

基于违约鉴别能力最大的信用等级划分方法^①

迟国泰¹, 于善丽^{2, 3}

1. 大连理工大学经济管理学院, 大连 116024;
2. 中国人民银行金融研究所博士后科研流动站, 北京 100800;
3. 银行间市场清算所股份有限公司, 上海 200002)

摘要: 信用等级划分旨在区分不同客户的风险水平, 然现有大多数信用等级划分研究要么无法严格保证划分的信用等级满足“信用等级越高, 损失率越低”标准, 要么无法保证尽可能地区分违约可能性不相似的客户, 而无论不符合哪种标准, 划分的信用等级都不能作为贷款决策的有效依据. 基于此, 以上述两个标准为目标划分信用等级. 创新与特色一是以客户的非违约累计频率与违约累计频率之差的最大绝对值的代数和最大为目标, 以后一个信用等级损失率大于前面信用等级损失率为主要约束, 确保划分的信用等级满足上述两个标准. 二是提出通过设置随机分割点的区间来划分信用的新算法, 避免随机赋予信用等级分割点时, 靠前的信用等级分割点落到靠后的客户中, 导致后面的信用等级无论怎样划分均划分不出来的弊端. 最后, 以中国某商业银行 3 045 笔小企业贷款样本进行实证, 结果表明本模型划分的信用等级满足“信用等级越高, 损失率越低”标准, 且其区分不同违约可能性客户的能力较强.

关键词: 信用等级划分; 损失率; 违约鉴别能力; K-S 检验统计量; 信用评级

中图分类号: F830.5/O224 **文献标识码:** A **文章编号:** 1007-9807(2019)11-0106-21

0 引言

信用等级划分是根据客户的信用状况对客户进行排序分等级的过程, 它是评级的核心环节, 可以为贷款、债券等投资决策以及金融资产定价中信用价差的确定提供参考依据.

信用等级划分至关重要, 信用等级的变动小到引起企业倒闭, 大到诱发金融危机. 例如, 2001 年标准普尔将美国安然公司由 BBB 级调至垃圾级, 造成其股价一天内跌去 85%^[1]. 2011 年标准普尔将美国的主权信用评级从 AAA 级降为 AA+ 后, 引起了全球金融市场的动荡^[2].

合理的信用等级一方面应满足“信用等级越高, 损失率越低”的信用评级本质属性, 以避免给评级高的好客户贷款时也面临较大的损失, 造成决策混乱. 另一方面应能最大程度地区分开违约可能性不同的客户, 否则势必导致某些信用好的客户评级偏低, 某些信用差的客户评级偏高, 造成好客户贷款利率高, 甚至贷款失败, 差客户反而贷款利率低, 划分信用等级的意义失效.

然而, 现有一些信用等级划分结果却破绽明显, 例如信用状况相似的客户被分到两个不同的信用等级, 较高的信用等级对应的损失率反而也高, 较低的信用等级对应的损失率反而也低. 这样

① 收稿日期: 2018-04-05; 修订日期: 2019-03-02.

基金项目: 国家自然科学基金重点资助项目(71731003; 71431002); 国家自然科学基金资助项目(71873103); 中国博士后科学基金资助项目(2018M641578).

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

的评级结果将造成产品定价等决策的失误。

同时 根据数轴上任意两点间有理数无穷多的原理 对于成千上万的客户 任意相邻两个信用等级临界点的调整 均会引起相邻信用等级客户数的变化 造成相邻信用等级损失率的改变 进而引起其他信用等级损失率发生变化 导致信用等级间损失率调整顾此失彼 造成信用等级划分难。且大数据环境下 样本数量众多 如何从大量样本中找到最优的信用等级划分临界点 无疑是信用等级划分中亟待解决的难题。

现有信用等级划分的研究现状主要如下。

1) 根据信用等级划分方法 主要分为如下 3 类。

一是基于聚类方法划分。代表性的有: Chai 等^[3]利用 FCM 模糊聚类算法将贷款企业划分为 9 个信用等级。Shi 等^[4]基于模糊聚类将客户划分为 9 个信用等级。张大斌等^[5]通过差分进化自动聚类模型 将客户划分为 3 个信用等级。张洪祥和毛志忠^[6]基于多维时间序列的灰色模糊聚类方法确定客户的信用等级。

二是基于非线性目标规划方法。代表性的有: Zhang 和 Chi^[7]基于双目标非线性规划模型将客户划分为 9 个信用等级; 迟国泰和石宝峰^[8]以不同信用等级违约损失率级差之和最小为目标 以信用等级越高 违约损失率越低为约束建立非线性目标规划模型 将客户划分为 9 个信用等级; Krink 等^[9]以所有信用等级内每个客户的违约概率与均值违约概率的距离最小为目标函数划分信用等级。

三是其他方法。代表性的有: JP 摩根基于 Credit Metrics 模型原理 将客户划为 8 个信用等级^[10]。琚春华等^[11]基于 K-核信用价值度量模型对客户风险进行分级。庞素琳等^[12]基于风险环境的企业多层交叉信用评分模型对客户风险进行分类。余乐安和汪寿阳^[13]基于核主元分析的带可变惩罚因子最小二乘模糊支持向量机模型 将客户的信用风险分为 2 类。Bai 等^[14]运用模糊粗糙集的方法将贷款客户划分到 4 个信用等级。

2) 根据信用等级划分标准 主要分为如下 3 类。

一是根据违约概率或信用得分划分信用等级。代表性的有: Krink 等^[9]、Lyra 等^[15]均以违

约概率为标准划分信用等级。Serrano-Cinca 和 Gutiérrez-Nieto^[16]以客户 FICO 分数为标准 将客户划分为 7 个信用等级。

二是根据客户数服从“中间信用等级户数多 两头信用等级户数少”的钟形分布或正态分布划分信用等级。代表性的有: 陈洪海^[17]按照信用等级划分户数服从钟形分布的思想 将客户初分为 9 个信用等级 然后根据银行的经营价值等因素 对初分信用等级进行调整 使之更符合实际。迟国泰等^[18]对小样本数据进行扩充后 将客户划分为满足不同信用等级客户数近似服从正态分布的 9 个信用等级。Zhi 和 Yang^[19]以 200 家企业贷款为样本 根据不同信用等级的客户数应该近似服从正态分布的原理划分为 9 个信用等级。程砚秋^[20]根据不同信用等级的客户数应该近似服从正态分布的原理 划分为 9 个信用等级。

三是根据损失率划分信用等级。代表性的有: 根据信用等级越高 违约损失率越低的思路 将客户划分为 9 个信用等级^[8, 21]。赵志冲等^[22]以 8 个信用等级临界点对应的信用评分差值之和最大为目标 以信用等级越高 违约损失率越低为约束划分信用等级。李战江^[23]首先通过等分法 将客户按照信用得分均分为 9 个信用等级 然后调整初分的信用等级 使最终划分的信用等级满足“较高信用等级对应较低违约损失率”。

现有信用等级划分研究虽已取得长足进展 然仍存在以下弊端: 仅根据客户的违约概率(或信用得分)^[9, 15-16]或者客户数分布^[17-20]划分信用等级 没有损失率的计算公式 无法严格保证划分的信用等级满足“信用等级越高 损失率越低”的标准 而根据“信用等级越高 损失率越低”标准划分等级的研究^[8, 21-23]中 有些由于没有信用等级违约鉴别能力的衡量公式^[8, 21, 23] 无法严格保证划分的信用等级能尽可能地把违约可能性不相似的客户区分开 有些则无法保证不同信用等级的损失率增减稳定 可能出现两个信用等级损失率相差极小的不合理现象^[22] 如表 5 第 7 列基于文献[22]划分的信用等级结果中 BBB 信用等级损失率为 6.824% BB 信用等级损失率为

6.892% ,可以看出,这两个信用等级的损失率级差仅为0.068%(6.892% - 6.824% = 0.068%) ,过于敏感. 而无论不符合哪种情况,划分的信用等级都不能作为贷款决策的有效依据.

鉴于此,本文以信用等级间违约鉴别能力最大为目标,以信用等级越高、违约损失越低为主要约束划分信用等级,以确保越高的信用等级对应损失率越低的前提下,划分的信用等级能最大程度地区分风险不同的客户.

1 原理

1.1 损失率的界定

1) 损失率的定义

贷款违约的损失率(loss rate, LR)指贷款违约后,银行的损失额占银行应回收总额的比例.

银行应回收总额是指应收本息,也即按照合同约定,贷款客户应返还银行的本金和利息.

银行的损失额是指应收未收本息,也即银行应收本息中未收回的本息,包括应收未收本金和应收未收利息两项,而非仅仅是应收未收本金. 这主要是因为:对银行而言,当一笔贷款发生违约时,银行的实际损失不仅包括未偿还的贷款本金,还包括银行贷出这笔未偿还款项需要对应支付给存款储户的存款利息和其他财务费用. 因为贷款本金往往是出自该银行储户的存款,储户的存款利息需要由贷款利息进行覆盖. 故一旦贷款出现违约,银行损失的是这笔贷款应收、但未收的本金和利息.

2) 本文中损失率与巴塞尔新资本协议中损失率的关系

①本文中损失率(LR)的计算

设:LR为本文中损失率;L为应收未收本息;R为应收本息.

则如前所述,本文中贷款违约的损失率LR为

$$LR = \frac{L}{R} \quad (1)$$

②巴塞尔新资本协议中损失率的计算

设:EL为预期损失额;EAD为违约风险暴露;

PD为违约概率;LGD为违约损失率.

则根据巴塞尔新资本协议,有

$$EL = EAD \times PD \times LGD \quad (2)$$

将式(2)的左右两端同时除以违约风险暴露EAD,得到损失率公式为

$$\text{损失率} = \frac{EL}{EAD} = PD \times LGD \quad (3)$$

显然,式(3)损失率的分子预期损失额EL和分母违约风险暴露EAD均为估计值,而非真实值.

③本文中损失率与巴塞尔新资本协议中损失率的公式转换.

当式(3)的分子预期损失额EL为贷款的实际损失额,即应收未收本息时,则有

$$EL = L \quad (4)$$

当式(3)的分母违约风险暴露EAD为应收本息时,则有

$$EAD = R \quad (5)$$

将式(4)和式(5)代入式(3),则本文中损失率为

$$LR = \frac{EL}{EAD} = \frac{L}{R} = PD \times LGD \quad (6)$$

式(6)表示本文中损失率与巴塞尔新资本协议中损失率的关系,也即当巴塞尔新资本协议中损失率的分子预期损失额EL和分母违约风险暴露EAD均为真实值,不再是估计值时,即为本文中贷款违约后的损失率LR.

显然,基于真实值计算得到的贷款违约的真实损失率LR,比贷款违约的预期损失率对贷款等金融产品的投资与定价更有参考价值.

3) 以损失率LR为信用等级划分依据的原因:

①由式(6)知,损失率LR的经济学含义中同时包括了违约概率PD和违约损失率LGD的含义,能更全面的反映贷款客户的违约损失情况.

②由式(6)知,损失率LR为应收未收本息与应收本息之比,可以通过实际中贷款客户的真实“应收未收本息”和“应收本息”值计算得到,是个可方便观测的真实值,可直接用于贷款等金融产品的定价参数计算中.

