数据跨境流转的风险测度与分析

——基于数据出境统计信息的实证研究

李 金 徐 姗 卓子寒 李建平

【摘 要】数据作为新的生产要素,在不断的跨境流动中贡献经济价值。然而,数据的无序跨境流动可能危害个人、社会和国家安全,需同步探索治理体系、防范化解风险挑战,实现安全与发展的统筹协调。本文基于制度失范理论和威慑理论分析数据跨境风险的影响机制,利用生物医疗行业脱敏处理的数据出境统计信息开展研究,通过数据跨境传输频数、风险值和间隔时间测度数据跨境传输行为的频次和风险。研究结果表明,机构所在省份的GDP、过往数据跨境传输量和采用互联网传输均正向影响数据跨境频数和风险,采用云存储正向影响数据跨境风险,相关法律的施行对数据跨境传输具有显著的威慑作用。本文定义一次数据跨境传输行为并设计风险测度方法,识别和分析影响数据跨境传输的因素及其作用,为我国数据跨境流动规制提供理论依据和管理启示。

【关 键 词】数据跨境;风险测度;风险分析;制度失范理论;威慑理论

【作者简介】李金、徐姗,西安交通大学管理学院;卓子寒,国家计算机网络应急技术处理协调中心;李建平(通讯作者),中国科学院大学经济与管理学院。

【原文出处】《管理世界》,2023.7.180~199,Ⅺ-1

【基金项目】本研究得到国家自然科学基金项目(71901169、T2293774)、中国科学院学部院士咨询评议重大项目(2022-ZW14-Z-027)、中国科协青年人才托举工程项目(YESS20200072)的资助。

一、引言

习近平总书记指出“数字经济发展速度之快、辐射范围之广、影响程度之深前所未有,正在成为重组全球要素资源、重塑全球经济结构、改变全球竞争格局的关键力量”(习近平,2022)。2019年数字经济成为我国经济增长的主要推动力,对我国经济的直接贡献达到6.6%(陈昌盛等,2020)。数字经济的蓬勃发展和信息技术与经济社会的充分交汇融合引发数据迅猛增长,数据的重要性愈加凸显并已成为一种基础性战略资源。数据跨境流动指数据在不同国家之间的动态流转和迁移,是跨国贸易的重要推力,也是全球经济发展的关键推动者(梅尔策,2015)。在2009年至2018年期间,全球的数据跨境流动对经济增长的贡献度超过10%,预计2025年其对全球经济增长的价值贡献有望突破11万亿美元(梅尔策、洛夫洛克,2018)。数据跨境流动在促进国家信息技术创新和数字经济产业发展的同时,也带来了国家安全、产业安全以及个人隐私权益等问题(覃庆玲,2019)。我国的数据跨境规模居世界首位,占比达到23%(商务部,2021)。《中华人民共和国国民经济和社会发展第十四个五年规划和2035年远景目标纲要》中明确提出“加强数据安全评估,推动数据跨境安全有序流动”①。

《全球数字治理白皮书(2020年)》指出,各个国家对数据跨境流动的安全隐患和风险管理日益关注,数据保护主义的态势呈上升趋势。据统计,2019年全球共发布超过200条针对数据跨境秩序的限制性政策,一定程度限制了数据要素价值的释放(中国信息通信研究院,2020)。2021年国家互联网信息办公室审议通过的《网络安全审查办法》规定网络平台运营者赴国外上市需申报网络安全审查②,对于涉及到我国数据安全的境外上市企业,需要严格把控其数据跨境流动问题。2021年国家互联网信息办公室等五部门联合发布《汽车数据安全管理若干规定(试行)》,要求不得违规向境外提供重要汽车数据。互联网和云计算的发展使跨境数据可以动态、实时地传输,数据跨境流动愈发常见(韩洪灵等,2021)。数据共享具有重大的战略意义,世界多国政府发起数据开放倡议,但这必须在确保国家安全和公共安全、社会稳定的前提下进行(孙等,2021)。

本国数据出境可能会涉及国家核心安全问题以及个人信息保护问题。为保障个人信息和重要数据安全,世界多个国家不断寻求平衡隐私保护、数据安全与数据跨境流动的方法,形成了以欧盟为代表和以美国为代

表的两种立法模式(孙等,2021)。欧盟于2018年出台《通用数据保护条例》(GDPR)严格限制个人数据的跨境转移。美国于2018年发布《澄清境外数据的合法使用法案》(CLOUD),规定了跨境调取海外公民的信息和通信数据等方面内容,同时推动构建多项贸易协定,推进国家间的数据自由流动。比如日本与美国在2019年签署的贸易协定中提出要确保各领域数据无障碍跨境传输。英国于2020年宣布了新的未来科技贸易战略,允许英国和亚太国家之间的数据自由流动,希望与日本等国达成比其作为欧盟成员国时期更进一步的数据协议。韩国于2018年修订《信息与通信网络法》,提出韩国监管机构可以对限制个人数据流出的国家进行对等的个人数据跨境转移限制。印度在《个人数据保护法案》(2019年国会审议稿)中明确要求印度公民的个人数据要在本地储存,印度公民的个人关键数据限制出境。

近年来,我国相继发布了《网络安全法》、《生物安全法》、《数据安全法》、《个人信息保护法》、《信息安全技术-数据出境安全评估指南(草案)》、《网络安全审查办法》、《数据出境安全评估办法》等一系列涉及数据跨境流动的政策法规、指南办法及相关标准,逐渐建立起较为完善的数据安全法律体系,为我国的数据跨境合规流转和风险管理提供了法律保障。其中,《生物安全法》规定"将我国人类遗传资源信息向境外组织、个人及其设立或者实际控制的机构提供或者开放使用的,应当向国务院科学技术主管部门事先报告并提交信息备份"。我国是数据跨境流动的重要参与者(胡炜,2018),亟需在保证数据安全的情况下,与其他国家或地区进行数据跨境交流与合作,增强我国在全球数据要素治理领域的影响力。

在国际博弈大环境和中国法律管理国情下,各地迫切需要破局数字经济与数据跨境安全的统筹协调发展。在保障数据安全和严控国家风险(孙晓蕾等,2015)的前提下,实现数据跨境自由流动并非易事(单文华、邓娜,2021)。当前研究多集中于剖析数字经济的发展、影响及规制(陈晓红等,2022;许恒等,2020;赵涛等,2020)、对比分析不同国家的数据跨境流动相关法律(胡炜,2018;马其家、李晓楠,2021a;沙利文,2019),以及聚焦在数据跨境流动规制的政策建议(王震,2022;魏远山,2021)等方面,鲜有研究对数据跨境流动的风险因素识别及测度进行定量分析。因此,本文利用生物医疗行业脱敏处理的数据跨境传输统计信息,从定量的角度识别影响生物医疗数据跨境传输的风险因素,并分析其影响效果。

以现有文献为基础,本文基于制度失范理论和威慑理论构建分析模型,以经过脱敏处理的跨境数据统计信息为支撑,使用数据跨境传输频数、传输风险值以及传输间隔时间3种测度方式刻画生物医疗行业的数据跨境传输行为,分析影响数据跨境传输频数及风险的因素,并进一步识别异常数据跨境传输行为、量化分析我国相关法律的实施以及欧盟GDPR违规案例公示对数据跨境传输行为的影响效果。通过实证分析识别影响数据跨境传输风险的关键要素,结果表明机构所在省份的国内生产总值(GDP)、过往数据跨境传输包数和采用互联网传输均与数据跨境传输频数及风险正相关,采用云存储正向影响数据跨境风险,而普通高校数量则与数据跨境传输风险负相关。基于上述实证结果,本文使用模拟仿真方法刻画了技术应用前后的数据跨境传输行为变化,为相关政策的制定提供决策支持。

本文的创新性及贡献主要体现在以下几个方面。第一,将制度失范理论和威慑理论拓展并应用于数据跨境风险管理领域。第二,以大数据为驱动,从定量分析视角厘清一次数据跨境传输行为的概念,使用数据跨境传输频数、风险值以及间隔时间3种测度方式分别刻画我国的数据出境情况。第三,基于我国部分地区的数据跨境传输记录,对影响数据跨境传输行为的因素开展计量分析,并针对异常数据跨境传输行为、我国相关法律实施的威慑作用以及欧盟GDPR违规案例的凸显作用做进一步分析。基于参数化的模型结果,模拟采用不同技术前后的机构以及我国部分地区整体数据跨境传输的变化情况,并通过进一步分析检验模型关键自变量的影响机制和模型结果的稳健性。第四,在数据跨境流动日益频繁的背景下,如何构建符合我国基本国情的数据跨境安全管理制度是亟需解决的问题,本文的研究结果可为相关政策的实施效果提供可参考的分析模式和研判依据。作为国内第一批开展的数据驱动下数据跨境流动分析研究,本文旨在为我国数据跨境流动的风险管理和合理规制提供量化分析依据,支持我国数字经济高质量健康发展。

二、相关文献和理论

近年来,越来越多的国家要求公司在本地存储特定类型的数据(罗丁、萨克斯,2020)。数据本地化在一定程度上是数字贸易的障碍,会阻碍企业做出最优的市场决策,但是出于隐私保护以及国家安全等需要,各个国家之间难以实现数据跨境的自由流动(盛斌、高疆,2020)。关于数据跨境流动的文献主要为政策类、定性分析以及少数定量分析文献,使用计量方法分析影响数据跨境传输行为的文献相对较少。

(一)数据跨境流动风险管理的相关文献

数据跨境流动的安全问题受到世界各国关注,探究数据跨境流动的驱动因素和影响机制、衡量数据跨境

流动的风险,对构建合理的数据跨境流动框架、促进数据跨境安全和自由流动至关重要。目前很少有学者针对数据跨境流动的风险测度及其影响因素展开定量研究。对于应当如何监管数据跨境流动以及影响数据跨境流动的关键因素尚未达成共识(阿伦森,2019)。

大数据驱动下的风险预见与评估监测是管理决策领域中的重要议题(陈国青等,2020),学术界从不同角度开展了对数据跨境流动的风险管理研究。面对重要数据的跨境流动所带来的风险挑战,对数据跨境流动风险进行量化分析并制定合理合规的风险防控应对机制显得尤为重要。在风险测度指标体系方面,李等(2022)从宏观和微观层面分别提炼数据跨境流动的风险因素,系统刻画了数据出境风险,构建数据出境风险指标体系和数据出境风险评价模型,为数据跨境流动的预警管理提供了可实践应用的量化方法。在数据跨境传输网络方面,李金等(2021b)整合数据传输节点的相关网络属性用以计算风险路径的风险值,引入网络节点的数据接收能力以优化预测算法,对我国重要行业跨境流动的数据开展实证分析,研判数据跨境传输的风险路径。在技术层面,拉赫曼等(2020)提出了一个基于区块链保障数据共享功能的可靠跨境数据共享平台。李等(2020)构建预防性的阳光互联网(Bright Internet)范式,基于欧盟 GDPR 的十项隐私权设计预防性网络安全措施以保护网络传输信息的隐私安全。

在保障数据安全的前提下进行数据跨境流动,需要重视数据泄漏风险的测度与管控。2022 年 9 月 1 日起施行的《数据出境安全评估办法》第八条中提及数据出境安全评估需要重点关注跨境数据出境中和出境后的泄漏和篡改等风险。诸多学者从新兴信息技术投资、技术采纳应用、高管是否具备信息技术背景等视角围绕数据泄漏风险开展相关研究(海斯利普等,2021;金、权,2019;森、博尔,2015)。森和博尔(2015)基于犯罪机会理论、制度失范理论和制度理论从不同层面剖析可能增加或降低数据泄漏风险的因素,发现在信息技术安全上的投资会导致相对更高的数据泄漏风险。金和权(2019)聚焦医疗行业,探究医院采用电子病历等数字化技术对不同类型的医院信息泄漏事件发生风险的影响。他们的研究结果表明,采用诸如电子病历等数字化技术在提高医疗质量的同时,也有可能增加信息泄漏风险(金、权,2019)。海斯利普等(2021)通过实证研究发现具有信息技术背景的高管可以显著降低所在企业的数据泄漏风险。

(二)数据跨境流动管理的相关政策分析

目前针对数据跨境流动的政策分析类文献多集中在不同法律条例的对比分析以及如何正确应对数据跨境流动所带来的各种风险挑战等方面。

数据跨境流动的规制面临“良好的数据保护”、“数据跨境自由流动”和各国政府“数据保护自主权”不可能同时达成的“三难选择问题”(黄宁、李杨,2017)。欧盟和美国两大规制体系构成了当今数据跨境流动规制的国际基本格局,二者在价值取向、规制路径和规制结果等方面均有不同(许多奇,2018)。欧盟极其重视隐私权,强调通过立法进行事前防范,对数据跨境流动进行了非常严格的规制;而美国的政策则相对宽泛和自由,主要通过行业自律的形式进行事后问责。《隐私盾协议》无效案是欧美在规制数据跨境流动冲突的具体表现,二者之间的根本矛盾在于“技术主权”之争(单文华、邓娜,2021)。在物联网背景下,沙利文(2019)认为与美国所倡导的亚太经合组织(APEC)跨境隐私保护规则体系(CBPR)相比,欧盟的 GDPR 在平衡数据保护和数据流动上更具优势。赛奥等(2017)定性地描述了 GDPR 如何影响物联网企业的成本,并基于统计方法和相关法律要求对成本进行了定量估计。从定性角度而言,GDPR 影响了企业的预防成本和法律成本;从定量角度而言,在 GDPR 实施后企业成本平均会增加 3 ~4 倍(赛奥等,2017)。鲍尔等(2016)对数据跨境流动的监管与国内经济绩效之间的关系进行计量分析,发现数据法规的颁布和实施导致欧盟的 GDP 降低了 0.48%。

