

P2P网络借贷中的借款描述质量、出借行为与信用风险

辛晨, 刘传哲(博士生导师), 夏雨霏

【摘要】 通过P2P网络借贷平台人人贷的真实数据探究借款描述质量与出借行为和真实信用风险水平的关系,结果表明:借款描述质量会显著影响出借人的决策行为,借款描述中语句通顺、长度适中、不包含负面情绪且蕴含人格更多的借款标的更容易满标;不同人格对于投资者决策的影响具有显著差异。另外,投资者以借款描述为媒介对于标的信用风险的认知与实际风险水平存在一定偏差,具有勤奋、追求生活品质、还款历史和尝试人格的借款人更倾向于如约还款。

【关键词】 P2P网络借贷; 借款描述; 投资者决策; 信用风险; IVprobit回归

【中图分类号】 F832

【文献标识码】 A

【文章编号】 1004-0994(2017)24-0104-8

一、引言

P2P网络借贷是指资金供需双方通过互联网平台实现的无抵押直接借贷。在P2P网络借贷中,借款人在平台上发布借款信息,并公布个人信息。出借人根据信息、风险偏好和主观判断决定是否出借及出借金额。资金暂时由平台冻结,若截止期前未获得足额投资,则交易失败,平台将资金全额退回;若借款人所要求的借款金额被全额满足,则称为“满标”,P2P网络借贷交易达成,平台将资金划拨至借款人账户,借款人之后按约定偿还本息。

目前降低P2P网络借贷信用风险的一个合理可行的思路是加强信息披露,大规模的信息可为评价借款人信用提供依据,从而缓解借贷行为中国有的信息不对称与道德风险。这些信息既包括经过验证的、标准的、可回溯的硬信息,如经平台审核的借款人身份信息、央行开具的信用报告等,也包括未经验证的、非标准的、借款人自愿披露的软信息,如年龄、性别、学历等。Klaft(2008)、Herzenstein et al.(2011)发现在P2P网络借贷中,决策者难以像金融中介那样获取大量的硬信息,只能更多地依赖软信息评价借款人的信用水平。Michels(2012)发现,尽管软信息难以验证,但是它仍然降低了借贷成本,能够有效地促进交易达成。

借款描述作为一种典型的软信息,是借款人所写的供潜在出借人阅读的,关于自身以及借贷事项的描述。借款描述通常包含借款人的借款经历和当前状况,可侧面反映借款人的某种品质和人格。Herzenstein et al.(2011)的研究表明,借款人有充足的动机依靠借款描述塑造出理想的人格(如守信、勤奋等品质),赢得出借人的信任,从而改变其投资行为。但是借款描述中的人格仅是衡量借款描述质量的因素之一,当前对借款描述质量的其他因素,如长度、错别字、语病和感情表达的研究还不多见。Gao and Lin(2015)发现Prosper平台上借款描述的可读性与贷款违约率关系密切;Dowling and Lucey(2005)发现证券市场上投资者的情绪受心情影响巨大,心情较好的投资者易做出乐观的投资决策。我国个人信用评级发展滞后,P2P网络借贷中能够使用的硬信息相对不足,那么借款描述中提供的软信息是否也会影响出借人的决策?如果能,其影响程度又如何?

二、假设提出

在本研究中,借款描述的质量主要包括错别字及病句、长度、情绪表达及人格等四方面。西方心理学研究表明,单词的错误拼写表明作者认知能力和教育水平的不足。Pynte et al.(2004)提出语病会使得语意含混,降低贷款满标的概率。但是在汉语环境

【基金项目】 中央高校基本科研业务费专项资金项目(项目编号:2017BSCXB20)

下针对借款描述可读性与投资者决策关系的研究还很少见。借款描述中错别字和语病的出现暗示借款人存在粗心、马虎的性格特征,也可被解读为缺乏借款诚意,本文推测存在错别字和语病的借款描述更难吸引出借人投标。据此,提出假设1a:

H1a: 如果借款描述中存在较多的错别字和语病,则难以获得贷款。

同时,错别字和语病还可视为体现借款人教育水平的有效代理变量。尽管廖理等(2015)指出P2P网络借贷的出借人并不青睐高学历借款人,但是教育水平提高了人的自我约束水平,受过高等教育的借款人如约还款的概率更高。据此,提出假设1b:

H1b: 如果借款描述中存在较少的错别字和语病,则较少违约。

借款描述长度是决定其可读性的另一个重要因素。但是长度对投资者决策的影响可能存在截然相反的影响途径:一方面,借款描述越长,越容易出现错别字和语病,根据H1a和H1b会降低满标和还款概率;P2P网络借贷中的出借人每笔投资金额较小,面对过长的借款描述,他们可能不愿意阅读而跳过这些标的;长借款描述同样体现了借款人啰嗦、不果断的性格特点,可能引起出借人的反感,从而降低出借人投资的意愿。另一方面,借款描述包含大量信息,能够详细说明借款人的情况,缓解信息不对称,同时也是借款人对自身信用情况比较自信的表现。综合借款描述长度潜在的正反向影响,本文推测借款描述长度对投资者决策的影响为倒U型,即长度适中的借款描述能提高满标率。而当前尚无针对借款描述与实际违约率间关系的研究,本文猜想二者关系同为倒U型。据此,提出假设2:

H2: 借款描述长度与出借人投资意愿及实际违约率的关系为倒U型。

在P2P网络借贷中,因为投资者可以直接观察到所有潜在的借款人,其投资决策更容易受到情绪上的影响。正面、乐观、讨喜的文字表述容易使出借人做出非理性的投资决策;Slovic et al.(2007)认为在充满不确定性的模糊环境中,行为主体倾向于依赖情绪和情感决策以弥补理性决策的不足。Bruton et al.(2015)在众筹领域验证了包含正面情绪的文字描述容易导致投资者过分自信,而负面、悲观的借款描述则暗示借款人遭遇困难,急需帮助,也会吸引出借人的投资。据此,提出假设3a和3b:

H3a: 使出借人产生正面情绪的借款描述与满标概率正向相关。

H3b: 使出借人产生负面情绪的借款描述与满标概率正向相关。

Herzenstein(2011)研究了借款描述中人格对投资人行为的影响,发现描述中所包含人格的数量越多,借款成功率越高,承担的利率越低,贷款违约率也越低。李焰等(2014)、王会娟和何琳(2015)使用了与Herzenstein相近的人格研究了我国P2P网络借贷中人格对投资人行为的影响,得出了类似的结论。据此,提出假设4a和4b:

H4a: 借款描述中的人格数量越多,则借款成功率越高,贷款违约率越低。

H4b: 借款描述中不同人格对出借人投资决策和贷款违约率的影响程度不同。

三、研究设计

1. 数据来源。本文采用人人贷平台2015年7月至9月间发生的20000笔交易(交易编号150000~169999)作为初始样本。由于实地认证标的和机构担保标的与P2P网络借贷纯线上、纯中介的特征相违背,因此本文剔除原始样本中实地验证标的、机构担保标的和重复记录,剩余9746条包含借款描述的交易记录。由于本文需要人工逐条读取识别借款记录的错别字、语病、传达的正面及负面情绪、人格等,因此从中随机抽取了3700条记录,然后剔除重复记录6条,剩余3693条记录。其中包含流标标的3458个,满标交易235个,不包含还款中的交易。满标交易中包含如约还款记录199个,违约记录36个。

2. 被解释变量。P2P网络借贷中出借人的行为具有较强的趋利性,满标与否体现了出借人对该标的信用风险的主观判断,而违约与否则是该标的信用风险的真实体现,主观与客观的偏差也能够验证投资者的非理性行为的存在。本文将满标与否作为除H1b外的其他七个假设的被解释变量;而违约与否则是假设1b和假设4b中的被解释变量。构建哑变量满标率(Success),表示该笔交易是否获得足额的投资。若交易状态显示为已还清、还款中或已垫付,则标记为1,否则记为0。构建哑变量如约还款(FP),表示该交易的借款人是否按照约定归还本息,若交易状态显示为已还清,则记为1,若显示已垫付,则记为0。

3. 解释变量。针对H1a和H1b,错别字和语病分别用SMR和FW表示。SMR表示错别字个数占借款描述总字数的比率;FW则是哑变量,借款描述中存在语病取1,否则取0。H2a和H2b中的借款描述长度(#word)由Excel中的LEN和LENB函数统计,

