

中国互联网金融的地域歧视研究^①

廖理¹ 李梦然² 王正位¹

(1. 清华大学五道口金融学院; 2. 清华大学经济管理学院)

【摘要】 本文借助一家P2P网络借贷平台的交易数据, 实证考察了互联网金融借贷中是否存在地域歧视。描述性统计表明各省份之间的订单成功率存在巨大差异, 在控制了订单信息和借款人信息之后, 该差异有所降低, 但仍在统计上显著, 这说明我国的P2P网络借贷的交易中存在地域歧视问题。进一步的研究表明: 被歧视的省份订单违约率并没有显著地高于其他省份, 这说明本文发现的地域歧视属于偏好性歧视——这是一种非理性行为。

关键词 互联网金融 地域歧视 P2P网络借贷 违约率

中图分类号 F830.5 **文献标识码** A **JEL分类号** G11, G21, G23

DOI:10.13653/j.cnki.jqte.2014.05.004

Regional Discrimination in Chinese Internet Finance

Abstract: Using data from a peer-to-peer online lending platform, this paper empirically examines whether the regional discrimination exists in the internet financial borrowing. The descriptive result shows that there is a huge difference among provinces in the probability of funding success. After controlling the borrowers-related and loan-related characteristics, the difference decreases, however, still is statistically significant, suggesting the existence of regional discrimination in China's P2P lending. Furthermore, the findings in this paper show that those discriminated provinces have lower default rates, and it might be natural to conclude that such evidence is suggestive of taste-based discrimination, which is a kind of irrational behavior.

Key words: Internet Finance; Regional Discrimination; P2P Lending; Default Rate

引言

“歧视”无论是在社会学研究还是经济学研究中都是一个非常重要的话题, 所谓歧视就是不平等对待, 在现实生活中, 每个人都可能会在不同程度、在不同方面存在歧视他人的行为。根据社会学的研究, 歧视源于社会中广泛存在的竞争压力、个人偏好以及人与人之间的严重的信息不对称。

^① 本文得到国家自然科学基金重点项目“中国城市居民家庭/消费者金融研究”(71232003)的资助。

在经济学领域,对于歧视研究已有了很长的历史。例如,劳动经济学就深入研究了那些非经济性质的个人特征(如种族、性别、肤色、年龄、家庭背景、民族传统、宗教、身体素质、户口所在地等)所引起的劳动者在就业、薪资、金融借贷等方面的不公平待遇,以及不能够得到平等的市场准入资格等问题。相对于经济学的研究,金融领域对歧视的研究还相对较少,其主要原因在于:要深入考察“歧视”问题,我们需要大量的基于微观个体层面的金融数据,但是长期以来金融学的研究缺乏该方面数据的积累(尤其是缺乏个体违约率方面的数据)^①。

近年来,一种新兴的信贷市场——P2P网络借贷平台的数据为我们研究歧视问题提供了良好的契机。Lin等(2009)、Bachmann等(2011)定义P2P网络借贷(Peer to Peer Online Lending)是指借款人与投资者之间点对点,通过网络借贷平台而不是传统的金融机构中介产生的无抵押贷款。这种模式源自英国,是将互联网技术与小额信贷相融合的新型民间借贷模式,平台上会披露借款人的各类信息,包括其信用记录,同时也包括各类未经证实的个人陈述等信息,它为个人与个人之间提供了公开透明的小额信用交易的可能。这种平台在达成交易的同时,也为我们积累了大量的微观个体层面的金融数据,尤其是积累了借款成功率和借款人违约率方面的数据。

但有学者指出,这种P2P网络借贷平台与传统的银行借贷相比,由于P2P网络借贷平台中的借贷双方不需要面对面的接触,因此对借款人而言显得更为友好(Pope和Sydnor, 2011)。因此,歧视的问题在网络借贷市场中可能在很大程度上得到了解决,例如,Barasinska和Schafer(2010)^②对欧洲最大的P2P借贷平台Smava进行了性别歧视的研究,发现在该平台上对于借款人性别没有偏好性歧视,男性和女性在贷款成功率上没有显著性差异,但女性通常会设定比男性更高的利率主要是由于女性认为自己会被歧视,因此设定更高的利率来吸引投资者。他们的研究表明在经济学中常见的性别歧视问题在互联网借贷中并不存在。但是即便如此,互联网借贷中,仍然存在其他形式的歧视,例如,从借贷成功率的角度研究借贷歧视问题,Ravina(2008)^③针对借款人与投资者之间的共同特征进行了研究,其结果发现借款人与投资者间的共同特征与订单成功率具有很强的正相关性,共同的地域、共同的种族以及共同的性别都会显著增加订单的成功率。Pope和Sydnor(2011)基于Prosper利率拍卖机制的数据发现,非裔美国人借贷成功的概率比与其信用评级相似的白人低25%~34%,同时非裔美国人最终的贷款利率比白人获得贷款的利率高0.6%~0.8%。Herzenstein等(2011)^④也证实了这一结论,他们发现非裔美国人相比其他种族的人确实具有更低的贷款成功率。Pope和Sydnor(2011)同时他们还发现了对于年龄歧视的问题,与35~60岁的群体相比,35岁以下的借款人具有更高的借款成功率,60岁以上的借款成功率更低。Barasinska和Schafer(2010)发现对于年龄有一定的歧视,年龄越大其借款成功

① 基于数据的可得性,目前金融领域少有的一些关于歧视的研究主要集中在对于传统银行借贷中性别歧视的研究(Schafer和Ladd, 1981; Ladd, 1982; Cavalluzzo和Wolken, 2005)和种族歧视的研究(Quercia和Stegman, 1992; Georing和Wienk, 1996; Blanchflower等, 2003)。

② Barasinska N., and D. Schafer, 2010, *Are Women More Credit-constrained than Men? Evidence From a Rising credit market?* [R], Working Paper, German Institute for Economic Research.

③ Ravina, E., 2008, *The Effect of Beauty and Personal Characteristics in Credit Markets* [R], Working Paper, Columbia University.

④ Herzenstein, M., S. Sonenshein, and U. M. Dholakia, 2011, *Tell Me a Good Story and I May Lend You My Money: The Role of Narratives in Peer-to-peer Lending Decisions* [R], SSRN working paper.

