

doi:10.19920/j.cnki.jmsc.2025.02.009

数字足迹与违约预测：基于消费信贷的证据^①

胡俊¹, 李强^{1*}, 曾勇¹, 田轩², 刘嵩³

(1. 电子科技大学经济与管理学院, 成都 611731; 2. 清华大学五道口金融学院, 北京 100083;
3. 四川新网银行股份有限公司, 成都 610094)

摘要: 本研究利用国内某金融科技借贷机构借款人层面的独特数据,在将借款人信息分为自我报告、央行征信、电商信用评分和网购行为共四类信息的基础上,重点考察电商信用评分和网购行为两类数字足迹的违约预测作用及其提高信贷可得性的内在机理。以自我报告和央行征信两类基础信息的预测能力为基准,研究结果表明:两类数字足迹信息能够将违约预测的准确性提高约50%,其中电商信用评分的预测能力明显高于其他三类信息,网购行为足迹不仅能达到与自我报告信息相当的预测效果,而且相对于电商信用评分还具有额外的信息含量。进一步,针对数字足迹提高信贷可得性机理的检验结果表明,利用两类数字足迹进行违约预测,不仅可为传统征信严重缺失的信用白户建立数字信用,还能对信用质量被传统征信错误低估的借款人进行更为准确的风险评估。

关键词: 金融科技; 数字足迹; 违约预测; 信贷可得性

中图分类号: F830.589 **文献标识码:** A **文章编号:** 1007-9807(2025)02-0115-25

0 引言

党的十九届五中全会明确指出要全面促进消费,“十四五”规划也提出要增强消费对经济发展的基础性作用,而依托金融科技提高金融普惠性是全面促进消费的重要手段。过去的十多年里,我国消费信贷规模快速增长^②,但由于小额、分散、无抵押等的贷款特征,加之我国个人征信体系尚不健全,消费信贷领域的信息不对称问题尤为严重,传统信贷模式也存在授信成本高、风控难度大等问题,这导致我国消费信贷市场长期存在结构性失衡、覆盖范围有限、信贷约束严重的问题^③。所幸的是,近年来以大数据和机器学习为代表的

数字技术与消费金融业务深度融合,为促进消费升级和畅通国内大循环注入了新动力。

在消费信贷市场,无论是新进的金融科技公司,还是在位的持牌金融机构,基于金融科技的借贷模式(简称:金融科技借贷, fintech lending)至少存在三个方面的独特优势。首先,申请、审批、放款、还款的全业务流程可实现线上化和自动化,没有或极少有人工参与^[1],拓展了服务边界^[2],提高了服务效率^[3]。其次,智能手机、传感器等智能设备终端及大数据、云计算、物联网等技术的广泛使用,大大降低了数据采集、存储和使用的成本,特别是网络购物和社交媒体等在线活动的普及,

① 收稿日期: 2020-07-23; 修订日期: 2023-02-03。

基金项目: 国家自然科学基金资助项目(71972027; 71790591); 国家杰出青年科学基金资助项目(71825002); 四川省软科学项目(2022JDR0290)。

通讯作者: 李强(1980—), 男, 山西吕梁人, 博士, 教授, 博士生导师。Email: liq@uestc.edu.cn

② 根据《2019 消费金融行业发展报告》数据显示: 2010 年—2019 年 10 月, 我国金融机构短期消费贷款规模由 0.96 万亿增长到 9.60 万亿, 增长了近 10 倍。

③ 《2019 年中国消费金融发展报告》指出, 我国消费金融总体覆盖率远低于发达国家, 长尾客户覆盖力度不足, 截至 2017 年, 我国有将近 40% 的成年人未获得过消费金融服务。

催生了大量以“数据资产化”为商业模式的大数据公司. 通过合作或购买等方式, 金融机构能够利用关于借款人画像的多维异构数据进行信用风险评估^[4]. 例如, 金融机构不仅可获取借款申请时借款人所用手机的型号、系统类型等设备数据, 还可从第三方购买综合的金融科技信用评级(如芝麻信用评级), 以及从电商平台获得更为底层的网购行为数据. 最后, 相对于 Logit 等传统线性模型, 机器学习算法的使用不仅能有效捕捉数据间的非线性关系, 还能处理高维稀疏的数据结构问题^[5, 6].

需要指出的是, 在消费信贷领域, 小额、分散、无抵押的贷款特征导致采用贷后监督来降低借款人道德风险的难度和成本较大, 而通过提升违约预测能力来降低贷前的逆向选择风险, 是实施风险管控和增加信贷供给的重要途径. 目前的中国情境下, 央行征信系统仅能有效覆盖大约四分之一的人群^[7], 存在大量因央行征信严重缺失而常被传统金融机构拒贷的信用白户, 而利用传统征信之外的非传统数据, 特别是用户在使用互联网过程中自然留下的易于获取、覆盖广泛的数字足迹, 为其建立数字信用, 有望成为提高信贷市场效率和促进普惠金融发展的重要手段^[8].

从已有研究来看, 如 Goldstein 等^[9]所言, 尽管“大数据 + 机器学习”的违约预测方法被广泛运用于实践, 但学术研究远滞后于业界应用, 其中关于数字足迹违约预测的作用及其机理的研究更是缺乏. 近期极少数文献利用独特的账户数据, 通过比较第三方金融科技信用评级和传统征信的违约预测效果, 提供了金融科技信用评级具有独特优势的证据^[10-12]. 然而, 当下的业界实践并不满足于仅使用第三方提供的综合信用评级, 而是还会大量使用更为底层的非传统数据, 但遗憾的是, 已有研究尚未能很好地打开二者如何相互作用以及怎样共同影响借贷效率的“黑箱”. Berg 等^[4]利

用德国某家具电商平台数据, 研究发现设备系统类型、网页访问方式、注册邮箱等数字足迹具有与传统信用评级同等的违约预测效果. 王正位等^[13]利用有无信用卡和有无“花呗”额度分别代表借款人的传统征信和消费行为, 研究发现有无“花呗”额度相对于传统征信具有额外的违约预测作用, 特别是能够提高对薄弱人群的风险识别效率, 但由于“花呗”额度仍属于与信用评分类似的综合评价结果, 同样无法回答其究竟是由于捕捉了什么样的消费偏好或行为特征而具有违约预测作用.

基于此, 本研究利用国内某领先的持牌金融科技借贷机构借款人层面的独特数据^④, 将用于违约预测的信息分为四大类: 申请借款时自愿填写的自我报告信息、反映央行征信的银行卡信息、来自第三方的电商信用评级、反映经济状况或个人特质的网购行为足迹. 其中, 前两类被界定为基础信息, 后两类信息分别来自国内两家头部的电商平台, 且无论是网购行为足迹, 还是电商信用评级依赖的购物、支付、转账、理财等底层数据, 都是用户在使用互联网过程中自然留下的痕迹, 被界定为数字足迹^⑤. 利用 AUC 值 (area under curve) 作为不同模型预测能力的评价指标, 本研究重点考察数字足迹相对于基础信息、网购行为足迹相对于电商信用评级的违约预测能力, 并进一步对数字足迹提高信贷可得性的内在机理和网购行为足迹能够识别违约风险的经济机制进行检验. 研究结果表明, 两类数字足迹均对违约概率具有显著的预测作用, 二者同时使用能够将预测准确性提高约 50%, 其中电商信用评级的预测能力最强, 网购行为足迹不仅具有与自我报告信息相当的预测能力, 还相对于电商信用评级具有额外的信息含量. 进一步针对信贷可得性机理的研究结果表明: 数字足迹是通过为传统征信严重缺失的信用白户建立数字信用, 以及为信用质量被传统

④ 需要特别强调, 出于隐私保护, 本研究所用数据中任何与客户隐私相关的信息, 均进行了加密或脱敏处理, 本研究无法从所用数据中识别出现实中的具体客户.

⑤ 数字足迹 (digital footprints) 指用户在使用互联网和智能终端过程中产生的所有关于线上活动的信息^[14]. 与此同时, 当前金融科技借贷机构经用户授权, 可从多个平台获取底层数据. 因此, 利用来自不同平台的数据, 不仅为本文提供了独特的研究机会, 而且也更加契合当前的金融科技借贷实践.

征信错误低估的借款人提供更为准确的风险评估而发挥作用。针对网购行为足迹用于违约预测的机制检验结果表明,网购金额一定程度上与个人的消费水平及经济状况有关,凌晨下单则能反映个人的自控力,取消订单和货到付款可以反映个人的履约能力和履约意愿。

本研究的创新与贡献概括为三点。第一,现有研究指出了金融科技信用评分在违约预测方面的良好表现,但大都是将其直接与传统征信的违约预测能力进行比较,而金融科技信用评分是算法输出的综合评价结果,该做法不但难以回答金融科技信用评分具有优势的原因何在,更是缺乏底层非传统数据和综合信用评分如何共同影响借贷效率的机理考察。本研究首次利用同时包含多种类型信息的独特数据,通过数字足迹相对于传统征信、网购行为相对于电商信用评分两个层面的违约预测效果比较,不仅揭示了数字足迹提升违约预测能力进而提高信贷可得性的机理,还指出了网购行为足迹和电商信用评分在违约预测中的互补作用,丰富和拓展了金融科技借贷领域的研究。

第二,创新性地探讨并定量考察了网购足迹能够用于违约预测的微观机制。以 Jiang 等^[11]和王正位等^[13]为代表的少数文献,基于网购数据被广泛用于构建信用评分的业界实践与经验推断,定性讨论了网购行为能够识别个人违约的可能原因。本研究则从订单金额体现消费收入水平、下单时间反映自控力、订单取消和支付方式选择反映履约能力与意愿等多个维度,理论上探讨了网购足迹相关指标通过反映消费偏好或行为特征而能够用于违约预测的经济机制,并提供了相应的实证证据,探索性地建立了非传统数据与个人行为特征之间的理论联系。

第三,关于网购行为足迹相对于电商信用评分具有额外违约预测能力的研究结论,可为金融科技借贷机构的风控模式选择提供决策依据。现实中,金融科技借贷机构不仅会直接购买第三方信用评分,还会购买大量更为底层的非传统数据。显然,究竟是采用直接购买第三方信用评分的外部模式,还是购买底层非传统数据来自建风控系

统的内部模式,取决于二者在违约预测中的替代或互补关系。本研究发现,更为底层的网购行为足迹和电商平台提供的综合信用评分具有互补作用,因此,借贷机构不能千篇一律地只依赖外部科技巨头的技术输出,而应充分利用非传统数据自建或优化内部风控体系,才是建立市场竞争优势的关键所在。

特别说明的是 Berg 等^[4]是与本研究最接近的研究,但本研究与之存在如下四点不同:1)从数据性质和信息特征来看,Berg 等使用的设备类型、访问方式和邮箱地址等数字足迹,均属于特定时点上的单次采样数据,本研究使用的借款申请前的逐笔网购订单数据,属于特定时段上的连续采样数据,具有时域多尺度性和信息噪声小的特点,对个体行为特征的刻画具有更好的稳定性和一致性,不仅有利于避免借款人的信息操纵行为,还能通过更好刻画借款人长期的习惯或行为特征而具有稳定的预测效果;2)从违约预测能力的考察层面来看,Berg 等只将数字足迹与传统信用评分进行了比较,本研究则不仅指出了两类数字足迹与传统征信的互补作用,还进一步比较了更为底层的网购足迹信息相对于电商信用评分的预测能力;3)从研究场景来看,在 Berg 等的电商购物场景下,商品和资金的提供主体并不独立,是否如期付款很大程度上还会受到商品质量、售后服务、广告营销等混淆因素的影响,并非严格意义上的借贷场景,本研究则直接考察真实的消费信贷场景;4)从中国情境下的现实应用来看,本研究关注的两类数字足迹更加契合金融科技借贷的业界实践。一方面,Berg 等构造的近一半特征变量依赖邮箱地址信息,但在邮箱普及率远不及欧美国家且非实名认证的中国情境下,该信息不但覆盖人群有限(尤其是长尾客户),而且用于身份验证的有效性也较低,从而实践中金融机构出于用户体验考虑通常不会收集该类信息。另一方面,手机 APP 已成为提供金融服务的主要渠道,但该渠道通常不会涉及网购场景下的页面访问方式(比价或广告网站导流)和设备类型(手机或电脑)等足迹。

1 文献综述与理论分析

Stigleitz 和 Weiss^[15] 的开创性研究指出,借贷双方信息不对称导致的逆向选择和道德风险,是产生信贷配给的根本原因. 随后的研究从软信息与关系型借贷^[16]、信息共享与信息中介^[17]等不同角度,探讨了传统借贷领域信息生产及处理机制对提高信贷效率的作用. 近年来,数字技术的飞速发展不仅使得数据呈现爆炸式增长,也加速了金融与科技的深度融合^[18]. 金融科技借贷机构不仅能够通过大数据技术批量采集和存储借款人在数字环境中留下的丰富痕迹,还可利用机器学习算法进行实时分析和预测^[19],以至于大数据与机器学习算法的运用成为缓解金融市场信息摩擦的重要途径^[20].

