DOI:10.3876/j.issn.1671-4970.2026.05.011

基于生成式大语言模型的数据资产赋能企业治理研究

尹　涛[1],王耀民[1],孟令超[2]

(1.华东政法大学商学院;2.对外经济贸易大学中国金融学院)

摘要:数据资产作为数字经济时代企业的核心战略资源与关键治理要素,能够通过信息传递与创新驱动机制提升企业治理水平。基于2011—2023年中国A股上市公司数据,借助人工采集语料微调大语言模型构建企业数据资产评价指标体系,并实证检验其对企业治理水平的影响。研究发现,数据资产能够显著提升企业治理水平,在控制企业特征及固定效应后,结论依然成立。进一步采用滞后变量、工具变量、倾向得分匹配、双重差分及替换核心解释变量等方法进行检验发现,研究结论具有较好的稳健性。相较于数字化转型整体水平,数据资产能更加突出数据资源的开发利用与价值转化,其边际效应显著高于数字化转型的整体影响效应。机制检验表明,数据资产主要通过降低信息不对称和提升创新能力两条路径改善企业治理水平,其中数据资产能够提高企业信息传递效率、缓解内外部信息壁垒,同时通过优化资源配置和提升创新效率增强企业治理能力。异质性分析进一步发现,上述效应在小规模、低现金流比率、高成长性及东部地区企业中尤为突出。相关结论丰富了数据资产经济后果与企业治理领域的研究,揭示了数据资产赋能企业治理的作用机制与边界条件,为数字化转型背景下企业利用数据资产提升治理水平提供了理论解释与实践启示,对推动数据资产价值转化、完善数据要素配置机制及健全现代公司治理体系具有重要政策参考价值。

关键词:数据资产;企业治理水平;大语言模型;信息不对称;创新能力

一、引　言

企业治理水平是保障资本市场稳健发展的核心要素。《中共中央办公厅　国务院办公厅关于完善中国特色现代企业制度的意见》明确提出要完善公司治理结构、提升企业科学管理水平,这为进一步健全现代企业制度和提升企业治理效能提供了重要政策指引。然而,当前我国企业治理在信息披露质量、中小股东参与及内部监督机制等方面仍存在较大优化空间。《中国上市公司治理指数(2024)》显示,我国上市公司整体治理水平稳步提升但增幅有所放缓,治理指数平均值由2023年的64.76提升至64.87;监事会治理指数下降至59.17,仍是公司治理体系中的相对薄弱环节。从治理评级来看,上市公司治理评级主要集中于BBB(34.63%)和BB(31.32%),A类样本中尚无AAA和AA等级企业,表明我国上市公司治理水平仍存在进一步提升空间[1]。由此可见,构建与现代企业制度相适配的治理能力提升机

引用本文:尹涛,王耀民,孟令超.基于生成式大语言模型的数据资产赋能企业治理研究[J].河海大学学报(哲学社会科学版),2026,28(5):115-129.

基金项目:国家自然科学基金项目(72503040);北京大学中央高校基础科研基金项目(数理经济数量金融实验室LMEQF)

作者简介:尹涛(1984—),男,讲师,博士,主要从事数据资产研究。E-mail:taoyin@ecupl.edu.cn

制,对于深化资本市场改革、促进实体经济高质量发展具有重要意义。作为数字经济时代的重要生产要素,数据资产凭借信息整合、价值挖掘和实时反馈等能力,能够为企业的战略决策、风险管理和运营监督等治理环节提供新的支持。《国务院办公厅关于做好金融“五篇大文章”的指导意见》《“十四五”数字经济发展规划》等政策文本均强调要发挥数据要素作用、提升数据治理能力,这为企业数据资产价值释放提供了政策基础。然而,《2025 中国企业数字化转型指数》表明,尽管中国企业正在加速生成式人工智能应用,但数据治理与价值转化能力仍存在不足,仅 9%的企业实现了由生成式人工智能驱动的业务价值显著提升[2],由此说明,企业数据资产价值释放仍有较大提升空间。

围绕数据资产的价值创造及其经济后果,已有研究主要沿两个维度展开:第一,数据资产价值创造研究。相关研究基于数据金字塔与数据价值链理论,指出数据资产能够通过参与企业经营决策、优化业务流程和提升资源配置效率实现价值增值[3-4]。同时,学者认为数据价值并非局限于货币化交易,非货币交易模式、大数据应用及创新驱动机制同样能够促进数据价值释放[5-8]。第二,数据资产价值传导机制及经济效应研究。学者从宏观与微观两个层面分析了数据资产的作用路径,认为数字经济规模测算框架与增加值结构分析为数据资产价值评估提供了方法基础[9-10],并从数据资本化视角拓展了数据资产价值研究边界,对中国数据资本形成总额与数据资本存量进行了系统测算,为刻画数据资产的资本属性及其经济价值提供了新的量化依据[11]。相关研究表明,数据要素能够通过优化资源配置、促进创新等机制推动经济增长[12-14],在企业层面则通过降低创新试错成本、提升市场响应能力等路径增强企业竞争优势[15]。此外,数字经济对“一带一路”沿线国家技术追赶的研究表明,数据资产具有时效性特征,且能够影响企业战略决策与技术进步过程[16]。已有研究揭示了数据资产的价值创造过程及传导机制,为进一步探讨其对企业治理水平的影响提供了理论基础。在数据资产价值研究不断深入的背景下,其定价、入表与测度成为重要研究方向,相关研究系统界定了数据资产边界特征,并提出了涵盖确认标准、计量模型和统计路径的核算体系,为数据资产价值评估提供了理论依据[17]。基于无形资产估值原理,学者进一步利用企业市场价值与账面价值差额对数据资产进行量化表征[18]。在测度方法方面,基于文本挖掘和机器学习算法构建关键词词库,能够有效识别企业数据资产信息披露特征,已成为重要测度路径[19],如何瑛等进一步结合 Python 与 Word2Vec 模型构建数据资产词典,利用年报文本词频从自用型和交易型两个维度测度企业数据资产水平,拓展了现有测度方法[20]。围绕企业治理研究,学者主要围绕治理主体、治理机制及企业特征等维度展开:治理主体方面,已有研究发现,中小股东参与会影响企业治理效果[21];治理机制方面,双层股权制度等治理机制会通过影响企业创新活动对公司治理产生影响[22]。此外,独立董事激励机制也会对公司治理效率产生影响[23]。总体来看,现有企业治理研究主要聚焦传统治理主体、治理机制及制度环境,对数据资产等新型生产要素如何影响企业治理水平的关注仍较为有限。随着数字经济发展,数据资产与企业治理之间的关系逐渐受到关注。相关研究指出,当前企业治理面临信息权益保护与要素市场化配置之间的平衡问题[24],市场驱动型环境治理框架为缓解信息权益保护与资源配置之间的矛盾提供了新的思路[25]。与此同时,数据资产凭借价值融合性、实时性和非竞争性等特征[17]能够促进信息整合与资源优化配置,为企业治理机制创新提供了新的可能。数据资产相关信息披露能够改善资本市场信息环境,提高市场定价效率,说明数据资产不仅具有内部经营价值,也具有外部治理价值[26]。

