拖尾性，偏自相关函数表现为在 k=1 和 11 时未落入二倍标准差范围之内，故本模型考虑 p 选择 1 或 2，q 选择 0。此外观察 k=12, 24, 36 时自相关和偏自相关函数情况，发现序列的自相关系数和偏自相关系数随时滞按季节周期的增加而衰减，序列可以用季节自回归模型描述，P 考虑选择 1，Q 考虑选择 1 或 2。根据参数估计结果，最终选择了 ARIMA(2, 1, 0) (1, 0, 1)¹² 模型。改进模型拟合结果见表 2-5。

表 2-5 改进模型参数估计结果

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0.525341	0.128550	-4.086667	0.0001
AR(2)	-0.403867	0.127022	-3.179516	0.0025
SAR(12)	0.932577	0.067740	13.76710	0.0000
MA(12)	-0.877139	0.038070	-23.04021	0.0000
R-squared	0.669130	Mean dependent var		2.856159
Adjusted R-squared	0.650401	S.D. dependent var		15.90937
S.E. of regression	9.406713	Akaike info criterion		7.388316
Sum squared resid	4689.771	Schwarz criterion		7.531688
Log likelihood	-206.5670	Hannan-Quinn criter.		7.444035
Durbin-Watson stat	2.060397			
Inverted AR Roots	.99	.86-.50i	.86+.50i	.50-.86i
	.50+.86i	.00+.99i	-.00-.99i	-.26+.58i
	-.26-.58i	-.50+.86i	-.50-.86i	-.86+.50i
	-.86-.50i	-.99		
Inverted MA Roots	.99	.86+.49i	.86-.49i	.49-.86i
	.49+.86i	.00-.99i	-.00+.99i	-.49-.86i
	-.49+.86i	-.86-.49i	-.86+.49i	-.99

4. 模型预测

使用改进模型对 2020 年 1 月至 6 月海南省接待国内过夜游客人数进行简单预测，估计结果分别为 539.91 万人、569.94 万人、548.53 万人、471.91 万人、471.70 万人和 429.25 万人。通过与实际统计数据对比，可以分析 1 月至 6 月出海南省接待国内过夜游客人数受疫情影响的程度导致实际人数仅达到 65.8%、9.4%、26.8%、47.9%、66.6%、79.3%，与上一模型预测水平大致相同。



图 2-11 海南省接待国内过夜游客人数实际值和模型拟合值

5. 序列确定性因素分解

以模型二为例，考察长期趋势、季节变动、周期波动和不规则变动等四个因素对 X 序列产生的影响。根据时序图 2-1 所示，接待国内过夜游客人数每年的振幅随着水平的提高而递增，此处考虑使用 X-11 乘法模型对其进行确定性因素分解，结果如下：

