

# 信息不对称下银行对中小微企业的最优信贷策略研究

## ——基于 Logistic 回归的违约率测算模型

孙雨忱

(中国人民大学统计学院, 北京 100872)

**摘要:** 通常认为, 中小微企业融资难的主要原因在于抵押担保能力不足、银行贷前及贷后管理成本高。但究其深层次根源, 则是在信息不对称情况下, 银行难以对中小微企业的违约率进行可靠测算, 从而难以估量这部分授信的资本占用, 加大了银行自身资本充足率管理的难度。本文利用银行较易获取的企业交易信息(发票信息)、信用评级、信贷记录等有限信息, 通过构建 Logistic 回归的违约率测算模型, 测算出不同信用水平下中小微企业的预期违约率。在此基础上, 以银行利润最大化为目标, 构建银行最优信贷策略的非线性规划模型, 并以此模型指导银行对潜在中小微企业客户开展授信活动。

**关键词:** 中小微企业; 违约率; 信贷; Logistic 回归

**中图分类号:** F830.5 **文献标识码:** B **文章编号:** 1674-2265 (2021) 06-0078-07

**DOI:** 10.19647/j.cnki.37-1462/f.2021.06.010

### 一、引言

中小微企业是我国构建现代化经济体系、推动经济高质量发展的重要基础, 是扩大就业、改善民生的重要支撑。中小微企业融资问题一直是我国信贷政策关注的重点, 近年来国家出台了很多推进中小微企业信贷业务的优惠政策, 但中小微企业融资难问题一直没有从根本上得到解决。

#### (一) 文献综述

学术界对中小微企业融资难问题的关注和研究由来已久。学者们普遍认为, 中小微企业抵押担保能力差、信用水平低、贷前调查和贷后管理成本高等原因, 导致银行对其授信的积极性不高, 影响了信贷政策的实施效果。其背后的深层次根源则是在信息不对称情况下银行的理性选择。在银行提供信贷服务的过程中, 信息的收集、处理和交流至关重要 (Goetzmann 等, 2013)<sup>[1]</sup>, 造成中小企业融资难的很

重要的原因就是信息不对称。由于信息不对称导致银行对中小微企业限贷或实施信贷配给, 中小微企业往往被排除在信贷市场之外 (Stiglitz 和 Weiss, 1981)<sup>[2]</sup>。为了获取融资, 中小微企业往往需要与所在地的银行建立长期、稳定的关系 (Petersen 和 Rajan, 1994)<sup>[3]</sup>, 否则很难获得信贷资金。信息技术的发展为解决银企之间信息不对称提供了便利, 极大降低了银行对企业硬信息的使用成本, 尤其节约了收集、加工乃至决策的人力成本, 但软信息仍需在银企间长期互动中收集、鉴别与解读, 因此, 获取软信息成本依然很高 (袁志刚等, 2021; 董良泉, 2020)<sup>[4,5]</sup>。

银行对企业授信的前提是尽量维护本金的安全, 因此, 需要根据企业的违约率来决定自己对企业的信贷政策。银行收集和分析企业信息的主要目的, 在于以此判断企业的偿付能力、盈利能力、久期经