③仅仅以信用得分、违约概率或者客户数分布为标准进行信用等级划分,不能直接衡量违约

后的损失,不能严格保证划分的信用等级满足“信用等级越高,损失率越低”的信用评级本质属性。

因此,以贷款客户违约的损失率 LR 为信用等级划分的主要依据,能确保越高的信用等级,对应的损失率越低,为客户贷款定价提供违约风险溢价参数参考。

1.2 信用等级划分的标准

信用等级划分的本质是揭示客户的违约风险大小,因此,合理的信用等级应满足如下 3 个标准。

标准 1 以损失率 LR 为信用等级划分依据。

如前所述,以损失率 LR 为信用等级划分依据,能更全面反映贷款客户的违约损失情况,而且可基于实际值方便地观测到,能为客户贷款定价提供更真实的违约风险溢价参数参考。

标准 2 信用等级越高,损失率越低。

好的信用等级应能确保越高的信用等级对应的损失率越小,越低的信用等级对应损失率越高。

若信用等级高,对应的损失率也高,则给高信用等级的客户贷款时也面临较大的损失,造成决策者决策混乱,划分信用等级的意义失效。且若信用等级高,对应的损失率也高时,可能会导致高信用等级的客户贷款定价高,也即好客户的贷款利率反而高,差客户的贷款利率反而低,造成定价混乱。

标准 3 信用等级间违约鉴别能力最大。

信用等级间违约鉴别能力是指划分的信用等级区分不同违约可能性大小的客户的程度。若划分的信用等级能把违约可能性大小不相似的客户区分开,也即违约可能性相似的客户都集中在 1 个信用等级内,则该信用等级间违约鉴别能力就越大,划分的信用等级就越好。

信用等级间违约鉴别能力最大标准,要求划分的信用等级能最大程度地区分开违约可能性不相似的客户,这也是信用等级划分的根本目的。若划分的信用等级不能最大程度地区分违约可能性不相似的客户,势必导致某些信用好的客户评级偏低,某些信用差的客户评级偏高,造成好客户贷款利率高,甚至贷款失败,差客户反而贷款利率低,划分信用等级的意义失效。

1.3 信用等级划分的难点

难点 1 如何确保划分的信用等级能最大程度地区分开违约可能性不相似的客户,使信用等级间具有最大的违约鉴别力。

难点 2 大数据环境下样本数众多,如何设置合理的信用等级划分算法,以避免无穷多次的划分调整。任意相邻两个信用等级临界点的调整,必然会使这两个信用等级客户数增加或减少,引起这两个信用等级损失率变化,进而而引起其他信用等级损失率变化。而由于样本数量巨大,这种调整是无穷多的,通过人工试凑或者计算机遍历都是不可能完成的,必须设置合理的划分算法。

1.4 突破难点的思路

突破难点 1 的思路: 信用得分对应的非违约客户累计频率和违约客户累计频率相差越大,说明违约客户和非违约客户越不能均匀地混合在一起,则信用得分越能把违约客户和非违约客户区分开,也即其违约鉴别能力就越大。故在违约客户累计频率与非违约客户累计频率的差值最大值对应的信用得分处,将客户划分到不同的两个信用等级,能确保以该信用得分为分割点的这两个信用等级能最大程度地区分开违约客户和非违约客户,信用等级间违约鉴别能力最大,此时,划分的信用等级内的客户大部分为同类客户,对应的 $K-S$ 检验统计量 D 值较大。

因此,以划分的所有信用等级的 D 均值最大为目标,建立目标规划模型划分信用等级,可以确保划分的信用等级能最大程度地区分违约可能性不相似的客户,使信用等级间具有最大的违约鉴别能力。

突破难点 2 的思路: 通过设置信用等级随机分割点的区间来划分信用等级,避免随机赋予信用等级分割点时靠前的信用等级分割点落到靠后面的客户中,导致后面信用等级无论怎样划分均划分不出来的弊端,减少了无效的信用等级划分次数,避免了无穷多次的调整。

基于违约鉴别能力最大的信用等级划分原理如图 1 所示。

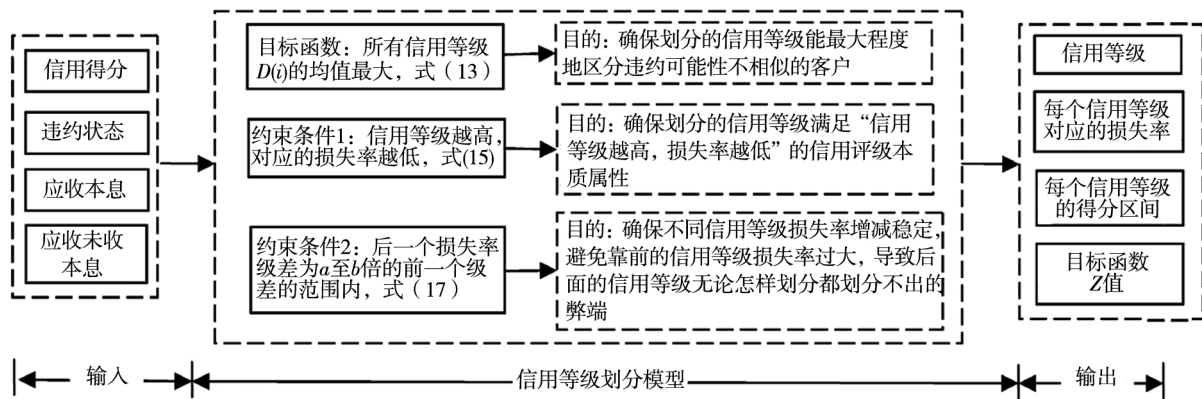


图1 基于违约鉴别能力最大的信用等级划分原理

Fig. 1 Credit grade division principle based on maximum default discrimination

2 信用等级划分模型

2.1 目标函数的建立

2.1.1 信用等级 i 的 K-S 检验统计量 $D(i)$ 值的计算

设:

$D(i)$ — 第 i 个信用等级的 K-S 检验统计量值 $i=1, 2, \dots, J$ J 为划分的信用等级总数;

$A_{1,h}(i)$ — 第 i 个信用等级内第 h 个信用得分对应的客户群内非违约客户数;

$N_{1,g}(i)$ — 第 i 个信用等级中第 g 个累加客户群内的非违约客户累计频数;

$N_{1,n}(i)$ — 第 i 个信用等级内最后一个客户群内的所有非违约客户频数;

$A_{2,h}(i)$ — 第 i 个信用等级内第 h 个信用得分对应的客户群内违约客户数.

$N_{2,g}(i)$ — 第 i 个信用等级内第 g 个累加客户群内违约客户的累计频数;

$N_{2,n}(i)$ — 第 i 个信用等级内最后一个客户群内的所有违约客户频数;

$Sn_{1,g}(i)$ — 第 i 个信用等级内第 g 个累加客户群内的非违约客户累计频率;

$Sn_{2,g}(i)$ — 第 i 个信用等级内第 g 个累加客户群内的违约客户累计频率.

则

$$N_{1,g}(i) = \sum_{h=1}^g A_{1,h}(i) \quad (7)$$

$$N_{2,g}(i) = \sum_{h=1}^g A_{2,h}(i) \quad (8)$$

$$Sn_{1,g}(i) = \frac{N_{1,g}(i)}{N_{1,n}(i)} \quad (9)$$

$$Sn_{2,g}(i) = \frac{N_{2,g}(i)}{N_{2,n}(i)} \quad (10)$$

1) 每个信用等级内的每个信用得分都对应着 1 个客户群, 将这些客户群按照信用得分降序排列.

每个信用得分都对应着 1 个或多个不同的客户, 把这 1 个信用得分对应的多个客户定义为 1 个客户群. 将这些客户群按照信用得分降序排列.

为清晰, 本处以客户信用得分为 1, 0.5, 0 为例, 说明 $D(i)$ 值的计算过程. 如表 1 第 2 列所示. 每个信用得分对应的客户群编号列入表 1 第 1 列.

2) 统计第 i 个信用等级内第 h 个信用得分对应的客户群内非违约客户数 $A_{1,h}(i)$ 和违约客户数 $A_{2,h}(i)$, 并将 $A_{1,h}(i)$ 和 $A_{2,h}(i)$ 分别列入表 1 第 3 列和第 4 列.

3) 计算第 i 个信用等级内第 g 个累加客户群内非违约客户累计频数 $N_{1,g}(i)$ 和违约客户累计频数 $N_{2,g}(i)$.

信用得分最高的那一个客户群, 每次都累加下一个较低信用得分对应的客户数, 会形成新的客户群, 即累加客户群.

例如第 1 个信用得分的客户形成第 1 个累加

客户群,前两个信用得分的客户形成第 2 个累加客户群,前三个信用得分的所有客户又形成第 3 个累加客户群.除了第 1 个累加客户群外,每次累加,都加上下一个群的客户,如此等等.

依次计算第 i 个信用等级内第 g 个累加客户群内的非违约客户累计频数 $N_{1,g}(i)$ 和违约客户累计频数 $N_{2,g}(i)$.

以表 1 为例,第 i 个信用等级内信用得分为 1

对应的累加客户群内、非违约客户累计频数 $N_{1,1}(i) = A_{1,1}(i)$ 列入表 1 第 5 列第 1 行.信用得分 0.5 对应的累加客户群内非违约客户累计频数 $N_{1,2}(i) = N_{1,1}(i) + A_{1,2}(i)$ 列入表 1 第 5 列第 2 行,其他累加客户群内非违约客户累计频数列入表 1 第 5 列其余行.

同理,违约客户累计频数 $N_{2,g}(i)$ 列入表 1 第 6 列对应行.

表 1 K-S 检验统计量 $D(i)$ 的计算

Table 1 Calculation of K-S test statistics $D(i)$

1	2	3	4	5	6	7	8	9	10
客户群号	信用得分	非违约客户数 $A_{1,h}(i)$	违约客户数 $A_{2,h}(i)$	非违约客户累计频数 $N_{1,g}(i) = \sum_{h=1}^g A_{1,h}(i)$	违约客户累计频数 $N_{2,g}(i) = \sum_{h=1}^g A_{2,h}(i)$	非违约客户累计频率 $Sn_{1,g}(i) = \frac{N_{1,g}(i)}{N_{1,n}(i)}$	违约客户累计频率 $Sn_{2,g}(i) = \frac{N_{2,g}(i)}{N_{2,n}(i)}$	累计频率差值 $D_g(i) = (7) - (8) $	D
1	1	$A_{1,1}(i)$	$A_{2,1}(i)$	$N_{1,1}(i) = A_{1,1}(i)$	$N_{2,1}(i) = A_{2,1}(i)$	$Sn_{1,1}(i) = N_{1,1}(i) / N_{1,3}(i)$	$Sn_{2,1}(i) = N_{2,1}(i) / N_{2,3}(i)$	$D_1(i) = Sn_{1,1}(i) - Sn_{2,1}(i) $	$\max D_g(i)$
2	0.5	$A_{1,2}(i)$	$A_{2,2}(i)$	$N_{1,2}(i) = A_{1,1}(i) + A_{1,2}(i)$	$N_{2,2}(i) = A_{2,1}(i) + A_{2,2}(i)$	$Sn_{1,2}(i) = N_{1,2}(i) / N_{1,3}(i)$	$Sn_{2,2}(i) = N_{2,2}(i) / N_{2,3}(i)$	$D_2(i) = Sn_{1,2}(i) - Sn_{2,2}(i) $	
3	0	$A_{1,3}(i)$	$A_{2,3}(i)$	$N_{1,3}(i) = A_{1,1}(i) + A_{1,2}(i) + A_{1,3}(i)$	$N_{2,3}(i) = A_{2,1}(i) + A_{2,2}(i) + A_{2,3}(i)$	$Sn_{1,3}(i) = N_{1,3}(i) / N_{1,3}(i)$	$Sn_{2,3}(i) = N_{2,3}(i) / N_{2,3}(i)$	$D_3(i) = Sn_{1,3}(i) - Sn_{2,3}(i) $	

4) 计算第 i 个信用等级内第 g 个累加客户群内的非违约客户累计频率 $Sn_{1,g}(i)$ 和违约客户累计频率 $Sn_{2,g}(i)$.

