

天津工业大学

本科生毕业论文

基于时间序列的影响居民消费价格指数组成要素分析

学 号：2010650410

姓 名：于劭

专 业：信息管理与信息系统

学 院：经济与管理学院

指导教师：吴中元

职 称：教授

完成日期：2024 年 6 月

诚信声明

本人郑重声明：

本人所呈交的毕业论文，是在论文指导教师的指导下，独立进行研究工作所取得的成果。论文由本人独立完成，除文中已经注明引用的内容外，本论文不含任何其他个人或集体已经发表或撰写过的作品成果。本人完全了解本声明的法律结果由本人承担。

毕业论文作者签名：


于劭

日

期： 2023 年 12 月 1 日

天津工业大学 2024 届本科毕业设计（论文）任务书

题目	基于时间序列的影响居民消费价格指数的组成因素的分析				
学生姓名	于劭	学号	2010650410	专业班级	信息 H2004 班
学院名称	经济与管理学院			课题类型	论文类
课题意义	<p>CPI 作为反映居民购买并用于消费商品和服务项目价格水平的变化趋势和变动幅度的统计指标，对经济生活的各个方面都有重要的影响。因此，理解 CPI 及其组成部分的动态变化对于制定经济政策具有重要的参考价值。例如，当 CPI 持续较快上涨时，政府可能需要分析上涨原因,通过调控货币政策或财政政策来稳定某几类商品的价格。该课题旨在准确把握居民消费价格指数的变动趋势，为经济政策制定提供参考依据。通过分析 CPI 的组成部分，可以更好地理解物价水平的变化规律，从而为政府和企业提供决策支持。</p>				
任务要求	<p>1. 2023. 10. 20-2023. 12. 4 确定指导老师，学生选定设计（论文）题目 2. 2023. 12. 5-2024. 1. 5 教师指导学生查阅文献、撰写开题报告和设计（论文）提纲，完成设计（论文）开题答辩 3. 2024. 1. 6-2024. 3. 14 开始撰写论文，提交设计（论文）初稿,完成中期检查 4. 2024. 3. 15-2024. 5. 12 在导师指导下修改设计（论文），完成查重 5. 2024. 5. 13-2024. 5. 25 毕业设计（论文）抽检 6. 2024. 5. 26-2024. 6. 6 参加答辩</p>				
推荐参考资料	<p>[1] 王振中, 陈松蹊, and 涂云东. "中国居民消费价格指数的动态结构研究及中美量化比较." 数理统计与管理 42. 01 (2023): 109-126. doi: 10.13860/j.cnki.sltj.20211130-018.</p> <p>[2] 栾惠德. "居民消费价格指数的实时监测——基于季节调整的方法." 经济科学 . 02 (2007): 59-67. doi: 10.19523/j.jjxx.2007.02.006.</p> <p>[3] 李隆玲, 田甜, and 武拉平. "中国居民消费价格指数预测." 农业展望 10. 07 (2014): 75-79.</p> <p>[4] 曹晓俞. "居民消费价格指数的时间序列模型分析." 华北金融 . 07 (2012): 8-10.</p>				
起止日期	2023. 10-2024. 6				
备注					

指导教师 


教研室主任 _____

主管院长 _____

天津工业大学 2024 届本科毕业设计（论文）开题报告表

学院：经济与管理学院

2023 年 12 月 4 日

姓名	于劭	学号	2010650410	专业班级	信息 H2004 班
题目	基于时间序列的影响居民消费价格指数的组成因素的分析			指导教师	吴中元
<p>一、与本课题有关的国内外研究情况、课题的目的、意义及主要内容：</p> <p>（一）与本课题有关的国内外研究情况：在国内外，有研究者对我国的居民消费价格指数数据进行了建模预测，以准确把握 CPI 的变动趋势。此外，还有实证分析研究了我国居民消费价格指数（CPI）的影响因素，指出消费者价格指数是一个主要的统计指标，反映一个国家或地区的宏观经济形势，可以反映不同时期的生活成本的变化。也有文章利用时间序列的趋势外推方法对某个国家目前居民消费价格指数（CPI）进行了建模分析和预测。</p> <p>（二）目的和意义：该课题旨在准确把握居民消费价格指数的变动趋势，为经济政策制定提供参考依据。通过分析 CPI 的组成部分，可以更好地理解物价水平的变化规律，从而为政府和企业提供决策支持。</p> <p>（三）课题研究的主要内容：1. 选取关键的 CPI 组成部分，如食品类、医疗保健类、烟酒类、交通通讯类等。2. 运用时间序列分析方法，描述历史数据随时间变化的规律。3. 建立预测模型，预测未来一段时间内这些组成部分的价格指数变化趋势。4. 结合宏观经济背景，分析这些组成部分价格指数变化对整体 CPI 的影响。</p>					
<p>二、进度及预期结果：</p>					
起止日期		主要内容		预期结果	
2023.10.20-2023.12.4		确定指导老师，学生选定论文题目。		选定论文题目	
2023.12.5-2024.1.5		教师指导学生查阅文献、撰写开题报告和论文提纲		完成论文开题答辩	
2024.1.6-2024.3.14		开始撰写论文，提交论文初稿		完成论文初稿	
2024.3.15-2024.5.11		在导师指导下修改论文，提交查重，完成定稿		完成定稿	
2024.5.12-2024.5.25		毕业论文抽检		完成论文抽检	
2024.5.26-2024.6.6		论文答辩		参加答辩	
完成课题的现有条件		1. 指导教师指导论文。 2. 利用图书馆查阅资料。 3. 通过互联网进行资料搜索、查找。 4. 查阅相关杂志、期刊等资料。			
审查意见		指导教师：  年 月 日			
学院意见		主管领导： 年 月 日			

摘 要

在国内外,有研究者对我国的居民消费价格指数数据进行了建模预测,以准确把握 CPI 的变动趋势。此外,还有实证分析研究了我国居民消费价格指数 (CPI)的影响因素,指出消费者价格指数是一个主要的统计指标,反映一个国家或地区的宏观经济形势和不同时期的生活成本的变化。也有文章利用时间序列的趋势外推方法对某个国家目前居民消费价格指数 (CPI)进行了建模分析和预测,CPI 作为反映居民购买并用于消费商品和服务项目价格水平的变化趋势和变动幅度的统计指标,对经济生活的各个方面都有重要的影响。因此,理解 CPI 及其组成部分的动态变化对于制定经济政策具有重要的参考价值。例如,当 CPI 持续较快上涨时,政府可能需要分析上涨原因、通过调控货币政策或财政政策来稳定某几类商品的价格。文章主要通过选取关键的 CPI 组成部分的商品及服务种类,如食品类、医疗保健类、烟酒类、交通通讯类等。运用时间序列分析方法,描述历史数据随时间变化的规律。并建立预测模型,预测未来一段时间内这些组成部分的价格指数变化趋势。结合宏观经济背景,分析这些组成部分价格指数变化对整体 CPI 的影响。该课题旨在提供一种新的预测验证的方法,以准确把握居民消费价格指数的变动趋势,为经济政策制定提供参考依据。通过分析 CPI 的组成部分,可以更好地理解物价水平的变化规律,从而为政府和企业提供决策支持。

关键词: CPI; 商品及服务种类; 时间序列分析; 组成要素; 变化趋势;

ABSTRACT

At home and abroad, researchers have modeled and predicted the consumer price index data of residents in China to accurately grasp the trend of CPI changes. In addition, empirical analysis has been conducted on the influencing factors of China's Consumer Price Index (CPI), pointing out that CPI is a major statistical indicator that reflects the macroeconomic situation of a country or region and changes in the cost of living at different times. Some articles have also used the trend extrapolation method of time series to model, analyze, and predict the current Consumer Price Index (CPI) of a country. CPI, as a statistical indicator reflecting the trend and magnitude of changes in the price level of goods and services purchased and used by residents, has an important impact on various aspects of economic life. Therefore, understanding the dynamic changes of CPI and its components is of great reference value for formulating economic policies. For example, when the CPI continues to rise rapidly, the government may need to analyze the reasons for the increase and stabilize the prices of certain commodities by regulating monetary or fiscal policies. The article mainly selects the types of goods and services that are key components of CPI, such as food, healthcare, tobacco and alcohol, transportation and communication, etc. Use time series analysis methods to describe the patterns of historical data changes over time. And establish a predictive model to predict the trend of price index changes of these components in the future. Based on the macroeconomic background, analyze the impact of price index changes in these components on the overall CPI. This project aims to accurately grasp the trend of changes in the consumer price index of residents and provide a reference basis for economic policy formulation. By analyzing the components of CPI, we can better understand the changing patterns of price levels, thereby providing decision-making support for the government and enterprises.

Keywords: CPI; Types of goods and services; Time series analysis; Constituent elements; Changing trend;

目 录

第一章 绪论	1
1.1 研究背景	1
1.2 研究意义	2
1.3 研究目的	2
1.4 研究方法	3
第二章 相关概念	4
2.1 CPI 的定义和计算	4
2.2 CPI 组成要素的商品种类和选取	4
第三章 国内各种消费价格指数的数据分析	7
3.1 软件及分析工具的选取	7
3.2 各种消费价格指数的数据选取	7
3.3 不同因素对居民消费价格指数的影响分析	8
3.4 拟合参数计算	27
3.5 结果分析与建议	27
第四章 总结	29
参考文献	30
附录	32
谢 辞	35

第一章 绪论

1.1 研究背景

居民消费价格指数（Consumer Price Index，简称 CPI）是衡量一篮子消费品和服务价格随时间变化的指标，它反映了居民购买的消费品和服务的价格水平变动情况。CPI 的历史背景和作为经济指标的重要性可以从以下几个方面进行详细描述：

历史背景：CPI 的概念可以追溯到 19 世纪末，当时为了研究工人生活成本的变化，一些国家开始编制初步的价格指数。20 世纪初，随着工业化进程的加快和城市化的发展，对价格变动的关注逐渐增加，CPI 开始被用作衡量通货膨胀的重要工具。第一次世界大战和第二次世界大战期间，CPI 在经济管理中的作用显著增强，成为政府制定价格控制政策和工资政策的重要依据^[1]。战后，随着经济全球化和市场经济的发展，CPI 成为衡量经济稳定性和预测经济走势的关键指标。20 世纪中叶，随着国际组织如联合国和国际劳工委员会（ICP）的成立，CPI 的编制方法开始标准化，以确保不同国家之间的可比性。各国统计局根据国际标准和本国实际情况，制定了各自的 CPI 计算方法和权重体系。

同时，CPI 作为经济指标也有相当大的重要性，CPI 可以对通货膨胀进行衡量：CPI 是衡量通货膨胀的主要指标之一。当 CPI 持续上升时，表明物价水平上涨，购买力下降，固定收入者（如退休人员）的实际购买力会下降，而投资者和债务人可能会从通货膨胀中获益或受损，取决于债务的利率和期限，这可能影响居民的生活水平和经济稳定。CPI 可以影响经济政策的制定：政府和中央银行使用 CPI 来制定和调整货币政策和财政政策。例如，中央银行可能会通过调整利率来控制通货膨胀，保持物价稳定，这也可能进一步影响经济增长速度和就业水平。CPI 可以影响居民的收入和福利：CPI 的变化直接影响工资、退休金、福利金等的调整^[2]。许多劳动合同和社会保障计划都与 CPI 挂钩，以保护收入不受通货膨胀的侵蚀，如果 CPI 的上升速度超过工资增长速度，可能会引发社会不满和抗议活动。政府需要通过社会保障和补贴政策来缓解这种压力，维持社会稳定。CPI 可以影响投资和储蓄的决策：投资者和储蓄者关注 CPI，因为它影响实际回报率和购买力。高通货膨胀可能导致投资者寻求更高回报的投资渠道来对冲，比如房地产和黄金，而储蓄者可能会寻找保值手段^[3]。CPI 可以在国际上进行科学的比较：CPI 是进行国际经济比较的重要工具。通过比较不同国家的 CPI，可以了解各国物价水平和生活成本的差异。CPI 可以评估经济政策预估经济形势：经济学家和研究人员使用 CPI 来分析经济趋势、评估经济政策的效果以及研究消费者行为的变化。

