



北京大学数字金融研究中心
Institute of Digital Finance, Peking University

北京大学数字金融研究中心工作论文系列

IDF Working Paper Series

NO.IDFWP2020001 (总第 20 期)

新冠肺炎疫情对线下微型商户短期冲击的定量估算 ——兼论数字金融缓解冲击的价值

王靖一、郭峰、李勇国*

摘要：新型冠状病毒肺炎疫情给国民经济运行，特别是线下微型商户的经营造成严重冲击，但由于数据缺失，相关定量评估分析比较困难。本文基于支付宝旗下支付工具“码商”的海量数据，利用机器学习方法推算了如果没有疫情发生，2020年春节后线下微型商户运行应该具有的“反事实结果”，进而定量估算了疫情造成的冲击。结果表明，2020年正月初七后的两周，线下微型商户活跃商户量与营业额较估算的没有疫情冲击的“反事实结果”，分别下降了约40.4%和52.4%。进一步分析显示数字金融发展可以缓解疫情对线下微型商户的冲击。本文的研究在结论上和方法论上，对当前其他国家应对疫情，以及我国未来可能发生的其他突发公共卫生事件，均具有非常好的借鉴价值。

关键词：新冠肺炎，线下微型商户，数字金融，机器学习

*王靖一：中央财经大学金融学院、北京大学数字金融研究中心；郭峰：上海财经大学公共经济与管理学院、北京大学数字金融研究中心；李勇国：蚂蚁金融服务集团研究院。通讯作者：郭峰，电话：021-65903586，邮箱：guo.feng@mail.sufe.edu.cn，地址：上海市杨浦区武川路111号上海财经大学凤凰楼521室。本文获得国家社会科学基金重大项目“数字普惠金融的创新、风险与监管研究”（课题号：18ZDA091）的资助，也得到了北京大学数字金融研究中心和蚂蚁金服集团有关部门以及黄益平、李振华、都阳、程志云、王芳、舒眉、张瑜等专家提供的宝贵意见和帮助。感谢国家地球系统科学数据中心的徐洋在地理信息方面给予的专业建议；感谢石庆玲、田莫帆、毛晨宇、王雪、魏亚杰的帮助。特此致谢。文责自负。

新冠肺炎疫情对线下微型商户短期冲击的定量估算

——兼论数字金融缓解冲击的价值

摘要：新型冠状病毒肺炎疫情给国民经济运行，特别是线下微型商户的经营造成严重冲击，但由于数据缺失，相关定量评估分析比较困难。本文基于支付宝旗下支付工具“码商”的海量数据，利用机器学习方法推算了如果没有疫情发生，2020年春节后线下微型商户运行应该具有的“反事实结果”，进而定量估算了疫情造成的冲击。结果表明，2020年正月初七后的两周，线下微型商户活跃商户量与营业额较估算的没有疫情冲击的“反事实结果”，分别下降了约40.4%和52.4%。进一步分析显示数字金融发展可以缓解疫情对线下微型商户的冲击。本文的研究在结论上和方法论上，对当前其他国家应对疫情，以及我国未来可能发生的其他突发公共卫生事件，均具有非常好的借鉴价值。

关键词：新冠肺炎，线下微型商户，数字金融，机器学习

一、引言

2019年年底开始爆发的新型冠状病毒肺炎疫情，迅速蔓延至全国所有省份，截至2020年3月底已致全国8万余人感染，病亡3000余人。这次新冠肺炎疫情，是新中国成立以来在我国发生的传播速度最快、感染范围最广、防控难度最大的一次重大突发公共卫生事件¹。疫情不仅严重危害国民的身心健康，还因为突然性的“封城”²、“闭户”³，在短时期内对国民经济运行造成严重的冲击。特别是广大线下微型商户，持续性的闭门休市，已经给千百万以此为生的家庭造成了严重影响。而在疫情缓解之后的复工复产中，与积极推动规模以上大企业和大型商超的复工复产相比，各地地方政府对推动个体经营者恢复营业的积极性明显不高。然而，遍布城乡各地的沿街商铺和流动商贩等线下微型商户⁴，不仅是很多家庭赖以维生的重要经济来源，也是市场经济运行的微观基础和国民经济发展的毛细血管，对宏观经济的健康发展和稳定运行具有非常重要的意义。2020年2月23日，习近平总书记专门强调要“解决个体工商户尽快恢复营业的问题”。2019年12月30日、2020年2月18日的两次国务院常务会议上，李克强总理也肯定了个体工商户、“小店经济”对经济可持续发展和实现包容性就业的重要价值。从实务出发，线下微型商户包括了个体工商户、少数未经注册开展营业的商户以及一小部分极其微小的公司，他们的共同特征是，收款账户是一个自然人而非法人账户。

不过，线下微型商户的重要价值虽然已经得到各界的广泛认可，但无论是政策界还是学术界，关于疫情对线下微型商户冲击的具体情况却甚少讨论，特别是缺乏量化的评估分析。

¹ 习近平总书记在2020年2月23日召开的《统筹推进新冠肺炎疫情防控和经济社会发展工作部署会议》会议上对本次疫情的定性，<http://cpc.people.com.cn/n1/2020/0223/c64094-31600380.html>

² 2020年1月23日凌晨，武汉市新型冠状病毒肺炎防控指挥部宣布自1月23日10时起，全市城市公交、地铁、轮渡、长途客运暂停运营，机场、火车站离汉通道暂时关闭，之后该政策被湖北省其他城市迅速效仿。

³ 2020年2月1日，浙江省温州市宣布自2月2日起，在全市范围实行村（居）民出行管控措施，全市每户家庭每两天可指派1名家庭成员出门采购生活物资，这一政策随后被全国很多地区效仿。

⁴ 从严格定义而言，除线下的沿街商铺和流动商贩等之外，线下微型商户还应该包括在各个网络平台上开展交易的微型商户，但在本文中，我们仅考虑线下的线下微型商户。

究其原因,就是因为作为一种非正式或半正式的商业活动,沿街商铺和流动商贩等线下微型商户的相关统计非常困难,因此在统计上会或多或少地被遗漏¹。在这一背景下,本文利用支付宝旗下的数亿级的“码商”数据和机器学习分析方法,定量评估了疫情对线下微型商户产生的冲击,并特别考察了数字金融在缓解疫情对线下微型商户冲击中起到的作用²。这些研究成为可能,完全是因为数字经济及数字金融已经广泛渗透到我们的生产生活中的方方面面。在数字经济时代,电子支付越来越成为中国人购物消费时的首选方式。根据中国人民银行金融消费者权益保护局的调研数据,2018年,全国使用电子支付的成年人比例为82.39%;农村地区使用电子支付的成年人比例也达到72.15%(中国人民银行金融消费者权益保护局,2019)。目前,不仅各种互联网线上的商业交易是通过互联网支付等线上支付手段进行的,甚至各类沿街商铺和流动商贩等线下微型商户也是通过支付宝、微信、银联支付等机构提供的二维码支付等电子支付手段来进行收款的,这就为这些极其微小的商户留下了重要的数据记录。而这些数据信息在这些线下收单机构内的积累形成的大数据,就可以产生单个数据点所无法产生的价值。例如根据这些收单流水等信息,收单机构就可以自己或联合其他放贷机构,向这些被传统金融机构完全忽略的小店小铺发放贷款,从而大大突破了传统金融机构放贷的边界,有利地促进这些国民经济毛细血管的成长。而另一方面,这些数据形成一定的积累后,还可以用来帮助我们较为宏观的层面分析这些非正式和半正式的微型商业活动的规模(王靖一等,2020),以及定量评估疫情等外生冲击对这些商户活动经营的影响。例如,2020年农历腊月廿六(1月20日)前两周时间内,疫情还没有被各地政府和民众重视,“码商”营业额较2019年同期(农历)增长了33.1%,但2020年正月初七(1月31日)后两周时间内,“码商”营业额较2019年同期(农历)则下降了39.4%。在此基础上,我们在图1中还将春节前后“码商”营业额的同比变化精确到全国地级市层面,从中也可以看出疫情对个体经营户的冲击是全面性的,全国没有什么地方能幸免:虽然初步看来,疫情之后,内蒙古、青海、西藏、广西等地区的“码商”营业额仍能维持正增长,但如果跟这些地区春节前同比80-90%的增速相比,冲击的幅度依然是非常巨大的。

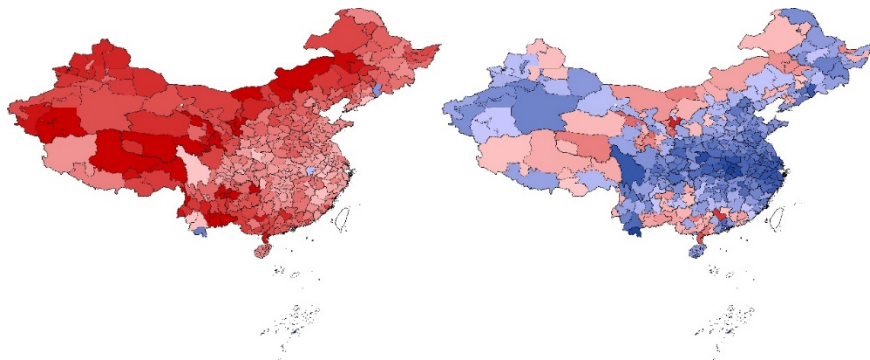


图 1: 2020 年春节前后“码商”营业额同比增幅对比

注:①左图为2020年腊月廿六前两周码商营业额与2019年农历同期的同比变化,右图为2020年正月初七后两周的同比变化;②红色代表增长,蓝色代表下降,颜色越深,增长(下降)幅度越大。③港澳台地区数据缺失。

通过这些数字的简单分析,我们就可以明白单纯用疫情发生后线下微型商户营业额与去

¹ 例如,第四次经济普查数据显示2018年年末全国有线下微型商户6295.9个,但根据王靖一等(2020)的估算,全国线下微型商户2018年实有9776.5万户。

² 本文的研究是和蚂蚁金服研究院共同完成的,在分工上,课题组外部成员负责了课题方案设计、数据分析、外部数据收集整理、论文执笔等工作;而来自蚂蚁金服的课题组成员则负责底层数据的整理和汇总等工作,外部研究人员自始至终没有接触过码商的底层微观数据。数据分析工作也一直在蚂蚁金服开发的数巢实验室当中进行。上述实验室和分工机制使得在保护消费者隐私和商业机密的前提下,又可以利用金融科技企业经营中形成的大数据积累,来分析国民经济运行中的各种宏微观问题。

