

农户银行数字足迹与贷款违约风险^{*}

郑海荣 张 洋 何 婧 穆争社

内容提要:大数据如何助力贷款风险管理是金融科技时代的重要问题。本文首次考察农户使用手机银行时留下的数字足迹大数据,能否有效帮助银行进行风险管理。基于信号传递博弈理论,本文运用农商银行百万条农户贷款数据,揭示了农户银行数字足迹对贷款违约风险的影响。研究发现,农户银行数字足迹有效降低了贷款违约风险,且降低硬信息丰富群体贷款违约风险的作用强于软信息丰富群体。进一步分析表明,农户银行数字足迹对生产、经营等传统硬信息传递了有效的信息增量,对传统软信息传递的有效信息增量有限。异质性分析发现,在银行数字化转型程度越高、地区农村信用体系建设成熟度越高及普惠金融改革试验区的样本中,农户银行数字足迹降低贷款违约风险的作用效果越强。本文为发挥银行数字足迹等金融类大数据在贷款风险管理中的作用提供了事实依据。

关键词:数字足迹 违约风险 信号传递 软信息 硬信息

作者简介:郑海荣,福建农林大学经济与管理学院副教授,350002;

张 洋(通讯作者),中国农业大学经济管理学院博士研究生,100083;

何 婧,中国农业大学经济管理学院教授,100083;

穆争社,中国农业大学经济管理学院博士生导师,100083。

中图分类号:F832.4 **文献标识码:**A **文章编号:**1002-8102(2025)04-0059-18

一、引言

防范化解金融风险是金融工作的永恒主题。农村商业银行等中小银行是防范化解金融风险的重点关注对象,提升农村商业银行的贷款风险管理水平,是降低其风险的重要方面,也是亟待解决的关键难题。随着科技在金融领域的运用不断深入,数据资源在银行风险评估中正日益发挥重要作用。中国人民银行发布的《金融科技发展规划(2022—2025年)》的通知中提到“健全自动化风险控制机制”,从事前、事中和事后三个维度对发挥大数据在银行风险管理中的作用做出

^{*} 基金项目:国家社会科学基金重大项目“数字普惠金融支持乡村振兴的政策与实践研究”(22&ZD123);国家社会科学基金一般项目“‘新基建’背景下中国农村普惠金融发展对策研究”(20BJY153)。作者感谢匿名审稿专家的宝贵意见,文责自负。张洋电子邮箱:oceansicau@163.com。

了规划。

数据资源,指在数字生活中由行为主体(如自然人)的行为、交流和交易等活动产生的信息集合。其中,个人或企业在使用移动设备和电脑时产生的数据信息,被形象地称为“数字足迹”(Weaver和Gahegan,2007)。银行高度重视“银行数字足迹”在风险管理中的重要作用,深挖数据价值、释放数据潜能,是数字化风控的重要方面。如青岛银行建立的“鹰眼360智能风险监控平台”结合银行账户交易渠道数据等数字足迹进行动态风险监控和风险评级,“银行数字足迹”等数据信息能够反映农户信用状况、资金流动情况及消费习惯,为风险管理提供重要依据。那么,“农户银行数字足迹”是否能够有效作用于贷款风险管理,降低农户贷款违约风险?回答上述问题,是在大数据时代开展贷款风险管理的必答题,也为如何运用数字金融深化农村金融服务提供了重要路径。

数字足迹在金融业的应用受到学者关注,但主要集中在非金融类数字足迹在金融机构贷款决策中的作用。如网页浏览记录等数字足迹(Berg等,2020)、网络社交平台活动等数字足迹(Jayasuriya等,2023)、电商消费行为信息等数字足迹(王正位等,2020)及公共政务数据等数字足迹(余楷文等,2024)被证实有助于降低违约风险。与非金融类大数据信息相比,银行数字足迹更能精准刻画客户金融行为变动,准确反映客户信用状况等关键信息,而且具有无须交叉验证和可信度高的特征,但鲜有研究关注农户银行数字足迹在防控贷款违约风险中的作用。

基于此,本文通过构建信号传递理论模型,揭示农户银行数字足迹对贷款违约风险的影响,并运用某省份农商银行百万条农户贷款数据进行实证检验。研究发现,农户银行数字足迹在防控贷款违约风险中发挥了信号传递作用,能够显著降低贷款违约风险,且降低硬信息丰富型农户群体贷款违约风险的作用强于软信息丰富型农户群体。机制分析表明,银行数字足迹为农户传统硬信息提供信息增量,强化了降低农户贷款违约风险的效果;而对农户传统软信息并未提供有效信息增量。异质性分析显示,在银行数字化转型程度越高、地区农村信用体系建设成熟度越高及地处普惠金融改革试验区的样本中,农户银行数字足迹降低贷款违约风险的作用效果越强。

与现有文献相比,本文贡献如下。第一,提出农户银行数字足迹的概念,并将数字足迹对金融机构经营管理的影响文献由电商数字足迹拓展至银行数字足迹。农户银行数字足迹是指农户在使用银行提供的线上服务时留下的数据信息。本文通过百万条银行贷款数据,发现农户银行数字足迹对银行经营管理的重要性与电商数字足迹类似,为数据资源在金融业的有效运用提供了新的证据。第二,揭示农户银行数字足迹影响贷款违约风险的新机制。本文从经典风险管理的信息包括硬信息与软信息的分析框架出发,剖析了农户银行数字足迹这一数据资源如何在硬信息和软信息的基础上发挥作用,发现农户银行数字足迹之所以能够降低贷款违约风险,主要是因为其对硬信息丰富型农户群体贷款违约风险防控提供了额外的信息增量,增强了银行风险管理能力。第三,本文为研究农户贷款违约风险提供了新视角。已有关于农户贷款违约风险的研究主要集中于农户的传统财务数据和实物抵质押品(李庆海等,2020;马九杰等,2021)、社会网络(Wang等,2022)等,本文将农户贷款违约风险的研究拓展至数字金融发展背景下,揭示了农户银行数字足迹成为影响农户贷款违约风险的关键因素之一,这为农户贷款风险管理的相关研究提供了新的事实依据。第四,本文还具有较强的实践意义。研究结论能够清晰回答银行内部大数据对提升风险管理水平的重要作用,对持续改进风险管理模式具有重要实践意义。

二、理论分析与研究假说

(一)数字足迹的概念

“数字足迹”的概念最早由 Weaver 和 Gahegan(2007)提出,指用户在使用互联网各项服务中留下的能反映其行为规律特征的信息。随着数字金融的发展,Berg 等(2020)提出“数字足迹”指客户使用互联网访问电商平台或在网站上注册而留下的行为信息。据此,本文首次提出“农户银行数字足迹”这一概念,指农户在使用数字银行服务(如手机银行 App、网上银行等)过程中产生的数据信息。农户银行数字足迹是农户金融交易直接相关的行为数据,包括交易数据信息、行为数据信息、偏好信息和社交信息四类。^①

(二)模型设定

银行与农户之间存在的信息不对称是贷款违约风险产生的根源,而信号传递是缓解信息不对称的关键措施(Spence, 1973)。既有研究揭示了学历(Huang 等, 2020)、声誉(张晓玫等, 2018)、客户活跃度(Tang 等, 2023)等因素可作为传递贷款违约风险的信号。本文基于信号传递理论,揭示农户银行数字足迹作为传递贷款违约风险重要信号的效果及作用机制。

1. 农户类型与策略选择

假设农户有两种类型(T):低贷款违约风险农户($T = L$)和高贷款违约风险农户($T = H$), $T \in \{L, H\}$ 。假设农户会采取两种行动策略(a):积累银行数字足迹($a = ABUN$)或不积累银行数字足迹($a = SHORT$), $a \in \{ABUN, SHORT\}$ 。

