

企业数据要素应用同群效应及其实现机制研究

赵 丽 胡植尧

内容提要:数字经济时代,数据要素成为重要的战略性资源,企业彼此竞相模仿应用数据要素。本文以2007—2021年中国上市公司数据为样本,实证研究企业数据要素应用的行业同群效应及其实现机制。研究结果显示,企业数据要素应用具有行业同群效应,即同行企业数据要素应用对企业数据要素应用具有促进作用,且经过内生性分析和稳健性检验后,研究结论依然成立。内部实现机制表明,企业代理成本越高、内部数据要素循环累积越多,企业数据要素应用的行业同群效应越强;外部实现机制表明,信息学习越强、行业竞争越激烈,企业数据要素应用的行业同群效应越强。进一步分析结果表明,企业数据要素应用行业同群效应遵循“先内后外”定律,即企业更倾向于模仿产权性质相同、规模类似的同行企业。本文的研究在一定程度上有利于降低企业决策成本、决策风险、转型风险以及不确定性,摆脱企业数据要素“不敢应用”的局面。

关键词:数据要素 数据要素应用 行业同群效应 实现机制 内外部影响

中图分类号:F270.7

文献标识码:A

文章编号:1000-7636(2025)08-0019-18

一、问题提出

数字经济时代,以人工智能、区块链、云计算、大数据为代表的“数字技术群”的蓬勃发展以及数字基础设施的广泛连接,促使数据要素成为关键生产要素、战略性资源。从政策层面看,2019年10月,党的十九届四中全会将数据要素纳入生产要素体系;2022年1月,国务院印发的《“十四五”数字经济发展规划》强调,数据要素是数字经济深化发展的核心引擎;2024年1月,国家数据局等部门联合印发《“数据要素×”三年行动计划(2024—2026年)》,推动充分发挥数据要素乘数效应。从实践层面看,根据《中国数字经济发展研究报告(2024)》,中国2023年数字经济规模达到53.9万亿元,占国内生产总值(GDP)的比重达到42.8%,有效支撑经济稳增长。一方面,以数字平台企业为代表的新型企业积极布局大数据发展战略,建立数据中心、增强数据学习效应,优化支撑数据价值释放的算力和算法,抢占数据要素发展的战略先机;另一方面,以制造业企业为代表的传统企业利用数据要素的低成本、大规模可得、即时生成、无

收稿日期:2024-09-03;修回日期:2025-06-16

基金项目:国家社会科学基金年度项目“‘技术-经济’范式下企业新质生产力发展机制与路径研究”(24CJY007)

作者简介:赵 丽 宁波大学商学院讲师,宁波,315211;

胡植尧 中国社会科学院大学经济学院博士研究生,通信作者,北京,102488。

作者感谢匿名审稿人的评审意见。

限迭代等特性,逐步渗透、赋能企业经营管理的各方面,带来包括企业思维模式、组织模式、研发模式、生产模式、营销模式、营运模式的创新^[1],实现数据要素由企业的单点渗透向全局应用转变,加强数据要素应用的频度、广度和深度,促进企业跨越式发展。在政策和实践的双重力量作用下,数据要素的嵌入和赋能成为企业发展的重要支点,蕴藏着巨大的商业价值。

数据要素应用将数据的价值属性与实体经济有效连接,通过要素驱动、融合激发、协同提升、决策反馈,促进企业模式创新、产业关联创造、智能决策实现^[2],有效激发数据活力,释放数据要素赋能实体经济的价值,引领企业高质量发展。已有研究验证了数据要素应用的经济价值^[3],包括提升制造业企业创新质量^[4]、促进企业市场价值实现^[5]、提高企业全要素生产率^[6]、变革决策范式^[7]、赋能供应链敏捷^[8]等。然而,对企业数据要素应用影响因素的实证研究尚不多见,现有研究主要采用理论分析方法,从生产力属性^[1]、“技术-经济”范式变革^[9]、外部市场环境和用户需求变化^[10]等方面考察企业数据要素应用的影响因素。

传统理论假设企业的经济决策是基于企业基本特征而做出的,不具有群体关联性。然而,现实是企业的经济决策往往模仿、学习同行企业的决策行为,即存在企业同群效应。现有关于企业同群效应的研究主要聚焦于同群效应识别研究、实现机制研究以及经济后果研究三方面,探究资本结构^[11-13]、违规行为^[14]、企业投资^[15]、金融资产配置^[16]、数字化转型^[17]等内容。陈庆江等(2021)认为企业内部资源基础和动态能力,以及外部环境的网络嵌入、市场竞争、环境不确定性条件都会使行业内企业数字化转型决策具有同群效应^[18]。虽有少数学者对数字经济时代,企业数字化转型的同群效应进行研究,但数字技术群是一个庞大且泛化的概念,而数据要素作为数字技术应用的新生产要素,具有重大的现实意义。

梳理上述文献可以发现,既往研究更偏重数据要素的经济价值,关于如何推动数据要素应用的影响因素的论述不足,更鲜有从社会关系和行为互动角度实证检验企业数据要素应用的影响因素。这在一定程度上不利于解释企业应用数据要素的决策逻辑和行为机理^[18]。企业或经营个体的经济决策在很大程度上受社会关系和行为互动的影 响,这就决定从地位类似、互动密切的同行业群体的角度探讨数据要素应用的影响因素很有必要。

本文以2007—2021年中国沪深A股上市公司为样本,实证检验企业数据要素应用同群效应及其实现机制。本文的边际贡献体现在三方面。一是实证分析行业数据要素应用对焦点企业数据要素应用的影响。这说明企业不仅考虑其自身特征,还综合参考同行企业的数据要素应用战略决策,形成系统化的决策过程,最大程度降低企业决策成本、决策风险、转型风险以及不确定性,摆脱企业数据要素“不敢应用”、数字化“不敢转型”的局面。二是分别从企业代理成本和内部数据要素循环累积以及从信息学习和行业竞争维度,探究企业数据要素应用行业同群效应的内部和外部实现机制。不同于聚焦企业外部实现机制,代理成本和内部数据要素循环累积更聚焦于企业内部实现机制。这不仅丰富了数据要素应用行业同群效应的实现机制,还为企业进一步制定数据要素应用战略提供理论和实践参考。三是对数据要素应用行业同群效应的研究能有效拓宽数据要素应用研究的范畴,使其不仅局限于数据要素应用的经济价值,也能从行业同群效应视角追溯数据要素应用的影响因素,将同行企业数据要素应用纳入焦点企业数据要素应用的理论分析框架,进而丰富行业同群效应的研究范围。

二、理论分析及研究假设

(一) 数据要素

数字经济时代,以人工智能、区块链、云计算、大数据为代表的数字技术群的蓬勃发展,促进数字技术与

实体经济深度融合,一切皆可数字化、数据化,数据的大规模可得性和价格低廉使得数据成为数字经济时代关键生产要素^[9]。广义的数据是指在统计、计算、科学研究中所依赖的数值资源或者以数字技术为核心的非物质经济资源。数据是对客观事物的逻辑归纳和表达^[19]。数据是一种符号记录^[20],不仅包括已经转化为数字形式的对现实世界的观察,还包括未经转化为数字形式的图像、录音、文字等。通常来说,转化为数字形式的数据,一般是以“0-1”形式存在的二进制信息单元,更易于被收集、存储、分析,从中提取出有价值的信息和知识^[21],而未经转化为数字形式的半结构化和非结构化数据需要将其转化为结构化的数字形式数据后方可进行存储和分析。可以看出数据是信息和知识的载体和原材料,它是能够提高预测精准性的信息^[21]。

