China P2P Online Loan Platform Risk Assessment and Risk Zoning

—An Empirical Study Based on Entropy Weight-TOPSIS and Information Gain

Lijun Cai^{1,a*} Guanhua Zhao^{2,b}

¹ Financial Department, Jiangsu University, Zhenjiang, Jiangsu, 212013, China

² Professor, School of Accounting, Shandong University of Finance, Jinan, Shandong, 250000, China

^{a*} cailijun88@163.com

^b sczgh2006@sina.com

ABSTRACT

China's P2P online loan industry has entered a period of stable development. However, due to external supervision and internal management irregularities, the problem platform has occurred from time to time, causing huge losses to the vast number of investors and seriously disrupting the market order of Internet finance. Therefore, how to effectively identify and evaluate the risks of existing P2P platform, and then effectively partition the risk level, so that the risk prevention can be targeted, and the probability of counter-regional and systemic financial risks is reduced, which becomes one of the important topics in the field of Internet finance. This paper adopts the Entropy Weight-TOPSIS method, using 628 problem platforms and 1127 normal platforms accumulated from 2015 to 2018 as research samples. This paper constructs the risk assessment system of P2P platform from the five aspects of platform operation foundation, executive background, profitability, risk control ability and netizen public opinion, and evaluates the risk influencing factors and risk levels of P2P online lending platform in China. Then the information gain method is used to partition the risk assessment results of the P2P platform, and finally the effect of the partition is evaluated.

Keywords: P2P online loan platform, Entropy Weight-TOPSIS, Information Gain, Risk assessment, Risk zoning.

中国 P2P 网贷平台风险评估与险情分区

——基于熵权-TOPSIS 和信息增益的实证研究

蔡丽君^{1, a*} 赵冠华^{2, b}

¹ 江苏大学财务处, 镇江, 江苏, 中国

² 山东财经大学会计学院, 济南, 山东, 中国

^{a*} cailijun88@163.com

^b sczgh2006@sina.com

摘要

我国的 P2P 网贷行业现已进入一个平稳发展期, 但由于外部监管和内部经营管理不规范等原因, 导致问题平台频频发生, 给广大的投资者造成了巨大损失, 严重扰乱了互联网金融的市场秩序。因此, 如何对现有运行的 P2P 平台风险进行有效的识别和评估, 进而对风险水平进行有效分区, 使风险的防范能够有的放矢, 减少反生

区域性、系统性金融风险的概率,就成了互联网金融领域研究的重要课题之一。本文采用熵权-TOPSIS 方法,利用 2015 年~2018 年累计发生的 628 家问题平台和 1127 家正常平台为研究样本,从平台的运营基础、高管背景、盈利能力、风控能力以及网民舆情五个层面构建了 P2P 平台的风险评估体系,并对我国的 P2P 网贷平台风险影响因素、风险水平进行评估,利用基于信息增益的方法对 P2P 平台风险评估结果进行险情分区,最后对分区的效果进行了评价。

关键词: P2P 网贷;平台熵权-TOPSIS 信息增益风险评估险情分区

1. 引言

P2P 网络借贷是互联网技术与民间借贷相结合的一种金融创新模式,是互联网金融的重要形式(韩丰霞等,2017)。2005 年,以 Zopa、Lending Club、Prosper 为代表的借贷平台相继在欧美出现。2007 年以来,P2P 网贷平台在中国兴起并快速发展,截至 2015 年,P2P 网贷行业已初具规模,2016 年,行业进入快速发展期。但 P2P 网贷行业由于监管政策缺失、征信系统不完善以及经营管理不规范等原因,频频出现了平台卷款跑路、提现困难、信息泄露、黑客入侵等事件。2016 年之前的中国 P2P 网贷行业发展可谓是“野蛮生长、竞争无序、监管缺位、风险倍增”(姜琪,2018)。为了控制行业风险,规范市场秩序,促进有效竞争,国家监管部门相继出台了一系列有关规定。2017 年 2 月,银监会发布《网络借贷资金存管业务指引》,8 月下发《网络借贷信息中介机构业务活动披露指引》。随着这些行业准则的颁布,我国的 P2P 问题平台数量显著减少,平台风险得到了一定的遏制。2018 年以来,P2P 网贷行业进入平稳发展期。但如何对现有运行的 P2P 平台风险进行有效的识别和评估,进而提出防控策略,就成了学者们探讨的热点。

为此,国内外学者对 P2P 平台风险进行了一系列研究。评估方法层面:现有文献主要有主观赋权法和客观赋权法两类。主观赋权法是根据决策者主观信息进行赋权的一类方法,如二项系数法、层次分析法等(孙晨辉等,2014;宋琳,2017;周莉等,2017)。运用该方法确定的各评估指标权重,反映了决策者的意向,决策或评估结果会打上主观烙印,导致同一家 P2P 评级结果相差很大,投资者不仅对评级结果感到困惑,有些甚至产生误导。客观赋权法是依据一定的规则对评估对象进行自动赋权的一类方法,如熵权法、因子分析法、CRITIC 法等(孙玥璠等,2015;王伟,2016;尹钧惠,2016;张高胜,2017)。它是从原始数据出发,从样本中提取信息,根据各个指标提供的信息量大小或各指标间的相互关系来确定权重,不依赖于人的主观判断,有较强的数学理论依据;评估对象层面:现有文献主要以特定平台上借款人的“信用风险”作为研究对象,以“平台风险”作为研究对象的实证研究

