

# 数据管理能力对企业生产率的影响效应

——来自中国企业—劳动力匹配调查的新发现

李 唐, 李 青, 陈楚霞

**[摘要]** 随着全球经济进入大数据时代,数据管理能力对于中国企业在风险预测、业务模式调整和价值发现等领域的重要性日益凸显。运用 2018 年中国企业—劳动力匹配调查数据,本文创新性地构建了数据管理能力得分指标,就数据管理能力对企业生产率的影响效应、内在机制进行了全面的实证检验。统计分析表明,与生产经营组织流程的管理效率相比,现阶段中国企业的数字管理能力明显偏低,已成为制约中国企业资源配置优化、生产率提高的重要瓶颈。研究发现,在其他条件不变的前提下,数据管理能力得分每提高 1 个标准差,企业全要素生产率平均增加 18.15%。异质性检验表明,数据管理能力的生产率效应对于信息不对称问题较为突出的企业分组相对偏高。进一步机制分析发现,数据管理能力将通过增强企业家风险预测能力、改善人力资本绩效激励、优化目标规划、提升企业创新绩效等 4 种中介渠道促进生产率增长。稳健性检验和工具变量估计表明,在有效处理遗漏变量、反向因果等内生性问题的前提下,数据管理能力对于企业生产率的影响效应满足因果推断的统计要求。本文的实证发现对大数据时代中国企业如何通过数字化转型实现生产率增长,具有重要的启示作用。

**[关键词]** 数据管理能力; 生产率; 中介效应; 工具变量; 中国企业—劳动力匹配调查

**[中图分类号]**F273 **[文献标识码]**A **[文章编号]**1006-480X(2020)06-0174-19

DOI:10.19581/j.cnki.ciejournal.2020.06.010

## 一、问题提出

近年来,随着全球经济迈入大数据时代,大数据作为互联网、物联网、3D 打印、云计算、人工智能等技术变革汇聚而成的颠覆性力量,正在对企业生产、运营和管理等产生深刻的重塑效应。为适应大数据所引致的中国经济发展基础性机制的变革,2015 年党的十八届五中全会首次提出国家大

**[收稿日期]** 2020-03-30

**[基金项目]** 国家社会科学基金青年项目“营商环境对中国企业全要素生产率影响的实证研究”(批准号 19CJL008);国家自然科学基金面上项目“‘一带一路’建设与中国制造:战略转型与价值链提升”(批准号 71573058)。

**[作者简介]** 李唐,广东外语外贸大学广东国际战略研究院副教授,经济学博士;李青,广东外语外贸大学广东国际战略研究院教授,博士生导师,管理学博士;陈楚霞,广东外语外贸大学广东国际战略研究院硕士研究生。通讯作者:李青,电子邮箱:gwliqing@163.com。本文所使用的 2018 年中国企业—劳动力匹配调查数据,得到中国社会科学院都阳教授、香港科技大学 Albert Park 教授和斯坦福大学李宏彬教授等项目发起人的授权。感谢匿名评审专家和编辑部提出的宝贵意见,当然文责自负。

数据战略,国务院印发《促进大数据发展行动纲要》,正式指出大数据是国家基础性战略资源,要加快建设数据强国。在此基础上,针对中国经济新常态所出现的增速趋缓、结构趋优、动能转换等新变化,2017—2019年政府工作报告连续提及大数据对于中国经济发展的重要作用,指出中国应实施大数据发展行动,通过“智能+”战略加快大数据与实体经济融合,为制造业转型升级赋能。因此,从微观企业视角入手,研究如何通过大数据应用实现企业生产率增长与价值链升级,具有重要的理论价值与现实意义。