2. 信号假设

假设农户银行数字足迹可作为农户传递自身风险类型的信号,原因如下。第一,农户在银行内部积累数字足迹的多与少,决定了银行可用于评估其贷款违约风险的信息量多少,农户信息量越充足,则银行与农户之间的信息不对称程度越低,银行可以相应调整对农户贷款决策,因而低违约风险借款农户有更强的动机积极积累银行数字足迹,以此传递自身信息。第二,农户积累银行数字足迹存在一定成本,且不同贷款违约风险农户传递相同信号的成本存在差异。一方面,农户积累银行数字足迹的成本高于不积累银行数字足迹的成本。农户积累的银行数字足迹越多,表明使用银行数字金融服务越频繁,在积累银行数字足迹中投入的成本越高。假设农户积累银行数字足迹付出的成本为 C ,其中,低贷款违约风险农户 L 积累银行数字足迹的成本为 C_1 ,不积累银行数字足迹的成本为 C_2 , $C_1 > C_2$ 。另一方面,高贷款违约风险农户 H 伪造与低违约风险农户 L 相同的银行数字足迹(即向银行传递相同信号),需要投入大量的时间和精力,成本极高。假设高贷款违约风险农户 H 积累银行数字足迹的成本为 $\kappa_1 C_1$,不积累银行数字足迹的成本为 $\kappa_2 C_2$,其中 $\kappa_1 > 1$, $\kappa_2 > 1$,且 $\kappa_1 C_1 > \kappa_2 C_2$ 。综上,农户银行数字足迹具有信号观测程度高和信号成本高的典型特征,满足作为信号的基本条件(Connelly 等, 2011)。

3. 银行行动策略

农户与银行之间存在信息不对称,银行作为信息劣势方,仅能获得农户违约风险类型的先验概率 $p(T)$ 。^②农户以银行数字足迹 Dig 作为传递其贷款风险违约类型的信号;银行在观测到农户

^① 受篇幅所限,未报告具体内容,请见线上附录。

^② 由于在未观测到信号时,银行无法判断农户违约风险高低,为便于分析,均设为 0.5。

银行数字足迹 Dig 后通过贝叶斯法则修正先验概率 $p(T)$, 得到农户贷款违约风险类型的后验概率 $\tilde{p}(T|Dig)$ 。据此, 银行的行动策略 (b) 有两种, 即终止贷款 ($b = STOP$) 和继续贷款 ($b = CONTINUE$), $b \in \{STOP, CONTINUE\}$ 。

4. 收益函数

假设农户初始禀赋的收益为 U_1 , 获得贷款的收益为 U_2 , 农户获得银行贷款额度为 q , r 为贷款利率, 到期还本付息 $q(1+r)$, $U_2 > q(1+r)$ 。农户违约的惩罚记为 $P(q)$, 假设对积累银行数字足迹的贷款违约农户的惩罚记为 P_a , 对不积累银行数字足迹的贷款违约农户的惩罚记为 P_s 。银行从惩罚中所获得的补偿记为 λ , $\lambda \in (0, 1)$, $\lambda P < q$ 。假设农户净收益为 $U(a, b, T)$, 银行净收益为 $\pi(a, b, T)$, 由农户是否积累银行数字足迹行为 a 、银行行动策略 b 和农户贷款违约风险类型 T 共同决定。农户和银行不同行动策略组合下的收益见表 1。^①

表 1 农户与银行不同行动策略下的收益函数

a. 低违约风险农户 L			
行动策略		银行	
		$CONTINUE$	$STOP$
农户	$ABUN$	$\{U_1 + U_2 - q(1+r) - C_1, rq\}$	$\{U_1 - C_1, 0\}$
	$SHORT$	$\{U_1 + U_2 - q(1+r) - C_2, rq\}$	$\{U_1 - C_2, 0\}$
b. 高违约风险农户 H			
行动策略		银行	
		$CONTINUE$	$STOP$
农户	$ABUN$	$\{U_1 + U_2 - \kappa_1 C_1 - P_a, \lambda P_a - q\}$	$\{U_1 - \kappa_1 C_1, 0\}$
	$SHORT$	$\{U_1 + U_2 - \kappa_2 C_2 - P_s, \lambda P_s - q\}$	$\{U_1 - \kappa_2 C_2, 0\}$

(三) 均衡分析与研究假说

在分离均衡下, 农户银行数字足迹丰富程度高低可作为银行准确判断农户违约风险高低的信号。实现分离均衡需满足两个前提条件, 一是低违约风险农户认为银行数字足迹可传递自身违约风险信息, 银行可根据其数字足迹做出风控决策, 即低违约风险农户积累银行数字足迹带来的净收益始终高于不积累银行数字足迹带来的净收益, 满足 $U_1 + U_2 - q(1+r) - C_1 > U_1 - C_2$; 二是高违约风险农户不积累银行数字足迹带来的净收益始终高于积累银行数字足迹带来的净收益, 满足 $U_1 - \kappa_2 C_2 > U_1 + U_2 - \kappa_1 C_1 - P_a$ 。综上, 需满足条件 1, 即 $\frac{\kappa_2 C_2}{\kappa_1} + \frac{U_2 - P_a}{\kappa_1} < C_1 < C_2 + U_2 - q(1+r)$, 才能形成分离均衡。分离均衡如下式所示:

$$(SE) \text{ 分离均衡: } \begin{cases} a^*(T=L) = ABUN, a^*(T=H) = SHORT \\ b^*(a=ABUN) = CONTINUE, b^*(a=SHORT) = STOP \\ \tilde{p}(T=L|a=ABUN) = 1, \tilde{p}(T=H|a=ABUN) = 0 \\ \tilde{p}(T=L|a=SHORT) = 0, \tilde{p}(T=H|a=SHORT) = 1 \end{cases} \quad (1)$$

^① 农户和银行不同行动策略组合详见线上附录。

命题1:若条件1成立,则存在唯一的分离均衡。在分离均衡中,银行数字足迹成为传递农户违约风险的信号,低违约风险农户选择积累银行数字足迹,而高违约风险农户选择不积累银行数字足迹。

条件1成立的关键是低违约风险农户积累银行数字足迹的成本 C_1 需在一定区间范围内。低违约风险农户积累银行数字足迹的成本 C_1 不能过高,否则会失去积累银行数字足迹的动力;同时, C_1 不能过低,否则高违约风险农户也有积累银行数字足迹的动力,难以形成分离均衡。此外,对于积累银行数字足迹的农户违约惩罚 P_d 应足够大,降低高违约风险农户伪装成低违约风险的动机,从而使得信号传递能够形成分离均衡,即农户一旦传递银行数字足迹信号,则为低违约风险农户;反之,则为高违约风险农户。^①

一方面,农户积累银行数字足迹具有一定成本 C_1 。农户的银行数字足迹积累越丰富,表明其使用银行数字金融服务越频繁,而积累数字足迹需要农户投入更多的时间、精力等成本。另一方面,对于积累银行数字足迹的农户而言,违约惩罚 P_d 更为严重。农户贷款违约可能会面临禁止使用数字金融服务等惩罚,对于积累充足银行数字足迹的农户而言成本更高。由于银行数字足迹的不可篡改特征和高度可观测特征,高违约风险农户伪造数字足迹的成本极高,从而形成了两类农户的分离均衡。如Björkegren和Grissen(2020)通过分析手机使用数据,如通话时长、短信频率、App使用情况等,发现数字足迹揭示了用户的社交活跃度、生活习惯及潜在需求,进而可用于评估其信用风险和还款意愿。因此,农户银行数字足迹可作为传递违约风险的信号。综上,本文提出如下研究假说。

