

被中国知网 (CNKI) 收录集刊

Journal of Corporate Finance 公司金融研究

济南大学金融研究院

2015卷第2辑 总第8期

中国P2P信贷市场中的借贷行为监管研究
——基于前景理论的PPDai数据挖掘

薪酬激励对过度投资的影响：CEO特质的调节作用

基于供应链视域下的生产运营计划制订和管控研究

金融系统性风险研究综述与反思

掠夺性信贷与金融消费者保护

区域金融创新对农业技术创新的影响分析
——以山东省17地市的面板数据为例

公司金融研究

(2015 卷第 2 辑 总第 8 期)

济南大学金融研究院



中国金融出版社

责任编辑：肖丽敏

责任校对：孙蕊

责任印制：陈晓川

图书在版编目 (CIP) 数据

公司金融研究 (Gongsì Jinrong Yanjiu). 2015 卷第 2 辑. 总第 8 期/济南大学金融研究院. —北京: 中国金融出版社, 2015. 6

ISBN 978 - 7 - 5049 - 7330 - 6

I. ①公… II. ①济… III. ①公司—金融—文集 IV. ①F276. 6 - 53

中国版本图书馆 CIP 数据核字 (2015) 第 187969 号

出版

发行

中国金融出版社

社址 北京市丰台区益泽路 2 号

市场开发部 (010)63266347, 63805472, 63439533 (传真)

网上书店 <http://www.chinafph.com>

(010)63286832, 63365686 (传真)

读者服务部 (010)66070833, 62568380

邮编 100071

经销 新华书店

印刷 北京市松源印刷有限公司

尺寸 185 毫米 × 260 毫米

印张 9

字数 185 千

版次 2015 年 6 月第 1 版

印次 2015 年 6 月第 1 次印刷

定价 30.00 元

ISBN 978 - 7 - 5049 - 7330 - 6/F. 6890

如出现印装错误本社负责调换 联系电话 (010) 63263947

编 委 会

主 编：路士勋

执行主编：孙国茂

副 主 编：董新兴

学术委员会主席：夏 斌

学术委员会委员(以汉语拼音为序)：

郭田勇：中央财经大学

贺 强：中央财经大学

胡金焱：山东大学

胡汝银：上海证券交易所

施光耀：中国上市公司市值管理研究中心

韦 森：复旦大学

王培志：山东财经大学

吴晓求：中国人民大学

夏 斌：国务院研究中心

易宪容：中国社会科学院

张卫国：山东省社会科学院

朱武祥：清华大学

编 委 会 主 任：范跃进

编 委 会 委 员(以汉语拼音为序)：

蔡先金 陈月辉 董新兴 范跃进

葛金田 贾 岩 李光红 路士勋

卢欣笙 孙国茂 邢乐成 岳云龙

封面题字：王天义

目录

盛浙湘 尹优平 盛 辉

中国 P2P 信贷市场中的借贷行为监管研究
——基于前景理论的 PPDai 数据挖掘

● 1

刘井建 郭文丽 纪丹宁

薪酬激励对过度投资的影响：CEO 特质的调节作用

● 24

董 鹏 吴炬文

基于供应链视域下的生产运营计划制订和管控研究

● 43

童中文 魏歆七 邹 静

金融系统性风险研究综述与反思

● 58

赵 煊

掠夺性信贷与金融消费者保护

● 69

苟成娟 李永平

区域金融创新对农业技术创新的影响分析
——以山东省 17 个地市的面板数据为例

● 80

孙国茂

互联网金融：本质、现状与趋势

● 91

朱相平

宏观经济因素影响下的上市公司失败：基于 1993—2010 年的实证分析

● 120

中国 P2P 信贷市场中的借贷行为监管研究

——基于前景理论的 PPDai 数据挖掘*

盛浙湘 尹优平 盛 辉^①

【摘要】 本文从前景理论的微观视角出发, 基于国内首家 P2P (个人对个人) 纯信用无担保网络借贷平台——PPDai. com 进行借贷行为的数据挖掘, 结果表明, 影响放贷行为的首要因素是借款人的历史违约率, 但超额补偿会导致资金涌向高违约风险的借款; 此外, 借款人的“硬信息”(hard information, 是指能用准确的硬指标表示的信息, 是正式的、精准的、符合逻辑的、可追溯的, 如业绩报告、任务指标、财务报表等) 能直接地造成贷款人的非理性投资行为, 借款人的社交网络信息能间接地造成贷款人的非理性投资行为。