从技术进步视角出发,现有文献就以信息与通信技术(ICT)、互联网和人工智能为代表的大数据技术对于企业生产率的影响效应和作用机理展开了较多关注。学者们认为,作为一项通用技术,大数据技术已摆脱了早期计算机技术单一维度进步所出现的“信息技术生产率悖论”问题(Milgrom and Roberts, 1990; Brynjolfsson and Hitt, 1996),其与实体经济的深入融合已对生产率增长产生了显著的促进作用(郭家堂和骆品亮, 2016)。具体而言,影响机制主要包括四个方面:①以ICT为代表的大数据技术可以有效促进工业设备数控化、生产工艺及生产流程的科学化,这将直接提高企业生产能力,促进企业投入—产出效率的改进(金碚, 2014)。②以互联网、ICT为代表的大数据技术可以通过技术外溢效应加快生产要素的重新组合(韩先锋等, 2014),推动企业生产组织流程的“精益管理”变革(Khuntia et al., 2018),提高企业生产制造的柔性化程度,从而实现企业生产范式的“熊彼特创新”(张三峰和魏下海, 2019)。③人工智能技术应用能够提高资本回报率从而加快资本积累,有效缓解劳动力成本上升所引致的生产率增速下行压力(陈彦斌等, 2019),并通过大数据技术与高技能人力资本的互补效应,实现全要素生产率增长(何小钢等, 2019)。④通过大数据技术的平台效应有效减少交易成本,促进创新网络深化(黄群慧等, 2019),并通过促进产业链分工水平提升减少资源错配(施炳展和李建桐, 2020)。

然而,大数据对于企业的影响绝不仅限于技术本身。越来越多的理论分析表明,随着大数据技术与实体经济日益广泛的融合,企业如何使用日趋庞大而多元的数据信息资源,已不再只是一个技术进步问题,也对企业管理实践创新提出了新的挑战。一方面,现有文献认为,随着数字化带来的变革,企业经营环境、主体行为、产品创造过程均发生了显著变化(Perrin and Jiang, 2018)。面对数字化程度的提高,企业管理需要从赋能向使能转变,即通过加强数据搜集、数据分享和数据分析在企业运营管理中的地位,实现大数据驱动的价值创新(陈剑等, 2020)。另一方面,部分文献提出,数字经济条件下,信息情境、决策主体、理念假设、方法流程等决策要素均呈现出显著变化,这将催生出新型的大数据管理决策范式。在此背景下,以数据资源综合应用为核心的决策范式将显著提高企业的行为洞察、风险预见和业务模式创新能力(陈国青等, 2020)。综上,已有研究的一个理论猜测在于,随着数字经济的深入,数据管理能力对于企业生产率提升的边际贡献或将更加凸显。然而,囿于微观实证数据的稀缺性,现有文献缺乏全面、准确测度企业数据管理能力的有效指标。除部分文献运用发达国家微观数据对上述问题做出了一定的实证分析外(Brynjolfsson et al., 2011; Brynjolfsson and McElheran, 2016a),来自中国的实证研究并不多见。

为此,本文运用2018年中国企业—劳动力匹配调查(China Employer—Employee Survey, CEES)数据,就数据管理能力对于企业生产率的影响效应、作用机理进行创新性的实证研究。本文的边际贡献主要有三点:①从决策数据可得性、决策数据依赖程度、企业数据搜集主体多样性、决策过程数据使用频率、工作过程数据使用频率、使用统计方法预测频率等6个维度出发,构建了能够全面反映数据管理能力的得分指标,有效规避了已有研究仅从单一维度研究数据管理问题的局限性(Bloom et al., 2013; Brynjolfsson et al., 2011);②运用多种识别策略,就数据管理能力对于企业生

生产率的影响效应进行了较为稳健的实证检验,从而丰富了现有文献对于管理实践如何影响企业生产率的理论认知(Bloom and Van Reenen,2007;Bloom et al.,2017);③从多个角度出发,研究数据管理如何影响企业生产率的内在机制,从而在一定程度上破解了数据管理生产率效应的“黑箱”。

本文剩余部分安排如下:第二部分是内在机制分析与研究假说,就数据管理能力对于企业生产率的影响效应和作用机理进行理论分析;第三部分是研究设计,分别对数据来源、描述性统计结果和基准回归模型进行介绍;第四部分是实证检验,首先基于基准回归方程报告数据管理能力对于企业生产率影响效应的估计结果,在此基础上,对数据管理能力的生产率效应分别进行异质性检验和稳健性检验;第五部分是进一步讨论,就数据管理能力对于企业生产率影响的内在机制进行检验,并运用工具变量法对内生性问题进行处理。最后是结论与政策启示。

