

中国经济区域GDP特征结构分析与建模

许嘉峻, 卢天棋, 侯欣然, 谢京京, 赵文丽, 刘 博, 白晓东

大连民族大学, 理学院、预科教育学院, 辽宁 大连

收稿日期: 2023年10月17日; 录用日期: 2023年12月4日; 发布日期: 2023年12月11日

摘 要

本文借用GDP增速这一指标研究了中国各区域经济的发展以及受疫情影响的程度。从2013年第三季度至2020年第四季度的六大经济区GDP增速及国家GDP增速的数据出发,建立了ARIMA模型,并由此对中国各区域的经济增速情况进行了分析,有助于掌握疫情防控背景下不同地区的经济恢复情况。

关键词

六大经济区, ARIMA, GDP增速

Analysis and Modeling of the Characteristic Structure of China's Economic Regional GDP

Jiajun Xu, Tianqi Lu, Xinran Hou, Jingjing Xie, Wenli Zhao, Bo Liu, Xiaodong Bai

School of Science and School of Pre-University, Dalian Minzu University, Dalian Liaoning

Received: Oct. 17th, 2023; accepted: Dec. 4th, 2023; published: Dec. 11th, 2023

Abstract

This paper uses the index of GDP growth rate to study the economic development of each region and the extent affected by the epidemic. The ARIMA model was established based on the data of the GDP growth rate of the six major economic zones and the national GDP growth rate from the third quarter of 2013 to the fourth quarter of 2020 to analyze the economic growth rate of various regions in China, and understand the economic recovery of different regions under the background of epidemic prevention and control.

Keywords

Six Economic Zones, ARIMA, GDP Growth Rate

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

近年来,国内区域经济研究的学者学科背景横跨经济学、金融学、地理学、城市规划等学科[1][2][3]。但应用 ARIMA 模型对区域 GDP 进行研究的还很少。2020 年疫情对我国各经济区造成了一定的冲击,对各经济区的具体影响程度也不尽相同。刘帅(2021)[4]、孙久文(2020)[5]对疫情影响进行了理论描述,从 GDP 增速角度进行 ARIMA 预测,得到疫情对经济造成的影响程度及疫情得到控制后各经济区域的恢复情况。

2. 研究内容和方法

根据 2013~2019 年季度数据,本文查阅了陈国艳(2003)[6]、陈璐(2019)[7]关于经济的研究,将全国共分为六大经济区,基于 R 语言画出各经济区以及国家 GDP 增速的时序图,对其建立 ARIMA 模型进行预测,并尝试分析近年来我国经济区域 GDP 的特征结构、各经济区受疫情影响以及后疫情时代各经济区恢复情况等问题。如表 1 所示。

Table 1. List of economic zones

表 1. 各经济区名单

经济区名称	省市
东北经济区	辽宁省、吉林省、黑龙江省
长江中游经济区	安徽省、江西省、湖北省、湖南省
黄河中游经济区	山西省、内蒙古自治区、河南省、陕西省
西北经济区	西藏、甘肃省、青海省、宁夏、新疆
沿海及发达经济区	北京市、天津市、河北省、山东省、上海市、江苏省、浙江省、福建省、广东省、海南省
西南经济区	广西壮族自治区、重庆市、四川省、贵州省、云南省

3. ARIMA 模型相关定义及步骤

本文构造 ARIMA 模型,实质上是差分运算与 ARMA 模型(自回归移动平均模型)的结合,意在于将序列通过一阶或多阶差分后能够达到平稳,方能使用 ARMA 模型进行拟合、预测。

设 $\{x_t, t \in T\}$ 是一个时间序列,称满足如下结构的模型为自回归移动平均模型,即为 ARMA(p, q):

$$x_t = \phi_0 + \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (3.1)$$

其中, $\phi_p \neq 0, \theta_q \neq 0$, ε_t 为均值为零的白噪声序列,且 ε_t 与 x_{t-j} ($j=1, 2, \dots$) 无关,即 $\forall s < t, E(x_s \varepsilon_t) = 0$ 。

借助延迟算子,ARMA 模型可简记为:

$$\Phi(B)x_t = \Theta(B)\varepsilon_t \quad (3.2)$$

构的模型称为 ARIMA(p, d, q)模型:

$$\begin{cases} \Phi(B)\nabla^d x_t = \Theta(B)\varepsilon_t \\ E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ E x_t \varepsilon_s = 0, \forall s < t \end{cases} \quad (3.3)$$

