
大数据征信与平台流量： 基于共享单车免押骑行的经验研究

曹光宇 刘畅 周黎安*

内容提要 基于大数据技术的个人征信服务是构建社会信用体系的一项重要内容。由于中国目前主要的大数据征信服务供应商是互联网平台企业的内嵌板块而非独立主体,其业务蓬勃发展的同时也引发了关于用户流量向头部平台集中的担忧。本文以芝麻信用分达到650分即可免押金骑行某品牌共享单车这一消费情境为例,利用消费者个体层面数据考察了大数据征信对互联网平台流量的影响。我们使用断点回归的识别策略发现,免押骑行政策导致用户在支付宝端口使用共享单车的比例提高了约12个百分点,而其主要竞争对手平台的流量均有不同程度的下滑。本文首次从计量上发现,互联网平台企业提供的大数据征信服务可能影响市场竞争结构,这为今后一个时期中国互联网平台企业反垄断规制和社会信用体系建设提供了参考。

关键词 大数据征信 共享单车 平台流量

一 引言

社会信用体系作为维系社会诚信的制度安排,是国家治理体系和社会主义市场经济体制的重要组成部分(国务院,2014;潘功胜,2014)。个人征信作为社会信用体

* 曹光宇:北京大学经济学院 电子邮箱:cgy1117@pku.edu.cn;刘畅(通讯作者):中国人民大学国家发展与战略研究院 电子邮箱:changliuecon@ruc.edu.cn;周黎安:北京大学光华管理学院 电子邮箱:zhoula@gsm.pku.edu.cn。

本文所用数据存储于合作企业的云服务器,其中的用户个人信息均已加密处理;作者仅能获得经验分析的结果,无法接触原始数据。作者感谢国家自然科学基金重大项目(72192844)的资助,感谢匿名审稿人的宝贵意见,文责自负。

系的一项核心内容,在建设良好市场秩序的过程中发挥了重要作用。蚂蚁金服旗下的芝麻信用以支付宝移动客户端(APP)所留存的用户数字足迹(digital footprint)作为数据基础,利用机器学习等前沿技术对个人信用做出评价,是当前中国用户数量最多、应用场景最广的大数据征信产品之一。芝麻信用分除了被金融机构用作授信依据之外,还在信用租赁等消费场景中发挥了重要作用。

以芝麻信用为代表的大数据征信行业的发展,虽然有助于社会信用体系建设,但同时引发了关于互联网平台流量集中的担忧。与费埃哲(Fair Isaac Corporation, FICO)等专营化征信机构不同,芝麻信用并非独立的市场活动主体,而是内嵌于支付宝平台的子板块,其应用场景和业务范围的扩大很可能影响用户流量在平台间的分布。以本文主要关注的共享单车免押骑行为例,该业务模式看似竞争中性(competition-neutral),实则可能挤压竞争对手,成为其母体平台(也即支付宝)吸引和锁定用户流量的工具。大数据征信为平台带来用户流量,用户在平台形成更多数据沉淀,而留存的数据则被进一步用于征信业务。这一“征信→流量→数据→征信”的正反馈闭环可能有助于互联网“巨头”攫取用户流量、构建市场势力。

本文利用“芝麻信用分达到650分即可免押金骑行某品牌共享单车”这一自然实验,通过断点回归(Regression Discontinuity, RD)的识别策略,探究大数据征信如何影响共享单车用户流量在不同端口之间的分布。与既有研究相比,本文的主要贡献体现在以下三个方面。首先,本文拓展了关于个人征信服务乃至整个社会信用体系的作用的认知。既有文献对以FICO信用分为代表的征信服务和各类声誉体系进行了广泛研究。其中,经验研究的核心关注点通常在于其对金融市场的影响(Keys *et al.*, 2009, 2010, 2012; Garmaise, 2012; Argyle *et al.*, 2020)或对个体行为的约束性(Cai *et al.*, 2014; Liberman, 2016; Homonoff *et al.*, 2021),而理论研究普遍关注评分(scoring)作为一种信息传递机制的效率及最优设计问题(Ball, 2019; Chatterjee *et al.*, 2020)。本文的切入点与上述研究均不相同,主要考察大数据征信作为技术手段对于互联网平台企业获取用户流量的作用。

其次,本文丰富了关于用户流量在不同互联网应用间传导方式的讨论。在为数不多的类似主题研究中, Li and Agarwal (2017) 关注 APP 间的业务整合, 发现 Facebook 与 Instagram 的整合显著提升了 Instagram 使用频次, 同时也带动了图片分享这一领域的总体发展; Lee *et al.* (2020) 聚焦于 APP 之间的交叉营销, 发现用户对 APP 多样性存在偏好, 会倾向于下载各类不同的 APP; Zheng *et al.* (2019) 则以微信为例讨论了超级 APP 对于其他应用的导流效应, 发现微信只对腾讯新闻等少数应用具备正向的导流效果, 流量分发

效果可能不及预期。与上述研究不同,本文主要关注超级APP的内嵌应用(芝麻信用)对于其母体平台(支付宝)的导流效果。就这一角度而言,本文的研究与Zheng *et al.* (2019)构成镜像对照,共同提供了关于超级APP与其内嵌应用间双向互动的证据链。

最后,在更广泛的意义上,本文从大数据征信这一具体的业务场景出发,探讨了数据对用户流量的潜在影响,加深了我们对数据这一新型生产要素的理解。既有研究多采用理论建模的方式刻画数据要素对企业竞争和市场结构的影响,但具体的关注点和机制设定则各有差异。例如,Condorelli and Padilla(2022)主要关注数据的跨市场价值,Hagi and Wright(2022)假定基于数据的产品改进可以提升消费者的支付意愿,Prüfer and Schottmüllerde(2021)则假设数据有助于降低生产成本^①。或许是囿于数据可得性和研究场景的匮乏,前述研究多停留在理论层面,至多提供了基于理论的模拟或数值解结果。本文利用个体微观数据和断点回归识别策略,为这一支文献提供了经验支持。

本文的剩余部分安排如下:第二部分介绍本文的主要研究背景,包括大数据征信服务、互联网平台流量和共享单车免押骑行等,并结合研究背景对相关文献进行介绍;第三部分介绍本文经验分析所用数据和识别策略,并检验识别策略的有效性;第四部分报告本文的基本结果;第五部分开展进一步分析;第六部分对全文进行总结。

二 研究背景及相关文献

(一)大数据征信服务

所谓征信,是指由专业化机构基于存量数据,利用云计算、人工智能等技术对市场主体的信用水平进行评估,针对其偿还意愿、偿还能力等进行画像,为授信、放贷等决策提供参考依据(Tang, 2019; Chatterjee *et al.*, 2020)。根据评估对象性质的不同,征信可以进一步分为企业征信和个人征信。在中国为数众多的个人征信服务中,覆盖人数最多、应用场景最广、社会影响最大的当属中国人民银行征信中心提供的个人信用报告和蚂蚁金服旗下芝麻信用管理有限公司推出的芝麻信用分^②。截至2015年4月,央

^① De Cornière and Taylor(2020)对此提出了一个更为一般性的理论框架,熊巧琴和汤珂(2021)、徐翔等(2021)对国内外关于数据生产要素的相关研究进行了系统性梳理。

^② 需要说明的是,芝麻信用并非互联网平台提供的唯一征信服务,腾讯以微信这一超级APP为依托也开展了类似业务。公开资料显示,腾讯于2018年1月开始内测“腾讯信用分”,而后在当年11月正式上线“微信支付分”并以此为基础推出了信用免押服务。但考虑到本文的样本期间为2018年1月至3月,腾讯的大数据征信服务尚处于起步阶段,我们倾向于认为这不会对本文结果构成实质挑战。

