

# 基于 LM 检验的小型工业企业债信评级模型及实证

迟国泰<sup>1</sup>, 于善丽<sup>1</sup>, 袁先智<sup>2</sup>

(1. 大连理工大学管理与经济学部, 辽宁 大连 116024; 2. 上海同济大学风险管理研究所, 上海 200092)

**摘要:** 债信评级是通过建立评级指标体系和评价方程来评价客户一笔债务信用等级的高低及债务违约损失的概率。它是信用风险管理极为重要的环节。小型工业企业财务信息更难获取, 数据的真实性更难考证, 导致小型工业企业债信评级更难。本文基于 LM 检验建立了小型工业企业债信评级模型。本文的创新与特色: 一是通过把每一个待遴选的指标、与前面已选定的  $j(j=0,1,2,\dots,m)$  个指标共同组成的  $j+1$  个指标组的所有数值、代入表示企业违约概率的拉格朗日函数, 得到  $j$  个拉格朗日统计量  $LM_j(j=0,1,2,\dots,m)$ ; 然后对这  $j$  个拉格朗日统计量  $LM_j$  进行  $\chi^2$  检验, 在所有通过  $\chi^2$  检验的统计量  $LM_j$  中、也即具有显著区分违约状态能力的指标中, 遴选  $LM_j$  值最大的指标为评级指标, 保证了遴选的指标对鉴别全部企业的违约状态具有显著影响。二是通过在相关系数大于 0.7 的一对指标中, 删除均方差数值小的指标, 既避免了指标体系的信息冗余、又避免了误删信息含量大的指标。三是对中国某商业银行 1814 笔小型工业企业贷款借据数据为样本的实证表明: 行业景气指数等宏观因素; 法人代表贷款违约记录等个人因素, 未偿还贷款总额占资产总额比等信用参数组成的 20 个指标评级体系, 能够显著区分小型工业企业的违约状态。

**关键词:** 工业企业; 小企业; 信用评级; 债信评级; 拉格朗日乘数检验

**中图分类号:** F830.56 **文献标识码:** A **文章编号:** 1004-6062(2019)01-0170-012

**DOI:** 10.13587/j.cnki.jieem.2019.01.017

## 0 引言

小型工业企业是指从业人员在[20, 300)人之间, 营业收入在[300, 2000)万元之内的工业企业<sup>[1]</sup>。由于小型工业企业数量众多, 分布广泛, 组织机构更加不健全, 财务信息更加不规范, 真实数据更难获取等特点, 使得对小型工业企业债信评级研究的难度增大, 导致小型工业企业债信评级的研究较少, 进而金融机构缺乏对小型工业企业贷款和融资的依据, 导致小型工业企业贷款难。

因此, 建立一套小型工业企业债信评级体系至关重要, 不仅可为银行进行小型工业企业贷款提供决策依据, 而且可帮助解决小型工业企业贷款难问题。

### (1) 评级指标体系的现状

美国 Dun and Bradstreet 信用评级公司主要从企业财务因素、员工数量、还款意愿等方面建立企业信用评级体系<sup>[2-3]</sup>。美国 Standard & Poor 信用评级公司主要从产品市场等企业经营状况方面和财务政策等企业内部财务因素等方面建立企业信用评级体系<sup>[3]</sup>。中国工商银行从股东、管理层、经营状况、信誉状况、宏观条件、发展前景和偿债能力等方面进行小企业信用评级<sup>[4]</sup>。中国建设银行从财务风险、账户行为、定性评级和特例调整事项等方面进行小企业信用评级<sup>[5]</sup>。中国银行从偿债能力、获利能力、经营管理、履约情况等方面进行信用评级<sup>[6]</sup>。阿里小贷通过淘宝和阿里平台等收集客户的历史交易记录、用户投诉等信息, 通过第三方

认证获取客户的海关、税务等数据, 通过阿里的心理测试系统分析企业主的性格、诚信度, 同时兼顾客户的婚姻状况、银行流水、资金利用率、利润率、及水电费等缴纳情况等元素进行信用评价<sup>[7-8]</sup>。Maik Dierkes 等建立了包括企业从业年限、企业前景、年产品销售量等指标的企业信用评级指标<sup>[9]</sup>。

### (2) 评级指标遴选方法的研究现状

迟国泰等运用偏相关方法删除信息冗余指标, 通过违约能力判别方法保留区分违约状态能力强的指标, 建立了信用评级体系<sup>[10]</sup>。邓超等利用贝叶斯界定折叠法, 解决了因样本有偏引起的小企业信用评分模型分类能力丧失问题, 建立了小企业信用评分模型<sup>[11]</sup>。范柏乃等通过隶属度分析保留对中小企业信用评级影响大的指标, 通过相关性分析删除评级体系内冗余指标, 通过鉴别力分析, 保留能鉴别不同中小企业差距的指标, 最终建立了中小企业信用评级体系<sup>[12]</sup>。Mu Zhang 等以 187 家上市公司为样本, 运用 Cox 模型遴选对企业信用风险影响显著的指标<sup>[13]</sup>。

### (3) 评级方法的研究现状

中国工商银行根据构建的打分标准或公式对相应指标进行打分, 运用线性回归模型对企业信用进行评级<sup>[4]</sup>。刘澄等运用 Vague 集理论, 用区间值同时测评定量与定性指标, 进行企业信用评级<sup>[14]</sup>。张目等根据 Topsis 方法思想, 对云重心评判法进行改进, 建立了基于改进的云重心评判法的企业信用评级模型<sup>[15]</sup>。

**收稿日期:** 2015-10-20 **修回日期:** 2017-10-31

**基金项目:** 国家自然科学基金资助重点项目(71731003); 国家自然科学基金资助项目(71171031、71471027); 国家社科基金资助项目(16BTJ017); 大连银行小企业信用风险评级系统与贷款定价项目(2012-01)

**作者简介:** 迟国泰(1955—), 男, 黑龙江海伦县人; 大连理工大学管理与经济学部教授, 博士生导师; 研究方向: 银行风险管理、商业银行评价、复杂系统评价。

Dimitrios Niklis 等运用支持向量机方法对非上市公司的信用得分进行求解<sup>[6]</sup>。Nebojsa Nikolic 等用蛮力逻辑回归模型对塞尔维亚企业求解信用评级得分<sup>[17]</sup>。Zarkic-Joksimovic 等基于 5500 笔秘鲁贷款机构的样本数据,运用神经网络模型进行客户信用评分<sup>[18]</sup>。

现有研究不足:一是现有评级体系<sup>[4-6,9-10]</sup>或是没有针对小型工业企业或者未以区分违约鉴别力为指标遴选标准,导致不能保证评级指标均能显著区分小型工业企业违约状态。二是现有评级体系<sup>[12]</sup>在两相关性大的指标对中,主观删除其中一个,没有根据指标的信息含量删除指标,不能较好的保留信息含量大的指标。三是现有评级体系<sup>[4,14-15]</sup>仅仅给出信用等级,却未给出每个等级对应的违约损失率,更没有直观反映年违约损失率与信用等级的匹配关系。

## 1 小型工业企业债信评级体系构建原理

### 1.1 小型工业企业债信评级的特点

(1) 信用信息、财务信息等不完善。由于小型工业企业管理不规范,信用档案缺失,缺乏规范健全的财务信息,且经营信息主要掌握在企业业主手中,信息不对称,给某些经典信用评级指标的数据获取带来困难,导致针对小型工业企业的债信评级体系建立难。

(2) 以抵押为主要贷款模式,信用贷款少。目前,小型工业企业主要通过抵质押担保方式进行贷款,而根据信用状况进行贷款的方式较少,贷款模式单一、贷款难度大,这主要是因为缺乏小型工业企业债信评级的合理体系。

(3) 依赖于地区的经济发展。由于小型工业企业大部分处在创业期,产品销售范围窄,市场占有率低,且为劳动密集型企业,这导致企业的发展在很大程度上依赖所在地区的经济发展。故不同地区的小型工业企业还款能力不一样,给针对小型工业企业债信的评级造成困难。

### 1.2 问题的难点

难点 1: 如何遴选能显著区分小型工业企业违约状态的债信评级指标。与企业违约状态鉴别能力无关的指标,不会组成合理的评级体系。

难点 2: 如何在相关性程度大的指标对中,保留信息含量大的指标。两个指标相关性大,说明指标反映信息重复,应该剔除其中一个,以保证评级体系的简洁有效。

### 1.3 突破难点的思路

#### 1.3.1 指标体系的遴选思路

通过把每一个待遴选的指标、与前面已选定的  $j(j=0,1,2,\dots,m)$  个指标共同组成的  $j+1$  个指标组的所有数值、代入表示企业违约概率的拉格朗日函数,得到  $j$  个拉格朗日统计量  $LM_j(j=0,1,2,\dots,m)$ ; 然后对这  $j$  个拉格朗日统计量  $LM_j$  进行  $\chi^2$  检验,在所有通过  $\chi^2$  检验的统计量  $LM_j$  中、也即具有显著区分违约状态能力的指标中,遴选  $LM_j$  值最大的指标为评级指标,保证了遴选的指标对鉴别全部企业的违约状态均具有显著影响。解决难点 1。

#### 1.3.2 信息冗余指标的剔除思路

通过在相关系数大于 0.7<sup>[9]</sup> 的一对信息冗余指标中,删除均方差数值较小、也即信息含量较少的指标,既避免了指标体系的信息冗余、又避免了误删信息含量大的指标。解决