收稿日期: 2021-05-25

作者简介: 孙雨忱, 女, 山东济南人, 中国人民大学统计学院, 研究方向为应用统计 (保险精算)。

营能力,归根结底是判断企业信贷的违约率。因此,在信息不对称情况下如何合理判断企业的违约率,就成为解决中小微企业融资难的瓶颈问题。学术界对企业贷款违约率的研究方法大致可分为三类:第一类是利用多种指标构建风险评价指标体系,这类方法需要企业全面的经济财务信息支持,如牛红红和王文寅(2008)<sup>[6]</sup>构建了企业历史信用记录、基础素质、发展潜力、财务状况四类指标为一体的综合性企业信用风险评估体系,乔薇(2011)<sup>[7]</sup>选取了流动资产周转率、存货周转率、净资产收益率、总资产收益率等指标建立了中小企业信用评价体系。第二类是基于Merton期权定价理论,运用KMV等数学模型对企业违约率进行定量分析(Collin-Dufresne和Goldstein, 2001)<sup>[8]</sup>,这类方法要求对企业价值及其波动率进行准确估计。受数据可获性的影响,这两类方法多运用在信息披露较规范的上市公司的信用分析中。第三类是基于各类企业的历史违约率,运用判别分析法(Altman, 1977)<sup>[9]</sup>、Logistic模型等来估算类似企业的预期违约率。这类方法的核心数据是企业的历史违约样本数据,对财务数据要求较少,因此,在非上市的中小微企业信贷风险测度方面应用较多。国内类似研究多集中在判别分析法和Logistic等模型的应用和拓展上(庞素琳和王燕鸣, 2006)<sup>[10]</sup>,徐晓萍和马文杰(2011)<sup>[11]</sup>运用判别分析法和决策树模型,对非上市中小企业的违约风险进行了分析,发现二者结合能较好地判断企业违约率;黄苒等(2018)<sup>[12]</sup>基于违约风险成分分析法,对中小企业违约风险进行分析,发现违约风险成分分析法能较好解释中小企业违约风险的相关性和差异性。

## (二) 研究思路及本文贡献

本文认为,在中小微企业融资难的原因中,缺少抵押担保能力等因素仅是表象。其深层次原因是,在信息不对称情况下,银行难以合理测算中小微企业的偿债能力,具体表现为难以测算中小微企业的预期违约率。在不掌握中小微企业预期违约率的情况下,受制于资本充足率管理要求,银行只能对中小微企业实施限贷或信贷配给。鉴于中小微企业信息披露不充分现象较为普遍,本文借鉴国内相关研究成果,尝试在有限信息条件下,通过构建基于Logistic回归的违约率测算模型,来测算中小微企业违约率数据,从而为银行建立面向中小微企业的最优信贷策略提供参考。本文的边际贡献在于为银行提供了一种利用较少数据对中小微企业违约率进行估

算的方法。银行利用较易获取的企业交易信息(发票信息)、信用评级、信贷记录等有限信息,估算出不同信用水平的中小微企业的预期违约率,并以此为依据完善银行的中小微企业信贷政策。

## (三) 数据来源

本文所需数据来源于2020年全国大学生数学建模大赛数据<sup>①</sup>,包括2017—2020年123家有信贷记录、302家无信贷记录的中小微企业进项和销项发票信息、企业类型、信用评级以及企业贷款利率与客户流失率关系等统计数据。

## 二、违约测算模型的构建

规模较大的企业通常具有较强的抵御风险能力。企业规模的大小可以用很多指标来体现,如资产总额、现金流、购销总额、员工数量等。进、销项发票的流水总额及其增长率指标,分别代表着企业原材料采购和产品销售规模及变化,无疑能反映企业的规模大小、经营实力和抗风险能力,对于银行而言则反映了企业偿付能力的大小。因此,企业的进销项发票信息与企业的预期违约率负相关。企业的信用等级是专业评级公司运用多种评级模型和一系列财务指标对企业实力和信用水平进行的综合评定,信用等级的高低无疑与企业的预期违约率正相关。信贷违约记录则是直接体现企业违约率大小的指标。

综上所述,本模型以违约率来量化中小微企业的信贷风险。将银行的123家现实客户(有信贷记录的中小微企业)数据进行训练集和测试集的分割,应用Logistic回归分析来构建违约率测算模型。

本模型基于以下假设。假设1:银行是否给予中小微企业放贷,取决于银行对企业违约率的预测;假设2:中小微企业信用评级所对应分值的分布近似为正态分布;假设3:中小微企业持续经营且经营效益是长期稳定的。

## (一) 数据处理

1. 利用Excel的预处理。(1)利用数据透视表,将票据信息按照企业做分类求和与计数处理。(2)将企业的“是否违约”变量转化为0-1变量,其中违约记为1,未违约记为0。(3)将企业的“信用评级”这一名义变量转化为数值变量。本文将信用评级对应的量化分值的分布近似视作标准正态分布,参考教学中等级成绩数量化的方法,先计算信用评级为A、B、C、D的各组企业数所占比率,各组企业数比率中点以下的累积概率所对应的z值就是各等级中点

