

# 基于 Probit 回归的小企业债信评级模型及实证<sup>①</sup>

迟国泰, 张亚京, 石宝峰

(大连理工大学管理与经济学部, 大连 116024)

**摘要:** 小企业债信评级系指评价一笔小企业的债务信用资质的高低和债务违约损失率的大小. 它事关银行贷款的风险管理以及合格的小企业能否得到融资. 因此, 债信评级的指标必须能直接鉴别小企业的违约状态. 以等级评分和排序为目的的现有的信用评级体系并不评价一笔债务违约损失率大小. 现有的信用评级体系也没有任何证据表明其评级指标体系与企业违约状态的鉴别能力有关. 本研究根据指标对违约状态鉴别能力的大小遴选指标体系, 建立了小企业债信评级体系. 本文的创新与特色一是在偏相关系数大于 0.7、反映信息重复的一对高度相关的指标中, 删除  $F$  值小、对小企业违约状态判别能力弱的指标, 既避免了第一次筛选后评级指标体系的信息冗余, 又避免了误删对违约状态影响大的指标, 改变了现有研究评级体系指标的遴选与指标违约状态的鉴别能力无关的状况. 二是通过求解违约状态变量与评价指标之间 Probit 回归方程的回归系数  $\beta$  和回归系数的标准误差  $SE_{\beta}$ , 构建 Wald 统计量对回归系数  $\beta$  的显著水平进行检验, 剔除对小企业违约状态影响小的、回归系数  $\beta$  不显著的指标, 保证了第二次筛选出的指标能显著区分企业的违约状态. 三是研究表明本文构造的指标体系的感受型曲线 ROC 的面积 AUC 大于 0.9, 不仅保证单个指标能有效区分违约状态, 同时确保了整体构建的指标体系对违约状态具有极强的鉴别能力. 四是研究结果表明, 满足信用等级越高、违约损失率越低的小企业债信评级指标体系的非财务指标权重为 56%. 五是通过 1 231 笔小企业贷款数据进行实证, 结果表明: 速动比率、总资产增长率、行业景气指数等 23 个指标能够显著区分小企业的违约状态.

**关键词:** 小企业; 债信评级体系; 指标筛选; Probit 回归

**中图分类号:** F830.56      **文献标识码:** A      **文章编号:** 1007-9807(2016)06-0136-21

## 0 引言

债信评级是指对客户或一笔贷款的信用资质、违约损失率进行评估. 银行参照客户信用等级的高低决策贷款的发放. 因此, 债信评级事关客户

能否从银行获得贷款, 及银行对一笔贷款所承担的风险.

据国家工商总局统计, 小企业已达 1 169.87 万家, 占企业总数的 76.57%, 且提供了 70% 的新增就业和再就业岗位<sup>[1]</sup>. 我国小企业数量庞大,

<sup>①</sup> 收稿日期: 2015-03-26; 修订日期: 2015-12-25.

基金项目: 国家自然科学基金资助项目(71171031; 71471027); 国家自然科学基金青年科学基金资助项目(71201018; 71503199); 教育部人文社会科学研究青年基金资助项目(11YJC790157); 辽宁经济社会发展重点课题资助项目(2015Lslktzdian-05); 辽宁省博士启动基金资助项目(20131017); 河北省自然科学基金青年科学基金资助项目(G2012501013); 教育部科学技术研究项目(2011-10); 中国银监会银行业信息科技风险管理项目(2012-4-005); 大连银行小企业信用风险评级系统与贷款定价项目(2012-01); 中国邮政储蓄银行总行小额贷款信用风险评价与贷款定价资助项目(2009-07). 本文入选“第十二届金融系统工程与风险管理年会”优秀论文(山西大学, 2014年8月).

作者简介: 迟国泰(1955—), 男, 黑龙江海伦人, 博士, 教授, 博士生导师. Email: chigt@dlut.edu.cn

已成为国民经济的重要支柱;也是安置新增就业人员的主要渠道。因小企业的发展需要,其贷款需求也高出大型和中型企业分别 17.2 个和 8.2 个百分点<sup>[2]</sup>。但是,小企业由于自身财务信息不完备、缺乏抵质押品等特点,导致其贷款融资难的问题长期存在,抑制了小企业发展。如何构建一套合理的小企业信用评级模型,帮助解决小企业融资难的问题,成为亟待解决的难题。

### 1) 企业信用评级指标体系的研究现状

“5C”原则是国际经典信用评级方法,主要指品德(character)、资本(capital)、能力(capacity)、担保(collateral)、经营环境(condition of business)五个方面<sup>[3]</sup>,从这五个方面对企业信用状况进行评估。美国穆迪(Moody's)评级公司从企业规模、销售增长、资本结构等方面对企业信用评级<sup>[4]</sup>。标普(S&P)评级机构主要从企业经营状况、财务因素两大方面对企业进行评价<sup>[5]</sup>。惠誉(Fitch)评级机构主要从企业结构、企业盈利、企业战略等方面对贷款企业的信用水平评价<sup>[6]</sup>。大公国际主要从宏观经济环境、行业特性、企业经营管理、企业财务状况、特殊事项风险及外部支持等要素,对工商企业信用质量做出的判断<sup>[7]</sup>。中国银行对企业的信用评级指标主要是现金比率、债务覆盖率、存货周准率等财务指标<sup>[8]</sup>。中国建设银行综合考虑宏观经济数据、客户信用记录、客户财务数据、客户基本面信息四个方面对公司类客户信用评级<sup>[9]</sup>。Li 等利用贷款目的、资产负债率等 17 个指标,对农户贷款的信用风险进行评价<sup>[10]</sup>。Psillaki 等从息税前利润、员工数、固定资产、营业收入等 16 个指标,对企业是否违约进行甄别<sup>[11]</sup>。

现有研究<sup>[4-11]</sup>企业信用评级指标体系是针对大中型企业,并不适合财务信息不完备的中国小企业信用评级研究。

### 2) 企业信用评级指标筛选方法的研究现状

指标筛选是指从大量指标中遴选出用于企业信用评级的指标。Li 等采用投影寻踪模型,遴选了适用于农户信用风险指标体系<sup>[10]</sup>。Shi 等通过显著性判别方法,构建了商户小额贷款信用评级指标体系<sup>[12]</sup>。马晓青等采用逻辑回归模型与因子分析相结合的方法,建立了用于企业信用评级的指标体系<sup>[13]</sup>。

现有研究<sup>[10-12-13]</sup>筛选指标的方法在剔除反

映信息冗余的指标时,往往都是人为主观删除一个指标,并没有考虑哪个指标对企业的违约状态影响显著。

### 3) 企业信用评分方法的研究现状

一是基于数理统计的信用评价方法。早期,Altman 通过线性多元判别法建立了 Z 分数判别函数对 33 家企业进行信用评级<sup>[14]</sup>。Maria 等使用了 Logistic 回归构建了适用于塞尔维亚公司的信用评分模型<sup>[15]</sup>。高丽君以 928 个企业为样本,通过贝叶斯平均生存模型、传统生存模型等三种模型对企业信用违约情况进行估计;并对三种模型的预测结果进行对比,实证表明贝叶斯模型平均生存模型对违约判别的准确性更高<sup>[16]</sup>。

二是基于人工智能方法信用评价方法研究。Hong 等以韩国科技型中小企业为样本,通过支持向量机方法进行信用评级,并将信用评级结果与反向传播神经网络方法所得结果进行对比,结果表明支持向量机的判别准确率更高<sup>[17]</sup>。Niccolò 以 3 100 个意大利制造业的中小企业为样本,通过遗传算法预测企业的违约状况<sup>[18]</sup>。BLANCO 等采用神经网络法,为秘鲁构建了小额信贷机构信用评分模型<sup>[19]</sup>。肖进等采用动态分类器集成选择模型 DCEM 对客户信用状况进行评价<sup>[20]</sup>。余高峰等将粗糙集与改进后的局部变权方法相结合,建立了企业质量信用评价模型<sup>[21]</sup>。

现有研究<sup>[16-19]</sup>的信用评价方法没有专门针对中国小企业的情况,提出适用于我国小企业的信用评价方法。

针对上述问题,本文基于 Probit 回归方法构建了小企业债信用评级模型,并以中国某区域性商业银行 1 231 笔小企业贷款数据为基础进行实证。本文选用 Probit 回归模型的主要依据在于:一是现有研究已经证明 Probit 回归模型在债务预警、个人信用评级、农户信用评级等领域都有良好的评估能力,且对二元离散因变量有较强的判别能力<sup>[22-23]</sup>;二是较少学者以中国商业银行小企业为实证对象,利用 Probit 回归模型构建小企业债信用评级体系,本文旨在利用我国商业银行小企业信贷数据,校验 Probit 回归模型的适用性;三是正如下文 3.6 所述,采用 Probit 回归模型构建的小企业信用评级指标体系,其 ROC 曲线对应的 AUC 值高达 0.987 6,表明该方法在小企业债信用评级

中,具有极高的违约判别能力.

### 1 小企业债信评级的构建原理

#### 1.1 小企业的贷款特点

1) 信息不健全且不易获取<sup>[24]</sup>. 小企业因企业规模小,管理不规范,很多财务信息与非财务信息的经典指标数据都不易获取,给评价小企业信用状况带来困难.

2) 小企业贷款单笔的额度小,还款的不确定性大,但同一企业贷款频次高、不同企业组合后的贷款业务量大,组合风险小.

3) 地区的宏观经济社会环境影响着小企业的还款能力<sup>[24-25]</sup>. 各省份在国家政策、经济发展水平、自然环境等方面都存在较大差异,故不同地区的宏观经济社会环境影响着小企业的偿债能力.

#### 1.2 问题的难点

难点1 在剔除反映信息重复的指标时,如何避免误删对企业违约状态鉴别能力强的指标.

难点2 如何确保遴选的指标能有效判别小企业违约状态.

#### 1.3 突破难点的思路

1) 在偏相关系数大于0.7、反映信息重复的一对高度相关的指标中,删除F值小的、对小企业违约状态判别能力弱的指标(F值的含义将在下文2.3“式(8)的经济学含义”中详细解释),既避免了评级指标体系的信息冗余、又避免了误删对违约状态影响大的指标.改变了现有研究评级体系指标的遴选与指标违约状态的鉴别能力无关的状况.解决难点1.

2) 通过求解违约状态变量与评价指标之间Probit回归方程的回归系数 $\beta$ 和回归系数的标准误差 $SE_{\beta}$ ,构建Wald统计量对回归系数 $\beta$ 的显著水平进行检验,剔除对小企业违约状态影响小的、回归系数 $\beta$ 不显著的指标,保证了筛选出的指标能显著区分企业的违约状态.解决难点2.

小企业债信评级的原理如图1所示.

