

# 基于分层逻辑回归的小企业信用评价模型

李战江

(内蒙古农业大学 经济管理学院, 呼和浩特 010019)

**摘要:** 由于小企业的财务信息不健全、贷款难与急、违约样本数据缺乏等特点, 建立小企业的信用评价模型对于小企业获得贷款具有重要的作用。文章通过将统计抽样理论中分层思想与逻辑回归模型相结合, 构造了基于分层逻辑回归的小企业信用评价模型, 并对中国小企业进行了实证分析。

**关键词:** 小企业; 信用评价; 评价模型; 分层逻辑回归

**中图分类号:** F832.42; O212.4

**文献标识码:** A

**文章编号:** 1002-6487(2016)07-0178-05

## 0 引言

小企业信用评价是指以小企业信用风险评价指标体系为基础从信用得分大小层面对小企业的信用风险进行评价。建立小企业信用评价模型是划分小企业信用等级的必备基础, 同时也是商业银行对小企业提供贷款的决策依据。

小企业的信用评价方法有三类: ①基于统计与计量方法的小企业信用评价模型; ②基于人工智能的企业信用评价模型; ③基于随机概率方法的企业信用评价模型。

现有的企业信用评价模型不足之处: 一是人工智能方法与随机概率方法无法对信用评价模型的经济意义进行直观解释, 这个弊端影响了该类方法在银行信贷系统的应用。二是逻辑回归计量模型尽管不需要正态分布的严格假定, 但当解释变量个数较多时, 逻辑回归计量模型存在着变量之间的多重共线性严重以及模型结果的经济意义检验无法通过的弊端, 这些弊端极大地影响了逻辑回归模型的评价效果, 降低了逻辑回归模型的预测准确度。针对逻辑回归计量模型的不足, 现有研究提出了主成分-逻辑回归信用评分模型以及因子分析-逻辑回归信用评分模型, 这两种方法使用主成分变量或公共因子变量作为逻辑回归模型的解释变量, 较好地克服了逻辑回归模型中解释变量间的多重共线性问题, 但主成分方法或因因子分析方法的引入导致违约状态与信用风险评价变量间的经济意义不合理的弊端, 此外主成分方法以及因子分析方法本身的使用需要正态分布的严格假定。

本文通过将统计抽样理论中分层思想与逻辑回归模型相结合, 构造了基于分层逻辑回归的小企业信用评价模型。本文通过单一评价指标的秩和统计量  $Z$  绝对值与准则层内所有指标的  $Z$  绝对值的比值确定单一评价指标的

重要程度, 反映了指标对违约企业样本与不违约企业样本差异越大则指标权重越大的权重测算思路, 进而解决了每个准则层的得分测算问题。通过分层逻辑回归模型建立以准则层得分变量为解释变量、以违约状态为被解释变量的小企业违约概率测算模型, 克服了现有逻辑回归模型经济意义不合理的弊端, 解决了小企业信用评分的问题。

## 1 小企业信用评价模型的构建方法

### 1.1 准则层内指标的秩和统计量赋权方法

#### 1.1.1 秩和统计量 $Z$ 的计算

设  $R_{ij}$  为第  $i$  个评价指标违约样本中的第  $j$  个样本值在混合样本中的秩,  $i=1, 2, \dots, m, j=1, 2, \dots, n_1$ ;  $Z_i$  为第  $i$  个评价指标的秩和统计量值;  $n_1$  为违约样本的样本数目;  $n_2$  为不违约样本的样本数目;  $n$  为评价指标的样本总数目;  $t_i$  为混合样本中第  $i$  个重复样本值的重复次数,  $i=1, 2, \dots, g$ ;  $g$  为混合样本中重复样本值的个数;  $m$  为评价指标的个数。则  $Z_i$  的计算公式为:

$$Z_i = \frac{\sum_{j=1}^{n_1} R_{ij} - \frac{n_1(n+1)}{2}}{\sqrt{\frac{n_1 n_2 (n+1)}{12} - \frac{n_1 n_2 (\sum_{i=1}^g t_i^3 - \sum_{i=1}^g t_i)}}{12n(n-1)}} \quad (1)$$