表1: 二元 Logistic 回归的变量选择

指标	回归系数	标准误	自由度	P值	OR值
Zscore(信誉评级)	-5.763	1.642	1	.000	.003
Zscore(进项价税合计)	-85.658	54.089	1	.113	.000
Zscore(销项价税合计)	5.138	5.720	1	.369	170.436
Zscore(总毛利率)	-13.605	7.499	1	.070	.000
Zscore(2020年毛利率)	52.876	31.538	1	.094	$9.197 \times 10^{22}$
Zscore(2020年毛利率同比增长率)	-49.76	30.2828	1	.100	.000
常量	-13.495	6.261	1	.031	.000

所对应的数量化分数。其中, A 对应 1.227826, B 对应 0.321321, C 对应 -0.430727, D 对应 -1.295574 (见图 1)。(4) 根据进销项的有效发票数量和作废发票数量求出有效发票比例; 根据进项价税合计与销项价税合计求出总毛利率。

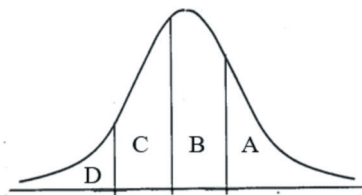


图1: 等级变量的数量化

2. Python 的再次处理。(1) 利用正则表达式提取出企业的性质, 将企业性质分为“个体经营”“有限(责任)公司”“子公司”三类, 分别记为 1、2、3, 并定义为“企业性质”指标。(2) 企业购销时差形成的资金流错期配置会诱发企业资金需求, 企业对资金的需求程度取决于其购买支出与销售收入的资金缺口。本文将企业的初始资金视为 0, 读取数据进行遍历操作, 按时间顺序对企业的入账及出账资金流进行计算, 从而求得在有记录的时间范围内企业曾出现过的最大资金缺口及其时间点, 如企业的入账资金始终大于出账资金, 则最大资金缺口记为 0。(3) 计算各企业年毛利率, 对于缺失数据则进行线性插值处理。

3. SPSS 的最终处理。经过前两次数据处理, 共取得 16 项企业有关指标, 分别是“是否违约( $Y_i$ )”“信用评级( $G_i$ )”“企业性质( $Q_i$ )”“进项价税合计( $T_{in,i}$ )”“销项价税合计( $T_{out,i}$ )”“进项有效发票数( $N_{inv,i}$ )”“销项有效发票数( $N_{out,i}$ )”“进项作废发票数( $N_{inc,i}$ )”“销项作废发票数( $N_{outc,i}$ )”“最大资金缺口( $G_{apmax,i}$ )”“总毛利率( $R_i$ )”“2017 年毛利率( $R_{2017i}$ )”“2018 年毛利率( $R_{2018i}$ )”“2019 年毛利率( $R_{2019i}$ )”“2020 年毛利率( $R_{2020i}$ )”“2020 年毛利率同比增长率( $D_{2020i}$ )”。(1) 对除名义变量“企业性质”之外的指标进行 Z-score 标准化处理。(2) 将上述指标作为自

变量(其中,“企业性质”转换为哑变量)、“是否违约”作为因变量进行二元 Logistic 回归分析。根据结果中各个变量的显著性, 将 P 值过大的不合意指标剔除, 保留 P 值较小或现实意义重要的变量作为二元 Logistic 回归的自变量(见表 1)。

## (二) 基于 Logistic 回归的违约率测算模型

根据企业进销项票据的相关信息可以评判贷款企业的实力和供求关系, 将其与企业所属信用等级的因素进行综合考虑, 即可衡量企业违约率的大小。根据上述数据的处理和显著性水平情况, 以有信贷记录的 123 家企业的“信用评级( $G_i$ )”“进项价税合计( $T_{in,i}$ )”“销项价税合计( $T_{out,i}$ )”“总毛利率( $R_i$ )”“2020 年毛利率( $R_{2020i}$ )”“2020 年毛利率同比增长率( $D_{2020i}$ )”6 个指标作为自变量, 以“是否违约( $Y_i$ )”为因变量, 构建基于 Logistic 回归的违约率测算模型。