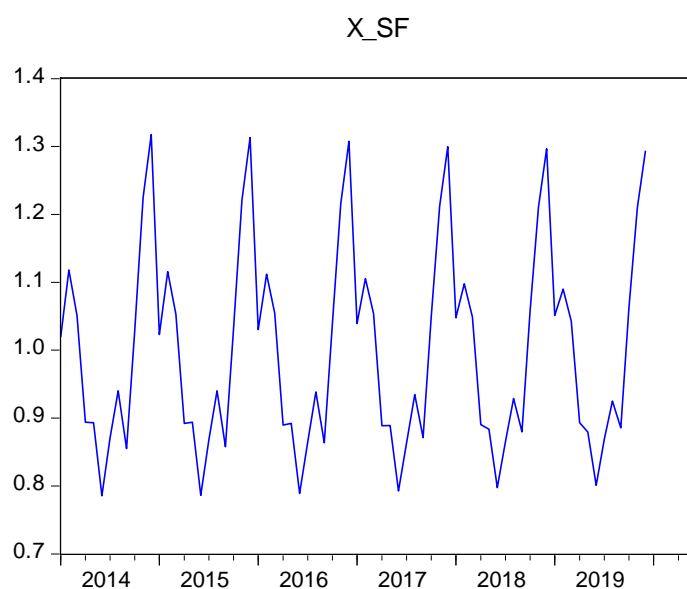


图 2-12 最终的季节因子影响

从季节指数图可以看到，海南省接待国内过夜游客人数具有显著的季节变动特征：每年 10 月至次年 2 月是旺季，接待过夜游客人数增加较为明显，其他时间则为淡季。

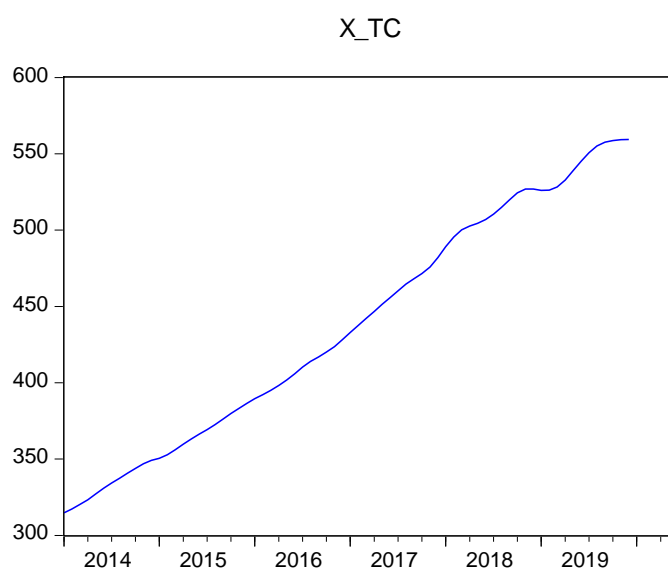


图 2-13 最终的循环因子影响

从图 2-13 可以看出，2014 年至 2019 年海南省接待国内过夜游客总量持续增加，发展势头良好，但 2018 年 4 月出台的海南省房地产“全域限购”调控政策对“购房+旅游”类型游客产生影响，减缓了来琼过夜游客的增长势头。