以表 1 为例,由于只有 3 行,其第 i 个信用等级内信用得分为 1 对应的累加客户群内非违约客户累计频率 $Sn_{1,1}(i) = N_{1,1}(i) / N_{1,3}(i)$ 见表 1 第 1 行第 7 列.其他累加客户群内非违约客户累计频率如表 1 第 7 列其余行.

同理,违约客户累计频率 $Sn_{2,g}(i)$ 如表 1 第 8 列.

5) 计算第 i 个信用等级内第 g 个累加客户群内的非违约客户累计频率 $Sn_{1,g}(i)$ 和违约客户累计频率 $Sn_{2,g}(i)$ 之差 $D_g(i)$,有

$$D_g(i) = | Sn_{1,g}(i) - Sn_{2,g}(i) | \quad (11)$$

例如,表 1 的信用得分为 1 对应的累加客户群内累计频率差值 $D_1(i) = | Sn_{1,1}(i) - Sn_{2,1}(i) |$,列入表 1 第 9 列第 1 行.同理,其他累加客户群内累计频数差值列入表 1 第 9 列第 2、第 3 行.

6) 第 i 个信用等级的 K-S 检验统计量 $D(i)$ 值的确定.

第 i 个信用等级的 K-S 检验统计量 $D(i)$ 是多个累加客户群中非违约累计频率与违约累计频率之差 $D_g(i)$ 的最大值,如表 1 第 10 列所示.也即^[24]

$$D(i) = \max D_g(i) = \max | Sn_{1,g}(i) - Sn_{2,g}(i) | \quad (12)$$

2.1.2 目标函数的建立

由上,目标函数为

$$\begin{aligned} \max Z &= \frac{1}{I} \left(\sum_{i=1}^I D(i) \right) \\ &= \frac{1}{I} \left(\sum_{i=1}^I \max | Sn_{1,g}(i) - Sn_{2,g}(i) | \right) \\ &= \frac{1}{I} \left(\sum_{i=1}^I \max \left| \frac{N_{1,g}(i)}{N_{1,n}(i)} - \frac{N_{2,g}(i)}{N_{2,n}(i)} \right| \right) \end{aligned} \quad (13)$$

式(13)进行最佳划分信用等级的思路如下.

将客户按照信用得分从高到低排列,任意确定第 1 个信用等级的信用得分临界点,根据式(7)~式(12),可以计算出第 1 个信用等级的 $D(1)$.任意确定第 2 个信用等级的得分临界点,可以计算出第 2 个信用等级的 $D(2)$.以此类推,

可以计算出第 I 个信用等级的 $D(I)$. 将 $D(1), D(2), \dots, D(I)$ 代入式 (13), 可以得到 1 个 Z 值.

同理, 重复上述步骤, 任意赋予不同信用等级的得分临界点 n 次, 可以确定 n 种不同信用等级的 $D(i)$ 及 Z 值, 其中 Z 值最大的信用等级划分即为最终的信用等级划分结果.

根据最终确定的最大的 Z , 可以得到相应的每个信用等级的得分分割点. 则此时最优信用等级划分完毕. 由式 (13) 知, 它保证了客户的非违约累计频率与违约累计频率之差的最大绝对值 $\max |Sn_{1,g}(i) - Sn_{2,g}(i)|$ 的代数和最大, 这就使信用等级的划分最大限度地区分了违约与非违约客户, 确保了信用等级间的违约鉴别能力最大.

2.2 约束条件的建立

2.2.1 约束条件 1 的建立

1) 单个信用等级损失率 LR_i 的确定

设:

LR_i —第 i 个信用等级的损失率, $i = 1, 2, \dots, I$ I 为信用等级总数;

n_i —第 i 个信用等级的客户总数;

L_{ij} —第 i 个信用等级中第 j 个客户的应收未收本息;

R_{ij} —第 i 个信用等级中第 j 个客户的应收本息.

则 LR_i 为

$$LR_i = \frac{\sum_{j=1}^{n_i} L_{ij}}{\sum_{j=1}^{n_i} R_{ij}} \quad (14)$$

式 (14) 是基于第 i 个信用等级内所有客户的真实应收未收本息 L_{ij} 和, 除以该信用等级内所有客户的真实应收本息 R_{ij} 和来测算 LR_i .

2) 约束条件 1 的确定

设 LR_i 为第 i 个信用等级的损失率, $i = 1, 2, \dots, I$, 其中 $i = 1$ 为信用最好信用等级, 依次递减, I 为信用等级总数, 则约束条件 1 为

$$0 < LR_1 < LR_2 < \dots < LR_I \quad (15)$$

在式 (15) 中, 后一个信用等级的损失率 LR_{i+1} 均严格大于前一个信用等级的损失率 LR_i ,

确保了划分的信用等级结果满足“信用等级越高, 损失率越低”的信用评级本质属性.

其中 $LR_1 > 0$ 的原因是实际中每个客户都存在着或大或小的违约风险, 不可能存在一个信用等级内的全部客户均不违约的情况. 因此, 根据每个信用等级内均会存在至少一个违约客户的客观实际, 令第 1 个信用等级的损失率大于 0.

2.2.2 约束条件 2 的建立

设 $\Delta LR_{i,j+1}$ 为第 i 个信用等级损失率 LR_i 与第 $i+1$ 信用等级损失率 LR_{i+1} 的级差, 则损失率级差 $\Delta LR_{i,j+1}$ 为

$$\Delta LR_{i,j+1} = LR_{i+1} - LR_i \quad (16)$$

式 (16) 的含义是相邻信用等级损失率的级差 $\Delta LR_{i,j+1}$ 为后一个信用等级的损失率 LR_{i+1} 减去前一个信用等级的损失率 LR_i .

设 a, b 是后一个损失率级差 $\Delta LR_{i,j+1}$ 为前一个级差 $\Delta LR_{i-1,j}$ 的倍数, 且 $a \leq b$. 则约束条件 2 为

$$\Delta LR_{i,j+1} \in [a, b] \times \Delta LR_{i-1,j} \quad (17)$$

a, b 过小, 损失率级差变化过于敏感, 即损失率少许有点变化, 就被划分为不同的信用等级了, 应该避免.

a, b 过大, 则会使得下一个损失率级差 $\Delta LR_{i,j+1}$ 过大, 下一个信用等级的损失率 LR_{i+1} 也过大, 导致下一个信用等级的违约客户数足够多, 而当样本的客户数确定的情况下, 样本中的违约客户数是固定的, 下一个信用等级的违约的客户数可能不会足够多, 这会导致下一个或下几个信用等级的损失率无法保证式 (15) 的条件成立, 导致划分信用等级的过程失败.

因此 a 和 b 的取值标准一是划分的信用等级损失变化不会过于敏感, 二是划分的信用等级时间不会过长. a 和 b 的确定可以经过计算机编程的计算过程调试, 找到最佳的数值范围. 例如下文“4 实证研究”的多次实证验证, 在一定的迭代次数内, 当损失率极差范围 $[a, b] = [0.1, 6]$ 时, 可以较快的划分出信用等级, 且损失率没有过度分散.

约束条件 2 的特色在于通过设置后一个损失率级差 $\Delta LR_{i,j+1}$ 为前一个级差 $\Delta LR_{i-1,j}$ 的 $[a, b]$ 倍的约束, 确保了划分的信用等级损失率级差在

一定范围内,保证了不同信用等级的损失率增减稳定,避免了划分的信用等级的损失率级差过于敏感或划分失效的现象。

2.3 本模型与现有研究的区别

本模型与现有研究的主要区别如表 2 所示,现分述如下。

表 2 本模型与现有研究的区别

Table 2 Difference between this model and the existing research

1	2	3	4	5		6	
				划分特点		优点	缺点
第一类 [9, 15-16]	信用得分或违约概率	以信用等级内每个客户的违约概率与违约概率均值的差距最小为目标划分信用等级 ^[9]	非线性目标优化法 ^[9] 等	能够保证每个信用等级内的客户违约概率均集中在其均值附近,违约概率离散程度最小	无法严格保证信用等级越高,损失率越低		
第二类 [17-20]	客户数	以不同信用等级间的客户数服从正态分布或钟形分布为目标进行信用等级划分 ^[17-20]	初分调整	能确保大量的客户集中在中间信用等级,少量的客户分布在最高和最低信用等级	既无法严格保证划分的信用等级间具有最大的违约鉴别能力,又无法严格保证信用等级越高,损失率越低		
第三类 [8, 21-23]	损失率	以相邻两个信用等级的评分差值之和最大为目标,以信用等级越高,违约损失率越低为约束划分信用等级 ^[22]	非线性目标规划 ^[8, 22] 、初分调整 ^[21, 23]	能确保信用等级越高对应的损失率越低	一是出现信用等级损失率相差过小的现象 ^[22] 。二是无法保证划分的信用等级具有最大的违约鉴别能力 ^[8, 21, 23]		
本模型	损失率	以信用等级间违约鉴别能力最大为目标,以信用等级越高,损失率越低为约束 1 以后一个损失率级差为前一个损失率级差的 a 至 b 倍为约束 2,划分信用等级	非线性目标规划	既能确保信用等级越高对应的损失率越低,且信用等级间损失率增减稳定,又能保证划分的信用等级间具有最大的违约鉴别能力	——		

2.3.1 第一类现有研究与本模型的区别

第一类现有研究是根据违约概率或信用得分划分信用等级。

本模型与第一类现有研究的主要区别是划分标准的区别。现有研究^[9, 15-16]以信用得分或者违约概率为标准划分信用等级,这种标准不能保证信用等级的违约损失符合“信用等级越高,损失率越低”的金字塔标准,因为这种标准跟损失率无关。本模型则以损失率为划分标准,通过式(15),确保划分的信用等级满足“信用等级越高,损失率越低”。

2.3.2 第二类现有研究与本模型的区别

第二类现有研究是根据客户数服从“中间等级户数多,两头信用等级户数少”的钟形分布或

正态分布划分信用等级。

本模型与第二类现有研究的主要区别如下。

一是划分标准的区别。现有研究^[17-20]以客户数为信用等级划分标准,这种标准跟损失率无关,亦不能保证信用等级的违约损失符合金字塔标准;

二是划分思路的区别。现有研究^[17-20]以不同信用等级间客户数服从钟形分布或正态分布为目标进行信用等级划分。这种划分的好处在于能确保大量的客户集中在中间信用等级,少量的客户分布在最高和最低的信用等级。

但这种划分因其信用等级划分模型中没有客户区分程度的公式,仅使“中间信用等级户数多、两头信用等级户数少”,无法严格保证划分的信

用等级能最有效区分好客户和差客户。

三是划分方法的区别,现有研究先初步划分信用等级,然后对初分的信用等级进行调整确定最终信用等级或者直接设置每个信用等级的客户比例划分信用等级。本模型为非线性目标规划方法。

2.3.3 第三类现有研究与本模型的区别

第三类现有研究是根据损失率划分信用等级。

本模型与第三类现有研究的主要区别是划分思路的区别。代表性研究^[8 21-23]或是以全部相邻信用等级损失率级差之间距离最小为目标,或是以信用等级临界点对应的信用评分差值之和最大为目标,或是在初分信用等级基础上进行调整,以找出满足“信用等级越高,损失率越低”的信用等级划分。这种划分的好处在于能确保高信用等级对应的损失率小,低信用等级对应的损失率高。

但相邻信用等级损失率级差间距最小仅能确保由高到低、不同信用等级的损失率递增速度的最小,信用等级临界点对应的信用评分差值之和最大,它对样本敏感,划分的信用等级稳定性不强。初分信用等级调整也仅能确保越高的信用等级损失率越低,不能使划分的信用等级对客户的区分能力最大。