1. 理论推导。在对回归模型拟合时, 响应变量描述了一个企业违约的概率。当企业违约率大于设定的阈值时, 则认为该企业会违约, 反之则不会违约。Sigmoid 函数的自变量取值范围极大、因变量的取值局限在  $[0, 1]$  范围内, 在较大自变量的区间里因变量的敏感程度较低。将 Sigmoid 函数值  $f(x)$  视为企业违约的概率, 较符合预测“企业票据信息和信用等级对企业违约与否”的影响系数的模型特点。

$$f(x) = \frac{1}{1 + e^{-x}} \quad (1)$$

对于某企业而言, 它违约的概率为  $f(x)$ , 不违约的概率为  $1 - f(x)$ 。

$$P(y=1|x) = f(x) \quad (2)$$

$$P(y=0|x) = 1 - f(x) \quad (3)$$

将上述两个分类变量的表达式整合如下:

$$P(y|x) = f(x)^y (1 - f(x)) \quad (4)$$

通过这两个分类变量  $P(y=1)$  和  $P(y=0)$ , 可以得到“违约与否”事件的概率:

$$g(x) = \ln\left(\frac{p}{1-p}\right) \quad (5)$$



违约率测算模型公式如下：

$$\ln\left(\frac{p}{1-p}\right) = g(x) = w_0 + w_1G_i + w_2T_{in_i} + w_3T_{out_i} + w_4R_i + w_5R_{2020i} + w_6D_{2020i} \quad (6)$$

其中， $p$ 为预测的违约率， $g(x)$ 为Sigmoid函数， $w_i$ 为 $x_i$ 的权重（ $i=1,2,\dots,n$ ）， $w_0$ 为截距项。

2. 模型拟合结果。将123家已知信贷记录的中小微企业数据进行训练集和测试集的分割，以训练集的自变量和因变量数据对Logistic回归模型进行拟合，将测试集的实际违约情况和模型预测违约情况进行对比，根据准确度和召回率可有效地评判模型的拟合程度。在数据集划分的方法中，选取StratifiedKFold实现五折交叉验证，以保证训练集与测试集的分类分布基本一致，减小划分方式带来的误差。

由于每次交叉验证选取的训练集和测试集不同，在进行模型拟合时，各自变量的系数和截距项可能有所不同，但准确度和召回率相对较高且固定。

此处，取准确度最高的一组系数数据（准确率0.9593、召回率1.0）建立违约率预测模型，并用其测算的违约率值来衡量企业的信贷风险，具体表示为：

$$\ln\left(\frac{p}{1-p}\right) = -2.4400G_i + 0.0931T_{in_i} - 0.3437T_{out_i} - 0.4337R_i + 0.3948R_{2020i} - 0.2579D_{2020i} - 2.2065 \quad (7)$$

其中， $p$ 为企业违约率； $G_i$ 为企业 $i$ 的信用评级对应分值， $T_{in_i}$ 为企业 $i$ 的进项价税总和， $T_{out_i}$ 为企业 $i$ 的销项价税总和， $R_i$ 为企业 $i$ 的总毛利率， $R_{2020i}$ 为企业 $i$ 在2020年的毛利率， $D_{2020i}$ 为企业 $i$ 在2020年的毛利率同比增长幅度。

### 三、最优信贷策略模型的构建

作为市场化经营主体，银行在向中小微企业贷款时必然谋取利润最大化。银行在固定的年度信贷总额内，必然会在信贷收益、客户违约率、贷款利率、客户流失率等指标之间进行综合衡量。本文以银行利润最大化为目标，以银行对企业是否放贷（ $w_i$ ）、企业违约率（ $P_i$ ）、单个企业贷款额度（ $A_i \sim [10, 100]$ 万元）、贷款期限（ $T_i=1$ 年）、贷款利率（ $I_i \sim [4\%, 15\%]$ ）和对应客户流失率（ $L_i$ ）以及总放贷金额小于最大贷款额度（A）为约束，建立单目标非线性规划模型。其中，银行是否放贷根据企业违约率是否超过设定阈值决定，企业违约率通过上文基于Logistic回归的违约率模型测算所得；不同信用评级对应客户的流失率通过曲线模拟与贷款利率建立联系。

#### （一）目标函数

银行在年度授信额度内，综合衡量客户违约风险、客户流失率，谋求利润最大化的目标函数如下：

$$\max Z = \sum_i [W_i A_i I_i T_i (1 - P_i) (1 - L_i) - A_i I_i - a A_i P_i] \quad (8)$$