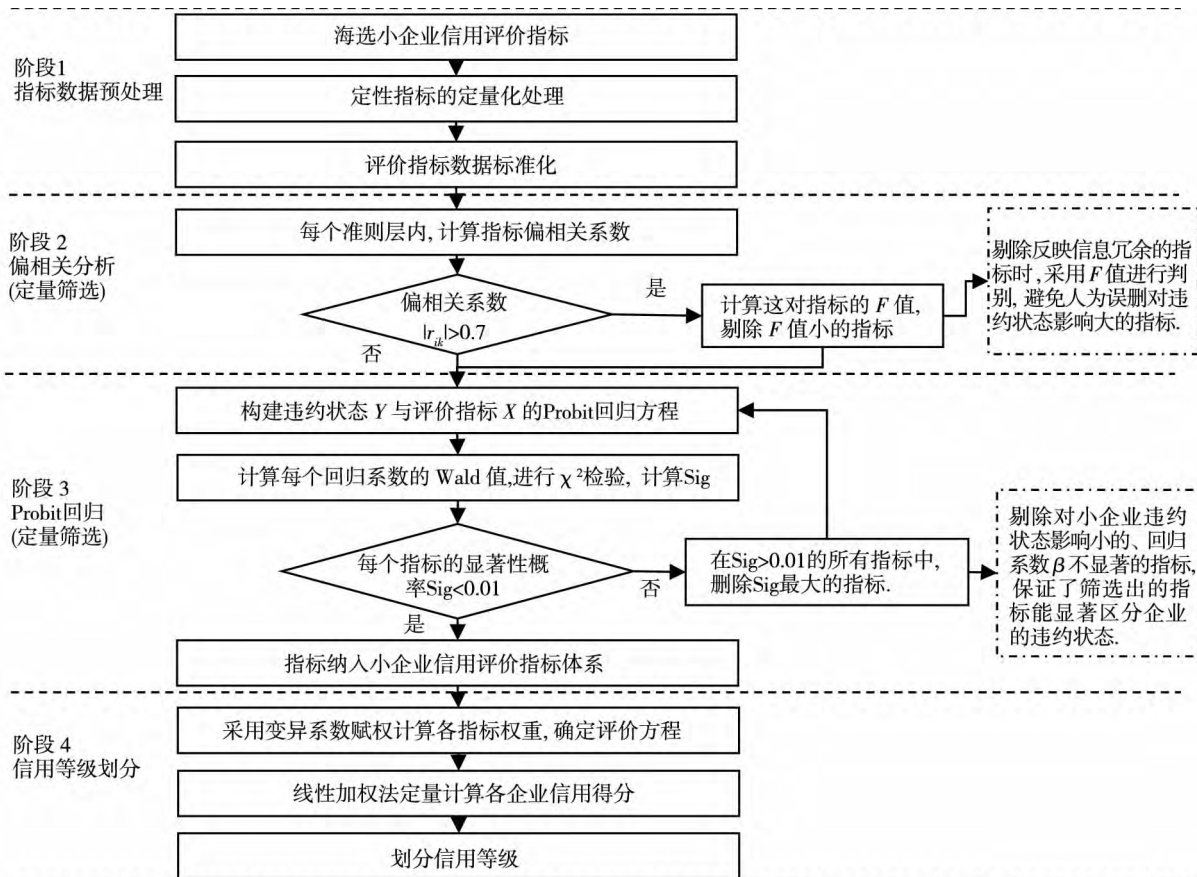


图1 小企业债信评级的原理

Fig. 1 The principle of indicator system of debt rating for small enterprises

## 2 小企业债信用评级模型的构建方法

### 2.1 债信用评级指标的海选

参考国际穆迪、标普等评级机构、及国内建设银行、工商银行等金融机构的经典信用评级指标,同时梳理与信用评价相关的经典文献<sup>[4-19]</sup>,海选小企业的信用评价指标,确保高频经典指标不遗漏。根据数据可观测性原则,将没有数据来源的海选指标删除,保证初步筛选后的指标体系可以量化。

### 2.2 指标的标准化

#### 1) 正向指标的标准化

正向指标: 数值越大表明信用状况越好的指标。设  $x_{ij}$  为第  $i$  个指标第  $j$  个企业标准化后的值;  $v_{ij}$  为第  $i$  个指标第  $j$  个企业的值;  $n$  为企业的总数。根据正向指标的打分公式  $x_{ij}$  为<sup>[26]</sup>

$$x_{ij} = \frac{v_{ij} - \min_{1 \leq j \leq n}(v_{ij})}{\max_{1 \leq j \leq n}(v_{ij}) - \min_{1 \leq j \leq n}(v_{ij})} \quad (1)$$

#### 2) 负向指标的标准化

负向指标: 数值越小信用状况越差的指标。负向指标的标准化公式为<sup>[26]</sup>

$$x_{ij} = \frac{\max_{1 \leq j \leq n}(v_{ij}) - v_{ij}}{\max_{1 \leq j \leq n}(v_{ij}) - \min_{1 \leq j \leq n}(v_{ij})} \quad (2)$$

式(2)中各符号含义与式(1)相同。

#### 3) 区间型指标的标准化

最佳区间型指标: 数值距离某一特定区间越近、表明小企业信用状况越好的指标。例如,“年龄”的最佳区间为  $[31, 45]$ <sup>[26]</sup>,表明企业负责人的年龄越偏离这个区间,这个指标的得分应该越小。

设  $q_1$  为指标最佳区间的左端点;  $q_2$  为指标最佳区间的右端点。根据最佳区间型指标的标准化公式,则  $x_{ij}$  为<sup>[26]</sup>

$$x_{ij} = \begin{cases} 1 - \frac{q_1 - v_{ij}}{\max(q_1 - \min_{1 \leq j \leq n}(v_{ij}), \max_{1 \leq j \leq n}(v_{ij}) - q_2)}, & v_{ij} < q_1 \quad (a) \\ 1 - \frac{v_{ij} - q_2}{\max(q_1 - \min_{1 \leq j \leq n}(v_{ij}), \max_{1 \leq j \leq n}(v_{ij}) - q_2)}, & v_{ij} > q_2 \quad (b) \\ 1, & q_1 \leq v_{ij} \leq q_2 \quad (c) \end{cases} \quad (3)$$

式(3)中其他符号含义与式(1)相同。

#### 4) 定性指标的标准化

由于同一个定性指标存在不同的状态,因此可以针对不同的状态进行打分,从而把定性指标量化。对于可观测的指标,例如下文的 23 个可观测定性指标进行评分,如表 1 所示。

表 1 定性指标打分标准

Table 1 The Standard of qualitative indicators

(1) 准则层	(2) 指标名称	(3) 选项号	(4) 选项内容	(5) 打分
企业内部非财务因素	1 相关从业年限	a	从业年限 ≥ 8 年	1.00
		b	5 年 ≤ 从业年限 < 8 年	0.70
		c	2 年 ≤ 从业年限 < 5 年	0.40
		d	0 < 从业年限 < 2 年 或数据缺失	0.00
	...	...	...	...
	8 是否审计	a	公司主营产品为国家级名牌产品	1.00
		b	公司主营产品为省级名牌产品	0.75
		c	公司主营产品为地市级名牌产品	0.50
		d	公司主营产品为区县级名牌产品	0.25
		e	其他 或数据缺失	0.00
...	...	...	...	
...	...	...	...	...
企业负责人基本情况	23 企业间合同违约次数	a	企业间合同违约 0 次	1.00
		b	企业间合同违约 1 次	0.60
		c	企业间合同违约 2 次	0.30
		d	企业间合同违约 3 次及其以上 或数据缺失	0.00
		...	...	...

### 2.3 基于偏相关分析的第一次筛选

目的 单独一个指标看似很好,但整个评价指标体系中可能会有其他指标与它反映的信息重复,就会造成评价指标体系因含有无用指标而过于庞杂。通过偏相关分析删除反映信息重复的指标。

在准则层内偏相关分析,而不对全部指标偏相关分析的原因:避免指标在数值上相关,而实际经济含义上不相关;因此选择在准则层内相关分析,保证指标若在数据上相关,在经济含义上也一定相关。

#### 1) 准则层内指标之间偏相关系数的计算

设  $x_{ij}$  为第  $i$  个指标第  $j$  个企业的值,  $x_{kj}$  为第  $k$  个指标第  $j$  个企业的值,  $r_{ik}$  为第  $i$  个指标与第  $k$  个指标之间的相关系数,则  $r_{ik}$ <sup>[27]</sup> 为

$$r_{ik} = \frac{\sum_{j=1}^n (x_{ij} - \bar{x}_i)(x_{kj} - \bar{x}_k)}{\sqrt{\sum_{j=1}^n (x_{ij} - \bar{x}_i)^2} \sqrt{\sum_{j=1}^n (x_{kj} - \bar{x}_k)^2}} \quad (4)$$

其中  $n$  表示企业总数,  $\bar{x}_i$  表示第  $i$  个指标的均值,  $\bar{x}_k$  表示第  $k$  个指标的均值.

设  $R$  为第  $i$  个指标和第  $k$  个指标的简单相关系数  $r_{ik}$  组成的相关系数矩阵,  $m$  为此准则层指标的个数. 相关系数矩阵  $R$  如式(5)所示<sup>[27]</sup>.

$$R = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1m} \\ r_{21} & r_{22} & \dots & r_{2m} \\ \vdots & \vdots & \dots & \vdots \\ r_{m1} & r_{m2} & \dots & r_{mm} \end{bmatrix} \quad (5)$$

相关系数矩阵  $R$  的逆矩阵为  $C$ <sup>[27]</sup>

$$C = R^{-1} = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1m} \\ c_{21} & c_{22} & \dots & c_{2m} \\ \vdots & \vdots & \dots & \vdots \\ c_{m1} & c_{m2} & \dots & c_{mm} \end{bmatrix} \quad (6)$$

根据偏相关系数的计算公式, 可得第  $i$  个指标和第  $k$  个指标的偏相关系数<sup>[27]</sup>

$$r'_{ik} = \frac{-c_{ik}}{\sqrt{c_{ii}c_{kk}}} \quad (7)$$

式(7)的经济学含义: 第  $i$  个指标和第  $k$  个指标的偏相关系数  $r'_{ik}$  越大表明第  $i$  个指标和第  $k$  个指标的相关性越强, 第  $i$  个指标和第  $k$  个指标的偏相关系数  $r'_{ik}$  越小两个指标的相关性越弱.

### 2) $F$ 值的计算

计算  $F$  值的目: 在高相关指标中, 为避免主观误删对违约状态影响显著的指标, 分别计算高相关指标的  $F$  值.

设  $F_i$  - 第  $i$  个指标的  $F$  值;  $\bar{x}_i^{(0)}$  - 非违约样本中第  $i$  个指标打分的均值;  $\bar{x}_i$  - 第  $i$  个指标的平均值;  $\bar{x}_i^{(1)}$  - 违约样本中第  $i$  个指标的均值;  $n^{(0)}$  - 非违约样本的总数;  $x_{ij}$  - 第  $j$  个样本第  $i$  个指标的;  $n^{(1)}$  - 违约样本的总数;  $n$  - 小企业样本的总数. 则<sup>[27]</sup>

$$F_i = \frac{(\bar{x}_i^{(0)} - \bar{x}_i)^2 + (\bar{x}_i^{(1)} - \bar{x}_i)^2}{\frac{1}{n^{(0)} - 1} \sum_{j=0}^n (x_{ij} - \bar{x}_i^{(0)})^2 + \frac{1}{n^{(1)} - 1} \sum_{j=1}^n (x_{ij} - \bar{x}_i^{(1)})^2} \quad (8)$$

其中  $\bar{x}_i^{(0)} = \frac{1}{n^{(0)}} \sum_{y_j=0} x_{ij}$ ,  $\bar{x}_i^{(1)} = \frac{1}{n^{(1)}} \sum_{y_j=1} x_{ij}$ ,  $\bar{x}_i = \frac{1}{n} \sum_{j=1}^n x_{ij}$ .

式(8)的经济学含义: 式(8)等号右边的分子是第  $i$  个指标违约、非违约样本均值与总体样本均值的距离之和, 反映了违约样本与非违约样本的差异, 差异越大表明该指标越能区分企业的违约状态. 分母是第  $i$  个评价指标违约样本内的方差与非违约样本内的方差之和, 反映了违约和非违约样本各自的离散程度, 离散程度越小表明违约、非违约样本内部的指标特征越集中. 式(8)中  $F_i$  反映了第  $i$  个指标对企业违约状态的鉴别能力,  $F_i$  越大表明鉴别能力越强, 指标对违约状态的影响越显著; 反之, 该指标对违约状态的影响越小.

偏相关分析好处: 多元相关分析时, 简单相关分析不能真实地反映出两个变量之间的相关程度, 因为它们可能受到不止一个变量的影响, 而偏相关分析的相关系数排除了其他指标的影响.

### 3) 偏相关分析删除指标的标准

两个评价指标的偏相关系数绝对值  $|r_{ik}| > 0.7$ , 认为这两个评价指标高度相关<sup>[27]</sup>, 反映的信息重复, 因此仅保留一个指标, 将另一指标删除. 为避免人为主观误删对违约状态影响显著的指标, 本文采用  $F$  值鉴别指标对违约状态的判别能力, 在偏相关系数绝对值大于 0.7 的指标集中保留  $F$  值大的指标.