难点 2。

基于拉格朗日乘数检验的小型工业企业债信评级模型构建原理如图 1 所示。

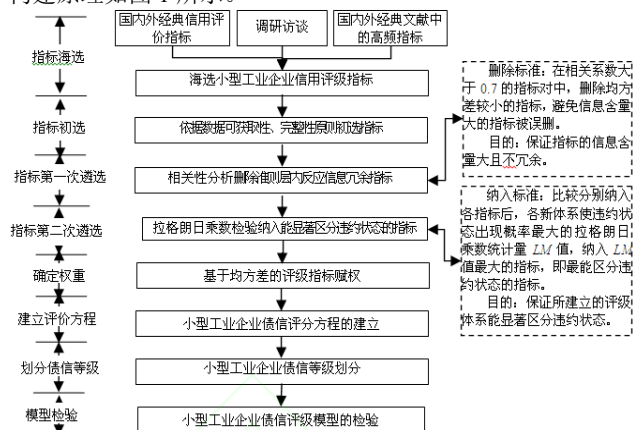


图 1 小型工业企业债信评级体系构建原理

## 2 小型工业企业债信评级方法

### 2.1 指标数据标准化方法

设:  $x_{ij}$ —第  $i$  个小型工业企业第  $j$  个指标的标准化值;  $v_{ij}$ —第  $i$  个小型工业企业第  $j$  个指标的原始数据;  $n$ —小型工业企业总数。

#### (1) 正向指标标准化

正向指标是数值越大,小型工业企业债信情况越好的指标,如“营业收入增长率”等指标。则  $x_{ij}$  为<sup>[10]</sup>:

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

#### (2) 负向指标标准化

负向指标是数值越小,小型工业企业债信情况越好的指标,如“资产负债率”等指标。则  $x_{ij}$  为<sup>[10]</sup>:

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

#### (3) 最佳区间型指标标准化

最佳区间型指标是指指标值在某一个特定区间内,小型工业企业债信情况较好的指标。如“年龄”等指标。

设:  $q_1$ —指标最佳区间左边界;  $q_2$ —指标最佳区间右边界。则  $x_{ij}$  为<sup>[10]</sup>:

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

式(3)的含义:若指标值  $v_{ij}$  在最佳区间  $[q_1, q_2]$  内,则该指标值最好,打 1 分;若指标值  $v_{ij}$  偏离最佳区间越远,说明这个指标越不好,标准化得分  $x_{ij}$  越低。

#### (4) 定性指标打分

通过对国内某些银行中熟悉债信评级的高层管理人员发放问卷、调查访谈等方式,对定性指标制定出适合小型工业企业债信评级的赋值标准,如表 1 所示。

表 1 定性指标打分标准

(1)准则层	(2)指标	(3)选项内容	(4)分值
企业的商业信誉	企业法律纠纷情况	无	1.00
		有, 且胜诉	0.50
		有, 且败诉	0.00
企业内部非财务因素	相关行业从业年限	...	...
		从业年限 $\geq 8$ 年	1.00
		5年 $\leq$ 从业年限 $< 8$ 年	0.70
		2年 $\leq$ 从业年限 $< 5$ 年	0.40
		0 $<$ 从业年限 $< 2$ 年, 或数据缺失	0.00
		...	...

## 2.2 基于相关性分析的第一次指标遴选

目的：删除同一准则层内反映信息冗余指标。若在不同的准则层内删除相关系数大的指标，则会误删数据相关，但经济意义无关的指标。

### (1) 相关系数的计算

设： $r_{kj}$ -第  $k$  个指标和第  $j$  个指标的相关系数； $x_{ik}$ -第  $i$  个小型工业企业的第  $k$  个指标值； $\bar{x}_k$ -第  $k$  个指标的平均值； $x_{ij}$ -第  $i$  个小型工业企业第  $j$  个指标值； $\bar{x}_j$ -第  $j$  个指标的平均值。则  $r_{kj}$ <sup>[19]</sup>为

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

式(4)的含义：相关系数越接近 1，说明两指标反应信息越重复，应该删除其中一个。

### (2) 相关性指标遴选标准

相关系数标准：相关系数绝对值大于 0.7<sup>[19]</sup>为显著相关，可以删除其中的一个指标。

信息含量标准：在两个相关的指标中，保留均方差值大，也即指标信息含量多的指标；删除另一个指标。这会避免信息含量大的指标被误删。

设： $\sigma_j$ -第  $j$  个指标的均方差； $x_{ij}$ -第  $i$  个小型工业企业的第  $j$  个指标值； $\bar{x}_j$ -第  $j$  个指标的平均值。

$$\sigma_j = \sqrt{\frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n-1}} \quad (5)$$

式(5)的含义：式(5)表示每个指标偏离各自均值的程度，偏离程度越大，该指标信息含量越多。

### (3) 相关性指标遴选特色

在相关系数大于 0.7 的一对指标中，删除均方差值较小、即信息含量较低的指标，既避免了指标体系的信息冗余、又避免了误删信息含量大的指标。

## 2.3 基于拉格朗日乘数检验的第二次指标遴选

目的：删除同一准则层内对小型工业企业违约状态影响不显著的指标，保证所建立的小型工业企业评级体系能显著区分违约状态。

### 2.3.1 第一个指标的遴选

每一次将一个指标代入式(6)，有  $m$  个指标，则代  $m$  次，如下文表 2 第 1 行第 2 至 4 列。

下面分别求这  $m$  个含一个指标的式(6)中的指标  $LM_j$  值。

Step1：构造假设，也即约束条件

$H_0$ ：第  $j$  个指标在式(8)的系数  $b_j=0$ ，即第  $j$  个指标对区

分违约状态没有显著影响；

$H_1$ ：第  $j$  个指标在式(8)的系数  $b_j \neq 0$ ，即第  $j$  个指标对区分违约状态有显著影响。

### Step2：构造 logistic 模型

$y_i$ -第  $i$  个小型工业企业的违约状态标识， $y_i=0$  表示非违约企业， $y_i=1$  表示违约企业； $x_{ij}$ -第  $i$  个小型工业企业第  $j$  个债信评级指标数值； $b_0$ -常数项； $b_j$ -第  $j$  个指标的回归系数； $p_i(y_i=1)$ -第  $i$  个小型工业企业违约的概率。则 logistic 模型<sup>[20]</sup>：

$$p_i = p_i(y_i = 1) = \frac{1}{1 + e^{-(b_0 + b_j x_{ij})}} \quad (6)$$

式(6)的含义：在不含指标的体系中纳入一个指标后，使第  $i$  个小型工业企业违约的概率。

由于每一个指标都分别代入式(6)，故此时有  $m$  个式(6)的表达式。

### Step3：构造对数似然函数

设： $\ln L(b_j)$ -关于参数  $b_j$  的对数似然函数。 $\ln L(b_j)$ 表达式<sup>[20]</sup>为：

$$\ln L(b_j) = \sum_{i=1}^n [y_i * \ln p_i + (1 - y_i) * \ln(1 - p_i)] \quad (7)$$

式(7)的含义：表示无约束条件下，所有小型工业企业发生违约的概率。

### Step4：构造拉格朗日函数

设： $\ln L^*(b_j)$ -关于参数  $b_j$  的拉格朗日函数，表示所有小型工业企业发生违约的概率； $\lambda_j$ -假设  $H_0$  对应的拉格朗日乘数<sup>[20]</sup>：

$$\ln L^*(b_j) = \ln L(b_j) + \lambda_j(b_j - 0) \quad (8)$$

式(8)的含义：表示在假设约束条件  $H_0: b_j = 0$  下，所有小型工业企业发生违约的概率。

### Step5：构造 LM 统计量

$LM_j$ -第  $j$  个指标下，所有小型工业企业发生违约的概率最大时，该指标的拉格朗日乘数检验统计量； $S(b_j)$ -表示式(8)的拉格朗日函数  $\ln L^*(b_j)$  在  $b_j=0$  处的一阶偏导值， $I(b_j)$ -表示式(8)的拉格朗日函数  $\ln L^*(b_j)$  在  $b_j=0$  处的二阶偏导值。 $LM_j$ 表达式<sup>[21]</sup>：

$$LM_j = [S(b_j)|_{b_j=0}]^2 * [-I(b_j)|_{b_j=0}] \quad (9)$$

式(9)的含义：表示在原假设约束条件下，在某个指标下，所有企业出现违约的概率最大时，该指标的拉格朗日乘数统计量值，用以在 Step6 中检验原假设是否成立。

将每一个指标的标准化数据  $x_{ij}$  代入式(6)，得到  $m$  个式(6)；将式(6)代入式(7)，得到  $m$  个式(7)；将式(7)代入式(8)，得到  $m$  个式(8)；然后对  $m$  个式(8)分别求在  $b_j=0$  处求一阶偏导值  $S(b_j)$  和二阶偏导值  $I(b_j)$ ，将  $S(b_j)$  和  $I(b_j)$  代入式(9)，得出  $m$  个指标的  $LM_j$  值，列入表 2 第 1 行第 5 列。

应该指出， $b_j=0$  时，单个指标下的所有企业违约概率最大。这是因为式(8)的一阶导为 0 时，才有  $b_j=0$ ；而式(8)的一阶导为 0 时，可以求出式(8)的违约概率极大值；因此， $b_j=0$  时，实际上得到的是在特定指标数值下的、式(8)的违约概率极大值；在这种情况下，特定指标下的所有企业违约概率最大。