其中，非负整数 d 为求和阶数， $\Phi(B) = 1 - \Phi_1 B - \dots - \Phi_p B^p$ ，为平稳可逆 ARMA(p, q)模型的自回归系数多项式； $\Theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ 为平稳可逆 ARMA 模型的移动平均系数多项式。 d 阶差分算子：

$$\nabla^d = (1 - B)^d = \sum_{k=0}^d C_d^k (-B)^k \tag{3.4}$$

通过参考建立时间序列相关书籍得到了 ARIMA 的建模步骤[8]：

- 1) 判断观察值序列的平稳性；
- 2) 对数据进行差分；
- 3) 对数据进行白噪声检验；
- 4) 模型识别，如表 2 所示；

Table 2. Model order determination principles
表 2. 模型定阶原则

模型	自相关系数	偏自相关系数
AR(p)	拖尾	p 阶结尾
MA(q)	q 阶结尾	拖尾
ARIMA(p, d, q)	结尾	拖尾

- 5) 模型有效性检验及模型的优化；
- 6) 预测序列未来走势。

4. 六大经济区 GDP 增速发展现状

收集六大经济区 2013 年第一季度至 2020 年第四季度的 GDP 增速数据并画出相应的时序图，得到六大经济区 GDP 增速的曲线图。

从图 1 沿海及发达经济区 GDP 增速时序图整体来看，这一经济区 GDP 增速较为平稳，但从 2019 年第一季度开始 GDP 增速呈现下降的趋势，在 2020 年第一季度时达到了最低点，从 2020 年第一季度开始呈现上升的趋势。

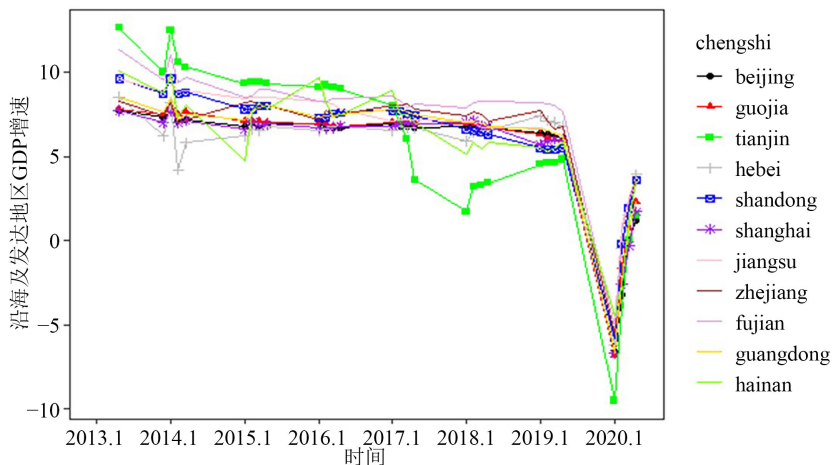


Figure 1. Time sequence chart of GDP growth rate in coastal and developed economic zones
图 1. 沿海及发达经济区 GDP 增速时序图

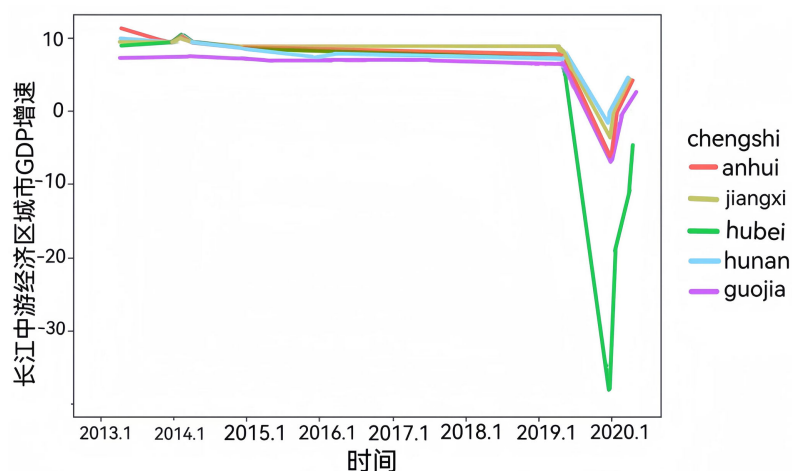


Figure 2. Time sequence chart of GDP growth rate of the economic zone in the middle reaches of the Yangtze River
图 2. 长江中游经济区 GDP 增速时序图

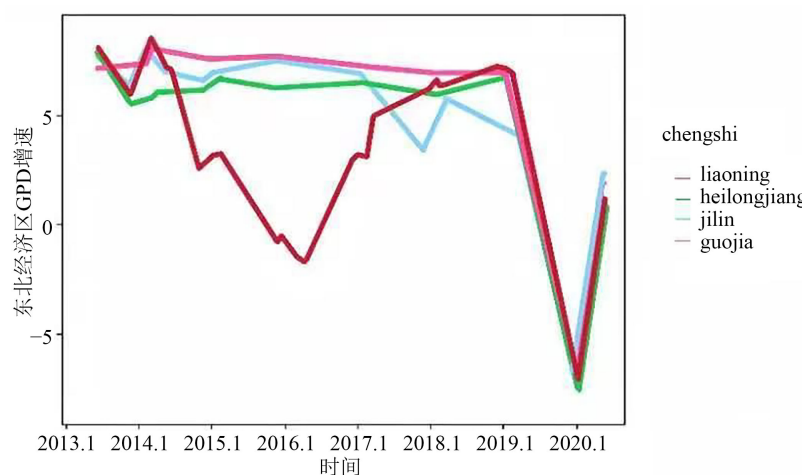


Figure 3. Time sequence chart of GDP growth rate in northeast economic zone
图 3. 东北经济区 GDP 增速时序图