标点符号不统计在内。H3a 和 H3b 中的正面情绪(pos)和负面情绪(neg)为哑变量,1表示阅读该借款描述会产生相应的情绪,0则表示该借款描述不存在情绪感染力。借款人认真填写的、具有感染力的正面的借款描述,可以引起出借人的共鸣,产生乐观、积极的情绪;而与之相反的是,敷衍了事的、充满负能量的借款描述则表明借款人缺乏诚意,容易引起出借人的反感。H4a 和 H4b 中的 N_identities 表示借款描述中包含的人格总数。本文提取了 3694 条交易记录中借款描述的关键词,结合李焰等(2014)的研究,确定了 9 个人格——诚信、勤奋、有家庭观念、追求生活品质、有事业心、稳定、还款历史、刷信用和尝试。具体各人格用哑变量 $X_1 \sim X_n$ 表示,借款描述中存在该人格取 1,否则取 0。其中诚信、勤奋、有家庭观念体现了中华民族的传统品质,而稳定、有事业心侧面体现了借款人的还款能力,追求生活品质体现了借款人当前的生活状况。还款历史、刷信用和尝试这三个人格则是在整理借款描述中总结提取的。与

口头承诺相比,提及还款历史的借款人显得更加有说服力。而刷信用和尝试则是我国 P2P 网络借贷平台上特有的,平台普遍对新手借款人的借款金额进行限制,只有完成一定数量的交易,借款人才能进行更大额度的借贷,因此为快速提升信用,一些借款人会发放小额短期标的,承诺不提现并提前偿还,因此此类标的较具有吸引力。P2P 网络借贷作为一种新兴事物,大量借款人第一次尝试使用,对流程不熟悉,在借款描述中也会透露相应的信息。结合人人贷中的实际借款描述,投资者情绪与借款人人格的定义、关键词和例子总结见表 1。

由于汉语词义较为灵活,仅使用关键词不能准确提取全部人格和情绪信息,且缺乏有效的中文错别字和病句识别软件,因此本文采取人工识别的方法获取实验所需错别字和语病、情绪和人格数据。识别工作由 4 名互不认识、不了解假设内容且与人人贷中的参与者年龄及学历接近的科研助理完成,4 名助理两两分组。将 3694 条交易记录平均分配给 4

表 1 投资者情绪与借款人人格描述

	定 义	关键词	例 子
诚信	借款人表明自己诚实守信,会按时还款	诚信(257)、信誉(251)、相信(71)	(#1018)信用第一,保证按时还款,诚信为主,绝对讲信用
勤奋	借款人承诺努力工作或同时具有多份工作	努力(31)、奋斗(7)	(#2727)我因为上班的时间比较短,所以闲暇之余想做点小生意,可是刚工作不久,也没存到什么钱,所以想暂时通过贷款用来做成本
有家庭观念	借款人所借款项用于家庭成员使用	父母(58)、妻子(50)、孩子(52)	(#16)放暑假期间,希望可以带儿子出去玩一圈,因没有太多闲钱,希望给我一点帮助,先谢谢啦,借 3000 元,3 个月后再还完
追求生活品质	借款人希望提升生活品质,如购买电子产品、买车、买房等	装修(405)、买房(105)、买车(56)	(#194)本人大学生一枚,想月供买个手机,2000 元就可以,一个月百把块钱本人还是能拿得出的,不会出现断钱
有事业心	借款人借款用于创业或发展自身事业	创业(250)、货(549)	(#10)现在店铺位于步行街内,经营了一年多了,因为步行街还没一家专业的美甲店,想扩大店面,服装与美甲一条龙,为女性服务
稳定	借款人具有稳定的工作与收入来源用于偿还贷款	稳定(432)、国企(38)	(#8)本人在绵阳东材科技上班,月工资 3100 元,扣除保险和公积金后实拿 2580 元,每月工资打卡
还款历史	借款人陈述个人过去的还款情况	逾期(59)	(#380)已经借款一次,12 个月,因一次周末外出旅游,周一回来还款的,所以导致了逾期一次的记录
刷信用	借款人借款为提升信用额度	提升(51)、额度(80)、信用(485)	(#585)刷信誉了,投资理财的跟上脚步
尝试	借款人第一次接触网贷,不太熟悉网贷流程	第一次(233)、体验(43)	(#2387)审核刚通过,借机尝试一下,希望大家支持
正面情绪	撰写认真,容易让人产生乐观、振奋、积极等正面情绪的借款描述	谢谢(712)、梦(22)、祝福(25)	(#121)需要一笔钱做短期周转用下,请好心人帮助渡过难关,万分感谢
负面情绪	撰写敷衍,容易让人产生消极、低沉、悲观等负面情绪的借款描述	病(23)、困难(100)、死(11)	(#771)看病急需钱,凑齐后立即还款,可能资料还不够齐,实在是没有别的办法了