### Step6：显著性检验

在  $H_0$  假设成立时，拉格朗日乘数检验统计量渐进服从自由度为约束条件个数  $n$  的  $\chi^2(n)$  分布<sup>[21]</sup>。因本文每次仅纳入一个指标，即约束条件  $n=1$ 。故拉格朗日乘数检验统计量服从  $\chi^2(n)=\chi^2(1)$  分布。

根据显著性水平  $\alpha$ ，查  $\chi^2$  分布统计表，得到拒绝域临界



值 $\chi^2_{\alpha}$ 。当 $LM_j > \chi^2_{\alpha}(1)$ 时, 拒绝假设 $H_0$ , 接受假设 $H_1$ , 即第 $j$ 个指标对区分违约状态有显著影响, 用 $\sqrt{\quad}$ 表示, 见表2第1行第6列。当 $LM_j < \chi^2_{\alpha}(1)$ 时, 接受假设 $H_0$ , 即第 $j$ 个指标对区分违约状态没有显著影响, 用 $\times$ 表示, 见表2第1行第6列。

#### Step7: 遴选指标标准

根据表2第1行第6列, 在 $LM_j > \chi^2_{\alpha}(1)$ 的所有指标中, 保留 $LM_j$ 值最大的一个指标 $X_j$ , 也即保留区分违约状态能力最强的指标 $X_j$ , 见表2第1行第8至9列。至此, 第一个指标遴选完毕。

#### 2.3.2 第二个指标的遴选

由于“①第一个指标的遴选”已经遴选了一个指标, 因此, 此时只剩下 $m-1$ 个指标。

此时进行含两个指标的一组指标测算: 其中, 一个指标是上文“2.3.1”已经选定的指标, 第二个指标是剩下的 $m-1$ 个指标中的任一个指标。重复上文“step1-Step7”, 遴选出第二个指标, 如表2第2行。

与第一个指标遴选过程的不同, 一是本过程一共得到 $m-1$ 个指标的 $LM_j$ 值, 二是这一步式(6)中由一个指标 $x$ 变为含两个指标 $x$ 的式(10), 即

$$p_i = p_i(y_i = 1) = \frac{1}{1 + e^{-(b_0 + b_j x_{i,j} + b_g x_{i,g})}} \quad (10)$$

表2 基于LM检验的指标遴选过程

(1)序号	(2)次数	(3)指标	(4)公式	(5)LM值	(6)是否通过检验	(7)最大的LM	(8)纳入指标	(9)最终体系
1	第1个指标遴选	$X_j(j=1,2,\dots,m)$	$m$ 个式(6)	$LM_j$	$IF(LM_j > \chi^2_{\alpha}(1)), \sqrt{\quad}$ ; or, $\times$	$LM_j = \max(LM_j   LM_j > \chi^2_{\alpha}(1))$	$X_j$	$\{X_j\}$
2	第2个指标遴选	$X_g(g=1,2,\dots,m, g \neq j)$	$m-1$ 个式(10)	$LM_g$	$IF(LM_g > \chi^2_{\alpha}(1)), \sqrt{\quad}$ ; or, $\times$	$LM_g = \max(LM_g   LM_g > \chi^2_{\alpha}(1))$	$X_g$	$\{X_j, X_g\}$
...	...	...	...	...	...	...	...	...
H	第h个指标遴选	$X_h(h=1,2,\dots,m, h \neq \dots \neq g \neq j)$	$m-h+1$ 个式(11)	$LM_h$	$LM_h$ 均小于 $\chi^2_{\alpha}(1)$ , $\times$	—	无	$\{X_j, X_g, \dots, X_h\}$

## 2.4 指标赋权与债信评分方程的建立

### 2.4.1 基于均方差的指标赋权

根据指标信息含量来赋权。指标偏离各自均值程度越大, 表明该指标信息含量越大, 则其权重也越大; 反之, 越小。

设:  $\sigma_j$ -第 $j$ 个指标的均方差, 公式见(5);  $w_j$ -第 $j$ 个指标的权重;  $m$ -指标个数。则:

$$w_j = \frac{\sigma_j}{\sum_{j=1}^m \sigma_j} \quad (12)$$

式(12)的含义: 式(12)将指标均方差值标准化为[0,1]之间的值, 以保证所有指标权重和为1。均方差值越大, 反应信息越多, 则指标的权重 $w_j$ 越大。

### 2.4.2 评分方程的建立

设:  $Z_i$ -第 $i$ 个小型工业企业债信得分;  $S_i$ -第 $i$ 个小型工业企业的债信标准化得分。则债信得分方程和债信标准化得分方程<sup>[10]</sup>分别为:

$$Z_i = \sum_{j=1}^m w_j x_{ij} \quad (13)$$

$$S_i = \frac{Z_i - \min_{1 \leq i \leq n} (Z_i)}{\max_{1 \leq i \leq n} (Z_i) - \min_{1 \leq i \leq n} (Z_i)} \times 100 \quad (14)$$

式(13)-(14)的含义: 式(13)表示第 $i$ 个小型工业企业的加权债信得分。因 $Z_i$ 值在[0,1]之间, 区分度较小, 故用式(14)将债信得分 $Z_i$ 标准化为[0,100]之间区分度大的得分值 $S_i$ , 以更精确的进行企业等级的划分。

式(10)的含义: 在含1个指标 $x_{ij}$ 的体系中纳入指标 $x_{ig}$ 后, 使第 $i$ 个小型工业企业违约的概率。

### 2.3.3 其余指标的遴选

其他指标的筛选, 仅需在“Step2: 构造 logistic 模型”中将式(6)变为:

$$p_i = p_i(y_i = 1) = \frac{1}{1 + e^{-(b_0 + \sum_{j=1}^m b_j x_{ij})}} \quad (11)$$

仿照第一个指标和第二个指标的遴选过程, 得到其余指标, 见表2其他行。

### 2.3.4 指标遴选的终止

若 $LM_j$ 值均小于 $\chi^2_{\alpha}(1)$ , 即所有指标均对区分违约状态没影响, 此时停止再纳入指标, 见表2最后一行。

拉格朗日乘数检验遴选指标的特色: 通过把每一个待遴选的指标、与前面已选定的 $j(j=0,1,2,\dots,m)$ 个指标共同组成的 $j+1$ 个指标组的所有数值、代入表示企业违约概率的拉格朗日函数, 得到 $j$ 个拉格朗日统计量 $LM_j(j=0,1,2,\dots,m)$ ; 然后对这 $j$ 个拉格朗日统计量 $LM_j$ 进行 $\chi^2$ 检验, 在所有通过 $\chi^2$ 检验的统计量 $LM_j$ 中、也即具有显著区分违约状态能力的指标中, 遴选 $LM_j$ 值最大的指标为评级指标, 保证了遴选的指标对鉴别全部企业的违约状态具有显著影响。

## 2.5 小型工业企业债信等级的划分

### 2.5.1 每个债信等级年违约损失率的计算

年违约损失率(LGD)指小型工业企业贷款违约时, 平均每年给银行造成的违约损失。

设:  $LGD_e$ -第 $e$ 个债信等级年违约损失率;  $f$ -第 $e$ 个债信等级内所有小型工业企业总数;  $L_{ie}$ -第 $e$ 个债信等级第 $i$ 个小型工业企业贷款年应收未收本息;  $R_{ie}$ -第 $e$ 个债信等级第 $i$ 个小型工业企业贷款的年应收本息。则<sup>[22]</sup>:

$$LGD_e = \sum_{i=1}^f L_{ie} / \sum_{i=1}^f R_{ie} \quad (15)$$

式(15)的含义: 每个债信等级内, 所有小型工业企业的年总应收未收本息占年总应收本息的比重。该种用同一债信等级内, 所有客户真实年应收未收本息和除以真实年应收本息和来测算年违约损失率的方法, 可更加客观的反映银行客户的真实年违约损失率。

### 2.5.2 债信等级划分

每个债信等级对应一定的债信得分区间, 通过对债信得分区间上下限的不断调整, 可以改变每个债信得分区间内的客户数, 进而改变每个得分区间对应的年违约损失率, 直至债信等级的划分结果满足“债信等级越高, 债信得分越高, 年违约损失率越低”。

由于该上下限调整过程的复杂性和工作量巨大特性, 该划分债信等级方法由计算机完成, 并为我们团队的国家发明专利<sup>[22]</sup>。下文实证将使用我们团队的国家发明专利<sup>[22]</sup>, 根据

小型工业企业的债信得分等参数，将所有小型工业企业划分为满足“债信等级越高，年违约损失率越低”的 9 个债信等级，为银行小型工业企业贷款提供决策依据。

2.6 基于 K-S 的债信评级模型区分度检验

检验目的：通过检验违约样本与非违约样本债信评分的分布是否有显著差异，来验证模型对违约状态的区分能力。

检验步骤：（1）构造原假设

原假设 H：非违约样本债信得分的分布与违约样本债信得分的分布不具有显著的差异。

（2）构造检验统计量 D

设： $D$ -非违约样本的观测累积分布函数与违约样本的观测累积分布函数值之差的最大值； $Sn_1(X)$ -非违约样本的观测累积分布函数； $Sn_2(X)$ -违约样本的观测累积分布函数； $n_1$ -非违约样本数； $n_2$ -违约样本数。则  $D$  的计算公式为<sup>[23]</sup>：

$$D=\max|Sn_1(X)-Sn_2(X)| \tag{16}$$

其中， $Sn_1(X)=K_1/n_1$ ， $K_1$  为非违约样本中不超过评分  $X$  的样本数； $Sn_2(X)=K_2/n_2$ ， $K_2$  为违约样本中不超过评分  $X$  的样本数。