率越低,并分析了主要原因是该平台的借款期限都是3~5年,因此投资者考虑了一定的生命周期成本在其中。

不过,甄别歧视是否理性对于我们的研究是有帮助的。事实上,在经济学理论中,学者们区分出了两类歧视:一类是Phelps(1972)和Arrow(1973)提出的有效统计歧视(Efficient Statistical Discrimination),另一类是Becker(1957)提出的非有效偏好歧视(Inefficient Tasted-based Discrimination)。具体到借款成功率这一话题,有效统计歧视是指群体的借款成功率“歧视”来源于群体背后的不同违约率差异——换句话说,这种“歧视”背后是有经济原因的;而非有效偏好歧视是指群体的借款成功率“歧视”完全是由投资者的好恶决定的,背后没有其合理的经济原因。例如,Pope和Sydnor(2008)发现,从订单的预期收益角度来看,非裔美国人的订单收益要明显低于白人,即使其通常会设定较高的借款利率,但仍无法弥补其相对于白人而言更高的违约概率。因此他们认为,尽管表面上看起来Prosper平台上存在对于种族的歧视问题,但这种歧视是一种有效统计歧视,换句话说,非裔美国人借贷成功的概率低,并不能说明Prosper平台存在对于种族的歧视,相反,相对于其违约率而言,这种歧视还远远不够。

具体到中国而言,一个比较有意思的话题是有关地域歧视问题。这主要是源于我国地理面积广阔,人口众多,各地区发达程度、文化传统等各方面差异化严重,因此地域歧视的问题在我国较为普遍,针对资本市场的大部分研究主要集中在股票市场地域偏好,然而对信贷市场的地域偏好研究并不多。本文想了解,在社会学中比较常见的地域歧视现象在金融领域是否也存在。具体来说,本文利用中国“人人贷P2P网络贷款平台”的交易数据,实证分析考察了这家P2P平台上是否存在地域歧视?进一步,如果存在地域歧视,那么这种歧视是理性的还是非理性的?

我们的研究表明,我国的P2P网络借贷市场中,不同的地域之间借款成功率存在显著差异,这种差异在控制了其他变量之后依然存在。进一步,对各省份订单的违约率进行研究发现,地域之间借款成功率差异与其违约率之间并无统计上的关系——这就意味着:我国P2P信贷市场中的地域歧视是存在的,而且这种歧视是一种非有效的偏好歧视。该结果一定程度上说明了我国P2P信贷市场中的非理性现象,对于国内外学者及投资者了解中国P2P信贷市场有一定的帮助,同时也丰富了中国新兴的互联网金融借贷市场中投资者的行为研究,对学术界、监管层及金融实务界理解和完善新兴借贷市场具有一定的参考意义。

一、实证设计

1. 实证思路

如前所述,本文的宗旨在于考察P2P网络借贷平台上是否存在地域歧视,并检验这种歧视是否为理性行为。为此,我们首先可以考察不同省份之间,其P2P借款成功率方面是否存在差异,如果存在差异,则说明P2P网络借贷中地域歧视的存在。

在此基础上,我们可以接着考察这种地域歧视是否合理。为此,我们可以借助不同省份间违约率的情况说明这一问题。如果借款违约率较低的地域其订单成功率反而较低,那么我们就可以推断这种歧视是偏好性歧视,属于非理性歧视的范畴;但是,如果借款成功率较低的区域其订单违约率相对较高,那么我们就可以推测是更高的违约风险导致较低的借款成功率,因此这种歧视可能是一种理性的歧视。

对于理性歧视，我们还可以通过 Phelps (1972)、Arrow (1973) 所提出的有效统计歧视理论，进一步检验违约所带来的损失与借款成功带来的预期收益之间的关系：如果由违约所带来的损失小于其有可能获得的预期收益，那么不但这种歧视属于理性歧视，而且我们还能推断这种歧视的“程度”还不够；相反，如果由违约所带来的损失大于其有可能获得的预期收益，那么我们就可以推断该歧视虽然“理性”，但存在过度歧视的嫌疑。

本文的核心实证设计流程图可以参见图 1 所示。

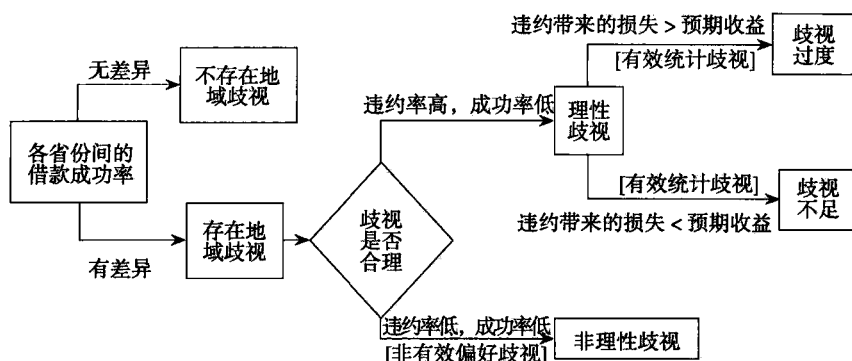


图 1 实证设计思路

2. 地域歧视实证模型设定

检验地域歧视存在性的一个直观逻辑是看各省份之间的借款成功率是否有显著差异。但是，这种简单的比较并不足以证明地域之间的贷款成功率差异是“稳健”存在的，其原因在于：订单的借款成功率除了可能受到地域因素影响之外，还有可能受到“借款人所披露的信息”以及“订单自身设定的信息”的影响。例如，如果其他因素都保持不变，借款利率更高的订单就更有可能是融资成功。因此我们在考察借款成功率地域间差异的时候，我们需要控制“借款人所披露的信息”以及“订单自身设定的信息”的影响，换句话说，我们应该使用“其他条件不变”效果 (Ceteris Paribus Effect) 检验模型进行检验。因此，本文对地域歧视的实证模型设定如下：

$$success_i = \alpha + \sum \beta_n \times province_n + \gamma \times X_i + \zeta \times W_i + \epsilon_i \quad (1)$$

其中， $success_i$ 是订单 i 的融资成功情况， $success=1$ 表示为订单融资成功， $success=0$ 为订单失败； $province_n$ 表示第 n 个省份；同时，我们将控制变量分成了两组，一组为订单自身的信息 X_i ，另一组为借款人的各类信息 W_i ； ϵ_i 表示误差项。

在公式 (1) 中，如果 β_n 是联合显著的，则我们就可以推断地域对于订单成功率是具有显著影响的，因此可以认为 P2P 市场中是存在地域歧视的。

3. 歧视是否理性的实证模型设定

如果 P2P 市场中存在地域歧视，我们需要进一步考察投资者的这种歧视是否为理性行为。按照图 1 的逻辑，我们可以检验各省份之间的借款成功率差异是否源于背后的违约率差异。因此，类似于公式 (1)，我们将首先检验不同区域之间的借款违约率差异：

$$default_i = \alpha' + \sum \beta'_n \times province_n + \gamma' \times X_i + \zeta' \times W_i + \epsilon'_i \quad (2)$$