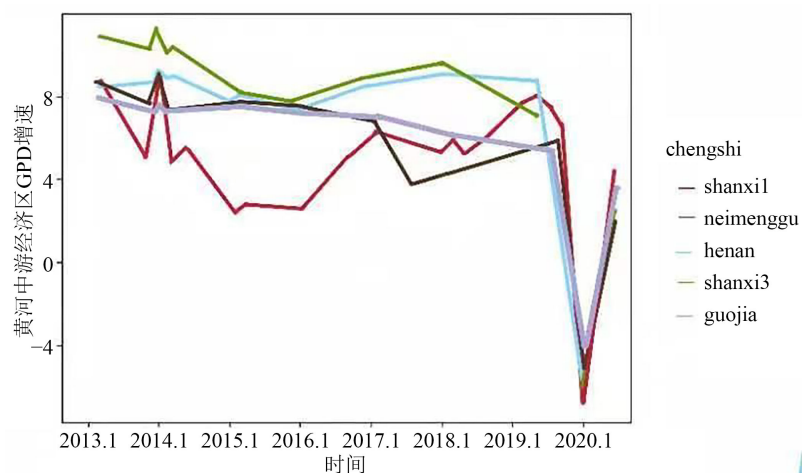


Figure 4. Time sequence chart of GDP growth rate of the economic zone in the middle reaches of the Yellow River
图 4. 黄河中游经济区 GDP 增速时序图



Figure 5. Time sequence chart of GDP growth rate in northwest economic zone
图 5. 西北经济区 GDP 增速时序图

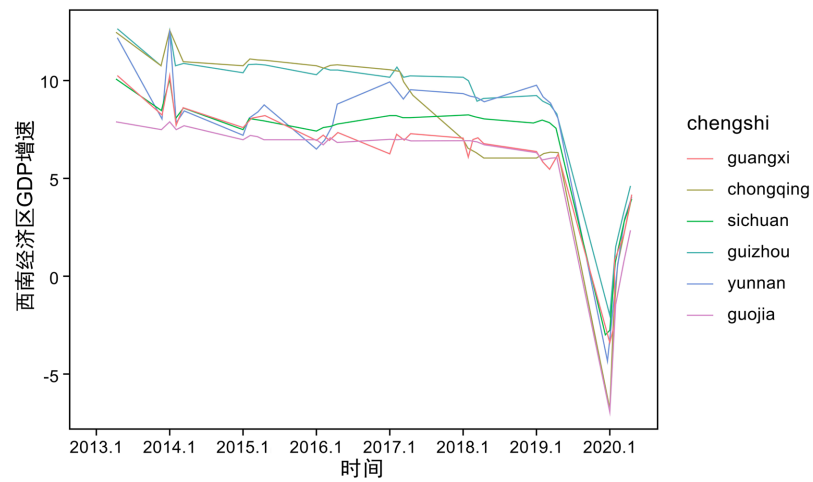


Figure 6. Time sequence chart of GDP growth rate in southwest economic zone
图 6. 西南经济区 GDP 增速时序图

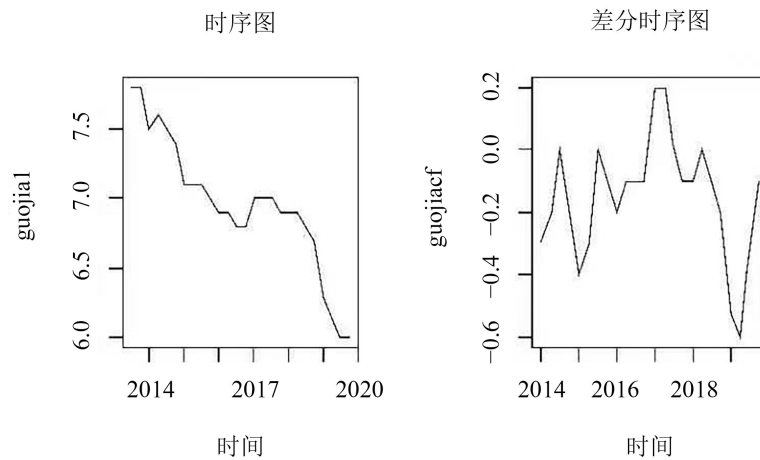


Figure 7. National GDP growth rate time sequence chart and difference time sequence chart
图 7. 国家 GDP 增速时序图及差分时序图