2.4 模型的求解算法

2.4.1 满足约束条件1和约束条件2的第一次信用等级划分

步骤1 客户排序。

将客户按照信用得分从大到小进行排序,当两个或两个以上的客户的信用得分 S_j 相同时,则随机排列,确保每一个客户有1个排序的序号。

步骤2 设置初始参数。

设置第1个信用等级得分分割点所在的初始区间。

设 m_1 为步骤1中得分排序后的得分序号数。设置第1个信用等级得分序号的分割点可能落入的初始区间为 $[1, m_1]$ 。随机选取 $[1, m_1]$ 区间内的任意一个数为分割点,因为当样本足够多的时候,样本遍历不完,因此应该随机选取分割点的

值,用足够多的随机取值的次数代替遍历的次数。

根据式(15),建议 m_1 的取值应该使第1个信用等级内至少有1个违约客户。因此 m_1 取值为“大于等于包括第1个违约客户的得分序号”,确保第1个信用等级至少包括1个违约客户,以减少无效的信用等级划分次数。

设置第1个信用等级分割点区间 $[1, m_1]$ 的好处在于这会避免分割点落到靠后面的客户中,造成样本中第1个信用等级损失率比较大,后面的违约客户较少,无法划分成满足“信用等级越高,损失率越低”的金字塔标准。

设置信用等级个数 C 、迭代次数 φ 、以及级差倍数 a 和 b 的值。

1) 确定第1个信用等级。

步骤3 确定第1个信用等级分割点及损失率 LR_1 。

在该区间 $[1, m_1]$ 内随机生成1个分割点,假设该分割点序号为 $m_{1,1}$ 。根据式(14),计算分割点序号1到 $m_{1,1}$ 对应的所有客户的损失率 $LR_{1,1}$ 。

根据式(15),若该损失率 $LR_{1,1} > 0$,则 $m_{1,1}$ 即设置为第1个信用等级分割点。若该损失率 $LR_{1,1} \leq 0$,则将 m_1 加1,将第1个信用等级分割点区间范围扩大为 $[1, m_1 + 1]$ 。

重复上述过程,在区间 $[1, m_1 + 1]$ 内随机生成1个分割点,假设该分割点序号为 $m_{1,2}$ 。则根据式(14),计算分割点序号1到 $m_{1,2}$ 对应的所有客户的损失率 $LR_{1,2}$ 。若该损失率 $LR_{1,2} > 0$,则 $m_{1,2}$ 即为第1个信用等级的分割点。若该损失率 $LR_{1,2} \leq 0$,则将 m_1 加2,将区间范围扩大为 $[1, m_1 + 2]$,重复上述过程,直至找到使损失率 $LR_{1,2} > 0$ 的点,作为第1个信用等级的分割点。

假设最终确定的第1个信用等级的分割点为 m_1 ,也即第1个信用等级的分割点序号区间为 $[1, m_1]$ 。计算的分割点序号1到 m_1 对应的所有客户的损失率为 LR_1 。

2) 确定第2个信用等级

步骤4 设置第2个信用等级分割点所在区间。

设置第2个信用等级分割点可能落入的初始

区间为 [第 1 个信用等级的分割点 + 1, 第 1 个信用等级的分割点 + 2] = $[m_1 + 1, m_1 + 2]$.

步骤 5 确定第 2 个信用等级分割点及损失率 LR_2 .

在该区间 $[m_1 + 1, m_1 + 2]$ 内随机生成一个分割点, 假设该分割点序号为 $m_{2,1}$.

根据式(14), 计算分割点序号 $m_1 + 1$ 到 $m_{2,1}$ 对应的所有客户的损失率 $LR_{2,1}$.

根据式(15), 若该损失率 $LR_{2,1} > LR_1$, 则 $m_{2,1}$ 即为所求的第 2 个信用等级的分割点.

若该损失率 $LR_{2,1} \leq LR_1$, 则将第 2 个信用等级区间右端点值加 1 变为 $m_1 + 3$, 将第 2 个信用等级区间范围扩大为 $[m_1 + 1, m_1 + 3]$. 在该区间内随机生成一个分割点, 假设该分割点序号为 $m_{2,2}$.

同理, 根据式(14), 计算分割点序号 $m_1 + 1$ 到 $m_{2,2}$ 对应的所有客户的损失率 $LR_{2,2}$. 若该损失率 $LR_{2,2} > LR_1$, 则 $m_{2,2}$ 即为所求的第 2 个信用等级的分割点.

若该损失率 $LR_{2,2} \leq LR_1$, 则将第 2 个信用等级区间右端点值再次加 1 变为 $m_1 + 4$, 将区间范围扩大为 $[m_1 + 1, m_1 + 4]$, 在该区间内随机生成 1 个分割点, 重复上述步骤, 直至确定出第 2 个信用等级的分割点.

假设最终确定的第 2 个信用等级的分割点为 m_2 , 也即第 2 个信用等级的分割点序号区间为 $[m_1 + 1, m_2]$. 计算的分割点序号 $m_1 + 1$ 到 m_2 对应的所有客户的损失率值为 LR_2 .

步骤 6 计算第 1 个信用等级和第 2 个信用等级的损失率级差 $\Delta LR_{1,2}$.

根据式(16), 计算前两个信用等级的损失率级差 $\Delta LR_{1,2} = LR_2 - LR_1$.

3) 确定第 3 个信用等级

步骤 7 设置第 3 个信用等级分割点所在区间.

第 3 个信用等级分割点可能落入的初始区间设置为 [第 2 个信用等级的分割点 + 1, 第 2 个信用等级的分割点 + 2] = $[m_2 + 1, m_2 + 2]$.

步骤 8 确定第 3 个信用等级分割点及损失

率 LR_3 .

在该区间 $[m_2 + 1, m_2 + 2]$ 内随机生成 1 个分割点, 假设该分割点序号为 $m_{3,1}$. 计算分割点序号 $m_2 + 1$ 到 $m_{3,1}$ 对应的所有客户的损失率 $LR_{3,1}$. 若该损失率 $LR_{3,1} > LR_2$, 则计算第 2 个信用等级和第 3 个信用等级的损失率级差 $\Delta LR_{2,3} = LR_{3,1} - LR_2$, 比较 $\Delta LR_{1,2}$ 和 $\Delta LR_{2,3}$ 是否满足式(17). 若满足, 则 $m_{3,1}$ 即为所求的第 3 个信用等级的分割点.

若该损失率 $LR_{3,1} \leq LR_2$, 或者 $\Delta LR_{1,2}$ 和 $\Delta LR_{2,3}$ 不满足式(17), 则将第 3 个信用等级区间右端点值加 1 变为 $m_2 + 3$, 将第 3 个信用等级区间范围扩大为 $[m_2 + 1, m_2 + 3]$. 在该区间内随机生成 1 个分割点, 假设该分割点序号为 $m_{3,2}$, 计算 $m_2 + 1$ 到 $m_{3,2}$ 对应的所有客户的损失率 $LR_{3,2}$. 若该损失率 $LR_{3,2} > LR_2$, 且 $\Delta LR_{1,2}$ 和 $\Delta LR_{2,3}$ 满足式(17), 则 $m_{3,2}$ 即为所求的第 3 个信用等级的分割点.

若该损失率 $LR_{3,2} \leq LR_2$, 或者 $\Delta LR_{1,2}$ 和 $\Delta LR_{2,3}$ 不满足式(17), 则将第 3 个信用等级区间右端点值再次加 1 变为 $m_2 + 4$, 将区间范围扩大为 $[m_2 + 1, m_2 + 4]$, 在该区间内随机生成 1 个分割点, 重复上述步骤, 直至确定出第 3 个信用等级的分割点.

假设最终确定的第 3 个信用等级的分割点为 m_3 , 也即第 3 个信用等级的分割点序号区间为 $[m_2 + 1, m_3]$. 计算的 $m_2 + 1$ 到 m_3 对应的客户损失率为 LR_3 . 第 2 个信用等级和第 3 个信用等级的损失率级差 $\Delta LR_{2,3} = LR_3 - LR_2$.

4) 确定剔除最后两个信用等级后的其余信用等级

以实证 $I=9$ (划分为 9 个信用等级) 时为例, 此处是确定第 4 个 ~ 第 7 个信用等级的分割点.

第 4 个 ~ 第 7 个信用等级的分割点、损失率、损失率级差的确定方式, 与第 3 个信用等级分割点的确定方式一样.

例如假设确定的第 7 个信用等级的分割点为 m_7 , 其分割点区间为 $[m_6 + 1, m_7]$.

分割点 $m_6 + 1$ 到 m_7 对应的客户损失率记为 LR_7 .

第6个和第7个信用等级的损失率级差为 $\Delta LR_{6,7}$.

5) 确定最后两个信用等级 $I-1$ 和 I

仍以实证中 $I=9$ 为例进行说明. 则最后两个信用等级为第9个信用等级和第8个信用等级. 这两个信用等级的确定比较特殊, 它只需要在剩余客户中确定1个分割点即可同时确定这两个信用等级, 且用于划分最后两个信用等级的剩余客户数已经较少, 不需再采用随机设置分割点的思路, 而是通过遍历的思路确定最后一个分割点. 具体如下:

首先以“信用等级7的分割点 m_7+1 ”为信用等级8的起始点, 以“信用等级7的分割点 m_7+2 ”为信用等级8的中止点. 则信用等级8的分割点区间为 $[m_{I-2}+1, m_{I-2}+2] = [m_7+1, m_7+2]$, 信用等级9的分割点区间为 $[m_7+3, m_I]$, m_I 为最后一个得分序号的分割点.

根据式(14), 计算 m_7+1 至 m_7+2 对应的客户的损失率 $LR_{8,1}$, m_7+3 至最后一个得分序号 m_I 对应的客户的损失率 $LR_{9,1}$.

根据式(16), 计算信用等级7和信用等级8损失率级差 $\Delta LR_{7,8}$, 及信用等级8和信用等级9的损失率级差 $\Delta LR_{8,9}$.

若损失率满足 $LR_7 < LR_{8,1} < LR_{9,1}$, 且 $\Delta LR_{6,7}$ 、 $\Delta LR_{7,8}$ 和 $\Delta LR_{8,9}$ 均满足式(17), 则 m_7+2 即为信用等级8的分割点, 也即信用等级8的信用得分分割点区间为 $[m_{I-2}+1, m_{I-2}+2] = [m_7+1, m_7+2]$.

若损失率不满足 $LR_7 < LR_{8,1} < LR_{9,1}$, 或者 $\Delta LR_{6,7}$ 、 $\Delta LR_{7,8}$ 和 $\Delta LR_{8,9}$ 不满足式(17), 则以“信用等级7的分割点 m_7+3 ”为信用等级8的中止点. 则信用等级8的分割点区间为 $[m_{I-2}+1, m_{I-2}+3] = [m_7+1, m_7+3]$, 信用等级9的分割点区间为 $[m_7+4, m_I]$, m_I 为最后一个得分序号的分割点.

同理, 计算 m_7+1 至 m_7+3 对应的损失率 $LR_{8,2}$, $m_7+4 \sim m_I$ 对应的损失率 $LR_{9,2}$, 以及损失率级差 $\Delta LR_{7,8}$ 和 $\Delta LR_{8,9}$.

若损失率满足 $LR_7 < LR_{8,2} < LR_{9,2}$, 且 $\Delta LR_{6,7}$ 、 $\Delta LR_{7,8}$ 、 $\Delta LR_{8,9}$ 均满足式(17), 则得分序号 m_7+3 即

为信用等级8的分割点, 也即信用等级8的分割点区间为 $[m_{I-2}+1, m_{I-2}+3] = [m_7+1, m_7+3]$.

