

消费者信心指数影响因素分析——实证研究

2021110809 权宰铉

摘要

本研究旨在分析影响消费者信心指数的主要因素，并评估这些因素在预测 CCI 中的有效性。研究首先采用皮尔逊相关系数、Lasso 回归及递归特征消除法（RFE）对众多潜在影响因素进行筛选，确定关键特征。随后，利用线性回归模型、决策树模型及随机森林模型对选定特征进行训练与预测，以考察不同模型的预测性能。进一步地，通过决策树和随机森林模型的特征重要性评分，评估各变量对 CCI 的贡献程度。最后，基于重要特征，采用长短期记忆网络（LSTM）和多层感知器（MLP）等深度学习模型作为基准模型，综合分析这些因素在预测消费者信心指数中的表现。研究结果将为理解消费者信心的驱动因素及其预测提供实证依据。

关键词： 消费者信心指数 机器学习模型

引言

消费者信心指数（Consumer Confidence Index, CCI）作为衡量消费者对当前经济状况及未来经济预期的重要指标，一直以来在经济研究与政策制定中扮演着关键角色。高水平的消费者信心通常预示着消费支出的增加，而消费支出则是推动经济增长的主要动力之一。因此，理解和分析影响消费者信心的因素，对于促进经济稳定与增长具有重要意义。近年来，尤其是在新冠疫情爆发之后，中国的消费者信心指数持续低于 100。这一现象反映出消费者对经济前景的担忧和不确定性，进而影响了消费行为和经济活力。疫情期间，尽管政府采取了一系列刺激经济的措施，但消费者信心恢复缓慢，部分原因在于疫情带来的长期经济影响、就业压力以及收入不稳定等问题。

根据国家统计局的数据，2023 年以来，中国的消费者信心指数多次低于历史平均水平，显示出消费者对未来经济发展的信心不足。这种低迷的信心不仅抑制了消费需求的释放，也对整体经济的有效需求产生了负面影响。有效需求的不足可能导致经济增长放缓，甚至引发通货紧缩的风险，进而影响社会稳定和民生改善。扩大有效需求、提振消费者信心已成为当前中国经济政策的重要目标。提高消费者信心有助于激发消费潜力，推动消费结构升级，进而促进经济的高质量发展。为实现这一目标，必须深入分析和理解影响消费者信心的多种因素，从而制定针对性的政策措施。

在研究方法上，学者采用的方法众多，包括机器学习模型和传统的经济模型等。机器学习方法包括：孙景（2021）等以网络搜索关键词指数数据和网络新闻文本数据为基础，运用自适应提升树回归、Lasso 回归和多层感知机回归等机器学习方法，通过滑动时间窗口筛选变量预测 CCI。研究发现不同模型在不同提前期各有优劣，且其提出的关键词滚动筛选及模型滚动更新方法能提升模型对互联网热词变化的适应性和预测准确率。唐晓彬

（2020）等基于 LSTM 神经网络模型结合网络搜索数据构建 LSTM&US 预测模型，用于 CCI 的长、中、短期预测。结果显示引入网络搜索数据可提高 LSTM 模型预测性能，LSTM&US 模型泛化能力佳且精度优于其他基准模型。邹鸿飞（2019）等建立 CEEMD-DEGWO-BPNN 模型预测 CCI，经检验该模型较对比模型精度更高、泛化能力更强，能更好捕捉 CCI 变化规律。也有很多学者考虑使用传统的计量方法，包括：王世文（2020）基于“有限理性”理论，运用向量自回归模型研究 CCI 与居民消费价格指数（CPI）关联效应，发现 CCI 对 CPI 有预测作用但影响有限，而 CPI 变动对 CCI 影响显著且与时间相关。王岱（2016）

采用向量自回归和方差分解等方法分析 CCI 影响因素，表明消费者价格指数有负向影响且有时滞，制造业和金融业发展有正向影响，农村居民收入影响更大。徐国祥（2018）对 CCI 与 CPI 月度数据进行格兰杰因果检验、交叉谱分析和 MS - VAR 模型探究，发现 CCI 是 CPI 的格兰杰原因且领先 3.11 个月，二者存在非线性关系和循环传导路径。陈晓菲（2020）运用列昂惕夫天际图分析文化产业结构升级与 CCI 关联，发现产业升级能提升消费者文化消费满意度和体验度，促进 CCI 正向攀升，且 CCI 状态影响产业结构升级战略重心。