然而,现有研究较少从企业治理视角考察数据资产的经济效应:一方面,数据资产相关研究主要聚焦价值创造、定价及测度等问题,对其微观治理效应关注相对不足;另一方面,企业治理研究主要考察治理主体、治理机制及企业特征等传统因素,对数据资产如何影响企业治理水平及其作用机制缺乏系统分析。鉴于此,本研究选取 A 股上市公司,在构建大模型测度数

据资产的基础上,系统考察数据资产对企业治理水平的影响,并从降低信息不对称和提升创新能力两个维度分析其作用机制,同时开展异质性分析,以拓展数据资产经济后果与企业治理研究。相较于已有文献,此次研究的边际贡献主要体现在以下方面:第一,在研究方法上,基于A股上市公司年报文本,运用生成式大语言模型(GLM-4)结合LoRA微调技术构建企业数据资产测度指标,从而弥补传统词频法在语义识别和上下文理解方面的不足,并为企业数据资产的微观测度提供新的研究工具。第二,在研究视角与研究内容上,从数据资产视角切入企业治理研究,拓展了数据要素经济后果研究边界,并进一步揭示数据资产影响企业治理水平的作用机制与异质性特征,丰富了数据资产赋能企业治理的理论研究。第三,在实践价值上,为企业挖掘数据资产价值、提升治理水平并推动数据要素市场化配置提供理论依据与实践启示。

二、理论分析与研究假设

1.数据资产对企业治理水平的直接影响效应

基于动态能力理论、制度理论与利益相关者理论,系统阐释数据资产对企业治理水平的作用机制:数据资产的形成和积累过程,主要借助动态能力理论来说明;数据资产嵌入企业组织、形成管理惯例的过程,则结合制度理论来展开;进入治理效应的分析之后,主要依赖委托代理理论和信息传递理论。相关研究提出,数据资产通过降低信息不对称与提升创新能力两条路径优化企业治理,并进一步从企业规模、现金流状况及区域差异等维度分析其异质性影响,为后续实证检验提供理论支撑。数字经济为数据资产提供了流通和增值的环境,而数据资产则是数字经济价值释放的关键桥梁。相较于传统生产要素,数据资产具有非竞争性、非消耗性和时效性等特性,能够成为治理框架中信息披露的核心载体,并通过实时数据共享机制强化股东对管理层的监督效能,降低代理成本[17]。就实践层面而言,数据资产的沉淀推动企业治理在决策流程、风险管控、激励机制等环节实现优化:依托大数据分析技术构建的智能决策系统,数据资产能够提升董事会决策的科学性与时效性;动态风险预警体系则能够助力管理层及时识别治理漏洞。基于动态能力理论,数据资产可通过实时采集市场动态、政策法规、竞争对手等多维度数据,帮助企业治理层及时捕捉行业趋势与潜在风险[6];数据资产作为跨部门的通用性资源,能够打破内部信息壁垒,从而增强企业治理的适应性与动态调整能力;基于制度理论,率先构建数据治理委员会且通过DCMM认证的企业,往往能形成“制度套利”优势,既满足《中华人民共和国数据安全法》等强制合规要求,又通过数据融合提升数据价值[27],实现在不完全制度环境下的治理创新。基于利益相关者理论,数据资产通过三维治理机制重构传统治理范式:信息整合维度通过融合多源数据构建利益主体诉求的精准识别系统;决策协同维度通过建立实时反馈回路降低委托代理冲突与交易成本;价值传导维度则通过资本市场的信息传递效应强化外部治理监督,最终促成治理效能从工具性改进向结构性跃迁的转变。从动态能力理论、制度理论及利益相关者理论3个理论视角的阐释表明,数据资产化通过优化企业治理结构、提升企业治理创新并加强各方主体对决策的认同,提高了企业治理水平。

基于以上分析,提出假说H1:数据资产有助于提升企业治理水平。

2.降低信息不对称的传导机制

由委托代理理论可知,信息不对称与激励不相容是企业治理面临的核心矛盾。由于企业的所有权与经营权分离,股东与管理层之间存在信息获取能力和信息处理能力差异,管理层可能会利用信息优势进行机会主义行为,从而增加代理成本并降低治理效率。数据资产具有通用性、非消耗性与动态性等特征[17],能够通过信息整合、实时反馈和价值挖掘改善企业信息环境,降低不同治理主体之间的信息偏差。一方面,数据资产的积累与应用能够促进企业内部信息资源整合,打通财务、生产、销售及运营等环节的信息壁垒,减少传统治理模式下因部门分割所形成的信息孤岛。通过对多维经营

数据的实时采集、分析和反馈,企业董事会、监事会及管理层能够更加全面及时地掌握企业运行状态,提高决策信息质量。同时,基于数据分析形成的智能监测和风险预警体系能使治理主体由传统的事后监督转向动态监督,及时识别经营风险和管理层机会主义行为,从而降低内部代理成本。另一方面,数据资产的规范化管理和可信流转能够提升企业外部信息披露质量,增强投资者、监管机构及其他利益相关者对企业经营状况的识别能力。在传统治理模式下,由于信息获取渠道有限,外部主体难以充分评价企业真实经营情况,易产生监督不足问题。而数据资产通过提高信息透明度和传递效率,能够构建企业与外部治理主体之间更加有效的信息沟通机制,强化资本市场监督和外部治理约束[28]。此外,数据资产的动态更新能力能够提高企业治理体系对外部环境变化的适应能力。相较于传统静态信息披露机制,数据资产能够持续反映企业经营变化、市场需求及风险因素,为治理主体提供更加及时的信息支持,推动企业治理由经验驱动向数据驱动转变。因此,数据资产能够通过改善内部信息传递效率、增强外部监督效能并提升治理响应能力,降低企业信息不对称程度,最终促进企业治理水平的提升。

基于以上分析,提出假说 H2:数据资产能够通过降低信息不对称进而提高企业治理水平。

3. 提升创新能力的传导机制

数据资产作为新型生产要素,能够通过赋能企业创新生态系统进而强化治理效能。动态能力理论认为,企业竞争优势来源于对内外部资源的持续感知、整合与重构能力。在数字经济环境下,数据资产突破了传统资源边界,能够帮助企业动态获取市场需求、技术趋势和竞争环境等信息,提高企业对外部变化的识别能力和内部资源配置效率,从而增强组织适应性和战略调整能力。一方面,数据资产能够推动企业创新资源的有效整合,提高创新活动效率。在研发、生产和经营过程中,数据资产通过连接企业内部各业务环节,实现信息共享与资源协同,降低因信息分散和资源错配引发的创新成本。同时,数据资产能够通过数据分析和智能决策工具提升企业对创新方向的判断能力,减少研发过程中的试错成本和协同创新成本[15]。随着数据资源的不断积累,企业可将原始数据转化为创新知识库、技术储备和研发资源,推动创新成果持续转化,拓展企业长期价值创造空间。另一方面,创新能力提升能够进一步改善企业治理环境。创新能力增强不仅能够提高企业产品竞争力和市场适应能力,还能够降低技术变化、市场波动等因素带来的经营不确定性,使企业经营结果更加稳定。在此基础上,股东与管理层之间由于未来发展预期差异产生的信息偏差和目标冲突得到缓解,代理成本随之降低。同时,创新能力提升也能够促进企业内部管理模式优化,提高资源配置效率,使治理主体能够基于更加科学的信息进行战略决策和监督管理。此外,数据资产赋能创新并非单纯提升企业研发产出,而是通过转变企业资源组织方式和价值创造模式,推动企业治理体系由被动监督向主动适应转变。创新能力的持续提升增强了企业应对环境变化的能力,使企业能够在复杂市场环境下长期保持竞争优势,并进一步促进治理机制完善。因此,数据资产通过提升企业创新能力,提高组织动态调整能力,降低经营不确定性并缓解代理冲突,最终提升企业治理水平。