年同期数据的比较，来刻画疫情的冲击是不严谨的：从学术角度谈，只有估算出“假如不发生疫情”的反事实结果，才能做出疫情与冲击之间严谨的因果联系；从数据角度谈，简单的同比、环比指标会因为业务本身处于增长扩张期而变得存在大幅度偏差。评估疫情等突发事件对经济运行的冲击，一个根本性的问题是只有发生了疫情后的实际数据，缺少如果疫情没有发生、经济会如何运行的信息。因此为求得冲击的科学估算，关键是推算出如果没有疫情发生的话，经济运行应该具有的“反事实结果”。我们使用这一方法，并不是把简单的问题复杂化，而是不这样做不足以解决复杂的现实问题。在本文，我们定量估算疫情对线下微型商户冲击的基本思路是基于 2019 年同期（农历）以及 2020 年 1 月 20 日（含）前两周¹的线下微型商户数据，用机器学习算法来“预测”如果没有疫情发生，在 2020 年 1 月 31 日（正月初七）-2 月 13 日（正月二十日）两周时间内，线下微型商户的活跃商户量、营业额等“反事实结果”，然后与 2020 年这两周时间内实际发生的活跃商户量、营业额等进行比较，两者相差就是疫情这个突发的外生冲击的真正影响。具体而言，根据我们的估算，在重点疫区湖北省，线下微型商户活跃商户量和营业额分别下降 59.3% 和 69.7%；而在全国其他地区，活跃商户量和营业额则分别下降约 40% 和 50%。而在数字金融缓解疫情冲击的分析中，我们基于“北京大学数字普惠金融指数”的底层指标（郭峰等，2020）与不同地区码商数据的匹配和回归，发现一个地区数字金融的发展可以显著缓解疫情的冲击。具体而言，基于数字技术精准发放的贷款每增长 1%，疫情对线下微型商户营业额的负面影响平均就减弱 2.57%；而如果一个地区基于数字技术精准发放的贷款，能从全国的均值发展到杭州的水平，会使得疫情的负面冲击下降 51%。

本文在研究视角和研究方法上具有如下几个创新性：第一，以往相关研究中，能将研究触角延伸至小微企业，已属不易，而本文则克服了数据缺失的障碍，从金融科技公司经营中积累的数据入手，定量评估了新冠肺炎疫情对线下微型商户的冲击，并考察了数字金融在缓解疫情冲击中的独特作用。第二，利用机器学习算法，估算了如果没有疫情这一突发外生冲击的发生，线下微型商户的经营在 2020 年春节后应该具有的“反事实结果”，从而为科学严谨地评估疫情的冲击，创造了条件。利用机器学习进行因果关系识别，目前已经成为经济学实证研究中的最新趋势，本文是这方面的一个最新尝试。第三，为了克服行政区划聚合数据过疏，栅格数据缺乏根据线下微型商户密度进行自适应调节的缺陷，本文以银行网点为中心点，将全国划分为 26.7 万个泰森多边形，从而解决了微观数据聚合尺度的问题。最后，从上述几点创新的总结中也可以看出，虽然本文考察焦点在于评估新冠肺炎疫情对线下微型商户造成的短期冲击，但评估的思路与方法、以及研究结论，都并不限于眼前，而是对当前其他国家应对疫情，以及我国未来可能发生的其他突发公共卫生事件，均具有非常好的借鉴价值。。

本文余下安排为：第二部分综述与本文研究主题相关的几支文献；第三部分介绍我们使用的数据和研究设计；第四部分报告我们对疫情冲击定量评估的结果，并分析数字金融在缓解疫情冲击中发挥的作用；最后一部分是结论总结与启示。

二、文献综述

新冠肺炎疫情发生后，各级政府、各个部门纷纷出台多种扶持政策，以求稳定经济和就业。然而，精准高效的扶持政策，有赖于对疫情冲击的科学评估。疫情冲击的评估，对有的行业可能比较容易，比如交通运输业，客运量、货运量数据非常容易获取。而对有些业态，评估疫情的冲击就难度较大，例如线下微型商户。对于沿街商铺、流动商贩等线下微型商户

¹ 本文提及某时间窗口的起止日期时，均包含该日，下同。

的性质,目前并没完全契合的中文学术词汇可以用来概括这个商业群体。英文中类似的词大概是 *informal sector* (非正式行业) 或 *informal economy* (非正式经济),但这个词一般是指没有在政府部门注册登记,也不缴纳税收或缴纳社保的业态 (Babbitt et al., 2015; Harriss-White, 2010; Maloney, 2004; La Porta and Shleifer, 2008)。当然,用非正式行业或非正式商业活动来定义我们的研究对象,也不完全准确。我们研究的这个群体虽然确实基本不怎么缴纳直接的税负,但很多还是经过了工商登记注册过的,例如个体工商户。当然,也存在一些没有经过工商注册,但事实上仍在从事一些商业活动的群体,这也在我们的研究范围。虽然难以有非常准确的定义,但这些线下微型商户一般具有如下几个特征:营业额一般都不高,一年的营业额一般不过几万元或几十万元;往往以经营者个人名义注册银行和其他第三方支付收款账户,即不区分经营实体和经营者自身;雇员规模较小或无雇员,以夫妻店、家庭经营的形式经营,店主或雇员往往都缺乏正式的社保和个税缴纳记录。因此,从文献上谈,我们研究的线下微型商户是一种非正式 (*informal sector*) 或半正式 (*quasi-formal sector*) 的商业活动。

非正式行业在非洲、拉美等欠发达国家非常普遍。例如,根据非洲开发银行的统计,在撒哈拉以南非洲国家,非正式经济能占到 GDP 的 55%,就业的 80% (AFDB, 2013)。又如,根据乌干达统计局的官方统计,非正式经济也占到乌干达经济总量的 43% (UBOS, 2012)。在印度,非正式经济的规模也非常庞大,有估计认为印度的非正式经济活动占到了印度 GDP 的 50% (Sen and Kolli, 2009)。在印度尼西亚,根据相关劳动力调查数据,非正式行业就业占到了全部就业人口的 61%-70% (Alatas and Newhouse, 2010)。而根据王靖一等 (2020) 的估算,中国的线下微型商户从业人员达到 2.31 亿人,占全国劳动力总人口的 28.8%。非正式行业一个显著的特点是其往往脱离正式的统计和监测 (Blackburn et al., 2012; Capasso and Jappelli, 2013; Schneider and Enste, 2000),从而使得财政、货币等宏观调控政策变得更加困难。Restrepo-Echavarria (2014) 的研究就发现,如果非正式经济的测度比较差,那么在国民经济波动的统计分析中,就会产生误导性的结论。但数字经济的发展,为监测和统计分析线下微型商户,包括评估外生事件对线下微型商户经营产生的冲击创造了条件。

过去几年,数字经济以及数字金融的发展对国民经济发展产生了深远的影响 (Goldfarb and Catherine, 2019; 黄益平和黄卓, 2018)。随着数字经济和数字金融的深入发展,数字支付以及其他数字金融服务已经广泛渗透到我们生活的方方面面 (Chen, 2016),在便利了我们的衣食住行之外,也为这些支付工具提供商积累了宝贵的数据,进而也为对数字金融使用者的统计分析创造了条件:郭峰等 (2020) 基于互联网支付、信贷、征信、保险等指标,构造了一个度量各地区数字金融发展现状的“数字普惠金融指数”;王靖一等 (2020) 则利用支付宝旗下的“码商”数据,估算了中国全国和各地区线下微型商户的总规模,以及对就业的贡献度等。当然,数字经济和数字金融的发展,对国民经济的价值并不限于便利了线下微型商户等微型商户的统计检测。数字经济和数字金融对经济社会生活影响是方方面面的,对此,现有文献已经给予了广泛讨论,例如现有文献研究了数字金融对银行运行的影响 (邱晗等, 2018)、对农村金融的影响 (傅秋子和黄益平, 2018)、对创新创业的影响 (谢绚丽等, 2018; 张勋等, 2019) 和对居民消费的影响 (易行健和周利, 2018), 等等。

本文关注的另一个焦点是数字经济和数字金融发展中积累的大数据,也为线下微型商户获得信贷等金融支持,创造了条件。关于中小微企业融资难问题,在经济学文献中已经得到了非常广泛的讨论。中小微企业融资难最重要的一些原因包括:金融体系结构失衡 (刘克崮, 2009)、地方债务挤压 (田国强和赵旭霞, 2019),以及金融机构对中小微企业服务上存在的获客、征信成本高企以及信息不对称等障碍 (李扬, 2017)。而为解决中小微企业融资难问题,文献也提出了很多建议:通过政府的定向扶持政策降低企业,特别是民营企业的融资难度 (徐光等, 2019)、通过担保物权制度改革,降低企业抵押融资难度和融资成本 (钱雪

松等, 2019)、发展定位于服务中小微企业的中小金融机构(刘克崮, 2009; 李广子等, 2016)、发展供应链金融(宋华和卢强, 2017), 等等。然而, 如果中小微企业融资都如此困难, 那么连“企业”都算不上的广大线下微型商户更是融资无门了, 上述关于解决中小微企业融资的建议, 也很难推广适用于从事非正式和半正式商业活动的线下微型商户身上。但线下微型商户在融资上的“边缘化”(孔祖根, 2002)就无法扭转吗? 并非如此, 数字经济和数字金融的发展和广泛渗透, 为线下微型商户解决融资问题, 进而提高抗风险能力, 提供了可能。一些数字金融机构利用支付等数字金融服务中积累的关于线下微型商户的大数据, 解决了获客难与风控难的问题, 进而服务了大量的线上与线下微型商户。关于数字金融机构利用大数据、机器学习等技术改进放贷策略的研究已有不少, 如 Björkegren and Grissen (2019)、Gambacorta et al. (2019)、邱晗等 (2020), 但一般性地讨论数字金融机构对线下微型商户的信贷支持, 并不是本文的主要目的。本文主要焦点是在评估疫情对线下微型商户冲击之后, 进一步考察一个地区数字金融发展水平更高, 是否能提高该地区线下微型商户的抗风险能力, 从而缓解疫情对该地区线下微型商户的冲击。

新冠肺炎疫情发生之后, 经济学家们投入了极大精力, 开展了对 COVID-19 的相关研究。经济学家们首先研究了经济和社会因素在影响 COVID-19 扩散中的作用 (Fang et al., 2020; Qiu et al., 2020; Zhan et al., 2020), 以及以前类似大流行病对经济政治的影响, 例如 1918 年大流感的经济后果 (Correia et al., 2020; Barro et al., 2020)、埃博拉病毒对美国政治选举的影响 (Campante et al., 2020)。Atkeson (2020) 基于 SIR 模型考察了 COVID-19 在未来 12-18 月对美国经济的可能影响, Eichenbaum et al. (2020) 利用类似的模型, 考察了 COVID-19 导致美国经济衰退的可能后果。这些研究都很有价值, 但对于本次 COVID-19 的经济后果, 还缺乏具体的量化评估。中国疫情与其他国家形成了数周的时间差, 因此来自中国的经济影响的评估和相关应对政策经验, 对其他国家都会有非常直接的启示。基于小样本的调查数据, 中国学者对中国中小企业遭受的冲击已经进行了一些研究 (朱武祥等, 2020; 金泉等, 2020)。通过抽样问卷调查开展研究可以得到一些直观的结论, 但又存在严重的弊端: 其一, 在疫情特殊时期, 问卷应答率可能不够理想, 而且很难核实, 而未响应偏差又是影响问卷调查结果科学性的重要因素 (Korinek et al., 2007); 其二, 问卷调查可能会存在一些主观回答导致的偏差, 这也是问卷调查中常见的问题 (Kleinjans and Van Soest, 2014)。特别是对于个体工商户为主的线下微型商户, 相关调查研究更是完全忽视 (王靖一等, 2020)。因此, 考察疫情对线下微型商户的影响, 迫切需要一个全新的数据来源, 而国家级金融科技公司经营中积累的大数据为此创造了可能。