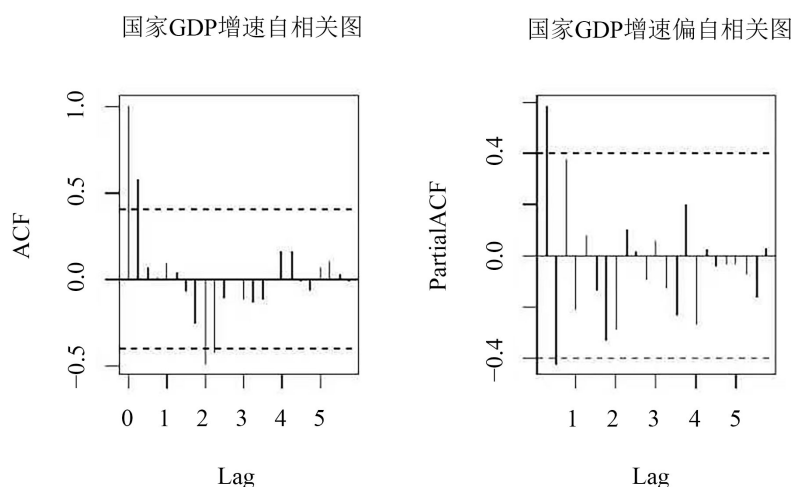


Figure 8. National GDP growth rate difference autocorrelation chart and partial autocorrelation chart
图 8. 国家 GDP 增速差分自相关图及偏自相关图

由图 2 长江中游经济区 GDP 增速时序图可知, 在 2019 年前, 这四省 GDP 增速均呈现平稳的趋势且增速略高于国家增速, 在 2019 年第一季度开始, 这一经济区 GDP 增速呈现下降的趋势。由图 3 东北经济区 GDP 增速时序图可知, 从总体来看, 黑龙江、吉林、辽宁 GDP 增速均低于国家 GDP 的增速且 GDP 增速呈现不稳定的趋势。由图 4 可知, 在 2019 年前除山西外黄河中游经济区各省份 GDP 增速均高于国家 GDP 增速。由图 5 可知, 西北经济区内各省份 GDP 增速均高于国家 GDP 增速。由图 6 可知, 西南经济区内各省份 GDP 增速发展水平均高于国家发展水平。

4.1. 数据预处理及平稳化

对六大经济区以及国家 GDP 增速共 32 组数据建立 ARIMA 模型, 对数据进行差分后画出经差分后的数据得到图 7, 由差分图可知, 差分后的时序图呈现平稳的趋势。将差分后的数据画出自相关图和偏自相关图, 如图 8 所示。

由图 8 可知, 国家 GDP 增速差分后的自相关图呈现二阶拖尾性, 偏自相关呈现一阶拖尾性。

4.2. 模型识别、检验及优化

根据国家差分自相关图和偏自相关图, 可得到式(4.1)。

根据表 3 以及信息准则法可知, ARIMA(2, 2, 1)为国家 GDP 增速进行预测时的最优模型。其余各省市拟合模型结果如表 4 所示。

Table 3. National GDP growth rate prediction model and corresponding AIC values

表 3. 国家 GDP 增速预测模型及相应 AIC 值

模型	AIC 值
ARIMA(1, 2, 1)	-6.71
ARIMA(2, 2, 1)	-13.29
ARIMA(1, 2, 2)	-8.64
ARIMA(2, 2, 2)	-11.32

Table 4. ARIMA model of the provinces and cities in six major economic zones
表 4. 六大经济区各省市 ARIMA 模型

城市	拟合模型 ARIMA(p, d, q)	残差值 (6 阶)	残差值 (12 阶)	AIC
北京市	(1, 0, 1)	0.6604	0.4672	0.73
天津市	(2, 2, 2)	0.6054	0.8967	72.93
河北省	(1, 0, 1)	0.8883	0.4503	95.35
山东省	(5, 0, 1)	0.979	0.992	28.3
上海市	(2, 0, 1)	0.9453	0.9881	19.09
江苏省	(1, 3, 2)	0.0941	0.0636	29.3
浙江省	(1, 0, 1)	0.7665	0.3406	32.97
福建省	(4, 0, 1)	0.4231	0.8335	49.44
广东省	(1, 0, 1)	0.4115	0.5375	31.3
海南省	(1, 0, 1)	0.8469	0.7898	92.5
安徽省	(2, 2, 1)	0.6547	0.5714	3.61
江西省	(1, 0, 1)	0.5671	0.8182	16.5
湖北省	(4, 0, 1)	0.127	0.2215	27.81
湖南省	(5, 0, 1)	0.8586	0.9903	34.22
山西	(1, 0, 2)	0.4091	0.7015	82.11
内蒙古自治区	(2, 0, 0)	0.5518	0.8383	54.42
河南省	(1, 1, 1)	0.8942	0.9941	34.33
陕西省	(1, 0, 1)	0.436	0.8948	72.04
辽宁	(1, 0, 1)	0.978	0.9061	97.1
黑龙江	(2, 1, 1)	0.481	0.5904	65.91
吉林	(3, 0, 1)	0.9024	0.7402	66.67
广西壮族自治区	(2, 1, 1)	0.1792	0.0603	56.29
重庆市	(1, 2, 2)	0.7946	0.935	51.57
四川省	(1, 0, 2)	0.5786	0.8623	39.06
贵州省	(2, 0, 1)	0.2888	0.6891	42.51
云南省	(2, 0, 1)	0.8182	0.6104	83.35
西藏自治区	(2, 0, 3)	0.7911	0.8288	69.71
甘肃省	(1, 0, 1)	0.5915	0.9475	84.1
青海省	(1, 0, 0)	0.575	0.8271	57.52
宁夏回族自治区	(1, 0, 1)	0.7895	0.5237	43.24
新疆自治区	(1, 1, 2)	0.2693	0.4356	69.79
国家	(2, 2, 1)	0.9355	0.5629	-13.29