琼斯和托内蒂(Jones & Tonetti, 2020)^[22]认为数据是除了知识和创意外的信息,它不能直接产生经济价值,需要通过数据采集、数据挖掘、数据分析、数据整理“提纯”出对国民经济运行和生产产品服务基本投入的信息要素^[23],即“提纯”为知识,进而赋能实体经济。根据上述观点,本文对数据要素应用进行界定,认为企业数据要素应用是指企业通过数据采集、数据存储、数据挖掘、数据分析、数据整理,从数据中“提纯”出有价值的信息和知识,并投入企业经济运行及生产经营过程中的必要的基本社会资源。需要指出的是,企业数据要素应用之前必须经历数据收集、存储、挖掘、分析以及整理。

(二) 同群效应

同群效应的研究最早可追溯到教育经济学领域,研究对象为个体,认为个体的行为和决策不仅受自身因素的影响,还受相同群体内其他个体的影响,表现出群体内个体行为和决策的趋同^[24]。同群效应经常被译为 peer effect 或 group effect^[15]。

随着研究的深入,对同群效应的研究逐渐从教育经济学向经济学、管理学领域延伸,同群效应研究的主体也从“个人”向“企业”转变,即认为企业的行为和决策不仅受企业自身因素的影响,还受周围参照群体内其他企业行为和决策的影响。在受参照群体内其他企业影响的过程中,参照群体的行为和决策改变了焦点企业的行为和决策,认为焦点企业在制定决策时,并非独立的,而是在与企业的社会互动过程中做出最优决策^[25]。其中,参照群体是由互相影响的个体组成的集合。焦点企业是指受到参照群体影响的个体企业^[25]。同群效应根据其群体划分差异,大致可分为行业同群效应和地区同群效应两种^[26]。其中,行业同群效应是指焦点企业的行为和决策受到同行业内除本企业外其他企业的行为和决策的影响。

根据上述分析,本文以数据要素应用为核心内容,研究数据要素应用的行业同群效应。鉴于此,本文对数据要素应用行业同群效应界定为,同行企业数据要素应用对焦点企业数据要素应用的影响。需要说明的是,虽然同群效应可分为行业同群效应和地区同群效应,而本文主要研究行业同群效应,考虑到地区内企业数据要素应用对焦点企业数据要素应用的影响,本文在后续的实证研究部分对此进行稳健性检验。

(三) 企业数据要素应用的行业同群效应

数字经济时代,数据要素解构并重构企业的组织模式、营销模式、商业模式、思维模式、战略方向,企业逐步实现数据化管理、数据化经营、数据化商业,以适应数字经济时代不断变化的市场环境和用户需求。数据要素应用不仅需要数据要素的累积、数字基础设施的建设、数字人才的培养、数据分析能力和数据处理能力的培育,还需要综合分析、审视市场竞争环境和业务发展阶段。因此,数据要素应用是一个循序渐进且极具不确定性和高失败风险的系统性工程。正是数据要素应用的系统复杂性加剧了企业决策风险、运营风险、战略风险以及应用结果的不确定性。然而,数据要素应用主体并非单独理性进行经济决策,也非完全取决于企业个体特征^[25]。行业内市场主体间的社会关系和行为互动使得企业间的经济决策和行为是互相影

响和互相学习的,尤其在信息不对称、行业竞争程度高的情况下,企业更倾向于在社会网络中寻找一个高度关联、互动密切的参照物^[18],通过获得更多公共信息和私人信息,降低决策风险、运营风险、战略风险以及应用结果的不确定性,以此做出最优的数据要素应用决策。据此,本文提出以下假设。

假设 1:企业数据要素应用存在行业同群效应,即同行企业数据要素应用促进焦点企业数据要素应用。

(四) 企业数据要素应用行业同群效应的实现机制

企业数据要素应用行业同群效应的实现机制主要包括内部实现机制和外部实现机制。

1. 内部实现机制

一是企业代理成本。代理成本理论认为经理人出于自身利益的考虑,经常与股东利益发生“偏离”,从而提高了企业的代理成本。经理人在能力不足或信息不对称时,为保留“卸责”空间和规避决策失败的负面声誉影响^[27],不惜牺牲股东利益,倾向忽略自身拥有的基本信息,通过复刻行业内更具信息优势的同行企业的数据要素应用决策行为,对决策信息进行“降噪”,达成同行企业数据要素应用决策“共识”,发出决策合规化“信号”,塑造数据要素应用“同行者”甚至“先行者”形象。据此,本文提出以下假设。

假设 2a:代理成本越大,数据要素应用的行业同群效应越强。

二是企业内部数据要素循环累积。组织学习理论认为,内外部经验学习是企业获得决策信息的有效途径^[28]。一方面,企业研发、制造、营销、营运等内部数据要素的循环累积可以通过机器学习和深度学习,发挥其学习效应和数据循环累积效应,缓解企业信息不对称性;另一方面,企业可以根据观察同行企业的决策行为,增强企业外部信息的供给,缓解企业信息不对称性。企业为更大程度增强信息的透明性和对称性,可能将数据要素内部循环累积对同行企业外部模仿的替代作用变为互补作用,形成“内部数据+外部数据”的互补效应,获得内外部完备的信息,增强企业信息的透明性和对称性,进而促进行业内数据要素应用决策的趋同。据此,本文提出以下假设。

假设 2b:数据要素资源循环累积越多,越会加强数据要素应用的行业同群效应。

2. 外部实现机制

一是基于信息学习的模仿。利伯曼和阿萨巴(Lieberman & Asaba, 2006)认为信息学习是企业模仿的主要原因之一^[29]。信息学习理论强调,在信息不对称的情况下,企业更倾向于学习、模仿具有信息优势的企业。一般而言,领导企业位于行业网络节点密集的中心位置,并且规模越大、年限越长、杠杆率越高、营业收入越高的企业,与外部关联度越高,信息获取渠道越广,越能有效累积更多信息,企业的信息对称性越强,在做决策时越有信息优势。数据要素应用是一个系统的复杂过程,规模越小、年限越短、杠杆率越低、营业收入越低的追随企业,与外部关联度越低,企业信息获取渠道越狭小,能够累积的信息有限,企业的信息对称性越弱,在做决策时信息劣势越明显。信息劣势企业可能无法在决策时,进行精准建模分析,难以充分考虑所有备选的方案,造成决策预期与决策结果的偏差,进而加大数据要素决策的风险和成本,决策的“先行概率”缺乏可以借鉴的成功经验,成功概率低。此时,信息劣势的追随企业出于对领导企业信息优势的考虑,通过领导企业决策行为传递出的决策信息,在进行数据要素应用决策时更倾向于“搭便车”,模仿领导企业数据要素应用的决策行为。并且,随着领导企业发挥数据要素应用“领袖”作用,领导企业会预测追随企业的决策行为,实时动态跟踪追随企业的数据要素应用决策变化,对追随企业的决策行为进行预知性地模仿和学习,以保持持续的领导地位。因此,领导企业和追随企业之间数据要素应用决策互相模仿的双向互动,加强数据要素应用行业同群效应。据此,本文提出以下假设。

