

数据资源信息披露赋能民营企业投资效率提升： 来自大语言模型的文本分析证据

罗进辉, 沈文轩, 宋欣悦

(厦门大学 管理学院, 福建 厦门 361005)

[摘要]在数字经济时代,数据是加快培育和发展新质生产力的新型生产要素。基于2010—2024年中国A股民营上市公司样本,利用FinBERT2大语言模型构建企业数据资源信息披露指标,实证检验数据资源信息披露对民营企业投资效率的影响。研究发现,数据资源信息披露能够显著降低民营企业的非效率投资水平,即显著提升其投资效率,且结论经多项稳健性检验后依然成立。机制分析表明,数据资源信息披露主要通过缓解信息不对称问题和抑制代理问题两条路径提升民营企业投资效率。异质性分析发现,数据资源信息披露对民营企业投资效率的积极影响在高科技、小规模、低资产密集型、高数字投资水平、市场化程度较高地区的企业中表现得更强。进一步分析表明,数据资源信息披露通过抑制非效率投资水平促进了民营企业的全要素生产率提升。研究不仅丰富了数据资源及其信息披露的经济后果和民营企业投资效率影响因素的研究文献,而且对于国家促进民营经济高质量发展的宏观实践具有重要的政策启示。

[关键词]数据资源;信息披露;投资效率;民营企业;大语言模型;高质量发展

[中图分类号]F239 **[文献标志码]**A **[文章编号]**2096-3114(2026)05-0060-12

一、引言

随着数字技术的快速发展与数据要素市场化配置改革的持续推进,数据已深度嵌入企业生产运营、技术创新和战略决策全过程,成为与土地、劳动力、资本和技术并重的生产要素。为规范数据资源的会计处理,财政部于2023年发布《企业数据资源相关会计处理暂行规定》,并自2024年起正式实施,标志着企业的数据资源开始进入入表实践阶段。与此同时,2026年全国两会通过的“十五五”规划明确要求深化数据资源开发,深入开展“数据要素×”行动^①。2026年,政府工作报告进一步提出“深化数据资源开发利用,健全数据要素基础制度,建设高质量数据集”,凸显了数据资源在经济发展中的战略地位^②。

在此背景下,数据资源的经济价值日益受到学术界关注。已有研究表明,数据资源及其信息披露能够有效促进企业的决策效率、创新能力、人力资本优化^[1-5]。然而,作为数智经济时代的一种战略资源,数据资源并不是对所有企业都具有同等重要的经济价值。特别地,相较于国有企业,民营企业长期面临更加突出的资源约束和代理问题,这些问题在很大程度上制约着我国民营经济的高质量发展。一方面,民营企业缺乏政府隐性担保和信用背书,在资本市场中往往面临更高的信息甄别成本和更低的市场关注度,外部利益相关者难以及时准确地获取企业经营状况、资源禀赋、发展前景等信息,更容易受到信息不对称的影响^[6]。另一方面,民营企业多由家族或自然人创立和控制,股权结构相对集中,公司治理规

[收稿日期]2026-05-12

[基金项目]国家社会科学基金重点项目(25AGL009)

[作者简介]罗进辉(1983—),男,福建连城人,厦门大学管理学院教授,博士生导师,博士,主要研究方向为资本市场财务会计与公司治理;沈文轩(2001—),女,山东济南人,厦门大学管理学院硕士生,主要研究方向为资本市场财务会计与公司治理;宋欣悦(1997—),女,江西赣州人,厦门大学管理学院博士生,主要研究方向为资本市场财务会计与公司治理,通信作者,邮箱:songxinyue732@163.com。

①详见:https://www.gov.cn/yaowen/liebiao/202603/content_7062633.htm。

②详见:https://www.gov.cn/yaowen/liebiao/202603/content_7062625.htm。

范性和信息披露质量总体弱于国有企业^[7],代理问题更为突出。

数据资源的入表或披露将在一定程度上增厚民营企业本来就相对薄弱的资产实力,同时也有助于更加全面地反映民营企业的数据资源价值和数字化能力,改善企业信息环境和治理环境,从而使民营企业获得更多更好的外部资源支持。这意味着,对于信息透明度相对较低、治理基础相对薄弱的民营企业而言,数据资源信息披露所带来的信息增量和治理增量更大。因此,本文认为相对于国有企业,民营企业对数据资源信息披露的经济价值将表现得更为敏感而有效。但是,遗憾的是目前鲜有文献专门探讨这一问题。进一步,民营企业普遍面临的资源约束问题和代理问题会直接影响民营企业的投资决策,是民营企业投资效率的重要决定因素。因此,数据资源信息披露是否会影响民营企业的投资效率,这是一个很值得探讨的重要研究课题。

鉴于此,本文以2010—2024年中国A股民营上市公司为研究对象,借助FinBERT2大语言模型科学构建了上市公司的数据资源信息披露指标,进而实证检验了数据资源信息披露对民营企业投资效率的潜在影响。相较于以往研究,本文的研究贡献主要体现在以下三个方面:首先,本文从投资效率视角拓展了数据资源信息披露经济后果研究,为理解数据资源信息披露如何影响企业资源配置提供了新的经验证据。其次,本文丰富了企业投资效率影响因素的研究。在经济高质量发展要求下,如何提高企业的投资效率是社会各界共同关注的重要问题,既有研究主要从公司治理结构^[8]、财务报告质量^[9]、非财务信息披露^[10]、外部制度环境^[11-12]等多个维度进行了广泛探讨。本文则从数据资源信息披露这一特殊的非财务文本信息切入,丰富了企业投资效率影响因素的研究文献。最后,本文在研究方法上也具有一定的创新。既有文献多采用文本词典法对数据资源信息披露进行度量^[2,4,13],本文借鉴罗进辉等的做法^[14],通过引入FinBERT2大语言模型来对企业年报中的数据资源表述进行识别与量化,该方法结合上下文语境较为准确地识别企业是否真实拥有披露的相关数据资源及其使用与经营情况,从而有效降低传统词频法的测度误差,提升相关指标体系的科学性,进而为未来相关研究提供有价值的参考。