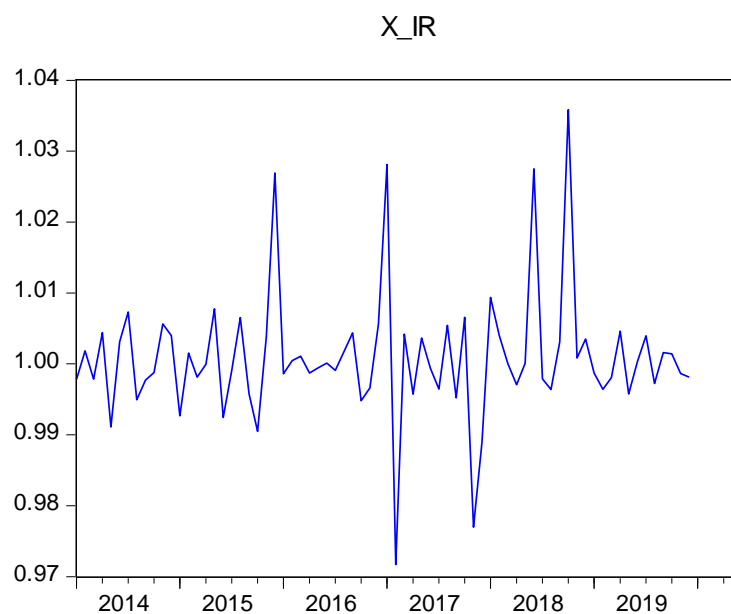


图 2-14 最终的不规则因子影响

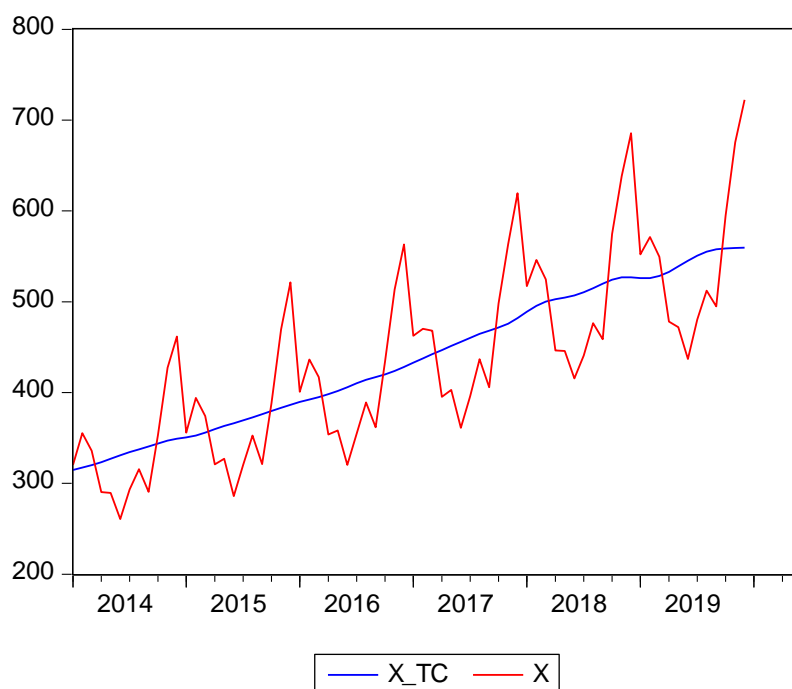


图 2-15 最终的季节调整后序列

三、酒店、旅行社业务活动预期指数与国内过夜游客人数的回归分析

本研究还针对海南酒店、旅行社的商务活动指数和业务活动预期指数与国内过夜游客人数之间的关系进行了分析。一共使用 5 种指标，其中 X1:酒店业务活动预期指数，X2: 旅行社业务活动预期指数，X3:酒店商务活动指数，X4:旅行社商务活动指数，Y: 海南国内过夜游客人数。这 5 个指标为月度数据，时间跨度为 2014 年 1 月-2020 年 7 月。

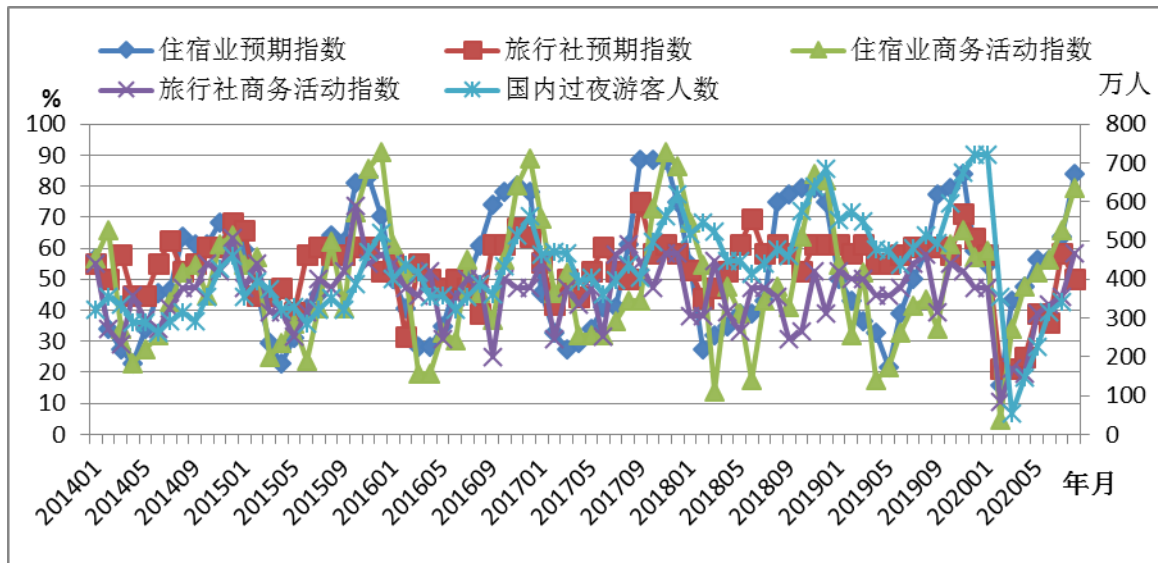


图 3-1 2014 年 1 月-2020 年 7 月海南各月酒店、旅行社业务活动预期指数和商务活动指数与国内过夜游客人数变化图

从图 3-1 可以看出，酒店、旅行社业务活动预期指数和商务活动指数与国内过夜游客人数的变化趋势基本一致，而且直观上看趋近于平稳。

（一）相关性检验

本文采用 Pearson 相关系数对酒店、旅行社业务活动预期指数和商务活动指数与国内过夜游客人数相关性进行检验。首次检验结果表明四类指数与国内过夜游客人数之间的相关性不高，最大的也没有超过 0.5，对五类指标取对数后的相关性依然不高；

