

基于Logistic回归的绿色信贷信用风险评估研究

龙海菊

贵州大学经济学院, 贵州 贵阳

收稿日期: 2024年9月13日; 录用日期: 2024年10月8日; 发布日期: 2024年11月28日

摘要

近年来, 中国大力推行绿色信贷政策, 引导金融机构加大绿色信贷规模, 为提高绿色信贷的审批效率与实施效果, 我国银行业也在积极探索构建绿色信贷实施准则与评估体系。本文以我国深交板主板上市的重污染企业2018~2022年的基本信息、财务数据和非财务数据为基础, 建立上市公司绿色信贷信用风险评估初步指标体系, 再通过特征筛选剔除得分低的11个指标后, 得到11个重要性程度高的指标重新建立上市公司绿色信贷信用风险指标体系。利用计量工具进行Logistic回归实证分析上市公司绿色信贷项目风险管理, 通过模型的混淆矩阵、准确率等评价指标结果分析模型的适用性, 最后得出结论并提出建议。

关键词

绿色信贷, 信用风险, 逻辑回归

Research on Credit Risk Assessment of Green Credit Based on Logistic Regression

Haiju Long

School of Economics, Guizhou University, Guiyang Guizhou

Received: Sep. 13th, 2024; accepted: Oct. 8th, 2024; published: Nov. 28th, 2024

Abstract

In recent years, China has vigorously promoted the green credit policy and guided financial institutions to increase the scale of green credit. In order to improve the approval efficiency and implementation effect of green credit, China's banking industry is also actively exploring the construction of green credit implementation criteria and evaluation system. Based on the basic information and non-financial data of the heavily polluting enterprises listed on the main board of Shenzhen Stock Exchange of China from 2018 to 2022, this paper establishes a preliminary index system for green credit risk assessment of listed companies, and then removes the last 11 indicators with low scores

文章引用: 龙海菊. 基于 Logistic 回归的绿色信贷信用风险评估研究[J]. 电子商务评论, 2024, 13(4): 5640-5654.

DOI: 10.12677/ecl.2024.1341802

through feature screening. 11 indicators of high importance are obtained to re-establish the green credit risk index system of listed companies. Logistic regression was used for empirical analysis of the risk management of green credit projects of listed companies. The applicability of the model was analyzed through the results of evaluation indicators such as confusion matrix, accuracy rate were drawn and suggestions were put forward.

Keywords

Green Credit, Credit Risk, Logistic Regression

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

当前，中国正处在经济结构转型和增长模式转变的重要阶段，对推动绿色产业发展和实现经济、社会可持续性的需求日益增长。2012 年，中国银监会发布的《绿色信贷指引》将绿色信贷纳入商业银行的考核体系，标志着金融领域对环保的重视。在党的二十大报告中，强调了实施全面节约战略、发展绿色低碳产业、推广绿色消费的重要性，并提出了在产业结构调整、污染治理、生态保护和应对气候变化等方面的统筹规划，以加快向绿色发展模式的转变，这需要金融行业的有力支持。“绿色金融”这一概念应运而生，反映了公众对环境保护的日益关注和对低碳资源开发的迫切需求。中国在碳达峰碳中和的第一次全国会议上提出，将力争在 2030 年前实现碳达峰，2060 年前实现碳中和。随着相关政策和法规的逐步推出、实施和完善，中国的商业银行正在加大绿色信贷的投放力度。

区别于传统信贷，绿色信用是指通过资本的导向效应，把资源投入到环境保护领域，从而降低了对重污染企业及工程的融资成本，从而达到资源“绿色配置”的目的。绿色信贷政策的推行，既可以优化银行的信贷资源，又可以促进重度污染者改变其生产模式。同时，由于相关制度的不健全，企业在社会责任认定和环境信息披露方面的投入不够，加之绿色信贷的特性与之前的传统信贷有很大的区别，所以无法直接套用传统的信用风险评价方法来评价绿色信贷的信用风险，从而影响了商业银行的绿色信贷发展。由此可见，迫切需要建立一套符合我国国情的绿色信贷制度，并对其进行评价。国内外的学者也都从不同的角度分析了绿色信贷的发展状况和功能，并强调了强化绿色信贷风险管理的重要意义。然而，在现有的研究中，很少有对绿色信贷的信用风险进行单独的分析，并且大部分的研究都停留在理论的层次上，很少有定量的研究。

因此本文在理论分析的基础上，以深交板主板上市的重污染行业企业为例，基于 Logistic 回归方法构建用以评估绿色信贷信用风险的模型，判断其对绿色信贷信用风险评估的准确性和有效性，并对此提出相应的建议。

2. 文献综述

2.1. 绿色信贷研究

可持续发展理念在我国最早确立，国家也大力倡导金融要主动服务实体经济，促进产业转型。绿色信贷是一种新兴的融资工具，它的重要性也在逐步提升。在绿色金融、可持续金融等理念的基础上，国内学者对绿色信贷的内涵进行了拓展和深化。Mark A. White (1996)认为，可持续融资需要金融机构对其

进行合理的资源分配,并在其风险评价中加入绿色企业的环境风险,从而达到生态保护与经济发展的双赢[1]。何德旭、张雪兰(2007)认为,绿色信贷可以通过两种方式实现资金:“绿色配置”:一是对环境保护有正向效应的企业提供长期的优惠利率,二是对重污染企业的授信额度、期限等进行约束[2]。Weber O, Fenchel M 和 Scholz R W (2010)提出,环保金融就是银行以环保为前提,通过对环保要求的项目进行社会资源配置,从而为环保工程提供福利性的借贷融资[3]。孙光林 *et al.* (2017)的研究表明,商业银行开展绿色信贷能够提高企业经营绩效,并能有效缓解由于消费者的环保违规行为而导致的信用风险[4]。罗霞和夏梦瑶(2022)等研究得出,现绿色信贷风险管理中仍存在风险评价标准与程序不够健全、对绿色信贷资金投入过度集中、有效控制绿色信贷风险的能力不够、相关金融专业人才缺失等情况[5]。

