

基于安徽省人口特征的产业结构升级预测研究

张起望

安徽财经大学, 中国·安徽 蚌埠 233030

【摘要】本文基于安徽省2001-2020相关人口特征数据,采用自适应增强算法预测了2021-2035产业结构升级态势的变化,结果显示,在保持当前人口空间分布、物质条件差异、年龄结构及人力资本变化前提下,产业结构升级速率的波动将有所收敛,虽然仍会受到衰退冲击,但与前期相比上升的稳定性加强,而产业转型步伐也将愈发稳健。纳入空间因素后,若城乡收入差距在地区间表现的不平衡持续加剧,会严重阻碍产业升级,甚至造成衰退。因此在解释产业升级特征前后区别的同时,提出缓解负面冲击因素的对策建议。

【关键词】人口特征; 产业结构升级; 自适应增强

Forecast of Industrial Structure Upgrading Based on Population Characteristics of Anhui Province

Zhang Qiwang

Anhui University of Finance and Economics, Bengbu, Anhui 233030

[Abstract] Based on the relevant population characteristic data of Anhui Province from 2001 to 2020, this paper uses the adaptive enhancement algorithm to predict the change of the upgrading trend of the industrial structure from 2021-2035. The results show that the fluctuation of the upgrading rate of the industrial structure will converge on the premise of maintaining the current spatial distribution of population, material differences, age structure and changes in human capital. Although it will still be impacted by the recession, the stability of the rise compared with the previous period will be strengthened. The pace of industrial transformation will also become more stable. After the spatial factor is included, if the urban-rural income gap between regions continues to increase, it will seriously hinder industrial upgrading, and even cause recession. Therefore, while explaining the differences between the characteristics of industrial upgrading before and after, the paper puts forward countermeasures and suggestions to mitigate the negative impact factors.

[Keywords] Demographic characteristics; Industrial structure upgrading; Adaptive enhancement

【基金项目】2022年度安徽财经大学本科生科研创新基金项目(XSKY22089); 居民杠杆率、住房价格与产业结构升级的联动机制研究。

1 引言

1.1 研究背景

产业结构的变动与人口迁移、居民生活水平差异、人力资源分布、人员素质等多种因素息息相关,通过挖掘人口数据的相关特征与产业转型升级的深层次关联,探索更为有效的预测路径,有利于从社会福利改善、产业政策创新、公共资金投入等角度,促进新兴工业化、农业现代化及城镇化进程,实现2035远景目标。

1.2 文献综述

从作用机理、差异来看,城镇化率的提高往往表现为大量劳动力在城市的集聚,人才集聚有利于生产技术的创新,从而推动城市产业结构优化(Gibbs, 1992),同时人口集聚也带来多样化的消费需求,促使第三产业发展(Kolko, 2010);产业结构升级有利于缩小城乡收入差距,而城乡收入差距的扩大不利于产业结构的高级化发展(黄可人等, 2016^[1]),抑或是产业结构调整与城乡收入差距呈现“倒U”形或“U”形关系(穆怀中等, 2016^[2]);人力资本通过学习效应产生动态规模经济促进产业结构升级(朱承亮等, 2011^[3]),而高级人力资本较中级人力资本对产业转型升级促进作用更强,低级人力资本则会产生抑制效果(李敏等, 2019^[4]);有学者认为人口老龄化通过减少创新型人才流动而阻碍产业结构的升级(卓乘风等, 2018^[5]),但也有人指出人口老龄化会倒逼企业用资本和技术替代劳动力从而推动了产业结构升级(汪伟, 2015^[6])。从实证方法来看,已有研究利用PVAR

模型研究人口老龄化、城乡收入差距和产业结构升级三者的动态增长关系(赵敏等, 2022^[7]);也有采用工具变量2SLS方法,分阶段、分区域分析了城乡收入差距对产业结构升级的影响(程玉鸿等, 2021^[8])。

1.3 方法创新

大多数学者聚焦于论证这几类变量的相互关联及影响路径,但较为缺乏基于现有数据对未来一定时间内的产业升级状况作出合理预测与解释的研究案例,本文以安徽省近二十年数据为例,采用监督学习算法,利用核心人口特征预测本省产业结构高级化进程,并就省域范围内地级市之间城乡收入差距差异化的变动以及城镇化率的不平衡如何影响产业优化的进程,做出假设推演。