## 二、内在机制分析与研究假说

作为数字经济时代一项重要的能力要素,本文认为数据管理能力不仅本身具有直接的生产率效应,而且在与其他生产要素存在互补性条件下将其引入企业生产活动,数据管理能力还将通过提高企业家风险预测能力、增强人力资本绩效激励作用、优化经营目标规划、提升创新绩效等影响渠道实现企业生产率增长。

### 1. 数据管理能力对于企业生产率的直接影响

作为一项重要的管理要素,提升数据管理能力对于企业生产率的直接影响主要体现在两方面:一方面,针对数字经济时代数据信息所具备的高速增长性、复杂性、价值稀疏性和功能多样性等特征(杨善林和周开乐,2015),企业基于数据驱动的决策能力对于企业边际产出的重要性日趋凸显(Brynjolfsson et al.,2011;Buffington et al.,2017)。在数字化商业情境下,随着实体与虚拟的不断融合,企业运营管理时间、空间和连接要素等将发生质的变化,管理决策、商业模式将积极调整以适应数字经济所引致的产品生命周期缩短、消费者需求长尾效应突出、供应链生态圈重构等新情况(陈剑等,2020)。在此背景下,只有在需求预测(黎波,2020)、产品设计(肖静华等,2018)、定价和库存管理(Huang and Van Mieghem,2014)、供应链管理(Kusiak,2017)等关键环节不断提高数据搜集、分析和应用能力,企业运营管理才能有效实现数字化赋能,从而把握住新的商业机会,实现价值创新(刘业政等,2020)。进一步地,部分研究发现:大数据时代数据资产存在显著的多源异构性,并不会“自动而必然”地促进企业生产率增长;为更好地运用数据资源,企业数据管理只有从数据间协同(Xu and Dukes,2019)、计算间协同(Lau et al.,2018)、分析间协同(Bertsimas and Kallus,2019)和人机间协同(Blumenstock,2018)等方面做出积极调整,才能创造出更高的商业价值(刘业政等,2020)。因此,根据上述文献的理论解释,数据管理能力对于企业生产率具有显著的促进效应。Brynjolfsson et al.(2011)运用麦肯锡公司2008年所做的一项针对数据管理能力的调查项目,研究发现,在给定物质资本、人力资本和IT使用状况的前提下,数据管理得分每提高1个标准差,以人均销售收入表征的企业劳动生产率将提高5%—6%。另一方面,数据管理能力具有和其他技术性要素相似的资源配置效应(Bloom et al.,2017)。通过对一系列大型跨国企业的案例分析,Davenport and Harris(2007)研究表明,无论对于传统制造业企业而言,还是对于金融集团或跨国零售企业,在其他条件相同情况下,数据获取能力和分析能力对于企业成功运营都十分重要。对于参与企业数据决策调查的32个跨国企业而言,主营业务收入的5年复合增长率与受访企业的数据获取能力、分析能力均存在显著的正相关性。Lavalle et al.(2010)研究发现,与同行业内数据信息分析能力较弱的竞争对手相比,在其他条件相同的情况下,对于数据信息分析能力较高的组织而言,其成长为行



业领军企业(Top Performers)的边际概率要高出2倍。运用一般均衡分析方法,部分研究进一步发现,在不确定性经济环境下,对于初始数据分析能力更强的企业而言,其具有更大的规模经济优势,能够更有效地吸引外部生产要素,实现要素集聚与规模收益递增(Hilbert,2017)<sup>①</sup>。综上,现有研究表明,对于数据管理能力更高的企业而言,竞争性优势与规模经济优势往往更为显著,其能通过优化资源配置促进生产率增长。