在金融科技借贷的相关研究方面,较早以实证研究为主的大量文献主要利用 P2P 网络借贷数据,综合运用心理学、社会学和传播学等学科的理论与方法,考察了借款人自我报告的各类信息通过缓解信息不对称而对借贷效率的影响. 廖理和张伟强^[21] 和李强^[22] 等对相关研究进行了较为全面的综述. Iyer 等^[23] 在将借款人的自我报告信息进行分类的基础上,以 AUC 值作为评价指标,给出了自我报告信息相对于传统信用评级具有额外信息含量的证据.

随着业界实践的不断探索,一些大数据公司、成功转型的 P2P 平台、大型电商平台等金融科技公司发展,凭借其在流量、数据、风控等方面的优势,纷纷采用“数据资产化”的商业模式,专注于面向持牌金融机构进行技术输出. 在此背景下,相关研究开始重点关注第三方提供的金融科技信用评级相对于传统征信在违约预测中的效果. Jagtiani 和 Lemieux^[10] 利用 Lending Club 的数据研究发现, Lending Club 的信用评级和 FICO 信用分之间的相关性从 2007 年的 80% 降至 2014 年—2015 年的 35%; 同时, Lending Club 的信用评级的违约预测效果更好,特别是能够筛选出那些 FICO 信用分低但真实违约风险未必高的借款人. Frost 等^[24] 利用阿根廷金融科技公司 Mercado Credit 的

数据, 同样也发现金融科技信用评级比传统信用评级对小微企业的违约预测效果更好. 无独有偶, Jiang 等^[11] 和 Gambacorta 等^[12] 分别针对我国两家金融科技提供的信用评级, 一致发现金融科技信用评级比央行征信或银行内部评分具有更好的违约预测效果. Jiang 等^[11] 更是明确指出, 金融科技信用评级使用了高频、实时且难以伪造的个人行为数据, 可通过提供更多的信息含量和修正错误报告的信息两种途径来提高其违约预测能力.

除了直接使用第三方信用评分外, 金融机构还可利用粒度细、维度多、可挖掘性强的底层数据来自建内部风控系统, 这引起了少数文献对底层非传统数据的进一步关注. Berg 等^[4] 利用德国某家“先买后付”家具电商平台的 270 399 个用户数据, 利用用户注册和购买商品时留下的设备系统(例如是否是 iOS 系统)、邮箱类型(如 Gmail 和 Yahoo)、邮箱地址(如大小写, 是否包含姓名)、页面访问方式、购买时间等数字足迹, 研究发现, 通过反映经济状况, 声誉、冲动消费和自控能力等个人特质, 数字足迹能够达到与传统信用评级相当的违约预测效果. Gong 等^[25] 利用国内一家信用卡公司的账户数据, 发现借款人游戏充值的频率越高、金额越大, 贷后违约风险越高, 而游戏充值行为可由借款人的自控力所解释.

整体而言, 无论是金融科技信用评级与传统征信的违约预测能力比较, 还是个人行为足迹在违约预测中的应用, 现有研究共同的重要启示在于: 传统征信信息的违约预测能力尚为有限, 大数据及机器学习算法在个人消费信贷领域具有重要作用; 少数文献也提供了金融科技信用评级或非传统数据相对于传统征信具有信息优势的典型证据. 然而, 从金融科技借贷实践来看, 金融机构原则上可以同时使用自我报告、传统征信、第三方征信、底层非传统数据等多种信息, 不同信息之间既可能相互补充, 也可能彼此冗余^[26], 这使得考察特别是比较每类信息相对于其他信息的额外预测能力尤为重要. 金融机构究竟是只购买第三方信用评分, 还是底层非传统数据来加工利用, 亦或二

者同时购买再叠加利用,取决于不同类型信息在违约预测中的边际贡献。

具体到本研究所使用的借款人数据,除了自我报告信息和反映央行征信的银行卡信息外,还包括了来自国内某大型电商平台的电商信用评分和来自另一家大型电商平台的网购订单数据。在自我报告信息方面,尽管P2P借贷领域的大量研究指出了其用于识别违约风险的积极作用,但此类信息存在错误报告、易被操纵等问题^⑥,甚至可能加剧贷前的逆向选择风险^[27]。在传统征信方面,尽管人民银行提供的央行征信报告也涵盖了个人社保、公积金、借款、还款、逾期等信息,但与欧美发达国家的个人征信体系的差距仍然较大。更为重要的是,无论是美国的FICO信用分还是央行征信,更多依赖的是以“借债还钱”记录为主的硬信息,不仅对征信记录缺失的“信用白户”的违约预测作用十分有限^[13],还由于过度关注负面记录,导致在自然灾害、金融危机、疫情等外部偶然因素的冲击下会进一步加剧借款人的信贷约束^[28]。

区别于传统征信信息,大型电商平台提供的信用评分通常使用的个人长期且无意间留下的数字足迹,具有频率高、覆盖广、伪造难度大等特点,能够通过反映借款人的偏好习惯、行为特质等而发挥良好的违约预测作用^[11, 12]。特别是对于“信用白户”而言,即便没有传统征信看重的借贷历史,他们也会在网购、社交等线上活动中留下丰富的数字足迹,利用这些信息进行违约预测,可以提高他们的信贷可得性。然而,由于实践中存在“数据孤岛”问题,提供金融科技信用评分的平台通常难以直接使用借款人在其他平台留下的数字足迹,即便特定平台提供的信用评分已经包含了丰富的信息含量,进一步使用其他平台的数字足迹依然有机会提供额外的信息含量。即使不同平台具有相似业务,但他们的数据收集、处理、分析能力也不完全相同,这也意味着来自不同平台的底层非传统数据和电商信用评分在违约预测方面可能具有互补作用。

相对于社交媒体信息等非传统数据,电商平台积累的大量网购数据不仅属于非刻意展示的数字足迹,而且还与消费偏好和个人特质^[11]、经济状况^[29]等高度相关。Gladstone等^[30]关于信用卡消费记录预测个人特质的研究发现,消费记录较之社交媒体信息对个人特质的预测作用在不同人群和不同时期都更加稳定。遗憾的是,由于数据所限,现有研究尚无法回答网购行为足迹能否用于违约预测,更是缺乏对其经济机制的考察。即便少数文献^[11, 12]指出了金融科技信用评分的违约预测作用与其使用了网购行为信息有关,但他们不仅没有提供金融科技公司如何利用网购行为足迹的直接证据,也无法揭示网购行为足迹为何能够用于违约风险识别的机理,更是不能回答网购行为足迹是否相对于金融科技信用评分具有额外信息含量。

理论上,基于网购订单数据构建相关变量,可通过反映经济状况和个人特质而发挥违约预测作用,主要包括如下几个方面:1)网购订单金额及次数,真实准确地记录了个人或家庭的消费支出与水平,而消费又与难以获知的真实收入有关。这是因为,消费是由个人劳动收入和财富决定^[29],消费越高通常收入也越高;而且,如白重恩等^[31]指出的那样,居民会出于隐私保护和策略性动机而瞒报收入,消费则是以“真实收入”而非“上报收入”为基础,可以更好地反映真实收入水平;2)网购的凌晨下单行为可以反映个人的自控力。Berg等^[4]针对线上购买家具、Gong等^[25]针对线上游戏充值、Djeundje等^[32]针对发送邮件的证据表明,在凌晨从事相关活动越频繁的借款人的自控力越差,从而违约概率也越大。同样是任意时段都可进行的线上活动,如果借款人频繁地在凌晨进行网购,那么差的自控力也会导致他们的还款表现更差;3)网购的取消订单行为与个人的履约能力有关。一切形式的市场交易都是以信用为基础的契约关系,而信用的本质是对契约关系中义务承担者履约能力和履约意愿的预期^[33]。网购订

⑥ 当前市场上已经存在熟悉金融机构风控流程的“黑产”中介,专门帮助信用质量较低的借款人包装、粉饰、甚至伪造用于风险评估的有关信息。

单正是购买人通过电商平台与商家签订的一份约定标的、数量、金额等要素且立即生效的契约,一个人经常取消网购订单一定程度上说明其履约能力进而信用质量较差,那么他在借贷交易中的违约概率也会更高;4)网购的付款方式选择与个人的履约意愿有关,在线支付和货到付款是两种可供购买人自愿选择的付款方式,从交易双方履约顺序的角度来看,在线支付意味着购买人主动选择先履约,货到付款则相当于购买人主动选择后履约,显然主动选择后履约的人的履约意愿更低,进而借贷交易中履行还款义务的意愿也更低;5)网购历史中收货地址和收货人电话的动态变化,不仅可以反映个人的真实居住环境,还与个人的稳定状态有关。与“居无定所”的借款人相比,无论是物理地址(居住地址)还是数字地址(电话号码),地址不常变化的借款人更加稳定,违约风险也相对更低^{[34]⑦}。

需要说明的是,尽管本研究探索性地对网购行为足迹用于违约预测的经济机制进行了上述分析,但如 Berg 等^[4]所言,某些数字足迹同时会与多种个人特质有关,很难明确给出唯一特定的经济解释,本研究也难以对网购行为足迹有关的每个变量的经济含义进行分析。例如,收货人个数一方面可以反映借款人的社会资本,社会资本越强,违约概率越低^[35];另一方面,收货人个数越多也可说明借款人赡养家庭成员的负担越重,违约概率反而越高^[32]。又如,Gu 等^[36]等利用工作时段信用卡消费来测度工作偷懒行为,研究发现房地产的财富效应会加剧个人偷懒行为。在网购交易中,如果工作时段频繁网购也能反映个人的偷懒行为,这一方面会通过降低工资收入或增加失业风险而提高违约概率,但另一方面偷懒行为受借款人财富状况影响,因财富状况好而工作偷懒的借款人的违约概率反而不见得高。

2 数据、变量与模型

2.1 数据与样本

本研究使用国内某持牌金融科技借贷机构借款人层面的数据。该公司主要依托大数据、人工智能等数字技术向借款人提供循环授信的线上消费类信用贷款。贷款的步骤大致如下:首先,借款人安装该公司的手机 APP 和使用实名认证的手机号注册账号,自愿填写性别、年龄、婚姻、学历、收入等个人信息,并授权公司从第三方获取与借款人风险画像相关的数据;其次,公司利用内部积累和外部获取的数据,采用违约预测模型进行风险评估,做出“贷不贷”“贷多少”“贷多贵”的决策;最后,根据批准的额度和利率,借款人在授信额度内使用借款,并按照约定日期及方式还款。

本研究的初始样本包括 2016 年 4 月—2017 年 4 月期间申请借款且有电商信用评分和网购订单信息的 48 275 个借款人,覆盖我国大陆 31 个省、自治区和直辖市的 296 个城市,因此样本具有很好的代表性。剔除关键变量缺失的样本后,最终样本包含 46 607 个观测值。针对每个借款人,除了申请时间和贷后还款表现外,还包括如下四类信息:1)借款申请时自我报告的性别、年龄、婚姻状况等人口统计特征,以及会员等级、身份证和手机号认证情况等信息^⑧;2)来自央行征信的银行卡信息,包括银行卡类型(储蓄卡或信用卡)、开户行名称等;3)从国内某大型电商平台获取的个人电商信用评分;4)从另一家大型电商平台获取的历史逐笔网购订单数据,包括订单的金额(精确到角)、下单时间(精确到秒)、付款方式(在线支付或货到付款)、订单状态(已完成或已取消)、收货地址(精确到乡镇或街道)等^⑨。

需要指出,第 3) 和第 4) 类的数据分别来自

⑦ 与申请借款时可获取的 GPS 定位信息和手机号码等静态信息相比,网购订单中的收货地址和收货人电话的动态变化,才能够反映借款人工作或生活的稳定状况,这些信息已被金融科技公司广泛用于构建信用评分,业界实践的相关报道参见:<https://finance.huanqiu.com/article/9CaKmJH4mb>; <https://techcrunch.com/2015/01/27/data-from-alibabas-e-commerce-sites-is-now-powering-a-credit-scoring-service/?guccounter=1>。

⑧ 该数据对所有涉及个人隐私的信息均做了加密或脱敏处理,本研究只能得到是否通过认证的状态。

⑨ 本研究只能获得借款人曾经使用过的所有收货地址信息,但无法将收货地址与每笔订单一一匹配,因此利用收货地址的个数和其中地址为乡镇的占比等进行收货地址的指标测度。

两家不同的大型电商平台,都具有很好的覆盖性和代表性.其中,第 3)类的电商信用评分不仅使用了网购有关的信息,还使用了支付、缴费、资产、理财等多个维度的信息.然而,即便该信用评分已经利用了平台内部的网购数据,但由于“数据孤岛”问题,该平台难以使用其他平台的数据.因此,利用借款人在另一家大型电商平台的网购订单数据,不仅为本研究的问题考察提供了独特的研究机会,更是符合金融科技借贷机构通常会从不同渠道获取数据的业界实践.

2.2 变量与指标

根据借款人贷后的还款表现,定义是否违约

的 0/1 虚拟变量 *Default*:超过约定还款日后的规定天数仍未还款,定义为违约 (*Default* = 1),否则为没有违约 (*Default* = 0)^⑩.