2 模型设定、变量处理及数据说明

2.1 模型设定

自适应增强算法实现了多棵基础决策树的加权运算,以分类问题为例,若设基底弱学习器为二叉树(CART),则模型为:

$$F(x) = \sum_{m=1}^M a_m f(x)_m = F(x)_m - 1 + a_m f(x)_m$$

其中 $F(x)$ 是由 M 颗基础决策树构成的最终提升树, $F(x)_{m-1}$ 表示经过 $m-1$ 轮迭代后的提升树, a_m 为第 m 棵基础决策树所对应的权重, $f(x)_m$ 为第 m 棵决策树。基于前一棵基础决策树的分类结果对样本点设置不同的权重,如果在前一棵基础

决策树中将某样本点预测错误,则会增大该样本点的权重,否则会降低样本权重,进而再构建下一棵基础决策树。

进一步构建指数型损失函数:

$$L(y, F(x)) = \exp(-yF(x)) = \exp(-y(F(x)_{m-1} + a_m f(x)_m))$$

若将所有训练样本点代入损失函数后一定存在最佳 a_m 和 $f(x)_m$, 使得函数取最小值, 将函数改写为:

$$(a_m, f(x)_m) = \arg \min_{a, f(x)} \sum_{i=1}^M P_{mi} \exp(-y_i a_m f(x_i)_m)$$

其中 $P_{mi} = \exp(-y_i F(x_i)_{m-1})$, $y_i f(x_i)_m = 1$, 当且仅当第 m 棵基础决策树准确预测时成立, 否则为 -1, 所以最佳 $f(x)_m$ 表示为:

$$Bestf(x)_m = \arg \min_f \sum_{i=1}^N P_{mi} I(y_i \neq f(x)_m)$$

其中 f 表示所有可用的基础决策树空间, I 表示预测结果与实际值不相等时返回 1, 进而将损失函数改写为:

$$L(y, F(x)) = (\exp(-a_m) - \exp(-a_m)) \sum_{i=1}^N P_{mi} I(y_i \neq$$

$$f(x)_m) + \exp(-a_m) \sum_{i=1}^N P_{mi}$$

对 a_m 求偏导, 令偏导数为 0 得:

$$Basica_m = \frac{1}{2} \log \frac{1 - e_m}{e_m} = \frac{\sum_{i=1}^N P_{mi} I(y_i \neq f(x)_m)}{\sum_{i=1}^N P_{mi}}$$

该式即为基础决策树权重, 其中 $e_m = \frac{\sum_{i=1}^N P_{mi} I(y_i \neq f(x)_m)}{\sum_{i=1}^N P_{mi}}$ 表示基础决策树 m 的错误率, 则经过单位化处理的第 $m+1$ 轮基础决策树中样本点权重为:

$$\omega_{m+1,i} = \frac{\omega_{mi} \exp(-y_i Basica_m Bestf(x)_m)}{\sum_{i=1}^N \omega_{mi} \exp(-y_i Basica_m Bestf(x)_m)}$$

2.2 变量处理

2.2.1 数据来源

由于 2022 年《安徽统计年鉴》尚未发布, 2021 年部分细节数据难以获取, 选取 2001-2020 年安徽省时间序列数据构建模型, 将 2021 年作为预测年份, 并加入预测数据集 2021-2035 年; 同时选取省内 16 个地级市 2010-2020 年的相关数据, 加入修正后的建模数据集中, 用于衡量预测结果波动。

2.2.2 解释变量

城镇化率 (Ur): 反映了人口由农村向城市的迁移趋势, 是体现人口分布地理特征的关键指标, 选取安徽省 2001-2020 年末常住人口城镇化率作为人口空间分布指标。

泰尔指数 (URIG):

$$URIG_{i,t} = \sum_{j=1}^2 \frac{Y_{ij,t}}{Y_{i,t}} \ln \left(\frac{\frac{Y_{ij,t}}{Y_{i,t}}}{\frac{N_{ij,t}}{N_{i,t}}} \right)$$

