

基于大数据的小微企业信用风险测度研究

■ 方匡南

新冠肺炎疫情发生以来,餐饮、旅游等大量小微企业抗风险能力弱,面临诸多困难,现金流吃紧、短期内贷款偿还能力下降、复工复产伴随更大的融资需求,成为社会关注的焦点。要帮助小微企业渡过难关,最核心是帮助小微企业解决融资需求问题。

我国小微企业数量庞大,据第四次全国经济普查结果显示,2018年末我国小微企业(含个体工商户)数量超过9000万家,占全部企业数量的99.8%,遍布全国不同地区的各个行业,贡献了全国60%的国民生产总值和50%以上的税收,满足了80%左右的城镇就业,因此发展小微企业对我国经济高质量发展具有十分重要的意义。近年来,在政府的大力推动和金融机构的业务创新下,我国小微企业金融服务质量和水平有了较大的提升,但是小微企业融资难、融资贵问题仍然十分突出,尤其是小微企业“首次贷”难的问题,表面上看是缺钱,实质上是缺信息、缺信用,根本原因是银行和小微企业之间存在信息不对称。政府除了出台对小微企业贷款的扶持政策,更重要的在于完善我国小微企业的信息披露机制,构建有中国特色的小微企业信用体系以及开发小微企业信贷风险指标体系和统计测度方法。

随着大数据、人工智能、区块链等金融科技在金融领域的应用,金融机构可以全方位地收集小微企业多维度信息,采用机器学习等人工智能方法对其贷前风险科学评估,贷后风险及时智能监测和预警,为金融机构有效服务小微企业开拓了一条新的路径。尤其是在重大疫情下,借助于大数据、人工智能等技术的纯线上“无接触”信用贷款模式成为眼下许多小微企业的救命稻草。因此,系统地研究我国小微企业的信用风险统计测度方法以及构建我国小微企业信用体系非常重要。

一、小微企业融资难、融资贵问题的原因

中小企业贷款难问题一直制约着我国中小企业的成长和发展,成为我国经济发展的瓶颈之一,这一问题长期以来都备受经济金融理论界的关注。相对于大企业,中小企业更为信息不透明,并缺乏可抵押的资产、市场竞争力弱,

因而中小企业的融资也更加困难。国外有学者研究认为,企业处在初创和成长阶段更加依赖从传统金融机构获取资金支持。此外,也有学者研究表明,信贷缺口规模受到小微企业的行业属性、组织形式、年龄等因素影响,中小企业借贷的约束性条款主要有定期提供财务信息、投向特定项目、缩短贷款期限、要求抵押和担保等。近年来,中国政府通过提供各种优惠政策和补贴、推动银企对接、鼓励发展中小金融机构和担保机构、加强中小企业信用体系建设等措施,缓解中小企业融资难问题。但是我国小微企业融资难、融资贵的问题仍然突出。其中,信息不对称是中小企业融资困境的重要原因。国内有学者建立了一个关于信息不对称和中小企业融资问题的分析框架,认为中小企业“融资难”是由于信息不对称导致较为严重的信贷配给约束,为了减少信用风险,银行等传统金融机构倾向于减少小微企业的信贷业务。国际经验表明,基于全面准确的小微企业信用信息而开发的信用评分能有效解决小微企业与金融机构之间的信息不对称,从而提升小微企业融资的可获得性。因此,发展和创新小微企业征信服务是破解小微企业融资约束的有效方法。比如美国著名的邓白氏(D&B)信用评级公司利用其庞大的全球数据库和客户资源网络,针对中小企业特点,开发出小微企业信用风险系统解决美国的小微企业融资难问题。另外,基于互联网金融等金融科技具有边际成本递减和边际效益递增的特点,对于解决小微企业融资具有天然优势,打破了传统金融的“二八定律”。

总的来说,我国的小微企业融资难融资贵问题带有鲜明的中国特色,既有市场机制原因又有体制因素,需要对此一问题加以系统分析,找出影响我国小微企业融资的关键因素,以及如何借助大数据、人工智能等技术解决我国小微企业融资难融资贵问题。

二、大数据时代小微企业信用风险统计测度方法与存在的问题

小微企业信贷风险管理的基础是风险识别问题。传统的信用决策系统主要依赖于专家的主观判断,难以做到科学、客观。为了降低信用决策中的主观因素,基于统计模型的信贷风险测度方法越来越多地被用到决策中。信用评分系统在衡量信用风险的内部评级中占有核

