

结合随机森林和数据包络的安徽省各地市绿色经济效率分析

陶无言 王迪迪 汪晓云 张 玥

(安徽工程大学 数理与金融学院, 安徽 芜湖 241000)

摘要: 2019年安徽省纳入长江三角洲区域一体化发展版图,在经济快速增长的同时,生态环境也在不断恶化。本文收集安徽省16个地市的2019年6月到2020年12月数据,利用随机森林筛选出对绿色经济发展影响程度较大的投入指标。接着,利用DEA模型对2018-2020年安徽省各地市绿色经济效率进行测算。研究结果对安徽省和全国其他地区构建绿色经济发展模式有一定的参考价值。

关键词: 绿色经济效率 随机森林 DEA-SBM

DOI: 10.12319/j.issn.2096-1200.2022.26.25

随着安徽经济结构的逐渐改善,经济与生态协调发展成为各市政府所面临的主要难题。党的十九大将生态文明建设纳入制度化、法制化轨道。生态文明建设进入力度最大、措施最实、推进最快、成效最好的时期。与此同时,学术界对于绿色经济发展也展开研究。

徐祥等^[1]运用非径向非角度的SBM模型对安徽省16个地区的绿色经济效率进行测算,有效分析了绿色经济效率的区域差异,同时结合Tobit模型探究了其影响因素。殷阿娜等^[2]采用“G1+主成分”组合赋权法,对2005-2018年期间中国绿色经济发展水平进行了科学评估与区域比较,同时构建动态面板模型,使用GMM估计方法实证研究了中国绿色经济发展的关键影响因素。李爽等^[3]基于长江经济带九省二市的105个地级市的面板数据,采用非期望SBM模型对各城市绿色效率进行了测算,并利用面板Tobit模型探索了绿色效率的影响因素。牛丹丹等^[4]对河南省民办高校2015-2020年毕业生的就业数据进行了收集,采用随机森林算法,对影响就业的特征进行重要性评估,找出影响就业的主要特征和部分成因。曹泽涛等^[5]先基于DEM进行水文分析将研究区域划分成不同的小流域,然后利用数字地形分析提取29个不同方面的特征来表征流域的形态,并基于随机森林算法进行了特征选择和参数标定。

本文就如何合理设置投入指标提出新的解决方法,即结合随机森林方法,对可能成为数据包络分析中的投入指标进行重要性排序,利用数据挖掘找到符合现状的投入指标,再进行数据包络分析测算绿色经济效率值。

一、变量选取

(一) 投入指标体系的建立

国内大多数研究者在确定影响区域绿色经济效率的投

入指标时,往往采取传统的经济代表指标,包括资本、能源和劳动,细分为资源投入和非资源投入。但由于社会资源、国家政策、国际环境的变化,这些传统投入指标存在一定的滞后性和偏差。遂采用随机森林的方法,利用收集的数据计算各指标的重要程度,选取重要的指标作为DEA的投入指标。

本文选取以下28个指标作为投入指标的备选方案,分别是本科生在校人数 X_1 、硕士研究生在校人数 X_2 、上市企业数量 X_3 、新能源汽车销量 X_4 、城市水质指数 X_5 、空气质量综合指数 X_6 、优良天数比例 X_7 、森林面积 X_8 、城市轨道交通已运营里程 X_9 、高新技术企业数量 X_{10} 、美丽乡村省级中心村个数 X_{11} 、水资源总量 X_{12} 、城镇失业率 X_{13} 、固定资产投资增长速度 X_{14} 、企业研究与开发经费 X_{15} 、城市化率 X_{16} 、乡村居民人均可支配收入 X_{17} 、从业人员数量 X_{18} 、全社会用电量 X_{19} 、全社会用水量 X_{20} 、全社会供气总量(煤气、天然气) X_{21} 、发明专利授权量、 X_{22} 、2A级以上自然景区数量 X_{23} 、公共交通客运总量 X_{24} 、人工造林面积 X_{25} 、省级绿色工厂数量 X_{26} 、平均降尘量 X_{27} 、是否为全国文明城市 X_{28} 。