### 4) 第一次指标筛选的特点

在偏相关系数大于 0.7、反映信息重复的一对高度相关的指标中, 删除  $F$  值小的、对小企业违约状态判别能力弱的指标, 既避免了评级指标体系的信息冗余、又避免了误删对违约状态影响大的指标. 改变了现有研究的评级体系指标的遴选与指标违约状态的鉴别能力无关的状况.

## 2.4 基于 Probit 回归的第二次筛选

目的 通过在每个准则层内 Probit 回归, 逐步剔除对小企业违约状态影响小的、回归系数  $\beta$  不显著的指标, 筛选出能显著区分小企业违约状

态的指标.

2.4.1 Probit 回归模型

Probit 模型是针对二分类因变量的回归模型, 本文用 Probit 模型筛选对因变量  $y_j$  (违约状态) 影响显著的指标. 因变量  $y_j$  只有两种选择, 指贷款小企业只有违约 ( $y_j = 1$ )、非违约 ( $y_j = 0$ ) 两种状态.

1) 测算潜在变量  $y_j^*$

引入潜在变量  $y_j^*$  的原因: 因企业真实违约状态  $y_j$  是 0、1 离散变量, 无法用线性回归方程直接给出, 则通过间接潜在变量 (企业的违约倾向)  $y_j^*$  与评价指标拟合线性回归方程. 当估算出的潜在变量  $y_j^* > 0$  时, 认为  $y_j = 1$ <sup>[26]</sup>, 即该企业违约; 相反, 当估算出的潜在变量  $y_j^* < 0$  时, 认为  $y_j = 0$ <sup>[26]</sup>, 即企业没有违约.

设  $y_j^*$  - 潜在变量、代表第  $j$  个企业的潜在违约状态;  $\alpha$  - 常数项;  $\beta_i$  - 第  $i$  个指标的回归系数;  $x_{ij}$  - 第  $i$  个指标第  $j$  个企业的值 ( $i = 1, 2, \dots, m, j = 1, 2, \dots, n$ );  $\varepsilon_j$  - 随机误差项, 且  $\varepsilon_j \sim N(0, \sigma^2)$ <sup>[28]</sup>

$$y_j^* = \alpha + \sum_{i=1}^m \beta_i x_{ij} + \varepsilon_j = \alpha + X_j \beta + \varepsilon_j \quad (9)$$

其中  $\beta = (\beta_0, \beta_1, \dots, \beta_m)^T$  为回归系数组成的列向量,  $X_j = (x_{1j}, x_{2j}, \dots, x_{mj})$  为第  $j$  个企业各个指标数值组成的行向量.

式(9)的含义: 由于小企业违约状态  $y_j$  为离散变量, 无法直接评价各指标  $X_j$  与违约状态  $y_j$  之间的非线性函数关系式. 本文通过式(9)建立潜在变量  $y_j^*$  与评价指标  $X_j$  之间的线性回归方程, 进而利用下文式(10)和式(11)、求得每个小企业对应的违约概率, 推导出潜在变量  $y_j^*$  与小企业真实违约状态  $y_j$  之间的非线性函数关系, 实质上也就间接地构建了各评价指标与违约状态  $y_j$  之间的非线性函数关系. 最后, 通过式(12)的对数似然估计求得待估参数  $\alpha, \beta$ .

2) 测算企业的违约概率  $P(y_j = 1 | X_j)$ .

利用式(9)潜在变量  $y_j^*$  求解每个企业样本的违约概率. 因式(9)中的  $\varepsilon_j \sim N(0, \sigma^2)$ , 则企业

的违约 ( $y_j = 1$ ) 概率为<sup>[28]</sup>

$$P(y_j = 1 | X_j) = P(y_j^* > 0 | X_j) = \Phi(\alpha + X_j \beta) \quad (10)$$

其中  $\Phi(\cdot)$  表示正态累计分布函数, 式中  $\Phi(\alpha + X_j \beta)$  函数中的  $\alpha, \beta$  由下文式(12)中求出.

相应可得到企业的非违约概率, 如式(11)所示

$$P(y_j = 0 | X_j) = 1 - P(y_j = 1 | X_j) = 1 - \Phi(\alpha + X_j \beta) \quad (11)$$

3) 通过极大似然估计求解待估计参数  $\alpha, \beta_i$ .

式(9)中的待估计参数  $\alpha, \beta_i$  可以通过极大似然估计法求解, 其对数似然函数为<sup>[28]</sup>

$$\max \ln L = \sum_{j=1}^n [y_j \ln(\Phi(\alpha + X_j \beta)) + (1 - y_j) \ln(1 - \Phi(\alpha + X_j \beta))] \quad (12)$$

式(12)的含义: 对数似然函数  $\ln L$  最大, 表明对违约状态的估计最准确, 此时  $\alpha, \beta$  的估计值即为所求.

4) 求解思路.

步骤 1 求解  $\alpha, \beta$ . 在式(12)中  $y_j$  是真实的违约状态, 是已知的;  $X_j$  是第  $j$  个企业的各指标值组成的向量也是已知的; 只有  $\alpha, \beta$  是未知的、待求解的; 所以每给定一个  $\alpha, \beta$  值, 都能对应计算出一个似然值  $\ln L$ . 给定  $\alpha, \beta_i$  参数一个初值, 将给定的  $\alpha, \beta_i$  代入式(12), 得到对数似然函数值  $\ln L$ . 若此时  $\ln L$  是最大值, 则  $\alpha, \beta_i$  即为所求. 否则, 给定新的  $\alpha, \beta_i$  值, 重复上述过程, 直至式(12)的似然函数  $\ln L$  最大. 一般上述思路通过拟牛顿法<sup>[28]</sup>迭代求解, 由计算机软件实现.

步骤 2 求解企业的违约概率  $P(y_j = 1 | X_j)$ . 将求解出的  $\alpha, \beta_i$  及每个企业的指标值向量  $X_j$  代入式(10), 确定各个企业的违约概率.

步骤 3 求解企业非违约概率  $P(y_j = 0 | X_j)$ . 将求解出的  $\alpha, \beta_i$  及每个企业的指标值向量  $X_j$  代入式(11), 可求解出各个企业的违约概率.

仿此, 每个企业的潜在违约状态  $y_j^*$  也可求出.

2.4.2 指标第二次筛选步骤

在每个准则层内进行 Probit 回归, 计算指标

的 Wald 统计量,进行  $\chi^2$  检验,在对应双尾显著性概率  $Sig > 0.01$  的指标中,删除 Sig 值最大的指标,保留能有效区分违约状态的指标.

1) 基于 Probit 回归第二次筛选指标的具体步骤

步骤 1 求解 Probit 回归系数. 将所有的  $m$  个评价指标与违约状态  $y_j$  的观测值,按照式 (9) - 式 (12) 构建 Probit 回归模型,并求解参数  $\alpha, \beta_1, \beta_2, \dots, \beta_m$  的估计值,及  $\beta_k$  的标准误差  $SE_{\beta_k}$ . 此计算过程由 Stata 软件实现.

步骤 2 求解每个指标的双尾显著性概率值 Sig. 构建 Wald 统计量,对每个指标系数的显著性进行假设检验,查表得到对应的 Sig 值.

假设  $H_0: \beta_k = 0$ . 若  $H_0$  成立,则第  $k$  个指标对违约状态影响不显著.

假设  $H_1: \beta_k \neq 0$ . 若  $H_1$  成立,则第  $k$  个指标对违约状态影响显著.

设  $W_k$  - 第  $k$  个评价指标 Wald 统计量值,  $\beta_k$  - 第  $k$  个指标的参数估计值,  $SE_{\beta_k}$  -  $\beta_k$  的标准误差 (standard error); 则<sup>[28]</sup>

$$W_k = (\beta_k / SE_{\beta_k})^2 \tag{13}$$

式 (13) 的作用 通过构造 Wald 统计量  $W_k$ , 检验  $\beta_k$  是否显著为 0. 在  $H_0$  假设成立的条件下,  $W_k$  服从自由度为 1 的  $\chi^2$  分布,查卡方分布表得到对应的双尾显著性概率 Sig 值.

①若  $Sig < 0.01$ , 则拒绝原假设  $H_0$ , 说明该指标对违约状态影响显著.

②若  $Sig > 0.01$ , 则接受原假设  $H_0$ , 说明第  $k$  个指标的系数  $\beta_k = 0$ , 即该指标对因变量没有影响.

步骤 3 在显著性概率  $Sig > 0.01$  的指标中,剔除 Sig 值最大的一个指标.

$Sig > 0.01$  的指标说明接受  $H_0$  假设,对违约状态影响不显著,在所有不显著的指标中,剔除 Sig 值最大、也就是删除对违约状态影响最小的一个指标. 这一个步骤不能一次全部剔除  $Sig > 0.01$  的指标,只能逐步一个一个剔除,因为一个变量受到多个变量的影响,因而去掉一个变量后再做 Probit 回归,其中有些变量就会变得显著.

步骤 4 将剩余的  $m - 1$  个指标与因变量违约状态  $y$  返回步骤 1 - 步骤 3,直至模型中所有剩余变量的系数都确保  $Sig < 0.01$ , 则停止.

2) 指标第二次筛选的特点

通过求解违约状态变量与评价指标之间 Probit 回归方程的回归系数  $\beta$  和回归系数的标准误差  $SE_{\beta}$ , 构建 Wald 统计量对回归系数  $\beta$  的显著性概率进行检验,剔除对小企业违约状态影响小的、回归系数  $\beta$  不显著的指标,保证了筛选出的指标能显著区分企业的违约状态.

### 2.5 基于 ROC 曲线检验指标体系的有效性

#### 2.5.1 ROC 曲线检验指标体系有效性的方法

目的是通过 ROC 曲线的 AUC 值,检验 Probit 模型筛选出的指标判断样本违约与否的正确率.

将违约样本 ( $y_j = 1$ ) 正确判定为违约的个数,记为 TP (true positive); 将违约样本误判为非违约的个数,记为 FN (false negative); 将非违约样本 ( $y_j = 0$ ) 正确判定为非违约的个数,记为 TN (true negative); 将非违约样本误判为违约的个数,记为 FP (false positive),如表 2 所示.

表 2 Probit 模型分类结果

Table 2 Credit scoring using Probit analysis

实际违约状态	模型分类结果		
	1( 违约)	0( 非违约)	合计
1( 违约)	实际违约样本被模型正确判定为违约的个数 true positive( TP)	实际违约样本被模型误判为非违约的个数 false negative( FN)	实际违约样本数 TP + FN
0( 非违约)	实际非违约样本被模型误判为违约的个数 false positive( FP)	实际非违约样本被模型正确判定为非违约的个数 true negative( TN)	实际非违约样本数 FP + TN
合计	模型判定为违约的样本数 TP + FP	模型判定为非违约的样本数 FN + TN	总样本数 T

ROC( Receiver Operating Characteristic Curve) 曲线, 也称为感受型曲线. ROC 曲线作图时需要用到两个指标: 灵敏度( sensitivity)、特异度( specificity), 如式(14)、式(15)<sup>[29]</sup>所示.

$$\text{sensitivity} = \text{TP} / (\text{TP} + \text{FN}) \quad (14)$$

$$\text{specificity} = \text{TN} / (\text{FP} + \text{TN}) \quad (15)$$

式(14) sensitivity 灵敏度, 是用正确判定为违约的个数 TP 除以实际所有的违约数 TP + FN, 指实际违约的样本被式(10)的 Probit 模型准确判定为违约样本的比率, 它为一种违约状态判别的正确率.

式(15) 特异度 specificity, 是用正确判定为非违约的个数 TN 除以所有的非违约数 FP + TN, 指实际非违约的样本被准确判定为非违约的比率, 它是一种非违约状态判别的正确率.

1 - 式(15)、即 1 - 特异度, 是将实际违约的样本误判为非违约的错误率, 它为一种非违约状态判别的错误率.

ROC 曲线是以 1 - 式(15)、即 1 - 特异度为横坐标, 以式(14) 灵敏度为纵坐标绘制而成的<sup>[29]</sup>.

当横坐标不变时, 纵坐标越大, 实际非违约的样本被准确判定为非违约的比率 TP / (TP + FN) 越大, 这个指标体系区分违约状态的能力越强. 此时, ROC 下边包围的面积 AUC 也就越大. 因此, AUC 面积越大, 则遴选的指标体系鉴别违约状态的能力越强.