三、数据来源与研究设计

(一) 数据来源

正式介绍数据和研究设计之前, 还需要对线下微型商户的概念进行说明, 特别是其与个体工商户之间的区别和联系。粗略来讲, 所谓个体工商户, 是指按照相关法律法规在工商部门注册登记过的微型商户; 而线下微型商户则不仅包含工商注册的, 还包括没有注册过, 但实际上在从事经营行为的微型商户。此外, 从严格定义而言, 除线下的沿街商铺和流动商贩等之外, 微型商户还应该包括在各个网络平台上开展小额交易的微型商户, 但在文中, 我们仅考虑线下的微型商户。而“码商”之所以能为本研究提供基础性数据, 首先是因为“码商”的定位和服务对象与线下微型商户高度重合。例如, “码商”基本上都是用个人的名义注册, 这与线下微型商户一般不区分经营实体与经营人的特点高度契合, 这样就排除了那些大型商

超、连锁店等大型商业机构，这些大型商业机构往往以公司名义注册和收款¹。而统计也显示，码商的营业额也普遍不高，99%的码商通过支付宝收款的年营业额在 40 万元之下，也与我们理解的线下微型商户的规模高度一致。其次，线下微型商户也有很强的积极性去注册、使用码商，这也是一个非常重要的前提。实际上，开通码商非常方便，只需要通过手机简单申请即可即时生成；而且这个商户也有很强的激励开通码商，因为使用收钱码而不是平常的收款二维码，可以减免提现手续费。最后，我们删去了那些交易不活跃、经营区域范围不明确的码商，这样就保证了码商后面确实是一个日常性经营的商户，而不是一种一次性的偶尔行为。当然，这里需要补充说的是，由于毕竟还有很多其他公司也在提供类似的支付服务，因此不能将支付宝一家公司的具体产品“码商”直接等同于全部的线下微型商户²，但是，我们在本文中，主要评估的是疫情对线下微型商户活跃商户量和营业额的冲击，是一个百分比的形式，线下微型商户同时使用多个收款方式，并不会影响我们的评估。况且支付手段的选择权更多地掌握在消费者手中，线下微型商户因其体量微小基本也不会牵扯几家巨头间的“选边站”，同时展示几种收款码，或者使用不同支付手段统一入口的“聚合码”，是一种极为常见的现象。因此上述“以偏概全”就不算是一个严重的问题，因为一方面疫情不会选择性地影响偏好某家服务的线下微型商户、消费者，另一方面疫情也不会让线下微型商户与消费者在选择两家产品时发生结构性的偏好改变。

当然，除了码商数据之外，我们还匹配了很多其他外部数据。选择这些数据的一个重要考量是因为线下微型商户“线下”与“微型”两个基本特征决定其服务对象往往是其经营场所附近的人群，因此他们的规模也与其所在地在的周围环境密切相关。因此，为更好地刻画线下微型商户与其周边环境的关系，我们选取了如下一些人口、经济和地理等方面的数据：

首先，我们通过 API 从高德地图中抓取了约 3500 万个兴趣点的数据。兴趣点是一个地理信息系统的概念，简单而言，一切可以在高德地图检索到的商店、饭店、学校等实体，都可以称为一个兴趣点。我们共获取了与本研究相关的 18 个一级分类³。同时，我们也获取了下文将要正式定义的每个泰森多边形中心点到所属区县中心、地市中心（市县中心的定义根据高德选取，为经济重心）的驾车距离。数据采集时间为 2019 年 12 月。

其次，为了获得颗粒度更高的经济发展水平、人口密度等信息，我们使用了 500 米精度的夜晚灯光亮度栅格数据，以及 1 公里级别的人口栅格数据，以及 30 米精度的高程数据。夜晚灯光数据来自美国国家大气与海洋局公布的 VIIRS 数据，使用的是 2018 年的年度数据。人口数据来自哥伦比亚大学，原始数据的颗粒度为 1 公里栅格，为了进行匹配，我们进行了简单平均的超分辨率处理。人口数据的得出是结合卫星遥感与普查数据，时间为 2015 年，这是能够找到的最新年份的人口数据。高程数据则主要用来生成海拔信息与一系列衡量该区域地理崎岖程度的变量，这也是一个线下微型商户辐射范围的重要影响因素。

再次，由于线下微型商户的经营对于天气情况极为敏感，特别是夜间经济受到气温的影响尤其明显，故而我们使用了来自欧洲中期天气预报中心的气象数据，主要包含气温、风力、湿度与降水量，数据精度为约 25 公里的栅格。此外，空气质量也可能对线下微型商户活跃程度有所影响，故而我们获取了中国大陆境内 1400 余个空气监测点数据，包括空气质量指数（AQI）以及 PM_{2.5}、PM₁₀、SO₂、CO、NO₂、O₃ 等单项污染物的浓度数据。同时，NO₂

¹ 在蚂蚁金服内部，商超、连锁店等作为 KA 客户（Key Account），其运维团队与我们研究使用的“码商”团队完全不同，不在我们的研究范围。在我们使用的“码商”数据中，公司名义注册的码商占比不超过 0.02%，已从我们研究样本中剔除。

² 不过，王靖一等（2020）确实用支付宝码商数据估算了全国线下微型商户的总规模，详细的讨论详见该文。

³ 按高德官方的划分和命名方法，具体包含：涵盖汽车服务、汽车销售、汽车维修、摩托车服务、餐饮服务、购物服务、生活服务、体育休闲服务、医疗保健服务、住宿服务、风景名胜、商务住宅、政府机构及社会团体、科教文化服务、交通设施服务、金融保险服务、公司企业与公共设施公共设施。

数据与汽车，CO 与工厂排放也均有着较为紧密的关系。

最后，为了衡量数字金融环境对于线下微型商户的影响，我们还提取了郭峰等（2020）编制的“北京大学数字金融普惠指数”的底层指标（取自然对数后进行标准化）。这些底层数据相较于指数化后的结果，可以更加清晰地衡量每个分项的影响力，同时也可以对影响的尺度进行清晰的量化解释。

（二）数据聚合

在正式进行建模分析之前，我们还需要将数千万的码商以及其他数据在某个地理层面进行聚合，即将数千万级的码商数据在某个层级的地理尺度进行加总¹。传统基于行政区划聚合数据会导致过于疏松：即便精确到县一级，全国也才 2800 个县域；而栅格划分方法又缺乏根据线下微型商户密度进行自适应调节的缺陷：核心城区一个栅格内可能有很多线下微型商户，而到了中西部地区和农村地区，又可能会出现大量空白栅格。一种理想的聚合方式应当可以根据当地的商业活力进行自适应调节：在那些人口密集，商业活跃的地区，区块应当划分得适当密集一些；而在那些商业活力较低、人烟稀少的区域，单个区块的面积应当适当大一些。这样，能够保证区块的数量在一个可控的范围，同时每个区块承载的线下微型商户数量和规模与其周边区域的人口、经济等特征之间的关系也更加紧密一些，从而使得后续的测算变得更为有效。这种区块划分的思路看似比较新奇，但实际上跟中国行政区划划分的历史沿革是高度一致的。在中国的地方行政区划中，县级行政区是最为重要的一级，保持了上千年的稳定历史，而县级行政区的设置逻辑就是人口密度高、经济活力强的地区，县域设置就相对稠密一些；而西部偏远地区，县域设置就稀疏一些，即政区幅员的尺度并非整齐划一（周振鹤，2013）。

根据上述思路，我们选择了一种被称为“泰森多边形”的方法（蒋晨琛等，2019；Jia et al., 2018）。在这个思路中，在平面上有一系列中心点，这些中心点生成的泰森多边形满足如下特性：某多边形内的任意一点到本多边形中心点的距离均小于该点到其他多边形中心点的距离。在那些中心点密集的区域，多边形小而稠密；在中心点稀疏的地方，多边形面积大而稀疏。泰森多边形划定的一个关键要素是如何确定每个区域的中心点。在研究犯罪行为时空分布问题时，蒋晨琛等（2019）将某市的派出所作为中心点，这是一种非常有启发性的中心点选择方法。而在本研究中，经过反复斟酌讨论后，我们将银行网点（含自助银行网点）作为划分泰森多边形的中心点。一方面，在商业活跃、经济发达的地区，银行网点的分布确实就非常稠密²；另一方面，国有银行在相当程度上履行了一些社会责任，内地 31 省中，只有非常特殊的西藏自治区双湖县、申扎县以及海南省的南沙群岛、中沙群岛等六个县域中没有检索到银行网点³。

最终，在对距离过近的银行网点进行去重⁴之后，课题组共获得中国内地 266793 个银行网点，并基于此构建泰森多边形，将内地 31 省（直辖市、自治区，以下简称“省”）进行了网格化分割。图 2 展示了杭州市西湖区经过网格化处理后的结果，每一个红点就是一个银行

¹ 即便抛开数据处理量的问题，我们也不能拿单个码商的数据进行分析，因为这涉及到用户隐私问题。

² 在我们的样本中，单个 100 米*100 米方格内银行网点数最多出现在香港某网格内：17 个。

³ 网点位置数据由研究团队使用 API 从高德地图中获取，采集时间为 2019 年 12 月。

⁴ 一些城市的繁华位置银行分布往往过于密集，因此为了数据的合理性，我们将聚集在一个 100 米*100 米方格的网点聚合为一个，其位置为这些银行网点坐标的平均值。做如此处理主要是基于两点原因：首先，受限于中国法律规定，任何在线地图提供的坐标值均不是国际通行的真实经纬度，而是按照 GC02 标准进行随机、不可逆加密后的坐标值，经过加密后，坐标的位置会发生随机不可控的几米到几十米的偏移；其次，根据 Okabe 的研究，在城市区域、几十米的尺度下，基于空间距离建立泰森多边形变得不合理，此时应当基于道路交通距离进行多边形的构建。

网点，细线便代表泰森多边形的边界。杭州市西湖区是一个很好的可视化样例，因为一方面它的东北区域是密集的城市商业区，银行网点密集；另一方面南部和东部又存在面积广大的风景区，只有山坳中的小镇，从而银行网点就稀疏一些。这一密一疏恰恰展现了基于银行网点的泰森多边形对不同商业活跃环境的自调节能力，从而达到了我们预期的效果。

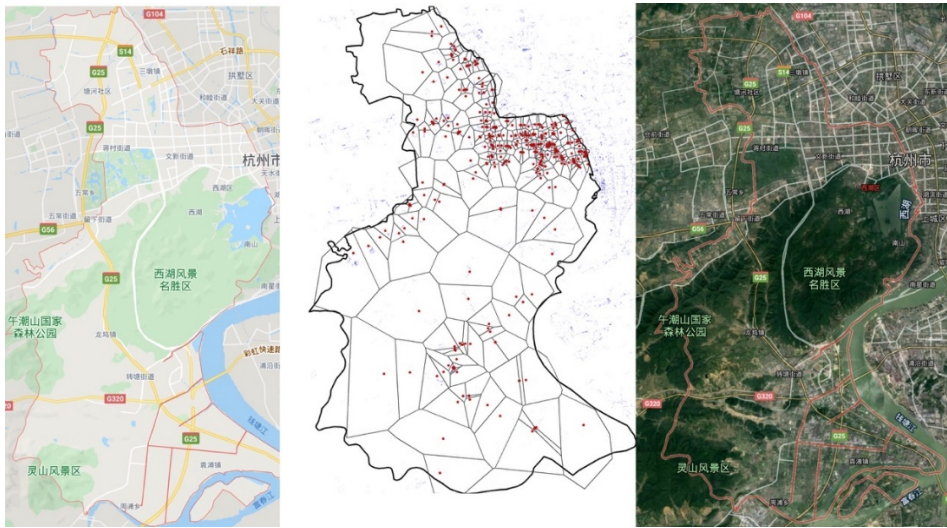


图 2：基于网点的泰森多边形：以杭州市西湖区为例

注：中图为研究者自绘，蓝点代表餐馆，位置信息来自高德；左右图截取自谷歌地图。二者形状上的差距系不同投影方式造成。

确定好度量尺度后，我们就将码商、兴趣点等数据在这一层面进行汇总聚合，同时也将灯光、人口等地理栅格数据在这个尺度上进行匹配。表 1 和表 2 分别汇报了所用的码商和外部特征数据的基本情况。