总之，CPI 不仅是一个反映价格变动的统计指标，更是一个影响经济政策、指导个人和企业决策、以及进行国际经济分析的关键经济指标。它的准确性和可靠性对于经济的健康发展至关重要。

1.2 研究意义

本文的研究对于政府制定宏观经济政策具有重要的指导意义，主要体现在以下几个方面：

1. 通货膨胀监测与控制：通过对 CPI 组成要素的深入分析，政府能够更准确地监测通货膨胀的风险，识别价格上涨的主要驱动因素，并据此制定相应的货币政策和财政政策举措，如调整利率或增加公共支出，以稳定物价水平。

2. 社会福利政策调整：研究结果能够帮助政府评估生活成本的变化对不同收入群体的影响，特别是对低收入和固定收入家庭的影响。基于这些信息，政府可以调整社会福利政策，如提高最低工资标准、调整退休金和补贴，以保障民生。

3. 经济预测与规划：CPI 的变化趋势对于经济预测至关重要。政府可以利用本研究的预测模型和分析结果，进行宏观经济预测和长期规划，为国家的经济发展制定战略和政策提供数据支持。

同时，本文的研究对企业经营决策和居民消费行为也有显著影响：

1. 企业战略规划：企业可以根据 CPI 的变化趋势和组成要素分析，预测未来成本和需求的变化，从而在产品定价、库存管理、市场策略等方面做出更合理的决策。

2. 居民消费选择：居民可以根据 CPI 的研究结果，更好地理解价格变动的趋势，调整个人和家庭的消费模式和储蓄计划。例如，居民可能会选择在价格上涨前购买某些商品，或者寻找价格更稳定的替代品。

还有，本文的研究对于学术界深入理解 CPI 组成要素动态变化有所贡献，主要体现在：

1. 理论发展：研究提供了新的视角和方法来分析 CPI 的组成要素和动态变化，有助于丰富和完善现有的价格指数理论。

2. 方法论创新：通过运用时间序列分析等统计方法，引入新的分析思路，区别于直接对 CPI 进行建模预测，而是先对其各个组成要素进行建模再对它们的预测值加权求和得到 CPI 的预测结果，并与直接建模方法得到的预测值相比对，以提高 CPI 分析的精确性和预测的可信度。

3. 跨学科研究促进：CPI 的研究涉及经济学、统计学、社会学等多个学科领域。本研究的成果可以促进不同学科之间的交流和合作，推动跨学科研究的发展。

4. 教育与培训：研究成果可以作为教育资源，用于高等教育和专业培训中，提高未来经济学家和政策制定者对 CPI 及其影响的理解。

综上所述，本论文的研究不仅对政府制定宏观经济政策具有重要的指导意义，也对企业经营决策和居民消费行为产生直接影响，同时为学术界提供了深入理解 CPI 组成要素动态变化的新视角和方法。

1.3 研究目的

本文的主要目的是探究在仅研究 CPI 组成要素的情况下能否做到精确把握 CPI 的变化

趋势并找到某段时间内影响价格指数的主要驱动因素从而帮助政府避免风险并对经济发展状况进行预测，进行社会福利政策的调整，帮助企业精准识别市场需求、合理安排生产计划，帮助居民调整个人和家庭的消费模式和储蓄计划。为了达到这个目的，就需要对第三章中拟合得到的模型的准确度进行检验，具体做法是将预测方法分为两组，对照组是直接对 CPI 数据进行时间序列分析和预测，实验组是分别对组成 CPI 的八个分类进行时间序列分析和预测，之后在通过分类的预测值求得总的 CPI 的预测值，如果第二种方法可行，那么其与第一种方法得到的结果在往后几期中应当十分接近。

1.4 研究方法

时间序列分析是统计学中的一种方法，用于分析按时间顺序排列的数据点。在本论文中，主要采用了以下几种时间序列分析方法：

1. ARIMA 模型（自回归积分滑动平均模型）：ARIMA 模型结合了自回归（AR）模型、差分（I）操作和滑动平均（MA）模型。它能够处理非平稳时间序列数据，通过差分操作使数据平稳。AR 部分表示当前值与其前面的值之间的关系；I 部分是指数据需要经过几次差分才能变得平稳；MA 部分表示当前值与前面误差项之间的关系。选择合适的 ARIMA 模型需要确定三个参数（ p, d, q ），其中 p 是自回归项的阶数， d 是差分次数， q 是滑动平均项的阶数^[4]。

2. 季节性分解：季节性分解是将时间序列分解为趋势、季节性和随机成分的过程。这通常通过季节性调整来实现，目的是从时间序列中移除季节性效应，以便更清楚地看到趋势和其他成分。常用的季节性分解方法包括经典分解和 STL（Seasonal and Trend decomposition using Loess）分解^[4]。

模型选择的依据主要来自以下几个方面：

1. 数据特性：首先，对时间序列数据进行可视化和初步分析，以识别数据的趋势、季节性和周期性等特性。使用 ACF（自相关函数）和 PACF（偏自相关函数）图来确定 AR 和 MA 模型的阶数。

2. 模型诊断：对拟合的模型进行诊断，检查残差是否符合白噪声的假设，即残差之间不存在相关性。使用信息准则（如 AIC 或 BIC）来比较不同模型的拟合优度。

3. 预测能力：通过交叉验证或保留一部分数据作为测试集，评估模型的预测能力。选择在验证集上具有最低预测误差的模型。

模型验证的方法为残差分析，检查残差的分布情况，确保残差服从正态分布且具有恒定的方差（同方差性）。通过残差图来识别任何可能的模式或结构，这可能表明模型未能捕捉到数据的所有特征。

通过上述方法，可以确保所选的时间序列模型能够有效地捕捉数据的动态特征，并对未来的值进行准确的预测。在本论文中，这些方法将被用来分析和预测 CPI 的组成要素，为经济政策制定提供科学依据。

第二章 相关概念

2.1 CPI 的定义和计算

CPI 的定义：居民消费价格指数（Consumer Price Index, CPI）是衡量一段时间内居民购买的一篮子消费品和服务价格水平变化的统计指标^[5]。CPI 旨在反映居民消费商品和服务的价格变动情况，是衡量通货膨胀或通货紧缩的关键经济指标之一。CPI 的变化直接影响到居民的实际购买力和生活成本，同时也是制定宏观经济政策的重要依据。

CPI 的计算涉及以下几个关键步骤：

1. 确定基期：选择一个时间点作为参照基准，通常是过去的某个特定年份或时期。
2. 构建商品篮子：根据居民消费调查，选择一组代表性的商品和服务，这些商品和服务应该能够全面反映居民的消费习惯。
3. 分配权重：根据各类商品和服务在居民总消费支出中的比例分配权重。权重通常基于家庭消费调查数据，这些数据反映了不同收入层次家庭的消费模式。
4. 收集价格数据：定期（如每月）收集选定商品和服务的价格信息。这些价格数据可以通过访问商场、超市、农贸市场等传统渠道，或者通过互联网电商平台来获取。
5. 计算价格指数：使用加权平均的方法计算价格指数。CPI 通常使用以下公式计算：
$$CPI = (\sum(\text{当前时期价格} \times \text{对应权重}) / \sum(\text{基期价格} \times \text{对应权重})) \times 100$$
，其中， \sum 表示求和，当前时期价格和基期价格分别对应于当前和基期的商品篮子价格^[6]。

尽管 CPI 是一个重要的经济指标，但它也存在一些局限性：

新产品和服务的引入：CPI 可能无法及时反映新产品和服务的引入对价格水平的影响。

质量变化：CPI 通常不考虑产品质量的改进或下降对价格的影响。

替代效应：CPI 不考虑消费者在面对价格变化时可能选择替代品的行为。

通常，我们可以采取以下方法来改进 CPI 的计算：

1. 更新商品篮子：定期更新商品篮子，以反映消费者偏好的变化和技术进步。
2. 考虑质量变化：开发质量调整的价格指数，如使用价格指数来衡量同一产品的质量变化。
3. 考虑替代效应：使用更复杂的模型来捕捉消费者的替代行为，如使用近乎理想需求系统（AIDS）模型^[7]。

2.2 CPI 组成要素的商品种类和选取

CPI 的组成要素通常包括数百种不同的商品和服务，按照其在居民消费中的比重分为不同的大类，如食品烟酒类、衣着类、居住、医疗保健类、生活用品及服务类、教育文化和娱乐类、其他用品和服务类。如今，按照统计制度规定，我国 CPI 每五年进行一次基期轮换。每次基期轮换后，调查分类目录、代表规格品和调查网点均有调整，分类权数也有变化，以反映居民消费结构的最新变动^[8]。2016 年 1 月-2020 年 12 月编制和发布的是以 2015

年为基期的 CPI。与前几轮基期相比，此轮基期 8 大类有显著变化，其中“食品”、“烟酒”合并为“食品烟酒”，“医疗保健和个人用品”拆分至“生活用品及服务”、“医疗保健”和“其他用品和服务”中，“娱乐教育文化用品及服务”拆分至“教育文化和娱乐”、“其他用品和服务”中，“家庭设备用品及维修服务”拆分至“生活用品及服务”和“其他用品及服务”中。此外，“食品”的指标内涵发生了变化，前几轮基期中的“食品”为大类，包括粮食、肉禽、鲜菜、鲜果、水产品、茶及饮料、在外餐饮等分类；此轮基期的“食品”为“食品烟酒”大类下的中类，仅包括粮食、畜肉、禽肉、鲜菜、鲜果、水产品等，不再包括“茶及饮料”和“在外餐饮”两项。2021 年 1 月开始编制和发布以 2020 年为基期的 CPI。本轮基期仍分为食品烟酒、衣着、居住、生活用品及服务、交通通信、教育文化娱乐、医疗保健、其他用品及服务 8 个大类，基本分类增加至 268 个。本文选取的数据来自于 2021 年到 2024 年 2 月，在新五年基期轮换、调查分类目录进行拆分合并以及相应权重的大背景下对 CPI 及其组成的八大类商品服务的价格指数进行尽可能具有时效性的研究^[8]。

其中，每个组成要素的权重和对 CPI 变动的贡献是根据居民消费支出调查数据确定的。例如，食品烟酒通常占有较高的权重，因为它们是居民日常消费的重要组成部分。居住和医疗保健也是重要的组成要素，因为它们在居民总消费中占据较大比重，并且价格变动对居民生活成本的影响显著。

选择特定商品和服务种类的原因通常基于它们的消费普遍性和代表性。选取的商品和服务应当能够全面反映居民的实际消费模式。此外，选取的商品和服务种类应当能够覆盖不同收入层次的家庭，以确保 CPI 的广泛适用性^[9]。不同商品和服务种类在 CPI 中的重要性不同，这取决于它们在居民总消费中的比重以及价格变动的敏感性。某些商品和服务可能对经济政策更为敏感，因此在分析和预测 CPI 变动时需要特别关注。例如，能源价格的波动可能会对交通和居住成本产生较大影响，从而对整体 CPI 产生显著影响^[9]。