式(1)反映了评价指标对违约状态的区分能力, 式(1)的绝对值越大则指标对违约状态的区分能力越强。其原因是:  $Z_i$  绝对值越大, 即秩和与其平均值的标准距离变大, 则拒绝违约样本中位数与不违约样本中位数相等的能力变大, 评价指标对违约状态的区分能力变强;  $Z_i$  绝对值越小, 即秩和与其平均值的标准距离变小, 则接受违约样本中位数与不违约样本中位数相等的能力变大, 评价指标对违约状态的区分能力变弱。

**基金项目:** 国家自然科学基金资助项目(71171031; 71471027); 国家社会科学基金资助项目(13XJY019); 中国博士后科学基金资助项目(2015M582754XB); 内蒙古畜牧业经济研究基地资助项目; 内蒙古农村牧区发展研究所资助项目

**作者简介:** 李战江(1977—), 男, 内蒙古乌海人, 博士, 讲师, 研究方向: 商业银行风险管理决策理论与模型。

### 1.1.2 秩和统计量Z的赋权方法

设 $a_{ij}$ 为第 $i$ 个准则层的第 $j$ 个评价指标的权重; $Z_{ij}$ 为第 $i$ 个准则层的第 $j$ 个评价指标的秩和统计量 $Z$ 取值, $i=1, 2, \dots, s; j=1, 2, \dots, k; s$ 为指标体系中准则层个数; $k_i$ 为第 $i$ 个准则层的指标个数。则第 $i$ 个准则层的第 $j$ 个评价指标的权重 $a_{ij}$ 的计算公式为:

$$a_{ij} = \frac{|Z_{ij}|}{\sum_{j=1}^{k_i} |Z_{ij}|} \quad (2)$$

式(2)的公式含义:第 $i$ 个准则层的第 $j$ 个评价指标的权重 $a_{ij}$ 等于该准则层的第 $j$ 个评价指标的秩和统计量 $Z$ 绝对值与该准则层的所有指标 $Z$ 绝对值之和的比值。

式(2)的经济意义:①每个准则层内的所有评价指标的权重和为1,且每个评价指标的权重为0到1内的取值;②反映了秩和统计量 $Z$ 绝对值取值越大则对应评价指标在准则层内权重越大。

### 1.2 准则层得分的确定方法

设 $S_i$ 为第 $i$ 个准则层的信用得分; $a_{ij}$ 为第 $i$ 个准则层的第 $j$ 个评价指标的权重; $Y_{ij}$ 为第 $i$ 个准则层的第 $j$ 个评价指标的标准化取值向量, $i=1, 2, \dots, s; j=1, 2, \dots, k; k_i$ 为第 $i$ 个准则层的指标个数; $s$ 为指标体系中准则层个数。

$$S_i = \sum_{j=1}^{k_i} a_{ij} Y_{ij} \quad (3)$$

其中权重系数 $a_{ij}$ 的取值区间为: $0 < a_{ij} < 1$ 。且 $a_{ij}$ 是标准化的权重系数。

式(3)的公式含义:第 $i$ 个准则层的信用得分等于第 $i$ 个准则层中所有信用风险评价指标的加权组合,其中权重和为1。

式(3)的作用:通过秩和统计量 $Z$ 赋权法确定评价准则层的信用得分,反映了准则层内评价指标对评价准则层信用得分的影响程度,解决了小企业信用评价中准则层信用得分的测算问题。