公式 (2) 中的样本均为成功筹款的订单，同时剔除了正在还款的样本，其原因在于我

们无法预知正在还款的样本其未来的违约状况，因此不作为本文的考察范围。公式(2)中 $default_i$ 为订单的违约情况， $default=1$ 表示订单违约，否则为0。同样，为了剔除其他因素的影响，我们在公式(2)中控制了 X_i 和 W_i 。

通过公式(1)和公式(2)，我们就可以得到控制了其他因素之后不同省份的平均借款成功率和平均违约率。根据图1的逻辑，进一步考察两者之间的关系，如果平均违约率越低的省份，其订单成功率也越低，那么我们判断这种歧视则被视为非理性的歧视；相反，如果违约率越高的省份，其成功率越低，那么我们就推断这种地域歧视在某种程度上可能是理性的。

二、样本数据与变量设定

1. 数据选取与网站特征描述

人人贷网站成立于2010年5月，由于该网站在成立初期的时候有一些测试数据，会对研究具有一定干扰作用，因此本文选取人人贷网站2011年1月1日至2013年4月27日期间发布的所有有效借款订单数据作为研究分析的全样本，共包含64539个有效订单，其中涉及11年的订单数为15409个，12年的订单数为29546个，13年的订单数为19584个。

人人贷的用户可以自愿将其个人信息披露于人人贷网站，包括工作城市、性别、年龄、单位规模(人数)、职业、工作时间、收入、毕业的学校、学历、入学时间、房产情况、车产情况及信息、房贷、车贷等信息，其中部分信息可以提供客观证明由人人贷网站进行审核后得以认证，成为被认证过的信息。

人人贷网站的投标种类分为四种，信用认证标、实地认证标、机构担保标以及智能理财标。从图2可以看到，本文的样本中，包括信用认证标52097个，实地认证标4121个，机构担保标2054个，智能理财标92个，其中信用认证标是最主要的订单种类，占到全样本的90.28%。由于实地认证标、机构担保标以及智能理财标其担保方式和信息披露方式与我们要考察的目标存在较大的差异，并且在我们的样本期间违约率均为零，因此本文主要对信用认证标进行深入研究，剔除了其他三类订单数据的样本。



图2 样本期间人人贷订单分布状况

根据人人贷网站规定,借款人在申请不同订单种类进行借款时,分别有相应的认证信息需要提交人人贷网站进行审核。对于信用认证标(现又叫工薪贷),必须提交审核的材料包括身份证复印件、个人征信报告、带公司公章的劳动合同或工作证明以及近3个月工资卡银行流水;人人贷网站目前的运营方式是不参与利率定价,2012年之前,对订单利率不设定下限,但需要根据国家有关条例对民间借贷的规定设定利率上限,2012年以后,人人贷网站出台了新的利率规定,利率范围定为10%~24%,信用认证标的利率是借款人自主设定的;人人贷网站规定借款金额必须在3000~500000元,且为50的倍数,借款期限可以为3、6、9、12、18、24个月,还款方式均为等额本息,按月还款。当一个借款人发起一个订单时,投资者可以进行投标,投标的最低金额是50元,因此对于同一个订单,可能会有多个投资者共同参与,每个投资者投资的额度也不尽相同。一个订单自挂出后,网站规定其筹标期限为7天,当所有投资者的出资额度相加等于借款人需要借款的数额时,表示该订单筹款成功,即使期限未到,订单也不能再继续接受投资者投标;如果在规定期限内未能筹到款或是所有投资者愿意投资的额度相加不能达到订单需要借款的数额时,则订单筹款失败。

2. 变量设置

本文剔除了香港、澳门、台湾三个地区(这三个地区订单数量较少),保留全国31个省份。不同地区的订单成交情况如表1所示。

表1 各省份贷款订单状态的描述性统计

| 省 份 | 失败 | 成功 | | | 总计 |
|-----|------|-----|-------|----|------|
| | | 已还清 | 正在还款中 | 违约 | |
| 安徽 | 1286 | 131 | 15 | 13 | 1445 |
| 北京 | 2061 | 317 | 16 | 5 | 2399 |
| 福建 | 3105 | 326 | 36 | 15 | 3482 |
| 甘肃 | 403 | 45 | 4 | 2 | 454 |
| 广东 | 8182 | 646 | 43 | 41 | 8912 |
| 广西 | 1405 | 68 | 11 | 8 | 1492 |
| 贵州 | 737 | 105 | 12 | 4 | 858 |
| 海南 | 389 | 28 | 0 | 1 | 418 |
| 河北 | 1773 | 134 | 20 | 6 | 1933 |
| 河南 | 1968 | 248 | 35 | 15 | 2266 |
| 黑龙江 | 965 | 67 | 6 | 2 | 1040 |
| 湖北 | 1804 | 135 | 13 | 10 | 1962 |
| 湖南 | 1787 | 146 | 11 | 8 | 1952 |
| 吉林 | 759 | 86 | 6 | 8 | 859 |
| 江苏 | 3288 | 646 | 52 | 39 | 4025 |
| 江西 | 1239 | 95 | 6 | 6 | 1346 |
| 辽宁 | 1441 | 118 | 7 | 8 | 1574 |
| 内蒙古 | 667 | 18 | 4 | 2 | 691 |

(续)

| 省 份 | 失败 | 成功 | | | 总计 |
|-----|-------|------|-------|----|-------|
| | | 已还清 | 正在还款中 | 违约 | |
| 宁夏 | 188 | 8 | 3 | 0 | 199 |
| 青海 | 90 | 5 | 1 | 0 | 96 |
| 山东 | 3778 | 494 | 74 | 69 | 4415 |
| 山西 | 1119 | 49 | 6 | 6 | 1180 |
| 陕西 | 1050 | 120 | 7 | 8 | 1185 |
| 上海 | 2061 | 194 | 10 | 17 | 2282 |
| 四川 | 2468 | 113 | 8 | 6 | 2595 |
| 天津 | 591 | 46 | 7 | 4 | 648 |
| 西藏 | 54 | 5 | 1 | 0 | 60 |
| 新疆 | 407 | 19 | 1 | 0 | 427 |
| 云南 | 1129 | 120 | 4 | 10 | 1263 |
| 浙江 | 4239 | 692 | 65 | 60 | 5056 |
| 重庆 | 1028 | 50 | 5 | 2 | 1085 |
| 总计 | 51461 | 5274 | 489 | 53 | 57599 |

