

基于机器学习及非参数统计的我国各上市银行 绿色信贷差异化影响因素研究

胜紫菡

中央民族大学, 中国·北京 100080

【摘要】“绿色信贷”是各大银行通过金融杠杆来遏制高耗能高污染产业盲目扩张的一种信贷政策。截至2021年9月末,国内21家主要银行机构的绿色信贷余额已达14.08万亿元,但各大银行的绿色信贷发展状况参差不齐,信贷余额也呈现出较大的差异。本文将基于《中国统计年鉴》、《中国保险年鉴》和国内各上市银行的《可持续发展报告》、《社会责任报告》中的数据,在常规线性回归模型的基础上,进一步采用随机森林回归与非参数回归的方法研究我国各上市银行绿色信贷差异化的影响因素。

【关键词】绿色信贷; 随机森林回归; 非参数回归; 影响因素

Research on the Influencing Factors of Green Credit Differentiation of Listed Banks in my Country Based on Machine Learning and Non-parametric Statistics

Victory Zihan

Minzu University of China, Beijing 100080, China

[Abstract] "Green credit" is a credit policy used by major banks to curb the blind expansion of high-energy-consuming and high-polluting industries through financial leverage. As of the end of September 2021, the green credit balance of 21 major domestic banking institutions has reached 14.08 trillion yuan, but the development of green credit among major banks is uneven, and the credit balance also shows great differences. Based on the data in China Statistical Yearbook, China Insurance Yearbook, and the Sustainability Report and Social Responsibility Report of various domestic listed banks, this paper will further use random forest regression on the basis of conventional linear regression model. And nonparametric regression method to study the influencing factors of green credit differentiation of listed banks in my country.

[Key words] green credit; random forest regression; nonparametric regression; influencing factors

1 模型及方法介绍

1.1 随机森林回归概述

随机森林回归为机器学习算法的一个分支,其在不断的学习与训练中不断提高模型的精度,从而使得到的决策树更加精准的反馈出各个自变量的影响力大小,提高模型拟合的精度。通俗来讲,随机森林是用随机的方式建立一个森林,森林中由很多的决策树组成。随机森林以Bagging集成为基础,以CART决策树为学习器,其与普通决策树的主要区别在于选择最优属性的过程:普通决策树直接在当前结点中选择将该结点分裂的最优属性,而随机森林算法,是在当前结点的属性集合中选择一个属性子集,再从该子集中选出最优划分属性。

1.2 非参数回归概述

非参数回归模型是指并不需要知道总的分布的情况下进行的一种统计推断方法。它是对当前路段和几条相邻路段的交通流信息对当前路段进行交通流预测的单条路段分析的扩展。该算法认为系统所有的因素之间的内在联系都蕴含在历史数据中,因此直接从历史数据中得到信息而不是为历史数据建立一个近似模型。非参数回归最为一种无参数、可移植、预测精度高的算法,它的误差比较小,且误差分布情况良好。尤其通过对搜索算法和参数调整规则的改进,使其可以真正达到实时交通流预测的要求。

2 回归自变量的选择

通过查阅现有文献,影响各大银行信贷余额的因素有:总资产收益率、不良贷款余额、银行总资产、成本收入比、权益乘数等等。并且通过对各变量间进行相关系数非零性检验,在 α

$=0.05$ 的情况下,对于绿色信贷余额而言,总资产收益率、银行总资产、期末贷款余额、资本充足率、营收增长率、权益乘数六个变量均拒绝了原假设,即与绿色信贷余额相关。因而最终确定各变量如下:

表1: 各变量含义及英文缩写

变量	含义	英文缩写
x^*_1	总资产收益率	ROTA
x^*_2	银行总资产	TA
x^*_3	期末贷款总额	TLATE
x^*_4	资本充足率	CAR
x^*_5	营收增长率	IROR
x^*_6	权益乘数	EM
y^*	绿色信贷余额	GCB