【关键词】 前景理论 P2P 信贷 数据挖掘 社交网络

Customer Behavior on China P2P Lending Platform -
From the view of Prospect theory and data mining on PPDai

Sheng Zhexiang Yin Youping Sheng Hui

Abstract: This paper investigates the priority factors of lender's decisions in peer-to-peer lending from the view of prospect theory. We apply data mining methods to identifying the priority of various factors. We find that borrower's historical default ratio is

* 本文感谢美国宾夕法尼亚州立大学黑泽尔分校 Eric Lu 副教授及 PPDai. com 提供的数据支持。

① **【作者简介】** 盛浙湘 (1989—), 男, 浙江平湖, 美国伊利诺伊大学香槟分校统计学硕士, 供职于美国第一资本金融集团。

尹优平 (1970—), 男, 山西新绛人, 金融学博士, 研究员, 供职于中国人民银行金融消费者权益保护局。

盛辉 (1961—), 男, 湖南益阳人, 金融学学士, 供职于中国人民银行平湖市支行。

paramount among all influential variables, and the priority factors of a lender's bidding decision are borrower's personal behavior, hard information, and borrower's social network. Maximum entropy algorithm verifies these results and proves the robustness of our model.

Keywords: Prospect Theory P2P Lending Data Mining Social Network

一、引言

网络 P2P 信贷是微观金融中一种重要的新型借贷模式。根据 Greiner 和 Wang 在 2009 年给出的定义, P2P 信贷指不以第三方金融机构为媒介, 而直接通过互联网发生个人对个人借贷和还款的融资行为, 它的出现使个人借贷双方可以不通过银行等金融机构直接发生借贷关系, 大大加速了金融资本的流动性, 丰富了小微信贷的资金渠道。目前在实业界, 我国已涌现出诸如拍拍贷、人人贷、你我贷等一批网络 P2P 信贷平台; 然而相关学术研究的展开与监管框架的形成却相当滞后。这种理论与实践的失衡可能直接导致形成错误的政策导向与机制框架, 因此对于我国现有的 P2P 信贷市场进行研究具有重要的学术价值和政策意义。

我国现有的 P2P 平台与国际上流行的 P2P 平台存在异质性, 即我国的 P2P 平台除了承担普通金融机构的借贷媒介角色外, 其中大部分还承担了担保职能, 甚至有一部分还开展了资金池业务。因此, 在我国的 P2P 实践中, 其功能和定位相当于异化的金融机构。尽管如此, 近期中国银监会要求网络 P2P 平台必须把握四个边界: 明确平台的中介性、明确平台本身不得提供担保、不得搞资金池、不得非法吸收公众存款。可见在政策导向上, 监管方希望网络 P2P 平台的主要职责仍是扮演借贷媒介角色, 使得个人借贷者能够在平台上发生互相间的直接借贷行为。因此对平台上借贷行为进行研究能够有助于了解 P2P 平台的运作机理, 发现目前 P2P 平台存在的问题。

网络 P2P 信贷平台的运行机制是通过在线竞拍实现的, 主要有两种方式: 第一种方式为投标金额达到设定的借款金额时, 竞标过程即刻终止; 第二种方式为在借款人设定的竞标期限内, 放贷人可持续投标, 利率低者胜出。完成竞标后, 借贷双方通过签署具有法律效应的电子合同保障资金借入者和资金借出者的利益, 同时借贷平台可收取一定比例的中介费, 借款人则在未来规定期限内以等额本息方式分期还款。目前我国的 P2P 平台主要采用第二种交易方式, 并且在竞标过程中平台会实时显示各借款请求的已有投标金额、投标次数、完成百分比等信息, 成为后续竞标人的重要参考指标。