同时，可以通过国内相关文献的研究工作查看商品种类选取的要点：

王双进,崔福鑫,孙淑敏在基于 CPI 数据的我国价格总水平分析一文，从不同层面分析 2023 年 CPI 运行特征及背后原因；接着，从粮食价格稳中有升、生猪价格低位回升、上下游产品价格传导、翘尾因素、国际市场价格影响等五个方面，分析研判 2024 年 CPI 走势变化。研究发现：（1）2023 年 CPI 涨幅比上年明显缩小，CPI 八大类价格“六涨一平一降”，猪牛羊肉价格降幅扩大，旅游价格涨幅扩大；（2）粮食价格稳中有升、生猪价格低位回升、国际市场价格波动，是影响 2024 年 CPI 上涨的主要因素；（3）上游产品价格持续下跌减缓下游产品价格上涨压力、翘尾因素，是影响 2024 年 CPI 下降的主要影响因素。针对 CPI 运行走势及其影响因素，从财政货币政策协调配合、民生商品保供稳价、改善房地产市场调控、密切关注全球经济形势等方面提出政策建议^[1]。

邓海云,张席,杨钧博,刘清,尹佳斌在我国城乡居民 CPI 的影响因素与预测研究一文中，从文中建立的四个模型综合分析可得，七个影响因素中对我国消费价格数起主要作用的是：食品、居住、娱乐教育文化用品及服务，建议我国政府应大力关注这三个因素在我国

的现状，尽力让社会生产力满足我国居民的生活需要，同时也要协调好其他影响因素之间的关系，从而提高我国居民的人均消费水平，实现经济稳步增长^[6]。

许光建,马祎明.在 CPI、PPI 两种价格指数走势背离研究一文中，从我国目前 CPI 和 PPI 的背离现象入手，通过实证分析研究得到了“剪刀差”现象的两大重要原因，一是食品价格（尤其是猪肉价格）是拖动 CPI 下行的主要动力，二是能源价格是拉动 PPI 上行的主要动力，该研究展示了两者的传导机制，对于更好认识价格水平运行规律、完善价格水平调控机制有着重要现实意义^[16]。

陶宏展,宋涛.在中国不同收入群体进口价格指数估算一文中本文，构建异质性视角下特定收入群体进口价格指数的结构性理论框架，利用详细的进口产品数据和城镇住户调查数据，测算了中国不同收入群体的进口价格指数，分析了不同收入群体进口价格通胀程度。研究发现：各收入群体进口价格指数整体呈现上升趋势，中等收入群体增长最高，低收入群体次之，而高收入群体最低，中等收入群体和低收入群体经历了更高的进口价格通胀，而工业制成品进口价格指数的快速增长是低收入和中等收入群体经历更高进口价格通胀的主要原因^[4]。

从国内相关文献中可以看出，在不同时期对 CPI 起关键性影响的商品种类并不相同，这就需要对历史数据进行分析研究，得到可靠的科学结论。受上述文献的启发，本文重视对 CPI 数据预测的准确性，选取了国家统计局在计算 CPI 时所用到的所有八种价格指数，对其一一进行时间序列建模，以求得到可靠结果。

第三章 国内各种消费价格指数的数据分析

3.1 软件及分析工具的选取

CPI 的组成商品的大类繁多，数据冗长，选取合适的统计工具至关重要，本文主要通过 R 语言对各类数据指标进行统计检验，运用 SPSS_AU 基于拟合结果优劣以及 AIC 最小准则选取合适的时间序列模型并生成拟合报告，本文的所有结果均可以通过文中提供的代码进行复现。

3.2 各种消费价格指数的数据选取

数据选取的详细说明：在本文中，数据主要来源于国家统计局，这是一个官方机构，负责收集、分析和发布国家层面的宏观经济数据。国家统计局发布的数据具有高度的权威性和可靠性，被广泛用于学术研究和政策制定。

在时间跨度上，数据覆盖的时间跨度根据研究目的和可获得性而定。通常，为了进行长期趋势分析和模型预测，会选择过去近 5 年的数据，或者更长的时间跨度，以捕捉可能的长期趋势和周期性变化。但是依据统计局每五年更换一次基期年份以及商品服务分类目录和相对应的权重，本文数据决定采用 2021 年到 2024 年 2 月份（至今）期间的数据来确保数据的准确性和时效性。在采集频率上，国家统计局通常按月发布 CPI 数据，这种频率足以捕捉到短期内的价格变动和季节性波动，同时也便于进行月度比较和年度比较。同时，国家统计局采用标准化的数据收集和处理流程，确保数据的准确性和一致性。此外，国家统计局还会定期进行数据审计和质量控制，以保证数据的高质量和可靠性。

在进行时间序列分析之前，需要对原始数据进行预处理，以确保数据的质量和适用性。以下是数据预处理的主要步骤：

1. 数据清洗：检查数据中是否存在异常值或错误，如价格指数的极端波动或不合理的数值。这些异常值可能是由于数据录入错误、计算错误或其他外部因素导致的。去除或更正这些异常值，以避免它们对分析结果产生不利影响。

2. 缺失值处理：识别数据中的缺失值。缺失值可能是由于数据收集过程中的问题或特殊情况（如自然灾害）导致的。根据数据的性质和缺失情况，选择合适的方法处理缺失值，如插值、使用平均值或最近邻值填充，或者在必要时删除含有缺失值的记录。

3. 数据转换：根据分析需要，可能需要对数据进行转换，如将原始价格指数转换为对数形式，以稳定方差和线性化关系。对于季节性分解，可能需要对数据进行差分处理，以消除趋势和季节性影响，使数据平稳。

4. 数据整合：如果数据来自不同的来源或涉及多个时间段，需要将它们整合到一个统一的数据集中。确保不同数据集之间的一致性和可比性，如统一时间格式、商品分类标准等。

通过上述预处理步骤，可以确保数据的质量和适用性，为后续的时间序列分析和模型

建立打下坚实的基础。在本论文中，这些预处理步骤将被用来准备 CPI 的组成要素数据，以便进行准确的分析和预测。

3.3 不同因素对居民消费价格指数的影响分析

3.3.1 模型选取

季节性 SARIMA 模型是在 ARIMA 模型（移动平均自回归模型）基础上多出一个季节性（seasonal）。比如某旅游景点的销售额数据，每年中有夏天的 6/7/8 共 3 月为旺季，但是其它时间是淡季，但每年整体的销售额均呈现出一定逐步上升趋势^[11]。在模型构建时就需要考虑该周期性因素，即此处的周期值 $S=12$ （1 年为 12 月）。SARIMA 模型正是处理此类带有明显周期性的时间序列数据而生，其它理论内容与 ARIMA 基本一致。除开上述中的参数值即周期值 $S=12$ ，SARIMA 模型还涉及另外 3 个参数值即 $P/D/Q$ ，此这 3 个参数值与 ARIMA 模型的 $p/d/q$ 意义保持一致^[11]。

对于 SARIMA 模型的构建时，需要我们先找出一些可能的参数值模型，然后记录下该模型的 AIC/BIC/均方根误差 RMSE 指标，并且针对多个模型进行对比，AIC/BIC 这两个信息准则值越小意味着模型越优，均方根误差 RMSE 指标表示模型残差的平均 gap 值，该指标也是越小越好，对比多个潜在模型然后先出最优模型^[12]；

首先需要知道 S 值，即季节性周期值，正常情况下可以凭借背景知识确定好，如果无法确定建议使用时序图直观查看和判断周期值。接着进行周期差分即得到 D 值，正常情况下 D 值为 1 或者 0 即可（0 是指不做季节性差分），如果 D 值过大会导致模型过于复杂且准确性下降（经验法则得到）^[12]。接着基于第 2 步后，进行平稳性判断和分析（时序图或者单位根检验），继而得到 d 值。接着基于 3 步之后数据，进行判断 $p/q/P/Q$ 值，需要注意的是 p 值和 q 值需要在一个周期 S 范围内观察，而 P/Q 值的判断和观察应该基于第 i 时刻，第 $i+S$ 时刻，第 $i+2S$ ，第 $i+3S$ 依次下去时刻的滞后阶数去判断 P/Q 值。以及多数情况下 P/Q 值应该尽量小，原因在于时间序列数据通常较少且 P/Q 值过大时，其对当前数据的预测作用减弱^[12]。在得到各个参数值后，主动设置参数值进行分析，得到预测结果等。本文选用 SPSS_AU 进行自动化建模和检验，可以减少时间成本并得到更为准确的拟合模型。

3.3.2 居民消费价格指数的时间序列分析及预测

表 3-1 SARIMA 模型参数结果

参数	值	参数	值
p	1	P	2
d	0	D	0
q	0	Q	0
		S	12

上表格展示依据信息准则最小原则找到的季节性 ARIMA 模型的 7 个参数值：

p 、 d 、 q 此 3 个参数值表示时间趋势性时对应的自回归阶数、差分阶数和移动平均阶

数； P 、 D 、 Q 此 3 个参数值表示周期性序列对应的自回归阶数、差分阶数和移动平均阶数； S 值表示周期性参数值，数据以 12 个月为一个周期，则数字 12。所以，模型为 SARIMA(1,0,0)(2,0,0)12

表 3-2 SARIMA 模型参数表

项	系数	标准误	z 值	p 值	95% CI
intercept	8.482	11.381	0.745	0.456	-13.825 ~ 30.789
ar.L1	0.960	0.056	17.250	0.000	0.851 ~ 1.070
ar.S.L12	-0.691	0.216	-3.204	0.001	-1.114 ~ -0.268
ar.S.L24	-0.438	0.205	-2.137	0.033	-0.840 ~ -0.036
sigma2	0.153	0.041	3.717	0.000	0.072 ~ 0.234

AIC 值: 56.414

BIC 值: 64.602

上表格展示模型构建结果，包括模型参数和信息准则。上表格中 ar 表示自回归，L1 表示自回归滞后 1 阶，S.L12 表示季节性自回归滞后 12 阶的意思，由于 $p=2$ ，因而会出现 L12/L24，再加上 sigma2，发现这几项的 $p\text{-value}<0.05$ ，说明它们都显著，再分析 z 值，z 值是系数除以其标准误。z 值的大小可以用于判断系数是否显著不为 0。通常，一个较大的 z 值（绝对值大于 1.96 或 2.58）表明系数显著不为 0，显然，上述几项的系数都显著不为 0，最后判断 intercept 截距项，发现其标准误差大，这个值表示截距项估计的不确定性。如果标准误很大，说明估计的不确定性很高，其 z 值=0.745 小于 1.96 说明其系数很有可能是 0。最后判断 $p\text{-value}=0.456>0.05$ ，通常意味着系数不显著，即我们可以认为系数为 0，但是通过观察其 95%置信区间：-13.825 到 30.789，这个区间包含了 0，这表明我们不能拒绝截距项等于 0 的假设。

综合以上信息，特别是考虑到 p 值为 0.456 远大于 0.05，以及 95%置信区间包含了 0，这表明截距项在统计上不显著。因此，从统计的角度来看，截距项可能不应该被保留在模型中，因为它没有提供显著的解释能力。然而，是否保留截距项还应该考虑其他因素，如模型的整体拟合度、理论背景、以及截距项在实际应用中的意义等。如果截距项在理论上或实际应用中有重要意义，即使统计上不显著，也可能需要保留。

表 3-3 预测值 (10 期)

后 1 期	后 2 期	后 3 期	后 4 期	后 5 期	后 6 期	后 7 期	后 8 期	后 9 期	后 10 期
99.436	98.989	99.340	99.603	99.682	99.471	99.505	99.556	99.846	100.468