$URIG_{i,t}$ 表示地区 i 在 t 时期的泰尔指数, $j=1, 2$ 分别表示城镇和农村地区, $Y_{ij,t}$ 表示 i 地区城镇 ($j=1$) 或农村 ($j=2$) 在 t 时期的总收入, 总收入由总人口与居民可支配收入之积表示, $Y_{i,t}$ 表示地区 i 在 t 时期的总收入, $N_{ij,t}$ 表示城镇 ($j=1$) 或农村 ($j=2$) 在 t 时期的总人口, $N_{i,t}$ 表示地区 i 在 t 时期的总人口。城乡收入差距可以衡量居民消费水平的地区差异, 可作为人口物质条件差异化指标, 而泰尔指数既考虑了城乡收入分布,

又兼顾了城乡人口结构, 数值与收入差距成正比。

老年抚养比 (Odr): 老年人口占劳动年龄人口比例, 衡量由于人口年龄结构变动而造成的人力资本负担变化、劳动力市场供给质量与数量的演变, 还可以体现相关市场规模的变动对总体产业造成的结构性需求冲击,

人均受教育年限 (Asy): 体现人口素质提高所产生的人力资本高级化倾向。

2.2.3 被解释变量

产业结构升级指数变动率 (Crisui):

$$Crisui_t = \frac{\Delta Isui_{t,t-1}}{Isui_{t-1}} = \frac{\Delta \left(\sum_{i=1}^3 L_i stru_i \right)_{t,t-1}}{\left(\sum_{i=1}^3 L_i stru_i \right)_{t-1}} (i=1, 2, 3)$$

产业结构升级指数 (Isui) 是反映产业结构高级化的状态性指标, 以各产业增加值在 GDP 增加值中的比重, 即三次产业拉动率与各产业的劳动生产率的乘积之和来表示, 而本文以年为数据尺度, 以该指数的环比增长速度作为新指数以衡量产业结构升级的速率和变动方向, 其中 L_i 为各产业的劳动生产率, 用各产业的增加值除以就业人数来衡量, $stru_i$ 为三次产业拉动率。

2.2.4 变量分析

对 2001-2020 年数据进行描述性统计如下, 结果取三位小数:

表 1 描述性统计

变量	个案数目	最小值	最大值	均值	标准差
Ur	20	0.294	0.583	0.438	0.089
URIG	20	0.082	0.171	0.128	0.032
Odr	20	12.09	22.83	16.356	2.618
Asy	20	7.009	9.350	8.275	0.794
Crisui	20	-0.471	1.877	0.242	0.632

对四组解释变量进行 Shapiro - Wilk 正态分布检验如下:

表 2 正态分布检验

	统计	自由度	显著性
Ur	0.962	20	0.589
URIG	0.882	20	0.019
Odr	0.964	20	0.622
Asy	0.926	20	0.127

由于原假设为随机变量服从正态分布, 备择假设为随机变量不服从正态分布, 四组变量的显著性 P 值仅有泰尔指数小于 0.05, 因此接受原假设, 并计算得到相关系数如下:

表 3 皮尔逊积矩相关系数

		Ur	URIG	Odr	Asy
Ur	coefficient	1	-0.840**	0.919**	0.992**
	Sig.		0	0	0
URIG	coefficient	-0.840**	1	-0.705**	-0.857**
	Sig.	0		0.001	0
Odr	coefficient	0.919**	-0.705**	1	0.878**
	Sig.	0	0.001		0
Asy	coefficient	0.992**	-0.857**	0.878**	1
	Sig.	0	0	0	

注: ** 在 0.01 级别 (双尾), 相关性显著。

各解释变量均在 99% 的置信水平上通过 t 检验且有的表现, 具有显著的线性相关关系, 而解释与被解释变量间并不具有明显的相关关系, 不适合使用线性回归对变量进行预测, 观察以年份为自变量的 $Crisui$ 与 $Isui$ 变化曲线, 虽然 $Isui$ 随时间演

变呈现出大致的波动上升倾向,但Crisui更多表现为振幅随机的无序波动,无法直接通过时间序列分析测得未来变动,若间接使用Isui预测Crisui走势,则容易由于拟合优度过低、回归系数显著性较弱而出现预测精度差的问题。但时间与解释变量间呈现显著相关,将年份转变为1-20的数量尺度以方便运算,分别对此拟合四条回归曲线:

$$Ur = 0.015Time + 0.2812, URIG = -0.0046Time + 0.1762 \\ Odr = 0.4054Time + 12.098, Asy = 0.1333Time + 6.8756$$