根据研究需要，本文设定核心研究变量如下：

订单成功 (*success*)：哑变量，若订单募集成功取值为 1，否则为 0。根据描述性统计可知，在本文样本中，所有订单中大约有 10% 的订单筹款成功。

订单违约 (*default*)：哑变量，若订单违约则取值为 1，否则为 0，由于正在还款的订单无法考量其最终的违约情况，因此不计算在内（当一个订单完成筹款后，如果未能按时还款超过约定时间 30 天后，人人贷网站将这种行为认定为严重逾期，网站对于该订单的状态会显示为“逾期”；另有一些订单逾期时间超过 90 天，人人贷网站将按照对投资者“本金保障”的约定，垫付本金，这时网站对于该订单的状态会显示为“已垫付”。本文将这两种状态下的订单都视作违约订单）。在本文样本中，人人贷网站已成功募集的订单中大约有 6% 左右的订单违约。

订单信息变量：

(1) 利率 (*interest rate*)：人人贷网站的订单利率是由借款人自行设定的，其中 2011 年利率的设定只规定上限不得超出银行同类贷款的 4 倍，2012 年后网站规定利率范围为 10%~24%。根据描述性统计可知，该样本中订单设定的平均利率为 15.7%。

(2) 借款期限 (*month*)：人人贷网站订单的借款期限是由借款人自行设定的，可以为 3、6、9、12、18、24 个月。

(3) 借款金额的对数值 \ln (*requested loan amount*)：人人贷网站规定，信用认证标的借款金额必须为 3000~500000 元的数，并且是 50 的倍数，其中，对于不同信用等级的借款人其最高借款数额需根据网站提供给他人的信贷额度所决定。

借款人信息变量：

(1) 性别：人人贷网站中性别不是必须要披露的信息，部分借款人未披露其性别信息（根据本文的统计，有 18% 的人没有披露其性别信息），因此我们对借款人是否披露其性别设定一个哑变量 (*sex_blank*)，未披露时取值为 1，否则为 0；此外，对于披露性别的借款人，我们设定一个性别哑变量 (*female*)，对女性赋值为 1，否则为 0，根据本文统计，所有借款人中女性借款者大约是 11%，

(2) 信贷额度 (*credit line*)：人人贷网站根据借款人的信用等级、信用评分、其交易记录及其他由借款人所提供的认证信息，综合评定后给每个借款者一个借贷额度，信贷额度主要是对借款人的借款上限做出了限制。

(3) 借款人风险水平 (*risk level*)：人人贷网站根据借款人披露的各类信息及交易记录为每位借款人进行信用评级，具体分为：AA、A、B、C、D、E、HR 七个级别。其中 HR 等级的占到全样本的 89.94%（HR 为信用最差的，相应的风险水平最高），因此我们对其做虚拟变量，信用等级为 HR=1，其他等级的为 0。

(4) 认证信息：人人贷网站中借款人可以提供相应的材料对其信息进行认证，我们将这些认证信息全部处理为虚拟变量，如果进行认证则=1，没有进行认证或认证未通过则=0，这部分变量包括提供身份证复印件认证身份信息 (*idcard_cert*)，提供央行开具的个人征信报告对其信用情况进行认证 (*credit_cert*)，提供工作证复印件或劳动合同对其工作情况进行认证 (*workpermit_cert*)，提供盖有单位公章的收入证明或工资卡银行流水对其收入情况进行认证 (*income_cert*)，提供结婚证认证其婚姻状况 (*merriage_cert*)，提供户口本认证其户籍及住址情况 (*household register_cert*)，提供个人视频 (*video_cert*)，提供手机号进行手机认证 (*mobile_cert*)，提供微博账号进行微博认证 (*weibo_cert*)，提供汽车产权证明对其购车情况进行认证 (*car_cert*)，提供房屋产权证明对其房产情况进行认证 (*house_cert*)。

表 2 给出了变量的描述性统计情况。

表 2 变量描述性统计

| 变量 | 变量个数 | 平均值 | 标准差 | 最小值 | 最大值 |
|------------------------------|-------|----------|----------|------|---------|
| <i>success</i> | 57599 | 0.106564 | 0.308561 | 0 | 1 |
| <i>default</i> | 5649 | 0.066 | 0.249 | 0 | 1 |
| <i>interest rate</i> | 57599 | 15.66793 | 4.271661 | 3 | 24.4 |
| <i>requested loan amount</i> | 57599 | 43499.95 | 102691.7 | 1000 | 1000000 |
| <i>months</i> | 57599 | 10.33856 | 7.51263 | 3 | 36 |
| <i>female</i> | 57599 | 0.113995 | 0.317808 | 0 | 1 |
| <i>sex_blank</i> | 57599 | 0.182712 | 0.386433 | 0 | 1 |
| <i>credit line</i> | 57579 | 8030.115 | 40211.19 | 0 | 1000000 |
| <i>risk level</i> | 57599 | 0.899408 | 0.30079 | 0 | 1 |
| <i>idcard_cert</i> | 57599 | 0.335145 | 0.472045 | 0 | 1 |
| <i>credit_cert</i> | 57599 | 0.025053 | 0.156286 | 0 | 1 |
| <i>workpermit_cert</i> | 57599 | 0.022032 | 0.146788 | 0 | 1 |

| (续) | | | | | |
|-------------------------------|-------|----------|----------|-----|-----|
| 变量 | 变量个数 | 平均值 | 标准差 | 最小值 | 最大值 |
| <i>income_cert</i> | 57599 | 0.021337 | 0.144507 | 0 | 1 |
| <i>merriage_cert</i> | 57599 | 0.47329 | 0.49929 | 0 | 1 |
| <i>householdregister_cert</i> | 57599 | 0.025851 | 0.158693 | 0 | 1 |
| <i>video_cert</i> | 57599 | 0.080088 | 0.271432 | 0 | 1 |
| <i>mobile_cert</i> | 57599 | 0.044515 | 0.206237 | 0 | 1 |
| <i>weibo_cert</i> | 57599 | 0.060331 | 0.238101 | 0 | 1 |
| <i>car_cert</i> | 57599 | 0.070939 | 0.256725 | 0 | 1 |
| <i>house_cert</i> | 57599 | 0.063942 | 0.244652 | 0 | 1 |

三、实证结果

1. 地域歧视的实证检验

首先,我们在表3中计算了各个省市的订单成功率,并对各省份成功的订单计算了违约率。我们将各省份的订单成功率按照由低到高的顺序进行排序,可以看到订单成功率最低的省份是内蒙古,只有3.47%,而最高的省份江苏却达到18.31%,对省份间的差异性进行检验可以发现省份之间的订单成功率有着显著差异。图3可以直观地看到以上基本结果。