由于国内过夜游客人数量纲与四类指数不一致，所以考虑计算各月游客人数指数，构成新的序列再做一次相关性检验，发现相关系数变得更低，甚至有一些出现负相关，对各月游客人数指数取对数后与其他四类取对数的指数之间相关性还是偏低；考虑到四类指数都是跟未来紧密联系的指标，将国内过夜游客取滞后一期构成新的序列再做一次相关性检验，发现相关系数变高，所有指标取对数后的相关性也都有所提高，其中旅行社业务活动预期指数的对数与滞后一期的游客人数序列之间的相关性最高（详见表）。

表 3-1 六次相关性检验结果

第一次	X1	X2	X3	X4
Y	0.35521	0.48908	0.44131	0.37324
第二次	lnx1	lnx2	lnx3	lnx4
lnY	0.24257	0.59648	0.27327	0.44954
第三次	X1	X2	X3	X4
$Y_t/Y_{(t+1)}$	0.20281	-0.1416	0.01547	-0.1156
第四次	lnX1	lnX2	lnX3	lnX4
$\ln[Y_t/Y_{(t+1)}]$	0.364	0.13835	0.36882	0.24276
第五次	X1	X2	X3	X4
$Y_{(t+1)}$	0.57319	0.60597	0.48529	0.4322
第六次	lnX1	lnX2	lnX3	lnX4
$\ln Y_{(t+1)}$	0.54113	0.7129	0.5773	0.64649

(二) 平稳性检验

当两组序列都平稳时，可以构建带输入变量回归的 ARIMAX 模型来拟合响应序列的变化，但是如果平稳性条件不满足的话，那就容易产生虚假回归的问题，所以需要对这些指标进行平稳性检验。由于这些指标取对数可以部分消除异方差的问题，另外，也可减少数据的波动，可以将可能的非线性关系转化为线性关系，减少变量的极端值和非正态分布。本研究用单位根检验方法对 $\{\ln Y_{(t+1)}\}$ 和 $\{\ln X1\}$ 、 $\{\ln X2\}$ 、 $\{\ln X3\}$ 、 $\{\ln X4\}$ 进行平稳性检验。借助编程在 SAS 中运行得到序列 $\ln Y$ 、 $\ln X1$ 、 $\ln X2$ 、 $\ln X3$ 、 $\ln X4$ 的平稳性检验结果如表 3-2 和表 3-3 所示，从检验的结果可以看出，各序列均为平稳序列，所以可以直接进行回归分析。

表 3-2 时间序列 $\ln Y$ 、 $\ln X1$ 、 $\ln X2$ 的 ADF 检验结果表

类型	延迟阶数	$\ln Y$ τ 检验统计量的值	$\Pr < \tau$ $\{\ln Y\}$	$\ln X1$ τ 检验统计量的值	$\Pr < \tau$ $\{\ln X1\}$	$\ln X2$ τ 检验统计量的值	$\Pr < \tau$ $\{\ln X2\}$
1: 无均值、无趋势模型	0	-0.23	0.6020	-0.35	0.5566	-0.45	0.5159
	1	-0.20	0.6102	-0.19	0.6156	-0.95	0.5391
2: 有均值、无趋势模型	0	-3.80	0.0043	-3.62	0.0074	-4.21	0.0012
	1	-4.21	0.0012	-4.42	0.0006	-3.73	0.0053
3: 有趋势模型	0	-3.82	0.0204	-3.64	0.0330	-4.27	0.0058
	1	-4.26	0.0060	-4.39	0.0040	-3.81	0.0211

表 3-3 时间序列 $\ln X_3$ 、 $\ln X_4$ 的 ADF 检验结果表

类型	延迟阶数	$\ln X_3$ τ 检验统计量的值	$\Pr<\tau$ $\{\ln X_3\}$	$\ln X_4$ τ 检验统计量的值	$\Pr<\tau$ $\{\ln X_4\}$
1: 无均值、无趋势模型	0	-0.65	0.4312	-0.48	0.5029
	1	-0.51	0.4894	-0.23	0.6019
2: 有均值、无趋势模型	0	-6.00	<.0001	-6.05	<.0001
	1	-5.08	0.0001	-4.32	0.0008
3: 有趋势模型	0	-5.97	<.0001	-6.12	<.0001
	1	-5.04	0.0005	-4.46	0.0033