根据信息来源及类型,前述第 1) ~ 第 4) 中用于违约预测的借款人信息划分为四类:申请借款时自我报告的人口特征信息、反映央行征信的银行卡信息^⑪、来自第三方的电商信用评分、反映网购行为的订单信息,依次简称为自我报告信息、央行征信信息、电商信用评分和网购行为足迹,各类信息的变量构成及定义如表 1 所示.为了消除极端值对结果的影响,对连续变量在 1% 和 99% 水平上进行了缩尾处理.

表 1 变量定义与指标测度
Table 1 Definition and measurement of variables

类型	名称及符号	定义及说明
自我报告信息	手机认证 (<i>Phone_Y</i>)	通过手机认证为 1, 否则为 0
	身份证认证 (<i>Idcard_Y</i>)	通过身份证认证为 1, 否则为 0
	会员等级 (<i>Account_G</i>)	黄金和钻石会员为 1, 其他为 0
	绑定 QQ (<i>QQ_bound</i>)	绑定 QQ 为 1, 否则为 0
	绑定微信 (<i>Wechat_bound</i>)	绑定微信为 1, 否则为 0
	年龄分类 (<i>Age_i</i>)	<i>i</i> = 1, 2, 3, 分别代表年龄小于 30 岁, 30 ~ 40 岁及 40 ~ 60 岁 取值为 1, 其他为 0, 基准为没有报告年龄
	性别分类 (<i>Sex_i</i>)	<i>i</i> = 1, 2, 分别表示男性、女性为 1, 其他为 0, 基准为没有 报告性别
	收入分类 (<i>Income_i</i>)	<i>i</i> = 1, 2, 分别表示收入 6 000 元以上和以下为 1, 其他为 0, 基准为没有报告收入
	婚姻分类 (<i>Merri_i</i>)	<i>i</i> = 1, 2, 分别表示未婚和已婚为 1, 其他为 0, 基准是没有 报告婚姻状况
	学历分类 (<i>Edu_i</i>)	<i>i</i> = 1, 2, 分别表示本科及以上, 本科以下为 1, 其他为 0, 基准是没有报告学历
央行征信信息	信息缺失比例 (<i>Null_num</i>)	在填写信息项中缺失项数除以总项数
	银行卡开户行个数 (<i>Bankname_num</i>)	所有银行卡的开户行个数
	绑定手机号个数 (<i>Bankphone_num</i>)	所有银行卡绑定不同手机号的个数
	信用卡张数 (<i>Creditcard_num</i>)	信用卡的张数
	储蓄卡张数 (<i>Depositcard_num</i>)	储蓄卡的张数
	信用卡开户行个数 (<i>Creditbank_num</i>)	信用卡开户行的个数
	储蓄卡开户行个数 (<i>Depositbank_num</i>)	储蓄卡开户行的个数
	国有开户行个数 (<i>Statbankname_num</i>)	所有银行卡的开户行中国有银行的个数

⑩ 在行业实践中,违约通常被定义为该借款人历史最大逾期天数超过一个门槛值,门槛值通常设定为 30 d 或者 60 d. 出于商业机密,本研究所用数据的借款人是否违约由公司定义,具体门槛值不便公开.

⑪ 数据所限,尽管银行卡信息并不能完全代表央行征信,但由于信用卡通常是基于央行征信而发放,因此,与王正位等^[13]采用是否有信用卡来衡量借款人央行征信信息的做法类似,本研究也利用银行卡的相关信息来反映央行征信信息.

续表 1

Table 1 Continues

类型	名称及符号	定义及说明
电 商 信 用 评 分 和 网 购 行 为 足 迹	电商信用评分($E_creditscore$)	电商平台对该借款人的信用评分
	网购开始年份($First_order_i$)	$i = 1, 2, 3$, 分别表示是否在 2013 年及以前, 2014 年—2015 年, 2016 年开始网购为 1, 其他为 0, 基准是 2017 年开始网购
	货到付款比例($Delivery_pay_ratio$)	借款前 3 年的订单中, 货到付款订单的比例
	取消订单比例($Order_cancel_ratio$)	借款前 3 年的订单中, 取消订单的比例
	凌晨下单比例($Order_night_ratio$)	借款前 3 年的订单中, 凌晨下单的比例
	工作时间下单比例($Order_workhour_ratio$)	借款前 3 年的订单中, 星期一至星期五的早上九点至下午五点下单的比例
	工作日下单比例($Order_workday_ratio$)	借款前 3 年的订单中, 星期一至星期五下单的比例
	借款前 3 年订单金额统计量 ($Order_*_year3$)	* 为通配符, 包括借款前 3 年的订单笔数, 交易金额的最大值、最小值、标准差、均值
	借款前半年订单金额统计量($Order_*_hyear$)	* 为通配符, 包括借款前半年的订单笔数, 交易金额的最大值、最小值、标准差、均值
	收货地址个数($Reciev_addr_num$)	历史收货地址的个数
	收货城市个数($Recieve_city_num$)	收货地址中不同城市的个数
	收货地址中乡镇比例($Town$)	收货地址为镇或乡的个数占收货地址总个数的比例
	收货电话个数($Recieve_phone_num$)	收货地址信息中收货人电话的个数
	收货人个数($Receiver_num$)	收货地址信息中收货人的个数

2.3 模型设定与评价指标

为了识别和比较不同类型信息的违约预测能力, 本研究建立如下的 Logit 回归模型

$$\begin{aligned} Default_i = & \alpha_0 + \beta_1 Selfreport_i + \beta_2 Centralcredit_i + \\ & \beta_3 Thirdscore_i + \beta_4 Behavior_i + \eta_i + \varepsilon_i \end{aligned} \quad (1)$$

其中下标 i 表示借款人; 被解释变量 $Default_i$ 是借款人 i 是否违约的虚拟变量; $Selfreport_i$ 、 $Centralcredit_i$ 、 $Thirdscore_i$ 、 $Behavior_i$ 分别表示自我报告信息、央行征信信息、电商信用评分、网购行为足迹的变量向量; η_i 控制了申请借款的月份固定效应; ε_i 为随机误差项。

参考 Berg 等^[4]等研究, 本研究采用 AUC 值评价基于不同类型信息的模型的预测能力。 AUC 值取值范围在 0.5 ~ 1 之间, 取值越大表示预测能力越好。 Iyer 等^[23]给出了违约概率预测模型 AUC 值的参考范围: 当信息含量较低时, 60% 是模型 AUC 值的目标; 当 AUC 值达到 70% 以上时, 表示信

息含量丰富、模型预测能力较好; AUC 值每增加 1% 对信用风险评估实践都具有重要经济价值。

3 实证结果及分析

3.1 描述性统计

表 2 给出了主要变量的描述性统计。 由表 2 可见, $Default$ 的均值仅为 2.61%, 可能存在样本不平衡的问题, 但仍高于 Berg 等^[4]的样本均值; 电商信用评分($E_creditscore$)的最小值和最大值分别为 164 和 605, 说明该评分对不同借款人具有良好的区分度。 网购行为方面, 借款人在借款前 3 年和前半年的平均网购订单数量($Order_num_year3$ 和 $Order_num_hyear$)分别为 42.198 8 笔和 19.002 2 笔, 平均的收货地址数($Reciev_addr_num$)为 3.405 3 个, 所涉及的城市($Recieve_city_num$)平均为 1.696 2 个, 平均村镇地址比例($Town$)为 0.340 8。

表 2 主要变量的描述性统计
Table 2 Summary statistics of key variables

变量名称	观测值	均值	中位数	标准差	最小值	最大值
<i>Default</i>	46 607	0. 026 1	0	0. 159 5	0	1
<i>Phone_Y</i>	46 607	0. 981 9	1	0. 133 2	0	1
<i>Idcard_Y</i>	46 607	0. 705 6	1	0. 455 8	0	1
<i>Null_num</i>	46 607	0. 795 9	0. 875 0	0. 286 5	0	1
<i>Creditcard_num</i>	46 607	1. 574 9	1	2. 349 5	0	36
<i>Depositcard_num</i>	46 607	2. 219 7	2	1. 568 3	0	28
<i>E_creditscore</i>	46 607	475. 042 1	490	78. 760 2	164	605
<i>Order_night_ratio</i>	46 607	0. 118 7	0. 040 0	0. 191 0	0	1
<i>Delivery_pay_ratio</i>	46 607	0. 084 7	0	0. 192 7	0	1
<i>Order_cancel_ratio</i>	46 607	0. 149 7	0. 052 5	0. 224 5	0	1
<i>Order_num_year3</i>	46 607	42. 198 8	16	63. 995 6	1	363
<i>Order_num_hyear</i>	46 607	19. 002 2	8	26. 214 1	1	140
<i>Reciev_addr_num</i>	46 607	3. 405 3	3	2. 813 2	1	35
<i>Recieve_city_num</i>	46 607	1. 696 2	1	1. 129 9	1	24
<i>Town</i>	46 607	0. 340 8	0. 142 9	0. 401 3	0	1

表 3 统计了主要变量的相关系数. 可以看出, 电商信用评分与 *Default* 的相关系数最大, 呈现与预期一致的负相关关系, 表明该变量具有很好的违约预测能力; 网购行为足迹的相关变量中, 取消订单比例和收货地址个数与 *Default* 的相关系数, 分别超过和接近信用卡数量与 *Default* 的相关系数, 表明网购行为足迹也具有较好的预测能力. 从电商信用评分与其他变量的相关系数来看, 一方面与反映央行征信的银行卡信息高度相关, 其与信用卡数量和储蓄卡数量的相关系数分别达到 0. 361 5 和 0. 293 9; 另一方面与网购订单数量和收货地址个数的相关系数也高达 0. 425 8 和 0. 394 9, 说明尽管本研究所用的电商信用评分和网购数据分别来自两家不同的平台, 但由于在不同平台的消费习惯或行为特征具有相似性, 电商信用评分也间接反映了借款人在另一家平台网购行为足迹的部分信息. 但是, 本研究也注意到, 凌晨下单比例、货到付款比例、取消订单比例等变量与电商信用评分的相关系数并不太高, 均未超过 0. 05, 说明即便不同来源的两类数字足迹的信息含量可能部分重合, 但二者相对彼此仍然具有额外信息含量.

3.2 多元回归分析

表 4 报告了以是否违约的 0/1 虚拟变量为被解释变量、不同类型信息为解释变量的 Logit 回归结果, 并给出了评价模型预测能力的 *AUC* 值. 表 4

第(1)列~表 4 第(4)列是四类信息依次单独加入模型的回归结果, 图 1 给出了相应的 ROC 曲线. 其中, 表 4 第(1)列和第(2)列基于自我报告和央行征信两类基础信息的模型 *AUC* 值分别为 68. 30% 和 63. 77%, 均在 1% 水平上显著(相对于 *AUC* 等于 50%). 特别地, 表 4 第(2)列基于央行征信的模型 *AUC* 值, 虽然低于 Berg 等^[4] 利用德国征信信用分预测电商用户违约风险的 *AUC* 值 68. 30%, 但高于 Iyer 等^[23] 采用益博睿信用评分(*experian scoreX plus credit score*)预测 Prosper 平台借款人违约风险的 *AUC* 值 62. 54%, 也高于 Gambacorta 等^[12] 利用我国央行征信信息预测某金融科技公司借款人违约风险的 *AUC* 值 59. 39%.

