

# 供应链金融大数据信用风险评估与应用研究

周 雷<sup>a</sup>, 赖姝牟<sup>a</sup>, 付煜棋<sup>b</sup>

(苏州市职业大学 a. 商学院; b. 计算机工程学院 江苏 苏州 215104)

**摘要:** 大数据在供应链金融中的应用,推动了供应链信用评估模式和风险管理方法的重大变革。通过将前沿大数据征信技术与评分卡方法相结合,以整车制造供应链为场景,对供应链金融信用风险进行测度。首先,借助 Python 软件,从“企查查”API 数据接口和 Wind 数据库获取相关数据,对 27 家核心企业 122 条供应链多维指标进行数据挖掘、WOE 编码和变量筛选,构建指标体系。然后,运用大数据和人工智能建模思路,建立涵盖 14 个特征解释变量的 Logistic 回归模型,并运用多种工具训练和改进模型,形成可应用于实务的 Logistic 评分卡。经实证检验,最终确定的信用评估模型区分能力强,风险预测准确率能达到 96.77%。基于大数据的 Logistic 评分卡将供应链信用等级数字化,相较于传统的信用评级更具有实用性,因此大数据技术的运用对提升供应链金融信用风险评估和管理水平具有重要价值。

**关键词:** 大数据; 供应链金融; 信用风险; Logistic 评分卡

**中图分类号:** F275; F832 **文献标志码:** A **文章编号:** 2095-0098(2023)01-0064-10

## 一、引言

随着金融科技的创新发展,大数据、区块链、人工智能等新一代信息技术驱动传统供应链金融向数字供应链金融转型,扩大了金融服务覆盖面,可以为供应链核心企业和上下游中小企业提供高效融资。但是,数字供应链金融模式存在交易主体多元化、交易场景虚拟化、交易关系复杂化等特点,使得对其信用风险的识别、评估比较困难。为了对供应链金融信用风险进行有效识别与管控,大多数学者从宏观理论分析与微观风险测度两方面展开研究。在宏观理论分析方面,付玮琼(2020)运用供应链金融网络拓扑结构及依赖共生模型来说明核心企业主导模式下的依赖性垄断现象及其对供应链金融风险的作用机理<sup>[1]</sup>。曹允春等(2020)指出大数据等金融科技技术在不断推动供应链金融智慧化,各交易主体跨界联动及其数据挖掘可以解决传统风险评估中的供应链关系识别、信用风险指标关联性分析等问题<sup>[2]</sup>。赵成国和江文歆(2021)基于供应链金融生态发展逻辑,从资源、组织、知识分析视角建立供应链金融生态风险管理内外协同共生机制,并提出了供应链金融生态风险评估和管理建议<sup>[3]</sup>。周雷等(2021)运用博弈论分析了区块链赋能供应链金融的作用机理,指出区块链的智能合约技术有助于提高信用风险评估效率<sup>[4]</sup>。在微观风险测度方面,大数据征信在很大程度上改变了传统征信对数据采集、加工和分析的方式<sup>[5]</sup>,特别是能够通过重新设计信用评级模型以及创新风险识别系统,更有效地控制和评估供应链金融风险<sup>[6]</sup>。因此,部分学者使用主成分分析和逻辑回归模型<sup>[7]</sup>、Rough 和 GADIMATEL 模型<sup>[8]</sup>、GA-BP 神经网络模型<sup>[9]</sup>、PSO-SVM 供应链金融预警模型<sup>[10]</sup>、

收稿日期: 2022-08-09

基金项目: 教育部人文社会科学研究基金项目“新时代大学生互联网金融风险认知、风险偏好与投资行为研究”(19YJCZH272); 江苏高校哲学社会科学研究项目“数字经济时代金融科技服务实体经济高质量发展研究”(2022SJYB1630); 苏州市职业大学校级研究性课程项目“数智化时代大数据征信应用研究”(SZDYKC-220201)

作者简介: 周 雷(1983—),男,江苏苏州人,讲师,硕士,研究方向为金融科技。

Copula 模型<sup>[11]</sup>等进行了供应链金融信用风险测度。

“十四五”规划指出,要加快构建全国一体化大数据中心体系,鼓励企业开放搜索、电商、社交等数据,加强信用信息归集、共享、公开和应用,推动供应链金融创新发展,强化监管科技运用和金融创新风险评估,提升产业链供应链现代化水平<sup>[12]</sup>。大数据技术在供应链金融信用风险评估中的应用,对于防控信用风险,提升供应链发展水平具有重要的现实意义。从作用机制看,大数据扩大了数据采集范围,提高了数据挖掘能力,有助于建立科学、合理的供应链信用风险评估指标体系;知识图谱与网络分析技术的发展,能更准确和全面地识别上下游核心企业与供应商等构成的复杂供应链关系网络,为防控信用风险,提高供应链管理水平奠定基础;大数据与人工智能建模技术的应用,能更充分地揭示指标间的关联性,提高模型的风险排序、评估和预测能力,赋能数字供应链金融风控与管理。但是,从已有的相关文献看,理论研究和机理分析较多,部分实证研究也大多从学术角度出发,未能提供直接可以实践应用的供应链金融信用风险评价工具;或者仅针对某个特定供应链企业案例展开,对整个行业场景的实践指导价值较弱。鲜有学者将实务中普遍运用的评分卡工具用于供应链金融信用风险测度,而笔者将前沿大数据征信模型与评分卡工具相结合,针对具体行业场景的供应链金融信用风险评估以应用为导向进行深入研究,具有一定的创新性和现实针对性。通过运用大数据和人工智能建模思路,以整车制造行业企业供应链金融为场景,借助 Python3.8 软件进行数据挖掘和指标体系构建,估计供应链金融信用风险评估 Logistic 模型,并运用多种指标和工具训练及检验模型,形成可用于实务的 Logistic 评分卡,将信用等级数字化,提升了供应链金融信用风险评估的准确率和有效性。最后,根据实证研究结论,从供应链金融信用风险治理的角度提出对策建议,以期推动大数据技术在供应链金融中的应用,服务供应链产业链升级,赋能双循环新发展格局。