H1:农户银行数字足迹能够降低贷款违约风险。

根据信号传递理论,信号传递效应会在信号发送者不同特征条件下发生变化(Spence,1973; Connelly等,2011)。因此,根据农户信息类型,本文考察了农户银行数字足迹对传统硬信息和传统软信息丰富型农户群体贷款违约风险的影响。

农户银行数字足迹为传统硬信息提供了有效的信息增量,与其形成互补关系。在传统硬信息丰富的农户贷款时,银行凭借其抵押物、提交的财务报表等低频硬信息做出贷款决策,银行数字足迹则为传统硬信息提供了更为高频、实时的动态量化信息。一是银行数字足迹包含农户数字金融交易累积的消费记录、交易金额等信息,银行可判断农户消费习惯和经营状况等,实时监控农户还款能力。二是通过二维码支付、手机银行转账及缴费等交易数据,收入支出情况等高频、实时信息,银行可以更加便捷地实时评估农户经营状况和信用水平,强化违约风险监测能力。三是通过对农户使用数字金融交易产生的消费流量和收单数据的分析,银行可以及时发现异常交易、欺诈行为等风险事件。因此,农户银行数字足迹为传统财务报表等硬信息提供了有效的信息增量,若银行观测到农户银行数字足迹更丰富,可能会根据此信号降低传统硬信息丰富型农户贷款违约风险的后验概率。

然而,农户银行数字足迹为传统软信息提供的信息增量有限,未与之形成有效的互补关系。农户的社会关系、声誉、个人品行等软信息往往依赖银行的本土优势、人际关系和经验来获取(Berger等,2017),因为这类信息与特定的地域、文化和社交环境紧密相关(Ge等,2017)。尽管农户手机银行使用习惯、产品和服务偏好等银行数字足迹能够提供大量信息,但银行数字足迹难以准确识别客户个人能力、品行等主观特征,导致信号可观测程度较低,银行仍需通过线下网点员工

^① 本文还对混同均衡情形进行分析,混同均衡不成立,详见线上附录。

与客户的日常互动、面对面的交流识别风险,通过人工经验判断、实地观察、人际沟通等方式进行交叉验证(秦志华等,2018;Liu,2022)。因此,银行数字足迹为传统软信息提供的信息增量有限。综上,农户银行数字足迹对不同传统信息类型的农户违约风险的影响存在差异,银行数字足迹降低传统硬信息丰富型农户贷款违约风险的作用强于传统软信息丰富型农户。据此,本文提出如下研究假说。

H2:农户银行数字足迹降低传统硬信息丰富型农户群体贷款违约风险的效果强于传统软信息丰富型农户群体。

三、数据来源、变量选取与模型构建

(一)数据来源

农商银行逐笔贷款数据来源于某省份农信系统,包括2018—2022年农户基本信息和贷款台账数据。农商银行经营及财务数据等通过国泰安数据库进行收集整理。县域、地市级经济社会发展数据来源于《中国县域统计年鉴》及《中国城市统计年鉴》。针对缺失数据,通过查询《国民经济和社会发展统计公报》进行补充。在对农户贷款台账重要指标空缺值进行删除和异常值进行1%双边缩尾处理后,最终获得2018—2022年1796646条农户贷款样本数据。^①

(二)变量定义与说明

1.被解释变量

本文的被解释变量为农户贷款是否产生违约风险(*Default*),为二元虚拟变量。参考余楷文等(2024)的研究,依据中国人民银行制定的《贷款风险分类指导原则》,若农户贷款属于五级分类中的“关注”“次级”“可疑”“损失”任一类,则认定为农户贷款产生违约风险,*Default*赋值为1,若属于“正常”则赋值为0。在稳健性检验中,本文以农户贷款是否属于“次级”“可疑”“损失”三类定义贷款违约风险(*NP*)。

2.核心解释变量

本文的核心解释变量为农户银行数字足迹,包含农户是否积累银行数字足迹(*In_Dig*)和农户银行数字足迹丰富程度(*Dig_amt*)两个维度。由于数字金融服务多通过手机应用程序直接触达农户,若农户开通且使用手机银行完成交易,则认为农户积累银行数字足迹,*In_Dig*赋值为1,否则为0。

农户银行数字足迹丰富程度最为直接的衡量方式为银行对每个农户多维度数字足迹进行综合度量,由于这涉及银行的核心涉密数据,且银行在实际运用数字足迹的过程中,通过大模型难以精准度量每个维度所占权重。因此,本文参考Jayasuriya等(2023)以网页使用程度衡量数字足迹的思路,以“农户使用手机银行交易金额/年度交易金额”(*Dig_amt*)衡量农户银行数字足迹的丰富程度。这是因为,以手机银行为例,农户的手机银行使用越多,自然会留下的痕迹越多,数字足迹越丰富。*Dig_amt*越高,意味着农户越倾向于使用手机银行进行支付、转账、理财等金融活动,就越能积累更为丰富的银行数字足迹。相比于使用频率、应用打开次数等指标,以*Dig_amt*作为衡量标准,能够量化反映农户在手机银行平台上的活跃程度和使用深度,直观体现农户银行数字足迹的有效信息量。

^① 样本代表性说明详见线上附录。

此外,结合已有关于农户贷款违约风险影响因素的研究(苏治、胡迪,2014;马九杰等,2021;郑海荣等,2023),本文对农户特征、银行特征以及地区特征变量进行控制。表2为主要变量说明。

表 2		主要变量说明					
变量名			变量含义	均值	标准差	最小值	最大值
被解释变量	贷款违约风险	Default	关注、次级、可疑、损失=1;正常=0	0.0586	0.2348	0	1
		NP	次级、可疑、损失=1;正常、关注=0	0.0174	0.1309	0	1
核心解释变量	农户是否积累银行数字足迹	In_Dig	是否开通并使用手机银行,是=1,否=0	0.7222	0.4479	0	1
	农户银行数字足迹丰富程度	Dig_amt	农户使用手机银行交易金额/年度交易金额	0.8195	0.2688	0	1
控制变量	年龄	Age	农户年龄(岁)	40.2003	9.2161	18	70
	性别	Gender	男性=1,女性=0	0.7846	0.4111	0	1
	受教育程度	Edu	受教育年限(年)	10.0678	2.5213	0	16
	收入	Income	家庭年收入(万元),取对数	21.4624	14.9105	4	100
	资产	Asset	家庭总资产(万元),取对数	112.2457	99.4058	5	600
	婚姻状况	Marriage	是否已婚,是=1,否=0	0.7955	0.4034	0	1
	政治面貌	Party	是否为中共党员,是=1,否=0	0.0149	0.1211	0	1
	个人品行	Conduct	较差=1,一般=2,中等=3,良好=4,优秀=5	4.1275	0.5229	1	5
	贷款利率	Rate	农户贷款月利率	0.0079	0.0019	0.0015	0.0180
	存贷比	LDR	贷款总额/存款总额	0.6492	0.0825	0.3900	0.8900
	产权形式	Reform	农信社是否改制,是=1,否=0	0.4940	0.5000	0	1
	经济发达程度	Economy	人均GDP(万元),取对数	11.4342	0.3380	9.2500	13.5400
	产业结构	Insdustry	第一产业增加值/GDP	0.1211	0.0751	0	0.3800
	金融发展水平	Finance	银行业金融机构贷款余额/GDP	0.7328	0.3371	0.1900	2.2500
	移动电话普及率	Mobile	移动电话用户数/户籍人口数	1.1082	0.1034	0.9100	1.5400
	银行竞争程度	HHI	地区银行业传统市场竞争的赫芬达尔-赫希曼指数	0.4085	0.0786	0.3092	1.9630