## 2. 数据管理能力对于企业生产率的其他影响渠道

除直接影响外,数据管理能力还将通过提高企业家风险预测能力、增强人力资本绩效激励作用、优化经营目标规划和提升创新绩效等渠道促进企业生产率提高。

(1)数字化商业情境下,数据管理对于降低供应链风险具有重要作用。随着决策数据可获得性与数据来源多样性的提高,企业家在供应链决策时将不再仅依据财务报表,而是可以通过分析企业内部数据、个人数据、政府数据、社交网络数据、第三方数据和利益相关者数据,从而更加全面地刻画供应商的内外风险(陈剑等,2020)。不仅如此,随着数字经济形势下消费者需求与过去相比变化速度更快、需求个性化特征更为明显,数据管理对于提高企业需求预测准确性的边际作用更为凸显。例如,除交易数据外,企业家还可以将客户在门户网站和电子商务平台的搜索、浏览、评价等数据进行综合分析,基于“涉入理论”(Involvement Theory)更为精准地预测消费者的多样化需求行为(He et al.,2019)。因此,现有文献分析表明,考虑到企业家有限的信息处理能力,数据管理能力能够通过解决信息不对称有效提高企业家经营风险的预测质量,减少不确定性对生产率增长的影响(Milgrom and Roberts,1990; Ben-David et al.,2013)。

(2)数据管理能够通过影响人力资本绩效激励实现企业生产率增长。运用厂商不完全契约理论,Brynjolfsson(1991)研究分析表明,随着技术进步所引致的企业信息资产(Information Asset)增长,企业管理组织将趋于去中心化(Decentralization);信息流程的改变将有效解决管理者与一线员工的信息不对称问题,降低委托—代理成本。基于实证数据,Cappelli(2017)研究发现,数据技术能够有效驱动人力资源管理水平提高。对于数据管理能力较高的企业而言,其能够更为便捷地发现员工的个体特征,从而对员工工作努力程度、离职倾向等问题进行动态检测,使员工薪酬、职位晋升与个体绩效的关系更为紧密。运用2010年管理组织实践调查数据,Brynjolfsson and McElheran(2016b)研究表明,2005—2010年,在考核监督领域运用数据驱动决策的美国制造业企业增长了3倍。其之所以能够成为备受企业欢迎的“最佳管理实践”(Best Practice),一个重要原因在于数据驱动决策可以有效改善绩效激励,从而使企业得以吸引更高质量人力资本,促进生产率增长。

(3)数据管理能够通过优化经营目标规划促进企业生产率提高。根据Bloom and Van Reenen(2007)、Buffington et al.(2017)等文献,目标规划是用于测度受访企业对于长期、短期经营目标设定准确程度的管理实践变量,其涉及企业对于未来生产经营的前瞻性分析以及当企业偏离经营目标之时的反馈优化。由于数据管理能力将影响企业用于决策、工作和风险预测的有效信息,对于数据管理能力较高的企业而言,其对于未来经营发展的前瞻性分析往往更为准确(Buffington et al.,2017)。随着数字经济的到来,稀缺的关键性资源不仅是数据本身,还包括处理和利用数据信息的能力。在相同的数据环境下,对于数据管理能力较高的企业,其能够将更多决策要素纳入目标制定过

① Bloom et al.(2013)研究发现,对于期初在考核监督领域应用更多数据驱动决策的企业而言,调查期间内的人均销售收入增速显著偏高。Brynjolfsson and McElheran(2016a)研究发现,考核监督领域的数字驱动决策对于企业生产率增速的影响存在异质性。对于存续年限较久、采用集团组织架构的企业分组而言,数据驱动决策对于人均销售收入增速的边际效应更大。

程之中,既能够提高目标制定过程的一致性与透明度,也能够通过科学化的统计方法对决策结果量化展示,在很多情形下避免目标制定者的主观认知偏差(Rahwan et al.,2019)。综上,数据管理能够有效促进管理决策全过程的信息分享,既包括经营目标的制定和选择,也涵盖效果跟踪、评估与反馈等其他关键环节,这将有效解决传统决策理论中决策者的“有限理性”问题,使目标规划不断优化,从而驱动企业生产率提高。