仍属空白(叶青等,2016)。在我国目前 P2P 平台众多、鱼龙混杂的格局下,仅仅关注借款人的“信用风险”是远远不够的,投资者面临的首要问题是平台选择问题。同时,我国政府部门要求防范风险的重点也主要是“平台风险”(叶青等,2016),但现有文献对 P2P“平台风险”评估进行系统性的研究明显不足;评估指标层面:现有文献大多采用平台的“运营情况”和借款人的“信用状况”等“硬指标”(王赛芳,2016;叶青,2016;孙宝文,2016;刘洋等,2017),缺乏对“平台网络舆情”等“软指标”的相关研究。

综上所述,已有文献对 P2P 平台风险进行了系统的研究,但也存在一些不足。首先,现有文献鲜有将客观赋权与 TOPSIS 结合,更未见到有将信息增益技术应用于风险分区的研究;其次,现有文献缺乏对“平台网络舆情”等“软指标”的相关研究;再其次,现有文献大多将反映借款人的“信用风险”作为研究对象,将“平台风险”作为研究对象的较少;最后,现有文献选择的研究样本偏少,不足以支撑我国 P2P 平台整体风险影响因素以及风险分布的研究。为此,本文共搜集了 1755 家问题平台和正常平台作为研究样本,增加了“网民舆情”等软指标,用熵权法确定各个评估指标的权重,用 TOPSIS 法对 P2P 平台风险进行排序,再用信息增益技术对风险排序结果进行险情分区。

2. P2P 平台风险评估与险情分区度量方法—熵权-TOPSIS 和信息增益

2.1. 熵权-TOPSIS 法原理

香农首次提出著名的信息熵公式,即:

$$H = -\sum_{i=1}^m p_i \ln p_i \quad (1)$$

式中, H 为信息熵, p_i 表示系统每种状态出现的概率, $i=1, 2, \dots, m$ 。

熵权法评价正是基于信息熵的原理,若把各评估指标看作是系统的每种状态,则当指标的信息熵越小,数据间的差异越大,所能提供的信息也就越多,在评估过程中应赋予较大的权重;反之,指标的信息熵越大,数据间的差异不明显,所能提供的信息就越小,

相应赋予较小权重（赵腾，2019）。

TOPSIS (Technique for Order Preference by Similarity to an Ideal Solution) 是一种逼近理想解的排序方法, 它通过计算评估对象到最优解、最劣解的距离进行排序。若评价对象距最优解最近、距最劣解最远, 则将其评为最佳值; 若评价对象距离最优解最远、最劣解最近, 则将其定为最差值。介于最佳值和最差值之间的若干评价对象, 根据其距离最优解或最差解的相对距离进行排序。所谓“最优解”是由各评估指标的最佳状态组成的理想解集, 也称“正理想解”; “最劣解”是由各评估指标最差状态组成的理想解集, 亦称“负理想解”。熵权-TOPSIS 法就是将熵权法的客观赋权与 TOPSIS 排序相结合, 从而实现 P2P 平台风险的评估。

2.1.1. 利用熵权法计算各指标的权重

第一步, 将初始评估矩阵标准化。设有 m 个评估对象, n 个评估指标, 则初始评估矩阵为:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \quad (2)$$

其中, x_{ij} ($i=1, 2, \dots, m; j=1, 2, \dots, n$) 表示第 i 个评估对象在第 j 项指标中的数值。

由于各指标的量纲单位均存在差异, 所以为消除因量纲不同对评估结果造成的影响, 需要对各指标进行标准化处理。一般采用极差变换法, 计算公式为:

$$x'_{ij} = \frac{x_{ij} - \max_i \{x_{ij}\}}{\max_i \{x_{ij}\} - \min_i \{x_{ij}\}} \quad (3)$$

$$x'_{ij} = \frac{\max_i \{x_{ij}\} - x_{ij}}{\max_i \{x_{ij}\} - \min_i \{x_{ij}\}} \quad (4)$$

其中, 公式(3)适用于“正向指标”, 公式(4)适用于“负向指标”。

第二步, 计算标准化评估矩阵的比重。第 i 个评估对象、第 j 项指标的比重为:

$$y_{ij} = \frac{x'_{ij}}{\sum_{i=1}^m x'_{ij}} \quad (5)$$

第三步, 计算第 j 项指标的信息熵。计算公式为:

$$e_j = -\frac{1}{\ln m} \sum_{i=1}^m y_{ij} \ln y_{ij} \quad (j=1, 2, \dots, n) \quad (6)$$