注:收入和资产仅在回归中取对数,表中的描述性统计为取对数前的原始值。

(三)模型构建

本文采用固定效应模型,检验农户银行数字足迹对贷款违约风险的影响,模型设定如下:

$$Default_{i,t} = \alpha_0 + \alpha_1 Dig_Inf_{i,t}^{Type} + \alpha_2 Control_{i,t} + \gamma_t + v_r + v_c + \varepsilon_{i,t} \tag{2}$$

$$Default_{i,t} = \beta_0 + \beta_1 Dig_Inf_{i,t}^{Type} + \beta_2 Dig_Inf_{i,t}^{Type} \times dum_Inf_{i,t} + \beta_3 dum_Inf_{i,t} + \beta_4 Control_{i,t} + \gamma_t + v_r + v_c + \varepsilon_{i,t} \tag{3}$$

其中, i 代表农户, t 为年份。 $Default_{i,t}$ 表示农户贷款违约风险, $Dig_Inf_{i,t}^{Type}$ 表示农户是否积累银

行数字足迹(In_Dig)与农户银行数字足迹丰富程度(Dig_amt), $Control_{i,t}$ 为控制变量。 γ_t 代表时间固定效应, v_r 和 v_c 分别为县域、城市的地区固定效应, $\varepsilon_{i,t}$ 为服从正态分布的随机扰动项。

式(3)加入农户银行数字足迹与是否为硬信息丰富型农户的二元虚拟变量($dum_Inf_{i,t}$)^①的交互项 $Dig_Inf_{i,t}^{Type} \times dum_Inf_{i,t}$ 进行回归,以考察农户银行数字足迹对传统硬信息和软信息丰富型农户群体贷款违约风险影响的差异,其他变量同式(2)。

四、实证结果分析

(一)基准回归

表3展示了农户银行数字足迹对贷款违约风险的回归结果。第(1)列和第(3)列仅控制时间和地区固定效应,结果显示农户银行数字足迹的系数均在1%的水平下显著降低贷款违约风险,初步验证了H1。第(2)列和第(4)列均在前一列基础上加入了控制变量,结果发现回归系数的显著性和方向基本未发生改变,再次验证了H1。具体地,农户银行数字足迹丰富程度(Dig_amt)每提升1个标准差,贷款违约风险将下降0.489~0.624个百分点。^②这与Campbell等(2022)和Huang等(2020)关于数字消费贷款、P2P网贷平台违约风险的研究发现一致,农户银行数字足迹使得银行能够更为便捷地通过手机银行监测农户行为,银行能够及时发现农户风险行为并采取风险防控措施,显著降低农户贷款违约风险。

表3 农户银行数字足迹对贷款违约风险的影响

	贷款违约风险(Default)			
	(1)	(2)	(3)	(4)
In_Dig	-0.0561*** (0.0012)	-0.0526*** (0.0013)		
Dig_amt			-0.0232*** (0.0015)	-0.0182*** (0.0016)
控制变量	否	是	否	是
时间、地区固定效应	是	是	是	是
样本量	1796631	1454858	1118857	922734
Adj.R ²	0.024	0.051	0.008	0.015

注:括号内为个体聚类稳健标准误。***、**和*分别表示在1%、5%和10%的水平下显著。如无特别说明,下同。

① 本文以农户贷款担保方式作为划分硬信息丰富型农户和软信息丰富型农户的标准(Angori等,2019;周鸿卫、田璐,2019;盛天翔等,2020),即若农户贷款为抵押、质押或保证贷款,则农户属于硬信息丰富型农户;若农户贷款为信用贷款,则农户属于软信息丰富型农户。这是由于在实际操作中,农商银行等中小银行的信用贷款是将关系型贷款技术和信用评级技术结合在一起而形成的贷款方式。对于缺乏抵质押物的农户,信息收集与验证需通过信贷员直接接触或通过熟人关系调查完成,信用评级结果本质上是对软信息处理的结果。降低对报表、抵质押物等硬信息的要求,重视一手获取软信息的贷款模式正是关系型贷款的核心。因此,农户信用贷款本质上可视为银行基于农户软信息决策的贷款。

② 根据表2, Dig_amt 的标准差为0.2688,以第(2)列和第(4)列的回归系数与 Dig_amt 的标准差相乘,得到边际效应。

(二)农户银行数字足迹对不同信息类型农户贷款违约风险的影响

表4为农户银行数字足迹对不同信息类型农户贷款违约风险的回归结果。第(1)、(2)列为交互项回归结果;第(3)、(4)列为对所有样本农户中软信息丰富型农户贷款违约风险的回归结果;第(5)、(6)列为对所有样本农户中硬信息丰富型农户贷款违约风险的回归结果。

第(1)列和第(2)列交互项的回归系数均显著为负,表明相较于软信息丰富型农户群体,农户银行数字足迹降低传统硬信息丰富型农户贷款违约风险的作用更强,初步证实了H2。此外,第(3)列回归系数的绝对值小于第(5)列回归系数的绝对值,表明农户银行数字足迹降低传统硬信息丰富型农户贷款违约风险的作用更强,进一步证实了H2。具体地,第(4)列和第(6)列的回归结果分别表明,农户银行数字足迹丰富程度(*Dig_amt*)每提升1个标准差,软信息丰富型农户的贷款违约风险将下降0.177个百分点,硬信息丰富型农户的贷款违约风险将下降0.315个百分点。^①这主要得益于银行数字足迹对分析传统硬信息丰富型农户的贷款违约风险提供了更为全面和动态的信息。传统硬信息,如财务报表等为历史信息,可能受到农户主观操作的影响,银行数字足迹能够实时反映农户的经济活动和信用状况,为银行提供更为全面和动态的评估信息,当银行同时拥有农户的传统硬信息和银行数字足迹时,二者互为补充,有助于银行对农户的信用状况进行动态监控和调整,加强贷后管理,降低贷款违约风险。

表4 农户银行数字足迹对不同信息类型农户贷款违约风险的影响

	交互项回归		软信息丰富型农户		硬信息丰富型农户	
	(1)	(2)	(3)	(4)	(5)	(6)
<i>In_Dig</i> × <i>dum_Inf</i>	-0.0256*** (0.0026)					
<i>Dig_amt</i> × <i>dum_Inf</i>		-0.0192*** (0.0031)				
<i>In_Dig</i>			-0.0205*** (0.0008)		-0.0321*** (0.0010)	
<i>Dig_amt</i>				-0.0066*** (0.0010)		-0.0117*** (0.0012)
控制变量	是	是	是	是	是	是
时间、地区固定效应	是	是	是	是	是	是

注:第(3)列和第(5)列回归的样本量一致,第(4)列和第(6)列回归的样本量一致,系数具有可比性。

(三)机制分析

为深入剖析农户银行数字足迹降低贷款违约风险的作用机制,本文基于经典风险管理的信息分析框架,检验农户银行数字足迹对农户传统硬信息与传统软信息的信息增量效果。

1. 农户银行数字足迹对传统硬信息的信息增量分析

农户的传统硬信息,如抵质押物、固定资产等,往往局限于静态的、历史的数据,难以全面反映农户当前的生产经营状况和未来的还款能力。农户银行数字足迹则为银行提供了关于农户生产、