假设 3a:信息学习加强了企业数据要素应用行业同群效应,领导企业和追随企业互相模仿。

二是基于行业竞争的模仿。利伯曼和阿萨巴(2006)认为行业竞争是企业模仿的另一原因^[29]。基于行业竞争的模仿是指企业为维持竞争地位或压制竞争对手,模仿、学习同行企业的决策行为。在行业竞争程度高的情况下,尤其在企业连续接收到同行企业释放的数据要素应用的竞争信号时,企业开始竞相实施数据要素应用行动和战略,抢占数据要素战略性资源,增强企业数据要素分析能力和预测能力,以爆炸式增长的用户自生成内容洞察用户个性化需求,为其提供柔性化生产服务,实现供需双方的精准匹配,增加用户黏性,增强竞争优势或维持竞争地位,实现行业内数据要素应用的趋同。相对于异质化数据要素应用决策的高风险、高复杂性、高成本以及高不确定性,企业更倾向于与竞争对手采取“行业平均水平”的同质化数据要素应用决策,以此降低风险、复杂性、搜索成本以及不确定性^[29]。同质化数据要素应用决策说明,企业通过不使任意一方变好或变坏方式,避免破坏制约均衡^[30],与竞争对手实行“匹配性”战略,维持竞争地位,即以“匹配性”数据要素应用战略实现“帕累托最优”,维持竞争状态的平衡。据此,本文提出以下假设。

假设 3b:行业竞争加强了企业数据要素应用行业同群效应。

三、样本、数据与变量

(一) 样本选取与数据来源

本文以 2007—2021 年中国 A 股上市公司作为研究样本,依次剔除如下样本:(1)ST、*ST 和 PT 公司;(2)首次公开募股(IPO)当年的观测值和已退市的公司;(3)净资产为负的观测值;(4)主要变量缺失的公司。此外,为消除异常值,对所有连续变量进行上下 1%的缩尾处理。企业数据要素应用的数据主要来自上市公司年报。其他数据主要来自深圳希施玛数据科技有限公司 CSMAR 中国经济金融研究数据库。行业分类以 2012 年中国证券监督管理委员会修订的《上市公司分类指引》规定的二位行业代码为标准,企业所在地区以上市公司注册地所在的省份为标准。经过上述筛选后,最终得到 32 435 个有效观测值,使用软件 Stata 17 进行处理。

(二) 变量定义

1. 被解释变量

数据要素在经济发展中的价值越来越凸显,然而数据作为一种新型生产要素,其虚拟性特征使其难以“入表”,因而难以以传统的“表上”数据来度量企业数据要素应用水平。企业数据要素应用是一个复杂的系统性工程,要准确刻画微观企业数据要素应用水平面临重大挑战。参照张叶青等(2021)^[5]的研究,本文采用文本分析法测度企业层面的数据要素应用水平,即根据上市公司历年年报进行文本分析,抓取与数据要素应用相关的关键词,以词频作为数据要素应用水平的代理变量。具体步骤如下:首先,在桑德斯和坦布(Saunders & Tambe, 2013)^[31]以及张叶青等(2021)^[5]的基础上,对关键词字典进行补充和完善,最终归纳出 129 个与数据要素应用密切相关的关键词;其次,爬取 2007—2021 年所有上市公司年报,并将其转换为文本格式;最后,将上述关键词与年报文本进行匹配,统计其在每家公司年报中出现的频数,汇总后得到关键词总词频,将其对数作为本文被解释变量(*data*)。

2. 核心解释变量

同行企业数据要素应用(*average*)。将同行企业定义为除企业本身外所有二位行业代码相同的企业。同行企业数据要素应用水平是指同行企业滞后一期数据要素应用水平的平均值,采用滞后一期的原因

是:相对于企业自身的决策行为,同行企业的决策行为和战略只能在信息披露和决策执行后通过显性或隐性方式传递信息,即同行企业的信息传递具有滞后性^[32]。

3. 企业层面控制变量

为降低焦点企业遗漏变量的干扰,本文参照张叶青等(2021)^[5]、吴娜等(2022)^[32]、赵丽和胡植尧(2024)^[33]的研究,控制企业层面控制变量,主要包括:企业规模(*Size*)、资产负债率(*Lev*)、营业收入增长率(*Growth*)、独董比例(*Indep*)、两职合一(*Dual*)、第一大股东持股比例(*Top1*)、产权性质(*SOE*)、成立年限(*FirmAge*)、无形资产比例(*inasset*)。

4. 行业层面控制变量

为降低同行企业带来的遗漏变量的干扰,本文参照吴娜等(2022)^[32]的研究,控制行业层面控制变量。同行企业的决策和行为具有滞后性,因此行业层面控制变量均取滞后一期平均值,主要包括:同行企业规模(*Size_average*)、同行企业资产负债率(*Lev_average*)、同行企业营业收入增长率(*Growth_average*)、同行企业独董比例(*Indep_average*)、同行企业两职合一(*Dual_average*)、同行企业第一大股东持股比例(*Top1_average*)、同行企业产权性质(*SOE_average*)、同行企业成立年限(*FirmAge_average*)、同行企业无形资产比例(*inasset_average*)。

(三)描述性统计

表1报告了主要变量的描述性统计结果。在2007—2021年,共有32 435个公司-年度观测值。其中,数据要素应用(*data*)的样本均值为1.656 4,标准差为0.941 6,最小值为0,最大值为6.327 9,说明不同上市公司的数据要素应用程度具有较大差异,并有部分企业尚未应用数据要素。同行企业数据要素应用(*average*)的样本均值为1.564 4,标准差为0.623 9,最小值为0,最大值为4.317 5,说明同行企业数据要素应用程度不一。其他变量的描述性统计与现有文献基本吻合。

表1 描述性统计结果

变量类型	变量名称	变量符号	均值	标准差	最小值	最大值
被解释变量	数据要素应用	<i>data</i>	1.656 4	0.941 6	0.000 0	6.327 9
解释变量	同行企业数据要素应用	<i>average</i>	1.564 4	0.623 9	0.000 0	4.317 5
企业层面控制变量	企业规模	<i>Size</i>	22.181 7	1.286 2	19.405 8	26.429 7
	资产负债率	<i>Lev</i>	0.433 4	0.202 7	0.027 4	0.924 6
	营业收入增长率	<i>Growth</i>	0.176 2	0.415 2	-0.659 7	4.330 4
	独董比例	<i>Indep</i>	0.374 1	0.053 5	0.250 0	0.600 0
	两职合一	<i>Dual</i>	0.258 5	0.437 8	0	1
	第一大股东持股比例	<i>Top1</i>	0.346 8	0.149 1	0.081 3	0.758 4
	产权	<i>SOE</i>	0.392 6	0.488 3	0	1
	成立年限	<i>FirmAge</i>	2.852 4	0.355 6	0.693 1	3.610 9
	无形资产比例	<i>inasset</i>	0.047 9	0.063 5	0.000 0	0.938 3
	同行企业规模	<i>Size_average</i>	22.052 4	0.674 1	19.414 9	26.062 7
行业层面控制变量	同行企业资产负债率	<i>Lev_average</i>	0.421 6	0.102 1	0.052 6	0.896 2
	同行企业营业收入增长率	<i>Growth_average</i>	0.179 6	0.142 2	-0.545 4	4.309 5
	同行企业独董比例	<i>Indep_average</i>	0.372 6	0.013 2	0.250 0	0.555 6
	同行企业两职合一	<i>Dual_average</i>	0.262 0	0.130 3	0	1
	同行企业第一大股东持股比例	<i>Top1_average</i>	0.353 3	0.053 7	0.090 9	0.758 4
	同行企业产权	<i>SOE_average</i>	0.391 1	0.246 1	0	1
	同行企业成立年限	<i>FirmAge_average</i>	2.789 0	0.225 4	1.098 6	3.434 0
	同行企业无形资产比例	<i>inasset_average</i>	0.046 7	0.030 4	0.000 0	0.337 0

四、实证分析

(一) 基准回归

为了检验企业数据要素应用是否存在行业同群效应,本文参考吴娜等(2022)^[32]设定的模型,将同行企业的数据要素应用水平和行业控制变量等变量滞后一期。主要原因在于:同行企业数据要素应用信息需要在其披露后可得,具有一定的时滞性。具体模型如下:

$$data_{i,j,t} = \alpha_0 + \alpha_1 average_{-i,j,t-1} + X'_{i,j,t} \alpha_2 + Z'_{-i,j,t-1} \alpha_3 + \mu_i + \vartheta_t + \tau_j + \varepsilon_{i,j,t} \quad (1)$$