上述交易机制保证了网络 P2P 信贷平台相比传统借贷的进入门槛更低,自由程度更高,但是,P2P 的低门槛准入也使得投资者面临比向金融机构借贷款更高的违约风险。因此,关于借款人违约风险和违约行为的研究也成为 P2P 借贷研究的重中之重:一方面在宏观上,违约率较低的平台交易量更大,市场信息的影响更显著,金融权益保护的渠道也更通畅(Sheng et al., 2014);另一方面在微观上,违约率较低的平台呈现出更理性的投资行为(Zhang and Liu, 2012; 陈冬宇等, 2012)。Greiner 和 Wang (2009) 基于全球交易规模最大的 P2P 信贷平台 Prosper 的数据发现,美国 Prosper 市场的整体违约率^①为 35%; 相比而言, Lu 等 (2012) 针对我国最大的 P2P 信贷平台 PPDai 的随机抽样数据发现,中国 PPDai 市场的整体违约率接近 50%。该数据说明,我国 P2P 市场的参与者面临更高的违约风险,其投资决策的收益也更为不确定。因此,本文针对我国 P2P 信贷市场中的借贷行为进行数据挖掘,以期找到违约风险不确定的情形下影响理性投资行为的关键因素,为制定 P2P 信贷的金融行为监管政策和保护框架提供学术依据。

对 P2P 信贷平台违约风险和理性投资行为的已有研究主要分为两类。第一类是基于借款人视角研究其违约行为,如 Iyer (2009) 以及 Lin 等 (2012) 基于美国市场的研究发现,P2P 市场的信息可分为“硬信息”(公开易得、易定量)和“软信息”(公开但难以捕捉,难以定量)两类,并且不单诸如信用等级、借款金额、性别、种族等“硬信息”会决定借款人是否违约(Berger and Gleisner, 2007; Kumar, 2007),诸如社交网络等“软信息”也会显著地影响借款人的违约行为(Lin, et al., 2012; Puro, et al., 2010)。Sheng 和 Lu (2012, 2013) 针对中国市场做了一系列的研究发现,在中国 P2P 市场上不仅社交网络等“软信息”会显著影响借款人的违约行为,并且社交网络中关系的紧密程度直接决定了影响程度的强弱。第二类是基于放贷人视角研究其理性投资行为,如 Zhang 等学者 (2012) 发现美国市场存在理性的“羊群效应”,即有较多已有投标的借款项目获得后续投标的概率也越大,并且这种“羊群效应”会给借贷双方带来显著的正效应。而陈冬宇等 (2012) 对中国市场的研究则表明,我国的 P2P 借贷市场存在非理性的“羊群效应”,即投标次数不能显著降低借款违约率。

尽管已有研究从借款人的视角分析了其违约行为,从放贷人的视角研究了其理性投资行为,但是上述研究均未对三个根本性问题作出回答。第一个问题是影响违约行为和理性投资行为的机制是什么。因为定量研究的方法限制,以往的研究多从

^① 根据已有研究,P2P 平台的违约率计算以每一单交易为基本单位。如 A 的一笔借款由 B、C、D、E 四人获得竞标(共 4 单交易),其中 B、C、D 违约,而 E 不违约,则违约率计为 75%。

宏观的视角测度出群体性行为,并以此作为解释个体行为方式的机制。然而事实上相同的群体性行为并不代表该群体中的个体都具有相同的行为逻辑。以“羊群效应”为例,当投资者因为获得相似的决策信息而做出趋同性的理性决策行为时,也会在宏观上表现出“伪羊群效应”,但是这种群体性行为的发生机制与实质的“羊群效应”却是截然不同的。因此,基于宏观视角的定量测度来解释个体的行为机制存在一定的逻辑缺陷。第二个问题是不同的影响变量之间的优先级如何。传统的定量研究往往给定若干条件不变的情况下,测度特定变量的影响是否显著。但是这种测度的逻辑是人为假定了待测度变量会对应变量造成影响,事实上当变量足够多的情形下,即使在统计意义上该变量显著,也不能说明该变量影响了投资决策。第三个问题是以往的研究没有将借款人的违约行为与放贷人的理性决策行为放到一个统一的框架下进行分析。