表 1：码商数据基本情况

变量类型	变量名称
总量数据	码商数量
	码商交易总笔数
	码商交易总金额
性别分类	女性码商数量
	女性码商交易总笔数
	女性码商交易总金额
年龄分类	55 岁以上码商数量
	24 岁以下码商数量
	55 岁以上码商交易总笔数
	24 岁以下码商交易总金额
	55 岁以上码商交易总笔数
	24 岁以下码商交易总金额
时段分类	周末交易总笔数
	周末交易总金额
	18: 00-23: 00 交易总笔数
	23: 00-06: 00 交易总笔数
	18: 00-23: 00 交易总金额
	23: 00-06: 00 交易总金额

地区分类	外省人码商数量
	外省人码商交易笔数
	外省人码商交易总金额

注：①这些数据均在蚂蚁提供的研究开放平台数巢实验室中与外部数据相匹配，确保数据不外泄，同时因为我们所使用的的变量有极强的时效性，安全等级要求更高，这里无法对码商的数据特征进行更详细的披露；②此时间段以北京为标准，其他地区的时间跨度按按其经度所在时区进行调整，中国的经度横跨五个时区；③地理颗粒度均为上述 26.7 万个泰森多边形。

表 2：外部特征数据基本情况

变量类型	变量名称	频率	层级	备注
灯光	夜间灯光密度	年度	多边形	由 500 米栅格数据汇总到多边形
人口	人口数量	截面	多边形	由 1 公里栅格数据汇总到多边形
兴趣点	18 个一级分类	截面	多边形	根据经纬度匹配到多边形
高程	平均海拔	截面	多边形	预先进行 1%-99%的缩尾处理，以避免极端值影响
	海拔差	截面	多边形	多边形内 99%分位数与 1%分位数的差
	最大梯度	截面	多边形	多边形内梯度值的 99%分位数，避免极端值影响
	中位数梯度	截面	多边形	30 米精度的高程（可简单认为是海拔）数据
天气	气温	月度	多边形	由 25 公里栅格数据匹配到多边形
	风力	月度	多边形	由 25 公里栅格数据匹配到多边形
	降水量	月度	多边形	由 25 公里栅格数据匹配到多边形
	湿度	月度	多边形	由 25 公里栅格数据匹配到多边形
空气质量	空气质量指数	月度	多边形	站点*小时级数据汇总、匹配到多边形
	PM2.5	月度	多边形	站点*小时级数据汇总、匹配到多边形
	PM10	月度	多边形	站点*小时级数据汇总、匹配到多边形
	SO ₂	月度	多边形	站点*小时级数据汇总、匹配到多边形
	NO ₂	月度	多边形	站点*小时级数据汇总、匹配到多边形
	CO	月度	多边形	站点*小时级数据汇总、匹配到多边形
	O ₃	月度	多边形	站点*小时级数据汇总、匹配到多边形
数字金融	支付、信贷、保险投资等	年度	县域	“北京大学数字金融普惠指数”的底层指标，共计 33 个，详见郭峰等（2020）的公式（1）

（三）评估策略

（1）反事实结果估计。在经济学中，进行外生冲击或某一个政策效果的评估中，一个根本性的问题是，如何进行“反事实结果”的构建。具体而言，我们只能够观测到受新冠疫情冲击后的实际数据，而不知道如果疫情没有发生，经济运行和经济数据将会是一个什么样子。因此，为了进行疫情冲击评估，最佳的方式是我们估算出来，如果没有新冠疫情发生，2020 年春节后线下微型商户经营状况应该是一个什么样子的“反事实结果”，那么疫情带来的冲击便是 2020 年春节后实际经营规模与构造出的 2020 年春节后的“反事实结果”之间的差值。因此，现在的核心问题便是如何估算出一个合理的没有疫情发生的 2020 年春节后的“反事实结果”。逻辑上，一个地区线下微型商户的活跃商户量、营业额等跟该区域的经济水平、人口密度、地理条件等有非常密切的关系，即：

$$OMB_{it} = F(X_{it})$$

其中, OMB_{it} 代表 i 地区 t 时间窗口内的线下微型商户活跃商户量 (营业额), 而 X_{it} 则代表该地区 t 时期的经济水平、人口密度、地理条件等影响线下微型商户的因素。如果是一个中长期的考察, 这样的方程已经足够, 但我们现在的目标主要是评估疫情的短期冲击, 特征变量 X_{it} 的时间颗粒度无法保障, 其中很多特征变量还是不随时间变化的截面数据。因此, 我们将春节前的线下微型商户数据也作为特征变量, 纳入模型:

$$OMB_{it} = F(OMB_{it-1}, X_{it})$$

其中, OMB_{it-1} 代表了 2020 年春节前的线下微型商户的一些特征¹, 通过将春节前线下微型商户数据纳入模型, 就解决了其他特征变量时间颗粒度不够精细的问题。如果没有疫情的发生, 春节前在某地工作生活的民众, 虽然春节期间会有探亲访友、旅游度假等行为, 但春节假期过后, 则基本上应该回到居住地, 因而增加了春节前线下微型商户的数据之后, 就可以大幅提高模型的预测精度。当然, 要完成上述推算, 还需要知道上述模型中的“参数”: 没有参数, 只有特征变量, 也无法求得响应变量。而这一参数, 我们需要通过机器学习方法, 从 2019 年的春节前后中获取。具体而言, 将 2019 年春节后线下微型商户与 2019 年春节前线下微型商户 (以及其他特征变量) 之间的映射关系, 通过梯度下降树 (GBDT) 回归这一机器学习方法, 训练得到一组参数, 然后泛化到 2020 年, 进而利用 2020 年春节前的线下微型商户数据和其他特征数据, 推测出 2020 年春节后的反事实结果。

这一处理的合理性在于线下微型商户的活跃水平与交易量在春节前后呈现出一定的规律性, 即如果没有类似于新冠肺炎这样的外生冲击, 每一个春节前后的情况大致呈现一种共同的规律性。即这里并不要求春节后和春节前的人口何经济活动一定是对称分布的, 而是只要求, 如果没有疫情发生, 2020 年春节前后的人口经济分布与 2019 年春节前后的人口经济分布保持稳定。此外, 通过控制气象、空气质量数据, 也可以很大程度上避免每年春节在阳历日期上的细微变化带来的影响。

具体而言, 2020 年 1 月 20 日 (农历腊月廿六) 可以视为疫情的一个重要转折点, 该日后, 全国各地逐步进入了被肺炎疫情影响的状态: 确诊人数增加、更严格的管理政策随之到来, 线下微型商户被严重抑制。而在 1 月 20 日之前, 包含线下微型商户在内的经济活力基本上没用受到明显的影响, 可以视为一个典型的春节前状态, 因此此前的线下微型商户经营数据可以用来度量春节前后该地区应该具有的状况。至于 2020 年春节后冲击评估的时间窗口的选择, 本文主要目的是评估疫情的短期冲击, 因此选择 2020 年正月初七后的 14 天作为评估时间窗口。如图 3 所示, 从大约正月廿一 (2 月 14 日) 开始, 全国 (除湖北外) 累计病例数基本达到一个收敛的状态, 新增病例显著下降, 各地也陆续开始推动复工复产。

¹ 为了获得更好的模型参数拟合能力和泛化能力, 作为特征变量的春节前线下微型商户数据, 包含的变量更加丰富, 例如不仅包含了春节前线下微型商户总规模数据, 还包含了女性商户、分年龄段商户、外省人商户、夜间商户的规模等变量。详见上文表 1。

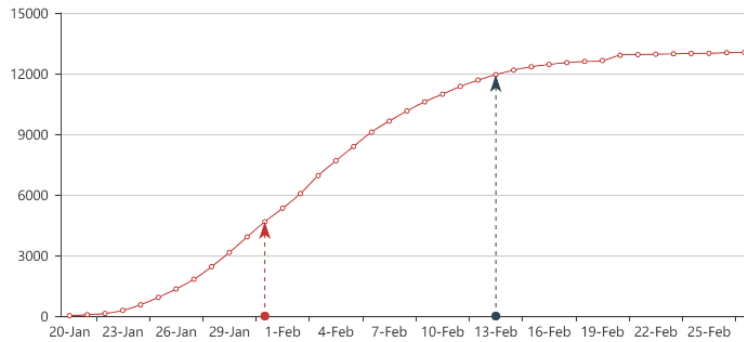


图 3：全国（除湖北外）新冠肺炎累计确诊病例趋势

综上所述，我们就利用了 2019 年同期（以农历记）数据和 2020 年春节前数据，推算了 2020 年春节后如果没有疫情发生的“反事实结果”。这里需要再次强调的是，这里并不是用 2019 年的线下微型商户的规模来预测 2020 年线下微型商户规模；这里只是将 2019 年春节前后线下微型商户（及其他特征变量）之间的映射关系，应用到 2020 年春节前后。这是一个典型的机器学习算法的思维。模型的训练样本为 2019 年的数据，具体而言，2019 年 1 月 18 日到 1 月 31 日（农历腊月廿六）的码商数据，以及上述外部获取的经济、人口、地理等，为训练模型的特征变量；2019 年 2 月 11 日（正月初七）到 2 月 24 日的码商活跃商户量（营业额）为模型的响应变量。使用上述响应变量和特征变量，在全国 26.7 万个泰森多边形的尺度上，进行机器学习算法训练。之后，将上述训练得到的参数，泛化到 2020 年，就可以根据 2020 年 1 月 7 日至 1 月 20 日（农历腊月廿六）的码商数据，以及其他外部特征数据，推测出如果没有疫情的发生，2020 年春节后应该具有的“反事实结果”。这种使用机器学习算法构造反事实结果，进而进行因果推断的思路，已经成为经济学实证研究的最新趋势（Kumar and Liang, 2019; Guo and Zhang, 2019; 陈硕和王宣艺, 2018）。

（2）机器学习算法介绍。对于具体机器学习算法，我们选取的是梯度下降树（Gradient Boosting Decision Tree, GBDT）回归。这是一种较为经典的机器学习方法，最初由 Friedman（2001）提出，并在后续一系列机器学习相关竞赛中展现出出色的表现，而 Chen（2016）所提出的 XGBoost 算法则将这种经典算法的性能进一步提升。即便是在深度学习概念火爆的今天，GBDT 算法与 XGBoost 算法仍然具有稳定、优越的表现，特别是在处理结构化数据方面，国内外一线的科技公司，依然将这一算法作为许多核心问题的解决方案。蚂蚁金服提供的研究开放平台上的 GBDT 算法事实上融合了部分 XGBoost 算法的优点，但依照平台的命名方式，我们在此处依然称其为 GBDT 回归。GBDT 回归的实现过程可以形象化如下表述：算法会随机生成一些使用较少参数、规则相对简单的小判别模型（树），之后算法会综合这些小模型的意见产生一个总体的估计值，通过比较估计值与真实值之间的偏差，对训练数据的采样、小模型的参数、小模型在最终计算中的权重进行调整，多次迭代至前后两次迭代残差小于给定阈值时，迭代停止。