(三) 回归分析

1. 探寻国内过夜游客人数滞后一期序列与酒店、旅行社业务活动预期指数和商务活动指数四个序列建立回归模型的可能性。

首先采用向前回归的方法,建立以 $\ln Y_{(t+1)}$ 为因变量, $\ln X_1$ 、 $\ln X_2$ 、 $\ln X_3$ 、 $\ln X_4$ 为自变量的两个有截距项和没有截距项的线性回归模型。

有截距项:

$$\ln Y_{(t+1)} = 1.34715 + 0.72038 \ln X_2 + 0.22256 \ln X_3 + 0.25492 \ln X_4 + \dots \quad (1)$$

没有截距项:

$$\ln Y_{(t+1)} = 0.99961 \ln X_2 + 0.22834 \ln X_3 + 0.31145 \ln X_4 + \dots \quad (2)$$

两个模型的拟合及诊断指标对比如表 3-3 所示。从模型的拟合优度和 F 检验值看,显然模型 2 拟合效果更好。从残差图(图 3-2)来看,残差分布较为均匀,都服从均值为 0,方差为 σ^2 的正态分布,所以模型 2 适合用来作预测。

表 3-4 向前回归法模型拟合及诊断对比

模型	变量	方程拟合 R^2	F 检验 ($\Pr>F$)
模型 1 (有截距)	$\ln X_2$	0.5082	78.54 (<.0001)
	$\ln X_3$	0.6242	62.3 (<.0001)
	$\ln X_4$	0.6476	45.32 (0.03)
模型 2 (无截距)	$\ln X_2$	0.998	38252.1 (<.0001)
	$\ln X_3$	0.9985	24632.8 (<.0001)
	$\ln X_4$	0.9986	17652.9 (<.0001)