部分学者认为数据流动贸易协定是管理数据跨境流动的适当途径(梅尔策,2015;阿伦森,2016;米切尔、米什拉,2019)。美国和欧盟利用贸易协定来推进和限制数据的自由流动(阿伦森,2015)。2020 年 11 月 15 日,中国和东盟 10 国、日本、韩国、澳大利亚、新西兰共 15 个亚太国家正式签署了《区域全面经济伙伴关系协定》(RCEP),彭德雷和张子琳(2021)将其与《全面与进步跨太平洋伙伴关系协定》(CPTPP)进行对比分析,并分别从政府层面和企业层面讨论 RCEP 所带来的影响。米切尔和米什拉(2019)建议世贸组织成员通过关于自由数据跨境流动和禁止数据本地化的约束性条款以及促进商业和消费者信任数字贸易的相关条款。国际贸易框架在减轻数据跨境流动所带来风险的同时,可以更有效地利用数据跨境流动所带来的收益(塞奇威克,2017)。我国则应以“一带一路”倡议为基础,积极参与制定国际规则,在未来贸易谈判中维护数据跨境流动相关的数据权益(胡炜,2018)。

数据跨境流动是推动我国数字经济发展的必然选择(马其家、李晓楠,2021b)。我国数据跨境流动管理相关的法律法规体系正在逐步建立并完善。从法律法规框架层面,一些学者认为我国应完善数据跨境流动治理

体系,在保护我国数据安全的前提下实现合规的数据跨境流动(李墨丝,2020;许多奇,2018)。为应对由数据跨境自由流动和禁止数据本地化等规则带来的重大挑战,我国应注重国内法治与国际规则的良性互动(李墨丝,2020),综合考虑我国产业发展现状、法律现状和风险管控能力以及我国与国际规则互动情况,建立数据分类审核制度(高山行、刘伟奇,2017;胡炜,2018;马其家、李晓楠,2021b)。例如马其家和李晓楠(2021b)认为可以将跨境数据按来源和重要程度划分为个人数据、商业数据、特种行业数据进行分类管理。除了通过完善立法和政府监管的方式外,行业监管部门亦可以根据行业自身特点具体落实规制数据跨境流动的规章制度(许多奇,2018)。

(三)本研究的理论基础

1. 制度失范理论

制度失范理论(Institutional Anomie Theory)主张经济规则在一种有利于发达社会中高犯罪率的制度结构下享有最高优先地位,其关注社会制度中所体现的文化和社会结构。梅斯纳和罗森菲尔德(2012)认为,美国的犯罪率高是由特定的文化价值观和制度之间的关系造成的。该理论认为经济上的主导地位意味着非经济机构的削弱,即促使个体遵守制度规范的主要动机在经济主导条件下很可能是薄弱的,进而导致违反规范的行为变得更加频繁和普遍(梅斯纳等,2008)。包括教育机构在内的一些社会机构,可能会加剧社会对物质成功的关注,进而加强通过非法手段获得物质财富的压力(梅斯纳、罗森菲尔德,1997,2012)。制度失范理论从宏观层面解释了影响违法违规行为的社会因素,并发展成为一个具有普适性的理论框架(梅斯纳、罗森菲尔德,2012;梅斯纳等,2008)。制度失范理论被广泛应用于解释一些犯罪及非道德行为的同时,也被应用于跨国背景下的社会文化和偏差行为相关研究之中(卡伦等,2004;森、博尔,2015;哈夫曼、梅斯纳,2021;修斯等,2015;纳姆等,2014)。偏差行为指大多数人不参与的行为或者结果与平均水平显著不同的行为(默滕斯等,2016)。在数据跨境传输的背景下,某一机构存在超量传输和新增传输链路等异常数据跨境传输行为均可看作偏差行为。纳姆等(2014)将创新看作一种积极偏差行为,以制度失范理论解释不同国家企业创新的异质性。除此之外,利他主义和企业家精神等都被认为是积极的偏差行为(赫克特,1998;卡伦等,2014)。在跨界创新领域,跨界指打破原有障碍的行为,包括跨行业、跨组织、跨国界等(张等,2022),跨界创新行为亦是一种积极偏差行为。数据跨境传输行为往往由跨境电商、跨境合作等跨境创新行为引发,因此合规的数据跨境传输行为属于积极偏差行为的范畴。基于制度失范理论,经济因素和社会文化因素将影响数据跨境传输行为。因此,本文在中国情境下研究制度失范理论中宏观层面的经济以及社会文化等因素是否会对数据跨境传输频数以及风险造成影响。

2. 威慑理论

一般而言,威慑作用指当个体或组织看到其他个体或组织因某种行为被惩罚后,会在未来避免类似行为(班杜拉、沃尔特斯,1977)。当对于未来实施犯罪的预期为负面结果时,个体实施此类行为的动机会随之降低(伊迈、克里希纳,2004)。威慑理论目前已被广泛应用到信息安全领域(斯特劳布,1990;林润辉等,2015)。当违反信息安全规定的代价增大时,员工遵守信息安全规定的意图会相应提高(林润辉等,2015)。基于威慑理论,当威慑确定性高以及威慑严重性高时,潜在犯罪者的反社会行为将被抑制(斯特劳布,1990),法律的存在会增加相关个体和组织对非法行为惩罚的感知风险(李、李,2002)。违法所需承担后果的严重程度反映了法律的威慑严重性程度,当潜在违法者所感知到的违法成本高于其违法的潜在收益后,会放弃实施违法行为。我国近年来不断完善关于数据跨境流动治理的法律体系,如《生物安全法》于 2021 年 4 月 15 日开始实施,针对生物信息出境做出了规制。相关组织机构及个体在从事数据跨境传输行为时,必须考虑自身行为是否符合我国法律规定并开展相应的跨境数据安全评估。在上述法律施行后,相关数据源机构可以预见违法的后果,进而降低实施违法行为的动机。因此,本文基于威慑理论,进一步探究数据跨境传输行为受相关法律施行的影响效果。

3. 理论机制分析

基于制度失范理论和威慑理论,本部分通过构建模型分析制度失范因素和威慑因素以及机构特征因素对相关机构数据跨境传输行为的影响机制。经济发展水平更高的省份,其数字经济发展水平也相应更高,数字产品以及数字服务的跨境交易也成为数字经济全球化的重要标志(中国信息通信研究院,2021)。因此,本部分假设制度失范因素中的经济因素会影响该地区的数字经济发展程度,进而影响该地区机构跨境传输活动的潜在收益。李金等(2021a)指出数据出境机构所在地区的整体受教育程度越高,其对数据跨境传输的合规性要求越高。因此,本部分假设教育因素会影响数据跨境传输活动的合规成本。机构出于获取利益或开展合作

等目的,向境外机构传输数据。与此同时,机构可以通过合规跨境传输以及违规跨境传输的方式向境外传输数据。假设机构的合规数据跨境传输量为 k,违规数据跨境传输量为 m,$k\geq 0$,$m\geq 0$。机构对于数字技术($Tech$)的采纳会使得数据跨境传输更为便利,进而影响数据跨境传输的收益,假设其对数据跨境传输所带来的收益为 $Tech\cdot\gamma$,$\gamma>0$,代表机构采纳数字技术所能获取的数据跨境传输附加收益。合规数据跨境传输需要付出一定的合规成本 CC,该传输成本与合规跨境传输量 k 正相关,即 $CC=\delta\cdot k^2$,$\delta>0$,代表随着合规数据跨境传输量的增加,机构所需付出的合规成本增加(卡罗等,2018)。此外,考虑到部分敏感数据仅能通过违规传输的方式出境,假定合规传输与违规传输的收益系数不同。机构通过合规跨境传输所能获得的收益函数为 $\pi(k)=(Tech\cdot\gamma+\alpha_1)k$,通过违规跨境传输所能获得的收益函数为 $\pi(m)=(Tech\cdot\gamma+\alpha_2)m$(伦诺克斯,2006),其中 $\alpha_1>0$,$\alpha_2>0$,代表合规跨境传输以及违规跨境传输的基础收益系数,$Tech\cdot\gamma$ 则代表了快捷的传输所带来的附加收益。机构选择违规传输时会面临被监管机构发现的风险以及需要承担违规被发现的后果。由于法律的施行会使组织对非法行为的感知风险增加(李、李,2002),潜在违规组织的违规意愿会随着威慑确定性和威慑严重性的增加而降低(斯特劳布,1990)。威慑确定性通过机构违规传输被发现的概率 p 表示,威慑严重性则通过违规后果 VC 表示。机构违规传输量 m 越大,其所感知到的违规成本(PC)越高。基于上述定义,机构的目标为使其通过数据跨境传输活动所获取的总效用 U 最大化,其中:

$$U=\pi(k)+\pi(m)-CC-PC \tag{1}$$

$$U=(Tech\cdot\gamma+\alpha_1)k+(Tech\cdot\gamma+\alpha_2)m-\delta\cdot k^2-p\cdot VC\cdot m^2$$

上述效用函数的一阶导为:

$$\frac{\partial U}{\partial k}=Tech\cdot\gamma+\alpha_1-2\delta\cdot k=0$$

$$\frac{\partial U}{\partial m}=Tech\cdot\gamma+\alpha_2-2p\cdot VC\cdot m=0$$

通过运算可得:

$$k=\frac{Tech\cdot\gamma+\alpha_1}{2\delta} \tag{2}$$

$$m=\frac{Tech\cdot\gamma+\alpha_2}{2p\cdot VC} \tag{3}$$

$$N=k+m=\frac{Tech\cdot\gamma+\alpha_1}{2\delta}+\frac{Tech\cdot\gamma+\alpha_2}{2p\cdot VC} \tag{4}$$

由式(2)、式(3)可知,$\partial k/\partial\alpha_1>0$、$\partial m/\partial\alpha_2>0$,随着跨境传输收益的增加,合规跨境传输以及违规跨境传输次数均增加,即数据跨境传输收益更高的地区数据跨境传输频次更高,且存在更多的潜在违规跨境传输风险。数字经济发展水平在经济发达地区相对较高(中国信息通信研究院,2021),经济发达地区的机构通过数据跨境传输所能获取的基础收益越高,进而导致其有更高的数据跨境传输频次和更高的数据跨境传输风险。基于式(3)可知,$\partial m/\partial p<0$、$\partial m/\partial VC<0$,即随着威慑确定性和威慑严重性的增加,相关机构的违规跨境传输次数降低。随着我国相关法律的实施,法律所带来的威慑确定性和威慑严重性均会随之增加,进而降低机构数据跨境传输频次和风险。式(4)展示了机构数据跨境传输总次数(N)与各因素之间的关系,其中 $\partial N/\partial Tech>0$,表明机构采纳数字技术会增加其数据跨境传输频次和风险,后文将从存储方式和数据传输方式来衡量机构对数字技术的采用情况。最后,考虑到机构进行决策时可能并非理性,其决策结果可能会受到其自身特征(如过往数据跨境传输经验)、环境特征(如所在地区的数据出境概况)、行业特征(如科研机构、公司机构、专业机构)、地域特征(如东部、中部、西部地区)等的影响,因此本文在实证模型中加入了包含上述因素的控制变量。

图 1 为本文的研究框架图。基于上述制度失范理论、威慑理论以及理论机制分析的结果,识别包括经济因素、教育文化因素在内的制度失范因素和威慑作用因素。分别使用数据跨境传输频数、数据跨境传输风险值以及相邻两次数据跨境传输时间间隔测度数据跨境传输频率及风险,并分析上述因素对数据跨境传输行为的影响效果。根据此模型,识别以超量传输、新增链路传输为代表的异常数据跨境传输风险事件,并分析上述因素的差异化影响效果。此外,本文针对我国法律的实施效果开展研究,分析《生物安全法》以及省级法规的实施对生物医疗数据跨境传输行为的威慑作用。基于凸显理论,本文进一步分析了欧盟 GDPR 违规案例对违规数据出境后果的凸显作用。基于对生物医疗数据跨境传输风险的整体测度结果,本文从存储方式和传输方式两个角度展开模拟仿真分析,对比不同技术应用前后的数据跨境传输行为,为政策制定提供决策支持和参

考。最后,本文通过改变一次数据跨境传输的定义、调整因变量测度方式等开展模型稳健性检验。

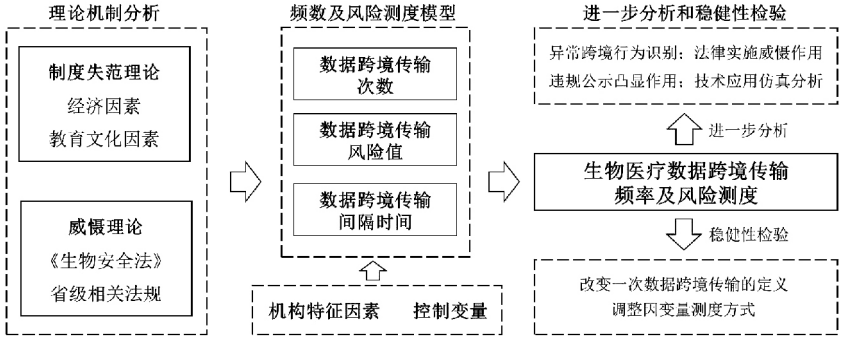


图 1 研究框架图

三、数据来源与研究设计

(一)数据来源与样本选取

本文基于调研得到的生物医疗行业经过脱敏处理的跨境数据统计信息开展研究,样本区间为 2019 年 7 月 ~ 2020 年 8 月。首先,一次数据传输可能会包含多条传输记录,这些传输记录的境内源单位与境外目的单位相同,并且其传输开始时间十分相近。因此,本文将具有相同的源单位与目的单位,且传输开始时间间隔小于一小时的传输记录定义为一次数据传输。其次,本文从国家统计局获取了全国各个省份在 2019 年和 2020 年间的 GDP 和进出口贸易额、高校数量等数据,并与数据跨境流动数据集进行匹配。最后,在此基础上本文补充了关于数据传输源单位的机构和数据传输相关信息,包括机构的数据存储方式、数据传输方式、机构所属行业及地区等在内的多维特征数据。综上,基于以上数据来源和定义,本研究的数据集共包含 175 个数据源单位在 428 天时间范围内的共计约 7.5 万条观察记录。