二、文献综述

随着数据要素市场化改革深入推进,围绕数据资源信息披露经济后果的研究逐渐兴起。由于当前数据资源信息披露仍属于企业自愿性披露范畴,上市公司主要通过年报中的表外文本披露数据资源相关信息^[14],已有研究主要从外部信息环境和企业经营发展两个维度展开探讨。在外部信息环境方面,数据资源信息披露能够提升分析师关注度及预测准确性、缓解银企之间的信息不对称,并影响企业审计监督^[2-3,13,15];在企业经营发展方面,数据资源信息披露有助于提升企业决策效率、促进创新产出、优化人力资本配置^[1,4-5]。总体而言,现有研究普遍认为数据资源信息披露能够改善企业信息环境和资源获取能力,并产生积极的经济后果。

企业投资效率通常表现为企业能否将有限资源配置于净现值为正的项目,从而实现资源的有效配置^[16]。已有研究主要从企业内部治理和外部制度环境两个方面探讨投资效率的影响因素。在企业内部治理方面,公司治理结构、财务报告质量以及ESG信息、关键审计事项等非财务信息披露均被证实能够改善企业信息环境、强化监督机制,从而提高企业投资效率^[8-10,17];在外部制度环境方面,政府数字治理改革、分析师跟踪以及财会监督等外部治理机制通过降低制度性交易成本和缓解信息不对称,能够有效抑制企业的非效率投资行为^[11-12,18]。面对数字经济浪潮,已有研究发现信息基础设施建设、数字创新等数字化因素能够促进企业投资效率提升,而数字经济发展与企业投资效率之间还可能存在非线性关系^[19-21]。

综合来看,现有研究分别围绕数据资源信息披露经济后果和企业投资效率影响因素形成了较为丰富的研究成果,但仍存在以下不足:第一,已有关于数据资源信息披露经济后果的研究主要聚焦于外部信息使用者反应和企业经营绩效,对于其是否以及如何影响企业投资效率,仍缺乏系统的理论分析与经

验证据。第二,关于企业投资效率影响因素的研究虽已从财务信息质量拓展至 ESG、数字经济等非财务信息和数字化因素,但对数据资源信息披露这一企业层面信息披露行为的资源配置效应关注不足。第三,现有研究较少关注民营企业这一特殊情境。相较于国有企业,民营企业通常面临更严重的信息不对称和代理问题,其投资效率对信息环境改善更为敏感。因此,本文聚焦民营企业,考察数据资源信息披露对企业投资效率的影响关系及其作用机制,从而弥补上述研究文献的不足。

三、理论分析与假设提出

传统投资理论认为,企业投资主要由投资机会所驱动。然而,在现实资本市场中,企业的投资决策通常面临信息不对称和代理问题的约束^[22-23],从而偏离最优投资水平。对于民营企业而言,其信息透明度与外部关注度普遍低于国有企业,控制权集中所引致的代理冲突也更为突出^[7],对投资决策的影响通常更为严重。随着数字经济的发展,数据资源已成为具有重要经济价值的关键生产要素。通过加强数据资源信息披露,企业不仅能够满足利益相关者的信息需求,还能够强化对管理层行为的内外部监督约束。本文认为,数据资源信息披露能够通过缓解信息不对称与抑制代理问题,提高民营企业的投资效率。

首先,数据资源信息披露有助于满足利益相关者对民营企业数据资源的信息需求,通过缓解信息不对称问题提升企业的投资效率。与厂房、设备等有形资产不同,数据资源具有无形性强、异质性高、价值实现依赖具体应用场景等鲜明特征,其经济价值难以通过传统财务报表充分反映。因此,外部利益相关者往往难以及时识别企业数据资源的规模、质量、开发利用情况及潜在价值,导致投融资双方之间存在显著的信息不对称问题^[5]。长期面临资源劣势的民营企业通过积极披露数据资源信息,能够提供标准化财务信息之外的重要增量信息,向外界传递关于数据资源规模、结构及其潜在价值的关键信息^[2],从而有效缓解信息不对称。这有助于利益相关者更准确地评估民营企业的数据资源禀赋、数字化能力及未来发展潜力,提高市场对企业价值和投资项目质量的判断能力,减少因信息偏差而导致的逆向选择和资源错配^[23],促使资金更多流向净现值为正的优质项目,从而提升企业投资效率。

其次,数据资源信息披露能够强化对民营企业管理层和大股东行为的内外部监督约束,通过抑制代理问题提升企业的投资效率。根据代理理论,现代企业治理中存在两类代理冲突:其一,管理层可能基于规模扩张或职位稳固等私人动机,将资源配置于净现值为负或回报偏低的项目,从而引发非效率投资行为^[22];其二,大股东凭借控制权优势可能通过关联交易、资金占用等“隧道挖掘”行为侵占公司资源,进而导致资源错配并降低投资效率^[24]。数据资源信息披露通过提升信息透明度能够对上述两类代理问题产生约束作用。一方面,数据资源信息披露提高了企业数据资源开发、管理和应用情况的透明度,使董事会、投资者及其他利益相关者能够更加全面地了解企业数据资源配置和利用情况,增强对管理层投资决策及受托责任履行情况的监督,从而抑制其机会主义行为。另一方面,作为关键战略资源的数据资源,其信息披露能够向外部利益相关者传递更多关于企业数字化经营和资源配置能力等关键信息,增强外部市场主体对大股东行为的监督,进而抑制其利益侵占行为^[2]。随着代理问题的缓解,民营企业将更加关注项目的真实回报与长期价值,减少资源错配,最终提升企业的投资效率^[11]。

综合上述理论分析,本文认为数据资源信息披露能够通过缓解信息不对称问题和抑制代理问题进而提高民营企业的投资效率,据此提出如下研究假设 H:

H:同等条件下,数据资源信息披露有助于提高民营企业的投资效率。

四、研究设计

(一) 样本选择与数据来源

本文选取 2010—2024 年沪深 A 股民营上市公司的年度观测数据作为研究样本。选择 2010 年作为

样本初始年份,主要基于两点考虑:一是我国自 2007 年起实施与国际趋同的企业会计准则,该制度变革可能对企业投资效率的衡量产生影响;二是 2008—2009 年国际金融危机对企业投资决策造成显著冲击。为了提高样本间的可比性,本文对初始得到的 38318 个公司-年度观测样本进行了如下筛选:(1)剔除 ST、*ST 的公司年度样本;(2)剔除金融行业的公司样本;(3)剔除上市不足 2 年的公司年度样本,因为本文关键变量的构造需要用到 $t-2$ 年的财务数据;(4)剔除相关变量数据缺失的公司年度样本。最终,本文得到了 24100 个有效的公司-年度观测样本。

