大数据应用如何抑制企业违规*

潘子成[1]　易志高[1]　柏淑嫄[2]

（1.南京师范大学商学院，江苏　南京　210023；
2.南京大学商学院，江苏　南京　210093）

内容提要：数字经济时代，大数据应用为公司治理创造机遇的同时也衍生出诸多挑战，其实践成效有待检验。本文以企业违规为切入点，基于2010—2020年沪深A股上市公司数据，系统探讨了大数据应用对企业违规的影响。实证研究发现，大数据应用会显著降低企业违规发生的概率和频次，以上结论经过一系列内生性分析和若干稳健性检验后仍然成立。而且，在区分大数据应用维度后，发现大数据应用深度、广度及二者联合效应可显著抑制企业违规，但深度的抑制效应最为明显。机制研究表明，大数据应用主要通过信息赋能和决策赋能两条路径抑制企业违规行为。异质性分析结果显示，大数据应用的违规抑制效应在非国有企业、高科技企业以及位于高水平营商环境地区的企业更为明显。经济后果检验还发现，违规企业再次发生违规行为的可能性更高，即企业违规存在“二皮脸”效应，但大数据应用则能有效弱化此效应。本文结论为数字经济时代为何要全力推进大数据发展应用以及如何优化公司治理提供了重要实践参考。

关键词：大数据应用　企业违规　大数据应用广度　大数据应用深度

中图分类号：F275.2　**文献标志码：**A　**文章编号：**1002—5766(2025)03—0126—23

一、引　言

当前，新一代数字技术正推动人类进入一个智能处理、精准决策的新时代，尤其是大数据技术的高速发展和迭代升级，使得数据的存储、传播、处理等呈指数级增长，奠定了大数据泛化应用和改变世界的基础。我国政府高度重视大数据发展，陆续颁布了《促进大数据发展行动纲要》等政策文件，旨在让大数据成为强信心、稳经济和促发展的新动力源。受此激励，众多企业主张依托大数据这一战略资源和新兴技术，着力推动大数据与经营业务融通融合、共促共进，藉此赋能企业高质量发展。与此同时，大数据应用给企业经营发展蓄力“赋能”的同时也带来了诸多“负能”，这根植于实践中所引发的组织结构冲突、大数据伦理失范、经营风险加剧等一系列问题。除此之外，众多企业在数字化应用过程中或多或少面临成本过高、人才短缺、技术鸿沟、算法寻租等诸多挑战，致使大数据对公司治理和价值增长的赋能效果并不明显。据Gartner所发布的报告显示，超过半数公司运用大数据等数字技术而未给组织带来相应效益。可见，大数据应用的实践成

收稿日期：2023-10-07

*** 基金项目：**国家社会科学基金重点项目“企业数据资产信息披露行为选择、经济后果与对策研究”(24AJY034)；国家自然科学基金项目“国际化经验是否提升了国际化绩效——一项多层次整合研究”(71972104)；教育部人文社科青年基金项目“企业数据资产对绿色创新的影响机理及策略研究：基于资源获取与资源配置的双重视角”(24YJC630161)。

作者简介：潘子成，男，讲师，管理学博士，研究领域为数字化与公司治理，电子邮箱：pzctgedu@163.com；易志高，男，管理学博士，教授，博士生导师，研究领域为企业数字化、公司财务与金融，电子邮箱：shuapi@126.com；柏淑嫄，女，博士研究生，研究领域为公司治理与社会责任，电子邮箱：1339243840@qq.com。通讯作者：易志高。

效还有待检验。

当前,既有文献从多维视角对大数据应用的经济后果展开了有益探索。研究表明,大数据应用能够优化经营决策(George 等,2014)[1],助推创新升级(张明超等,2021)[2],增加市场价值(张叶青等,2021)[3]等;但也有学者指出,大数据应用会带来信息过载(Saxena 和 Lamest,2018)[4],引发数据伦理失范(凡景强和邢思聪,2023)[5],且因“信息茧房”、大数据依赖,致使创新决策失灵(Ghasemaghaei 和 Calic,2020)[6]等。尽管上述文献为大数据应用经济后果研究提供了相对丰富的素材,但关于大数据应用的有效性尚未达成一致意见,且鲜有文献专门讨论大数据应用对企业违规的影响,这一理论缺口为本研究提供了切入点。实际上,财务造假、股价操纵等违规行为在全球范围内持续存在并有蔓延趋势,其中,奥林巴斯财务造假案、康得新虚构利润案、瑞幸咖啡财务舞弊案等案件的曝光,更是引起了资本市场对企业违规的热切关注。在此情形下,如何防范和治理企业违规行为成为业界不容忽视的重要议题。实际上,面对大数据快速迭代和广泛应用的现实情境,将大数据等新兴数字技术融入公司治理的各个环节或许是现代企业管理完善和优化的前进方向。譬如,埃森哲、海尔等知名企业纷纷借助大数据前沿技术搭建财务共享平台和智能风控系统,辅助企业开展违规核查与风险自检,旨在实现各类违规行为的数字化管理。在此背景下,并结合当前研究结论尚未统一的现状,本文提出以下问题:大数据应用对企业违规的影响究竟如何?其背后的作用机制是什么?又会受到哪些因素影响?

为回答以上问题,本文以 2010—2020 年沪深 A 股非金融类上市公司为研究样本,系统探讨了大数据应用对企业违规行为的影响效应。本文可能的边际贡献如下:第一,当前相关文献更多探讨大数据的宏观经济效用,而探讨其微观经济行为的文献相对不足,且研究结论关于大数据应用对公司治理是否有效尚无定论。本文创新性地从企业违规视角探索了大数据应用的公司治理效果,并从信息赋能和决策赋能两个层面解释了大数据应用抑制企业违规的影响机制。以上结论为大数据应用的经济后果和企业违规影响前因提供了新的文献补充,并为大数据应用的公司治理有效性贡献了可信的经验证据。第二,既往文献更倾向于从整体层面论证大数据应用的影响效应,本文打破了将大数据应用进行同质化处理的研究惯性,将大数据应用细分为广度和深度两个维度,并探索了二者及其联合效应对企业违规的差异化影响,这不仅是对大数据应用研究的有益补充,且增进人们对于大数据应用的全景式认知。第三,关于企业违规前因的探索是多维复合的,相关研究从高管特征、公司治理、外部监管等多方视角讨论了企业违规的影响因素。本文则将大数据应用这一前沿数字技术纳入企业违规的研究框架,系统探讨了大数据应用对企业违规机会的抑制效应。另外,本文还发现,违规企业再次发生违规行为的可能性更高,即企业违规具有“二皮脸”效应,而大数据应用则可抑制此效应。上述研究结论丰富了企业违规的前因和后果,并对业界关于为何以及如何抑制企业违规行为贡献了新的技术线索。

二、文献综述与研究假设

1.文献综述

(1)大数据应用概念及经济后果。当前,业界对于大数据的概念尚未达成一致意见。Gartner 咨询公司将大数据定义为海量、高速、多样化的信息资产。我国学者则将大数据定义为在各种新兴信息技术支撑下,针对海量、复杂、散乱的数据,利用全新的数据分析处理方法来智能发现有价值信息的技术统称(史丹,2019)[7]。不难发现,大数据既是一类数据,也是一项技术。基于数据特征视角,大数据是具有独特性质的数据集合,具备规模性(volume)、多样性(variety)、高速性(velocity)、价值性(value)、准确性(veracity)的“5V”特征(McAfee 等,2012[8];安小米等,2023[9])。基于技术视角,大数据作为新一代信息系统架构和技术,能够对数量巨大、来源分散、格式多样的数

据进行采集、存储，并进行关联性分析（Oussous等，2018）[10]。总的来看，数据特征视角强调在传统数据概念基础上附加新的特征，用于交易并创造出更大的价值，属于静态观分析。技术视角则主张通过技术方式把大数据转化为竞争力，属于动态观分析（迟考勋等，2022）[11]。

当前，诸多学者就数据应用的价值作用展开了系统性研究。一方面，部分学者从技术或数据等层面肯定了大数据的“赋能”效应，认为大数据作为划时代的战略资源和核心技术，具有容量大、速度快和高价值的典型特征，在降低信息获取成本、优化经营决策、推动技术进步等方面作用显著（George等，2014[1]；Li等，2022[12]；任曙明和马橙，2023[13]），从而赋能微观企业价值增长和宏观经济高质量发展。譬如，杨俊等（2022）[14]构建了一个多部门的熊彼特质量阶梯模型，理论阐述了大数据通过乘数作用和研发模式转型驱动技术进步，进而持续推动经济增长。肖静华等（2020）[15]基于案例研究发现，大数据为企业与用户互动创新提供了新的创新环境、工具和方法，从而引导创新生态。张叶青等（2021）[3]基于A股上市公司的研究数据发现，企业大数据应用程度能够提高公司生产效率和研发投入从而显著提高企业市场价值。另一方面，也有部分学者认为，大数据应用是一项复杂的、系统的工程，需构建与大数据应用相适应的制度、流程、架构、技术等（张叶青等，2021[3]；杨德明等，2020[16]）。但在实践过程中，大数据应用会冲击企业旧有惯例，引发组织不确定性，形成信息茧房，带来技术利维坦风险①等，倘若处置不当则可能成为大迷惑、大忽悠、大混乱（李金昌等，2020）[17]，从而给组织带来诸多“负能”（刘宇琪和秦宗文，2023）[18]。譬如，杨德明等（2020）[16]明确指出，大数据应用会给企业带来包括战略风险、运营风险和财务风险等。凡景强和邢思聪（2023）[5]认为，大数据往往会引发数据鸿沟、算法偏见、隐私权侵犯、数据垄断等一系列伦理失范问题，从而冲击各行业固有的伦理价值体系。张海丽等（2022）[19]认为，过高的大数据可能会导致信息过载、增加创新信息吸收和利用难度、提升创新复杂性等，从而降低产品创新度。

（2）企业违规的影响因素研究。企业违规本质上是一种机会主义行为，涉及虚构利润、虚列资产、虚假记载、推迟披露、操纵股价等多个方面，不仅侵犯了大小投资者的利益，同时阻碍了资本市场的健康发展。基于舞弊三角理论，违规行为通常是由动机（压力）、机会以及借口三个要素共同作用的结果（Homer，2020）[20]。其中，动机指的是舞弊决策者的逐利倾向和意愿；机会指的是实施舞弊行为的条件和环境；借口指的是实施舞弊行为时找到的合理化理由。舞弊三角理论为企业违规行为研究构建了基础分析框架，后续研究者沿袭此理论逻辑，从三个维度、多方视角为违规行为的形成动因和应对措施寻求理论解释。具体地，从机会维度看，部分学者基于组织监管环境、管理层权力等视角具体探索了企业违规舞弊行为的影响因素，如巡回法院设立（Cao等，2023）[21]、分析师预测偏差（Ren等，2021）[22]、社会信用环境（黄卓等，2023）[23]、高管本地任职（吴雅倩等，2024）[24]。从动机维度看，部分学者基于责任、利益、声誉、高管个体特征视角深度剖析了企业违规舞弊行为的前置因素，如企业社会责任（Liao等，2019）[25]、高管股权激励（Hass等，2016）[26]、员工持股计划（张学志等，2022）[27]、外部薪酬差距（朱沛华等，2024）[28]等。相对于机会维度和动因维度，借口维度的研究相对孱弱，相关学者主要从道德辩护、责任转移等视角为企业舞弊行为寻求原因线索，如审计师变更（Parlindungan等，2017）[29]、组织认同程度（陈邑早等，2020）[30]等。此外，也有学者结合多个维度，从董事选举得票率（周泽将等，2022）[31]、长征精神（吴武清和洪振瀚，2024）[32]等研究视角分析企业违规舞弊现象产生的根源。