### 1.3 分层逻辑回归的构建方法

#### 1.3.1 分层逻辑回归模型的表达式

设 $P$ 为小企业的违约概率; $\beta_0$ 为分层逻辑回归模型的常数项; $\beta_i$ 为分层逻辑回归模型的第 $i$ 个回归系数, $i=1, 2, \dots, k; S_i$ 为小企业信用评价模型的第 $i$ 个准则层; $k$ 为指标体系中准则层个数。则分层逻辑回归模型的表达式为:

$$P = \frac{1}{1 + e^{-(\beta_0 + \sum_{i=1}^k \beta_i S_i)}} \quad (4)$$

式(4)的公式含义:小企业的违约概率等于准则层信用得分的非线性组合,其中 $\beta_i$ 是组合系数。

式(4)的适用条件:分层逻辑回归对小企业指标分布与类型无任何限制条件,适用于指标分布情况未知、指标类型任意情况下的小企业信用评价模型。

式(4)的经济意义:①反映了每一个准则层信用得分变量对小企业违约状态的非线性影响。②回归系数 $\beta_i$ 的取值符号反映了准则层信用得分变量对小企业违约状态的

非线性影响方向。③回归系数 $\beta_i$ 的取值绝对数大小反映了准则层信用得分变量对小企业违约状态的非线性影响程度。

#### 1.3.2 分层逻辑回归模型的参数估计

通过最大似然估计法确定分层逻辑回归模型的未知参数,得到小企业的违约概率测算模型。

设 $L$ 为分层逻辑回归的对数似然函数; $n$ 为第 $i$ 个小企业的违约状态取值; $\beta_0$ 为分层逻辑回归模型的常数项; $\beta$ 为分层逻辑回归模型的回归系数行向量, $\beta = [\beta_1, \beta_2, \dots, \beta_k]$ ;  $k$ 为准则层个数; $T_i$ 为第 $i$ 个小企业的准则层得分列向量, $T_i = [s_{i1}, s_{i2}, \dots, s_{ik}]^T$ ;  $n$ 为小企业样本个数。则分层逻辑回归的对数似然函数 $L$ 计算公式为:

$$L = \sum_{i=1}^n [r_i (\beta_0 + \beta T_i) + \ln(1 - \frac{e^{\beta_0 + \beta T_i}}{1 + e^{\beta_0 + \beta T_i}})] \quad (5)$$

对 $L$ 两边关于常数项 $\beta_0$ 以及回归系数行向量 $\beta$ 求偏导数,结果为<sup>[8]</sup>:

$$\begin{cases} \frac{\partial L}{\partial \beta_0} = \sum_{i=1}^n (r_i - \frac{e^{\beta_0 + \beta T_i}}{1 + e^{\beta_0 + \beta T_i}}) \\ \frac{\partial L}{\partial \beta} = \sum_{i=1}^n [r_i - \frac{e^{\beta_0 + \beta T_i}}{1 + e^{\beta_0 + \beta T_i}}] T_i \end{cases} \quad (6)$$

令偏导数为0,得到对数似然函数 $L$ 的最大值点为<sup>[8]</sup>:

$$\begin{cases} \sum_{i=1}^n (r_i - \frac{e^{\beta_0 + \beta T_i}}{1 + e^{\beta_0 + \beta T_i}}) = 0 \\ \sum_{i=1}^n [r_i - \frac{e^{\beta_0 + \beta T_i}}{1 + e^{\beta_0 + \beta T_i}}] T_i = 0 \end{cases} \quad (7)$$

式(7)的作用:通过求解式(7)的联立方程组确定分层逻辑回归模型(4)的参数估计值。

#### 1.3.3 分层逻辑回归模型的经济意义合理性检验

经济意义合理性检验的重要地位:在检验顺序中经济意义合理性检验是排名第一的检验。如果模型无法通过经济意义的合理性检验,则其它检验无效;如果模型通过经济意义的合理性检验,则进行其它检验。

分层逻辑回归模型的经济意义合理性检验是指对准则层信用得分变量对小企业违约状态的非线性影响方向进行合理性检验,即对式(4)中回归系数 $\beta_i$  ( $i=1, 2, \dots, k$ )的取值符号进行经济意义合理性检验,常数项 $\beta_0$ 不需要进行经济意义合理性检验。