本文用于构建企业数据资源信息披露指标的公司年报文本数据来自巨潮资讯网,上市公司注册所在地的市场化指数来自王小鲁等编制的中国分省份市场化指数^[25],其他财务数据和公司治理数据则来自 CSMAR 数据库。

(二) 变量定义

1. 被解释变量:企业的投资效率($INVEFF$)。借鉴申志轩等和 Richardson 的做法^[12,16],本文使用计量回归模型估计企业的期望投资水平,使用企业实际投资水平与期望投资水平的差值即回归模型残差衡量企业的非效率投资水平,反向反映企业的投资效率水平。具体地,本文用于估计企业的期望投资水平的计量回归模型如式(1)所示:

$$INVEST_{i,t} = \alpha_0 + \alpha_1 SIZE_{i,t-1} + \alpha_2 LEV_{i,t-1} + \alpha_3 AGE_{i,t-1} + \alpha_4 CASH_{i,t-1} + \alpha_5 GROWTH_{i,t-1} + \alpha_6 RETURNS_{i,t-1} + \alpha_7 INVEST_{i,t-1} + \sum Ind + \sum Year + \varepsilon_{i,t} \quad (1)$$

其中, $INVEST$ 为企业的新增投资支出,采用(购建固定资产、无形资产和其他长期资产支付的现金净额+取得子公司及其他营业单位支付的现金净额-处置固定资产、无形资产和其他长期资产收回的现金净额-处置子公司及其他营业单位收到的现金净额)/年初总资产衡量; $SIZE$ 为企业规模,采用企业年末总资产的自然对数衡量; LEV 为资产负债率,采用企业年末总负债与总资产的比值衡量; AGE 为企业年龄,采用企业创立年限加 1 后取自然对数衡量; $CASH$ 为企业现金持有量,采用年末现金及现金等价物之和与总资产的比值衡量; $GROWTH$ 为营业收入增长率,采用企业本年与上一年营业收入的差异与上一年营业收入的比值衡量; $RETURNS$ 为公司股票收益率,采用考虑现金红利再投资的年个股回报率衡量; $Year$ 和 Ind 表示年度固定效应和行业固定效应。本文使用残差 $\varepsilon_{i,t}$ 的绝对值($|\varepsilon_{i,t}|$)反映企业的非效率投资水平, $|\varepsilon_{i,t}|$ 越大,表明企业的非效率投资程度越高,投资效率越低。

2. 解释变量:数据资源信息披露水平($ARDAF$)。由于数据资源正式实现资产入表面临着现行会计准则对确认与计量的严格要求,现实中企业通过资产负债表确认数据资源的比例较低,例如截至 2025 年第三季度仅有 102 家企业实现了数据资产入表,绝大多数企业仍主要通过非结构化的年报文本形式对其拥有的数据资源进行表外披露^[14]。因此,本文参考罗进辉等的做法^[14],采用 FinBERT2 大语言模型对上市公司年报中与数据资源相关的表述进行识别和量化,据此衡量企业拥有的数据资源及其披露水平。相较于传统基于关键词词频的词典方法,该方法能够结合上下文语境识别企业是否真实披露其数据资源的持有、使用与经营情况,从而在一定程度上缓解词典覆盖不足与语义误判带来的测度误差。

具体地,第一步,对上市公司年报文本进行预处理。依据句号、感叹号、问号等终止符对年报文本进行语句切分,并剔除页眉页脚、格式化文本、表格内容以及过长或过短的语句,以提高文本质量。第二步,借鉴何瑛等的做法^[3],运用 Word2Vec 方法,以“数字”“数据”“信息”“网络”等词语为种子词扩展数据资源相关词汇,据此从年报文本中初步筛选涉及数据资源表述的语句。第三步,结合生成式人工智能辅助识别与人工复核,对语句是否体现企业对数据资源的持有、使用或经营进行标注,并在此基础上构建训练样本,进而使用 FinBERT2 模型进行监督学习训练。在模型训练过程中,将标注样本按照 8:1:1 的比例划分为训练集、验证集和测试集,对 FinBERT2 模型进行微调,以提升其在数据资源语

句识别任务中的分类性能。结果表明,微调后的模型在测试集上的准确率、精确率和召回率分别达到 94.2%、89.8%、94.4%,具有较高的识别有效性。第四步,本文利用训练完成的 FinBERT2 模型对年报文本进行逐句分类识别,并以识别出的数据资源相关语句数量占年报总语句数量的比例衡量企业的数据资源信息披露水平指标(*ARDAF*)。该指标的数值越大,表明企业披露拥有的数据资源水平越高。

3. 控制变量。参考已有关于企业投资效率的研究文献^[11-12],本文控制了一系列可能影响企业投资效率的公司治理变量、财务特征变量,具体包括:公司规模(*SIZE*)、负债水平(*LEV*)、盈利能力(*ROA*)、企业年龄(*AGE*)、固定资产比率(*PPE*)、经营现金流(*CASH*)、有形资产比率(*TANG*)、管理层持股比例(*MSHARE*)、管理层薪酬(*COMPS*)、独立董事占比(*INDEP*)、董事会规模(*BOARD*)、两权分离度(*SEP*)、两职合一(*DUAL*)、第一大股东持股(*TOP1*)、审计意见(*OPIN*)、地区市场化程度(*MARKET*)。这些变量的具体定义如表 1 所示。

表 1 变量定义

变量名称	变量符号	变量定义
企业投资效率	<i>INEFF</i>	模型(1)残差的绝对值,数值越大,投资效率越低
数据资源信息披露	<i>ARDAF</i>	年报中包含数据资源表述的句数占比
公司规模	<i>SIZE</i>	公司年末总资产的自然对数
负债水平	<i>LEV</i>	公司年末负债总额与资产总额的比值
盈利能力	<i>ROA</i>	公司年末净利润与资产总额的比值
企业年龄	<i>AGE</i>	公司成立年限加 1 之和的自然对数
固定资产比率	<i>PPE</i>	公司年末固定资产与总资产的比值
经营现金流	<i>CASH</i>	公司年末经营活动产生的现金流量净额与总资产的比值
有形资产比率	<i>TANG</i>	公司年末有形资产与总资产的比值
管理层持股比例	<i>MSHARE</i>	公司董监高持有股份数与总股数的比值
管理层薪酬	<i>COMPS</i>	管理层前三名薪酬总额与管理层薪酬总额的比值
独立董事占比	<i>INDEP</i>	公司独立董事人数与董事会总人数的比值
董事会规模	<i>BOARD</i>	公司董事会总人数的自然对数
两权分离度	<i>SEP</i>	实际控制人拥有上市公司控制权与所有权的差
两职合一	<i>DUAL</i>	哑变量,若公司的董事长兼任总经理,则取值为 1,否则为 0
第一大股东持股	<i>TOP1</i>	公司第一大股东持有股份数与总股数的比值
审计意见	<i>OPIN</i>	哑变量,若公司年报审计获得标准无保留意见,则取值为 1,否则为 0
地区市场化程度	<i>MARKET</i>	王小鲁等编制的中国各省或直辖市的市场化指数 ^[25]