（3）检验标准

当  $D$  统计量小于临界值时，接受原假设，即非违约样本债信评分的分布与违约样本债信评分的分布不具有显著的差

异。当  $D$  统计量大于临界值时，拒绝原假设，即非违约样本债信评分的分布与违约样本债信评分的分布具有显著的差异。

其中，临界值通过查 K-S 检验临界值表来确定。

3 小型工业企业债信评级实证研究

3.1 评价指标的选取

3.1.1 指标海选

通过梳理国内外经典文献中的高频评级指标，检索标普、邓白氏等国内外权威机构的评级指标，共海选了 107 个债信评级指标，分属于 2 个一级准则层：还款能力和还款意愿，及 7 个二级准则层：企业内部财务因素、企业内部非财务因素等，其中企业内部财务因素准则层下有 4 个三级准则层：偿债能力、盈利能力、营运能力和成长能力。具体如表 3 第 2-5 列所示。表 3 第 6 列是指标类型，第 7 列是指标对应的参考文献来源，第 8 列为指标遴选结果，具体遴选步骤将在下文“3.1.2-3.5”中给出。

3.1.2 指标初选

根据数据的可获取性及完整性进行指标初选。从 107 个海选指标中剔除经济环境、国家政策等 26 个数据不可获取的指标，在表 3 第 8 列中用“不可观测删除”标识。共保留了 81 个指标，如表 4 第 1 至 81 行第  $c$  列所示。

表 3 小型工业企业债信评级海选指标集

(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)指标来源	(8)遴选结果
1				资产负债率	负向	[3-4],[6],[8-9],[11-13]	LM 检验删除
2			偿债能力	流动负债经营活动净现金流比率	正向	[8]	相关性删除
...				...	...	...	...
28			贷款用途	...	定性	[12]	不可观测删除
29		企业	盈利	净资产收益率	正向	[3],[8],[12]	相关性删除
...		内部	能力	...	...	...	...
44		财务		每股收益	正向	[3]	不可观测删除
45		因素	营运	应收账款周转速度	正向	[3],[6],[8-9],[11],[13]	LM 检验删除
...			能力	...	...	...	...
54	还款			现金循环周期	正向	[3]	相关性删除
55	能力		成长	营业收入增长率	正向	[4],[8-9],[11]	LM 检验删除
...			能力	...	...	...	...
63				工资、福利增长率	正向	[8],[11]	不可观测删除
64				行业景气指数	正向	[3-4]	保留
...		企业外部宏观条件		...	...	...	...
72				经济环境	定性	[3-4],[12]	不可观测删除
73				相关行业从业年限	定性	[4],[7-9],[12]	保留
...		企业内部非财务因素		...	...	...	...
85				管理水平	定性	[3-4],[8-9],[11-12]	不可观测删除
86				法人代表学历	定性	[9-10]	LM 检验删除
...		法人代表基本情况		...	...	...	...
98				业主素质	定性	[4],[8-9]	不可观测删除
99	还款			企业到位注册资金类别	定性	[12]	相关性删除
...	意愿	企业基本信用情况		...	...	...	...
102				客户投诉率	定性	[8]	不可观测删除
103				企业纳税记录	定性	[4],[9],[12]	相关性删除
...		企业的商业信誉		...	...	...	...
106				企业间合同违约次数	定性	[3-4],[10]	保留
107		抵质押担保因素		抵押/质押/担保	定性	[4],[10]	保留

3.2 样本及数据来源

（1）样本来源

本文选取某全国性商业银行 1814 笔小型工业企业贷款数据作为实证样本。其中有 1799 个非违约小型工业企业，15 个违约小型工业企业。

（2）数据来源

数据来源于某全国性商业银行总行的小企业贷款系统数据库。

指标原始数据如表 4 第  $E$  列前 81 行所示，对应的标准化数据列于第  $F$  列，计算过程将在下文“3.3”中给出。表 4 第 1 列表头的 200809280094 是借据号。表 4 第 82 行是小型工业企业违约状态标识，其中违约用 1 标识，非违约用 0 标识。

3.3 指标数据标准化

（1）正向指标标准化

根据表 4 第  $d$  列指标类型，将正向指标原始数据、指标最大值、指标最小值，代入式(1)，得到正向指标标准化打分。

结果列入表 4 第  $F$  列对应行。

(2) 负向指标标准化

根据表 4 第  $d$  列指标类型, 将负向指标原始数据、指标最大值、指标最小值, 代入式(2), 得到负向指标标准化打分。结果列入表 4 第  $F$  列对应行。

(3) 区间型指标标准化

根据表 4 第  $d$  列指标类型, 将区间型指标原始数据代入式(3)进行标准化打分。

以表 4 第 71 行“年龄”指标为例, 说明区间型指标打分过程。因年龄在 31 和 45 岁之间的企业法人还款能力最强, 故年龄的最佳区间为[31,45]。

将表 4 第 1 列第 71 行  $v_{71,1}=37$ , 位于最佳区间内, 故由式(3)中(c)知  $X_{71,1}=1$ , 结果列入表 4 第 71 行第 1815 列。

同理, 将表 4 第  $d$  列其他区间型指标对应的第  $E$  列原始数据、最大值、最小值及  $q_1$ 、 $q_2$  代入式(3), 得到标准化值, 列入表 4 第  $F$  列对应行。

(4) 定性指标标准化打分

根据表 1 的打分标准及表 4 第  $d$  列指标类型, 对表 4 第  $c$  列中的定性指标进行标准化打分, 结果列入表 4 第  $F$  列对应行。

以表 4 第 55 行“相关行业从业年限”为例, 说明定性指标打分过程。

根据表 4 第 55 行第 1 列“相关行业从业年限”为 15, 及表 1 第 2 至 4 列“相关行业从业年限”打分标准“从业年限 $\geq 8$ 年, 为 1 分”知, 借据号为 200809280094 的企业“相关行业从业年限”打 1 分。结果列入表 4 第 55 行第 1815 列。

同理, 根据表 4 第 55 行第 2-1814 列“相关行业从业年限”指标原始数据, 及表 1 第 2 至 4 列“相关行业从业年限”打分标准, 得“相关行业从业年限”指标下其余 1813 个标准化值, 列入表 4 第 55 行第  $F$  列对应行。

根据表 4 第  $c$  列其余定性指标对应的第  $E$  列原始数据, 及表 1 第 2 至 4 列打分标准, 得到其余定性指标标准化值, 列入表 4 第  $F$  列对应行。

