

基于ARIMA模型的上海市GDP分析与预测

谭梓怡¹, 何文²

¹上海理工大学, 管理学院, 金融系, 上海

²上海理工大学, 管理学院, 国际经济与贸易系, 上海

收稿日期: 2023年12月6日; 录用日期: 2024年1月9日; 发布日期: 2024年1月17日

摘要

本GDP是衡量国家或地区经济发展水平和人民生活状况的重要指标, 分析与预测GDP未来趋势可以为国家制定和检验经济政策提供参考。上海市作为我国的经济、金融、贸易和航运中心, 对全国经济发展起着引领带动作用。因此, 文章选取上海市1991~2022年的GDP年度统计数据建立ARIMA模型, 对上海市2023~2025年的GDP进行预测。研究结果表明, 2023~2025年上海市GDP将持续稳步增长, 预测2023~2025年依次为4.97万亿元、5.44万亿元、5.94万亿元。最后, 基于模型预测结果并结合上海市经济发展特点为上海市未来经济发展规划提出参考建议。

关键词

GDP, ARIMA模型, 时间序列数据, 经济预测

Analysis and Forecasting of Shanghai's GDP Based on ARIMA Model

Ziyi Tan¹, Wen He²

¹Finance Department, School of Management, University of Shanghai for Science and Technology, Shanghai

²Department of International Economy and International Trade, School of Management, University of Shanghai for Science and Technology, Shanghai

Received: Dec. 6th, 2023; accepted: Jan. 9th, 2024; published: Jan. 17th, 2024

Abstract

As an important indicator to measure the level of economic development of a country or region and the living conditions of the people, analyzing and predicting the future trend of GDP can provide a reference for the country to formulate and test economic policies. As the economic, financial, trade and shipping center of China, Shanghai plays a leading role in the national economic devel-

文章引用: 谭梓怡, 何文. 基于 ARIMA 模型的上海市 GDP 分析与预测[J]. 理论数学, 2024, 14(1): 79-86.

DOI: 10.12677/pm.2024.141009

opment. Therefore, the paper selected the annual statistics of Shanghai's GDP from 1991 to 2022 and established an ARIMA model to forecast Shanghai's GDP from 2023 to 2025. The results of the study show that the GDP of Shanghai will continue to grow steadily from 2023 to 2025, and it is predicted to be 4.97 trillion yuan, 5.44 trillion yuan, and 5.94 trillion yuan in sequence from 2023 to 2025. Finally, recommendations are made for the future economic development planning of Shanghai based on the model prediction results and the characteristics of Shanghai's economic development.

Keywords

GDP, ARIMA Model, Time Series Data, Economic Prediction

Copyright © 2024 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

国内生产总值(Gross Domestic Product, GDP)是指一个国家或地区在一定时期内全部经济活动所创造的总价值,是衡量国家或地区经济发展规模的重要经济指标。一方面,它代表着整个国家或地区的发展和人民生活状况;另一方面,又代表着国民经济的增长、市场价格的变动和社会发展的步伐。因此,分析与预测 GDP 未来趋势既有利于人们全面了解宏观经济运行状况和国民福利水平,同时也为国家制定和检验经济政策提供了重要依据[1]。上海市作为全国性的经济、金融、贸易中心,经济总量持续保持第一,对全国经济发展有着引领带动作用。近年来,上海已成为全球金融业的重要中心之一,同时依托长江三角洲的产业链和地理位置优势,其在国家现代化建设大局和全方位开放格局中具有举足轻重的战略地位。但是自 2019 年以来,受新冠肺炎疫情影响,经济发展的外部挑战性上升,上海市 GDP 增速也受到一定影响。此外,上海领先第 2 位城市北京的优势也越来越小,2022 年上半年还曾一度被北京超越。2023 年是全面贯彻落实党的二十大的开局之年,实施“十四五”规划承上启下的关键一年,新一届政府开展各项工作的新起点,做好上海市经济发展工作意义重大。本文通过选取上海市 1991~2022 年 GDP 年度数据进行分析和建模,并基于 ARIMA 模型对未来三年上海市 GDP 进行预测,为上海市未来经济发展规划提供科学的参考依据。