均方根误差 RMSE: 0.4387

均方误差 MSE: 0.1925

平均绝对误差 MAE: 0.3228

平均绝对百分比误差 MAPE: 0.0032

上表格和下图展示 SARIMA 模型展示数据拟合情况，以及数据向后预测情况；

第一：表格中展示具体的向后预测值数据信息；

第二：图中仅展示原始时间序列最近 50 期的实际值和拟合值、以及预测值及预测值 95% 置信区间，各个种类的误差统计量均小于 1，说明该模型拟合效果良好，可以采用。

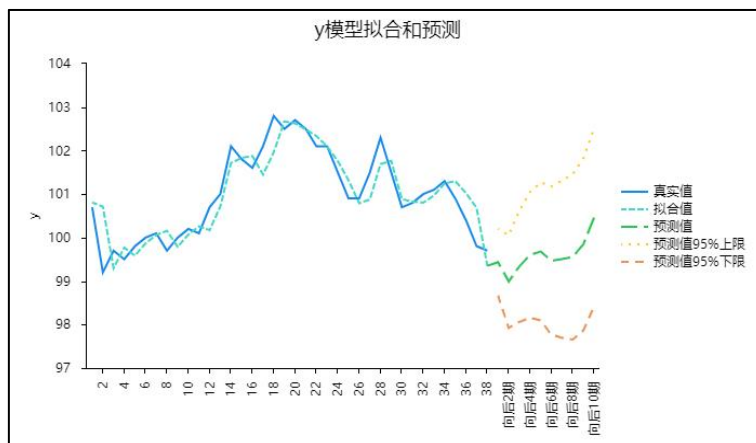


图 3-1 模型拟合和预测

在时间序列分析中，残差检验是用来评估模型拟合效果的重要步骤。残差是指实际观测值与模型预测值之间的差异。一个好的模型应该能够最小化这些残差，并使残差呈现出随机性，即白噪声的特性，下表格展示出 SARIMA 模型各类残差检验指标：

表 3-4 模型残差检验结果

LM 序列检验	值	Ljung-Box 白噪声检验	值
F 统计量	6.988	Ljung-Box 值	0.553
p 值	0.430	p 值	0.457
$T * R^2$ 统计量	0.956	Jarque-Bera 正态性检验	值
p 值	0.485	JB 值	4.846
Heteroskedasticity 异方差检验	值	p 值	0.089
H 值	1.159	残差偏度 Skew	-0.656
p 值	0.794	残差峰度 Kurtosis	4.158

拉格朗日乘数检验(Breush-Godfrey LM 检验)用于检验模型残差序列是否存在序列相关，其原假设 H_0 为序列不存在序列相关，备选假设 H_1 为序列存在序列相关，LM 检验提供两个统计量分别是 F 和 $T * R^2$ 方,通常使用 F 统计量即可；

Ljung-Box 白噪声检验用于检验模型残差是否满足白噪声，其原假设 H_0 为序列为白噪声，备选假设 H_1 为序列非白噪声；

Heteroskedasticity 异方差检验用于检验模型残差是否存在异方差，其原假设 H_0 为序列不存在异方差，备选假设 H_1 为序列存在异方差；

Jarque-Bera 正态性检验用于检验模型残差是否满足正态性，其原假设 H_0 为序列满足正态性，备选假设 H_1 为序列非正态性；

通常情况下希望序列均接受原假设 H_0 即残差序列不存在序列相关、满足白噪声、不存在异方差问题并且满足正态性。通过观察表格，我们可以得到以下结论：

LM 序列检验 (Lagrange Multiplier Test) : F 统计量为 6.988，这是用来检验残差中是否存在自相关性的。 p 值为 0.430，这个值大于 0.05，表明我们没有足够的证据拒绝残差中不存在自相关性的原假设。

Ljung-Box 白噪声检验: Ljung-Box 值为 -0.553, 这个值是用于检验残差是否为白噪声的统计量。p 值为 0.457, 同样大于 0.05, 意味着我们没有证据拒绝残差是白噪声。

T * R² 统计量, 值为 0.956, 这个统计量用于检验残差的自相关性。p 值为 0.457, 大于 0.05, 我们认为残差中不存在显著的自相关性。

Jarque-Bera 正态性检验: JB 值为 4.846, 这个统计量用于检验残差是否服从正态分布。p 值为 0.485, 大于 0.05, 表明我们没有证据拒绝残差服从正态分布的假设。

异方差 (Heteroskedasticity) 检验: H 值为 1.159, 这个值用于检验残差的方差是否恒定, 如果 H 值显著大于 1, 这可能表明存在异方差性, 即残差的方差随着时间的变化而变化。然而, H 值是 1.159, 略高于 1, 这可能表明需要进一步的调查, 但单独这个值并不足以断定存在显著的异方差性。p 值为 0.089, 大于 0.05, 这意味着我们没有足够的证据拒绝残差方差恒定的原假设, 不能断定残差存在显著的异方差性。

残差偏度 (Skewe): 值为 -0.656, 偏度衡量数据分布的不对称性。p 值为 0.794, 大于 0.05, 表明残差的偏度不显著, 分布相对对称。

残差峰度 (Kurtosis): 值为 4.158, 峰度衡量数据分布的尖峭程度或平坦程度, 认为残差的峰度并不十分显著。

综上所述, 根据这些检验结果, 我们可以得出以下结论:

1. 残差不存在显著的自相关性。
2. 残差可以被认为是正态分布的。
3. 残差的方差不存在异方差问题。
4. 残差的偏度和峰度不显著, 分布相对对称且不极端。

白噪声检验是用来验证一个时间序列是否是随机的, 即序列中的值是否独立且同分布, 通常假设为均值为零, 方差为常数的正态分布。在时间序列分析中, 我们希望模型的残差 (误差项) 表现为白噪声, 这表明模型已经尽可能地解释了数据中的信息, 残差中不再包含可预测的结构。在进行白噪声检验时, 应该注意以下几点:

1. p 值: 检验结果中的 p 值用于判断统计显著性。如果 p 值小于常用的显著性水平 (如 0.05), 则拒绝原假设, 认为序列不是白噪声。
2. 检验的滞后阶数: 检验时需要选择合适的滞后阶数, 这通常取决于序列的长度和模型的复杂度。
3. 多重检验问题: 如果同时进行多个检验, 需要考虑多重比较的问题, 可能需要调整显著性水平。
4. 图形分析: 除了统计检验外, 还可以通过图形分析来辅助判断, 例如绘制残差的自相关函数 (ACF) 和偏自相关函数 (PACF) 图。
5. 模型诊断: 如果白噪声检验的结果显示残差不是白噪声, 可能需要重新考虑模型的选择和参数设定

表 3-5 模型 Q 统计量表格

项	统计量	p 值
Q1	0.262	0.609
Q2	0.348	0.840
Q3	0.838	0.840
Q4	0.887	0.926
Q5	0.940	0.967
Q6	2.227	0.898
Q7	2.471	0.929
Q8	2.860	0.943
Q9	2.941	0.967
Q10	3.544	0.966
Q11	3.568	0.981
Q12	3.862	0.986
Q13	3.862	0.993
Q14	5.792	0.971
Q15	5.813	0.983

* $p < 0.05$ ** $p < 0.01$

上表格展示模型 Q 统计量信息，包括统计量值和 p 值：

SARIMA 模型要求模型残差为白噪声，即残差不存在自相关性，可通过 Q 统计量检验进行白噪声检验（原假设：残差是白噪声），比如 Q6 用于检验残差前 6 阶自相关系数是否满足白噪声，通常其对应 p 值大于 0.1 则说明满足白噪声检验（反之则说明不是白噪声），常见情况下可直接针对 Q6 进行分析即可；如果拒绝白噪声假定($p < 0.05$),意味着模型可能拟合不佳，反之通常意味着模型可正常使用。

本节对拟合报告中的各项检验及指数都做了解释，由于篇幅限制后续对余下几个价格指数进行建模预测时均按照此流程，只展示模型参数和拟合预测的结果，不再做详细说明。

3.3.3 对食品烟酒类居民消费价格指数的时间序列分析及预测

以下是对模型参数估计和拟合预测的图表展示：

表 3-6 SARIMA 参数模型结果

参数	值	参数	值
p	1	P	0
d	0	D	0
q	0	Q	1
		S	6

模型为 SARIMA(1,0,0)(0,0,1)₆

图 3-7 SARIMA 模型参数表及预测值表

项	系数	标准误	z 值	p 值	95% CI
intercept	21.401	8.906	2.403	0.016	3.945 ~ 38.857
ar.L1	0.787	0.089	8.882	0.000	0.613 ~ 0.961
ma.S.L6	0.796	0.291	2.732	0.006	0.225 ~ 1.367
sigma2	1.311	0.375	3.493	0.000	0.575 ~ 2.046

AIC 值：133.253

BIC 值：139.803

预测值（10 期）

后 1 期	后 2 期	后 3 期	后 4 期	后 5 期	后 6 期	后 7 期	后 8 期	后 9 期	后 10 期
101.025	100.680	100.724	100.664	100.084	101.088	100.952	100.845	100.761	100.694

均方根误差 RMSE: 1.2551

均方误差 MSE: 1.5754

平均绝对误差 MAE: 0.9044

平均绝对百分比误差 MAPE: 0.0090

可以看到这几项参数的 z 值较大，p 值小于 0.05 且不在 95%置信区间内，说明系数显著，可以保留，各类型误差较小说明拟合效果良好

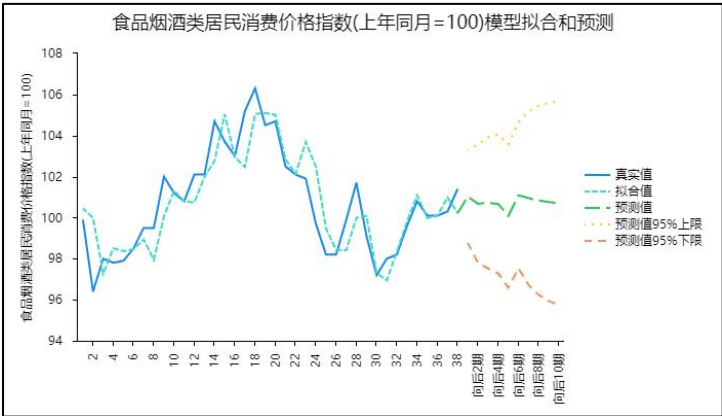


图 3-2 食品烟酒类居民消费价格指数(上年同月=100)模型拟合和预测

表 3-8 模型残差检验结果

LM 序列检验	值	Ljung-Box 白噪声检验	值
F 统计量	10.204	Ljung-Box 值	1.792
p 值	0.177	p 值	0.181
$T * R^2$ 统计量	1.612	Jarque-Bera 正态性检验	值
p 值	0.182	JB 值	0.657
Heteroskedasticity 异方差检验	值	p 值	0.720
H 值	0.579	残差偏度 Skew	-0.250
p 值	0.336	残差峰度 Kurtosis	3.405

综合这些检验结果，我们可以得出以下结论：

1. 残差不存在显著的自相关性。
2. 残差可以被认为是正态分布的。
3. 残差的方差是恒定的，不存在异方差性。
4. 残差的分布相对对称，偏度不显著。
5. 残差的峰度值为 3.405，略高于正态分布的峰度值 3，表明残差分布的中心峰可能略尖一些，但是否显著需要 p 值来判断。

这些检验结果表明，模型的拟合效果较好，残差表现出了白噪声的特性，这通常意味着模型已经捕捉了数据中的大部分信息，残差中没有明显的可预测的结构。然而，峰度略高可能提示需要进一步的模型诊断或考虑数据的其他特性。