数据来源为安徽各市统计局、环境生态厅等官方网站,数据时间为2020年以及2019年部分月度数据。

上述变量中, X_{28} 为分类变量,取值为1或者2,其余均为连续变量。由于本文探求的是绿色经济效率的影响因素,故设置期望产出指标为绿色经济综合指数,计算公式为:

$$\text{绿色经济综合指数} = \text{GDP} / \text{城市绿地面积} \quad (1)$$

接下来以绿色经济综合指数为被解释变量探求各备选投入指标的重要性大小。

(二) 投入与产出指标的选取

在随机森林训练过程中,通过Bootstrap抽样,从训练集

中随机选取70%的记录作为决策树的训练样本，剩下30%作为测试样本。

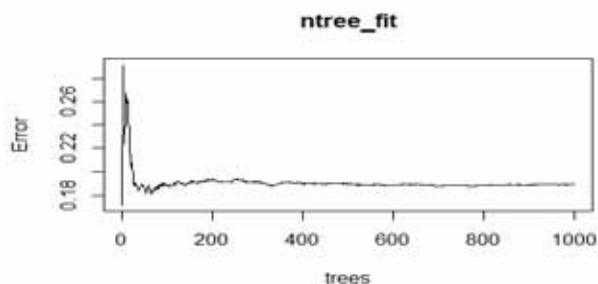


图1 OOB误分率与随机森林规模变化关系

由图1可以看出，当决策树数量达到500时，模型误差基本稳定，该随机森林模型的分类准确率为80.74%，表明该模型具有良好的识别度。

选取Mean Decrease Accuracy为评判指标重要性的依据，值越大表示该变量的重要性越大，该因素对绿色经济发展的效率影响就越大。本文选取绿色经济综合指数影响因素重要性前八名的指标作为DEA模型的投入指标，选取的投入指标按照重要程度依次为城市水质指数 X_5 、固定资产投资增长速度 X_{14} 、全社会用电量 X_{10} 、社会从业人员数量 X_{18} 、乡村居民人均可支配收入 X_{17} 、发明专利授权量 X_{22} 、空气质量综合指数 X_6 、优良天数比例 X_7 。

期望产出指标选取绿色经济综合指数、非期望产出指标选取工业颗粒物排放量、工业二氧化硫排放量、工业氮氧化物排放量。

二、安徽省各地市绿色经济效率值测算

通过DEA测算得到2018—2020年安徽省16个地市的绿色经济发展效率（表1），由经济效率值可以看到：

表1 2018—2019年安徽省各地区绿色经济效率值

| 地区 | 2018 | 2019 | 2020 | 均值 |
|-----|--------|--------|--------|--------|
| 合肥 | 1.0000 | 0.4598 | 0.2404 | 0.5667 |
| 芜湖 | 1.0000 | 1.0000 | 0.2240 | 0.7413 |
| 马鞍山 | 0.3128 | 0.3947 | 0.2825 | 0.3300 |
| 滁州 | 1.0000 | 1.0000 | 0.3162 | 0.7721 |
| 阜阳 | 0.4801 | 0.4222 | 1.0000 | 0.6341 |
| 安庆 | 1.0000 | 0.7266 | 0.4559 | 0.7275 |
| 宿州 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 蚌埠 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 亳州 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 六安 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 宣城 | 0.4109 | 0.4931 | 0.4118 | 0.4386 |
| 淮南 | 1.0000 | 0.5100 | 0.2918 | 0.5673 |
| 淮北 | 0.3027 | 0.3294 | 0.2706 | 0.3009 |
| 铜陵 | 0.1655 | 1.0000 | 0.1091 | 0.4249 |
| 池州 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 黄山 | 0.1469 | 0.1701 | 0.0980 | 0.1383 |
| 均值 | 0.7387 | 0.7129 | 0.5438 | 0.6651 |