2. 文献综述

对 GDP 的分析与预测是众多学者持续关注的重点问题,国内已有诸多研究使用 ARIMA 模型对 GDP 进行分析与预测并表示具有良好的预测效果。其中,李振亮和乐昕雨[2]通过选取北京市 1978~2020 年的 GDP 年度统计数据,运用 R 软件构建 ARIMA(2,2,1)模型对北京市 2021~2025 年 GDP 进行预测,相对误差在 7% 以内。杨忠裕和薛紫玥[3]选取甘肃省 1992~2021 年 GDP 数据建立 ARIMA(0,2,1)模型,对未来两年甘肃省 GDP 做出短期预测。张梓[4]通过选取贵州省 1978~2020 年 GDP 数据,运用 Python 构建 ARIMA(0,1,1)模型来对贵州省 2021~2025 年 GDP 进行预测,相对误差控制在 7% 以下。夏如玉和王梓桥[5]选取重庆市 2009~2020 年 GDP 数据建立 ARIMA(0,1,0)模型,对重庆市未来 12 年的 GDP 指数进行了预测。潘典雅[6]根据吉林省 1993~2017 年的 GDP 数据建立最优模型 ARIMA(2,1,1)对 2018~2019 年内的 GDP 做出短期预测,并对吉林省制定经济决策提出建议。王芳芳和王倩[7]通过分析陕西省 2000~2018 年

GDP 历史数据, 利用 SPSS 软件建立 ARIMA(5,2,1)模型, 对 2019~2021 年的 GDP 值进行预测, 拟合度较好。郑梦琪和朱家明[8]根据浙江省 1978~2018 年 GDP 数据建立 ARIMA(5,1,5)模型对浙江省未来三年的 GDP 指数进行预测。严彦文[9]根据山东省 1975~2015 年 GDP 数据, 运用统计学原理, 建立 ARIMA(1,1,1)模型对 2016~2020 年 GDP 进行了预测。

3. ARIMA 模型介绍

ARIMA 模型全称为差分自回归移动平均(Auto Regressive Integrated Moving Average Model)模型, 是由自回归模型(Auto Regression, AR)、移动平均模型(Moving Average, MA)和差分法结合而来的时间序列预测模型, 常用于非平稳时间序列的分析和预测。其中, AR 部分用于处理时间序列的自回归部分, 它考虑了过去若干时期的观测值对当前值的影响; MA 部分用于处理时间序列的移动平均部分, 它考虑了过去的预测误差对当前值的影响。ARIMA 模型表达式为 ARIMA(p,d,q), 其中 p 是自回归项数, q 是移动平均项数, d 是差分次数。当 $d = q = 0$ 时, ARIMA 模型即为 AR 模型; 当 $p = d = 0$ 时, ARIMA 模型即为 MA 模型。本文中 ARIMA 模型的建模步骤包括: ① 平稳性检验。对选取的上海市 1991~2022 年 GDP 时间序列数据进行绘图和单位根检验, 结果表明为非平稳序列。因此, 对原序列进行取对数和一阶差分处理, 使其符合平稳时间序列特征。② 选定参数 p, d, q 并拟合模型。通过绘制自相关 ACF 图和偏自相关 PACF 图, 初步判断 $p = 1, q = 0$ 。为了提高精度, 进一步建立多个 ARIMA 模型进行拟合并根据 AIC 和 BIC 准则确定最优模型为 ARIMA(1,1,2)。③ 模型检验。通过绘制残差序列 Q-Q 图和白噪声检验, 保证模型的残差满足独立正态分布, 不具有序列相关性。④ 模型预测。通过最优模型 ARIMA(1,1,2)对上海市 2023~2025 年 GDP 进行预测分析, 作为政策建议的理论依据。

4. 基于 ARIMA 模型的实证分析

4.1. 数据来源与初步分析

为保证数据分析的准确性, 从《上海统计年鉴》中选取上海市 1990~2022 年 GDP 年度统计数据作为原始时间序列, 使用 R 软件对数据进行预处理, 画出时间序列图。由图 1 可以看出, 1991~2022 年上海市 GDP 呈现指数型增长趋势, 其中 2019~2020 年受疫情影响, 经济增长速度减缓, 2020~2021 年经济形势快速反弹, 持续恢复。2022 年上海市 GDP 总值为 44652.8 亿元, 是 1991 年的 50 倍左右, 经济总量实现了巨大飞跃。