3.3.4 衣着类居民消费价格指数的时间序列分析及预测

表 3-9 SARIMA 模型参数结果

参数	值	参数	值
p	0	P	0
d	1	D	0
q	0	Q	1
		S	6

模型为 SARIMA(0,1,0)(0,0,0)6，在对衣着类价格指数进行建模时，S 参数设定为 12 结果效果不理想，设定为 4 时也不理想，为 6 时模型效果有了很大改善，可能是由于数据中存在的季节性特征与预期的周期性不匹配，或者是由于其他外部因素的影响，可以从以下两个方面思考，1.季节性变化的复杂性：实际数据中的季节性可能并不总是遵循标准的季节性周期（如每年的 12 个月或每季度的 3 个月）。可能存在其他非标准的季节性模式，例如周期为半年气候变化等，大致来讲，秋装与冬装的价格普遍比春装夏装贵，而我国居民有错季购买服装的习惯，这些都可能弱化价格指数中理应 12 个月的周期特征，导致 6 个月的周期性更为显著。2，数据中的异常值：在特定的月份，可能会出现异常值，这些异常值可能会影响模型的拟合效果。例如，如果在每年的某个季节出现了异常的消费高峰或低谷，而这个季节特征的周期并不是 12 个月或 4 个月，那么以 12 月或 4 月为周期的模型就难以

捕捉到这种模式，例如电商兴起后衍生出的购物狂欢节 618、双 11，两个节日的时间周期正好为 6 个月。

从 3-10 模型参数表中可以看到截距项、sigma2 项参数的 z 值（绝对值）较大，p 值小于 0.05 且不在 95%置信区间内，说明系数显著，可以保留，但是季节性滑动平均滞后 6 项的系数并不显著，需要结合拟合效果决定是否保留。

此外，各类型误差较小说明拟合效果良好。

表 3-10 SARIMA 模型参数表及预测值

项	系数	标准误	z 值	p 值	95% CI
intercept	-0.041	0.009	-4.656	0.000	-0.059 ~ -0.024
ma.S.L6	-0.849	0.559	-1.519	0.129	-1.943 ~ 0.246
sigma2	0.015	0.007	2.211	0.027	0.002 ~ 0.028

AIC 值: -37.653

BIC 值: -32.821

预测值 (10 期)

预测	后 1 期	后 2 期	后 3 期	后 4 期	后 5 期	后 6 期	后 7 期	后 8 期	后 9 期	后 10 期
值	99.615	99.589	99.623	99.526	99.848	99.646	99.604	99.563	99.521	99.480

均方根误差 RMSE: 0.1287

均方误差 MSE: 0.0166

平均绝对误差 MAE: 0.0986

平均绝对百分比误差 MAPE: 0.0010

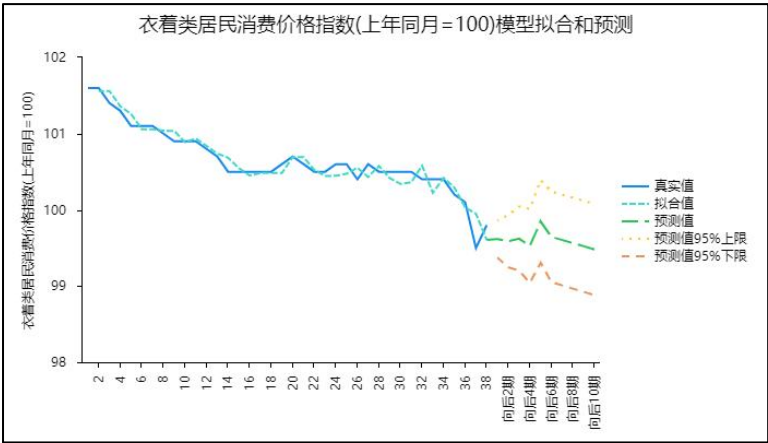


图 3-3 衣着类居民消费价格指数(上年同月=100)模型拟合和预测

表 3-11 模型残差检验结果

LM 序列检验	值	Ljung-Box 白噪声检验	值
F 统计量	5.707	Ljung-Box 值	3.287
p 值	0.574	p 值	0.070
$T * R^2$ 统计量	0.738	Jarque-Bera 正态性检验	值
p 值	0.642	JB 值	13.915
Heteroskedasticity 异方差检验	值	p 值	0.001
H 值	7.230	残差偏度 Skew	-0.974
p 值	0.002	残差峰度 Kurtosis	5.287

综合这些检验结果，我们可以得出以下结论：

1. 残差不存在显著的自相关性，尽管 Ljung-Box 检验的 p 值接近显著性水平。
2. 残差可以被认为是正态分布的，Jarque-Bera 检验的 JB 值较高，但 p 值大于 0.05。
3. 残差存在显著的异方差性，这表明模型可能未能充分解释数据中的变异性，或者数据中存在其他未被模型捕捉的结构。
4. 残差存在显著的偏度，表明分布不对称，这可能指示模型需要进一步调整以更好地捕捉数据的不对称特性。
5. 残差的峰度值显著高于 3，表明残差分布比正态分布更尖锐，极端值出现的概率较高。

这些检验结果表明，尽管模型在自相关性和正态性方面表现良好，但存在显著的异方差性和偏度，这可能提示模型需要进一步的改进或考虑其他潜在的数据结构。

3.3.5 居住类居民消费价格指数的时间序列分析及预测

表 3-12 SARIMA 模型参数结果

参数	值	参数	值
p	2	P	1
d	0	D	0
q	1	Q	0
		S	12

模型为 SARIMA(2,0,0)(1,0,0)₁₂

表 3-13 SARIMA 模型参数表及预测值

项	系数	标准误	z 值	p 值	95% CI
intercept	6.977	null	null	null	null ~ null
ar.L1	1.826	null	null	null	null ~ null
ar.L2	-0.872	null	null	null	null ~ null
ma.L1	-0.742	null	null	null	null ~ null
ar.S.L12	-0.499	null	null	null	null ~ null
sigma2	0.016	0.003	5.069	0.0001	0.010 ~ 0.022

AIC 值: -32.170

BIC 值: -22.344

预测值 (10 期)

后 1 期	后 2 期	后 3 期	后 4 期	后 5 期	后 6 期	后 7 期	后 8 期	后 9 期	后 10 期
99.399	99.283	99.252	99.452	99.579	99.630	99.797	99.976	100.213	100.401

均方根误差 RMSE: 0.1351

均方误差 MSE: 0.0183

平均绝对误差 MAE: 0.1061

平均绝对百分比误差 MAPE: 0.0011

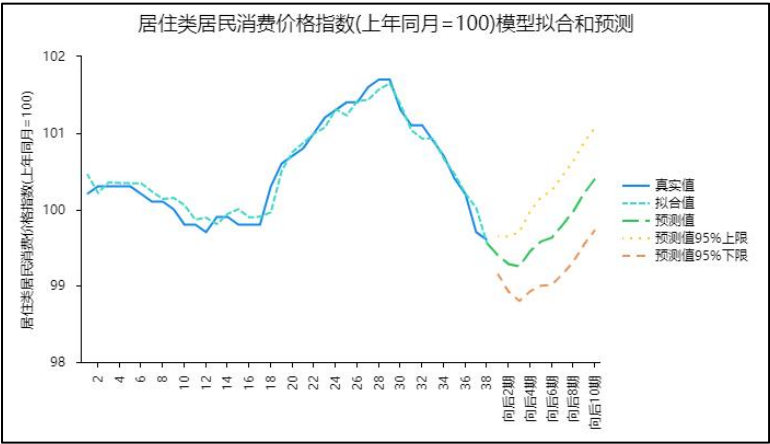


图 3-4 居住类居民消费价格指数模型拟合和预测

3.3.6 生活用品及服务类居民消费价格指数的分析及预测

表 3-14 SARIMA 模型参数结果

参数	值	参数	值
p	1	P	0
d	0	D	0
q	0	Q	1
		S	12

模型为 SARIMA(1,0,0)(0,0,1)12

表 3-15 SARIMA 模型参数表及预测值

项	系数	标准误	z 值	p 值	95% CI
intercept	11.054	9.292	1.190	0.234	-7.157 ~ 29.266
ar.L1	0.890	0.092	9.642	0.000	0.709 ~ 1.071
ma.S.L12	-0.728	0.549	-1.326	0.185	-1.804 ~ 0.348
sigma2	0.076	0.034	2.252	0.024	0.010 ~ 0.142

AIC 值: 27.265

BIC 值: 33.815

预测值 (10 期)

后 1 期	后 2 期	后 3 期	后 4 期	后 5 期	后 6 期	后 7 期	后 8 期	后 9 期	后 10 期
100.000	100.328	100.144	100.395	100.407	100.525	100.556	100.463	100.516	100.800

均方根误差 RMSE: 0.3201

均方误差 MSE: 0.1025

平均绝对误差 MAE: 0.2423

平均绝对百分比误差 MAPE: 0.0024

从表 3-15 中可以看出截距项和 L12 项的系数不够显著，需结合拟合效果进一步判断，此外，各类型误差较小，说明拟合效果良好。

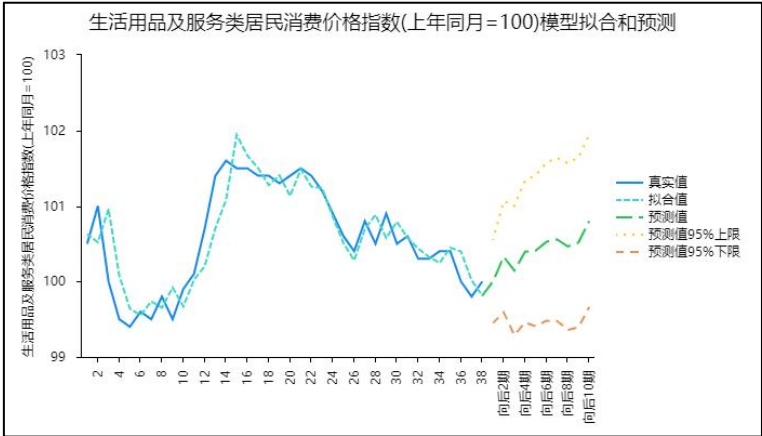


图 3-5 生活用品及服务类居民消费价格指数(上年同月=100)模型拟合和预测

表 3-16 模型残差检验结果

LM 序列检验	值	Ljung-Box 白噪声检验	值
F 统计量	5.941	Ljung-Box 值	0.213
p 值	0.547	p 值	0.645
$T * R^2$ 统计量	0.779	Jarque-Bera 正态性检验	值
p 值	0.611	JB 值	0.739
Heteroskedasticity 异方差检验	值	p 值	0.691
H 值	0.344	残差偏度 Skew	-0.285
p 值	0.065	残差峰度 Kurtosis	3.376

综合这些检验结果，我们可以得出以下结论：

1. 残差不存在显著的自相关性。
2. 残差可以被认为是白噪声，即模型已经捕捉了数据中的大部分信息。
3. 残差可以被认为是服从正态分布的。
4. 残差的方差是恒定的，不存在显著的异方差性。
5. 残差存在轻微的偏度，可能提示分布的不对称性，但这需要进一步的调查。
6. 峰度值略高于 3，表明残差分布的中心峰较尖，这可能表明存在一些极端值，但是否显著需要 p 值来判断。

这些检验结果表明，模型的拟合效果较好，残差表现出了白噪声的特性，这通常意味着模型已经捕捉了数据中的大部分信息，残差中没有明显的可预测的结构。然而，残差偏度的 p 值接近显著性水平，可能提示需要进一步的模型诊断或考虑数据的其他特性。