表 4 第(3)列和第(4)列分别是单独包括电商信用评分和网购行为两类数字足迹的模型预测结果, *AUC* 值分别达到了 73. 05% 和 68. 62%, 二者分别高出列第(2)列的模型 *AUC* 值 9. 28 和 4. 85 个百分点. 进一步按照 Iyer 等^[23] 和 Jiang 等^[11] 的做法, 若以央行征信信息的预测能力为对照基准, 利用电商信用评分和网购行为足迹的违约预测准确性分别提高 $(0. 730 5 - 0. 5) / (0. 637 7 - 0. 5) - 1 = 67. 39\%$ 和 $(0. 686 2 - 0. 5) / (0. 637 7 - 0. 5) - 1 = 35. 22\%$. 表 4 最后两行关于不同模型 *AUC* 值的差异性检验结果及图 1 的 ROC 曲线也一致显示: 电商信用评分的违约

表 3 主要变量的相关系数表
Table 3 Correlation coefficients for key variables

变量名称	Default	Phone_Y	Idcard_Y	Creditcard_num	Depositcard_num	E_creditscore	Order_night_ratio	Delivery_pay_ratio	Order_cancel_ratio	Order_num_year3	Order_num_hyear	Reciev_addr_num	Town
Default	1												
Phone_Y	-0.0151	1											
Idcard_Y	0.066 0	0.131 9	1										
Creditcard_num	-0.046 0	0.027 0	-0.053 5	1									
Depositcard_num	-0.029 7	0.056 3	-0.064 4	0.376 9	1								
E_creditscore	-0.134 7	0.044 5	-0.163 4	0.361 5	0.293 9	1							
Order_night_ratio	0.03 76	0.001 7	0.016 7	-0.019 2	-0.001 6	-0.086 7	1						
Delivery_pay_ratio	0.032 1	-0.031 7	0.004 3	-0.063 4	-0.040 7	-0.031 3	-0.001 7	1					
Order_cancel_ratio	0.065 2	-0.023 6	0.032 5	-0.025 9	0.047 2	-0.171 5	0.053 7	-0.045 4	1				
Order_num_year3	-0.020 1	0.025 1	-0.040 7	0.460 8	0.382 8	0.425 8	-0.009 3	0.012 6	-0.023 3	1			
Order_num_hyear	-0.023 2	0.028 2	-0.037 0	0.343 6	0.328 7	0.368 4	0.000 9	-0.051 8	-0.017 4	0.822 9	1		
Reciev_addr_num	-0.041 6	0.020 9	-0.024 9	0.321 6	0.314 2	0.394 9	-0.043 7	0.014 6	-0.059 5	0.446 8	0.376 9	1	
Town	0.017 3	-0.002 0	0.008 8	-0.068 6	-0.031 7	-0.066 9	0.008 1	0.035 1	0.017 0	-0.035 0	-0.032 9	-0.062 5	1

表 4 不同类型信息影响是否违约的 Logit 回归结果
Table 4 Results for the Logit regression of default likelihood on different types of information

变量名称	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
自我报告信息	是	否	否	否	否	是	是	是	是
央行征信信息	否	是	否	否	否	是	是	是	是
E_creditscore			-0.009 6 *** (-25.814 8)		-0.010 4 *** (-25.855 5)		-0.010 1 *** (-24.516 9)		-0.010 4 *** (-24.089 2)
网购行为足迹：									
Delivery_pay_ratio				0.954 6 *** (7.894 1)	0.777 9 *** (6.495 2)			0.820 9 *** (6.547 7)	0.779 2 *** (6.271 4)
Order_cancel_ratio				1.211 7 *** (11.777 5)	0.674 2 *** (6.425 3)			1.066 8 *** (10.098 5)	0.639 3 *** (5.906 0)
Order_night_ratio				0.763 2 *** (5.883 7)	0.554 1 *** (4.478 0)			0.700 6 *** (5.320 2)	0.513 6 *** (3.983 7)

续表 4

Table 4 Continues

变量名称	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>Order_workhour_ratio</i>				-0.023 8 (-0.174 7)	-0.036 1 (-0.293 5)			-0.042 2 (-0.310 4)	-0.042 6 (-0.335 6)
<i>Order_workday_ratio</i>				-0.195 7 (-1.531 7)	-0.126 5 (-1.083 4)			-0.184 1 (-1.437 5)	-0.132 0 (-1.090 4)
<i>Order_num_year3</i>				0.000 0 (0.026 7)	0.000 5*** (3.007 4)			0.000 3* (1.953 8)	0.000 4*** (2.668 0)
<i>Order_num_hyear</i>				-0.002 1 (-1.109 5)	-0.000 2 (-0.485 2)			-0.001 1 (-0.844 9)	-0.000 3 (-0.428 9)
<i>Reciev_addr_num</i>				-0.272 9*** (-8.830 2)	-0.158 2*** (-5.381 3)			-0.231 7*** (-7.403 0)	-0.139 5*** (-4.656 9)
<i>Recieve_city_num</i>				0.322 7*** (6.762 3)	0.257 3*** (5.643 8)			0.260 2*** (5.404 8)	0.203 4*** (4.329 7)
<i>Town</i>				0.155 0** (2.268 3)	0.124 4* (1.802 7)			0.117 2* (1.684 1)	0.108 6 (1.537 5)
<i>Recieve_phone_num</i>				0.119 8 (1.478 6)	0.117 4 (1.526 2)			0.100 8 (1.277 9)	0.088 9 (1.146 8)
<i>Receiver_num</i>				0.012 4 (0.152 7)	-0.050 2 (-0.656 0)			0.021 4 (0.274 8)	-0.021 0 (-0.276 4)
常数项	控制	控制	控制	控制	控制	控制	控制	控制	控制
月份固定效应	控制	控制	控制	控制	控制	控制	控制	控制	控制
观测值	46 607	46 607	46 607	46 607	46 607	46 607	46 607	46 607	46 607
伪 R^2	0.050 1	0.029 2	0.078 9	0.050 0	0.104 0	0.064 1	0.115 6	0.089 2	0.135 6
<i>AUC</i>	0.683 0***	0.637 7***	0.730 5***	0.686 2***	0.771 4***	0.707 0***	0.783 3***	0.746 7***	0.806 4***
(SE)	0.007 0	0.007 9	0.007 5	0.007 7	0.006 8	0.006 8	0.006 0	0.006 4	0.005 7
<i>AUC</i> 差异比较 ^b	(1) - (2) 0.045 3***		(3) - (1) 0.047 5***	(4) - (1) 0.003 2	(5) - (3) 0.040 9***	(6) - (1) 0.024 0***	(7) - (6) 0.076 3***	(8) - (6) 0.039 7***	(9) - (6) 0.099 4***

注: “网购行为足迹”还包括了以下两类变量: ①将借款人从 2017 年开始网购的年份作为基准, 依次定义借款人网购开始年份的三个 0/1 虚拟变量, 分别为在 2013 年以前、2014 年 - 2015 年、2016 年开始网购为 1, 其他为 0; ②分别针对借款申请的前 3 年和前半年, 利用借款人所有网购订单计算收货人个数和订单金额的若干统计量 (均值、最大值、最小值和标准差)。^b 倒数第二行的 (i) ~ (j) 表示对应的下方单元格的第 (i) 列与第 (j) 列模型 *AUC* 的差异性检验结果。表中的 ***、** 和 * 分别代表在 1%、5% 和 10% 的水平上显著, 下同。

预测能力显著强于其他三类信息;网购行为足迹的预测能力与自我报告信息的预测能力没有显著差异,但仍然强于央行征信信息^⑫。

从相关变量回归系数的经济含义来看,表4第(3)列的电商信用评分回归系数在1%水平上显著为负,说明借款人的电商信用评分越高,违约的可能性越小。表4第(4)列的货到付款订单比例、取消订单比例和凌晨下单比例三个网购行为变量,均在1%水平上与违约概率显著正相关,这与前述理论分析的预期一致,说明这些变量可以通过反映个人特质而对违约概率具有解释作用;收货地址中城市个数和乡镇比例则与违约概率显著正相关,这是因为收货地址涉及的城市个数越多,说明借款人过往经历中生活状态越不稳定,乡镇比例则反映了借款人的真实居住环境,高的比例说明借款人居住在农村的可能性较大。

尽管电商信用评分和网购行为足迹分别单独对违约概率具有预测效果,但是电商信用评分可能已经捕捉到了借款人在其他平台网购的行为特征,因此,两类数字足迹之间的替代或互补关系还需进一步检验。借鉴 Berg 等^[4]和 Iyer 等^[23]等的检验思路,表4第(5)列将两类数字足迹的相关变量同时加入模型,结果显示:模型的 AUC 值提高到 77.14%,比单独加入二者的模型 AUC 值分别高出 4.09 和 8.52 个百分点,且提高效果在 1%水平上显著,这说明同时使用两类数字足迹能够显著地提高违约预测效果。从回归系数来看,不仅表4第(4)列中统计上显著的变量在第(5)列依然显著,而且无论是绝对值大小还是显著性水平,表4第(5)列中电商信用评分的回归系数都有所提高,这说明加入网购行为足迹反而进一步增强了电商信用评分的违约预测能力,从而网购行为足迹相对于电商信用评分具有额外的信息含量。

现实中,金融借贷机构原则上可以同时使用以上四类信息,两类数字足迹对违约预测的边际贡献需要基于自我报告和央行征信两类基础信息的预测能力来判断。表4第(6)列同时利用两类基础信息的结果显示,70.70%的 AUC 值显著高于表4第(1)列和第(2)列,进一步,以同时利用两类基础信息的违约预测效果为基准,表4第

(7)列和第(8)列分别加入电商信用评分和网购行为足迹,结果显示:AUC 值可以分别高达 78.33%和 74.67%,与表4第(6)列的 AUC 值相比,分别提高了 7.63 和 3.97 个百分点。因此,相对于两类基础信息的预测能力,无论是电商信用评分,还是网购行为足迹,都对违约预测效果具有明显的提升作用。表4第(9)列将四类信息同时加入回归的结果显示:AUC 值可以高达 80.64%,大幅高于表4的其他模型。特别地,与作为对照基准的表4第(6)列相比,同时使用两类数字足迹可以将违约预测效果提高约 50% ($(0.8064 - 0.5) / (0.7070 - 0.5) - 1 = 48.02\%$),这很好地体现了数字足迹对于金融科技借贷违约预测的重要作用。

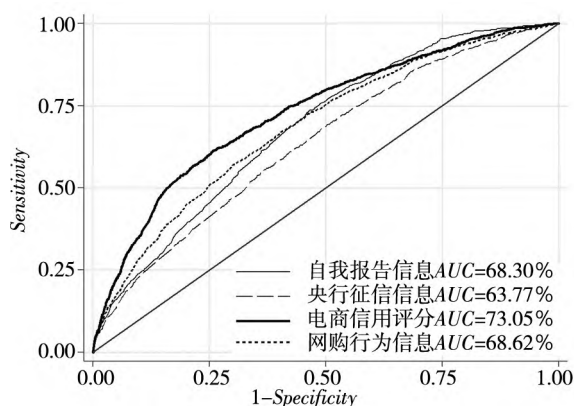


图1 不同类型信息单独用于违约预测的 ROC 曲线

Fig. 1 ROC curves for default prediction separately using different types of information

3.3 稳健性检验

3.3.1 样本外预测

前述结果均属于样本内预测,可能存在过拟合问题。为了保证结论的可靠性,本小节采用训练样本外预测和时间样本外预测两种方法进行稳健性检验。针对训练样本外预测,采用机器学习中的 K 折交叉验证方法,表5第(1)列和第(2)列分别给出了 2 折和 5 折交叉验证方法下各预测模型的 AUC 值。以 5 折交叉验证为例,具体过程包括四个步骤。第一步,将全样本随机划分为 5 个子样本。第二步,将其中一个子样本作为测试集,利用其余 4 个子样本进行表4的 Logit 回归,根据回归得到的估计模型,对测试集中所有观测值进行预测,并

^⑫ 此外,本研究剔除身份证和手机两项认证信息,基于自我报告信息的模型 AUC 值仅为 62.44%,降低了 5.86 个百分点。篇幅所限,需要此部分内容,可以联系作者索取。

计算出各个预测模型在该测试集的 *AUC* 值。第三步,依次将5个子样本分别作为测试集,为每个模型得到5个 *AUC* 值并求均值作为一次5折交叉验证的结果。第四步,重复上述步骤100次,对每次结果再求平均得到各个模型最终的 *AUC* 值。由表5可见:无论采用2折还是5折交叉验证的样本外预测,尽管各模型的 *AUC* 值与表4相比均有1个百分点左右的下降,但不同模型预测能力的相对排序并没有改变,说明训练样本外的预测结果稳健。

针对时间样本外预测,通常的做法是采用历史样本建立模型来预测当前样本,因此,将最后一个月(即2017年4月)的子样本作为测试集,利用之前所有月份的子样本建立预测模型,而后对测试集样本进行预测,结果如表5第(3)列所示。总体来看,除了基于网购行为足迹的预测模型 *AUC* 值略小于基于自我报告信息的模型 *AUC* 值外,不同模型预测能力的相对排序也没有变化,其中,同时利用两类数字足迹和同时利用四类信息的模型 *AUC* 值分别高达74.31%和78.66%,时间样本外的预测结果同样稳健^⑬。

3.3.2 随机抽样样本的检验

本研究所使用数据的违约样本比例仅为2.61%,尽管数值上大于Berg等^[4]的样本统计结果(有无征信信用评分的两个样本中,违约样本的占比分别为0.9%和0.25%),但依然存在样本不平衡问题。为了保证结果的稳健性,本研究从未违约样本中随机抽取一定比例的子样本,将其与所有违约的样本共同构成抽样样本,考察了违约样本比例分别为50%、30%、10%、5%的不同抽样样本的预测结果。篇幅所限未展示的结果显示,无论何种比例的随机抽样,也无论是单个模型的预测能力,还是不同模型预测能力的相对排序,结果都与表4结果类似,很好地通过稳健性检验。

3.3.3 其他稳健性检验

本研究还进行了下述三个方面的稳健性检验:1)预测模型的其他评价指标。除了利用 *AUC*

值评价模型的预测能力外,参考Roa等^[34]和Wu等^[37]的做法,*KS*(Kolmogorov-Smirnov)值和*GINI*系数也是违约预测中常用的模型评价指标,其数值越大,表示预测能力越好。针对表4的所有模型,在新的两个评价指标下,除了基于网购行为足迹的模型 *KS* 值略小于基于自我报告信息的模型 *KS* 值外,不同模型违约预测能力的相对排序与表4采用 *AUC* 值的结果基本一致,实证结果稳健。