## 二、场景选择与变量筛选

### (一) 场景选择

大数据供应链金融信用风险评估模型的构建首先需要选择合适的研究对象和模型训练场景。汽车行业具有典型的供应链协同特征,该行业一般以整车制造企业作为核心企业,上游汽车零配件供应商与下游整车经销商围绕整车制造核心企业产生合作,特别是供应商数量多、交易脉络复杂、应收账款和资金需求量较大,而应收账款质押融资模式发展较为成熟,是供应链金融的主流模式。因此,以整车制造供应链为场景探索大数据技术在供应链金融信用风险评估中的应用具有典型性和代表性,供应链样本具体选择条件如下:(1) 核心企业为整车制造的上市公司,属于国民经济行业分类中的“汽车制造业——汽车整车制造业”;(2) 中小融资企业选取自核心企业的上游供应商,考虑数据可得性,根据证监会《上市公司行业分类指引》将创业板、原中小板<sup>①</sup>中的“制造业——汽车制造业”供应商与核心企业共同构成样本供应链。根据以上筛选条件,共选取 27 家上市公司核心企业和 122 条样本供应链作为实证研究对象。

在选择场景的基础上,评估指标设计参考中国知网(CNKI)的 35 篇相关文献,并咨询大数据技术应用与供应链金融领域的专家,初步确定了 42 个原始指标供筛选。指标主要包括六个维度:中小融资企业情况、核心企业情况、供应链关系、宏观经济状况、信用评级情况、信息披露质量。指标数据来源为供应链关系数据通过“企查查”关系图谱 API 接口获取,数据爬取前申请 Key,以 JSON(JavaScript Object Notation)半结构化数据形式返回;由于上市公司的主体信用评级数据有限,因此采用华政 ESG 信用评级作为代理变量;核心企业和中小融资企业的信息披露质量,也选自华政 ESG 信用评级中的信息披露选项;其他指标主要通过 Wind 数据库整理导出。上述指标数据选取的时间点为 2020 年 12 月 31 日,均使用 Python3.10 统计软件进行数据挖掘和建模分析。

在整车制造供应链场景下,设定被解释变量为汽车行业上市公司主体的好坏,对好样本赋值为 0,坏样本赋值为 1。由于我国汽车行业上市公司截至 2021 年 10 月没有发生实质的退市,通常上市公司总体评级

<sup>①</sup> 2021 年 4 月 6 日,深交所主板与中小板正式合并,但原中小板股票代码不变,仍以“002”开头,因此将股票代码“002”开头的,作为核心企业供应商的汽车制造业上市公司也认定为所选样本供应链的供应商。

参考标准为 A 级以上,因此将 A 级作为好坏样本的切分点,评级为 A 级及以上的样本标签为好样本,A 级以下(不含 A 级)的为坏样本。根据采集的数据,122 个汽车行业供应链样本中,好样本 95 个,占比 77.87%;坏样本 27 个,占比 22.13%。为解决数据平衡性问题,使用 Python 进行了 SMOTE 过采样算法处理。

## (二) 变量筛选

1. 分箱与 WOE 编码。运用大数据技术构建 Logistic 评分卡模型,首先需要对各原始特征变量进行分箱与 WOE 编码处理,然后再筛选出满足 Logistic 回归模型要求的解释变量。分箱(Bin)的本质是连续变量的离散化,即对不同属性的变量,根据不同的类别打上不同的特征值。分箱最终的目的是让每一组内差异尽可能小,组间的差异尽可能大。分箱之后需要对变量进行编码。WOE 编码是评分卡模型常用的编码方式。WOE 称为证据权重(Weight of Evidence),计算公式如(1)式所示。WOE 在信贷业务中是衡量违约概率的指标,其本质是坏客户与好客户的比例的对数。WOE 越大,该分箱的坏客户比例越高,也就是说通过对变量进行 WOE 编码,并将 WOE 作为特征值,则特征值不仅代表了变量的一个分类,还代表了 this 分类对预测违约概率的权重或者说对信用风险评估的贡献度。

$$WOE_i = \ln \left( \frac{p_{i1}}{p_{i0}} \right) = \ln \left( \frac{*B_i / *B_T}{*G_i / *G_T} \right) \quad (1)$$