$data_{i,j,t}$ 表示 i 企业 j 行业 t 年的数据要素应用水平,是本文的被解释变量; $average_{-i,j,t-1}$ 表示除 i 企业本身 j 行业在 $t-1$ 年的数据要素应用水平均值,是本文的核心解释变量; $X'_{i,j,t}$ 表示 i 企业 j 行业在 t 年的企业层面控制变量向量; $Z'_{-i,j,t-1}$ 表示除 i 企业 j 行业在 $t-1$ 年的行业层面控制变量向量; μ_i 、 ϑ_t 、 τ_j 分别为企业、年份、行业固定效应; $\varepsilon_{i,j,t}$ 为随机扰动项。本文重点关注 α_1 的符号和显著性,以此检验数据要素应用行业同群效应的存在。

表2报告了企业数据要素应用行业同群效应的基准回归结果。列(1)是在控制了控制变量基础上考察同行企业对焦点企业数据要素应用的回归结果。列(2)是在列(1)基础上加入企业、行业以及年份固定效应的回归结果^①。从列(1)和列(2)回归结果可知,同行企业数据要素应用($average$)的回归系数分别为0.8687和0.4310,均在1%的置信水平下显著,表明同行企业数据要素应用对焦点企业数据要素应

表2 基准回归结果

变量	(1)	(2)
<i>average</i>	0.868 7 *** (0.022 8)	0.431 0 *** (0.031 1)
<i>Size</i>	0.043 0 *** (0.009 0)	0.133 4 *** (0.012 6)
<i>Lev</i>	-0.148 3 *** (0.049 7)	-0.114 0 ** (0.046 7)
<i>Growth</i>	0.043 3 *** (0.011 3)	0.011 9 (0.008 1)
<i>Indep</i>	0.321 7 (0.198 9)	-0.135 6 (0.116 3)
<i>Dual</i>	0.046 7 ** (0.019 6)	-0.008 3 (0.013 5)
<i>Top1</i>	-0.284 4 *** (0.067 4)	-0.331 7 *** (0.082 3)
<i>SOE</i>	-0.158 7 *** (0.024 4)	-0.032 3 (0.031 9)
<i>FirmAge</i>	-0.082 6 *** (0.030 0)	0.165 9 ** (0.078 3)
<i>inasset</i>	0.105 9 (0.123 3)	0.071 6 (0.119 7)
<i>Size_average</i>	-0.022 2 (0.021 9)	-0.081 4 *** (0.031 2)
<i>Lev_average</i>	-0.165 3 (0.121 4)	0.349 4 ** (0.145 5)
<i>Growth_average</i>	0.078 5 ** (0.034 9)	0.059 2 ** (0.026 8)
<i>Indep_average</i>	1.001 6 (0.665 7)	0.397 6 (0.517 4)
<i>Dual_average</i>	0.034 1 (0.094 9)	0.049 7 (0.077 8)
<i>Top1_average</i>	-0.257 2 (0.206 2)	0.070 1 (0.224 5)
<i>SOE_average</i>	0.108 9 * (0.062 3)	-0.041 6 (0.067 3)
<i>FirmAge_average</i>	0.140 1 ** (0.057 7)	-0.075 5 (0.089 1)
<i>inasset_average</i>	-0.424 6 (0.313 1)	-0.580 3 * (0.337 9)
常数项	-0.498 9 (0.379 8)	-0.525 6 (0.799 3)
企业固定效应	未控制	控制
行业固定效应	未控制	控制
年份固定效应	未控制	控制
观测值	32 435	32 049
$\overline{R^2}$	0.413 6	0.750 4

注:括号内为企业层面聚类的稳健标准误;***、**、* 分别表示1%、5%、10%置信水平。后表同。

① 样本期间,部分企业所属行业发生了变动,因此,在企业 and 年份固定效应的基础上,本文进一步控制了行业固定效应,保证回归系数始终为同一行业的数据要素应用水平对焦点企业数据要素应用的影响,避免行业变化带来的干扰。

用具有正向影响。这意味着,企业在数据要素应用决策时并非单独决策,更倾向于与行业内主体进行互动,模仿、学习同行业企业的数据要素应用决策行为,以此降低决策成本和决策风险,促进行业内焦点企业更大范围、更广程度地应用数据要素。因此,本文研究假设 1 得到实证结果的支持。考虑到研究结果的可靠性,下文的回归模型均加入了企业、行业、年份固定效应。

(二) 内生性分析

本文参考利里和罗伯茨(Leary & Roberts,2014)^[11]、孙锦萍和董志勇(2024)^[34]的研究,选择滞后一期同行企业股票特质收益率($\alpha_average$)作为同行企业数据要素应用的工具变量^①。工具变量必须满足相关性和排他性要求,即同行企业股票特质收益率只能通过同行企业数据要素应用,进而影响焦点企业数据要素应用。在相关性方面,股票特质收益率越高的企业越会为数据要素应用提供资金支持,加大对数据要素的投入^[18],进而发挥“数据要素×”作用,即同行企业股票特质收益率能够促进同行企业数据要素应用。在排他性方面,参照利里和罗伯茨(2014)^[11]的研究,利用资产定价模型,将股票特质收益率从股票收益率中分离,模型的残差是回报冲击,即认为股票特质收益率是一个外生冲击,得到的股票特质收益率不包含整个市场和行业等系统性因素;并且公司的各个时间序列不具有相关性,仅与公司自身特征相关^[11]。可以说,同行企业股票特质收益率与焦点企业数据要素应用不存在相关性,以同行企业股票特质收益率作为工具变量满足排他性要求,最终形成“同行企业股票特质收益率—同行企业数据要素应用—焦点企业数据要素应用”的作用路径。

表 3 报告了工具变量采用两阶段最小二乘(2SLS)法的回归结果。由第一阶段回归结果可知,工具变量($\alpha_average$)的回归系数显著为正,说明同行企业股票特质收益率促进了同行企业数据要素应用,与预期相符;由第二阶段回归结果可知,同行企业数据要素应用($average$)的回归系数在 1%的置信水平下显著为正,说明数据要素应用存在行业同群效应。此外,Kleibergen-Paap rk LM 统计量在 1%的置信水平下显著,证明了工具变量($\alpha_average$)不存在识别不足问题;Kleibergen-Paap Wald rk F 统计值为 1 267. 800 1,远大于 10,说明工具变量($\alpha_average$)不存在弱工具变量问题。

(三) 稳健性检验

为保证基准回归结果的可靠性,本文进行如下方式的稳健性检验。

1. 更换核心解释变量

本文参照吴娜等(2022)^[32]的做法,改

表 3 内生性分析回归结果

变量	第一阶段	第二阶段
$average$		0. 545 1 *** (0. 036 7)
$\alpha_average$	0. 800 0 *** (0. 022 5)	
常数项	-0. 525 0 (0. 367 7)	
企业层面控制变量	控制	控制
行业层面控制变量	控制	控制
企业固定效应	控制	控制
行业固定效应	控制	控制
年份固定效应	控制	控制
Kleibergen-Paap rk LM	657. 503 6 ***	
Kleibergen-Paap rk F	1 267. 800 1	
观测值	27 369	27 369
$\overline{R^2}$	0. 984 5	0. 043 1