表 4 小型工业企业 81 个指标的原始数据及标准化数据

(a)序号	(b)准则层	(c)指标	(d)指标类型	(E)1814 笔借据的指标原始数据 $v_{ij}$		(F)1814 笔借据的指标标准化打分结果 $x_{ij}$					
						15 笔违约借据		1799 笔非违约借据			
				(1)200809280094	(1814)200501310008	(1815)200809280094	(1829)200411120136	(1830)E2010063000103	(3628)200501310008		
1	C <sub>1</sub> 偿债能力	X <sub>1</sub> 资产负债率	负向	0.390	...	0.235	0.599	0.454	0.819	...	0.763
2		X <sub>2</sub> 流动负债经营活动净现金流比率	正向	1.433	...	-0.029	0.478	0.472	0.478	...	0.475
3		X <sub>3</sub> 速动比率	正向	6.000	...	6.000	0.145	0.000	0.232	...	0.682
...		...	...	...	...	...	...	...	...	...	...
20		X <sub>20</sub> EBITDA 与总负债比率	正向	2,451	...	0.232	0.075	0.052	0.051	...	0.038
21	C <sub>2</sub> 盈利能力	X <sub>21</sub> 净资产收益率	正向	0.051	...	0.071	0.285	0.232	0.068	...	0.070
...		...	...	...	...	...	...	...	...	...	...
33		X <sub>33</sub> 经营活动产生现金流入小计	正向	63829387.42	...	24607700	0.000	0.005	0.000	...	0.001
34		X <sub>34</sub> 应收账款周转速度	正向	3.000	...	0.130	0.001	0.015	0.017	...	0.000
...		...	...	...	...	...	...	...	...	...	...
43	C <sub>3</sub> 营运能力	X <sub>43</sub> 现金循环周期	负向	0.41	...	-13.670	0.503	0.502	0.487	...	0.498
44		X <sub>44</sub> 营业收入增长率	正向	-0.33	...	0.000	0.132	0.197	0.201	...	0.197
...	C <sub>4</sub> 成长能力	...	...	...	...	...	...	...	...	...	...
48		X <sub>48</sub> 留存收益增长率	正向	0.355	...	0.144	0.510	0.519	0.509	...	0.511
49		X <sub>49</sub> 行业景气指数	正向	135.700	...	135.970	0.704	0.695	0.648	...	0.707
50		X <sub>50</sub> 城乡居民人均储蓄年末余额	正向	31770.900	...	25877.780	0.341	0.239	0.564	...	0.269
51		X <sub>51</sub> GDP 增长率	正向	17.500	...	0.162	0.473	0.356	0.346	...	0.407
52	C <sub>5</sub> 企业外部环境	X <sub>52</sub> 居民消费价格指数	区间	104.000	...	102.600	1.000	0.976	0.951	...	1.000
...		...	...	...	...	...	...	...	...	...	...
54		X <sub>54</sub> 恩格尔系数	负向	37.600	...	41.500	0.729	0.651	0.716	...	0.559
55		X <sub>55</sub> 相关行业从业年限	定性	15	...	10	1.000	0.000	1.000	...	1.000
...		...	...	...	...	...	...	...	...	...	...
63	C <sub>6</sub> 企业内部非财务因素	X <sub>63</sub> 企业通过本银行回笼贷款占比	正向	0.000	...	0.646	0.000	0.000	0.400	...	0.646
64		X <sub>64</sub> 学历	定性	专科	...	中专	0.900	0.700	1.000	...	0.700
...	C <sub>7</sub> 法人代表基本情况	...	...	...	...	...	...	...	...	...	...
71		X <sub>71</sub> 年龄	区间	37	...	50	1.000	1.000	0.758	...	0.848
...		...	...	...	...	...	...	...	...	...	...
74	C <sub>8</sub> 企业基本信用信息	X <sub>74</sub> 担任该职务时间	定性	1	...	13	1.000	0.000	0.400	...	0.400
75		X <sub>75</sub> 企业到位注册资金类	定性	资金型	...	资金型	1.000	0.800	1.000	...	1.000
76		X <sub>76</sub> 近三年企业授信情况	定性	有授信记录、无转贷	...	有授信记录、无转贷	1.000	0.000	0.500	...	1.000
77	C <sub>9</sub> 企业的商业信誉	X <sub>77</sub> 企业纳税记录	定性	纳税三年以上、无拖欠	...	纳税三年以上、无拖欠	1.000	0.250	1.000	...	1.000
...		...	...	...	...	...	...	...	...	...	...
80		X <sub>80</sub> 企业间合同违约次数	定性	违约 0 次	...	违约 0 次	1.000	1.000	1.000	...	1.000
81	C <sub>10</sub> 抵质押担保因素	X <sub>81</sub> 抵质押担保得分	定性	现金质押	...	担保公司保证	0.150	0.669	0.100	...	0.700
82		是否违约 $v_i$	——	1.000	...	0.000	1.000	1.000	0.000	...	0.000

3.4 基于相关性分析的第一次指标遴选

第一次遴选目的: 删除信息冗余指标, 使所建债信评级

体系更简洁。

删除指标标准: 在相关系数绝对值大于 0.7<sup>[19]</sup>的指标对中,

删除信息含量少的指标，也即删除均方差值相对小的指标。

Step1：相关系数矩阵的求解。

任意两行数代入式(4)，可求得相关系数。

在表 4 第 F 列组成的多列矩阵中，对于前 81 行，任取 2 行数，均可代入式(4)求得相关系数。

由此会得到一个相关系数矩阵。

Step2：遴选相关系数>0.7 的指标对。

将“Step1”的相关系数矩阵中、相关系数>0.7<sup>[19]</sup>的每一对指标列入表 5 的同一行中，两相关指标对应的关系系数列入表 5 的第 7 列，其中将指标对中即将被删除的那个指标列入

表 5 第 2 列。由此得到了表 5 的 17 对、34 个指标。

Step3：计算相关性指标的均方差

在表 4 的第 F 列数据组成的矩阵中，找到与表 5 相同的 34 个指标，也即找到了 34 行。把这 34 行的数据分别代入式(5)，得到 34 个均方差，结果列入表 5 第 3 列和第 6 列对应位置。

Step4：相关性指标的删除

删除表 5 相关性指标对中均方差值较小的指标，删除结果列入表 5 第 8 列对应位置。

经过相关性分析，共删除 17 个指标，剩余 64 个指标，删除结果列入表 3 第 8 列对应位置。

表 5 基于相关性分析的指标第一次遴选

(1)序号	(2)评级指标	(3)均方差	(4)序号	(5)评级指标	(6)均方差	(7)相关系数	(8)相关分析删除的指标
1	流动负债经营活动净现金流比率	0.095	2	全部资产现金回收率	0.132	0.90	流动负债经营活动净现金流比率
...	...	...	...	...	...	...	...
13	企业到位注册资金类别	0.292	14	近三年企业授信情况	0.297	0.80	企业到位注册资金类别
...	...	...	...	...	...	...	...
33	城市居民人均可支配收入	0.118	34	城乡居民人均储蓄年末余额	0.156	0.83	城市居民人均可支配收入

### 3.5 基于拉格朗日乘数检验的第二次指标遴选

对第一次遴选后剩余的 64 个指标，进行基于拉格朗日乘数检验的第二次遴选。

第二次遴选目的：删除同一准则层内不能显著区分违约状态的指标。

以“企业财务准则层”为例说明指标遴选过程。“企业财务准则层”经相关性删除后剩余 36 个指标。

#### 3.5.1 第一个指标的遴选

每一次将一个指标代入式(6)，由于“企业财务准则层”内共有 36 个指标，故需要代 36 次。

将需要代入式(6)的指标列入下文表 6 第 I 行第 4 列。应指出，由于相关分析已删除指标  $X_2$ ，故列在表 6 第 I 行中第 2 行的是指标  $X_3$ ，而非  $X_2$ 。其它指标列法同理。

以纳入“ $X_1$  资产负债率”指标为例，说明  $LM_j$  值的求解过程。

Step1：构造假设

$H_0$ ：“ $X_1$  资产负债率”指标在式(8)的系数  $b_1=0$ ，即“ $X_1$  资产负债率”指标对区分违约状态没有显著影响；

$H_1$ ：“ $X_1$  资产负债率”指标在式(8)的系数  $b_1 \neq 0$ ，即“ $X_1$  资产负债率”指标对区分违约状态有显著影响。

Step2：构造 logistic 模型

由于在表 4 中，所有标准化数据在 1815 至 3628 列，共 1814 列数据。将“ $X_1$  资产负债率”这一行对应的第 1815 至 3628 列，共计 1814 列数据  $x_{ij}$  代入式(6)，得 1814 个 logistic 方程：

$$\begin{cases} p_1 = p_1(y_1 = 1) = \frac{1}{1 + e^{-(b_0 + 0.599b_1)}} \\ p_2 = p_2(y_2 = 1) = \frac{1}{1 + e^{-(b_0 + 0.454b_1)}} \\ \dots\dots \\ p_{1814} = p_{1814}(y_{1814} = 0) = \frac{1}{1 + e^{-(b_0 + 0.763b_1)}} \end{cases} \quad (17)$$

Step3：构造对数似然函数

将式(17)及表 4 第 82 行第 F 列数值  $y_i$  代入式(7)，构造对数似然函数：

$$\ln L(b_1) = 1 * \ln p_1 + 1 * \ln p_2 + \dots + (1-0) * \ln(1-p_{1814}) \quad (18)$$

Step4：构造拉格朗日函数

将式(18)代入式(8)，得拉格朗日函数：

$$\ln L^*(b_1) = \ln L(b_1) + \lambda_1(b_1 - 0) \quad (19)$$

Step5：构造 LM 统计量

对式(19)求一阶偏导，得到  $S(b_1)$ 。求二阶偏导，得到  $I(b_1)$ 。

将  $S(b_1)$  和  $I(b_1)$  代入式(9)，得“ $X_1$  资产负债率”指标的拉格朗日乘数检验统计量 LM 表达式：

$$LM_1 = [S(b_1)|_{b_1=0}]^2 * [-I(b_1)|_{b_1=0}]^{-1} \quad (20)$$

令式(20)中的  $b_1=0$ ，可得到“ $X_1$  资产负债率”指标的  $LM_1$  值为 9.135，结果列入表 6 第 1 行第 5 列对应位置。

整个过程的计算可运用 spss 软件方便地得到。

Step6：显著性检验

根据显著性水平 0.05，查  $\chi^2$  分布统计表，得到拒绝域临界值  $\chi^2_{0.05}(1) = 3.841$ 。因  $LM_1 = 9.135 > \chi^2_{0.05}(1) = 3.841$ ，故拒绝假设  $H_0$ ，接受假设  $H_1$ ，即“ $X_1$  资产负债率”指标对区分违约状态有显著影响。

将“ $X_1$  资产负债率”指标的显著性检验结果列入表 6 第 1 行第 6 列对应位置。

Step7：其它 35 个指标的 LM 值计算

如前所述，“企业财务准则层”中共有 36 个指标，已经计算出一个指标“ $X_1$  资产负债率”的 LM 值，还剩 35 个 LM 值未计算。重复“Step1-Step6”，计算其他 35 个指标的  $LM_j$  值。然后进行显著性检验，结果列入表 6 第 5 至 6 列对应行。

Step8：遴选第一个指标

根据上文“Step1-Step6”遴选的第一个指标和“Step7”中遴选的其余指标，及表 6 第 I 行第 5 列的 LM 值，在  $LM > \chi^2_{0.05}(1) = 3.841$  的所有指标中，指标“ $X_6$  息税前利润流动负债比”的  $LM_6 = 26.760$  为最大，也即“ $X_6$  息税前利润流动负债比”指标区分违约状态的能力最强。故将“ $X_6$  息税前利润流动负债比”指标保留。