本研究旨在通过实证分析，探讨影响中国消费者信心指数的主要因素，并评估这些因素在预测消费者信心中的有效性。研究将采用皮尔逊相关系数、Lasso 回归、递归特征消除法等多种特征选择方法，筛选出关键影响因素；随后，利用线性回归、决策树、随机森林等多种模型进行预测，并通过特征重要性分析确定各因素的贡献程度。最后，采用深度学习模型如 LSTM 和 MLP 进一步验证这些因素在时间序列预测中的表现。本研究期望能够为理解和提升消费者信心提供实证依据，为政策制定者制定有效的经济刺激措施提供参考，从而促进中国经济的稳定与可持续发展。

方法和数据集

总体来讲，我们的思路是首先利用特征工程筛选出与消费者信心指数密切相关的一些特征，之后再利用决策树模型等进行特征重要性排序进一步优化特征的筛选，最后再利用深度学习模型进一步拟合这些因素对于 CCI 的非线性影响关系。

（一）总体方法思路

一、特征筛选

- 1、皮尔逊相关系数：计算各潜在影响因素与消费者信心指数之间的相关性，初步筛选相关性较高的变量。
- 2、Lasso 回归：利用 Lasso 回归的正则化特性，进一步筛选出对 CCI 具有显著影响的特征，减少模型复杂度。
- 3、递归特征消除法（RFE）：通过递归地训练模型并消除最不重要的特征，最终确定最优的特征子集。

二、模型训练与预测

- 1、线性回归模型：作为基线模型，评估线性关系在预测 CCI 中的表现。
- 2、决策树模型：捕捉非线性关系，评估其在预测中的效果。
- 3、随机森林模型：通过集成多棵决策树，提升预测的准确性和稳定性。

对比各模型的预测性能，选取表现最佳的模型作为进一步分析的基础。

三、特征重要性分析

- 1、利用决策树和随机森林模型的特征重要性评分，评估各变量在预测 CCI 中的贡献。
- 2、排序并识别出最具影响力的关键因素，为后续深度学习模型的特征选择提供依据。

四、深度学习模型应用

- 1、长短期记忆网络（LSTM）：处理时间序列数据，捕捉 CCI 的时序特征。
- 2、多层感知器（MLP）：构建非线性映射关系，提升预测的复杂性和准确性。
- 3、基于重要特征，训练 LSTM 和 MLP 模型，评估其在预测 CCI 中的综合表现，并与传统模型进行对比分析。

（二）数据来源与变量处理

一、Akshare API 接口调用

通过 Python 编程环境，利用 Akshare 提供的 API 接口获取各类相关经济指标和市场数据，这些数据涵盖消费者信心指数及其潜在影响因素，包括 GDP 数据、失业率、CPI、PMI、上证 50 收益率、进出口情况、固定资产投资、社会零售品销售额等多角度的宏观数据。数据时间是从 2016 年 1 月到 2024 年 9 月的数据。

二、数据预处理

对获取的数据进行清洗和整理，主要包括缺失值处理、数据标准化及时间序列对齐，确保数据的完整性和一致性。其中除了 GDP 数据是季度数据、股票收益率数据是月度数据外，其余数据均为月度数据，因此必须要对齐时间序列。对于 GDP 这种季度数据，我们选择向前填充，来补齐数据；而对于股票收益率数据，我们选择取每个月的平均收盘价作为当月的上证 50 指数的收益率。同时，也对数据进行差分等处理（简便起见选择取增长率，没有选择对数增长率），从而保证我们的模型更加稳健。最后对所有的特征值都进行标准化操作，防止因为单位差异造成的预测性能的偏差。

三、数据集构建

根据我们的需要，将处理后的数据划分为训练集和测试集，其中训练集为 80% 的数据，测试集为剩余 20% 的数据，利用程序随机抽取，确保模型训练与评估的科学性。数据类型均为连续型变量，不存在类别变量，描述性统计如表 1: 描述性统计量所示。其中 CCI 表示消费者信心指数，CSI 表示消费者满意指数，CEI 表示消费者预期指数，CPI 表示消费者价格指数，UCPI、RCPI 分别表示城市 CPI 和农村 CPI，D1ind、D2ind、D3ind 分别表示第一产业 GDP 环比增长率、第二产业 GDP 环比增长率、第三产业 GDP 环比增长率。Retail 表示社会零售品销售额、dexp 和 dimp 分别表示进出口的环比增长率，dclose 表示上证 50 指数收盘价变化率。