LuoSumei 等(2021)利用双重差分法(DID)和双重差分法(DDD)分析了绿色贷款对我国商业银行的综合竞争能力的影响,结果表明,高信用和高信誉的银行可以更好地提高其核心竞争力[6]。陈琪和张广宇(2019)以 2016~2017 年两个剩余企业的实证数据为基础,从负债融资成本的视角,对企业的经营风险与财务业绩进行了实证研究,得出了样本公司的财务业绩有下降的趋势[7]。廖果平和李颖(2020)使用 KMV 对我国 A 股绿色产业与两高产业上市企业的信用风险进行了测算,结果表明,绿色产业的信用风险要比“两高”企业稍小,并没有太大的优势。接着,本文对两个产业中的信用风险因素进行了深入的实证研究,结果表明,负债规模对于环境保护企业的信用风险具有重要的作用,而对于“两高”企业则没有什么作用[8]。王馨(2021)等的研究发现,实施和实施绿色信贷政策可以有效地降低企业的环境污染行为,促进企业绿色低碳的技术进步与创新[9]。舒利敏和廖菁华(2022)提出,绿色贷款能够促进重污染产业企业的绿色转型和投资决策,并体现出绿色贷款对企业资本流向的导向效应[10]。

2.2. 信用风险评估研究

随着时间的推移和科技的飞速发展,机器学习逐渐走入了金融业。随着现代信用风险评价方法的提出和运用,企业信用风险评价的准确性得到了显著提高。吴世农与卢贤义(2001)利用多元线性回归、Fisher 鉴别和二元 Logistic 回归等方法识别上市公司的违约风险,结果表明,二元逻辑回归具有最佳的拟合优度,且错误发生率最低[11]。吴彬(2008)通过对不同的信贷风险评价方法的比较,构建了基于 Logistic 模型的实证分析,发现该模型具有良好的预测能力[12]。Belhadi Amine 等(2021)将旋转森林和 Logit 提升算法相结合,建立新型的融合机器学习模型,对农业 4.0 (SCF)下的中小企业进行信贷风险的预测,以改善其信贷风险的预测精度,并对其进行评估[13]。邵川(2020)运用灰色关联分析和回归分析等方法,探讨绿色信贷和绿色贷款结构对产业结构优化的作用机理,提出相应的政策建议[14]。王姝怡(2021)运用 Logistic 回归模型能够有效判别节能环保企业是否违约[15]。尹露颖(2022)通过参数调优后的集成学习算法实现了对上市公司绿色信贷违约损失的分析与预估[16]。王丹迪(2018)通过收集和整理问卷数据得到环保指标权重,结合财务和非财务指标,利用层次分析法建立绿色信贷信用风险评估模型,并选取代表企业进行实际应用[17]。

黄颖利、李友慧和王宇(2018)认为 KMV 模型主要是考虑财务因素来对信用风险进行度量,将 KMV 模型度量结果与道德风险、可持续发展水平等相关绿色影响因素一起代入 BP 神经网络模型,结果表明能更准确的预测绿色信贷信用风险大小[18]。Feng Yuan (2022)通过主成分分析和因子分析建立反向传播神经网络(BPNN)模型,通过实证发现,基于 BPNN 的绿色信用风险评估模型具有 95%的准确率,且 99%的样本企业风险较低。然而煤炭行业的大多数公司的风险较高[19]。张李红(2020)采用 Logistic 回归方法对绿色信贷违约率进行测算,结果表明模型判别正确率较好[20]。

2.3. 文献评述

在对国内外相关研究进行梳理的基础上,我们可以看到,国外学者对“可持续金融”这一概念有了

较为明确的界定,而国内学者对“可持续金融”的定义则更侧重于对“绿色信用”的界定。由于绿色信贷具有一定的风险性,因此,对其进行风险管理已成为国内外学术界研究的热点。无论是国内外的研究人员,还是国内的学者,都一致认同了绿色贷款和绿色金融在可持续发展中的重要作用。近年来,许多学者将研究的重点放在了将机器学习和绿色信用风险相结合的研究上。从文献的梳理中还可以看出,符合我国国情的 Logistic 回归方法具有很好的适用性和预测能力。然而,目前关于绿色信贷信用风险评价的研究,大多数都是基于 BP 神经网络、KMV 等,很少使用 Logistic 回归模型。为此,本项目拟利用深交所重污染行业上市公司的基本信息和财务数据,采用 Logistic 回归方法对其进行评价,为我国企业绿色信贷的信用风险评价提供理论和经验支撑。

3. 相关理论与模型构建

3.1. 相关理论

3.1.1. 赤道原则

赤道原则是 2002 年由世界银行下属的国际金融公司与荷兰银行共同提出的一项企业贷款准则,由国际金融公司和荷兰银行等 9 家银行在伦敦召开会议,会后提出一套针对融资项目中的环境和社会风险指南,这就是赤道原则。这一原则的提出,是为了降低银行因投资可能对环境和社会造成损害的项目而面临的信用风险和声誉损失。赤道原则强调,商业银行在进行项目投资前,应优先评估项目对环境和社会的潜在影响,确保投资活动符合环境友好和可持续发展的要求。在国际项目融资领域,赤道原则已经成为衡量和管理环境与社会风险的重要标准。其中,原则声明是赤道原则的核心,列举了采取赤道原则的金融机构在做出投资决定时需遵守的十条特别条款和准则。国外商业银行在开展绿色信贷过程中融入了赤道原则并沉淀出了诸多宝贵的实践经验,典型代表的银行包括美国花旗银行、荷兰银行、日本瑞穗银行[21]。以下表格总结了国外的以上三大杰出赤道原则践行银行的绿色信贷项目风险管理实践经验,以及从中总结出对我国商业银行有指导意义的启示。