心的地位,比如《巴塞尔新资本协议》鼓励银行采用内部评级法评估信用风险,即鼓励银行基于自身业务构建内部信贷风险测度模型。银行对大中型企业信贷风控技术相对比较成熟,对小微企业的信贷风控技术比较欠缺,小微企业由于其特殊性,不能简单套用大中型企业的信贷风险测度方法。另外,在实践中传统信贷风险测度模型也随着时间推移慢慢失效,预测效果变差,导致违约损失增加。信贷风险测度模型不是一成不变,会随社会经济环境、数据类型、法律政策环境、行为习惯等变化而变化。比如国外有学者基于生存分析模型考虑了宏观经济变量的变化,发现个人信用评分结果随宏观经济状态变化。

大数据及相关技术的发展,为小微企业融资服务和风险管理带来了新的契机。通过获取小微企业的经营、财务、税务、工商、水电等数据,可以全方位、多维度评估小微企业的信贷风险。但从统计建模角度来看,小微企业信用风险测度存在的问题有:(1)由于数据来源多、数据结构复杂、存在缺失值、数据非平衡等问题,对传统的信用风险测度方法带来了极大的挑战;(2)由于我国信息共享机制不完善以及随着隐私保护要求趋严,数据分散在各个数据源机器上,不能向外提供详细的原始数据,只能在数据源本机上进行分析,向外提供数据的分析结果,所以,在隐私保护机制下传统的信贷风险测度方法就不能用了;(3)贷后的风险管理一直是信贷领域的难点,基于信用风险测度技术,借助于大数据和人工智能等技术及时跟踪和判断客户的风险状况,智能监测和预警违约风险是小微企业信贷风险管理的重点。

综上所述,在多元大数据和隐私保护机制下,需要研究新的信贷风险统计测度方法,促进信贷风险统计测度模型的完善和优化,实现对小微企业信贷风险的识别、评估和贷后预警,提高风险管理的智能化水平。

三、我国小微企业信用风险测度的研究展望

(1)打破数据孤岛,构建动态更新的小微企业数据库

数据是信用风险测度的基础,除了企业直接提供的财务、管理、营销等企业内部信息外,银行可通过调研、实地考察等方式对企业信息

进行人工采集,也可利用网络爬虫、文本挖掘等技术在各新闻网站、行业网站、商业数据库中自动采集企业的外部信息,从而保证信息的全面和可靠。同时,由于市场的快速变化,小微企业的信用风险特征随时可能改变。因此,数据库要能够及时更新,为风险测度模型的构建、贷后风险的动态评估提供实时的数据保障。此外,当企业的风险状况发生变化时,需要对数据的收集方式和收集频率进行灵活的调整,为银行的风险决策提供更有效的信息支持。

(2)提前识别欺诈风险,减少信用风险测度误差

对于欺诈的客户,由于其对财务指标等有意作假,传统的小微企业信用风险测度方法可能没法识别出来,所以往往在信用评估之前先进行欺诈识别。对于欺诈风险识别方法:一方面,由于小微企业的特殊性,银行针对大中型企业的欺诈风险识别技术无法直接套用,但两个业务之间仍存在一定的相关性,因此,可通过迁移学习的方法将银行在大中型企业业务中积累的经验和数据迁移到小微企业的欺诈风险识别建模中,有效改善小微企业信贷业务标签或样本不足的问题,从而提高模型的识别能力;另一方面,在标注能力有限的情况下,可利用离群值检测方法、聚类算法和图算法等无监督算法发现异常模式,辅以有限人工干预进行识别,也可利用主动学习的方法筛选出较“难”分类的小微企业样本交给反欺诈岗位人员审核,提高标注的效率。

(3)创新统计方法,融合多源数据测度信用风险

随着大数据的发展,数据来源越来越多,针对相同或者相似的任务(或研究)能够收集到多个数据集。这些不同来源的数据,虽然是针对相同或者相似任务收集的,但是由于测量环境、测量标准或统计口径等因素的差异,不能简单将多数据集合并分析。对这种来源多样的复杂数据集如何有效地整合在一起分析是统计学和机器学习研究的新发展方向之一。在众多多源数据融合方法中,其中整合分析是最有前景的方法之一。整合分析方法起源于20世纪60年代,可以把不同来源、不同格式、不同特点的数据进行融合建模。相对于单一数据集模型,该方法整合了更多的原始信息,能解决因不同来源数据的差异而引起