(二)研究设计

1. 因变量

为刻画机构数据跨境传输行为的频数及风险,本文分别将数据传出机构 i 在第 t 天数据跨境传输次数的对数($\ln NumCross_{it}$)、机构每日风险值的对数($\ln Risk_{it}$)以及机构 i 的第 s 个两次数据跨境传输之间的时间间隔($TimeInterval_{is}$)作为因变量。具体的定义描述如下。

第一,机构每日数据跨境传输频数衡量了其开展数据跨境传输活动的频率,在进行生物医疗数据跨境流转的风险管理时需要重点关注频繁开展数据跨境传输活动的机构。此外,对机构日常数据跨境传输频数的刻画,可为后续进一步识别异常数据跨境传输行为提供基础条件。为避免个别极端值的影响,本文使用机构每日数据跨境传输次数的对数($\ln NumCross_{it}$)为因变量刻画数据跨境流转的频数。

第二,为衡量数据跨境传输风险,本部分将综合考虑风险发生可能性以及风险发生后果的影响(李金等,2022)。结合李金等(2022)基于复杂网络的数据流动路径风险量化算法,本部分使用机构在第 t 天之前一周的历史跨境传输数据计算数据传出机构 i 的数据跨境流动路径风险值,并进一步计算数据传出机构 i 在第 t 天的单日总风险值。在刻画机构数据跨境传输风险时,本文选取机构每日风险值的对数($\ln Risk_{it}$)作为因变量。

第三,上述两个因变量可从整体上分别衡量数据跨境传输的态势及风险。在进行生物医疗数据跨境流转的风险预警时,发生下一次异常数据跨境传输行为的风险比过去所发生的数据跨境传输行为更值得关注。两次事件之间的时间间隔被应用于信息安全领域以测度信息安全事件的风险,如使用多个信息间的间隔时间来识别消息注入式攻击(宋等,2016)、使用两次数据泄漏事件之间的间隔时间刻画数据泄漏事件风险(森、博尔,2015)等。相邻两次数据跨境流转之间的时间间隔越短,说明发生数据跨境流转事件的可能性越大。基于时间间隔的分布规律和动态变化趋势可以进一步检测生物医疗数据跨境传输的风险,如基于此时间间隔开展生存分析以衡量数据源单位下一次向境外传输生物医疗数据的概率,并进行技术应用模拟仿真等,具备一定现实意义。因此,为动态刻画数据跨境传输频率及风险,本文选取相邻两次数据跨境传输之间间隔的时间($TimeInterval_{is}$)作为因变量。

2. 自变量

制度失范因素。基于本研究理论基础部分的分析结果,制度失范理论表明经济因素和教育文化因素将会对数据跨境传输这一偏差行为带来影响。

针对制度失范理论的经济因素,本文选取数据跨境传输源单位所在省份的 GDP 代表该地区的市场机构

力量,分析经济因素对于数据跨境传输行为的影响效果。制度失范理论主张市场机构的主导力量会削弱遵守规范的动机(梅斯纳等,2008),而机构所在省份的 GDP 是衡量当地市场机构力量的重要指标(森、博尔,2015),更高的 GDP 可能会导致更高的数据跨境传输风险。此外,作为衡量经济表现的指标(森、博尔,2015),GDP 衡量了该地区的经济发展水平,经济发展水平较高的省份,数字经济发展水平也相应更高(中国信息通信研究院,2021)。各类数字化平台积累了海量数据,数字产品的跨境传输和数字化服务的跨境提供成为数字经济全球化的标志(中国信息通信研究院,2021)。数字经济的发展离不开科技实力的推动,数字化技术是数字经济发展的基础(田秀娟、李睿,2022)。数字经济发展水平更高的省份拥有更多数字资产,经济发展水平更高的地区对前沿技术的研发投入也相对更高。因此,经济发展水平更高的地区具备更多可供交易的数字资产,也具备更加强大的数字化技术,其数据跨境传输频次也会相应更高。综上,地区经济发展水平会影响当地数据跨境传输源单位的数据跨境传输频次和风险,在分析数据跨境传输行为时必须考虑经济发展程度差异所带来的影响。

针对制度失范理论的教育文化因素,本文在模型中加入数据跨境传输源单位所在省份的高校数量来衡量教育文化因素对数据跨境传输行为的影响效果。基于制度失范理论,强大的教育系统与对物质财富的渴望密切相关(纳姆等,2014),教育机构等社会机构会通过加强对物质成功的关注而增加以非法手段获得物质财富的压力(梅斯纳、罗森菲尔德,1997,2012),进而增加以违规手段实现数据跨境传输的动机,影响数据跨境传输风险。李金等(2021a)指出数据出境机构所在地区的整体受教育程度越高,其对数据跨境传输的安全性和合规性的要求也就越强,强大的教育系统可能会通过提升当地对数据跨境传输安全性和合规性的重视而降低数据跨境传输风险。因此,教育文化因素可能会对数据跨境传输风险同时造成两个方向的影响。高校数量与当地教育水平正相关(瓦莱罗、范里嫩,2019),本文使用数据跨境传输源单位所在省份的普通高等学校数量来表示当地的教育文化水平,通过定量分析探究制度失范理论中的教育文化因素与数据跨境传输风险的关系。

机构特征因素。包括机构类型、数据传输量等在内的机构特征因素会影响数据传出方的数据跨境传输风险(李金等,2022)。机构过往的数据跨境传输特征可能会影响其未来的数据跨境传输活动。在过去数据跨境传输活动中传输更大量数据到境外的机构,其未来也更有可能开展较多的数据跨境传输活动。不同机构在数据存储与数据传输方式上具有差异性。部分机构选择采用云存储而非本地存储,也有机构选择使用互联网传输而非内网传输。机构之间在技术采用上的差异性也会影响其数据跨境传输行为。因此,有必要在模型中整合包括机构传输特征以及机构自身技术特征在内的机构特征因素变量。

3. 控制变量

本文所选取的控制变量为环境因素变量,包括机构所在省份的进出口总额以及跨境传输机构占比。数据跨境传输行为与跨国贸易息息相关,进出口总额衡量了该地区对外贸易的规模。某一地区存在数据跨境传输行为的机构数量越多,该地区的数据跨境传输行为相对而言可能会更加普遍。因此在模型中加入了该省份数据跨境传输机构占所有数据跨境传输机构数量的比例来控制该省份的数据跨境传输环境因素的影响。

综上,本文所使用的因变量、自变量和控制变量的符号及定义如表 1 所示。

表 1　　变量定义

变量	定义
ln*NumCross*	数据源机构每日数据跨境传输频数的对数
ln*Risk*	数据源机构每日风险值的对数
TimeInterval	数据源机构相邻两次数据跨境传输之间的间隔时间(分钟)
GDP	数据源机构所在省在对应年度的 GDP(亿元)
NumUniversity	数据源机构所在省的普通高等学校数(个)
NumPackets	数据源机构过往数据跨境传输的加权总数据包数(个)
DataStorage	虚拟变量,数据源机构采用云存储取 1,采用本地存储取 0
DataTransmission	虚拟变量,数据源机构采用互联网传输取 1,采用内网传输取 0
ImportExport	数据源机构所在省在对应年度的进出口贸易总额(亿美元)
UnitRate	数据源机构所在省的跨境机构数与跨境机构总数的比值

4. 模型构建

本文使用3种不同的测度方式刻画数据跨境传输频率及风险:数据跨境传输频数、数据跨境传输风险值及相邻两次数据跨境传输时间间隔。数据跨境传输频数和数据跨境传输风险值可分别刻画数据跨境传输的频率和风险;而相邻两次数据跨境传输时间间隔可在刻画数据跨境传输频率的同时,动态反映发生下一次数据跨境传输的风险。

首先,本文构建了数据跨境传输频数测度模型和数据跨境传输风险测度模型。由于自变量中包含机构特征变量,因此不适用于个体固定效应。本模型采用行业固定效应以捕捉行业之间的差异,采用地区固定效应以捕捉地区之间的差异。此外,为了解决如技术进步等随着时间而变化的遗漏变量问题,进一步考虑了时间固定效应。本文设定的数据跨境传输频数测度模型及数据跨境传输风险测度模型如式(5)所示。

$$DV_{it}=\beta_0+\beta_1 GDP_{it}+\beta_2 NumUniversity_i+\beta_3 NumPackets_{it-1}+\beta_4 DataStorage_i+\beta_5 DataTransmission_i+\beta_6 ImportExport_{it}+\beta_7 UnitRatei+I_i+A_i+T_t+\varepsilon_{it} \quad (5)$$

其中 i 表示数据源单位,t 表示时间(天)。DV_{it} 的取值分别为 $\ln NumCross_{it}$ 和 $\ln Risk_{it}$。其中,$\ln NumCross_{it}$ 表示数据源单位 i 在第 t 天的数据跨境传输次数的对数,$lnRisk_{it}$ 表示数据源单位 i 在第 t 天的数据跨境风险值的对数。I_i 为行业固定效应,A_i 为地区固定效应,T_t 为时间固定效应。解释变量的系数分别表示其对数据跨境传输频数和风险值的影响。

为刻画数据跨境传输行为的频率变化以及发生下一次数据跨境传输的风险,本文进一步以两次数据跨境传输之间的时间间隔为因变量,构建负二项回归模型,并通过马尔可夫蒙特卡洛算法(MCMC)进行参数估计(森、博尔,2015)。具体的参数估计步骤详见《管理世界》网络发行版附录。机构的数据跨境传输时间间隔服从如式(6)所示的负二项分布。

$$TimeInterval_{is}\sim NBD(\mu_{is},v) \quad (6)$$

其中 $TimeInterval_{is}$ 表示数据源单位 i 的第 s 个时间间隔,μ_{is} 为负二项分布的均值,v 为离散参数。均值 μ_{is} 会随源单位 i 和数据跨境事件的发生而变化,其表达式如下:

$$\ln(\mu_{is})=\beta_0+\beta_1 GDP_{is}+\beta_2 NumUniversity_i+\beta_3 NumPackets_{is}+\beta_4 DataStorage_i+\beta_5 DataTransmission_i+\beta_6 ImportExport_{is}+\beta_7 UnitRate_i+I_i+A_i+\varepsilon_{is} \quad (7)$$

式(7)中解释变量的系数表示其对下一次发生数据跨境传输风险的影响。在负二项回归和使用马尔可夫蒙特卡洛算法估计各参数时,在其余因素不变的情况下,如果系数 $\beta>0$,表示对应自变量的取值增加1个单位时,数据跨境传输的时间间隔均值将会增加 e^{β} 倍,即发生数据跨境传输的风险降低;反之,若系数 $\beta<0$,表示对应自变量的取值增加1个单位时,数据跨境传输的时间间隔均值将会减少 e^{β} 倍,即发生数据跨境传输的风险增加。

考虑到各变量的量纲不同,本文对变量 GDP、$NumUniversity$、$NumPackets$、$ImportExport$ 做归一化处理。此外,对于过往数据传输包数($NumPackets$)做加权求和处理,给予近期传输包数更大权重。其中 $NumPackets_{it}=0.95\times NumPackets_{it-1}+Packets_{it}$,$Packets_{it}$ 表示数据源机构 i 在第 t 天的数据跨境传输总包数;类似地,$NumPackets_{is}=0.95\times NumPackets_{is-1}+Packets_{is}$,$Packets_{is}$ 表示数据源单位 i 的第 s 个时间间隔所对应的数据跨境传输总包数。

四、实证检验与结果分析

(一)描述统计

自2019年7月1日至2020年8月31日的数据跨境传输样本情况如图2所示,日均数据跨境传输次数波动明显且有上升趋势。我国数据大多传往美国、英国等国家或地区,其中目的机构在美国的传输次数最多,高达1.5万次,占总传输次数的59.27%;英国次之,共0.7万次,占总传输次数的28.80%。数据跨境传输的源机构多位于经济较为发达的省份,且这些省份的总传输次数也高于经济相对欠发达的省份。其中,广东省存在数据跨境传输行为的机构数最多,占样本总单位数的14.87%;北京次之约占12.31%,但北京的总数据跨境传输次数高于广东省,约占总传输次数的26.32%。该样本数据集中共包含3类发生数据跨境传输行为的机构,分别为公司机构、专业机构和科研机构。其中科研机构所发生的数据跨境传输次数最多,高达2.3万次,公司机构和专业机构则分别为0.3万次和0.1万次。

表2为因变量和部分主要自变量的描述性统计结果。此数据集中共涉及我国部分地区共计29个省/市的单位,这些省份的GDP、进出口总额、普通高等学校数均表现出较大的差异性。

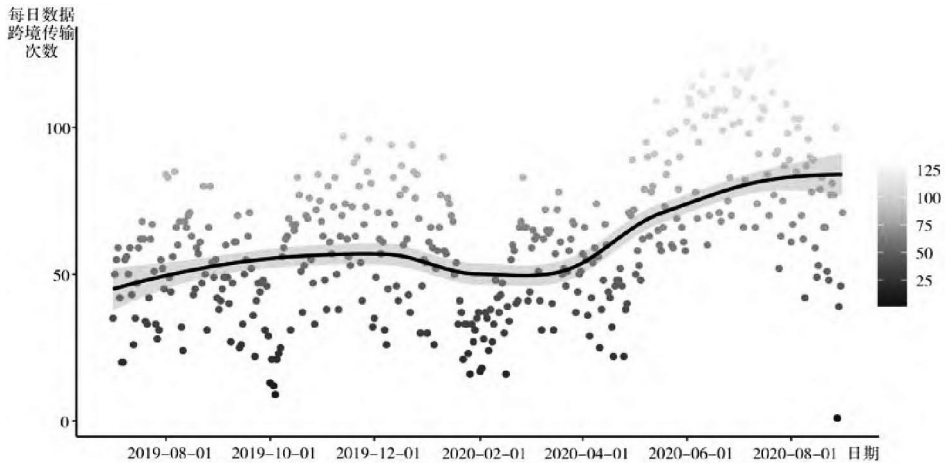


图 2　数据跨境传输状况

表 2　　变量描述性统计结果

变量	均值	标准差	最小值	最大值
ln*NumCross*	0.15	0.43	0	3.40
ln*Risk*	9.88	2.03	1.20	15.39
TimeInterval	2932.33	1637.91	0	500954
GDP	34197.69	25716.34	2941.1	11076.09
NumUniversity	91.79	38.72	12	167
NumPackets	15984.09	96410.52	1	2581806
DataStorage	0.04	0.2	0	1
DataTransmission	0.01	0.11	0	1
ImportExport	1589.29	2338.38	3.3	10366.27
UnitRate	0.03	0.04	0.002	0.15