2)控制样本选择性偏差的检验。由于偏好或习惯不同,不同借款人长期选择的网购平台也会不同,这会导致部分借款人因缺失电商信用评分或网购订单数据而未能进入本研究的研究样本,从而可能存在样本选择性偏差问题。因此,参考赵宸宇^[38]等,采用Heckman两阶段模型来控制样本选择性偏差问题。由于样本要求借款人既有一家平台的电商信用评分,又有另一家平台的网购订单数据,因此本研究分别针对两家平台的样本选择性偏差进行两阶段回归。以是否选择提供网购订单数据的平台为例:第一阶段回归中,以是否选择该平台的0/1虚拟变量为被解释变量,构建选择模型,针对每个样本计算逆米尔斯比率(inverse mills ratio);第二阶段回归中,将逆米尔斯比率作为控制变量加入表4的各个模型,分别重新进行Logit回归,实证结果显示,无论是针对提供网购订单数据的平台选择,还是针对提供电商信用评分的平台选择,控制选择性偏差后的结果稳健。

3)控制其他固定效应的检验。表4的回归模型仅控制了月份固定效应,但影响借款人还款表现的因素还包括资产状况、工作经验与所处行业等个人特征^[23],以及所处区域的经济社会环境^[39]。结合数据可得性,本研究进一步同时控制月份和省份固定效应,控制月份与省份的交互固定效应,以及继续控制行业固定效应^⑭,分别重新对表4的各个模型进行回归,实证结果依然稳健。篇幅所限,未展示的具体结果留存备索。

⑬ 未报告的结果中,本研究发现不包含两项认证的自我报告信息模型的 *AUC* 值仅为52.11%,远小于基于网购行为足迹的模型 *AUC* 值。

⑭ 由于本研究样本中仅有不到10%的借款人自愿报告了所属行业的信息,并且涉及的行业多达45类,从而每类行业的样本数十分有限,因此表4的回归中未控制行业变量。

表 5 样本外预测的稳健性检验
Table 5 Robust tests for out-of-sample prediction

变量类型	(1)	(2)	(3)
	训练样本外预测 (2 折交叉验证)	训练样本外预测 (5 折交叉验证)	时间样本外预测
自我报告信息	0.670 0	0.673 1	0.651 2
央行征信信息	0.626 0	0.629 3	0.563 2
电商信用评分	0.726 2	0.727 4	0.694 8
网购行为足迹	0.670 0	0.674 2	0.636 8
电商信用评分 + 网购行为足迹	0.761 3	0.764 0	0.743 1
自我报告信息 + 央行征信信息	0.693 1	0.696 8	0.676 6
自我报告信息 + 央行征信信息 + 电商信用评分	0.772 2	0.775 6	0.765 2
自我报告信息 + 央行征信信息 + 网购行为足迹	0.726 7	0.732 0	0.722 1
自我报告信息 + 央行征信信息 + 电商信用评分 + 网购行为足迹	0.790 4	0.795 0	0.786 6

3.4 网购行为足迹用于违约预测的机制检验

尽管实践中网购数据已被广泛用于个人征信评估,但数据所限,现有文献尚缺乏关于网购行为足迹的违约预测效果及其作用机制的经验证据.结合数据的可得性,本小节利用反映网购行为的订单金额、凌晨下单、取消订单、货到付款共四个变量,通过一系列的异质性检验,探索性地为网购行为足迹用于违约预测的经济机制提供证据.

3.4.1 网购订单金额

网购金额一定程度上反映了个人或家庭的消费水平和经济状况,而二者都与违约风险直接有关.直观而言,女性既是线上购物的主力,也相对男性在家庭中承担更多的消费角色^[40],从而女性借款人的网购金额更加能够反映个人或家庭的真实消费与收入水平.据此推断,可以预期:借款人历史订单的平均金额越大,违约概率越低,且这种负向关系在女性借款人子样本中更加明显.

利用性别数据完整的子样本,分别利用男性和女性两个子样本,表 6 面板 A 的回归结果显示:网购金额与违约概率具有显著的负向关系,且该负关系只在女性子样本中显著.此外,针对网购金额与收入水平的关系,本研究还以借款人收入是否高于 6 000 元的 0/1 虚拟变量作为被解释变量,以网购订单平均金额为解释变量,篇幅所限未展示的结果也显示,订单平均金额越高的借款人,属于超过 6 000 元的高收入人群的概率也越大^⑮.

3.4.2 凌晨下单比例

已有文献利用购买家具^[4]、游戏充值^[25]、邮件使用^[32]等个人线上活动的数据,一致发现:相

对于其他时段,借款人在 0 点 ~ 6 点的线上活动越频繁,其自控力越差,从而借款违约的概率也越大.类似地,本研究利用借款人在 0 点 ~ 凌晨 6 点的下单比例反映其自控力,表 6 面板 B 的第(1)列显示,凌晨下单比例与违约概率在 1% 水平上显著正相关.进一步,由于工作日的凌晨下单更会影响当天的正常工作,从而更能通过反映借款人的自控力而与违约概率具有更强的正向关系.为此,借鉴 Gong 等^[25]的做法,分别定义工作日和休息日凌晨下单比例(*Order_night_ratio_work*, *Order_night_ratio_week*)两个指标,考察与违约概率关系的异质性,表 6 面板 B 的第(2)列仅将星期六和星期日界定为休息日,表 6 面板 B 的第(3)列进一步考虑法定节假日重新进行划分,回归结果一致表明,只有工作日的凌晨下单比例才会与违约概率呈现显著的正向关系.

3.4.3 取消订单比例

网购交易属于“先下单后交付”的延期交易,购买人下单即意味着与商家签订了一份立即生效的商品买卖合同,商家便开始履行合同.尽管下单后购买人可通过取消订单来中止合同,但频繁取消订单会给商家造成配货、仓储和物流等多方面的经济损失,可视为履约能力差的一种表现.进一步, Fay 等^[39]和 Gupta^[41]指出,除了个人因素,履约能力还与个人所在区域整体的信用环境有关.如果借款人不仅频繁取消订单,而且还处于整体信用环境相对差的区域,那么取消订单行为为更可能通过反映个人的履约能力而影响违约概率.

表 6 面板 C 第(1)列关于是否违约对取消订

⑮ 本研究的数据只提供了每个人的收入范围,而无法获知具体的收入数值.

单比例的回归结果显示,借款人取消订单的比例越高,其在借贷交易中的违约概率也越高. 进一步,根据全国城市信用监测平台首次发布的“2017 年城市信用综合指数排行”,将排行前 20 的省会城市和前 50 的地级市界定为城市信用指数高组,其他城市为低组,表 6 面板 C 第(2)列和面板 C 第(3)列的对比结果表明,取消订单比例与违约概率的正向关系在城市信用指数低的子样本中更明显. 此外,本研究也根据“北京大学数字普惠金融指数”中的城市信用指数的中位数,将样本划分为城市信用指数高和低两组子样本,表 6 面板 C 第(4)列和面板 C 第(5)列的对比结果同样表明,处于低信用指数城市的借款人的取消订单比例对违约概率更具解释能力.

表 6 网购行为足迹用于违约预测的机制检验

Table 6 Tests for the mechanisms underlying the roles of online shopping footprints in predicting default likelihood

面板 A：网购订单平均金额					
变量名称	(1)	(2)	(3)		
	报告性别的子样本	男性子样本	女性子样本		
<i>Order_mean_hyear</i>	-0.066 6 *	-0.045 9	-0.294 6 **		
	(-1.915 2)	(-1.278 2)	(-2.084 7)		
自我报告信息、央行征信信息、常数项	控制	控制	控制		
月份和省份固定效应	控制	控制	控制		
观测值	22 085	18 139	3 243		
伪 <i>R</i> ²	0.093 0	0.091 3	0.130 0		
组间系数差异	-0.248 7 *				
<i>χ</i> ²	(2.909 5)				
面板 B：凌晨下单比例					
变量名称	(1)	(2)	(3)		
		不考虑法定节假日 划分休息日和工作日	考虑法定节假日划 分休息日和工作日		
<i>Order_night_ratio</i>	0.631 7 ***				
	(5.754 2)				
<i>Order_night_ratio_week</i>		0.180 8	0.116 4		
		(1.354 7)	(0.848 6)		
<i>Order_night_ratio_work</i>		0.604 1 ***	0.639 1 ***		
		(5.509 7)	(5.840 7)		
自我报告信息、央行征信信息、常数项	控制	控制	控制		
月份和省份固定效应	控制	控制	控制		
观测值	46 583	46 583	46 583		
伪 <i>R</i> ²	0.070 2	0.070 2	0.070 4		
面板 C：取消订单比例					
变量名称	(1)	(2)	(3)	(4)	(5)
	全样本	城市信用综合指数		数字信用发展程度	
		低	高	低	高
<i>Order_cancel_ratio</i>	1.118 6 ***	1.308 4 ***	0.912 4 ***	1.285 1 ***	0.932 9 ***
	(11.158 2)	(9.170 8)	(6.384 0)	(9.058 0)	(6.508 3)
自我报告信息、央行征信信息、常数项	控制	控制	控制	控制	控制
月份和省份固定效应	控制	控制	控制	控制	控制
观测值	46 583	19 627	26 884	18 959	27 522
伪 <i>R</i> ²	0.076 3	0.096 4	0.074 4	0.099 8	0.073 5
组间系数差异	0.396 0 **			0.352 2 *	
<i>χ</i> ²	(3.845 5)			(3.050 8)	

续表 6

Table 6 Continues

面板 D: 货到付款比例						
变量名称	(1)	(2)	(3)	(4)	(5)	(6)
	全样本	物流时效性		支付工具普及率		网购订单 金额大小
		低	高	低	高	
<i>Delivery_pay_ratio</i>	0.599 2 *** (4.861 8)	0.232 3 (1.214 4)	0.884 4 *** (5.488 0)	0.111 9 (0.414 4)	0.726 5 *** (5.260 5)	
<i>Delivery_pay_ratio_big</i>						0.230 9 (1.591 9)
<i>Delivery_pay_ratio_small</i>						0.414 3 *** (2.878 8)
自我报告信息、央行征信信息、常数项	控制	控制	控制	控制	控制	控制
月份和省份固定效应	控制	控制	控制	控制	控制	控制
观测值	46 583	18 059	28 368	8 504	37 921	46 583
伪 R^2	0.069 4	0.083 6	0.071 9	0.096 3	0.071 4	0.069 2
组间系数差异		0.652 1 ***		0.614 6 **		
χ^2		(6.800 0)		(4.108 5)		

3.4.4 货到付款比例

网购付款方式的选择,既与物流时效、支付工具使用、订单标的金额等客观因素有关,也与个人主动选择事先付款的履约意愿有关。因此,如果付款方式的选择不是由客观因素所致,那么其更加能够通过反映履约意愿而与违约行为有关。利用所处城市的经济发达程度和是否为省会城市来体现物流时效性^⑩,数字金融发展指数来衡量支付工具的普及程度,表6 面板 D 进行子样本分组回归,以及根据订单金额高低重新定义不同的比例指标,来检验货到付款能够预测违约的内在机制。

表6 面板 D 第(1)列的回归结果显示,借款人网购历史中选择货到付款的比例越高,违约概率越大。表6 面板 D 第(2)列与面板 D 第(3)列、面板 D 第(4)列与面板 D 第(5)列的对比结果一致表明:在网购物流效率高和支付工具普及程度高的子样本中,选择货到付款受客观因素的影响较小,从而更能通过反映借款人较低的履约意愿而与违约概率更具有正向关系。表6 面板 D 第(6)列根据每笔网购订单金额是否大于300元,对大额和小额订单分别定义货到付款比例的两个指标,结果发现,只有金额较小订单的货到付款比例才对违约概率具有解释能力,即小额订单依然选择货到付款的借款人违约概率更高是与较低的

履约意愿有关。

3.5 数字足迹的经济价值分析

尽管电商信用评分和网购行为两类数字足迹都能提升违约预测效果,但获取这些足迹数据都需付出成本。借鉴 Jiang 等^[11]关于百融云创信用评分违约预测能力的经济价值分析方法,以两类基础信息的违约预测能力为基准,通过数值模拟,依次计算进一步仅使用电商信用评分和同时使用两类数字足迹的经济增加值(economic value added, EVA),具体步骤如下:

首先,参照 Jiang 等^[11]的简化设定,假设成功借款后每个借款人的借款合同(金额、利率、期限)相同,每笔借款在期末一次性还本付息,且借款人一旦违约,本金和利息全部无法收回。对于特定一个成功借款的借款人,给定出借人的资金成本,计算得到借款人违约和不违约两种情形下出借人提供借款所能获得的净现值(net present value, NPV),分别记为 NPV^1 和 NPV^0 。

其次,给定一批借款申请者,出借人基于特定预测模型首先估计每个申请者的违约概率,而后根据事先设定的违约概率门槛值,向违约预测概率低于门槛值的借款人提供贷款。记借款申请者的平均通过率为 $Fundrate$,成功借款的借款人的平均真实违约概率为 $Defaultrate$,那么,每个申请者能为出借人带

^⑩ 考虑到电商平台在全国重要城市或区域建设物流仓库,本研究将全部省会城市以及 GDP 水平排名前 50 的非省会城市作为物流时效性高组,其他城市为物流时效性低组。

来的期望净现值可以表示为 $Fundrate \times [NPV^0 \times (1 - DefaultRate) + NPV^1 \times DefaultRate]$. 显然, 设定的门槛值越高, 借款申请的平均通过率越高, 但贷后的平均违约概率也会越高, 因此出借人需要选择最优门槛值来最大化每个申请者能够带来的期望净现值, 将最大期望净现值记为 $MaxNPV$.