正是由于这些特征，GBDT 回归具有一系列的优点：首先，它可以非常灵活地描述变量之间的各种线性与非线性关系，同时参数又有相对较好的可解释性。通过提升小模型（树）的数量，变量之间各种关系得以充分考虑。同时，GBDT 依然保留了最终结果与变量之间的映射关系，各个变量对最终结果的影响，即变量重要性得以保留并可观测。其次，GBDT 对于变量具有较强的稳健性，一方面我们不需要对变量进行预先的放缩处理，这样可以避免在进行子样本训练时的信息偏移或者污染，另一方面我们也无需担心多个变量之间可能存在的共线性，算法会自动将那些可以被其他变量充分预测、不单独贡献解释力的变量剔除掉。最

后，在合理的超参数选择下，GBDT 具有较其他算法更好的抗过拟合性与更小的预测置信区间，所得到的预测值能够更有效地使用。

四、实证结果

(一) 疫情对线下微型商户的冲击

(1) 总量估算。汇报正式的估算结果之前，我们介绍一下 GBDT 算法的训练和泛化过程。首先，我们使用 2019 年春节前的数据作为特征变量（解释变量），2019 年春节后的数据作为响应变量（被解释变量），在 26.7 万个泰森多边形上进行训练。每一个 2019 年春节后的指标对应着一个模型，即有一个模型来预测总体的营业额，另一个模型来预测总体的活跃商户数量。单个模型的特征数量为 107，这完全在 GBDT 算法可以接受的范围之内¹。其次，我们将所有数据进行 9:1 的切分，保留 10% 作为最终的测试集，另外的 90% 继续进行 4:1 的拆分，通过交叉验证（cross-validation）的方式进行超参数的调参。最终，将调节好的最佳超参数与最优模型在测试集进行测试，测试集存在的意义便是这部分数据从来未被模型学习过。就结果而言，在作为样本外的测试集上，营业额的准确度（ R^2 ）均在 0.86 以上；活跃商户数量的预测准确度（ R^2 ）均在 0.92 以上。之后，我们将最佳的超参数的组合，应用于全体数据，再进行一次拟合²，此时拟合的参数就是往 2020 年春节前后泛化时使用的参数。

而就具体的估算结果而言，在全国总体层面上看，受疫情影响，在 2020 年正月初七后的两个星期时间内，线下微型商户活跃商户较机器学习算法推算出的没有疫情冲击的“反事实结果”减少 40.4%，而营业额则降低了 52.4%。根据王靖一等（2020）的估算，2018 年年末，全国线下微型商户总量约为 9776.5 万户，涉及就业量约 2.3 亿人，全年营业额约为 13.1 万亿元。因此结合上述估算结果，疫情使得这两周内的活跃商户量减少约 3949.7 万户，减少营业额约 2640 亿元人民币，涉及 0.92 亿的就业人员³。而且，不同于大企业职工，线下微型商户这种小铺小店如果没有营业，其就业人员基本就没有“工资”可言。由此可见疫情对于线下微型商户的冲击是相当巨大的。

(2) 异质性估算。本文所使用的大数据不仅可以就线下微型商户受疫情冲击的总体情况进行估算，还可以从很多维度进行异质性估算：就算是线下微型商户，也不是“铁板一块”，疫情对不同地区、不同性质的线下微型商户的冲击也可能不同。为此，本部分我们从以下几个角度对疫情冲击的异质性进行分析。

疫情严重程度的异质性。为了考察不同地区疫情严重程度不同时，线下微型商户所遭遇的冲击幅度是否有明显的差异，例如是否疫情越严重，对线下微型商户的冲击就越严重，我们按照截至 2020 年 2 月 13 日时的累计确诊病例数，将内地 31 个省的疫情划分为五个等级，进行分别评估⁴。图 4 报告了不同疫情等级地区的疫情冲击估计结果。从中可以看出：湖北

¹ 一般而言，工程实践的经验表明，特征数量在 200 以下，GBDT 均能有较好的表现。

² 将最优超参数代回并在全集数据上进行训练，是工程实践常见的一个操作步骤。但是很多学术论文（即使是计算机）并不会提及这一步，因为他们更多地是需要证明自己提出的算法、特征对于预测准确度的改进，而不是对于预测的结果进行进一步的应用。而在实际实践之中，这一做法往往带来更优的效果。

³ 这里需要补充说明的是：我们考察的“码商”仅包含线下电子支付部分，网上消费、外卖性消费不在研究范围，而疫情可能导致线下消费转移到线上，因此这里对线下微型商户损失的估计，并不完全是整个社会的福利损失。

⁴ 具体而言，等级的划分标准如下：截至 2020 年 2 月 13 日，即我们评估的时间窗口截止日，湖北省疫情最为严重，为第一等级，累计确诊超过 5 万例；河南省、安徽省、浙江省、湖南省、江西省、广东省为第二等级，累计病例数均超过 800 例；黑龙江省、北京市、河北省、陕西省、山东省、四川省、云南省、广西壮族自治区、江苏省、上海市、重庆市、福建省、海南省为第三等级，累计病例数均超过 150 例；新疆维吾尔自治区、甘肃省、宁夏省、内蒙古自治区、山西省、贵州省、吉林省、辽宁省为第四等级，累计

省作为疫情最严重地区，疫情对线下微型商户造成了严重的冲击：窗口期内湖北省线下微型商户活跃商户下降 59.3%，营业额下降 69.7%；而对于其他等级的地区，图 4 则显示，疫情的冲击并没有非常明显的差异，特别是并没有出现疫情越严重，冲击越严重的证据：其他地区窗口期内活跃商户下降约 40%，营业额下降约 50%。这样的结果，也在情理之中，疫情发生后，由于对疫情认识还不成熟，疫情形势还存在较大的不确定性，各地迅速启动了一级响应¹，对疫情的防控措施异常严厉，且高度一致，在节后的复工复产中，虽然国务院将春节假期延长至 2 月 2 日，但由于疫情形势依然严峻，各地推动复工复产的节奏均较慢。受上述因素影响，虽然各地疫情各有不同，但各地管制政策高度统一，从而导致疫情对湖北以外各地区线下微型商户的冲击幅度大体相同。随着对疫情认识的更加深入，各地根据疫情严重程度，切实采取分类施策后，疫情冲击的恢复有望出现差异化的步伐。

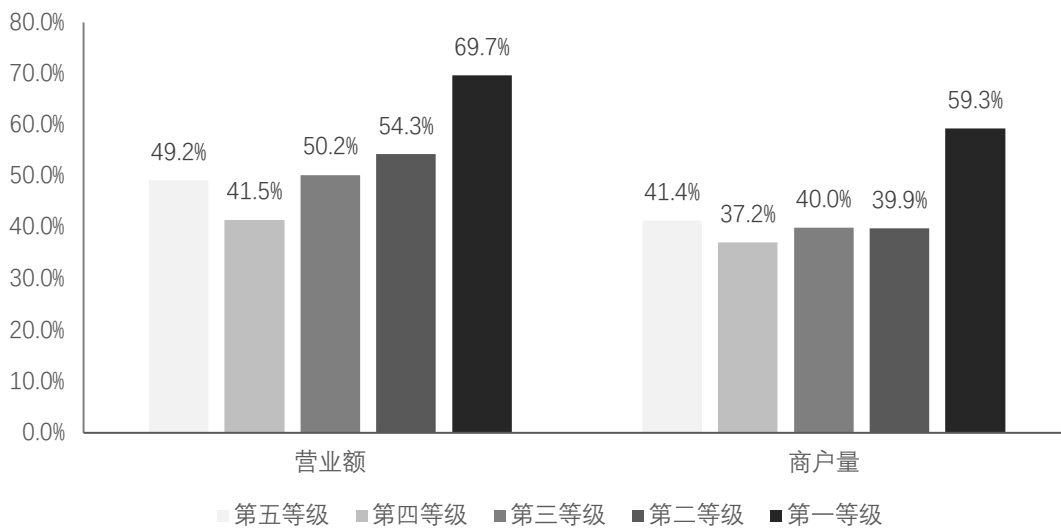


图 4：疫情等级与疫情冲击程度

城乡异质性。由于春节前人口流动是从城市往农村流动，而春节后则是从农村往城市“返程”，疫情形势在春节前夕突然恶化，大量人口滞留农村，无法于春节后返城。因此逻辑上，疫情对城市线下微型商户的冲击应该大于农村。为了检验我们的估计是否符合上述推断，我们下面对城市和农村线下微型商户遭遇的冲击进行对比分析。由于传统衡量城乡的指标比较粗糙，为了更严谨地对比城乡之间的差异，我们根据夜晚灯光数据，将一个区域的城建水平划分为四档：乡村、郊区与城区、城区、核心城区。图 5 汇报了此时的结果，从中可以看出，城市比农村受到疫情冲击的程度更高：疫情对线下微型商户的冲击程度从农村到郊区与镇区，再到城区和核心城区依次递增。具体而言，疫情使得窗口期内的乡村商户量和营业额分别下降 30.0%和 38.8%；而疫情对核心城区商户量和营业额的冲击幅度则分别达到 45.7%和 56.4%。

病例数均超过 20 例；青海省和西藏自治区为第五等级，累计病例数小于 20。

¹ 2020 年 1 月 23 日，浙江省率先启动突发公共卫生事件一级响应，之后各地纷纷跟进，西藏自治区 1 月 29 日启动一级响应，是最后启动一级响应的省级政府。

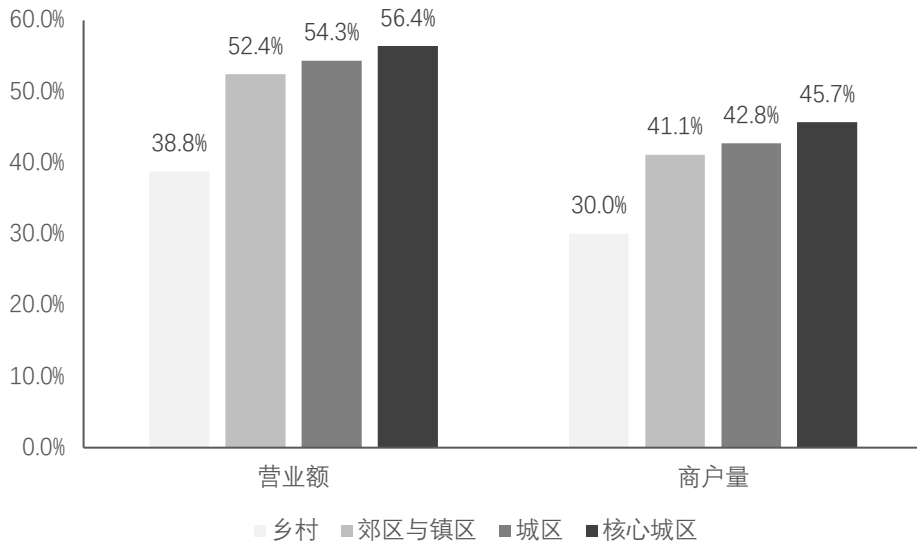


图 5：城市与农村的疫情冲击程度

商户特征的异质性。一个沿街商铺往往就是一个家庭的主要收入来源，不能开市营业，等于断掉了这个家庭的生计，时间一长，必然引发社会不稳定、金融风险等一连串不良反应。而越是脆弱的家庭，抗冲击能力可能就越弱。因此，为了进一步考察疫情对弱势群体商户经营的冲击，我们在上述分析的基础上，进一步将以下几个特征的商户单独识别出来：注册用户年龄小于 24 岁的青年商户、注册用户年龄大于 55 岁的中老年商户、女性商户，以及注册人来源于外省的商户。图 6 的结果显示，注册用户年龄 55 岁以上的中老年人商户、女性商户、外省人商户，相对于总体样本，受到了更大幅度的疫情冲击。青年人经营的商户受疫情冲击则相对较小。换言之，本次疫情对本来就弱势的线下微型商户产生了严重冲击，而且弱势群体中的弱势群体的冲击还要更加严重。