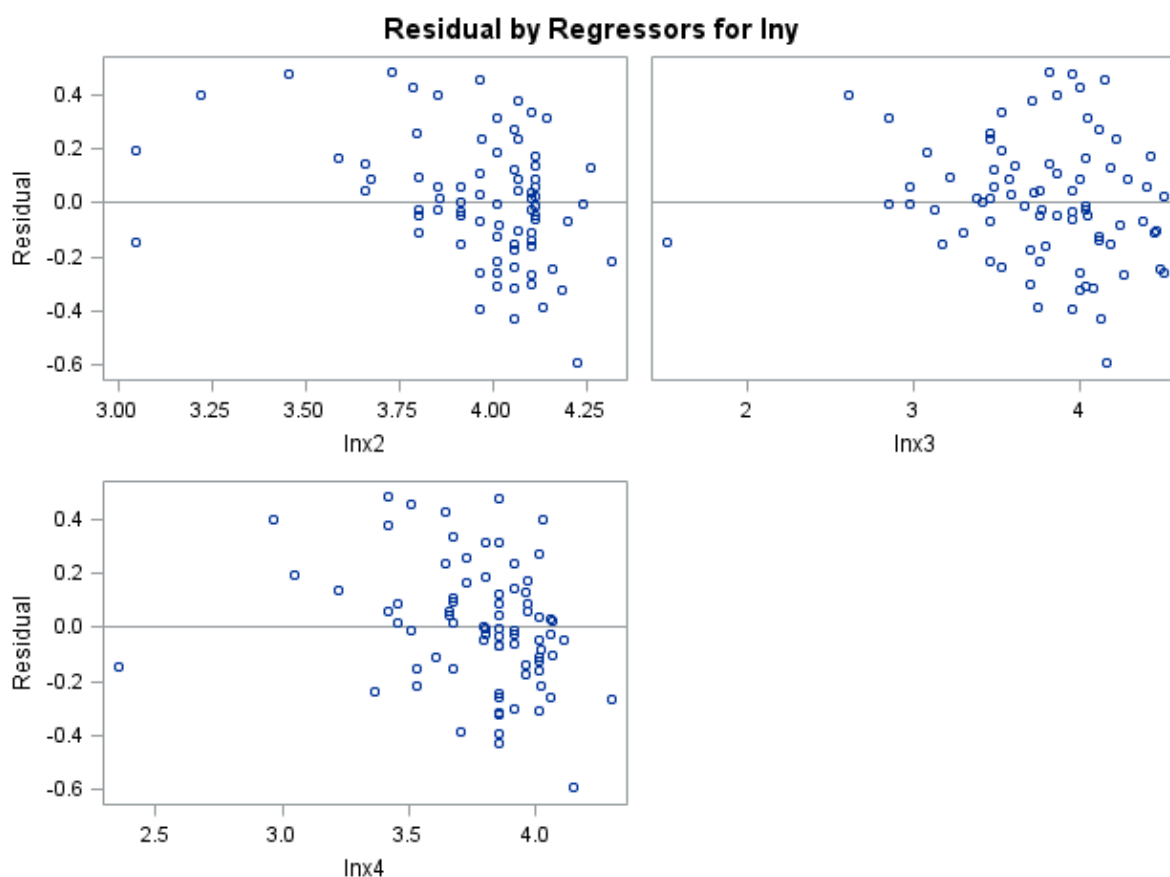


图 3-2 模型 2 拟合残差图

2. 探寻国内过夜游客人数滞后一期序列与旅行社业务活动预期指数序列建立回归模型的可能性。

根据前面的相关性检验结果来看,可以考虑建立以 $\ln Y_{(t+1)}$ 为因变量, $\ln X_2$ 为自变量的两个线性回归模型 (3) 和 (4), 有关的模型拟合指标及诊断情况如表 3-4 所示。从模型的拟合优度、F 检验和参数的 t 检验来看, 模型 4 拟合效果远远好过模型 1。而且从模型 4 残差拟合图 (图 3-3) 来看, 有 95% 以上的点在 $e=-0.5$ 和 $e=+0.5$ 的两条直线之间随机分布, 表明残差是服从均值为 0, 方差为 σ^2 的正态分布。再从预测值拟合图 (图 3-4) 来看, 基本上所有的预测值都在预测限范围内, 故回归模型 (4) 也可以用来对海南国内过夜游客人数进行预测。

$$\text{有截距项: } \ln Y_{(t+1)} = 1.68054 + 1.09123 \ln X_2 \dots\dots\dots (3)$$

$$\text{没有截距项: } \ln Y_{(t+1)} = 1.51363 \ln X_2 \dots\dots\dots (4)$$

表 3-5 单变量回归模型拟合及诊断对比

模型分类	方程拟合 R^2	F 检验 ($Pr>F$)	$\ln X_2$ 参数 t 检验
模型 3 (有截距)	0.5082	78.54 (<.0001)	8.86 (<.0001)
模型 4 (无截距)	0.998	38252.1 (<.0001)	195.58 (<.0001)