① 根据表2,*Dig_amt*的标准差为0.2688,以第(4)列和第(6)列回归系数与*Dig_amt*的标准差相乘,得到边际效应。

经营等硬信息的重要信息增量。一方面,农户银行数字足迹实现了银行对农户生产及经营活动的实时监测。农户银行数字足迹中的交易记录可以反映农户的生产经营活动,如购买生产资料、销售农产品等交易数据,有助于银行更准确地评估农户当前的生产能力和市场前景,为银行评估农户违约风险提供了信息增量。另一方面,农户银行数字足迹有助于银行全面了解农户生产和经营资金流变动状况,这为评估农户的还款能力和贷款风险提供了重要的信息增量。综上所述,农户银行数字足迹对传统硬信息提供了重要的信息增量,特别是在生产及经营等硬信息方面,这些增量信息有助于银行更全面地动态了解农户的生产经营活动状况及资金流状况,降低贷款违约风险。

本文从农户职业、收入来源、结息频率三个维度,区分农户生产及经营相关硬信息的丰富程度,采用如下固定效应模型进行回归,检验农户银行数字足迹对硬信息农户贷款违约风险的影响:

$$Default_{i,t} = \delta_0 + \delta_1 Dig_Inf_{i,t}^{Type} + \delta_2 Dig_Inf_{i,t}^{Type} \times dum_Hard_{i,t}^h + \delta_3 dum_Hard_{i,t}^h + \delta_4 Control_{i,t} + \gamma_t + v_r + v_c + \varepsilon_{i,t} \quad (4)$$

式(4)中加入农户银行数字足迹与代表不同维度硬信息丰富型农户的二元虚拟变量的交互项 $Dig_Inf_{i,t}^{Type} \times dum_Hard_{i,t}^h$ 进行回归, $h = \{Farmer, Business_inc, Paid\}$,以考察农户银行数字足迹对不同维度的硬信息丰富型农户贷款违约风险影响的差异,^①其他变量与式(2)一致。

表5为农户银行数字足迹对不同维度硬信息丰富型农户贷款违约风险的影响,反映了银行数字足迹为传统硬信息提供的信息增量。回归结果表明,农户银行数字足迹对农户生产、经营、交易信息等硬信息提供的信息增量更多,强化了降低农户贷款违约风险的效果。具体而言,第(1)~(2)列回归结果表明,相较于其他农户,农户银行数字足迹对降低新型职业农民贷款违约风险的效果更强。类似地,第(3)列回归结果表明,农户银行数字足迹降低经商收入等硬信息丰富型农户群体的贷款违约风险作用更强。第(5)列回归结果表明,相较于结息频率较低的农户,结息频率更高的农户产生的交易记录等农户银行数字足迹有力补充了其在线下或其他渠道交易形成的硬信息,降低农户贷款违约风险。从上述结果可以看出,农户银行数字足迹并非完全替代硬信息,而是补充了农户生产和经营等硬信息的更新频率与信息含量,提供了信息增量,进而强化了降低农户贷款违约风险的效果。

表5 农户银行数字足迹对传统硬信息的信息增量

	贷款违约风险(Default)					
	(1)	(2)	(3)	(4)	(5)	(6)
$In_Dig \times Farmer$	-0.0124*** (0.0039)					
$Dig_amt \times Farmer$		-0.0144*** (0.0049)				

① 农户职业(*Farmer*)为是否为新型职业农民的虚拟变量,以区分农户经营信息及生产信息等硬信息含量高低。若农户为新型职业农民,则表明经营信息及生产信息相对丰富,普通农户则表明相对不丰富。这是由于新型职业农民通常具备较高的文化素质和专业技能,能够运用现代科技手段进行农业生产和管理,生产信息更为丰富。收入来源(*Business_inc*)为农户收入是否以经商收入为主的虚拟变量,这是由于相较于工薪、打工等固定收入来源,以经商收入为主的农户能够提供的经营资金流等硬信息更加丰富。结息频率(*Paid*)为农户是否为按月结息的虚拟变量,相较于按季度结息、按年结息,农户按月结息时产生的银行交易记录等硬信息更为丰富。

续表 5

	贷款违约风险 (Default)					
	(1)	(2)	(3)	(4)	(5)	(6)
$In_Dig \times Business_inc$			-0.0162** (0.0082)			
$Dig_amt \times Business_inc$				0.0024 (0.0097)		
$In_Dig \times Paid$					-0.1042*** (0.0293)	
$Dig_amt \times Paid$						-0.0757 (0.0691)
控制变量	是	是	是	是	是	是
时间、地区固定效应	是	是	是	是	是	是

2. 农户银行数字足迹对传统软信息的信息增量分析

传统软信息主要涉及农户与银行之间的长期关系、农户在当地的声誉和地位,以及与其他金融机构、商业伙伴的关系等,这些信息对于银行评估农户信用状况和违约风险至关重要。然而,农户银行数字足迹主要记录的是农户在使用银行线上设备时留下的行为和交易记录,如支付习惯、交易频率和金额等,这些信息并不直接反映农户与银行的关系质量或农户在当地的声誉和地位。因此,数字足迹在关系类软信息方面提供的信息增量有限,导致降低贷款违约风险的作用弱于硬信息丰富型农户群体。本文从农户本地联结、人际关系两个维度度量农户传统软信息差异,将式(4)中的 $dum_Hard_{i,t}^h$ 替换为代表不同维度软信息丰富型农户的二元虚拟变量 $dum_Soft_{i,t}^s$, $s = \{Live_time, Social_exp\}$, 并进行回归,以考察银行数字足迹对不同维度的软信息丰富型农户贷款违约风险影响效应的差异。^①

表6为农户银行数字足迹对不同维度软信息丰富型农户贷款违约风险的影响,反映了银行数字足迹对农户不同类别的传统软信息提供的信息增量。回归结果表明,农户银行数字足迹为农户本地联结、人际关系等软信息提供的有效信息增量较少,表现为交互项的回归系数多为不显著。具体而言,第(1)列农户银行数字足迹对关系类软信息未提供显著的信息增量,相较于本地联结程度较低的农户,农户银行数字足迹对本地联结程度较高农户的贷款违约风险的影响无显著差异,仅在第(2)列交互项的回归系数在10%的水平下显著为负,表明农户银行数字足迹丰富程度提高,能够为关系类软信息提供部分信息增量。第(3)、(4)列的回归结果证实,农户银行数字足迹对人际关系较为丰富农户的贷款违约风险的影响效应,与人际关系相对不丰富的农户相比也不存在显著差异。从上述结果可以看出,农户银行数字足迹对传统软信息未产生显著的信息增量,这是导致农户银行数字足迹降低传统软信息丰富型农户违约风险作用较弱的根本原因。这一研究结论与社交媒体、电商平台浏览记录等信息有助于识别违约风险的研究结论存在差异(Berg等,2020; Wang等,2022),主要是由于农户在使用手机银行时留下的社交类信息相当有限,且银行凭借地缘、人缘等本土优势收集的软信息已能够满足风险识别的信息需求。

① 农户本地联结(Live_time)定义为农户在本地居住时间在6年(含)以上赋值为1,5年(含)以下赋值为0;农户人际关系(Social_exp)则以年交际应酬支出中位数为高低两组,高组赋值为1,低组赋值为0。

表6 农户银行数字足迹对传统软信息的信息增量

	贷款违约风险 (Default)			
	(1)	(2)	(3)	(4)
$In_Dig \times Live_time$	-0.0029 (0.0101)			
$Dig_amt \times Live_time$		-0.0163* (0.0099)		
$In_Dig \times Social_exp$			0.0005 (0.0037)	
$Dig_amt \times Social_exp$				-0.0025 (0.0041)
控制变量	是	是	是	是
时间、地区固定效应	是	是	是	是