纵观上述文献，现有研究可能存在以下不足。一是国内外学者对大数据应用经济后果的探索

① 技术利维坦风险是指技术力量形成的高度集中化、强大且可能失控的权力体系，从而对社会、个体和伦理带来的系统性威胁。

主要是以理论思辨和定性研究居多，关于大样本的实证探索还相对薄弱，而基于公司治理视角对企业大数据展开大样本实证研究的文献更是明显不足。而且，当前较多学者对大数据应用的有效性进行了多方法、多视角检验，但对其究竟带来“数字赋能”效应还是“数字负能”效应并未达成一致意见。特别地，我国正处于工业经济迈向数字经济的关键时期，而大数据作为数字经济最为关键的要素之一，对这一问题的模糊可能不利于后疫情时代数字经济的高速发展。二是关于企业违规前因的探索是多维复合的，区别于监管环境、区域文化等外部因素以及公司特征、高管特征等内部因素等以局部、短期或者不连贯的方式影响企业违规行为，真正低成本、高效率、可持续约束企业违规行为的底层逻辑和长期动能在于技术的发展和进步。大数据作为数智时代最重要的技术内核，学术界却尚未关注到大数据应用对公司违规行为的影响，更鲜有研究关注二者关系的内在机制和权变情境。尤其是在我国大数据应用场景不断增多和企业违规现象屡禁不止的现实背景下，上述研究疏漏亟需补充。

2.研究假设提出

基于舞弊三角理论，高管作为企业关键决策主体，其是否违规以及如何违规取决于动机、机会和借口三要素（Homer，2020）[20]。面对日趋严苛的监管环境，行为人即便存在舞弊压力和合理化缘由，但受制于机会因素也不会轻易从事舞弊行为，即机会被视为引发违规行为的关键因素。研究指出，违规机会取决于内部管理机制失效和外部监管机制失灵（王伊攀和朱晓满，2022）[33]。本文认为，大数据作为新一代信息数字技术，能够对数量巨大、来源分散、格式多样的数据进行采集、存储并进行关联性分析（Oussous等，2018）[10]，因而在应用实践过程中，可有效实现信息赋能和决策赋能，弥补内部治理和外部监管的不足，进而减少管理者实施违规行为的机会。

一是大数据应用提高了信息披露质量，实现信息赋能。从数据流动的视角看，大数据应用能够突破时间空间限制，提供从数据获取、整合、分析到应用的一揽子服务，并协同企业以较低的成本实现数据全生命周期管理，从而形成更完整的数据集市（Jin等，2015[34]；张叶青等，2021[3]）。譬如，大数据技术可通过对用户行为、社交关系等多维数据开展分析，更好地探查用户特征与个性化需求，藉此形成企业专有的“数据湖”。而企业内部数据的横向集成及企业间数据的纵向聚集，带来了数据的完整性、及时性、准确性，致使组织信息环境和治理模式发生诸多变化（McAfee等，2021）[8]。具体表现为：一方面，企业内外信息传递成本程度显著降低，信息处理效率不断提高，利益相关者可实施信息互联互通，实现数据集合共建共享，从而降低信息不对称程度；另一方面，大数据应用使得组织监管由人力监督转变为数据技术监督，由事后监督转变为过程监督，投资者等监管主体可依靠及时、完整的信息精准把握公司经营状况，倒逼企业管理者把权力关进“数据牢笼”。

舞弊三角理论指出，不对称的信息环境为管理层、股东等内部人利益侵占或违规行为提供了机会，他们会低估违规行为被发现可能性，并倾向性地将违规动机转化为违规事实（董盈厚等，2021）[35]。而基于大数据应用驱动的内外信息环境改善，不仅会促使利益相关者及时掌握监管动态并对监管行为形成有效预期，从而提升内部控制质量（李文贵和邵毅平，2022）[36]；而且会协助监管主体更为充分、全面地了解公司经营投资等方面的重点事项，进而参与公司治理并形成有效监督（Ren等，2021）[22]。内控质量和监管力度的提升，不仅能激发或唤醒个体权责意识，强化责任主体的道德观念，进一步明确权力边界；还能促使监管主体有效挖掘、识别、追踪企业财务舞弊、隐藏信息等违规行为，发挥事前威慑、事后惩罚效果，进而降低高管违规意愿和收益（张学志等，2022）[27]。此时，基于理性人假设，高管或股东有动机从源头减少违规行为，从而降低违规被稽查的概率。

二是大数据应用提升了企业决策质量，实现决策赋能。大数据能够帮助企业实现以事实为中

心的经营模式，逐步从对静态现象的分析预测，过渡到针对场景提供动态的决策建议，从而实施精准决策。首先，信息理论指出，充足信息能够提高企业决策效率。面对海量、多源的数据源，除了通过大数据技术进行数据采集、清洗外，还可对数据进行分析和挖掘，解构出数据间的关联，预测事物后续的发展趋势并形成行动路线规划，从而为管理者精准决策提供效率支撑（Janssen等，2017[37]；Merendino等，2018[38]）。譬如，在营销领域，企业可通过数据挖掘、机器学习等大数据技术，分析顾客消费行为，萃取客户价值主张，并迭代消费评估模型，由此制定精准营销方案。此外，大数据分析模型如数据分析模型、机器学习模型等被广泛使用，这些模型可通过逐条规则的设立及其关联性分析得到决策评估预测，并在实施过程中对决策方案进行实时感知、跟踪监测、态势分析及反馈优化，从而不断提高决策和执行效能（Li等，2012）[12]。可见，大数据应用可助力管理者植入大数据思维，摆脱决策信息有限的束缚，从而大大提高决策质量。

既有研究指出，低效率、过大偏误或无效决策经过逐级积累、放大，易引发组织运转紊乱、委托代理冲突加剧等，最终会增加内部人“钻空子”的机会，并引发披露不实、延迟披露等寻租套利行为（Jing，2023）[39]。诸多研究均表明，过度投资、过度金融化等低效投资决策易引发内部人盈余操纵、信息捂盘等违规行为（Jing，2023[39]；彭俞超等，2018[40]）。由此可知，决策质量或效率的提升，则可有效遏制内部人因决策不当而从事的违规行为，并进一步推动组织治理体系的系统性优化。其结果是：一方面有助于制衡管理层自由裁量权，减少内部人对以自身利益最大化为目标的决策方案的偏向性选择，既而约束管理层自利行为（Rashid等，2022）[41]；另一方面有助于曝光、识别内部人在决策管理过程中存在的可疑违规操作，强化外部监管者对于决策不当或者违规经营的预判和防控，这在实现有效决策监督的同时可降低高管违规机会。此时，内部人则有意愿规范和约束自身言行，倾向性减少虚构利润等违规行为的发生。综上，本文提出如下假设：

H_1：大数据应用能够抑制企业违规行为。

企业大数据应用是一项系统工程，面临多方面的痛点和难点，这倒逼企业需打出一套“点、面”结合的系统推进组合拳，进而更为充分地释放大数据潜能。其中，“点”思考的是突出重点，强调大数据技术在各业务场景和流程中的渗透性和融合程度，体现的是大数据技术应用的深度。“面”考虑的是顾及全局，强调在各业务场景和流程中所采用大数据技术的覆盖范围，体现的是大数据技术应用的广度。本文认为，大数据应用广度和深度分别代表了企业大数据技术的横向覆盖维度和纵向垂直维度，会对企业违规行为产生重要影响。

大数据应用广度反映企业应用大数据技术的种类和组合，体现了大数据技术的多样性和跨界性。一般而言，大数据应用广度高的企业对算力、数据挖掘等大数据技术都有涉及，多样性的大数据技术使得企业不再拘泥于一种技术路径，而是能尝试多元技术路径捕捉来自不同渠道的数据信息，并实现企业信息的交叉验证、可追溯等。而数据多源性和准确性的提升，有助于缓解信息不对称问题（Calic和Ghasemaghaei，2021）[42]，增强监管主体对企业违规活动的监督力度，进而及时捕捉管理层异常行为和识别潜在风险。另外，当企业掌握更多大数据技术组合，其更可能打破技术壁垒，实现技术互补交叉融合，在技术迭代演进中探索出全新的应用模式，最终产生更为明显的大数据赋能效应。在此情形下，除缓解信息不对称外，企业还可实现跨领域关联分析，制定出更明智的决策，从而减少因决策偏误而引发的违规套利行为。

大数据应用深度反映企业应用大数据技术的熟悉程度，体现了大数据技术的渗透性和专业性。当企业专注于某些特定的大数据技术并在实践中得以应用、发展和积累后，往往具备更为明显数据优势和技术便利，并可逐步沉淀为专业化的信息搜集和处理能力（陈德球和胡晴，2022）[43]。此时，企业可高效地收集、处理各类信息数据，改善组织内外部信息环境，从而降低高管实施违规行为的机会。而且，大数据应用深度也体现出企业更为强大的数据整合和关联分析能力，依托前

沿的大数据算法和模型，企业可实现更深入的数据挖掘和精细化的数据处理，并据此揭示出数据背后的复杂关系和细微模式。这种深度分析可辅助管理层面优化经营决策，降低主观判断和人为操纵，进而提高合规监管的效果。此外，大数据应用深度还体现其持续的自主学习与自适应能力，依托大数据模型，通过在不同场景下对大量历史数据的训练和学习，企业可逐渐提升其违规风险的预测、识别能力，不断强化违规风险预警的实时性和敏捷性，从而帮助企业揭示出潜在的违规风险点。

大数据应用广度和深度除单独对企业违规产生影响外，也存在共同作用于企业违规活动的可能性。究其原因，在实践过程中，大数据应用广度和深度绝非割裂存在的，而是相互作用和相互影响的。基于大数据开放性、渗透性、共享性和协同性等特征，大数据应用深度和广度的双重提升，一方面能够通过技术扩散和资源互补，为大数据应用实践提供更为充分的技术供给；另一方面能够重塑技术范式，推动技术间渗透融合，增强组织内大数据技术多样性和渗透性的协同效应，进而充分释放大数据应用对公司治理的赋能效果。在此情形下，大数据应用则可在降低信息不对称、优化组织决策等方面发挥更为积极的作用，由此呈现出显著的大数据应用违规抑制效应。