分层逻辑回归模型的经济意义合理性检验原理:

(1)如果回归系数 $\beta_i > 0$ ,则准则层信用得分 $S_i$ 得分越高则小企业信用状况越差,此时经济意义不合理。其不合理的作用机理是:如果回归系数 $\beta_i > 0$ ,则准则层信用得分 $S_i$ 得分与式(4)的分母负相关,从而准则层信用得分 $S_i$ 得分与违约概率正相关,最终准则层信用得分 $S_i$ 得分与信用评分负相关,即准则层信用得分 $S_i$ 得分越高则企业信用评分越低,企业信用状况越差,这与准则层信用得分 $S_i$ 的经济意义矛盾。

(2)如果回归系数 $\beta_i < 0$ ,则准则层信用得分 $S_i$ 得分越高则小企业信用状况越好,此时经济意义合理。其合理的作用机理是:如果回归系数 $\beta_i < 0$ ,则准则层信用得分 $S_i$ 得分

与式(4)的分母正相关,从而准则层信用得分 $S_i$ 得分与违约概率负相关,最终准则层信用得分 $S_i$ 得分与信用评分正相关,即准则层信用得分 $S_i$ 得分越高则企业信用评分越高,企业信用状况越好,这与准则层信用得分 $S_i$ 的经济意义一致。

分层逻辑回归模型的经济意义合理性检验标准:若分层逻辑回归模型的所有回归系数 $\beta_i(i=1, 2, \dots, k)$ 都小于0,则小企业的分层逻辑回归模型通过经济意义合理性检验,模型的经济意义合理;反之,若小企业的分层逻辑回归模型存在某一个回归系数 $\beta_i(i=1, 2, \dots, k)$ 大于0,则小企业的分层逻辑回归模型没有通过经济意义合理性检验,模型的经济意义不合理。

### 1.3.4 分层逻辑回归模型的系数显著影响检验

分层逻辑回归模型的系数显著影响检验的原假设:第 $i$ 个准则层变量对小企业的违约状态无影响,即 $\beta_i=0$ ,其中 $i=1, 2, \dots, k$ 。备择假设是:第 $i$ 个准则层变量对小企业的违约状态有影响,即 $\beta_i \neq 0$ 。

分层逻辑回归模型的系数显著影响检验统计量的计算公式:设 $W_i$ 为第 $i$ 个准则层变量的系数显著影响检验的Wald统计量; $\hat{\beta}_i$ 为第 $i$ 个准则层变量的回归系数估计值; $SE_i$ 为第 $i$ 个准则层变量的回归系数估计值 $\hat{\beta}_i$ 的标准误差。则 $W_i(i=0, 1, \dots, k)$ 计算公式为:

$$W_i = \left( \frac{\hat{\beta}_i}{SE_i} \right)^2 \quad (8)$$

分层逻辑回归模型的系数显著影响检验过程:将 $W_i$ 统计量与临界值 $\chi^2(1)$ 比较大小的过程转化为 $W$ 统计量的检验概率 $P$ 值与显著性水平 $\alpha$ 比较大小的过程。若 $P$ 值小于 $\alpha$ ,则拒绝第 $i$ 个准则层变量对小企业的违约状态无影响的原假设,表明第 $i$ 个准则层变量对小企业的违约状态影响显著;若 $P$ 值大于等于 $\alpha$ ,则接受第 $i$ 个准则层变量对小企业的违约状态无影响的原假设,表明第 $i$ 个准则层变量对小企业的违约状态无显著影响。

分层逻辑回归模型系数显著影响检验的标准:如果分层逻辑回归模型的所有准则层系数(不包括常数项)都通过显著影响检验,则分层逻辑回归模型通过系数的显著影响检验;如果分层逻辑回归模型有一个准则层系数(不包括常数项)不能通过显著影响检验,则分层逻辑回归模型没有通过系数的显著影响检验。