ROC 图中的对角线代表判断违约状态的正确率等于判断非违约错误率. 当将样本判断为违约的正确率远大于将样本判断为非违约的错误率, 即 ROC 曲线在对角线之上如图 2 所示时, 表明对违约与否的区分效果较好.

ROC 曲线以下的面积大小, 如图 2 阴影部分所示, 记为 AUC( Area Under Curve), 其判别效果分为五档<sup>[29]</sup>: (1) 当 AUC = 1 时为理想值, 其判别效果最好. (2) AUC = 0.9 以上时, 违约判别效果较好. (3) 当 AUC 在 0.7 ~ 0.9 之间违约判别效果中等. (4) AUC 在 0.5 ~ 0.7 之间违约判别效果较差. (5) 当 AUC 在 0 ~ 0.5 之间违约判别效果极差.

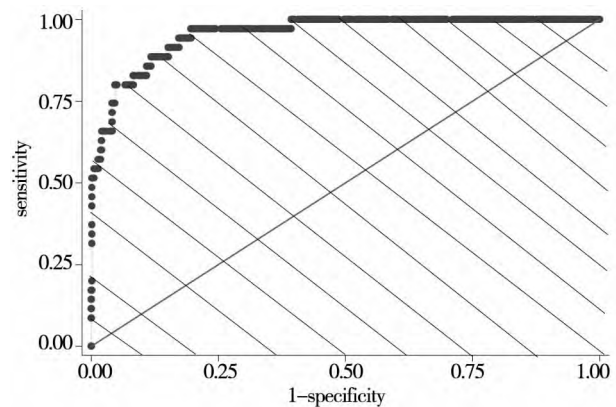


图 2 ROC 曲线

Fig. 2 ROC curve

若上文 2.4 中经过 Probit 模型筛选出的指标体系, 其违约状态判别的 AUC 值能够达到 0.9 以上, 则该指标体系对违约状态有较好的区分效果<sup>[29]</sup>.

研究表明下文构造的感受型曲线 ROC 的面积 AUC 大于 0.9 的指标体系, 确保了整体构建的指标体系对违约状态具有极强的鉴别能力. 通过违约样本被正确判定为违约的个数 TP 占全部实际违约样本的个数 (TP + FN) 的比率, 与非违约样本正确地判定为非违约的个数占实际全部非违约样本的个数 (FN + TN) 的对比关系, 构造感受型曲线 ROC 的面积 AUC 大于 0.9 的指标体系, 确保整体构建的指标体系对违约状态具有极强的鉴别能力.

指标体系的整体鉴别能力, 而不是单个指标的鉴别能力及其重要程度. 因为单个鉴别能力强的指标并不一定能组合成鉴别能力也强的指标体系.

流行的指标不一定组成一个好的指标体系, 因为这个指标的组合, 或这个指标体系的 AUC 值不一定大. 并不流行的指标组成的体系或许反而好, 因为这个指标组合, 或这个指标体系的 AUC 值可能会很大.

### 2.5.2 几点说明

1) 检验整体指标体系有效性的原因. 第一次偏相关分析筛选指标、第二次 Probit 回归筛选指标都是针对单个指标, 只能检验单个指标对违约状态的鉴别能力, 并不能保证最终筛选出的整个



指标体系对违约状态.而实际中为企业进行信用评级使用的整个指标体系,而非单个指标,所以进行整个指标体系的检验是必要的.

2) 指标体系的调整.当债信评级指标体系的 AUC 值较小时,变换筛选指标的方法重新建立指标体系,直至最终构建的指标体系的 AUC 值达到大于 0.9 的标准.本文中使用的指标筛选方法是偏相关分析和 Probit 回归分析,也可使用其他指标筛选方法,如聚类分析、主成分分析、Logit 回归分析等方法去构建债信评级指标体系.直至最终构建的指标体系的 AUC 值达到大于 0.9 的标准.

### 2.6 债信评级指标权重的确定与评价方程的建立

#### 2.6.1 指标权重的确定

设  $q_i$  - 第  $i$  个指标的变异系数;  $n$  - 小企业样本数;  $x_{ij}$  - 第  $i$  个指标第  $j$  个企业的标准化数值;  $\bar{x}_i$  - 第  $i$  个评价指标的均值,则<sup>[26]</sup>

$$q_i = \sqrt{\frac{1}{n-1} \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2} / \bar{x}_i \quad (16)$$

式(16)的含义:第  $i$  个指标的变异系数,反映该指标信息含量的大小,变异系数越大表明信息含量越大,权重也应越大.

设  $w_i$  - 第  $i$  个评价指标的变异系数权重,对式(16)中的  $q_i$  作归一化处理,则<sup>[26]</sup>

$$w_i = q_i / \sum_{i=1}^m q_i \quad (17)$$

#### 2.6.2 评价方程的建立

设  $z_j$  - 第  $j$  个企业的加权得分,则<sup>[26]</sup>

$$z_j = \sum_{i=1}^m w_i x_{ij} \quad (18)$$

式(18)计算所得加权得分  $z_j$  在  $[0, 1]$  区间内,为了符合人们的习惯,将  $z_j$  化为 0 - 100 分之间的得分,如式(19)所示.

设  $S_j$  - 第  $j$  个企业的信用得分,则<sup>[26]</sup>

$$S_j = 100 \times (z_j - \min(z_j)) / (\max(z_j) - \min(z_j)) \quad (19)$$

### 2.7 信用等级的划分

本文将客户划分为 AAA、AA、A、BBB、BB、B、CCC、CC、C 九个信用等级,从 AAA 到 C 信用等级

依次降低,代表客户信用依次变差.

应该指出,因现实中几乎每个信用等级都会有违约的客户出现,所以本文在划分信用等级时,为了使划分结果更接近实际,要求每个等级至少有一笔违约的贷款,以保证每个等级的违约损失率 LGD 大于 0.

#### 1) 信用等级的初步划分.

将客户按信用得分由高到低排列,将客户划分为 9 个等级,每个等级都要有违约的客户,并根据客户的应收本息和应收未收本息,计算每个等级违约损失率.

#### 2) 信用等级的调整.

不断调整每个等级的界限,即调整每个等级的企业数,使等级划分的结果满足信用等级越高违约损失率越低的原则.

等级界限调整过程中,一旦一个等级的上下限改变,就会引起相邻等级的违约损失率的变化,使相邻等级的界限也会被迫进行调整.实际调整的过程比较复杂,且不属于本文的研究范围与创新点,本文不再赘述,具体参见课题组研究成果<sup>[26]</sup>.

## 3 小企业债信评级的实证分析

### 3.1 指标的海选

以穆迪公司、标普公司等国外权威机构及中国建设银行等国内金融机构的经典高频指标为基础<sup>[4-13]</sup>,结合权威研究文献<sup>[14-19]</sup>,海选出用于小企业债信评价的 107 个指标,如表 3 第 5 列所示.指标主要涵盖了还款能力、还款意愿两个一级准则层,企业内部财务因素、企业外部宏观条件、抵押担保因素等 7 个二级准则层,其中,企业内部财务因素又细分为偿债能力、盈利能力等 4 个三级准则层,如表 3 第 2 列 ~ 第 4 列所示.每个指标的类型如表 3 第 6 列所示,指标对应的文献来源如表 3 第 7 列所示.

根据数据可观测性原则,剔除“还款来源”、“每股现金流”、“工资、福利增长率”等 26 个数据不可获得的指标,剔除的指标在表 3 第 8 列用“数据不可获得删除”标注.删除 26 个指标后,剩余 81 个债信评级指标,如表 5 第 e 列所示.

表 3 小企业债信用评级海选指标集

Table 3 Index set of credit risk evaluation chosen extensively for small enterprises

(1) 序号	(2) 一级准则	(3) 二级准则	(4) 三级准则	(5) 指标名称	(6) 类型	(7) 参考文献	(8) 筛选方式
1	A1 还款能力	B1 企业内部财务因素	C1	资产负债率	负向	[6-11], [21-22]	偏相关分析删除
...			...	...	...	...	...
24			偿债能力	还款来源	定性	[6-8]	数据不可获得删除
25			C2	净资产收益率	正向	[6], [9-11], [24]	Probit 删除
...			...	...	...	...	...
40			盈利能力	每股现金流量	正向	[10], [24], [26]	数据不可获得删除
41			C3	应收账款周转速度	正向	[6-11], [24]	保留
...			...	...	...	...	...
50			营运能力	现金循环周期	正向	[10], [24], [26]	偏相关分析删除
51			C4	营业收入增长率	正向	[6-11]	Probit 删除
...			...	...	...	...	...
59			成长能力	工资、福利增长率	正向	[24], [26]	数据不可获得删除
60			B2 企业外部宏观条件	行业景气指数	正向	[9-11]	保留
...				...	...	...	...
70				行业发展趋势	定性	[6-11], [24]	数据不可获得删除
71			B3 企业内部非财务因素	相关行业从业年限	定性	[9], [24], [26]	保留
...				...	...	...	...
83	新产品开发能力	正向		[10], [24], [26]	数据不可获得删除		
84	A2 还款意愿	B4 企业负责人基本情况	学历	定性	[6-11]	Probit 删除	
...			...	...	...	...	
97			管理团队素质	定性	[6-11], [24], [26]	数据不可获得删除	
98		B5 企业基本信用情况	企业到位注册资金类别	定性	[6-10]	偏相关分析删除	
...			...	...	...	...	
102			利息按期偿还率	正向	[6-9], [26]	数据不可获得删除	
103	B6 企业的商业信誉	企业纳税记录	定性	[8-10], [24]	偏相关分析删除		
...		...	...	...	...		
106		企业间合同违约次数	定性	[6-11]	保留		
107	B7 抵质押担保因素		抵押/质押/担保/信用	定性	[6-11], [24], [26]	保留	

3.2 样本选取和数据来源

1) 样本选取

根据中华人民共和国工业和信息化部发布的《中小企业划型标准规定》<sup>[30]</sup>, 批发业中从业人员 5 人及以上, 且营业收入 1 000 万元及以上的为小型企业. 零售业、建筑业等其他行业也按照《中小企业划型标准规定》中要求的从业人数和营业收入选出小企业样本.

按照工信部行业划分标准<sup>[30]</sup>, 将所有小企业样本划分为批发业、零售业等 12 个行业, 行业分布如表 4 第 2 列所示; 每个行业的违约样本数及总样本数如表 4 第 3 列、第 4 列所示.