行征信中心数据库共收录自然人 8.7 亿,其中 3.7 亿有信贷记录;支付宝 2017 年披露的全民账单则显示,芝麻信用自创立以来已累计为 4150 万用户减免超过 400 亿的押金。

本文主要考察芝麻信用管理有限公司提供的“芝麻信用分”这一个人信用分数。该分数由企业运用云计算、深度机器学习等技术,通过逻辑回归、模型提升决策树和随机森林等算法,从信用历史、行为偏好、履约能力、身份特质和人脉关系等 5 个维度对用户进行信用评估,最终以 350-950 之间的具体分值形式进行呈现^①。其主要数据来源为支付宝 APP 所留存的海量用户数字足迹。支付宝作为兼具转账支付、消费金融、电商购物和日常缴费等功能的超级 APP,不仅可以收集用户借款、还款等传统金融信贷行为的相关信息,更能够掌握消费者在网购、外卖和出行等方面的日常消费决策。芝麻信用分作为支付宝平台主推的个人征信产品,在个人小额信贷、汽车和房屋租赁等各类情境下均有广泛应用。

作为一项大数据征信服务产品,芝麻信用分具备以下两项重要特征。第一,芝麻信用分的评估方式和具体算法不公开,个体无法通过单方面行为迅速提升自身分数。虽然芝麻信用对外披露了有助于提升分数的若干行为,如按时还清信用卡账单等,但并未给出行为与分数间的精确对应关系,个体无法通过“刷分”对芝麻信用分进行精确操纵(precise manipulation)。这也是本文以芝麻信用分作为断点回归驱动变量的前提。第二,芝麻信用分在个体、时间两个维度均有差异,即不同个体的分值存在横截面上的差异,同一个体的分值也会随时间变化而有所不同。这为后续的经验分析提供了更为丰富的考察维度:我们不仅可以比较不同个体在横截面上的差异,还可以考察同一个体在芝麻信用分跨越断点前后的不同表现。

(二)互联网平台流量

关于流量价值的理论雏形,最早可以追溯到 Metcalfe and Boggs (1976)在探讨以太网价值时提出的“梅特卡夫定律”,即网络的价值与其中节点数的平方成正比。伴随着“互联网+”的快速推进和一大批互联网企业崛起,互联网企业的价值评估成为资产定价领域的重要问题。“流量”本身虽然不是有形的、可直接买卖的实物资产,但可以通过多种形式间接创造价值,也即互联网行业所谓的“流量变现”。具体而言有以下几种渠道和方式:其一是掌握流量的平台向用户进行信息投放,扮演宣传媒介、广告中介的角色;其二是撮合用户交易并参与分成,发挥交易中介的功能;其三是挖掘用户在本平台的数据沉淀,开发征信等衍生产品和服务。现有研究虽然在具体估值

^① 参见芝麻信用官方网站:www.zmxy.com.cn。

模型上尚未达成共识,但基本指向是一致的,即评估互联网企业的价值时需要充分考虑平台流量、用户基数以及由此而来的长期价值(Gupta *et al.*, 2004; Ho *et al.*, 2011)。

用户流量的独特价值对互联网平台企业的决策产生了巨大影响。诸多互联网“巨头”在进行战略投资、业务布局时,在传统的成本收益分析和经济模型核算之外,会额外关注用户流量的价值。以近年来蓬勃发展的共享单车行业为例,虽然其头部企业迄今为止尚未实现营收平衡,持续处于亏损状态,但这并不妨碍阿里巴巴、腾讯、美团和滴滴出行等互联网平台企业以投资、自营或并购等方式参与其中^①。这些企业的核心目的之一即在于依托共享单车用车的高频特性来为自身平台导流。

因此,在中国情境下讨论个人征信服务,不能简单孤立地分析这个行业,而是要充分把握平台企业的行为逻辑,特别是其对于用户流量的关注。这是由中国个人征信行业独特的业务模式、产业结构和发展路径决定的。以本文的主要研究对象——芝麻信用——为例,其作为提供个人征信服务的最大民营市场主体,与发达国家专营化征信机构(如美国的FICO等)存在显著差异:芝麻信用并非以盈利为目的的独立主体,而是隶属于蚂蚁金服的子公司。作为大型金融科技集团的子板块,芝麻信用的产品思维、推广策略和经营导向都与专营化征信企业不同。例如,芝麻信用并没有独立的APP,而是以生活号的形式嵌入在支付宝APP当中。因此,在研究芝麻信用代表的征信行业发展时,要充分考虑其作为平台企业的子板块对母体平台的作用。

(三)共享单车与免押骑行

共享单车诞生于2015年。经过2016–2017年间的迅猛扩张和高速增长,2018年以后共享单车行业逐渐进入调整期。初代车辆进入报废周期、线下运维效率低下和用户流量变现遇阻等难题逐渐凸显,前期过快增长积累的历史问题浮出水面。但毋庸置疑,共享单车为“最后一公里”出行需求提供了解决方案,创造了一种高频次、低额度的消费场景,对于城市交通、房产价格和空气质量等产生了广泛的影响(Zhang and Mi, 2018; Cao *et al.*, 2021; Chu *et al.*, 2021)。

在共享单车行业发展之初,为有效约束用户的用车行为,各共享单车平台普遍采取收取押金的方法。用户需首先向共享单车平台缴纳押金并完成实名认证,而后方可用车;当用户决定终止使用某品牌的共享单车时,可以申请退还押金。由于共享单车的高频特性,用户通常不会频繁缴纳和退还押金,而是在初次缴纳押金后保持账户

^① 例如,阿里巴巴和蚂蚁金服先后投资了永安行、ofo小黄车和哈啰单车,并促成永安行与哈啰单车合并;腾讯投资了摩拜单车;滴滴出行投资了ofo小黄车,自营青桔单车,并以托管方式接收了小蓝单车;美团创始人王兴以个人身份参与对摩拜单车的投资,而后美团全资收购了摩拜单车并将其更名为美团单车。

处于“已支付押金”的状态,以便于日常骑行。

为降低共享单车的使用门槛、推广芝麻信用分的使用场景,芝麻信用与 ofo 小黄车、哈啰单车等多家共享单车企业联合推出了“芝麻信用免押骑行”,即芝麻信用分达到 650 分的用户可以免押金使用共享单车。为本文提供数据的共享单车企业与蚂蚁金服达成协议,从 2017 年 3 月开始在上海、广州和深圳等多个城市分批次开通芝麻信用免押骑行服务。具体而言,免押骑行包括免押认证和免押用车两个步骤:首先,共享单车用户需经由支付宝端口,授权共享单车企业查询其芝麻信用分,确认其符合信用免押资格(也即芝麻信用分在 650 分及以上);其次,共享单车用户通过上述认证后,可以通过各类端口免押金使用共享单车,包括共享单车 APP(根据手机类型又可分为苹果手机的 iOS 客户端和其他各类手机的安卓客户端)、微信小程序、支付宝生活号和地图 APP 内嵌接口等。

据芝麻信用官方网站介绍,芝麻信用分的主要应用场景之一即为“信用免押”,包括房屋、服装、汽车、相机和充电宝等多种物品的租赁,免押骑行仅为其中之一^①。提供免押场景的共享经济或分时租赁企业(如共享单车)可以获取由芝麻信用带来的用户流量,消费者则可以免去支付押金的烦冗步骤并对相关资金进行合理配置。在上述“双赢”表象之外,有一个问题额外值得关注:作为个人征信服务供应商的芝麻信用,在此过程中的收益是什么?结合本部分第(二)节所提到的互联网平台企业对流量的关注,我们提出:

假说:芝麻信用推出免押骑行政策将会增加其母体平台(支付宝)的用户流量。

三 数据描述与识别策略

(一)数据描述与样本选择

我们从某共享单车企业处获取了 2018 年 1 月 6 日至 3 月 5 日间 25 个城市全部用户脱敏数据的临时使用权。样本覆盖了在上述时间内至少使用过一次共享单车、且已向共享单车平台授权查询其芝麻信用分的用户。我们能够使用的信息包括上述用

^① 其他信用免押服务与免押骑行存在以下几方面差异:第一,从业务规模来说,共享单车市场应该是各类信用免押服务中用户量最大、订单量最多的场景之一,汽车租赁等服务的业务体量与之存在数量级上的差异。第二,从推广时间来说,免押骑行是芝麻信用与共享经济结合的先行领域,本研究的样本期间早于其他共享经济企业开展信用免押的时间。第三,从规则设置上来说,各类信用免押服务的门槛存在较大差异,从 550 分至 900 分不等。因此,其他信用免押服务不会对本文的核心结论构成实质性挑战。

户的部分个体特征(年龄、性别、是否为学生、注册时间和所在城市)以及在该时段内的全部用车记录。

由于在本文样本期间内芝麻信用分仅在每月6日进行调整,月内维持不变,我们将数据在个体-月度层面加总处理,计算出每人单月当中在支付宝、微信、iOS、安卓、地图和其他端口等6类端口用车行为的占比^①。最终我们得到7343 490名用户的9438 321条个人-月度层面的观测值,相关变量的描述性统计如表1所示。

表1

描述性统计

	样本数	均值	标准差	最小值	最大值
Panel A 个体层面变量					
年龄(岁)	7343 490	30.537	8.230	18	60
性别(男性为1,女性为0)	7343 490	0.561	0.496	0	1
是否为学生(是为1,否为0)	7343 490	0.004	0.066	0	1
累积注册天数(天)	7343 490	270.595	113.031	9	909
Panel B 个体-月度层面变量					
免押骑行(是为1,否为0)	9438 321	0.858	0.349	0	1
芝麻信用分	9438 321	701.957	51.639	350	946
支付宝端口用车占比(%)	9438 321	78.434	39.667	0	100
微信端口用车占比(%)	9438 321	3.067	16.522	0	100
iOS端口用车占比(%)	9438 321	7.070	24.583	0	100
安卓端口用车占比(%)	9438 321	11.119	30.261	0	100
地图端口用车占比(%)	9438 321	0.206	2.515	0	100
其他端口用车占比(%)	9438 321	0.080	4.357	0	100
违规用车占比(%)	9438 321	14.415	27.122	0	100

(二)识别策略

由于某品牌共享单车用户能否免押骑行唯一取决于其芝麻信用分的分值(即是否大于等于650分),我们采用清晰断点回归(Sharp RD)的识别策略来评估免押骑行政策对消费者用车流量分布的影响。具体而言,我们估计如下方程:

$$Y_{im} = \alpha + \beta D_{im} + f(Z_{im}) + X_i + \lambda_c + \delta_m + \varepsilon_{im} \quad (1)$$

其中 Y_{im} 为*i*个体在*m*月份通过支付宝端口用车次数占总骑行次数的百分比; Z_{im} 为*i*用户在*m*月的芝麻信用分; D_{im} 为表示免押骑行状态的虚拟变量,当 $Z_{im} \geq 650$ 时该变量

^① 同理,由于芝麻信用分调整日期为每月的6日,此处所说的“月度”指2018年1月6日至2月5日和2018年2月6日至3月5日,与自然月的起止时间并不完全一致。后文凡涉及月份和月度,除另行说明外,含义与此处相同。

取1,否则取0; $f(\cdot)$ 为控制芝麻信用分及其与免押骑行虚拟变量交互项的一个函数。在基准模型设定中,我们采用局部线性回归(local linear regression)方法,选择矩形核函数(rectangular kernel function),依据Imbens and Kalyanaraman(2012)方法确定最优带宽为15(1K带宽)。在后续分析中,我们也使用了三角核函数(triangular kernel function)并改变带宽等模型设定,对基准模型进行了一系列的稳健性检验。 X_i 为表1的Panel A中所列示的个人特征变量; λ_c 代表城市固定效应, δ_m 代表月份固定效应; α 为截距项; ε_{im} 为误差项。 β 为我们关心的核心参数,即免押骑行对消费者用车端口选择的影响。

(三)断点回归有效性

断点回归有效的首要前提是驱动变量不能被精确操纵(Lee and Lemieux, 2010; Cattaneo *et al.*, 2019)。如果样本个体可以精确控制驱动变量、决定自身处置状态(treatment status),则会导致断点两侧的样本存在系统性差异,进而影响断点回归的有效性。需要注意的是,断点回归关于“驱动变量不能被精确操纵”的假设并不等价于“个体无法影响驱动变量”:断点回归允许个体通过自身行为对驱动变量施加影响;只要无法实现“精确操纵”,断点回归这一方法就依然适用。

在既有文献中,关于该条件成立与否的讨论主要从两个方面展开。一是从制度背景出发,讨论微观个体能够在何种程度上影响驱动变量、是否有可能实现精确操纵,提供最底层、最基础的事实支撑。二是对不受处置变量影响的协变量(covariate)进行平衡检验(balance test),确认其在断点附近的连续性。接下来我们将结合本文的研究主题,对以上两方面分别进行讨论。

1. 制度背景。本文第二部分第(一)节已经介绍了芝麻信用分的技术原理和基本特点。其核心特征之一在于,由于信用分的评估方式和具体算法不公开,个体无法对分数进行精确操纵。我们可以从以下三个方面来理解芝麻信用分的不可完全操纵性:第一,从个人征信服务的本质出发,该特征是征信服务有效可信的前提。以共享单车为代表的各类市场主体之所以愿意采信芝麻信用分,正是因为该分数能够较为客观公正地刻画个体信用水平。一旦个体能够精确决定自身信用分的具体分值,信用分就会完全失去参考价值,芝麻信用作为征信服务供应商的意义也就不复存在。第二,从技术基础的角度出发,芝麻信用分是基于深度学习、强化学习等复杂算法计算得出的一个分值。这意味着个体行为与信用分数之间并非一个确定性函数(deterministic function)关系。机器学习算法的“黑箱”特质决定了个体对自身信用分的精确操纵在技术上不具有可行性。第三,通过对征信行业从业者进行访谈,结合在

各大互联网平台(如百度、微博和知乎等)的搜索结果,我们未发现任何证据支持“个体可精确操纵芝麻信用分”。虽然有诸多关于提升芝麻信用分的经验法则(如完善个人信息认证、提升在天猫或淘宝的网购频次等),但这些做法均无法确保实现对信用分的精确控制。综上所述,从制度背景出发,芝麻信用分的技术原理和基本特征保证了芝麻信用分不能被精确操纵。

2. 协变量平衡检验。此项检验的底层逻辑是:假如微观个体无法精确控制自身驱动变量的取值,那么其驱动变量究竟落在断点哪一侧就具有一定的局部随机性;而这种局部随机性理应保证可观测协变量在断点附近的连续性。具体而言,我们对630-670分之间的样本按照芝麻信用分进行分组(分数相同的样本归入一组),并在图1的Panel A-D中分别报告了年龄、性别、累计注册天数以及是否为学生用户4个协变量的分组均值及相应的95%置信区间。从图中我们可以看出,断点两侧的用户在个人特征方面高度相似,并未呈现出显著差异。上述证据进一步夯实了我们对本文识别策略有效性的信心。