注:“关键词”一栏中,各词语后的数字表示该词汇在整个样本记录中出现的次数;“例子”一栏中的序号表示样本编号。

名科研助理,由助理独立对分配的任务进行处理。为减少错误并保证数据的一致性,错别字和病句的数量由同组成员间交换审查,如产生争议,则交由另一组的科研助理做出最终结论;而情绪和人格数据则进行一致性分析,用来判断不同主体间对同一借款描述的主观判断是否存在偏差,本文使用kappa值作为判断一致性程度的指标。

4. 控制变量。控制变量主要分为借款标的信息与借款人信息两类。借款标的信息包括借款金额(amount)与利率(vate);借款人信息包括借款人的年龄、婚姻状况、教育水平、通过审核数量等。年龄(age)为借款人填写的本人年龄,为连续变量;本文将婚姻状况(marry)作为哑变量,0表示借款人未婚或离异,1为已婚;教育水平(edu)从低到高分四项,高中及以下取0,大专取1,本科取2,研究生及以上取3;对于借款人收入(income),15000元/月以下为1,15000~20000元/月为2,20000元/月以上为3。人人贷平台提供了借款人房产认证、车产认证等服务,借款人可自愿上传证件进行认证,若证件经平台审核为真实,则会在借款标的页面显著位置标识以证明借款人资信。由于审核种类众多,将借款人通过审核数量(certification)作为控制变量。需要说明的是,李焰等(2014)的研究将平台信用等级和借款期限作为控制变量,但是我们发现借款期限与利率具有较强的相关性,信用等级与通过审核数量有较强的相关性。为消除共线性对模型的影响,本文将平台信用等级与借款期限从控制变量中剔除。

5. 回归方法。由于实验使用的被解释变量Success和FP均为二元变量,本文基于Stata 12软件,使用probit作为基准回归方法。对于不同的借款标的*i*,出借人决策 Y_i 只有投资(取值为1)或不投资(取值为0),且其决策由潜变量 y_i^* 决定。而 y_i^* 则由假设中的*n*个解释变量 $x_1 \sim x_n$ 、控制变量 $x_{control}$ 与随机扰动项 ε_i 决定:

$$y_i^* = \alpha_1 x_n + \beta_1 x_{control} + \varepsilon_i$$

$$Y_i = \begin{cases} 1, y_i^* < 0 \\ 0, y_i^* \geq 0 \end{cases}$$

其中:随机扰动项 ε_i 服从正态分布; α_1 与 β_1 分别是解释变量与控制变量的待估计参数。

本文借款标的的利率(vate)由借款人根据自身承受能力自行拟定,且A.Mild(2015)认为借款人可能根据出借人先前投标情况有策略地制定利率水平以提升获取贷款的概率,存在与被解释变量相互影响的可能,vate可能具有内生性。因此,使用工具变

量法解决内生性问题,参考Rivers and Vuong(1988)的做法,使用最大似然估计法对联立工具变量probit回归模型进行拟合。工具变量要求同存在内生性的解释变量高度相关,而不与其他解释变量和回归模型的随机扰动项相关。参考Dorfleitner et al.(2016)的做法,使用无风险利率(r_{free})作为工具变量,本文将央行公布的不同期限的定期基准存款利率作为无风险利率。联立工具变量probit回归模型:

$$y_i^* = vate_i' \gamma + \alpha_1 x_n + \beta_1 x_{control} + \mu_i$$

$$vate_i = \alpha' r_{free} + \varepsilon_i$$

$$Y_i = \begin{cases} 1, y_i^* < 0 \\ 0, y_i^* \geq 0 \end{cases}$$

其中: x_n 为工具变量; $x_{control}$ 为控制变量; μ_i 和 ε_i 分别为结构式和简约式的随机扰动项。

四、实证结果分析

1. 基准回归结果。表2中模型(1)~模型(5)分别表示借款描述质量对满标率影响的probit回归结果,表3中模型(6)~模型(10)则表示借款描述质量与如约还款关系的probit回归结果。为更直观地说明借款描述质量的各组成部分的影响,表中给出的是各变量的平均边际效应而非回归系数。

模型(1)的结果部分验证了H1a。借款描述的错别字率并不会显著影响满标率,但是病句则会对满标率产生负向影响。近年来随着网络用语发展迅速,一些有别于传统语法规范词组的产生加剧了借款描述中的错别字问题。P2P网络借贷的参与者长期接触互联网,对网络用语比较熟悉,从而降低了错别字对出借人行为的影响,加之拼音输入法的普及使错别字的读音相近,不会影响投资者对句意的理解。但是语病的存在则会比较明显地影响借款描述的通顺程度和句意表达,不利于投资者从借款描述中汲取需要的信息,甚至可能引起出借人的反感。