其中， $W_i$ 表示是否放贷， $A_i$ 为贷款额度， $I_i$ 为贷款利率， $T_i$ 为贷款期限， $P_i$ 为违约率， $L_i$ 为客户流失率。 $I_c=0.015$ 为一年期资金成本，由2020年度人民银行一年期存款基准利率表示； $a$ 为违约客户本金损失率，以中国银保监会公布的2020年度银行业整体不良率1.84%代替。

#### （二）约束条件

1. 银行放贷门槛约束。设定阈值 $P_0$ 来表示银行向企业放贷的门槛：若企业 $i$ 的违约率 $P_i$ 高于 $P_0$ ，则不予放贷；若企业 $i$ 的违约率 $P_i$ 等于或低于 $P_0$ ，则向该企业放贷。

$$W_i = \begin{cases} 1, & P_i \leq P_0 \\ 0, & P_i > P_0 \end{cases} \quad (9)$$

通常，银行会根据自身的风险容忍度和年度授信总额等情况，确定年度授信比率（获贷企业占申贷企业的百分比）。本文对123家有信贷记录的中小微企业违约率进行排名（违约率越低的企业排名越高），并按照各企业违约率排序情况和银行的授信率水平来确定提供贷款的阈值 $P_0$ ，以下是部分数据结果（见表2）。

表2：部分有信贷记录企业的违约率及其排名

企业代号	违约率	信用等级	排名
E1	0.000392657	A	1
E2	0.002098581	A	3
E3	0.174940955	C	67
E4	0.069422849	C	65
E5	0.042051171	B	49
E6	0.002909161	A	9
E7	0.002051814	A	2
E8	0.002629443	A	7
E9	0.002366421	A	4
E10	0.029903172	B	28
E11	0.274319641	C	91
E12	0.035707836	B	29

按照违约率进行排名之后，计算得到信用等级为A、B、C的有信贷记录的中小微企业排名分别位于前21.95%、前52.85%、前80.49%。由于不同银行的信贷政策和风险容忍度存在较大差异，本文将设定几组不同的授信率分类考察最优信贷策略。由于银行对信用评级为D的企业一般不予放贷，本文仅需考虑将阈值 $P_0$ 设在B和C等级对应的违约率之间即可。

2. 贷款利率与客户流失率函数关系约束。通过绘制信用等级为A的企业贷款利率 $I_i$ 与不同信用等级的

客户流失率  $L_i$  的散点图 (见图2), 可判断  $L_i$  与  $I_i$  的关系近似为线性分布或对数分布。由于自变量为 0.04 时因变量为 0, 排除分布为指数小于 1 的幂函数的可能。在此基础上, 利用 SPSS 对  $I_i$  与不同信用评级下的  $L_i$  分别进行曲线拟合。经验证, 对数拟合较线性拟合效果更优。因此, 选取对数拟合。综上, 信用评级为 A 的企业拟合关系如下:

$$L_{iA} = 0.669 \ln(I_i) + 2.239 \quad (10)$$

同理, 对于信用评级为 B、C 的企业, 其拟合关系分别如下:

$$L_{iB} = 0.651 \ln(I_i) + 2.158 \quad (11)$$

$$L_{iC} = 0.659 \ln(I_i) + 2.168 \quad (12)$$

3. 银行年度信贷总额及企业贷款额度约束。在银行年度授信总额 A 固定、银行对单个企业的授信区间为 [10, 100] 万元、贷款年利率区间为 [4%, 15%] 的情况下, 银行利润最大化约束条件如下:

$$\sum_i W_i A_i (1 - L_i) \leq A \quad (13)$$

$$100000 < A_i < 1000000 \quad (14)$$

$$0.04 < I_i < 0.15 \quad (15)$$

(三) 规划模型

综上所述, 得到银行最优信贷策略的非线性规划模型:

$$\max Z = \sum_i [W_i A_i I_i T_i (1 - P_i) (1 - L_i) - A_i I_i - a A_i P_i] \quad (16)$$

$$\text{s.t. } W_i = \begin{cases} 1, & P_i \leq P_0 \\ 0, & P_i \geq P_0 \end{cases} \quad (17)$$

$$L_i = \begin{cases} 0.669 \ln(I_i) + 2.239 \\ 0.651 \ln(I_i) + 2.158 \\ 0.659 \ln(I_i) + 2.168 \end{cases} \quad (18)$$