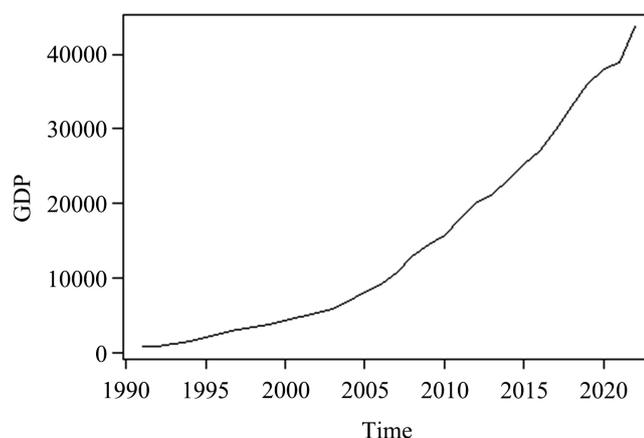


Figure 1. Timing chart of GDP in Shanghai from 1991 to 2022

图 1. 1991~2022 年上海市 GDP 时间序列图

4.2. 平稳性检验

上海市 GDP 值呈现指数型增长, 未出现周期性和季节性的波动, 可以初步判断为非平稳时间序列。而 ARIMA 模型要求时间序列是平稳的, 所以进一步对原始序列进行 ADF 单位根检验, 结果如表 1 所示。P = 0.99, $p > 0.05$, 因此可以判断原始序列为非平稳时间序列。

Table 1. ADF unit root test results

表 1. ADF 单位根检验结果

Augmented Dickey-Fuller Test		
Data: GDP		
Dicler-Fuller = 0.56689	Lag order = 3	p-value = 0.99
Alternative hypothesis: stationary		

针对呈现指数型增长趋势的非平稳时间序列, 首先利用 R 软件中 log 函数取对数得到 lnGDP 以消除指数趋势, 然后利用 diff 函数对 lnGDP 进行一阶差分处理, 得到差分后的 lnGDP 折线图(见图 2), 可以看出差分后的时间序列围绕固定值上下波动。同时, 单位根检验结果显示 $p = 0.0368 < 0.05$, 有充分理由拒绝原假设, 表明一阶差分后的序列符合平稳时间序列特征。

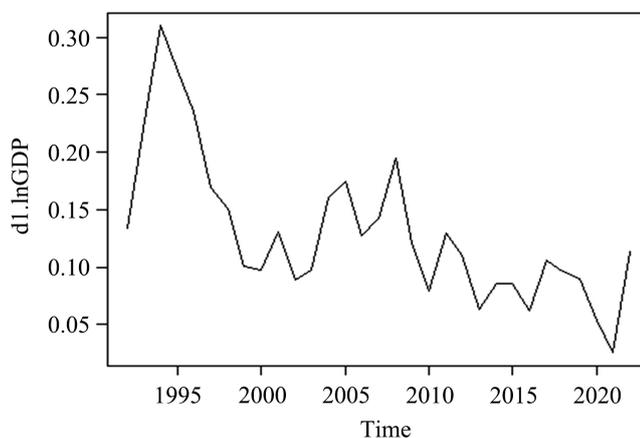


Figure 2. Line chart for stage I differential lnGDP

图 2. 一阶差分 lnGDP 折线图

4.3. ARIMA 模型定阶

考虑到原始数据通过取对数和一阶差分处理后变为平稳序列, 故选择 ARIMA(p,d,q)模型进行分析与预测, 其中 $d = 1$ 。

利用 R 软件中 acf 和 pacf 函数绘制出一阶差分后的平稳序列的自相关 ACF 图(如图 3)和偏自相关 PACF 图(如图 4), 通过观察图形变化初步确定 p 值和 q 值。

通过自相关 ACF 图可知, 自相关系数显示拖尾; 通过偏自相关 PACF 图可知, 偏自相关系数在一阶后迅速减小, 呈现截尾状态, 所以可以初步判断 $p = 1$, $q = 0$ 。为了保证 p, q 的选择更加准确合理, 同时建立多个 ARIMA 模型进行拟合, 结果如表 2 所示。根据 AIC 准则和 BIC 准则(越低越好)可以发现, 模型 ARIMA(1,1,2)的两项指标均出现最低值, 因此可选作最优模型。