(四)内生性讨论与稳健性检验

1. 内生性讨论

(1) 准自然实验法。本文构建农户是否使用手机银行的虚拟变量 (Dig_dum) 与首次开始使用手机银行的时间虚拟变量 ($Time$) 的交互项, 替代银行数字足迹变量进行回归。^① 回归结果见表7 Panel A 第(1)列和第(2)列。实证结果显示, 交互项 $Dig_dum \times Time$ 的系数显著为负, 证实了农户银行数字足迹能够显著降低贷款违约风险, 不存在显著的反向因果关系。为检验农户银行数字足迹对软信息、硬信息丰富型农户贷款违约风险影响效应的差异, 构建 $Dig_dum \times Time$ 与是否为硬信息丰富型农户 (dum_Inf) 的三重交互项进行回归, 回归结果见第(3)列和第(4)列, 实证结果显示, 交互项的系数显著为负, 进一步证实了农户银行数字足迹降低硬信息丰富型农户贷款违约风险的作用更强。

(2) 工具变量法。本文参考尹志超和张诚(2019)的研究, 选用同年份同营业网点同年龄段农户的平均银行数字足迹丰富程度 (\overline{Dig})、^② 农户所在县域北京大学数字普惠金融指数 ($DFII$) 作为农户银行数字足迹丰富程度的工具变量进行 2SLS 估计。农户银行数字足迹的积累易受到同地区周围同龄人的影响, 满足工具变量与内生解释变量的相关性要求; 但 \overline{Dig} 、 $DFII$ 并不会直接影响农户贷款违约风险, 满足工具变量的外生性要求。回归结果见表7 Panel B。DWH F 值均在 1% 的水平下显著, 即存在内生性。一阶段 F 值以及 Kleibergen-Paap rk LM 值均远大于临界值, 表明两个工具变量满足相关性。在加入控制变量后, Hansen J 检验的 P 值大于 0.1, 工具变量的排他性得到满足。回归系数显著为负, 表明在考虑内生性问题后, 基准回归结果仍成立。

① 具体地, 本文以样本期内在多笔借贷中从不使用手机银行到使用手机银行的样本农户作为实验组, 赋值为 1; 以样本期内始终不使用手机银行的样本农户作为控制组, 赋值为 0。获得样本量为 1629932 个, 其中控制组样本量为 1485427 个, 实验组样本量为 144505 个。

② 参考何婧和李庆海(2019)的处理方法, 根据农户借贷时的年龄划分为 5 个年龄段, 分别为 18 ~ 29 岁、30 ~ 39 岁、40 ~ 49 岁、50 ~ 59 岁和 60 岁及以上。

(3)Heckman两阶段选择模型。^①本文参考何婧等(2017)的研究,采用 Heckman两阶段选择模型进行回归,以解决潜在的样本选择偏差问题。回归结果显示,银行数字足迹变量的系数均为负,表明基准回归结果不受样本选择性偏误影响。

(4)交互固定效应模型。^②本文在回归中控制了时间与地区交互固定效应,以控制同时随时间和地区变化的不可观测的其他因素的潜在影响。回归结果显示,核心解释变量的回归系数显著为负,表明不存在显著的遗漏变量问题。

2.稳健性检验

为证实基准回归结果的稳健性,一是改变被解释变量度量方式,使用 *NP* 替换基准回归中的被解释变量 *Default*。二是替换核心解释变量,将银行数字足迹丰富程度的测度指标分别更换为“农户使用手机银行交易笔数/年度交易笔数”(*Dig_cnt*)以及“农户使用网上银行交易笔数/年度交易笔数”(*Dig_cnt_alt*)进行回归。三是排除新冠疫情冲击的影响,构建衡量新冠疫情冲击的交互项变量 *COVID* × *Time_c*,^③作为控制变量加入回归中。四是缩小样本区间,考虑到地级市与县域地区存在较大差异,本文剔除地级市的农户样本后重新进行回归检验。五是更换回归模型。考虑农户贷款违约风险为二元虚拟变量,存在取值受限的问题,参考吴雨等(2020)的做法,换用 Probit 模型进行回归。以上回归结果均显示,核心解释变量的回归系数仍显著为负,结果依然稳健。^④

表 7 内生性讨论

Panel A 准自然实验				
	贷款违约风险(<i>Default</i>)			
	(1)	(2)	(3)	(4)
<i>Dig_dum</i> × <i>Time</i>	-0.0340*** (0.0017)	-0.0282*** (0.0022)		
<i>Dig_dum</i> × <i>Time</i> × <i>dum_Inf</i>			-0.0249*** (0.0028)	-0.0254*** (0.0037)
控制变量	否	是	否	是
时间、地区固定效应	是	是	是	是
Panel B 工具变量法				
	贷款违约风险(<i>Default</i>)			
	(1)	(2)		
<i>Dig_amt</i>	-0.0720*** (0.0050)	-0.0509*** (0.0055)		
控制变量	否		是	

① 具体地,在第一阶段通过 Probit 模型,选择农户贷款额度、个人银行信用记录作为排他性约束量,估计农户银行数字足迹对农户是否违约的影响;在农户已存在违约的情况下加入逆米尔斯比率(λ)进行第二阶段回归,检验银行数字足迹对农户贷款违约风险(*NP*)的影响。本文采用最大似然估计法(MLE)估计 Heckman 两阶段选择模型,Wald 检验值为 13131.41,显著异于 0,说明 Heckman 两阶段选择模型有效。限于篇幅,未报告 Heckman 两阶段选择模型的实证回归结果,留存备案。

② 限于篇幅,未报告交互固定效应模型的实证回归结果,留存备案。

③ 其中,*COVID* 为某省份各地级市 2020 年末每万人中新冠确诊病例数,*Time_c* 为时间虚拟变量,2020 年及之后年份取值为 1,其余年份取值为 0。

④ 限于篇幅,回归结果未展示,请见线上附表。

续表 7

Panel B 工具变量法	贷款违约风险 (Default)	
	(1)	(2)
时间、地区固定效应	是	是
DWH F 值	132.821***	58.271***
一阶段 F 值	1251.96***	949.44***
Kleibergen-Paap rk LM 值	1272.064***	1014.324***
Hansen J P 值	0.0038	0.1614

五、进一步分析

(一)异质性分析^①

农户银行数字足迹能否降低贷款违约风险不仅取决于银行对数字信息的甄别能力,即银行自身数字化转型程度,还取决于地方政府对农村信用体系建设和普惠金融发展的重视程度。

1.银行数字化转型程度

银行数字化转型程度反映了银行数据处理水平以及对大数据、云计算等数字技术的运用能力(张一林等,2021;杨涛,2022),这会 影响银行数字足迹对农户贷款违约风险的影响效果。本文构建银行数字化转型程度指数(*Dif_dig*)^②与农户是否积累银行数字足迹(*In_Dig*)、农户银行数字足迹丰富程度(*Dig_amt*)的交互项并进行回归,考察银行数字足迹影响农户贷款违约风险的异质性特征。回归结果显示,交互项的回归系数显著为负,表明银行数字化转型强化了农户银行数字足迹降低贷款违约风险的作用效果。

2.农村信用体系建设成熟度

随着农村信用体系建设成熟度的提高,其为银行提供了更丰富的数据支持(李明贤、陈艳,2021),这将在更大程度上发挥农户银行数字足迹的信息增量作用。本文以农户信用档案建档率中位数对所有地区进行分组回归,并比较组间系数差异。回归结果显示,*In_Dig*、*Dig_amt*对农村信用体系建设较成熟地区农户贷款违约风险均具有显著负向影响,且影响效应显著强于欠成熟地区,表明在农村信用体系建设较成熟地区强化了农户银行数字足迹降低贷款违约风险的作用效果。