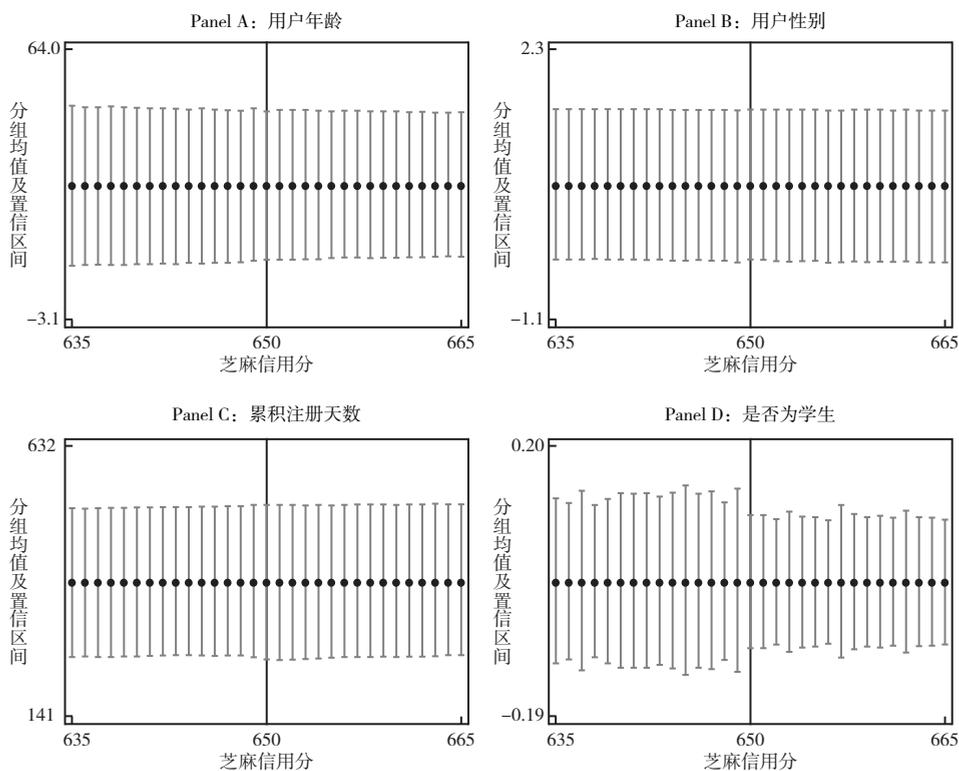


图1 650分断点两侧用户个人特征的平衡检验

3. 关于断点回归有效性的进一步讨论。在基于制度背景的讨论和平衡检验之外,既有文献中还有一种常见的关于断点回归有效性的补充检验,即对驱动变量的核密度分布进行检验、观察其在断点附近有无跳跃。这一检验的内在逻辑是:假如个体能够精确操纵其驱动变量取值、使之高于或低于特定门槛,驱动变量的分布函数就可能在断点附近出现跳跃。但这一检验方法之所以被称为一种“补充性”手段(Lee and Lemieux, 2010),是因为驱动变量分布在断点处的连续性,既不是保证断点回归有效的充分条件,也不是必要条件。一方面,就“非充分性”而言,即便驱动变量的分布函数在断点处平滑、连续,也不能排除微观个体精确操纵驱动变量的可能性(Lee and Lemieux, 2010)。另一方面,就“非必要性”而言,即便驱动变量的分布函数在断点处出现跳跃,也不一定源于个体对驱动变量的精确操纵(Imbens and Lemieux, 2008)。例如,相关政策可能对符合标准的个体具有吸引力,导致断点一侧的样本被观测到的概率增加。此时观察到的驱动变量分布函数的跳跃并非源于个体的精准操纵,而是因为政策的激励效应所导致的样本量偏移。

以信用分数作为断点回归驱动变量的研究多对应后一种情况。例如,美国金融界存在诸多基于FICO信用分的规则:在进行贷款资质审核时,对于信用分在特定分值以上的人,其审查会更加宽松,或者不再进行额外审查,直接授信;对于信用分低于某限值的人,其申请可能被直接拒绝。上述规则决定了断点两侧的样本数量很容易不均衡,反映为FICO信用分的分布在断点处出现跳跃。但这并不意味着个体能够精确操纵其FICO分数,也不妨碍研究者利用断点回归开展相关研究^①。

在本文的情境下,由于免押骑行政策的出现,芝麻信用分达到650分及以上的用户应当更倾向于授权共享单车平台查询其信用分数。从数据上来看,在后文回归分析的基准带宽范围内(也即635-665分),[635, 650]区间内样本数为356 649,[650, 665]区间内样本数为823 317,二者占比分别为30.225%和69.775%。与前述利用FICO信用分开展的研究类似,该现象不会对本文中“芝麻信用分数不能被精确操纵”这一核心假设构成实质性挑战。在后续回归分析中,我们还会采用空心断点回归(Donut-Hole RD)等方法进行稳健性检验。

^① 比如,Keys *et al.* (2010)利用“金融机构一般不向FICO信用分低于620分的人授信”这一经验规则,探究了授信审查、资产证券化与2008年金融危机之间的关系。此外,同样的分析方法还被用于研究优质贷款与次级贷款的差异性表现(Keys *et al.*, 2012)、灵活抵押贷款的优缺点(Garmaise, 2012)和搜寻成本对金融市场的影响(Argyle *et al.*, 2020)等一系列重要问题。

四 基本结果

(一)断点回归的图形证据

我们以IK带宽下、采用矩形核函数、不添加任何协变量的局部线性回归模型作为基准模型。在该设定下,断点回归图形如图2所示。

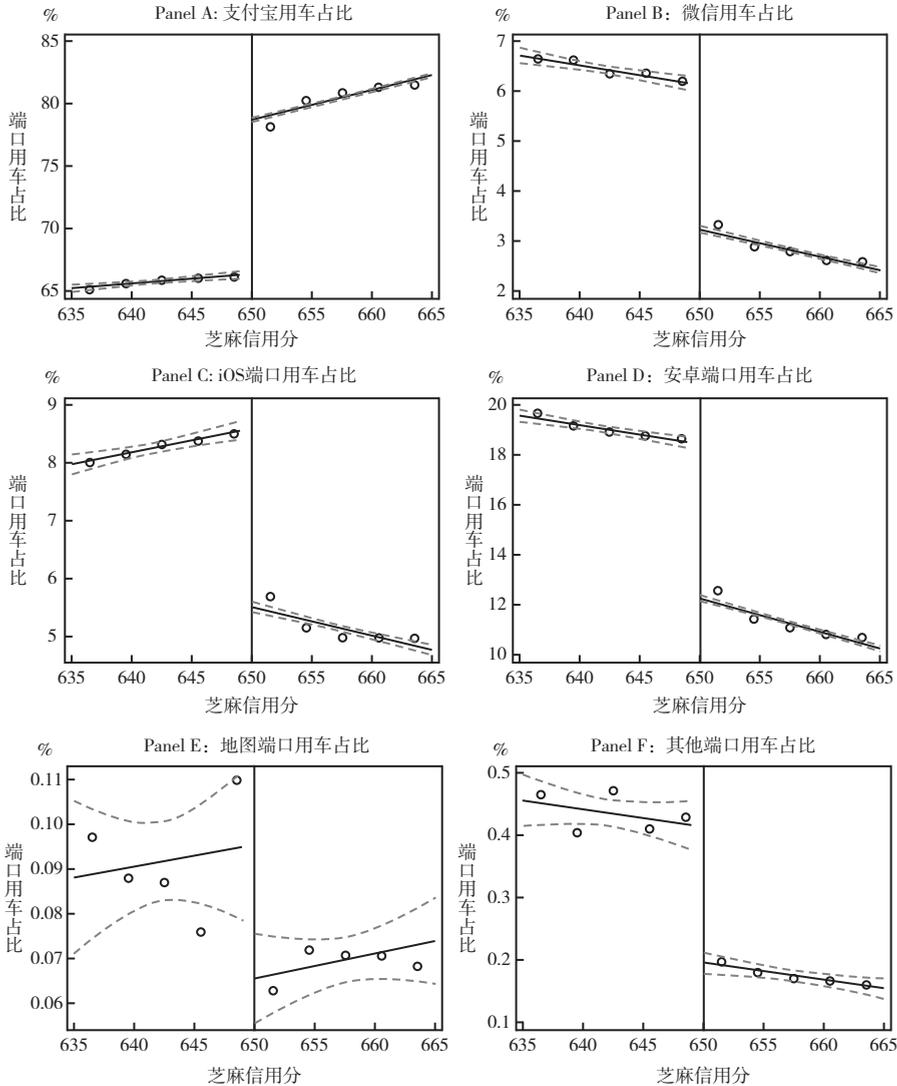


图2 650分断点两侧的各端口用车流量占比

说明:每个空心圆圈表示的是每个5分箱体(bin)内部的因变量均值。实线所示为IK最优带宽下的局部线性拟合,虚线表示的是95%显著性水平下的置信区间。下文图5、6当中的图形含义与此相同。