（二）结果分析

以频数和风险值为因变量的模型分析结果如表 3 所示，其中第（1）列和第（3）列分别为未引入数据源机构的技术采纳特征的数据跨境传输频数测度模型和风险测度模型的结果，第（2）列和第（4）列分别为引入数据源机构技术采纳特征后的模型分析结果。GDP 对机构的数据跨境传输频数和数据跨境传输风险的影响均显著为正，表明经济更发达地区的数据源机构不仅更频繁地开展数据跨境传输活动，而且也存在更高的数据跨境传输风险。普通高等学校数（*Num University*）在数据跨境频数测度模型中的影响不显著，但在数据跨境风险测度模型中的系数显著为负，说明地区教育发展程度更高省份的机构其数据跨境传输风险相对更低。过往数据跨境传输包数（*Num Packets*）的影响均显著为正，说明过去数据跨境传输量的大小与当日数据跨境传输次数及风险值正相关。整体而言，数据存储方式（*DataStorage*）对每日数据跨境传输频数的影响不显著，与数据跨境传输风险显著正相关。数据传输方式（*Data Transmission*）与数据跨境传输频数和风险均显著正相关，表明使用互联网传输的机构更有可能参与数据跨境传输活动且具有更高的风险发生可能性和更严重的风险发生后果。进出口总额（*ImportExport*）的系数显著为负，说明位于对外贸易量较高省份的机构其数据跨境传输频次和风险均更小。跨境单位占比（*Unit Rate*）的系数显著为正，说明高跨境单位占比省份的机构其数据跨境传输频数和风险均相对更高。

以相邻两次数据跨境传输时间间隔为因变量的模型分析结果如表 4 所示。本部分采用线性回归模型、负二项回归模型进行参数估计；同时考虑机构异质性，进一步使用 MCMC 方法进行负二项回归模型的参数估计。其中各模型第（1）列不包含行业及地区固定效应，第（2）列包含上述固定效应。负二项回归模型的 AIC 值和

BIC 值均小于线性回归模型的结果,拟合效果更优;基于 MCMC 做参数估计的模型其 AIC 值与 BIC 值与负二项回归模型相近,进一步说明使用负二项分布刻画数据跨境传输时间间隔的合理性。

表 3　　主模型估计结果(因变量:ln*Num Cross* 和 ln*Risk*)

变量	(1)	(2)	(3)	(4)
	ln*NumCross*	ln*NumCross*	ln*Risk*	ln*Risk*
截距项	−0.1229 (0.0862)	−0.1278 (0.0825)	5.3857*** (0.6823)	5.8764*** (0.6639)
GDP	0.2979** (0.1449)	0.3831*** (0.1407)	1.0095*** (0.3773)	4.6670*** (0.3930)
NumUniversity	0.1166 (0.1469)	0.0684 (0.1394)	−0.9190*** (0.2740)	−3.4345*** (0.2832)
NumPackets	0.0814*** (0.0275)	0.0816*** (0.0275)	1.0706*** (0.1807)	0.9382*** (0.1751)
DataStorage		0.1694 (0.1041)		0.3615*** (0.1043)
DataTransmission		0.6805*** (0.1822)		2.3032*** (0.1480)
ImportExport	−0.8862*** (0.0789)	−0.8629*** (0.0784)	−1.6873*** (0.2788)	−3.3516*** (0.2782)
UnitRate	3.6659*** (0.5073)	3.1697*** (0.4760)	9.1857*** (0.9349)	9.7806*** (0.9087)
观测值	74725	74725	10476	10476
Industry FE	YES	YES	YES	YES
Area FE	YES	YES	YES	YES
Time FE	YES	YES	YES	YES
R^2	0.0365	0.0365	0.1981	0.2041

注:*、**、***分别代表 10%、5%、1% 显著性水平,括号内为标准误。

表 4　　主模型估计结果(因变量:*TimeInterval*)

变量	线性回归		负二项回归		MCMC	
	(1)	(2)	(1)	(2)	(1)	(2)
截距项	3066.0728*** (445.1266)	15337.8362*** (857.9457)	8.8115*** (0.0505)	10.6405*** (0.0822)	8.8102*** (0.0497)	10.6339*** (0.0831)
GDP	−25656.4274*** (3181.3020)	−25658.0733*** (3458.0188)	−2.2515*** (0.3178)	−3.3824*** (0.3388)	−2.1812*** (0.3103)	−3.3999*** (0.2946)
NumUniversity	18541.4507*** (2031.8023)	17606.4650*** (2308.1404)	1.1883*** (0.2039)	2.5172*** (0.2266)	1.1504*** (0.2007)	2.5265*** (0.1993)
NumPackets	−1590.7079 (1002.1956)	−770.1685 (990.8184)	−0.5601*** (0.0832)	−0.1809** (0.0897)	−0.5532*** (0.0827)	−0.1742* (0.0902)
DataStorage	−1718.9515*** (498.6218)	−1677.2969*** (634.8895)	−0.5847*** (0.0543)	−0.7149*** (0.0665)	−0.5818*** (0.0545)	−0.7177*** (0.0652)

续表4

变量	线性回归		负二项回归		MCMC	
	(1)	(2)	(1)	(2)	(1)	(2)
DataTransmission	−1654.3293*** (568.4955)	−699.5723 (735.7679)	−1.5442*** (0.0595)	−1.7373*** (0.0759)	−1.5486*** (0.0600)	−1.7338*** (0.0757)
ImportExport	16645.5760*** (2529.8115)	18117.1342*** (2634.0330)	2.5110*** (0.2550)	2.4752*** (0.2606)	2.4468*** (0.2490)	2.4912*** (0.2297)
UnitRate	−58432.4922*** (7289.9087)	−66565.6549*** (8132.0452)	−15.6304*** (0.7579)	−14.5580*** (0.8135)	−15.4346*** (0.7547)	−14.5935*** (0.7689)
观测值	26303	26303	26303	26303	26303	26303
Industry FE	NO	YES	NO	YES	NO	YES
Area FE	NO	YES	NO	YES	NO	YES
对数似然	−292378.07	−292045.08	−209200.88	−207239.07	−209200.93	−207239.09
AIC	584772.14	584114.16	418419.76	414504.14	418419.85	414504.17
BIC	5848373.56	584212.29	418493.36	414610.45	418493.45	414610.48

注：*、**、***分别代表10%、5%、1%显著性水平，线性回归和负二项回归结果中括号内为标准误，MCMC结果中括号内为标准差。

制度失范因素和机构特征因素均对机构的数据跨境传输行为具有显著影响。根据表4中负二项回归模型以及MCMC模型的结果，GDP、过往机构传输包数（*NumPackets*）、数据存储方式（*DataStorage*）和数据传输方式（*DataTransmission*）的系数均显著为负，与相邻两次数据跨境传输之间的间隔时间负相关，与数据跨境传输频率及数据跨境传输风险正相关。普通高等学校数（*Num University*）的系数显著为正，说明地区教育发展程度更高省份的机构其数据跨境传输风险相对更低。上述模型结果与使用每日数据跨境传输频数和数据跨境传输风险值为因变量的模型结果一致。

五、进一步分析与稳健性检验

（一）进一步分析

上述实证分析结果表明了制度失范和机构特征等各类因素对数据源单位数据跨境传输频数和风险的影响。本部分将从以下5个方面开展进一步分析。第一，将存在超量传输以及新增传输链路的数据跨境传输行为定义为异常数据跨境传输的潜在风险事件，进一步分析各风险因素的影响效果；第二，收集新的数据集以分析我国相关法律的实施对生物医疗数据跨境传输行为的影响；第三，分析欧盟GDPR违规案例对违规数据跨境传输不良后果的凸显作用；第四，检验关键变量的影响机制；第五，基于参数化的实证模型，模拟仿真不同技术应用前后数据跨境传输行为的变化情况。

1. 识别异常数据跨境传输行为

本部分通过识别异常数据跨境传输量与数据源单位新增数据跨境传输目的单位，从数据集中提取异常数据跨境传输事件并进行分析。网络安全领域常通过分析监测到的网络流量负载、故障和包大小等是否与网络行为的正常模式一致来进行异常检测（熊等，2014）。数据跨境传输行为中的超量传输和新增传输链路都是值得关注的异常因素。超量传输指一次数据跨境传输行为中所传输的数据包数远高于常规水平，而新增传输链路则是指在数据集中第一次出现由某一源单位将数据传往某一境外单位时的情况。

本部分考虑如下两种异常数据跨境传输行为的情况：第一，当传输数据包数（*Num Packets*）大于该单位的数据跨境传输包数平均值加3个标准差时；第二，当数据跨境传输源单位第一次将数据传往某一境外目的单位时。本部分使用与上文相同的分析模型以异常数据跨境传输次数的对数（ln*NumCross_Abn*）以及异常数据跨境传输间的时间间隔（*TimeInterval_A bn*）为因变量刻画其频次及风险。由于在基于二分网络计算每日风险值时，需要综合考虑节点活跃时长、活跃强度等（李金等，2022），仅使用异常数据跨境传输行为计算机构风险值不能衡量其真实风险水平，因此本部分未使用每日风险值作为因变量。基于以上定义，本部分的数据集共包

含 162 个数据源单位的异常数据跨境传输记录。

异常数据跨境传输行为的模型分析结果如表 5 所示,其中第(1)、(2)列为以每日数据跨境传输频数为因变量的主模型及异常数据跨境传输行为模型的分析结果。除 GDP 外,其余因素对异常数据跨境传输风险的影响均与主模型一致。GDP 与异常数据跨境传输频数负相关,经济发展程度更高地区的异常数据跨境传输频数更低,即异常数据跨境传输风险更低。表 5 的第(3)、(4)列为以相邻两次时间间隔为因变量的主模型及异常数据跨境传输行为模型的分析结果。与数据跨境传输频次模型分析结果类似,GDP 与异常数据跨境传输时间间隔正相关,表明经济发展程度更高地区的异常数据跨境传输风险更低。此外,普通高等学校数的系数显著为负,表明教育程度较发达地区的异常数据跨境传输时间间隔更短,风险相对较高。在大数据驱动下,可考虑引入其他跨域信息,通过大数据融合分析全面识别异常因素进一步扩展此模型,更加精确且全面地识别出异常数据跨境传输行为,为保障数据跨境安全、自由合规流转管理提供支持。

表 5　　异常数据跨境传输行为分析结果

变量	(1)	(2)	(3)	(4)
	ln*NumCross*	ln*NumCross_Abn*	*TimeInterval*	*TimeInterval_Abn*
截距项	−0.1278 (0.0825)	0.0658*** (0.0177)	10.6405*** (0.0822)	14.1226*** (0.2586)
GDP	0.3831*** (0.1407)	−0.0883** (0.0396)	−3.3824*** (0.3388)	5.2750*** (0.9232)
NumUniversity	0.0684 (0.1394)	0.0542 (0.0335)	2.5172*** (0.2266)	−2.6354*** (0.6136)
NumPackets	0.0816*** (0.0275)	0.0899*** (0.0097)	−0.1809** (0.0897)	−0.1881 (0.1590)
DataStorage	0.1694 (0.1041)	0.0097 (0.0188)	−0.7149*** (0.0665)	−0.7443*** (0.2774)
DataTransmission	0.6805*** (0.1822)	0.2811*** (0.0329)	−1.7373*** (0.0759)	−2.5041*** (0.3000)
Controls	YES	YES	YES	YES
观测值	74725	69174	26303	2152
Industry FE	YES	YES	YES	YES
Area FE	YES	YES	YES	YES
Time FE	YES	YES	NO	NO
R^2	0.0365	0.0123	0.0211	0.0204

注:*、**、***分别代表 10%、5%、1% 显著性水平,括号内为标准误。

2. 法律实施的威慑作用

威慑理论表明法律的存在会增加相关组织对非法行为惩罚的感知风险(李、李,2002;马西米诺等,2017)。通过提升违规行为的感知成本,威慑因素可削弱参与违规行为的动机(陈等,2012;马西米诺等,2017)。《生物安全法》是针对生物行业的专门法律,于 2020 年 10 月 17 日审议通过,并于 2021 年 4 月 15 日起正式施行。《生物安全法》对人类遗传资源信息出境以及利用我国生物资源开展国际科学研究合作均做出了相关规定。相关组织机构在从事生物医疗数据跨境传输行为时,需要考虑自身行为是否符合我国《生物安全法》的规定并获取有关部门批准。本文所使用的数据集来自生物医疗行业,适用于《生物安全法》中的相关规定。基于威慑理论,《生物安全法》的实施会增加与生物行业相关机构对违规生物医疗数据跨境传输行为的感知风险,进而对潜在违规生物医疗数据跨境传输行为造成威慑作用,降低整体数据跨境传输频率。因此,本部分分析基于威慑理论,探究我国数据跨境传输行为受《生物安全法》施行的影响效果。在《生物安全法》施行前,许多省份施行

了一系列与数据安全相关的地方性法规,如《海南省大数据开发应用条例》、《贵州省大数据安全保障条例》等。此类地方性法规没有针对生物医疗数据跨境传输行为做出针对性规制,但其涉及数据安全管理领域。本部分使用数据源单位所在省份与数据安全相关的省级地方性法规的数量衡量地方性非专门法规的威慑作用。