(三) 计量回归模型

为了检验研究假设 H 有关数据资源信息披露对民营企业投资效率的经验影响关系,本文构建了如下式(2)的计量回归模型:

$$INEFF_{i,t} = \beta_0 + \beta_1 ARDAF_{i,t} + \beta_2 CONTROL_{i,t} + \sum Firm + \sum Year + \varepsilon_{i,t} \quad (2)$$

其中,被解释变量是企业的非效率投资水平(*INEFF*),解释变量是数据资源信息披露水平(*ARDAF*),本文在回归模型中同时引入了公司固定效应和年度固定效应,并对回归标准误进行了公司层面的聚类调整。此外,为了缓解极端数据的干扰影响,本文对所有连续型变量进行了上下 1% 的 Winsorize 缩尾处理。根据假设 H 的理论预期,回归系数 β_1 应该显著为负,即数据资源披露有助于降低民营企业的非效率投资水平,提高企业投资效率。

五、实证结果分析

(一) 描述性统计

表2列示了主要变量的描述性统计结果。可以看到,企业非效率投资变量(*INEFF*)的均值和标准差分别为0.042和0.049,表明不同民营企业样本间的投资效率水平存在较大差异。数据资源信息披露变量(*ARDAF*)的均值和标准差分别为0.005和0.007,中位数为0.002,说明样本企业的数据资源信息披露水平总体较低,而且不同企业间的差异也比较大。此外,其他控制变量的取值不存在明显的异常情况。

(二) 基准回归

表3列示了数据资源信息披露影响民营企业投资效率的OLS多元回归分析结果。其中,列(1)为仅包含核心解释变量*ARDAF*的单变量回归结果;列(2)加入企业财务特征、公司治理特征及地区市场化程度等控制变量;列(3)进一步控制公司固定效应和年份固定效应。根据列(3)的结果,数据资源信息披露变量(*ARDAF*)得到了5%统计显著的负回归系数,而且,根据回归系数可以计算得到,样本企业的数据资源信息披露水平每增加1个标准差,企业的非效率投资水平相对样本均值将降低3.78% ($-0.227 \times 0.007 / 0.042$)。这些结果表明,数据资源信息披露对民营企业投资效率的促进作用同时具有统计显著性和经济显著性,假设H得到了经验证据支持。

(三) 稳健性检验^①

为缓解反向因果、样本自选择及样本选择偏误等问题对估计结果的干扰,本文依次采用工具变量法、处理效应模型和PSM-DID方法进行内生性检验;同时,为排除外生事件冲击、测度口径干扰及遗漏变量偏误对估计结果的影响,本文分别通过特殊样本筛选、替换变量度量方法,以及进一步控制行业年度数字投资水平和省份-行业交互固定效应等方式开展稳健性检验。上述检验结果均表明,数据资源信息披露变量的估计系数仍显著为负,说明数据资源信息披露能够提升民营企业投资效率的研究结论具有较强的稳健性。

表2 主要变量的描述性统计结果

变量	样本量	均值	标准差	最小值	中位数	最大值
<i>INEFF</i>	24100	0.042	0.049	0.001	0.028	0.300
<i>ARDAF</i>	24100	0.005	0.007	0.000	0.002	0.043
<i>SIZE</i>	24100	21.999	1.083	19.648	21.885	25.339
<i>LEV</i>	24100	0.404	0.201	0.056	0.392	0.959
<i>ROA</i>	24100	0.030	0.080	-0.372	0.035	0.211
<i>AGE</i>	24100	2.970	0.315	1.386	2.996	4.234
<i>PPE</i>	24100	0.197	0.134	0.002	0.176	0.578
<i>CASH</i>	24100	0.048	0.070	-0.172	0.047	0.250
<i>TANG</i>	24100	0.920	0.090	0.526	0.952	1.000
<i>MSHARE</i>	24100	0.182	0.199	0.000	0.104	0.680
<i>COMPS</i>	24100	0.507	0.139	0.254	0.488	0.918
<i>INDEP</i>	24100	0.380	0.053	0.333	0.364	0.571
<i>BOARD</i>	24100	2.071	0.189	1.609	2.197	2.485
<i>SEP</i>	24100	0.046	0.071	0.000	0.001	0.286
<i>DUAL</i>	24100	0.379	0.485	0.000	0.000	1.000
<i>TOP1</i>	24100	0.307	0.134	0.081	0.290	0.689
<i>OPIN</i>	24100	0.953	0.212	0.000	1.000	1.000
<i>MARKET</i>	24100	10.435	1.815	4.667	10.560	13.867

表3 基准回归分析结果

	(1)	(2)	(3)
	<i>INEFF</i>	<i>INEFF</i>	<i>INEFF</i>
<i>ARDAF</i>	-0.445 *** (-9.106)	-0.337 *** (-6.330)	-0.227 ** (-2.088)
<i>SIZE</i>		-0.002 *** (-4.061)	0.003 ** (2.388)
<i>LEV</i>		0.013 *** (5.483)	0.020 *** (4.255)
<i>ROA</i>		0.069 *** (14.063)	0.059 *** (10.511)
<i>AGE</i>		-0.014 *** (-10.059)	-0.020 ** (-2.443)
<i>PPE</i>		0.024 *** (7.433)	-0.060 *** (-7.498)
<i>CASH</i>		-0.037 *** (-6.774)	-0.021 *** (-3.526)
<i>TANG</i>		-0.124 *** (-20.037)	-0.134 *** (-13.863)
<i>MSHARE</i>		0.003 (1.085)	-0.001 (-0.157)
<i>COMPS</i>		0.016 *** (4.816)	0.007 (1.302)
<i>INDEP</i>		0.013 (1.289)	-0.027 * (-1.824)
<i>BOARD</i>		0.000 (0.082)	-0.001 (-0.180)
<i>SEP</i>		0.003 (0.398)	-0.004 (-0.264)
<i>DUAL</i>		0.001 (0.862)	-0.000 (-0.022)
<i>TOP1</i>		0.010 *** (3.173)	0.033 *** (4.118)
<i>OPIN</i>		0.001 (0.438)	0.008 *** (3.453)
<i>MARKET</i>		-0.001 *** (-4.623)	0.001 (0.582)
截距项		0.227 *** (15.400)	0.150 *** (3.539)
公司固定效应	未控制	未控制	控制
年份固定效应	未控制	未控制	控制
样本量	24100	24100	24100
调整 R ²	0.004	0.080	0.201