综上，本文提出如下假设：

H_2：大数据应用广度、深度及其联合效应能够抑制企业违规行为。

三、研究设计

1.样本选择与数据来源

本文以2010—2020年沪深A股市上市公司为初选样本①，并按照以下标准进行样本筛选：剔除ST、PT类样本公司；剔除金融、保险类样本公司；剔除资不抵债的样本公司；剔除关键数据缺失和极端异常的样本公司。最终，获得25531个公司-年度观测值。大数据应用数据由Python软件从企业年报中抓取关键词获取，上市公司违规及其他财务数据主要来源于CSMAR和CNRDS数据库。本文还对连续变量进行了上下1%水平的缩尾处理以排除异常值影响。

2.变量定义

（1）企业违规。根据监管部门惩处公告看，违规行为主要包含包括虚构利润、虚假记载等15种违规类型②。借鉴Ren等（2016）[22]、张学志等（2022）[27]的做法，本文采用是否违规哑变量（*Fraud_Dum*）和违规次数（*Fraud_Fre*）来衡量企业违规。*Fraud_Dum*为虚拟变量，定义为若目标公司当年发生违规行为赋值为1，否则为0。*Fraud_Fre*为连续变量，定义为当年公司发生的违规次数之和并取自然对数。

（2）大数据应用。本文参考张叶青等（2021）[3]、赵宸宇等（2021）[44]的做法，以与大数据应用相关的关键词在年报中出现的词频来衡量大数据应用。具体地，首先，借助Python爬虫技术归整样本企业年报信息，并通过Java PDFBox文档协作类库提取所有文本内容，为后续特征词的筛选提供基础数据池。其次，基于大数据的政策文件和管理实践，并参考吴非等（2021）[45]、赵宸宇等（2021）[44]的思路，从大数据技术开发和大数据技术运用两个维度总结归纳与“大数据”有关的关键词，其中，与大数据技术开发相关的关键词包括：大数据、数据挖掘、文本挖掘、数据可视化、异构数据等；与大数据技术运用相关的关键词包括：数据管理、数据网络、数据平台、数字金融等，以此构

① 为缓解可能存在的互为因果问题而导致的内生性偏误，本文使用下一期（t+1期）的企业违规作为被解释变量，因此企业违规的样本区间为2011—2021年。

② 根据证监会规定，违规行为主要包括信息披露违规、经营违规和管理层违规违规三种违规类型，其中，信息披露违规包括虚构利润、虚列资产、虚假记载、推迟披露、重大遗漏、披露不实、一般会计处理不当等7个次类；经营违规包括出资违规、擅自改变资金用途、占用公司资产、违规担保和其他等5个次类；领导人违规包括内幕交易、违规买卖股股票、操纵股价等3个次类。

建特征词库。再次，基于基础数据池，对上述关键词进行搜索、匹配，并删除特征词前存在“没”“不”“否”等否定词语的表述，删除供应商、客户及高管简介部分含有大数据关键词的语句。最后，统计出关键词在年报中出现的频次并取自然对数，以此作为大数据应用水平的测度指标，记为*Bigdata*。在此基础上，本文采用年报中出现大数据应用关键词的种类多少衡量大数据应用广度，记为*Bd_Breath*；采用年报中与大数据技术运用相关的关键词词频的自然对数衡量大数据应用深度，记为*Bd_Depth*。

借鉴张叶青等（2021）[3]的思路，本文还对所构建的大数据应用指标的有效性进行检验。从应用投入看，企业为更好地收集和处理客户营销、生产流程端等产生的数据资源，会增加软件购买、网络平台搭建、管理智能系统等数字化投资支出。换言之，企业需加大数字化投资力度夯实大数据应用的技术基础。从应用成果看，大数据在应用过程中势必会涌现出各种新兴技术，引发不同规模的商业运用。为提升商业价值，企业则有动机通过专利申请保护来构筑市场壁垒。参考陈楠等（2024）[46]的思路，第一，本文基于国家知识产权专利数据库，对上市公司大数据应用相关技术专利的IPC分类号①进行手工整理和筛选，获取相应的大数据专利，构建基础专利库；第二，基于前文做法，总结归纳与大数据有关的关键词，构建词库；第三，结合此词库，对基础专利库中专利申请书的摘要、说明书和权利要求书进行文本分析；第四，核算出样本企业各个年度大数据专利申请量。在此基础上，本文拟将大数据应用指标与数字化投资支出、大数据专利申请量进行相关性分析，结果表明，相关系数分别为0.270、0.320，且均达到了1%显著性水平。总的来看，本文构建的大数据应用指标具有一定的科学性和有效性。

（3）控制变量。参考以往研究（Cao等，2023[21]；李文贵和邵毅平，2022[36]），本文控制以下变量：净资产收益率、负债水平、企业规模、独立董事比例、第一大股东持股比例、机构投资者持股比例、公司成长性、两职合一、固定资产比重、现金流量、审计师类型、股权制衡度、公司年限、两权分离。另外，本文还控制了时间、行业和省份固定效应。

具体变量定义如表1所示。

表1　　主要变量定义

变量名	变量符号	变量说明
违规倾向	*Fraud_Dum*	如若目标公司发生违规行为，则赋值为1，否则为0
违规频次	*Fraud_Fre*	目标公司违规次数之和，并取对数
大数据应用	*Bigdata*	大数据相关关键词在年报中出现的次数，并取自然对数
大数据应用广度	*Bd_Breath*	年报中出现大数据应用关键词的种类
大数据应用深度	*Bd_Depth*	年报中与大数据技术运用相关的关键词词频，并取自然对数
净资产收益率	*Roe*	净利润/平均总资产
负债水平	*Lev*	负债/资产总额
公司规模	*Size*	公司总资产，取自然对数
独董比例	*Indep*	独立董事人数/董事总人数
第一大股东持股比例	*First*	公司第一大股东持有公司股票数量/公司总股数
机构持股比例	*Inst*	机构投资者持股数量/公司总股数
公司成长性	*Growth*	营业总收入增长率
两职合一	*Dual*	董事长同时兼任CEO赋值为1，否则为0.

① 结合国家统计局发布的《数字经济及其核心产业统计分类（2021）》以及国家知识产权局发布的《国际专利分类定义（2023.01版）》，本文识别出大数据应用所属的技术领域及其对应的IPC代码，最终总结出包括A61B34/10、G06F21/00、G06Q10/04、G16C20/10、G08B13/196、G10L13/00、H04L12/18、H04W64/00等在内的16种IPC技术小类、44种IPC技术小组分类。

续表 1

变量名	变量符号	变量说明
固定资产比重	*PPE*	固定资产总额/总资产
现金流量	*Cashflow*	企业营业活动产生的现金流量净额/营业收入
审计师类型	*Big*4	公司当年被国际四大会计师事务所审计取值为 1,否则为 0
股权制衡度	*Balance*	第二到第五大股东持股比例之和/第一大股东持股比例
审计费用	*AuditFee*	公司公开披露的审计费用,并自然对数
公司年限	*Age*	公司成立的年份数,取自然对数
两权分离	*Seperate*	实际控制人控制权与现金流权的差值

3.模型设计

为检验大数据应用对企业违规行为的影响,构建以下模型:

$$Fraud = \alpha_0 + \alpha_1 Bigdata + Controls + \sum industry + \sum year + \sum region + \varepsilon \quad (1)$$

其中,*Fraud* 为企业违规,分别采用 *Fraud_Dum* 和 *Fraud_Fre* 进行衡量。其中,当以 *Fraud_Dum* 和 *Fraud_Fre* 作为被解释变量时,分别采用 Logit 回归和 OLS 回归。*Bigdata* 为企业大数据应用。*Controls* 为控制变量。$\sum industry$、$\sum year$ 和 $\sum region$ 分别为行业、年份和省份虚拟变量。ε 表示残差。另外,考虑到互为因果而导致的内生性偏误以及大数据应用效果的时滞性,本文将所有解释变量滞后一期。

四、实证结果分析

1.描述性统计

表 2 为描述性统计结果。*Fraud_Dum* 的均值为 0.193,表明在样本期间,19.3% 的样本公司存在违规行为。*Fraud_Fre* 的均值为 0.226,标准差为 0.503,标准差明显大于均值,说明样本企业违规频次存在较大差异。*Bigdata* 的均值为 1.026,中位数为 0.693,说明大数据应用水平呈右偏分布,大部分样本企业大数据应用程度相对较低。控制变量的值均在合理范围内。

表 2　　描述性统计结果

变量	均值	标准差	中值	最小值	最大值
Fraud_Dum	0.193	0.395	0.000	0.000	1.000
Fraud_Fre	0.226	0.503	0.000	0.000	2.079
Bigdata	1.026	1.111	0.693	0.000	4.522
Bd_Breath	1.247	1.510	1.000	0.000	7.000
Bd_Depth	0.748	0.894	0.693	0.000	3.738
Roe	0.065	0.123	0.070	−0.570	0.358
Lev	0.432	0.205	0.426	0.055	0.884
Size	22.197	1.258	22.033	19.878	26.040
Indep	0.375	0.054	0.357	0.333	0.571
First	0.348	0.147	0.327	0.093	0.743
Inst	0.454	0.248	0.473	0.003	0.931
Growth	0.172	0.423	0.104	−0.563	2.732
Dual	0.262	0.440	0.000	0.000	1.000
PPE	0.215	0.161	0.183	0.002	0.698

续表 2

变量	均值	标准差	中值	最小值	最大值
Cashflow	0.046	0.069	0.045	−0.159	0.238
*Big*4	0.053	0.225	0.000	0.000	1.000
Balance	0.695	0.587	0.530	0.026	2.706
AuditFee	13.649	0.636	13.554	12.468	15.825
Age	2.876	0.328	2.944	1.792	3.497
Seperate	4.819	7.467	0.000	0.000	28.249

2.大数据应用与企业违规基准回归分析

表 3 列示了针对假设 H_1 的回归结果。为提高实证结果可信度，本文在纳入核心解释变量和多维固定效应的基础上，采取逐步增加控制变量方式进行参数估计。结果显示，无论是以 *Fraud_Dum* 还是 *Fraud_Fre* 作为被解释变量，*Bigdata* 的回归系数均显著为负。从经济意义上看，以第(2)列和第(4)列为例，保持其他变量不变，年报中与大数据应用相关的词频每增加一个单位①，下一年违规行为发生的概率将下降 0.099%，违规频次将下降 0.168%。以上证据表明，随着大数据应用强度的提升，企业发生违规行为的概率和频次显著下降，即大数据应用对企业违规行为能够产生“赋能”效果。假设 H_1 得以验证。