4.3. 模型预测

通过模型的识别、检验和优化可以得到基于 ARIMA 的国家 GDP 增速预测模型为:

$$\begin{aligned} x_t &= 0.0986 x_{t-1} + 0.1195 x_{t-2} + \varepsilon_t - \varepsilon_{t-1} \\ \varepsilon_t &\sim N(0, 0.01848) \end{aligned} \quad (4.1)$$

根据模型对数据进行预测得到国家 2020 年四个季度的 GDP 增速分别为: 5.932%、5.869%、5.798%、5.727%。其余各省市 GDP 增速预测结果见表 5。

4.4. 模型精度检验

依据现有模型, 利用 2013 年第三季度至 2019 年第二季度的数据预测 2019 年第三、四季度的数据, 通过|真实值 - 预测值|/真实值来检验模型的精度得到表 6。

Table 5. Forecast values for the four quarters of 2020 for the six major economic zones

表 5. 六大经济区 2020 年四个季度的预测值

GDP 增速城市	2020 年第一季度(%)	2020 年第二季度(%)	2020 年第三季度(%)	2020 第四季度(%)
北京市	6.156	6.186	6.216	6.244
天津市	4.623	4.328	4.039	3.742
河北省	6.941	7.02	7.094	7.163
山东省	5.033	5.069	5.147	5.36
上海市	5.978	6.185	6.142	6.331
江苏省	6	5.874	5.747	5.617
浙江省	7.249	7.482	7.562	7.59
福建省	7.71	7.861	7.998	7.907
广东省	6.359	6.46	6.551	6.633
海南省	6.044	6.219	6.371	6.503
安徽省	7.479	7.342	7.205	7.092
江西省	8.243	8.273	8.301	8.329
湖北省	7.735	7.828	7.881	7.712
湖南省	7.571	7.344	7.832	7.761
山西	6.73	6.611	5.908	5.796
内蒙古自治区	5.251	5.419	5.6	5.766
河南省	7.31	7.342	7.345	7.346
陕西省	6.31	6.589	6.827	7.029
辽宁	5.425	5.363	5.309	5.263
黑龙江	4.231	4.27	4.265	4.251
吉林	3.263	4.062	4.073	4.697
广西壮族自治区	5.853	5.888	5.878	5.88
重庆市	6.161	5.974	5.762	5.537
四川省	7.633	7.962	8.182	8.119
贵州省	8.227	8.415	8.339	8.523

Continued

云南省	8.351	8.62	8.75	8.813
西藏	8.723	8.928	8.92	9.046
甘肃省	6.376	6.527	6.66	6.777
青海省	6.422	6.537	6.644	6.743
宁夏回族自治区	6.848	6.968	7.063	7.138
新疆自治区	5.661	5.517	5.374	5.231
国家	5.932	5.869	5.798	5.727

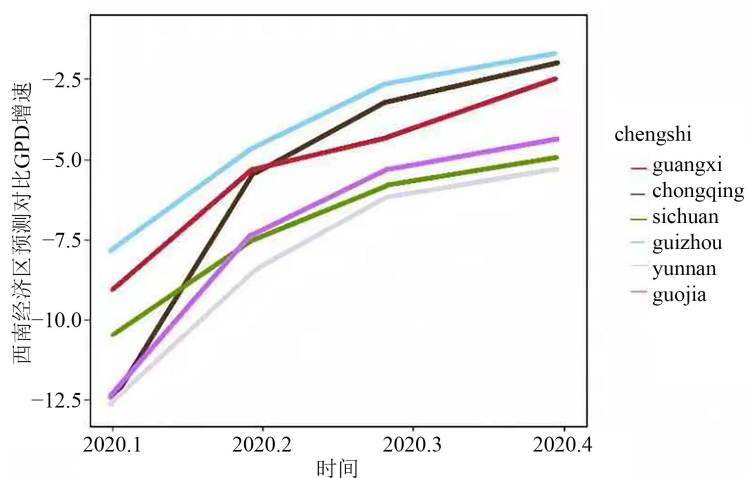


Figure 9. Time sequence chart of GDP growth rate forecast comparison in coastal and developed economic zones
图 9. 沿海及发达经济区 GDP 增速预测对比时序图