$$\sum_i W_i A_i (1 - L_i) \leq A \quad (19)$$

$$100000 < A_i < 1000000 \quad (20)$$

$$0.04 < I_i < 0.15 \quad (21)$$

$$I_c = 0.015 \quad (22)$$

#### 四、模型应用

以上研究隐含的前提条件是银行面对着两类中小微企业群体: 一类是现实客户群体 (有信贷记录、有限信息), 另一类是潜在客户群体 (无信贷记录、有限信息)。银行利用现实客户有限数据构建基于 Logistic 回归的违约率测算模型, 取得现实客户的既有违约率与进销项发票等有限信息之间的函数关系

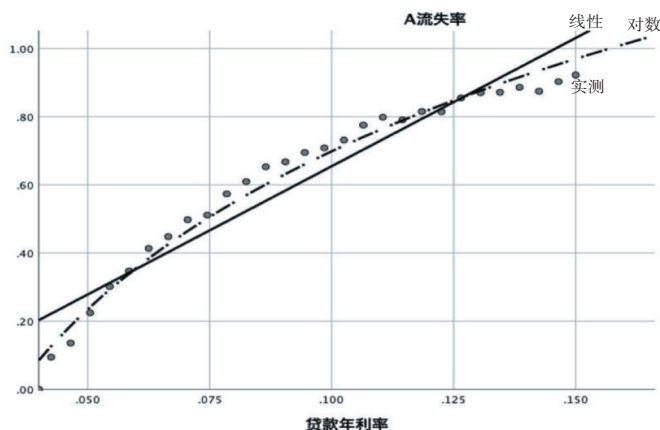


图 2: 信用评级为 A 企业的贷款利率与客户流失率散点图

后, 就可以将这种先验概率与有限信息之间的逻辑关系运用到潜在客户群体中, 在利润最大化目标下进一步拓展授信客户群体。本文以 302 家潜在客户 (无信贷记录、有限信息) 为研究对象, 探讨银行如何运用违约率测算模型计算其预期违约率, 并对银行的最优策略模型进行求解。

值得注意的是, 无信贷记录的中小微企业同样没有信用评级记录、且只具有有限信息 (进销项发票信息), 因此, 需要依据进销项发票信息对其信用评级作出估计, 并代入现有模型进行运算。

(一) 基于主成分分析的信用评级模型和违约率测算

对 302 家无信贷记录企业的 “进项价税合计” “销项价税合计” “总毛利率” “2020 年毛利率” 和 “2020 年毛利率同比增长率” 5 项指标数据进行主成分分析, 总方差解释结果显示该 5 项变量能够很好地描述 302 家企业的性质 (见表 3)。

表 3: 主成分分析总方差解释

成分	初始特征值			提取载荷平方和		
	总计	方差百分比	累积百分比	总计	方差百分比	累积百分比
1	1.810	36.196	36.196	1.810	36.196	36.196
2	1.665	33.303	69.499	1.665	33.303	69.499
3	1.317	26.335	95.834	1.317	26.335	95.834
4	0.199	3.986	99.821			
5	0.009	0.179	100.000			

利用 SPSS 显示的成分矩阵、特征根和方差, 分别计算 3 个主因子  $y_i (i=1,2,3)$  线性组合的系数, 满足如下关系式:

$$y_1 = u_{11}x_1 + u_{12}x_2 + u_{13}x_3 + u_{14}x_4 + u_{15}x_5 \quad (23)$$

$$y_2 = u_{21}x_1 + u_{22}x_2 + u_{23}x_3 + u_{24}x_4 + u_{25}x_5 \quad (24)$$

$$y_3 = u_{31}x_1 + u_{32}x_2 + u_{33}x_3 + u_{34}x_4 + u_{35}x_5 \quad (25)$$

其中  $x_j (j=1,2,3,4,5)$  为进行主成分分析的 5 项指标数据,  $u_{ij}$  为各指标所占的权重。

用变量  $x_j (j=1,2,3,4,5)$  在 3 个主因子  $y_i (i=1,2,3)$  中占的权重  $u_{ij}$  乘以各主因子解释变量的权重, 可得到这 5 项指标在解释 302 家无信贷记录企业特征时的权重, 记为  $\mu_1, \mu_2, \mu_3, \mu_4, \mu_5$ 。以此权重对 302 家企业的各指标进行加权求和, 计算综合分值, 并作为按比例划分企业信用等级的依据。假设信用评级分别为 A、B、C、D 的企业占比是固定的, 且所有企业信用评级对应分值的分布近似为正态分布, 则可根据得分计算出各企业的信用评级, 并可将其转化为数值变量。