(4)数据管理能够通过提高创新绩效实现生产率增长。Brynjolfsson(2011)研究发现,数据驱动决策促进了创新模式的改善,使得企业创新业务活动能够实时动态追踪;并能通过创新想法更广泛、更便捷的分享机制,加快商业模式与生产流程的改进,这是美国企业自20世纪90年代中期以来生产率增长较快的重要原因。运用动态能力理论,郭海和韩佳平(2019)研究了数字化技术应用对于中国企业开放式创新的影响。研究表明,数字化技术的发展及其在企业管理实践中的有效应用,能够增强企业连接外部环境与整合外部知识的能力,使企业能够利用与外部环境主体的稳定连接与共同协作实现价值创新。与之相似,黄群慧等(2019)研究了互联网发展对于中国制造业生产率的影响机制。研究发现,城市互联网发展指数每提高1%,制造业企业全要素生产率会提高0.3%;内在机制检验表明,互联网技术发展能够通过降低交易成本、减少资源错配促进制造业分工,从而促进企业创新。考虑到互联网是数据管理的重要技术基础,上述文献从一个新的侧面印证了数据管理能够促进企业创新从而推动生产率增长。

综合以上分析,本文提出:

假说1:其他条件不变下,企业在生产运营中数据管理能力的提高,会促进生产率增长。

假说2:企业数据管理能力的提高将通过提高企业家风险预测能力、改善人力资本绩效激励、优化目标规划等途径提高资源配置效率,并通过促进企业创新实现生产率增长。

### 三、研究设计

#### 1. 数据来源

本文使用的数据来源于中国社会科学院、香港科技大学和斯坦福大学等科研机构联合开展的2018年CEES数据。以国家市场监督管理总局2018年中国工商企业年报系统的全部制造业企业为抽样总体,2018年CEES采用随机分层抽样方法,搜集了广东、湖北、江苏、四川和吉林五省<sup>①</sup>1978家制造业企业2015—2017年有关企业基本情况、CEO、生产、创新、销售与出口、营商环境、人力资源和财务绩效等维度的1030项调查指标;在此基础上,CEES按照30%中高层管理员工、70%其他员工的抽样比例,匹配性地搜集了受访企业共计15646名员工有关工资、奖金、教育、工作时长、工作经历、工作任务、社保福利和人格特征等不同维度的443项调查指标。更为重要的是,2018年CEES参照美国2015年管理组织实践调查的问卷设计思路,首次搜集了中国制造业企业有关决策过程数据可得性、决策过程数据依赖程度、企业数据搜集主体多样性、决策过程数据使用频率、工作过程数据使用频率、使用统计方法预测频率等6个维度的11项细分指标,获得1942个有效样本数据管理能力的相关测度<sup>②</sup>。在此基础上,2018年CEES还搜集了企业家对于2019年销售收入的经营风险预测指标,并获取了绩效激励、目标规划和新产品销售收入占比等内在机制检验的重要变量。

#### 2. 描述性统计

在进入计量设定与实证检验之前,本文运用2018年CEES数据,就中国企业的数据管理能力、

① 样本累计覆盖上述五省60个地级市、99个县区行政单位。

② 本文的数据管理能力得分指标为受访企业2015—2017年数据管理能力的平均值。

数据管理能力与企业生产率的相关性等问题进行初步的统计分析。表1给出了2018年CEES全部受访样本数据管理能力各维度的描述性统计结果。一方面,研究发现:按0—1得分计算的中国企业整体数据管理能力平均为0.4886,较Bloom et al.(2018)运用相同数据围绕组织管理的目标规划、绩效激励、考核监督和管理实施4个维度计算的管理得分(Management Score)的平均值0.6318要显著偏低22.67%。这表明,与中国企业的组织管理实践相比,数据管理能力更是制约中国企业管理提升的一个重要短板。另一方面,数据管理能力的分项指标统计表明,决策过程数据可得性(0.5861)、决策过程数据使用频率(0.5690)、决策过程数据依赖程度(0.5453)3个维度的均值要高于数据管理能力的整体均值;而工作过程数据使用频率(0.4742)、使用统计方法预测频率(0.4725)和企业数据搜集主体多样性(0.2843)3个维度的均值要低于数据管理能力的整体均值。分项统计表明,相较于决策过程的数据管理,数据搜集主体多样性、工作过程的数据使用、统计方法的使用和预测是造成中国企业数据管理能力偏低的重要因素。