其中, $0 \leq e_j \leq 1$, 并规定当 $y_{ij} = 0$ 时, $y_{ij} \ln y_{ij} = 0$ 。

第四步, 计算第 j 项指标的权重。计算公式为:

$$w_j = \frac{1-e_j}{\sum_{j=1}^n (1-e_j)} \quad (j=1, 2, \dots, n) \quad (7)$$

2.1.2. 利用 TOPSIS 法计算各评估对象与最优解的接近程度 C

第一步, 初始评估矩阵趋同化处理。对初始评估矩阵(2)做趋同化处理, 即将“负向指标”转化为“正向指标”, 处理方法是按照公式(8)对“负向指标”取倒数, 并且可适当扩大或缩小一定比例来转换数据。

$$x'_{ij} = 1/x_{ij} \quad (8)$$

第二步, 趋同化数据规范化处理。对公式(8)按照下式做规范化处理:

$$z_{ij} = \frac{x'_{ij}}{\sum_{i=1}^m x'_{ij}} \quad (i=1, 2, \dots, m; j=1, 2, \dots, n) \quad (9)$$

第三步, 构建加权规范化矩阵。即将公式(7)计算出的指标权重 w_j 与公式(9)计算出的规范化矩阵 z_{ij} 相乘。

$$r_{ij} = w_j \times z_{ij} \quad (i=1, 2, \dots, m; j=1, 2, \dots, n) \quad (10)$$

第四步, 确定正、负理想解的集合 R^+ 和 R^- 。即将公式(10)计算出的矩阵 r_{ij} 每列中的最大值构成正理想解集合 R^+ , r_{ij} 每列中的最小值构成负理想解集合 R^- 。

$$R^+ = (R_1^+, R_2^+, \dots, R_n^+) \quad (11)$$

$$R^- = (R_1^-, R_2^-, \dots, R_n^-) \quad (12)$$

第五步, 计算各评估对象到正、负理想解的欧氏距离 D^+ 和 D^- 。具体计算如公式(13)、(14)所示。

$$D_i^+ = \sqrt{\sum_{j=1}^n (r_{ij} - R_j^+)^2} \quad (i=1, 2, \dots, m; j=1, 2, \dots, n) \quad (13)$$

$$D_i^- = \sqrt{\sum_{j=1}^n (r_{ij} - R_j^-)^2} \quad (i=1, 2, \dots, m; j=1, 2, \dots, n) \quad (14)$$

第六步, 计算各评估对象与“正理想解”的接近程度 C_i 。公式如下:

$$C_i = \frac{D_i^-}{D_i^- + D_i^+}, \quad 0 \leq C_i \leq 1 \quad (i=1, 2, \dots, m) \quad (15)$$

第七步, 对待评估对象排序。依据接近程度 C_i 大小对各评估对象进行降序排序, 确定评估结果。 C_i 值越大, 评估对象越接近“正理想解”, 评估对象越优, 风险越大; 反之, C_i 值越小, 评估对象越接近“负理想解”, 评估对象越劣, 风险越小。

2.2. 信息增益险情分区原理

本文的信息增益险情分区指的是利用前面计算出的各评估对象与“正理想解”的接近程度 C_i 值进行

¹ “正向指标”指对 P2P 平台风险起促进作用的指标, 方向为正; “负向指标”指对 P2P 平台风险起抑制作用的指标, 方向为负。

“有指导的分箱”，即将连续型变量值 C_i 离散化后变成若干个分区。这种分箱处理不是主观、随意的，而是基于分箱结果对被解释变量的解释能力最强即信息增益最大化这一指导原则下进行的，这一原则就是 MDLP 原则²。限于篇幅，这里仅给出该方法的计算步骤。

第一步，计算信息熵 $Ent(S)$ 。信息熵定义为：

$$Ent(S) = -\sum_{i=1}^k P(C_i, S) \log_2 P(C_i, S) \quad (16)$$

式中， S 为样本集合， C 为输出变量（被解释变量）， k 为输出变量 C 的取值类别数，本文的 P2P 平台除了问题平台就是正常平台，只有两种状态， $k=2$ ， $P(C_i, S)$ 为 C 取第 i 个类别时的概率。

第二步，计算信息增益 $Gains(A, T; S)$ 。计算公式为：

$$Gains(A, T; S) = Ent(S) - Ent(A, T; S) \quad (17)$$

式中， $Ent(S)$ 是分箱前的信息熵， $Ent(A, T; S)$ 是分箱后的信息熵，可通过公式 (18) 计算。

$$Ent(A, T; S) = \frac{|S_1|}{|S|} Ent(S_1) + \frac{|S_2|}{|S|} Ent(S_2) \quad (18)$$

式中， T 为分箱的组限， S_1 、 S_2 分别是将样本集 S 按照组限 T 分出的两个样本子集， $|S_1|$ 、 $|S_2|$ 、 $|S|$ 为样本集 S_1 、 S_2 、 S 的样本量。