表 4 按工信部标准分类的各行业小企业样本数

Table 4 The distribution of small enterprises by the standard of Ministry of Industry and Information

(1) 序号	(2) 行业划分	(3) 违约样本数	(4) 总样本数
1	批发业	6	495
2	零售业	0	192
3	建筑业	3	154
4	其他企业	7	151
5	交通运输业	0	101
6	租赁和商务服务业	14	81
7	软件和信息技术服务业	0	20
8	仓储业	0	16
9	住宿、餐饮业	0	11
10	信息传输业	4	7
11	房地产开发经营行业	1	3
12	工业	15	1 814

表5 小企业债信用评级指标筛选原始数据

Table 5 The original data of indicator of credit risk evaluation for small enterprises

(a) 序号	(b) 一级准则	(c) 二级准则	(d) 三级准则	(e) 指标名称	(f) 指标类型	贷款借据原始数据 $v_{ij}$				(g) 最大值 $\max(v_{ij})$	(h) 最小值 $\min(v_{ij})$
						1 196 笔非违约借据		35 笔违约借据			
						(1) 200410270004	(1 196) X2012060800099	(1 197) 200410090044	(1 231) X2010121600004		
1	A1 还款能力	B1 企业财务状况因素	C1 偿债能力	$X_1$ 资产负债率	负向	0.643	0.556	0.640	0.197	0.95	0.01
2				$X_2$ 流动负债经营活动净现金流比率	正向	0.306	0.010	1.125	0	4.09	-3.75
...					...	...	...	...	...	...	...
20				$X_{30}$ EBITDA 与总负债比率	正向	0.043 4	0.003	0.031	0.000	6.15	0.00
21			C2 盈利能力	$X_{31}$ 净资产收益率	正向	0.078 4	0.003	0.054 6	0.000	1.02	0.00
...					...	...	...	...	...	...	...
34			C3 营运能力	$X_{34}$ 应收账款周转速度	正向	5	13.194	1.937	0.048	335.03	0.05
...					...	...	...	...	...	...	...
43			C4 成长能力	$X_{43}$ 现金循环周期	正向	-1 412.309	7.500	-2.546	0	1 398.07	-1 412.31
44				$X_{44}$ 营业收入增长率	正向	0	0.023	0	0	4.09	-1.00
...			...	...	...	...	...	...	...		
48		$X_{48}$ 留存收益增长率	正向	0.076 0	1.251	1.810	6	75.54	-78.55		
49	A2 还款意愿	B2 企业外部宏观条件	$X_{49}$ 行业景气指数	正向	137.45	139.5	127.96	130.1	165.10	65.71	
50			$X_{50}$ 城乡居民人均储蓄年末余额	正向	23 388.05	64 380.629	23 388.05	50 113.69	86 009.91	3 718.05	
51			$X_{51}$ GDP 增长率	正向	15.2	13.5	15.2	15	27.86	8.20	
52			$X_{52}$ 居民消费价格指数	区间	100.6	105.4	100.6	100.2	121.4	98	
53			$X_{53}$ 城市居民人均可支配收入	正向	9 101.35	24 276	9 101.35	19 014	36 230.00	3 058.00	
54			$X_{54}$ 恩格尔系数	负向	39.4	37	39.4	37.9	54.30	31.40	
55	B3 企业内部非财务因素	$X_{55}$ 相关行业从业年限	定性	从业年限 $\geq 8$ 年	从业年限 $\geq 8$ 年	从业年限 $\geq 8$ 年	2 年 $\leq$ 从业年限 $< 5$ 年	—	—		
...			...	...	...	...	...	...	...		
64	B4 企业负责人基本情况	$X_{64}$ 学历	定性	专科	本科及以上学历	其他 或数据缺失	本科及以上学历	—	—		
...			...	...	...	...	...	...	...		
71	B5 企业基本信用情况	$X_{71}$ 年龄	区间	46	50	78	47	78	15		
...			...	...	...	...	...	...	...		
75	B6 企业的商业信誉	$X_{75}$ 企业到位注册资金类别	定性	资金型注册资金	资金型注册资金	数据缺失	资金型注册资金	—	—		
...			...	...	...	...	...	...	...		
77	B7 抵押担保因素	$X_{77}$ 企业纳税记录	定性	纳税历史三年及以上、且无拖欠税款记录	纳税历史三年及以上、且无拖欠税款记录	数据缺失	纳税历史三年及以上、且无拖欠税款记录	—	—		
...			...	...	...	...	...	...	...		
81	B7 抵押担保因素	$X_{81}$ 抵押/质押/担保/信用	定性	贷款金额 5 000 000 元, 商品房(非住宅) 产权现房抵押 5 000 000 元	无抵押	贷款金额 12 000 000 元, 其他企业保证 12 000 000 元	无抵押	—	—		
82		借据违约与否的标识 $y_j$	—	0	0	1	1	—	—		

多个非工业行业样本合并一起评级的原因: 上文 2.7 节信用等级划分中, 要求每个信用等级至少有一个违约样本, 本文划分 9 个信用等级, 则样本中必须有 9 个以上的违约样本. 从表 4 第 12 行可知, 工业行业的违约样本数及样本总数都较大, 适合单独债信用评级, 能够更有针对性地分析工业行业的债信用评级特性, 另有专文详述. 从表 4 前 11 行可知, 批发业等许多行业的违约样本小于 9 个, 零售业等行业甚至没有违约样本, 为了有足够的违约样本, 本文将批发业、零售业等 11 个非工业行业的样本数据合并, 共有 1 231 个小企业样本, 在这些数据的基础上为小企业债信用评级.

## 2) 数据来源

选取中国某区域性商业银行分布在京、津、沪、渝等分支行的 1 231 笔小企业贷款借据做为实证样本.

1 231 个小企业的原始数据, 如表 5 第 1 - 1 231 列所示. 其中, 包含 1 196 个非违约借据, 如表 5 第 1 - 1 196 列所示; 35 个违约借据, 如表 5 第 1 197 - 1 231 列所示. 每个企业对应的违约状态  $y_j$  如表 5 第 82 行所示, 0 代表非违约借据, 1 代表违约借据.

## 3.3 指标数据标准化

### 1) 定量指标数据的标准化

通用指标数据的确定: 定量指标包含正向、负向、区间型指标. 对表 5 的前 81 行的 58 个定量数据, 每一行的数据都确定出最大值和最小值, 列入其对应行的最后两列.

根据表 5 中第  $f$  列指标类型的标注, 将正向型指标数据代入式 (1)、负向型指标数据代入式 (2)、区间型指标数据代入式 (3), 得到标准化处理后 0 - 1 间的数值, 结果列入表 6 对应行.

### 2) 定性指标打分

经过 3.3 (1) 定量指标标准化后, 在表 5 第  $f$  列的剩余的“定性”的指标, 按照表 1 的定性指标打分标准进行打分, 打分后的值已在 0 - 1 之间, 不需要再通过正向、负向等打分公式转化到 0 - 1

之间, 将结果列入表 6 对应的行.

## 3.4 基于偏相关性分析的第一次指标筛选

在准则层内进行偏相关分析, 避免指标只在数据上相关, 在经济意义上不相关.

以“C1 偿债能力”准则层为例, 说明偏相关分析筛选指标的过程.

将表 6 前 20 行 20 个指标数据代入式 (4) - 式 (7), 计算偏相关系数, 结果如表 7 所示.

由表 7 可知, 共有 4 对指标的偏相关系数绝对值大于 0.7, 说明这 4 对指标高度相关, 反映信息冗余. 将这 4 对指标名称及对应偏相关系数列入表 8 第 2、4、6 列.

每对冗余指标中需要删除一个, 计算其  $F$  值. 以“ $X_1$  资产负债率”、“ $X_9$  股东权益比率”为例: 由表 8 第 1 行第 6 列知这两个指标的偏相关系数绝对值为  $0.997 > 0.7$ , 二者反映信息重复, 需计算两指标的  $F$  值.

将表 6 中第 1 行、第 9 行数据代入式 (8), 可以得到指标  $X_1$ 、 $X_9$  的  $F$  值, 列入表 8 第 1 行第 3 列与第 5 列, 因  $F_1 = 0.0297 < F_9 = 0.0299$ , 根据删除  $F$  值较小的指标的标准, 删除对违约状态影响不显著的指标“ $X_1$  资产负债率”, 如第 7 列所示.

同理删除其他三对重复指标中  $F$  值较小的那一个指标, 亦如第 7 列所示.

表 8 中的“C1 偿债能力”准则层前后共删除了 4 个反映信息重复的指标, 剩余 16 个指标.

其他准则层同理, 通过偏相关分析 -  $F$  值判别, 共剔除了“ $X_1$  资产负债率”、“ $X_{54}$  恩格尔系数”等 12 个反映信息冗余的指标. 所有剔除的指标在表 3 第 8 列用“偏相关分析删除”标注.

经过第一次偏相关分析筛选, 81 个指标中剔除 12 个指标, 共剩余 69 个指标.

经过第一次偏相关分析筛选, 找到反映信息重复的指标, 并通过  $F$  值进行判别, 剔除  $F$  值较小的、即对违约状态影响小的指标, 避免人为误删对违约状态影响大的指标.

表6 小企业信用指标筛选标准化数据  
Table 6 The data a fer standardized of indicator of credit risk evaluation for small enterprises

(a) 序号	(b) 一级 准则	(c) 二级 准则	(d) 三级 准则	(e) 指标名称	(f) 指标 类型	评价指标标准化后数据 $x_{ij}$					
						非违约客户贷款借据号			违约客户贷款借据号		
						(1) 200410270004	(1 196) X2012060800099	(1 197) 200410090044	(1 231) X2010121600004	...	...
1	A1 还款 能力	B1 企业 内部财 务因素	C1 偿债 能力	$X_1$ 资产负债率	负向	0.327	...	0.419	0.328	...	0.804
2				$X_2$ 流动负债经营活动净现金流比率	正向	0.517	...	0.480	0.622	...	0.478
...				...	...	...	...	...	...		
20				$X_{20}$ EBITDA 与总负债比率	正向	0.007	...	0.001	0.005	...	0.000
21			C2 盈利 能力	$X_{21}$ 净资产收益率	正向	0.077	...	0.003	0.053	...	0.000
...				...	...	...	...	...	...		
33				$X_{33}$ 经营活动产生现金流小计	正向	0.005	...	0.028	0.003	...	0.000
34				C3 营运 能力	$X_{34}$ 应收账款周转速度	正向	0.015	...	0.039	0.006	...
...		...	...		...	...	...	...			
43		$X_{43}$ 现金循环周期	正向		0.000	...	0.505	0.502	...	0.503	
44		C4 成长 能力	$X_{44}$ 营业收入增长率		正向	0.197	...	0.201	0.197	...	0.197
...			...	...	...	...	...	...			
48			$X_{48}$ 留存收益增长率	正向	0.510	...	0.518	0.521	...	0.549	
49			B2 企业外部 宏观条件	$X_{49}$ 行业景气指数	正向	0.722	...	0.742	0.626	...	0.648
...		...		...	...	...	...	...			
52		$X_{52}$ 居民消费价格指数		区间	0.976	...	0.976	0.976	...	0.951	
...	...	...		...	...	...	...	...			
54	B3 企业内部非 财务因素	$X_{54}$ 恩格尔系数	负向	0.651	...	0.755	0.651	...	0.716		
55		$X_{55}$ 相关行业从业年限	定性	1.000	...	1.000	1.000	...	0.400		
...		...	...	...	...	...	...	...			
63	A2 还款 意愿	B4 企业负责人 基本情况		$X_{63}$ 企业通过本银行回笼贷款总额占比	正向	0.667	...	1.000	0.000	...	0.000
64				$X_{64}$ 学历	定性	0.900	...	1.000	0.000	...	1.000
...				...	...	...	...	...	...	...	
71				$X_{71}$ 年龄	区间	0.970	...	0.848	0.000	...	0.939
...	B5 企业基本 信用情况			...	...	...	...	...	...	...	
74				$X_{74}$ 担任该职务时间	定性	0.400	...	0.400	0.000	...	0.400
75				$X_{75}$ 企业到位注册资金类别	定性	1.000	...	1.000	0.000	...	1.000
76				$X_{76}$ 近三年企业授信情况	定性	1.000	...	1.000	0.000	...	1.000
77	B6 企业的 商业信誉			$X_{77}$ 企业纳税记录	定性	1.000	...	1.000	0.000	...	1.000
...				...	...	...	...	...	...		
80				$X_{80}$ 企业间合同违约次数	定性	1.000	...	1.000	0.000	...	1.000
81	B7 抵押担保因素			$X_{81}$ 抵押/质押/担保/信用	定性	0.650	...	0.000	0.100	...	0.000
82	借据违约与否的标识 $y_j$				—	0	...	0	1	...	1

3.5 基于 Probit 回归的第二次指标筛选

在偏相关分析的基础上,通过 Probit 回归筛选各准则层中剩余的指标,遴选对违约状态有显著鉴别能力的指标。

以“C1 偿债能力”准则层为例,说明 Probit 回归删除指标的过程。

1) 第 1 个指标的剔除过程. 将“C1 偿债能力”准则层经第一次偏相关分析后,剩余的 16 个指标在表 6 对应行中的数据,及表 6 最后一行的违约状态代入式(9)~式(13),Stata 软件可以自动完成 Probit 回归过程,得到 Probit 回归参数,如表 9 第 3 列~第 6 列所示。

表 7 “偿债能力”准则层内指标的偏相关系数

Table 7 The partial correlation coefficient of indicators in “solvency” criterion layer