而模型(6)的回归结果说明错别字率及语病与如约还款间均无显著关系,表明投资者对错别字及病句的风险认识存在偏差。值得注意的是,控制变量中教育水平对还款有着明显的正向促进作用,即学历高的借款人倾向于如约还款,这也说明Dorfleitner et al.(2016)关于错别字率与病句是教育水平代理变量的推断值得商榷,笔者将这种错别字与病句同教育水平“脱钩”现象归结为高校语文教育的缺失和语言腐蚀。目前我国高等教育中语文边缘化严重,使得高校毕业生的语言表达能力出现下滑。另外,控制教育水平后计算年龄与错别字率和病句的偏相关系数,发现二者存在小而显著的正相关关系(偏相关系

表 2 借款描述质量与满标率关系的probit回归结果

	模型(1)	模型(2)	模型(3)	模型(4)	模型(5)
	被解释变量:Success				
SMR	-1.0282 (-1.51)				
FW	-0.0608** (-1.91)				
#word		0.0002 (0.98)			
#word ²		-0.000001 (-0.89)			
pos			0.0014 (0.18)		
neg			-0.1034*** (-3.01)		
N_identities				0.009495*** (2.68)	
X ₁					-0.0023 (-0.34)
X ₂					0.0563*** (4.69)
X ₃					-0.0380 (-1.57)
X ₄					0.0152 (1.57)
X ₅					0.0258*** (3.08)
X ₆					0.0088 (1.22)
X ₇					0.0530*** (3.91)
X ₈					-0.0022 (-0.19)
X ₉					-0.0229** (-2.07)
age	0.0028*** (6.17)	0.0029*** (6.20)	0.0028*** (6.17)	0.0029*** (6.20)	0.0027*** (6.15)
edu	-0.0051 (-1.28)	-0.0055 (6.26)	-0.0047 (6.11)	-0.0058 (6.3)	-0.0042 (5.99)
certification	0.0158*** (12.41)	0.0163*** (-1.35)	0.0159*** (-1.16)	0.0159*** (-1.43)	0.0152*** (-1.06)
income	0.0342*** (6.97)	0.0338*** (12.66)	0.0333*** (12.2)	0.0347*** (12.29)	0.0338*** (11.9)
amount	-0.0273*** (-9.8)	-0.0271*** (6.83)	-0.0274*** (6.82)	-0.0266*** (7.09)	-0.0289*** (6.79)
rate	-1.1574*** (-9.8)	-1.1739*** (-9.68)	-1.1717*** (-9.85)	-1.1801*** (-9.51)	-1.1793*** (-9.95)
marry	-0.0062 (-0.82)	-0.0071 (-10.53)	-0.0061 (-10.58)	-0.0068 (-10.63)	-0.0074 (-10.65)
constant	1.2994*** (3.18)	1.0661*** (-0.94)	1.4014*** (-0.81)	0.9764*** (-0.9)	1.5136*** (-1.01)
N	3690	2.49	3.36	2.34	3.4
Log likelihood	-512.36				

注:***、**和*分别代表在1%、5%和10%的水平上显著,括号内为相关统计量,下同。

数分别0.05和0.064),佐证了语言腐蚀现象的存在,即某种语言因使用减少而产生语言能力减弱的现象,通常受年龄、外语、腐蚀时间等因素影响。

模型(2)与模型(7)均对应H2。结果显示借款描述长度及其平方与满标率和如约还款不存在显著关系,H2未得到验证。

模型(3)对应H3a与H3b,在控制各种因素后,借款描述中的正面情绪同投资者决策不存在显著影响,即H3a不成立。这可能说明“谨慎乐观主义”在P2P网络借贷的借款决策中起着重要的作用。在P2P网络借贷的高风险下,违约造成的本金损失相比利息收入而言要大得多,即便违约概率不高,处于积极情绪状态的出借人也不愿意去冒险投资而是倾向于保持现状。而同时模型(3)的结果与H3b相反,阅读借款描述后使投资者产生负面情绪的标的会阻碍投资者的投资行为,这与情绪保持论所持的负面情绪下人们试图改变当前情绪而采取冒险行为的论断不一致。我们认为这是由于不同类型的负面情绪对投资者决策的影响方向不同,P2P网络借贷中借款描述使潜在出借人产生的负面情绪多源自不认真的填写及抱怨,以厌恶为主。而根据徐四华(2015)的实证研究,具有厌恶情绪的实验对象会更有效地利用当前信息做出风险规避行为。模型(8)的结论则拒绝了H3a和H3b,借款描述中的正面与负面情绪与标的违约情况不存在显著关系。