在得到企业信用评级数据之后, 代入前文建立的违约率测算模型, 计算出 302 家无信贷记录企业的违约率。需要注意的是, 尽管这样可能会导致模型出现多重共线性问题, 但不会影响对违约率的预测效果。

## (二) 最优信贷策略求解

本文将在银行年度授信总额固定为 1 亿元、银行对单个企业的授信区间为 [10, 100] 万元、贷款年利率区间为 [4%, 15%] 的情况下, 给出银行的最优信贷策略。此时, 只需要讨论银行不同放贷门槛 (阈值  $P_0$ ) 设置对于最优解的影响。银行是否同意向某企业放款, 依赖于企业实力和以往信用等级综合反映出的信贷风险大小。

对无信贷记录的 302 家中小微企业数据进行处理, 在按照预测违约率进行排名后, 可得信用等级为 A、B、C 的企业分别位于排名的前 21.85%、前 52.65% 和前 80.46%。考虑将阈值设在 B 和 C 等级所对应的违约率之间, 将允许借贷的企业百分比拟值分别设为 55%、60%、65%、70%、75% 和 80%, 通过对对应位次前后的违约率求均值, 分别设置阈值为 0.2738、0.2857、0.2901、0.2953、0.2969 和 0.2990。

通过 Lingo 软件计算在贷款总额度给定情况下, 银行在不同放贷门槛下的最优解。将确定的某个阈值代入前文建立的最优信贷策略模型, 求解出银行对不同企业的贷款额、贷款利率的分配, 以及银行在最优策略下的最大收益 (见表 4)。

从表 4 可知, 假设银行年度信贷总额为 1 亿元, 且不考虑外界因素对企业的冲击。当银行允许企业借贷的百分比逐渐增大, 即违约率阈值逐渐增大时, 银行采用最佳借贷策略的收益是逐步增大的, 但收益增量呈现边际递减效应。鉴于信用等级为 D 的企业被排除在信贷市场之外, 银行对企业授信的百分比和阈值的选取是有限制的, 会受 D 类企业占比的影响。因此, 违约率阈值一般不应该大于 0.3。

表 4: 银行最优信贷策略

允许放贷百分比	违约率阈值	银行最大收益(元)
55%	0.2738	3388397
60%	0.2857	3487118
65%	0.2901	3539240
70%	0.2953	3569623
75%	0.2969	3599685
80%	0.2990	3629583

综上, 银行对上述样本的中小微企业的最优策略是: 在满足资本充足率管理要求的前提下, 在国家信贷政策和银行内控制度规范内, 将授信比率设置在 75%—80% 附近、违约率阈值设置在 0.3 附近, 从而实现在风险可控条件下获取较高利润 (约 360 万元)。

## 五、结论

尽管近年来技术进步迅速、银行信贷管理制度不断健全, 但由于信息不对称、经济环境快速变化等原因, 银行业对中小微企业违约率进行合理判断一直存在很大困难。首先, 本文利用现实客户 (123 家有信贷记录的中小微企业) 的交易信息 (发票信息)、信用评级、信贷记录等有限信息, 构建出基于 Logistic 回归的违约率测算模型, 以测度该类企业的违约率与有限信息之间的函数关系; 其次, 以银行利润最大化为目标, 在拟合客户流失率与贷款利率关系的前提下, 构建起银行最优信贷策略的非线性规划模型; 最后, 利用违约率测算模型对潜在客户 (302 家无信贷记录的中小微企业) 的预期违约率进行度量, 并将潜在客户的预期违约率和其他有限信息代入银行最优信贷策略模型, 在限定授信额度、利率区间和单个企业授信额度区间的约束条件下, 对模型求解, 从而得到银行对潜在客户的最优信贷策略。