在剔除省份与二位码行业差异的前提下,数据管理能力对于企业生产率具有显著的正向效应。本文遵循现有文献对于CEES数据的处理思路(Cheng et al., 2019a),采用固定效应模型所估计的对数线性C-D生产函数的残差项作为全要素生产率(TFP)自然对数值的代理指标<sup>①</sup>。在此基础上,图1给出了数据管理能力得分与TFP的散点拟合图<sup>②</sup>。测算发现,在控制省份、行业差异的前提下,数据管理能力得分对于TFP的半弹性系数为1.7478,在1%显著性水平上拒绝零假设。这表明,在其他条件相同情况下,数据管理能力得分每增加1个标准差(0.1534),企业TFP平均将提高26.81%(1.7478×0.1534)。因此,图1的统计分析表明,数据管理能力存在显著的生产率效应。其他条件不变的前提下,企业生产运营过程中的数据管理能力越高,TFP显著偏高。描述性统计结果支持了本文研究假设1。表2进一步给出了将2018年CEES数据调整为2015—2017年面板数据之后的主要变量描述性统计结果。

### 3. 计量模型设定

借鉴已有研究(Brynjolfsson et al., 2011; Brynjolfsson and MaElheran, 2016b),本文基准回归方程设定为:

$$\ln y_{ijdt} = \beta_0 + \beta_1 DM_{ijdt} + Z_{ijdt} \theta + \gamma_j + \gamma_d + \gamma_t + \varepsilon_{ijdt} \quad (1)$$

其中,被解释变量 $\ln y_{ijdt}$ 为第 $j$ 个行业、第 $d$ 个城市的第 $i$ 个企业在 $t$ 期以C-D生产函数残差项为表征的TFP自然对数值。核心解释变量 $DM_{ijdt}$ 为根据表1所列出6个维度、11个细分项指标算数平均值所计算的数据管理能力得分。考虑到数据管理能力得分随所有制、企业存续年限与出口特征不同所出现的统计差异,控制向量组 $Z_{ijdt}$ 分别包括所有制类型、存续年限和是否出口等变量。值得说明的是,基准回归方程(1)中的所有制变量分别为国有企业(State-owned Enterprises, SOE)、港澳台企业(Hong Kong/Taiwan/Macao-invested Firms, HTM)以及外资企业(Foreign)3类虚拟变量,对照组样本为民营企业(Private)。为进一步解决样本选择性偏误问题,方程(1)中还分别引入所处二位码行业、城市和年份的固定效应( $\gamma_j$ 、 $\gamma_d$ 和 $\gamma_t$ ),以控制企业样本不随时间变化的行业、地区特征以及不

① 在对数线性C-D生产函数模型中,被解释变量为受访企业增加值的自然对数值,解释变量分别为固定资产净值、劳动力人数的自然对数值。值得说明的是,由于2018年CEES涵盖了受访企业较为准确的资产负债表、成本利润表和现金流量表等相关会计指标,本文采用收入法将固定资产折旧、劳动者报酬、生产税净额、营业盈余等4个细分项指标相加以求出受访企业的增加值。

② 为更好地呈现数据结果,本文按Cheng et al.(2019b)的处理思路,将全部受访企业按数据管理能力得分划分为20分位,在此基础上制作散点拟合图。

表 1 企业数据管理能力指标的描述性统计

变量名称	样本量	平均值	标准差	最小值	最大值
决策过程数据可得性(0—1 比值)	1942	0.5861	0.2620	0.0000	1.0000
决策过程数据依赖程度(0—1 比值)	1942	0.5453	0.2374	0.0000	1.0000
企业数据搜集主体多样性(0—1 比值)	1942	0.2843	0.1830	0.1667	1.0000
决策过程数据使用频率(0—1 比值)	1942	0.5690	0.2539	0.0000	1.0000
来自生产技术/工具的绩效指标(0—1 比值)	1942	0.5972	0.2996	0.0000	1.0000
来自管理层的正式/非正式反馈(0—1 比值)	1942	0.5789	0.2782	0.0000	1.0000
来自一线工人的正式/非正式反馈(0—1 比值)	1942	0.5980	0.3205	0.0000	1.0000
来自企业外部数据(0—1 比值)	1942	0.5018	0.2734	0.0000	1.0000
工作过程数据使用频率(0—1 比值)	1942	0.4742	0.2570	0.0000	1.0000
新产品和服务的设计(0—1 比值)	1942	0.4398	0.2988	0.0000	1.0000
需求预测(0—1 比值)	1942	0.4837	0.2757	0.0000	1.0000
供应链管理(0—1 比值)	1942	0.4992	0.2947	0.0000	1.0000
使用统计方法预测频率(0—1 比值)	1942	0.4725	0.2974	0.0000	1.0000
数据管理能力得分(0—1 比值)	1942	0.4886	0.1534	0.0278	0.9583