第三步，将分箱“收益”与“成本”进行比较。计算公式为：

$$Gains(A, T; S) > \frac{\log_2(N-1)}{N} + \frac{\Delta(A, T; S)}{N} \quad (19)$$

式中， $Gains(A, T; S)$ 为分箱获得的“收益”即信息增益， $\frac{\log_2(N-1)}{N} + \frac{\Delta(A, T; S)}{N}$ 为分箱的“成本”， N 为样本总量，只有公式 (19) 成立即分箱“收益”大于“成本”时，分箱才有意义。其中， $\Delta(A, T; S) = \log_2(3^k - 2) - [kEnt(S) - k_1Ent(S_1) - k_2Ent(S_2)]$ 。上述计算过程不断重复进行，直至找到最佳组限 T 为止。

3. P2P 平台风险评估和险情分区实证分析

3.1. 样本选取与数据来源

本文从网贷之家共选取了截止到 2018 年 7 月 31 日正常运营的 P2P 平台共 1700 家，问题平台选取的

²MDLP (Minimum Description Length Principle) 即基于最短描述长度原则，是 1993 年由法亚德 (Fayyad) 和埃拉尼 (Irani) 提出的一种有指导的数据分箱方法。其基本思想是，如果分箱后的解释变量对被解释变量解释能力显著高于分箱之前，则这样的分箱才是有意义的，所以，待分箱变量应在被解释变量的指导下进行。

时间段是 2015 年-2018 年，四年累计问题平台数 1974 家，其中，2015 年底、2016 年底和 2017 年底发生的问题平台数分别是 867 家、557 家和 219 家，2018 年 7 月底发生的问题平台数 331 家。剔除数据不全和极端值样本，最终得到的正常平台 1127 家、问题平台 628 家，合计 1755 家，详见表 1。

表 1 研究样本选取情况表

平台类型	统计日期	原始平台个数	筛选后平台个数	原始累计平台个数	筛选后累计平台个数	筛选后累计平台占比(%)	合计
正常平台	截止 2018. 7. 31	1700	1127	1700	1127	64. 2	1755
问题平台	2015. 1～ 2015. 12	867	166	1974	628	35. 8	
	2016. 1～ 2016. 12	557	134				
	2017. 1～ 2017. 12	219	93				
	2018. 1～ 2018. 7	331	235				

3.2. 构建评估指标体系

由于 P2P 平台出现问题后，一些交易数据无法获得，所有能够获得的数据均来自网贷之家和网贷天眼，最终共选取了 18 个评估指标，如表 2 所示。

表 2 P2P 平台风险评估指标的选取及含义

指标大类	指标名称	指标含义	指标属性	指标符号
运营基础	注册资本 (x1)	平台依照法律流程在工商局登记注册的资本额 (单位: 万元) 取对数	正向指标	+
	平台运营时间 (x2)	平台成立到统计截止日或出现问题的有效运营时间 (单位: 月)	负向指标	-
	平台背景 (x3)	如果平台有风险投资、银行、国企或上市公司背景, x3=1, 否则 x3=0	——	——
	平台所处地域 (x4)	按照六大行政区域取值: 1: 华东; 2: 华南; 3: 西南; 4: 华北; 5: 西北; 6: 东北	负向指标	-
	人均 GDP (x5)	平台所在省份的人均 GDP 水平	正向指标	+
高管背景	高管人数 (x6)	平台高管团队中高管总人数	正向指标	+
	硕士以上人数 (x7)	平台高管团队中拥有硕士以上学历的人数	负向指标	-
	金融从业经验人数 (x8)	平台高管团队中拥有金融从业经验的人数	负向指标	-
盈利能力	参考收益% (x9)	平台的平均利率 (%)	正向指标	+
	投资期限 (x10)	平台投资起息日到结息日的平均时间	负向指标	-

		单位: 月		
风控能力	银行存管 (x11)	用户资金是否交由银行存管, 是 x11=1, 否则 x11=0	-----	-----
	监管协会 (x12)	平台是否加入某监管协会, 是 x12=1, 否则 x12=0	-----	-----
	保障模式加总 (x13)	平台有无引入第三方担保机构, 有则记 1 分, 无则记 0 分; 平台有无风险准备金, 有则记 1 分, 无则记 0 分; 平台有无平台垫付措施, 有则记 1 分, 无则记 0 分; x13 取以上三项指标加总衡量。	负向指标	-
网民舆情	网友点评 (x14)	网友对平台的综合评估, 数值为 1-5 分	负向指标	-
	提现 (x15)	数值区间在 1 ~ 5, 表示提现速度从超过三天到半小时之内的度量	正向指标	+
	站岗 (x16)	数值区间在 1 ~ 5, 表示资金站岗现象持续的时间从超过两周到一天内	正向指标	+
	服务 (x17)	数值区间在 1 ~ 5, 表示网民对平台服务质量的评估	负向指标	-
	体验 (x18)	数值区间在 1 ~ 5, 表示网民对网站整体体验度从非常不满意到非常满意	正向指标	+