的建模不稳定,在模型解释性和预测方面都具有显著优势。多源数据的整合分析,可以分为两类,一类是“变量”方向的整合,即不同数据集样本相同,而变量不同,整合不同来源的变量信息,比如信用评分时整合了贷款记录数据、网购数据、公安法院数据等;另一类是“样本”方向的整合,即样本是不同的,但是变量是一样的,整合不同来源的样本信息,比如城市和农村村民的征信数据。

(4)重视信息安全和隐私、发展联邦学习等新技术

随着2021年9月1日起施行《中华人民共和国数据安全法》和2021年11月1日起施行《中华人民共和国个人信息保护法》,我国对数据安全和个人信息保护越来越重视。在新的法律要求下,小微企业信用风险测度该如何建模?需要发展能够解决数据安全和个人信息保护问题的新技术和新方法,联邦学习是一种比较有前景的方法。联邦学习是一种分布式的机器学习技术,在此框架下,各数据源本机仅需交换并更新梯度或参数等中间计算结果就能实现联合建模,从而避免了原始信息的交互,有效地控制了隐私泄露。然而,在真实的网络环境中,模型反演攻击、模型推理攻击等恶意攻击手段层出不穷,直接提供中间计算结果仍存在敏感信息被获取的风险。如何利用差分隐私、安全多方计算、同态加密等技术防范恶意分析者对敏感信息的反演攻击,并同时保证通信效率和模型精度是研究的重心。此外,由于数据储存在不同的地区或部门,各数据集的分布往往存在差异,如何挖掘数据间的异质特征也是所研究模型需要考虑的内容。

(5)借助于大数据和人工智能技术,实时监测贷后信用风险

贷后风险监测是金融信贷的难点之一,但在大数据和人工智能时代,实时监测贷后信用风险成为可能。首先基于多源复杂的数据构建科学的小微企业贷后信用风险指标体系,然后建立贷后风险识别模型,捕捉小微企业经营风险的变化情况,实时跟踪和判断风险状况,精确预测小微企业贷后违约的时点,并利用机器学习等人工智能方法实时动态智能监测小微企业的风险变化情况,实现风险被动防御向主动识别、监测和拦截转变。

(作者单位:厦门大学经济学院)

区块链技术助力统计监督的机制与路径探析

■ 李金昌 程开明

党的十九大报告提出“完善统计体制”,党的十九届四中全会强调“发挥统计监督职能作用”,近日中央全面深化改革委员会第二十一次会议审议通过《关于更加有效发挥统计监督职能作用的意见》,充分体现了党和政府从制度层面加强统计监督的决心。区块链融合分布式架构、链式数据验证与存储、点对点网络协议、加密算法、共识算法、身份认证、智能合约、云计算等多种技术,具备去中心化、数据完整透明、可追溯等特征,特别适合应用于统计监督。利用区块链技术提升统计监督和统计保障能力,对于解决当前统计工作面临的重大问题,提高统计数据质量,提升国家治理效能意义重大。

一、再认识统计监督的重要性

统计工作是经济社会发展的基础性工作,统计数据是宏观调控和科学决策的重要依据,统计能力是国家治理能力的重要组成部分。因此,新时代加强统计监督职能作用,明晰强化统计监督的保障机制,提出切实可行的建议与举措,有利于提升统计数据质量,完善党和国家监督体系,推进国家治理体系和治理能力现代化。

强化统计监督有利于保障与提升统计数据质量。数据质量是统计工作的核心,是统计职能发挥的前提和基础,直接影响到经济社会发展的重大决策,影响到治国理政的战略判断。当前我国统计数据质量仍存在一些现实问题,少数领导或企业负责人为了个人利益、局部利益或政治荣誉,违反统计法的有关规定虚报瞒报,在统计上弄虚作假,“以权定数”“以数谋私”,严重影响统计数据的真实性准确性。

统计监督的基本目标是确保统计数据真实准确、完整及时,因此强化统计监督职能作用有利于充分利用法治、行政等手段排除非法干扰和人为因素的影响,防范和惩治统计造假、弄虚作假,防止浮夸风的蔓延,避免失真失实统计数据干扰决策判断,为提高统计数据质量营造良好的内外环境。可见,统计监督是提高统计数据质量、保障政府统计权威性的有力武器。

强化统计监督有利于完善党和国家监督体系。党的十九届四中全会将统计监督纳入党和国家监督体系,凸显出统计监督作为国之重器在党和国家监督体系中不可替代的重要作用。十九届中央纪委四次全会提出,促进党内监督与人大监督、民主监督、行政监督、司法监