表 1: 描述性统计量

变量	计数	mean	std	min	25%	50%	75%	max
cci	106	109.2	85.5	89.2	116.1	122.2	127.6	15.1
dcci	106	-0.1	-23.4	-1.1	0.2	1.4	4.7	3.2
csi	106	105.3	84.2	88.9	110.6	117.2	121.0	13.3
dcsi	106	-0.1	-21.6	-1.0	0.2	1.3	4.1	2.9
cei	106	105.3	84.2	88.9	110.6	117.2	121.0	13.3
dcei	106	-0.1	-25.3	-1.4	0.2	1.4	7.0	3.5
cpi	106	101.6	99.2	100.7	101.8	102.3	105.4	1.2
dcpi	106	0.1	-1.2	-0.2	0.1	0.4	1.6	0.5
ucpi	106	101.6	99.2	100.8	101.8	102.3	105.1	1.1
ducpi	106	0.1	-1.2	-0.2	0.1	0.4	1.6	0.5

rcpi	10 6	101.6	99.2	100.5	101.7	102.3	106.3	1.5
drcpi	10 6	0.1	-1.3	-0.2	0.1	0.4	1.6	0.5
dgdg	10 5	5.6	-6.9	4.8	6.1	6.8	18.3	3.9
dlin	10 5	3.7	-3.1	3.1	3.5	4.1	8.1	1.9
d2in	10 5	5.5	-9.7	4.3	5.8	6.1	24.4	4.9
d3in	10 5	6.0	-5.4	4.6	7.2	7.7	15.6	3.7
retail	10 6	34107.6	24645.8	30415.9	34237.0	37913.4	45396.0	4862.6
dretail	10 6	0.8	-13.9	-2.3	1.5	5.1	15.5	6.4
dgdzctz	10 6	-2.7	-65.2	-10.1	-3.5	6.5	32.1	18.3
dexp	10 6	2.0	-62.0	-2.4	1.8	6.2	129.6	19.5
dimp	10 6	0.9	-31.2	-4.6	1.0	6.2	39.2	11.8
close	10 6	3133.0	2558.4	2949.6	3135.1	3292.5	3627.2	247.2
volume	10 6	2615988 9266.3	1130715 2368.4	1818184 6261.4	2600502 9138.4	3264031 6875.1	6967811 3377.8	9909275 067.3

实证建模及结果分析

（一）特征选择

为了研究消费者信心的影响因素，我们借鉴机器学习领域的特征工程的办法，首先考虑利用皮尔逊相关系数进行特征的筛选，计算变量两两之间的相关系数，并绘制出热力图如图 1 所示。我们选择其中最优的 8 个特征，得到如下结果：'cpi'，'dcpi'，'ucpi'，'rcpi'，'drcpi'，'dgdg'，'d2ind'，'d3ind'。第二步再利用递归消除法（RFE）进行特征的选择，得到如下的结果：'cpi'，'dcpi'，'ducpi'，'drcpi'，'dgdg'，'d2ind'，'d3ind'，'dclose'。第三步再利用 Lasso 回归，即使用带 L1 惩罚项的逻辑回归作为基模型筛选出如下的特征：'ucpi'，'ducpi'，'drcpi'，'dlin'，'d2ind'，'d3ind'，'dretail'，'dgdzctz'，'dexp'，'dimp'。为了尽可能多地保留信息，我们选择使用这三者的并集作为初步的特征选择结果，即 CPI、CPI 环比增长率、第一产业 GDP 环比增长率、第二产业 GDP 环比增长率、第三产业 GDP 环比增长率、上证 50 指数收盘价增长率、出口环比增长率、GDP 环比增长率、固定资产投资环比增长率、进口增长率、农村 CPI 环比增长率、社会零售品综合环比增长率、城镇 CPI 环比增长率、农村 CPI、城市 CPI 环比增长率等共 15 个变量。