最后, 针对基于“自我报告信息 + 央行征信信息”“自我报告信息 + 央行征信信息 + 电商信用评分”“自我报告信息 + 央行征信信息 + 电商信用评分 + 网购行为足迹”的三个违约预测模型, 通过上述步骤, 分别计算每个模型下出借人的最大期望净现值, 依次记为 $MaxNPV_1$ 、 $MaxNPV_2$ 和 $MaxNPV_3$. 以 $MaxNPV_1$ 为基准, 进一步使用电商信用评分的经济增加值为 $EVA_1 = MaxNPV_2 - MaxNPV_1$, 同时使用电商信用评分和网购行为足迹的经济增加值为 $EVA_2 = MaxNPV_3 - MaxNPV_1$.

参考 Jiang 等^[11] 的参数设定, 设定借款金额、(年化) 借款利率、借款期限分别为 3 万元、12% 和 3 年, 出借人的(年化) 资金成本为 3%~7%, 图 2 给出了使用两类数字足迹进行违约预测所带来的经济增加值随资金成本的变化. 可以看出: 若进一步使用电商信用评分, 每笔借款申请的经济增加值 EVA_1 为 230.642 9 元~639.669 1 元; 若同时使用两类数字足迹, 经济增加值 EVA_2 为 343.416 0 元~1 069.288 2 元, 从而网购行为足迹的经济增加值 $EVA_2 - EVA_1$ 为 112.773 1 元~429.619 1 元, 远大于实践中购买相应数据的成本^①.

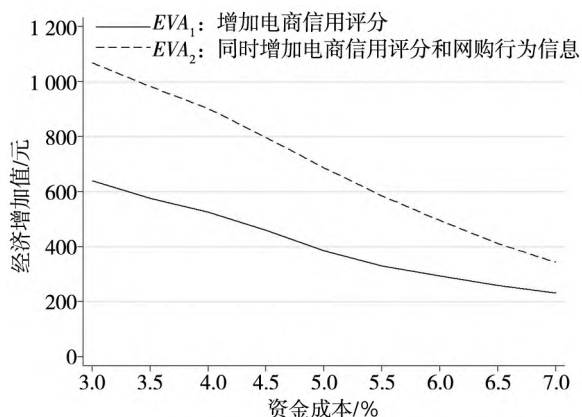


图 2 两类数字足迹用于违约预测的经济增加值

Fig. 2 Economic value added of additionally using two types of digital footprints in default prediction

4 数字足迹提高信贷可得性的机理分析

鉴于电商信用评分和网购行为两类数字足迹在违约预测中的良好表现, 本小节将从两个方面来分析数字足迹提高消费信贷可得性的机理所在. 一方面, 针对央行征信严重缺失的“信用白户”, 即便仅利用电商信用评分或网购行为足迹, 也可对这些因信用难以查证而被拒贷的全新客群进行风险评估. 另一方面, 针对那些被传统征信错误认定为劣质、但真实质量较高的老客群, 利用数字足迹信息能够进行更为准确的风险评估, 也可有效缓解其信贷约束.

4.1 信用白户的信贷可得性分析

《中国个人征信行业现状和趋势展望》报告数据显示, 截至 2017 年 11 月, 央行征信中心收录 9.5 亿自然人信息, 但其中仅有 4.8 亿人有借贷记录. 显然, 对于数量巨大、但征信历史缺失的信用白户而言, 如果能够利用覆盖广泛的数字足迹对其进行风险评估, 不仅能够极大地提高信贷资源配置效率, 还能有效地促进消费升级和推动普惠金融发展.

根据样本数据中借款人是否有信用卡, 本研究将无信用卡的子样本界定为传统征信缺失严重的信用白户, 表 7 第(1)列和第(2)列给出了利用有无信用卡两类子样本分别进行回归的模型 AUC 值, 结果显示: 在无信用卡的子样本中, 尽管所有模型的预测能力与表 4 相比均有一定程度下降, 但是单独加入电商信用评分和网购行为足迹的两个模型的 AUC 值分别只下降了 0.9 和 1.5 个百分点; 同时加入两类数字足迹和加入所有信息的两个模型的 AUC 值分别高达 76.19% 和 78.81%, 说明违约预测效果良好. 类似地, 根据借款人是否拥有五大国有银行发行的信用卡划分子样本, 表 7 第(3)列和第(4)列的结果同样显示, 对于无信用卡的白户客群, 加入两类足迹信息和加入所有信息的两个模型的 AUC 值依然分别超过 76% 和 79%, 同样具有很好的违约预测效果. 此外, 针对

① Jiang 等^[11] 将借款金额设定为 15 万元, 考虑到本文所研究的消费信贷具有小额、分散的特点, 因此将借款金额设定为 3 万元. 同时 Jiang 等^[11] 指出, 现阶段的金融科技借贷实践中, 出借机构针对每个申请者从第三方平台购买信用评分的成本通常不超过 50 元.

有无信用卡的两类子样本,本研究也进行了如表 5 的样本外预测,限于篇幅未展示的交叉验证结果显示,在无信用卡的子样本中,加入两类足迹信息和加入所有信息的模型 AUC 值分别超过 74% 和 76%,违约预测能力依然良好.

需要特别指出的是,针对有无信用卡的两类子样本,进一步对比表 7 第 5 行基于“电商信用评分+网购行为足迹”和表 7 第 6 行基于“自我报告信息+央行征信信息”的违约预测效果,并参考 Iyer 等^[23]和 Jiang 等^[11]的比较方法,可以得到数字足迹提升信用白户信贷可得性的证据.在表 7 第(2)列和第(4)列的有信用卡的子样本中,第

5 行使用两类数字足迹的预测准确性比第 6 行同时使用两类基础信息的预测准确性分别高出 $(0.773\ 1 - 0.5)/(0.728\ 4 - 0.5) - 1 = 19.57\%$ 和 $(0.791\ 5 - 0.5)/(0.763\ 8 - 0.5) - 1 = 10.50\%$. 与之对照,在表 7 第(1)列和第(3)列的无信用卡的子样本中,预测准确性的提高幅度分别高达 $(0.761\ 9 - 0.5)/(0.674\ 2 - 0.5) - 1 = 50.34\%$ 和 $(0.763\ 3 - 0.5)/(0.688\ 1 - 0.5) = 39.98\%$,大幅高于数字足迹对有信用卡子样本违约预测的提升作用.因此,对于因信用历史缺失而难以获得信贷服务的信用白户而言,数字足迹对违约预测的效果提升更明显.

表 7 基于有无信用卡的子样本检验

Table 7 Tests for the subsamples with or without credit card

变量类型	(1)	(2)	(3)	(4)
	信用卡		国有银行信用卡	
	无	有	无	有
自我报告信息	0.671 0	0.705 7	0.675 6	0.736 1
央行征信信息	0.600 6	0.640 3	0.617 1	0.667 8
电商信用评分	0.721 5	0.724 8	0.720 4	0.744 5
网购行为足迹	0.670 9	0.697 7	0.680 8	0.710 3
电商信用评分+网购行为足迹	0.761 9	0.773 1	0.763 3	0.791 5
自我报告信息+央行征信信息	0.674 2	0.728 4	0.688 1	0.763 8
自我报告信息+央行征信信息+电商信用评分	0.766 1	0.793 9	0.769 4	0.824 5
自我报告信息+央行征信信息+网购行为足迹	0.719 3	0.768 9	0.734 0	0.797 6
自我报告信息+央行征信信息+电商信用评分+网购行为足迹	0.788 1	0.820 2	0.795 2	0.846 7

4.2 “劣质”客群的信贷可得性分析

对于因央行征信记录不良或受外部冲击而被传统征信评估模式认定为“劣质”的老客群,利用数字足迹进行更为准确的信用评估,也能提高此类客群的信贷可得性.将有信用卡的借款人作为老客群子样本,本研究首先分别利用表 4 第(6)列仅包含自我报告信息和央行征信信息的估计模型,和表 4 第(9)列进一步包含了电商信用评分和网购行为足迹的估计模型,对子样本的每个借款人进行违约概率预测.其次,根据两个模型预测的违约概率的三分位数,对子样本进行 3×3 的二维分组,统计各组样本真实的违约概率及借款人数,如表 8 所示.最后,对比是否使用两类数字足迹进行违约预测两种情形下的借款人分布和真实违约概率,可以得到数字足迹提高信贷可得性的如下三个方面证据:

第一,给定使用数字足迹进行违约预测的每个分组,未使用数字足迹进行违约预测的低、中、高三组的真实违约概率并非单调递增,或者不同组间的差异并不明显;相反,给定未使用数字足迹进行违约预测的每个分组,使用数字足迹进行违约预测情形下低、中、高三组的真实违约概率不仅都一致地单调递增,而且三组的真实违约概率具有明显差别,说明两类基础信息之外进一步使用数字足迹的违约预测模型更加能够准确区分客群的真实质量.

第二,假设全样本中风险最高的三分之一的借款人无法获得借款^⑩,表 8 第 3 行数据显示:如果不使用数字足迹信息进行违约预测,违约概率被预测为高组的 8 967 个借款人将被拒贷;然而,如果进一步使用两类数字足迹进行违约预测,违

⑩ 假设处于一个信贷总供给给定,每个借款人有相同借款需求的理想环境.

约概率被预测为低组的 192 个和中组的 2 689 个借款人将获得借款,即合计占比为 $(192 + 2\ 689) / 8\ 967 = 32.13\%$ 的借款人因重新被评估为非高风险而可以获得借款.更为重要的是,对于未使用数字足迹进行违约预测而被拒贷的借款人,表 8 第 3 行数据还显示:利用数字足迹被预测为低组和中组的借款人的真实违约概率分别仅为 0.52% 和 1%,都比全样本总体的违约比例 1.94% 还低,即数字足迹在违约预测中的有效应用能够提高此类被“错杀”老客群的信贷可得性,这也是金融科技借贷服务的重点客群.

第三,在不使用数字足迹进行违约概率预测的情形下,表 8 第 1 行和第 2 行中共三分之二的

借款人可获得借款;如果进一步使用数字足迹信息进行违约预测,表 8 第 1 行和第 2 行中被重新预测为高组的两类共计 2 881 (332 + 2 549) 个借款人将不能获得借款.然而,对照真实的贷后违约比例可以发现,利用数字足迹进行违约预测被拒贷的两组借款人,正是传统金融机构“一直想要拒绝,但又无从甄别”的客群:二者的违约概率分别为 2.71% 和 4.16%,明显高于全样本的违约比例 1.94%;而且,表 8 第 2 行中进一步使用数字足迹被拒贷的 2 549 个借款人的违约概率 (4.16%) 甚至比未使用数字足迹进行违约预测而被拒贷的表 8 第 3 行的平均违约概率 (3.75%) 还要高.

表 8 是否使用数字足迹进行违约预测二维分组下的真实违约概率及借款人人数统计

Table 8 Statistics for real default probability and the number of borrowers sorted on predicted default probabilities with and without using digital footprints for default prediction

未使用数字足迹 预测的违约概率	使用数字足迹预测的违约概率			
	低	中	高	全样本
低	0.001 9 [6 882]	0.009 1 [1 754]	0.027 1 [332]	0.004 2 [8 968]
中	0.001 6 [1 894]	0.009 1 [4 524]	0.041 6 [2 549]	0.016 7 [8 967]
高	0.005 2 [192]	0.010 0 [2 689]	0.050 6 [6 086]	0.037 5 [8 967]
全样本	0.001 9 [8 968]	0.009 4 [8 967]	0.047 2 [8 967]	0.019 4 [26 902]

注：中括号内的数据表示该组样本的借款人人数；全样本指有信用卡的借款人样本.

上述三个方面的证据与 Berg 等^[4]针对电商平台采纳数字足迹后所观察到的实验结果一致:整体而言,尽管进一步使用数字足迹进行违约预测以后,被拒贷的借款人数并不一定小于重新获得贷款的借款人数量(例如,在三分之二获得借款的假设下,两部分借款人数刚好相等),但金融机构信贷资产的总体质量会显著提高,而且由于“拒伪识真”概率的提高,信贷资源整体的配置效率也会提高,这正是金融科技对普惠金融的赋能效果.

5 数字足迹违约预测能力的进一步分析

5.1 电商信用评分和网购行为足迹的互补作用

尽管本研究使用的电商信用评分和网购行

为足迹分别来自不同的两家电商平台,提供第三方信用评分的电商平台不能直接利用另一家平台的网购数据,但由于借款人在不同平台的网购习惯或行为特质具有一定相似性,两类数字足迹之间也具有一定相关性.因此,参考 Iyer 等^[23]分解用于违约预测的各类信息的做法,将电商信用评分作为因变量,将网购行为足迹的所有变量作为自变量,进行 OLS 回归得到回归残差 $E_creditscore_r$,用来测度电商信用评分中与网购行为无关的信息含量.类似地,将网购行为为足迹有关的每个变量作为因变量,分别对电商信用评分进行回归,同样提取出各个变量的回归残差(变量符号记为 $*_r$, $*$ 表示网购行为的相关变量),用于测度各个变量与电商信用评分无关的信息含量.