### 1.4 小企业信用评分的构建方法

小企业建立信用评分模型的构建思路是:通过对小企业的违约概率进行线性变换,得到小企业的百分制信用评分。

设 $SS_i$ 为第 $i$ 个小企业的信用评分; $P_i$ 为第 $i$ 个小企业的违约概率。则第 $i$ 个小企业的信用评分 $SS_i$ 为:

$$SS_i = (1 - P_i) \times 100 \quad (9)$$

式(9)的经济意义:小企业的信用评分与小企业的违约概率负相关,即企业违约概率越大则信用评分越低,或企业违约概率越小则信用评分越高。

式(9)的合理性:① $P_i$ 表示小企业的违约概率, $1-P_i$ 表示小企业按时还清贷款的,即信用得分;②在 $1-P_i$ 基

础上乘以100,表示将小企业的信用得分取值范围从 $[0, 1]$ 变换为 $[0, 100]$ 。

## 2 小企业信用评价的实证

### 2.1 指标与数据

使用国内某大型商业银行2013年信贷数据库的小企业客户作为原始样本,标准化后的样本数据见表1第1-113列。小企业的违约界定为在合同约定到期日企业不能还清全部本息,其中违约用1标识,不违约用0标识。

表1 小企业信用评价模型的指标与数据

序号	(a) 准则层	(b) 指标名称	违约企业样本数据				不违约企业样本数据			
			(1)企业1	...	(26)企业26	(27)企业27	...	(113)企业113		
1	企业基本情况	资产负债率	0.411	...	0.726	0.504	...	0.443		
...		...	...	...	...	...	...	...		
13	企业外部宏观环境	抵质押担保得分	0.190	...	0.190	0.700	...	0.717		
14		法人代表本地居住年限	0.000	...	0.000	1.000	...	1.000		
15	企业外部宏观环境	法人代表信用卡记录	0.000	...	0.000	1.000	...	1.000		
16		城市居民人均可支配收入	0.221	...	0.155	0.452	...	0.493		
17	企业外部宏观环境	恩格尔系数	0.559	...	0.576	0.838	...	0.742		
18		违约状态	1	...	1	0	...	0		

### 2.2 计算准则层内评价指标的权重

将表1前17行第1-113列评价指标标准化数据代入式(1),计算得到每个评价指标的秩和统计量 $Z_i$ 取值,结果填入表2第3列。

将表2第3列的秩和统计量 $Z_i$ 取值代入式(2),依次得到各个准则层内的评价指标权重 $a_{ij}$ ,结果填入表2第4列。

表2 秩和统计量权重的计算

序号	(1) 准则层	(2) 指标名称	(3) 统计量 $Z_i$	(4) 权重 $a_{ij}$
1	企业基本情况	资产负债率	-3.139	0.095
...		...	...	...
13	企业外部宏观环境	抵质押担保得分	-2.892	0.088
14		法人代表本地居住年限	-3.551	0.489
15	企业外部宏观环境	法人代表信用卡记录	-3.716	0.511
16		城市居民人均可支配收入	-0.517	0.197
17	企业外部宏观环境	恩格尔系数	-2.107	0.803

### 2.3 计算各准则层的信用得分

将表1第1-113列的标准化数据的转置矩阵代入式(3),同时将表3第4列权重取值列向量代入式(3),依次得到各个准则层的信用得分 $S_i$ , $S_i$ 的计算结果填入表3第1-3列。

表3第4列的小企业违约状态数据来源于表1第18行。

表3 评价准则层信用得分的计算

序号	(1) 企业基本情况 $S_1$	(2) 法人代表基本情况 $S_2$	(3) 外部宏观环境 $S_3$	(4) 违约状态 $R$
1	0.196	0.000	0.492	1
...	...	...	...	...
26	0.246	0.000	0.493	1
27	0.456	1.000	0.762	0
...	...	...	...	...
113	0.635	1.000	0.693	0