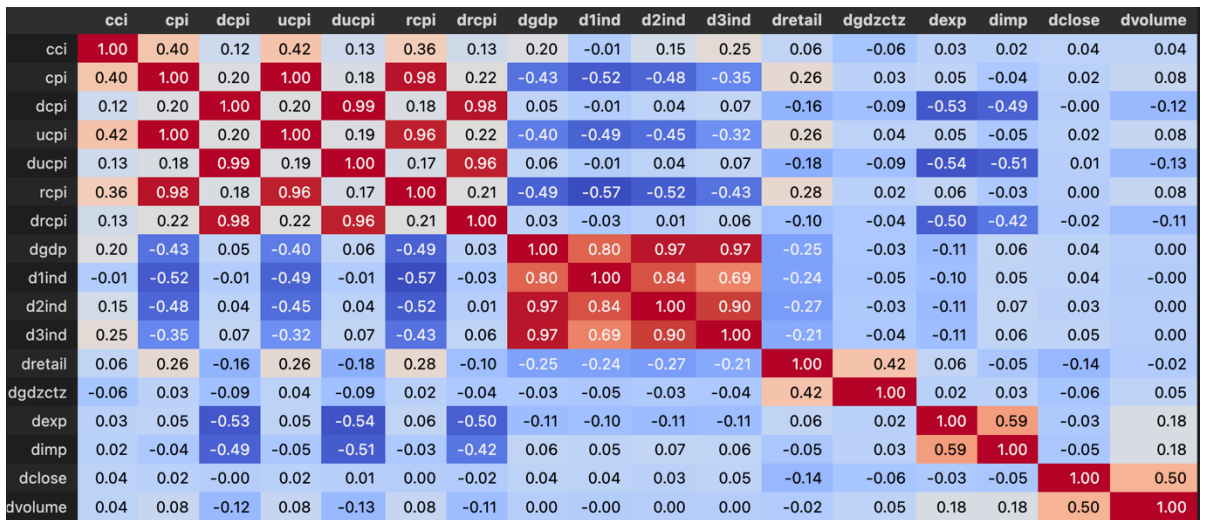


图 1： 变量之间相关系数（热力图）

（二）模型训练结果

下面利用我们前面所选择的 15 个特征，进行模型的训练。由于是回归任务，我们选取 MSE（均方误差）和 R 方（拟合优度）作为主要的考察指标。对于模型的选择，主要是机器学习模型包括线性回归、决策树回归、随机森林回归、K 邻近回归、SVM 回归以及深度学习模型包括前馈神经网络和 LSTM。

对于模型的训练结果，如表 2 所示，我们可以明显看出传统机器学习模型中的决策树回归和随机森林回归模型有较强的表现，深度学习模型中的 LSTM 有较好的表现。OLS 回归的拟合能力较差，这也在一定程度表明 CCI 和他的影响因素之间并非完全的线性关系，要考虑这些非线性关系的影响。

表 2：模型训练结果

模型分类	模型名称	均方误差	拟合优度	综合排名
机器学习模型	决策树回归模型	15.03	0.93	1
	随机森林回归模型	34.78	0.83	2
	线性回归模型	220.11	-0.09	4
	K 邻近回归	236.47	-0.17	5
	SVM 回归模型	243.58	-0.21	6
深度学习模型	LSTM	103.49	-52.06	3
	MLP	449.24	-1.23	7

同时我们也对决策树回归模型和随机森林回归模型中的特征重要性进行了打分，如表 3 所示，我们发现第三产业 GDP 环比增长率、第一产业 GDP 环比增长率、GDP 环比增长率、

第二产业 GDP 环比增长率、社会零售品销售总额环比增长率、CPI、农村 CPI 环比增长率、上证 50 指数月度增长率、固定资产投资环比增长率这 9 个因素重要性得分较高，因此我们下一步主要考虑利用这 9 个因素进一步改进模型，考察拟合效果是否有所提升，同时我们也不会再考察那么多模型，而会选择聚焦于重点关注的这三个表现良好的模型，即决策树回归模型、随机森林回归模型、LSTM 模型。