Table 6. Average relative error of ARIMA model in the provinces and cities of the six major economic zones
表 6. 六大经济区各省市 ARIMA 模型平均相对误差

城市	平均相对误差	城市	平均相对误差
北京市	0.010648	内蒙古自治区	0.067988
天津市	0.125497	河南省	0.075077
河北省	0.020395	陕西省	0.017006
山东省	0.006149	辽宁	0.009133
上海市	0.019992	黑龙江	0.0298
江苏省	0.016995	吉林	0.082823
浙江省	0.099335	广西	0.061808
福建省	0.053985	重庆	0.049945
广东省	0.081467	四川	0.062474
海南省	0.071237	贵州	0.063438
安徽省	0.019391	云南	0.067580
江西省	0.047389	西藏	0.097009
湖北省	0.037085	甘肃省	0.038116

Continued

湖南省	0.047036	青海省	0.038138
山西	0.096552	宁夏回族自治区	0.062692
国家	0.034012	新疆自治区	0.091042

根据表 6 可知，模型平均误差均在 0.2 以内，模型精度较高。

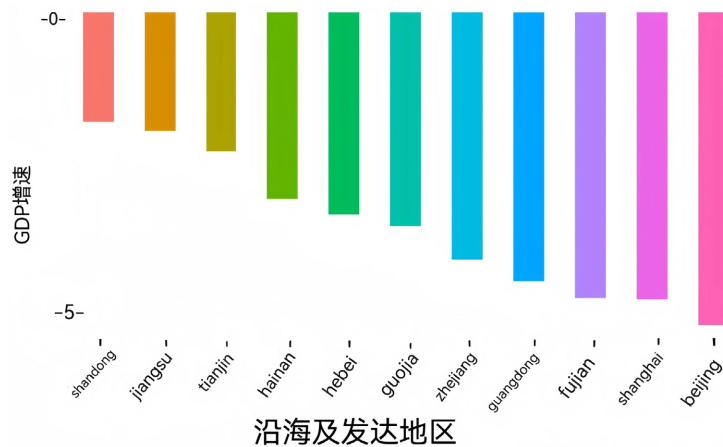


Figure 10. Bar chart of GDP growth rate forecast comparison in the fourth quarter of coastal and developed economic zones
图 10. 沿海及发达经济区 GDP 增速第四季度预测对比柱形图

5. 六大经济区预测数据与实际数据对比分析

通过用实际 GDP 增速的数据减去预测得到的全国各省市 GDP 增速的数据得到了六大经济区受疫情影响的 GDP 增速的数据。根据得到的数据是正是负来判断疫情对各省市 GDP 增速影响的大小。

由图 9、图 10 可知，沿海及发达地区受疫情影响情况以及恢复情况与国家发展相近。可以看出，疫情后山东江苏和天津基本已经恢复到了疫情前的 GDP 增速，经济恢复水平高于国家经济恢复水平。其中，天津在疫情之经济活动受影响较大，但根据柱形图显示，天津经济恢复水平要高于国家总体水平，恢复情况较好。

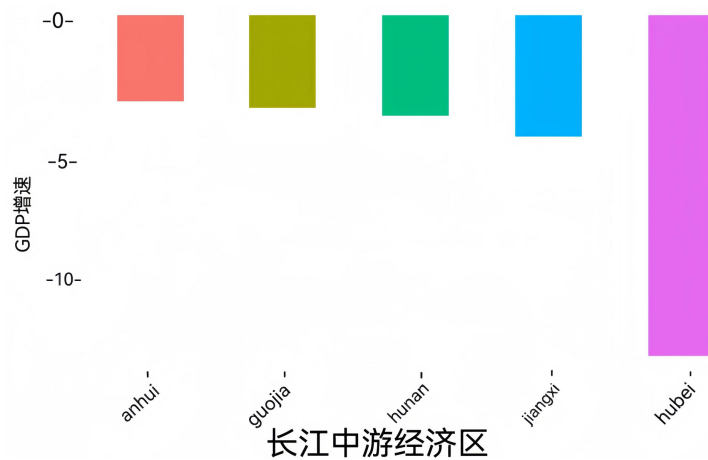


Figure 11. Time sequence chart of GDP growth forecast comparison of the economic zone in the middle reaches of the Yangtze River
图 11. 长江中游经济区 GDP 增速预测对比时序图