式(1)中  $p_{i1}$  是第  $i$  分箱中坏客户占有所有坏客户的比例,  $p_{i0}$  是第  $i$  分箱中好客户占有所有好客户的比例;  $*B_i$  是第  $i$  分箱中坏客户数;  $*B_T$  是所有坏客户数;  $*G_i$  是第  $i$  分箱中好客户数;  $*G_T$  是所有好客户数。

以资产负债率为例,该变量为连续变量,经分箱处理后,可以划分为三档区间,分别为低于 43.60%、介于 43.60% 与 64.52% 之间以及高于 64.52%。然后运用采集的数据,根据公式(1)可以计算出资产负债率变量各分箱对应的 WOE,如表 1 所示。根据表 1,资产负债率各分箱的 WOE,与坏客户率成正比,坏客户率越高,WOE 越大。因此,WOE 不仅可以作为资产负债率变量离散化处理后各分箱类别的特征值,而且反映了该类别对预测违约概率的权重。采用同样的方法,可以对本文各评估指标变量都进行分箱与 WOE 编码,作为变量的特征值用于后续的分析、筛选与建模。

表 1 资产负债率变量分箱与 WOE 计算

分箱类别	客户数	* $B_i$	* $G_i$	坏客户率	好客户率	* $B_i / * B_T$	* $G_i / * G_T$	WOE
<43.60	42	3	39	7.14%	92.86%	11.11%	41.05%	-1.30695
[43.60, 64.52]	41	10	31	24.39%	75.61%	37.04%	32.63%	0.12677
>64.52	39	14	25	35.88%	64.12%	51.85%	26.32%	0.67803
合计	122	27	95	22.13%	77.87%	100%	100%	0

2. 相关性分析与 IV 变量筛选。经过分箱与 WOE 编码,类别变量实现了赋值,连续变量也进行了离散化处理并取得了特征值,从而可进行相关性分析与 IV 变量筛选。两变量间的线性相关性可用皮尔森相关系数来衡量,该系数的取值范围为  $[-1, 1]$ 。相关系数越接近 0,说明两变量线性相关性越弱;越接近 1 或 -1 则两变量线性相关性越强。当指标间两两相关系数的绝对值大于 0.7 时,说明两者的相关性较强,会影响其对被解释变量的可解释性。因此,当两指标间的相关系数绝对值大于 0.7 时,选择信息增益值 IV 较高的指标作为备选解释变量,去除区分能力较低的指标。将连续变量离散化,必然伴随信息的损失,并且分箱越少,信息损失越大。为衡量特征变量上的信息量以及特征对预测函数的贡献度,定义了信息增益值 IV(Information Value),如式(2)所示。

$$IV = \sum_{i=1}^n (bad\% - good\%) \times WOE_i \quad (2)$$

式(2)中  $n$  代表分箱个数,  $i$  代表第  $i$  个分箱,  $bad\%$  是这个分箱的坏客户,即存在违约风险,标签为 1 的客户占整个特征中所有坏客户的比例,  $good\%$  是这个分箱的好客户,即标签为 0 的客户,占整个特征中所有好客户的比例。

IV 通过在 WOE 基础上乘以系数,不仅避免了负数出现,而且考虑了分箱中样本占总体样本的比例及其预测贡献。对分箱而言,IV 越大代表了这个分箱的坏客户越多。对整个变量而言,IV 代表了特征变量上的

信息量对模型的贡献度和排序的区分度。换言之,代表了分箱后模型对违约概率的预测效率。根据信息增益值 IV 的判定标准,一般需要大于 0.03,模型预测方为有效,因此在两两相关系数大于 0.7 的各组变量中去除 IV 较低指标的基础上,还需要去除虽然相关系数小于 0.7,但是 IV 小于 0.03 的指标。这些指标包括:营业利润、信息披露质量、公司治理能力、GDP 增长率、产业景气指数 AIC、核心企业资产负债率。

3. 多重共线性检验。多重共线性指解释变量之间由于存在相关关系而导致回归模型失真的情况。选取方差膨胀系数 VIF 来衡量多重共线性,如公式(3)所示,其中  $R_i$  是变量  $i$  与其他变量的复相关系数。

$$VIF_i = \frac{1}{1 - R_i^2} \quad (3)$$

式(3)中,方差膨胀系数 VIF 值越大代表解释变量多重共线性程度越严重。将 VIF 阈值设置为 10,当  $VIF > 10$ ,被视为出现多重共线性的情况。因此,筛选出所有方差膨胀系数小于 10 的变量,这些变量之间没有严重的多重共线性,可以构建 Logistic 回归模型。

### 三、Logistic 回归模型

#### (一) 模型建立

Logistic 模型是指被解释变量为二级计分或二类评定的回归分析,一般形式如式(4)。

$$\ln \frac{p}{1-p} = w_0 + w_1 X_1 + w_2 X_2 + \cdots + w_k X_k + \varepsilon \quad (4)$$

式(4)中, $w_0$  表示常数项, $X_k$  代表各解释变量, $w_k$  表示对应的回归系数, $\varepsilon$  为随机误差项。被解释变量为企业违约发生比的自然对数,其中  $p$  为违约概率,越接近 0 表示企业违约风险越低,越接近 1 表示违约风险越高; $\frac{p}{1-p}$  被定义为发生比 odds,表示客户违约的相对概率。