## 2.4 分层逻辑回归模型的参数估计

表3第4列是小企业分层逻辑回归模型的被解释变量R的建模数据,表3第1-3列是小企业分层逻辑回归模型的解释变量 $S_i(i=1, 2, 3)$ 的建模数据。小企业分层逻辑回归模型的解释变量和被解释变量数据均已标准化数据,不需要进行标准化处理。

将表3第1-4列数据代入联立方程组(7),通过迭代方法求解联立方程组(7)的解,得到分层逻辑回归模型的参数估计值 $\beta_i(i=0, 1, 2, 3)$ ,  $\beta_i$ 的计算结果填入表4第2列,  $\beta_i$ 的标准误差SE<sub>i</sub>计算结果填入表4第3列。分层逻辑回归模型的参数估计通过SPSS统计软件内置程序完成。

表4 分层逻辑回归模型参数估计结果

序号	(1) 解释变量 $S_i$	(2) 参数估计值 $\beta_i$	(3) 标准误差SE <sub>i</sub>	(4) 统计量 $W_i$	(5) 检验概率P
1	常数项 $S_0$	6.603	3.298	4.008	0.045
2	企业基本情况 $S_1$	-4.778	2.114	5.109	0.024
3	法人代表基本情况 $S_2$	-2.443	1.227	3.962	0.047
4	企业外部宏观环境 $S_3$	-9.494	4.781	3.944	0.047

将表4第2列的参数估计值 $\beta_i$ 代入式(4),得到分层逻辑回归模型的表达式,结果如下:

$$P = \frac{1}{1 + e^{-(\beta_0 + \sum_{i=1}^k \beta_i S_i)}} = \frac{1}{1 + e^{-(6.603 - 4.778S_1 - 2.443S_2 - 9.494S_3)}} \quad (10)$$

$$= \frac{1}{1 + e^{-6.603 + 4.778S_1 + 2.443S_2 + 9.494S_3}}$$

式(10)就是初步建立的基于分层逻辑回归模型的小企业违约概率测算模型。下面要对式(10)进行经济意义的合理性检验、模型与数据的匹配程度检验以及系数的显著影响检验。

## 2.5 模型的检验

### 2.5.1 经济意义合理性检验

#### (1)系数 $\beta_1$ 取值的合理性检验

表4第2行第2列显示:企业基本情况准则层 $S_1$ 的系数 $\beta_1$ 取值为-4.778,  $\beta_1$ 小于0。按照上文的分层逻辑回归模型的经济意义合理性检验方法,企业基本情况准则层 $S_1$ 的系数 $\beta_1$ 取值通过经济意义合理性检验。

#### (2)系数 $\beta_2$ 取值的合理性检验

表4第3行第2列显示:法人代表基本情况准则层 $S_2$ 的系数 $\beta_2$ 取值为-2.443,  $\beta_2$ 小于0。按照上文的分层逻辑回归模型的经济意义合理性检验方法,法人代表基本情况准则层 $S_2$ 的系数 $\beta_2$ 取值通过经济意义合理性检验。

#### (3)系数 $\beta_3$ 取值的合理性检验

表4第4行第2列显示:企业外部宏观环境准则层 $S_3$ 的系数 $\beta_3$ 取值为-9.494,  $\beta_3$ 小于0。按照上文的分层逻辑回归模型的经济意义合理性检验方法,企业外部宏观环境准则层 $S_3$ 的系数 $\beta_3$ 取值通过经济意义合理性检验。

#### (4)分层逻辑回归模型的经济意义合理性检验

结合3个准则层系数的检验结果,按照上文的分层逻辑回归模型的经济意义合理性检验标准,本文建立的分层逻辑回归模型通过经济意义检验,本文建立的分层逻辑回归模型的经济意义合理。其中,分层逻辑回归模型中常数