若损失率不满足 $LR_7 < LR_{8,2} < LR_{9,2}$, 或者 $\Delta LR_{6,7}$ 、 $\Delta LR_{7,8}$ 和 $\Delta LR_{8,9}$ 不满足式(17), 则以“信用等级7的分割点 m_7+4 ”为信用等级8的中止点.

重复上述过程, 直至损失率 LR_7 、 LR_8 、 LR_9 满足式(15), 损失率级差 $\Delta LR_{6,7}$ 、 $\Delta LR_{7,8}$ 和 $\Delta LR_{8,9}$ 满足式(17), 此时所有9个信用等级划分完成.

若信用等级8的分割点取遍剩余所有得分序号值, 仍有损失率 LR_7 、 LR_8 、 LR_9 不满足式(15), 或者损失率级差 $\Delta LR_{6,7}$ 、 $\Delta LR_{7,8}$ 、 $\Delta LR_{8,9}$ 不满足式(17), 则返回第1信用等级, 重复“1) ~ 5)”, 直至划分出满足约束条件1和2的第1种信用等级划分结果.

6) 计算第1次信用等级划分的目标函数 Z 值

第1次划分了 I 个信用等级, 所以有 I 个 $D(i)$ 值. 将这 I 个 $D(i)$ 值代入式(13), 得到第1次信用等级划分结果的目标函数 Z 值.

2.4.2 满足约束条件1和约束条件2的 n 次信用等级划分

重复“2.4.1, 1) ~ 2.4.1, 6)”中所有步骤, 可以划分出满足约束条件1和约束条件2的 n 种信用等级划分结果, 并得到 n 个信用等级划分结果的目标函数 Z 值.

2.4.3 确定最终的信用等级

比较这 n 次信用等级划分结果的 Z 值, 其中最大的 Z 值对应的信用等级划分结果即为最终划分的信用等级.

当设置迭代次数 n 特别大时, 满足约束条件1和约束条件2的各种信用等级划分情况会尽可能被遍历到, 则最终划分的信用等级的目标函数 Z 值越接近最优解.

2.4.4 信用等级划分输出的结果

1) 客户的信用等级及每个信用等级对应的损失率 LR_i .

2) 每个信用等级的信用得分区间确定如下:

①第1个信用等级的得分区间 = $[S_j(1), 100]$.

- ②第 i 个信用等级的得分区间 = $[S_j(i), S_j(i - 1))$ 其中 $i=2, 3, \dots, I-1$.
 - ③最后一个信用等级 I 的得分区间 = $[0, S_j(I-1))$.
 - 3) 目标函数 Z 值.
- 信用等级划分算法的流程如图 2 所示.

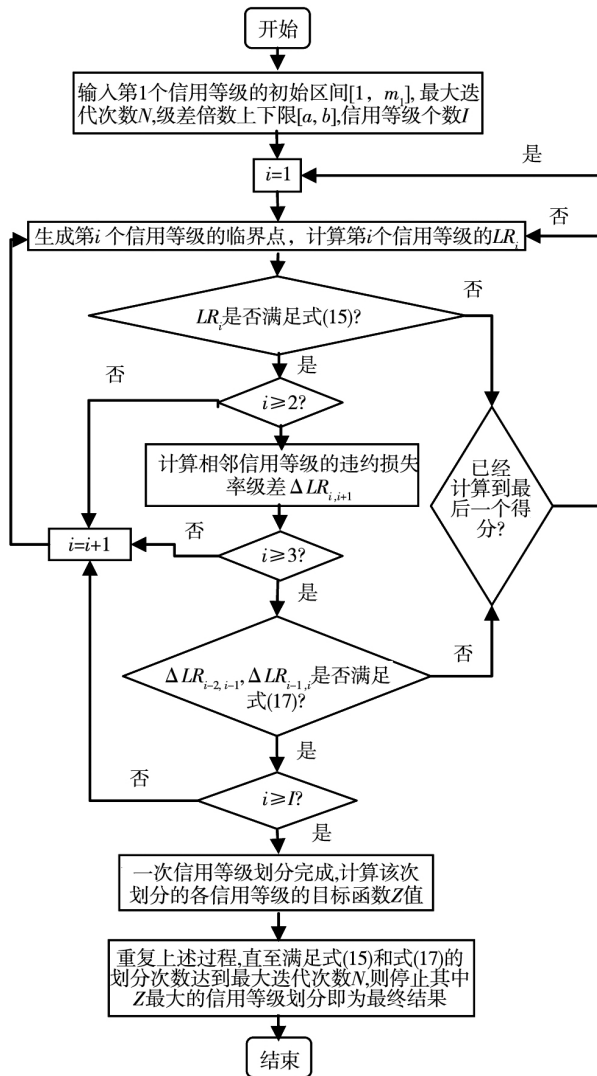


图 2 信用等级划分模型求解的流程图
Fig. 2 Solution of credit rating model

3 实证研究

3.1 样本及数据

从中国某商业银行的小企业信贷数据库中选取近 20 年、涉及京津沪渝等 28 个城市的 3 045 笔小企业贷款数据, 其中非违约小企业 2 995 笔, 违

约小企业 50 笔.

表 3 的数据均来自于文献 [25], 其中:

第 1 列为第 2 列所对应的序号. 由于第 2 列不同的得分有 1 878 个, 故第 1 列序号最后一个取值为 1 878.

第 2 列是从大到小进行排序后的客户信用得分 S_j . 信用得分 S_j 的确定不是本文的主要工作, 不赘述. 请注意表 3 的每一行为 1 个客户, 当两个或两个以上的客户的信用得分 S_j 相同时, 则相同得分的客户序号随机排列.

第 3 列为贷款客户号.

第 4 列和第 5 列分别为第 3 列所对应的应收本息 R_j 和应收未收本息 L_j , 数据来源于中国某商业银行的小企业信贷数据库 [26].

第 6 列客户的违约状态(违约 = 1, 非违约 = 0), 是与第 3 列客户对应的违约状态.

表 3 信用等级划分的数据

Table 3 Data used for credit rating

1	2	3	4	5	6
序号	信用得分 S_j	客户号	应收本息 R_j /元	应收未收本息 L_j /元	违约状态
1	100	1	302 625	0	0
⋮	⋮	⋮	⋮	⋮	⋮
4	94.94	13	203 276	0	0
5	92.90	14	51 336.5	0	0
⋮	⋮	⋮	⋮	⋮	⋮
773	62.38	1 376	960 244.48	631 142.51	1
774	62.36	1 377	1 868 738.28	0	0
⋮	⋮	⋮	⋮	⋮	⋮
1 878	0.00	3 045	78 904.72	78 904.72	1

3.2 信用等级的划分

通过参考中国人民银行信用评级管理指导意见的企业信用等级设置原则 [27], 设置小企业信用等级为 9 级, 即令 $I=9$.

3.2.1 确定满足约束条件 1 和约束条件 2 的第 1 次划分

步骤 1 如前所述, 将所有客户信息: 信用得

分 S_j 、应收未收本息 R_j 、应收本息 L_j 和违约状态，并按信用得分 S_j 从大到小排列成表(如表 3)。

步骤 2 设置初始参数. 设置第 1 等级分割点所在的初始区间为 $m_1 = 1\ 000$.

这样设置的原因在于由表 3 第 1、6 列知, 第 1 个违约客户出现在得分排序的第 773 行, 根据上文“2.4.1, 1) m_1 取值为“大于等于第 1 个违约客户的得分序号, 确保第 1 个信用等级至少包括 1 个违约客户”. 因此本文设置, 即 $[1, m_1] = [1, 1\ 000]$.

取信用等级个数 $I = 9$, 迭代次数 $m = 1\ 000$, 级差倍数下限 $a = 0.1$, 级差倍数上限 $b = 6$.

1) 确定第 1 个信用等级

步骤 3 确定第 1 个信用等级的分割点及损失率 LR_1 . 在区间 $[1, m_1] = [1, 1\ 000]$ 内随机生成得分序号分割点, 设该得分分割点序号为 772, 对应的客户号为 1 375. 则根据式(14), 计算客户 1 到客户 1 375 的损失率 LR_1 , 为

$$LR_1 = \frac{\sum_{j=1}^{1\ 375} L_{1j}}{\sum_{j=1}^{1\ 375} R_{1j}} = \frac{0 + \dots + 0}{302\ 625 + \dots + 3\ 006\ 372} = 0$$

由于该损失率 $LR_1 = 0$ 不满足式(15), 该得分分割点不是所求的第 1 个信用等级的分割点。

如 2.4.1, 1) 所述, “当样本足够多的时候, 样本遍历不完, 因此应该随机选取分割点值”. 故不能在得分分割点的基础上再加常数 1, 必须重设为区间 $[1, m_1 + 1] = [1, 1\ 000 + 1] = [1, 1\ 001]$ 在这个区间内随机选取 1 个数, 进行式(15)的验证, 以此类推.

因此, 将第 1 个信用等级区间右端点值 1 000 加 1, 将第 1 个信用等级分割点区间范围扩大为 $[1, 1\ 001]$.

在该区间内随机生成 1 个得分分割点, 重复上述过程, 直至找到使损失率大于 0 的点, 作为第 1 个信用等级的分割点.

当第 1 个信用等级分割点为 904 时, 对应客户号为 1 570, 根据式(14), 计算客户 1 到客户 1 570 的损失率 LR_1

$$LR_1 = \frac{\sum_{i=1}^{1\ 570} L_{1i}}{\sum_{i=1}^{1\ 570} R_{1i}} = \frac{0 + \dots + 631\ 142.505 + \dots + 0}{1\ 011\ 200 + \dots + 960\ 244.475 + \dots + 3\ 249\ 387.75} = 0.000\ 1$$

由于该损失率 $LR_1 > 0$, 满足式(15), 因此得分序号 904 即为此次所求的第 1 个信用等级的分割点.

则第 1 个信用等级的得分序号区间为 $[1, 904]$, 得分序号 904 在表 3 第 2 列中对应的信用得分为 60.239, 得分序号 1 在表 3 第 2 列中对应的信用得分为 100.

因此, 第 1 个信用等级的得分区间 $= [S_j(1), 100] = [60.239, 100]$, 列入表 5 第 3 列第 1 行备用.

相应的损失率 $LR_1 = 0.000\ 1$, 列入表 5 第 4 列第 1 行备用.

2) 确定第 2 个信用等级

确定第 2 个信用等级分割点可能落入的初始区间, 即为 $[m_1 + 1, m_1 + 2] = [904 + 1, 904 + 2] = [905, 906]$.

在该区间 $[905, 906]$ 内随机生成 1 个分割点, 假设该分割点序号为 906, 计算该得分序号 $[905, 906]$ 对应的客户 1 571 到 1 572 的损失率 LR_2

$$LR_2 = \frac{\sum_{i=1}^2 L_{2i}}{\sum_{i=1}^2 R_{2i}} = \frac{0 + 0}{10\ 816\ 927.26 + 203\ 726} = 0$$

由于该损失率 $LR_2 = 0 < LR_1 = 0.000\ 1$, 不满足式(15), 故 906 不是第 2 信用等级的分割点.

将第 2 个信用等级区间右端点值 906 加 1, 信用等级区间范围扩大为 $[905, 907]$. 在该区间 $[905, 907]$ 内随机生成 1 个分割点, 重复上述步骤, 当第 2 个信用等级的得分分割点序号为 965 时, 计算第 2 个信用等级对应的客户 1 571 到客户 1 669 的损失率 $LR_2 = 0.008\ 5 > LR_1 = 0.000\ 1$, 满足式(15), 故 965 是第 2 个信用等级的分割点.