基于以上分析,提出假说 H3:数据资产通过提升创新能力进而提升企业治理水平。

三、研究设计

为系统考察数据资产对企业治理水平的影响及作用机制,本研究基于 2011—2023 年 A 股上市公司数据,构建包含数据资产测度、企业治理水平评价及控制变量的综合研究框架。通过微调大语言模型(GLM-4)实现数据资产的精准量化,并采用主成分分析法构建企业治理水平综合指标,结合多元回归模型、中介效应检验及异质性分析,实证检验数据资产的治理效应及其边界条件,为数字经济时代企业治理优化提供理论依据与实证支持。

1. 样本选取与数据来源

选取 2011—2023 年中国 A 股上市公司作

为研究样本。样本期间起始于2011年,主要基于两方面考虑:一方面,随着数字经济发展和数据要素价值逐步受到重视,2011年后企业数据资源积累和数字化应用进入持续发展阶段;另一方面,企业年报文本、财务指标及治理相关数据在该阶段具有较好的连续性和可获得性,能够满足实证研究需要。样本截至2023年,以保证数据资产测度与企业治理指标之间具有较长的观察周期。样本处理方面,剔除ST、*ST、PT类公司样本,剔除金融行业企业,并进一步删除关键变量存在缺失值的样本,最终获得17 686个企业-年度观测值。为降低极端值对回归结果的影响,对连续变量进行1%和99%分位数缩尾处理。

数据来源方面,企业财务数据主要来自CSMAR数据库、Wind数据库和Choice数据库。数据资产指标所需文本数据来源于上市公司年度报告中的"管理层讨论与分析"(MD&A)部分,本研究基于人工标注数据对GLM-4模型进行LoRA微调,并从财务披露、项目实践、战略重视度、应用深度和价值转化共5个维度对企业数据资产水平进行量化测度。企业治理水平指标则依据监督、激励和决策等维度构建综合评价体系,并采用主成分分析法提取综合指标。

2. 变量定义

(1)被解释变量:企业治理水平(*GOVERNANCE*)

企业治理水平的测度采用主成分分析法[29-30],从监督、激励、决策等多个维度构建综合性指标。具体选用高管薪酬(*Mana_Pay*)与高管持股比例(*Mana_Share*)作为激励机制的代理变量,独立董事比例(*Outratio*)与董事会规模(*Board*)衡量董事会监督作用,机构持股比例(*Inst_Share*)与股权制衡度(*Share_Balance*)表征股权结构的监督作用,董事长与总经理是否两职合一(*Dual*)反映决策权力。对上述7个指标进行主成分分析,提取第一主成分作为企业治理水平综合指标,并取对数以提升变量正态性。

(2)解释变量:数据资产指标(*DS*)

数据资产的定量测度是当前研究的前沿难点,现有财务数据法样本有限且具有滞后性,传统词典法在语义理解与上下文语境方面存在不足。鉴于此,采用微调大语言模型的方法构建数据资产测度指标。首先,研究团队人工标注3 000条涉及数据资产的年报文本片段,由9名研究人员分组完成标注并经三重验证确保质量,将标注结果转换为JSON格式作为训练数据;其次,基于该标注数据集,采用LoRA(低秩自适应)技术对智谱清言(GLM-4)进行参数微调,使模型具备对数据资产相关信息的精准识别能力;最后,以企业年报中的"管理层讨论与分析"(MD&A)章节为分析载体,运用微调后的模型从财务披露、项目实践、战略重视度、应用深度及价值转化5个维度对企业数据资产进行量化评分(1~100分制),取对数后得到最终指标*DS*。选取MD&A章节而非全文,一方面保证了跨企业的可比性,另一方面可有效规避超长文本导致的模型注意力稀释问题。此外,将词频文本分析法所得指标(*DN*)取对数后作为替换变量进行稳健性检验。

测度结果显示(表1),*DS*全样本均值为68.29分,中位数为70分,2010—2023年数据资产发展呈三阶段特征:平稳增长期(2010—2016年)、加速攀升期(2017—2021年)和调整巩固期(2022—2023年),政策驱动效应显著(图1)。词频法所构建的数据资产测度指标能够有效识别并测度企业的数据资产特征,为相关实证研究提供了基础性度量工具[31]。参考何瑛等的测度框架[20],将词频法构建的数据资产指标(取对数处理)与数据资产得分(取对数处理)进行面板回归分析。指标有效性检验表明,*DS*与词频法指标*DN*的弹性系数为0.51($p<0.01$),二者在1%水平上显著正相关(表2),验证了本测度体系的有效性。

(3)中介变量

选取信息不对称(*ASY*)和创新能力(*NEW*)作为中介变量。信息不对称参考已有研究[32-33],构建流动性比率(*LR*)、非流动性比率(*ILL*)和收益率反转指标(*GAM*)3个股票流动性指标,并采用主成分分析法提取综合指标;创新能力参考杨林等的方法[34],以研发投入强度和技术人员比例两项指标标准化后加总,构建创新能力综合指数。

表 1　数据资产测度结果

变量	变量名称	样本量	均值	标准差	最小值	中位数	最大值
数据资产得分	*DS*	29 664	68.29	15.00	25	70	85
财务披露指数	*FDIS*	28 015	74.62	15.85	20	80	90
项目驱动指数	*PDIS*	27 914	66.78	21.47	10	75	95
战略定位指数	*SPIS*	27 895	64.44	21.02	10	70	95
应用深度指数	*ADIS*	27 764	55.46	20.26	5	60	90
变现能力	*MAS*	27 865	58.56	18.32	5	60	90
句频	*KSF*	27 943	15.90	11.16	1	15	70
正面比例/%	*PR*	29 408	48.67	22.01	5	60	80
负面比例/%	*NR*	29 387	23.98	17.14	0	20	90