表 3　　大数据应用与企业违规实证回归结果

变量	*Fraud_Dum*	*Fraud_Dum*	*Fraud_Fre*	*Fraud_Fre*
	(1)	(2)	(3)	(4)
Bigdata	−0.033*	−0.045**	−0.009**	−0.011***
	(−1.79)	(−2.40)	(−2.49)	(−3.06)
Roe		−1.957***		−0.650***
		(−13.36)		(−16.30)
Lev		1.039***		0.225***
		(9.47)		(10.47)
Size		−0.218***		−0.043***
		(−8.72)		(−9.13)
Indep		−0.058		0.047
		(−0.18)		(0.78)
First		−0.946***		−0.115***
		(−4.76)		(−3.39)
Inst		−0.303***		−0.059***
		(−3.15)		(−3.45)
Growth		0.043		0.007
		(1.00)		(0.78)
Dual		0.171***		0.027***
		(4.46)		(3.50)

① 由结果可知，大数据应用词频每增加 1%，下一年违规行为发生的概率值将下降 0.0065%（$e^{0.045}-1$）。统计发现，样本企业大数据应用词频的均值为 6.539，由此可推，大数据应用词频每增加一个单位，则下一年违规行为概率则降低 0.0994%（=1÷6.539×0.0065%×100）。同理，大数据应用词频每增加一个单位，违规频次下降 0.1682%。

续表 3

变量	*Fraud_Dum*	*Fraud_Dum*	*Fraud_Fre*	*Fraud_Fre*
	(1)	(2)	(3)	(4)
PPE		−0.225 (−1.60)		−0.080*** (−3.18)
Cashflow		−0.803*** (−2.85)		−0.069 (−1.32)
*Big*4		−0.647*** (−6.32)		−0.102*** (−8.12)
Balance		0.107*** (2.69)		0.031*** (3.90)
AuditFee		0.466*** (10.45)		0.097*** (11.42)
Age		−0.105 (−1.64)		−0.010 (−0.89)
Seperate		0.008*** (3.07)		0.001*** (3.44)
常数项	−1.614*** (−5.95)	−3.002*** (−5.00)	0.080 (0.64)	−0.318** (−2.04)
年份/行业/省份固定效应	是	是	是	是
观测值	25531	25531	25531	25531
伪 R^2/R^2	0.029	0.069	0.029	0.084

注：***、**和*分别表示在1%、5%和10%水平下显著；括号中为*t*值。下同

3.大数据应用的不同维度与企业违规回归分析

为检验假设 H_2，本文在基准模型(1)基础上，将解释变量更换为大数据应用广度、深度及其交互项，具体实证结果如表4所示①，当以 *Fraud_Dum* 为被解释变量时，*Bd_Breath*、*Bd_Depth* 和 *Bd_Breath*×*Bd_Depth* 的回归系数均显著为负，系数大小分别为−0.013、−0.023、−0.013；当以 *Fraud_Fre* 为被解释变量时，*Bd_Breath*、*Bd_Depth* 和 *Bd_Breath*×*Bd_Depth* 的回归系数分别为−0.019、−0.029、−0.013，且均至少达到了5%的显著性水平。这些经验证据表明，大数据应用广度、深度及其联合效应对企业违规行为均具有抑制作用，假设 H_2 得以验证。

需要说明的是，Wald检验表明，*Bd_Depth* 与 *Bd_Breath*、*Bd_Breath*×*Bd_Depth* 系数差异的经验*p*值均小于0.1，即拒绝回归系数相同的原假设。以上结果表明，大数据应用深度对企业违规的抑制效应要大于广度以及广度和深度的联合效应。究其原因，一是大数据应用广度体现了大数据技术应用的多样性、跨界性，虽能覆盖多个领域，解决大数据应用的普惠问题，但面对企业违规行为日趋多样化、复杂化和隐蔽化的新形势，其在企业违规风险识别、内控、监管等方面的精准性和有效性上可能稍逊一筹。二是在实践操作中，由于数据类型多样、组织资源约束以及数字技术局限等，在广度和深度间找到平衡点并实现有效协同交互并非易事，这种不平衡性很可能在抑制企业违规过程中导致联合效应无法充分发挥，甚至在某些情况下被大数据应用深度分析的精准性所超越。

① 为消除不同变量间因性质、量纲、数量级等属性的差异而带来的影响，且为更好比较大数据应用广度和深度对企业违规的影响程度差异，本文模型将所有变量进行标准化处理。

表4　　大数据应用广度、深度及其联合效应与企业违规实证回归结果

变量	*Fraud_Dum*	*Fraud_Dum*	*Fraud_Dum*	*Fraud_Fre*	*Fraud_Fre*	*Fraud_Fre*
	（1）	（2）	（3）	（4）	（5）	（6）
Bd_Breath	−0.013* （−1.65）			−0.019** （−2.31）		
Bd_Depth		−0.023*** （−3.25）			−0.029*** （−4.01）	
Bd_Breath×*Bd_Depth*			−0.013*** （−2.74）			−0.013*** （−2.79）
控制变量	控制	控制	控制	控制	控制	控制
年份/行业/省份固定效应	是	是	是	是	是	是
观测值	25531	25531	25531	25531	25531	25531
伪 R^2/R^2	0.068	0.068	0.068	0.084	0.084	0.084

4.内生性分析

（1）工具变量法。为缓解基准模型中可能存在的互为因果偏误，本文采用工具变量法进行再检验。借鉴黄群慧等（2019）[47]、王海等（2023）[48]的做法，本文选取1984年①企业所在城市每万人邮局数量（*Post_office*）、光缆线路长度（*Fiberl*）作为工具变量。数字基础设施是数字技术发展应用的基础，也是传统通讯手段的技术延伸。在现代通讯系统应用前，邮局是社会群体最为常用的通讯方式，其分布不仅会影响所在区域的信息流通效率、互联网接入机会，也会影响所在地居民的通信偏好、信息技术素养，因而对地区数字基础设施发展有着重要影响，满足工具变量相关性的要求。而随着通讯技术的快速发展，四十年前的邮局已经逐渐消失，历史上的邮局数量对当下企业经营决策的影响几乎不存在，满足工具变量排他性假设。另外，考虑到各城市1984年邮局数量为截面数据，本文参考黄群慧等（2019）[47]的做法，使用滞后一期的全国互联网上网人数分别与1984年各地级市每万人邮局数量的交乘项作为当期大数据应用程度的工具变量。光缆作为数字信息基础设施的关键组成部分，被视为“数字经济的血管”。一般而言，光缆线路长度，伴随着更广泛的数据连接，更高效、稳定的数据传输，其与网络覆盖广泛性、数据传输容量、数据处理速度等密切相关，深刻影响着大数据等数字技术的发展与应用。因此，光缆线路长度有助于完善数字信息基础设施和打通数字信息“大动脉”，从而影响区域内企业大数据应用水平，满足了工具变量选取的相关性要求。另外，光缆的建设和布局通常由政府规划和市场需求驱动，服务于整个地区和国家的经济发展，从理论上并不会直接影响企业违规行为，满足外生性条件。

表5　　内生性检验结果

变量	OLS	IVProbit	2SLS	Probit	Heckman第二阶段		PSM	
	Bigdata	*Fraud_Dum*	*Fraud_Fre*	*Bigdata_D*	*Fraud_Dum*	*Fraud_Fre*	*Fraud_Dum*	*Fraud_Fre*
	（1）	（2）	（3）	（4）	（5）	（6）	（7）	（8）
Bigdata		−0.493*** （−3.20）	−0.198*** （−3.89）		−0.053*** （−2.71）	−0.013*** （−3.48）	−0.060** （−2.07）	−0.012** （−2.16）

① 我国最早统计邮电业务量的年鉴是从1985年开始，统计的即为上一年的邮电业务数据信息。因此从纵向看，1984年的区域邮局数量包含着最久远的邮局业务分布信息。

续表 5

变量	OLS	IVProbit	2SLS	Probit	Heckman 第二阶段		PSM	
	Bigdata	*Fraud_Dum*	*Fraud_Fre*	*Bigdata_D*	*Fraud_Dum*	*Fraud_Fre*	*Fraud_Dum*	*Fraud_Fre*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post_office	0.098*** (6.00)			0.151*** (3.77)				
Fiberl	0.058*** (7.01)			0.109*** (5.31)				
IMR					0.525*** (2.51)	0.126*** (3.14)		
控制变量	控制	控制	控制	控制	控制	控制	控制	控制
年份/行业/省份固定效应	是	是	是	是	是	是	是	是
观测值	22929	22929	22929	22929	22929	22929	6236	6236
伪 R^2/R^2	0.381	—	0.145	0.032	0.069	0.083	0.092	0.119

本文分别采用IVProbit、2SLS模型进行回归，具体结果见表5的第(1)~(3)列。对于IVProbit回归，Wald检验结果表明，p值=0.0014<0.01，拒绝不存在内生性的原假设。第一阶段CLR值、AR值分别为101.06、11.07，p值均小于0.01，表明内生变量与工具存在相关性，通过了弱工具变量检验。对于2SLS回归，DWH检验结果显示，p值均小于0.01，拒绝了大数据应用为外生变量的原假设。Kleibergen-Paap rk LM统计量为125.007(p值为0.0000)，拒绝工具变量识别不足的原假设；Cragg-Donald Wald F统计量为62.566，大于Stock-Yogo弱工具变量识别检验在10%显著性水平上的临界值(19.93)，拒绝弱工具变量的原假设。Hansen J统计量为0.234(p值为0.629)，接受工具变量外生性要求。另外，第(1)列的第一阶段结果显示，*Post_office*和*Fiberl*的系数均在1%水平上显著为正。第(2)~(3)列为第二阶段回归结果，*Bigdata*的估计系数依然显著为负。表明在采用工具变量处理内生性问题后，研究结论未发生变化。

(2)Heckman两阶段法。为缓解选择性偏误带来的内生性问题，本文拟采用Heckman两阶段模型进行再检验。首先，设定虚拟变量*Bigdata_D*，若企业大数据应用程度大于行业-年度中位数，赋值为1，否则为0；其次，以上文*Post_office*、*Fiberl*作为排除性约束变量，其他控制变量与上文基本一致，采用Probit模型进行估计得到逆米尔斯比率(*IMR*)；最后，再将*IMR*代入基准模型进行回归测试。表5中的第(4)列为Heckman第一阶段回归结果，*Post_office*、*Fiberl*的回归系数分别为0.151、0.109，且达到了1%的显著性水平。表5第(5)~(6)列为Heckman第二阶段回归结果，*IMR*的估计系数均在1%水平上显著为正，表明模型存在样本选择偏误问题。在控制*IMR*这一偏差后，大数据应用均在1%水平下显著为负，研究结论保持不变。