注:(1)***、**、*分别代表统计显著水平为1%、5%和10%;(2)括号内为经过公司层面聚类调整的T值。下同。

^①考虑篇幅,稳健性检验结果未列示,留存备索。

(四) 影响机制检验

1. 缓解信息不对称的机制分析

在假设 H 的理论分析中,本文指出数据资源信息披露有助于投资者等利益相关者了解民营企业拥有的数据资源等重要价值资源,缓解企业内外部的信息不对称问题,从而帮助民营企业的投资活动得到更好的外部支持,进而提高投资效率。基于此,本文将进一步检验民营企业数据资源信息披露的信息不对称缓解机制。具体地,参考陈丽莉等的做法^[26],本文使用上市公司股票交易的微观结构特征构建信息不对称程度的代理变量,选取流动性比率、非流动性比率、收益率反转指标这三个能够反映信息不对称的指标进行主成分分析,提取第一主成分作为企业的综合信息不对称指标(ASY1)。变量 ASY1 数值越大,代表企业的信息不对称程度越高。然后,参考丁杰和黄金波的研究^[27],使用 KV 指数衡量企业信息不对称(ASY2)。KV 指数反映了投资者对股票交易量信息的依赖程度,公司信息披露越充分,投资者对交易量信息的依赖就越低。ASY2 取值越大,表示上市公司信息披露质量越低,信息不对称问题越严重。

本文以信息不对称指标作为因变量进行的回归分析结果如表 4 列(1)和列(2)所示。可以看到,当以 ASY1 和 ASY2 作为因变量时,数据资源信息披露变量(ARDAF)均得到了 1% 统计显著水平的负回归系数,表明数据资源信息披露能够显著降低民营企业的信息不对称问题,支持了本文的缓解信息不对称机制。

2. 抑制代理问题的机制分析

本文假设逻辑的另一个影响机制是抑制代理问题机制,即数据资源信息披露能够强化对民营企业管理层与大股东机会主义行为的内外部监督,通过抑制代理问题提高民营企业的投资效率。基于此,本文将进一步检验民营企业数据资源信息披露的代理问题抑制机制。具体地,参考罗进辉的做法^[28],本文采用营业收入与总资产的比值反向衡量股东与管理层之间的第一类代理成本(AC1),以其他应收款与总资产的比值衡量大股东与中小股东之间的第二类代理成本(AC2)。AC1 的取值越小,表明企业面临的第一类代理问题越严重。AC2 的取值越大,则表明企业面临的第二类代理问题越严重。

本文以上述两个代理成本变量作为因变量进行的回归分析结果如表 4 列(3)和列(4)所示。可以看到,当以第一类代理成本(AC1)为因变量时,数据资源信息披露(ARDAF)得到了 1% 统计显著的正回归系数;当以第二类代理成本(AC2)为因变量时,数据资源信息披露(ARDAF)则得到了 10% 统计显著的负回归系数。这些结果表明,数据资源信息披露能够抑制民营企业的两类代理问题,支持了本文的抑制代理问题机制。

表 4 机制检验结果

	(1)	(2)	(3)	(4)
	ASY1	ASY2	AC1	AC2
ARDAF	-2.240 *** (-3.240)	-0.927 *** (-3.153)	2.442 *** (3.193)	-0.149 * (-1.727)
控制变量	控制	控制	控制	控制
公司固定效应	控制	控制	控制	控制
年份固定效应	控制	控制	控制	控制
样本量	24083	24020	24100	24087
调整 R ²	0.731	0.624	0.744	0.410

(五) 异质性分析

1. 企业科技属性

在数字经济时代,高科技企业通常拥有更多可观的数据资源,其生产经营活动更依赖数据要素与持续创新,因而数据资源对于这些企业具有更突出的战略价值^[26]。与科技企业相比,传统的非科技企业对数据资源的依赖程度相对较低,数据资源信息披露所带来的边际信息增量可能相对有限。基于此,本文预期,数据资源信息披露对民营企业投资效率的促进作用可能更多体现在高科技企业中。

为了检验上述理论预期,本文基于企业的科技属性进行分组回归分析。具体地,本文分别借鉴吴卫红等、牛华等的分类做法将样本公司划分为高科技企业和非高科技企业^[29-30],其中吴卫红等把行业代码为 C25、C26、C27、C37、C38、C39、C40、C42、D44、I63、I64、I65、M73、N77 的 14 个行业划分为高科技行

业企业^[29],牛华等使用企业研发投入占营业收入的比例衡量研发投入强度^[30],并根据研发投入强度的年度中位数将样本划分为高科技企业和非高科技企业。本文据此进行的分组回归分析结果如表5所示,可以看到,无论采用行业分类还是采用研发投入强度进行划分,在高科技企业子样本中,ARDAF都得到了统计显著的负回归系数。虽然组间系数差异检验未达到边际统计显著水平,不能严格认为两类企业之间存在显著差异,但上述结果仍在一定程度上表明,数据资源信息披露对投资效率的促进作用可能更多体现在数据要素依赖程度较高的高科技民营企业中。

2. 企业规模

企业规模是影响民营企业资源禀赋、治理结构的一个重要因素。相较于大型民营企业,中小民营企业通常面临更为严重的资源约束、治理不完善等问题。相对而言,数据资源信息披露帮助企业获得更为优质的外部融资、改善信息环境并强化治理约束的作用,在中小民营企业中可能表现得更强。基于此,本文预期,数据资源信息披露对投资效率的促进作用可能在中小民营企业中表现得更为明显。