注：***、**、* 分别表示在 1%、5%和 10%的水平上显著。下同。

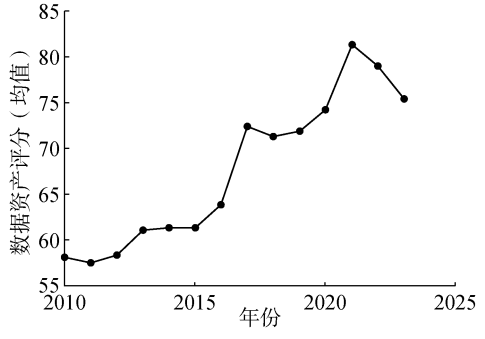

图 1　数据资产发展时序图

表 2　指标相关性检验结果

变量	*DS*
DN	0.510***(0.003)
_cons	4.082***(0.011)
固定效应	YES
R^2	0.210
N	17 686

(4)控制变量

参考已有做法[35]，选取如下控制变量：企业规模(*Size*)、总资产周转率(*ATO*)、现金流比率(*CashFlow*)、托宾 *Q*(*TobinQ*)、员工规模(*Employee*)。其中，企业规模控制了不同规模企业在治理结构和资源配置方面的系统性差异；总资产周转率控制了企业运营效率对治理效果的潜在影响；现金流比率控制了企业财务流动性对治理投入的约束作用；托宾 *Q* 值控制了市场估值预期带来的外部监督效应；员工规模则控制了企业人力资本规模对治理架构的塑造作用。这些控制变量的引入有效隔离了数据资产的独立影响，确保研究能够准确捕捉数据资产对企业治理水平的净效应，为研究结论的可靠性提供了坚实保障。

3. 模型设定

为检验假说 H1，设置式(1)考察数据资产对企业治理水平的影响：

$$GOVERNANCE_{i,t} = \alpha + \beta DS_{i,t} + \sum_{k=0}^{n} \gamma_k Controls_{k,i,t} + \varepsilon_{i,t} \quad (1)$$

式中：i 为企业；t 为时间；*GOVERNANCE* 为被解释变量，采用主成分分析法综合多个治理指标度量企业治理水平；*DS* 为解释变量，通过人工标注与大模型微调方法量化企业数据资产；*Controls* 为控制变量组，包括企业规模、总资产周转率、现金流比率等；ε 为随机误差项；α 为截距项；β 为核心系数，预期显著为正；γ_k 为各控制变量回归系数。

为有效检验假说 H2 和假说 H3，参考江艇的中介效应做法[36]，设定式(2)(3)分别考察信息不对称和创新能力作用机制的影响：

$$ASY_{i,t} = \alpha + \beta DS_{i,t} + \sum_{k=0}^{n} \gamma_k Controls_{k,i,t} + \varepsilon_{i,t} \quad (2)$$

$$NEW_{i,t} = \alpha + \beta DS_{i,t} + \sum_{k=0}^{n} \gamma_k Controls_{k,i,t} + \varepsilon_{i,t} \quad (3)$$

式中：ASY_{it} 和 NEW_{it} 为被解释变量，分别代表 i 企业 t 时期的企业信息不对称程度与创新能力。根据委托代理理论，信息不对称是制约企业治理效能的核心障碍。数据资产通过整合企业内外部信息、提升信息披露透明度，能够有效缓解管理层与股东之间的信息鸿沟，进而优化治理结构。因此，将信息不对称纳入中介变量，参考已有研究[33-34]，构建流动性比率(*LR*)、非

流动性比率(*ILL*)和收益率反转指标(*GAM*)3个股票流动性指标,采用主成分分析法提取综合指标 *ASY* 加以度量。根据动态能力理论可知,创新能力是企业将数据资产转化为治理效能的关键传导机制。数据资产通过降低创新试错成本、优化研发资源配置,持续提升企业创新水平,进而推动治理体系的现代化升级。因此,将创新能力纳入中介变量,参考杨林等的方法[34],以研发投入强度与技术人员比例两项指标标准化后加总构建综合指数 *NEW* 加以度量。

四、实证分析

本部分通过基准回归、稳健性检验与机制分析,系统考察数据资产对企业治理水平的影响。结果表明,数据资产显著提升了企业治理水平,且这一效应在小规模、低现金流及高成长性企业中更为突出。进一步分析揭示,降低信息不对称与增强创新能力是数据资产发挥作用的关键路径,为数字化转型背景下的企业治理优化提供了实证依据。

1. 描述性统计

根据表3的描述性统计,被解释变量企业治理水平(*GOVERNANCE*)的均值为0.171,标准差为0.983,表明不同企业间的治理水平存在较大差异。解释变量数据资产方面,基于大语言模型测度的 *DS* 与词频法测度的 *DN* 在均值与标准差上有所不同,体现了两种测量方法的差异性。信息不对称(*ASY*)的均值为负,标准差为0.618,个体间波动较为明显。控制变量中,企业规模(*Size*)分布相对集中,总资产周转率(*ATO*)、现金流比率(*CashFlow*)、托宾 *Q* 值(Tobin*Q*)及员工规模(*Employee*)等变量均呈现不同程度的离散特征,反映样本企业在营运效率、财务结构、市场估值和人力资本规模等方面存在显著差异。这些变量的异质性为后续的分组检验提供了分析基础。

2. 数据资产对企业治理水平的直接影响

本研究的基准回归模型控制了企业固定效应与年份固定效应,其中,企业固定效应用以吸收企业层面不随时间变化的异质性特征,年份固定效应则用于控制宏观政策调整与市场周期波动等时间趋势的共时冲击。为校正潜在的异方差与自相关问题,所有回归的标准误均在公司层面进行了聚类调整,以提升统计推断的稳健性。主效应分析采用多元线性回归模型,检验数据资产对企业治理水平的直接影响。基准回归结果表4列(1)显示,数据资产(*DS*)的系数显著为正,且在加入控制变量和固定效应后,如表4列(2)(3)所示,R^2 得到了明显的提升,系数依然稳健(0.310 至 0.364,均在1%水平显著),验证了假说H1。

考虑到可能存在的时间滞后性,将解释变量滞后一期和两期处理,重新进行的回归结果见表5列(1)(2),结果显示解释变量滞后一期和滞后两期均显著为正。基于滞后回归结果分析,解释变量滞后一期与滞后两期的系数估计值分别为0.368和0.403,这一实证发现表明,数据资产对企业治理水平的影响具有显著的时滞特征,其作用效果并非即时显现,而是呈现动态累积效应。滞后两期系数的较滞后一期有所

表3 描述性统计

变量	变量名称	均值	标准差	最小值	最大值
企业治理水平	*GOVERNANCE*	0.171	0.983	-1.752	2.374
数据资产	*DS*	4.249	0.217	3.401	4.50
数据资产	*DN*	3.283	0.720	1.791	5.799
信息不对称	*ASY*	-0.243	0.618	-8.270	1.991
全要素生产率	*TFP*	5.593	0.773	2.723	7.900
企业规模	*Size*	22.195	1.242	20.014	26.054
总资产周转率	*ATO*	0.635	0.381	0.102	2.344
现金流比率	*Cashflow*	0.050	0.066	-0.142	0.241
托宾 *Q* 值	Tobin*Q*	2.087	1.309	0.789	16.647
员工规模	*Employee*	4 890.180	9 202.339	72.000	71 736.000

表 4　基本回归说明结果

变量	(1) *GOVERNANCE*	(2) *GOVERNANCE*	(3) *GOVERNANCE*
DS	0.310***(0.034)	0.577***(0.029)	0.364***(0.043)
Constant	−1.149***(0.144)	7.447***(0.193)	8.197***(0.360)
控制变量	NO	YES	YES
固定效应	NO	NO	YES
R^2	0.004 7	0.251	0.278
N	17 686	17 686	17 686