表3 各省份订单的成功率及违约率

| 省 份 | 成功率 | 违约率 | 省 份 | 成功率 | 违约率 |
|-----|--------|--------|-----|--------|--------|
| 安徽 | 11.00% | 9.03% | 辽宁 | 8.45% | 6.35% |
| 北京 | 14.09% | 1.55% | 内蒙古 | 3.47% | 10.00% |
| 福建 | 10.83% | 4.40% | 宁夏 | 5.53% | 0.00% |
| 甘肃 | 11.23% | 4.26% | 青海 | 6.25% | 0.00% |
| 广东 | 8.19% | 5.97% | 山东 | 14.43% | 12.26% |
| 广西 | 5.83% | 10.53% | 山西 | 5.17% | 10.91% |
| 贵州 | 14.10% | 3.67% | 陕西 | 11.39% | 6.25% |
| 海南 | 6.94% | 3.45% | 上海 | 9.68% | 8.06% |
| 河北 | 8.28% | 4.29% | 四川 | 4.89% | 5.04% |
| 河南 | 13.15% | 5.70% | 天津 | 8.80% | 8.00% |
| 黑龙江 | 7.21% | 2.90% | 西藏 | 10.00% | 0.00% |
| 湖北 | 8.05% | 6.90% | 新疆 | 4.68% | 0.00% |
| 湖南 | 8.45% | 5.19% | 云南 | 10.61% | 7.69% |
| 吉林 | 11.64% | 8.51% | 浙江 | 16.16% | 7.98% |
| 江苏 | 18.31% | 5.69% | 重庆 | 5.25% | 3.85% |
| 江西 | 7.95% | 5.94% | | | |

注:成功率的计算:各省份成功订单数/各省份订单总数;违约率的计算:各省份违约订单数/(各省份违约订单数+各省份已还清订单数)。

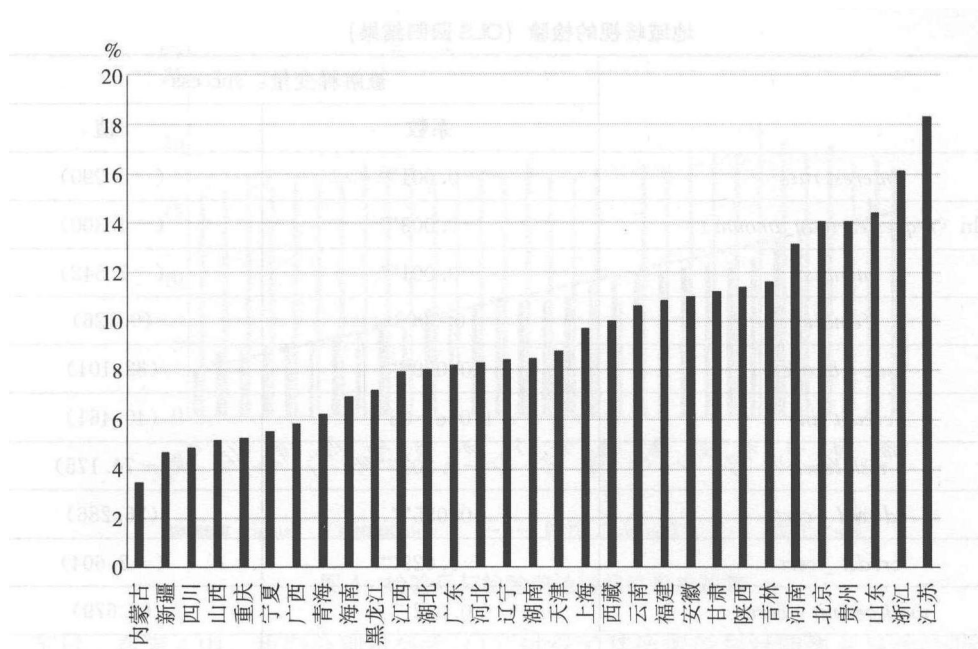


图3 各省份订单成功率

但是如前所述，这种初步的推断并不足以证明借款成功率的地域歧视。为了控制“借款人所披露的信息”以及“订单自身设定的信息”造成的影响，我们需要用公式（1）中的模型设定进行进一步检验。

表4展示了公式（1）的检验结果。从中可以看到该模型的 R^2 达到58.6%，说明本文的变量选取对订单成功率具有较强的解释力度。

从具体的结果来看，订单自身的信息——利率、订单的借款金额以及借款期限对于订单的成功率具有显著为负的作用。该结果说明：当借款人设定的利率越高时，其隐含的违约风险也相对越高，因此订单的成功率就相对越低；借款金额越高其违约的概率越高，因此订单的成功率会随之下降；借款期限越短，订单的成功率越高。

在借款人信息中，我们发现没有披露性别的人对于订单成功率具有显著为正的影 响，结合披露女性信息对于订单没有显著影响，我们认为投资者无法判断其性别的情况下更愿意进行投资的原因是在P2P网络借贷平台中，投资者对于性别没有歧视。借款人信贷额度越高，订单越容易成功，因为借款人信贷额度在某种程度上反映了借款者的信用水平；借款人的信用等级越差，订单的成功率越低；披露的各类认证信息中，对于身份证信息的认证、工作证信息的认证、收入情况的认证、婚姻情况的认证、视频和微博的认证、以及车产和房产的认证对于订单的成功率具有显著为正的效应，说明投资者披露这些能够有效提高其借款订单的成功概率；令人惊奇的是，提交信用征信报告的借款人，往往会降低其借款成功率。

表4 地域歧视的检验 (OLS 回归结果)

| | 被解释变量: success | |
|-----------------------------------|----------------|-----------|
| | 系数 | t 值 |
| <i>interest rate</i> | -0.001*** | (-5.990) |
| <i>ln (requested loan amount)</i> | -0.003*** | (-4.300) |
| <i>months</i> | -0.001*** | (-7.542) |
| <i>female</i> | 0.0003 | (0.126) |
| <i>sex_blank</i> | 0.079*** | (32.101) |
| <i>credit line</i> | 1.05e-06*** | (40.461) |
| <i>risk level</i> | -0.303*** | (-74.175) |
| <i>idcard_cert</i> | 0.055*** | (26.286) |
| <i>credit_cert</i> | -0.029*** | (-2.604) |
| <i>workpermit_cert</i> | 0.042*** | (2.679) |
| <i>income_cert</i> | 0.001 | (0.038) |
| <i>merriage_cert</i> | 0.020*** | (11.234) |
| <i>householdregister_cert</i> | -0.045*** | (-7.061) |
| <i>video_cert</i> | 0.290*** | (70.777) |
| <i>mobile_cert</i> | -0.007 | (-1.354) |
| <i>weibo_cert</i> | 0.062*** | (15.186) |
| <i>car_cert</i> | 0.100*** | (22.465) |
| <i>house_cert</i> | 0.083*** | (18.042) |
| <i>province</i> | Controlled | |
| <i>Constant</i> | 0.360*** | (39.042) |
| Observations | 57, 579 | |
| R ² | 0.586 | |