3.1.2. 可持续发展理论

在 1987 年,联合国世界环境与发展委员会首次提出了“持续发展”的理念。12 年后,中国科学院发布的《2009 中国可持续发展战略报告》进一步明确了符合我国国情的可持续发展策略。该理论主张在满足当代人需求的同时,不损害后代的生存环境。可持续发展追求经济、社会和环境的和谐发展,与当前流行的 ESG (环境、社会、治理)概念相契合。

实现可持续发展需要遵循三个基本原则:公平性、持续性和共同性。公平性原则强调在发展过程中要实现当代人之间的平等,并考虑与后代的公平关系。持续性原则要求人类在自然承受能力范围内合理利用资源,以维护生态系统的平衡。共同性原则则强调实现可持续发展是全人类的共同责任,需要全球社会的共同努力,而非仅依赖个别国家或组织。

可持续发展理论指出,传统经济增长模式可能导致人与自然关系的紧张,进而破坏生态环境。商业银行作为经济增长的重要参与者,在推动可持续发展的过程中,可以通过绿色信贷等手段,促进资源的合理分配,提升落后产能,推动节能减排,从而在经济增长与环境保护之间找到平衡点。

3.1.3. 信用风险管理理论

信用风险管理是一种综合运用多种策略来评估、监控和控制贷款过程中的潜在风险的方法。其核心目标是在确保风险可控的同时,实现收益最大化,从而提高贷款业务的经济效益。

该理论强调,银行需要准确识别信贷业务中的各种风险,并进行有效的量化,以确保风险与收益之间的平衡。这有助于银行在竞争激烈的市场中增强自身的盈利能力和竞争力。为此,银行应深入分析自

身的风险管理体系，明确风险管理的目标，并对风险进行分类和结构化。其次，要选取合适的风险指数，采取科学的计量手段，并结合历史数据，建立一种有效的风险管理框架。在此基础上，提出了相应的风险管理对策，构建了内部和外部的风险交流机制，以及应对风险的计划。

通过这些措施，银行可以更好地管理信用风险，提高贷款业务的经济效益，同时在激烈的市场竞争中保持竞争力和盈利能力。

3.2. 模型构建

Logistic 回归是一种可以用于预测某件事情发生的概率及解释其与导致该概率发生的因素之间关系的模型。该类模型适用于解决绿色信贷项目风险管理问题原因如下：首先，Logistic 回归模型既不要求样本服从同分布或方差不变的前提，并且不论自变量是连续或离散类型都不会影响模型的最后结果。此外，使用该模型所求得的预测概率值会一致保持在 0 到 1 的范围内，这样对因变量为二分类变量的结果解释也会更加容易。

Logistic 模型的基本原理如下所示：

$$P(y=1) = \frac{1}{1 + e^{-(\alpha + \sum \beta_j x_j)}} \quad (1)$$

$$P(y=0) = 1 - p = 1 - \frac{1}{1 + e^{-(\alpha + \sum \beta_j x_j)}} = \frac{1}{1 + e^{(\alpha + \sum \beta_j x_j)}} \quad (2)$$

其中 y 表示因变量，在本文中为绿色信贷项目是否会产生违约风险，当 y 为 0 时表示绿色信贷项目不会出现违约即正常的概率，当 y 为 1 时表示绿色信贷项目会出现违约即高风险的概率。 α 与 β 都表示回归系数， X 为自变量。

4. 数据来源、指标构建与优化

4.1. 数据来源

根据 2010 年环保部发布的《上市公司环境信息披露指南》，选择了火电、钢铁、水泥、煤炭、冶金和化工等 16 个行业，并选择了深交板主板上市的重污染企业作为样本，在 2018~2022 年期间进行，在随机抽样过程中，难免会出现同一公司在不同年度的数据被同时抽取的情形，为了确保结果的正确性，我们对同一公司进行了两次以上的人工控制，并且从各个上市公司的年报中获取了财务方面的资料。在此基础上，根据有关规范和标准加以处理，将金融类公司以及有关信息中存在严重遗漏的公司等加以删除，在保证各个指标数据准确性和完整性的基础上，最终得出 410 有效数据样本。

4.2. 指标选取

4.2.1. 财务指标

在传统的信用风险评价中，财务指标往往是最受重视的一个方面。通常情况下，公司的财务风险是由公司的偿债能力，运营能力，盈利能力，发展能力及其对应的子指标来衡量的。具体如下：

(1) 偿债能力。它反映了公司对到期债务的承担能力，也就是公司对到期债务的担保程度。根据流动比率、速动比率、资产负债比等因素，对企业的经营状况进行分析。

(2) 盈利能力。是指企业获取利润的能力。由基本资产报酬率、净资产收益率、营业利润率、毛利润率等来综合决定。

(3) 运营能力。是指企业经营和使用经济资源的效率，以及资本运作的周转性。由应收账款周转天数、

存货周转天数、流动资产周转率等来综合决定。

(4) 发展能力。是指企业扩大规模、壮大实力的潜在能力。由净利润增长率、营业总收入增长率、总资产增长率等来综合决定。

4.2.2. 非财务指标

(1) 基础信息指标

为了重污染行业能掌握绿色信贷项目的总体情况，本部分指标选择公司的经营年限以及总资产作为基本信息指标。其中，经营年限反映了公司在行业经营时间的长短，一般而言，公司在行业运作的时间越长，开展业务的经验越丰富，抗风险能力也就越大，公司违约的信用风险也就越低。而资产情况一定程度反映了企业的规模大小。