指标名称	(1) $X_1$ 资产 负债率	(2) $X_2$ 流动负债经营 活动净现金流比率	(3) $X_3$ 速动 比率	(6) $X_6$ 息税前利润 流动负债比	(12) $X_{12}$ 净资产与年 末贷款余额比率	(20) $X_{20}$ EBITDA 与 总负债比率
(1) $X_1$ 资产负债率	-1.000	—	—	—	—	—
(2) $X_2$ 流动负债经营活动净现金流比率	-0.011	-1.000	—	—	—	—
(3) $X_3$ 速动比率	-0.006	-0.072	-1.000	—	—	—
...	...	...	...	...	...	...
(9) $X_9$ 股东权益比率	0.997	0.015	0.007	-0.024	—	—
...	...	...	...	...	...	...
(11) $X_{11}$ 净利润现金含量	0.001	0.007	-0.008	0.036	—	—
(12) $X_{12}$ 净资产与年末贷款余额比率	-0.069	-0.110	0.114	0.163	-1.000	—
(13) $X_{13}$ 资本固定化比率	-0.031	-0.025	-0.054	0.091	-0.056	—
(14) $X_{14}$ 现金比率	0.003	0.077	0.716	-0.049	-0.107	—
(15) $X_{15}$ 长期资产适宜率	0.001	0.023	0.064	0.003	-0.016	—
...	...	...	...	...	...	...
(19) $X_{19}$ 总负债经营活动净现金流比率	0.007	0.934	0.084	-0.327	0.056	—
(20) $X_{20}$ EBITDA 与总负债比率	-0.006	-0.278	-0.206	0.727	-0.121	-1.000

表 8 偏相关分析删除的指标

Table 8 The deleted indicators by partial correlation coefficient

(1) 序号	偏相关系数绝对值大于 0.8 的指标对				(6) 偏相 关系数 $r_{ij}$	(7) 偏相关分析删除的指标
	(2) 评价指标 1	(3) 指标 1 的 F 值	(4) 评价指标 2	(5) 指标 2 的 F 值		
1	$X_1$ 资产负债率	0.029 7	$X_9$ 股东权益比率	0.029 9	0.997	$X_1$ 资产负债率
2	$X_2$ 流动负债经营活动净现金流比率	0.029 5	$X_{19}$ 总负债经营活动净现金流比率	0.012 0	0.934	$X_{19}$ 总负债经营活动净现金流比率
3	$X_3$ 速动比率	0.167 9	$X_{14}$ 现金比率	0.000 0	0.716	$X_{14}$ 现金比率
4	$X_6$ 息税前利润流动负债比	0.047 8	$X_{20}$ EBITDA 与总负债比率	0.034 4	0.727	$X_{20}$ EBITDA 与总负债比率

根据上文 2.4.2(1) 中“在显著性概率 Sig > 0.01 的指标中,剔除 Sig 值最大的一个指标”的标准,在表 9 第 6 列所有 Sig > 0.01 的指标中,“ $X_2$  流动负债经营活动净现金流比率”的显著性概率 Sig 值最大,表明该指标对违约状态影响最不显著.第一轮将其剔除,则“C1 偿债能力”准则层 16 个指标中还剩余 15 个指标.

2) 其余指标的删除过程

剔除“ $X_2$  流动负债经营活动净现金流比率”后,“C1 偿债能力”准则层中剩余 15 个指标按照上述方法进行筛选,直至剩余所有指标的显著性概率 Sig 均小于 0.01,即可停止,指标筛选过程结果如表 10 第 1 列~第 16 列所示.

“C1 偿债能力”准则层最终剩下“速动比率”、“超速动比率”两个指标的 Sig 值均小于 0.01,如表 10 第 15 列所示.

其余准则层同理,采用 Probit 回归筛选对违约状态影响显著的指标.

第一次偏相关筛选后剩余的 69 个指标,通过 Probit 回归分析进行第二次筛选,剔除了“ $X_2$  流动负债经营活动净现金流比率”、“ $X_{21}$  净资产收益率”等 46 个对违约状态影响不显著的指标,剔除的指标在表 3 第 8 列用“Probit 删除”标注.最终保留了“速动比率”、“应收账款周转速度”等 23 个指标,在表 3 第 8 列标注“保留”.由这 23 个指标构建的小企业债信用评级指标体系,如表 12 第 b 列~第 e 列所示.

第二次指标筛选通过求解违约状态变量与评价指标之间 Probit 回归方程的回归系数  $\beta$  和回归系数的标准误差  $SE_{\beta}$ ,构建 Wald 统计量对回归系数  $\beta$  的显著性进行检验,剔除对小企业违约状态影响小的、回归系数  $\beta$  不显著的指标,保证了筛选出的指标能显著区分企业的违约状态.

表9 Probit 回归的指标参数

Table 9 The parameters of indicators in Probit model

(1) 序号	(2) 指标名称	(3) 回归系数 $\beta$	(4) 标准误差 $SE_{\beta}$	(5) Wald 值	(6) Sig 值(双尾)
1	$X_2$ 流动负债经营活动净现金流比率	-0.183	1.771	0.01	0.918
2	$X_3$ 速动比率	1.343	0.384	12.23	0.000
3	$X_4$ 流动比率	-0.784	0.713	1.21	0.272
4	$X_5$ 主营业务收入现金比率	-0.720	0.833	0.75	0.388
5	$X_6$ 息税前利润流动负债比	0.279	1.693	0.03	0.869
6	$X_7$ 全部资本化率	-1.306	0.863	2.29	0.130
7	$X_8$ 全部资产现金回收率	0.423	1.038	0.17	0.684
8	$X_9$ 股东权益比率	0.882	0.510	2.99	0.084
9	$X_{10}$ 超速动比率	-3.987	1.232	10.48	0.001
10	$X_{11}$ 净利润现金含量	-0.349	0.617	0.32	0.572
11	$X_{12}$ 净资产与年末贷款余额比率	-0.423	0.502	0.71	0.399
12	$X_{13}$ 资本固定化比率	0.308	0.589	0.27	0.601
13	$X_{15}$ 长期资产适宜率	-3.374	2.878	1.37	0.241
14	$X_{16}$ 未偿还贷款总额占净资产比	0.832	0.646	1.66	0.198
15	$X_{17}$ 未偿还贷款总额占资产总额比	-0.292	1.012	0.08	0.773
16	$X_{18}$ 非流动负债经营活动净现金流比率	-1.876	1.961	0.91	0.339

表10 Probit 回归剔除指标的过程

Table 10 The delete process of indicators using Probit model

序号	指标名称	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
		第1次删除	第2次删除	第3次删除	第4次删除	第5次删除	第6次删除	第7次删除	第8次删除	第9次删除	第10次删除	第11次删除	第12次删除	第13次删除	第14次删除	第15次删除
1	$X_2$ 流动负债经营活动净现金流比率	<b>0.918</b>	—	—	—	—	—	—	—	—	—	—	—	—	—	—
2	$X_3$ 速动比率	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	$X_4$ 流动比率	0.272	0.261	0.239	0.236	0.263	0.297	0.212	0.179	0.102	0.076	<b>0.109</b>	—	—	—	—
4	$X_5$ 主营业务收入现金比率	0.388	0.391	0.389	0.376	0.373	0.320	0.324	0.165	0.026	0.013	0.021	0.018	0.026	<b>0.032</b>	—
5	$X_6$ 息税前利润流动负债比	0.869	<b>0.898</b>	—	—	—	—	—	—	—	—	—	—	—	—	—
6	$X_7$ 全部资本化率	0.130	0.130	0.133	0.073	0.081	0.072	0.057	0.051	0.047	0.013	0.028	0.097	<b>0.235</b>	—	—
7	$X_8$ 全部资产现金回收率	0.684	0.588	0.591	<b>0.621</b>	—	—	—	—	—	—	—	—	—	—	—
8	$X_9$ 股东权益比率	0.084	0.084	0.084	0.072	0.079	0.080	0.058	0.086	0.077	0.100	0.071	<b>0.180</b>	—	—	—
9	$X_{10}$ 超速动比率	0.001	0.001	0.001	0.001	0.001	0.000	0.000	0.001	0.000	0.001	0.000	0.000	0.000	0.000	0.000
10	$X_{11}$ 净利润现金含量	0.572	0.555	0.565	0.540	<b>0.656</b>	—	—	—	—	—	—	—	—	—	—
11	$X_{12}$ 净资产与年末贷款余额比率	0.399	0.403	0.402	0.414	0.412	0.414	<b>0.394</b>	—	—	—	—	—	—	—	—
12	$X_{13}$ 资本固定化比率	0.601	0.594	0.591	0.598	0.636	<b>0.696</b>	—	—	—	—	—	—	—	—	—
13	$X_{15}$ 长期资产适宜率	0.241	0.241	0.239	0.240	0.238	0.241	0.246	<b>0.254</b>	—	—	—	—	—	—	—
14	$X_{16}$ 未偿还贷款总额占净资产比	0.198	0.198	0.199	0.186	0.189	0.191	0.168	0.153	0.166	<b>0.262</b>	—	—	—	—	—
15	$X_{17}$ 未偿还贷款总额占资产总额比	0.773	0.781	<b>0.806</b>	—	—	—	—	—	—	—	—	—	—	—	—
16	$X_{18}$ 非流动负债经营活动净现金流比率	0.339	0.340	0.343	0.231	0.260	0.241	0.257	0.249	<b>0.340</b>	—	—	—	—	—	—

### 3.6 基于 ROC 曲线检验指标体系的有效性

将上文 3.5 节中筛选出的 23 个信用评级指标在表 6 对应行中的数据,代入式(9)~式(12),通过 Probit 回归模型,求解出式(10)中每个企业的违约概率  $P(y_j = 1 | X_j)$ 。当  $P(y_j = 1 | X_j) \geq 0.5$  时,判为违约;当  $P(y_j = 1 | X_j) < 0.5$ ,判为非违约。相应得到违约分类结果如表 11 所示。该过程可由 Stata 软件得到结果。

表 11 中: Probit 回归模型将实际违约的样本判定为违约的个数为 21,如表 11 第 1 行第 1 列所示。将实际非违约的样本判定为非违约的个数为 1 193,如表 11 第 2 行第 2 列所示。则判断准确率 =  $(21 + 1\ 193) / 1\ 231 = 98.62\%$ 。从分类结果可知,筛选出的 23 个指标对违约状态的鉴定准确性较高。

表 11 23 个指标通过 Probit 回归对违约状态的分类结果

Table 11 Classification using Probit regression by 23 indexes

实际违约状态	Probit 回归分类结果		
	(1) 违约	(2) 非违约	(3) 合计
(1) 违约	21	14	35
(2) 非违约	3	1 193	1 196
(3) 合计	24	1 207	1 231

ROC 曲线是由多个临界值  $C_0$  对应的灵敏度与特异度构成的。当样本的违约概率大于临界值  $C_0$  时,即  $P(y_j = 1 | X_j) \geq C_0$  时,Probit 模型判定该样本为违约,反之则判为非违约。

表 11 是以临界值为 0.5 时的分类结果,将表 11 第 1 行数据代入式(14),可得,灵敏度 =  $21/35 = 0.600$ ;将第 2 行数据代入式(15),可得,特异度 =  $1\ 193/1\ 196 = 0.997$ ;对应得到 ROC 曲线上的一个点  $(1 - \text{特异度}, \text{灵敏度}) = (1 - 0.997, 0.600)$ ,即  $(0.003, 0.600)$ 。

取不同的临界值,可得到多组灵敏度、特异度,对应可得到多个 ROC 曲线上的点。以  $1 - \text{特异度}$  为横轴、以灵敏度为纵轴画出 ROC 曲线,如图 3 的粗实线所示。可由 Stata 软件得到 ROC 曲线图。

图 3 中,粗实线 ROC 曲线下方的面积,即反映指标体系鉴别能力的  $AUC = 0.987\ 6 > 0.9$ 。所以,由筛选出的 23 个指标组成的债信用评级指标体系对违约状态有较好的区分效果。