3.普惠金融改革试验区

国家级普惠金融改革试验区通过打破信息壁垒,丰富了银行可获取的农户信息含量(李晓龙、郝晓婧,2024),有助于强化农户银行数字足迹降低贷款违约风险的作用效果。本文以是否处于普惠金融改革试验区对所有农户样本进行分组回归,并比较组间系数差异。回归结果显示,加入控制变量后,*In_Dig*、*Dig_amt*对普惠金融改革试验区农户贷款违约风险均具有显著负向影响,且影响

^① 限于篇幅,回归结果未展示,请见线上附表。
^② 具体处理过程为:建立银行数字金融技术应用水平关键词词库,基于百度搜索引擎中的百度资讯,搜索样本银行和关键词共同出现的新闻,获取各家银行在各年度符合条件的新闻数量,最后汇总检索量到“银行-年份”层面,对数据处理后获得各家银行数字化转型程度指数 *Dif_dig*。

效应显著强于非试验区,表明在普惠金融改革试验区,农户银行数字足迹降低贷款违约风险的作用效果得到强化。

(二)拓展分析:农户银行数字足迹与贷款额度

农户银行数字足迹能够降低贷款违约风险,那么对于银行而言,农户银行数字足迹丰富程度提高是否有助于提高对农户的贷款额度?这一问题目前存在争议。既有研究表明,随着信息量增多,贷款额度也会增加(Lu等,2023)。然而,部分研究却发现,贷款机构在通过学历信息识别风险时存在偏差,尽管高学历者违约风险较低,但贷款机构并未为其提高贷款额度(廖理等,2015)。鉴于此,本文实证检验了农户银行数字足迹对贷款额度的影响。

表8回归结果显示,农户是否积累银行数字足迹(*In_Dig*)降低了农户的贷款额度,但随着农户银行数字足迹丰富程度(*Dig_amt*)的提高,农户的贷款额度有所提升。其中,农户银行数字足迹对软信息丰富型农户群体贷款额度降低,但提升了对硬信息丰富型农户的贷款额度。这是由于软信息的收集与处理仍依赖农商银行的本土优势,引入银行数字足迹后使得原本依赖软信息发放贷款的决策更加谨慎,导致对软信息丰富型农户贷款额度有所减少,这也与Dong等(2023)和余楷文等(2024)的研究结论一致。农户银行数字足迹为硬信息丰富型农户的风险识别提供了高频、有效的信息增量,从而提升了贷款额度。

表 8 农户银行数字足迹、农户信息类型与贷款额度

	全部样本		软信息丰富型农户		硬信息丰富型农户	
	(1)	(2)	(3)	(4)	(5)	(6)
<i>In_Dig</i>	-0.0889*** (0.0112)	-0.1370*** (0.0114)	-0.3244*** (0.0317)	-0.3071*** (0.0358)	0.2354*** (0.0350)	0.1701*** (0.0387)
<i>Dig_amt</i>	0.3441*** (0.0240)	0.1933*** (0.0241)	-0.9695*** (0.0819)	-0.8795*** (0.0881)	1.3136*** (0.0878)	1.0727*** (0.0932)
控制变量	否	是	否	是	否	是
时间、地区固定效应	是	是	是	是	是	是

六、结论与启示

本文主要结论如下。(1)农户银行数字足迹在防控贷款违约风险中发挥了信号传递作用,降低了贷款违约风险。(2)农户银行数字足迹降低硬信息丰富型农户群体贷款违约风险的作用强于软信息丰富型农户群体。机制在于,农户银行数字足迹为农户传统硬信息提供信息增量,强化了降低农户贷款违约风险的效果,而对传统软信息提供的有效信息增量有限。(3)在银行数字化转型程度越高、地区农村信用体系建设成熟度越高以及地处普惠金融改革试验区的样本中,农户银行数字足迹降低贷款违约风险的作用效果越强。据此,本文提出如下政策建议。

第一,银行应重视农户银行数字足迹在降低贷款违约风险中发挥的信号传递作用,将其纳入贷款风险评估体系的核心部分。一是银行应建立完善的数据集成系统,整合来自手机银行、网上银行、线上支付等不同渠道的农户银行数字足迹,发挥其对降低贷款违约风险的独特价值。二是

优化贷款违约风险评估模型,将农户银行数字足迹作为贷款风险评估模型重要的数据信息来源,提高风险评估结果的准确性和可靠性。三是实时跟踪农户银行数字足迹变化,强化农户贷款风险动态监测,及时采取风险防范措施。

第二,银行应充分发挥农户银行数字足迹与传统信息在贷款违约风险防控中的互补作用,提升银行数字足迹对传统信息的信息增量效应。一是银行既要认识到农户银行数字足迹的信息价值,也不可忽视传统信息在风险评估中的基础作用,特别是传统软信息的作用,发挥好农户银行数字足迹和传统信息在贷款违约风险防控中的优势互补作用,筑牢风险防控安全网。二是银行需提升银行数字足迹对传统信息的信息增量。银行应深化对农户多样化数字足迹的挖掘与分析,可通过拓展场景金融,接入农户在电商平台、网络社交平台等留下的数字足迹,将多渠道收集的银行数字足迹与现有传统信息形成互补,通过持续优化算法模型,提高信息增量效果。

第三,优化丰富农户银行数字足迹等大数据的基础设施与制度保障,为农商银行等中小银行高效应用数字足迹提升风险管理能力保驾护航。一是农商银行应与金融科技大平台合作,完善数据收集、处理、存储及安全保护等基础设施建设,确保银行数字足迹的完整性与安全性。二是政府部门应通过政策引导、资金投入和技术支持,推动建立健全数据收集系统,为农商银行应用数字足迹提供坚实的制度保障。

参考文献:

- 1.何婧、李庆海:《数字金融使用与农户创业行为》,《中国农村经济》2019年第1期。
- 2.何婧、田雅群、刘甜、李庆海:《互联网金融离农户有多远——欠发达地区农户互联网金融排斥及影响因素分析》,《财贸经济》2017年第11期。
- 3.李明贤、陈艳:《金融科技、授信方式改进与涉农金融机构普惠能力提升》,《经济体制改革》2021年第2期。
- 4.李庆海、陈金鹏、郁杨成:《抵押与农户信贷违约风险:逆向选择还是道德风险?》,《世界农业》2020年第1期。
- 5.李晓龙、郝晓婧:《国家金融改革试验区设立能否提升金融服务实体经济效率?——来自准自然实验的经验证据》,《农村金融研究》2024年第12期。
- 6.廖理、吉霖、张伟强:《借贷市场能准确识别学历的价值吗?——来自P2P平台的经验证据》,《金融研究》2015年第3期。
- 7.马九杰、崔恒瑜、王雪、董翀:《设立村镇银行能否在农村金融市场产生“鲶鱼效应”?——基于农信机构贷款数据的检验》,《中国农村经济》2021年第9期。
- 8.秦志华、林莹、吴畏、王冬冬:《软信息对于网络信贷的影响机理研究——基于中国农业银行数据网贷的探索性案例分析》,《管理评论》2018年第11期。
- 9.余楷文、申宇、赵绍阳:《大数据对银行信贷行为的影响——来自数字社会信用平台的证据》,《经济研究》2024年第3期。
- 10.盛天翔、朱政廷、李伟雯:《金融科技与银行小微企业信贷供给:基于贷款技术视角》,《管理科学》2020年第6期。
- 11.苏治、胡迪:《农户信贷违约都是主动违约吗?——非对称信息状态下的农户信贷违约机理》,《管理世界》2014年第9期。
- 12.王正位、周从意、廖理、张伟强:《消费行为在个人信用风险识别中的信息含量研究》,《经济研究》2020年第1期。
- 13.吴雨、李成顺、李晓、弋代春:《数字金融发展对传统私人借贷市场的影响及机制研究》,《管理世界》2020年第10期。
- 14.杨涛:《金融科技与中小银行数字化转型研究》,《农村金融研究》2022年第7期。
- 15.尹志超、张诚:《女性劳动参与对家庭储蓄率的影响》,《经济研究》2019年第4期。
- 16.张晓玫、梁洪、卢露:《网络借贷中信息不对称缓解机制研究——基于信号传递和双边声誉视角》,《经济理论与经济管理》2018年第2期。
- 17.张一林、郁芸君、陈珠明:《人工智能、中小企业融资与银行数字化转型》,《中国工业经济》2021年第12期。
- 18.郑海荣、马九杰、王馨:《地理距离、银行数字化转型与金融机构农户信贷供给规模——来自F省农信系统的证据》,《华南师范大学学报(社会科学版)》2023年第11期。
- 19.周鸿卫、田璐:《农村金融机构信贷技术的选择与优化——基于信息不对称与交易成本的视角》,《农业经济问题》2019年