模型(4)、模型(5)分别对应H4a和H4b。从回归结果可以看出,借款描述中包含的人格数与满标率存在显著的正相关关系。具体来看,勤奋、有事业心和还款历史人格与满标率呈正相关关系,而尝试人格与满标率则呈负相关关系,其他人

表3 借款描述质量与如约还款关系的probit回归结果

	模型(6)	模型(7)	模型(8)	模型(9)	模型(10)
	被解释变量:FP				
SMR	22.4435 (0.9)				
FW	-0.6442 (-1.29)				
#word		0.0017 (0.74)			
#word ²		0.0000 (-0.61)			
pos			0.0384 (0.71)		
neg			-0.2279 (-1.15)		
N_identities				0.0089 (0.36)	
X ₁					-0.0149 (-0.32)
X ₂					0.1554** (2.16)
X ₃					-0.2464* (-1.87)
X ₄					0.1960** (2.51)
X ₅					0.0630 (1.12)
X ₆					-0.1175** (-2.47)
X ₇					0.1965* (1.82)
X ₈					-0.0222 (-0.26)
X ₉					0.2017* (1.81)
age	0.0020 (0.66)	0.0021 (0.68)	0.0020 (0.66)	0.0023 (0.71)	0.0017 (0.62)
edu	0.0543* (1.83)	0.0513* (0.68)	0.0571* (0.66)	0.0520* (0.75)	0.0467* (0.58)
certification	0.0335*** (2.62)	0.0359*** (1.66)	0.0341*** (1.88)	0.0350*** (1.72)	0.0374*** (1.59)
income	0.0453 (1.23)	0.0528 (2.76)	0.0531 (2.62)	0.0547 (2.7)	0.0408 (2.98)
amount	-0.0214 (-0.76)	-0.0220 (1.42)	-0.0205 (1.44)	-0.0201 (1.47)	-0.0166 (1.09)
rate	-3.0537*** (-3.30)	-2.8860*** (-0.74)	-2.8136*** (-0.72)	-2.8798*** (-0.69)	-3.1023*** (-0.59)
marry	-0.0979* (-1.76)	-0.1172** (-3.1)	-0.1062* (-3.04)	-0.1142** (-3.07)	-0.0978* (-3.4)
constant	1.9776 (1.46)	1.3805 (-2.09)	1.6004 (-1.91)	1.5351 (-2.05)	1.5725 (-1.84)
N	235	0.96	1.19	1.12	1.07
Log likelihood	-83.051267				

格与满标率无显著关系。勤奋、有事业心塑造了借款人有能力偿还贷款的正面形象,比起口头的还款承诺,过往的还款历史能更有力地说明借款人真实的信用水平。而尝试人格表明借款人对P2P网络借贷的机理和流程不够熟悉且缺乏足够的信用背书,因此投资人对新手标的的投资也更加谨慎。

模型(9)、模型(10)对人格数目及不同类型人格与如约还款的关系进行了验证,发现人格数与如约还款不存在显著关系。具体到各种人格上,具有勤奋、追求生活品质、还款历史和尝试人格的借款人更倾向于如约还款,而稳定和有家庭观念人格则对如约还款具有显著的负向作用。这表明投资者在对借款描述中表现出的不同人格的识别上与真实情况存在较大偏差。尽管基准回归的结果表明投资人更愿意借款给具有勤奋人格的借款标的、排斥具有尝试人格的标的,但是这两种人格对如约还款的影响方向与借款人的主观认识恰恰相反,这佐证了Mild et al. (2015)关于P2P网络借贷参与者存在非理性行为的论断。

2. 内生性问题的处理。尽管模型基于probit回归方法对借款描述质量与满标率、如约还款的关系得出了结论,但该结论有可能遭受内生性问题的干扰,内生性主要来源于控制变量利率与满标率之间的相互作用与影响。为排除内生性干扰,本文使用IVprobit方法重新对模型进行回归,表4为回归结果。

使用Wald检验判断借款利率变量外生性,Wald检验的零假设为该变量是外生的。由表4可知,四个模型对应的Wald统计量值均在1%的显著性水平上拒绝原假设,即接受IVprobit的结果。使用工具变量后借款利率的平均边际效应较probit回归中对满标率的负向影响更加明显。