通过指标体系构建,并运用相关性分析、多重共线性检验、IV 等多种方法进行变量筛选,最终根据机器学习梯度下降树算法 GBDT<sup>[13]</sup> 重要程度排序,确定被纳入 Logistic 模型的具体变量,共包括 14 个特征解释变量。为区分核心企业与中小融资企业的财务指标变量,将核心企业的指标名称加上后缀“.H”,同时所有解释变量都经过了分箱(Bin)和 WOE 编码。WOE 编码可以把原始指标与坏客户率之间可能存在的非线性特征转换为线性,这为构建包括 Logistic 回归在内的广义线性模型奠定了基础。然后,通过估计 Logistic 回归,得到大数据供应链金融信用风险评估模型的系数,如表 2 所示。需要说明的是,由于 WOE 编码始终与坏客户率成正相关,即变量的 WOE 编码值与供应链金融违约概率成正比,因此所有的模型系数均为正数。但是,对于指标本身的原始实际值与违约概率成负相关关系的负向指标,在后续制作 Logistic 评分卡与解释模型结果时,需要将其还原,以方便理解。例如,负向指标营业利润率,该指标越高,坏客户比例越低,WOE 编码就越小,则预测违约概率也越低,因此该指标的模型系数虽然也为正数,但实际是负相关关系。

表 2 大数据供应链金融信用风险评估模型系数

变量名称	指标名称	指标方向	模型回归系数
$X_1$	资产负债率_Bin_WOE	正向	1.0095
$X_2$	营业利润率_Bin_WOE	负向	1.0906
$X_3$	净资产增长率_Bin_WOE	负向	0.7376
$X_4$	现金流动负债比率_Bin_WOE	负向	0.1685
$X_5$	流动资产周转率_Bin_WOE	负向	0.9003
$X_6$	销售成本率_Bin_WOE	正向	0.0590
$X_7$	总资产增长率_Bin_WOE	负向	1.0791
$X_8$	存货周转率_Bin_WOE	负向	0.5784
$X_9$	总资产周转率_Bin_WOE	负向	0.8582

变量名称	指标名称	指标方向	模型回归系数
X <sub>10</sub>	应收账款周转率_Bin_WOE	负向	0.8688
X <sub>11</sub>	销售利润率_Bin_WOE	负向	0.2954
X <sub>12</sub>	净利润增长率_Bin_WOE	负向	0.6833
X <sub>13</sub>	销售成本率_H_Bin_WOE	正向	0.4271
X <sub>14</sub>	营业利润率_H_Bin_WOE	负向	0.4520

由于 Logistic 回归模型的被解释变量是对数发生比  $\ln(p/(1-p))$  ,即发生比取自然对数 ,因此每个特征解释变量的估计系数都可以解释为对该对数发生比的作用。将表 2 中的模型系数 ,用对数发生比模型表示 ,如式(5)所示。该模型即是本文构建的大数据供应链金融信用评估模型 ,各变量的回归系数均至少在 5% 的水平上显著。

$$\ln \frac{p}{1-p} = -1.1376 + 1.0095X_1 + 1.0906X_2 + 0.7376X_3 + 0.1685X_4 + 0.9003X_5 + 0.0590X_6 + 1.0791X_7$$
$$+ 0.5784X_8 + 0.8582X_9 + 0.8688X_{10} + 0.2954X_{11} + 0.6833X_{12} + 0.4271X_{13} + 0.4520X_{14} + \varepsilon \quad (5)$$

(二) 模型检验

根据大数据建模原理 ,需要在混淆矩阵基础上 ,综合运用多种指标和评价工具对构建的 Logistic 回归模型进行实证检验 ,并绘制 ROC 曲线 ,直观反映模型的风险排序效果和信用评估能力。对于二分类变量的混淆矩阵涉及以下四种情况:实际坏客户预测为坏客户 TP、实际好客户预测为好客户 FN、实际好客户预测为坏客户类 FN、实际坏客户预测为好客户 FP ,如表 3 所示。

表 3 二分类混淆矩阵

	预测值为 1	预测值为 0	总体 N
实际值为 1	TP	FN	TP + FN
实际值为 0	FP	TN	FP + TN
总体 N	TP + FP	FN + TN	TP + FN + FP + TN

根据混淆矩阵的定义 ,通过对构建的 Logistic 模型设置不同的概率阈值 ,不断测试模型 ,计算出各评估指标的最优值 ,结果如表 4 和图 1 所示 ,其中准确率、精确率、召回率和 LIFT 均为最大值 ,相应地 ,误分类率为最小值。通过表 4 和图 1 可知 ,构建的大数据信用风险评估模型预测的最大准确率为 96.77% ,AUC 值约为 0.88 ,接近于 1 ,模型对供应链金融信用风险的预测效果较好;经模型排序后“分箱”中坏客户浓度最大可达总体的 4.01 倍 ,风险排序能力和区分度强;实际为坏客户却预测为好客户的概率仅为 3.23% ,误分类率较低。

表 4 基于混淆矩阵的大数据模型检验结果

混淆矩阵评估指标	最优值
准确率( TP + TN) /N	96.77%
误分类率( FP + FN) /N	3.23%
精确率 TP/( TP + FP)	96.00%
召回率 TP/( TP + FN)	96.04%
LIFT 提升度	4.01 倍