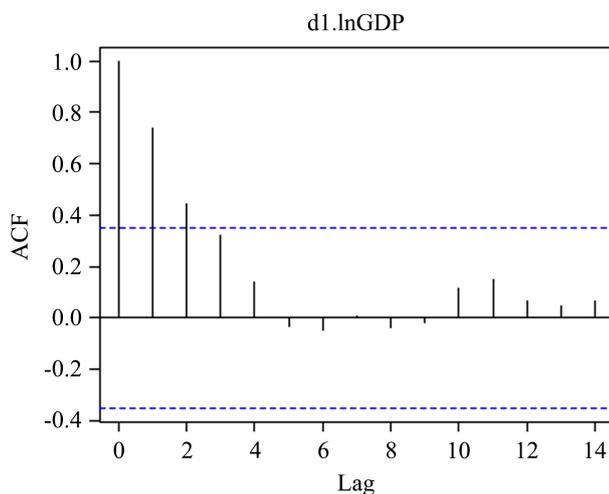


Figure 3. Autocorrelation graph after stage I difference
图 3. 一阶差分后的自相关图

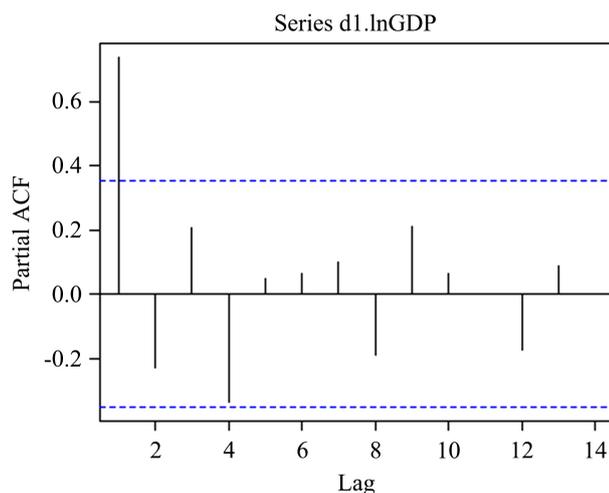


Figure 4. Self-autocorrelation graph after stage I difference
图 4. 一阶差分后的偏自相关图

Table 2. Comparison of ARIMA models
表 2. ARIMA 模型比较

	ARIMA (1,1,0)	ARIMA (1,1,1)	ARIMA (1,1,2)	ARIMA (2,1,0)	ARIMA (2,1,1)	ARIMA (2,1,2)	ARIMA (3,1,0)	ARIMA (3,1,1)	ARIMA (3,1,2)
AIC	-97.77	-97.05	-103.14	-95.96	-99.55	-101.14	-98.45	-100.82	-98.97
BIC	-94.91	-92.75	-97.41	-91.66	-93.81	-93.98	-92.72	-93.65	-90.37

4.4. ARIMA 模型检验

4.4.1. 拟合优度

利用 forecast 包里的 accuracy()函数对模型进行拟合优度的度量(见表 3), 结果显示该模型拟合较为准确。

Table 3. Measurement results of GFI

表 3. 拟合优度量度结果

	ME	RMSE	MAE	MPE	MAPE	MASE
Training set	-0.0012	0.0360	0.0272	0.0012	0.3091	0.2099

4.4.2. 残差白噪声检验

一般来说, 合适的模型的残差应该满足均值为 0 的正态分布, 并且对于任意的滞后阶数, 残差自相关都应该为 0。本文通过绘制残差序列的 Q-Q 图(见图 5)可以观察到残差近似的落在一条直线上, 可以初步判断该序列是随机的正态分布。