分别进行拟合优度检验得到四组分别为0.9988、0.7203、0.9862、0.8392,据此求得2021-2035的15组估计值。

3 实证分析

3.1 分类识别与趋势估计

利用哑变量对2001-2020已知的Crisui数据集进行标注,大于等于0的值记为1,小于0的值记为0,在python3.7中调用AdaBoostClassifier类,使用CART分类树作为基础弱分类器,构造指数型损失函数:

$$e_{ki} = -\exp\left(\frac{-y_i + f_k(x_i)}{E_k}\right) (E_k = \max|y_i - f_k(x_i)|)$$

其中,表示第k个基础分类器在训练样本点上的最大误差,以25%的测试集比例将数据集拆分为训练集和测试集并进行拟合,返回评估结果如表所示:

表4 模型评估结果

评估类别	计算方式	实际数值
预测准确率 (Accuracy)	$TN+TP/TN+FN+FP+TP$	81.25%
正例覆盖率 (Sensitivity)	$TP/FN+TP$	32%
正例命中率 (Precision1)	$TP/FP+TP$	68%
负例覆盖率 (Specificity)	$TN/TN+FP$	96%
负例命中率 (Precision2)	$TN/TN+FN$	83%

注:TN:正确预测为负例的样本个数,FN:预测为负例但实际为正例的样本个数,FP:预测为正例但实际为负例的样本个数,TP:正确预测为正例的样本个数

以1-Specificity作为x轴,Sensitivity作为y轴,绘制不同阈值下的受试者工作特征曲线,其AUC为0.79,模型效果一般,通过绘制自变量的重要性对比图进行特征筛选,发现Odr与URIG的数值均>0.3,将其提取拟合新模型并进行参数寻优,经过5重交叉验证后,确定树的最大深度为6,最佳基础分类器个数为100,迭代步长为0.01,调优后的模型ROC曲线AUC为0.93,拟合效果较好。

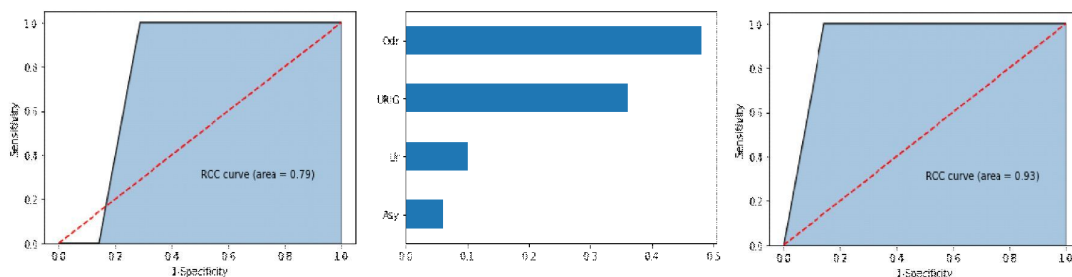


图1 修正前后的ROC曲线及因子重要性

使用修正后的分类模型预测2021-2035年Crisui数值正负性的变化趋势,结合已有数据绘制瀑布图,其中哑变量0和1分别转化为-1与1以突出正负相抵与累计效果,曲线走势方向代表振动区间内比例较大的正负性指标及区间末端指标的累和结果,可以看出该省产业结构衰退与升级呈现一定范围内的交替性,但交替频率逐渐降低且正向化趋势明显,说明在解释变量保持目前趋势变化的前提下,未来产业结构稳步升级的可能性加强。

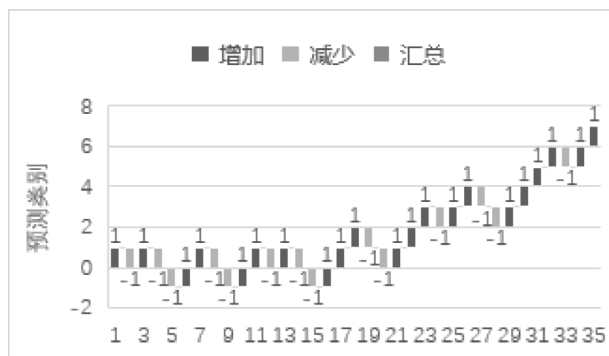


图2 瀑布图

3.2 回归预测与变量修正

调用AdaBoostRegressor类,使用CART回归树作为弱学习器,构造平方损失函数:

$$e_{ki} = \frac{(y_i - f_k(x_i))^2}{E_k} (E_k = \max|y_i - f_k(x_i)|)$$

分别选取三类弱学习器个数和四类树的最大深度作为测试参数,使用2001-2020数据作为训练集,2021-2035数据作为预测集,将训练样本及条拟合曲线可视化如下:

观察发现,某一参数控制前提下,另一参数的提升,总体上会使得曲线拟合效果越好,但由于采样限制导致训练样本量较少,为避免产生过拟合现象,确定迭代次数为100后绘制不同最大深度下的均方误差图,由此确认当树的最大深度为5时,该参数效果最佳。

结合图像可知,Crisui的回归值正负性与二分类结果基本一致,其在多种解释变量的共同作用下在长期内仍保持原有的波动惯性,但相较于前期,振幅有一定程度的缩小且负值所占比重下降,产业结构高级化态势经过了近五年的不规律增长后受到衰退因素的冲击,但在短期内迅速回正后迎来一段较稳定的增长,于2035年前继续保持良性增长。纳入空间因素的考量后,编制2010-2020年全省各地级市间泰尔指数及城镇化率的变异系数,测量人口指标的不均衡发展对回归结果可能产生的影响,首先利用线性建模绘制2010-2035跨度上的拟合曲线:

$$Ur_{\sigma} = -0.0045Time + 9.2646, URIG_{\sigma} = 0.0043Time - 8.4784$$

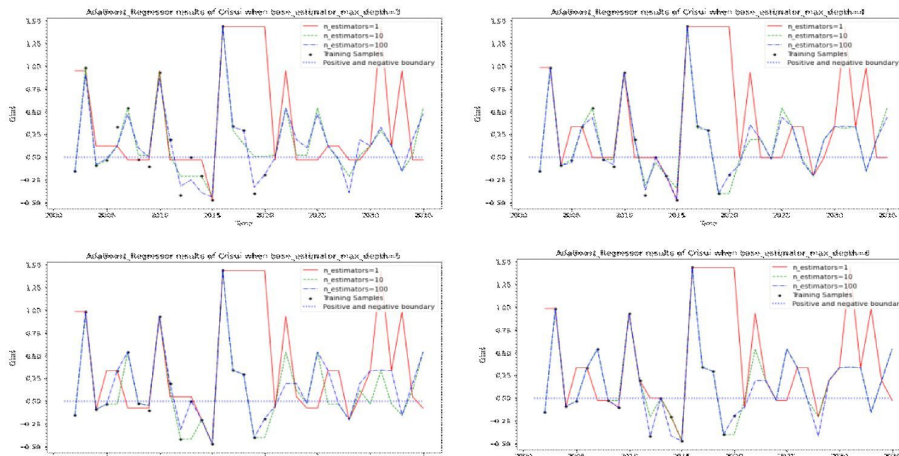


图3 最大深度依次为3-6的拟合曲线簇

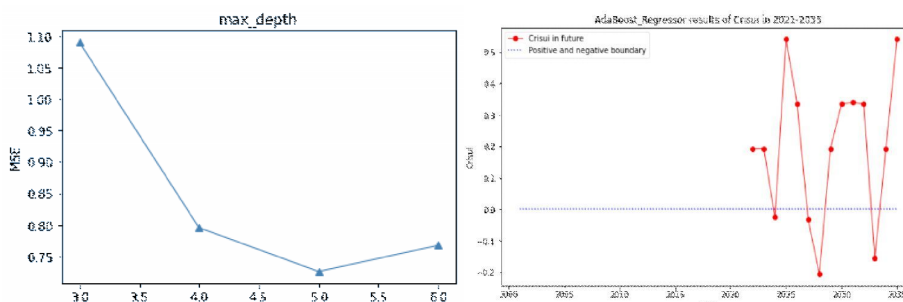


图4 均方误差(左)、基本预测结果(右)

模型检验如下:

表5 检验汇总

变异系数	R^2	T检验 ($\alpha=0.05$, p值)		F检验 (p值)
Ur_{σ}	0.8273	Intercept	0.000088	0.000103
		X	0.000103	
$URIG_{\sigma}$	0.3974	Intercept	0.0425	0.037591
		X	0.0376	

方程和回归系数均通过假设检验,但对于而言,拟合优度表现欠佳,将两组变量加入原数据集拟合新曲线如下:

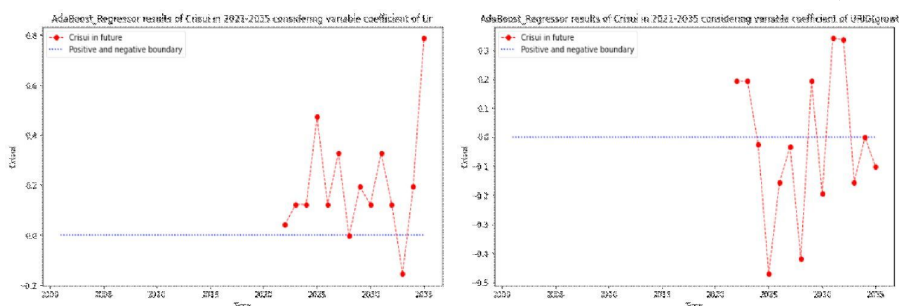


图5 修正的预测结果

可见地区间城镇化率进程的逐步统一有平抑预测前期数值波动、降低间歇性衰退概率以及抬高末端极大值的效果,而城乡收入差距地区差异化的扩大,会产生较为严重的负向冲击,并阻碍末端数值回正的倾向。

4 结论与政策建议

4.1该省预期产业结构升级速率的不规则波动和正负交替的长期存在,是对历史数据特征的考量与提取,反映了其对训练样本显著特征的合理摘取和基于预测数据集的修正行为。

4.2对于振幅趋缓有以下推测:城镇化率在当前增长模式下将会从60%上涨至预测年限末尾的80%,越发与成熟的发达经济体产业结构高级化所需的人才流动规模、技术进步要求、生产技术效率及高层次消费市场相适应,因而增长的常态化与突变的偶发性占据主流;且产业结构愈发完整,产业链韧性就越强,更能抵御外来的风险冲击,这也是曲线负向移动比例逐步降低的内在原因;安徽作为农业大省,城镇化率地区间表现差异的缩小,有助于农村剩余劳动力向城市转移,促使皖北地区的农业向着集约化、智能化方向发展,积聚产业升级进发的动力。

4.3在产业发展由“量变”向“质变”迈进的同时,城乡收入差距缩小意味着居民整体消费水平的提高,多样化需求促进了长期的技术革新和商业竞争,对产业升级速率造成持久的正向冲击,庞大而稳定的消费群体,在经济萧条时也能缓解产业结构过度衰退。但安徽省应当积极应对阜阳等人口大市,六安、铜陵等经济落后城市城乡收入差距改善滞后于总体进程的问题,避免在未来出现社会边际消费

倾向过低导致内需增长受限,以及消费断层阻碍高端产品市场形成,产业转型动力缺乏、衰退概率上升的状况。据此提出三点建议:

4.3.1贯彻“产业兴农”、“科技兴农”方针,加强农村基础设施建设,扶持农村电商,促进皖北农业区农业生产效率进步、农民生活条件改善;

4.3.2完善农村职业教育培训体系,通过技术扶持,提高农村人力资本投入;

4.3.3因地制宜谋发展,加强皖南等乡土文化繁荣地区资源优势转变力度,发挥要素禀赋。

参考文献:

- [1]黄可人,韦廷染. 经济增长、产业结构变迁与城乡居民收入差距——基于PVAR模型的动态分析[J]. 工业技术经济, 2016(4): 145-152
- [2]穆怀申,吴鹏. 城镇化、产业结构优化与城乡收入差距[J]. 经济学家, 2016(5): 37-44
- [3]朱承亮,师萍,岳宏志等. 人力资本、人力资本结构与区域经济增长效率[J]. 中国软科学, 2011(2): 110-119.

[4]李敏,张婷婷,雷育胜. 人力资本异质性对产业结构升级影响的研究——“人才大战”引发的思考[J]. 工业技术经济, 2019, 38(11): 107-114

[5]卓乘风,邓峰. 人口老龄化、区域创新与产业结构升级[J]. 人口与经济, 2018(1): 48-60

[6]汪伟,刘玉飞,彭冬冬. 人口老龄化的产业结构升级效应研究[J]. 中国工业经济, 2015(11): 47-61

作者简介:

张起望,男,安徽财经大学金融学院2020级本科生。