模型(11)对应H1a。其中错别字与满标率的关系仍不显著,而病句的边际效应在10%的显著性水平上为负。从数值上看,病句的平均边际效应为-0.07,其绝对值要显著大于probit回归所得的边际效应。根据这一数值,病句上升一个样本标准差时,满标率要降低7%,应该说这一数字具有较强的经济意义。

模型(12)对应H2。IVprobit与probit回归的结果具有较大的分歧,考虑到Wald检验结果,IVprobit的结果更加可信。从模型的结果来看,借款描述长度与满标率在1%的显著性水平上存在正向相关关系,而借款描述长度的平方则与满标率在5%的显著性水平上负相关,尽管边际效应不大,但是该结果验证了H2,说明借款描述长度与满标率存在倒U型关系。

模型(13)对应H3a和H3b。模型(13)IVprobit的回归结果与probit的回归结果相似,负面情绪的边际效应在1%的水平上显著为负,进一步强化了probit回归的结论。从数值上看,模型中负面情绪的边际效应为-0.168,其绝对值要显著大于probit回归结果。

模型(14)的结果进一步验证了H4a,且边际效应绝对值为0.025,大于probit的回归结果,说明借款描述中的人格数会显著增加借款人投资该标的的意愿。具体到各人格来看,模型(15)IVprobit回归中勤奋、有事业心、还款历史和稳定人格在1%的水平上显著提高满标率,而尝试人格同样在5%的水平上对满标率产生负向影响。同模型(5)中的probit回归结果相比,模型(15)的回归结果存在些许不同,稳定人格出现在积极影响投资者决策的人格之中。稳定人格体现了借款人具有稳定的工作与收入来源,为如约还清贷款奠定了坚实的基础。

3. 控制变量的回归结果。在借

表4 借款描述质量与满标率关系的IVprobit回归结果

	模型(11)	模型(12)	模型(13)	模型(14)	模型(15)
	被解释变量:Success				
SMR	-1.0130 (-1.15)				
FW	-0.0707* (-1.81)				
#word		0.0011*** (2.93)			
#word ²		-0.000003** (-2.24)			
pos			-0.0061 (-0.48)		
neg			-0.1682*** (-4.01)		
N_identities				0.02515*** (4.35)	
X ₁					-0.0071 (-0.64)
X ₂					0.0674*** (2.75)
X ₃					0.0109 (0.31)
X ₄					0.0106 (0.47)
X ₅					0.0585*** (4.47)
X ₆					0.0450*** (3.81)
X ₇					0.0483*** (3.22)
X ₈					-0.0028 (-0.08)
X ₉					-0.0395** (-2.12)
age	0.0026*** (2.75)	0.0026*** (2.76)	0.0025*** (2.7)	0.0026*** (2.69)	0.0021** (2.2)
edu	-0.0323*** (-4.65)	-0.0332*** (-4.73)	-0.0317*** (-4.48)	-0.0340*** (-4.85)	-0.0326*** (-4.69)
certification	0.0186*** (6.35)	0.0190*** (6.35)	0.0188*** (6.35)	0.0182*** (5.96)	0.0174*** (5.65)
income	0.0256** (2.43)	0.0242** (2.27)	0.0246** (2.33)	0.0268** (2.52)	0.0276** (2.59)
amount	-0.0219*** (-3.21)	-0.0226*** (-3.35)	-0.0234*** (-3.42)	-0.0207*** (-3.01)	-0.0253*** (-3.52)
rate	-7.2011*** (-11.16)	-7.1809*** (-10.89)	-7.1593*** (-10.54)	-7.2107*** (-10.97)	-7.4670*** (-13.27)
marry	-0.0310*** (-2.64)	-0.0317*** (-2.7)	-0.0314*** (-2.66)	-0.0317*** (-2.69)	-0.0335*** (-2.87)
constant	4.6570*** (22.39)	4.4340*** (19.76)	4.7630*** (22.36)	4.4490*** (20.28)	4.7505*** (22.84)
N	3690				
Log likelihood	6585.47	6588.38	6595.63	6590.15	6640.29
Wald statistic	42.59***	41.36***	39.98***	40.91***	47.54***