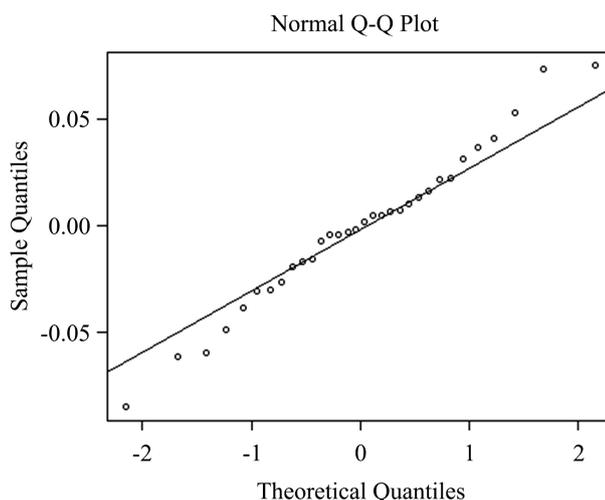


Figure 5. Q-Q Graph of residual sequences

图 5. 残差序列的 Q-Q 图

为了得出更准确的结论, 利用 R 软件中 `Box.test` 函数进一步开展白噪声检验, 如果统计量的 P 值大于给定的显著性水平, 则模型的残差没有通过显著性检验, 即我们认为残差的自相关系数为零。检验结果如表 4 显示: $p = 0.5164$, $p > 0.05$, 说明 ARIMA(1,1,2)模型的残差序列为白噪声序列。因此可以判定 ARIMA(1,1,2)合理有效, 几乎可以反映所有有用的信息, 无需对模型进行优化, 可以作为理想的预测模型。

Table 4. Test results of white noise

表 4. 白噪声检验结果

X-squared	df	P-value
0.4211	1	0.5164

4.5. ARIMA 模型结果分析

经过上述分析, 对上海市 GDP 时间序列数据建立的 ARIMA(1,1,2)模型通过了各项检验, 能较好地拟合数据, 最终模型如表 5 所示。

Table 5. ARIMA(1,1,2) model
表 5. ARIMA(1,1,2)模型

	ar1	ma1	ma2
	0.9929	0.2433	-0.7567
s.e.	0.0123	0.2151	0.1860

sigma² estimated as 0.001336: log likelihood = 55.57 aic = -103.14

4.6. ARIMA 模型预测

经过上述分析, 对上海市 GDP 时间序列数据建立的 ARIMA(1,1,2)模型通过了各项检验, 能较好地拟合数据, 最终模型如表 5 所示。利用 R 软件中 forecast 函数对上海市 2023~2025 年 GDP 进行预测, 并绘制 95%置信水平下的预测图形, 结果如表 6 所示。

Table 6. Shanghai's GDP forecast for 2023~2025 and its 95% confidence interval
表 6. 上海市 2023~2025 年 GDP 预测值及其 95%置信区间

时间	GDP 预测值(亿元)	95%置信区间	
		Lo 95	Hi 95
2023	49702.47	46214.55	53453.64
2024	54362.51	45580.71	64836.24
2025	59422.01	45716.29	77236.69

从预测结果来看, 模型短期预测效果良好, 预测结果显示上海市未来 3 年的 GDP 依次为 4.97 万亿元、5.44 万亿元、5.94 万亿元。由此可见, 上海市 GDP 在未来将持续呈现稳步增长的态势和较高的增长速度, 且有望在 2024 年突破 5 万亿元。上海市经济回稳向好态势不断巩固, 有助于提振社会预期和发展信心, 未来上海市也将继续发挥引领带动作用推动我国经济持续向好发展。

4.7. 模型评价

1) ARIMA 模型中的参数可以反映时间序列的特定性质。因此, 可以通过对模型参数的解释, 更好地了解时间序列数据的特征。

2) ARIMA 模型只需要考虑序列自身内在规律, 而不需要借助于其他外生变量。

3) 模型构建的基础是时间序列数据通过平稳性检验, 若不满足要求, 则需要使用取对数和差分等方法对数据进行处理使之通过检验。

4) ARIMA 模型只能针对线性时间序列数据进行建模, 并不能捕捉非线性关系。若 ARIMA 模型序列变量的未来取值并不符合先前观察值和随机误差间的线性函数关系, 则就不可以构建该模型, 需要采用其他更合适的方法。