本模型综合考虑了现金流、风险、客户流失等多种影响银行中小微企业信贷决策的因素, 立足于银企之间信息不对称的实际状况, 利用银行较易获取的中小微企业的发票信息等有限信息, 进行了具有针对性的建模, 从而为银行防范中小微企业信贷风险、推进中小微企业信贷业务、实现利润最大化等目标提供了有效工具。

注:

①建模数据来源: [http://www.mcm.edu.cn/html\\_cn/node/10405905647c52abfd6377c0311632b5.html](http://www.mcm.edu.cn/html_cn/node/10405905647c52abfd6377c0311632b5.html)

参考文献:



- [1]Goetzman W N, Ravid S A, Sverdllove R. 2013. The pricing of Soft and Hard Information: Economic Lessons from Screenplay Sales [J].Journal of Cultural Economics, 37 (2).
- [2]Stiglitz E, Weiss A. 1981. Credit Rationing in Markets with Imperfect Information [J].The American Economic Review, 71 (3).
- [3]Petersen M A, Rajan R G. 1994. The Benefits of Lending Relationship: Evidence from Small Business Data [J].The Journal of Finance, 49 (1).
- [4]袁志刚, 郭学琦, 葛劲峰. 中小企业融资与中小银行发展——由中国现实问题引发的理论与实证分析 [J]. 上海经济研究, 2021, (3).
- [5]董良泉. 中小企业融资需求背景下推进供应链金融的调查与思考 [J]. 金融发展研究, 2020, (11).
- [6]牛红红, 王文寅. 企业信用评估体系研究 [J]. 经济问题, 2008, (1).
- [7]乔薇. 中小企业信用评级指标体系与模型的构建 [J]. 开封大学学报, 2011, (4).
- [8]Pierre Collin-Dufresne, Robert S Goldstein. 2001. Do Credit Spreads Reflect Stationary Leverage Ratios? [J]. The Journal of Finance, Volume 56, Issue5, October.
- [9]Edward I. Altman, Robert G. Haldeman, P Narayanan. 1977. Zeta Analysis: A New Model to Identify Bankruptcy Risk of Corporations [J]. The Journal Banking and Finance, Volume1, Issue1, June.
- [10]庞素琳, 王燕鸣. 判别分析模型在信用评级中的应用 [J]. 南方经济, 2006, (3).
- [11]徐晓萍, 马文杰. 非上市中小企业贷款违约率的定量分析 [J]. 金融研究, 2011, (3).
- [12]黄苒, 范群, 郭峰. 中小企业违约风险系统性和异质性测度——基于违约风险成分分析法的研究 [J]. 中国管理科学, 2018, (3).
- [13]张磊, 姜孟瑞, 刘文贤. 等级成绩数量化与数量成绩等级化 [J]. 山东师范大学学报 (自然科学版), 2006, (2).
- [14]文舒. 信用评级违约率测算方法比较分析 [J]. 金融纵横, 2020, (5).
- [15]司守奎, 孙兆亮. 数学建模算法与应用 [M]. 北京: 国防工业出版社, 2020.

## Research on Banks' Optimal Credit Strategy for MSMEs Under Information Asymmetry

### ——Default Rate Measurement Model Based on Logistic Regression

Sun Yuchen

(School of Statistics, Renmin University of China, Beijing 100872, China)

**Abstract:** It is generally believed that the main reasons for the financing difficulties of micro, small and medium-sized enterprises are insufficient mortgage guarantee capabilities and high pre- and post-loan management costs of banks. However, the deep root cause is that under the situation of information asymmetry, it is difficult for the banks to reliably measure the default rate of small, medium and micro enterprises, which makes it hard to estimate the capital occupation of this part of the credit, increasing the difficulty of the bank's own capital adequacy ratio management. This paper uses the limited information such as corporate transaction information (invoice information), credit ratings, and credit records that are easily available to banks, and builds a logistic regression model of default rate calculation to calculate the expected default rate of small, medium and micro enterprises at different credit levels. On this basis, with the goal of maximizing bank profits, a non-linear programming model of the bank's optimal credit strategy is constructed, and this model is used to guide the bank to carry out credit activities for potential small, medium and micro enterprise customers.

**Key Words:** micro, small and medium-sized enterprises, default rate, credit, Logistic regression

(责任编辑 关 健; 校对 GJ, WY)