针对上述三个问题,本文从前景理论的视角出发,运用数据挖掘的方法测度我国 P2P 信贷市场中借贷行为的影响机制。前景理论(Kahneman and Tversky, 1979)认为在风险不确定的 P2P 市场中,放贷人的理性决策取决于该笔投资的预期可得收益与目标收益之间的关系:若预期收益高于目标收益则选择投资,反之则放弃投资。即前景理论认为市场信息对投资决策影响的传导机制为:投资者行为不直接受市场信息影响,而是市场信息通过改变预期投资收益与目标收益之间的关系,从而间接地影响了投资者行为。在风险不确定环境下对于投资收益与目标收益大小关系的判断,投资者不再使用绝对收益进行衡量,而是通过一个两阶段的估值行为将所有不确定事件纳入一个可比较的框架中:第一阶段为编辑阶段(editing phase),核心是通过已有信息确定期望目标收益(参照点),即机会成本;第二阶段为估值阶段(evaluating phase),核心是评估随机发生概率下,预期收益的期望价值。具体而言,在参照点不变的情形下,借款人还款违约率的上升会降低预期投资收益,从而降低放贷人的投资概率;另外,在投资预期收益不变的情形下,参照点的上升也会降低放贷人的投资概率。基于此,前景理论将借款人的违约行为和放贷人的理性决策行为放到了一个统一的研究框架下,通过借款人的违约行为改变贷款人的收益预期,从而有效地解释了违约行为与理性投资行为的运作机理。此外,数据挖掘的方法避免了传统量化研究的先验设定,该方法不假定投资者是否理性,不假定何种变量在决策中的优先级更高,仅通过行为数据进行挖掘,因而在此类已有研究不足的探索性问题上适用性更强。

本文余下部分的结构如下:第二部分为理论框架与数据构建,第三部分为模型设计,第四部分为实证结果,第五部分为讨论分析与稳健性检验,第六部分为结论与政策建议。

二、理论框架与数据构建

(一) 理论框架

前景理论是一种解释在风险不确定情形下人为判断和决策的理论，最早由 Kahneman 和 Tversky 于 1979 年首先提出。他们发现预期效用理论不能有效解释在风险不确定情形下的投资者行为，特别是下述五种被大量行为实验验证的情形：1) 确定效应，即在确定收益和不确定收益之间，理性人偏好确定收益；2) 反射效应，即在确定损失和可能不损失之间，理性人会尝试可能不损失的路径；3) 风险规避，即等量损失带来的负效应超过等量收益带来的正效应；4) 赌徒心理，当有绝对超额收益时，即使发生概率很低仍有人愿意参与；5) 参照依赖，平均收入 4 万元时拥有 5 万元收入的心理满足程度超过平均收入 7 万元时拥有 6 万元的心理满足程度。

基于上述情形，Kahneman 和 Tversky 将前景理论刻画为一双阶段模型。第一阶段为编辑阶段 (editing phase)，核心是通过已有信息确定目标收益 (参照点)。第二阶段为估值阶段 (evaluating phase)，核心是评估预期收益的总价值。总价值可表示成单项事件的个体价值函数与相应决策权重函数的组合，其数学表达式为：

$$V(x, p; y, q) = \pi(p)v(x) + \pi(q)v(y)$$

其中， $v(\cdot)$ 是价值函数相较参照点的收益或亏损，一般在收益时表现为凹函数 (concave)，在损失时表现为凸函数 (convex)，且损失区间的斜率大于收益区间。 π 为决策权重， p 为给定参照点约束下收益的概率， q 约束下损失的概率。Kahneman 和 Tversky (1992) 又将该形式推广到连续形式，以累积加权函数替代离散决策权重函数 (见图 1)。从数学角度来看，若以参考值为原点，则在参照点附近当收益 (黑