Panel A 表明,免押骑行使得用户在支付宝端口用车的流量占比提升了约 12 个百分点。这意味着通过实施芝麻信用免押骑行,共享单车用户在选择用车端口时,对于支付宝端口的依赖程度大幅提升,信用免押确实对母体平台产生了明显的导流作用。

一个随之而来的问题是,用户对于支付宝端口的黏性上升,是同时挤压了其他各类用车端口的流量占比还是只对个别端口产生了替代作用?为回答这一问题,我们分别计算了微信、iOS、安卓、地图和其他端口用车的流量占比并将其作为因变量,在基准模型下进行断点回归,相应图形结果如图 2 的 Panel B-F 所示。从中可以看出,其他 5 类端口用车的流量占比均出现了显著下降;特别是长期以来作为支付宝主要竞争对手的微信,其占比几乎降低了一半。

(二)断点回归的估计结果

表 2 列(1)为基准模型下的回归结果,表明信用免押将支付宝端口用车占比提升了约 12.3 个百分点。列(2)-(6)对列(1)的基准回归结果进行了一系列的稳健性检验。其中,列(2)将矩形核函数替换为三角核函数,从而在回归中赋予离断点更近的样本以更高的权重,这一模型设定下的 IK 带宽为 19;列(3)使用 Calonico *et al.*(2014)

表 2 断点回归基本估计结果

因变量	(1)	(2)	(3)	(4)	(5)	(6)
	支付宝端口用车占比					
免押骑行	12.344 (0.180)*** [0.766]*** {0.417}***	12.178 (0.178)*** [0.773]*** {0.410}***	11.592 (0.307)*** [0.206]*** {1.921}***	12.424 (0.178)*** [0.742]*** {0.415}***	11.898 (0.176)*** [0.759]*** {0.435}***	11.961 (0.175)*** [0.732]*** {0.431}***
带宽选择标准	IK	IK	IK	IK	IK	IK
带宽值	15	19	15	15	15	15
核函数	矩形	三角形	矩形	矩形	矩形	矩形
城市固定效应	未控制	未控制	未控制	控制	未控制	控制
月份固定效应	未控制	未控制	未控制	控制	未控制	控制
个人特征	未控制	未控制	未控制	未控制	控制	控制
样本观测值	1179 966	1413 805	1179 966	1179 966	1179 966	1179 966

说明:圆括号、方括号和大括号中分别报告的是异方差稳健标准误、聚类到芝麻信用分的稳健标准误和聚类到城市的稳健标准误。*、**、***分别表示 10%、5%、1% 的显著性水平。后表同。

建议的局部二阶多项式(local quadratic polynomial)方法进行非参数估计^①;列(4)加入了城市和月份固定效应;列(5)控制了用户的年龄、性别、是否校园用户和累计注册天数等个人特征;列(6)则同时控制了城市固定效应、月份固定效应和个体特征。各列系数估计值的大小稳定在12个百分点左右,并且在使用不同类型标准误进行统计推断时保持高度显著。这一结果初步表明,芝麻信用免押骑行政策极大地提升了支付宝平台在共享单车这一具体消费情境下的流量占有率。

(三)结果解读与政策启示

在解读结果和讨论政策启示之前,我们需要明确几项基本事实。第一,各类用车端口对用户而言是无差异的,均能以极为相似的产品形态满足用户的用车需求;各端口的收费规则和优惠政策也完全相同。第二,用户在完成免押骑行的认证后,依然可以通过所有端口使用共享单车,并非仅在支付宝端口用车才可以享受免押政策。第三,单就信用免押政策本身而言,确实可以起到为消费者免除押金、为共享单车企业导流的积极影响;接受相关政策系出于消费者和单车企业的自主选择,而非由支付宝强加。在上述事实背景下,我们尝试从以下两个角度解读本文的结果。

首先是关于互联网平台企业攫取用户流量的隐蔽性。本文所发现的流量转移,并非由于支付宝平台或共享单车企业的强制锁定,而纯粹源于消费者的自发行为;信用免押政策亦未采取欺骗性策略或呈现出明显的市场垄断意图。这正是本文的核心关注点之一:基于大数据征信的信用免押政策看似不违背公平竞争原则,实则极大地影响用户流量分布和市场竞争格局。

其次是关于互联网平台企业收益的多样性。此处我们仅观察到用户流量的转移,支付宝并未从中获取直接的经济收益。但需要强调的是,即便支付宝不向消费者或企业收取芝麻信用分查询费,信用免押依然为其带来巨大的隐性收益。一方面,由于用户对支付宝的黏性提升、打开频次增加,调用支付宝其他服务(如生活缴费、信用卡还款等)的概率可能会有所提升。虽然由于数据限制我们无法直接观测到用户的上述行为,但互联网平台对于“流量”的争夺以及我们自身作为用户的经验可以佐证这一论断。另一方面,由于用户在支付宝APP的停留时间和操作行为的增加,芝麻信用得以收集更多数字足迹以提升芝麻信用分的精确度和征信服务的总体水平,从而实现“征信→流量→数据→征信”的正反馈。

^① Gelman and Imbens(2019)指出,断点左右两侧拟合函数的多项式阶数最好不高于二阶。

在上述两方面解读之外,还有三点问题需要说明。第一,支付宝作为超级APP,搭载了诸多服务接口,除共享单车之外还包括支付、外卖、理财和缴费等。由于数据可得性的限制,本文所讨论的“流量”概念,是仅就共享单车这一市场而言的,并不涉及其他场景下的用户行为。第二,就消费者行为而言,本文的回归结果可能存在多种解释。例如,消费者可能不知晓免押后其他端口也可以用车;或者消费者注意力有限、只愿意使用少数几个APP等等。我们不掌握用户除共享单车用车行为以外的其他信息,也无法对用户进行实验或调研以了解他们的心理活动。但消费者行为层面的具体机制如何,不会影响本文研究结果的核心意涵。第三,本节所讨论的政策启示、特别是在反垄断规制方面的建议,是对既有规制框架的补充而非替代。本文意在强调的是,在对互联网企业进行反垄断规制时要充分把握其行业特点和企业特质,及时调整和升级规制框架;但这并不意味着忽略垄断协议等传统垄断行为,而是要实现对各种形式垄断行为的协同治理。

(四)稳健性检验

1. 空心断点回归。在本文第三部分第(三)节中,我们从个人征信服务的本质、芝麻信用的技术基础、来自访谈和互联网搜索的支持性证据等角度出发,论证了芝麻信用分不会被个体精确操纵;又通过平衡检验,验证了断点两侧样本的可比性。以上讨论基本可以保证本文断点回归的有效性,但部分读者可能对此仍有疑虑。为此,我们进一步借鉴 Barreca *et al.* (2011)的思路,利用空心断点回归进行稳健性检验。该回归设计的逻辑是:假如个体确实能够操纵驱动变量的取值,并且该操纵行为需要付出一定的成本,那么断点附近区域内驱动变量被操纵的概率更大;因此,可以去除断点邻域内的样本、进行子样本回归,从而进一步缓解关于驱动变量被操纵的担忧。