（一）按时间分析各地绿色经济发展效率

从总体时间发展上来看，2018—2020年安徽省各地市绿色发展效率值整体呈下降态势，2018年各地市绿色发展效率均值为0.7387，还处于较高水平，但发展到2020年，绿色经济效率就降低至0.5438，说明社会经济的可持续发展在走下坡路。

从地区时间发展上分析，2018—2020年，除宿州、蚌埠、亳州、池州、六安、阜阳以外的其余十个城市绿色发展效率值有着不同程度的下降。其中马鞍山、宣城、淮北、铜陵绿色发展效率值先上升后下降，且上浮幅度在0.02到0.84之间，下降幅度在0.05到0.89之间。阜阳市绿色发展效率值先下降后上升，其上升幅度较大，为0.5118。剩下城市绿色发展效率值一直下降，降幅最大达到了0.78。

就代表性城市而言：宿州、蚌埠、亳州、六安、池州五市绿色经济效率值连续三年为1，其中少不了经济发展较好地区的帮扶。扶贫攻坚在这三年间进入最后冲刺阶段，国家以及全省对于经济欠佳城市的扶助力度达到顶峰，这五个城市的经济在这三年间获得长足发展，且生态环境保护到位，绿色经济得到很大程度的发展。黄山市作为安徽省旅游胜地，绿色发展效率持续3年位于全省最后一名，且效率值均在0.2以下。究其原因，发现黄山市虽然环境良好，绿地面积较大，但是缺少工业企业、高新企业，全市经济发展处于较低水平。另外，2020年疫情来袭，为配合疫情防控工作，全国旅游景点存在不同程度的关停，以旅游业为主要产业的黄山市经济发展严重停滞，从而导致该市绿色经济发展效率不高。

（二）按空间分析各地绿色经济发展效率

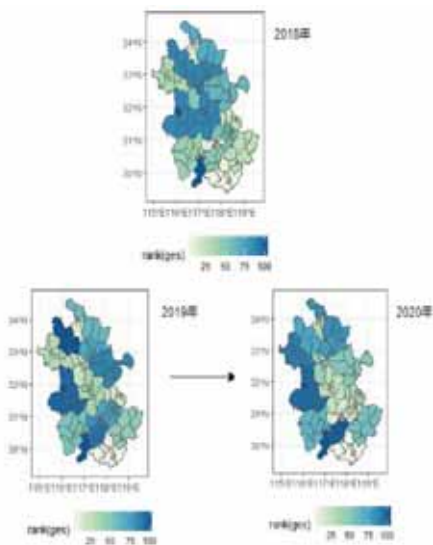


图2 安徽省各地市2018—2020年绿色经济效率空间演变

从空间分布上分析,根据图2可得,安徽省各地市的绿色经济效率值呈不规则分布,体现安徽省各地市的绿色经济发展水平分配不平衡的特点。总的来说,南北城市绿色经济发展水平逐渐提高,而中部城市逐渐降低。整体呈“空心化”发展。根据分析,中部城市经济发展较好,产业结构、城市规划基本稳定,绿色经济发展达到瓶颈,绿色经济效率很难维持在先前水平;而北部南部城市经济发展稍微落后全省,但其拥有较为丰富的森林资源,环境保护到位,在不损害环境的基础上发展经济会使得绿色经济效率明显增长。

(三)按地区发展均值分析各地绿色经济发展效率

根据文献研读,把绿色经济发展效率分为四个等级水平:0.8000-1.0000,0.7000-0.8000,0.6000-0.7000,0.6000以下,根据三年来不同城市的绿色经济发展水平均值,分为高效率水平、中高效率水平、中等效率水平、低效率水平,结果如表2所示。

表2 绿色经济发展水平在各区域的分布表

| 地区 | 第一等级 | 第二等级 | 第三等级 | 第四等级 |
|----|----------|--------|--------|----------|
| | 高效率水平 | 中高效率水平 | 中等效率水平 | 低效率水平 |
| 皖北 | 宿州、蚌埠、亳州 | — | 阜阳 | 淮北、淮南 |
| 皖中 | 六安 | 滁州 | — | 合肥、马鞍山 |
| 皖南 | 池州 | 芜湖、安庆 | — | 宣城、铜陵、黄山 |