为了检验上述理论预期,本文根据企业规模进行分组回归分析。具体地,本文参考申志轩等的做法^[12],以企业总资产的自然对数衡量企业规模,并根据企业规模的年度中位数将样本划分为规模较大的子样本和规模较小的子样本,据此进行的分组回归分析结果如表6列(1)和列(2)所示。可以看到,在规模较小的子样本中,ARDAF得到了1%统计显著水平的负回归系数。这些结果表明,数据资源信息披露对投资效率的促进作用在规模较小的民营企业中表现得显著更强。

3. 企业资产结构

除企业规模外,资产结构是影响企业资源配置方式与投资决策逻辑的另一个重要因素。由于数据资源正式入表还在探索中,数据资源资产化的占比还非常低,使得其并不能真正有效地扩大企业的资产规模,但是公开透明地披露具有可观价值的数据资源仍然具有重要的现实意义,这种意义对于更依赖数据资源等无形要素参与价值创造的固定资产密集度较低的企业将显得更为重要。基于此,本文预期,数据资源信息披露对投资效率的促进作用将在低资产密集型民营企业中表现得更强。

为了检验上述理论预期,本文根据企业资产结构进行分组回归分析。具体地,本文以固定资产净值占总资产的比例衡量企业资产结构,并根据企业资产结构的年度中位数将样本划分为高资产密集型子样本与低资产密集型子样本,据此进行的分组回归分析结果如表6列(3)和列(4)所示。可以看到,在低资产密集型子样本中,ARDAF得到了5%统计显著水平的负回归系数。

表5 企业科技属性的异质性检验结果

	(1)	(2)	(3)	(4)
	INEFF	INEFF	INEFF	INEFF
	基于吴卫红等的分类		基于牛华等的分类	
	高科技企业	非高科技企业	高科技企业	非高科技企业
ARDAF	-0.283 ** (-2.150)	-0.088 (-0.456)	-0.241 * (-1.743)	-0.294 (-1.596)
控制变量	控制	控制	控制	控制
公司固定效应	控制	控制	控制	控制
年份固定效应	控制	控制	控制	控制
样本量	12755	11314	13286	10373
调整 R ²	0.221	0.210	0.221	0.222
组间系数差异(P值)	0.174		0.379	

注:组间系数差异P值用于检验ARDAF组间系数差异的显著性,利用费舍尔组合检验(Fisher's Permutation test)重复取样1000次得到。下同。

表6 企业规模及资产结构的异质性检验结果

	(1)	(2)	(3)	(4)
	INEFF	INEFF	INEFF	INEFF
	企业规模较大	企业规模较小	高资产密集型	低资产密集型
ARDAF	-0.133 (-0.786)	-0.409 *** (-2.675)	0.167 (0.771)	-0.268 ** (-2.120)
控制变量	控制	控制	控制	控制
公司固定效应	控制	控制	控制	控制
年份固定效应	控制	控制	控制	控制
样本量	11915	11870	11812	11771
调整 R ²	0.253	0.195	0.190	0.260
组间系数差异(P值)	0.078 *		0.015 **	

4. 企业数字投资水平

数字投资是企业数字理念支持下对数字资源的资金注入与配置,对企业数字化发展和数据要素利用能力的提升有重要意义。相比于数字投资水平较低的企业,数字投资水平较高的企业通常积累了更多的数据资源,其数据资源信息在外部利益相关者判断企业风险和收益的过程中具有更强的价值相关性,并进一步作用于企业投资机会识别和资源配置过程^[31]。基于此,本文预期,数据资源信息披露对民营企业投资效率的促进作用可能更多体现在数字投资水平较高的企业中。

为了检验上述理论预期,本文参考孙忠娟等的做法^[32],使用数字投资总额与营业收入的比值衡量企业数字投资水平。其中,数字投资总额由企业财务报告中数字固定资产与数字无形资产期末金额加总得到。根据数字投资水平的年度中位数将样本划分为高数字投资子样本和低数字投资子样本,据此进行的分组回归分析结果如表7列(1)和列(2)所示。可以看到,在高数字投资水平子样本中,ARDAF得到了5%统计显著水平的负回归系数。虽然组间系数差异检验未达到边际统计显著水平,不能严格认为两组企业之间存在显著差异,但上述结果仍在一定程度上表明,数据资源信息披露对投资效率的促进作用可能更多体现在数字投资水平较高的民营企业中。

5. 地区市场化程度

市场化环境是影响企业信息环境和资源配置效率的重要外部制度因素。一般而言,市场化程度较高的地区具有更加完善的市场机制与法治环境,资本市场参与者能够更有效地识别、解读和利用企业披露的信息,从而增强信息披露的治理效应^[10]。因此,数据资源信息披露与良好的外部制度环境具有互补性,本文预期其对民营企业投资效率的促进作用将在市场化程度较高的地区更加明显。

为了检验上述理论预期,本文根据企业所处地区的市场化程度进行分组回归分析。具体地,本文使用王小鲁等编制的市场化指数衡量地区市场化程度^[25],并根据地区市场化程度的年度中位数将样本划分为高市场化程度子样本和低市场化程度子样本,据此进行的分组回归分析结果如表7列(3)和列(4)所示。可以看到,在

表7 企业数字投资水平及地区市场化程度的异质性检验结果

	(1)	(2)	(3)	(4)
	INVEFF	INVEFF	INVEFF	INVEFF
	高数字投资水平	低数字投资水平	高市场化程度	低市场化程度
ARDAF	-0.327 ** (-2.399)	-0.105 (-0.610)	-0.375 ** (-2.388)	-0.094 (-0.650)
控制变量	控制	控制	控制	控制
公司固定效应	控制	控制	控制	控制
年份固定效应	控制	控制	控制	控制
样本量	11764	11793	13370	10707
调整 R ²	0.227	0.217	0.211	0.201
组间系数差异(P值)		0.129		0.068 *

高市场化程度子样本中,ARDAF得到了5%统计显著水平的负回归系数。这些结果表明,数据资源信息披露对高市场化程度地区民营企业投资效率的促进作用表现得更强。

(六) 企业投资效率提升的方向

企业投资效率的提升既可能表现为抑制过度投资,也可能表现为缓解投资不足。投资过度通常源于管理层基于规模扩张、职位稳固等机会主义动机,将资源配置于净现值较低甚至为负的项目。相比之下,投资不足既可能受到管理层风险规避、职业声誉顾虑等代理因素影响,也可能受到企业融资能力以及投资机会识别不足等因素制约。数据资源信息披露通过提升信息透明度与强化监督约束,更有助于抑制管理层的机会主义行为,从而缓解过度投资^[10]。然而,对于融资约束、风险规避等因素导致的投资不足,数据资源信息披露难以直接改变企业的资金约束和风险偏好,其改善作用相对有限。基于此,本文预期,数据资源信息披露主要是抑制了民营企业的过度投资问题。