项 $S_0$ 并不进行经济意义检验,因此常数项 $S_0$ 的取值符号与经济意义检验无关。

### 2.5.2 系数显著影响检验

将表4第2列的参数估计值 $\beta_i$ 以及表4第3列的标准误差SE<sub>i</sub>代入式(8),得到检验统计量 $W_i$ ,结果填入表4第4列。

在显著性水平 $\alpha$ 为0.05下, $W_i$ 统计量的检验概率P值小于0.05,根据以上分层逻辑回归模型的系数显著影响检验方法, $\beta_i$ 通过系数显著影响检验,即所有准则层变量都对小企业的违约状态有显著影响。

### 2.5.3 本文最终建立的分层逻辑回归模型

综上所述,式(10)完全通过了模型的经济意义检验以及模型的系数显著影响检验,因此得出结论:式(10)就是本文最终建立的基于分层逻辑回归的小企业的信用评价模型。

### 2.5.4 小企业的信用评分

表5第1列的小企业违约状态取值来源于表3第4列。将表3第1-3列数据按行代入式(10),得到小企业的违约概率值 $P_i(i=1, 2, \dots, 113)$ ,结果填入表5第2列。

将表5第2列代入式(9),得到小企业的信用评分,结果列入表5第3列。表5第3列就是本文所有小企业的信用得分。

表5 小企业信用评分的计算

序号	(1)违约状态 $R_i$	(2)违约概率 $P_i$	(3)信用评分SS <sub>i</sub>
1	1	0.729598	27.040
...	...	...	...
26	1	0.678196	32.180
27	0	0.005199	99.480
...	...	...	...
113	0	0.004254	99.575

## 3 对比分析

### 3.1 与逻辑回归模型的实证对比

小企业的逻辑回归模型参数估计的结果(不包括常数项)见表6。

表6 经典逻辑回归模型的参数估计结果

序号	(1) 指标名称及代码	(2) 参数估计值	(3) 标准误差	(4) W值	(5) P值
1	资产负债率 $Y_1$	22.741	13.075	3.025	0.082
...	...	...	...	...	...
13	抵押得分 $Y_{13}$	-4.547	5.150	0.780	0.377
14	法人代表本地居住年限 $Y_{14}$	-2.660	3.995	0.443	0.506
15	法人代表信用卡记录 $Y_{15}$	9.016	2016.789	0.000	0.996
16	城市居民人均可支配收入 $Y_{16}$	-28.970	21.570	1.804	0.179
17	恩格尔系数 $Y_{17}$	2.020	50.509	0.002	0.968

表6第2列显示:17个解释变量中有8个解释变量的系数取值为负,而有9个解释变量的系数取值为正。按照上文的经济意义合理性检验,仅当全部解释变量的系数为负值时经济意义才合理,故小企业的逻辑回归模型无法通过经济意义的合理性检验。

### 3.2 与主成分-逻辑回归模型的实证对比

对表1第b列的17个评价指标以及指标数据进行主成分分析,得到累积贡献率超过85%时对应的8个主成分变量 $C_i$ 。

将17个指标的标准化数据代入每一个主成分表达式,得到主成分 $C_i$ 的得分数据。以8个主成分 $C_i$ 为解释变量,以小企业违约状态为被解释变量,建立小企业的主成分-逻辑回归模型,主成分-逻辑回归模型的参数估计结果见表7。

表7第2列显示:8个主成分变量中有5个主成分变量的系数取值为负,而有3个主成分变量的系数取值为正。按照上文的经济意义合理性检验,仅当全部解释变量的系数为负值时经济意义才合理,故小企业的主成分-逻辑回归模型无法通过经济意义的合理性检验。