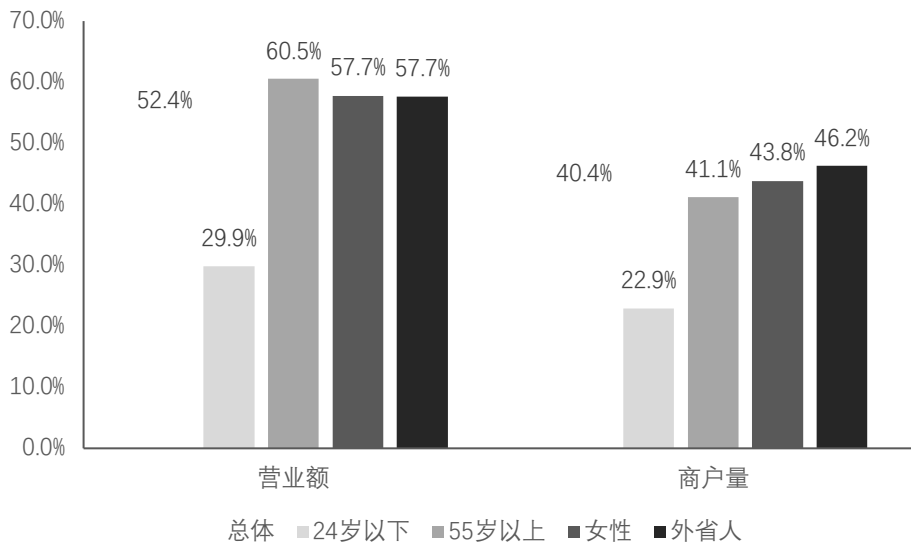


图 6：弱势群体商户与疫情冲击程度

（二）数字金融在缓解疫情中的作用

在本部分，我们分析数字金融在缓解疫情对线下微型商户冲击中的作用。之所以进行这样的分析，是因为规模极其微小的线下微型商户，基本都不在传统金融机构的服务范围，

无法依赖传统金融机构的金融支持来度过难关。但数字金融机构凭借其经营中获得的海量数据积累，通过精准放贷和其他金融手段，也许可以缓解疫情对线下微型商户的冲击。为了严谨考察这一点，我们将上述 26.7 万个泰森多边形数据与“北京大学数字普惠金融指数”的底层指标（郭峰等，2020）进行了匹配。对于数字普惠金融指数，我们使用的是 2018 年 2800 多个县域（县、县级市、市辖区）的底层指标。具体的研究设计为：

$$weakened_{pci} = \beta_0 + \beta_1 * fintech_c + \beta_2 * X_i + \lambda_p + \varepsilon_{pci}$$

其中，小标 i 代表泰森多边形， c 代表县， p 则代表省。 $weakened_{pci}$ 代表 p 省 c 县 i 多边形疫情冲击的缓解程度，具体为 $\log(\text{真实结果}) - \log(\text{反事实结果})$ 。 $fintech_c$ 代表 c 县数字金融的发展水平，度量变量为数字普惠金融指数的底层指标¹。具体而言，某项指标的形式为 $\frac{\log X - \log X^L}{\log X^H - \log X^L} * 100$ ，其中， x 为该业务变量的具体值（如人均信贷量等），而 X^H 与 X^L 都是与 2018 年数据无关的常数值²，故而我们此处使用底层指标可以清晰地度量某项底层指标取 \log 后乘以 100，这也是读者在后续进行参数解读时需要注意的。 X_i 则分别代表多边形层面的其他控制变量，对此在上文已有详细的介绍。 ε_{pci} 为残差项。 β_1 是我们关注的主要系数，其反映了数字金融对疫情冲击的缓解程度，具体含义为保持其他条件不变，当前业务每增长 1%，疫情的冲击幅度将减少多少。

表 3 报告了关于营业额的总体回归结果以及疫情不同等级地区的分样本回归结果³，在这其中，我们重点关注的是前三行变量，即人均信贷量、花呗在支付中的占比与人均保险量。而其他的数字金融变量（如人均余额宝、人均支付宝支付量等）指标，则与多边形的夜晚灯光亮度、人口、POI 信息等一起作为控制变量使用。这样进行区分的理由是，信贷、信用消费（花呗）、保险三项指标具有较强的政策含义，也在逻辑上与线下微型商户的运营、营收结合更为紧密。而人均支付量、余额宝量则更像是对于这一地区当前发展情况的度量，通过相关政策或者商业活动，政府或公司可以较为及时地对于信贷、信用消费、保险额等变量进行调整，但是人均支付量、支付宝渗透率等指标，则与兴趣点数量、夜晚灯光强度类似，很难在短时间人为改变，在回归中更多地是对多边形经济情况的控制。在后文中，我们只汇报前三项变量的回归系数，其余变量一并记为控制变量。

表 3 第（1）列的回归结果表明，保持其他变量不变，基于数字技术释放的信贷每增加 1%，疫情对于线下微型商户的营业额的冲击平均便减小 2.57%。为了对数字金融的这个缓解作用有更直观的感受，我们比较了基于数字技术精准释放的信贷发展水平较好的代表性城市杭州市与全国平均发展水平之间的发展差异，据此可以得到如下结论：如果一个地区的基于数字技术精准释放的信贷发展水平可以从全国平均水平上升到杭州市的发展水平，则可以使得疫情冲击幅度下降约 51%，即疫情的冲击减弱约一半。表 3 其他几列是按照疫情严重程度的分地区回归结果，从中可以看出从疫情较轻的第四等级到疫情最为严重的第一等级，数字信贷的缓解程度分别为：0.82%、3.04%、2.31%和 1.73%。即数字信贷的缓解作用是一种全国、整体上的缓解，特别是对于经济、人口占比很高的第二等级、第三等级区域。这些

¹ 具体变量及其含义可以参见郭峰等（2020）的表 1。

² 指数编制时，基期年份该业务量的最大值和最小值。

³ 由于西藏和青海两省的数据太少，回归结果不够稳健，因此没有报告这两个省份（疫情第五等级地区）的回归结果。

区域的尽快恢复，不仅是取得战“疫”工作的有力保障，更是涉及国计民生的大事。同时我们发现，信用消费（花呗占比）与人均保险额两项指标，也都起到了统计意义上的显著的正向作用，即它们也对于缓解冲击有所帮助。但是在影响程度上，则不及信贷。一个可能的解释是，这两项指标相对而言，发生作用机制相对间接、迟滞。

表 3：总体回归结果：营业额

样本范围	全国	四级疫情	三级疫情	二级疫情	一级疫情
被解释变量	营业额	营业额	营业额	营业额	营业额
人均信贷量	0.0257*** (0.00437)	0.00819*** (0.00072)	0.0304*** (0.00594)	0.0231*** (0.00676)	0.0173*** (0.0033)
支付中花呗的占比	0.00134*** (0.000695)	0.00041* (0.000225)	0.00175* (0.000989)	0.00314*** (0.00117)	1.00E-05 (0.00242)
人均保险额	0.00331*** (0.000713)	0.00178* (0.00106)	0.00051*** (0.000115)	0.00703*** (0.000843)	-0.00332 (0.00288)
信贷用户占比	0.000285 (0.00347)	-0.00598 (0.00422)	-0.00764 (0.00474)	0.00831 (0.00625)	0.0135 (0.0117)
信贷利率	-0.00113 (0.00163)	0.00366 (0.00390)	0.000376 (0.00230)	-0.00165 (0.00141)	-0.00809 (0.00594)
信用生活用户占比	-0.00393 (0.00388)	0.00117 (0.00375)	0.000255 (0.00276)	0.00129 (0.00406)	-0.0114* (0.00607)
人均信用生活调用次数	0.00125* (0.000747)	-0.000402 (0.00108)	-0.00178* (0.000938)	0.00305*** (0.000813)	-0.00250 (0.00158)
保险用户占比	0.00329** (0.00142)	-0.00183 (0.00154)	-0.000569 (0.00137)	0.00698*** (0.00256)	-0.000592 (0.00673)
支付宝用户占比	-0.000561 (0.000669)	0.00182 (0.00174)	-0.00153 (0.00102)	-0.00208 (0.00221)	-0.000763 (0.00260)
人均支付金额	-0.00378* (0.00204)	-0.00162 (0.00320)	-0.000530 (0.00283)	-0.0129*** (0.00404)	0.0126 (0.00920)
人均余额宝	-0.00557*** (0.00113)	-0.00515*** (0.00161)	-0.00514*** (0.00186)	-0.00227 (0.00240)	-0.00773 (0.00478)
夜晚灯光亮度加和	-5.46e-07 (7.28e-07)	5.32e-05** (2.42e-05)	-4.94e-07 (5.63e-07)	-3.47e-05 (2.71e-05)	-2.15e-05 (2.12e-05)
人口（取 log）	0.0284*** (0.00203)	0.0247*** (0.00345)	0.0278*** (0.00322)	0.0266*** (0.00301)	0.0145** (0.00595)
到区县中心的距离（取 log）	0.0342*** (0.00261)	0.0333*** (0.00504)	0.0353*** (0.00377)	0.0362*** (0.00562)	0.0221*** (0.00500)
到地市中心的距离（取 log）	0.00903** (0.00354)	0.0105** (0.00512)	0.00786 (0.00575)	-0.00186 (0.00470)	0.0259*** (0.00624)
海拔（米）	5.05e-06 (1.53e-05)	2.90e-05* (1.66e-05)	-3.58e-05* (1.82e-05)	0.000318*** (7.26e-05)	-5.17e-05 (6.93e-05)
省份固定效应	控制	控制	控制	控制	控制
POI 信息点	控制	控制	控制	控制	控制
N	266793	49224	123243	81589	9688
Adj-R ²	0.12	0.113	0.121	0.127	0.092

注：①POI 信息点包含 18 个一级分类；②（）内数值为回归系数在省级层面聚类稳健标准误；③*、

和*分别表示 10%、5%和 1%的显著性水平。下同。

表 4 则汇报了以活跃商户量为被解释变量的回归结果。从中我们发现数字信贷减弱疫情对活跃商户量的冲击也有着非常明显的作用。从全国而言,保持其他变量不变,人均信贷量每上升 1%,疫情对于活跃商户量的冲击便减小 1.46%。如果一个地区的基于数字技术精准释放的信贷发展水平可以从全国平均水平上升到杭州市的发展水平,则可以使得疫情对活跃商户数的冲击幅度下降约 29%。同时,值得特别注意的便是,数字信贷对于疫情最为严重的湖北省,起到了相对而言最好的保护效果。信用消费与保险的作用有所降低,但仍呈现正面的作用。