既有文献关于空心断点回归中邻域的“半径”大小并无定论。为充分检验经验结果的稳健性,我们从1开始,逐步扩大空心的半径至5,相应的经验结果汇报在表3中。其中,列(1)对应的空心半径为1,也即要删除芝麻信用分为649和650的样本;列(2)对应的空心半径为2,也即要删除芝麻信用分在[648, 651]区间内的样本……其余各列以此类推。随着空心半径的扩大,虽然样本量逐渐减少,但估计结果始终保持高度显著,系数大小也保持稳定。这进一步确认了断点回归的有效性。

2. 针对不同带宽的稳健性检验。本文基准回归报告的是IK带宽下的估计结果。为了在更广泛的意义上验证估计结果的稳健性,我们采用表2列(6)所对应的模型设

表3 空心断点回归估计结果

因变量	(1)	(2)	(3)	(4)	(5)
	支付宝端口用车占比				
免押骑行	12.188*** (0.202)	12.621*** (0.233)	12.964*** (0.272)	12.875*** (0.322)	13.220*** (0.293)
空心半径	1	2	3	4	5
样本观测值	1106 669	1034 473	960 783	886 466	861 788

说明:表中各列的模型设定均与表2列(6)相同。后表同。

定,分别在 $\pm 10, \pm 11, \dots, \pm 20$ 的带宽内进行估计。不同带宽下的回归系数与相应的95%置信区间如图3所示^①。从图中可知,当带宽从10逐渐增加到20时,估计系数的大小始终保持在11.5-12.5之间且高度显著。表2当中的回归结果在不同带宽选择下保持稳健。

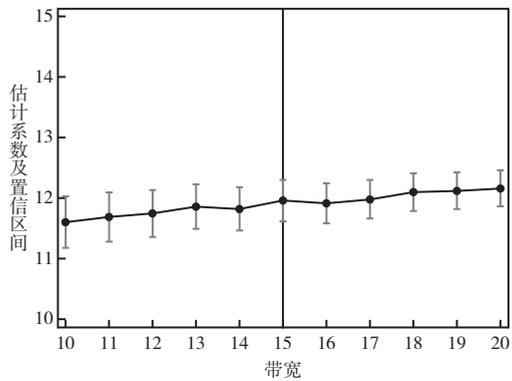


图3 不同带宽下的稳健性检验

3. 伪造断点检验。通过支付宝APP“信用生活”板块可以查询,除以650分为门槛的免押骑行外,还有以600分为门槛的“先用后买”、以700分为门槛的“信用买车”等诸多基于芝麻信用分的商业合作。基准结果由此面临的一项潜在挑战在于,给定散列于各分值的众多免押场景,我们观察到的共享单车免押骑行对支付宝的导流是否仅仅是由650分这一断点所驱动的?借鉴Chay *et al.* (2005)和Sun and Zhao (2016)的做法,我们针对635-665之间的所有芝麻信用分数进行一系列伪造断点检验,从而反向验证本文主要发现的可靠性。

我们通过以下步骤进行伪造断点检验。首先,针对635-665之间的每一芝麻信用分值 S ,选定 $S \pm 15$ 区间内的样本作为回归样本。其次,以 S 为断点, $Z_{im} \geq S$ 的样本 D_{im} 取1,否则取0。最后,利用以 S 为断点定义的 D_{im} 、采用表2列(6)的模型设定进行回归分析,记录估计系数 β 的 t 检验统计量的绝对值。伪造断点检验的 t 统计量绝对值如图4所示。从中可知,使用650以外的任意芝麻信用分作为断点,所得系数的 t 检验统

① 使用断点回归的文献在带宽选择上的另一常见标准为CCT方法(Calonic *et al.*, 2014)。我们根据CCT方法计算所得的最优带宽为18,包含在图3稳健性检验的带宽范围当中。

计量绝对值都小于使用真实断点所得到的数值。这意味着基于真实数据生成过程的模型设定比其他模型设定对数据的拟合效果更好,从而反向验证了本文回归结果的可靠性。

4. 使用样本期内芝麻信用分发生变化的子样本。尽管我们在图1中已经对若干消费者特征在断点两侧的平衡性进行了检验,但我们无法穷尽所有影响APP使用的个体特征。因此,仍可能存在一些不可观测的遗漏变量影响本文结果,例如不同消费者对支付宝、微信等APP偏好的异质性。如本文第二部分第(一)节所述,芝麻信用分在个体、时间两个维度均有差异,同一用户的分值会随时间变化而有所不同。受这一特征启发,我们遴选如下的用户数据构成子样本:用户芝麻信用分在本研究样本期内发生变化,且变化前后的分数分别位于650分两

侧。基于该子样本进行回归分析时,我们比较的是同一消费者在信用免押前后的行为变化,从而在最大程度上保证了消费者偏好等无法观测的特征的稳定性。虽然样本量大幅减少,但图5的图形证据与图2和表2中的基准结果依然类似,消费者在信用免押之后确实提升了在支付宝端口的用车占比,尽管其系数的绝对值大幅下降(略高于2个百分点)。

五 进一步讨论

本文以上部分主要考察了免押骑行对于芝麻信用母体平台——支付宝——的导流作用。接下来,我们将对回归结果开展进一步讨论,探索流量变动的来源、导流作用在不同特征用户间的异质性以及信用免押的潜在负面效应。

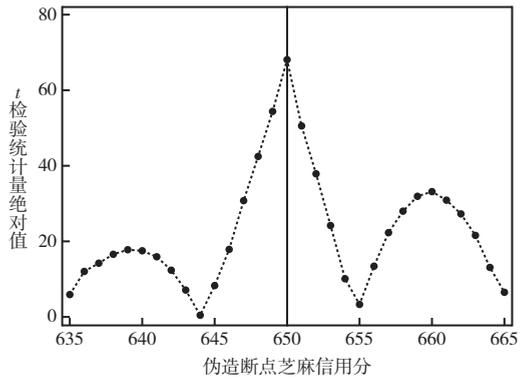


图4 伪造断点回归 t 检验统计量绝对值

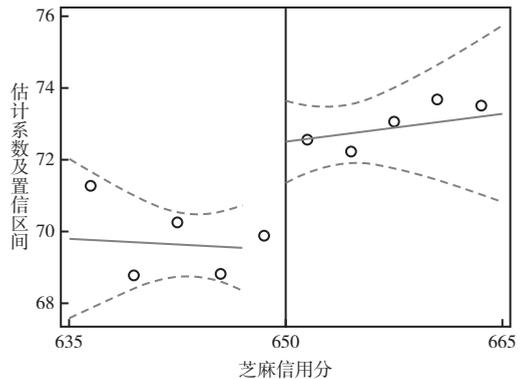


图5 样本期内芝麻信用分跨越650分的子样本

(一)流量变动的来源

在前文分析中,我们倾向于认为免押骑行提升了用户在支付宝端口的用车频次、降低了用户在其他端口的用车次数,也就意味着基准结果主要由存量流量的再分配所驱动。与此同时也存在另一种可能,即免押骑行增加了用户在支付宝端口的用车次数,但并未影响用户其他端口的用车行为,此时基准结果主要由增量流量的非均衡分布所驱动。为排除后一种可能性,我们在表4中报告了以不同端口的订单量水平值作为因变量的估计结果。从中可知,免押骑行提高了用户通过支付宝端口的用车总量,而其他各类与支付宝存在竞争关系的端口的订单量均有不同程度的下滑。这一发现表明,免押骑行对平台流量所产生的影响来源于存量流量的再分配、而非增量流量的非均衡分布^①。