至此，第一个指标遴选结束，结果如表 6 第 I 行第 7 至 9 列所示。



表 6 基于拉格朗日乘数检验的指标第二步遴选

(1)序号	(2)次数	(3)指标	(4)公式(6)中代入的指标	(5)LM 值	(6)是否通过检验	(7)最大的 LM	(8)纳入指标	(9)最终体系
I	第 1 个指标	1	{X <sub>1</sub> }	9.135	√	26.760	X <sub>6</sub> 息税前利润流动 负债比	{X <sub>6</sub> }
	2	2	{X <sub>3</sub> }	8.082	√			
	...	...	...	...	...			
	5	5	{X <sub>6</sub> }	26.760	√			
	...	...	...	...	...			
II	35	35	{X <sub>47</sub> }	0.009	×	12.297	X <sub>30</sub> EBITDA 利润率	{X <sub>6</sub> , X <sub>30</sub> }
	36	36	{X <sub>48</sub> }	2.561	×			
	第 2 个指标	1	{X <sub>6</sub> , X <sub>1</sub> }	2.627	×			
	2	2	{X <sub>6</sub> , X <sub>3</sub> }	11.074	√			
	...	...	...	...	...			
...	30	30	{X <sub>6</sub> , X <sub>30</sub> }	12.297	√	...	...	...
	...	...	...	...	...			
	34	34	{X <sub>6</sub> , X <sub>47</sub> }	0.000	×			
	35	35	{X <sub>6</sub> , X <sub>48</sub> }	0.824	×			
	...	...	...	...	...			
VII	第 7 个指标	1	{X <sub>6</sub> , X <sub>30</sub> , X <sub>3</sub> , X <sub>36</sub> , X <sub>17</sub> , X <sub>46</sub> , X <sub>1</sub> }	0.001	×	—	无	{X <sub>6</sub> , X <sub>30</sub> , X <sub>3</sub> , X <sub>36</sub> , X <sub>17</sub> , X <sub>46</sub> }
	2	2	{X <sub>6</sub> , X <sub>30</sub> , X <sub>3</sub> , X <sub>36</sub> , X <sub>17</sub> , X <sub>46</sub> , X <sub>4</sub> }	2.830	×			
	...	...	...	...	...			
	29	29	{X <sub>6</sub> , X <sub>30</sub> , X <sub>3</sub> , X <sub>36</sub> , X <sub>17</sub> , X <sub>46</sub> , X <sub>47</sub> }	0.016	×			
	30	30	{X <sub>6</sub> , X <sub>30</sub> , X <sub>3</sub> , X <sub>36</sub> , X <sub>17</sub> , X <sub>46</sub> , X <sub>48</sub> }	0.020	×			

## 3.5.2 第二个指标的遴选

由于上文“3.5.1”已经遴选了一个指标，因此，此时只剩下 35 个指标。

此时进行含两个指标的一组指标测算：其中，一个指标是上文“3.5.1”已经选定的指标“X<sub>6</sub> 息税前利润流动负债比”，其余 35 个指标分别与之组合成 35 组指标，分别为(X<sub>6</sub>, X<sub>1</sub>), (X<sub>6</sub>, X<sub>3</sub>), ..., (X<sub>6</sub>, X<sub>48</sub>)，如表 6 第 II 行第 4 列所示。将每一组指标代入式(10)，共得 35 个方程。重复上文“step1-Step8”，遴选出第二个指标“X<sub>30</sub>EBITDA 利润率”，如表 6 第 II 行。

与第一个指标遴选过程的不同，一是本过程一共得到 35 个指标的 LM<sub>f</sub> 值，二是这一步式(17)中由一个指标变为含两个指标的式(21)，即

$$\begin{cases} p_1 = p_1(y_1 = 1) = \frac{1}{1 + e^{-(b_0 + 0.599b_1 + b_2X_{1,g})}} \\ p_2 = p_2(y_2 = 1) = \frac{1}{1 + e^{-(b_0 + 0.454b_1 + b_2X_{3,g})}} \\ \dots\dots \\ p_{1814} = p_{1814}(y_{1814} = 0) = \frac{1}{1 + e^{-(b_0 + 0.763b_1 + b_2X_{1,g})}} \end{cases} \quad (21)$$

## 3.5.3 其余指标的遴选

其他指标的筛选，仅需在“Step2: 构造 logistic 模型”中将式(6)变为：

$$p_i = p_i(y_i = 1) = \frac{1}{1 + e^{-(b_0 + \sum_{j=1}^m b_j x_{ij})}} \quad (22)$$

仿照第一个指标和第二个指标的遴选过程，得到其余 X<sub>3</sub>, X<sub>36</sub> 等指标。

## 3.5.4 指标遴选的终止

在对第 7 个指标遴选时，剩余的 30 个指标的 LM<sub>f</sub> 值均小于 0.384，也即均不通过检验，此时停止再纳入指标，{X<sub>6</sub>, X<sub>30</sub>, X<sub>3</sub>, X<sub>36</sub>, X<sub>17</sub>, X<sub>46</sub>} 即为遴选的指标。见表 6 第 VII 行。

其余准则层的指标遴选过程同理。经过拉格朗日乘数检

验，最终遴选了 20 个指标，结果见表 7 第 E 列。

## 3.6 指标赋权和债信评分方程建立

## 3.6.1 基于均方差的指标权重确定

以“X<sub>3</sub> 速动比率”为例，说明权重确定过程。

将表 7 第 1 行第 1 至 1814 列数据、小型工业企业总数 1814 及  $\bar{X}_3 = 0.193$  代入式(5)，可得 X<sub>3</sub> 的均方差为 0.214，结果列入表 7 第 1 行第 1815 列。

同理，将表 7 第 2-20 行第 1 至 1814 列指标数据代入式(5)，得其余指标的均方差，结果列入表 7 第 1815 列对应行。

将表 7 第 1815 列数据代入式(12)，得“X<sub>3</sub> 速动比率”的权重为 0.048，结果列入表 7 第 1816 列。

同理，将表 7 第 1815 列数据代入式(12)，得其余 19 个指标权重，结果列入表 7 第 1816 列对应行。

## 3.6.2 评分方程的建立

将表 7 第 1816 列指标权重代入式(13)，得评分方程：

$$Z_i = 0.048x_{i3} + \dots + 0.016x_{i30} + \dots + 0.079x_{i181} \quad (23)$$

将表 7 第 1 列借据“200809280094”的标准化数据代入式(23)，得债信得分：

$$Z_1 = 0.048 \times 0.145 + \dots + 0.079 \times 0.100 = 0.564$$

将结果列入表 7 第 21 行第 1 列。

同理，将表 7 第 1 至 1814 列其余 1813 个小型工业企业的数据 x<sub>ij</sub> 代入式(23)，得到其余小型工业企业评级得分 Z<sub>i</sub>，列入表 7 第 21 行对应列。

将表 7 第 21 行第 1 列评级得分 0.564，1814 个小型工业企业债信得分最大值 0.831，最小值 0.151 代入式(14)，得标准化得分：

$$S_1 = (0.564 - 0.151) / (0.831 - 0.151) \times 100 = 60.735$$

结果列入表 7 第 22 行第 1 列。

同理，将表 7 第 21 行其余列 1813 个小型工业企业的评级得分 Z<sub>i</sub> 代入式(14)，得其余小型工业企业的标准化评级得分 S<sub>i</sub>，列入表 7 的第 22 行其余列。



表 7 小型工业企业债信评级指标体系

(a) 序号	(b) 一级 准则	(c) 二级 准则	(d) 三级 准则	(E) 指标层	1814 列借据的标准化数据 $x_{ij}$			(1815) 均方差 $\sigma_j$	(1816) 权重 $W_j$	(1817)“5C” 对应
					(1)200809280094	...	(1814)200501310008			
1	还款能力	企业内部财务因素	C <sub>1</sub> 偿债能力	X <sub>3</sub> 速动比率	0.145	...	0.682	0.214	0.048	资本
...			...	...	...	...	...	...	...	...
4		C <sub>2</sub> 盈利能力	C <sub>2</sub> 盈利能力	X <sub>30</sub> EBITDA 利润率	0.768	...	0.096	0.072	0.016	能力
...			...	...	...	...	...	...	...	...
7	还款意愿	C <sub>5</sub> 企业外部宏观环境	...	X <sub>49</sub> 行业景气指数	0.704	...	0.707	0.095	0.022	环境
...			...	...	...	...	...	...	...	...
10		C <sub>6</sub> 企业内部非财务因素	...	X <sub>55</sub> 相关行业从业年限	1.000	...	1.000	0.310	0.070	能力
...			...	...	...	...	...	...	...	...
14	还款意愿	C <sub>7</sub> 法人代表基本情况	...	X <sub>65</sub> 法人代表贷款违约记录	0.700	...	1.000	0.338	0.076	道德
...			...	...	...	...	...	...	...	...
18		C <sub>8</sub> 企业基本信用情况	...	X <sub>76</sub> 近三年企业授信情况	1.000	...	1.000	0.297	0.067	道德
...			...	...	...	...	...	...	...	...
20	C <sub>10</sub> 抵押担保因素	---	...	X <sub>81</sub> 抵押担保得分	0.100	...	0.700	0.348	0.079	担保
21			---	Z <sub>i</sub> 评级得分	0.564	...	0.585	---	---	---
22			---	S <sub>i</sub> 标准化评级得分	60.735	...	63.824	---	---	---