表 3：特征重要性得分

特征	排名	决策树模型	随机森林模型	平均分
d3ind	1	0.485268	0.235364	0.360316
d1ind	2	0.396715	0.15101	0.273863
dgdg	3	0.084826	0.344436	0.214631
d2ind	4	0.025606	0.09372	0.059663
dretail	5	0.001323	0.029376	0.01535
cpi	6	0	0.025271	0.012636
rcpi	7	0	0.023937	0.011969
dclose	8	0.004128	0.01499	0.009559
dgdzctz	9	0	0.01877	0.009385
ducpi	10	0.002133	0.013395	0.007764
dimp	11	0	0.012936	0.006468
dexp	12	0	0.011721	0.00586
dcpi	13	0	0.009237	0.004618
ucpi	14	0	0.009042	0.004521
drcpi	15	0	0.006795	0.003398

（三）改进模型预测表现

改进了特征选择之后，我们选择了最为关键的 8 个要素进行对模型的训练，我们发现决策树回归模型和随机森林模型的改进都比较明显，均方误差和拟合优度都有所提高。相比于这两个模型，LSTM 的效果并没有明显改变，这也可以解释，因为我们是从决策树模型和随机森林的上一步结果中筛选出的重要特征，所以可能对于 LSTM 的改进意义不大。从绝对数值来看，LSTM 的预测能力不如树模型，这也启示我们并非越复杂的模型的效果越好，要综合尝试并结合实际问题进行比较。