由于平台背景 (x3)、银行存管 (x11) 和监管协会 (x12) 三个指标为二分类型变量 (非数值型变

量), 应从评估体系中剔除, 实际参加风险评估的指标为 15 个。

3.3. P2P 平台风险影响因素分析

TOPSIS 风险评估中需要判断哪些指标是正向指标和负向指标, 为此, 需用 Logit 回归模型确定这些指标 (影响因素) 的显著性水平和方向。P2P 平台风险的影响因素很多, 考虑到数据的可获取性, 结合研究需要, 本文从平台的运营基础、高管背景、盈利能力、风控能力、网民舆情五个方面共选取了 18 个指标进行影响因素分析, 详见表 2。

由于被解释变量是二分类型变量, 本文构建 Logit 回归模型进行 P2P 平台风险的影响因素分析。公式为:

$$P = \frac{1}{1 + \exp(-(\beta_0 + \sum_{i=1}^k \beta_i x_i))} \quad (20)$$

该式是一个 Sigmoid 函数, 其中, x_i 为解释变量 ($i=1, 2, \dots, k$), β_0 为常数项, β_i 为解释变量 x_i 的系数, P 为事件发生的概率, $P \in [0, 1]$, 如果计算出的概率值 $P > 0.5$, 被解释变量预测为 1 类, 对应问题平台, 否则, 如果 $P \leq 0.5$, 被解释变量预测为 0 类, 对应正常公司。

将筛选后获得的 1755 个研究样本和 18 个评估指标 (见表 1、表 2) 构建一个 1755*18 的数据集, 利用 Clementine 软件进行 Logit 回归, 结果见表 3。

表 3 P2P 平台风险影响因素 Logit 回归结果

变量属性	变量名称	系数	Wald 统计量	P 值	预期符号	实际符号
运营基础	注册资本 x1	0.495***	18.184	0.000	-	+
	平台运营时间 x2	-0.064***	83.868	0.000	-	-
	平台背景 x3	-0.348	1.947	0.163	-	-
	平台所处地域 x4	0.544	0.609	0.505	+	+
	人均 GDPx5	0.427	1.355	0.244	-	+
高管背景	高管人数 x6	0.840***	77.419	0.000	-	+
	硕士以上人数 x7	-0.117	0.885	0.347	-	-
	金融从业经验人数 x8	-2.000***	230.256	0.000	-	-
盈利能力	参考收益 x9	0.253***	61.492	0.000	+	+
	投资期限 (月) x10	-0.085**	4.244	0.039	-	-
风控能力	银行存管 x11	0.416*	3.429	0.064	-	+
	监管协会 x12	-0.479	2.574	0.109	-	-
	保障模式加总 x13	-0.191	1.641	0.200	-	-
网民舆情	网友点评 x14	-1.006***	10.325	0.001	-	-
	提现 x15	0.063	0.075	0.784	+	+
	站岗 x16	0.711**	6.606	0.010	+	+
	服务 x17	-1.119**	6.642	0.010	-	-
	体验 x18	0.017	0.002	0.968	-	+

注: *表示在 10%水平上显著; **表示在 5%水平上显著; ***表示在 1%水平上显著。

从表 3 可以看出，（1）平台运营基础：平台运营时间 x2、平台背景 x3 与平台风险负相关，但后者不显著。平台所处地域 x4 与平台风险正相关，但不显著。注册资本 x1、人均 GDPx5 与平台风险正相关，但后者不显著，与预期符号相反。这可能是注册资本雄厚和经济越发达地区的平台往往疏于内部管理，导致风险增加；（2）平台高管背景：硕士以上人数 x7、金融从业经验人数 x8 与平台风险负相关，前者不显著。高管人数 x6 与平台风险正相关，与预期符号相反。这可能是高管人数越多的平台，意见难以统一，出现“三个和尚没水喝”的现象，导致平台风险增加；（3）盈利能力：参考收益 x9 和投资期限 x10 分别与平台风险显著正相关和负相关；（4）风控能力：监管协会 x12、保障模式加总 x13 与平台风险负相关，但不显著，银行存管 x11 与平台风险显著正相关，与

预期符号相反。这可能是加入银行存管的平台，反而不重视内部管理，增加了平台风险；（5）网民舆情：网友点评 x14 和服务 x17 与平台风险显著负相关，提现 x15、站岗 x16 和体验 x18 分别于平台风险正相关，但体验 x18 与预期符号相反。

3.4. 风险评估过程

3.4.1. 利用熵权法求出各个评估指标的权重

首先，将 1755 个研究样本和 15 个评估指标构建一个 1755*15 的数据矩阵，其次，对该矩阵标准化处理、计算指标比重和信息熵，最后，按照公式 (7) 计算指标的权重，所有计算均采用 Matlab 编程实现，计算结果如表 4 所示。