(2) 技术指标

企业在新技术研发方面的投资和科技实力对于绿色贷款风险评价具有重要意义。从企业的科研经费支出、研发人员所占的比例等方面可以看出企业对科研投入的关注程度。为此，论文选取了研发人员费用占营业收入的比重、研发人员比率、研发人员增长率这三个指标来进行研究。研发人员投资是指企业从事科研和技术发展活动所付出的成本和费用。但是，由于企业的规模和技术发展程度存在着一定的差别，所以本文选取了研发费用与营业收入的比例作为衡量企业技术投资的指标。科技人才是科学研究发展的根本依据，而科技人才的比例则是一个企业科技发展的重要指标。变量的基本情况如表 1 所示。

Table 1. Descriptive statistical results for indicators

表 1. 指标描述性统计结果

指标	均值	标准差
流动比率 X1	1.449	1.096
速动比率 X2	1.133	0.976
EBIT 利息保障倍数 X3	8.305	17.367
资产负债率 X4	0.527	0.186
权益乘数 X5	2.505	2.662
资产报酬率 X6	0.057	0.097
净资产收益率 X7	0.064	0.347
营业毛利率 X8	0.252	0.195
营业净利率 X9	0.033	0.287
应收账款周转天数 X10	67.187	101.688
存货周转天数 X11	95.182	134.256
流动资产周转率 X12	1.886	1.363
总资产周转率 X13	0.716	0.473
净资产收益率增长率 X14	4.154	79.846
净利润增长率 X15	2.685	59.054
营业收入增长率 X16	0.180	0.890
总资产增长率 X17	-0.699	1.915

续表

总资产 X18	9.846	4.092
经营年限 X19	23.840	5.415
研发人员数量占比 X20	10.531	6.451
研发投入占营业收入比例 X21	2.787	2.601
研发人数增长率 X22	1.363	3.107

4.2.3. 风险评级划分

通过查阅过往研究者的文献资料发现，绿色信贷具有较高的风险和较长的投资周期，一旦发生违规行为，无论对上市公司的长远发展，还是对其有序开展和绿色信贷项目的管理，都将产生巨大的影响。为此，我国应加大对上市公司绿色信用风险的分区力度。因此，本文将 AA+级及以上的公司定为低风险即信用正常，而其余则属于高信用风险类型公司。具体风险评级划分如表 2 所示：

Table 2. Risk rating division

表 2. 风险评级划分

风险评级	风险评级划分
AA+级以上	低风险或正常，表示为 0
AA 级以下	高风险或违约，表示为 1

4.3. 指标优化

4.3.1. Shapiro-Wilk 检验

由于描述性统计分析只是概括性地描述数据特征，具体样本分布仍是未知，为了进一步判断样本具体的总体分布情况，所以本部分利用 SPSS 来对上述的 22 个指标进行 Shapiro-Wilk 检验，做进一步检验和分析。检验结果如表所示，根据表 3 所示，Shapiro-Wilk 检验结果显示，上述全部指标数据均拒绝原假设，即数据不满足正态分布。因此 22 个指标都适合直接采用 Mann-Whitney 秩和检验。

Table 3. Shapiro-Wilk test results

表 3. Shapiro-Wilk 检验结果

指标	统计	自由度	显著性
流动比率 X1	0.748	410	0.000
速动比率 X2	0.713	410	0.000
EBIT 利息保障倍数 X3	0.584	410	0.000
资产负债率 X4	0.875	410	0.000
权益乘数 X5	0.256	410	0.000
资产报酬率 X6	0.638	410	0.000
净资产收益率 X7	0.352	410	0.000
营业毛利率 X8	0.884	410	0.000
营业净利率 X9	0.463	410	0.000
应收账款周转天数 X10	0.558	410	0.000

续表

存货周转天数 X11	0.462	410	0.000
流动资产周转率 X12	0.815	410	0.000
总资产周转率 X13	0.828	410	0.000
净资产收益率增长率 X14	0.037	410	0.000
净利润增长率 X15	0.040	410	0.000
营业收入增长率 X16	0.213	410	0.000
总资产增长率 X17	0.025	410	0.000
总资产 X18	0.577	410	0.000
经营年限 X19	0.960	410	0.000
研发人员数量占比 X20	0.968	410	0.000
研发投入占营业收入比例 X21	0.804	410	0.000
研发人数增长率 X22	0.128	410	0.000

4.3.2. Mann-Whitney U 检验

由于本次实证研究中，各指标下的样本数据不服从正态分布、方差齐性等特点，因而不能采用独立样本 T 检验，故使用 Mann-Whitney U 方法来检验不同风险类别在总体均值上是否有明显差异。基于以上 Mann-Whitney U 检验的检验结果，每股收益等 7 个指标均未通过 0.05 的显著性水平检验，因此本文最终选择了其中通过 0.05 显著性水平检验的 15 个解释变量指标作为以下进一步分析的变量。检验分析结果见下表(表 4):