其他省份与对照组差异性检验: $F(30, 57530) = 4.57$ Prob > F = 0.00

其他省份间差异性检验: $F(29, 57530) = 4.46$ Prob > F = 0.00

注: 我们在模型设定中以江苏为对照组。限于篇幅, 各省份的回归系数没有详细列示, 读者可以通过图4看到各省份之间的借款成功率差异。***代表在1%水平显著; **代表在5%水平显著; *代表在10%水平显著。

我们提取控制其他因素之后式(1)中各省份的系数^①, 并与描述性统计数据中的成功率进行了对比, 结果发现地域间的差异有所降低。在图4中, 虚线为柱状图的趋势线, 从中可以看到, 在控制了“借款人所披露的信息”以及“订单自身设定的信息”之后, 省份之间借款成功率的差异明显降低。

① 由于表4中使用的是OLS回归, 因此各省份哑变量的系数从经济意义上就表示了其与对照组之间借款成功率的差异。

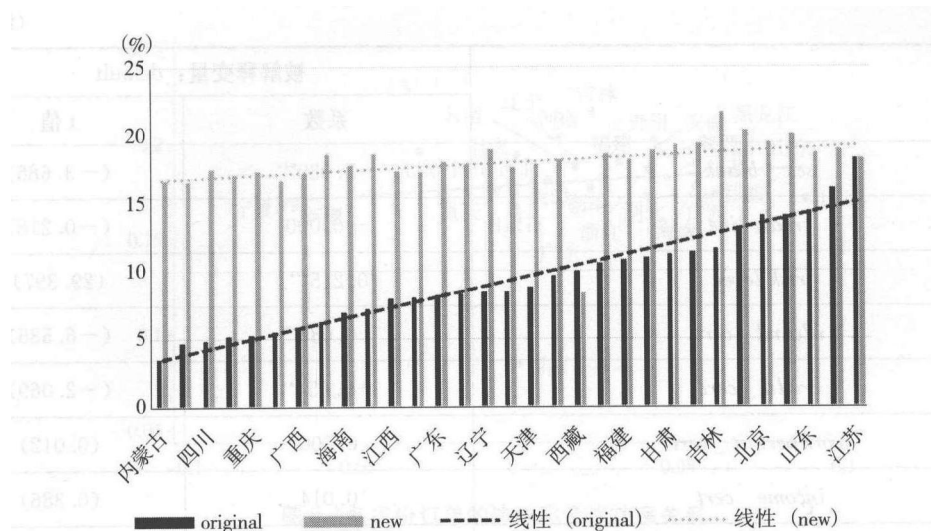


图4 检验后与检验前地域间订单成功率

不过，在表4中，我们分别对公式(1)进行了其他省份与对照组差异性检验以及省份间差异性检验，结果发现省份间订单的成功率仍然有显著性差异。也就是说，在控制了其他因素之后，我国P2P网络借贷平台上仍存在有明显的地域歧视。

2. 地域歧视是否理性

表4的结果说明P2P网络借款中是存在地域歧视的。根据图1的实证设计思路，我们将进一步考察这种地域歧视是否为理性行为——该结果通过公式(2)中的模型设定进行进一步检验。

表5展示了公式(2)的检验结果。从订单的信息来看，借款金额越高、借款期限越长的订单其违约概率越高。从借款人的信息来看，未披露性别的借款人其违约的概率相对较低^①；风险水平越高即借款人信用越差，则其违约的概率越高；借款人披露的认证信息中，认证过身份证、征信报告以及视频的借款人，其违约率相对更低，而披露手机和微博认证的借款人其违约率相对较高。进一步地，在表5中进行其他省份与对照组差异性检验以及省份间违约率差异性检验，我们发现省份间的违约率也具有显著性差异。

表5 地域歧视是否理性的检验 (OLS 回归结果)

| | 被解释变量: default | |
|-----------------------------------|----------------|----------|
| | 系数 | t 值 |
| <i>interest rate</i> | 0.001 | (0.572) |
| <i>ln (requested loan amount)</i> | 0.027*** | (7.513) |
| <i>months</i> | 0.007*** | (10.570) |
| <i>female</i> | -0.004 | (-0.295) |

① 这与前文中未披露性别的借款人借款成功率更高相一致。

(续)

| | 被解释变量: default | |
|-------------------------------|----------------|----------|
| | 系数 | t 值 |
| <i>sex_blank</i> | -0.030*** | (-3.686) |
| <i>credit_line</i> | -0.000 | (-0.218) |
| <i>risk_level</i> | 0.215*** | (29.397) |
| <i>idcard_cert</i> | -0.086*** | (-6.536) |
| <i>credit_cert</i> | -0.054** | (-2.069) |
| <i>workpermit_cert</i> | 0.000 | (0.012) |
| <i>income_cert</i> | 0.014 | (0.386) |
| <i>merriage_cert</i> | 0.005 | (0.695) |
| <i>householdregister_cert</i> | 0.010 | (0.941) |
| <i>video_cert</i> | -0.016** | (-2.255) |
| <i>mobile_cert</i> | 0.031*** | (3.724) |
| <i>weibo_cert</i> | 0.021*** | (2.578) |
| <i>car_cert</i> | -0.000 | (-0.034) |
| <i>house_cert</i> | 0.003 | (0.434) |
| <i>province</i> | Controlled | |
| <i>Constant</i> | -0.218*** | (-5.248) |
| Observations | 5645 | |
| R ² | 0.211 | |

其他省份与对照组差异性检验: $F(30, 5596) = 2.91$ Prob > F = 0.0000

其他省份间差异性检验: $F(29, 5596) = 2.99$ Prob > F = 0.0000

注: 该模型设定中仍以江苏为对照组。限于篇幅, 各省份的回归系数没有详细列示, 但读者可以通过图 5 看到各省份之间的违约率差异。*** 代表在 1% 水平显著; ** 代表在 5% 水平显著; * 代表在 10% 水平显著。

表 4 和表 5 分别为我们提供了到控制了其他因素之后不同省份的平均借款成功率和平均违约率。接下来, 我们将比较各省份的违约率和借款成功率之间的关系。

图 5 通过散点及趋势线的方式为我们直观地提供了各省份平均违约率与平均成功率之间的关系。根据图 1 的逻辑框架, 对于理性的地域歧视, 我们应该看到违约率和借款成功率之间有显著负相关关系。但是从图 5 中可以看到, 二者之间的趋势线呈现正相关关系, 该结果说明违约率越高的省份其借款成功率也越高, 这意味着不同省份之间借款成功率的差异并非源于其违约率差异, 因此可以判定前文所得到的地域歧视为一种非理性的歧视行为。