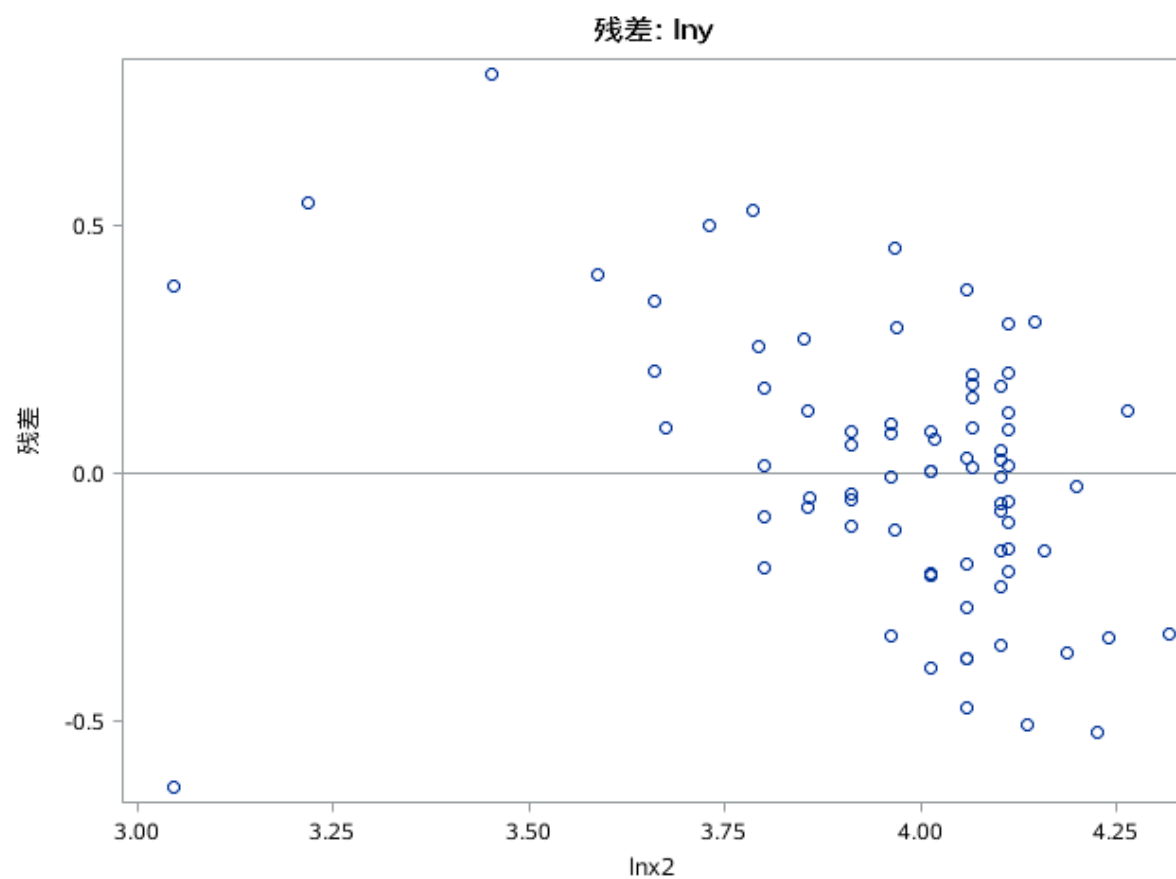


图 3-3 模型 4 拟合残差图

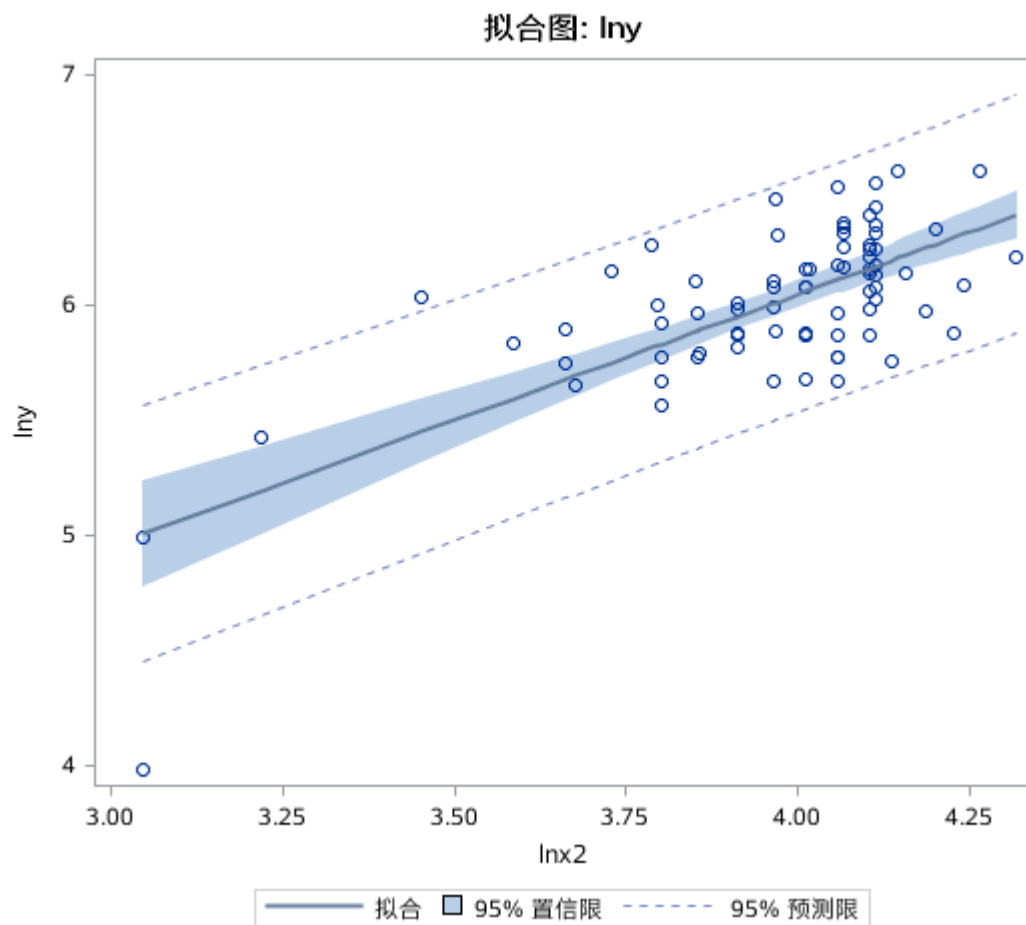


图 3-4 模型 4 拟合图

3. 回归模型预测

由于 7 月份和 8 月份旅行社对业务活动预期指数和酒店、旅行社商务活动指数已知，分别将其代入模型（2）和模型（4），可以分别得到两组预测值（表 3-6）。根据往年的经验，模型（2）的预测效果更好。

表 3-6 回归预测结果对比（万人）

预测值	模型 2	模型 4
8 月	494	471
9 月	481	373

四、主要结论和存在问题

（一）主要结论

1. 通过建立时间序列分析模型，以选定指标的模型预测值与实际值的差异来度量新冠疫情对海南省旅游业产生的影响，可以看到疫情发生后海南省旅游业接近“暂停”，在 2 月时仅为正常水平的一成左右，而随着复工复产推进，在 6 月

时已恢复至正常水平的七成。显然在旅游行业已恢复营业的今天，如何提高人们出游积极性、恢复旅行计划，如何从七成恢复至正常水平，是当前亟待思考的问题。

2. 分类采购经理指数长期内能较好地对接国内过夜游客人数进行预测，主要是由于采购经理对本行业相关信息掌握较为充分，而采购经理指数是采购经理经综合分析考量后的判断结果，可靠性较高，特别是在遇到新冠疫情这类突发事件时，回归模型也具有较好的拟合效果。根据回归模型结果，我们预测八月接待国内游客人数将上升，而九月接待国内游客人数会下降。待九月采购经理指数统计后，我们即可对后续月份接待国内游客人数进行预测。

（二）研究中存在的问题

1. “接待国内过夜游客人数”数据主要来源于海南省旅游和文化广电体育厅，虽然统计口径基本一致，但由于时间跨度较长可能有部分变动未被察觉。例如 2019 年 10 月起“接待国内过夜游客人数”中新纳入了“候鸟游客”，本文在数据使用前做了相应处理，但可能存在其他未标明的变动影响建模结果。