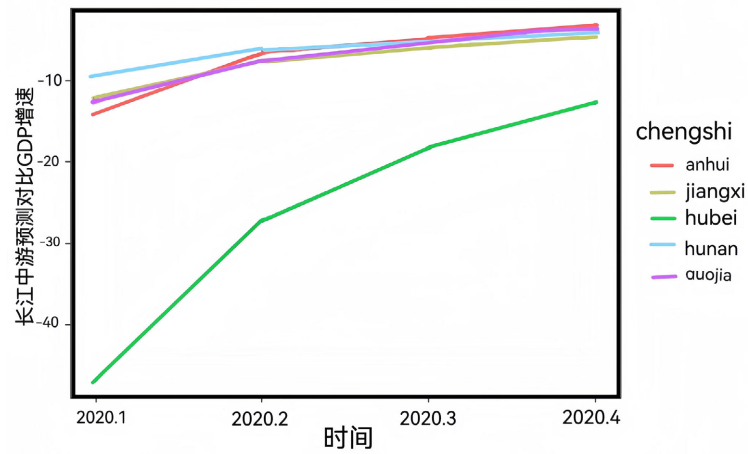


Figure 12. Bar chart of GDP growth rate forecast comparison in the fourth quarter of the economic zone in the middle reaches of the Yangtze River
图 12. 长江中游经济区第四季度 GDP 增速预测对比柱形图

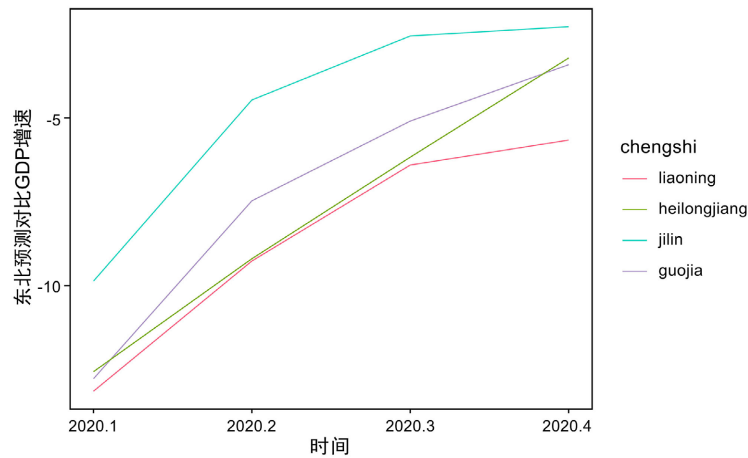


Figure 13. Time sequence chart of GDP growth forecast comparison in northeast economic zone
图 13. 东北经济区 GDP 增速预测对比时序图

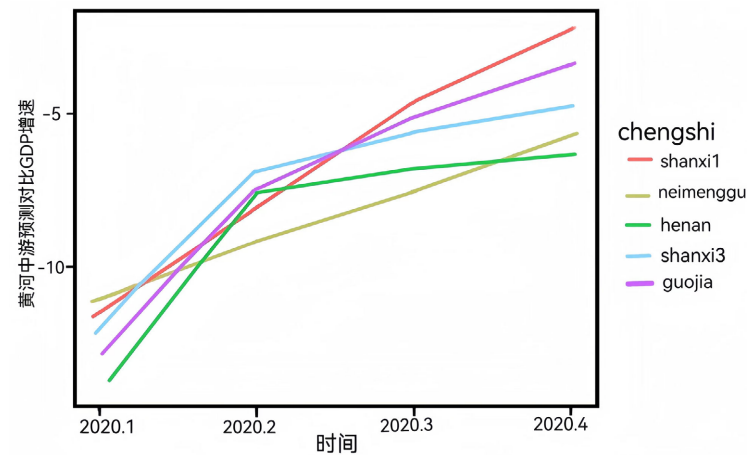


Figure 14. Time sequence chart of GDP growth forecast comparison of the economic zone in the middle reaches of the Yellow River
图 14. 黄河中游经济区 GDP 增速预测对比时序图

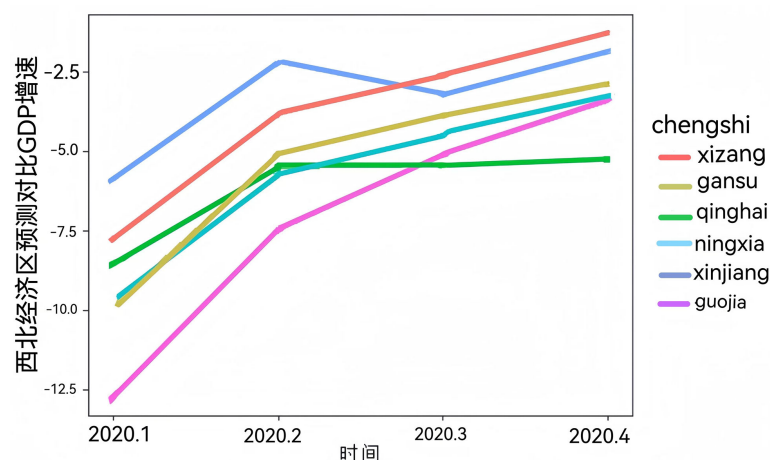


Figure 15. Time sequence chart of GDP growth forecast comparison in northwest economic zone
图 15. 西北经济区 GDP 增速预测对比时序图