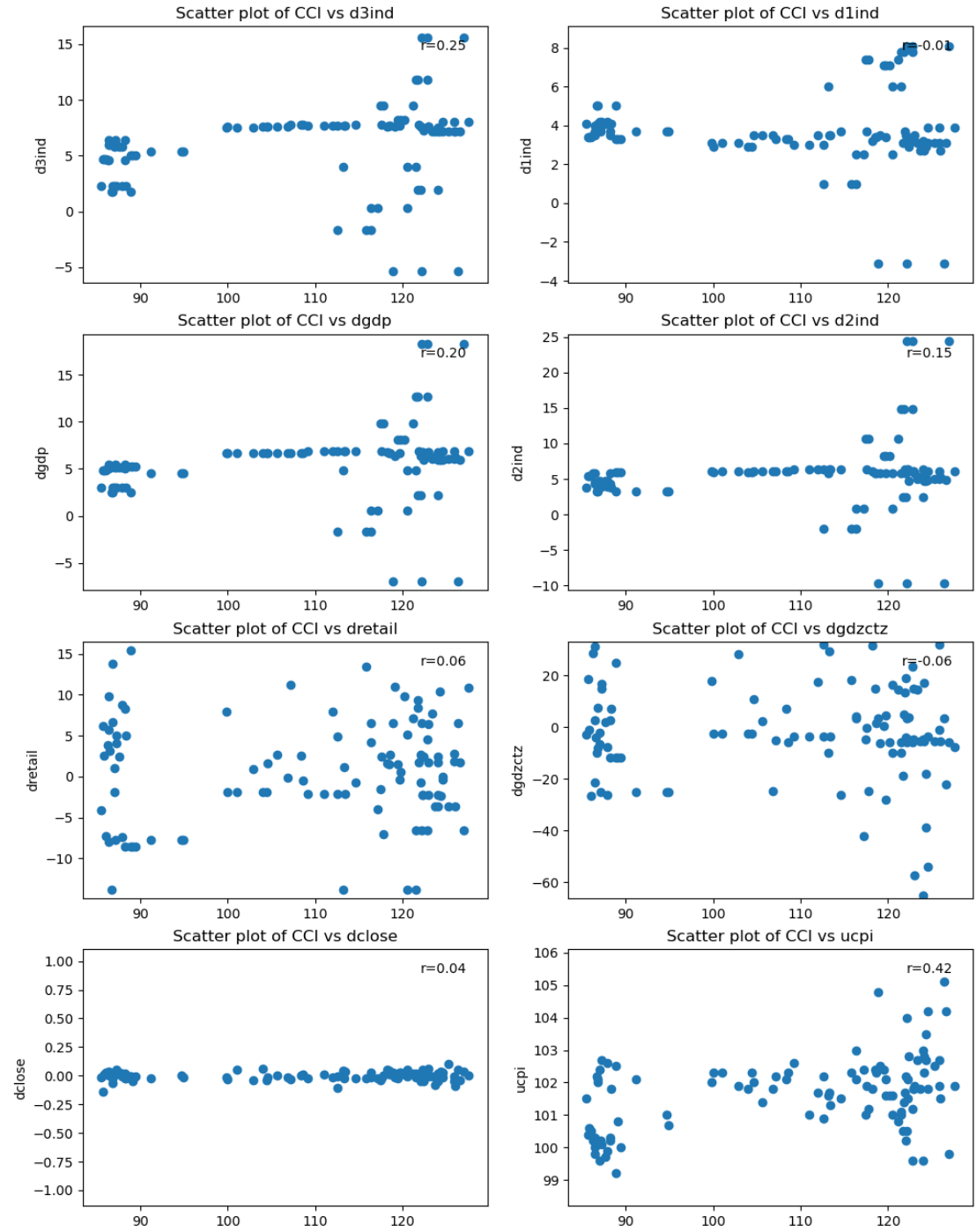
表 4:改进模型预测表现

模型名称	均方误差 1	拟合优度 1	均方误差 2	拟合优度 2
决策树回归模型	15.03	0.93	12.69	0.94
随机森林回归模型	34.78	0.83	26.88	0.87
LSTM	103.49	-52.06	103.72	-52.30

（四）相应的政策建议

我们首先需要考察 CCI 与这些因素之间的关系，主要是看是正相关还是负相关，下面我将绘制出相应的散点图以及皮尔逊相关系数。我们可以看到 CCI 和 CPI、GDP、GDP 环比

增速等为正相关，与固定资产投资为负相关，但这种相关性并不明确。总体来讲，GDP 增速、第三产业 GDP 环比增长率以及城镇 CPI 环比增长率之间的正相关性较为明显。因此从这个角度出发，为了提振居民消费者信心，应当主要考虑提升 GDP 增长能力以及适当的通胀。GDP 增速越快，反映居民收入相应增高；CPI 高反映市场上流动性更多，政府出台了更多的有利的政策，从而提升消费者信心。



结论

通过应用机器学习领域中的特征工程的相关方法，我们筛选出了影响消费者信心指数的几个重要因素，主要包括 CPI、GDP、进出口等，并且通过比较线性回归、决策树模型、KNN 回归等传统机器学习模型以及 LSTM 和 MLP 等深度学习模型，发现决策树模型、随机森

林模型、LSTM 在拟合 CCI 上有较好的表现。这些因素的筛选有助于政策制定部门更加合理地预测消费者信心指数，能够为提振消费者信心提供强有力的支持。

参考文献

- [1] 孙景, 朱建霖, 李挽澜, 等。基于互联网数据的消费者信心指数滚动预测研究 [J]. 西安交通大学学报 (社会科学版), 2021, 41 (06): 68 - 77. DOI:10.15896/j.xjtuskxb.202106007.
- [2] 唐晓彬, 董曼茹, 张瑞。基于机器学习 LSTM&US 模型的消费者信心指数预测研究 [J]. 统计研究, 2020, 37 (07): 104 - 115. DOI:10.19343/j.cnki.11 - 1302/c.2020.07.009.
- [3] 邹鸿飞, 王建州。一种基于差分灰狼算法的消费者信心预测指数设计 [J]. 数量经济技术经济研究, 2019, 36 (02): 120 - 134. DOI:10.13653/j.cnki.jqte.2019.02.007.
- [4] 王世文, 沈阳。消费者信心指数与居民消费价格指数关联效应研究 —— 基于 “有限理性” 理论的分析 [J]. 价格理论与实践, 2020, (03): 83 - 86. DOI:10.19851/j.cnki.cn11 - 1010/f.2020.03.374.
- [5] 王岱, 程灵沛, 祝伟。我国消费者信心的影响因素分析 [J]. 宏观经济研究, 2016, (04): 48 - 61. DOI:10.16304/j.cnki.11 - 3952/f.2016.04.006.
- [6] 徐国祥, 刘璐。中国消费者信心指数与居民消费价格指数的关系研究 [J]. 统计与决策, 2018, 34 (23): 5 - 10. DOI:10.13546/j.cnki.tjyjc.2018.23.001.
- [7] 陈晓菲, 韩平。文化产业结构升级与消费者信心指数的关联分析 [J]. 统计与决策, 2020, 36 (02): 89 - 92. DOI:10.13546/j.cnki.tjyjc.2020.02.019.