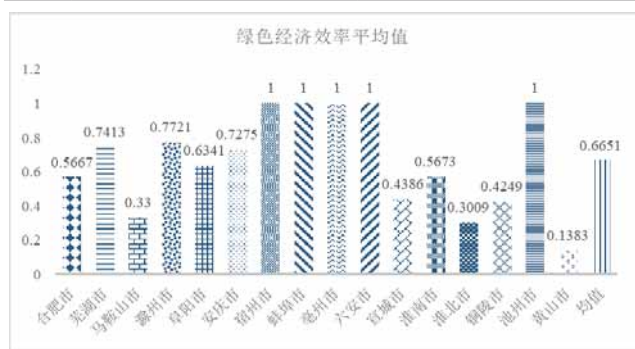


图3 2018—2020年安徽省各地市绿色经济效率平均值对比图

从地区发展均值层面分析,如图3,2018年到2020年间芜湖市、安庆市、滁州市、宿州市、蚌埠市、亳州市、六安市、池州市的绿色经济发展效率高于全省平均水平0.6651,绿色经济发展处于中高等水平;合肥市、马鞍山市、阜阳市、宣城市、淮南市、淮北市、铜陵市、黄山市绿色经济发展效率低于全省平均水平,其中马鞍山市和黄山市分别为0.33和0.1383;全省有7个城市绿色经济发展效率处于低效率水平,其中不免有经济发展不错的城市,如合肥。对于这种城市,应该更加重视地区绿色经济发展,要金山银山,更要绿水青山;对于老牌工业城市,如马鞍山

等,要做好企业转型,加大绿色贷款投入,加快本地区绿色发展速度。

三、政策建议

第一,优化产业结构。安徽省各城市绿色经济影响因素中,第二产业和第三产业与第一产业比值对绿色经济发展影响显著,故应该优化产业结构,不断提高城市绿色经济发展水平。2020年习近平总书记在联合国大会上提出争取在2060年前实现碳中和的目标,意义深远。因此安徽省各城市在优化产业结构时,要紧紧抓住降低碳含量这一主线,优化产业结构。

第二,重新建立对GDP的理解与认识。在改革开放前,我国经济发展衡量标准以前苏联的计划经济国民经济核算为蓝本,而改革开放以后,换成以西方国家国民经济核算为模板。在之前很长一段时间,全社会以GDP总量作为衡量政绩的重要指标,这受到了当时国际环境的约束,现在看来不够合理。地方政府在推动以提高经济总量为目的的相关政策时,忽略了企业的因生产造成的污染与资源浪费,在新时代社会经济发展中,各级地方政府对GDP要做到重而不唯。

第三,合理规划教育支出,加大科技创新投入。安徽省在经济发展过程中教育短板明显,分析原因可知,安徽省各城市教育资源分配不够合理,许多城市仅一所本科院校。并且安徽省人才流失严重,许多高校毕业生毕业后前往南京、上海等地寻求发展。安徽省应该充分利用高校平台,开设绿色经济学科,发展绿色经济人才,利用大数据整合社会资源,建立完备的全省资源信息数据库,减少在经济发展过程中造成的不必要资源浪费,发展绿色循环经济。

参考文献

[1]徐祥,朱敏.绿色经济效率区域差异及影响因素研究——以安徽省16个地市为例[J].重庆科技学院学报(社会科学版),2019,(02):44-47.

[2]殷阿娜,吕俊峰.中国绿色经济发展评价及影响因素研究[J].金融理论与教学,2021,(05):58-63.

[3]李爽,周天凯,樊琳梓.长江经济带城市绿色发展及影响因素分析[J].统计与决策,2019,35(15):121-125.

[4]牛丹丹,刘筠筠.基于随机森林算法的就业预警机制研究——以河南省民办高校为例[J].现代信息科技,2021,5(22):101-103.

[5]曹泽涛,方子东,姚瑾,熊礼阳.基于随机森林的黄土地貌分类研究[J].地球信息科学学报,2020,22(03):452-463.