本部分分析使用2021年1月至12月脱敏处理的数据跨境传输统计信息检验《生物安全法》(*BiosecurityLaw*)的实施对数据跨境传输行为的影响效果,并结合2019年7月至2020年8月的数据跨境传输统计信息检验地方性法规数量(*NumLocalLaw*)对数据跨境传输行为的影响效果。由于数据本身的局限性,缺少2021年数据跨境传输记录的接收节点的详细信息,无法构建二分网络并计算机构每日风险值,且两次数据跨境传输间的时间间隔会包含法律生效时间,不能准确界定其影响效果。因此,本部分以每日数据跨境传输次数为因变量,测度并分析数据跨境传输频次受我国法律施行的影响效果。结果见表6所示,其中第(1)列为使用2019年7月至2020年8月的数据的分析结果,第(2)和(3)列为使用2021年1月至12月数据的模型分析结果。第(1)列结果表明,与数据安全相关的地方性法规可以对当地数据跨境传输行为造成威慑作用,显著降低数据跨境传输量。第(2)和(3)列结果表明,与地方性非针对性法规相比,《生物安全法》对于生物医疗数据跨境传输行为的威慑作用更强。

表6　　我国相关法律实施对数据跨境传输行为的影响分析结果

变量	(1)	(2)	(3)
	ln*NumCross*	ln*NumCross*	ln*NumCross*
BiosecurityLaw		−0.1379*	−0.1379*
		(0.0757)	(0.0757)
NumLocalLaw	−0.0272***		−0.0045
	(0.0040)		(0.0120)
Controls	YES	YES	YES
观测值	74725	46780	46780
Industry FE	YES	YES	YES
Arca FE	YES	YES	YES
Time FE	YES	YES	YES
R^2	00370	00169	00169

注:*、**、***分别代表10%、5%、1%显著性水平,括号内为标准误。

3. 违规公示的凸显作用

凸显性在心理学中指当个体的注意力被吸引到环境中的某一部分时,个体在进行后续决策时会赋予此部分信息更高的权重(泰勒、汤姆森,1982)。机构组织所经历的或者历史中的突出事件会影响其对环境的认知以及对不同要素的关注程度(德赛,2014;霍夫曼、奥卡西奥,2001;马奇、夏皮罗,1992)。例如,行业报刊等媒体中公开报道的问责案例会成为相应行业内的焦点,并被业内成员广泛关注(霍夫曼、奥卡西奥,2001)。媒体对特定事件的报道会影响该事件的凸显性,进而影响同类组织的决策(德赛,2014)。

《通用数据保护条例(GDPR)》是欧盟于2018年5月25日出台的条例,该条例具有全球性的影响(李等,2020)。违反欧盟GDPR的机构会面临高额的罚款,而这些罚款案件的公开会吸引全球涉及数据跨境流转机构的关注(赛曼等,2022),进而凸显违规的严重不良后果。我国近年来在不断完善与数据跨境流动相关的法律体系,在2019年7月之前先后发布了《个人信息和重要数据出境安全评估办法(征求意见稿)》、《个人信息出境安全评估办法(征求意见稿)》等多版征求意见稿,反映出我国保障数据跨境安全自由流动的顶层设计,也传达出将管控违规数据跨境传输行为的信息。GDPR罚款事件会凸显违法向境外传输数据的严重后果,进而促使我国相关机构关注违规严重后果。因此,本部分在模型中引入GDPR的累积罚款金额变量来反映违规进行数据跨境传输的严重后果的凸显强度,进而衡量其对于数据跨境传输行为的影响作用。

本部分分析使用GDPR违规惩罚的平均强度,即截止到数据跨境传输事件发生日的平均罚款金额,衡量过往罚款事件对违规后果的凸显作用。表7第(1)、(2)、(3)列分别为使用每日数据跨境传输次数、数据跨境传输风险值以及两次数据跨境传输时间间隔为因变量的模型结果。

表 7 违规公示对数据跨境传输行为的影响分析结果

变量	(1)	(2)	(3)
	ln*NumCross*	ln*Risk*	*TimeInterval*
截距项	−0.0620 (0.0820)	9.7550*** (0.3416)	10.4930*** (0.0843)
AvgFine	−0.0677* (0.0358)	−3.9930*** (0.7508)	0.3582*** (0.0426)
Controls	YES	YES	YES
观测值	74725	10476	26303
Industry FE	YES	YES	YES
Area FE	YES	YES	YES
Time FE	YES	YES	NO
R^2	0.0365	0.2041	0.0213

注:*、**、***分别代表 10%、5%、1% 显著性水平,括号内为标准误。

表 7 分析结果表明随着凸显强度的增加,机构每日数据跨境传输次数减少,风险值降低,数据跨境传输时间间隔增加,出现下一次数据跨境传输的风险降低。该部分分析的结果表明欧盟 GDPR 惩罚事件可通过凸显违规数据出境的不良后果影响相关机构的数据跨境传输行为。

4. 关键变量的影响机制检验

前文的模型结果表明经济发展程度与数据跨境传输频数和风险值均显著正相关。本部分将针对 GDP 对数据跨境传输频次及风险的具体作用机制展开分析。在本文的理论基础以及研究设计部分阐述了 GDP 会通过提升该地区的研发投入进而推动该地区的数字经济发展,最终影响数据跨境传输的频率及风险。本部分采用数据源机构所在省份的规模以上工业企业的 R&D 经费与其 GDP 之比(*RDIntensity*)来衡量该地区研发投入的强度。经济更加发达的地区的企业为维持市场核心竞争优势,会增加其研发支出(郑登攀等,2017)。而研发投入则与数字技术的发展息息相关,进而影响数据跨境传输行为。由于 3 种测度方式所使用的样本构造方式以及样本量存在差异,本文分别使用上述 3 种测度方式对应的数据集进行关键变量的影响机制检验。表 8 为分别使用 3 种测度方式的模型中以定义的研发经费密度(*RDIntensity*)为被解释变量的模型分析结果。表 8 中列(1)、(2)、(3)分别对应原本使用数据跨境传输次数、数据跨境传输风险值、数据跨境传输间隔时间为因变量的模型机制检验结果。各列结果均显示 GDP 的系数为正值,且在 1% 的水平下显著,说明经济发达地区的研发投入密度(*RDIntensity*)也相对更高。

表 8 GDP 对数据跨境传输行为影响机制检验结果

变量	(1)	(2)	(3)
	因变量:ln*NumCross*	因变量:ln*Risk*	因变量:*TimeInterval*
	RDIntensity	*RDIntensity*	*RDIntensity*
截距项	0.0062*** (0.0010)	0.0095*** (0.0010)	0.0152*** (0.0001)
GDP	0.0037*** (0.0001)	0.0417*** (0.0006)	0.0274*** (0.0006)
NumUniversity	0.0056*** (0.0009)	−0.0187*** (0.0004)	−0.0149*** (0.0004)
NumPackets	−0.0000** (0.0000)	0.0004 (0.0003)	−0.0002 (0.0002)

续表8

变量	(1)	(2)	(3)
	因变量:ln*NumCross*	因变量:ln*Risk*	因变量:*TimeInterval*
	RDIntensity	*RDIntensity*	*RDIntensity*
DataStorage	-0.0007 (0.0014)	-0.0017*** (0.0002)	-0.0009*** (0.0001)
DataTransmission	-0.0001 (0.0024)	-0.0006*** (0.0002)	-0.0013*** (0.0001)
Controls	YES	YES	YES
观测值	74725	10476	26303
Industry FE	YES	YES	YES
Area FE	YES	YES	YES
Time FE	YES	YES	NO
R^2	0.8834	0.0060	0.7989

注:*、**、***分别代表10%、5%、1%显著性水平,括号内为标准误。

5.技术应用的仿真分析

当前我国已出台多项政策推动不同行业、不同企业进行数字化转型,鼓励企业上云(黄丽华等,2021)。云服务在各行业中日益普及(张等,2020),有必要考虑数据跨境传输机构在未来更加广泛地使用云存储而非本地存储给数据跨境传输行为带来的影响。上文的实证结果表明,使用云存储的机构数据跨境传输更频繁(见表3),且两次数据跨境传输的时间间隔相对较短(见表4),发生数据跨境传输的风险会相对更高。基于数据跨境传输频次及风险测度的模型分析结果,当机构从本地存储转换为云存储后,其平均每日数据跨境传输次数增加约17%,风险值增加约36%。基于以两次数据跨境传输之间的时间间隔为因变量估计出的参数化模型,本文进一步模拟了某一机构分别使用云存储和使用本地存储时的数据跨境传输时间间隔,并据此绘制生存曲线(图3左侧)。类似地,实证结果表明,机构使用互联网传输替代内网传输后会增加其进行数据跨境传输的频次。基于数据跨境传输频次及风险测度的模型分析结果,当机构从内网传输转换为互联网传输后,其平均每日数据跨境传输次数增加约68%,风险值增加超过200%。根据估计出的参数,本文模拟了某一机构分别使用互联网传输和使用内网传输时的数据跨境传输时间间隔,并绘制生存曲线(图3右侧)。图3表明在采用云存储以及采用互联网传输后,该机构的生存曲线下降速度变快,发生数据跨境传输的风险增加。

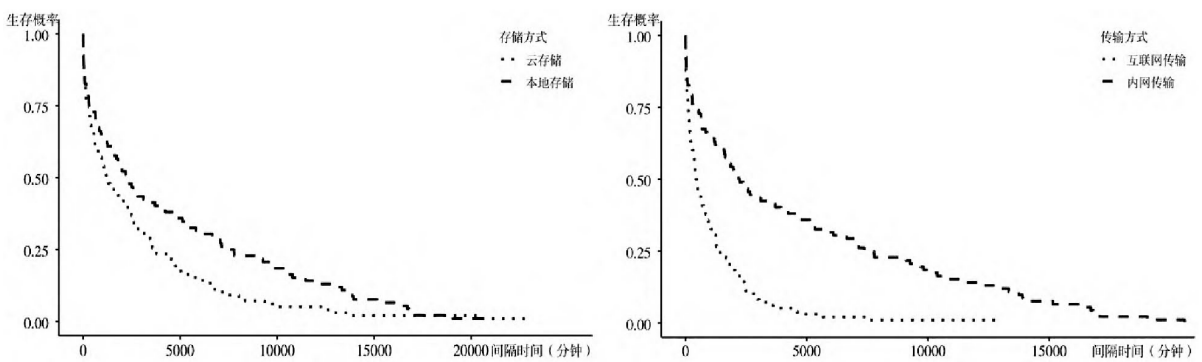


图3 存储方式和传输方式变化前后的生存曲线

当部分机构改变存储方式或传输方式后,我国部分地区整体的每日数据跨境传输次数也会发生相应改变。由上述针对某一机构的技术应用的模拟分析可知,使用云存储或互联网传输后其两次数据跨境传输行为之间的时间间隔缩短。因此,当有部分机构由使用本地存储转换为使用云存储或由内网传输转换为互联网传输后,我国部分地区整体的每日数据出境次数随之增加。基于估计出的参数及原始数据,图4为模拟不同比

例的机构由使用本地存储转为使用云存储后我国部分地区专业机构在2019年7月至2019年12月的每日数据出境次数的变化情况。其中散点代表估计出的技术变动前后的每日数据出境次数,曲线和阴影部分是根据估计出的每日数据出境次数拟合的曲线及标准误。当原本使用本地存储的机构转为使用云存储后,我国部分地区每日数据出境次数显著增加。其中,当25%的机构由本地存储转为使用云存储后,转换后拟合的每日数据出境曲线的取值大多在40次到45次之间,而转换前的拟合曲线的取值则在35次左右。随着转换比例的增大,转换前后的每日数据出境次数的拟合曲线之间的差距愈加明显。当转换比例达到100%时,即样本数据中所有机构均采用云存储时,我国部分地区每日数据出境次数的拟合曲线的取值基本维持在50次以上。在推动企业上云时,需考虑企业使用云服务技术后我国部分地区整体数据跨境传输风险会增加的情况,综合制定相应的风险防范对策,确保企业上云后的跨境数据安全。

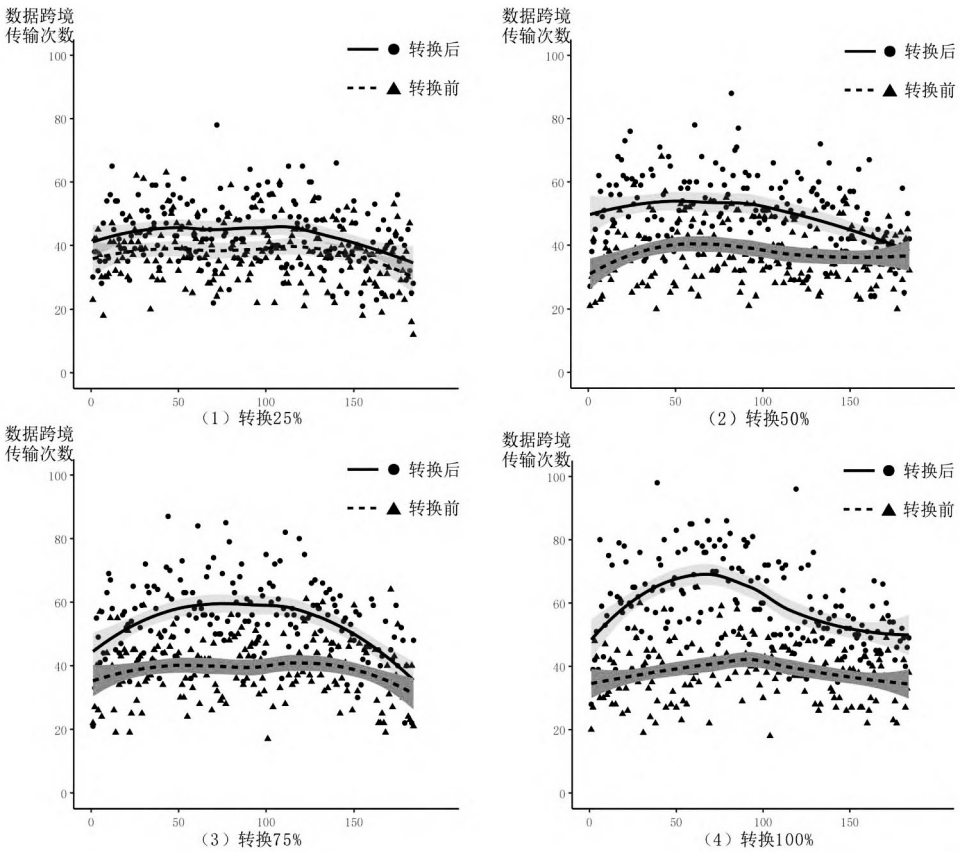


图4 改变存储方式前后整体数据出境次数的变化情况

与上述结果类似,图5为模拟不同比例的机构由使用内网传输转为使用互联网传输后,我国部分地区在2019年7月至2019年12月期间内的每日数据出境次数的变化情况。当原本使用内网传输的机构转为使用互联网传输后,每日数据出境次数显著增加。当25%的机构由内网传输转为使用互联网传输后,拟合的转换后每日数据出境次数大多为45次左右,而转换前的拟合曲线的取值则在35次左右。随着转换比例的增加,转换前后的每日数据出境次数的拟合曲线之间的差距愈加明显。当转换比例达到100%时,即样本数据中所有机构均采用互联网传输时,我国部分地区每日数据出境次数的拟合曲线的取值达到了约120次。采取不同数据跨境传输方式的数据源机构的数据跨境传输次数存在显著差异,相关行业监管部门可根据数据传输方式的不同,优化分配监管资源,重点开展高频数据跨境传输行为机构的风险管理。