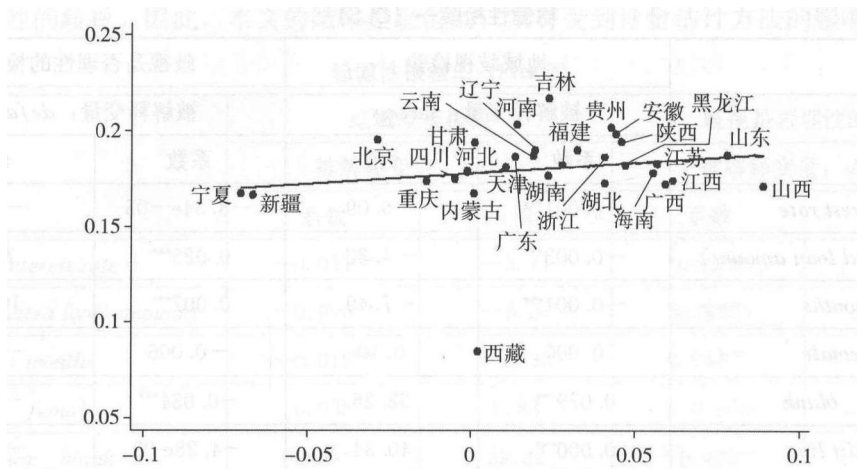


图5 各省份订单的违约率及成功率关系

实际上，我们可以通过回归的方式来检验订单成功率与违约率之间的关系，表6向我们展示了这一结果。从中可以看到，表6的模型解释力度仅为0.2%，而各省份平均违约率对其平均借款成功率没有显著的解释能力。该结果再次说明前文所看到的地域歧视更多地表现为一种偏好性的歧视，其背后没有明显的经济原理，是一种非理性行为。

表6 各省份订单成功率与违约率间的相关性检验

| | 被解释变量: $success_n$ | |
|--------------|--------------------|---------|
| | 系数 | t 值 |
| $default_n$ | 0.111 | (1.03) |
| Constant | 0.178*** | (39.78) |
| Observations | 31 | |
| 调整后的 R^2 | 0.002 | |

注：下标 n 表示不同的省份；*** 代表在 1% 水平显著；** 代表在 5% 水平显著；* 代表在 10% 水平显著。

3. 稳健性检验

作为对上文实证分析的稳健性检验，我们将地域分为被歧视组和未被歧视组两组。被歧视的一组按歧视程度的严重性排序为西藏、新疆、内蒙古、海南、宁夏、江西、山西、广西、湖南、广东、湖北、重庆、四川、河北、天津，共计 15 个，针对这些省份，我们设置哑变量 $discrimination=1$ ，其他省份 $discrimination=0$ 。

在表7的稳健性检验中，我们使用 OLS 回归对订单成功率及订单违约率进行解释。从中可以看到，在控制了其他变量之后，对于受到歧视 ($discrimination=1$) 的省份，其订单成功率的确更低，这说明我们之前选取的被歧视省份在 P2P 网络借款时平均来说的确遭遇了地域歧视问题。

另一方面，对于歧视是否理性的检验表明，控制了其他因素之后， $discrimination$ 显著为负，说明受到地域歧视的组其违约率反而更低，该结果说明地域歧视是一种非理性的歧视——这说明我们前面的结果是稳健的。

表 7 稳健性检验— (OLS)

| | 地域歧视检验 | | 歧视是否理性的检验 | |
|-------------------------------------|-----------------------|--------|-----------------------|-------|
| | 被解释变量: <i>success</i> | | 被解释变量: <i>default</i> | |
| | 系数 | t 值 | 系数 | t 值 |
| <i>interest rate</i> | -0.001*** | -6.09 | -8.34e-05 | -0.08 |
| ln (<i>requested loan amount</i>) | -0.003*** | -4.32 | 0.025*** | 7.18 |
| <i>months</i> | -0.001*** | -7.49 | 0.007*** | 10.55 |
| <i>female</i> | 0.000 | 0.10 | -0.006 | -0.46 |
| <i>sex_blank</i> | 0.079*** | 32.26 | -0.034*** | -4.25 |
| <i>credit line</i> | 0.000*** | 40.34 | -4.28e-09 | -0.11 |
| <i>risk level</i> | -0.303*** | -74.28 | 0.214*** | 29.49 |
| <i>idcard_cert</i> | 0.055*** | 26.35 | -0.085*** | -6.53 |
| <i>credit_cert</i> | -0.030*** | -2.64 | -0.034 | -1.33 |
| <i>workpermit_cert</i> | 0.041*** | 2.64 | 0.005 | 0.15 |
| <i>income_cert</i> | 0.002 | 0.10 | -0.009 | -0.24 |
| <i>merriage_cert</i> | 0.020*** | 11.23 | 0.013* | 1.78 |
| <i>householdregister_cert</i> | -0.045*** | -7.08 | 0.008 | 0.76 |
| <i>video_cert</i> | 0.289*** | 70.90 | -0.012* | -1.68 |
| <i>mobile_cert</i> | -0.007 | -1.45 | 0.032*** | 3.81 |
| <i>weibo_cert</i> | 0.062*** | 15.38 | 0.021*** | 2.57 |
| <i>car_cert</i> | 0.100*** | 22.57 | 0.001 | 0.07 |
| <i>house_cert</i> | 0.083*** | 18.06 | 0.005 | 0.73 |
| <i>discrimination</i> | -0.014*** | -8.07 | -0.018*** | -2.74 |
| Constant | 0.353*** | 46.20 | -0.205*** | -5.52 |
| Observation | 57579 | | 5645 | |
| R ² | 0.585 | | 0.200 | |

注:***代表在 1%水平显著;**代表在 5%水平显著;*代表在 10%水平显著。

此外,本文之前的结果都是基于 OLS 的回归结果,其目的在于让我们更方便地使用 OLS 的回归系数。但是,事实上,订单的借款成功率和违约率都是二元选择变量,这种变量更适合 probit 模型。为了排除计量方法选取的偏误,我们仍和表 7 一样将地域分为被歧视组和未被歧视组两组,使用 probit 模型来检验 *discrimination* 对订单成功情况以及违约情况进行检验。

表 8 描述了 probit 模型的估计结果,从表中可以看到:当被解释变量为订单是否募集成功时 (*success*), *discrimination* 显著为负,说明我们人为划分的“被歧视组”其借款成功率显著更低,这意味着确实存在地域歧视现象;另一方面,对歧视的理性程度检验表明, *discrimination* 也显著为负,说明被地域歧视的组其违约率更低,该结果再次证明地域歧视是