四、Logistic 评分卡构建与应用

为直观展示 Logistic 回归信用风险评估模型 ,进一步将回归模型转换为 Logistic 评分卡。Logistic 评分卡分值可以定义比率对数线性表达式 ,如式(6)所示。

$$score = A - B \times \ln(odds) = A - B \times \ln \frac{p}{1-p}$$
(6)

式(6)中 A 与 B 是常量 ,根据希望评分卡模型最终输出的总分区间设定 ,且一般均大于 0 ,B 前面的负号表示违约概率越低 ,信用得分越高。通常情况下 ,即分值越高代表风险越低 ,分值越低代表风险越高。将

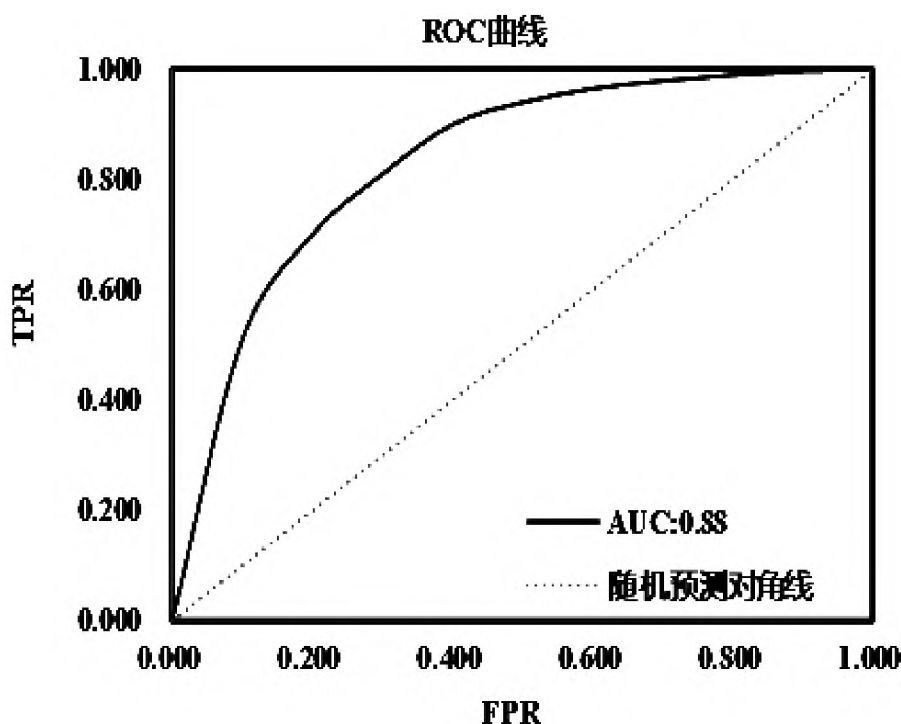


图1 大数据信用风险评估模型的 ROC 曲线

构建的 Logistic 模型式(4)代入式(6),省略随机误差项,可以得到可实际应用的 Logistic 评分公式,如(7)式所示。

$$\text{score} = A - Bw_0 - Bw_1X_1 - Bw_2X_2 - \cdots - Bw_kX_k \quad (7)$$

式(7)中, $A - Bw_0$ 为基础分, $X_k$ 是以WOE编码为特征值的各解释变量, $w_k$ 表示对应的 Logistic 回归模型系数。

转换过程中,首先将构建并通过检验的 Logistic 回归模型系数以及设定的常量  $A$ 、 $B$  代入评分公式(7),可以得到基础分 564 分。然后把各变量每个分箱的特征值 WOE,分别代入公式(7)中该变量对应的  $X$ ,得到每个分箱特征值的评分结果。最后,将各分箱对应的 WOE 编码还原为指标的实际值区间,得到指标各分箱类别的 Logistic 评分。

仍以资产负债率  $X_1$  为例,Logistic 评分计算过程如表 5 所示。

表5 资产负债率 Logistic 评分计算过程

分箱类别	计算过程	得分(取整)
基础分	$A - Bw_0 = 490 - 65 * (-1.1376) = 563.94$	564
$< 43.60$	$-Bw_1 * WOE_{11} = -65 * 1.0095 * (-1.30695) = 85.76$	86
$[43.60, 64.52]$	$-Bw_1 * WOE_{12} = -65 * 1.0095 * 0.12677 = -8.32$	-8
$> 64.52$	$-Bw_1 * WOE_{13} = -65 * 1.0095 * 0.67803 = -44.49$	-44

可以采用与表 5 相同的方法,计算出其余指标的 Logistic 评分,然后汇总、整理,形成完整的供应链金融信用风险评分卡,如表 6 所示。评分卡将每个指标的实际值划分为三档,评分时,在基础分上逐一加上每个评级指标对应的分档得分(负分为减去),得到最终的供应链金融信用风险评分。结果分数越高,代表该条供应链的信用资质越好,出现信用风险的概率越低。基于 Logistic 回归模型转换和生成的供应链金融信用风险评分卡,不仅具有与原 Logistic 模型一致的显著性水平,而且具有很强的可解释性和实际操作应用价值,能够使供应链金融业务人员无需深入了解大数据建模原理,即可根据评分卡直接得出客户的信用评分,以辅助授信决策和风险管理。