(3)倾向匹配得分法。为缓解一些不可观察的宏观经济、制度等因素对研究结论的干预，本文拟采用倾向匹配得分法(PSM)进行再检验。具体地，第一，将大数据应用强度按照大小划分为五等份，将强度最高的20%样本设置为实验组，其余为参照组；第二，将两样本混合后，通过对上文控制变量、年份、行业、区域等协变量进行Logit回归，获取倾向性得分；第三，依照最近邻匹配法构建控制组，并按1∶1的比例进行配对。核密度分布如图1所示，经过核匹配后，实验组和对照组之间的核密度曲线重合面积显著提高。总的来看，匹配后实验组与对照组无系统差异。实证结果见表5第(7)~(8)列，*Bigdata*与*Fraud_Dum*、*Fraud_Fre*均显著负相关，前文结论仍然成立。

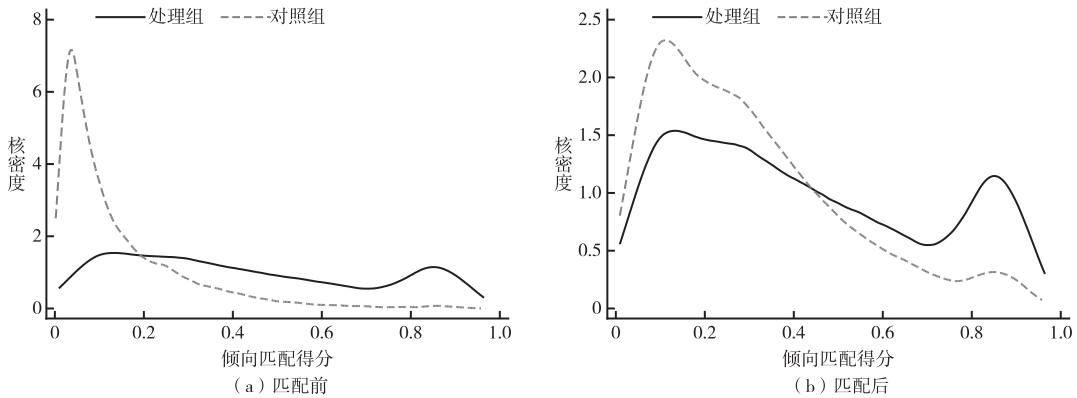

图1　核密度分布图

5.稳健性检验

（1）排除替代性解释。第一，排除策略性信息披露的替代性解释。在信息披露过程中，企业可能存在策略性炒作以及蹭热度、蹭概念的嫌疑，致使年报披露的大数据关键词与实际不符。为了排除这种可能的解释，本文参照肖土盛等（2022）[49]的做法，进行如下操作：①鉴于信息披露质量高的公司实施信息操纵的概率较低，剔除样本公司中信息披露考评结果为不合格的样本公司；②剔除样本期内因信息披露等问题受过证监会或证券交易所处罚的公司样本。表6第（1）~（4）列显示，*Bigdata*的估计系数均显著为负，且p值均小于0.05，研究结论并不受策略性信息披露行为的影响。

第二，排除媒体关注度的替代性解释。数字经济时代，大规模实施大数据应用的企业更易引发新闻媒体的关注。诸多研究表明，新闻媒体具有典型的信息传递和监督治理职能，能够抑制违规行为（Alkebsee和Habib，2021）[50]。可能的情形是，大数据应用程度高的企业之所以违规行为少是因为其受到的媒体关注度大，而非大数据应用的信息赋能和决策赋能。为排除上述替代性解释，本文根据企业媒体关注度①的年度-行业均值将研究样本细分为高低两组，分组进行检验。如果该上述解释成立，预期本文结论在高媒体关注的样本中更为明显。表6第（5）~（8）列显示，在媒体关注度低的企业，*Bigdata*的回归系数显著为负；在媒体关注度高的企业，*Bigdata*的回归系数则不显著。这排除了媒体关注度高的替代性解释。

表6　　排除替代性解释检验结果

变量	排除策略性媒体披露的替代性解释				排除高媒体关注度的替代性解释			
	剔除考评结果不及格样本		剔除信息披露违规样本		媒体关注低组		媒体关注度高组	
	Fraud_Dum	*Fraud_Fre*	*Fraud_Dum*	*Fraud_Fre*	*Fraud_Dum*	*Fraud_Fre*	*Fraud_Dum*	*Fraud_Fre*
	（1）	（2）	（3）	（4）	（5）	（6）	（7）	（8）
Bigdata	−0.054***	−0.012***	−0.062**	−0.004**	−0.049**	−0.012***	−0.039	−0.010
	（−2.65）	（−3.08）	（−2.08）	（−2.34）	（−2.38）	（−2.92）	（−0.83）	（−1.27）
控制变量	控制	控制	控制	控制	控制	控制	控制	控制
年份/行业/省份固定效应	是	是	是	是	是	是	是	是
观测值	19890	19890	22235	22245	20054	20054	5474	5477
伪R^2/R^2	0.060	0.070	0.041	0.022	0.069	0.086	0.097	0.096

① 本文采用CNRDS媒体数据库中涉及样本公司的财经新闻总量（标题出现该公司的新闻总数）的自然对数作为媒体关注的测度指标，该数据不仅囊括《证券时报》等传统报刊媒体，还包含有和讯网等大量主流网站，涵盖范围更为广泛。

(2)更换被解释变量。前文指出,从违规类型来看,企业违规可细分为信息披露违规、经营违规以及高管违规三类。本文进一步检验大数据应用对这三种违规行为的影响。分组结果如表7所示。结果表明,无论是以*Fraud_Dum*还是以*Fraud_Fre*作为被解释变量,除列(1)外,*Bigdata*的回归系数均显著为负。可见,大数据应用对企业信息披露违规、经营违规以及领导人违规均具有抑制作用。

表7　　更换被解释变量检验结果

变量	信息披露违规		经营违规		高管违规	
	Fraud_Dum	*Fraud_Fre*	*Fraud_Dum*	*Fraud_Fre*	*Fraud_Dum*	*Fraud_Fre*
	(1)	(2)	(3)	(4)	(5)	(6)
Bigdata	−0.033 (−1.44)	−0.006** (−2.12)	−0.044* (−1.88)	−0.005** (−2.54)	−0.085*** (−2.85)	−0.004*** (−3.09)
控制变量	控制	控制	控制	控制	控制	控制
年份/行业/省份固定效应	是	是	是	是	是	是
观测值	25531	25531	25531	25531	25501	25531
伪R^2/R^2	0.082	0.076	0.070	0.061	0.048	0.021

(3)更换解释变量。第一,为消除行业差异,本文将经过行业调整的大数据应用(目标企业大数据应用减去年度—行业均值)指标衡量企业大数据应用程度,记为*Bigdata_Ind*,重复基准模型。第二,为消除量纲的影响,本文以总词频除与年报文本的总词数的比值重新计量企业大数据应用程度,记为*Bigdata_R*。重复基准模型,实证结果见表8中第(1)~(4)列,无论以*Bigdata_R*或者*Bigdata_Ind*衡量企业大数据应用,回归系数均显著为负,研究结论保持不变。

表8　　更换解释变量和研究样本的检验结果

变量	更换解释变量				更换研究样本			
	Fraud_Dum	*Fraud_Fre*	*Fraud_Dum*	*Fraud_Fre*	*Fraud_Dum*	*Fraud_Fre*	*Fraud_Dum*	*Fraud_Fre*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Bigdata					−0.054*** (−2.71)	−0.014*** (−3.34)	−0.033 (−1.60)	−0.008** (−2.03)
Bigdata_Ind	−0.040** (−2.09)	−0.010*** (−2.63)						
Bigdata_R			−8.065*** (−4.41)	−1.535*** (−4.67)				
控制变量	控制	控制	控制	控制	控制	控制	控制	控制
年份/行业/省份固定效应	是	是	是	是	是	是	是	是
观测值	25531	25531	25531	25531	20867	20867	20611	20611
伪R^2/R^2	0.069	0.084	0.070	0.084	0.064	0.085	0.065	0.082

(4)更换研究样本。第一,2013年《国务院关于推进物联网有序健康发展的指导意见》出台之后,我国开始大规模推动企业的数字化转型。因此,这一政策冲击很可能对企业大数据应用与企业行为决策内在关系产生影响,从而对研究结论产生干扰。本文剔除2013年以前的研究样本重新进行回归检验,实证结果见表8第(5)~(6)列。第二,鉴于直辖市具有较大的经济、政治特殊性,在

大数据应用上存在一定的先发优势，对此，本文剔除四个直辖市的样本数据重新进行检验，实证结果见表8第(7)~(8)列。不难发现，除第(7)列外，*Bigdata*的回归系数均在1%水平上显著为负，研究结论未发生根本变化。

(5)其他稳健性检验。①为缓解公司层面不随时间变化的遗漏变量对研究结论的干预，本文在基准模型基础上进一步控制了公司固定效应，重新估计基准模型(1)，结果见表9第(1)~(2)列。②考虑到企业违规行为与行业的时变特征较为相关，本文在基准模型中纳入了行业-时间交互固定效应并进行重新估计，结果见表9第(3)~(4)列。③考虑到企业违规哑变量和违规次数符合归并数据特征，本文拟采用Tobit模型和负二项回归模型重新进行回归检验，结果见表9第(5)~(6)列。结果显示，*Bigdata*的系数均显著为负，检验结果与基准回归一致。综上，本研究结论具有较高的稳健性。

表9　　其他稳健性检验实证结果

变量	*Fraud_Dum*	*Fraud_Fre*	*Fraud_Dum*	*Fraud_Fre*	*Fraud_Dum*	*Fraud_Fre*
	(1)	(2)	(3)	(4)	(5)	(6)
Bigdata	-0.105***	-0.022***	-0.007**	-0.011***	-0.007**	-0.046***
	(-3.63)	(-3.28)	(-2.36)	(-3.07)	(-2.44)	(-3.00)
控制变量	控制	控制	控制	控制	控制	控制
年份/行业/省份固定效应	是	是	是	是	是	是
公司固定效应	是	是	否	否	否	否
年份×行业	否	否	是	是	否	否
观测值	20867	20867	25531	25531	25531	25531
伪R^2/R^2	—	0.085	—	—	0.072	0.066

五、拓展性研究

1.内在机制分析

依照假设的理论逻辑，大数据应用主要是通过提高信息披露质量(信息赋能)和组织决策质量(决策赋能)两条机制来抑制企业违规行为。本文采用交互模型对上述内在机制进行检验，具体模型如下：

$$Fraud = \gamma_0 + \gamma_1 Bigdata + \gamma_2 Mediaor + \gamma_3 Bigdata \times Mediaor + Controls + \sum industry + \sum year + \sum region + \varepsilon \quad (2)$$