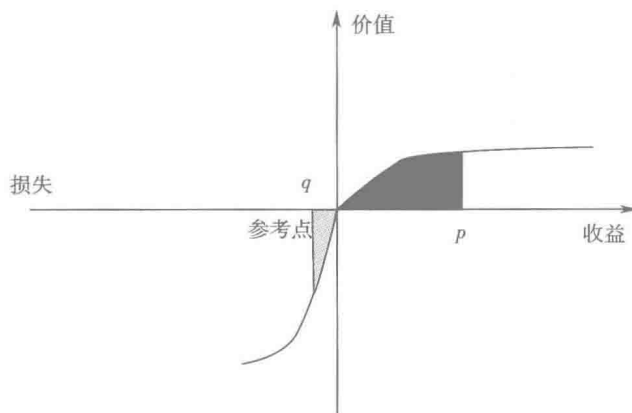


图 1 前景理论的收益

色)部分积分面积超过损失(灰色)部分面积时,选择投资;反之则选择不投资。而实际操作中,由于投资机会均为离散形式,因此往往采用累加形式来表达积分面积。

上述离散和连续的情形均被行为实验证明在风险未知的情形下有效(Laughunn and Crum, 1984; Lowenstein, 1988; Schroeder, 2005)。特别地,在金融领域,当一个理性投资者面临诸多不确定选择时,他会选择可投资金约束下超过目标收益的那些项目。具体而言,在P2P信贷市场中,一个潜在放贷人会面临诸多投资机会,在编辑阶段放贷人会根据已有市场信息确立自身的目标收益(参考点),然后在估值阶段判断某一特定借款的预期收益是否大于目标收益。显然,最终决策可以视为一个0-1属性变量,即预期收益大于目标收益则理性选择应为参与放贷($Y=1$),否则理性决策为不参与放贷($Y=0$)。

(二) 数据构建

1. 应变量构建

根据上述理论,放贷决策取决于两个因素,即心理参考值的确定与预期收益的确定(Sheng et al., 2014),其中本文采用的心理参考值为某一特定时点,某一特定投资者在P2P及其他市场所能获得的各类投资收益的平均值,即参考值 $R(I_M | T = t)$ 是在给定 t 时刻,市场可得平均收益 I_M 的函数;本文采用的预期收益为某一笔特定借款的收益与该借款不违约的乘积,即预期收益 $ER[R_c, P(D^c)]$ 是合约约定还款利率 R_c 与不违约概率 $P(D^c)$ 的函数。综上,我们得出理性决策的表达式:

$$\begin{cases} Y = 1, ER[R_c, P(D)^c] > R(I_M | T = t) \\ Y = 0, ER[R_c, P(D)^c] \leq R(I_M | T = t) \end{cases} \quad (1)$$

其中, R_c 为借款合同签订时的利率,易从数据库中直接获得,因此关键在于确定不违约概率。需要注意的是,本文在构造最终违约状态的变量和预测违约概率时使用了两种不同的方法。即,最终违约状态的变量是一个非随机的确定值,借款人在违约还款日未按时偿还合同规定的本息即视为违约;而在现实中预测违约概率时,我们无法观测到未来时点的真实情况,因此我们往往基于借款人的历史违约率进行预测。特别地,已有研究已经证明在P2P信贷平台上,借款者的违约行为不单单取决于其自身因素,还受到如图2所示的社交网络的影响(Lu, et al., 2012)。