表 6 供应链金融信用风险评分卡

指标名称	指标挡位	分档参考区间( 实际值)	得分
	基础分		564
资产负债率	第 1 挡	<43.60	86
	第 2 挡	[43.60 64.52]	-8
	第 3 挡	>64.52	-44
营业利润率	第 1 挡	<1.87	-95
	第 2 挡	[1.87 6.19]	20
	第 3 挡	>6.19	82
净资产增长率	第 1 挡	<1.25	-97
	第 2 挡	[1.25 2.66]	-45
	第 3 挡	>2.66	85
现金流负债比率	第 1 挡	<0.11	-65
	第 2 挡	[0.11 0.15]	-1
	第 3 挡	>0.15	73
流动资产周转率	第 1 挡	<1.64	-37
	第 2 挡	[1.64 2.05]	-15
	第 3 挡	>2.05	67
销售成本率	第 1 挡	<81.56	10
	第 2 挡	[81.56 90.28]	-27
	第 3 挡	>90.28	-37
总资产增长率	第 1 挡	<-3.46	-72
	第 2 挡	[-3.46 2.81]	35
	第 3 挡	>2.81	56
存货周转率	第 1 挡	<1.37	-30
	第 2 挡	[1.37 1.73]	5
	第 3 挡	>1.73	21
总资产周转率	第 1 挡	<0.19	-39
	第 2 挡	[0.19 0.25]	30
	第 3 挡	>0.25	52
应收账款周转率	第 1 挡	<1.57	-85
	第 2 挡	[1.57 2.16]	35
	第 3 挡	>2.16	73
销售利润率	第 1 挡	<1.05	-5
	第 2 挡	[1.05 5.00]	12
	第 3 挡	>5.00	30
净利润增长率	第 1 挡	<49.50	-31
	第 2 挡	[49.50 154.73]	14
	第 3 挡	>154.73	18
销售成本率. H	第 1 挡	<84.27	25
	第 2 挡	[84.27 85.28]	0
	第 3 挡	>85.28	-40
营业利润率. H	第 1 挡	<0.95	-15
	第 2 挡	[0.95 1.14]	22
	第 3 挡	>1.14	90

运用上述 Logistic 评分卡,对 122 条供应链的信用风险进行评分,评分区间的频数分布如图 2 所示。从总体上看,汽车行业供应链金融信用风险得分集中在 [200 800] 区间,其中 [700 800] 区间最多,而高于 900 分的情况较少。进一步以一家供应链核心企业——北汽福田汽车股份有限公司(A 股简称:福田汽车)为例,将福田汽车及其上游中小融资企业真实的原始数据代入 Logistic 评分卡模型,计算福田汽车上游中小融

资企业的信用评分结果如图3所示。根据模型输出结果,这10家供应链上游融资企业的信用风险评分均属于中上部水平。其中评分最高的金力泰为福田汽车上游重要供应商,从事汽车原厂涂料的生产和销售,在垂直细分领域具有极强的竞争力;评分最低的通合科技为新能源电动汽车充电电源及配套设备供应商,在创业板上市,尚未形成稳定的盈利能力,但评分仍略高于基础分,总体信用风险可控,若能抓住“双碳”机遇,加大技术投入和产品研发,仍具有广阔的发展前景。