其中，*Mediator*为机制变量，具体包括信息披露和组织决策，其余变量定义与模型(1)一致。本文将重点关注交互项的回归系数γ_3，以此判定上述机制是否成立。

(1)信息赋能路径。参考一般做法，本文采用可操纵盈余(*DA*)和分析师关注度(*Analyst*)刻画企业信息披露质量。前者由修正的Jones模型计算所得，其绝对值越大，信息披露质量越低；后者由年度内关注该企业的分析师人数并取自然对数进行衡量。结果如表10中的Panel A所示。第(1)~(8)列显示，*Bigdata*的系数均显著为负，该结果再次支持了假设H_1；第(1)列和第(3)列显示，*DA*的系数均在1%水平上显著为正；第(5)~(8)列显示，*Analyst*的系数均在1%水平上显著为负；这些证据印证了信息披露质量能够抑制企业违规的论点。交互项*Bigdata*×*DA*的回归系数显著为正，*Bigdata*×*Analyst*的回归系数显著为负，说明随着信息披露质量的提升，大数据应用对企业违规行为的抑制作用更为明显，即信息披露质量在企业大数据应用降低违规行为过程中发挥了促进作用。信息赋能路径成立。

（2）决策赋能路径。本文以企业管理决策效率（*MDE*）和投资效率（*INV*）作为高管决策效率的测度指标。原因在于，企业管理作为一种实践，包含组织、计划、领导、控制等职能活动，而决策贯穿和渗透于以上各项活动。因此，管理效率可在一定程度体现企业决策效率高低，二者具有一致性。而投资决策是企业决策的重要体现，准确识别投资机会和把握投资价值是测评高管团队优劣以及评估决策是否有效的关键考量。对于测量，本文参考杨继生和阳建辉（2015）[51]的做法，采用经行业调整的管理费用和销售费用之和与营业总收入的比值衡量管理决策效率，比值越大，管理效率越低；采用Richardson模型估计的残差绝对值衡量企业投资效率，残差绝对值越大，企业投资效率越低。结果如表10中的Panel B所示。结果表明，*MDE* 和 *INV* 的回归系数均在1%水平上显著为正，这说明管理效率、投资效率越高，企业违规的概率和频次越低，即企业决策质量能够抑制企业违规行为。另外，第（2）列、第（4）列，以及第（6）列、第（8）列显示，交互项 *Bigdata*×*MDE* 和 *Bigdata*×*INV* 的回归系数均显著为正，这表明，当企业非效率决策更低时，大数据应用对企业违规的抑制作用会增强，可见大数据应用会弱化企业管理、投资等非效率决策从而进一步降低企业违规行为。决策赋能路径成立。

表10　　内在机制检验结果

Panel A 信息赋能路径

变量	*Fraud_Dum*	*Fraud_Fre*	*Fraud_Dum*	*Fraud_Fre*	*Fraud_Dum*	*Fraud_Fre*	*Fraud_Dum*	*Fraud_Fre*
	（1）	（2）	（3）	（4）	（5）	（6）	（7）	（8）
Bigdata	−0.043**	−0.082***	−0.011***	−0.020***	−0.039**	−0.009	−0.010***	−0.003
	（−2.30）	（−3.29）	（−2.96）	（−4.21）	（−2.06）	（−0.39）	（−2.78）	（−0.49）
DA	0.886***	0.329	0.249***	0.112				
	（3.22）	（0.91）	（4.37）	（1.56）				
Analyst					−0.112***	−0.084***	−0.018***	−0.012***
					（−6.21）	（−3.67）	（−5.541）	（−2.87）
Bigdata×*DA*		0.534**		0.139***				
		（2.39）		（2.80）				
Bigdata× *Analyst*						−0.025**		−0.006**
						（−2.03）		（−2.44）
控制变量	控制	控制	控制	控制	控制	控制	控制	控制
年份/行业/省份固定效应	是	是	是	是	是	是	是	是
观测值	25531	25531	25531	25531	25531	25531	25531	25531
伪 R^2/R^2	0.069	0.070	0.085	0.085	0.070	0.071	0.085	0.085

Panel B 决策赋能路径

变量	*Fraud_Dum*	*Fraud_Fre*	*Fraud_Dum*	*Fraud_Fre*	*Fraud_Dum*	*Fraud_Fre*	*Fraud_Dum*	*Fraud_Fre*
	（1）	（2）	（3）	（4）	（5）	（6）	（7）	（8）
Bigdata	−0.049***	−0.050***	−0.012***	−0.012***	−0.044**	−0.068***	−0.011***	−0.020**
	（−2.60）	（−2.68）	（−3.34）	（−3.25）	（−2.34）	（−2.91）	（−3.02）	（−4.38）
MDE	1.000***	0.509**	0.216***	0.129***				
	（6.08）	（2.30）	（6.07）	（2.93）				
INV					0.888***	0.559*	0.188***	0.070
					（3.85）	（1.79）	（4.30）	（1.24）

续表 10

Panel B 决策赋能路径

变量	*Fraud_Dum*	*Fraud_Fre*	*Fraud_Dum*	*Fraud_Fre*	*Fraud_Dum*	*Fraud_Fre*	*Fraud_Dum*	*Fraud_Fre*
	（1）	（2）	（3）	（4）	（5）	（6）	（7）	（8）
Bigdata×*MDE*		0.446***		0.072**				
		（3.38）		（2.42）				
Bigdata×*INV*						0.359*		0.133***
						（1.70）		（2.94）
控制变量	控制	控制	控制	控制	控制	控制	控制	控制
年份/行业/省份固定效应	是	是	是	是	是	是	是	是
观测值	25531	25531	25531	25531	25531	25531	25531	25531
伪 R^2/R^2	0.070	0.071	0.085	0.086	0.069	0.070	0.084	0.085

2.异质性分析

（1）产权异质性。在中国特有的经济体制下，国有企业和非国有企业承担的目标、所获取的资源并不相同。因此，企业的产权性质很可能会影响大数据应用与企业违规的既定关系。本文基于样本公司实际控制人属性将样本细分为国有企业组和非国有企业组，分组检验大数据应用对企业违规的影响差异。表 11 结果显示，无论是以 *Fraud_Dum* 还是 *Fraud_Fre* 衡量企业违规行为，在非国有企业组，*Bigdata* 的系数在 1% 水平上显著为负；而在国有企业组，*Bigdata* 的系数则不显著。而且，本文采用基于 Bootstrap 的费舍尔组合检验法并将抽样次数设定为 500 次进行组间系数差异检验，结果显示，*Bigdata* 的组间差异 *p* 值均小于 0.1。这说明，大数据应用对非国有企业违规行为的抑制效果更显著。这可能是因为，一是非国有企业内部监管参差不齐，合规机制并不健全，大数据应用能够实现对企业活动的全面、实时、精准监控，从而有效抑制企业违规。而国有企业行为决策会受到企业制度和党规党纪的双重约束，不断完善的监管机制和问责机制使得国有企业违规行为能够更早被察觉，因而对大数据技术的依赖程度相对较低。二是非国有企业数字化起步晚，数字化应用水平低，引入和应用大数据能够产生更为显著的违规抑制效应。而国有企业基于特殊的制度背景，更倾向于贯彻实施国家提出的数字发展战略，其现有数字化体系已形成了常态化的违规防控机制，此时大数据应用对违规行为的边际抑制效应则相对较低。

表 11　　基于产权性质的异质性检验结果

变量	非国有企业	国有企业	非国有企业	国有企业
	Fraud_Dum	*Fraud_Dum*	*Fraud_Fre*	*Fraud_Fre*
	（1）	（2）	（3）	（4）
Bigdata	−0.070***	−0.003	−0.018***	−0.002
	（−3.20）	（−0.08）	（−3.77）	（−0.30）
控制变量	控制	控制	控制	控制
年份/行业/省份固定效应	是	是	是	是
观测值	15809	9703	15809	9720
伪 R^2/R^2	0.066	0.073	0.099	0.062
费舍尔组合检验	0.065		0.010	

（2）行业异质性。高科技和非高科技企业间的禀赋特征、竞争环境等对大数据的应用程度并不相同，这势必会对企业违规产生异质性影响。本文参考彭红星和毛新述（2017）[52]的做法，将行

业大类代码为C25-C29、C31、C32、C34-C41、I63-I65、M73的行业定义为高科技行业，并据此将样本细分为高科技行业组和非高科技行业组。分组实证结果如表12所示。结果显示，在高科技行业组，*Bigdata*与*Fraud_Dum*、*Fraud_Fre*均在1%水平上显著负相关。在非高科技行业组，*Bigdata*的回归系数则不显著。费舍尔组合检验结果显示，*Bigdata*系数均在1%水平上通过了组间差异检验。可见，大数据应用对企业违规行为的抑制作用主要体现在高科技企业中。原因在于，高技术行业内企业具有高成长性、高知识密集度、高研发投入的典型特征，其发展模式与数字技术有着天然内生需求和适配优势。因此，高科技企业对大数据技术更新和战略转型的接受程度相对较高，所面临的变革压力和风险相对较小，所遭遇的转型冲击也相对较弱。而且，高科技企业前期的创新成果、前沿技术积累较为丰富，并不需要太多的时间对大数据开发和应用进行摸索。此时，大数据在企业各业务流程和环节的渗透效果更快、赋能作用更强，进而产生更为明显的违规抑制效应。

表12　　基于行业特征的异质性检验结果

变量	非高科技行业	高科技行业	非高科技行业	高科技行业
	Fraud_Dum	*Fraud_Dum*	*Fraud_Fre*	*Fraud_Fre*
	(1)	(2)	(3)	(4)
Bigdata	0.059*	-0.092***	0.003	-0.017***
	(1.79)	(-3.96)	(0.41)	(-3.76)
控制变量	控制	控制	控制	控制
年份/行业/省份固定效应	是	是	是	是
观测值	11013	14518	11013	14518
伪R^2/R^2	0.079	0.067	0.088	0.086
费舍尔组合检验	0.000		0.000	

(3)地区异质性。本文参考杨仁发和魏琴琴(2021)[53]的做法测算区域营商环境指数，并根据其年度中位数，将样本企业划分为营商环境高、低两组，分组结果如表13所示。在低营商环境样本组，*Bigdata*的回归系数均不显著；在高营商环境样本组，*Bigdata*的回归系数均在1%水平上显著为负。另外，基于费舍尔组合检验方法得到的经验p值均小于0.05，说明*Bigdata*的回归系数在营商环境高、低两组间存在显著差异。可见，大数据应用对企业违规行为的抑制作用因营商环境存在显著差异。究其原因，在营商环境好的地区，与大数据发展相关的法律法规更为健全、监管机制更为完善、引领政策更为精准，这可有效约束算法控制、篡改数据等技术寻租问题，破除阻碍大数据发展的机制障碍，进而营造良好的制度环境。而且，营商环境好的地区，要素市场发育相对健全，具有更完备的市场基础设备、更为充裕的资源禀赋、更高质量的政府服务等(宋林霖和陈志超，2023)[54]，这些可与大数据技术形成互补，促进技术外溢，从而更好发挥其赋能增值效应。因此，在营商环境好的地区，大数据应用对企业违规行为的抑制作用更为显著。