① 企业股票特质收益率的具体计算方法可参照利里和罗伯茨(2014)^[11]以及吴娜等(2022)^[32]的研究。

变同行企业的定义,即将同行企业的定义由原来的二位行业代码相同的企业变为一位行业代码相同的企业,并生成新的同行企业数据要素应用水平变量(*average1*)。由表 4 可知,*average1* 的回归系数在 1%的置信水平下显著为正,说明在更换核心解释变量后,企业数据要素应用行业同群效应依旧存在。

2. 更换被解释变量

参考张叶青等(2021)^[5]的研究,在本文原有词典的基础上,缩小数据要素应用关键词词典范围,选取 6 个与数据要素应用相关的关键词构建关键词词典,生成数据要素应用指标(*data1*),替换本文的被解释变量数据要素应用(*data*)。由表 4 可知,*average* 的回归系数在 1%的置信水平下显著为正,说明在更换被解释变量后,企业数据要素应用行业同群效应依旧存在。

3. 控制省份同群效应影响

考虑到企业数据要素应用水平受所在相同省份、不同行业数据要素应用水平的影响,本文在基准回归基础上,加入滞后一期省份同群企业数据要素应用的平均值,生成省份同群效应变量(*province_average*),控制相同省份、不同行业数据要素应用对企业数据要素应用的影响。据此,为排除某省份内企业数据要素应用的互相影响,特此在模型(1)的基础上加入基于相同省份的同群企业数据要素应用滞后一期均值变量(*province_average*),控制省份同群效应的影响。由表 4 可知,*average* 的回归系数在 1%的置信水平下显著为正,表明在控制省份同群效应后,企业数据要素应用依旧存在行业同群效应。

4. 控制省份-行业同群效应影响

考虑到企业受同省份、同行业企业的“制度等价”效应影响,企业可能会增强模仿和学习同省份、同行业企业数据要素应用的动机^[35]。据此,为排除相同省份、相同行业内企业数据要素应用的互相影响,在模型(1)基础上加入基于相同省份、相同行业的同群企业数据要素应用滞后一期均值变量(*indprovince_average*),以此控制“省份-行业”企业数据要素应用对焦点企业数据要素应用的影响。由表 4 可知,*average* 的回归系数显著为正,表明在控制“省份-行业”同群效应后,企业数据要素应用依旧存在行业同群效应。

表 4 稳健性检验回归结果

变量	更换核心解释变量	更换被解释变量	省份同群效应	省份-行业同群效应
<i>average</i>		0.657 7*** (0.039 1)	0.416 7*** (0.031 3)	0.406 0*** (0.033 7)
<i>average1</i>	0.560 5*** (0.042 6)			
<i>province_average</i>			0.218 5*** (0.041 0)	
<i>indprovince_average</i>				0.042 8*** (0.015 7)
常数项	-0.264 2 (0.784 7)	-3.431 1*** (0.871 8)	-0.762 7 (0.797 4)	-0.273 1 (0.846 5)
企业层面控制变量	控制	控制	控制	控制

表4(续)

变量	更换核心解释变量	更换被解释变量	省份同群效应	省份-行业同群效应
行业层面控制变量	控制	控制	控制	控制
企业固定效应	控制	控制	控制	控制
行业固定效应	控制	控制	控制	控制
年份固定效应	控制	控制	控制	控制
观测值	32 049	32 047	32 049	29 545
$\overline{R^2}$	0.749 9	0.713 0	0.751 0	0.759 9

五、实现机制分析

(一) 内部实现机制分析

1. 企业代理成本

在不同代理成本水平下,数据要素应用的行业同群效应不同。本文借鉴陈冬华等(2005)^[36]、夏子航(2021)^[37]的做法,以企业管理费用作为企业代理成本(*Agent_cost*)的代理变量,研究企业代理成本对数据要素应用行业同群效应的影响。据此,本文构建模型(2)进行实证检验。回归结果如表5列(1)和列(2)所示。其中,列(2)是在列(1)基础上,控制行业层面控制变量的结果,交互项(*average*×*Agent_cost*)的回归系数均显著为正,即在企业数据要素应用决策时,企业管理者“理性”地为避免声誉损失、增大卸责空间,增强了管理者与股东意愿的“背离”,增加了企业代理成本,进而增强了企业数据要素应用决策对同行企业的模仿和学习,验证了研究假设2a。

$$data_{i,j,t} = \alpha_0 + \alpha_1 average_{-i,j,t-1} + \alpha_2 average_{-i,j,t-1} \times Agent_cost_{-i,j,t-1} + \alpha_3 Agent_cost_{-i,j,t-1} + X'_{i,j,t} \alpha_4 + Z'_{-i,j,t-1} \alpha_5 + \mu_i + \vartheta_t + \tau_j + \varepsilon_{i,j,t} \tag{2}$$

2. 企业内部数据要素循环累积

相对于外部模仿学习,企业内部数据要素累积是支持企业决策的重要方式。随着企业内部数据的指数型增长,在算法的支撑下,企业通过机器学习、深度学习、边缘学习,发挥数据要素循环累积的学习效应,增强信息对称性和协同性,深度赋能企业决策行为。企业内部的数据要素学习效应和外部的行业同群效应是内外部学习的有效途径。本文借鉴万良勇等(2016)^[25]的做法,利用企业过去三年是否连续应用数据要素(当近三年连续使用数据要素时,取值1,否则为0),构造企业内部数据要素循环累积变量(*Innerlearn*),研究企业内部数据要素循环累积对数据要素应用行业同群效应的影响。据此,本文构建模型(3)进行实证检验。回归结果见表5列(3)和列(4),列(4)是在列(3)基础上控制行业层面控制变量的结果,交互项(*average*×*Innerlearn*)的回归系数均显著为正,即企业内部数据要素循环累积越强,数据要素应用的行业同群效应越强,企业内部数据要素循环累积并未“挤出”对同行企业数据要素应用的模仿、学习。对此,可能的原因在于:企业出于内外部信息互补性考虑,在内部数据要素循环累积的基础上,模仿学习外部决策行为,增强企业信息的完整性和全面性,进而增强数据要素应用决策的科学性,验证了研究假设2b。

$$data_{i,j,t} = \alpha_0 + \alpha_1 average_{-i,j,t-1} + \alpha_2 average_{-i,j,t-1} \times Innerlearn_{-i,j,t-1} + \alpha_3 Innerlearn_{-i,j,t-1} + X'_{i,j,t} \alpha_4 + Z'_{-i,j,t-1} \alpha_5 + \mu_i + \vartheta_t + \tau_j + \varepsilon_{i,j,t} \tag{3}$$

表 5 代理成本、数据要素循环累积与数据要素应用行业同群效应回归结果

变量	代理成本		数据要素循环累积	
	(1)	(2)	(3)	(4)
<i>average</i>	0.374 8*** (0.030 4)	0.359 1*** (0.030 1)	0.323 5*** (0.040 3)	0.314 4*** (0.040 5)
<i>average</i> × <i>Agent_cost</i>	0.104 8*** (0.021 7)	0.105 3*** (0.021 6)		
<i>Agent_cost</i>	-0.131 1*** (0.035 8)	-0.132 7*** (0.035 7)		
<i>average</i> × <i>Innerlearn</i>			0.102 9*** (0.032 5)	0.099 3*** (0.032 7)
<i>Innerlearn</i>			-0.045 8 (0.041 6)	-0.041 5 (0.041 8)
常数项	-1.952 2*** (0.355 3)	-0.127 5 (0.800 2)	-1.236 3*** (0.419 5)	0.706 3 (0.857 5)
企业层面控制变量	控制	控制	控制	控制
行业层面控制变量	未控制	控制	未控制	控制
企业固定效应	控制	控制	控制	控制
行业固定效应	控制	控制	控制	控制
年份固定效应	控制	控制	控制	控制
观测值	32 049	32 049	27 162	27 162
$\overline{R^2}$	0.750 6	0.751 0	0.762 0	0.762 5