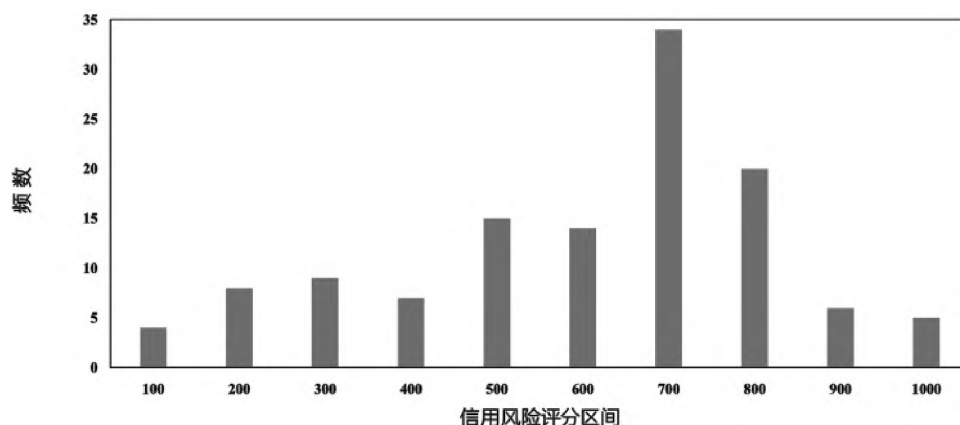


图2 122条供应链信用风险评分结果分布

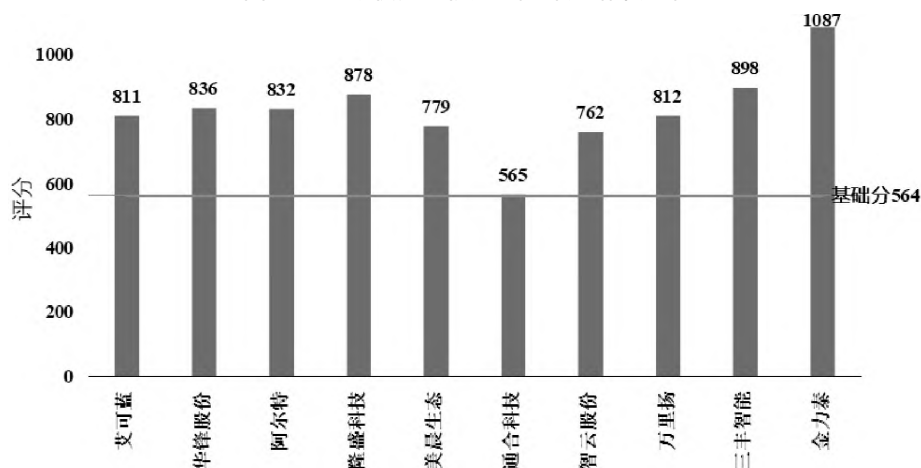


图3 福田汽车上游供应商的融资信用评分结果

根据华政 ESG 评分,上述各家企业的信用评级均为 BBB 级,与全文构建的大数据供应链金融信用风险评分卡模型的评分结果较为吻合,但是本文模型的区分度更高,更加精准地刻画了汽车行业供应链金融的信用风险状况,即实现了信用等级的数字化。传统风险评估模型的底层指标筛选和建模逻辑大多采用德尔菲法、层次分析法等依赖专家经验判断的方法,指标选择、区间分级和权重确定存在一定的主观性,评级结果以信用等级形成给出,精确度和可解释性有限。而基于大数据技术建立的供应链信用风险评估模型,不事先设定指标区间和权重,通过数据挖掘和机器学习构建和优化评估体系,不仅能给出定量、客观的预测违约概率,而且通过将模型转换为 Logistic 评分卡,建立起原始指标与评分结果间的对应关系,实现了信用等级数字化、可解释的呈现,能够清晰获知被评估对象的精确得分和每项指标的特征分。

综上,笔者建立的 Logistic 回归模型对供应链金融信用风险的预测准确率高,基于该模型构建的评分卡方便业务人员操作应用,且相较于传统第三方评级机构 10 个等级的信用评级结果,评分更加精确,更具实用性和参考价值,能够具体解释和分析评分结果与指标特征间的定量关系。

## 五、结论与建议

运用大数据技术和人工智能建模思路,借助 Python3.8 软件,对整车制造企业供应链金融信用风险进行



测度,从“企查查”API 数据接口和 Wind 数据库挖掘相关数据,将 27 家核心企业 122 条供应链指标数据分箱、WOE 编码和变量筛选,建立了涵盖 14 个特征解释变量的 Logistic 回归模型,并运用多种工具验证模型,生成了 Logistic 回归评分卡。实证检验结果表明:本文最终确定的 Logistic 回归模型对供应链金融的信用风险区分能力较强,预测准确率能够达到 96.77%;基于上述模型构建的 Logistic 评分卡将不同信用等级数字化,相较于已有的第三方信用评级工具更加具有大数据征信参考价值和实用性。因此,运用大数据技术测度供应链金融信用风险对提升供应链融资风险管理水平具有重要的现实意义。

具体而言,根据构建的大数据信用风险评估模型回归系数以及最终生成的供应链金融信用风险 Logistic 评分卡,核心企业与中小企业自身的财务指标和表现,均会对供应链金融信用风险产生显著影响,主要结论如下:

首先,核心企业的经营状况、盈利能力对供应链金融信用风险有重要影响。在 5% 的显著性水平下,核心企业的销售成本率与供应链金融违约概率成正相关关系;营业利润率与供应链金融违约概率成负相关关系。(1) 销售成本率指标刻画核心企业经营过程中的成本费用控制能力,核心企业成本费用控制能力越强,销售成本率越低,信用评分越高,供应链金融违约的概率越小。(2) 营业利润率刻画核心企业的盈利能力和发展前景,核心企业营业利润率越高,盈利能力越强,发展前景就越好,整条供应链的信用评分就越高,越能带来更多的融资便利,违约风险就越低。

其次,反映供应链上下游中小融资企业的偿债能力、盈利能力和运营效率的财务指标对供应链金融信用风险的影响均较大。(1) 在 5% 的显著性水平下,中小融资企业的资产负债率与供应链金融信用风险成正相关关系。资产负债率是综合反映企业偿债能力的重要指标,资产负债率越高,企业债务负担越重,信用评分就越低,债务违约概率就越大。(2) 在 5% 的显著性水平下,中小融资企业的营业利润率、总资产增长率与供应链金融违约概率成负相关关系。其中,营业利润率衡量中小融资企业的盈利能力,该指标越大,表明企业的盈利能力越强,从而保障企业还款付息,信用评分就越高,供应链金融链条的违约风险也就越小;总资产增长率反映了中小企业的增长潜力,资产增长越快,相应地评分就越高,信用风险越低。(3) 在 5% 的显著性水平下,中小融资企业的应收账款周转率、总资产周转率、流动资产周转率均与供应链融资违约概率成负相关关系。周转率指标衡量企业的运营效率和资产的流动性,上述指标越大,表明企业的运营效率越高,资产变现能力越强,相应地信用评分就越高,供应链融资违约概率就越小。