Table 4. Mann-Whitney U test results
表 4. Mann-Whitney U 检验结果

原假设	显著性	结论
正常公司与违约公司两个不同的样本群中，流通比率 X1 没有差别。	0.000	拒绝原假设
正常公司与违约公司两个不同的样本群中，速动比率 X2 没有差别。	0.000	拒绝原假设
正常公司与违约公司两个不同的样本群中，利息保障倍数 X3 没有差别。	0.321	接受原假设
正常公司与违约公司两个不同的样本群中，资产负债率 X4 没有差别。	0.002	拒绝原假设
正常公司与违约公司两个不同的样本群中，权益乘数 X5 没有差别。	0.000	拒绝原假设
正常公司与违约公司两个不同的样本群中，资产报酬率 X6 没有差别。	0.010	拒绝原假设
正常公司与违约公司两个不同的样本群中，净资产收益率 X7 没有差别。	0.001	接受原假设
正常公司与违约公司两个不同的样本群中，营业毛利率 X8 没有差别。	0.252	拒绝原假设
正常公司与违约公司两个不同的样本群中，营业净利率 X9 没有差别。	0.002	拒绝原假设
正常公司与违约公司两个不同的样本群中，应收账款周转天数 X10 没有差别。	0.000	拒绝原假设
正常公司与违约公司两个不同的样本群中，存货周转天数 X11 没有差别。	0.000	拒绝原假设
正常公司与违约公司两个不同的样本群中，流动资产周转率 X12 没有差别。	0.000	接受原假设
正常公司与违约公司两个不同的样本群中，总资产周转率 X13 没有差别。	0.608	接受原假设
正常公司与违约公司两个不同的样本群中，净资产收益率增长率 X14 没有差别。	0.063	接受原假设

续表

正常公司与违约公司两个不同的样本群中，净利润增长率 X15 没有差别。	0.315	拒绝原假设
正常公司与违约公司两个不同的样本群中，营业收入增长率 X16 没有差别。	0.015	接受原假设
正常公司与违约公司两个不同的样本群中，总资产增长率 X17 没有差别。	0.475	拒绝原假设
正常公司与违约公司两个不同的样本群中，总资产 X18 没有差别。	0.000	拒绝原假设
正常公司与违约公司两个不同的样本群中，经营年限 X19 没有差别。	0.000	拒绝原假设
正常公司与违约公司两个不同的样本群中，研发人员数量占比 X20 没有差别。	0.021	拒绝原假设
正常公司与违约公司两个不同的样本群中，研发投入占营业收入比例 X21 没有差别。	0.000	接受原假设
正常公司与违约公司两个不同的样本群中，研发人数增长率 X22 没有差别。	0.063	拒绝原假设

5. 实证部分

根据多次试验结果，本文从样本中随机抽取 30%作为测试集合，其余 70%作为训练集合，删除存在缺失值以及重复的数据，最终保留数据共 410 条，具体分布如下(表 5):

Table 5. Individual sample size
表 5. 个样本容量

样本分组	正常企业	违约企业	合计
训练组	153	134	287
测试组	60	63	123
总样本	213	197	410

对于分类问题，永远都逃不过的一个痛点就是样本不均衡问题。从上表中可以看出，整合的重污染行业上市公司样本虽然两类样本差距不大，但仍然是不均衡的，可能会导致少数类被错误分类的可能性很大，从而导致分类效果不理想。本文采用 SMOTE 过采样算法对数据进行均衡处理，通过 SMOTE 过采样处理后我们得到信用危险样本 271 条，信用正常样本 271 条，总样本 542 条。处理后的数据样本容量见表 6。

Table 6. SMOTE sample size information after oversampling
表 6. SMOTE 过采样后各样本容量信息

样本分组	正常企业	违约企业	合计
训练组	149	149	298
测试组	64	65	128
总样本	213	214	426

5.1. 相关性与多重共线性分析

根据 Mann-Whitney U 的检验结果，最终选取了其中通过 0.05 显著性水平检验的 15 个指标作为 Logistic 回归模型建模使用的输入自变量变量。在进行 Logistic 回归模型建立之前，需要对数据进行相关性与多重共线性分析，因为本文为全面研究影响深交所主板市场上市公司绿色信贷项目风险表现的因素，

故而选取了较多的指标变量。但是，不同的解释变量之间存在一定的相关性，在没有考虑这些因素的情况下，就会引入大量的噪音，使得对解释变量的显著性检验变得毫无意义。另外，由于模型中的一些重要解释变量也有可能被排除，这将对最终的解释能力和预测能力产生很大的影响。经过 Shapiro-Wilk 检验，发现各指标均不符合正态分布，因此，采用 Spearman 相关分析方法，对相关系数较大的解释变量进行了处理，在剔除变量之前的热力图可见下图 1。

结合 Spearman 相关性系数与相关强度的定义，把 P 值大于 0.8 的解释变量找出并去掉其变量。经过处理后，保留了 11 个解释变量。

其次，是方差膨胀因子 VIF 的筛选。VIF 值是表示自变量相互之间是否存在共线性以及共线性程度大小的一个指标数值。一般 VIF 大于 10 就认为共线性很强，因此对于此类的自变量予以删除。由图 2 中可以看出，解释变量相互之间的相关性均为弱相关或极弱相关，因此没有删除变量，最终同样保留了 11 个解释变量。具体指标显示如图 2 所示。

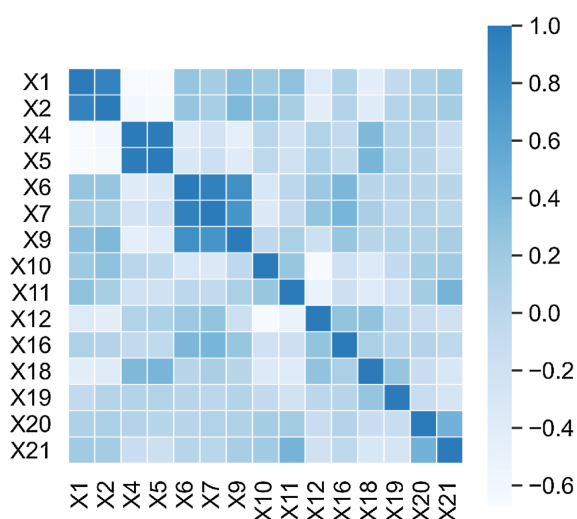


Figure 1. Spearman correlation coefficient thermal maps before high correlation and multicollinearity screening
图 1. 高相关性及多重共线性筛选前的 Spearman 相关性系数热力图