表 4: 总体回归结果: 活跃商户数量

样本范围	全国	四级疫情	三级疫情	二级疫情	一级疫情
被解释变量	商户量	商户量	商户量	商户量	商户量
人均信贷量	0.0146*** (0.00249)	0.00207 (0.00271)	0.0165*** (0.00249)	0.0129*** (0.00424)	0.0240*** (0.00425)
支付中花呗的占比	0.00175*** (0.000446)	0.000176 (0.000496)	0.000925** (0.000440)	0.00309*** (0.000707)	0.00101 (0.000799)
人均保险额	0.000408 (0.000379)	0.000729 (0.000570)	0.00116*** (0.000419)	0.000387 (0.000542)	0.00292*** (0.000706)
控制变量	控制	控制	控制	控制	控制
省份固定效应	控制	控制	控制	控制	控制
N	266793	49224	123243	81589	9688
Adj-R ²	0.289	0.241	0.283	0.398	0.455

注: 本表中的控制变量包括表 3 中列示的所有控制以及 POI 兴趣点变量。下同。

表 5 分析了在不同城乡区域内数字金融对疫情冲击的缓解作用。从结果而言,在整体上保持对于疫情缓解正向作用的大前提之下,数字金融在不同城乡区域上表现了有趣的结构性差异。对于核心城区的线下微型商户而言,帮助最大的是信贷;对于郊区与镇区,以及城区而言,信用消费起到了显著高于整体值的帮助效果;而对于农村而言,更为有效的帮助来自保险。这一结构上的差异与我们对于不同区域线下微型商户在细分行业、经营策略、面向人群的上面的认知是一致的。故而在疫情稳定与结束后的恢复发展过程中,城乡之间的细微差异应当引起足够的重视,不仅仅是在防疫中要做到社区与村执行过程中的因地制宜,在之后的恢复生产生活中,也应避免一刀切的情况发生。

表 5: 分城乡回归结果: 营业额

样本范围	乡村	郊区与镇区	城区	核心城区
被解释变量	营业额	营业额	营业额	营业额
人均信贷量	0.0100*** (0.00263)	0.00281*** (0.00113)	0.0171*** (0.00504)	0.0223*** (0.00558)
支付中花呗的占比	0.00101 (0.000945)	0.00242*** (0.000895)	0.00284*** (0.000775)	0.00108*** (0.000317)
人均保险额	0.00208* (0.00107)	0.00109 (0.000895)	0.00033 (0.000704)	0.000743*** -0.000109
控制变量	控制	控制	控制	控制
省份固定效应	控制	控制	控制	控制
N	78252	33541	45932	99957
Adj-R ²	0.146	0.065	0.063	0.071

而在活跃商户数量的分析上，表 5 所体现的结构性差异在表 6 上基本保留，特别是数字保险对于农村线下微型商户的保护作用。在某种程度上讲，农村的个体经营者由于其场所、商品种类以及服务人群的相对固定，自身对于风险的应对意识相对于城区经营者可能较差，对于数字信贷等需要更深入技术知识与金融认知的产品接受度也相对较低，这种情景下保险可能是相对而言最为有效、稳妥的支持方式。

表 6：分城乡回归结果：活跃商户数量

样本范围	乡村	郊区与镇区	城区	核心城区
被解释变量	商户量	商户量	商户量	商户量
人均信贷量	0.0163*** (0.00227)	0.0171*** (0.00293)	0.0180*** (0.00296)	0.0136*** (0.00348)
支付中花呗的占比	0.00182*** (0.000354)	0.00205*** (0.000509)	0.00241*** (0.000499)	0.00184*** (0.000695)
人均保险额	0.00112*** (0.000374)	-0.000158 (0.000459)	4.40e-05 (0.000451)	0.000603 (0.000580)
控制变量	控制	控制	控制	控制
省份固定效应	控制	控制	控制	控制
N	78252	33541	45932	99957
Adj-R ²	0.246	0.186	0.184	0.168

表 7 展现了我们对于不同业主身份的线下微型商户进行回归的结果。一方面，从中我们发现数字信贷对于 55 岁以上人群具有更强的帮助，这是一件好事情，毕竟如上文的发现，他们是受到冲击相对更重的弱势群体；而从另一个角度而言，尤其是外省人相对而言的较弱保护，可能是信贷发放过程中的风险评估有所倾斜，当然在没有更多证据与内部信息的情况下，这只能是一个猜测。而我们能够看到另一点有趣现象是，24 岁以下年轻人与外省人从保险中获得了最为显著的帮助，这可能与年轻人在保险意识、数字产品理解上有较高的水平相关。而外省人与本地人士相比对于风险也有更高的警觉。

表 7：分业主身份回归结果：营业额

样本范围	24 岁及以下	55 岁及以上	女性	外省人
被解释变量	营业额	营业额	营业额	营业额
人均信贷量	0.0196*** (0.00743)	0.0330*** (0.00737)	0.0196*** (0.00421)	0.00723** (0.00240)
支付中花呗的占比	0.00134*** (0.000427)	0.00244* (0.00134)	0.00072*** (0.000277)	0.00168*** (0.000427)
人均保险额	0.00829*** (0.00143)	0.00164 (0.00162)	0.0022 (0.00678)	0.00456*** (0.000466)
控制变量	控制	控制	控制	控制
省份固定效应	控制	控制	控制	控制
N	266793	266793	266793	266793
Adj-R ²	0.096	0.088	0.124	0.116

对于活跃商户量的分析，表 8 展示的回归结果基本保持了表 7 所得出的关键结论，我们注意到外省人仍然是受到信贷支持较低的一部分群体。中国的发展，特别是大型城市的繁荣，离不开这些远离家乡打拼的人，疫情对于他们的短期冲击不言自喻，而在长期上不会产生结构性、根本性转变，还需要未来基于更长时期数据的中长期评估。

表 8：分业主身份回归结果：活跃商户数量

样本范围	24 岁及以下	55 岁及以上	女性	外省人
被解释变量	商户量	商户量	商户量	商户量
人均信贷量	0.00851*** (0.00256)	0.0123*** (0.00221)	0.0142*** (0.00235)	0.00605** (0.00240)
支付中花呗的占比	0.00151*** (0.000456)	0.00110** (0.000435)	0.00158*** (0.000416)	-8.83e-05 (0.000427)
人均保险额	0.00102** (0.000515)	0.000333 (0.000434)	-0.000352 (0.000350)	-0.000736 (0.000466)
控制变量	控制	控制	控制	控制
省份固定效应	控制	控制	控制	控制
N	266793	266793	266793	266793
Adj-R ²	0.075	0.136	0.247	0.172

五、结论与启示

经过全国界的共同努力，付出了极大牺牲，到 2020 年 3 月底，中国本土新冠肺炎疫情传播已基本阻断，但与此同时，疫情却又在其他国家蔓延开来。中国为防控疫情所做出的努力，有哪些是其他国家可以借鉴的，包括未来本次疫情可能的反弹或者其他可能的突发公共卫生事件，应该从本轮疫情防控中汲取什么经验和教训，本文的研究对此有重要的启发：

首先，疫情对线下微型商户确实造成了一定的短期冲击，但这种冲击是可承受的。根据我们的估算，在重点疫区湖北省，线下微型商户活跃商户量和营业额，较根据机器学习算法推算出的如果没有疫情发生的“反事实结果”，分别下降 59.3%和 69.7%；而在全国其他地区，活跃商户量和营业额则分别下降约 40%和 50%。并且，湖北之外，疫情对线下微型商户的冲击，并不因疫情不同程度而不同，这是因为疫情初期各地采取的应对措施高度一致。即疫情对经济的影响，并不是疫情本身的影响，而是防控疫情的政策带来的，这是主动、可控的。随着疫情的缓解，中国政府采取了按照疫情风险等级分批、有序地推动复工复产，疫情的冲击自然也就随之消解。从而做到了该管起来就能够迅速地管起来，该放开又能够有序地放开。

其次，疫情的冲击是不均衡的，因此政府救助政策也应该做到有的放矢。主动而为的封城闭户政策，虽然不会对经济产生全面的、长远的冲击，但确实会对部分行业造成一定的影响。例如，本文所研究的对象线下微型商户就是遭遇冲击最严重的业态之一，因此相关救助政策也应该有针对性地扶持这些受冲击对象，而不是大水漫灌地撒钱和救市。在救助线下微型商户上，中国政府采取了一些针对性的措施，例如减免中小企业和个体工商户的税负；最初是微型企业缓缴社保，随后改为减免部分社保；疫情期间，全国高速公路全部免费通行，切实降低商户物流成本；要求国有房地产企业和物业公司减免商户租金，鼓励其他民营企业同步跟进；要求电、水、通讯等国有企业对个体工商户做到无论是否欠费，都不得停电、停水或停机；支持金融机构加大对微型企业的放贷力度。所有这些有针对性的政策，都对缓解疫情冲击，提高线下微型商户抗风险能力，有非常大的帮助。

再次，将资金直接推送到离个体经营者最近的金融机构，同时充分发挥数字金融的作用。线下微型商户不容易获得金融服务，主要还是因为金融机构缺少渠道与抓手。一些数字金融机构利用移动支付等数字技术，解决了获客难与风控难的问题，服务了大量的线上、线下线下微型商户。但数字金融机构往往缺乏资金，因此，在突发事件发生时，特别是线下金融机构业务面临冻结状态时，政府应该通过央行再贷款、专项债及资产证券化等工具把资金推送到直接服务线下微型商户的网络银行与其它金融机构，同时鼓励它们通过贷款展期、息费减

免等手段帮助线下微型商户维持经营，共度难关。

最后，通过发展数字经济和数字技术，将微型企业纳入统计分析框架。线下微型商户是中国经济的重要组成部分，对于消费、就业甚至社会稳定都是至关重要的。但问题是，很多线下微型商户并未计入官方数据，政府要支持它们也缺乏抓手。而数字金融在缓解疫情冲击中的作用提供了一个重要的启示，就是政府可以通过数字技术把近亿的线下微型商户的信息集中到一起，既能及时了解这个行业的动态，同时也为政策调控特别是政策扶持提供一个重要的抓手。而且，中国数字经济的发展经验，对于广大人口密度大的发展中国家，也有非常大的借鉴价值。