督、审计监督、财会监督、统计监督、群众监督和舆论监督贯通融合、协调协同。

当前统计监督内涵的深化、内容的拓展,充分体现出新时代完善党和国家监督体系的制度自觉、制度自信和制度创新。发挥统计监督职能作用有利于把统计监督与党内监督、国家监察、群众监督结合起来,同法律监督、审计监督、司法监督、舆论监督等协调协同,构造全方位的国家监督体系,形成监督合力,推动我国制度优势更好地转化为治理效能。

强化统计监督有利于深化与推进国家治理体系和治理能力现代化。随着中国特色社会主义进入新时代,作为推进国家治理的重要工具,统计工作在经济社会发展中的“晴雨表”“指示器”“风向标”功能更加凸显,对于各级党政领导准确分析研判形势、审时度势科学决策的作用更为彰显。数据驱动型决策与管理日益成为政府治理的主要方式,准确判断经济形势、精准调控宏观政策、制定改革发展规划、推进基本公共服务等,都离不开真实准确、完整及时的统计数据和统计监督。

充分发挥统计监督职能作用,健全“用数据说话、用数据管理、用数据决策、用数据创新”的治理方式,是各级政府科学管理、精准调控的重要手段,有利于强化统计对经济监测预警能力,优化统计对宏观决策的服务水平,大力推进国家治理体系和治理能力现代化。

二、区块链推动统计监督的机制及路径

(一)区块链保证统计数据质量,减轻统计监督工作量

区块链将数据按时间进行分块,通过链的方式组织起来,保证数据的连续性和可追溯性;借助于链条的前后块哈希值及多副本机制,数据一旦加入链中就难以被篡改,保证了数据的完整和真实。区块链还利用密码学知识对数据进行加密处理,交易双方都通过虚拟地址的方式进行加密,数据交换、数据更新都通过虚拟地址的方式进行,保证了数据的隐私性。在统计系统中引入区块链技术,统计数据质量能够得到有效保障,为统计监督创造便利条件。

将区块链技术应用于日常统计调查或大型普查中,当企业上报数据或者住户记账数据被录入后,新增数据将实时更新到整个区块链网络之中;记录一旦生成,各级调查单位只能依据权限查询、汇总相关数据,无论哪个节点都无法对数据进行修改或删除操作,确保统计数据的唯一性和真实性。

区块链网络内各节点都存储数据,数据不会被篡改或丢失,即使存在通信故障和蓄意攻击等问题,仍能保证数据存储的准确性。运用区块链存储技术把统计调查系统接入的存储节点池化,建成一个规模巨大的统一、共享存储池,有利于数据存储安全、查询和监督。去中心化分布式存储使数据存储到成千上万个节点上,有效避免单点故障带来的负面影响,使统计从中心化向分布式转变,保障数据质量的同时使统计监督更为方便。

基于区块链技术构建一套区块链统计系统,给每个企业分配一定数量的虚拟地址,企业将自己的经营活动和交易记录数据通过区块链系统进行交互,企业之间不知道区块中每笔交易数据的含义,但统计人员可从中提取相关数据用于校验逻辑合理性和完整性,且区块链系统中数据的真实性和准确性由共识机制等手段加以保证,大大减少统计监督的复核验证工作量。

(二)区块链促进多源数据融通共享,提升协同监督效果

利用区块链去中心化、分布式账本、信任透明等优势能够整合各个政府部门的数据资源,有利于实现数据共享互通。基于区块链技术在不同利益主体之间构建去中心化分布式数据库系统,数据通过全网快速广播至各个主体,保证共享互通数据的真实性和及时性。区块链还能够加强政府部门数据的交换和融合,对各部门的数据进行技术集成,丰富数据内容,扩大数据范围,推动利用多源数据有效开展统计监督。

数据的开放共享有利于协同监督,除了各方的私有信息被加密外,区块链数据对接人的所有用户开放,任何用户都可以通过公开接口查询区块链数据和开发相关应用,系统信息高度透明。通过区块链可以打破不同政府部门之间的信息孤岛,促进不同部门数据共享,强化统计监督与其他监督的协同。

传统统计数据采集从乡镇到县、市、省,直至国家存在多个层级,也涉及公安、房管、税务等各个部门,数据口径、人为干扰等因素都影响着数据质量。利用区块链技术将各个层级、不同行政部门的数据实现共享互通,全程监控每一个在链内流动的数据,各个政府部门数据平台相互校验,保证链内流动的数据高透明度、高追溯性和高真实性,推动形成一个具有共识共建共享、互联互通互动的统计监督体系。