3.3.7 交通和通信类居民消费价格指数的时间序列分析及预测

表 3-17 SARIMA 模型参数结果

参数	值	参数	值
p	1	P	2
d	1	D	0
q	0	Q	0
		S	12

模型为 SARIMA(1,1,0)(2,0,0)₁₂

表 3-18 SARIMA 模型参数表

项	系数	标准误	z 值	p 值	95% CI
intercept	-0.121	0.224	-0.542	0.588	-0.560 ~ 0.317
ar.L1	0.479	0.151	3.177	0.001	0.183 ~ 0.774
ar.S.L12	-0.882	0.185	-4.782	0.000	-1.244 ~ -0.521
ar.S.L24	-0.691	0.132	-5.249	0.000	-0.949 ~ -0.433
sigma2	0.865	0.283	3.055	0.002	0.310 ~ 1.419

AIC 值: 129.272

BIC 值: 137.327

预测值 (10 期)

后 1 期	后 2 期	后 3 期	后 4 期	后 5 期	后 6 期	后 7 期	后 8 期	后 9 期	后 10 期
94.111	91.190	91.192	90.977	90.343	88.382	87.453	89.071	89.158	91.349

均方根误差 RMSE: 1.2150

均方误差 MSE: 1.4762

平均绝对误差 MAE: 0.9317

平均绝对百分比误差 MAPE: 0.0092

1. 截距项在统计上不显著，可能表明基线水平对因变量的影响不显著。
2. arL1、ar. S. L12 和 ar. S. L24 项显著，表明这些滞后项对因变量有显著影响。
3. sigma2 显著，表明模型对残差方差的估计也是可信的。
4. 各类型误差较小，说明模型拟合效果良好。

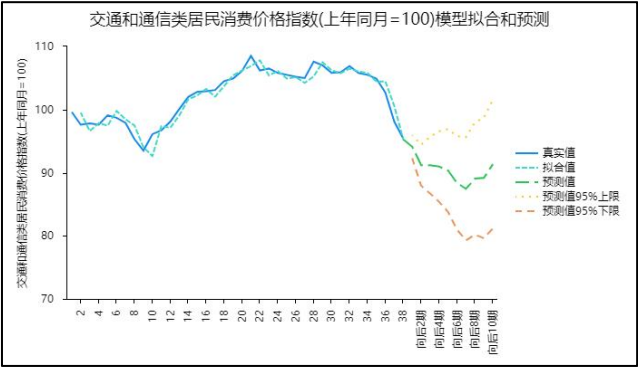


图 3-6 交通和通信类居民消费价格指数(上年同月=100)模型拟合和预测

表 3-19 模型残差检验结果

LM 序列检验	值	Ljung-Box 白噪声检验	值
F 统计量	5.728	Ljung-Box 值	0.010
p 值	0.572	p 值	0.921
$T * R^2$ 统计量	0.742	Jarque-Bera 正态性检验	值
p 值	0.640	JB 值	1.701
Heteroskedasticity 异方差检验	null	p 值	0.427
H 值	1.534	残差偏度 Skew	-0.100
p 值	0.470	残差峰度 Kurtosis	4.031

综合这些检验结果，我们可以得出以下结论：

1. 残差不存在显著的自相关性。
2. 残差可以被认为是白噪声，即模型已经捕捉了数据中的大部分信息。
3. 残差可以被认为是服从正态分布的。
4. 残差的方差是恒定的，不存在显著的异方差性。
5. 残差偏度不显著，表明分布相对对称。
6. 残差峰度显著高于 3，表明残差分布比正态分布更尖锐，可能存在一些极端值。

这些检验结果表明，模型的拟合效果较好，残差表现出了白噪声的特性，这通常意味着残差中没有明显可预测的结构。但差峰度值较高，可能需要进一步的模型诊断或考虑数据的其他特性，例如是否存在异常值或需要对模型进行调整以更好地捕捉极端值。

3.3.8 教育文化和娱乐类居民消费价格指数的分析及预测

表 3-20 SARIMA 模型参数结果

参数	值	参数	值
p	1	P	1
d	0	D	0
q	1	Q	1
		S	12

模型为 SARIMA(1,0,1)(1,0,1)12

表 3-21 SARIMA 模型参数表

项	系数	标准误	z 值	p 值	95% CI
intercept	7.621	null	null	null	null ~ null
ar.L1	0.948	null	null	null	null ~ null
ma.L1	-0.083	0.219	-0.380	0.704	-0.512 ~ 0.346
ar.S.L12	-0.444	null	null	null	null ~ null
ma.S.L12	-0.699	0.444	-1.575	0.115	-1.569 ~ 0.171
sigma2	0.161	0.071	2.273	0.023	0.022 ~ 0.299

AIC 值: 67.645

BIC 值: 77.471

预测值 (10 期)

后 1 期	后 2 期	后 3 期	后 4 期	后 5 期	后 6 期	后 7 期	后 8 期	后 9 期	后 10 期
100.334	100.543	100.630	100.296	100.297	100.734	101.895	102.057	102.124	102.861

均方根误差 RMSE: 0.6045

均方误差 MSE: 0.3654

平均绝对误差 MAE: 0.3996

平均绝对百分比误差 MAPE: 0.0039

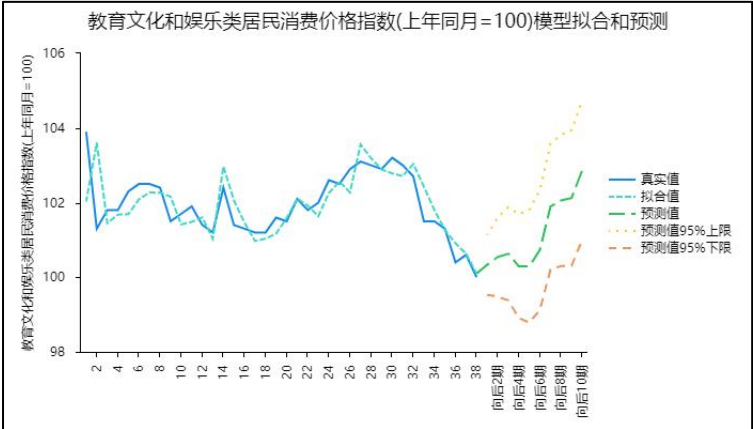


图 3-7 教育文化和娱乐类居民消费价格指数(上年同月=100)模型拟合和预测

表 3-22 模型残差检验结果

LM 序列检验	值	Ljung-Box 白噪声检验	值
F 统计量	6.705	Ljung-Box 值	0.243
p 值	0.460	p 值	0.622
$T * R^2$ 统计量	0.907	Jarque-Bera 正态性检验	值
p 值	0.518	JB 值	18.476
Heteroskedasticity 异方差检验	值	p 值	0.0001
H 值	0.651	残差偏度 Skew	-1.234
p 值	0.450	残差峰度 Kurtosis	5.361

3.3.9 医疗保健类居民消费价格指数的时间序列分析及预测

表 3-23 SARIMA 模型参数结果

参数	值	参数	值
p	0	P	0
d	1	D	0
q	0	Q	0
		S	12

模型为 SARIMA(0,1,0)(0,0,0)12

表 3-24 SARIMA 模型参数表

项	系数	标准误	z 值	p 值	95% CI
intercept	-0.030	0.015	-1.953	0.051	-0.060 ~ 0.000
sigma2	0.009	0.002	4.301	0.0003	0.005 ~ 0.012

AIC 值: -67.075

BIC 值: -63.854

预测值 (10 期)

后 1 期	向后 2 期	后 3 期	后 4 期	后 5 期	后 6 期	后 7 期	后 8 期	后 9 期	后 10 期
值 100.370	100.341	100.311	100.281	100.251	100.222	100.192	100.162	100.132	100.103

均方根误差 RMSE: 0.0926

均方误差 MSE: 0.0086

平均绝对误差 MAE: 0.0748

平均绝对百分比误差 MAPE: 0.0007

根据表 3-24 可以发现,此截距项显著、系数也显著,应该保留。sigma2 显著,表明模型对残差方差的估计也是可信的。各个种类的误差项较小,说明拟合效果良好。从预测图片以及向后 10 期的预测值可以发现,我国医疗保障类居民消费价格指数季节差异不大,并且没有明显的周期性,而且价格指数呈现逐年稳步下降的特点,这也说明了国家大力发展的医保体系大大减少了居民在医疗上的花费。

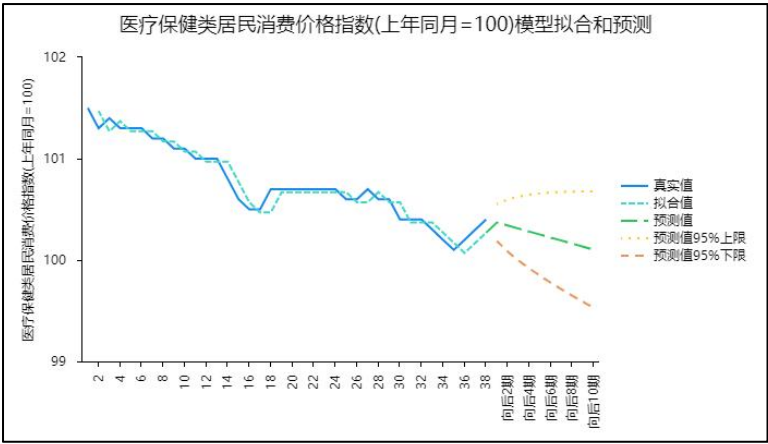


图 3-8 医疗保障类居民消费价格指数模型拟合和预测

表 3-25 模型残差检验结果

LM 序列检验	值	Ljung-Box 白噪声检验	值
F 统计量	9.359	Ljung-Box 值	0.150
p 值	0.228	p 值	0.699
$T * R^2$ 统计量	1.425	Jarque-Bera 正态性检验	值
p 值	0.245	JB 值	0.025
Heteroskedasticity 异方差检验	值	p 值	0.988
H 值	1.675	残差偏度 Skew	0.009
p 值	0.384	残差峰度 Kurtosis	2.874

综合这些检验结果,我们可以得出以下结论:

1. 残差不存在显著的自相关性。
2. 残差可以被认为是白噪声，即模型已经捕捉了数据中的大部分信息。
3. 残差可以被认为是服从正态分布的。
4. 残差的方差是恒定的，不存在显著的异方差性。
5. 残差偏度不显著，表明分布相对对称。
6. 残差峰度略高于 3，但这个值通常不被认为是显著偏离正态分布。

这些检验结果表明，模型的拟合效果较好，残差表现出了白噪声的特性，这通常意味着模型已经捕捉了数据中的大部分信息，残差中没有明显的可预测的结构。

3.3.10 其他用品和服务类居民消费价格指数的分析及预测

表 3-26 SARIMA 模型参数结果

参数	值	参数	值
p	0	P	0
d	1	D	0
q	2	Q	1
		S	12

模型为 SARIMA(0,1,2)(0,0,1)12

根据下表 3-27 可以得出以下结论：

1. 截距项在统计上显著，表明模型中包含了重要的常数效应。
2. Ma.L2 项显著，表明滞后 2 期的移动平均效应对模型有贡献。
3. Ma.L1 和 Ma.SL12 项不显著，表明这些项对模型的贡献不大，可能在模型选择时考虑剔除。
4. 残差方差 sigma2 的显著性处于边缘状态，可能需要进一步考察。