资料来源:根据 2018 年中国企业—劳动力匹配调查数据进行整理。以下各表同。

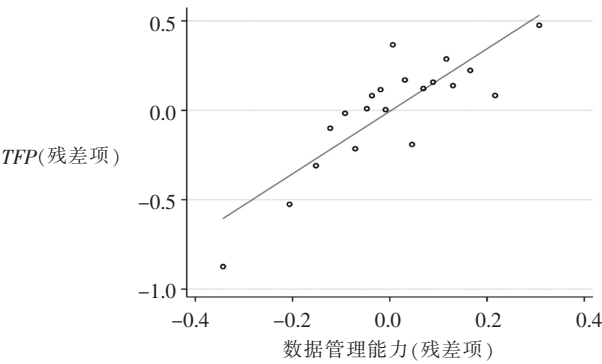


图 1 数据管理能力得分与 *TFP* 的线性拟合

注:线性拟合控制了省份和二位数行业固定效应。估计值(标准误)=1.7478(0.1702)。

表 2 2015—2017 年企业面板数据主要变量的描述性统计

变量名称	样本量	平均值	标准差	最小值	最大值
全要素生产率( <i>TFP</i> )	5322	0.0022	1.3605	-6.2925	6.7074
企业存续年限(年)	5826	11.6802	7.1331	0.0000	58.0000
出口企业(虚拟变量)	5733	0.4008	0.4901	0.0000	1.0000
国有企业( <i>SOE</i> ,虚拟变量)	5826	0.0911	0.2878	0.0000	1.0000
民营企业( <i>Private</i> ,虚拟变量)	5826	0.7523	0.4317	0.0000	1.0000
港澳台企业( <i>HTM</i> ,虚拟变量)	5826	0.0855	0.2796	0.0000	1.0000
外资企业( <i>Foreign</i> ,虚拟变量)	5826	0.0711	0.2569	0.0000	1.0000

随横截性单元变化的时间特征。 $\varepsilon_{ijt}$  为随机误差项。在方程(1)中, $\beta_1$  是本文最关心的参数,如果  $\beta_1$  在统计上显著为正,则表明企业在生产运营中提高数据管理能力,将有助于促进生产率提升。



## 四、实证检验

### 1. 基准回归结果

表3报告了对应回归方程(1)的基准回归估计结果。可以发现,数据管理能力对于企业 *TFP* 具有显著的正向效应。运用2018年CEES数据,表3第(1)列回归结果表明,在简单OLS回归前提下,数据管理能力对于 *TFP* 的半弹性系数估值为1.8582,并在1%显著性水平上拒绝零假设。这说明,在不考虑其他因素的前提下,数据管理能力得分每提高1个标准差(0.1534),企业 *TFP* 平均而言将提高28.50%。进一步地,表3第(2)、(3)列回归结果表明,依次将企业存续年限、出口、所有制类型等因素引入回归方程(1)之后,数据管理能力对于 *TFP* 的半弹性系数估值从1.8582分别下降为1.4986和1.3724,而计量模型的拟合优度则从0.0444分别提升到0.1391和0.1639。对比表3第(1)和第(3)两列的回归结果,这表明,受访企业在存续年限、出口、所有制类型等维度的异质性不仅能够有效解释数据管理能力对于 *TFP* 边际效应的26.14%,并且上述因素能够整体解释11.95%的企业间生产率差异。另外,表3第(4)列回归结果表明,引入行业、城市和年份固定效应之后,数据管理能力对于 *TFP* 的半弹性系数估值从1.3724下降到1.1834,仍在1%显著性水平上拒绝零假设。这表明,虽然行业、地区和年份等样本差异能够解释数据管理能力生产率效应的13.77%,但在充分引入其他条件的前提下,数据管理能力每提高1个标准差,企业 *TFP* 平均而言仍将显著提高18.15%。