5) 因为采用了数学模型的限制条件, 所以 ARIMA 模型对于噪声的影响比较敏感, 容易造成较大误差。

6) 模型的预测结果无法考虑未来重大突发事件对 GDP 发展的影响, 可能会导致预测值出现较大偏差, 影响模型预测效果。

5. 结论与建议

本研究通过选取上海市 1991~2022 年 GDP 年度统计数据, 利用 R 软件构建出 ARIMA(1,1,2)模型, 并对未来 3 年上海市的 GDP 值做出预测。研究结果表明, 2023~2025 年上海市 GDP 呈现出稳步增长的趋势, 其中 2023 年上海市 GDP 预测值为 49702.47 亿元, 增速约 11%, 并预计上海市 GDP 将于 2024 年突破五万亿元大关。

2023 年以来, 疫情对企业经营的影响明显减弱甚至基本消除, 但是未来上海经济发展依然面临重大挑战。为实现“十四五”规划、市第十二次党代会和新一届政府提出的目标, 现基于预测模型所展示出的经济发展态势, 对上海市经济建设提出以下政策建议:

一是落实国家重大战略任务, 提升经济首位度和引领带动效应。上海作为我国经济实力最强的城市, 应当全面深化浦东引领区建设, 持续深入实施“三大任务”, 着力提升“五个中心”能级, 加快打造双向开放“三大平台”, 加快建设具有世界影响力的社会主义现代化国际大都市, 打造具有世界影响力的核心功能, 在新征程上继续当好发展的先行者[10]。

二是推动城市数字化转型, 打造国际数字之都。面对数字化转型新机遇, 上海市应当加快发展数字经济核心产业和基础设施建设, 打造高端数字产业集群和数字赋能体系, 推进 5G 网络建设和深度覆盖。推动构建精准、普惠的数字生活服务体系, 从医疗、教育、交通等角度丰富生活数字化场景应用。聚焦“高效办成一件事”, 完善城市运行数字体征系统, 提高治理数字化应用时效。

三是实施创新驱动发展战略, 建设现代化产业体系。通过推进在沪国家实验室高水平运行, 加快高能级创新平台和新型研发机构建设布局, 进一步改善创新创业环境, 来更大程度的提升科技创新策源能力。同时, 着力打造高端制造业增长极, 促进服务业高质量发展, 加快培育新兴产业集群, 推进高水平人才高地建设。

四是鼓励绿色低碳转型, 树立美丽上海新风貌。在推进落实碳达峰碳中和的背景下, 上海应当积极相应国家号召, 打造上海碳中和产业园, 坚决遏制高耗能、高排放、低水平项目盲目发展, 建立和完善碳普惠体系, 倡导绿色低碳生活新风貌。

五是优化疫情防控工作, 打造安全韧性城市。准确把握疫情防控新阶段新任务, 持续加强医疗资源储备, 保障群众及时的就医用药。加大科技投入, 加快创新药研发转化, 为重大公共卫生事件提供科技支撑。此外, 切实保障能源和粮食安全, 防范化解金融风险, 做好各项安全保障工作, 为经济发展提供坚实基础。

参考文献

- [1] 肖丹. 基于 ARIMA 模型的四川省 GDP 分析与预测[J]. 生产力研究, 2023(10): 62-66.
- [2] 李振亮, 乐昕雨. 基于 ARIMA 模型的北京市 GDP 分析与预测[J]. 中小企业管理与科技, 2023(1): 153-155.
- [3] 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.
- [4] 张梓. 基于 ARIMA 模型的贵州省 GDP 分析与预测[J]. 国土与自然资源研究, 2022(5): 39-41.
- [5] 夏如玉, 王梓桥. 基于 ARIMA 模型对重庆市 GDP 预测分析[J]. 中国储运, 2022(8): 93-94.
- [6] 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.
- [7] 王芳芳, 王倩. 基于 ARIMA 模型的陕西省 GDP 分析及预测研究[J]. 产业创新研究, 2020(19): 22-24.
- [8] 郑梦琪, 朱家明. 基于 ARIMA 模型对浙江省 GDP 预测分析[J]. 哈尔滨师范大学自然科学学报, 2020, 36(3): 56-61.
- [9] 严彦文. 基于 ARIMA 模型的山东省 GDP 的分析与预测[J]. 数学的实践与认识, 2018, 48(4): 285-292.
- [10] 邓智团. 进一步提升上海经济首位度研究[J]. 科学发展, 2023(5): 14-22.