该社交网络共分为四个层级。第一级为某个特定借款人 i ;第二级为该借款人发出的第 j 笔借款;因为P2P平台大多采用分期付款的方式,故第三级为该 j 笔借款所处的第 t_i 个到期还款日,在每个独立的到期还款日,该借款人 i 都可选择是否对该笔借款 j 违约。因此,在每个 i, j, t 所对应的社交网络未必相同。例如,借款人 i 在 t_1 到 t_2 之间结识了新朋友 k ,则 k 会对 i 在 t_2 的决策产生影响,并不会对 t_1 的决策产

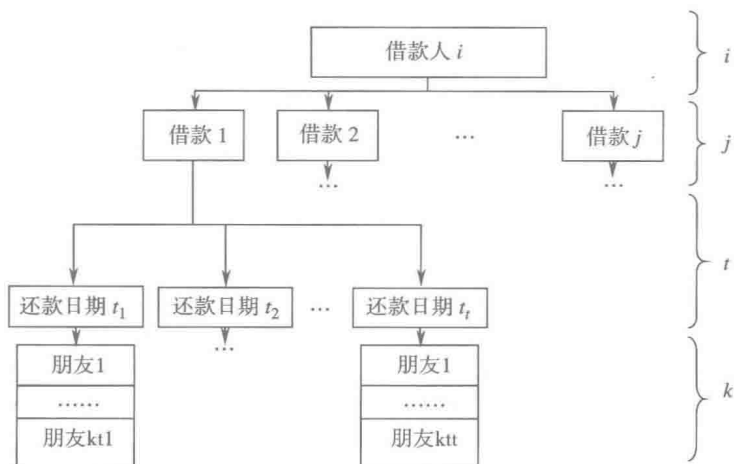


图2 影响违约行为的社交网络

生影响。因此，该四层次的社交网络模型会借款人是否违约产生影响，从而影响了本文的违约概率预测模型。

对于放贷人，通常需要预测某特定借款人的违约行为，从而做出理性的投资决策。由于借款合同与借款人具有一一对应的关系，因此放贷人所面临的违约概率预测模型具有如图3所示结构：

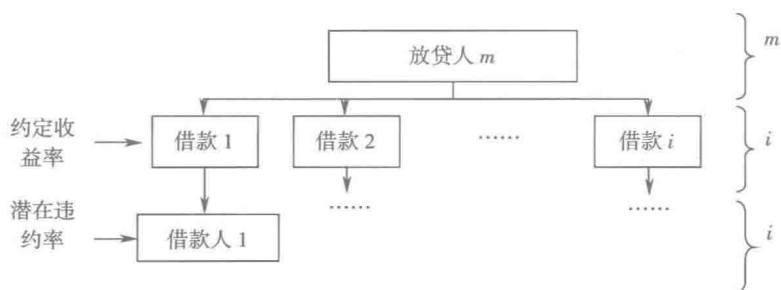


图3 放贷人决策树

上述的讨论确定了构建，而对于另一决定因素——参考值的确定，通常有三种常用指标：同期存款利率、同期贷款利率、同期债券利率。由于我国为了限制民间高利贷，对个人放贷有严格的监管，因此同期贷款利率并不是一个合理的指标。同期债券利率对于美国市场较为常用，因为其具有很高的流动性和较大的交易规模，而在我国对投资金额较小的个人投资者而言，债券通常不是理财组合的首选。相比而言，个人投资者选择银行定期存款的规模远远大于债券，因此采用与 P2P 借款期限相同的同期存款利率来构造中国 P2P 市场的参考值相对合理。此外，P2P 市场有别于其他市场的显著特征是该市场的参与者可以选择投资其他借款合同。由于考虑

到搜寻成本和估值成本,一个个体投资者难以评估出市场最佳投资作为其参考值,一般只能对市场信息有一个大致的掌握,因此我们采取该时点的市场历史投资收益均值作为参考值,即投资者的参考值函数:

$$R(I_M | T = t) = \max[IR_{deposit} \overline{IR}_{P2P} \times \overline{P}(D^c)]$$

(2)