参考文献

- [1] 陈硕、王宣艺，《机器学习在社会科学中的应用：回顾及展望》，复旦大学经济学院工作论文，2018。
- [2] 傅秋子、黄益平，《数字金融对农村金融需求的异质性影响：来自数字普惠金融指数与中国家庭金融调查的证据》，《金融研究》，2018年第11期，第68-84页。
- [3] 郭峰、王靖一、王芳、孔涛、张勋、程志云，《测度中国数字普惠金融发展：指数编制与空间特征》，《经济学季刊》，2020年第19卷第2期。
- [4] 黄益平、黄卓，《中国的数字金融发展：现在与未来》，《经济学季刊》，2018年第1卷第4期，第205-218页。
- [5] 孔祖根，《系统性风险下的差别待遇：丽水市个体工商户金融服务个案研究》，《金融研究》，2002年第7期，第130-134页。
- [6] 蒋晨琛、霍宏涛、刘克俭、冯琦，《空间数据驱动的B市主城区犯罪时空分布及其影响因素分析》，《科学技术与工程》，2019年第26期，第1671-1814页。
- [7] 金泉、李辉文、苏庆新、马文杰，《新冠肺炎疫情突发事件对中小微企业企业家信心的影响与对策——基于中国企业创新创业调查(ESIEC)数据库的分析》，《产业经济评论》，2020年第2期，第47-55页。
- [8] 李扬，《“金融服务实体经济”辨》，《经济研究》，2017年第6期，第4-16页。
- [9] 李广子、熊德华、刘力，《中小银行发展如何影响中小企业融资？——兼析产生影响的多重中介效应》，《金融研究》，2016年第12期，第78-94页。
- [10] 刘克崧，《建设中国草根金融体系促进草根经济发展和城乡就业——兼论我国小企业融资难的破解》，《管理世界》，2009年第11期，第1-5页。
- [11] 钱雪松、唐英伦、方胜，《担保物权制度改革降低了企业债务融资成本吗？——来自中国《物权法》自然试验的经验证据》，《金融研究》，2019年第7期，第115-134页。
- [12] 邱晗、黄益平、纪洋，《金融科技对银行行为的影响：基于互联网理财的视角》，《金融研究》，2018年第11期，第17-30页。
- [13] 邱晗、纪洋、黄益平、朱光耀、王芳，《房价、消费信贷与居民消费——来自蚂蚁金服抽样数据的证据》，北京大学数字金融研究中心工作论文，2020。

- [14] 宋华、卢强,《什么样的中小企业能够供应链金融中获益?——基于网络和能力的视角》,《管理世界》,2017年第6期,第104-121页。
- [15] 田国强、赵旭霞,《金融体系效率与地方政府债务的联动影响——民企融资难融资贵的一个双重分析视角》,《经济研究》,2019年第8期,第4-20页。
- [16] 王靖一、郭峰、李勇国,《中国线下微型商户总量估算及影响因素度量:来自金融科技公司的证据》,北京大学数字金融研究中心工作论文,2020。
- [17] 谢绚丽、沈艳、张浩星、郭峰,《数字金融能促进创业吗:来自中国的证据》,《经济学季刊》,2018年第17卷第4期,第1557-1580页。
- [18] 徐光、赵茜、王宇光,《定向支持政策能缓解民营企业的融资约束吗?——基于民营企业债务融资支持工具政策的研究》,《金融研究》,2019年第12期,第187-206页。
- [19] 易行健、周利,《数字普惠金融发展是否显著影响了居民消费——来自中国家庭的微观证据》,《金融研究》,2018年第11期,第47-67页。
- [20] 张勋、万广华、张佳佳、何宗樾,《数字经济、普惠金融与包容性增长》,《经济研究》,2019年第8期,第71-86页。
- [21] 中国人民银行金融消费者权益保护局,《中国普惠金融指标分析报告(2018年)》,2019年10月。
- [22] 朱武祥、张平、李鹏飞、王子阳,《疫情冲击下中小微企业困境与政策效率提升——基于两次全国问卷调查的分析》,《管理世界》,2020年第4期,第13-25页。
- [23] 周振鹤,《中国历史政治地理十六讲》,中华书局,2013年12月。
- [24] AfDB, “Recognizing Africa’s Informal Sector. In: AfDB, 27 March.<<http://www.afdb.org/en/blogs/afdb-championing-inclusive-growth-across-Africa/post/recognising-informal-sector-11645>>Accessed 19.05.16, 2013.
- [25] Alatas, V., and Newhouse, D., “Indonesia Jobs Report: Toward Better Jobs and Security for All”, Washington, DC: Main Report, World Bank, 2010.
- [26] Atkeson, A. G., “What will be the Economic Impact of COVID-19 in the US? Rough Estimates of Disease Scenarios”, NBER Working Paper No. w26867, 2020, Available at <http://www.nber.org/papers/w26867>.
- [27] Babbitt, L. G., Brown, D., and Mazaheri, N., “Gender, Entrepreneurship, and the Formal Informal Dilemma: Evidence from Indonesia”, *World Development*, 2015, 72, 163-174.
- [28] Björkegren, D., and Grissen, D., “Behavior Revealed in Mobile Phone Usage Predicts Loan Repayment”, *The World Bank Economic Review*, lhz006, 2019.
- [29] Blackburn, K., Bose, N., and Capasso, S., “Tax Evasion, the Underground Economy and Financial Development”, *Journal of Economic Behavior & Organization*, 2012, 83(2), 243-253.
- [30] Barro, R. J., Ursúa, J. F., and Wang, J., “The Coronavirus and the Great Influenza Pandemic: Lessons from the ‘Spanish Flu’ for the Coronavirus’s Potential Effects on Mortality and

- Economic Activity ’ , NBER Working Paper No. w26866, 2020, Available at <http://www.nber.org/papers/w26866>.
- [31] Capasso, S., and Jappelli, T., “Financial Development and the Underground Economy”, *Journal of Development Economics*, 2013, 101, 167-178.
- [32] Correia, S., Luck, S., and Verner, E., “Pandemics Depress the Economy, Public Health Interventions Do Not: Evidence from the 1918 Flu” , Working Paper, 2020, Available at SSRN: <https://ssrn.com/abstract=3561560>.
- [33] Campante, F. R., Depetris-Chauvin, E., Durante, R., “The Virus of Fear: The Political Impact of Ebola in the U.S.”, NBER Working Paper No.w26897, 2020, Available at <http://www.nber.org/papers/w26897>.
- [34] Chen, L., “From Fintech to Finlife: The case of Fintech Development in China”, *China Economic Journal*, 2016, 9(3), 225-239.
- [35] Chen, T., and Guestrin, C., “Xgboost: A Scalable Tree Boosting System”, In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016,785-794.
- [36] Eichenbaum, M., Rebelo, T., and Trabandt, M., “The Macroeconomics of Epidemics” , NBER Working Paper No. w26882, 2020, Available at <http://www.nber.org/papers/w26882>.
- [37] Fang, H., Wang, L., Yang, Y., “Human Mobility Restrictions and the Spread of the Novel Coronavirus (2019-nCoV) in China”, NBER Working Paper No. w26906, 2020, Available at <http://www.nber.org/papers/w26906>.
- [38] Friedman, J. H., “Greedy Function Approximation: a Gradient Boosting Machine”, *Annals of statistics*, 2001, 29(5), 1189-1232.
- [39] Gambacorta, L., Huang, Y., Qiu H., and Wang, J., “How do Machine Learning and Non-traditional Data Affect Credit scoring? New Evidence from a Chinese Fintech Firm”, BIS Working Papers No 834, 2019.
- [40] Goldfarb, A., and Tucker, C., “Digital Economics”, *Journal of Economic Literature*, 2019, 57(1), 3-42.
- [41] Guo, J., and Zhang, Z., “Does Renaming Promote Economic Development? New Evidence from a City-renaming Reform Experiment in China, *China Economic Review*, 2019, 57, 101344.
- [42] Harris, J.R., and Todaro, M.P., “Migration, Unemployment and Development: A Two Sector Analysis”, *American Economic Review*, 1970, 60(1), 126-142
- [43] Harriss-White, B., “Work and Wellbeing in Informal Economies: The Regulative Roles of Institutions of Identity and the State”, *World Development*, 2010, 38, 170-183.
- [44] Jia, R., Khadka, A., and Kim, I., “Traffic Crash Analysis with Point-of-interest Spatial Clustering”, *Accident Analysis & Prevention*, 2018, 121, 223-230.

- [45] Kleinjans, K., J., and Van Soest, A., “Rounding, Focal Point Answers and Nonresponse To Subjective Probability Questions”, *Journal of Applied Econometrics*, 2014, 29(4), 567-585.
- [46] Korinek, A., Mistiaen, J., A., and Ravallion, M., “An Econometric Method of Correcting for Unit Nonresponse Bias in Surveys”, *Journal of Econometrics*, 2007, 136(1), 213-235.
- [47] Kumar, A., and Liang, C., “Credit Constraints and GDP Growth: Evidence from a Natural Experiment”, *Economics Letters*, 2019, 181, 190–194.
- [48] La Porta, R., and Shleifer, A., “The Unofficial Economy and Economic Development”, *Brookings Papers on Economic Activity*, 2008, 275.
- [49] Maloney, W. F., “Informality Revisited”, *World Development*, 2004, 32 (7), 1159-1178.
- [50] Neyman, J., “Sur Les Applications de La Théorie Des Probabilités Aux Experiences Agricoles: Essai Des Principes”, *Roczniki Nauk Rolniczych*, 1923,10, 1-51.
- [51] Okabe, A., Boots, B., Sugihara, K., and Chiu, S. N., “Spatial Tessellations: Concepts and Applications of Voronoi Diagrams”, John Wiley & Sons, 2009, 501.
- [52] Qiu, Y., Chen, X.; Shi, W., “Impacts of Social and Economic Factors on the Transmission of Coronavirus Disease 2019 (COVID-19) in China”, GLO Discussion Paper, No. 494, Global Labor Organization (GLO), Essen, 2020.
- [53] Restrepo-Echavarria, P., “Macroeconomic Volatility: The Role of the Informal Economy”, *European Economic Review*, 2014, 70, 454-469.
- [54] Sen, P., and Kolli, R., “Delhi Group on Informal Sector – Contribution and Present Status”, In: Paper Prepared for the Special IARIW-SAIM Conference on “Measuring the Informal Economy in Developing Countries”. Kathmandu, Nepal, 2009.
- [55] UBOS, *2012 Statistical Abstract*, Kampala: Uganda Bureau of Statistics, 2012.
- [56] Zhan, C., Tse, C., Fu, Y., Lai, Z. and Zhang, H., “Modelling and Prediction of the 2019 Coronavirus Disease Spreading in China Incorporating Human Migration Data, Working paper, 2020, Available at <https://www.medrxiv.org/content/10.1101/2020.02.18.20024570v1>.

Quantitative Estimation of the Impact of COVID-19 Epidemic on Offline Micro Business in China

—And an Investigation of the Digital Finance's Value on Mitigating Shock

WANG Jingyi^{1,2}, GUO Feng^{3,2}, LI Yongguo⁴

(¹Central University of Finance and Economics, ²Peking University, ³Shanghai University of Finance and Economics, ⁴Ant Financial Services Group)

Abstract: In early 2020, the COVID-19 epidemic, which has turn into the most serious public health emergency since the founding of People's Republic of China, has caused serious impact on the national economy, especially the offline micro business (OMB). Using the data of hundreds of millions of the payment tool- *QR merchants*, affiliation to Alipay, and the method of machine learning, this study calculates the "counterfactual results" that the OMB should have after the Spring Festival in 2020, if there is no outbreak of the COVID-19 epidemic. The results show that, during January 31, 2020 to February 13, 2020, comparing with the estimated "counterfactual result", the number of active OMB owners and turnover of their decrease by about 40.4% and 52.4%, respectively. Further analysis shows that the development of digital finance has played a certain role in alleviating the impact of the COVID-19 epidemic on OMBs. As far as the research conclusion and methodology are concerned, this study has a very good reference value for other countries to deal with this epidemic and for our country to deal with other public health emergencies in the future.

Keywords: COVID-19, Offline Micro Business, Digital Finance, Machine Learning