为进一步识别数据资源信息披露提升企业投资效率的具体方向,本文借鉴 Richardson 的模型^[16],根据回归残差的符号分别构造投资过度(*OVERINV*)和投资不足(*UNDERINV*)变量。其中,残差大于0定义为投资过度,残差小于0定义为投资不足,并取绝对值处理。回归分析结果如表8所示,可以看到,*ARDAF*显著降低了投资过度,但对投资不足的影响未达到统计显著水平。这些结果说明,与本文的预期相符,数据资源信息披露对民营企业投资效率的促进作用主要表现为有效抑制了企业的过度投资问题。

(七) 经济后果分析

根据前文的研究结论,数据资源信息披露能够提高民营企业的投资效率。那么,数据资源信息披露通过提高投资效率能否最终赋能民营企业的高质量发展?为了回答这个经济后果问题,本文进一步检验了数据资源信息披露对民营企业投资效率的影响如何作用于企业的全要素生产率。具体地,本文借鉴鲁晓东和连玉君的做法^[33],分别选取FE法和OLS法计算企业的全要素生产率,分别记为*TFP_FE*和*TFP_OLS*,构建了如下式(3)的计量回归模型:

$$\frac{TFP_FE_{i,t}}{TFP_OLS_{i,t}} = \lambda_0 + \lambda_1 ARDAF_{i,t} \times INVEFF_{i,t} + \lambda_2 ARDAF_{i,t} + \lambda_3 INVEFF_{i,t} + \lambda_4 CONTROL_{i,t} + \sum Firm + \sum Year + \varepsilon_{i,t} \quad (3)$$

其中,交乘项*ARDAF* × *INVEFF*的估计系数是本文关注的重点,当估计系数 λ_1 显著为负时,表明数据资源信息披露对投资效率的促进作用有助于提高民营企业的全要素生产率。

回归分析结果如表9所示。可以看到,列(1)和列(2)中交乘项*ARDAF* × *INVEFF*均得到了统计显著水平的负回归系数,意味着数据资源信息披露通过提升民营企业投资效率促进了企业全要素生产率提升,说明数据资源信息披露是赋能民营企业高质量发展的一项重要举措。

表8 企业投资效率提升方向检验

	(1)	(2)
	<i>OVERINV</i>	<i>UNDERINV</i>
<i>ARDAF</i>	-0.634 ** (-2.183)	-0.063 (-0.889)
控制变量	控制	控制
公司固定效应	控制	控制
年份固定效应	控制	控制
样本量	8437	14654
调整 R ²	0.227	0.163

表9 数据资源信息披露经济后果的回归分析结果

	(1)	(2)
	<i>TFP_FE</i>	<i>TFP_OLS</i>
<i>ARDAF</i> × <i>INVEFF</i>	-17.247 ** (-1.963)	-16.012 * (-1.810)
<i>ARDAF</i>	4.180 *** (3.711)	4.117 *** (3.669)
<i>INVEFF</i>	-0.452 *** (-4.968)	-0.457 *** (-5.150)
控制变量	控制	控制
公司固定效应	控制	控制
年份固定效应	控制	控制
样本量	23799	23799
调整 R ²	0.927	0.920

六、结论性评述

本文基于2010-2024年中国A股民营上市公司数据,利用FinBERT2大语言模型构建企业数据资源信息披露指标,系统考察了其对民营企业投资效率的影响及作用机制。研究发现:第一,数据资源信息披露能够显著降低民营企业的非效率投资水平,即显著提升其投资效率,且结论经多项稳健性检验后依然成立。第二,数据资源信息披露主要通过缓解信息不对称和抑制代理问题提升投资效率,其作用主要表现为抑制企业的过度投资。第三,数据资源信息披露对民营企业投资效率的促进作用在高科技企业、小规模企业、低资产密集型企业、高数字投资水平企业以及市场化程度较高地区的企业中更为明显。进一步分析表明,数据资源信息披露通过提升投资效率,促进了民营企业的全要素生产率提升。

本文的研究结论具有重要的政策启示和实践意义。首先,民营企业在当前数字经济时代应充分认识数据资源和数据资产的战略价值,积极完善数据资源管理与信息披露机制。本文研究表明,民营企业应围绕数据资源的权属界定、开发利用与安全治理,构建高质量的数据资源信息披露

体系,并将其嵌入企业投资决策、技术创新与战略管理中,以充分释放数据要素价值并缓解企业的非效率投资问题。其中,高科技民营企业可重点披露数据资源在技术创新和价值创造中的应用,中小民营企业则应结合数字化转型进程,逐步完善数据资源管理和信息披露机制。其次,政府监管部门应进一步健全数据确权、计量、披露、交易与安全治理等配套规则,顺应我国数据要素市场建设的发展方向,逐步建立统一的数据资源信息披露标准。短期应完善数据资源信息披露指引和监督机制,引导企业规范披露;中长期应健全数据资源价值评估、产权保护及流通交易制度,持续推进地区市场化改革,强化数据资源信息披露的治理效应。特别地,应重点关注高科技企业、小规模企业及低资产密集型企业,加强对这些企业的数据资源信息披露的监督激励。最后,资本市场投资者和证券分析师应将企业披露的数据资源信息纳入企业价值评估与投资决策体系,提高投资决策和盈余预测质量。分析师可将企业数据资源规模、开发利用能力及数据治理水平纳入估值分析框架,投资者可结合企业年报中关于数据资源应用场景、商业模式及相关风险的披露,更加全面地评估企业长期发展潜力,尤其是对于数据要素依赖程度较高的高科技企业和数字化企业,从而提升投资决策的有效性。

参考文献:

- [1] Hannila H, Silvola R, Harkonen J, et al. Data-driven begins with DATA; Potential of data assets[J]. Journal of Computer Information Systems, 2022, 62(1): 29-38.
- [2] 危雁麟,张俊瑞,汪方军,等.数据资产信息披露与分析师盈余预测关系研究——基于文本分析的经验证据[J].管理工程学报,2022(5):130-141.
- [3] 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.
- [4] 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(8):64-75.
- [5] 赵雪琦,孙小哲,李纪珍.数据资产信息披露对数字技术创新的影响研究[J/OL].科学学研究,https://doi.org/10.16192/j.cnki.1003-2053.20250714.005
- [6] 张璇,刘贝贝,汪婷,等.信贷寻租、融资约束与企业创新[J].经济研究,2017(5):161-174.
- [7] 钱爱民,肖亦忱,吴春天.国有股东委派董事能否改善民营企业信息不对称[J].金融研究,2024(8):132-149.
- [8] 黄炳艺,黄雨婷.职工董事影响企业投资效率吗——基于中国资本市场的经验证据[J].会计研究,2022(5):77-91.
- [9] Chen F, Hope O K, Li Q, et al. Financial reporting quality and investment efficiency of private firms in emerging markets[J]. The Accounting Review, 2011, 86(4): 1255-1288.
- [10] 李泽浩,杨振,陈传龙.ESG信息披露提升了企业投资效率吗?——非财务信息披露的财务作用[J].审计与经济研究,2024(4):71-80.
- [11] 余一品,郭王非,袁璐,等.政府财会监督对企业投资行为的治理效应研究[J].会计研究,2025(10):134-151.
- [12] 申志轩,祝树金,文茜,等.以有为政府赋能有效市场:政府数字治理与企业投资效率[J].世界经济,2025(2):166-195.
- [13] 张俊瑞,危雁麟,尹兴强,等.企业数据资源信息披露影响审计收费吗?——基于文本分析的经验证据[J].审计研究,2023(3):60-71.
- [14] 罗进辉,王毓泽,易炫彤,等.企业数据资源信息披露现状与测度——基于大语言模型的指标构建[J].财会月刊,2026(9):8-19.
- [15] 牛彪,于翔,苑泽明,等.数据资产信息披露与审计师定价策略[J].当代财经,2024(2):154-164.
- [16] Richardson S. Over-investment of free cash flow[J]. Review of Accounting Studies, 2006, 11(2-3): 159-189.
- [17] Chan D K, Liu N. The effects of critical audit matter disclosure on audit effort, investor scrutiny, and investment efficiency[J]. The Accounting Review, 2023, 98(2): 97-121.
- [18] Langberg N, Rothenberg N. Audit quality and investment efficiency with endogenous analyst information[J]. The Accounting Review, 2023, 98(4): 247-272.
- [19] 孔东民,陶云清.信息基础设施建设与企业投资——基于“宽带中国”试点政策的准自然实验[J].经济科学,2023(2):106-124.
- [20] 王化成,李雪晨,李海彤.数字创新与企业投资效率——基于专利文本分析的证据[J].会计研究,2023(7):55-71.

- [21] 刘亦文,谭慧中,陈熙钧,等. 数字经济发展对实体经济投资效率提升的影响研究[J]. 中国软科学,2022(10):20-29.
- [22] Jensen M. Agency costs of free cash flow, corporate finance, and takeovers[J]. American Economic Review, 1986, 76(2): 323-329.
- [23] Biddle G C, Hilary G, Verdi S R. How does financial reporting quality relate to investment efficiency? [J]. Journal of Accounting and Economics, 2009, 48(2-3): 112-131.
- [24] 袁明硕,魏成龙. 中小投资者保护制度创新与上市公司非效率投资缓解——基于投服中心持股行权的证据[J]. 证券市场导报,2024(8):37-47.
- [25] 王小鲁,胡李鹏,樊纲. 中国分省份市场化指数报告(2021)[M]. 北京:社会科学文献出版社,2021.
- [26] 陈丽莉,张若琪,戎珂. 数据要素赋能企业创新:基于内外部资源视角[J]. 管理评论,2024(12):15-25.
- [27] 丁杰,黄金波. 银企数字化促进企业绿色转型的协同效应研究[J]. 系统工程理论与实践,2024(1):102-122.
- [28] 罗进辉. 媒体报道的公司治理作用——双重代理成本视角[J]. 金融研究,2012(10):153-166.
- [29] 吴卫红,刘颖,张爱美. 股权激励能促进企业创新吗——基于激励对象和激励模式异质性的视角[J]. 会计研究,2024(4):98-111.
- [30] 牛华,连梦瑶,朱莉. 银行金融科技能延长企业信贷期限吗? ——基于银行信贷配给视角[J]. 财经研究,2025(1):154-168.
- [31] 张龙,张俊瑞,陈怡欣,等. 数据资产信息披露影响了资本市场资源配置效率吗? ——以权益资本成本为视角[J]. 管理工程学报,2026(4):63-76.
- [32] 孙忠娟,刘凯月,冯佳林,等. 如何实现数字竞争优势:“专精特新”企业数字投资与专业投资的协同逻辑[J]. 中国工业经济,2025(4):156-173.
- [33] 鲁晓东,连玉君. 中国工业企业全要素生产率估计:1999—2007[J]. 经济学(季刊),2012(2):541-558.

[责任编辑:高 婷]

Data Resources Information Disclosure Empowers the Investment Efficiency Improvement of Private Firms: Evidence from LLM-Based Textual Analysis

LUO Jinhui, SHEN Wenxuan, SONG Xinyue

(School of Management, Xiamen University, Xiamen 361005, China)

Abstract: In the era of digital economy, data have become a new factor of production that accelerates the cultivation and development of new quality productive forces. Using a sample of Chinese A-share listed private firms from 2010 to 2024, this study employs the FinBERT2 large language model to construct a measure of corporate data resource information disclosure and empirically examines its impact on the investment efficiency of private firms. The results show that data resource information disclosure significantly reduces inefficient investment and improves investment efficiency, and these findings remain robust across a series of robustness tests. The mechanism analysis indicates that data resource information disclosure enhances investment efficiency of private enterprises mainly by alleviating information asymmetry and mitigating agency problems. The heterogeneity analysis further shows that the positive effect of data resource information disclosure on investment efficiency is more pronounced among private firms operating in high-tech industries, smaller firms, firms with lower asset intensity, firms with higher levels of digital investment, and firms located in regions with higher levels of marketization. Further analysis suggests that the inhibitory effect of data resource information disclosure on inefficient investment contributes to improving the total factor productivity of private firms. These findings provide empirical evidence for understanding the economic consequences of data resources and related information disclosure, as well as the determinants of investment efficiency in private firms, and offer important policy implications for promoting the high-quality development of China's private economy.

Key Words: data resources; information disclosure; investment efficiency; private firms; large language model; high-quality development