表 4 熵权法各指标权重计算表

目标层	第一指标层	第一指标层权重	第二指标层	第二指标层权重	总权重
P2P 平台风险评估	运营基础 (YY1)	0.1090	注册资本(x1)	0.1149	0.0125
			平台运营时间(x2)	0.2264	0.0247
			平台所处地域(x4)	0.3085	0.0336
			人均 GDP(x5)	0.3503	0.0382
	高管背景 (GG1)	0.4816	高管人数(x6)	0.1700	0.0819
			硕士以上人数(x7)	0.7726	0.3721
			金融从业经验人数(x8)	0.0575	0.0277
	盈利能力 (YL1)	0.0135	参考收益%(x9)	0.5906	0.0080
			投资期限(x10)	0.4094	0.0055
	风控能力 (FK1)	0.1669	保障模式加总(x13)	1.0000	0.1669
	网民舆情 (WM1)	0.2291	网友点评(x14)	0.1964	0.0450
			提现(x15)	0.1772	0.0406
			站岗(x16)	0.1300	0.0298
			服务(x17)	0.3676	0.0842
			体验(x18)	0.1288	0.0295

3.4.2. 利用熵权-TOPSIS 法对平台风险进行评估

首先，利用 (8)、(9) 式对初始评估矩阵 (2) 做趋同化处理和规范化处理，其次，将规范化矩阵 (9) 与公式 (7) 计算出的权重相乘构建加权规范化矩阵 (10)，再其次，确定正、负理想解集合并计算各个

评估对象与正负理想解间的距离，最后，计算各评估对象与正理想解的接近程度 C (详见表 5)，再将 C 值按照降序排序。其中，排序名次靠前的是那些 C 值较大的 P2P 平台。 C 值越大，评估对象相对于正理想解的接近程度越大，风险就越大，通过对 C 值降序排序最终实现 P2P 平台风险的评估，如表 5 所示。

表 5 熵权-TOPSIS 法距离正、负理想解距离及正理想解接近程度计算表³

序号	平台名称	是否 问题平台 Y	距正理想解 距离 D^+	距负理想解 距离 D^-	正理想解 接近程度 C
1	稳赚宝	1	0.0115	0.0118	0.5068
2	58 车贷	1	0.0114	0.0119	0.5115
3	招财猫理财	1	0.0115	0.0118	0.5065
4	金储宝	1	0.0090	0.0139	0.6077
5	票票喵	1	0.0145	0.0076	0.3443
6	温商贷	1	0.0148	0.0073	0.3308
7	钱保姆	1	0.0147	0.0075	0.3371
8	链金所	1	0.0091	0.0138	0.6032
9	中投融	1	0.0115	0.0118	0.5067
10	财蜂发财树	1	0.0089	0.0139	0.6097
...	...	...	...	...	...
1746	义贷网	0	0.0147	0.0073	0.3323
1747	汇贷天下	0	0.0096	0.0138	0.5879
1748	陇 e 贷	0	0.0118	0.0118	0.4990
1749	今金贷	0	0.0146	0.0076	0.3415
1750	妙银理财	0	0.0156	0.0034	0.1804
1751	财富观	0	0.0163	0.0021	0.1126
1752	如意财富	0	0.0165	0.0017	0.0951
1753	沃联财富	0	0.0117	0.0118	0.5033
1754	海享贷	0	0.0165	0.0018	0.0965
1755	沃顿金服	0	0.0163	0.17	0.948

³限于篇幅，表 5 仅列出前 10 家和后 10 家 P2P 平台的距离正、负理想解距离 D^+ 、 D^- 和正理想解接近程度 C 的计算值。

3.5. 险情分区过程

利用风险评估过程中构建的 1755*15 的数据矩阵, 采用 Clementine 中的“字段选项/分箱”功能, 以表 5 中的正理想解接近程度 C 为分箱变量, 以“是

否问题平台”为分箱的指导变量, 最终将 C 分为五个区, 对应的险情分区域分别命名为: 高风险区、风险区、灰色区、正常区高值和正常区, 表 6 给出了部分 P2P 平台的险情分区结果。

表 6 熵权-TOPSIS 法 P2P 平台险情分区结果计算表

风险降序 排序序号	平台名称	是否问 题平台	正理想 解接近 程度 C	分区样本 个数及占 比	险情分区组 限	风险区问题 平台个数及 占比	风险区正常 平台个数及 占比	险情分 区
1	敬友财富	1	0.6941	178	C≥0.6068	124	54	高风险 区
2	商赢金融	1	0.6928					
3	慧达财富	1	0.6877	10.1%		19.8%	4.7%	
4	盛世典当	1	0.6802					
5	博源 e 贷	1	0.6653					
...	...	...	...	...	...	...	...	...
291	车聚宝	0	0.6067	137	0.5851≤ C < 0.6068	31	106	风险区
292	萌小薪	1	0.6066					
293	聚宝珠	0	0.6066	7.8%		5.0%	9.3%	
294	车来贷	1	0.6062					
295	知合金服	0	0.6062					
...	...	...	...	...	...	...	...	...
415	万财金服	1	0.5806	326	0.5091≤ C < 0.5851	232	94	灰色区
416	汇银融通	0	0.5637					
417	三湘金融	1	0.5616	18.6%		36.9%	8.2%	
418	智慧树金融	1	0.5615					
419	宏量财富	1	0.5562					
...	...	...	...	...	...	...	...	...
1566	稳通金融	0	0.509	810	0.1185≤ C < 0.5091	206	604	正常区 高值
1567	马甲袋	0	0.509					
1568	工程惠贷	0	0.509	46.2%		32.8%	53.5%	
1569	一休理财	0	0.509					
1570	利得行	1	0.5089					
...	...	...	...	...	...	...	...	...
1566	赢贷通	0	0.1184	303	C< 0.1185	35	269	正常区
1567	一点成圆	0	0.1183					
1568	大拇指财富	0	0.118	17.3%		5.5%	24.8%	
1569	满益网	0	0.118					
1570	橙天金融	0	0.1179					