## 3.7 小型工业企业债信等级的划分

将表 7 第 22 行标准化评级得分  $S_i$  降序排列，结果列入表 8 第 2 列。表 8 第 3 至 4 列数据来自该银行总行贷款系统。

将表 8 第 3 至 4 列数据代入我们团队的信用等级划分国家发明专利<sup>[22]</sup>，将 1814 个小型工业企业样本划分为符合“债信等级越高，年违约损失率越低”原则的 9 个债信等级，如表 8 第 6 列所示。同时得到每个债信等级对应的样本数，列于表 8

第 5 列，以及债信等级对应的得分区间，列入表 8 第 8 列。

把表 8 第 3 列前 1107 行年应收本息的代数和作为分母、第 4 列中前 1107 行年应收未收本息的代数和作为分子代入式(15)，得到 AAA 级的年违约损失率  $LGD=0.002\%$ ，列入表 8 第 7 列的对应行。同理，将表 8 第 3 至 4 列数据代入式(15)，得其他等级的年违约损失率，列入表 8 第 7 列对应行。

表 8 1814 个小型工业企业债信等级划分数据及结果

(1)序号	(2)债信评分	(3)应收本息(元)	(4)应收未收本息(元)	(5)样本数	(6)债信等级	(7)年 LGD(%)	(8)得分区间
1	100	2278652.00	0.00	1107	AAA	0.002	(62.141,100]
...	...	...	...				
1107	62.14	195220.02	109744.79				
...	...	...	...	326	A	1.811	(52.571,60.860]
1163	60.85	4902661.70	0.00				
...	...	...	...				
...	...	...	...	27	BB	2.480	(48.490,50.656]
1543	50.66	3068984.72	2793982.47				
...	...	...	...				
...	...	...	...	44	CCC	3.447	(30.461,35.564]
1671	35.56	5911691.56	5279490.32				
...	...	...	...				
...	...	...	...	62	C	4.788	(0,21.972]
1753	21.88	11182133.92	0.00				
...	...	...	...				
1814	0.00	9847525.99	9336840.86				

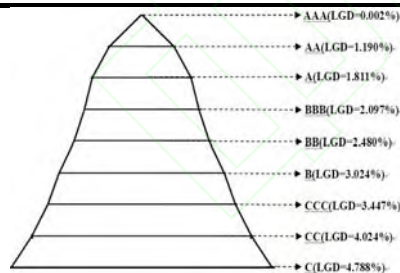


图 2 小型工业企业债信等级分布

为直观的体现“债信等级越高，年违约损失率越低”原则，以年违约损失率  $LGD$  为横坐标，以 9 个债信等级为纵坐标，得债信等级与年违约损失率  $LGD$  的分布图，如图 2 所示。

债信等级划分的特色：一是本研究的债信等级划分国家发明专利<sup>[22]</sup>不仅可以划分债信等级，而且可以给出每个债信等级对应的年违约损失率  $LGD$ ；二是本研究得到的年违约损失率分布符合“债信等级越高，年违约损失率越低”的客观实际。

## 3.8 基于 K-S 的债信评级模型区分度检验

## (1) 构造原假设

原假设  $H$ ：非违约样本债信评分的分布与违约样本债信评分的分布不具有显著的差异。

(2) 计算检验统计量  $D$ 

表 9 第 2 列债信得分区间值来自表 8 第 8 列。将每个债信得分区间内的非违约样本数列入表 9 第 3 列，违约样本数列入表 9 第 4 列。将样本累计频数分别列入表 9 第 5 至 6 列。

将每个债信得分区间对应的非违约样本累计频数除以非违约样本总数 1799，得到非违约样本累计频率，结果列入表 9 第 7 列。同理，违约样本累计频率列入表 9 第 8 列。

债信得分区间(62.141,100]对应的频率差值=非违约样本累计频率 0.615-违约样本累计频率 0.067=0.548，结果列入表 9 第 1 行第 9 列。同理，计算其他得分区间对应的频率差值，结果列入表 9 第 9 列对应行。

由表 9 第 9 列知，K-S 检验统计量  $D$  为频率差值的最大值 0.625，即  $D=0.625$ 。

## (3) 检验标准

通过查 K-S 检验临界值表，在 0.05 显著性水平、非违约样本数 1799 及违约样本数 15 下，得临界值  $D_{0.05}=0.353$ 。

即  $D$  统计量值 0.625 大于临界值 0.353，故拒绝原假设，即非违约样本债信评分的分布与违约样本债信评分的分布存在显著差异，也即模型具有显著区分违约状态的能力。

表 9 K-S 检验统计量的计算

(1) 序号	(2)债信得分 区间	(3)非违约样 本数	(4)违约样 本数	(5)非违约样本 累计频数	(6)违约样本累 计频数	(7)非违约样本 累计频率 $Sm_1$	(8)违约样本累 计频率 $Sm_2$	(9)频率 差值
1	(62.141,100]	1106	1	1106	1	0.615	0.067	0.548
2	...	...	...	...	...	...	...	...
3	(52.571,60.860]	325	1	1485	3	0.825	0.200	0.625
4	...	...	...	...	...	...	...	...
5	(48.490,50.656]	26	1	1564	5	0.869	0.333	0.536
6	...	...	...	...	...	...	...	...
7	(30.461,35.564]	43	1	1707	7	0.949	0.467	0.482
8	...	...	...	...	...	...	...	...
9	(0.21,972]	60	2	1799	15	1	1	0.000

3.9 小型工业企业债信评级模型特点分析

(1) 与“5C”原则相对应

本文建立的小型工业企业债信评级体系涵盖了经典“5C”原则的 5 方面：资本(capital)、能力(capacity)、品德(character)、经营环境(condition)、担保(collateral)。对应关系列入表 7 第 1817 列。

通过“速动比率”、“总资产周转速度”共 2 个指标反应“5C”原则的资本原则。通过“产品销售范围”、“相关行业从业年限”等 10 个指标反应“5C”原则的能力原则。通过“近三年企业授信情况”、“企业间合同违约次数”等 4 个指标反应“5C”原则的道德原则。通过“城乡居民人均储蓄年末余额”、“GDP 增长率”等 3 个指标反应“5C”原则的经营环境原则。通过“抵质押担保因素”指标反应“5C”原则的担保原则。

(2) 企业内部财务因素涵盖了 Altman 五因素

本文建立的小型工业企业债信评级体系中，企业内部财务因素指标涵盖了 Altman 的五因素：流动性(Liquidity)、盈利性(Profitability)、杠杆比率(Leverage)、偿债能力(Solvency)和活跃性(Activity)<sup>[24]</sup>。对应关系如表 10 第 3 到 4 列所示。

表 10 与 Altman 五因素的对应关系

(1)序号	(2)准则层	(3)指标层	(4)五因素
1	企业内部财务因素	速动比率	反映流动性
2		息税前利润流动负债比	反映偿债能力
3		未偿还贷款总额占资产总额比	反映杠杆比率
4		EBITDA 利润率	反映盈利性
...		...	...
6		总资产周转速度	反映活跃性

通过速动比率反映 Altman 的流动性，通过息税前利润流动负债比反映 Altman 的偿债能力，通过未偿还贷款总额占资产总额比反映 Altman 的杠杆比率，通过 EBITDA 利润率反映 Altman 的盈利性，通过总资产周转速度等 2 个指标反映 Altman 的活跃性。

(3) 鉴别违约状态能力强

本文建立的小型工业企业债信评级体系能更好的区分违约状态。剔除看似很好却不能较好的区分违约状态的指标，保证指标体系具有显著区分违约状态的能力。

例如，由表 6 第 1 行第 5 列可知，仅含“资产负债率”指标时，“资产负债率”的 LM 值为 9.135，高于 $\chi^2$ 分布的临界值 $\chi^2_{0.05}(1)=3.841$ ，表明指标“资产负债率”对小型工业企业的违约状态有显著鉴别能力。

(4) 非财务因素重要程度远远超过财务因素。

由表 7 第 E 列知，本文建立的小型工业企业债信评级体系中共有 20 个指标，其中企业财务因素指标有 6 个，占比 30%，非财务因素有 14 个，占比 70%，非财务因素个数占比远远超过财务因素个数占比。

由表 7 第 1816 列知，财务因素指标总权重为 0.188，非

财务因素指标总权重为 0.812，非财务因素指标总权重远远超过财务因素总权重。所以非财务因素重要程度远远超过财务因素。

由于小型工业企业的财务信息获取难，且真实性可考证性差，故为保证评级结果的可靠性，企业非财务因素占比超过财务因素才是合理的。

(5) 企业外部宏观环境指标不容忽视

小型工业企业规模小，市场占有率低，且大部分处在创业期，在很大程度上依赖所在地区的经济发展，禁不起地区经济的较大波动，抗风险弱。所以，在对小型工业企业进行评级时，不能忽视企业外部宏观环境指标。

由实证结果表 7 第 E 列可知，本文建立的小型工业企业债信评级体系中，用“行业景气指数”、“城乡居民人均储蓄年末余额”、“GDP 增长率”三个典型的宏观经济指标来反映企业外部宏观环境。