表 5　滞后回归结果

变量	(1) *GOVERNANCE*	(2) *GOVERNANCE*
L. *DS*	0.368***(0.053)	
L2. *DS*		0.403***(0.057)
_cons	7.710***(0.458)	7.219***(0.486)
控制变量	YES	YES
固定效应	YES	YES
R^2	0.264	0.332
N	8 648	7 245

提升,这一递增趋势反映了数据资产价值释放的渐进性——随着时间推移,数据通过整合、分析与应用过程的逐步完善,其对治理结构的优化效应得以强化,从而提升企业治理水平。从方法论角度看,尽管滞后模型导致样本量缩减,但系数的统计显著性反而有所提升,这与表 4 基准回归结果形成稳健性互证,共同支持“数据资产提升企业治理水平”的假说 H1。研究结果提示,企业需认识到其治理改善效应具有明显的时序特征,并建立数据资产的长效积累机制,以逐步提升企业的治理水平。

3. 利用工具变量缓解内生性

数据资产可能与企业治理水平互为因果,或存在其他未观测因素同时影响两者。为此,选取企业总部所在地级市与杭州市之间的球面距离自然对数作为工具变量。杭州市作为全国数据要素市场的核心枢纽,集聚了阿里巴巴、浙江大数据交易中心等关键平台,地理距离直接影响企业获取数据要素的便利性,满足工具变量的相关性要求。而地理距离作为历史性变量,与企业治理水平无直接关联,满足排他性约束。

采用两阶段最小二乘法进行估计。第一阶段验证地理距离对数据资产的显著影响,第二阶段估计数据资产对企业治理水平的净效应。模型控制了企业规模、资产周转率等变量,并加入年份和行业固定效应,标准误聚类到企业层面。通过工具变量回归,可有效识别数据资产对企业治理水平的因果效应。表 6 列(1)显示,工具变量回归结果为企业所在地级市与杭州市的球面距离(*DISTANCE*)在第一阶段的回归系数显著为负,满足工具变量的相关性要求,并验证了数据资产会随着与杭州距离的增加而在一定程度上减小。第二阶段回归的结果如表 6 列(2)所示,数据资产的系数同样显著为正,与前文基准回归结论一致。这一结果表明,在使用工具变量控制潜在内生性问题后,数据资产对企业治理水平的正向促进作用依然稳健,进一步验证了研究结论的可靠性。

表 6　工具变量的检验结果

变量	(1)第一阶段	(2)第二阶段
DS		2.979**(1.490)
DISTANCE	-1.1×10^{-5}*(3.94×10^{-6})	
控制变量	YES	YES
固定效应	YES	YES
LM 统计量		7.597***
N	17 686	17 686

4. 稳健性检验

(1)PSM 匹配后的样本再回归

为降低样本选择偏误,采用倾向得分匹配方法(PSM),以控制变量为协变量进行 1∶1 近邻匹配,最终获得 12 072 个观测值。匹配后样本再回归结果如表 7 列(1)所示,结论依然显著为正,结果稳健。

(2)PSM-DID

以 2016 年实施的《大数据产业发展规划(2016—2020 年)》作为外生冲击,以各行业 2016 年数据资产中位数划分处理组与对照组,设置时间虚拟变量 *Post*。PSM-DID 回归结果如表 7 列(2)所示,结论显著为正;平行趋势

检验表明政策前系数不显著,满足了平行趋势假设。

安慰剂检验通过 500 次随机抽样回归,估计系数均分布在 0 附近(图 2),排除了其他干扰因素的影响。上述检验共同支持了核心结论的稳健性。

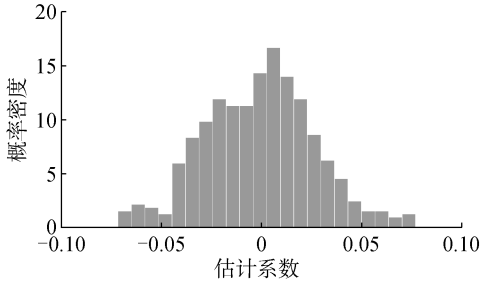

图 2　安慰剂检验结果

表 7 报告的 PSM-DID 回归因变量为企业治理水平的行业均值。考虑到政策冲击主要作用于数据要素市场环境,其对微观企业治理的影响路径具有显著的行业同质性,为准确评估政策冲击的行业净效应并避免企业个体层面的噪音干扰,本研究在双重差分估计中将样本压缩至“行业–年份”维度。经行业聚合及倾向得分匹配后,有效样本量为 72 个。

表 7　PSM、PSM-DID 稳健性检验结果

变量	PSM	PSM-DID
DS	0.452***(0.055)	
DID		0.457**(0.209)
_cons	7.668***(0.414)	6.464(13.101)
控制变量	YES	YES
固定效应	YES	YES
R^2	0.283	0.966
N	12 072	72

为保证结论稳健,在随后的平行趋势检验与安慰剂检验中,还原到企业层面进行分析,以充分验证微观主体在政策实施前后的动态一致性。因此,表 8 平行趋势检验的报告样本量(4 170 个)为企业–年份层面观测值。

5. 其他稳健性检验

(1)替换解释变量为词频法测度数据资产

由于数据资产指标(*DS*)基于大语言模型构建,可能存在方法依赖性。为验证核心指标的测量误差是否影响结果,分别将传统的词频法文本分析得到的数据资产指标作为解释变量,其余部分均不变,再次进行回归检验。首先实施两种不同方法得到的数据资产指标的回归分析,然后用词频法对企业治理水平进行回归分析。从表 9 列(1)的结果可以看出,词频法数据资产对 *DS* 具有显著正向影响,可作为替换的解释变量使用;表 9 列(2)的结果表明词频法数据资产对企业治理水平均具有显著正向影响,证实了数据资产可以驱动企业治理水平的提升。传统词频法仅统计关键词频次,而 GLM-4 模型通过微调和语义分析,能识别语境中隐含的数据资产信息(如数据中台建设、智能决策系统等复合概念),更全面地捕捉企业数据资产的实际价值,这也在系数的提升上得到了体现。

表 8　平行趋势检验结果

变量	PSM-DID
DID	0.125**(2.46)
pre_2013	0.00571(0.05)
pre_2014	0.116(1.17)
pre_2015	0.0487(0.54)
_cons	8.951***(0.652)
控制变量	YES
固定效应	YES
R^2	0.287
N	4 170

(2)替换解释变量为零样本测度数据资产

由于前文的数据资产指标(*DS*)是通过特定的提示词模板与模型微调构建的,可能存在一定的方法依赖或过拟合偏误。为验证结论是否受制于这种特定的测度手段,采用零样本(Zero-shot)提示法重新提取的数据资产指标(*DSzero*)作为替换的解释变量,其余控制变量与固定效应保持不变,再次进行回归检验。从表 9 列(3)的结果可以看出,零样本测度的数据资产指标对企业治理水平依然具有显著的正向影响(系数为 0.779,且在 1% 的水平上显著),再次证实了数据资产可以驱动企业治理水平的提升。相较于经过专门微调的模型,零样本测度完全依赖大语言模型原生的泛化能力与基础语义理解,未向模型注入特定领域的先验知识。在这种更为严苛且无干预的基础测量模式下,解释变量依然保持高度显著,充分说明数据资产赋能公司治理的逻辑是客观存在且高