(二) 外部实现机制分析

1. 基于信息理论

借鉴陆蓉等(2017)^[13]、李秋梅和梁权熙(2020)^[38]、吴娜等(2022)^[32]的做法,分别根据公司年龄、公司规模、资产负债率、营业收入将企业划分为行业领导者和行业追随者,检验行业领导者和行业追随者的互相影响。具体步骤如下:一是将上市公司在行业内的公司年龄(*ListAge*)、公司规模(*Size*)、资产负债率(*Lev*)、营业收入(*Income*)从大到小排序,排名前30%和后30%的企业分别定义为行业领导者和行业追随者;二是重新计算滞后一期除企业本身外同行领导者(追随者)数据要素应用平均值,进而生成新的领导者(追随者)同行数据要素应用变量;三是在研究行业追随者(领导者)数据要素应用对领导者(追随者)数据要素应用反应时,利用行业追随者(领导者)样本数据,将行业领导者(追随者)作为同行企业数据要素应用水平,与行业追随者(领导者)数据样本进行匹配。据此,再次对模型(1)进行实证检验。

表6报告了行业追随者数据要素应用决策对行业领导者数据要素应用决策的反应结果,同行领导者数据要素应用(*average*)的回归系数均显著为正,说明行业追随者在数据要素应用决策时,更倾向于模仿企业年龄大、规模大、杠杆率高、营业收入高的行业领导者,行业领导者与外部主体连接更多、更频繁,具有更多显性和隐性信息,行业追随者在对行业领导者的数据要素应用决策的学习和模仿中能获得更多信息,增强信息优势,降低行业追随企业数据要素应用决策的成本、风险以及不确定性。

表6 行业追随者对行业领导者的反应结果

变量	企业年龄	企业规模	资产负债率	营业收入
<i>average</i>	0.176 2 ^{***}	0.225 5 ^{***}	0.319 5 ^{***}	0.281 0 ^{***}
	(0.050 1)	(0.038 7)	(0.045 0)	(0.031 8)
常数项	0.978 9	-2.661 8 ^{**}	-2.481 7 [*]	-3.852 3 ^{***}
	(1.594 7)	(1.204 4)	(1.383 9)	(1.298 6)
企业层面控制变量	控制	控制	控制	控制
行业层面控制变量	控制	控制	控制	控制
企业固定效应	控制	控制	控制	控制
行业固定效应	控制	控制	控制	控制
年份固定效应	控制	控制	控制	控制
观测值	9 299	9 454	9 342	9 645
$\overline{R^2}$	0.779 6	0.769 6	0.778 1	0.771 2

表7报告了行业领导者数据要素应用决策对行业追随者数据要素应用决策的反应结果,同行追随者数据要素应用(*average*)的回归系数均显著为正,说明行业领导者在数据要素应用决策时,基于“反馈动机”,针对行业追随者决策的模仿和学习行为进行自身的数据要素应用预期性或适应性调整,以此保持信息优势,验证了研究假设3a。综上所述,行业领导者和行业追随者在数据要素应用决策时互相影响。

表7 行业领导者对行业追随者的反应结果

变量	企业年龄	企业规模	资产负债率	营业收入
<i>average</i>	0.180 5 ^{***}	0.206 3 ^{***}	0.300 7 ^{***}	0.225 9 ^{***}
	(0.037 7)	(0.048 1)	(0.041 2)	(0.050 5)
常数项	-0.813 3	4.061 7 ^{**}	2.070 6	3.322 9 [*]
	(2.040 4)	(1.738 4)	(1.558 2)	(1.782 2)
企业层面控制变量	控制	控制	控制	控制
行业层面控制变量	控制	控制	控制	控制

表7(续)

变量	企业年龄	企业规模	资产负债率	营业收入
企业固定效应	控制	控制	控制	控制
行业固定效应	控制	控制	控制	控制
年份固定效应	控制	控制	控制	控制
观测值	9 015	8 962	8 803	8 970
$\overline{R^2}$	0. 747 6	0. 773 5	0. 766 5	0. 762 3

2. 基于行业竞争

维持竞争优势是企业模仿、学习同行企业决策的主要原因。行业竞争程度不同,对企业数据要素应用行业同群效应影响不同。本文采用两种方法度量行业竞争:一是利用行业赫芬达尔指数(*HHI*)表示行业集中度,其值越大,行业竞争越小。据此研究行业集中度对数据要素应用行业同群效应的影响。本文构建模型(4)进行实证检验。回归结果见表 8 列(1)和列(2),列(2)是在列(1)基础上控制行业层面控制变量的结果,交互项(*average*×*HHI*)的回归系数均在 1%置信水平上显著为负,即行业集中度越高,数据要素应用行业同群效应越弱。

$$data_{i,j,t}=\alpha_0+\alpha_1average_{-i,j,t-1}+\alpha_2average_{-i,j,t-1}\times HHI_{-i,j,t-1}+\alpha_3HHI_{-i,j,t-1}+X'_{i,j,t}\alpha_4+Z'_{-i,j,t-1}\alpha_5+\mu_i+\vartheta_t+\tau_j+\varepsilon_{i,j,t}$$

(4)

二是借鉴陈志斌和王诗雨(2015)^[39]的做法,将行业内企业数量作为行业竞争程度(*Num*)的代理变量,其值越大,行业竞争程度越大。据此,研究行业竞争程度对数据要素应用行业同群效应的影响。据此,本文构建模型(5)进行实证检验。回归结果见表 8 列(3)和列(4),列(4)是在列(3)基础上控制行业层面控制变量的结果,交互项(*average*×*Num*)的回归系数均在 1%置信水平上显著为正,即行业竞争程度越高,数据要素应用行业同群效应越强,说明基于竞争理论,企业为获得竞争优势或者保持竞争地位的动机越强,越可能模仿、学习同行企业数据要素应用决策行为,验证研究假设 3a。

$$data_{i,j,t}=\alpha_0+\alpha_1average_{-i,j,t-1}+\alpha_2average_{-i,j,t-1}\times Num_{-i,j,t-1}+\alpha_3Num_{-i,j,t-1}+X'_{i,j,t}\alpha_4+X'_{-i,j,t-1}\alpha_5+\mu_i+\vartheta_t+\tau_j+\varepsilon_{i,j,t}$$

(5)

表 8 基于行业竞争的回归结果

变量	行业集中度		行业竞争程度	
	(1)	(2)	(3)	(4)
<i>average</i>	0. 496 6 *** (0. 035 0)	0. 480 2 *** (0. 034 4)	0. 357 3 *** (0. 030 3)	0. 339 1 *** (0. 030 6)
<i>average</i> × <i>HHI</i>	-0. 351 0 *** (0. 106 8)	-0. 331 3 *** (0. 105 5)		
<i>HHI</i>	0. 558 6 *** (0. 169 8)	0. 543 6 *** (0. 168 2)		

表8(续)

变量	行业集中度		行业竞争程度	
	(1)	(2)	(3)	(4)
<i>average</i> × <i>Num</i>			0.000 8 *** (0.000 2)	0.000 9 *** (0.000 2)
<i>Num</i>			-0.001 3 ** (0.000 5)	-0.001 7 *** (0.000 5)
常数项	-2.321 6 *** (0.345 7)	-0.673 9 (0.792 8)	-2.042 7 *** (0.346 6)	-0.648 8 (0.805 2)
企业层面控制变量	控制	控制	控制	控制
行业层面控制变量	未控制	控制	未控制	控制
企业固定效应	控制	控制	控制	控制
行业固定效应	控制	控制	控制	控制
年份固定效应	控制	控制	控制	控制
观测值	32 049	32 049	32 049	32 049
$\overline{R^2}$	0.750 5	0.750 9	0.750 9	0.751 2