进一步,将是否违约的 0/1 虚拟变量作为被

解释变量,将上述回归得到的电商信用评分和网购行为足迹互不相关的信息含量(即回归残差)作为解释变量,Logit 回归结果如表 9 所示. 分别对比表 4 中的第(3)列、第(4)列、第(7)列和第(8)列,表 9 第(1)列~表 9 第(4)列结果表明:两类数字足迹中相互补充的信息含量依然具有很好

的违约预测能力,各个模型的 AUC 值相对于表 4 对应模型的 AUC 值并未大幅下降,这表明尽管两类数字足迹具有一定相关性,但也仍然具有彼此不相关的信息含量,这进一步强化了网购行为足迹相对于电商信用评分具有额外信息含量的结论.

表 9 电商信用评分和网购行为足迹的互补作用检验

Table 9 Tests for complementary effect between e-commerce credit score and online shopping footprints

变量名称	(1)	(2)	(3)	(4)
$E_creditscore_r$	-0.009 5***		-0.008 9***	
	(-24.464 5)		(-21.148 1)	
网购行为足迹 $_r$	否	是	否	是
自我报告信息	否	否	是	是
央行征信信息	否	否	是	是
常数项	控制	控制	控制	控制
月份固定效应	控制	控制	控制	控制
观测值	46 607	46 607	46 607	46 607
伪 R^2	0.063 6	0.036 4	0.101 3	0.089 1
AUC	0.722 5	0.656 3	0.769 8	0.748 9

5.2 网购行为足迹信息含量的进一步挖掘

鉴于网购行为足迹对违约概率的良好预测效果,深度挖掘网购订单数据中关于借款人消费习惯和行为特质的信息含量,无疑会进一步提高预测效果. 结合数据可得性,本小节尝试构建如下几个变量:1)借款前最后一笔网购订单的金额及其距离借款申请日的天数;2)借款前 3 年借款人进行网购的天数和相邻两笔订单的最大间隔天数;3)根据借款人手机号(原始数据隐藏了中间的四位数字)的前三位数字构建移动、联通、电信运营商的分类变量,以及根据

第二位数字构建分类变量,来反映手机号码的使用时长;4)根据每笔订单中加密的商品名称^⑨,利用订单笔数与商品名称个数的比值衡量所购商品的多样性,以及构建是否复购的虚拟变量;5)借款人在“618”“双 11”“双 12”等电商购物节的订单笔数(金额)占所有订单笔数(金额)的比例,来衡量借款人对电商促销活动的敏感性.

将第 1)~第 5)的相关变量作为进一步衡量网购行为足迹的指标,重新进行与表 4 类似的回归,各模型的 AUC 值如表 10 所示.

表 10 网购行为足迹信息含量的进一步挖掘

Table 10 Further exploration of information from the online shopping footprints

变量类型	(1)	(2)	(3)
	样本内预测	训练样本外预测 (5 折交叉验证)	时间样本外预测
网购行为足迹	0.711 0	0.688 0	0.650 7
电商信用评分 + 网购行为足迹	0.783 9	0.768 5	0.738 2
自我报告信息 + 央行征信信息 + 网购行为足迹	0.765 7	0.742 3	0.727 0
自我报告信息 + 央行征信信息 + 电商信用评分 + 网购行为足迹	0.815 7	0.797 7	0.785 2

⑨ 出于隐私保护,商品名称的原始数据是经过 MD5 加密的字符串,本研究无法获知具体的商品名称. 根据借款人所购商品的属性(耐用品、非耐用品、生活必需品)和用途(奢侈品、图书、母婴、家具建材、游戏等)等,未来可进一步捕捉借款人的消费习惯和行为特征,这是提高违约预测效果的重要方向.

从表 4 第(1)列可以看出：单独使用网购行为足迹进行预测的模型 *AUC* 值由 0.686 2 提高到 0.711 0，提高了 2.5 个百分点，尽管仍然低于电商信用评分的违约预测能力，但远高于自我报告信息的违约预测能力。进一步，同时加入四类信息的模型 *AUC* 值高达 0.815 7，两类数字足迹相对于两类基础信息的违约预测效果提高了 52.51% ($((0.815\,7 - 0.5)/(0.707\,0 - 0.5) - 1 = 52.51\%)$)。表 4 第(2)列和第(3)列的样本外预测结果也显示，加入进一步衡量网购行为足迹的相关变量后，单独包括网购行为足迹的模型 *AUC* 值均提高了 1.4 个百分点。

表 11 是否有电商平台信用额度的子样本检验

Table 11 Tests for the subsamples with or without credit limit granted by e-commerce platform

面板 A：无信用额度子样本							
变量名称	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<i>E_creditscore</i>	-0.010 6 ***		-0.011 2 ***		-0.010 4 ***		-0.010 7 ***
	(- 16.332 6)		(- 15.593 9)		(- 14.835 4)		(- 14.207 0)
网购行为足迹	否	是	是	否	否	是	是
自我报告信息	否	否	否	是	是	是	是
央行征信信息	否	否	否	是	是	是	是
常数项	控制	控制	控制	控制	控制	控制	控制
月份固定效应	控制	控制	控制	控制	控制	控制	控制
观测值	19 395	19 395	19 395	19 395	19 395	19 395	19 395
伪 <i>R</i> ²	0.070 6	0.051 9	0.097 8	0.056 5	0.102 4	0.086 8	0.125 1
<i>AUC</i>	0.715 5	0.687 2	0.758 4	0.684 6	0.759 0	0.734 5	0.786 2
面板 B：有信用额度子样本							
变量名称	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<i>E_creditscore</i>	-0.008 6 ***		-0.009 5 ***		-0.009 5 ***		-0.010 0 ***
	(- 17.182 5)		(- 18.492 4)		(- 17.506 4)		(- 17.779 3)
网购行为足迹	否	是	是	否	否	是	是
自我报告信息	否	否	否	是	是	是	是
央行征信信息	否	否	否	是	是	是	是
常数项	控制	控制	控制	控制	控制	控制	控制
月份固定效应	控制	控制	控制	控制	控制	控制	控制
观测值	27 212	27 212	27 212	27 212	27 212	27 212	27 212
伪 <i>R</i> ²	0.062 5	0.044 6	0.093 2	0.068 3	0.114 2	0.093 3	0.138 5
<i>AUC</i>	0.709 5	0.683 8	0.763 5	0.722 7	0.794 9	0.760 7	0.817 9

由表 11 可见：无论是哪类子样本，也无论是何种违约预测模型，电商信用评分的回归系数均在 1% 水平上显著为负。仅考虑电商信用评分和网购行为两类数字足迹，对比表 11 面板 A 和面板 B 前三列中每个模型的 *AUC* 值，其差值分别仅为 0.60%、0.34%、-0.51%，表明数字足迹对是否有电商信用额度两类子样本的违约预测能力并无明显差异，说明电商信用评分具有违约预测能力的主要原因是该评分利用了用户

5.3 电商平台信用授信的影响分析

电商信用评分之所以具有良好的违约预测效果，既可能是因使用了大量且独特的非传统数据，也可能是由于客户使用了电商平台提供的“类信用卡”业务，导致电商信用评分还利用了与央行征信类似的“借债还钱”信息。那么，借款人在电商平台的“消费信贷”使用情况，是否会影响电商信用评分的预测能力？在本研究的样本中，超过 40% 的借款人尽管有电商信用评分，但没有电商平台的信用额度。为此，本研究根据是否有信用额度，将样本划分为有无额度两个子样本，控制其它三类信息，表 11 给出了不同子样本中电商信用评分的违约预测效果。

的其他信息，而非与央行征信类似的借债还钱信息。进一步，针对子样本中数字足迹对违约预测能力的提升作用，以同时加入自我报告和央行征信两类基础信息的第(4)列为基准，对比两个面板的最后四列可以看出：表 11 面板 A 中第(5)列至第(7)列模型的预测能力较之表 11 面板 A 第(4)列分别提高了 40.30%、27.03%、55.04%，分别与面板 B 中对应的提升效果 32.42%、17.06%、42.75% 相比更高，即两类数

字足迹对违约预测的提升效果在无电商信用额度的子样本中反而更明显.

5.4 网购频率的影响分析

网购行为足迹对违约概率的预测作用还可能受网购频率的影响. 根据借款人过去 3 年网购订单数的中位数, 将样本划分为网购频率高和低两组子样本, 回归结果如表 12 所示. 对比表 12 面板 A 和面板 B 中各类模型的 AUC 值可以发现: 就两类数字足迹单独的预测能力而言, 电商信用评分对网购频率低组子样本的预测效果更好, 而网购行为足迹对网购频率高组子样本的预测效果更

好; 但是, 如果将两类数字足迹同时加入模型, 高低两组子样本的预测效果并无明显差异, 甚至低组的 AUC 值略高, 这说明网购频率并不影响两类数字足迹的违约预测能力. 与此同时, 针对数字足迹提升违约预测效果的作用, 还可以看出: 表 12 面板 A 中第 (5) 列至第 (7) 列模型的预测效果较之表 12 面板 A 第 (4) 列分别提高了 47.79%、27.16%、57.97%, 分别与面板 B 中对应的提升效果 30.64%、21.72%、44.19% 相比, 说明在网购频率低的子样本中数字足迹更加能够提升违约预测效果.

表 12 网购频率高低分组的子样本检验

Table 12 Tests for the subsamples split according to online shopping frequency

面板 A: 网购频率低组							
变量名称	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$E_creditscore$	-0.012 6 ***		-0.012 6 ***		-0.012 0 ***		-0.011 8 ***
	(-20.320 5)		(-18.957 8)		(-17.992 2)		(-16.873 5)
网购行为足迹	否	是	是	否	否	是	是
自我报告信息	否	否	否	是	是	是	是
央行征信信息	否	否	否	是	是	是	是
常数项	控制	控制	控制	控制	控制	控制	控制
月份固定效应	控制	控制	控制	控制	控制	控制	控制
观测值	23 334	23 334	23 334	23 334	23 334	23 334	23 334
伪 R^2	0.098 4	0.	0.118 7	0.062 6	0.124 8	0.092 1	0.143 0
AUC	0.755 7	0.690 1	0.782 7	0.694 4	0.787 3	0.747 2	0.807 1
面板 B: 网购频率高组							
变量名称	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$E_creditscore$	-0.009 4 ***		-0.009 2 ***		-0.009 6 ***		-0.009 8 ***
	(-18.798 9)		(-17.075 2)		(-17.324 0)		(-16.660 7)
网购行为足迹	否	是	是	否	否	是	是
自我报告信息	否	否	否	是	是	是	是
央行征信信息	否	否	否	是	是	是	是
常数项	控制	控制	控制	控制	控制	控制	控制
月份固定效应	控制	控制	控制	控制	控制	控制	控制
观测值	23 273	23 273	23 273	23 273	23 273	23 273	23 273
伪 R^2	0.071 0	0.065 0	0.106 9	0.071 6	0.117 3	0.108 9	0.149 2
AUC	0.732 7	0.710 6	0.781 1	0.726 5	0.795 9	0.775 7	0.826 6

6 结束语

当前,尚不完善的个人征信体系严重制约了我国消费信贷市场的发展,而大数据与机器学习等数字技术向金融领域的全面赋能,不仅通过为数以亿计的信用白户建立数字信用极大地推动了

普惠金融发展,而且通过促进消费信贷市场增长有效助力了消费升级国家战略的实施.然而,金融科技借贷领域的现有研究整体滞后于业界实践,第三方征信和个人行为足迹等非传统数据能否及如何用于个人信用风险评估,以及怎样提高特定客群的信贷可得性等问题,既缺乏相应的理论分析,也缺乏必要的经验证据.