为 P2P 市场平均预期收益与同期存款利率中较大的一个。

2. 自变量选择

根据上述论述,贷款人的理性决策行为取决于对借款人违约行为的预测,而在 P2P 平台中借款人的违约行为不单单取决于其“硬信息”,行为数据,也取决于社交网络的影响。因此,本文选择我国首家 P2P(个人对个人)纯信用无担保网络借贷平台——PPDai(拍拍贷)作为考察对象。PPDai 成立于 2007 年。截至 2013 年 8 月末,PPDai 注册用户超 160 万户,总交易规模达 7 亿元人民币。

PPDai 不但公布了包括根据自身信用评价算法得出的信用等级,还公布了年龄、性别、教育背景、婚姻状况等借款人信息;同时也披露了诸如借款金额、设定利率、还款期限、总竞标数、已有竞标金额等借款信息;诸如还款次数、还款金额等还款信息;此外,PPDai 还公布了社交网络信息,并将社交网络分为线上朋友和线下朋友两类,其中线上朋友包括 P2P 朋友与其他朋友,线下朋友包括亲属、密友、同事、同学、普通朋友、好友等。基于此,本文的自变量选择分为三个层次(见表 1):

表 1

自变量

硬信息 (H)	借款人行为 (B)	社交网络 (S)
设定利率	历史违约率	社交网络是否有违约记录
借款金额	是否存在违约记录	社交网络违约率
投标金额	还款次数	密友违约率
有效投标金额	借款目的	好友违约率
竞标数量		
完成竞标比例		其他朋友违约率
竞标利率		P2P 朋友违约率
市场基准利率		一般朋友违约率
教育背景		
孩子数量		
年龄		
性别		

续表

硬信息 (H)	借款人行为 (B)	社交网络 (S)
是否已婚		
借入信用得分		
放贷信用得分		
总信用等级		
是否手机验证		
是否视频验证		
是否学历验证		
担保方式		

注：本文随机抽取了 2010 年 9 月至 2012 年末的交易数据作为研究对象。

三、模型设计

由于本文的研究目的是为了测度我国 P2P 市场中借贷行为的影响机制，因此在此部分，我们选取两种数据挖掘的方法：Logistic 分类法和随机森林（random forest）。其中，Logistic 分类法用于测度该 P2P 市场的行为是否符合前景理论假定；随机森林算法用于找出在前景理论假定下，影响决策的变量优先级。

（一）Logistic 分类法

相比其他数据挖掘的方法，Logistic 模型不但能给出分类结果，还能同时测度待分类变量（应变量）与影响变量（自变量）之间的关系。因此，该方法提供了一种直观的途径去观测各影响变量是否符合前景理论假定及各自的经济意义。本文采用逐步建模的方法，在基础模型中仅考虑硬信息的影响（模型 1），测度借款人的财务指标与验证信息是否会对放贷人的竞标决策产生显著影响。即，检验 1）财务指标是否是信贷决策的显著信号；2）诸如性别、教育背景等属性变量是否能有效地将借款人分为不同的组别。

在此基础上，我们在模型中加入借款人行为（模型 2）。加入借款人行为变量后，本文主要考察两个方面：1）行为变量的系数大小与显著性，即借款人行为是否能显著影响放贷人的决策；2）模型 1 中的系数变化，即行为变量的加入是否会削弱“硬信息”的影响。为了进一步考察行为变量是否符合前景理论的基本设定，本文采用半参数 Logistic 模型（模型 3）来测度行为变量的非线性特征，即在模型 3

中对“硬信息”保持线性约束，而不对行为变量设定线性约束，以观察该 P2P 市场是否存在“低风险—低利率”的保守投资（确定效应）和“高风险—高收益”的激进投资（赌徒风险）。