这些参数的显著性表明 SARIMA 模型能够捕捉到时间序列数据中的某些动态特性。对于不显著的参数，可能要考虑是否需要在模型中包含这些项，或者是否有其他因素需要考虑。

表 3-27 SARIMA 模型参数表

项	系数	标准误	z 值	p 值	95% CI				
intercept	-0.120	0.044	-2.706	0.007	-0.207 ~ -0.033				
ma.L1	0.003	0.184	0.018	0.985	-0.357 ~ 0.364				
ma.L2	-0.473	0.234	-2.024	0.043	-0.931 ~ -0.015				
ma.S.L12	-0.782	0.705	-1.109	0.268	-2.163 ~ 0.600				
sigma2	0.387	0.231	1.676	0.094	-0.066 ~ 0.840				
AIC 值: 89.997									
BIC 值: 98.052									
预测值 (10 期)									
后 1 期	后 2 期	后 3 期	向后 4 期	向后 5 期	向后 6 期	向后 7 期	向后 8 期	向后 9 期	后 10 期
99.311	99.158	98.853	100.251	100.493	98.900	98.347	97.868	97.823	97.742

均方根误差 RMSE: 0.6958

均方误差 MSE: 0.4841

平均绝对误差 MAE: 0.4979

平均绝对百分比误差 MAPE: 0.0049

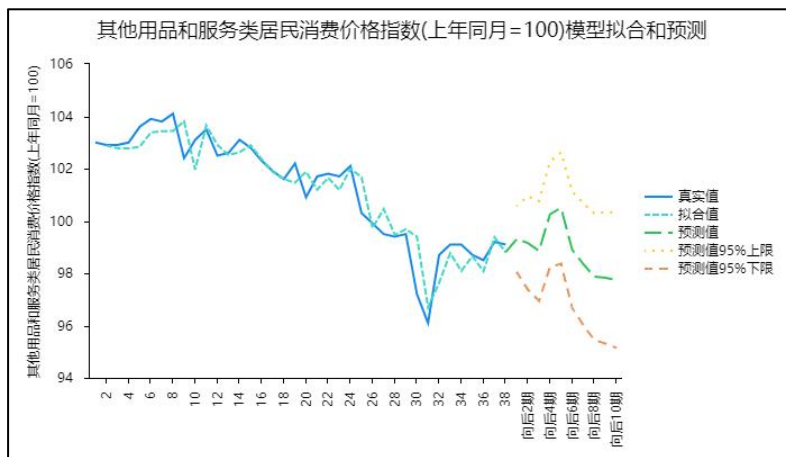


图 3-9 其他用品和服务类居民消费价格指数模型拟合和预测

表 3-28 模型残差检验结果

LM 序列检验	值	Ljung-Box 白噪声检验	值
F 统计量	3.203	Ljung-Box 值	0.104
p 值	0.866	p 值	0.747
$T * R^2$ 统计量	0.376	Jarque-Bera 正态性检验	值
p 值	0.907	JB 值	16.028
Heteroskedasticity 异方差检验	值	p 值	0.000
H 值	2.539	残差偏度 Skew	-1.239
p 值	0.120	残差峰度 Kurtosis	5.063

综合这些检验结果，我们可以得出以下结论：

1. 残差不存在显著的自相关性。
2. 残差可以被认为是白噪声，即模型已经捕捉了数据中的大部分信息。
3. 残差可以被认为是服从正态分布的，尽管 Jarque-Bera 检验的 JB 值相对较高，但 p 值大于 0.05。
4. 残差存在显著的异方差性，这表明模型可能未能充分解释数据中的变异性，或者数据中存在其他未被模型捕捉的结构。
5. 残差偏度的统计显著性不强，但需要进一步关注。
6. 残差峰度显著高于 3，表明残差分布比正态分布更尖锐，这可能表明存在一些极端值。

这些检验结果表明，尽管模型在自相关性和正态性方面表现良好，但存在显著的异方差性和较高的峰度值，这可能提示模型需要进一步的改进或考虑其他潜在的数据结构。特别是异方差性的显著性提示我们可能需要考虑变换数据或使用其他模型技术来解决方差非恒定的问题。

3.4 拟合参数计算

下图展示了中国 CPI 八大子类权重。由于我国并没有官方公布的 CPI 权重数据，图中的结果是由 CPI 对其八大子类做线性回归得到的，在线性回归的过程中取消对截距 intercept 的拟合（因为 CPI 是各大类消费指数的加权平均之和，不存在截距），数据区间是 2021 年 1 月到 2024 年 2 月，结果显示八大商品服务种类的权重系数分别为：

0.278796,0.00561,0.208466,0.078379,0.137838,0.109822,0.158001,0.22796

R-squared 值以及调整后的 R-squared 值均为 1，说明模型能够解释所有实际值，结果可靠。

```
> sourcedata <- readxl::read_xls("数据(2).xls")
> fit = lm(y ~ .-1, data = sourcedata)
> summary(fit)
```

Call:
lm(formula = y ~ .-1, data = sourcedata)

Residuals:
Min 1Q Median 3Q Max
-0.07279 -0.01776 -0.00332 0.02399 0.04656

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
食品烟酒类居民消费价格指数(上年同月=100)	0.278796	0.003952	70.549	< 2e-16 ***
衣着类居民消费价格指数(上年同月=100)	0.005610	0.034691	0.162	0.872606
居住类居民消费价格指数(上年同月=100)	0.208466	0.022307	9.345	2.16e-10 ***
生活用品及服务类居民消费价格指数(上年同月=100)	0.078379	0.014568	5.380	7.96e-06 ***
交通和通信类居民消费价格指数(上年同月=100)	0.137838	0.004527	30.449	< 2e-16 ***
教育文化和娱乐类居民消费价格指数(上年同月=100)	0.109822	0.010508	10.451	1.62e-11 ***
医疗保健类居民消费价格指数(上年同月=100)	0.158001	0.048334	3.269	0.002709 **
其他用品和服务类居民消费价格指数(上年同月=100)	0.022796	0.005152	4.425	0.000117 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03417 on 30 degrees of freedom
Multiple R-squared: 1. Adjusted R-squared: 1
F-statistic: 4.148e+07 on 8 and 30 DF, p-value: < 2.2e-16

图 3-10 回归代码及结果

3.5 结果分析与建议

在以下表格中，第一行预测总指数是指在 3.1 节中对总的 CPI 往后十期的预测，以下的八个分类依次代表 3.2 到 3.9 节中分析预测的八个 CPI 组成要素，加权平均的总指数是指将往后每一期八大分类的预测值乘以第三章开头回归分析得到的各自的系数再求和，之后再与预测总指数相减求绝对值以查看直接对 CPI 进行预测和对主要分类进行预测后再相加的值是否一致，从下图中我们可以看到，在第一期和第二期这两种预测方法得到的结果几乎相同，但在之后它们之间的差别越来越大，所以我们可以得到结论，第二种预测方法在至少往后两期的结果上是准确的，这种预测方法是可行的。

	第一期	第二期	第三期	第四期	第五期	第六期	第七期	第八期	第九期	第十期	系数
预测总指数	99.436	98.989	99.34	99.603	99.682	99.471	99.505	99.556	99.846	100.468	None
预测分类1	101.025	100.68	100.724	100.664	100.084	101.088	100.952	100.845	100.761	100.694	0.278796
预测分类2	99.615	99.589	99.623	99.526	99.848	99.646	99.604	99.563	99.521	99.48	0.00561
预测分类3	99.399	99.283	99.252	99.452	99.579	99.63	99.797	99.976	100.213	100.401	0.208466
预测分类4	100	100.328	100.144	100.395	100.407	100.525	100.556	100.463	100.516	100.8	0.078379
预测分类5	94.111	91.19	91.192	90.977	90.343	88.382	87.453	89.071	89.158	91.349	0.137838
预测分类6	100.334	100.543	100.63	100.296	100.297	100.734	101.895	102.057	102.124	102.861	0.109822
预测分类7	100.37	100.341	100.311	100.281	100.251	100.222	100.192	100.162	100.132	100.103	0.158001
预测分类8	99.311	99.158	98.853	100.251	100.493	98.9	98.347	97.868	97.823	97.742	0.022796
加权平均的总指数	99.39682448	98.91427844	98.9039894	98.90889686	98.68991446	98.72536875	98.70656607	98.93168479	98.97517496	99.39423023	None
与预测值的差值	0.039175524	0.074721565	0.436010603	0.694103144	0.992085542	0.745631251	0.798433934	0.624315215	0.87082504	1.073769771	None

图 3-11 两种预测方式的差值

在当前全球化和信息化的时代背景下，CPI 作为衡量居民消费价格水平的重要指标，对政策制定者、企业以及普通消费者都具有深远的影响。因此，本文不仅提出了对政府、企业和学术界的建议，还从更广泛的视角出发，探讨了如何利用 CPI 数据推动社会整体福祉的提升和经济的可持续发展。

首先，对于政策制定者而言，理解 CPI 的动态变化是至关重要的。这不仅包括对 CPI 本身的分析，还应扩展到 CPI 背后反映的经济和社会问题。例如，当某一类商品或服务价格上涨时，可能反映了供应链的问题或是生产成本的增加。政府可以通过深入研究这些因素，采取相应的措施来缓解这些问题，如优化供应链管理、提高生产效率或实施价格补贴政策等。此外，政府还可以借助 CPI 数据来评估其宏观经济政策的有效性，如货币政策是否达到了预期的通货膨胀控制目标，财政政策是否促进了经济的稳定增长等。

对于企业来说，CPI 提供了宝贵的市场信息。企业可以利用这些数据来调整自己的商业策略，比如定价策略、库存管理和市场定位。通过分析 CPI 的变化趋势，企业可以更好地预测市场需求和成本变动，从而做出更加精准的商业决策。同时，企业也可以通过创新和改进产品与服务来应对价格竞争，提升自身的市场竞争力。

对于消费者而言，CPI 的透明度可以帮助他们更好地规划个人和家庭的财务。随着生活水平的提高和物价水平的波动，消费者需要对收入、支出和储蓄进行合理的规划。了解 CPI 的变化趋势可以帮助消费者做出更为明智的消费选择和投资决策，比如何时购买耐用品以规避短期内的价格上升风险。

最后，学术界应继续深化对 CPI 及其组成要素的研究。除了传统的时间序列分析和模型预测方法外，还可以探索更多新兴的分析手段和技术，如大数据分析、人工智能在经济预测中的应用等。这些技术的应用可以提高分析的准确性和效率，为政策制定和学术研究提供更为丰富和深入的支持。

总之，CPI 不仅是一个经济指标，更是连接政府、企业、消费者和学术界的重要桥梁。通过对 CPI 的深入研究和应用，我们可以更好地理解经济和社会的变化趋势，从而为促进社会的和谐发展和经济的健康增

第四章 总结

本文通过对中国居民消费价格指数（CPI）的深入分析，旨在探讨 CPI 各组成部分的变化趋势及其对经济政策、企业经营和居民消费的影响。文章首先定义了 CPI 并详细介绍了其计算方法，强调了 CPI 作为衡量通货膨胀和经济稳定性的关键指标的重要性。通过对 2016 年 1 月至 2024 年 2 月期间的数据进行时间序列分析，包括 ARIMA 模型和季节性分解等高级统计方法，详细分析了 CPI 的各个组成部分，如食品烟酒类、衣着类、居住类等，预测了这些部分的价格指数变化趋势，并通过加权平均的方法计算了总的 CPI。