表 4 第 1 个信用等级的 K-S 检验统计量 $D(1)$ 计算

Tabel 4 K-S test statistics $D(1)$ of the first credit rating

1	2	3	4	5	6	7	8	9	10
客户 群号	信用 得分	非违约客户数 $A_{1,h}(c)$	违约客户数 $A_{2,h}(c)$	非违约客户 累计频数 $N_{1g}(c) = \sum_{h=1}^g A_{1,h}(c)$	违约客户 累计频数 $N_{2g}(c) = \sum_{h=1}^g A_{2,h}(c)$	非违约客户 累计频率 $Sn_{1g}(c) = \frac{N_{1g}(c)}{N_{1n}(c)}$	违约客户 累计频率 $Sn_{2g}(c) = \frac{N_{2g}(c)}{N_{2n}(c)}$	累计 频率差值 $d_g(c) = (7) - (8) $	$D(c) = \max(d_g)$
1	100	4	0	4	0	0.003	0	0.003	0.876
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
772	62.385	1	0	1 375	0	0.876	0	0.876	
773	62.380	0	1	1 375	1	0.876	1	0.124	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
904	60.239	1	0	1 569	1	1	1	0	

则第 2 个信用等级的得分序号区间为 $[905, 965] = (904, 965]$,在表 3 第 2 列中对应的信用得分区间 $= [S_j(2), S_j(1)] = [59.355, 60.239)$,列入下文表 5 第 3 列第 2 行备用.

相应的损失率 $LR_2 = 0.0085$,列入表 5 第 4 列第 2 行备用.

根据式(16) ,前两个信用等级的损失率级差 $\Delta LR_{1,2}$ 为

$$\Delta LR_{1,2} = LR_2 - LR_1 = 0.0085 - 0.0001 = 0.0084$$

结果列入表 5 第 5 列第 1 行备用.

3) 确定第 3 个信用等级

确定第 3 个信用等级分割点可能落入的初始区间 ,即为 $[m_2 + 1, m_2 + 2] = [965 + 1, 965 + 2] = [966, 967]$.

在该区间 $[966, 967]$ 内随机生成 1 个分割点 ,假设该分割点序号为 967 ,同理 根据式(14) ,计算对应的损失率 $LR_3 = 0$.

由于损失率 $LR_3 \leq LR_2$,不满足式(15) ,则将第 3 个信用等级区间右端点值 967 加 1 ,区间范围扩大为 $[966, 968]$. 重复上述步骤 ,当第 3 个信用等级的得分分割点序号为 1 573 时 ,计算第 3 个信用等级对应的客户 1 670 到客户 2 516 的损失率 $LR_3 = 0.0156 > LR_2 = 0.0085$,满足式(15) .

根据式(16) ,计算第 2 个信用等级和第 3 个信用等级的损失率级差 $\Delta LR_{2,3}$,则有

$$\Delta LR_{2,3} = LR_3 - LR_2 = 0.0156 - 0.0085 = 0.0071$$

由于 $\Delta LR_{2,3} = 0.0071$,为 0.85 倍的 $\Delta LR_{1,2}$ (0.0084) ,满足式(17) ,故 1 573 即为第 3 个信用等级的分割点.

故第 3 个信用等级的得分序号区间为 $(965, 1 573] = [966, 1 573]$ 在表 3 第 2 列中对应的得分区间 $= [S_j(3), S_j(2)] = [43.986, 59.355)$,列入表 5 第 3 列第 3 行备用.

相应的损失率 $LR_3 = 0.0156$,列入表 5 第 4 列第 3 行备用.

第 2 个信用等级和第 3 个信用等级的损失率级差为 $\Delta LR_{2,3} = 0.0071$,列入表 5 第 5 列第 2 行备用.

4) 确定第 4 个信用等级 ~ 第 7 个信用等级

与“3.2.1 3)”中第 3 个信用等级的确定方式一样 ,得到信用等级 4 ~ 信用等级 7 的临界点及损失率.

① 第 4 个信用等级的得分序号区间为 $(1 573, 1 788] = [1 574, 1 788]$,在表 3 第 2 列中对应的信用得分区间 $= [S_j(4), S_j(3)] = [32.354, 43.986)$,列入表 5 第 3 列第 4 行备用.

相应的损失率 $LR_4 = 0.0548$,列入表 5 第 4 列第 4 行备用.

第 3 个信用等级和第 4 个信用等级的损失率级差为 $\Delta LR_{3,4} = 0.0392$,列入表 5 第 5 列第 3 行备用.

②第 5 个信用等级的得分序号区间为 $(1788, 1809] = [1789, 1809]$ 在表 3 第 2 列中对应的信用得分区间 = $[S_j(5), S_j(4)] = [30.307, 32.354]$ 列入表 5 第 3 列第 5 行备用.

相应的损失率 $LR_5 = 0.0707$,列入表 5 第 4 列第 5 行备用.

第 4 个信用等级和第 5 个信用等级的损失率级差为 $\Delta LR_{4,5} = 0.0159$,列入表 5 第 5 列第 4 行备用.

③第 6 个信用等级的得分序号区间为 $(1809, 1817] = [1810, 1817]$ 在表 3 第 2 列中对应的信用得分区间 = $[S_j(6), S_j(5)] = [29.227, 30.307]$ 列入表 5 第 3 列第 6 行备用.

相应的损失率 $LR_6 = 0.0778$,列入表 5 第 4 列第 6 行备用.

第 5 个信用等级和第 6 个信用等级的损失率级差为 $\Delta LR_{5,6} = 0.0071$,列入表 5 第 5 列第 5 行备用.

④第 7 个信用等级的得分序号区间为 $(1817, 1854] = [1818, 1854]$ 在表 3 第 2 列中对应的信用得分区间 = $[S_j(7), S_j(6)] = [17.303, 29.227]$ 列入表 5 第 3 列第 7 行备用.

相应的损失率 $LR_7 = 0.0914$,列入表 5 第 4 列第 7 行备用.

第 6 个信用等级和第 7 个信用等级的损失率级差为 $\Delta LR_{6,7} = 0.0136$,列入表 5 第 5 列第 6 行备用.

5) 确定最后两个信用等级 8 和信用等级 9

以“信用等级 7 的分割点 $m_7 + 1$ ”为信用等级 8 的起始点,即为 $1854 + 1 = 1855$,以“信用等级 7 的分割点 $m_7 + 2$ ”为信用等级 8 的第 1 次分割点,即 $1854 + 2 = 1856$.计算 $LR_8 = 0$,由于 $LR_8 \leq LR_7$,不满足式(15),则将第 8 个信用等级区间右端点值 1856 加 1 将第 8 个信用等级区间范围扩大为 $[1855, 1857]$.重复上述步骤,当第 8 个信

用等级的分割点为 1858 时,第 9 个信用等级的得分序号区间为 $(1858, 1878] = [1859, 1878]$ 分别计算这两个信用等级的损失率,则有 $LR_8 = 0.1678$ 和 $LR_9 = 0.3516$.

损失率 $LR_9(0.3516) > LR_8(0.1678) > LR_7(0.0914)$,且第 7 个信用等级和第 8 个信用等级的损失率级差 $\Delta LR_{7,8} = LR_8 - LR_7 = 0.1678 - 0.0914 = 0.0764$,为 5.62 倍的 $\Delta LR_{6,7}(0.0136)$,满足式(17),第 8 个信用等级和第 9 个信用等级的损失率级差 $\Delta LR_{8,9} = LR_9 - LR_8 = 0.3516 - 0.1678 = 0.1838$,为 2.41 倍的 $\Delta LR_{7,8}(0.0764)$,也满足式(17),故 1858 即为所求的第 8 个信用等级的分割点.

则第 8 个信用等级的得分序号区间为 $(1854, 1858] = [1855, 1858]$ 在表 3 第 2 列中对应的信用得分区间 = $[S_j(8), S_j(7)] = [16.531, 17.303]$,列入表 5 第 3 列第 8 行备用.

相应的损失率 $LR_8 = 0.1678$,列入表 5 第 4 列第 8 行备用.

第 7 个信用等级和第 8 个信用等级的损失率级差为 $\Delta LR_{7,8} = 0.0764$,列入表 5 第 5 列第 7 行备用.

第 9 个信用等级的得分序号区间为 $(1858, 1878] = [1859, 1878]$ 在表 3 第 2 列中对应的信用得分区间 = $[0, S_j(8)] = [0, 16.531]$,列入下文表 5 第 3 列第 9 行备用.

相应的损失率 $LR_9 = 0.3516$,列入表 5 第 4 列第 9 行备用.

第 8 个信用等级和第 9 个信用等级的损失率级差为 $\Delta LR_{8,9} = 0.1838$,列入表 5 第 5 列第 8 行备用.

在计算过程中,若损失率 LR_7, LR_8 和 LR_9 不满足式(15),或者损失率级差 $\Delta LR_{6,7}, \Delta LR_{7,8}$ 和 $\Delta LR_{8,9}$ 不满足式(17),则重复上述过程,若遍历完所有客户,仍不满足式(15)和式(17),则返回第 1 个信用等级,重复上文“2.4.1.1) ~ 2.4.1.5)” ,直至划分出满足式(15)和式(17)的第 1 种信用等级划分结果.

表 5 本模型的信用等级划分结果及其对比分析

Table 5 Results of model this paper established and its comparative analysis

1	2	本模型			对比模型 1 - 文献[22]模型			对比模型 2 - 文献[18]模型		
		3	4	5	6	7	8	9	10	11
序号	信用等级	信用得分区间	损失率 $LR(\%)$	损失率级差 $\Delta LR_{i,i+1}(\%)$	信用得分区间	损失率 $LR(\%)$	损失率级差 $\Delta LR_{i,i+1}(\%)$	信用得分区间	损失率 $LR(\%)$	损失率级差 $\Delta LR_{i,i+1}(\%)$
1	AAA	[60.239,100]	0.01	0.84	[62.223,100]	0.011	0.227	[80.718,100]	0	0
2	AA	[59.355,60.239)	0.85	0.71	[59.355,62.223)	0.238	0.282	[69.607,80.718)	0	0.086
3	A	[43.986,59.355)	1.56	3.92	[44.745,59.355)	0.520	6.303	[58.496,69.607)	0.086	-0.067
4	BBB	[32.354,43.986)	5.48	1.59	[33.345,44.745)	6.824	0.068	[47.385,58.496)	0.019	7.766
5	BB	[30.307,32.354)	7.07	0.71	[30.025,33.345)	6.892	1.994	[36.274,47.385)	7.785	-2.738
6	B	[29.227,30.307)	7.78	1.36	[13.391,30.025)	8.886	32.681	[25.162,36.274)	5.047	0.029
7	CCC	[17.303,29.227)	9.14	7.64	[6.844,13.391)	41.568	53.247	[14.051,25.162)	5.076	39.693
8	CC	[16.531,17.303)	16.78	18.38	[3.109,6.844)	94.814	5.186	[2.940,14.051)	44.769	55.231
9	C	[0,16.531)	35.16	—	[0,3.109)	100	—	[0,2.940)	100	—

重复上述过程中, 尽管第 1 个分割点至少包括 1 个违约客户的原则没变, 但是第 1 个信用等级的分割点区间右端点 m_1 值发生了变化, 导致第 1 个信用等级的分割点随机取值时发生了变化, 客户数发生了变化, 第 1 个信用等级的损失率也发生变化.