(二)稳健性检验

本部分将从以下两个方面开展稳健性检验:一是改变一次数据跨境传输的定义,再次进行实证分析;二是设定不同的因变量测度方式,说明本文模型结果的稳健性。

1.改变一次数据跨境传输的定义

在上述分析中,将具有相同的源单位与目的单位,且传输开始时间相差小于一小时的传输记录定义为一次

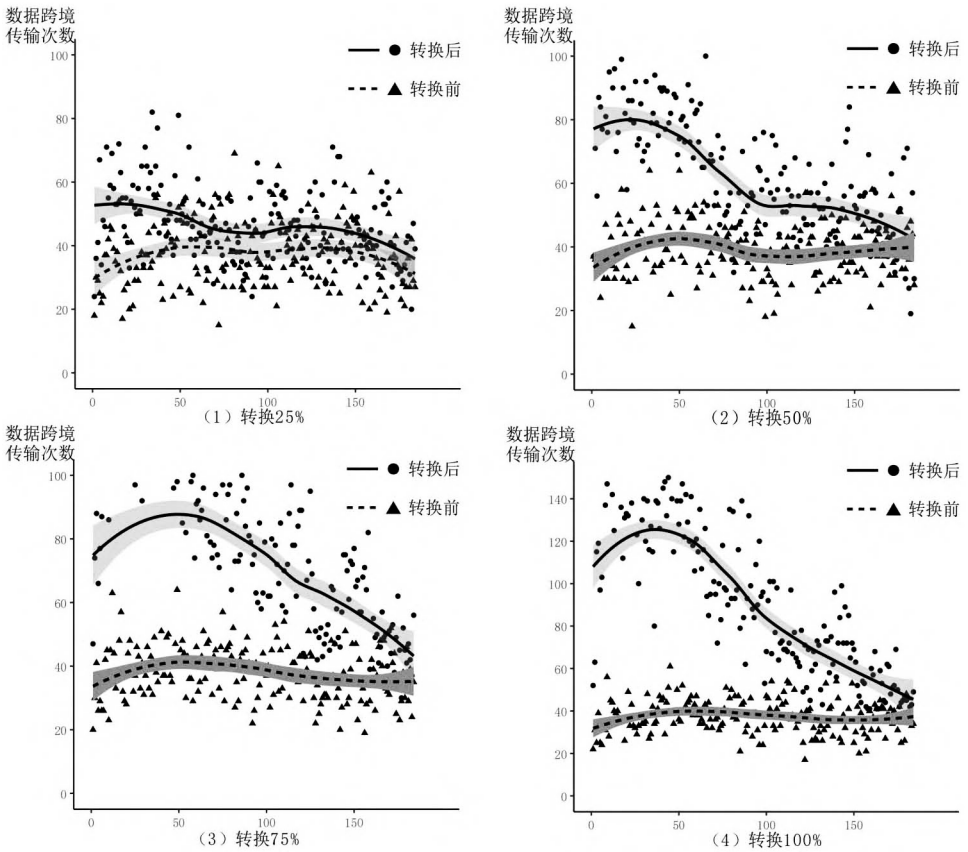


图5 改变传输方式前后整体数据出境次数的变化情况

数据传输，数据跨境传输次数与数据跨境传输时间间隔会受到此定义的影响。考虑具体定义的影响，表9中第(1)列将一次数据跨境传输行为定义改为具有相同的源单位与目的单位，且传输开始时间相差小于六小时的传输记录；第(2)列将定义改为具有相同的源单位且传输开始时间相差小于一小时的传输记录。第(1)与(2)列的因变量分别为根据更改定义后计算的每日数据跨境传输次数的对数(ln*NumCross*)与相邻两次数据跨境传输时间间隔(*TimeInterval*)。具体结果如表9所示，各参数的估计结果与原模型的结果基本保持一致。

表9 改变一次数据跨境传输的定义稳健性检验结果

变量	(1)		(2)	
	ln*NumCross*	*TimeInterval*	ln*NumCross*	*TimeInterval*
截距项	−0.0830 (0.0639)	11.2032*** (0.0964)	−0.0833 (0.0649)	10.2974*** (0.0849)
GDP	0.3421*** (0.1084)	−1.3393*** (0.3779)	0.3594*** (0.1208)	−3.1387*** (0.3387)
NumUniversity	0.0208 (0.1076)	1.1057*** (0.2563)	0.0326 (0.1143)	2.0911*** (0.2319)
NumPackets	0.1184*** (0.0211)	−0.3932*** (0.1078)	0.0641** (0.0253)	−0.3317*** (0.0766)
DataStorage	0.1456* (0.0807)	−0.7449*** (0.0766)	0.1608** (0.0794)	−0.6736*** (0.0657)
DataTransmission	0.6128*** (0.1412)	−1.8729*** (0.0881)	0.2135 (0.1390)	−0.5505*** (0.0856)

续表9

变量	(1)		(2)	
	ln*NumCross*	*TimeInterval*	ln*NumCross*	*TimeInterval*
Controls	YES	YES	YES	YES
观测值	74725	16643	74725	18960
Industry FE	YES	YES	YES	YES
Area FE	YES	YES	YES	YES
Time FE	YES	NO	YES	NO
R^2	0.0373	0.0225	0.0360	0.0098

注：*、**、***分别代表10%、5%、1%显著性水平，括号内为标准误。

2. 调整因变量测度方式

主模型中分别使用每日数据跨境传输次数和机构每日风险值的对数形式衡量机构数据跨境传输的频次和风险，在本部分中则针对上述因变量修改测度方式，使用原始数据以及对应的二元变量衡量频次和风险。具体来说，针对数据跨境传输频次测度模型，表10的列(1)和列(2)分别为使用每日数据跨境传输次数(*NumCross*)以及每日是否发生数据跨境传输(*CrossDummy*)为因变量的稳健性检验结果。针对数据跨境传输风险测度模型，列(3)和列(4)分别为使用机构每日风险值(*Risk*)和每日是否属于高风险机构(*HighRisk*)为因变量的稳健性检验结果。其中，每日风险值高于第三个四分位数的机构被定义为高风险机构。如表10所示，稳健性检验结果与原模型结果一致。

表10 调整因变量测度方式稳健性检验结果

变量	(1)	(2)	(3)	(4)
	NumCross	*CrossDummy*	*Risk*	*HighRisk*
截距项	−0.5216* (0.2748)	−0.0256 (0.0588)	−21841.4244 (62153.1967)	−0.0233 (0.0986)
GDP	0.7683* (0.4442)	0.4171*** (0.1110)	144664.2019*** (36798.0007)	0.1481** (0.0584)
NumUniversity	0.4200 (0.4517)	−0.0683 (0.1042)	−145428.4850*** (26513.6278)	−0.2324*** (0.0421)
NumPackets	0.4500*** (0.0839)	0.0200 (0.0235)	141308.9578*** (16396.8328)	0.1528*** (0.0260)
DataStorage	0.3097 (0.3517)	0.1701** (0.0715)	−8176.3860 (9768.8412)	−0.0005 (0.0155)
DataTransmission	4.2017*** (0.6157)	0.1568 (0.1252)	285293.3262*** (13854.6939)	0.5608*** (0.0220)
Controls	YES	YES	YES	YES
观测值	74725	74725	10476	10476
Industry FE	YES	YES	YES	YES
Area FE	YES	YES	YES	YES
Time FE	YES	YES	YES	YES
R^2	0.0210	0.0380	0.1068	0.4229

注：*、**、***分别代表10%、5%、1%显著性水平，括号内为标准误。

六、结束语

数据跨境流转关乎国家数据安全,围绕数据跨境传输行为进行定量分析,对保障数据安全和自由合规流转、促进数字经济的统筹发展具有十分重要的意义。基于制度失范理论和威慑理论,本文分别从发生频数和风险两个视角探究制度失范因素、机构特征因素以及历史传输特征因素等对机构数据跨境传输行为的影响,并进一步使用负二项分布刻画了数据跨境传输的时间间隔。本文通过识别超量传输和新增链路两种异常因素,综合分析了不同因素对异常数据跨境传输行为的影响作用,并分析了地方性非专门法规和《生物安全法》实施以及欧盟 GDPR 惩罚案例对数据跨境传输行为的影响效果。本文依据模型结果进行技术应用模拟分析,展现出采用不同技术时机构的数据跨境传输风险变化情况。最后,依据模型结果开展机制检验以及稳健性检验。

本文的研究结果表明,数据跨境传输行为的频数及风险受制度失范因素、威慑因素、机构特征因素的等影响,且影响效果具有差异性。具体而言,制度失范因素中的经济因素,即机构所在地区的 GDP,与数据跨境传输频率及风险均正相关,表明经济发展水平更高的地区面临更高的数据跨境传输风险。普通高等学校数量代表制度失范因素中的教育文化因素,与数据跨境传输风险负相关,表明强大的教育系统更有可能通过增强地区对数据安全的重视程度降低数据跨境传输风险,而非通过增加以非法手段获得物质财富的压力而增加数据跨境传输风险。我国《生物安全法》针对人类遗传资源信息出境做出规定,其实施可显著降低生物医疗数据跨境传输频率,且与数据安全相关的省级法规的施行也会影响生物医疗数据跨境传输行为,表明相关法律的存在可以对潜在违规数据跨境传输行为造成威慑作用。本文研究结果从实证的角度量化了我国生物和数据安全相关法律施行的影响效果。代表机构特征因素的机构过往传输包数和采用互联网传输也与数据跨境传输频率及风险正相关,而采用云存储则仅与数据跨境传输风险正相关。欧盟 GDPR 惩罚案例的公示也会凸显违规数据出境的不良后果,进而影响我国相关机构的数据跨境传输行为。此外,异常数据跨境传输行为受各因素的影响效果与上述数据跨境传输行为受各因素影响效果不同。因此,政府和行业监管机构可对具有不同传输频率以及风险的数据跨境传输行为采取差异化的规制方法,并针对具体管控目标完善治理数据跨境传输风险的法律及政策。

本文具有以下理论贡献。第一,本文从管理视角提出了一次数据跨境传输定义,并首次从频数、风险值、间隔时间等多个维度系统性地构建了风险测度方法,拓展应用于异常数据跨境传输的行为识别。首先,本文将在特定时间区间内的具有相同传出方和接收方的数据跨境传输记录定义为一次数据跨境行为,更加符合行业监管实际需求。其次,基于生物医疗行业机构经过脱敏处理的跨境数据统计信息,本文探索性地使用数据跨境传输频数、计算的风险值以及间隔时间分布等测度方式系统性地刻画数据跨境传输频次和风险。最后,本文提出的风险测度方法具有一定稳健性,并且可以应用于数据跨境的异常行为风险识别、技术应用仿真分析等行业监管和政策施行的重点关注领域。因此,本文提出的风险测度方法,可以从整体上反映数据跨境传输的态势和风险,为数据跨境流转的风险预警精细化管理提供风险量化方法。第二,本文融合了制度失范理论、威慑理论等理论并将其拓展应用于数据跨境风险管理领域,为数据跨境流转的风险管理提供可参考借鉴的理论依据。随着数字技术和企业数字化转型的发展,数据跨境安全和风险管理问题愈发重要,与此相关的风险管理实践、政策制定施行等均亟需理论层面的参考依据和支撑。因此,本文结合数据跨境风险管理的具体情境,刻画了制度失范理论中的经济和教育文化等制度失范因素、威慑理论中的法律威慑因素以及机构特征因素,构建了数据传出机构效用函数以分析上述因素的影响机制,并通过实证研究识别并量化了其影响效果,丰富了数据跨境流转风险管理研究的理论基础。第三,本文是使用计量方法开展数据跨境传输行为分析的首批工作之一,构建了数据跨境风险研判的量化分析框架,丰富了数据跨境传输风险管理领域的定量研究文献。首先,围绕数据跨境管理的现有研究多以政策分析和定性探索为主,本文基于数据出境的统计信息开展实证研究,量化分析了影响数据跨境传输行为发生的因素。其次,基于构建的风险测度模型和方法,本文进一步模拟分析了云存储等技术应用前后的数据跨境传输行为变化,直观地表现出新兴技术采纳对于潜在数据跨境传输风险的影响效果。因此,本文不仅从定量分析视角回答了何为影响数据跨境流动关键因素的研究问题,也为诸如推动企业上云等相关政策施行效果的模拟仿真提供了量化分析和风险研判的研究框架。