表 9 替换自变量、加入省份固定效应与数字化转型作为控制变量的回归结果

变量	(1) *DS*	(2) *GOVERNANCE*	(3) *GOVERNANCE*	(4) *GOVERNANCE*	(5) *GOVERNANCE*
DN	0.05***(0.003)	0.162***(0.022)			
DSzero			0.779***(0.124)		
DS 加入省份固定效应				0.301***(0.043)	
DS 加入数字化转型					0.295***(0.042)
Constant	3.427***(0.049)	8.930***(0.322)	9.060***(0.340)	8.042***(0.360)	8.467***(0.357)
控制变量	YES	YES	YES	YES	YES
固定效应	YES	YES	YES	YES	YES
R^2	0.234	0.280	0.273	0.301	0.282
N	17 686	17 686	13 449	17 686	17 686

度稳健的,其并非特定大模型微调策略带来的偶然结果。列(3)样本量减少至 13 449,是零样本测度方法对部分样本未能生成有效评分所致。

(3)加入省份固定效应与数字化转型

为控制地区不可观测的异质性,加入省份层面固定效应重新回归,结果如表 9 列(4)所示,数据资产对企业治理水平的正向影响依然显著,凸显其突破地域限制的特性。此外,为排除数字化转型整体趋势对结果的干扰,将数字化转型纳入控制变量再次回归,表 9 列(5)结果表明,即便控制数字化转型后,数据资产仍是驱动企业治理水平提升的关键要素,进一步验证了结论的稳健性。

五、进一步研究

在基准分析基础上,本部分通过异质性检验深入考察数据资产治理效应的边界条件,揭示其在企业规模、财务特征及区域分布等维度的差异化表现,并基于信息不对称与创新能力双重视角解析作用机制,为完善数据要素市场化配置与企业治理现代化提供政策启示。

1. 作用机制检验

(1)信息不对称的作用机制检验

表 10 的信息不对称机制检验结果表明,数据资产对企业治理水平的提升作用存在显著的中介路径。表 10 列(1)显示,数据资产对企业治理水平的直接效应显著为正,证实数据要素本身具有治理优化功能;列(2)显示,数据资产能显著降低信息不对称程度,而信息不对称的改善又进一步促进了治理水平提升;列(3)显示,当同时控制数据资产和信息不对称时,数据资产的系数虽有所降低但仍保持高度显著,这完整揭示了“数据资产→降低信息不对称→提升治理水平”的作用机制,表明数据资产能够通过降低信息不对称提升企业治理水平,假说 H2 成立。具体而言,数据资产通过增强信息披露透明度、优化信息传递效率等途径,有效缓解了企业内外部信息壁垒,进而减少了代理成本、强化了监督机制,最终实现治理效能提升。这一发现不仅验证了信息不对称在数据资产与公司治理关系中的关键中介作用,也为企业数字化转型提供了重要启示。为此,应当重点培育数据要素的信息整合能力,通过构建智能化信息

表 10 信息不对称的机制检验结果

变量	(1) *GOVERNANCE*	(2) *ASY*	(3) *GOVERNANCE*
DS	0.364***(0.043)	−0.177***(0.016)	0.326***(0.042)
ASY			−0.212***(0.023)
Constant	8.197***(0.360)	8.035***(0.210)	9.902***(0.393)
控制变量	YES	YES	YES
固定效应	YES	YES	YES
R^2	0.27	0.59	0.28
N	17 686	17 686	17 686

平台来系统性改善治理结构,同时监管机构要推动数据治理标准建设,为数据驱动的治理模式创新提供制度保障。

(2)创新的作用机制检验

表 11 的创新能力机制检验揭示了数据资产提升企业治理水平的关键路径。研究显示,数据资产不仅对治理水平具有直接促进作用,还能显著提升企业创新能力。如表 11 列(3)所示,当同时引入数据资产和创新能力变量时,创新能力表现出显著的正向影响,而数据资产系数虽从 0.363 降至 0.320,但仍保持高度显著性,完整呈现了"数据资产→加强创新能力→提高治理水平"的传导链条,表明数据资产能够通过提升创新能力提升企业治理水平,假说 H3 成立。具体而言,这种传导机制主要通过 3 个维度实现:首先,数据驱动的流程再造激发了组织创新活力。通过实时监测生产运营数据,企业能够快速识别创新机会点,将原先用于纠错的管理资源转化为创新投入,形成"数据反馈-快速迭代"的创新闭环。其次,数据技术赋能研发体系升级,通过智能算法辅助研发决策、预测性分析缩短创新周期,创新效率的提升直接增强了企业的价值创造能力,为治理改进提供可持续的资金支持。最后,数据要素与人才团队的深度融合催生了"数据+创新"的共生生态,这种生态通过知识共享平台和数字化协作机制,持续培育创新型治理文化。

2. 异质性分析

采用异质性分析方法,通过多维度比较不同企业特征下的作用效应,系统考察数据资产对企业治理水平影响的差异化表现。在数字化转型背景下,企业治理面临机遇与挑战,数据资产作为核心战略资源,其对治理水平的影响受到多重因素制约。为此,重点从企业规模、营运效率、现金流比率及企业区位 4 个关键维度,深入解析数据资产对企业治理的差异化影响机制,以期为提高企业治理水平提供精准的理论支撑和实践指导。

(1)基于企业规模的异质性

《中共中央办公厅　国务院办公厅关于完善中国特色现代企业制度的意见》强调,要根据企业规模、发展阶段、所有制性质等分类施策。因科层制管理束缚,大型企业管理层更依赖传统经验,规模优势形成的风险缓冲带也使数据资产的边际贡献被摊薄;中小企业组织结构扁平、生存压力更大,更需借助数据资产快速修正治理偏差。本研究以企业市值中位数划分大、小规模企业分组回归,验证数据资产治理效应的规模差异。

表 12 列(1)(2)的实证结果显示,相较于大规模企业,数据资产对小规模企业治理水平的提升效应更为显著,两类企业的回归系数均在 1%的统计水平上显著,验证了该结论的可靠性与稳健性。综上,数据资产对企业治理水平有显著的正面影响,并且这种影响在大规模与小规模企业之间存在显著差异。

(2)基于现金流比率的异质性

《国务院关于进一步提高上市公司质量的意见》提出,要优化现金流管理,提高资金使用效率。高现金流比率企业因资金冗余易产生管理松懈与路径依赖,往往将数据资产仅作为辅助工具,对数据驱动的流程再造缺乏紧迫感;低现金流比率企业迫于经营压力,容错空间有限,必须借助数据资产精准识别资金浪费点与价值增长点,更积极地将其转化为治理优化工具。本研究以现金流比率中位数为基准划分高、低现金流组别进行回归分析,验证数据资产对不同财务特征企业治理水平的差异化影响。

表 11　创新能力的机制检验结果

变量	(1) *GOVERNANCE*	(2) *NEW*	(3) *GOVERNANCE*
DS	0.363*** (0.043)	0.207*** (0.012)	0.320*** (0.042)
NEW			0.205*** (0.052)
Constant	8.211*** (0.361)	−0.292*** (0.100)	8.271*** (0.358)
控制变量	YES	YES	YES
固定效应	YES	YES	YES
R^2	0.27	0.45	0.28
N	17 637	17 637	17 637