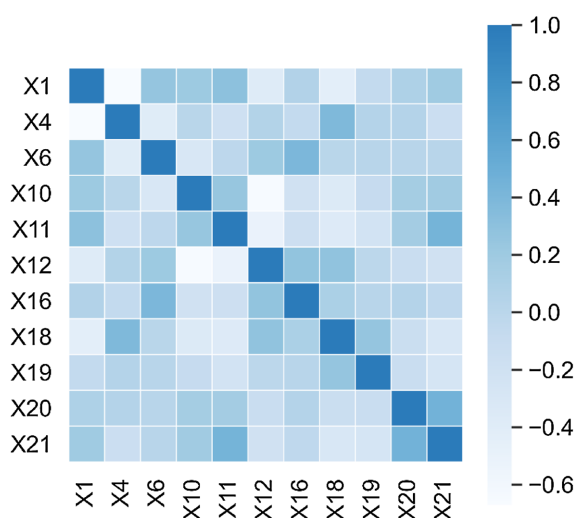


Figure 2. Spearman correlation coefficient heat maps after high correlation and multicollinearity screening
图 2. 高相关性及多重共线性筛选后的 Spearman 相关性系数热力图

在经过相关性与共线性诊断之后，具体的指标选取如下表所示(表 7)。

Table 7. Explanatory variables after high correlation and multicollinearity screening
表 7. 高相关性及多重共线性筛选后的解释变量

指标名称	编号
流动比率	X1
资产负债率	X4
资产报酬率	X6
应收账款周转天数	X10
存货周转天数	X11
流动资产周转率	X12
营业收入增长率	X16
总资产	X18
经营年限	X19
研发人员数量占比	X20
研发投入占营业收入比例	X21

5.2. 回归分析结果

使用以上方法所得出的 11 个指标作为自变量来构建 Logistic 回归模型。由表 8 可知，在自由度为 8、显著水平为 0.05 时，卡方临界值为 12.592。该模型下的卡方值大于以上查表得出的卡方临界值，并且对应的显著性小于 0.05，所以，可以得出迭代过程中的卡方值和显著性均通过检验，故模型整体有效。

Table 8. The comprehensive coefficient test of the model
表 8. 模型的综合系数检验

	卡方	自由度	显著性
步骤	14.155	1	0.000
块	393.022	8	0.000
模型	393.022	8	0.000

霍斯默 - 莱梅肖试验(Horsmer-Leamshaw)试验可以通过计算估算的参数和实测数据的差值来检验模型适合性，如果二者的差值比较大，则不能忽略这一点，因此该模型的拟合效果很差。由表 9 可知，P 值为 0.937，显著性为 0.05，远远超过 0.05，可以忽略检验结果与真实数据之间的差别，拟合得很好。

Table 9. Hosmer-Lemeshaw test
表 9. 霍斯默 - 莱梅肖检验

卡方	自由度	显著性
12.461	8	0.142

表 10 为 Logistic 回归结果，4 个自变量的系数均是具有显著性，该结论充分说明了构建的 Logistic 回归模型是显著有效的，且以上 4 个变量对模型都具有一定的解释力度。因此可以得出，Logistic 回归模型

能够有效地识别关键风险解释变量并对违约风险结果起到一定解释作用。

解释变量 X4、X18 相关系数都为正，证明上市公司的资产负债率和总资产越大，其绿色信贷违约风险概率也就相对较大；解释变量的相关系数都为 X12 和 X19 相关系数都为负，证明上市公司的流动资产周转率和经营年限得分越大，其绿色信贷违约风险概率相对较小。通过上表显示，得到 4 个解释变量均满足以上强影响因素的条件，从而说明了该 4 个解释变量对公司的绿色信贷违约风险均有较强影响，由此表明了资产负债率、总资产、流动资产周转率和经营年限四种指标相比于其他因素对上市公司的绿色信用风险的影响更大。因此，商业银行在给上市公司提供绿色信贷过程中应对以上几个指标的重点进行动态监测，这可以作为判断影响上市公司是否会发生绿色信贷项目违约风险的重要参考。

Table 10. Regression coefficients and significance results for each explanatory variable

表 10. 各解释变量的回归系数和显著性结果

变量	回归系数	标准误差	瓦尔德	自由度	显著性
X4	0.301	0.276	1.189	1	0.000
X12	-0.622	0.235	6.994	1	0.008
X18	0.033	0.040	0.681	1	0.000
X19	-0.148	0.046	10.257	1	0.001
常量	3.667	1.967	3.473	1	0.062

5.3. 模型预测结果分析

为了检验该模型的预测准确率，把测试组的数据样本代入以上构建的 Logistic 回归模型，通过运算就可以得出绿色信贷违约风险概率 p 。本文的二分类模型中，概率划分规则如下：

Table 11. Accuracy of default cases in the model test group

表 11. 模型测试组违约情况的准确率

		预测	
实测		0	1
0	52	8	86.66%
1	10	58	85.29%
总体百分比	-	-	85.98%

对于所选测试组样本，由表 11 可知，Logistic 模型中对正常公司的预测准确率为 86.66%，而对违约公司的预测准确率为 85.29%，总体样本的预测准确率为 85.98%。可见，该 Logistic 模型对上市公司的绿色信贷项目风险的总体预测有较好的准确率，但是违约上市公司的预测准确率仍然有待提高。