针对不同来源数据的类型和标准不一致等

问题,基于区块链和智能合约制定统一的数据类型和标准规则,智能合约被存储和同步在区块链各个节点,区块链根据智能合约上的代码自动执行验证,执行过程公开透明且执行结果可检验,能够提高多源数据的共享效率。运用区块链技术的信息管理平台实现政府部门信息共享,及时了解最新监督进度,各部门各司其职,协同工作开展多种形式的监督,减少政府部门的重复监督,增强监督效果。

(三)区块链实现数据可追溯溯源,提高统计监督针对性

区块链具有不可篡改的时间戳,强化数据的可追溯性,有助于提升统计监督在经济社会信息工作中的职能作用。利用区块链的去中心性和不可篡改性,记录数据的获取和共享情况,可以实施追溯溯源,并结合策略承诺、违反检测和隐私审计,以溯源问责的方式保护隐私,为评估数据质量和开展统计监督提供良好的技术支持。

在统计监督中利用区块链技术,有助于开展数据跟踪和防止统计欺诈。区块链作为去中心化分布式数据库,支持数据的历史状态查询,用以确认当前数据状态是否正确,有效保证数据的完整性和可追溯性。如果在统计执法检查中发现数据,可利用区块链技术在取证过程中准确识别联系人和工作人员,为统计监督提供有力证据。

使用区块链的安全日志记录、单向数据管理和身份验证加密等特征,可以安全地分配数据、跟踪所有删除和更改、监控数据流,整个数据生产过程中都能对数据进行跟踪、管理,有效保证数据的安全性和可靠性,使得统计监督做到有的放矢。

区块链中每存储一个数据就同步带带时间戳特征,留存记录者信息,可溯源的便捷性有利于统计监督检查。在具体数据质量的统计监督中,哪个企业在何时何地录入数据,哪个记账户何时何地录入数据,都在区块链中留存唯一的时间节点标记。各级统计机构可通过区块链中的网络标记方便快捷地查询到相关数据的时间和记录者信息,增强统计监督的针对性,为统计监督节省大量的人力物力时间。

(四)区块链优化统计执法检查,增强统计监督有效性

运用区块链存证溯源、安全记录和单向加密等方式管理和验证统计数据,使数据在分布式数据库中跨库安全共享,任何删改变化都可追溯,数据流动过程可监控,数据录入生成全流

程痕迹管理,有效提升统计数据的公信力。面对海量数据,区块链的追溯成本远低于传统数据传输模式,有效保障数据的安全性和真实性,提高统计执法效率。

传统统计监督模式下,统计监督机构在执行统计监督活动时有可能受到人为干扰,影响统计监督的客观性和独立性。结合区块链技术开展统计监督,区块链的共识机制和智能合约使得统计机构和统计人员实施统计监督时具有匿名性,方便监督机构和人员开展监督业务,减少人为干涉对统计监督的影响,提高统计监督的独立性。

引入区块链技术对企业数据进行前期筛查,将企业统计报表数据与跨平台数据进行比对,考察其是否与其他政府部门记录数据匹配、企业纵向数据是否互悖、数据修改是否异常等,精准锁定可能存在问题的相关企业后,再进行针对性的统计执法检查,实现统计执法成本最大化,降低执法检查成本。

区块链强化统计的去中心化,增强地方和部门统计机构的独立性,一定程度上重构国家统计局和地方统计局之间、统计局和部门所属统计机构之间的管理和监督体制,将自上而下和自下而上的统计监督有机结合起来。区块链使得统计数据更易于访问和使用,每个节点都可以参与监督,也共同为其作证,有效提升统计监督的参与度和监督力度。

三、结语

当前,我国统计工作面临前所未有的新机遇,统计改革发展也处于历史最好时期,但也面临着诸多挑战和新要求。党中央将统计监督纳入党和国家监督体系,对加强统计监督职能作用提出新要求,赋予新使命。依据加强统计监督职能作用的总体要求,针对当前我国统计监督存在的突出问题,依据区块链技术与统计监督的结合点,未来应切实利用区块链技术强化统计监督职能作用。

为推进国家治理体系和治理能力现代化,需不断拓展统计监督领域,健全统计监督范围,丰富统计监督内容,实现统计监督的全覆盖,以现代信息技术为支撑,探索区块链、大数据等在统计监督中的应用,形成强大的统计服务能力和统计监督功能。健全数据驱动型的治理方式,将统计监督作为科学执政、精准调控的重要手段,充分发挥统计监督职能作用,切实提高党和国家的治理效能。

(作者单位:浙江财经大学、浙江工商大学)