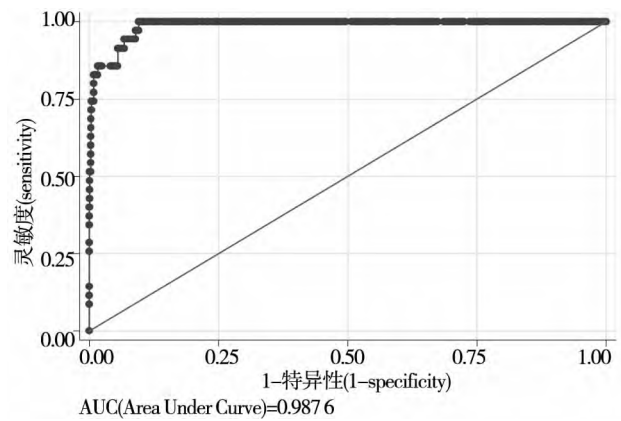


图 3 ROC 曲线图

Fig. 3 ROC curve

### 3.7 小企业债信用评级指标权重的确定与评价方程的建立

#### 1) 债信用评级指标权重的确定

将最终筛选出的 23 个指标在表 6 对应数据  $x_{ij}$  列入表 12 第 1 列 - 第 1 231 列中,并将数据代入式(16) - 式(17),得到变异系数为指标赋权结果  $w_i$ ,如表 12 第 f 列所示。将每个准则层的指标权重相加,为该准则层的权重,如表 12 第 b 列 ~ 第 d 列所示。

其中,表 12 第 1 行 - 第 8 行财务指标的权重之和为  $0.056 + \dots + 0.027 = 0.44$ ,表 12 第 9 行 - 第 23 行非财务指标的权重之和为  $0.008 + \dots + 0.024 = 0.56$ ,即财务指标的重要程度为 44%,非财务指标的重要程度为 56%。

可见在小企业债信用评级指标体系中,非财务指标的重要程度大于财务指标,这与传统信用评级指标体系中常常注重财务指标有所不同。原因是小企业财务指标更加难获取,而且小企业的发展受外界环境的影响与冲击更大,所以非财务指标相对于财务指标更加重要。

#### 2) 评价方程的建立

将最终筛出的 23 个指标在表 12 第 1 行 ~ 第 23 行第 f 列的指标权重代入式(18),得到小企业的信用评价方程

$$z_j = 0.065x_{1j} + \dots + 0.027x_{46j} + \dots + 0.024x_{81j} \quad (20)$$

将表 12 第 1 列 ~ 第 1 231 列数据,代入评价方程(20),得到 0 至 1 之间的评价得分  $z_j$ ,结果如表 13 第 3 列所示。再将表 13 第 3 列评价得分  $z_j$  代入式(19),化为 0 到 100 之间的信用得分  $S_j$ ,并按照信用得分降序排列,结果如表 13 第 4 列所示。



表 12 小企业信用评价指标指标权重及标准化后数据

Table 12 The weight of information of indicators and the data after standardized of credit rating for small enterprises

(a) 序号	(b) 一级准则	(c) 二级准则	(d) 三级准则	(e) 指标	(f) 权重 $w_j$	标准化后数据 $x_{ij}$		5C 原则对应关系					
						(1) 200410270004	(1 231) X2010121600004	(g) 品质 (character)	(h) 能力 (capacity)	(i) 资本 (capital)	(j) 担保 (collateral)	(k) 经营环境 (condition)	
1	A1 还款能力	B1 企业内部财务因素	C1 偿债能力 0.105	$X_3$ 速动比率	0.056	0.000	...	1.000			√		
2				$X_{10}$ 超速动比率	0.049	0.070	...	0.013			√		
3			C2 盈利能力 0.173	$X_{24}$ 总资产报酬率	0.061	0.038	...	0.000		√			
...				...	...	...	...	...	...	...	...	...	...
6				$X_{32}$ 经营活动产生的现金流量净额	0.007	0.552	...	0.498		√			
7				C3 营运能力 0.135	$X_{34}$ 应收账款周转速度	0.135	0.015	...	0.000			√	
8		C4 成长能力 0.027	$X_{40}$ 总资产增长率	0.027	0.271	...	0.283		√				
9		B2 企业外部宏观条件 0.025	$X_{49}$ 行业景气指数	0.008	0.722	...	0.648					√	
...			...	...	...	...	...	...	...	...	...	...	
11			$X_{55}$ 城市居民人均可支配收入	0.014	0.182	...	0.481					√	
12		B3 企业内部非财务因素 0.226	$X_{55}$ 相关行业从业年限	0.024	1.000	...	0.400		√				
...			...	...	...	...	...	...	...	...	...	...	
14			$X_{61}$ 产品销售范围	0.035	0.500	...	0.000		√				
15		A2 还款意愿	B4 企业负责人基本情况 0.233	$X_{65}$ 法人代表贷款违约记录	0.027	1.000	...	0.700	√				
...	...			...	...	...	...	...	...	...	...	...	
20	$X_{74}$ 担任该职务时间		0.042	0.400	...	0.400		√					
21	B5 企业基本信用情况 0.025	$X_{76}$ 近三年企业授信情况	0.025	1.000	...	1.000	√						
22	B6 企业的商业信誉 0.027	$X_{80}$ 企业间合同违约次数	0.027	1.000	...	1.000	√						
23	B7 抵押担保因素 0.024	$X_{81}$ 抵押/质押/担保/信用	0.024	0.650	...	0.000					√		

注: 第 1 列 ~ 第 1 231 列数据来自表 6.

### 3.8 小企业的信用等级划分

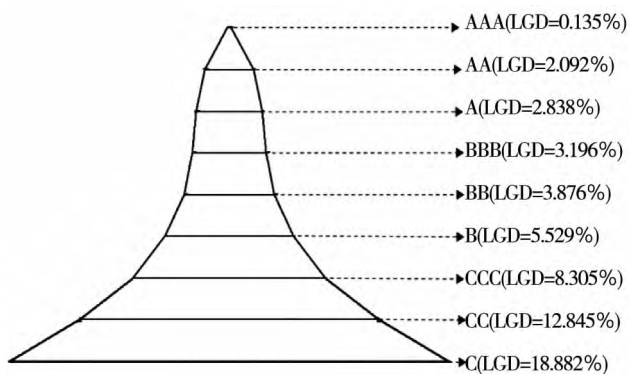


图 4 1 231 个小企业违约损失率分布

Fig. 4 The LGD distribution of 1 231 small enterprises

根据上文 2.7 中的信用等级划分的思路,通

过课题组的研究成果<sup>[26]</sup>,对 1 231 个小企业划分信用等级,结果如表 13 第 5 列 ~ 第 8 列所示.

信用得分区间的确定过程,以 AAA 级为例说明.由表 13 第 5 列可知,AAA 级共 202 个样本,表 13 前 202 行为 AAA 等级;AAA 级第 1 个样本得分为 100 分,最后一个样本为 55.66 分,则 AAA 等级的信用得分区间为  $55.66 \leq S \leq 100$ .同理,可以得到 AA、A、...、C 其余 8 个等级的信用得分区间,结果如表 13 第 8 列所示.

为了更直接的检验各等级的违约损失率 LGD 是否呈现金字塔的形状,对表 13 中第 5 列、第 7 列数据做如下处理:

以表 13 第 7 列违约损失率 LGD 大小为底边,

以表 13 第 5 列等级为高度作图 结果如图 4 所示. 图 4 显示: 越高的信用等级 其违约损失率越低. 表明该信用评价指标体系满足“信用等级越高、违约损失率越低”的原则.

表 13 1 231 个小企业信用等级划分结果

Table 13 The credit rank result of 1 231 small enterprises

(1) 序号	(2) 借据号	(3) 得分 $z_j$	(4) 信用得分 $S_j$	(5) 信用等级	(6) 样本数	(7) 违约损失率 $LDG$	(8) 信用得分区间
1	200902020002	0.62	100	AAA	202	0.135%	$55.66 \leq S \leq 100$
...	...	...	...				
202	X2010010500014	0.37	55.66				
...	...	...	...	...	...	...	...
843	X2010052500006	0.26	35.61	A	13	2.838%	$35.27 \leq S < 35.61$
...	...	...	...				
855	X2011080900034	0.25	35.27				
856	X2010082700003	0.25	35.25	BBB	93	3.196%	$30.77 \leq S < 35.27$
...	...	...	...				
948	E2011070600042	0.23	30.77				
...	...	...	...	...	...	...	...
1 138	E2011082400024	0.14	14.06	B	31	5.529%	$12.09 \leq S < 14.22$
...	...	...	...				
1 168	E2011052000002	0.12	12.09				
1 169	E2010092500037	0.12	11.84	CCC	10	8.305%	$11.37 \leq S < 12.09$
...	...	...	...				
1 178	200507120023	0.12	11.37				
...	...	...	...	...	...	...	...
1 195	X2011010600003	0.12	10.69	C	37	18.882%	$0 \leq S < 10.84$
...	...	...	...				
1 231	200411040001	0.06	0.00				

3.9 小企业债信用评级指标体系合理性分析

1) 小企业债信用评级指标体系与 5C 原则的对应

从表 12 第 g 列 ~ 第 k 列的对应关系可知, 该指标体系的 23 个指标涵盖了“5C 评价原则”的五个方面, 每个方面都有指标与之对应.

一是通过法人代表信用卡记录、企业间合同违约次数等 4 个指标, 反映了 5C 原则的品德 (character) 属性, 如表 12 第 g 列所示.

二是通过总资产报酬率、营业利润率等 12 个指标, 反映了 5C 原则的能力 (capacity) 属性, 如表 12 第 h 列所示.

三是通过速动比率、超速动比率等 3 个指标, 反映了 5C 原则的资本 (capital) 属性, 如表 12 第 i 列所示.

四是通过指标抵质押担保因素, 反映了 5C 原则的担保 (collateral) 属性, 如表 12 第 j 列所示.

五是通过行业景气指数、居民消费价格指数

等 3 个指标, 反映了 5C 原则的经营环境 (condition) 属性, 如表 12 第 k 列所示.

2) 债信用评级指标体系反映小企业的贷款特点

一是反映小企业贷款容易受到外部环境的影响. 最终构建的债信用评级指标体系的 23 个指标中, “行业景气指数”、“居民消费价格指数”、“城市居民人均可支配收入”三个指标反映了外部宏观条件对小企业的还款的影响, 如表 12 第 9 行 ~ 第 11 行所示. 因为小企业规模小, 容易受到市场波动的影响, 当行业不景气时, 小企业的还款风险也会增大.

二是反映小企业贷款受到非财务因素的影响较大. 由上文 3.7(1) 中可知, 对小企业债信用评级时, 非财务因素的权重为 56%, 可见非财务因素更能反映小企业的信用资质. 小企业财务信息不完善, 且财务信息中容易有虚假信息, 所以小企业进行债信用评级时, 非财务指标所占权重比例更高.