由于 P2P 市场不同于一般投资市场的主要特征为 P2P 市场存在社交网络的影响。尽管已有研究验证了这种影响不但存在于美国市场（Lin et al.，2009，2012），也同样存在于中国市场（Lu et al.，2012，2013），但是已有研究的结论均是基于未知投资者是否理性得出的。即，即使研究者认为此类社交网络因素会影响理性放贷人的投资决策，在实际竞标过程中非理性放贷人未必将其视为影响因素。而本文的设定规避了需要事先假定投资者是否理性的这一约束，仅通过前景理论构造出纯理性的决策数据（包含所有社交网络因素），而后通过数据挖掘的方法对真实数据逐一做出理性或非理性的分组，以此判断各个投资者是否遵从了理性投资决策（模型 4）。同样的，本文还对社交网络因素进行了非线性影响的考察，并对比不同亲密程度的社交网络对违约行为的影响方式有何异同（模型 5，见表 2）。

表 2 模型设计

模型	Logistic	随机森林
1	$\text{logit}(\text{bid}) = H\beta_H$	H
2	$\text{logit}(\text{bid}) = H\beta_H + B\beta_B$	H + B
3	$\text{logit}(\text{bid}) = H\beta_H + s(B)$	H + B + S
4	$\text{logit}(\text{bid}) = H\beta_H + B\beta_B + S\beta_S$	
5	$\text{logit}(\text{bid}) = H\beta_H + S(B) + s(S)$	

(二) 随机森林

尽管本文通过 Logistic 模型已经基本确定了理性决策的潜在影响因素，但由于在实践中存在信息搜寻成本与估值成本，一个个体放贷人不可能对所有影响因素都相同对待，因此，本文还需通过其他数据挖掘的方法来考察 P2P 市场中影响因素的优先级排序，为信息披露机制的建立提供可靠的参考。在数据挖掘中，分类树和随机森林是比较有效的可以直接考察指标重要性排序的两种方法，但由于分类树模型存在不稳健的固有缺陷，因此本文采用随机森林的方法对影响因素的重要性进行排序，为了保证稳健性，本文构造的随机森林大小为 1000。为了保证前后的可比性，随机森林方法同样采用逐步加入影响因素的建模思想。在模型 1 中仅包含“硬信息”，在模型 2 中加入借款人行为，最后考虑社交网络的影响（见表 2）。

四、实证结果

(一) 描述性统计量

本文的数据集包含 1300234 条数据, 平均借款额度为 11870.04 元。基于前景理论, 若预期收益超过参考值则理性投资者应当参与竞标, 理性竞标率为 59.93%。对于借款人的违约行为, 本文数据集显示最高的历史违约率为 10%, 最低的为 0。对于社交网络中所涉及的所有参与人, 信用等级分布在 B、C、D、E、HR 五个等级上。表 3 列出了主要变量的描述性统计量。

但由于计算机运算能力的限制, 我们对每种方法均采用随机抽样的方法, 每次从原数据集中抽取 13 万条, 并重复验证 10 次。根据机器学习的普遍约定, 我们将每次抽取的 13 万条数据中的 80% (104000 条数据) 作为训练集, 余下 20% (26000 条数据) 作为测试集 (Hastie, et al., 2011)。

表 3 主要变量的描述性结果

变量名	均值	标准差
用户信息		
性别 (1 = 女, 2 = 男)	1.30	0.4580
年龄	31.82	5.2159
孩子数	1.52	0.5897
放贷信用得分	93.17	18.4976
借款信用得分	29.23	15.06
是否已婚 (2 = 已婚, 1 = 未婚)	1.71	0.4548
借款信息		
借款金额	11870.04	9364.46
月偿还额度	2080.87	4212.6788
现行利率	16.66	3.4934
借款人信用等级 (AA = 0, A = 1, B = 2, C = 3, D = 4, E = 5, HR = 6)	3.78	0.6299
竞标数量	54.70	34.5313
借款人行为		
历史违约记录 (1 = 有违约, 0 = 无违约)	0.03	0.1765
历史违约率	0.01	0.0200
社交网络违约记录 (1 = 有违约, 0 = 无违约记录)	0.56	0.4961
还款次数	8.40	3.3595
放贷人行为		
竞标 (1 = 应竞标, 0 = 不应竞标)	0.60	0.4900