表13　　基于区域特征的异质性检验结果

变量	营商环境低组	营商环境高组	营商环境低组	营商环境高组
	Fraud_Dum	*Fraud_Fre*	*Fraud_Dum*	*Fraud_Fre*
	(1)	(2)	(3)	(4)
Bigdata	-0.011	-0.104***	-0.001	-0.024***
	(-0.46)	(-3.51)	(-0.25)	(-4.56)
控制变量	控制	控制	控制	控制
年份/行业/省份固定效应	是	是	是	是
观测值	14487	11032	14489	11042
伪R^2/R^2	0.063	0.090	0.077	0.103
费舍尔组合检验	0.010		0.000	

3.经济后果分析

车笑竹和苏勇（2018）[55]研究发现，违规企业在受到监管处罚后，有动机采取一些符合社会期望的行为，如改善公司治理、变更涉事高管等，减小合法性差距，以此提升企业形象。这些证据表明，对违规行为的监管处罚具有积极作用。但部分学者认为，我国监管制度并不健全、违规成本相对较低，违规成本和违规收益的长期失衡导致一些违规“前科”企业有恃无恐，进而难以发挥违规监管处罚的震慑作用（于博等，2022）[56]。对此，本文提出问题：当违规事件发生后，企业是“改过自新”减少违规行为，以提高组织合法性；还是“变本加厉”违规再犯，以攫取违规收益？另外，大数据应用在其中扮演何种角色？为回答此问题，本文构建如下模型：

$$Punish_agaN = \theta_0 + \theta_1 FraudDum + \theta_2 Bigdata + \theta_3 FraudDum*Bigdata + Controls + \sum industry + \sum year + \sum region + \varepsilon \quad (3)$$

其中，*Punish_agaN*为企业受罚后未来n年（n=2，3）发生违规行为的频次。*FraudDum*为企业当年是否存在违规行为，若存在赋值为1，否则为0。其他变量与模型（1）基本一致。由表14可知，*FraudDum*的回归系数均显著为正，这说明违规次数越多的企业，在未来两年和三年内再次发生违规行为的可能性越高，即企业违规存在“二皮脸”效应。原因在于，违规企业具有更明显的逐利属性，在多数情形下，违规公司仅受到警告、谴责、责令整改等处罚，所遭受的违规损失与获取的巨额违规收益是极其不对称的。因此，即便在违规事件被稽查或者处罚后，违规企业为实现自身利益最大，依然有动机铤而走险来实施违规行为，只是违规手段或者方式更为多样或者隐蔽。第（4）列结果显示，交互项*FraudDum×Bigdata*的回归系数显著为负；在第（3）列中，交互项虽不显著，但依然为负。这些证据基本证明，大数据应用能够降低违规企业再犯的可能性。可能的原因在于，在大数据技术赋能之下，利益相关者获取信息成本显著下降，并能快速实现数字信息的连接共享和互联互通（Oussous等，2018）[10]。当企业受到监管处罚后，监管主体可依靠大数据技术精准把握违规公司真实经营状况等，倒逼企业规范公司治理。在此约束和震慑之下，违规企业则会减少违规短视行为，避免企业形象进一步受损。

表14　　大数据应用、企业违规与是否再次违规实证检验结果

变量	*Punish_aga*2	*Punish_aga*3	*Punish_aga*2	*Punish_aga*3
	（1）	（2）	（3）	（4）
FraudDum	0.421***	0.279***	0.468***	0.398***
	（6.12）	（3.12）	（5.13）	（3.50）
Bigdata			0.033	0.060*
			（1.57）	（1.88）
FraudDum×Bigdata			−0.041	−0.124*
			（−0.74）	（−1.74）
控制变量	控制	控制	控制	控制
年份/行业/省份固定效应	是	是	是	是
观测值	18676	15573	18677	15573
伪R^2/R^2	0.026	0.024	0.026	0.025

六、结论和启示

1.结论

基于企业大数据泛化应用以及违规现象屡禁不止的现实情境，本文以2010—2020年沪深A股非金融类上市公司为研究样本，系统探索了大数据应用对企业违规行为的影响效应和作用机理。

主要研究结论如下：大数据应用能够抑制企业违规行为，该结果在通过若干内生性分析和稳健性检验后依然成立。这说明，依托大数据高速性、强渗透性、泛时空性等特征，企业应用大数据能够有效缓解内部管理机制失效和外部监管机制失灵等问题，从而降低企业违规行为发生的概率和频次。在区分大数据应用维度后，研究发现大数据应用广度、深度及其联合效应能够抑制企业违规行为，从影响程度看，大数据应用深度对企业违规的抑制效应要大于广度以及广度和深度的联合效应。机制分析发现，提高信息披露质量和组织决策质量是大数据应用抑制企业违规行为的重要渠道，即大数据应用可通过信息赋能和决策赋能，从而减少高管实施违规行为的机会。异质性分析发现，大数据应用对企业违规的抑制作用主要存在于非国有企业、高科技企业以及位于高质量营商环境地区的企业。这反映出我国企业大数据应用存在数字资源投入不均衡、数字基础建设不充分、制度配套不完善等问题，从而为相关部门进一步调整大数据政策、释放数据潜能提供学理参考。经济后果发现，违规次数越多的企业再次发生违规行为的可能性越高，这一结果说明企业违规具有“二皮脸”效应，但大数据应用能够削弱此效应。

2.实践启示

数智时代，企业既要把握大数据技术带来的治理革新机遇，又需警惕技术异化引发的潜在风险。为此，本文基于上述研究结论，并结合数智治理的实践逻辑，从技术理性与治理效能的辩证关系出发，提出针对性的经验启示，旨在为企业构建兼具技术适配性与治理有效性的大数据应用方案提供决策参照。具体如下：

第一，在数智时代“赢者通吃”的竞争格局下，企业需紧跟时代步伐积极拥抱大数据，着力补齐大数据技术短板，不断强化大数据收集、整合、分析能力，深度挖掘大数据在增强信息赋能和决策赋能的核心驱动作用，以求优化信息环境，提高企业决策质量，进而助推大数据应用在抑制企业违规等方面的优势。需强调的是，在实践过程中，大数据应用也可能带来经营风险、算法寻租、决策失灵等问题，一定程度上削弱了大数据应用的赋能效果。为此，企业需辩证性看待大数据技术应用，以守住经营安全为底线，有计划地、有次序地制定大数据战略规划，避免贸然实施大数据战略给组织增加诸多转型风险，以求充分释放大数据技术带来的公司治理红利。

第二，充分发挥大数据应用广度、深度以及协同作用，应成为数字时代优化公司治理的重要举措。一方面，企业需逐步引入或开发多元性的大数据技术或组合，推动其弹性扩展和技术服务，打好大数据技术应用“组合拳”。而且，当前数字技术应用呈多方向、宽前沿、集群式等趋势，企业也需积极开展跨界合作，驱动数据资源互补和数字技术共享，稳步扩大大数据服务与应用覆盖范围。另一方面，需根据自身需求，锚定业务主攻方向，聚力深挖大数据核心价值，推动大数据应用做实、做精和做深。与此同时，要加快由表及里、由浅入深，深化“大数据+”模式应用，积极促进数据、技术、场景有效融合，进一步促进大数据技术场景延伸，以此更有效发挥其赋能增值效应。

第三，在实践过程中，企业需因企制宜、因业制宜和因地制宜，实行大数据差异化应用。对非国企而言，需加强顶层设计，积极了解和申请相关数字化财税补贴，逐步配备专业的人才和技术，最大程度展现其大数据应用的灵活性和适应性，进而提升大数据应用的赋能效果。对高科技企业而言，需充分发挥其发展模式对数字技术天然的内生需求与适配优势，分阶段、有步骤地加大在大数据等前沿技术领域的研发投入，积极搭建产学研合作平台，助力自身大数据应用的强变革和快发展。对高水平营商环境地区企业而言，需关注大数据落地应用的痛点难点堵点，紧握所在区域政务服务效率高、流程简化等优势，加强大数据开发、应用和管理能力，持续推进自身大数据发展。另外，当地政府也应积极探索优化外部营商环境的多维路径，譬如，

推动“一网通办”再升级、深化“放管服”改革等，为大数据技术平稳落地、做大做强提供营商环境支持。

参考文献

[1]George，G.，M.R.Haas，and A.Pentland.Big Data and Management[J].Academy of Management Journal，2014，57，(2)：321-326.

[2]张明超，孙新波，王永霞.数据赋能驱动精益生产创新内在机理的案例研究[J].天津：南开管理评论，2021，(3)：102-116.

[3]张叶青，陆瑶，李乐芸.大数据应用对中国企业市场价值的影响——来自中国上市公司年报文本分析的证据[J].北京：经济研究，2021，(12)：42-59.

[4]Saxena，D.，M.Lamest.Information Overload and Coping Strategies in the Big Data Context：Evidence from the Hospitality Sector[J].Journal of Information Science，2018，44，(3)：287-297.

[5]凡景强，邢思聪.大数据伦理研究进展、理论框架及其启示[J].西安：情报杂志，2023，(3)：167-173.

[6]Ghasemaghaei，M.，and G.Calic.Assessing the Impact of Big Data on Firm Innovation Performance：Big Data is Not Always Better data[J].Journal of Business Research，2020，108：147-162.

[7]史丹.基于大数据的经济分析[M].北京：中国社会科学出版社，2019.

[8]McAfee，A.，E.Brynjolfsson，T.H.Davenport，et al.Big Data：The Management Revolution[J].Harvard Business Review，2012，90，(10)：60-68.

[9]安小米，黄婕，许济沧，王丽丽，洪学海，王志强，韩新伊.全景式大数据质量评估指标框架构建研究[J].天津：管理科学学报，2023，(5)：138-153.

[10]Oussous，A.，F.Z.Benjelloun，A.and A.Lahcen，et al.Big Data Technologies：A Survey[J].Journal of King Saud University-Computer and Information Sciences，2018，30，(4)：431-448.

[11]迟考勋，邵月婷，苏福.大数据价值来源、价值内容与价值创造机理——基于2011—2021年管理类和商业类SSCI期刊分析[J].武汉：科技进步与对策，2022，(22)：151-160.

[12]Li，L.，J.Lin，and Y.Ouyang，X.et al.Evaluating the Impact of Big Data Analytics Usage on the Decision-Making Quality of Organizations[J].Technological Forecasting and Social Change，2022，175，121355.

[13]任曙明，马橙.大数据应用与中国企业创新资源错配[J].北京：科学学研究，2024，(1)：194-204.

[14]杨俊，李小明，黄守军.大数据、技术进步与经济增长——大数据作为生产要素的一个内生增长理论[J].北京：经济研究，2022，(4)：103-119.

[15]肖静华，胡杨颂，吴瑶.成长品：数据驱动的企业与用户互动创新案例研究[J].北京：管理世界，2020，(3)：183-205.