本研究利用国内某头部金融科技借贷机构借款人层面的数据,根据数据来源及类型,将用于违约预测的信息分为自我报告、央行征信、电商信用评分和网购行为共四种类型,实证考察并重点比较了电商信用评分和网购行为两类数字足迹在违约预测中的作用。利用 *AUC* 值衡量不同违约预测模型的预测能力,以自我报告和央行征信两类基础信息的预测能力为对照基准,研究发现:电商信用评分和网购行为两类数字足迹都能很好地预测消费信贷的违约概率,其中电商信用评分的预测能力最强,网购行为足迹则不仅与自我报告信息的预测能力无显著差异,而且还相对于电商信用评分具有额外的信息含量,同时利用两类足迹信息能使违约预测的准确性提高约 50%。经过样本外预测、随机抽样、使用模型预测能力的不同评价指标、考虑样本选择性偏差、控制其他固定效应等一系列检验,实证结果依然稳健。进一步的机制检验结果表明,有效利用个人数字足迹信息进行违约预测,不仅能够帮助传统征信缺失严重的信用白户获得消费信贷,还能提高信用质量被错误低估的借款者的信贷可得性。此外,本研究还考察了不同子样本中和不同指标测度时网购行为相关变量对违约概率解释力的异质性,为网购行为足迹能够预测违约的经济机制提供了探索性的经验证据。

本研究的理论贡献与现实意义包括三个层

面。第一,从数字足迹用于违约预测的角度扩展了借贷市场的相关研究,不仅指出了数字足迹提高信贷可得性的内在机理,还探索性地建立了网购行为和违约行为之间的理论联系。研究结论可进一步推广到小微、三农等金融供给薄弱的其他领域,为科技赋能背景下金融增强服务实体的能力提供参考依据;第二,违约预测方面的少数文献注意到了第三方信用评分、底层非传统数据与传统征信的互补作用,本研究则利用逐笔网购订单数据,比较并发现了更为底层的网购行为足迹相对于电商信用评分的额外预测作用,这不仅理论上丰富了金融科技借贷的相关研究,还能为数字化转型背景下金融机构对外部购买和内部自建的风控模式选择提供决策参考;第三,本研究发现数字足迹能够提高被传统征信忽视的特定客群的信贷可得性,但个人行为相关的数字足迹的获取及使用极易触及隐私安全问题。未来的研究可考虑个人数字信用和隐私保护,进一步探讨金融科技借贷中的数据隐私保护机制和数据治理模式,以及考察不同机制和模式对市场效率的影响。此外,尽管本研究利用网购行为足迹构建了与经济状态和个人特质有关的多个变量,但本研究只针对部分变量能够预测违约的经济机制提供了初步证据,后续研究可基于数字足迹与个人行为的关系进行更进一步的深入探讨,扩展金融科技借贷和相关领域的交叉研究。

参考文献:

- [1] Buchak G, Matvos G, Piskorski T, et al. Fintech, regulatory arbitrage, and the rise of shadow banks[J]. *Journal of Financial Economics*, 2018, 130(3): 453–483.
- [2] Erel I, Liebersohn J. Can fintech reduce disparities in access to finance? Evidence from the paycheck protection program[J]. *Journal of Financial Economics*, 2022, 146(1): 90–118.
- [3] Fuster A, Plosser M, Schnabl P, et al. The role of technology in mortgage lending[J]. *Review of Financial Studies*, 2019, 32(5): 1854–1899.
- [4] Berg T, Burg V, Gombovic A, et al. On the rise of fintechs: Credit scoring using digital footprints[J]. *Review of Financial Studies*, 2020, 33(7): 2845–2897.
- [5] Butaru F, Chen Q, Clark B J, et al. Risk and risk management in the credit card industry[J]. *Journal of Banking and Finance*, 2016, 72(11): 218–239.
- [6] 陆瑶, 张叶青, 黎波, 等. 高管个人特征与公司业绩——基于机器学习的经验证据[J]. *管理科学学报*, 2020, 23(2): 120–140.
Lu Yao, Zhang Yeqing, Li Bo, et al. Managerial individual characteristics and corporate performance: Evidence from a machine learning approach[J]. *Journal of Management Sciences in China*, 2020, 23(2): 120–140. (in Chinese)
- [7] Liao L, Martin X, Wang N, et al. Carrot and Stick: Two Natural Field Experiments on Lender Information Sharing[R]. SSRN Working Paper, 2019. <https://ssrn.com/abstract=3182671>.
- [8] 杨佳, 陆瑶, 李纪珍, 等. 数字时代下普惠金融对创业的影响研究——来自中国家庭微观调查的证据[J]. 管

理科学学报, 2022, 25(11): 43–68.

Yang Jia, Lu Yao, Li Jizhen, et al. The impact of inclusive finance on entrepreneurship in the digital age: Evidence from Chinese Household Micro Survey[J]. *Journal of Management Sciences in China*, 2022, 25(11): 43–68.

- [9] Goldstein I, Jiang W, Karolyi G A. To fintech and beyond[J]. *Review of Financial Studies*, 2019, 32(5): 1647–1661.
- [10] Jagtiani J, Lemieux C. The roles of alternative data and machine learning in fintech lending: Evidence from the lending club consumer platform[J]. *Financial Management*, 2019, 48(4): 1009–1029.
- [11] Jiang J, Liao L, Lu X, et al. Deciphering big data in consumer credit evaluation[J]. *Journal of Empirical Finance*, 2021, (62): 28–45.
- [12] Gambacorta L, Huang Y, Qiu H, et al. How do Machine Learning and Non-traditional Data Affect Credit Scoring? New Evidence from a Chinese Fintech Firm[R]. SSRN Working Paper, 2020. <https://ssrn.com/abstract=3506945>.
- [13] 王正位, 周从意, 廖理, 等. 消费行为在个人信用风险识别中的信息含量研究[J]. *经济研究*, 2020, 55(1): 149–163. Wang Zhengwei, Zhou Congyi, Liao Li, et al. Informational content of consumption behavior in consumer credit risk evaluation[J]. *Economic Research Journal*, 2020, 55(1): 149–163. (in Chinese)
- [14] Dai L, Han J, Shi J, et al. Digital Footprints as Collateral for Debt Collection[R]. SSRN Working Paper, 2023. <https://ssrn.com/abstract=4135159>.
- [15] Stiglitz J E, Weiss A. Credit rationing in markets with imperfect information[J]. *American Economic Review*, 1981, 71(3): 393–410.
- [16] Berger A N, Miller N H, Petersen M A, et al. Does function follow organizational form? Evidence from the lending practices of large and small banks[J]. *Journal of Financial Economics*, 2005, 76(2): 237–269.
- [17] Padilla A J, Pagano M. Endogenous communication among lenders and entrepreneurial incentives[J]. *Review of Financial Studies*, 1997, 10(1): 205–236.
- [18] 黄益平, 黄卓. 中国的数字金融发展: 现在与未来[J]. *经济学(季刊)*, 2018, 17(4): 1489–1502. Huang Yiping, Huang Zhuo. The development of digital finance in China: Present and future[J]. *China Economic Quarterly*, 2018, 17(4): 1489–1502. (in Chinese)
- [19] 曹伟威, 刘圣, 李宜威, 等. 结伴出行推动消费升级: 基于铁路大数据的分析[J]. *管理科学学报*, 2020, 23(2): 18–38. Cao Weiwei, Liu Sheng, Li Yiwei, et al. Travelling companion stimulates consumption upgrade: An analysis based on railway big data[J]. *Journal of Management Sciences in China*, 2020, 23(2): 18–38. (in Chinese)
- [20] 黄益平, 邱晗. 大科技信贷: 一个新的信用风险管理框架[J]. *管理世界*, 2021, 37(2): 12–21. Huang Yiping, Qiu Han. Bigtech lending: A new credit risk management framework[J]. *Journal of Management World*, 2021, 37(2): 12–21.
- [21] 廖理, 张伟强. P2P网络借贷实证研究: 一个文献综述[J]. *清华大学学报(哲学社会科学版)*, 2017, 32(2): 186–196. Liao Li, Zhang Weiqiang. Empirical research on online P2P lending: A literature review[J]. *Journal of Tsinghua University (Philosophy and Social Sciences)*, 2017, 32(2): 186–196. (in Chinese)
- [22] 李强. P2P网络借贷的研究综述与现实发展[J]. *电子科技大学学报(社科版)*, 2019, 21(5): 37–44. Li Qiang. P2P lending: Literature review and practical development[J]. *Journal of University of Electronic Science and Technology of China (Social Sciences Edition)*, 2019, 21(5): 37–44. (in Chinese)
- [23] Iyer R, Khwaja A I, Luttmer E F P, et al. Screening peers softly: Inferring the quality of small borrowers[J]. *Management Science*, 2016, 62(6): 1554–1577.
- [24] Frost J, Gambacorta L, Huang Y, et al. Bigtech and the changing structure of financial intermediation[J]. *Economic Policy*, 2019, 34(100): 761–799.
- [25] Gong S, Levine R, Lin C, et al. Debtors at play: Gaming behavior and consumer credit risk[J]. *Management Science*, 2024, 70(9): 5691–5708.
- [26] Hibbeln M, Norden L, Usselman P, et al. Informational synergies in consumer credit[J]. *Journal of Financial Intermediation*, 2020, (44): 1–16.
- [27] Garmaise M J. Borrower misreporting and loan performance[J]. *Journal of Finance*, 2015, 70(1): 449–484.
- [28] Garmaise M J, Natividad G. Consumer default, credit reporting, and borrowing constraints[J]. *Journal of Finance*, 2017, 72(5): 2331–2368.
- [29] Deng Y, Liao L, Yu J, et al. Capital leakage, house prices, and consumer spending: Quasi-experimental evidence from house purchase restriction spillovers[J]. *Review of Financial Studies*, 2022, 35(6): 3060–3099.
- [30] Gladstone J J, Matz S C, Lemaire A. Can psychological traits be inferred from spending? Evidence from transaction data[J]. *Psychological Science*, 2019, 30(7): 1087–1096.
- [31] 白重恩, 唐燕华, 张琼. 中国隐性收入规模估计——基于扩展消费支出模型及数据的解读[J]. *经济研究*, 2015, 50(6): 4–18.

- Bai Chongen, Tang Yanhua, Zhang Qiong. An extended expenditure: Based estimation of income underreporting and size of black economy in China[J]. *Economic Research Journal*, 2015, 50(6): 4–18. (in Chinese)
- [32] Djeundje V B, Crook J, Calabrese R, et al. Enhancing credit scoring with alternative data[J]. *Expert Systems with Applications*, 2021, (163): 1–12.
- [33] 吴汉洪, 徐国兴. 信用本质的经济学分析[J]. *中国人民大学学报*, 2004, (4): 56–62.
Wu Hanhong, Xu Guoxing. An economic thinking about the essence of credit[J]. *Journal of Renmin University of China*, 2004, (4): 56–62. (in Chinese)
- [34] Roa L, Correa-Bahnsen A, Suarez G, et al. Super-app behavioral patterns in credit risk models: Financial, statistical and regulatory implications[J]. *Expert Systems with Applications*, 2021, (169): 1–15.
- [35] 廖理, 李梦云, 王正位. 借款人社会资本会降低其贷款违约概率吗——来自现金贷市场的证据[J]. *中国工业经济*, 2020, (10): 5–23.
Liao Li, Li Mengyun, Wang Zhengwei. Can borrowers' social capital reduce its loan default rates: Evidence from cash loan market[J]. *China Industrial Economics*, 2020, (10): 5–23. (in Chinese)
- [36] Gu Q, He J, Qian W, et al. Housing Booms and Shirking[R]. SSRN Working Paper, 2022. <https://ssrn.com/abstract=3189933>.
- [37] Wu W, Xu D, Zhao Y, et al. Do consumer internet behaviors provide incremental information to predict credit default risk? [J]. *Economic and Political Studies*, 2020, 8(4): 482–499.
- [38] 赵宸宇. 数字化发展与服务化转型——来自制造业上市公司的经验证据[J]. *南开管理评论*, 2021, 24(2): 149–163.
Zhao Chenyu. Digital development and servitization: Empirical evidence from listed manufacturing companies[J]. *Nankai Business Review*, 2021, 24(2): 149–163. (in Chinese)
- [39] Fay S, Hurst E, White M J. The household bankruptcy decision[J]. *American Economic Review*, 2002, 92(3): 706–718.
- [40] Rao M B, Hymavathi C L, Rao M M. Factors affecting female consumer's online buying behavior[J]. *Academy of Marketing Studies Journal*, 2018, 22(2): 1–20.
- [41] Gupta A. Foreclosure contagion and the neighborhood spillover effects of mortgage defaults[J]. *Journal of Finance*, 2019, 74(5): 2249–2301.

Digital footprints and default prediction: Evidence from consumer credit

HU Jun¹, LI Qiang^{1*}, ZENG Yong¹, TIAN Xuan², LIU Song³

1. School of Management and Economics, University of Electronic Science and Technology of China, Chengdu 611731, China;
2. PBC School of Finance, Tsinghua University, Beijing 100083, China;
3. Sichuan XW Bank Corp., Ltd., Chengdu 610094, China

Abstract: Using a unique data set obtained from a Chinese fintech lending company, this paper categorizes the information used for loan default prediction into four types: Self-reported information, central bank credit information, e-commerce credit score, and online shopping behaviors. The roles played by the two types of digital footprints, e-commerce credit score and online shopping behaviors, in predicting loan default are investigated and the underlying mechanisms through which digital footprints improve credit availability are explored. Our results show that the digital footprints can help lenders to improve the accuracy in predicting a borrower's default likelihood by about 50%. The prediction power of the e-commerce credit score is significantly the highest among the four types of information. Nevertheless, online shopping behaviors can not only provide additionally useful information beyond the e-commerce credit score, but also perform as well as the self-reported information disclosed by borrowers. Furthermore, the preliminary evidence on the mechanisms through which digital footprints improve credit availability shows that e-commerce credit score and online shopping behaviors can be used to establish credit scores for “unbanked customers” and accurately evaluate the credit worthiness of borrowers whose credit quality is always wrongly underestimated.

Key words: fintech; digital footprints; default prediction; credit availability