表 12 异质性分析检验结果

变量	(1)小规模	(2)大规模	(3)低现金流比率	(4)高现金流比率	(5)低成长性	(6)高成长性	(7)东部	(8)非东部
DS	0.369***	0.300***	0.430***	0.265***	0.305***	0.422***	0.367***	0.271***
	(0.046)	(0.046)	(0.043)	(0.049)	(0.045)	(0.048)	(0.041)	(0.054)
Constant	7.153***	6.859***	7.491***	8.321***	9.059***	6.808***	7.879***	7.414***
	(0.488)	(0.348)	(0.296)	(0.306)	(0.326)	(0.313)	(0.258)	(0.366)
控制变量	YES	YES	YES	YES	YES	YES	YES	YES
固定效应	YES	YES	YES	YES	YES	YES	YES	YES
R^2	0.21	0.11	0.26	0.29	0.34	0.19	0.27	0.26
N	8 843	8 843	8 843	8 843	8 843	8 843	12 864	4 818

回归结果如表 12 列(3)(4)所示,低现金流比率企业的回归系数显著高于高现金流比率企业的回归系数,表明数据资产对低现金流企业的治理提升效果更为突出。这一发现通过了1%水平的显著性检验,结果具有较好的统计可靠性。综上,数据资产对企业治理水平有显著的正面影响,并且这种影响在低现金流比率与高现金流比率之间存在显著差异。

(3)基于托宾 *Q* 的异质性

《国务院关于进一步提高上市公司质量的意见》为上市公司通过科技创新提升市场价值提供了政策指引。高托宾 *Q* 企业融资条件优越、管理层更重视数据资产的战略价值,治理赋能作用得以充分释放;低托宾 *Q* 企业则受制于资本成本高与短期业绩压力,易碎片化应用数据资产,治理效能提升相对受限。因此,本研究以托宾 *Q* 中位数为基准划分高、低估值企业分组回归,验证数据资产治理效应的估值差异。

回归结果如表 12 列(5)(6)所示,高托宾 *Q* 企业的回归系数显著高于低托宾 *Q* 企业的回归系数,表明数据资产对高托宾 *Q* 的治理提升效果更为突出。这一发现通过了 1%水平的显著性检验,结果具有较好的统计可靠性。综上,数据资产显著提升了企业治理水平,且该提升效应在高、低托宾 *Q* 企业间存在显著差异。

(4)基于地区的异质性

《长江三角洲区域一体化发展规划纲要》等数字经济相关政策共同塑造了东部地区领先的数字化发展环境,凭借完善的市场机制、数字基础设施与人才资源,叠加东部企业深度嵌入全球价值链的竞争压力,使数据资产对东部地区企业治理效能提升的边际贡献显著大于非东部地区。以企业注册地划分东部与非东部样本进行回归分析,验证数据资产治理效应的区域差异。

回归结果如表 12 列(7)(8)所示,东部企业的回归系数显著高于非东部企业的回归系数,表明数据资产对东部企业的治理提升效果更为突出。这一发现通过了 1%水平的显著性检验,结果具有较好的统计可靠性。综上,数据资产显著提升了企业治理水平,且该提升效应在东部、非东部企业间存在显著差异。

六、结论与建议

以 2011—2023 年中国 A 股上市公司为研究样本,基于生成式大语言模型(GLM-4)结合 LoRA 微调技术构建企业数据资产测度指标体系,系统考察数据资产对企业治理水平的影响及作用机制。基准回归结果表明,数据资产能够显著提升企业治理水平,该结论在替换核心变量、内生性处理及一系列稳健性检验后依然成立。机制分析表明,数据资产主要通过两条传导路径促进企业治理水平提升:一是降低信息不对称,数据资产通过整合企业内部经营信息、提升信息披露质量和强化外部监督机制,提高企业信息透明度,缓解管理层与股东之间的代理冲突;二是提升创新能力,数据资产通过优化创新资源配置、降低创新试错成本及增强市场响应能力,推动企业创新能力提升,并进一步促进治理体系优化。异质性分析发现,数据资产的治理赋能效应在小规模、低现金流比率、高成长性及东部地区企业中表现更为显著,表明企业内部资源约束、成长潜力及外部数字化环

境是影响数据资产治理价值释放的重要因素。进一步分析表明,数据资产并非简单的信息资源,而是通过“数据资源积累-能力形成-治理优化”的路径实现价值转化,为数字经济背景下企业治理模式创新提供了新的经验证据。

基于上述研究结论,提出以下政策建议:第一,企业应加强数据资产管理体系建设,推动数据资源向治理能力转化。研究表明,数据资产能够通过改善信息环境、提高信息透明度降低企业内部和外部治理主体之间的信息不对称。因此,企业应建立完善的数据治理体系,加强财务、运营、市场等多源数据整合,推动信息共享机制建设,提高董事会、管理层及投资者对企业经营状况的识别能力。同时,应结合企业自身发展阶段探索数据资产应用场景,将数据资源嵌入战略决策、风险管理和内部监督等治理环节,充分发挥数据资产在提升治理效率方面的作用。

第二,企业应深化数据资产与创新体系融合,释放数据要素的长期价值。机制分析表明,数据资产能够通过提升企业创新能力促进治理水平优化。因此,企业应推动数据资产与研发管理、生产运营及市场分析深度结合,利用数据分析技术优化创新资源配置,降低研发过程中的试错成本,提高创新活动效率。同时,应加强数据人才培养和组织协同建设,促进技术、管理与业务部门之间的信息融合,推动企业形成数据驱动的创新模式。

第三,监管部门应完善数据资产相关制度建设,提升企业数据治理规范化水平。当前企业数据资产仍面临确权、计量、披露等制度问题,影响数据价值的充分释放,为此,监管部门应进一步完善数据资产信息披露规则和评价体系,引导上市公司规范披露数据资源管理、开发利用和价值创造情况,提高资本市场对企业数据资产价值的识别能力,为企业数据资产治理提供制度保障。

第四,政府应优化数据要素发展环境,促进数据资产价值均衡释放。异质性分析表明,数据资产治理效应受到企业规模、资源条件及区域数字化水平影响,为此,政府应进一步完善数字基础设施建设,推动数据流通和共享机制完善,缩小区域间数字发展差异。同时,应针对中小企业数据治理能力不足的问题,提供公共数据服务、技术支持和数字化转型指导,降低企业应用数据资产的门槛,促进不同类型企业共享数字经济发展成果。

参考文献:

[1] 南开大学中国公司治理研究院公司治理评价课题组. 2024 年上市公司治理水平稳步提升[N]. 上海证券报,2024-11-14(04).

[2] 埃森哲. 2025 埃森哲中国企业数字化转型指数[R]. 2025:3-18.

[3] Rassier D G, Kornfeld R J, Strassner E H. Treatment of data in national accounts[R]. Paper Prepared for the BEA Advisory Committee, Bureau of Economic Analysis, 2019:5-7.