注: 限于篇幅, 表 6 仅列出每一个分区的前五名的相关数据。

由表 6 可以看出, 风险排名越靠前的平台, C 值越大, 评估对象距正理想解距离越小, 风险越大, 反之, 排名越靠后, C 值越小, 平台风险越小。处于“正常区”和“正常区高值”的样本占总样本的比例为 64%, 与表 2 中统计的正常平台占比 64.2% 十分接近。剩下的 36% 样本分别分布在“高风险区”、“风险区”和“灰色区”, 也与表 2 统计的问题平台占比十分接

近。从各个分区正常平台和问题平台的占比来看, 在 1127 个正常样本中, 有 4.7% 的正常样本落在了“高风险区”, 9.3% 的正常样本落在了“风险区”, 还有 8.2% 的正常平台落在了“灰色区”。说明了目前还在正常经营的平台中, 有 22.2% 即 1/5 多的正常平台存在不同程度的风险。从表 6 还可以看出, 分别有 32.8% 和 5.5% 的问题平台分别处于“正常区高值”和“正

常区”，这些平台从评估指标上看与正常平台并无明显差异，但可能因为某种外部突发事件比如平台大面积违约导致资金链断裂，从而使平台倒闭或提现困难。

为了说明熵权-TOPSIS 法在 P2P 网贷平台风险评估中的应用效果，这里以普通 TOPSIS 法做为参照，比较两种方法在几个层面上的应用效果，详见表 7。

4. P2P 平台风险评估和险情分区应用效果评价

表 7 熵权-TOPSIS 和 TOPSIS 在 P2P 平台风险评估中的应用效果比较

熵权-TOPSIS 法					TOPSIS 法				
分区样本个数及占比	险情分区组限	风险区问题平台个数及占比	风险区正常平台个数及占比	险情分区	分区样本个数及占比	险情分区组限	风险区问题平台个数及占比	风险区正常平台个数及占比	险情分区
178	$C \geq 0.6068$	124	54	高风险区	290	$C \geq 0.0968$	260	30	高风险区
10.10%		19.80%	4.70%		16.50%		41.50%	2.70%	
137	$C \geq 0.5851, C < 0.6068$	31	106	风险区	124	$C \geq 0.0902, C < 0.0968$	73	51	风险区
7.80%		5.00%	9.30%		7.10%		11.60%	4.50%	
326	$C \geq 0.5091, C < 0.5851$	232	94	灰色区	1151	$C \geq 0.0582, C < 0.0992$	285	866	灰色区
18.60%		36.90%	8.20%		65.60%		45.30%	76.80%	
810	$C \geq 0.1185, C < 0.5091$	206	604	正常区高值	190	$C < 0.0582$	10	180	正常区
46.20%		32.80%	53.50%		10.80%		1.60%	16.00%	
303	$C < 0.1185$	35	269	正常区					
17.30%		5.50%	24.80%						

(1) 从险情分区个数来看，熵权-TOPSIS 法将整个险情区域分为五个区，TOPSIS 法将整个险情区域分为四个区。这主要是因为熵权-TOPSIS 法在求接近程度 C 时考虑了各个指标对评估权重的不同，而 TOPSIS 法将各个指标的权重视为同一个值。因此，熵权-TOPSIS 法对 P2P 平台风险的评估结果更科学，分区更明细，也更合理；(2) 从风险区问题平台占比来看，熵权-TOPSIS 法的高风险区和风险区问题平台占比为 25%，TOPSIS 法的高风险区和风险区问题平台占比为 53%，前者与表 2 的实际问题平台占比 36% 更接近；(3) 从正常区正常平台占比来看，熵权-TOPSIS 和 TOPSIS 分别为 78% 和 16%，前者与表 2 的实际正常平台占比 64% 更接近。进一步说明无论是风险评估还是险情分区，熵权-TOPSIS 法效果更好；(4) 从风险区正常平台占比来看，熵权-TOPSIS 为 14%，TOPSIS 为 7%，说明我国正常运营的 P2P 平台中尚有一定比例的平台处于风险区和高风险区，随时有可能出现提现困难、卷款跑路等现象，这与我国 P2P 平台的整体运营情况相符。