3 模型的应用

3.1 随机森林回归的应用

本文从全国36家上市银行的绿色信贷余额相关数据出发,将36条数据按照7:3的比例,每次随机划分为训练集与测试集。其中训练集用于模型的建立,测试集用于模型的检验。首先,为了选择合适的决策树个数,用800棵决策树得到的模型进行预测,通过平均残差平方与决策树的数量关系,确定当决策树的数量达到500后,曲线基本趋于平稳。因此,此处选择决策树的数量为500,从而在保证拟合效率的同时,尽可能的提高模型的精度。将数据经处理后导入R软件,得到在该模型中各自变量的重要程度:

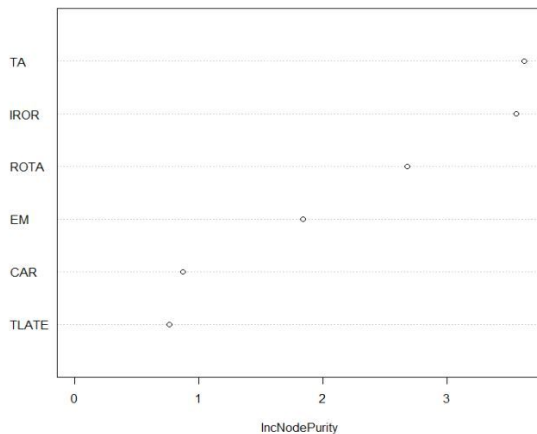


图 1:随机森林回归中各自变量重要程度示意图

由上图可见,重要程度:TA> IROR> ROTA> EM> CAR> TLATE,也即,银行总资产与营收增长率是影响银行绿色信贷余额的首要因素,其影响程度显著高于其他因素;其次是总资产收益率与权益乘数,二者的影响程度适中;最后是资本充足率与期末贷款总额,二者的影响力最小。从而得到表征各自变量重要程度大小的回归方程为:

$$y^* = 2.521x_1^* + 3.379x_2^* + 0.797x_3^* + 0.981x_4^* + 3.332x_5^* + 1.897x_6^*$$

进一步,使用训练样本集与测试样本集评估预测的精度与性能,大部分训练样本集样本点与测试样本集样本点在直线左右只有微小的波动,说明模型的拟合效果较好,预测的精度较高。

3.2 非参数回归的应用

本文从全国36家上市银行的绿色信贷余额相关数据出发,采用权函数的方法,利用R软件,依次建立因变量:绿色信贷余额,关于6个自变量的非参数回归模型。为了减少线条交叉同时方便观察,本文依据在随机森林回归中得到的结论,将6个自变量按照其重要程度相对大小分为3组进行绘图:

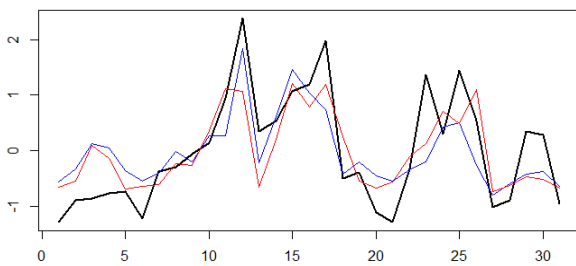


图 2:TA、IROR 与 GCB 非参数回归拟合结果

【注释】黑色-GCB 红色-IROR 蓝色-TA

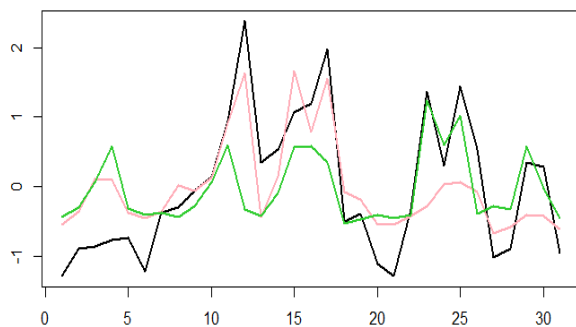


图 3:ROTA、EM 与 GCB 非参数回归拟合结果

【注释】黑色-GCB 粉色-ROTA 绿色-EM

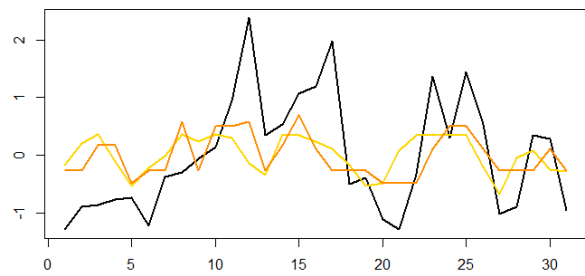


图 4:CAR、TLATE 与 GCB 非参数回归拟合结果

【注释】黑色-GCB 黄色-CAR 橙色-TLATE

在以上三组图中,第一组图,蓝色TA、红色IROR与黑色GCB的拟合度最高;第二组图,粉色ROTA、绿色EM与黑色GCB的拟合度也较高,但明显不如第一组图拟合度高;第三组图,黄色CAR、橙色TLATE与黑色GCB的拟合度就相对较低了。换句话说,这也印证了在随机森林回归中各个自变量重要性大小:TA> IROR> ROTA> EM> CAR> TLATE

4 模型的评价

相比传统的线性回归模型,随机森林回归与非参数回归则表现出显著的优越性。在本文中,通过随机森林回归,得到各自变量的重要程度排序为:TA > IROR > ROTA > EM > CAR > TLATE。同时,在非参数回归中也得到了与此基本一致的结论。另外,通过模型拟合,二者对原始数据的拟合效果明显优于传统的线性回归模型。另一方面,相比传统的线性回归模型,二者的使用条件更加宽松,而不需要对 Gauss—Markov 假设进行严格的限制。如此,也在解决实际问题中,降低了对数据的要求,使得分析的问题可以更加广泛。但同时我们也要注意,随机森林回归与非参数回归尽管在回归效果方面存在很大的优越性,但其算法复杂度要远高于传统的线性回归模型。在随机森林回归中,我们需要不断的训练与测试,才能得到理想的决策森林。在非参数回归中,我们也要首先确定好带宽与核函数的形式,才能进行进一步的回归分析。

5 结论与建议

5.1 结论

通过以上回归模型的建立及回归结果的可视化描述,影响各大银行之间绿色信贷余额差异的首要因素为:银行资产总额与营收增长率,这与现实情况也是相符合的。银行规模越大,即资产总额越多,与其对接的环境污染企业或环保个人也就越多,从而信用贷款的总额度也就越高,在相同的贷款额度下,余额也就自然更高。另外,资本充足率与期末贷款总额对绿色信贷余额的影响较小。

5.2 建议

“绿色信贷”是协调经济发展与环境保护的一项重要金融政策,也是时代的潮流。对于总资产量较大、规模较大的银行来说,应该继续发挥自己的规模优势,限制非绿色企业及个人的非绿色行为。对于资本总量较小、规模较小的银行来说,应该创新信贷政策,吸引越来越多的企业加入到环境保护的行列当中。在环保的大背景下不断推动经济的发展。各大银行响应绿色号召,结合自身特点,创新信贷方案,主动承担社会责任,以实现“绿色信贷”的可持续发展。

参考文献:

- [1]胡颖超,肖浩伟,孔晨曦,廖俊娟.互联网金融背景下大学生金融风险意识的影响因素研究——基于随机森林与多元回归的对比分析[J].湖北经济学院学报(人文社会科学版),2022,19(01):37-41.
- [2]吴赛楠,白涵.绿色信贷对商业银行经营绩效影响研究——基于国有控股商业银行分析[J].商讯,2021(35):89-91.
- [3]陈涛,徐珊珊.绿色信贷对我国商业银行盈利性的影响研究[J].长春金融高等专科学校学报,2021(06):58-65.