第 1 个信用等级分割点值的变化, 导致第 2 个信用等级分割点区间发生变化, 同理, 会引起第 2 个信用等级分割点的变化, 及客户数的变化, 导致第 2 个信用等级的损失率也发生变化. 以此类推, 引起后面其余信用等级的变化, 直至能找到符合式(15)和式(17)的信用等级划分结果.

至此, 划分出了 1 次满足约束条件 1 和约束条件 2 的信用等级结果.

6) 计算第 1 次信用等级划分的目标函数 Z 值

根据“2.1.1 步骤 1) ~ 步骤 6)”计算第 1 次划分的每个信用等级的 $D(i)$ 值.

以第 1 个信用等级的 $D(1)$ 值计算为例说明.

①由 3.2.1, 1) 知, 第 1 个信用等级的客户数为 1 570, 也即对应表 3 第 1 行 ~ 第 1 570 行.

这第 1 570 行客户的信用得分中, 不同的信用得分有 904 个, 将这 904 个不同的信用得分列入表 4 第 2 列.

同一个信用得分可能对应一个或多个客户, 这一个或多个客户称为 1 个客户群, 故 1 570 个

客户按照信用得分合并为 904 个客户群, 将对应的客户群编号列入表 4 第 1 列.

②根据表 3 前第 1 570 行第 2 列的不同得分, 一个得分的客户群统计非违约的客户数, 于是根据表 3 第 6 列的数据, 统计第 1 个信用等级内不同信用得分对应的客户群内非违约客户数 $A_{1,h}(1)$ 列入表 4 第 3 列.

同理, 可以统计出违约客户数 $A_{2,h}(1)$ 列入表 4 第 4 列.

③将表 4 第 3 列的 $A_{1,h}(1)$ 分别代入式(7), 得到第 1 个信用等级内不同信用得分对应的累加客户群内非违约客户累计频数 $N_{1,g}(1)$ 列入表 4 第 5 列.

同理, 将表 4 第 4 列 $A_{2,h}(1)$ 分别代入式(8), 得到第 1 个信用等级内不同信用得分对应的累加客户群内违约客户累计频数 $N_{2,g}(1)$ 列入表 4 第 6 列.

④将表 4 第 5 列每一行的非违约客户累计频数 $N_{1,g}(1)$, 第 1 个信用等级的非违约客户总数 $N_{1,n}(1) = 1 569$ 分别代入式(9), 得到第 1 个信用等级内不同信用得分对应的累加客户群内非违约客户累计频率 $Sn_{1,g}(1)$ 列入表 4 第 7 列.

同理, 将表 4 第 6 列每一行违约客户累计频数 $N_{2,g}(1)$, 第 1 个信用等级的违约客户总数 $N_{2,n}(1) = 1$ 分别代入式(10), 得到第 1 个信用等级内不同信用得分对应的累加客户群内违约客户

累计频率 $Sn_{2,g}(1)$ 列入表4第8列.

⑤将表4第7列 $Sn_{1,g}(1)$ 及表4第8列 $Sn_{2,g}(1)$ 代入式(11),得到第1个信用等级内不同信用得分对应的累加客户群内非违约客户累计频率 $Sn_{1,g}(1)$ 和违约客户累计频率 $Sn_{2,g}(1)$ 之差 $d_g(1)$ 列入表4第9列.

⑥表4第9列 $d_g(1)$ 的最大值为0.876,即有第1个信用等级 K-S 检验统计量 $D(1) = 0.876$ 列入表4第10列.

同理计算其他信用等级的 $D(c)$ 值,由此得到了 $D(1) D(2) ; \dots D(9)$ 等参数的数值.

将这9个 $D(c)$ 值代入式(13),得到第1次信用等级划分结果的目标函数 $Z_D = 0.759$.

3.2.2 满足约束条件1和约束条件2的n次信用等级划分

由于设置满足约束条件1和约束条件2的信用等级划分迭代次数为1000,故重复上文步骤,可以划分出满足约束条件1和约束条件2的1000种信用等级划分结果,并得到1000个信用等级划分结果的目标函数 Z 值.

3.2.3 确定最终的信用等级

1) 目标函数 Z 值的确定

比较“3.2.2”中1000次信用等级划分结果的 Z 值,其中 $Z = 0.759$ 为最大,对应的信用等级划分结果即为最终划分的信用等级.

上述过程通过 Matlab 编程求解.

2) 信用等级得分临界点的反推

目标函数 $Z = 0.759$ 对应的最优信用等级即是计算出 $Z = 0.759$ 时假设的信用等级得分分割点.而这些分割点已在3.2.1中给出了来龙去脉,这些分割点划分的信用等级得分区间如表5第3列所示.因此,表5第3列~第5列就是本模型的最佳信用等级划分结果.

3.2.4 信用等级划分结果分析

为更直观地展示损失率和信用等级的对应关系,以信用等级为纵坐标,以每个信用等级对应的损失率为横坐标画出信用等级划分的金字塔图,如图3~图5所示.

由图3~图5可以看出,本模型划分的9个

信用等级符合“信用等级越高,损失率越低”的金字塔标准.

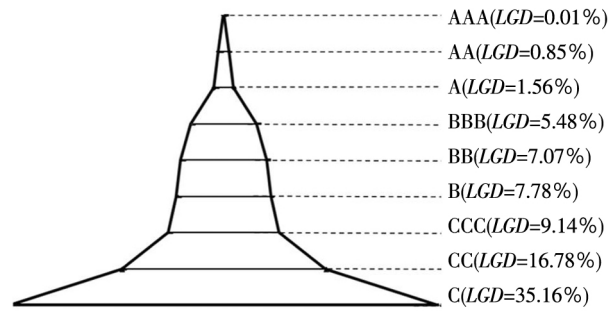


图3 本模型信用等级与损失率的金字塔图

Fig.3 Pyramid diagram of this model

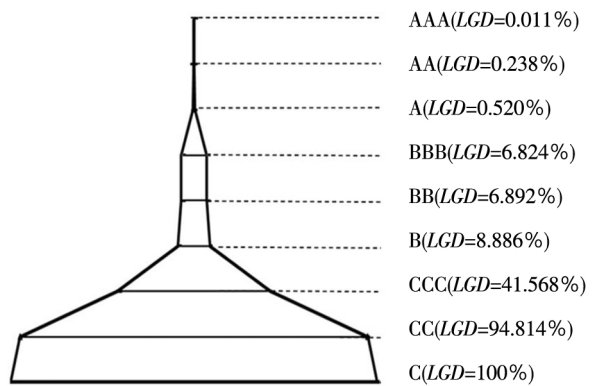


图4 对比模型1^[22]信用等级与损失率的金字塔图

Fig.4 Pyramid diagram of contrast model 1^[22]

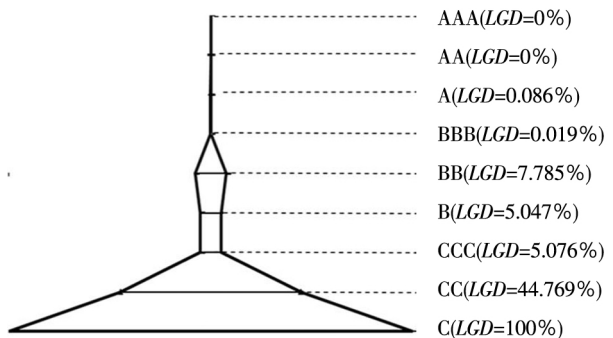


图5 对比模型2^[18]信用等级与损失率的金字塔图

Fig.5 Pyramid diagram of contrast model 2^[18]

3.3 对比分析

对比分析的数据采用3.1中的3045笔小企业的信用得分 S_j 、应收本息 L_j 、应收未收本息 R_j 数据,具体如表3所示.

对比分析的标准:一是“信用等级越高,损失率越低”的金字塔标准.若划分的信用等级不满足“信用等级越高,损失率越低”,则该信用等级划分模型失效;二是信用等级间损失率级差变化

稳定性标准. 信用等级间损失率级差应该满足式 (17), 避免损失率级差变化过于敏感, 导致损失率少许有一点变化, 就被划分为不同的信用等级的弊端.

应指出, 由于对比模型^[18 22]划分的信用等级特点, 无法计算目标函数 Z 值, 故不对衡量信用等级间违约鉴别能力的目标函数 Z 值进行对比.

3.3.1 对比模型 1

1) 与本模型的主要区别

目标函数的不同. 对比模型 1 来自于文献 [22], 它以 8 个信用等级临界点对应的信用评分差值之和最大为目标, 能确保信用评分差异大的客户被划分到不同信用等级. 本模型以划分的信用等级间的违约鉴别能力 $D(i)$ 均值最大为目标, 可确保划分的信用等级能最大程度的区分违约可能性不相似的客户.

约束条件的不同. 对比模型 1 仅有“信用等级越高, 损失率越低”约束, 没有本模型中信用等级间损失率级差的约束, 这样虽能确保越高的信用等级, 损失率越低, 然而不能避免信用等级间损失率级差变化过于敏感的弊端.

信用等级划分的算法不同. 对比模型 1 首先设置第 1 个信用等级的临界点, 然后在第 1 个信用等级的临界点上每次增加 1 个客户, 找到符合约束条件的 1 次信用等级划分. 通过依次改变第 1 个信用等级临界点, 得到不同的信用等级划分结果, 从中找到满足目标函数的信用等级划分结果. 这种算法虽然能找到合理的信用等级划分结果, 然样本数特别多的时候, 可能会导致第 1 个信用等级临界值取值范围特别大, 通过取遍第 1 个信用等级所有临界值情况以找到信用等级划分的全局最优解, 势必导致信用等级划分时间较长. 本模型运用带约束的随机算法划分信用等级, 在样本数多的情况下, 也能较快找到最优信用等级划分结果.

2) 模型结果与分析

将表 3 第 2 列 ~ 第 4 列数据代入对比模型 1^[22]中, 得到对比模型 1 的信用等级划分结果, 如表 5 第 6 列和第 7 列所示.

表 5 第 8 列结果为根据表 5 第 7 列数值和式 (16) 计算得到.

结果表明: 一是本模型和对比模型 1 均符合信用等级越高, 损失率越低标准. 由图 3 本模型金字塔图和图 4 对比模型 1 的金字塔图可以明显看出, 随着信用等级的升高, 损失率严格递减.

本模型不同信用等级间的损失率级差变化更稳定. 由表 5 第 7 列第 4 行的 BBB 信用等级损失率为 6.824%, 第 5 行的 BB 信用等级损失率为 6.892%, 可看出, 对比模型 1 的损失率少许有一点变化, 就被划分为不同的信用等级了, 此时这两个信用等级的损失率级差仅为 0.068%, 过于敏感. 而本模型由于设置了信用等级间损失率级差的约束, 避免了损失率级差过于敏感的弊端.

3.3.2 对比模型 2

1) 与本模型的主要区别

对比模型 2 是以“中间信用等级户数多, 两头信用等级户数少”为标准进行信用等级划分^[18], 它能确保大量的客户集中在中间信用等级, 少量的客户分布在最高和最低的信用等级.

然由于其没有以损失率为标准进行信用等级划分, 不能确保划分的信用等级满足信用等级越高, 损失率越低的标准.

2) 模型结果与分析

将表 3 第 2 列数据代入对比模型 2^[18], 得到对比模型 2 的信用等级划分结果, 如表 5 第 9 列所示. 表 5 第 10 列结果为根据表 5 第 9 列数值和式 (14) 计算得到. 表 5 第 11 列结果为根据表 5 第 10 列数值和式 (16) 计算得到.

由表 5 第 9 列 ~ 第 11 列可以看出: 基于对比模型 2^[18]划分的信用等级中, AAA、AA 信用等级的损失率均为 0, 因为其不以损失率为信用等级划分标准, 致使划分结果出现信用等级发生了变化, 而损失率却没有变化的荒谬现象, 导致信用等级划分失效, 不能有效地区分客户风险.