第5期。

20. Angori, G., Aristei, D., & Gallo, M., Lending Technologies, Banking Relationships, and Firms' Access to Credit in Italy: The Role of Firm Size. *Applied Economics*, Vol.51, No.58, 2019, pp.6139–6170.
21. Berg, T., Valentin, B., Ana, G., & Manju, P., On the Rise of FinTechs: Credit Scoring Using Digital Footprints. *The Review of Financial Studies*, Vol.33, No.7, 2020, pp.2845–2897.
22. Berger, A. N., Bouwman, C. H. S., & Dasol, K., Small Bank Comparative Advantages in Alleviating Financial Constraints and Providing Liquidity Insurance over Time. *The Review of Financial Studies*, Vol.30, No.10, 2017, pp.3416–3454.
23. Björkegren, D., & Grissen, D., Behavior Revealed in Mobile Phone Usage Predicts Credit Repayment. *The World Bank Economic Review*, Vol.34, No.3, 2020, pp.618–634.
24. Campbell, D., Grant, A., & Thorp, S., Reducing Credit Card Delinquency Using Repayment Reminders. *Journal of Banking & Finance*, Vol.142, No.9, 2022, pp.1–16.
25. Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R., Signaling Theory: A Review and Assessment. *Journal of Management*, Vol.37, No.1, 2011, pp.39–67.
26. Dong, Y., Qin, G., & Qiu, H., Big Tech Credit Score and Default Risk —Evidence from Loan-Level Data of a Representative Microfinance Company in China. *China Economic Review*, Vol.81, No.10, 2023, pp.1–14.
27. Ge, R., Feng, J., & Zhang, P., Predicting and Deterring Default with Social Media Information in Peer-to-Peer Lending. *Journal of Management Information Systems*, Vol.34, No.2, 2017, pp.401–424.
28. Huang, W., Qian, Y., & Xu, N., The Signaling Effects of Education in the Online Lending Market: Evidence from China. *Economic Modelling*, Vol.92, No.11, 2020, pp.268–276.
29. Jayasuriya, D. D., Ayaz, M. & Williams, M., The Use of Digital Footprints in the US Mortgage Market. *Accounting & Finance*, Vol.63, No.1, 2023, pp.353–401.
30. Liu, M., Assessing Human Information Processing in Lending Decisions: A Machine Learning Approach. *Journal of Accounting Research*, Vol.60, No.2, 2022, pp.607–651.
31. Lu, L., Wei, J., Wu, W., & Zhou, Y., Pricing Strategies in BigTech Lending: Evidence from China. *Financial Management*, Vol.52, No.2, 2023, pp.333–374.
32. Spence, M., Job Market Signaling. *The Quarterly Journal of Economics*, Vol.87, No.3, 1973, pp.355–374.
33. Tang, X., Zhu, J., He, M., & Feng, C., How Can We Learn from a Borrower's Online Behaviors? The Signal Effect of a Borrower's Platform Involvement on Its Credit Risk. *Electronic Commerce Research and Applications*, Vol.59, No.5, 2023, pp.1–15.
34. Wang, Y., Drabek, & Wang, Z., The Role of Social and Psychological Related Soft Information in Credit Analysis: Evidence from a Fintech Company. *Journal of Behavioral and Experimental Economics*, Vol.96, No.2, 2022, pp.1–19.
35. Weaver, S. D., & Gahegan, M., Constructing, Visualizing, and Analyzing a Digital Footprint. *Geographical Review*, Vol.97, No.3, 2007, pp.324–350.

Farmers' Digital Footprints and Loan Default Risk

ZHENG Hairong (Fujian Agriculture and Forestry University, 350002)

ZHANG Yang, HE Jing, MU Zhengshe (China Agricultural University, 100083)

Summary: Information asymmetry is a critical barrier that hinders banks' management of farmers' loan risks and underpins their reluctance to lend to farmers, and the difficulty and high cost associated with such lending. Banks increasingly recognize the importance of “digital footprints” in risk control, and exploring the value of data and unlocking its potential has become an important part of digital risk management. Compared with traditional bank information, can digital footprints provide additional information, effectively manage credit risk, and reduce lending risks? Answering this question is essential for managing loan risk in the era of Big Data and offers a pathway for how to use digital finance to deepen rural financial

services.

Based on this, we constructed a signaling game model and empirically tested it using millions of farmers' loan records from rural commercial banks to identify the impact of digital footprints on farmers' loan default risk. The results show that digital footprints play a signaling role in preventing and controlling loan default risk, significantly reducing farmers' loan default risk. Furthermore, this risk-reducing effect is more pronounced for farmers with abundant hard information than for groups of farmers with abundant soft information. Mechanism analysis shows that digital footprints provide incremental information on farmers' production, operations and other flows of hard information, which strengthens the risk-reducing effect. However, farmers' digital footprints do not provide meaningful additional insights for traditional soft information. Moreover, in areas with a higher degree of digital transformation among banks, a more mature rural credit system, and in inclusive finance pilot zones, the risk-reducing effect of digital footprints is stronger. This study reveals the role of digital footprints in identifying and monitoring farmers' loan default risk and provides a factual basis for better exploiting financial big data in the credit market.

This paper contributes to the existing literature in the following ways. First, it introduces the concept of farmers' digital footprints, extending the literature on the impact of digital footprints on financial institutions' operations from e-commerce to banking. Second, it uncovers a new mechanism by which the farmers' digital footprints affect their loan default risk. By distinguishing hard and soft information, this paper uses millions of farmers' loan records to analyze how the data of farmers' digital footprints reduces the risk of loan defaults mainly by providing incremental information for hard information-rich farmers, thereby enhancing the data effectiveness. Third, this paper provides a new perspective for studying farmers' loan default risk. While previous research has focused on traditional financial data and physical collateral, this paper expands the scope of research to the context of digital finance, identifying digital footprints as a key factor affecting the loan default risk, which provides a new factual basis for the related research on farmers' loan risk management.

The findings have strong practical implications. They enable banks to better segment farmer groups and fully play the complementary roles of digital footprints and banks' traditional soft and hard information in loan risk management.

Keywords: Digital Footprints, Default Risk, Signaling, Soft Information, Hard Information

JEL: G14, G21, D82

责任编辑:诗 华