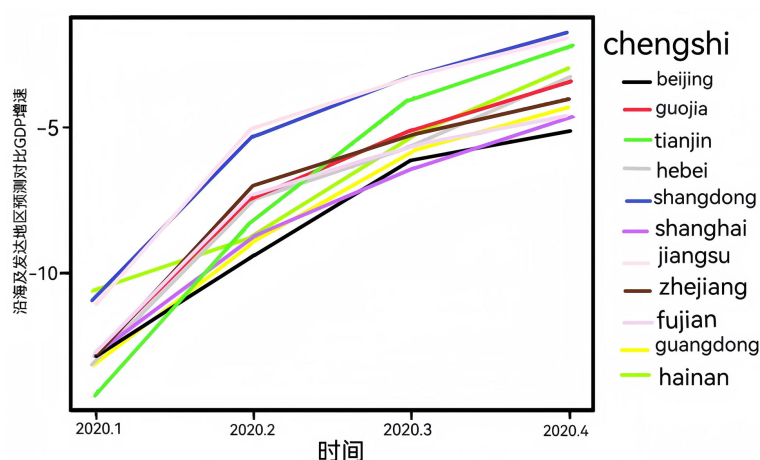


Figure 16. Time sequence chart of GDP growth forecast comparison in southwest economic zone
图 16. 西南经济区 GDP 增速预测对比时序图

在长江中游经济区中由图 11 可以看出,在疫情之初,湖北受疫情影响最严重,几乎比预期 GDP 增速低了 50%,但经济恢复情况较好。由图 12 可知,除湖北外,长江中游经济区各省份受疫情影响情况与国家总体受疫情影响情况相似,除安徽外,其余三个省份 2020 年第四季度 GDP 增速低于国家总体恢复水平,长江中游经济区经济恢复水平低于国家总体水平。

由图 13 可知,在疫情之初,东北经济区受疫情影响较大。经过调整恢复,东北经济区恢复情况还是略高于国家总体水平。由图 14 可知,在疫情之初,黄河中游经济区总体来看受疫情影响较小,但黄河中游经济恢复情况要低于国家总体水平。

由图 15 西北经济区 GDP 增速预测对比时序图可知,从总体来看,西北经济区各省份 GDP 增速受影响水平要低于国家总体水平,根据 2020 年第四季度的数据表明,除青海外,西北经济区各省份 GDP 增速基本恢复正常,均高于国家总体水平。由图 16 发现,西南经济区受疫情影响较小,经全面复工复产,经济恢复水平要高于国家总体水平。

6. 结论

总体来看,六大经济区中沿海及发达经济区与国家 GDP 增速发展水平相似,而长江中游经济区等其

他区 GDP 增速均高于国家 GDP 增速, 只有东北经济区 GDP 增速略低于国家 GDP 增速。在疫情之初, 沿海及发达经济区与国家总体情况相似, 而黄河中游经济区、西北经济区、西南经济区受疫情影响较小, 长江中游经济区受疫情影响较为严重。另外, 东北经济区受疫情影响较大。在疫情防控下, 西北经济区、东北经济区经济恢复情况略高于国家总体恢复情况, 沿海及发达经济区、西南经济区与国家总体恢复水平相近, 黄河中游经济区、长江中游经济区经济恢复情况低于国家总体恢复水平。西南经济中的重庆和沿海及发达地区中的天津经济恢复较快, 恢复水平较好。在面对新冠疫情这类不利的冲击时, 可借鉴重庆和天津这两座城市的相关政策。

研究的不足: 首先, 建立 ARIMA 模型的过程中具有主观性, 只能对短期内的数据进行预测。其次, 由于知识储备量的有限, 对各经济区发展水平等相关问题描述以及分析得不到位。

基金项目

国家级大创项目资助(项目编号: 202312026043); 校级大创项目资助(项目编号: 202312026457)。

参考文献

- [1] 骆达荣, 黄灏然, 郭开仲. 基于 SVR 的 RAR 区域经济预测模型[J]. 数学的实践与认识, 2013, 43(19): 36-42.
- [2] 刘艳春, 孙亮. 人口因素与区域经济的多重协整关系研究——基于沈阳市时间序列数据的分析[J]. 数理统计与管理, 2016, 35(2): 309-318.
- [3] 胡晶. 数据分类及其在中国经济区域划分问题上的应用[D]: [硕士学位论文]. 长春: 东北师范大学, 2003.
- [4] 刘帅. 新冠肺炎疫情对中国区域经济的影响[J]. 地理研究, 2021, 40(2): 310-325.
- [5] 孙久文. 新冠肺炎疫情对中国区域经济发展的影响初探[J]. 区域经济评论, 2020, 36(2): 8-11.
- [6] 陈国艳. 基于中国各省 GDP 的经济区域划分统计方法的研究[D]: [硕士学位论文]. 长春: 东北师范大学, 2003.
- [7] 陈璐. 环境保护税费的异质性影响研究: 基于我国六大经济区面板数据的实证分析[D]: [硕士学位论文]. 成都: 西南财经大学, 2019.
- [8] 白晓东. 应用时间序列分析[M]. 北京: 清华大学出版社, 2017.