表4 免押骑行对不同互联网平台端口订单总量的影响

因变量 端口类别	(1)	(2)	(3)	(4)	(5)	(6)
	各平台端口用车订单量					
	支付宝	微信	iOS	安卓	地图	其他
免押骑行	0.184*** (0.046)	-0.133*** (0.017)	-0.286*** (0.036)	-0.620*** (0.079)	-0.001* (0.001)	-0.006*** (0.002)
样本观测值	1179 966	1179 966	1179 966	1179 966	1179 966	1179 966

(二)异质性分析

接下来我们将探索免押骑行的导流作用在不同子样本群体间可能存在的异质性,具体而言包括用户性别、年龄和所在城市规模特征(是否一线城市)等三个维度。表5列(1)和列(2)的回归结果显示,免押骑行对不同性别群体的影响差异不大。列(3)和列(4)的回归结果显示,45岁以下的人群更容易受到免押骑行政策的影响。考虑到芝麻信用的普及、渗透程度在年轻群体中更高,这一结果也比较符合预期。最后,列(5)和列(6)的回归结果表明,免押骑行政策在样本中的三个一线城市(上海、广州和深圳)对支付宝平台的导流作用更为明显^②。

① 应提供数据的共享单车企业要求,为保护企业商业机密,我们不能报告表4因变量的描述性统计量。

② 关于城市规模的异质性分析还有另一重含义。经过查询公开资料我们发现,在上海、深圳等一线城市,还存在若干区域性的信用免押政策。例如,上海图书馆曾尝试面向常住人口推广免押金信用阅读。因此,表5列(6)基于非一线城市样本的回归结果同时表明,本文的主要分析结果并非由上海等超大城市的区域性信用免押政策所驱动。

表5 芝麻信用免押的异质性影响

因变量	(1)	(2)	(3)	(4)	(5)	(6)
	男性	女性	45岁以下	45岁以上	一线城市	非一线城市
免押骑行	11.884*** (0.221)	12.078*** (0.286)	12.071*** (0.184)	10.507*** (0.580)	14.108*** (0.271)	10.349*** (0.229)
样本观测值	724 012	455 954	1070 638	109 328	524 432	655 534

(三)信用免押的潜在负面效应

在免押骑行的情境下,另一个问题同样值得关注,即信用免押对消费者行为的约束作用。共享单车平台收取押金的初衷是通过押金形成威慑,以约束用户破坏单车、违规停放和不规范用车等行为。实行信用免押之后,“罚没押金”的威慑被“下调信用分数”的威胁所替代。一个随之而来的问题在于,用户对自身信用记录的关注与关注,是否能够有效约束用户行为?

由于共享单车平台无法监控用户用车的全过程,难以精确地识别所有非规范用车行为。但平台通过跟踪车辆后续使用情况、发掘用户历史行为等信息,利用机器学习、人工智能等技术,可以对非规范用车行为进行预测。据提供数据的共享单车企业内部评估,该判断的准确率约为85-90%,已初步满足企业运营需求。具体而言,针对每一次骑行,该预测模型都会生成一个虚拟变量以表征用车状态,若预测为非规范用车则取1,否则取0。我们依然将数据在个体-月度层面加总,计算得出*i*个体*m*月份全部骑行当中的非规范用车行为占比,将其作为断点回归的因变量。图6显示免押状态下非规范用车行为占比会上升约0.7个百分点,这也意味着信用免押对于用户行为所形成的约束力小于收取押金所形成的威慑。这一发现说明,在社会信用体系建设刚起步、配套措施不完备的情况下,需要合理评估失信惩罚对于个体行为的威慑效力,尽快建立和完善“守信激励、失信惩戒”相关机制。

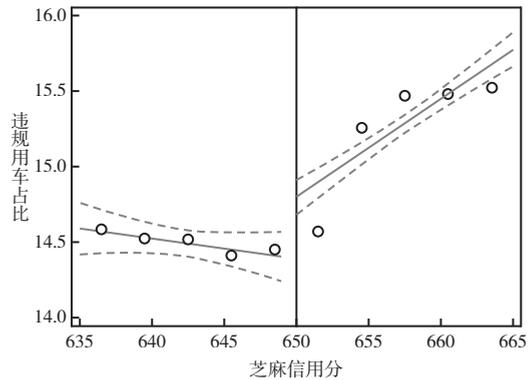


图6 免押骑行对违规用车行为的影响

六 结论

社会信用体系是社会主义市场经济体制重要组成部分,但其建设是一个牵一发而动全身的系统性工程。由于中国个人征信服务供应商多为互联网平台企业的内嵌板块,随着数据要素逐步向头部互联网平台企业集中,大数据征信很可能被母体平台用作导流工具,致使用户流量在不同平台间呈现非均衡分布、进而影响市场竞争格局。

本文以芝麻信用分达到650分即可免押金骑行共享单车这一消费情境为例,首次利用微观个体数据考察了大数据征信对平台企业用户流量的影响。我们基于断点回归的识别策略发现,免押骑行政策导致用户在支付宝端口使用共享单车的比例提高了约12个百分点,而其主要竞争对手平台的流量均有不同程度的下滑。进一步分析表明,上述发现主要源于存量流量的再分配而非增量流量的非均衡分布,不随用户性别而发生变化,但在年轻人群和一线城市更明显。此外,信用免押对于用户不良用车行为的约束力弱于收取押金所形成的威慑力。

本文的发现具有重要的政策含义。本文研究成果对今后一个时期的平台规制有启发意义。大数据征信可能通过克服传统市场情境下广泛存在的信息不对称,实现用户、平台主体和相关企业共同受益的格局。然而,互联网平台提供的大数据征信服务同样可能加剧用户流量在不同平台之间的非均衡分配。本文的研究结果表明,平台企业基于信用分实施的优惠政策可以提高平台用户流量,构建“征信→流量→数据→征信”的正反馈闭环。在当前互联网平台快速增长的背景下,规制机关也需更新监管思路,充分把握中国在数字经济方面独特的商业模式和发展路径,对平台企业利用大数据征信等技术手段影响公平竞争的潜在风险保持高度警惕。

本文也对今后一个时期中国社会信用体系建设具有借鉴意义。现阶段中国征信行业发展和信用体系建设同时依靠公共部门和市场主体的力量,二者的产品形态分别以中国人民银行征信中心提供的个人信用报告和蚂蚁金服旗下芝麻信用分为代表。长期以来,以芝麻信用分为代表的由企业提供的征信服务,由于应用场景广泛、用户体验良好而获得了社会各界的普遍关注;与此同时,部分研究机构和学者从居民隐私和数据产权的角度出发,对政府参与数据收集、开发和利用表达了隐忧(Beraja *et al.*, 2022)。本文的研究结果启发我们,考虑到互联网平台企业可能以大数据征信服务为抓手损害市场竞争的公平性,今后中国在推进社会信用体系建

设的过程中,仍要充分发挥公共部门的作用,统筹协调政府、企业和个人所扮演的角色。

参考文献:

- 国务院(2014):《社会信用体系建设规划纲要(2014—2020年)》。
- 潘功胜(2014):《中国需要建立发达的征信市场》,《人民日报》2014年10月13日第18版。
- 熊巧琴、汤珂(2021),《数据要素的界权、交易和定价研究进展》,《经济学动态》第2期。
- 徐翔、厉克奥博、田晓轩(2021),《数据生产要素研究进展》,《经济学动态》第4期。
- Argyle, B.; Nadauld, T. D. and Palmer, C. “Real Effects of Search Frictions in Consumer Credit Markets.” *NBER Working Paper No.26645*, 2020.
- Ball, I. “Scoring Strategic Agents.” *Working Paper available at arXiv:1909.01888*, 2019.
- Barreca, A. I.; Guldi, M.; Lindo, J. M. and Waddell, G. R. “Saving Babies? Revisiting the Effect of Very Low Birth Weight Classification.” *Quarterly Journal of Economics*, 2011, 126(4), pp. 2117–2123.
- Beraja, M.; Yang, D. Y. and Yuchtman, N. “Data-Intensive Innovation and the State: Evidence from AI Firms in China.” *Review of Economic Studies*, 2022, forthcoming.
- Cai, H.; Jin, G. Z.; Liu, C. and Zhou, L. A. “Seller Reputation: From Word-of-Mouth to Centralized Feedback.” *International Journal of Industrial Organization*, 2014, 34, pp. 51–65.
- Calonico, S.; Cattaneo, M. D. and Titiunik, R. “Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs.” *Econometrica*, 2014, 82(6), pp. 2295–2326.
- Cao, G.; Jin, G. Z.; Weng, X. and Zhou, L. A. “Market Expanding or Market Stealing? Competition with Network Effects in Bike-Sharing.” *RAND Journal of Economics*, 2021, 52(4), pp. 778–814.
- Cattaneo, M. D.; Idrobo, N. and Titiunik, R. *A Practical Introduction to Regression Discontinuity Designs: Foundations*. UK: Cambridge University Press, 2019.
- Chatterjee, S.; Corbae, D.; Dempsey, K. P. and Fios-Rull, J. V. “A Quantitative Theory of the Credit Score.” *NBER Working Paper No. w27671*, 2020.
- Chay, K. Y.; McEwan, P. J. and Urquiola, M. “The Central Role of Noise in Evaluating Interventions that Use Test Scores to Rank Schools.” *The American Economic Review*, 2005, 95(4), pp. 1237–1258.
- Chu, J.; Duan, Y.; Yang, X. and Wang, L. “The Last Mile Matters: Impact of Dockless Bike Sharing on Subway Housing Price Premium.” *Management Science*, 2021, 67(1), pp. 297–316.
- Condorelli, D. and Padilla, J. “Data-Driven Predatory Entry with Privacy-Policy Tying.” *Working Paper available at SSRN 3600725*, 2022.
- De Cornière, A. and Taylor, G. “Data and Competition: A General Framework with Applications to Mergers, Market Structure, and Privacy Policy.” *CEPR Working Paper No. DP14446*, 2020.
- Garmaise, M. J. “The Attractions and Perils of Flexible Mortgage Lending.” *Review of Financial Studies*, 2012, 26(10), pp. 2548–2582.
- Gelman, A., and Imbens, G. “Why High-Order Polynomials Should Not Be Used in Regression Discontinuity

- Designs.” *Journal of Business & Economic Statistics*, 2019, 37(3), pp. 447–456.
- Gupta, S.; Lehmann, D. R. and Stuart, J. A. “Valuing Customers.” *Journal of Marketing Research*, 2004, 41(1), pp. 7–18.
- Hagiü, A. and Wright, J. “Data-Enabled Learning, Network Effects and Competitive Advantage.” *RAND Journal of Economics*, 2022, forthcoming.
- Ho, C. T.; Liao, C. K. and Kim, H. T. “Valuing Internet Companies: A DEA-Based Multiple Valuation Approach.” *Journal of the Operational Research Society*, 2011, 62(12), pp. 2097–2106.
- Homonoff, T.; O’Brien, R. and Sussman, A. B. “Does Knowing Your FICO Score Change Financial Behavior? Evidence from a Field Experiment with Student Loan Borrowers.” *Review of Economics and Statistics*, 2021, 103(2), pp. 236–250.
- Imbens, G. and Kalyanaraman, K. “Optimal Bandwidth Choice for the Regression Discontinuity Estimator.” *Review of Economic Studies*, 2012, 79(3), pp. 933–959.
- Imbens, G. and Lemieux, T. “Regression Discontinuity Designs: A Guide to Practice.” *Journal of Econometrics*, 2008, 142(2), pp. 615–635.
- Keys, B. J.; Mukherjee, T.; Seru, A. and Vig, V. “Financial Regulation and Securitization: Evidence from Subprime Loans.” *Journal of Monetary Economics*, 2009, 56(5), pp. 700–720.
- Keys, B. J.; Mukherjee, T.; Seru, A. and Vig, V. “Did Securitization Lead to Lax Screening? Evidence from Subprime Loans.” *Quarterly Journal of Economics*, 2010, 125(1), pp. 307–362.
- Keys, B. J.; Seru, A. and Vig, V. “Lender Screening and the Role of Securitization: Evidence from Prime and Subprime Mortgage Markets.” *Review of Financial Studies*, 2012, 25(7), pp. 2071–2108.
- Lee, G. M.; He, S.; Lee, J. and Whinston, A. B. “Matching Mobile Applications for Cross Promotion.” *Information Systems Research*, 2020, 31(3), pp. 865–891.
- Lee, D. S. and Lemieux, T. “Regression Discontinuity Designs in Economics.” *Journal of Economic Literature*, 2010, 48(2), pp. 281–355.
- Li, Z. and Agarwal, A. “Platform Integration and Demand Spillovers in Complementary Markets: Evidence from Facebook’s Integration of Instagram.” *Management Science*, 2017, 63(10), pp. 3438–3458.
- Liberman, A. “The Value of a Good Credit Reputation: Evidence from Credit Card Renegotiations.” *Journal of Financial Economics*, 2016, 120(3), pp. 644–660.
- Metcalf, R. M. and Boggs, D. R. “Ethernet: Distributed Packet Switching for Local Computer Networks.” *Communications of the ACM*, 1976, 19(7), pp. 395–404.
- Prüfer, J. and Schottmüller, C. “Competing with Big Data.” *Journal of Industrial Economics*, 2021, 69(4), pp. 967–1008.
- Sun, A. and Zhao, Y. “Divorce, Abortion, and the Child Sex Ratio: The Impact of Divorce Reform in China.” *Journal of Development Economics*, 2016, 120, pp. 53–69.
- Tang, H. “The Value of Privacy: Evidence from Online Borrowers.” *Working Paper available at SSRN 3880119*, 2019.

Zhang, Y. and Mi, Z. "Environmental Benefits of Bike Sharing: A Big Data-Based Analysis." *Applied Energy*, 2018, 220, pp. 296–301.

Zheng, J.; Qi, Z.; Dou, Y. and Tan, Y. "How Mega Is the Mega? Exploring the Spillover Effects of WeChat Using Graphical Model." *Information Systems Research*, 2019, 30(4), pp. 1343–1362.

Big-Data Credit Investigation and Platform User Traffic: An Empirical Study of Deposit-Free Bike-Sharing

Cao Guangyu; Liu Chang; Zhou Li'an

Abstract: Personnel credit investigation services based on big-data technology play a key role in the construction of the social credit system. However, the current major providers of big-data credit investigation services in China are mostly integrated into internet platform companies rather than independent entities, and therefore their rapid growth has also raised concerns regarding the concentration of user traffic on major platforms. This paper uses the policy adopted by a leading bike-sharing platform where a consumer can use its shared bikes without any cash pledge if his/her Zhima Credit score reaches 650 points. Using individual consumer-level data aided by a regression discontinuity identification strategy, the paper finds that the credit-based deposit-free policy increases the proportion of bike-sharing usage through the Alipay APP by about 12%, while all other competing platforms experience decreases in user traffic. The paper provides novel empirical evidence showing that internet platform companies could go so far as to leverage big-data credit investigation technology to affect the competitive structure of the market. This insight offers rich policy implications for anti-trust regulation of internet platforms and the construction of the social credit system in China.

Key words: big-data credit investigation, bike-sharing, platform user traffic

JEL codes: L41, G28, D12

(截稿:2022年6月 责任编辑:郭若楠 吴海英)