结合以上实证研究结论,可以从大数据技术赋能供应链金融信用风险评估和治理的角度提出以下建议:

第一,供应链上的核心企业和小微企业要运用大数据等新一代信息技术提高财务管理水平和信用风险治理能力,从而提升供应链金融质效和整条供应链的财务健康度。根据实证检验结果,供应链上的核心企业与小微企业具有密切的信用联系,两者的财务指标和经营状况均会对供应链信用风险产生显著影响。在数智化时代,要运用大数据、人工智能等新一代信息技术,精准评估供应链信用风险,同时引导核心企业和上下游企业加大技术投入,推进企业智能化改造和数字化转型,改进财务管理,降低运营成本,提升财务绩效和风控能力,从而获得更高的信用评分,提升供应链融资可获得性,满足核心企业和中小企业的各类金融服务需求,提高整条供应链的财务健康度和协同运营绩效,实现可持续发展。

第二,构建供应链数据化生态圈,实现供应链上核心企业、上下游中小企业的融资交易流程透明化。传统供应链金融的相关利益方主体存在信息不对称现象,运用大数据、区块链等技术构建供应链数据化生态圈,能够使得供应链各主体的运营和金融活动中所发生的数据和信息均被不可篡改地完整记录在平台上,实现信息流、资金流、物联和商流的全程可追溯,金融机构、大数据征信机构等相关主体经授权可随时查询数据,采集动态、可信的各类供应链企业财务指标与全周期贸易信息,从而为全面、准确评估供应链金融信用风险奠定基础,在有效识别、评估和管控信用风险的前提下,拓展数字供应链金融新生态,赋能产业链供应链健康发展。

参考文献:

[1] 付玮琼. 核心企业主导的供应链金融模式风险机理研究[J]. 企业经济, 2020(1): 136-143.

- [2]曹允春,林浩楠,李彤. 供应链金融创新发展下的风险变化及防控措施[J]. 南方金融, 2020(4): 36-44.
- [3]赵成国,江文歆. 金融生态视角下供应链金融风险管理体系构建[J]. 财会通讯, 2021(6): 130-133+171.
- [4]周雷,邓雨,张语嫣. 区块链赋能下供应链金融服务小微企业融资博弈分析[J]. 金融理论与实践, 2021(9): 21-31.
- [5]周雷,陈音,张璇,等. 金融科技底层技术创新及其对金融行业的赋能作用[J]. 金融教育研究, 2020(1): 24-33.
- [6]Tian Z, Hassan A. F. S, Razak N. H. A. Big Data and SME financing in China[J]. Journal of Physics: Conference Series, 2019(1): 1-7.
- [7]余得生,李星. 供应链金融模式下的中小企业信用风险评估: 以电子制造业为例[J]. 征信, 2019(10): 72-77.
- [8]周茜,谢雪梅,张哲. 供应链金融下科技型小微企业信用风险测度与管控分析: 基于免疫理论[J]. 企业经济, 2019(8): 146-154.
- [9]孙中叶,徐晓燕. 农业供应链金融风险评估研究: 基于 GA-BP 神经网络模型[J]. 技术经济与管理研究, 2021(8): 78-82.
- [10]李健,张金林. 供应链金融的信用风险识别及预警模型研究[J]. 经济管理, 2019(8): 178-196.
- [11]胡海青,陈迪,张丹,张琅. 基于 Copula 的供应链金融质物组合价格风险测度研究[J]. 运筹与管理, 2020(3): 77-90.
- [12]中华人民共和国国民经济和社会发展第十四个五年规划和 2035 年远景目标纲要[N]. 人民日报, 2021-03-13(1).
- [13]许文彬,张佳韬. 我国城投债信用风险测算: 基于 Logistic 回归的评分卡模型[J]. 数理统计与管理, 2020(4): 705-720.

## An Empirical Study on Credit Risk Assessment in Supply Chain Finance Based on Big Data

ZHOU Lei<sup>a</sup>, LAI Shumou<sup>a</sup>, FU Yuqi<sup>b</sup>

(a. School of Business; b. School of Computer Engineering,  
Suzhou Vocational University, Suzhou, Jiangsu 215104, China)

**Abstract:** The application of big data in supply chain finance has driven a major change in the supply chain credit assessment mode and risk management methodology. By combining the cutting-edge big data credit technology with the scorecard method, we measure the credit risk of supply chain finance by taking the whole vehicle manufacturing supply chains as the scenario. Firstly, with the help of Python, relevant data were obtained from the “Qichacha” API and Wind database, data mining, WOE coding and variable screening were carried out for 122 supply chain multidimensional indicators of 27 core enterprises to construct an indicator system. Then, using big data and AI modeling technologies, a logistic regression model covering 14 characteristic explanatory variables is established, and multiple tools are used to train and improve the model to form a logistic scorecard that can be used in practice. After empirical testing, the finalized credit assessment model has strong differentiation ability and can reach 96.77% risk prediction accuracy. The logistic scorecard based on big data digitizes the supply chain credit rating, which is more practical compared to the traditional credit rating, so the use of big data is of great value to improve the credit risk assessment and management of supply chain finance.

**Key words:** Big data; Supply chain finance; Credit risk; Logistic scorecard

(责任编辑: 黎芳)