[16]杨德明，夏小燕，金淞宇，林丹滢，马晴.大数据、区块链与上市公司审计费用[J].北京：审计研究，2020，(4)：68-79.

[17]李金昌.大数据应用的质量控制[J].北京：统计研究，2020，(2)：119-128.

[18]刘宇琪，秦宗文.大数据在行政决策中应用的反思与优化——基于科层制运行视角的考察[J].哈尔滨：行政论坛，2023，(4)：118-126.

[19]张海丽，王宇凡，Michael Song.大数据驱动创新过程提高数字创新绩效的路径[J].北京：科学学研究，2023，(6)：1106-1120.

[20]Homer.，E.M.Testing the Fraud Triangle：A Systematic Review[J].Journal of Financial Crime，2020，27，(1)：172-187.

[21]Cao，Y.，Z.Zhang，and K.Peng，et al.The Establishment of Circuit Courts and Corporate Fraud[J].Emerging Markets Finance and Trade，2023，59，(8)：2600-2616.

[22]Ren，L.，X.Zhong，and L.Wan.Missing Analyst Forecasts and Corporate Fraud：Evidence from China[J].Journal of Business Ethics，2021，181：171-194.

[23]黄卓，陶云清，王帅.社会信用环境改善降低了企业违规吗？——来自“中国社会信用体系建设”的证据[J].北京：金融研究，2023，(5)：96-114.

[24]吴雅倩，李吉园，王成方.克己复礼：董事长本地任职与企业违规[J].北京：会计研究，2024，(4)：63-79.

[25]Liao，L.，G.Chen，and D.Zheng.Corporate Social Responsibility and Financial Fraud：Evidence from China[J].Accounting & Finance，2019，59，(5)：3133-3169.

[26]Hass，L.H.，M.Tarsalewska，and F.Zhan.Equity Incentives and Corporate Fraud in China[J].Journal of Business Ethics，2016，138：723-742.

[27]张学志，李灿权，周梓洵.员工持股计划、内部监督与企业违规[J].北京：世界经济，2022，(3)：185-211.

[28]朱沛华，李方方，李军林.高管外部薪酬差距的激励扭曲效应——以企业违规为例的研究[J].北京：经济管理，2024，(4)：

113-131.

[29] Parlindungan, R., F. Africano, and P. Elizabeth. Financial Statement Fraud Detection Using Published Data Based on Fraud Triangle Theory[J].Advanced Science Letters,2017,23,(8):7054-7058.

[30]陈邑早,陈艳,于洪鉴.会计舞弊行为决策中的合理化因素:分析与展望[J].上海:外国经济与管理,2019,(7):85-98.

[31]周泽将,王浩然,万明华.董事选举得票率与企业违规行为[J].上海:外国经济与管理,2022,(5):19-32

[32]吴武清,洪振瀚.革命文化的时代价值:来自长征精神抑制企业违规的证据[J].北京:世界经济,2024,(3):3-29.

[33]王伊攀,朱晓满.政府大客户能够制约上市公司违规吗?[J].上海财经大学学报,2022,(3):76-91.

[34]Jin, X., B.W.Wah, and X.Cheng, et al.Significance and Challenges of Big Data Research[J].Big Data Research,2015,2,(2):59-64.

[35]董盈厚,马亚民,董馨格,韩亮亮.金融资产配置与盈余价值相关性——"有效市场"抑或"功能锁定"[J].北京:会计研究,2021,(9):95-105.

[36]李文贵,邵毅平.监管信息公开与上市公司违规[J].北京:经济管理,2022,(2):141-158.

[37]Janssen, M., H.van der Voort, and A.Wahyudi.Factors Influencing Big Data Decision-Making Quality[J].Journal of Business Research,2017,70:338-345.

[38]Merendino, A., S.Dibb, and M.Meadows, et al.Big data, Big Decisions: The Impact of Big Data on Board Level Decision-Making[J].Journal of Business Research,2018,93: 67-78.

[39]Jing, Z., S.Lu, and Y.Zhao, et al.Economic Policy Uncertainty, Corporate Investment Decisions and Stock Price Crash Risk: Evidence from China[J].Accounting & Finance,2023,63:1477-1502.

[40]彭俞超,倪骁然,沈吉.企业"脱实向虚"与金融市场稳定——基于股价崩盘风险的视角[J].北京:经济研究,2018,(10):50-66.

[41]Rashid, M.A., A.Al-Mamun, and H.Roudaki, et al.An Overview of Corporate Fraud and its Prevention Approach[J].Australasian Accounting Business & Finance Journal,2022,16,(1):101-118.

[42]Calic, G., and M.Ghasemaghaei.Big Data for Social Benefits: Innovation as A Mediator of the Relationship between Big Data and Corporate Social Performance[J].Journal of Business Research,2021,131:391-401.

[43]陈德球,胡晴.数字经济时代下的公司治理研究:范式创新与实践前沿[J].北京:管理世界,2022,(6):213-240.

[44]赵宸宇,王文春,李雪松.数字化转型如何影响企业全要素生产率[J].财贸经济,2021,(7):114-129.

[45]吴非,胡慧芷,林慧妍等.企业数字化转型与资本市场表现——来自股票流动性的经验证据[J].北京:管理世界,2021,(7):130-144,10.

[46]陈楠,蔡跃洲,马文君.企业数据技术能力对市值和利润的影响——基于大数据专利的机制检验和互补性投入分析[J].北京:财贸经济,2024,(1):106-123.

[47]黄群慧,余泳泽,张松林.互联网发展与制造业生产率提升:内在机制与中国经验[J].北京:中国工业经济,2019,(8):5-23.

[48]王海,郭冠宇,尹俊雅.数字化转型如何赋能企业绿色创新发展[J].北京:经济学动态,2023,(12):76-91.

[49]肖土盛,孙瑞琦,袁淳,孙健.企业数字化转型、人力资本结构调整与劳动收入份额[J].北京:管理世界,2022,(12):220-237.

[50] Alkebsee, R. H., and A. Habib. Media Coverage and Financial Restatements: Evidence from China [J]. Asian Review of Accounting,2021,29,(4): 505-524.

[51]杨继生,阳建辉.行政垄断、政治庇佑与国有企业的超额成本[J].北京:经济研究,2015,(4):50-61,106.

[52]彭红星,毛新述.政府创新补贴、公司高管背景与研发投入——来自我国高科技行业的经验证据[J].北京:财贸经济,2017,(3):147-161.

[53]杨仁发,魏琴琴.营商环境对城市创新能力的影响研究——基于中介效应的实证检验[J].北京:调研世界,2021,(10):35-43.

[54]宋林霖,陈志超.温度、测度、制度:中国营商环境治理结构及其调试[J].北京:中国行政管理,2023,(5):6-13.

[55]车笑竹,苏勇.企业违规对社会责任报告及其价值效应的影响[J].北京:经济管理,2018,(10):58-74.

[56]于博,林龙斌,安邦.证券市场监管处罚避免企业违规再犯了吗?[J].西安:人文杂志,2022,(5):77-88.

How the Application of Big Data Inhibits Corporate Fraud

PAN Zi-cheng[1], YI Zhi-gao[1], BAI Shu-yuan[2]
(1.School of Business, Nanjing Normal University, Nanjing, Jiangsu, 210023, China;
2.School of Business, Nanjing University, Nanjing, Jiangsu, 210093, China)

Abstract: In the era of digital economy, the application of big data brings both development opportunities and challenges to corporate governance. However, the practical effectiveness of such application still needs to be examined. A literature review reveals that although the aforementioned studies provide relatively abundant materials for researching the economic consequences of big data application, there is still no consensus on its effectiveness. Moreover, few studies have specifically discussed the relationship between big data application and corporate fraud, which provides an important breakthrough for this research. In recent years, China has witnessed a surge in various violations such as financial fraud, stock price manipulation, and false disclosures. In fact, in the face of the rapidly evolving and thriving reality of big data, incorporating emerging digital technologies such as big data into various aspects of corporate governance may be the direction for improving and optimizing modern enterprise management. Against this background, this article raises the following questions: How does the application of big data affect corporate fraud? What are the underlying mechanisms? And what factors may influence it?

Taking corporate fraud as the starting point, this paper systematically explores the impact of big data application on corporate fraud based on data from 2010-2020 A-share listed companies in Shanghai and Shenzhen. Empirical results show that the application of big data can significantly reduce the probability and frequency of enterprise violations. The above conclusions are still valid after several robustness tests, such as instrumental variable method, Heckman two-stage method and propensity matching score, replacing core variables, excluding alternative explanations, and changing research samples. Furthermore, after distinguishing the dimensions of big data applications, it is found that both the depth and breadth of big data applications, as well as their combined effects, have negative impacts on corporate violations, but the inhibitory effect of depth is the most pronounced. Mechanism research shows that big data application mainly inhibits corporate fraud through two paths: "information empowerment" and "decision empowerment". Heterogeneity analysis shows that the deterrent effect of big data application to corporate fraud is more pronounced in non-state-owned enterprises, high-tech enterprises, and enterprises located in regions with good business environments. Economic consequence analysis finds that the likelihood of repeated corporate fraud is higher for non-compliant firms, indicating a "two-face" effect of corporate fraud, but big data application can mitigate this to some extent.

The possible research significance of this article is as follows: Firstly, current literature exploring the economic consequences of big data application at the micro-level of enterprises mainly focuses on areas such as innovation, market value, and business decision-making, with no consensus on its effectiveness. This article, however, examines the value of big data application from the perspective of corporate fraud, studying its positive and negative effects. It provides new empirical evidence for the effectiveness of big data application. Secondly, previous literature has tended to demonstrate the impact of big data applications from an overall perspective. This paper breaks the research inertia of treating big data applications homogeneously by subdividing them into two dimensions: breadth and depth. It also explores the differentiated impacts of these two dimensions and their combined effects on corporate violations, thereby providing a further supplement to the research on big data applications. Thirdly, this article explores the path of the deterrent effect of big data application on corporate fraud from the perspectives of information disclosure quality and management decision-making efficiency. It further refines the situational context of this effect from three dimensions, namely ownership nature, industry characteristics, and regional characteristics. The research findings unveil the "black box" mechanism of how big data application can prevent corporate fraud, and enhance people's comprehensive understanding of the relationship between the two. Lastly, this article focuses on the economic outcomes and analyzes in detail whether companies involved in fraud choose to reform or escalate their violations after being inspected. It also discusses the important role played by big data application in this process. This not only reveals the "negative spiral" effect of corporate fraud but also provides new insights for the industry on why and how to restrain corporate fraud. In summary, this research provides important practical references for why the digital economy era should promote high-quality development of big data and how to optimize corporate governance.

Key Words: big data application; corporate fraud; breadth of big data application; depth of big data application
JEL Classification: L86, M16
DOI: 10.19616/j.cnki.bmj.2025.03.007

(责任编辑:张任之)