基于对比模型 2^[18]划分的信用等级中, A 信用等级的损失率 0.086%, 大于较低信用等级 BBB 的损失率 0.019%. 同样, BB 信用等级的损失率 7.785%, 大于较低信用等级 B 损失率 5.047% 和信用等级 CCC 的损失率 5.076%, 也即信用等级高, 损失率也高, 这不符合“信用等级越高, 损失率越低”的信用评级本质属性, 显然不合

理,这是由于其不以“信用等级越高,损失率越低”为信用等级划分标准所致。

因此这种不以损失率为标准的模型划分信用等级,会出现信用等级发生了变化、而损失率却没有变化、或者出现随着信用等级的升高,损失率反而也升高的荒谬现象。

而如前所示,本模型则能确保划分的信用等级符合“信用等级越高,损失率越低”的金字塔标准。

4 结束语

4.1 主要结论

1) 信用等级划分应该满足信用等级间违约鉴别能力最大的原则,也即划分的信用等级能够最大程度的区分不同客户的信用风险,避免信用好的客户评级偏低,信用差的客户评级偏高的弊端。

2) 信用等级划分应该满足信用等级越高,损失率越低的客观实际,以避免不同信用等级的损失率相同、或者高信用等级的客户损失率反而高的荒谬现象。

3) 以中国某商业银行的 3 045 笔小企业贷款

数据为样本进行实证,结果表明,本模型划分的信用等级不仅满足信用等级越高、损失率越低标准,还能确保划分的信用等级能将不同违约可能性的客户更大程度区分开。

4.2 主要创新

根据信用等级越能把违约可能性大小不相似的客户区分开,则信用等级的违约鉴别能力就越大思路,以非违约客户累计频率与违约客户累计频率之差的绝对值最大值的代数和最大为目标,以后一个信用等级的损失率均严格大于其前面信用等级的损失率为约束 1,以后一个损失率级差 $\Delta LR_{i,j+1}$ 为前一个级差 $\Delta LR_{i-1,j}$ 的 $[a,b]$ 倍为约束 2,建立非线性目标规划模型划分信用等级,确保了划分的信用等级满足“信用等级越高,损失率越低”且能尽可能区分开违约可能性不同的客户标准,改变了大部分现有模型不能严格保证划分的信用信用等级同时满足上述标准的弊端。

提出了通过设置随机分割点的区间来划分信用等级的新算法,避免随机赋予信用等级分割点时,靠前的信用等级分割点落到靠后面的客户中,导致后面信用等级无论怎样划分均划分不出来的弊端,减少了无效的信用等级划分次数。

参 考 文 献:

- [1]张海洋. 信用评级影响很大[N]. 环球时报,2002,01—07(10).
Zhang Haiyang. Credit ratings have a big impact[N]. Global Times,2002,01.07(10). (in Chinese)
- [2]孙祁祥,刘宇飞,段 誉. 美国主权信用评级下调的影响[J]. 中国金融,2011,(16): 45-46.
Sun qixiang, Liu yufei, Duan yu. The impact of us sovereign credit rating downgrading[J]. China Finance,2011,(16): 45-46. (in Chinese)
- [3]Chai N N, Wu B, Yang W W, et al. A multicriteria approach for modeling small enterprise credit rating: Evidence from China[J]. Emerging Markets Finance and Trade,2019,55(11): 2523-2543.
- [4]Shi B F, Chen N, Wang J. A credit rating model of microfinance based on fuzzy cluster analysis and fuzzy pattern recognition: Empirical evidence from Chinese 2,157 small private businesses[J]. Journal of Intelligent and Fuzzy Systems,2016,31(6): 3095-3102.
- [5]张大斌,周志刚,许 职,等. 基于差分进化自动聚类的信用风险评价模型研究[J]. 中国管理科学,2015,23(4): 39-45.
Zhang Dabin, Zhou Zhigang, Xu Zhi, et al. Study on credit risk assessment model based on automatic clustering using an differential evolution algorithm[J]. Chinese Journal of Management Science,2015,23(4): 39-45. (in Chinese)
- [6]张洪祥,毛志忠. 基于多维时间序列的灰色模糊信用评价研究[J]. 管理科学学报,2011,14(1): 28-37.
Zhang Hongxiang, Mao Zhizhong. Research of multidimensional time series credit evaluation based on gray-fuzzanalysis

- model[J]. *Journal of Management Sciences in China*, 2011, 14(1): 28–37. (in Chinese)
- [7] Zhang Y J, Chi G T. A credit rating model based on a customer number bell-shaped distribution[J]. *Management Decision*, 2018, 56(5): 987–1007.
- [8] 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: ZL201210201461.6. 2015. 08. 19.
Chi Guotai, Shi Baofeng. Credit rating system and method based on the matching of credit rating and default loss rate[P]. Chinese Patent: ZL201210201461.6, 2015.08.19. (in Chinese)
- [9] Krink T, Paterlini S, Resti A. The optimal structure of PD buckets[J]. *Journal of Banking and Finance*, 2008, 32(10): 2275–2286.
- [10] Basel Committee on Banking Supervision. The new basel capital accord[R]. Basel: Bank for International Settlements, 2006.
- [11] 琚春华, 邹江波, 傅小康. 融入在线社会资本的个人信用价值度量模型[J]. *管理科学学报*, 2017, 20(11): 114–126.
Ju Chunhua, Zou Jiangbo, Fu Xiaokang. Credit value measurement model incorporating personal social capital[J]. *Journal of Management Sciences in China*, 2017, 20(11): 114–126. (in Chinese)
- [12] 庞素琳, 何毅舟, 汪寿阳, 等. 基于风险环境的企业多层交叉信用评分模型与应用[J]. *管理科学学报*, 2017, 20(10): 57–69.
Pang Sulin, He Yizhou, Wang Shouyang, et al. Multilayer crossing credit scoring models for enterprise and application based on risk environment[J]. *Journal of Management Sciences in China*, 2017, 20(10): 57–69. (in Chinese)
- [13] 余乐安, 汪寿阳. 基于核主元分析的带可变惩罚因子最小二乘模糊支持向量机模型及其在信用分类中的应用[J]. *系统科学与数学*, 2009, 29(10): 1311–1326.
Yu Lean, Wang Shouyang. Least squares fuzzy support vextor machine methodology with variable penalty factors for credit classification[J]. *Journal of Systems Science and Mathematical Sciences*, 2009, 29(10): 1311–1326. (in Chinese)
- [14] Bai C G, Shi B F, Liu F, et al. Banking credit worthiness: Evaluating the complex relationships[J]. *Omega*, 2019, 83: 26–38.
- [15] Lyra M, Paha J, Paterlini S, et al. Optimization heuristics for determining internal rating grading scales[J]. *Computational Statistics & Data Analysis*. 2010, 54(11): 2693–2706.
- [16] Serrano-Cinca C, Gutiérrez-Nieto B. The use of profit scoring as an alternative to credit scoring systems in peer-to-peer (P2P) lending[J]. *Decision Support Systems*, 2016, 89(C): 113–122.
- [17] 陈洪海. 基于信息敏感性的小企业信用评级模型研究[D]. 大连: 大连理工大学, 2016.
Chen Honghai. Small enterprise credit rating model based on information sensitivity [D]. Dalian: Dalian University of Technology, 2016. (in Chinese)
- [18] 迟国泰, 潘明道, 齐菲. 一个基于小样本的银行信用风险评级模型的设计及应用[J]. *数量经济技术经济研究*, 2014, (6): 102–116.
Chi Guotai, Pan Mingdao, Qi Fei. A credit rating model for analyzing bank customers based on small sample[J]. *The Journal of Quantitative and Technical Economics*, 2014, (6): 102–116. (in Chinese)
- [19] Zhi H Y, Yang Z Y. Research on credit rating of SMEs based on combination evaluation[C]// 2011 International Conference on Business Management and Electronic Information, 2011: 661–664.
- [20] 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: 20–84.
Chen Yanqiu. Research on evaluation and decision of small amount loans for farmers based on support vector machines [D]. Dalian: Dalian University of Technology, 2011: 20–84. (in Chinese)
- [21] 迟国泰, 李鸿禧, 潘明道. 基于违约鉴别能力组合赋权的小企业信用评级——基于小型工业企业样本数据的实证分析[J]. *管理科学学报*, 2018, 21(3): 105–126.
Chi Guotai, Li Hongxi, Pan Mingdao. Small enterprises credit rating based on default identification ability of combination weighting: Based on an empirical analysis of small industrial enterprises [J]. *Journal of Management Sciences in China*, 2018, 21(3): 105–126. (in Chinese)

- [22]赵志冲,迟国泰,潘明道. 基于信用差异度最大的信用等级划分优化方法[J]. 系统工程理论与实践, 2017, 37(10): 2539–2554.
Zhao Zhichong, Chi Guotai, Pan Mingdao. Optimal method of credit rating division based on maximum credit difference degree[J]. Systems Engineering: Theory and Practice, 2017, 37(10): 2539–2554. (in Chinese)
- [23]李战江. 租赁和商务服务业小企业的信用评价研究[D]. 大连: 大连理工大学, 2014.
Li Zhanjiang. Research on credit evaluation for small enterprises of leasing and business services[D]. Dalian: Dalian University of Technology, 2014. (in Chinese)
- [24]薛 薇. 统计分析与 SPSS 的应用[M]. 北京: 中国人民大学出版社, 2011.
Xue Wei. The Application of Statistical Analysis and SPSS[M]. Beijing: China Renmin University Press, 2011. (in Chinese)
- [25]Yu S L, Chi G T, Jiang X. Credit rating system for small businesses using the K-S test to select an indicator system[J]. Management Decision, 2018, 57(1): 229–247.
- [26]Shi B F, Meng B, Yang H F, et al. A novel approach for reducing attributes and its application to small enterprise financing ability evaluation[J]. Complexity, 2018, 2018(2): 1–17.
- [27]中国人民银行. 中国人民银行信用评级管理指导意见[R]. 银发[2006]95号. 2006.03.02.
People's Bank of China. The people's bank of China's credit rating management guidelines[R]. People's Bank of China issued[2006]95. 2006–03–02. (in Chinese)

Credit rating division method with maximum default identification capability

CHI Guo-tai¹, YU Shan-ti^{2,3}

1. School of Economics and Management, Dalian University of Technology, Dalian 116024, China;
2. Postdoctoral Research Station, Financial Research Institute, The People's Bank of China, Beijing 100800, China;
3. Inter Bank Market Clearing House Co., Ltd, Shanghai 200002, China

Abstract: Credit rating is to differentiate the risk of different customers. However, most existing methods either fail to guarantee that the credit rating meets the criteria “the higher the credit rating, the lower the loss rate”, or fail to differentiate the customers with different default possibilities to the greatest possible degree. Therefore, the existing credit rating method cannot be used as an effective tool for loan decisions. This study gives credit grades according to the above criteria. The innovations are: firstly, credit grades are divided by taking the maximum difference between the cumulative frequency of non-defaulting customers and that of defaulted customers as the objective function and taking “the higher the credit rating, the lower loss rate” as the main constraint, so that the credit grades meet the above dual criteria; Secondly, a new algorithm for credit rating by setting the intervals of random segmentation points is proposed. The algorithm avoids that problem that a front division point may be designated to customers in the back, resulting in the failure of grading customers in the back. Finally, this study uses 3 045 small businesses from a Chinese bank for empirical study, and the results show that the credit grades thus derived meet the above dual criteria.

Key words: credit grade division; loss rate; discriminating ability; K-S test statistics; credit rating