在二元分类问题上，通常把预测值为 1、真实值为 1 记作 TP，预测值为 1、真实值为 0 记作 FP，预测值为 0、真实值为 0 记作 TN，预测值为 0、真实值为 1 记作 FN。准确度、精确度、召回率分别以 Accuracy、Precision、Recall 表示。其中，精确度和召回率通过计算又可产生另一指标为 F1_score。具体公式如下：

$$Accuracy = (TP + TN) / (TP + TN + FN + FP) \quad (3)$$

$$Precision(P) = TP / (TP + FP) \quad (4)$$

$$Recall(R) = TR / (TP + FN) \quad (5)$$

$$F1_score = 2 * (P * R) / (P + R)$$

(6)

ROC 曲线是指通过二分类方法，分别以假正率和真正率为横纵坐标、反映敏感性与特异性关系的曲线。ROC 曲线越靠近左上角，即 ROC 曲线下方的面积越大，也就代表模型的准确度就越好，这个准确度可用 AUC 值来表示。

Logistic 回归模型的 ROC 曲线如图 3 所示：

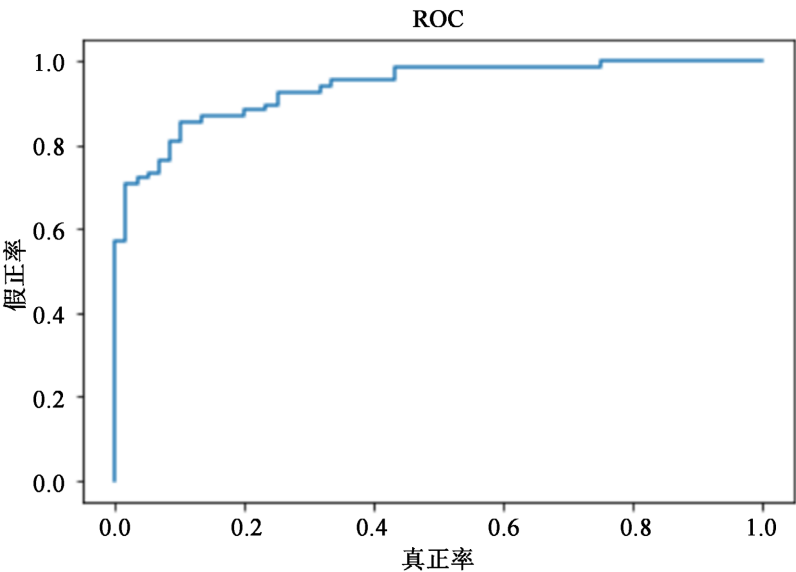


Figure 3. ROC curve of Logistic regression model
图 3. Logistic 回归模型的 ROC 曲线

综上，本部分经过 Logistic 回归建模后得到的性能显示，发现总体预测准确率 Accuracy 较高。模型对上市公司样本绿色信用风险的判别效果良好，对绿色信贷项目信用风险的评估在解释力上具有一定的实际指导意义。并且可以发现，其余指标都是大于 80% 以上，说明违约公司整体的预测都比较好，Logistic 模型整体准确率比较高。Logistic 回归性能评价数据汇总如下(见表 12)：

Table 12. Logistic regression performance evaluation data summary
表 12. Logistic 回归性能评价数据汇总

对比项	Logistic 回归
Accuracy	85.94%
F1_score	87.36%
Precision	88.64%
Recall	85.84%
AUC	93.75%

6. 结论与政策建议

当前，我国正在大力提倡发展绿色经济，以推动经济与环境的可持续发展，在各种绿色金融工具中，绿色信贷的发展是最早的，并且在相关政策的推动下，在商业银行的积极实施下，其在所有的绿色金融产品中都占据了很大的比重，为了确保绿色信贷的更广泛、更好地开展，才能充分地发挥它的作用，因

此,必须加强对绿色信贷的风险管理。在这样的背景下,本文选取了我国重污染上市公司作为研究对象,对我国绿色信贷的信用风险进行了实证研究。在对本文的研究内容进行梳理的基础上,得出了以下几点结论:

(1) 本文从公司的财务指标和非财务指标两个方面,初步选定了 22 个指标建立上市公司绿色信贷信用风险评估初步指标体系。再通过特征筛选剔除得分低的后 11 个指标,以 11 个重要性程度高的指标重新建立上市公司绿色信贷信用风险指标体系。

(2) 本文选取 2018~2022 年深交板主板上市的重污染企业作为数据集,进行缺失值处理、min-max 标准化、SMOTE 过采样处理(随机数种子设置为 0)等一系列数据预处理,通过实验结果分析比较,选出了最适合本文数据集的训练集测试集划分,使用 70%的数据作为训练集,30%的数据作为测试集。

(3) 建立 Logistic 模型进行实证分析,得到混淆矩阵、分类模型的评估报告和 ROC 曲线等一系列评价指标的实验结果,实验结果发现,模型的准确率比较高,准确率、精准率、召回率、f1_score、ROC 曲线和 AUC 值等值都表现不错。