表7 主成分-逻辑回归模型的参数估计结果

序号	(1)指标名称及代码	(2)参数估计值	(3)标准误差	(4)W值	(5)P值
1	主成分 $C_1$	-2.903	2.122	1.872	0.171
2	主成分 $C_2$	-2.200	0.965	5.194	0.023
3	主成分 $C_3$	0.828	6.593	0.016	0.900
4	主成分 $C_4$	-3.255	3.103	1.100	0.294
5	主成分 $C_5$	6.784	11.063	0.376	0.540
6	主成分 $C_6$	-0.333	4.263	0.006	0.938
7	主成分 $C_7$	4.306	6.894	0.390	0.532
8	主成分 $C_8$	-13.365	7.158	3.486	0.062

### 3.3 与因子分析-逻辑回归模型的实证对比

对表1第b列的17个评价指标以及指标数据进行因子分析,按照累积贡献率超过85%选取8个公共因子,并对载荷矩阵进行方差最大正交旋转。

将17个指标的标准化数据代入每一个公共因子表达式,得到公共因子 $F_i$ 的得分数据。以8个公共因子 $F_i$ 为解释变量,以小企业违约状态为被解释变量,建立小企业的因子分析-逻辑回归模型,因子分析-逻辑回归模型的参数估计结果见表8。

表8第2列显示:8个公共因子变量中有7个公共因子变量的系数取值为负,而有1个公共因子变量的系数取值为正。按照上文的经济意义合理性检验,仅当全部解释变量的系数为负值时经济意义才合理,故小企业的因子分析-逻辑回归模型无法通过经济意义的合理性检验。

表8 因子分析-逻辑回归模型的参数估计结果

序号	(1)指标名称及代码	(2)参数估计值	(3)标准误差	(4)W值	(5)P值
1	公共因子 $F_1$	-2.625	1.460	3.234	0.072
2	公共因子 $F_2$	-1.303	0.513	6.451	0.011
3	公共因子 $F_3$	-0.103	0.741	0.019	0.889
4	公共因子 $F_4$	-2.842	2.911	0.953	0.329
5	公共因子 $F_5$	-0.662	0.359	3.410	0.065
6	公共因子 $F_6$	-0.778	0.434	3.214	0.073
7	公共因子 $F_7$	-2.552	1.719	2.204	0.138
8	公共因子 $F_8$	0.706	0.496	2.029	0.154

分层逻辑回归模型与逻辑回归模型的对比分析结果显示:逻辑回归模型经济意义不合理,而分层逻辑回归模型经济意义合理。分层逻辑回归模型与主成分-逻辑回归模型的对比分析结果显示:主成分-逻辑回归模型的经济意义不合理,而分层逻辑回归模型经济意义合理。分层逻辑回归模型与因子分析-逻辑回归模型的对比分析结果显示:因子分析-逻辑回归模型经济意义不合理,而分层逻辑回归模型经济意义合理。

### 参考文献:

- [1]Yu J, Zhou Z, Zhong H, et al. An Improved Fuzzy ISODATA Algorithm for Credit Risk Assessment of The EIT enterprises [J]. Modern Economy, 2012, 3.
- [2]Khashman A. Credit Risk Evaluation Using Neural Networks: Emotional Versus Conventional Models [J]. Applied Soft Computing, 2011, 11(8).
- [3]Danenas P, Garsva G, Gudas S. Credit Risk Evaluation Model Development Using Support Vector Based Classifiers[J]. Procedia Computer Science, 2011, 4.
- [4]曾诗鸿,王芳.基于KMV模型的制造业上市公司信用风险评价研究[J].预测, 2013, 32(2).
- [5]郭富霞,路兰,高齐圣.基于主成分的Logistic模型的上市公司信用风险评价分析[J].青岛大学学报(自然科学版), 2012, 25(2).
- [6]陆芳园.基于Logistic模型的中小企业信用风险管理研究[D].成都:西南财经大学, 2011.
- [7]吴喜之.非参数统计[M].北京:中国统计出版社, 2006.
- [8]毛毅,陈稳霖,郭宝龙等.基于密度估计的逻辑回归模型[J].自动化学报, 2014, 40(1).

(责任编辑/刘柳青)

## 4 结论