六、异质性分析

企业属性差异带来同行企业数据要素应用对焦点企业数据要素应用影响的差异。陈仕华和卢昌崇(2013)认为企业出于经济政策环境和用户群体的相似性,更倾向模仿和学习企业属性相似的企业,并对此类企业的模仿、学习程度更强,即企业在模仿、学习时遵循“先内后外”规律^[40]。本文从企业产权性质、企业规模视角探讨同行企业数据要素应用对焦点企业数据要素应用的模仿规律。

(一) 基于企业产权性质差异的异质性分析

根据企业产权性质差异,将样本企业分为国有企业和非国有企业样本组,为了检验产权性质对数据要素应用行业同群效应的异质性影响,参考万良勇等(2016)^[25]的做法,在基本回归基础上构建了两个新变量:同行国有企业数据要素应用(*state_average*)和同行非国有企业数据要素应用(*private_average*),前者是滞后一期同行国有企业数据要素应用平均值(除企业本身外),后者是滞后一期同行非国有企业的数据要素应用平均值(除企业本身外)。由表9可知,同行国有企业(*state_average*)和非国有企业数据要素应用(*private_average*)的回归系数分别为0.327 5和0.098 3,且在1%的置信水平下显著,说明国有企业在数据要素应用决策时既模仿国有企业决策,也模仿非国有企业决策,但更大程度受同行国有企业影响(0.327 5>0.098 3),说明国有企业更倾向于模仿、学习相同产权性质的同行企业,即遵循“先内后外”规律^[25]。由表9可知,同行国有企业(*state_average*)和非国有企业数据要素应用(*private_average*)的回归系数分别为0.154 9和0.247 7,且均在1%的置信水平下显著,说明非国有企业在数据要素应用决策时尽管受同行国有企业和非国有企业数据要素应用的双重影响,但受同行非国有企业数据要素应用的影响程度更大(0.247 7>0.154 9),说明非国有企业

更倾向模仿、学习与其地位类似的同行非国有企业。

(二) 基于企业规模差异的异质性分析

根据企业规模的中位数,将样本企业分为大规模企业和小规模企业样本组。为了检验企业规模对数据要素应用行业同群效应的异质性影响。参考万良勇等(2016)^[25]的做法,在基本回归的基础上生成两个新变量:同行大规模企业数据要素应用(*big_average*)和同行小规模企业数据要素应用(*small_average*),前者是滞后一期同行大规模企业数据要素应用平均值(除企业本身外),后者是滞后一期同行小规模企业数据要素应用平均值(除企业本身)。表9结果表明,大规模企业在模仿、学习数据要素应用决策时更倾向“对标”大规模企业(大规模企业和小规模企业回归系数分别为0.238 5和0.078 3,且0.238 5>0.078 3),小规模企业在模仿、学习数据要素应用决策时更倾向于与小规模企业达成决策“共识”(大规模企业和小规模企业的回归系数分别为0.151 7和0.172 2,且0.172 2>0.151 7),再次证明数据要素应用的行业同群效应遵循“先内后外”规律。

表 9 异质性分析回归结果

变量	产权性质		企业规模	
	国有企业	非国有企业	大规模企业	小规模企业
<i>state_average</i>	0.327 5*** (0.041 1)	0.154 9*** (0.025 6)		
<i>private_average</i>	0.098 3*** (0.032 8)	0.247 7*** (0.033 9)		
<i>small_average</i>			0.078 3*** (0.027 2)	0.172 2*** (0.032 0)
<i>big_average</i>			0.238 5*** (0.038 4)	0.151 7*** (0.026 8)
常数项	1.196 9 (1.282 0)	-2.015 1* (1.108 3)	2.799 1** (1.144 6)	-3.854 3*** (1.242 4)
企业层面控制变量	控制	控制	控制	控制
行业层面控制变量	控制	控制	控制	控制
企业固定效应	控制	控制	控制	控制
行业固定效应	控制	控制	控制	控制
年份固定效应	控制	控制	控制	控制
观测值	12 133	18 774	15 590	15 307
$\overline{R^2}$	0.687 4	0.780 6	0.768 8	0.791 1

七、结论与启示

数字经济时代,数据要素成为重要的战略性资源,企业竞相应用数据要素,以此激活数据活力,释放数据价值。本文主要利用中国A股上市公司披露的年报,抓取与数据要素应用相关的关键词,构造企业层面的数据要素应用指标,实证检验同行企业数据要素应用对焦点企业数据要素应用的影响。研究结果表明,同行企业数据要素应用对焦点企业数据要素应用具有正向影响,即存在数据要素应用的行业同群效应。在经过替换核心解释变量、被解释变量,控制省份同群效应、控制省份-行业同群效应、内生性检验后,研究结果依旧稳健。企业代理成本越高、内部数据要素循环累积越强,企业数据要素应用的行业同群效应越强;信息学习越强、行业竞争越激烈,企业数据要素应用行业的同群效应越强。此外,企业在数据要素应用决策时更倾向模仿与产权性质、企业规模相同或类似的同行企业,即遵循“先内后外”规律。

根据上述研究结论,本文提出以下研究启示:

第一,鼓励发挥行业同群效应,加强企业数据要素应用。数据要素应用并非单独行为,而是深受同行企业决策影响的系统性工程。政府应牵头推动构建行业数据共享平台,鼓励同行企业间的经验交流和技术合作。特别是通过线上线下相结合的方式,通过案例分享、技术讨论、培训研讨等活动,帮助焦点企业快速获取同行经验,降低数据应用的不确定性和风险。同时,行业协会应推动制定行业标准,规范数据要素应用的流程和模式,降低企业进入数据要素领域的门槛,释放更多数据价值。

第二,优化企业内外部环境,增强数据要素应用的机制作用。从内部而言,支持企业通过政策补贴和专项资金,增强内部数据要素循环能力,促进数据要素积累和分析能力提升。同时,为高代理成本的企业提供管理咨询和外部监督服务,降低决策层与实际执行层之间的信息不对称,减少内部摩擦,从而提高数据要素应用的效率。从外部而言,推动信息学习渠道建设,增强企业间的技术流动性,降低信息获取的成本。针对行业竞争激烈的领域,鼓励企业以数据要素应用为核心竞争力进行技术迭代。

第三,因企制宜,强化重点企业有针对性的模仿能力。国有企业应优先模仿产权和治理结构类似的同行企业,非国有企业则更注重向同类企业学习,以增强决策的现实性和可操作性。大规模企业可以通过对标行业龙头企业的先进经验,优化其资源配置效率;而小规模企业应更注重与同类企业的合作和模仿,提升灵活性和适应能力。

参考文献:

- [1]李海舰,赵丽.数据成为生产要素:特征、机制与价值形态演进[J].上海经济研究,2021(8):48-59.
- [2]王谦,付晓东.数据要素赋能经济增长机制探究[J].上海经济研究,2021(4):55-66.
- [3]谢康,夏正豪,肖静华.大数据成为现实生产要素的企业实现机制:产品创新视角[J].中国工业经济,2020(5):42-60.
- [4]陶长琪,丁煜.数据要素何以成为创新红利?——源于人力资本匹配的证据[J].中国软科学,2022(5):45-56.
- [5]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.
- [6]杨向阳,徐从才.数据要素与流通企业高质量发展[J].商业经济与管理,2024(3):5-17.
- [7]陈国青,张瑾,王聪,等.“大数据—小数据”问题:以小见大的洞察[J].管理世界,2021,37(2):203-213.