## 4 结束语

### 4.1 主要结论

1) 通过对中国某区域性商业银行分布在京、津、沪、渝等分支行的 1 231 笔小企业贷款借据进行实证,结果表明:速动比率、净资产收益率、行业景气指数、年龄、抵质押担保等 23 个指标能够显著区分小企业的违约状态。

2) 实证结果表明,本文构建的小企业的债信评级指标体系符合“信用等级越高、违约损失率越低”的原则,且满足 5C 原则。

### 4.2 主要创新与特色

1) 通过偏相关分析,在高度相关的指标中,删除  $F$  值小、对小企业违约状态判别能力弱的指标,既保证了筛选的指标反映信息不重复,也避免

了现有删除重复指标时、对违约状态影响大的指标可能被误删的弊端。且采用偏相关分析弥补了简单相关分析仅仅考虑两个变量间的相关程度,忽略了其余变量影响的不足。

2) 通过求解违约状态变量与评价指标之间 Probit 回归方程的回归系数  $\beta$  和回归系数的标准误差  $SE_{\beta}$ , 构建 Wald 统计量对回归系数  $\beta$  的显著性进行检验,剔除对小企业违约状态影响小的、 $\beta$  不显著的指标,保证了筛选出的指标能显著区分企业的违约状态。

3) 通过构造小企业债信评级指标体系的感受型曲线 ROC,测算出 ROC 曲线下方的面积 AUC 值,用于检测整个指标体系区分违约状态的有效性。弥补了单个指标对违约状态影响显著,而构建的整个指标体系可能对违约状态鉴别能力不强的弊端。

### 参考文献:

- [1] 中华人民共和国工商行政管理总局. 国家工商总局发布小微企业发展情况报告 [EB/OL]. 中华人民共和国工商行政管理总局网站. [http://www.saic.gov.cn/ywdt/gsyw/zjyw/xxb/201403/t20140331\\_143496.html](http://www.saic.gov.cn/ywdt/gsyw/zjyw/xxb/201403/t20140331_143496.html). 2014-03-31.  
State Administration for Industry & Commerce of the People's Republic of China. The small enterprises development report [EB/OL]. State Administration for Industry & Commerce of the People's Republic of China Website. [http://www.saic.gov.cn/ywdt/gsyw/zjyw/xxb/201403/t20140331\\_143496.html](http://www.saic.gov.cn/ywdt/gsyw/zjyw/xxb/201403/t20140331_143496.html). 2014-03-31. (in Chinese)
- [2] 中国人民银行调查统计司. 2014 年第 1 季度银行家调查问卷 [EB/OL]. 中国人民银行网站. [http://www.pbc.gov.cn/publish/diaochatongjisi/193/2014/20140321155104356260357/20140321155104356260357\\_.html](http://www.pbc.gov.cn/publish/diaochatongjisi/193/2014/20140321155104356260357/20140321155104356260357_.html). 2014-04-25.  
Statistics and analysis department of the People's Bank of China. The questionnaire report from bankers in the first quarter of 2014 [EB/OL]. [http://www.pbc.gov.cn/publish/diaochatongjisi/193/2014/20140321155104356260357/20140321155104356260357\\_.html](http://www.pbc.gov.cn/publish/diaochatongjisi/193/2014/20140321155104356260357/20140321155104356260357_.html). 2014-04-25. (in Chinese)
- [3] Ileana Nicula. Some aspects concerning the measurement of credit risk [J]. *Procedia Economics and Finance*, 2013, 6: 668-674.
- [4] 吴青. 《巴塞尔协议 II》、内部信用评级及小企业贷款 [J]. *国际金融研究*, 2007, (5): 44-53.  
Wu Qing. The Basel II, Internal credit rating and small-business loans [J]. *Studies of International Finance*, 2007, (5): 44-53. (in Chinese)
- [5] Standard & Poor's. *Standard & Poor's Guide to Credit Rating Essentials* [R]. New York: Standard & Poor's, 2011.
- [6] Fitch Ratings. *Fitch Ratings Global Corporate Finance 2012 Transition and Default Study* [R]. New York: Credit Market Research-Fitch Ratings, 2013.
- [7] 大公资信评估有限公司. 大公信用评级方法及流程 [EB/OL]. 大公资信评估有限公司网站, <http://www.dagongcredit.com/ratingSystem/contents1.html>. 2012-09-24.  
Dagong Global Credit Rating Co. Credit rating method and process [EB/OL]. Dagong Global Credit Rating Co. website. <http://www.dagongcredit.com/ratingSystem/contents1.html>. 2012-09-24. (in Chinese)
- [8] 中国银行. 中国银行股份有限公司客户信用评级指标体系与评分标准说明 [R]. 中国银行有限公司, 2010.  
Bank of China. Credit Rating Index System and the Definition of Scoring Standard for Custom in Bank of China [R]. Bank of China Ltd. 2010. (in Chinese)
- [9] 中国建设银行. 中国建设银行小企业客户评价办法 [R]. 北京: 中国建设银行, 2007.

- China Construction Bank. Enterprise Customer's Credit Rating Rules of CCB [R]. Beijing: China Construction Bank, 2007. (in Chinese)
- [10] Li Z, Chi G. Factors Study of Credit Risks of Farmer Loans Based on Projection Pursuit [C]// Dongguan: Service Operations and Logistics, and Informatics (SOLI) 2013 IEEE International Conference on, 2013: 274–277.
- [11] Psillaki M, Tsolas I E, Margaritis D. Evaluation of credit risk based on firm performance [J]. European Journal of Operational Research, 2010, 201(3): 873–881.
- [12] Shi B F, Chi G T. A credit risk evaluation index screening model of petty loans for small private business and its application [J]. Advances in Information Sciences and Service Sciences, 2013, 5(7): 1116–1124.
- [13] 马晓青, 刘莉亚, 胡乃红, 等. 小企业信用评估的模型构建与实证分析 [J]. 财经研究, 2012, 38(5): 28–37.  
Ma Xiaqing, Liu Liya, Hu Naihong, et al. The credit rating model for small enterprise and empirical study [J]. Journal of Finance and Economics, 2012, 38(5): 28–37. (in Chinese)
- [14] Altman E I. Predicting Financial Distress of Companies: Revisiting the Z-score and ZETA Models [R]. New York: Stern School of Business, New York University, 2000.
- [15] Maria D C, Antonio B O, Rafael P M, et al. Improving the management of microfinance institutions by using credit scoring models based on statistical learning techniques [J]. Expert Systems with Applications, 2013, 40(17): 6910–6917.
- [16] 高丽君. 基于贝叶斯模型平均生存模型的中小企业信用风险估计 [J]. 中国管理科学, 2012, S1: 327–331.  
Gao Lijun. The estimation of credit risk of SMEs based on Bayesian model averaging survival model [J]. Chinese Journal of Management Science, 2012, S1: 327–331. (in Chinese)
- [17] Hong Sik Kim, So Young Sohn. Support vector machines for default prediction SMEs based on technology credit [J]. European Journal of Operational Research, 2010, 201(3): 838–846.
- [18] Niccolò Gordoni. A genetic algorithm approach for SMEs bankruptcy prediction: Empirical evidence from Italy [J]. Expert Systems with Applications, 2014, 41(14): 6433–6445.
- [19] Blanco A, Pino-mejías R, Lara J, et al. Credit scoring models for the microfinance industry using neural networks: Evidence from Peru [J]. Expert Systems with Applications, 2013, 40(1): 356–364.
- [20] 肖进, 刘敦虎, 顾新, 等. 银行客户信用评估动态分类器集成选择模型 [J]. 管理科学学报, 2015, 18(3): 114–126.  
Xiao Jin, Liu Dunhu, Gu Xin, et al. Dynamic classifier ensemble selection model for bank customer's credit scoring [J]. Journal of Management Sciences in China, 2015, 18(3): 114–126. (in Chinese)
- [21] 余高锋, 刘文奇, 石梦婷. 基于局部变权模型的企业质量信用评估 [J]. 管理科学学报, 2015, 18(2): 85–94.  
Yu Gaofeng, Liu Wenqi, Shi Mengting. Credit evaluation of enterprise quality based on local variable weight model [J]. Journal of Management Sciences in China, 2015, 18(2): 85–94. (in Chinese)
- [22] 朱钧钧, 谢识予, 许祥云. 基于空间 Probit 面板模型的债务危机预警方法 [J]. 数量经济技术经济研究, 2012, 10: 100–114.  
Zhu Junjun, Xie Shiyu, Xu Xiangyun. Debt crisis early warning method: Application of a spatial probit panel model [J]. The Journal of Quantitative & Technical Economics, 2012, 10: 100–114. (in Chinese)
- [23] Abdou H A, Pointon J. Credit scoring, statistical techniques and evaluation criteria: A review of the literature [J]. Intelligent Systems in Accounting, Finance and Management, 2011, 18(2–3): 59–88.
- [24] 刘新华, 线文. 我国中小企业融资理论述评 [J]. 经济学家, 2005, (2): 105–111.  
Liu Xinhua, Xian Wen. The reviews on Chinese SME financing theory [J]. Economist, 2005, (2): 105–111. (in Chinese)
- [25] 徐广军, 倪晓华, 肖运香. 标普、穆迪、邓白氏企业信用评价指标体系比较研究 [J]. 浙江金融, 2007, (3): 51–52.  
Xu Guangjun, Ni Xiaohua, Xiao Yunxiang. The comparative study among S&P, Moody's, Dun & Bradstreet enterprise credit evaluation index system [J]. Zhejiang Finance, 2007, (3): 51–52. (in Chinese)
- [26] 大连理工大学迟国泰课题组. 中国邮政储蓄银行农户小额贷款信用风险决策评价研究 [R]. 大连: 大连理工大学, 2011.  
Chi Guotai research group in DLUT. Credit Risks Evaluation and Decision System of Microcredit for Farmers for Postal Savings Bank of China [R]. Dalian: Dalian University of Technology, 2011. (in Chinese)
- [27] 程砚秋. 基于支持向量机的农户小额贷款决策评价研究 [D]. 大连: 大连理工大学, 2011.

- Cheng Yanqiu. Research on Evaluation and Decision System of Small Amount Loans for Farmers Based on Support Vector Machines [D]. Dalian: Dalian University of Technology, 2011. (in Chinese)
- [28] 威廉·H. 格林. 计量经济分析 [M]. 第六版. 北京: 中国人民大学出版社, 2011: 759 – 781.
- Greene W H. Econometric Analysis [M]. 6th Edition. Beijing: China Renmin University, 2011: 759 – 781. (in Chinese)
- [29] 张建国. ROC 曲线分析的基本原理以及在体质与健康促进研究中的应用 [J]. 体育科学, 2008, (6): 62 – 66.
- Zhang Jianguo. The basic principle of ROC curve analysis and application in physical fitness and health promotion studies [J]. China Sport Science, 2008, (6): 62 – 66. (in Chinese)
- [30] 工业和信息化部, 国家统计局, 国家发展和改革委员会, 财政部. 中小企业划型标准规定 [R]. 北京: 工业和信息化部, 国家统计局, 国家发展和改革委员会, 财政部, 2011.
- Ministry of Industry and Information Technology, National Bureau of Statistics, National Development and Reform Commission. Standard type division formiddle and small-sized enterprises [R]. Beijing: Ministry of Industry and Information Technology, 2011. (in Chinese)

## The debt rating for small enterprises based on Probit regression

CHI Guo-tai, ZHANG Ya-jing, SHI Bao-feng

Faculty of Management and Economics, Dalian University of Technology, Dalian 116024, China

**Abstract:** Debt rating of small enterprises means a system that evaluates the size of a small enterprise's credit level and LGD (Loss Given Default). It concerns the risk management of bank loans and whether small enterprises can get financing. Therefore, the indicators of credit rating must be able to identify the status of non-compliance of small enterprises. Existing credit rating system, aiming to rate scale and sort, does not evaluate the size of a debt LGD, and does not give evidence that its rating system is relevant to identify non-compliance. This article selects a index system according to the ability to identify the default status of enterprises, and establishes a credit rating system of small enterprises. This article has some innovations and features. Firstly, the paper deletes an index whose F-value is small and whose ability to distinguish the status of non-compliance of small enterprises is weak between one pair of highly related indicators which have partial correlation coefficients greater than 0.7 and reflect the duplication of information. This avoids information redundancy of the rating system after first section and at the same time avoids mistakenly deleting the index that has a big influence on the status of non-compliance. This paper improves the situation that existing credit rating system's selection is unrelated to the ability to identify non-compliance. Secondly, the paper builds the Wald statistic to test for the significant of the Probit regression coefficient,  $\beta$ , through solving the regression coefficient  $\beta$ , and the standard error,  $SE_{\beta}$  of the regression coefficient between the default state variables and evaluation, and deletes the index that has a small influence on the status of non-compliance and whose regression coefficient,  $\beta$ , is not significant. The index can distinguish significantly the status of non-compliance of enterprises after the second selection. Thirdly, the AUC area of the indicator system's experience curve built by this paper is greater than 0.9, and this guarantees the index system has a strong identification ability for the status of non-compliance. Fourthly, the results show the weight of the non-financial indicators, for small enterprises having a higher the credit rating and lower LGD, in the credit rating system is 56%. Finally, the empirical analysis for 1 231 small enterprises shows there are 23 indicators, involving the quick ratio, total asset growth, industry sentiment index, which can be used to distinguish the status of non-compliance of small enterprises.

**Key words:** small businesses; debt rating; data mining; Probit regression