[4] Fernandez R C, Subramaniam P, Franklin M J. Data market platforms: trading data assets to solve data problems[J]. Proceedings of the VLDB Endowment, 2020,13(12):1933-1947.

[5] Goodridge P, Haskel J. How much is UK business investing inbigdata? [R]. Working Papers 25159, Imperial College Business School, 2015:10-11.

[6] Ahmad N, Vande Ven P. Recording and measuring data in the system of national accounts [R]. OECDWorking Party on National Accounts, 2018: 7-13.

[7] 杨东,高一乘. 赋能新质生产力发展:企业数据资产“确权”的三重维度[J]. 商业经济与管理,2024 (4) : 83-93.

[8] 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]. 经济研究,2021,56(12):42-59.

[9] 许宪春, 张美慧. 中国数字经济规模测算研究——基于国际比较的视角[J]. 中国工业经济,2020(5):23-41.

[10] 蔡跃洲,牛新星. 中国数字经济增加值规模测算及结构分析[J]. 中国社会科学,2021(11):4-30.

[11] 许宪春,雷泽坤,胡亚茹. 中国数据资本形成总额与数据资本存量测算[J]. 经济研究,2025(3): 54-68.

[12] Lu Xiaomeng, Lai Yali, Zhang Yong. Digital financial inclusion and investment diversification: evidence from China [J]. Accounting & Finance, 2023, 63 (S2):2781-2799.

[13] 许中缘,郑煌杰. 赋能新质生产力:数据要素资产化的法律配置[J]. 湖北大学学报(哲学社会科学版),2024,51(5):136-146.
[14] 刘涛雄,张亚迪,戎珂,等. 数据要素成为中国经济增长新动能的机制探析[J]. 经济研究,2024,59(10):19-36.
[15] 徐翔,赵墨非,李涛,等. 数据要素与企业创新:基于研发竞争的视角[J]. 经济研究,2023,58(2):39-56.
[16] 刘友金,张颖斌,蔡翔. 从技术鸿沟到技术追赶:数字经济发展效应研究——来自“一带一路”沿线国家的证据[J]. 经济研究,2024,59(11):192-208.
[17] 许宪春,张钟文,胡亚茹. 数据资产统计与核算问题研究[J]. 管理世界,2022,38(2):16-30.
[18] Veldkamp L. Valuing data as an asset[J]. Review of Finance,2023,27(5):1545-1562.
[19] Hu Changyu,Li Yuetong,Zheng Xiaojia. Data assets, information uses, and operational efficiency [J]. Applied Economics,2022,54(60):6887-6900.
[20] 何瑛,陈丽丽,杜亚光. 数据资产化能否缓解“专精特新”中小企业融资约束 [J]. 中国工业经济,2024(8):154-173.
[21] 郑志刚,李邈,雍红艳,等. 中小股东一致行动改善了公司治理水平吗? [J]. 金融研究,2022(5):152-169.
[22] 石晓军,王骜然. 独特公司治理机制对企业创新的影响——来自互联网公司双层股权制的全球证据[J]. 经济研究,2017,52(1):149-164.
[23] 张天舒,陈信元,黄俊. 独立董事薪酬与公司治理效率[J]. 金融研究,2018(6):155-170.
[24] 汪浩,张俊妮. 连带责任、数据交易与信息权益保护[J]. 经济研究,2024,59(9):118-135.
[25] 张云,吕纤,韩云. 机构投资者驱动企业绿色治理:监督效应与内在机理[J]. 管理世界,2024,40(4):197-221.
[26] 李世刚,邵宏彬,方芳,等. 企业数据资产信息披露与资本市场定价效率[J]. 中国工业经济,2025(7):138-155.
[27] Li W C Y, Nirei M, Yamana K. Value of data: there's no such thing as a free lunch in the digital economy[R]. Tokyo,Japan:RIETI,2019:19-23.
[28] 陶克涛,张术丹,赵云辉. 什么决定了政府公共卫生治理绩效? ——基于 QCA 方法的联动效应研究[J]. 管理世界,2021,37(5):128-138.
[29] 周茜,许晓芳,陆正飞. 去杠杆,究竟谁更积极与稳妥? [J]. 管理世界,2020,36(8):127-148.
[30] 顾乃康,周艳利. 卖空的事前威慑、公司治理与企业融资行为——基于融资融券制度的准自然实验检验[J]. 管理世界,2017(2):120-134.
[31] 苑泽明,于翔,李萌,等. 数据资产促进了中国企业人力资本水平提升吗? ——基于文本分析法的经验证据[J]. 南开管理评论,2025,28(8):64-75.
[32] 宋敏,周鹏,司海涛. 金融科技与企业全要素生产率——“赋能”和信贷配给的视角[J]. 中国工业经济,2021(4):138-155.
[33] 于蔚,汪森军,金祥荣. 政治关联和融资约束:信息效应与资源效应[J]. 经济研究,2012,47(9):125-139.
[34] 杨林,和欣,顾红芳. 高管团队经验、动态能力与企业战略突变:管理自主权的调节效应[J]. 管理世界,2020,36(6):168-188.
[35] 翟淑萍,韩贤,张晓琳,等. 数字金融能降低企业债务违约风险吗[J]. 会计研究,2022(2):117-131.
[36] 江艇. 因果推断经验研究中的中介效应与调节效应[J]. 中国工业经济,2022(5):100-120

Data Assets Empowering Corporate Governance Based on Generative Large Language Models /

Yin Tao[1], Wang Yaomin[1], Meng Lingchao[2] (1. Business School, East China University of Political Science and Law; 2. China School of Banking and Finance, University of International Business and Economics)

Abstract: Data assets, as core strategic resources and key governance elements in the digital economy era, can enhance corporate governance through signaling and innovation-driven mechanisms. Based on data from Chinese A-share listed companies from 2011 to 2023, this study employs a large language model fine-tuned with a manually collected corpus to construct a corporate data asset evaluation index system and empirically examines the impact of data assets on corporate governance. The results show that data assets significantly improve corporate governance, and the findings remain robust after controlling for firm characteristics and fixed effects. Robustness tests, including lagged variable estimation, instrumental variable approach, propensity score matching, difference-in-differences analysis, and replacement of core explanatory variables, confirm the robustness of the conclusions. Compared with the overall effect of digital transformation, data assets place greater emphasis on the development, utilization, and value realization of data resources, and their marginal governance benefits significantly exceed the overall effect of digital transformation. Mechanism analysis indicates that data assets enhance corporate governance through two channels: reducing information asymmetry and improving innovation capability. Specifically, data assets improve information transmission efficiency, alleviate internal and external information barriers, optimize resource allocation, and enhance innovation efficiency, thereby strengthening corporate governance. Heterogeneity analysis reveals that these effects are more pronounced among smaller firms, firms with lower cash flow ratios, high-growth firms, and firms located in eastern China. This study enriches literature on the economic consequences of data assets and corporate governance, reveals the mechanisms and boundary conditions through which data assets empower corporate governance, provides theoretical explanations and practical implications for how enterprises leverage data assets to improve governance in the context of digital transformation, and offers policy implications for promoting the value realization of data assets, improving data factor allocation mechanisms, and advancing modern corporate governance systems.

Key words: data assets; corporate governance; large language models; information asymmetry; innovation capability