- [8]孙新波,钱雨,张明超,等.大数据驱动企业供应链敏捷性的实现机理研究[J].管理世界,2019,35(9):133-151.
- [9]蔡跃洲,马文君.数据要素对高质量发展影响与数据流动制约[J].数量经济技术经济研究,2021,38(3):64-83.
- [10]陈国青,曾大军,卫强,等.大数据环境下的决策范式转变与使能创新[J].管理世界,2020,36(2):95-105.
- [11]LEARY M T, ROBERTS M R. Do peer firms affect corporate financial policy? [J]. The Journal of Finance, 2014, 69(1): 139-178.
- [12]钟田丽,张天宇.我国企业资本结构决策行为的“同伴效应”——来自深沪两市A股上市公司面板数据的实证检验[J].南开管理评论,2017,20(2):58-70.
- [13]陆蓉,王策,邓鸣茂.我国上市公司资本结构“同群效应”研究[J].经济管理,2017,39(1):181-194.
- [14]PARSONS C A, SULAEMAN J, TITMAN S. The geography of financial misconduct[J]. The Journal of Finance, 2018, 73(5): 2087-2137.
- [15]李佳宁,钟田丽.企业投资同伴效应的研究述评与展望[J].管理学报,2023,20(2):308-316.
- [16]吴娜,魏智佳,白雅馨.经济周期下企业金融资产配置的同群效应研究[J].云南财经大学学报,2023,39(2):45-64.
- [17]李倩,王诗豪,邓沛东,等.企业数字化转型的同群效应[J].科技进步与对策,2023,40(17):1-12.
- [18]陈庆江,王彦萌,王茂丰.企业数字化转型的同群效应及其影响因素研究[J].管理学报,2021,18(5):653-663.
- [19]徐翔,厉克奥博,田晓轩.数据生产要素研究进展[J].经济学动态,2021(4):142-158.
- [20]白永秀,李嘉雯,王泽润.数据要素:特征、作用机理与高质量发展[J].电子政务,2022(6):23-36.
- [21]FARBOODI M, VELDKAMP L. A model of the data economy[Z]. NBER Working Paper No. 28427, 2021.
- [22]JONES C I, TONETTI C. Nonrivalry and the economics of data[J]. American Economic Review, 2020, 110(9): 2819-2858.
- [23]王胜利,樊悦.论数据生产要素对经济增长的贡献[J].上海经济研究,2020(7):32-39.
- [24]宋广蕊,马春爱,肖榕.研发投入同群效应促进了企业创新“增量提质”吗? [J]. 外国经济与管理,2023,45(4):137-152.
- [25]万良勇,梁婵娟,饶静.上市公司并购决策的行业同群效应研究[J].南开管理评论,2016,19(3):40-50.
- [26]孙锦萍,李亚飞,侯麟科.上市公司创新决策的区域同群效应研究[J].经济与管理研究,2022,43(7):115-131.
- [27]SCHARFSTEIN D S, STEIN J C. Herd behavior and investment[J]. American Economic Review, 1990, 80(3): 465-479.
- [28]HUBER G P. Organizational learning: the contributing processes and the literatures[J]. Organization Science, 1991, 2(1): 88-115.
- [29]LIEBERMAN M B, ASABA S. Why do firms imitate each other? [J]. Academy of Management Review, 2006, 31(2): 366-385.
- [30]CHEN M J, SMITH K G, GRIMM C M. Action characteristics as predictors of competitive responses[J]. Management Science, 1992, 38(3): 439-455.
- [31]SAUNDERS A, TAMBE P. A measure of firms' information practices based on textual analysis of 10-K filings[Z]. University of Pennsylvania Working Paper, 2013.
- [32]吴娜,白雅馨,安毅.主动模仿还是被动反应:商业信用同群效应研究[J].南开管理评论,2022,25(3):149-161.
- [33]赵丽,胡植尧.数据要素、动态能力与企业全要素生产率——破解“数据生产率悖论”之谜[J].经济管理,2024,46(7):55-72.
- [34]孙锦萍,董志勇.大企业引领了创新发展吗? ——基于中国上市公司示范效应的研究[J].经济与管理研究,2024,45(11):104-125.
- [35]马骏,李书娴,李江雁.被动模仿还是主动变革? ——上市公司互联网涉入的同群效应研究[J].经济评论,2021(5):86-101.
- [36]陈冬华,陈信元,万华林.国有企业中的薪酬管制与在职消费[J].经济研究,2005(2):92-101.
- [37]夏子航.企业金融化同群效应:“取长补短”抑或“盲目跟风”? [J]. 中南财经政法大学学报,2021(4):74-88.
- [38]李秋梅,梁权熙.企业“脱实向虚”如何传染? ——基于同群效应的视角[J].财经研究,2020,46(8):140-155.
- [39]陈志斌,王诗雨.产品市场竞争对企业现金流风险影响研究——基于行业竞争程度和企业竞争地位的双重考量[J].中国工业经济,2015(3):96-108.
- [40]陈仕华,卢昌崇.企业间高管联结与并购溢价决策——基于组织间模仿理论的实证研究[J].管理世界,2013(5):144-156.

Peer Effect of Enterprise Data Factor Applications and Its Realization Mechanism

ZHAO Li¹, HU Zhiyao²

(1. Ningbo University, Ningbo 315211;

2. University of Chinese Academy of Social Sciences, Beijing 102488)

Abstract: In the era of the digital economy, data factors have emerged as a vital strategic resource. Enterprises actively compete to leverage data factors to unlock data value. Previous studies have predominantly emphasized the economic value of data factors, with limited exploration of the influencing factors that drive data factor applications, particularly from the perspectives of social relations and behavioral interactions. Therefore, this paper conducts an empirical analysis using data from Chinese listed companies from 2007 to 2021, investigating the industry peer effect of enterprise data factor applications and the underlying mechanisms.

The main findings indicate that enterprise data factor applications show an industry peer effect. Specifically, data factor applications by peer enterprises positively influence focal enterprises' data practices. This result remains valid after a series of endogeneity and robustness tests. Internal mechanisms reveal that the peer effect is more pronounced in enterprises with higher agency costs and greater accumulation of internal data resources. External mechanisms indicate that stronger information learning and more intense industry competition amplify the peer effect. Further analysis confirms that the peer effect of data factor applications follows an internal-to-external pattern: enterprises are more inclined to imitate peer enterprises with similar ownership structures and comparable sizes.

The potential contributions are threefold. First, this paper provides empirical evidence on how industry-level data factor applications influence the behaviors of focal enterprises, encouraging them to consider the strategic decisions of their peers, thereby forming a more systematic decision-making process that minimizes costs and risks. Second, it examines the internal and external mechanisms of the peer effect of enterprise data factor applications, offering theoretical and practical insights for firms to refine their data element application strategies. Third, the investigation of the peer effect of data factor applications expands the scope of research on data factor applications and broadens the theoretical horizon of peer effect research.

Based on these findings, this paper proposes the following policy implications. First, the government should take the lead in establishing industry-level data-sharing platforms. At the same time, industry associations should promote the development of standards that regulate processes and models for data factor applications, thereby lowering the entry threshold for enterprises. Second, optimizing enterprises' internal and external environments is essential to reinforce the mechanisms of data factor applications. Third, enterprises should adopt tailored strategies to enhance targeted imitation. State-owned and non-state-owned enterprises, as well as large and small enterprises, should place greater emphasis on learning from and imitating peer enterprises with similar characteristics to enhance the practicality and implementability of their strategic decisions.

Keywords: data factor; data factor application; industry peer effect; realization mechanism; internal and external influence

责任编辑:李 叶;魏小奋