款描述质量与满标率的关系方面,比较probit回归和IVprobit回归中各控制变量的边际效应可以发现,各变量符号均未发生改变,其中借款人年龄、通过审核数量和收入都与满标率在1%的水平上正相关;借款利率及金额则在1%的显著性水平上与满标率负相关。通过审核数量作为平台缓解P2P网络借贷中信息不对称及硬信息缺乏的重要手段,在实践中对投资者的决策和信用风险识别起着关键的作用;而利率水平的高低决定着借款人的利息支持程度,对投资者决策的利益回报具有关键影响。值得注意的是,两种回归方法下教育水平和婚姻状况存在明显的差别。在IVprobit模型中,教育水平及婚姻状况都会对满标率产生显著负向影响,即在控制其他变量的前提下,高学历或单身借款人均不受投资者青睐,关于教育水平的结论同廖理等(2015)的研究结论相吻合。而从社会资本角度来看,单身的借款人失去了由于婚姻带来的配偶方家族的横向关系和纵向的亲子关系,也有可能丧失亲子关系衍生出的其他关系,因此是否具有配偶极大地影响着借款人的社会资本,进而影响其还款能力。

4. 稳健性检验。为验证以上结论的稳健性,本文从两方面进行了稳健性检验:第一,将全部解释变量放入模型之中,使用probit与IVProbit方法进行重新回归,研究结论保持不变;第二,考虑到IVProbit中借款人婚姻与教育水平的回归结果与probit不同,仿照廖理等(2015)的做法对这些变量分组进行回归,结果显示probit回归与IVprobit回归所得的结论依旧稳健。限于篇幅,在这不予赘述。

五、结论

本文从借款描述中的错别字与病句、长度、情绪及人格等四方面研究了借款描述质量在P2P网络借贷中的作用。从借款描述质量对投资者决策的影响上看,病句的存在会影响出借人的阅读体验,降低出借意愿;借款描述对出借人的投资意愿的影响呈倒U型曲线,即过短或过长的借款描述均不利于满标。此外,从借款描述包含的情绪上看,负面情绪的存在会显著降低满标率。在出借人看来,借款描述中的人格数目的增加缓解了P2P网络借贷中的信息不对称,丰富了标的信息,有利于提高借款人的借款成功率。而具体到人格,勤奋、稳定、有事业心及还款历史人格对满标率的影响为正,而新手人格的影响为负。而从借款描述质量对标的违约的影响上看,除勤奋、

追求生活品质、还款历史和尝试人格的影响为正、稳定和家庭观念人格的影响为负外,其他借款描述质量因素对真实的还款行为不存在显著影响。而教育水平、婚姻状况、借款利率和通过审核数量则可作为识别标的信用水平的有效媒介。

本文研究还表明:一方面,借款描述质量对投资者决策具有显著影响,借款人可通过提高借款描述质量提高贷款成功率;而另一方面,投资者对贷款标的信用风险的认知存在偏差,大部分借款描述质量的组成因素与标的违约与否不相关,体现了P2P网络借贷参与者在硬信息缺乏的大背景下行为趋向于非理性。

依据本研究的结论又可做如下政策引申:第一,某些借款描述质量因素及某些控制变量可作为评价信用风险高低的重要指标,这一研究结果既有利于帮助出借人正确识别信用风险,也有助于平台内部信用评分系统的完善;第二,借款描述中的一些人格对信用风险识别的积极作用为“互联网+”环境下征信系统的建立提供了新的思路,可考虑基于大数据视角,从社交网络中深入挖掘相关数据,扩大征信系统的数据来源;第三,出借人对标的信用风险识别的偏差表明其仍未对P2P网络借贷的高风险产生清醒的认识,因此有必要进一步加强出借人的风险意识教育,帮助其理性投资,维护P2P网络借贷秩序和产业的持续健康发展。

主要参考文献:

Michels J.. Do unverifiable disclosures matter? Evidence from peer-to-peer lending [J]. The Accounting Review, 2012(4).

李焰,高弋君,李珍妮等. 借款人描述性信息对投资人决策的影响——基于P2P网络借贷平台的分析[J]. 经济研究, 2014(S1).

王会娟,何琳. 借款描述对P2P网络借贷行为影响的实证研究[J]. 金融经济研究, 2015(1).

徐四华. 趋近和回避动机情绪对不确定性决策的影响[J]. 中华行为医学与脑科学杂志, 2015(9).

吴彩霞,李艳,靳小怡. 农村大龄未婚男性社会资本研究——基于借贷网络的视角[J]. 人口与经济, 2012(1).

作者单位:中国矿业大学管理学院,江苏徐州 221116