基于以上结论,提出以下建议:第一,要构建和健全我国绿色信贷的信贷风险评价机制。中国央行、环境保护部、银保监会等部门积极推动绿色信用建设,但至今仍没有形成统一的实施标准和评价指标。虽然 17 家商业银行已按照国家有关规定出台了相应的绿色信用标准,但是由于各银行执行情况的差异,导致了对绿色信用的评价比较困难。如何制定统一的绿色信用政策,并构建相应的风险评价系统,是制定绿色信用政策的基础,也是合理评估绿色信用政策执行结果的关键。第二,要加强对绿色信用的风险管理。同时,要加强对绿色信贷的激励。政府有必要对绿色信贷进行优化。近年来,我国出台了一系列具有指导性意义的绿色信贷政策,但在实际执行中仍然存在着难以实现、缺乏有效监管等诸多问题。商业银行是一个逐利的金融机构,其对环境项目的长期性和长期性关注较少,这就造成了绿色贷款的发展滞后。在此基础上,通过政策指导、强化监管、加大金融、税收等方面的扶持,推动商业银行开展绿色信贷,强化风险管理,提升企业的环境治理能力。在有限的管制条件下,政府应该采取强有力的激励措施,以促进金融机构开展绿色信贷业务,实现产品多元化。在此基础上,进一步完善政府信息公开制度,搭建信息共享平台。通过实证分析,我们认为,在环境、社会和公司治理等方面的信息披露不足,增加了银行审批绿色贷款的难度。第三,要加强对绿色信用的审查。在推行绿色信贷的过程中,银行必须对企业的贷款前、中、后等信息进行充分的认识,保证企业能够及时地向企业提供真实的信息以及贷款的使用情况,有助于银行对资本的流动进行监督,避免资金被挪用到污染工程中,降低因信息不对称引起的逆向选择与道德风险。为此,必须构建一套从贷前审核、贷中监管到贷后监管的严密审核机制,保证绿色信贷业务的良性发展。

参考文献

- [1] White, M.A. (1996) Environmental Finance: Value and Risk in an Age of Ecology. *Business Strategy and the Environment*, 5, 198-206. [https://doi.org/10.1002/\(sici\)1099-0836\(199609\)5:3<198::aid-bse66>3.0.co;2-4](https://doi.org/10.1002/(sici)1099-0836(199609)5:3<198::aid-bse66>3.0.co;2-4)
- [2] 何德旭, 张雪兰. 对我国商业银行推行绿色信贷若干问题的思考[J]. 上海金融, 2007(12): 4-9.
- [3] Weber, O., Fenchel, M. and Scholz, R.W. (2006) Empirical Analysis of the Integration of Environmental Risks into the Credit Risk Management Process of European Banks. *Business Strategy and the Environment*, 17, 149-159. <https://doi.org/10.1002/bse.507>
- [4] 孙光林, 王颖, 李庆海. 绿色信贷对商业银行信贷风险的影响[J]. 金融论坛, 2017(10): 31-40.
- [5] 罗霞, 夏梦瑶, 陈新建. 商业银行发展绿色信贷的风险与管理策略[J]. 征信, 2022, 40(3): 72-80.
- [6] Luo, S., Yu, S. and Zhou, G. (2021) Does Green Credit Improve the Core Competence of Commercial Banks? Based on Quasi-Natural Experiments in China. *Energy Economics*, 100, Article 105335. <https://doi.org/10.1016/j.eneco.2021.105335>

-
- [7] 陈琪, 张广宇. 绿色信贷对企业债务融资的影响研究——来自重污染企业的经验数据[J]. 财会通讯, 2019(8): 36-40.
- [8] 廖果平, 李颖. 绿色信贷政策下企业信贷风险测度及影响因素识别[J]. 财会通讯, 2020(2): 121-123.
- [9] 王馨, 王营. 绿色信贷政策增进绿色创新研究[J]. 管理世界, 2021, 37(6): 173-188.
- [10] 舒利敏, 廖菁华. 末端治理还是绿色转型?——绿色信贷对重污染行业企业环保投资的影响研究[J]. 国际金融研究, 2022(4): 12-22.
- [11] 吴世农, 卢贤义. 我国上市公司财务困境的预测模型研究[J]. 经济研究, 2001(6): 46-55.
- [12] 吴彬. 商业银行信用风险管理研究[D]: [硕士学位论文]. 上海: 复旦大学, 2008.
- [13] Belhadi, A., Kamble, S.S., Mani, V., Benkhalti, I. and Touriki, F.E. (2021) An Ensemble Machine Learning Approach for Forecasting Credit Risk of Agricultural SMEs' Investments in Agriculture 4.0 through Supply Chain Finance. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-021-04366-9>
- [14] 邵川. 绿色信贷、风险管理与产业结构调整优化[J]. 江汉论坛, 2020(10): 12-19.
- [15] 王姝怡. 基于 Logistic 回归的绿色信贷信用风险评估研究[D]: [硕士学位论文]. 合肥: 安徽大学, 2021.
- [16] 尹露颖. 基于集成算法的上市公司绿色信贷违约风险预测研究[D]: [硕士学位论文]. 石家庄: 河北地质大学, 2022.
- [17] 王丹迪. 商业银行绿色信贷信用风险评估研究[D]: [硕士学位论文]. 西安: 西安理工大学, 2018.
- [18] 李友慧, 黄颖利, 王宇. 造纸业上市公司信贷风险研究——基于绿色因素分析[J]. 林业经济问题, 2018, 38(3): 46-50, 104.
- [19] Feng, Y. (2022) Bank Green Credit Risk Assessment and Management by Mobile Computing and Machine Learning Neural Network under the Efficient Wireless Communication. *Wireless Communications and Mobile Computing*, 2022, Article 3444317. <https://doi.org/10.1155/2022/3444317>
- [20] 张李红. 基于纳入环境因子的 RAROC 模型的绿色信贷定价方案策划[D]: [硕士学位论文]. 上海: 上海师范大学, 2020.
- [21] 周艳, 陈虎. ESG 信息披露机制下的绿色信贷发展研究——基于企业披露视角[J]. 西南金融, 2022(11): 72-83.