4 结 论

4.1 主要结论

对中国某商业银行 1814 个小型工业企业样本的实证结果表明：

(1) 行业景气指数、本地城乡居民人均储蓄年末余额等宏观因素；法人代表贷款违约记录、企业家年龄、性别和任职时间等个人因素，未偿还贷款总额占资产总额比、企业间合同违约次数、贷款抵押担保情况等信用参数组成的 20 个指标评级体系，能够显著区分小型工业企业的违约状态。

(2) 企业非财务因素重要程度远远超过财务因素。由表 7 第 1816 列知，企业财务因素指标总权重为 0.188，企业非财务因素指标总权重为 0.812，由表 7 第 E 列知，企业财务因素指标占比 30%，非财务因素占比 70%。可知，非财务因素指标的重要程度远远超过财务因素。

(3) 通过 K-S 检验知，本文所求的非违约样本债信评分分布与违约样本债信评分的分布存在显著的差异，也即模型具有显著区分违约状态的能力。

(4) 划分的债信等级符合“违约损失率越高，债信等级越低”标准。由图 2 可直观看出，根据本团队的国家发明专利<sup>[22]</sup>划分的债信等级不仅给出每个等级的年违约损失率而且符合“债信等级越高，年违约损失率越低”的客观实际。

4.2 主要创新与特色

(1) 通过把每一个待遴选的指标、与前面已选定的  $j(j=0,1,2,...m)$  个指标共同组成的  $j+1$  个指标组的所有数值、代入表示企业违约概率的拉格朗日函数，得到  $j$  个拉格朗日统计量  $LM_j(j=0,1,2,...m)$ ；然后对这  $j$  个拉格朗日统计量  $LM_j$  进行  $\chi^2$  检验，在所有通过  $\chi^2$  检验的统计量  $LM_j$  中、也即具有

显著区分违约状态能力的指标中, 遴选  $LM_j$  值最大的指标为评级指标, 保证了遴选的指标对鉴别全部企业的违约状态具有显著影响。

(2) 通过在相关系数大于 0.7<sup>[19]</sup> 的一对指标中, 删除均方差数值小的指标, 既避免了指标体系的信息冗余、又避免了误删信息含量大的指标。

### 参考文献

- [1] 工业和信息化部, 国家统计局, 发展改革委, 财政部. 中小企业划型标准规定[EB/OL]. 工业和信息化部网站, [http://www.gov.cn/jwqk/2011-07/04/content\\_1898747.htm](http://www.gov.cn/jwqk/2011-07/04/content_1898747.htm), 2011-6-18.
- [2] Dun and Bradstreet. Dun and Bradstreet Rating[EB/OL]. <http://www.dandb.com/glossary/d-b-rating/>, 2014.
- [3] 徐广军, 倪晓华, 肖运香. 标普、穆迪、邓白氏企业信用评级指标体系比较研究[J]. 浙江金融, 2007, (3): 51-52.
- [4] 中国工商银行. 关于印发《中国工商银行小企业法人客户信用等级评定办法》[R]. 中国工商银行, 工银银[2005]185 号.
- [5] 中国建设银行. 关于印发《中国建设银行小企业客户信用评级办法(试行)》[R]. 中国建设银行, 建总发[2009]101 号.
- [6] 中国银行. 关于印发《国内机构公司客户信用评级管理办法》[R]. 中国银行, 中银信[2010].
- [7] 中国人民银行征信中心与金融研究所联合课题组, 纪志宏, 王小明, 曹凝蓉等. 互联网信贷、信用风险管理与征信[J]. 金融研究, 2014, 10:133-147.
- [8] 黄飘. 网络时代的产业链中小贷—网络新金融系列专题报告之二: 阿里小贷与线上供应链金融[R]. 北京: 长城证券, 2013.4-8.
- [9] Maik Dierkes, Carsten Erner, Thomas Langer, *et al.* Business credit information sharing and default risk of private firms [J]. *Journal of Banking & Finance*, 2013, 37(8): 2867-2878.
- [10] 迟国泰, 潘明道, 程砚秋. 基于综合判别能力的农户小额贷款信用评级模型[J]. 管理评论, 2015, 06:42-57.
- [11] 邓超, 胡梅梅, 曾文潮. 基于贝叶斯界定折叠法的小企业信用评分模型研究[J]. 管理工程学报, 2015, 04:162-170.
- [12] 范柏乃, 朱文斌. 中小企业信用评级指标的理论遴选与实证分析[J]. 科研管理, 2003, 06:83-88.
- [13] Mu Zhang, Ying He, Zong-fang Zhou. Study on the Influence Factors of High-Tech Enterprise Credit Risk: Empirical Evidence from China's Listed Companies [J]. *Procedia Computer Science*, 2013, 17: 901-910.
- [14] 刘澄, 胡巧红, 孙莹, 武鹏, 黄翔. 基于 Vague 集理论的商业银行客户信用等级测定[J]. 统计与决策, 2013, 04:43-45.
- [15] 张目, 周宗放. 改进的云重心评判法在高新技术企业信用评级中的应用[J]. 数学的实践与认识, 2010, 19:37-44.
- [16] Dimitrios Niklis, Michael Doumpos, Constantin Zopounidis. Combining market and accounting-based models for credit scoring using a classification scheme based on support vector machines [J]. *Applied Mathematics and Computation*, 2014, 234(15): 69-81.
- [17] Nebojsa Nikolic, Nevenka, Iva Joksimovic. The application of brute force logistic regression to corporate credit scoring models: Evidence from Serbian financial statements [J]. *Expert Systems with Applications*, 2013, 15(40): 5932-5944.
- [18] Zarkic-Joksimovic, Djordje Stojanovski, Antonio Blanco, *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.
- [19] 尼尔·J·萨金德. 爱上统计学[M]. 重庆: 重庆大学出版社, 2010.
- [20] 巴克豪斯等. 多元统计分析方法[M]. 上海: 格致出版社, 2009.
- [21] 张定胜. 计量经济学[M]. 武汉: 武汉大学出版社, 2000.
- [22] 迟国泰, 石宝峰. 基于信用等级与违约损失率匹配的信用评级系统与方法[P]. 中国: CN102779317B, 2015-08-19.
- [23] 黄海等. spss10.0 for Windows 统计分析[M]. 北京: 人民邮电出版社, 2001.
- [24] 徐秀渠. Altman's Z-Score 模型在企业风险管理中的应用研究[J]. 经济经纬, 2010, 04:103-106.

## Facility rating model and empirical for small industrial enterprises based on LM test

CHI Guo-tai<sup>1</sup>, YU Shan-li<sup>1</sup>, YUAN Xian-zhi<sup>2</sup>

(1. Faculty of Management and Economics, Dalian University of Technology, Dalian 116024, China;

2. Institute of Risk Management, Tongji University, Shanghai 200092, China)

**Abstract:** Facility rating evaluates the creditworthiness and the LGD (Loss Given Default) of debt by establishing indicator system and evaluation equation. It plays an extremely important role in credit risk management. However, the difficulty in getting financial information of small industrial enterprises and validating the authenticity of the data makes credit rating harder. Taking 1814 small industrial enterprises loans from a commercial bank of China as empirical samples, this paper establishes credit rating for small industrial enterprises based on LM test.

First, by reviewing the international literature and credit rating system from various international financial agencies (Standard & Poor's etc.), we selected 107 small industrial enterprises credit rating indicators, and deleted 26 indicators from them and then standardize the remaining 81 indicators. Second, we removed indicators with smaller mean squared error between indicators whose the correlation coefficient is greater than 0.7, which reduces the information redundancy of facility rating indicator system and avoids eliminating indicators with more information content. Third, according to the criterion that the indicator with bigger discriminatory power should be reserved, we substituted  $j+1$  indicators into Lagrangian function, which is composed of one undetermined indicator and the previous selected  $j$  ( $j = 0, \dots, m$ ) numerical indicators, and selected the indicator  $x_j$  whose  $LM_j$  is the maximum and passes  $\chi^2$  test. By this way, we can ensure that the selected indicator  $x_j$  has a significant discriminatory power. Four, we weighed indicators by mean squared error method according to the principle that the indicator with more information should be endowed with a larger weight, and computed customer's facility score based on the linear weighted method. Last, according to the principle that customers with higher credit rating should correspond to lower LGD, this paper finally obtains 9 credit ratings and calculates the LGD of each credit rating.

This paper constructs facility rating indicator system for small industrial enterprises. The system includes 20 indicators, including macroeconomic, personal and credit factors. Macroeconomic factors involve industry climate index, etc. Entrepreneurs' personal factors involve legal representative default record etc. Credit parameters involve outstanding loans accounted for total assets ratio. Furthermore, non-financial factors are more important than financial



factors in the indicator system, which can be shown by the empirical result that total weight of financial factors is 0.188, whose quantity accounts for 30% of total quantity. The total weight of non-financial factors is 0.812, whose quantity accounts for 70%. What's more, this paper obtains 9 credit ratings, which satisfy the principle that customers with higher credit rating should correspond to lower LGD, and calculates the LGD of each credit rating to provide loan decision for banks.

Facility rating indicator system must have significant discriminatory power. In addition, for small industrial enterprises' facility rating, we should pay more attention to non-financial factors. Moreover, the credit rating must satisfy the criterion that customer with higher credit rank should correspond to lower LGD.

**Key words:** Industrial enterprises; Small enterprises; Credit rating; Facility rating; Lagrange multiplier test

中文编辑: 杜 健; 英文编辑: Charlie C. Chen