2. 使用 ARIMA 模型对“接待国内过夜游客人数”进行模型拟合要求数据或差分后的数据是平稳的，但游客人数常受到政策、宣传等因素影响，例如 2010 年海南推进“国际旅游岛”建设，2018 年海南全岛“自由贸易试验区”建设，都对后期游客人数产生明显影响，模型拟合尚未能充分体现这些影响。基于 ARIMA 模型的特性，拟合模型的预测误差会随着时间的延长而变大。

3. 基于数据的可获取性，本文在 ARIMA 建模和回归预测阶段均使用了省一级的指标数据，但现实中可能海口、三亚等城市受疫情影响接待国内游客减少量较大，而其他城市影响相对较小，如能获取更微观层面的数据，细化分析结果将使建模分析更具实用性。

参考文献

- [1]Goh Carey, *et al.* Modeling and forecasting tourism demand for arrivals with stochastic nonstationary seasonality and intervention. *Tourism Management*. 2002, 23: 499~ 510
- [2]朱迎波, *et al.* SARS 对中国入境旅游人数影响的研究. *地理研究*. 2003 年 9

月, Vol. 22, No. 5: 551-558

[3]Zheng Cao, *et al.* Modelling the interdependence of tourism demand: The globalvector autoregressive approach. Annals of Tourism Research 67(2017)1-13

[4]王慧蕾. 基于灰色 - 灰色残差修正的广东旅游业收入预测. 农家参谋, 2017年 12:257-258

[5]Hossein Hassani, *et al.* Forecasting accuracy evaluation of tourist arrivals. Annals of Tourism Research 63(2017)112-117

[6]Xin Li, *et al.* Forecasting tourism demand with composite search index. Tourism Management 59(2017)57-66