本文研究结论具有一定现实指导意义,可以为引导和促进数据自由跨境流转、统筹数字经济和数据安全协调发展提供以下管理启示和政策建议。

第一,逐步构建并动态拓展基于多源数据整合的数据跨境流转风险要素集。本文研究结果表明,制度失范因素、威慑因素以及机构特征因素等均会影响数据跨境传输频率和风险,意味着数据跨境管理过程中存在

风险要素种类多样的问题。为实现数据跨境安全风险防控与监测预警,识别网络数据跨境传输的重要特征,有必要基于多源数据构建数据跨境流转的风险要素集。该风险要素集可在本文实证分析结论的基础上扩充纳入更多的异常风险因素,进一步涵盖数据收发方机构属性信息、地理位置信息、传输路径信息、传输内容信息、机构间的合作关系信息以及相关法律法规出台施行信息等,为实现数据跨境安全风险预警提供数据库基础和可供参考的风险测度指标。

第二,加强数据跨境安全风险的精准管控,优化行业的监管资源配置。本文从实证角度发现不同地区和机构类型的数据跨境传输行为存在显著差异性,并且识别了数据跨境传输行为受各类风险要素的影响效果。因此,行业监管部门可基于本文识别的风险要素和异常数据跨境传输特征,采用本文构建的风险测度方法评估机构的风险水平,加强对高风险机构的数据跨境传输行为的精准管理,优化行业监管部门的监管资源配置。此外,基于本文模型以及未来不断拓展丰富的风险要素集,可进一步建立数据跨境安全态势测度指标体系,完善当前的数据跨境安全评估方法,探索构建数据跨境安全风险预警模型。在风险预警强时效性的要求下,解决提前且高效识别高风险数据跨境行为并采取针对性安全管理的关键瓶颈问题。推动监管模式由全面管控转为针对重大风险的精准管控,促进数据跨境传输的自由合规流动。

第三,完善数据跨境风险管理的法律体系,推动建立数据违规跨境公示系统。本文研究结果表明,相关法律的施行会对潜在违规机构造成威慑,降低数据跨境传输风险。加快相关法律法规和办法条例中涉及的诸如数据分类分级管理制度等的落地执行,一方面可以提升法律威慑的确定性,降低违规数据的出境风险;另一方面可以给予合规数据跨境传输机构明确指导,促进数据自由合规流转。与此同时,除法律法规的威慑作用外,本文研究结果也表明公示违规案例能够凸显违规数据跨境传输的不良后果,进而降低数据跨境传输风险。随着我国数据跨境传输法律体系的不断完善,可针对具体管控目标,试点探索通过国家企业信用信息公示系统、“信用中国”等网站平台全面公示数据出境违规案例及处罚结果,凸显违规数据出境的后果,降低潜在违规数据出境的频率和风险,进而保障国家数据安全,护航数字经济发展。

本文存在一定的局限性。首先,本文所使用的测度指标有一定适用性,但并不能刻画所有种类的数据跨境传输风险,如数据传出单位互联网协议地址突然变动、传输目的地数量激增且分散分布等潜在异常数据跨境传输风险,未来需综合使用多种测度方法和技术进行系统性风险管理。其次,本文所使用的数据中未包含关于数据跨境传输的详细信息,如数据传输目的、数据传输双方的合作关系等,未来可以从风险相关性视角开展研究。最后,本文所使用的数据为以中国部分地区的机构为数据传出方的数据跨境传输统计信息。本文的模型、方法和结果可以为其他国家和地区提供借鉴,并可在未来开展不同国家数据跨境传输风险管理的对比分析③。

文责自负。

《数据跨境流转的风险测度与分析——基于数据出境统计信息的实证研究》附录

使用马尔科夫链蒙特卡洛方法估计模型参数的步骤

本文使用了马尔科夫链蒙特卡洛方法(MCMC)估计以相邻两次数据跨境传输时间间隔为因变量的模型参数。数据源机构 i 的第 s 个数据跨境传输时间间隔 $TimeInterval_{is}$(后文简写为 T_{is})服从均值为 μ_{is},离散参数为 v 的负二项分布,其中未知参数服从如下先验分布:

$$\beta_k \sim N(0,100),k=0,1,2\cdots7 \tag{A1}$$

$$v \sim U(0,50) \tag{A2}$$

根据正文中式(6)、式(7)定义的模型函数以及上述未知参数的先验分布,使用 Gibbs 采样方法生成样本,参数采样具体步骤如下。

(1)各参数的满条件分布似然函数为:

$$L(\beta,v)=\prod_{i=1}^{N}\prod_{s=1}^{M}\frac{\Gamma(T_{is}+v^{-1})}{\Gamma(v^{-1})\Gamma(T_{is}+1)}\left(\frac{1}{1+v\mu_{is}}\right)^{v^{-1}}\left(\frac{\mu_{is}}{1+v\mu_{is}}\right)^{T_{is}}=\prod_{i=1}^{N}\prod_{s=1}^{M}\frac{\Gamma(T_{is}+v^{-1})}{\Gamma(v^{-1})\Gamma(T_{is}+1)}\left(\frac{1}{1+ve^{\beta X_{is}}}\right)^{v^{-1}}\left(\frac{ve^{\beta X_{is}}}{1+ve^{\beta X_{is}}}\right)^{T_{is}} \tag{A3}$$

将似然函数与参数先验相乘获得联合概率密度:

$$p(T,\beta,v)=L(\beta,v)\prod_{k=0}^{7}\left[\frac{1}{10\sqrt{2\pi}}\exp\left(-\frac{\beta_k^2}{200}\right)\right]\frac{1}{50} \tag{A4}$$

$$p(\beta,v|T)=\frac{p(T,\beta,v)}{p(T)}\propto p(T,\beta,v) \tag{A5}$$

参数 β_k 的满条件分布如下：

$$p(\beta_k|\beta_{-k},v,T)=\frac{p(\beta,v|T)}{p(v|T)p(\beta_{-k}|v,T)}\propto p(\beta,v|T)\propto p(T,\beta,v)\propto\prod_{i=1}^{N}\prod_{s=1}^{M}\left(\frac{1}{1+ve^{\beta_k X_{kis}+\beta_{-k}X_{-kis}}}\right)^{v^{-1}+T_{is}}\times e^{T_{is}\beta_k X_{kis}-\beta_k^2/200} \tag{A6}$$

其中，$k=0,1,2\cdots7$，$X_{0is}=1$。同理，参数 v 的满条件分布如下：

$$p(v|\beta,T)=\frac{p(\beta,v|T)}{p(\beta|T)}\propto p(\beta,v|T)\propto p(T,\beta,v)\propto\prod_{i=1}^{N}\prod_{s=1}^{M}\frac{\Gamma(T_{is}+v^{-1})}{\Gamma(v^{-1})\Gamma(T_{is}+1)}\left(\frac{1}{1+ve^{\beta x_{is}}}\right)^{v^{-1}+T_{is}}\times v^{T_{is}} \tag{A7}$$

(2)随机初始化参数状态，$\beta_k^{(0)}$，$k=0,1,2\cdots7$；$v^{(0)}$。其中参数 β_k 的初值为服从标准正态分布的随机数，参数 v 的初值为服从标准均匀分布的随机数。

(3)从 $t=0$ 到 $t=n$，分别：

A)从条件概率分布 $p(\beta_0|\beta_{-0}^t,v^t,T)$ 中采样得到样本 β_0^{t+1}。

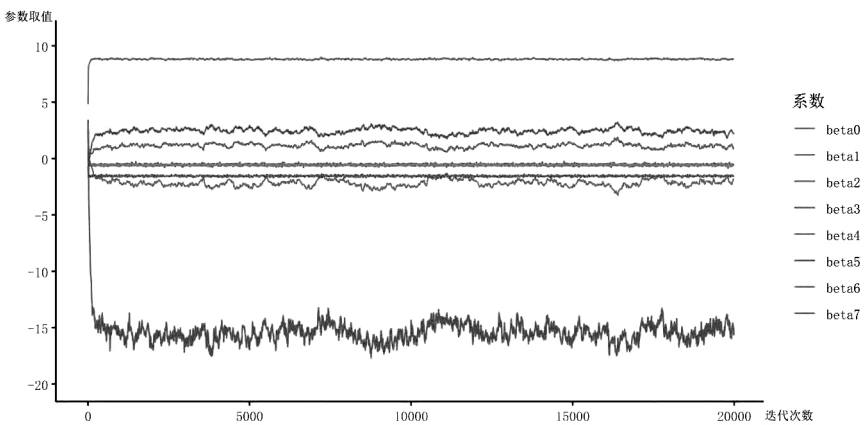
B)从条件概率分布 $p(\beta_1|\beta_{-1}^t,v^t,T)$ 中采样得到样本 β_1^{t+1}。

……

H)从条件概率分布 $p(\beta_7|\beta_{-7}^t,v^t,T)$ 中采样得到样本 β_7^{t+1}。

I)从条件概率分布 $p(v|\beta^t,T)$ 中采样得到样本 v^{t+1}。

样本集 $\{(\beta_k^{(0)},k=0,1,2\cdots7;v^{(0)}),\cdots\cdots,(\beta_k^{(n)},k=0,1,2\cdots7;v^{(n)})\}$ 即为 Gibbs 采样所获取的样本集。在迭代 20000 次后，获取了每个参数的 20000 个样本。由于初值是随机生成的，所以马尔可夫链的初始部分的值与最终收敛值可能存在较大差异，因此仅选取后 17000 个样本进行参数估计。每个参数样本的均值作为该参数的估计值，参数样本的标准差作为该参数的标准差。附图 1 是各参数的采样轨迹图。各参数在经历 3000 次迭代后达到较为稳定的状态，并围绕其均值上下波动。



附图 1　参数采样轨迹

注释：

①参见《中华人民共和国国民经济和社会发展第十四个五年规划和 2035 年远景目标纲要》，中国政府网，https://www.gov.cn/xin-wcn/2021-03/13/contcnt_5592681.htm。

②《网络安全审查办法》第七条：掌握超过 100 万用户个人信息的网络平台运营者赴国外上市，必须向网络安全审查办公室申报网络安全审查。

③中外文人名(机构名)对照：梅尔策(Meltzer)；洛夫洛克(Lovelock)；孙(Sun)；沙利文(Sullivan)；罗丁(Rawding)；萨克斯(Sacks)；阿伦森(Aaronson)；李(Li)；拉赫曼(Rahman)；李(Lee)；海斯利普(Haislip)；金(Kim)；权(Kwon)；森(Sen)；博尔(Borle)；赛奥(Seo)；鲍尔(Bauer)；米切尔(Mitchell)；米什拉(Mishra)；塞奇威克(Sedgewick)；梅斯纳(Messner)；罗森菲尔德(Rosenfeld)；卡伦(Cullen)；哈夫曼(Hövermann)；修斯(Hughes)；纳姆(Nam)；默滕斯(Mertens)；赫克特(Heckert)；张(Zhang)；班杜拉(Bandura)；沃尔特斯(Walters)；伊迈(Imai)；克里希纳(Krishna)；斯特劳布(Straub)；卡罗(Caro)；伦诺克斯(Lenox)；宋(Song)；瓦莱罗(Valero)；范里嫩(Van Reenen)；熊(Xiong)；马西米诺(Massimino)；陈(Chen)；泰勒(Taylor)；汤姆森(Thompson)；德赛(Desai)；霍

夫曼(Hoffman);奥卡西奥(Ocasio);马奇(March);夏皮罗(Shapira);赛曼(Saemann)。

参考文献:

[1]陈昌盛、许伟、兰宗敏、江宇:《"十四五"时期我国发展内外部环境研究》,《管理世界》,2020 年第 10 期.

[2]陈国青、曾大军、卫强、张明月、郭迅华:《大数据环境下的决策范式转变与使能创新》,《管理世界》,2020 年第 2 期.

[3]陈晓红、李杨扬、宋丽洁、汪阳洁:《数字经济理论体系与研究展望》,《管理世界》,2022 年第 2 期.

[4]高山行、刘伟奇:《数据跨境流动规制及其应对——对《网络安全法》第三十七条的讨论》,《西安交通大学学报(社会科学版)》,2017 年第 2 期.

[5]韩洪灵、陈帅弟、刘杰、陈汉文:《数据伦理、国家安全与海外上市:基于滴滴的案例研究》,《财会月刊》,2021 年第 15 期.

[6]胡炜:《跨境数据流动立法的价值取向与我国选择》,《社会科学》,2018 年第 4 期.

[7]黄丽华、朱海林、刘伟华、窦一凡、王今朝、蔡莉、陈煜波、廖貅武、吴晓波、谢康、叶强、张兮、陈文波:《企业数字化转型和管理:研究框架与展望》,《管理科学学报》,2021 年第 8 期.

[8]黄宁、李杨:《"三难选择"下跨境数据流动规制的演进与成因》,《清华大学学报(哲学社会科学版)》,2017 年第 5 期.

[9]李金、李建平、孙晓蕾、申苏浩、彭俊皓:《我国数据跨境传输的风险分析及应对策略》,《2020 社会管理重点问题研究报告》,科学出版社,2021 年 a.

[10]李金、申苏浩、孙晓蕾、邢潇:《重要数据跨境流动背景下风险路径的识别与分级》,《中国管理科学》,2021 年 b 第 3 期.

[11]李金、张黎明、李建平、邢潇:《跨境数据传出机构的风险分类管控和影响因素分析》,《系统科学与数学》,2022 年第 9 期.

[12]李墨丝:《CPTPP+数字贸易规则、影响及对策》,《国际经贸探索》,2020 年第 12 期.

[13]林润辉、谢宗晓、吴波、李大辉:《处罚对信息安全策略遵守的影响研究——威慑理论与理性选择理论的整合视角》,《南开管理评论》,2015 年第 4 期.

[14]马其家、李晓楠:《国际数字贸易背景下数据跨境流动监管规则研究》,《国际贸易》,2021 年 a 第 3 期.