5. 研究结论与启示

本文的研究结果表明：第一，有“国资”背景的、有银行存管的以及经济发达地区的平台，平台出现问

题的风险反而加大；第二，高管中硕士以上学历和金融从业经验人数越多，平台风险越小，而高管人数的增加反而加大了平台风险；第三，网友对平台的各项评价恰如其分的反映了平台的风险水平，网民舆情与平台风险有显著关联性；第四，从险情分区视角，由于熵权-TOPSIS 法在求接近程度 C 时考虑了各个指标对评估权重的不同，因此，对 P2P 平台风险的评估结果更科学，分区更接近现实，也更合理；第五，从风险区问题平台占比来看，熵权-TOPSIS 法的高风险区和风险区问题平台占比与实际问题平台占比更接近，证实了该法应用于 P2P 平台风险评估更有效；第六，从正常区正常平台占比来看，熵权-TOPSIS 的评估结果也与实际正常平台占比更接近，进一步印证了上述结论；第七，从风险区正常平台占比来看，有一定比例的正常平台被划归为风险区，说明我国正常运营的 P2P 平台中仍有一定比例的平台处于风险区和高风险区，随时有可能出现提现困难、卷款跑路等现象，这与我国 P2P 平台的整体运营情况相符。

本文的研究结果具有一定的政策启示。第一，重点加强“国资”背景和经济发达地区平台的监管力度，强化平台内部控制和治理水平；第二，从平台高管团队层面，应重点引进学历层次高以及具有金融从业经验的人员，而不是盲目扩大高管队伍人数；第三，外

部监管部门和平台内部应及时关注网民舆情，从媒体关注的视角判定平台风险的大小及走向；第四，监管部门在对平台风险进行评估时，应考虑各个评估指标权重的不同，采用客观赋权法与评估模型相结合才能对平台风险进行更加科学、合理的评估；第五，对那些尚在正常运营的平台，监管部门也应加强防控和监管，采用外部监督和内部控制相结合的方式，防范P2P平台风险，将平台风险扼杀在萌芽状态，减少发生区域性、系统性金融风险的概率。

项目基金

本研究为国家社科基金《金融大数据分析视角下P2P网贷平台风险识别、评估及防控路径研究》（17BJL078）的阶段性成果之一

REFERENCES

- [1] Han Fengxia, Xiao Hanjie et al. Research on the driving force of green finance development under the New normal of economy -- Based on the tripartite game relationship between government, banks and enterprises [J]. Economics and Management Review, 2017 (5).
- [2] Jiang Qi. Research on efficiency difference and Trading Volume influencing Factors of P2P online loan platform in China [J], Quantitative & Technical Economics Research, 2018 (6).
- [3] Sun Fan, Bao Na, Zhang Ruolan. Risk assessment of material misstatement under information environment based on entropy weight method [J]. Friends of Accounting, 2015 (7).
- [4] Wang Wei. Risk assessment of P2P online lending platform in China -- Based on entropy weight method and CRITIC method [J]. Finance Theory and Practice, 2016 (12).
- [5] Zhang GAosheng. Construction of credit risk assessment model for small and micro enterprises -- Based on factor analysis-entropy weight assessment model [J]. Finance and Accounting Communications, 2017 (17).
- [6] Sun Chenhui, LI Fufu. Judgment and evaluation of private financial risk based on AHP and fuzzy comprehensive evaluation [J]. Economic Management, 2014 (2).
- [7] Song L. Research on risk assessment of P2P network loan industry based on fuzzy analytic Hierarchy Process [J]. Dongyue Cong, 2017 (10).
- [8] Zhou Li, ZHANG Kai, QIU Yulian. Audit risk assessment of P2P online loan platform based on fuzzy comprehensive evaluation [J]. Finance and Accounting Communications, 2017 (9).
- [9] Yin Junhui, XIONG Ranran. Internal risk assessment of P2P online lending platform based on factor analysis [J]. Monthly Finance and Accounting, 2016 (33).
- [10] Ye Qing, LI Zengquan, XU Weihang. Research on risk identification of P2P online lending platform [J]. Accounting Research, 2016 (6).
- [11] Wang Saifang. Analysis of characteristics of Online loan problem platform in China based on mean comparison and T-test [J]. Financial Development Research, 2016 (9).
- [12] Sun Baowen, NiuChaoqun, ZHAO Xuankai, Jing Wenjun. Financial distress identification: Research on risk characteristics of P2P platforms in China [J]. Journal of Central University of Finance and Economics, 7th issue, 2016.
- [13] Liu Yang, Wang Huizhan. Risk assessment and control of Private lending in China from the perspective of Internet: A case study of P2P platform [J]. Macroeconomic Research, 2017 (3).
- [14] Zhao Teng, Yang Shizhong. Application of entropy weight TOPSIS method in enterprise financial risk assessment: A Case study of Jiujiu Liquor Company [J]. Finance and Accounting Monthly, 2019 (3).