一种非理性的歧视。因此，本文的结果是稳健的，没有受到计量估计方法的影响。

表 8 稳健性检验二 (Probit)

| | 地域歧视检验 | | 歧视是否理性的检验 | |
|-------------------------------------|-----------------------|--------|-----------------------|--------|
| | 被解释变量: <i>success</i> | | 被解释变量: <i>default</i> | |
| | 系数 | z 值 | 系数 | z 值 |
| <i>interest rate</i> | -0.011*** | -3.71 | 0.027** | 2.14 |
| ln (<i>requested loan amount</i>) | -0.098*** | -8.37 | 0.382*** | 7.36 |
| <i>months</i> | -0.019*** | -9.32 | 0.061*** | 8.37 |
| <i>female</i> | 0.035 | 0.83 | -0.159 | -1.13 |
| <i>sex_blank</i> | 0.809*** | 29.32 | -0.478*** | -4.49 |
| <i>credit line</i> | 0.000*** | 23.87 | -1.13e-07 | -0.17 |
| <i>risk level</i> | -0.923*** | -31.37 | 2.420*** | 18.5 |
| <i>idcard_cert</i> | 0.990*** | 31.98 | -0.813*** | -6.11 |
| <i>credit_cert</i> | -0.053 | -0.47 | -0.936* | -1.65 |
| <i>work permit_cert</i> | 0.024 | 0.17 | -0.036 | -0.05 |
| <i>income_cert</i> | 0.106 | 0.72 | 0.109 | 0.13 |
| <i>merriage_cert</i> | 0.284*** | 11.58 | 0.015 | 0.18 |
| <i>householdregister_cert</i> | -0.158*** | -3.10 | 0.068 | 0.44 |
| <i>video_cert</i> | 1.037*** | 33.48 | -0.304*** | -3.33 |
| <i>mobile_cert</i> | -0.034 | -0.85 | 0.343*** | 2.92 |
| <i>weibo_cert</i> | 0.020 | 0.59 | 0.103 | 0.89 |
| <i>car_cert</i> | 0.266*** | 7.85 | -0.126 | -1.4 |
| <i>house_cert</i> | 0.231*** | 6.58 | 0.034 | 0.38 |
| <i>discrimination</i> | -0.141*** | -5.93 | -0.189** | -2.33 |
| Constant | -0.701*** | -6.16 | -6.577*** | -11.22 |
| Observation | 57579 | | 5645 | |
| Pseudo R ² | 0.622 | | 0.4365 | |

注:***代表在1%水平显著;**代表在5%水平显著;*代表在10%水平显著。

四、结论与启示

“歧视”是金融学中重要的研究领域。对于新兴的互联网金融市场而言，一个重要的话题是：这种非面对面形式的网络借贷是否消除了歧视，营造了更为公平的借贷环境？

本文把研究聚焦在“地域歧视”这一富有中国特色的特殊歧视上，利用人人贷 P2P 网络贷款平台的交易数据，对地域歧视的存在性以及合理性进行了深入研究。我们的研究发现：在控制了订单信息以及借款人信息之后，各省份之间的订单成功率仍然存在显著的地域差异，说明人人贷 P2P 借贷平台的交易中存在地域歧视问题。

进一步,我们研究了这种地域歧视是理性的还是非理性的。本文研究发现,平台上那些被歧视的省份,其订单的违约率并没有显著地高于其他未被歧视的省份,这表明该平台上表现出来的地域歧视本质上属于“非有效的偏好歧视”,是一种非理性的行为。

本文的研究针对新兴的互联网金融领域 P2P 网络借贷行为,本文的结果丰富了互联网金融的相关研究,对于国内外学者和投资者了解中国 P2P 网络信贷市场具有一定实际意义。同时,本文所揭示的一些非理性结果对于 P2P 网络平台更为合理地设计其风险评估模型也提供了一定的数据支持和理论依据。

此外,本文的局限之处在于尚未进一步探讨造成地域歧视的成因。在未来的研究中,可以从区域间的经济发达程度以及文化差异等角度探讨“地域歧视”的根源,进而寻找解决之道。

参考文献

- [1] Arrow K. , 1973, *The Theory of Discrimination* [M], In *Discrimination in Labor Markets*, ed. Ashenfelter Orley and Albert Rees, Princeton University Press.
- [2] Bachmann A. , A. Becker, D. Buerckner, M. Hilker, M. Lehmann and P. Tiburtius, 2011, *Online Peer-to-peer Lending-a Literature Review* [J], *Journal of Internet Banking and Commerce*, 16, 1~18.
- [3] Becker G. , 1957, *The Economics of Discrimination* Chicago [M], University of Chicago Press.
- [4] Blanchflower D. G. , Levine P. B. and D. J. Zimmerman, 2003, *Discrimination in the Small-business Credit Market* [J], *Review of Economics and Statistics*, 85, 930~943
- [5] Cavalluzzo K. S. , and J. D. Wolken, 2005, *Small Business Loan Turndowns, Personal Wealth and Discrimination* [J], *Journal of Business*, 78, 2153~2177.
- [6] Duarte J. , S. Siegel, and L. Young, 2012, *Trust and Credit: the Role of Appearance in Peer-to-peer Lending* [J], *The Review of Financial Studies*, 25, 2455~2484.
- [7] Goering J. and R. Wienk, 1996, *Mortgage Lending, Racial Discrimination and Federal Policy* [M], Urban Institute Press, Washington D. C.
- [8] Ladd H. F. , 1982, *Women and Mortgage Credit* [J], *American Economics Review*, 72, 166~170.
- [9] Lin M. , N. R. Prabhala, and S. Viswanathan, 2013, *Judging Borrowers by the Company They Keep: Social Networks and Adverse Selection in Online Peer-to-peer Lending* [J], *Management Science*, 59, 17~35.
- [10] Phelps E. , 1972, *The Statistical Theory of Racism and Sexism* [J], *American Economic Review*, 62, 659~661.
- [11] Pope D. G. and J. R. Sydnor, 2011, *What's in a Picture? Evidence of Discrimination from Prosper.Com* [J], *The Journal of Human Resources*, 46, 53~92.
- [12] Quercia R. and M. , Stegman, 1992, *Residential Mortgage Default: A Review of the Literature* [J], *Journal of Housing Research*, 3, 341~79.
- [13] Schafer R. and H. F. Ladd, 1981, *Discrimination in Mortgage Lending* [M], Massachusetts Institute of Technology Press, Cambridge.

(责任编辑:王喜峰)