[15]马其家、李晓楠:《论我国数据跨境流动监管规则的构建》,《法治研究》,2021 年 b 第 1 期.

[16]彭德雷、张子琳:《RCEP 核心数字贸易规则及其影响》,《中国流通经济》,2021 年第 8 期.

[17]覃庆玲:《数据出境安全评估若干问题探究》,中国网信网,http://www.cac.gov.cn/2019-06/26/c_1124675681.htm,2019 年.

[18]单文华、邓娜:《欧美跨境数据流动规制:冲突、协调与借鉴——基于欧盟法院"隐私盾"无效案的考察》,《西安交通大学学报(社会科学版)》,2021 年第 5 期.

[19]商务部:《中国数字贸易发展报告 2020》,2021 年 9 月.

[20]盛斌、高疆:《超越传统贸易:数字贸易的内涵、特征与影响》,《国外社会科学》,2020 年第 4 期.

[21]孙晓蕾、姚晓阳、杨玉英、吴登生、李建平:《国家风险动态性的多尺度特征提取与识别:以 OPEC 国家为例》,《中国管理科学》,2015 年第 4 期.

[22]田秀娟、李睿:《数字技术赋能实体经济转型发展——基于熊彼特内生增长理论的分析框架》,《管理世界》,2022 年第 5 期.

[23]王震:《陈晓红委员:加强我国跨境数据流动监管》,人民网,2022 年 3 月.

[24]魏远山:《博弈论视角下跨境数据流动的问题与对策研究》,《西安交通大学学报(社会科学版)》,2021 年第 5 期.

[25]习近平:《不断做强做优做大我国数字经济》,求是网,http://www.qsthcory.cn/dukan/qs/2022-01/15/c_1128261632.htm,2022 年 1 月.

[26]许多奇:《个人数据跨境流动规制的国际格局及中国应对》,《法学论坛》,2018 年第 3 期.

[27]许恒、张一林、曹雨佳:《数字经济、技术溢出与动态竞合政策》,《管理世界》,2020 年第 11 期.

[28]赵涛、张智、梁上坤:《数字经济、创业活跃度与高质量发展——来自中国城市的经验证据》,《管理世界》,2020 年第 10 期.

[29]郑登攀、张洁音、章丹:《区县政府科技投入、GDP 增长对企业研发支出的影响研究》,《科技管理研究》,2017 年第 18 期.

[30]中国信息通信研究院:《全球数字治理白皮书(2020)》,2020 年.

[31]中国信息通信研究院:《中国数字经济发展白皮书》,2021 年.

[32]Aaronson,S.,2015,"Why Trade Agreements Are Not Setting Information Free:The Lost History and Reinvigorated Debate over Cross-Border Data Flows,Human Rights,and National Security",World Trade Review,14(4),pp.671~700.

[33]Aaronson,S.A.,2016,The Digital Trade Imbalance and Its Implications for Internet Governance,Centre for International Governance Innovation.

[34]Aaronson,S.A.,2019,"Data Is Different,and That's Why the World Needs a New Approach to Governing Cross-Border Data Flows",Digital Policy,Regulation and Governance,21(5),pp.441~460.

[35]Bandura,A. and Walters,R.H.,1977,Social Learning Theory,Prentice Hall:Englewood cliffs.

[36]Bauer,M.,Ferracane,M.F. and Van Der Marel,E.,2016,Tracing the Economic Impact of Regulations on the Free Flow of Data and Data Localization,Centre for International Governance Innovation and Chatham House.

[37] Caro, F., Chintapalli, P., Rajaram, K. and Tang, C. S., 2018, "Improving Supplier Compliance Through Joint and Shared Audits with Collective Penalty", Manufacturing & Service Operations Management, 20(2), pp. 363 ~ 380.

[38] Chen, Y., Ramamurthy, K. and Wen, K. W., 2012, "Organizations´ Information Security Policy Compliance: Stick or Carrot Approach?", Journal of Management Information Systems, 29(3), pp. 157 ~ 188.

[39] Cullen, J. B., Johnson, J. L. and Parboteeah, K. P., 2014, "National Rates of Opportunity Entrepreneurship Activity: Insights from Institutional Anomie Theory", Entrepreneurship Theory and Practice, 38(4), pp. 775 ~ 806.

[40] Cullen, J. B., Parboteeah, K. P. and Hoegl, M., 2004, "Cross-National Differences in Managers´ Willingness to Justify Ethically Suspect Behaviors: A Test of Institutional Anomie Theory", Academy of Management Journal, 47(3), pp. 411 ~ 421.

[41] Desai, V. M., 2014, "The Impact of Media Information on Issue Salience Following Other Organizations´ Failures", Journal of Management, 40(3), pp. 893-918.

[42] Haislip, ., Lim, J. H. and Pinsker, R., 2021, "The Impact of Executives´ IT Expertise on Reported Data Security Breaches", Information Systems Research, 32(2), pp. 318 ~ 334.

[43] Heckert, D. M., 1998, "Positive Deviance: A Classificatory Model", Free Inquiry in Creative Sociology, 26(1), pp. 23 ~ 30.

[44] Hoffman, A. J., and Ocasio, W., 2001, "Not All Events Are Attended Equally: Toward a Middle-Range Theory of Industry Attention to External Events", Organization Science, 12(4), pp. 414 ~ 434.

[45] Hövermann, A. and Messner, S. F., 2021, "Institutional Imbalance, Marketized Mentality, and the Justification of Instrumental Offenses: A Cross National Application of Institutional Anomie Theory", Justice Quarterly, 38(3), pp. 406 ~ 432.

[46] Hughes, L. A., Schaible, L. M. and Gibbs, B. R., 2015, "Economic Dominance, the ´American Dream´ and Homicide: A Cross-National Test of Institutional Anomie Theory", Sociological Inquiry, 85(1), pp. 100 ~ 128.

[47] Imai, S. and Krishna, K., 2004, "Employment, Deterrence, and Crime in a Dynamic Model", International Economic Review, 45(3), pp. 845-872.

[48] Kim, S. H. and Kwon, J, 2019, "How Do EHRs and a Meaningful Use Initiative Affect Breaches of Patient Information?", Information Systems Research, 30(4), pp. 1184 ~ 1202.

[49] Lee, J. K., Chang, Y., Kwon, H. Y. and Kim, B., 2020, "Reconciliation of Privacy with Preventive Cybersecurity: The Bright Internet Approach", Information Systems Frontiers, 22(1), pp. 45 ~ 57.

[50] Lee, J. and Lee, Y., 2002, "A Holistic Model of Computer Abuse within Organizations", Information Management & Computer Security, 10(2), pp. 57 ~ 63.

[51] Lenox, M. J., Rockart, S. F. and Lewin, A. Y., 2006, "Interdependency, Competition, and the Distribution of Firm and Industry Profits", Management Science, 52(5), pp. 757 ~ 772.

[52] Li, J, Dong, W., Zhang, C. and Zhuo, Z., 2022, "Development of a Risk Index for Cross-border Data Movement", Data Science and Management, 5(3), pp. 97 ~ 104.

[53] March, J. G. and Shapira, Z., 1992, "Variable Risk Preferences and the Focus of Attention", Psychological Review, 99(1), pp. 172 ~ 183.

[54] Massimino, B., Gray, J. V. and Boyer, K. K., 2017, "The Effects of Agglomeration and National Property Rights on Digital Confidentiality Performance", Production and Operations Management, 26(1), pp. 162 ~ 179.

[55] Meltzer, J. P., 2015, "The Internet, Cross-Border Data Flows and International Trade", Asia & the Pacific Policy Studies, 2(1), pp. 90 ~ 102.

[56] Meltzer, J. P. and Lovelock, P., 2018, Regulating for a Digital Economy: Understanding the Importance of Cross Border Data Flows in Asia, Washington, DC: Brookings Institution

[57] Mertens, W., Recker, J., Kohlborn, T. and Kummer, T. F., 2016, "A Framework for the Study of Positive Deviance in Organizations", Deviant Behavior, 37(11), pp. 1288 ~ 1307.

[58] Messner, S. F. and Rosenfeld, R., 1997, "Political Restraint of the Market and Levels of Criminal Homicide: A Cross-National Application of Institutional-Anomie Theory", Social Forces, 75(4), pp. 1393 ~ 1416.

[59] Messner, S. F. and Rosenfeld, R., 2012, Crime and the American Dream, Belmont, CA: Wadsworth.

[60] Messner, S. F., Thome, H. and Rosenfeld, R., 2008, "Institutions, Anomie, and Violent Crime: Clarifying and Elaborating Institutional-Anomie Theory", International Journal of Conflict and Violence (IJCV), 2(2), pp. 163 ~ 181.

[61] Mitchell, A. D. and Mishra, N., 2019, "Regulating Cross-Border Data Flows in a Data-Driven World: How WTO Law Can Contribute´", Journal of International Economic Law, 22(3), pp. 389 ~ 416.

[62] Nam, D., Parboteeah, K. P., Cullen, J. B. and Johnson, J. L., 2014, "Cross-National Differences in Firms Undertaking Innovation Initiatives: An Application Of Institutional Anomie Theory", Journal of International Management, 20(2), pp. 91 ~ 106.

[63] Rahman, M. S., Al Omar, A., Bhuiyan, M. Z. A., Basu, A., Kiyomoto, S. and Wang, G., 2020, "Accountable Cross-Border Data Sharing Using Blockchain Under Relaxed Trust Assumption", IEEE Transactions on Engineering Management, 67(4), pp. 1476 ~ 1486.

[64] Rawding, M. and Sacks, S., 2020, "The Balkanization of the Cloud Is Bad for Everyone", MIT Technology Review, https://www.

technologyreview. . com/2020/12/17/1014967/balkanization-cloud-computing-bad-everyone/.

[65]Saemann, M. , Theis, D. , Urban, T. and Degeling, M. , 2022, "Investigating GDPR Fines in the Light of Data Flows", Proceedings on Priwacy Enhancing Technologies, 2022(4), pp. 314 ~ 331.

[66]Sedgewick, M. B. , 2017, "Transborder Data Privacy as Trade", California Law Review, 105(5), pp. 1513 ~ 1542.

[67]Sen, R. and Borle, S. , 2015, "Estimating the Contextual Risk of Data Breach: An Empirical Approach", Journal of Management Information Systems, 32(2), pp. 314 ~ 341.

[68]Seo, J. , Kim, K. , Park, M. and Lee, K. , 2017, "An Analysis of Economic Impact on IoT Under GDPR", 2017 International Conference on Information and Communication Technology Convergence (ICTC), pp. 879-881.

[69]Song, H. M. , Kim, H. R. and Kim, H. K. , 2016, "Intrusion Detection System Based on the Analysis of Time Intervals of CAN Messages for In-vehicle Network", 2016 International Conference on Information Networking (ICOIN), pp. 63 ~ 68.

[70]Straub, D. W. , 1990, "Effective IS Security: An Empirical Study", Information Systems Research, 1(3), pp. 255 ~ 276.

[71]Sullivan, C. , 2019, "Eu Gdpr or Apec Cbpr? A Comparative Analysis of the Approach of the Eu and Apec to Cross Border Data Transfers and Protection of Personal Data in the IoT Era", Computer Law & Security Review, 35(4), pp. 380 ~ 397.

[72]Sun, L. , Zhang, H. and Fang, C. , 2021, "Data Security Governance in the Era of Big Data: Status, Challenges, and Prospects", Data Science and Management, 2, pp. 41 ~ 44.

[73]Taylor, S. E. and Thompson, S. C. , 1982, "Stalking the Elusive 'Vividness' Effect", Psychological Review, 89(2), pp. 155 ~ 181.

[74]Valero, A. and Van Reenen, J, 2019, "The Economic Impact of Universities: Evidence from Across the Globe", Economics of Education Review, 68, pp. 53 ~ 67.

[75]Xiong, W. , Hu, H, Xiong, N. , Yang, L. T. , Peng, W. C. , Wang, X. and Qu, Y. , 2014, "Anomaly Secure Detection Methods by Analyzing Dynamic Characteristics of the Network Traffic In Cloud Communications", Information Sciences, 258, pp. 403 ~ 415.

[76]Zhang, Y. , Gao, C. and Wang, J. , 2022, "Financing Constraints and Innovation Performance: The Moderating Role of the Network Location of Cross-border Innovation Cooperation Among Internet Enterprises", European Journal of Innovation Management, ahead-of-print.

[77]Zhang, Z. , Nan, G. and Tan, Y. , 2020, "Cloud Services Vs. On-Premises Software: Competition under Security Risk and Product Customization", Information Systems Research, 31(3), pp. 848-864.

Risk Measurement and Analysis for Cross-Border Data Flow: An Empirical Study Based on Statistics of Outbound Data

Li Jin Xu Shan Zhuo Zihan Li Jianping

Abstract: As a new production factor, data contributes economic value through the constant cross-border data flows. However, the uncontrolled cross-border data flows may jeopardize personal, social and national securities. Thus, it needs to explore the governance system and resolve risk challenges in parallel, in order to better balance the economic development and security protection. Based on institutional anomie theory and deterrence theory, this study analyzes the influencing mechanism of cross-border data flows, and conducts empirical analyses through statistics of desensitized cross-border data in biomedical industry. It depicts the frequency and risk of cross-border data transmission behaviors by using three distinct measurement methods, namely the number, calculated risk value and time intervals of cross-border data transmission. The results show that the GDP of the province, the past cross-border data flows and the use of internet transmission positively affect both the frequency and risk of cross-border data flows, and the cloud storage adoption positively affects the risk value. Meanwhile, the implementation of relevant laws presents a significant deterrence effect on cross-border data transmission. This study defines the concept for a cross-border data transmission, designs the risk measurement methods, and identifies factors influencing the risk of cross-border data flows. It provides a theoretical basis and managerial insights for the regulation of cross-border data in China.

Key words: cross-border data; risk measurement; risk analysis; institutional anomie theory; deterrence theory