数据分析结果显示，CPI 的各组成部分在观察期间表现出不同的变化模式。例如，医疗保健类显示出逐年稳步下降的特点，这可能反映了国家医保体系改进带来的积极影响。而交通和通信类则显示了较为显著的周期性特征。此外，通过回归分析得出的各商品服务种类的权重系数，表明不同类别在 CPI 构成中的重要性存在差异，其中食品类和居住类占据了较大的权重，这与它们在居民总消费支出中占比较高的现实相吻合。

进一步地，本研究还对比了直接对 CPI 数据进行时间序列分析和先对组成 CPI 的各个分类进行时间序列分析再求和的两种预测方法。结果表明后者至少在两期之后提供了更为准确的预测结果，验证了分别分析各个组成部分并进行综合评估的有效性。这一发现为未来的 CPI 研究和预测提供了新的思路。

总体而言，通过对 CPI 及其组成要素的深入分析和研究，本论文不仅为宏观经济政策制定者提供了宝贵的数据支持和建议，同时也为企业的市场策略调整和消费者个人及家庭财务规划提供了实证依据。此外，本文的研究亦为未来关于 CPI 研究的方法论提供了重要的参考价值。

参考文献

- [5] 王双进,崔福鑫,孙淑敏.基于 CPI 数据的我国价格总水平分析: 2023 年回顾及 2024 年展望[J/OL].价格与实践理论:1-6.
- [6] 黄鹏. 国际油价波动对 APEC 居民消费价格指数的影响研究[D].江西财经大学,2024.
- [7] 邵莉.我国居民消费价格指数数据质量评估研究[D]杭州电子科技大学, 2024.
- [8] 陶宏展,宋涛.中国不同收入群体进口价格指数估算[J].当代财经,2022(12):111-123.
- [9] Zhang Y. The Economic Linkage Between China and US: An Empirical Evidence from CPI[C]//Department of Economics, King's Business School, King's College London, Cardiff Business School, Business School. Proceedings of the 2nd International Conference on Business and Policy Studies (part15) .2023:12.
- [10] 邓海云,张席,杨钧博,等.我国城乡居民 CPI 的影响因素与预测研究[J].数学的实践与认识,2023,53(02):74-85.
- [11] 蔡敏.基于 SARIMA 和 BP 神经网络模型的江苏省 CPI 预测分析[D].广西师范大学,2023.
- [12] 王琳惠. 基于 SARIMAX-EEMD-LSTM 模型的 CPI 指数预测[D].山东大学,2023.
- [13] 黎庆莹,廖基源.基于 ARIMAX 模型的 PPI 与 CPI 传导机制研究[J].仲恺农业工程学院学报,2021,34(04):54-59.
- [14] 唐萍,王丰效.基于小波分析的 ARIMA-GRNN 模型的新疆 CPI 分析[J].兰州文理学院学报(自然科学版),2021,35(04):25-29+98.
- [15] 李树,汪飞,王丰效.基于 ARIMA 模型的青岛市居民消费价格指数实证分析[J].兰州文理学院学报(自然科学版),2020,34(06):11-16.
- [16] 张洋洋,邓伟,宋长钰.基于时间序列分析的山东省 CPI 预测[J].山东师范大学学报(自然科学版),2020,35(03):329-336.
- [17] 于寒梅. CPI 的组合预测实证研究[D].东北财经大学,2022.
- [18] Xin L. On the Basic Situation and Influencing Factors of CPI in China[C]//Institute of Management Science and Industrial Engineering. Proceedings of 2019 2nd International Workshop on Advances in Social Sciences (IWASS 2019). [出版者不详],2019:3.
- [19] Wan F, Zhu Y, Hu H, et al. DeepCPI: A Deep Learning-based Framework for Large-scale in silico Drug Screening[J]. Genomics, Proteomics & Bioinformatics,2019,17(05):478-495.
- [20] 许光建,马祎明.CPI、PPI 两种价格指数走势背离研究[J].价格理论与实践,2021(11):9-15.
- [21] 陈希睿. 基于时间序列分析的 CPI 预测[D].清华大学,2021.
- [22] WANG Y. Analysis for Sales Revenue and Economic Index: CPI with Business Statistics Methods[C]//Science and Engineering Research Center. Proceedings of 2018 International Conference on Education Reform, Management and Applied Social Science (ERMAS 2018). DEStech Publications,2018:5.

- [23] 赵伟霖,旷永鑫,郭兴方.基于 ARMA 模型的南阳市 CPI 预测分析[J].河南教育学院学报(自然科学版),2018,27(03):18-21.
- [24] Liu M, Duan C. The Analysis of China CPI Trend Forecast Based on ARIMA Model[C]//Wuhan Zhicheng Times Cultural Development Co., Ltd. Proceedings of 2018 International Conference on Management, Economics, Education and Social Sciences (MEESS 2018). Atlantis Press,2018:6.

附录

英文原文：

In addition, this article believes that further explanation is needed in the following aspects:

1. Data source and collection: The data used in this study mainly comes from the National Bureau of Statistics, which is responsible for collecting, analyzing, and publishing macroeconomic data at the national level. To ensure the authority and reliability of the data, data from January 2016 to February 2024 were used in the study. The selection of this time period is based on the availability and timeliness of data, aiming to capture long-term trends and cyclical changes. In addition, in order to comprehensively analyze the dynamic changes of CPI, this article also referred to data and analysis results from other relevant literature.

2. Detailed description of data analysis methods: In the process of data analysis, this article adopts various statistical models and methods. Time series analysis is one of the main methodologies, especially the ARIMA (Autoregressive Integral Moving Average) model, which combines the characteristics of autoregressive (AR), differential (I) operations, and moving average (MA) models, making it suitable for processing non-stationary time series data. Seasonal decomposition is the process of breaking down time series into trends, seasonality, and random components to clearly display the basic structure of data. In addition, to determine the optimal model parameters, extensive parameter estimation and model diagnostic tests were conducted in this paper, including residual analysis, white noise test, and heteroscedasticity test.

3. Specific analysis of the contribution of each component to CPI: By classifying and analyzing different categories of goods and services, the study found significant differences in the contribution of each component to CPI. For example, due to its fundamental position in household consumption, the price changes of food products have a particularly significant impact on CPI. The healthcare sector shows different trends due to the nature of its services and the influence of government policies. In addition, residential properties, as another important component, also exhibit price patterns closely related to economic cycles. These findings not only reveal the status of different goods and services in household consumption and their sensitivity to price fluctuations, but also provide a basis for formulating targeted economic policies.

4. Comparison of Prediction Methods and Models In order to evaluate the accuracy and practicality of prediction methods, this article compares two prediction methods: directly conducting time series analysis on CPI data and first conducting time series analysis on the various classifications that make up CPI before summing up. The results showed that the latter provided fairly accurate predictions for at least two periods. This conclusion validates the effectiveness of analyzing each component separately and conducting comprehensive evaluations, providing new ideas and methods for future research and prediction of CPI.

5. Research limitations and future research directions Although this study has achieved certain results, there are also some limitations. Firstly, the data relied upon by the research institute mainly comes from official statistical agencies and may be subject to limitations in data collection and publication. Secondly, although multiple time series analysis methods are used, the applicability and predictive ability of the model are still limited by the specific characteristics of the selected methods and data. Therefore, future research can consider introducing more data sources, such as social media data or market research data, as well as exploring more advanced

prediction models and technologies, such as machine learning algorithms, to improve the accuracy and applicability of predictions. In addition, in-depth research on the micro mechanisms and influencing factors of CPI is also an important direction for future research.

中文翻译：

此外，本文认为还需要在以下几个方面做出进一步解释说明：

1. 数据来源和收集：本研究所使用的数据主要来源于国家统计局，该机构负责收集、分析和发布国家层面的宏观经济数据。为确保数据的权威性和可靠性，研究中使用了 2016 年 1 月至 2024 年 2 月期间的数据。这一时间段的选择基于数据的可获得性和时效性，旨在捕捉长期趋势和周期性变化。此外，为了全面分析 CPI 的动态变化，本文还参考了其他相关文献中的数据和分析结果。

2. 数据分析方法详述：在数据分析过程中，本文采用了多种统计模型和方法。时间序列分析是主要的方法论之一，特别是 ARIMA（自回归积分滑动平均）模型，它结合了自回归（AR）、差分（I）操作和滑动平均（MA）模型的特点，适合处理非平稳时间序列数据。季节性分解则通过将时间序列分解为趋势、季节性和随机成分的过程，以清晰地展示数据的基本结构。此外，为了确定最佳的模型参数，本文还进行了广泛的参数估计和模型诊断测试，包括残差分析、白噪声检验和异方差性检验等。

3. 各组成部分对 CPI 贡献的具体分析：通过对不同商品和服务类别进行分类分析，研究发现各组成部分对 CPI 的贡献存在显著差异。例如，食品类由于其在居民消费中的基础性地位，其价格变动对 CPI 的影响尤为显著。医疗保健类则因其服务性质和政府政策的影响而表现出不同的变化趋势。此外，居住类作为另一个重要组成部分，也显示出与经济周期密切相关的价格变化模式。这些发现不仅揭示了不同商品和服务在居民消费中的地位及其对价格波动的敏感性，也为制定针对性的经济政策提供了依据。

4. 预测方法和模型比较 为了评估预测方法的准确性和实用性，本文对比了直接对 CPI 数据进行时间序列分析和先对组成 CPI 的各个分类进行时间序列分析再求和的两种预测方法。结果显示后者至少在两期之内提供了相当准确的预测结果。这一结论验证了分别分析各个组成部分并进行综合评估的有效性，为未来 CPI 的研究和预测提供了新的思路和方法。

5. 研究局限性和未来研究方向 尽管本研究取得了一定的成果，但也存在一些局限性。首先，研究所依赖的数据主要来源于官方统计机构，可能受到数据收集和发布的限制。其次，虽然采用了多种时间序列分析方法，但模型的适用性和预测能力仍受限于所选方法和数据的特定特征。因此，未来的研究可以考虑引入更多的数据源，如社交媒体数据或市场调查数据，以及探索更多先进的预测模型和技术，如机器学习算法等，以提高预测的准确性和适用性。此外，对于 CPI 的微观机制和影响因素的深入研究也是未来研究的重要方向。

谢 辞

在本论文即将圆满结束之际，我满怀感激之情，向所有在学习和研究旅程中给予我支持与帮助的人表达我诚挚的谢意。

首先，我要向尊敬的导师吴中元教授致以崇高的敬意和深深的感谢。吴教授不仅以其严谨的学术态度和深厚的专业知识为我树立了学习的榜样，而且在论文撰写的每一个阶段都给予了我极具价值的指导和建议。他耐心地解答了我提出的每一个问题，帮助我克服了研究过程中的重重困难，让我在学术探索的道路上不断前行。

我还要向同宿舍的伙伴们还有我所遇到的不同专业不同年级的同学们表达我的感激之情。在这段共同生活和学习的日子里，你们不仅在学术上给予了我巨大的帮助，更在生活上给予了我温暖的关怀和支持。我们共同经历的点点滴滴，将成为我人生中最宝贵的记忆和财富。

同时，我要向家人表达我最深切的感激。感谢你们始终如一的爱与理解，是你们的支持和鼓励让我得以心无旁骛地投入到学业中。你们的信任和支持是我不断前进的动力源泉。

此外，我还要感谢天津工业大学为我提供了一个充满机遇和挑战的学习环境。学校的优质教育资源和先进的设施，让我得以在知识的海洋中自由探索，为我未来的职业发展奠定了坚实的基础。

在未来的岁月里，我将携带着这份感激和收获，继续努力，不断进步，以期不辜负所有给予我帮助和期望的人。

最后，我再次向所有关心和帮助过我的人表示衷心的感谢，是你们让我的学习之路不再孤单。感谢你们！

+