基准回归结果表明,在引入企业存续年限、出口、所有制类型、行业、地区和年份特征的前提下,数据管理能力的生产率效应均显著为正,本文研究假设1得以基本证实。本文的基准回归结果与Brynjolfsson et al.(2011)、Brynjolfsson and McElheran(2016b)采用美国制造业微观数据所得出的结果基本一致;但对现阶段中国制造业而言,数据管理能力对于生产率的边际效应相对更高。对此,可能原因主要有两点:①由于数据限制,已有文献对于企业数据管理能力的分析仅能围绕单一维度(Brynjolfsson and McElheran,2016a),而本文能够从数据可得性、数据使用频率、搜集主体多样性、统计方法使用和预测等多个维度全面刻画企业数据管理能力,从而使得数据管理能力的测度更为准确。②考虑到边际收益递减规律的存在,由于中国制造业企业的数据管理能力与美国相比更为薄弱,这或将引致数据管理能力对于生产率边际效应相对偏高。因此,本文基准回归结果表明,对于当前中国经济而言,数据管理能力对于企业 *TFP* 的提升作用具有较强的经济显著性。

### 2. 异质性检验

已有研究表明,数据管理能力与生产率存在明显的企业异质性(Syverson,2011;Brynjolfsson and McElheran,2016b;Pugna et al.,2019)。为此,运用交互项回归,表4给出了数据管理能力对于企业生产率影响的异质性检验结果。本文研究发现:

(1)对于存续年限较长的企业,数据管理能力的生产率效应显著偏高。将全部受访企业按存续年限中位值(11年)划分为高、低两组,表4第(2)列回归结果表明,在其他条件相同的前提下,对于存续年限较短( $\leq 11$ 年)的企业分组而言,数据管理能力对于企业 *TFP* 的半弹性系数估值为0.7755。这表明,对于上述企业分组而言,数据管理能力得分每提高1个标准差(0.1587),企业 *TFP* 将提升12.31%。进一步地,表4第(2)列表明,对于存续年限较长( $>11$ 年)的企业,数据管理能力与存续年限较长分组虚拟变量的交互项分别为1.0369,且在1%显著性水平上拒绝零假设。这表明,对于存续年限较长的企业分组而言,数据管理能力每提高1个标准差(0.1470),数据管理能力对于 *TFP* 的影响效应较对照组偏高15.2424个百分点。因此,交互项回归表明,数据管理能力对于企业生产率的促进效应在存续年限较长的企业分组相对偏高。



表 3 数据管理能力得分对于企业生产率的影响效应(OLS 回归)

	TFP			
	(1)	(2)	(3)	(4)
数据管理能力得分	1.8582*** (0.1155)	1.4986*** (0.1100)	1.3724*** (0.1106)	1.1834*** (0.1144)
企业存续年限(自然对数值)		0.4183*** (0.0291)	0.3595*** (0.0293)	0.3317*** (0.0303)
出口企业		0.5498*** (0.0351)	0.4636*** (0.0363)	0.4498*** (0.0375)
国有企业			0.5062*** (0.0687)	0.5365*** (0.0699)
港澳台企业			0.3676*** (0.0633)	0.2897*** (0.0698)
外资企业			0.6937*** (0.0686)	0.4622*** (0.0673)
行业固定效应	否	否	否	是
城市固定效应	否	否	否	是
年份固定效应	否	否	否	是
样本量	5268	5268	5268	5268
R <sup>2</sup>	0.0444	0.1391	0.1639	0.2617

注:括号内数值为稳健标准误,\*\*\*、\*\*和\*分别表示在1%、5%和10%水平上统计显著。以下各表同。

(2)对于非出口企业而言,数据管理能力的生产率效应更具经济显著性。表4第(4)列回归结果表明,在其他条件相同的前提下,对于非出口企业而言,数据管理能力对于企业 *TFP* 的半弹性系数估值为 1.3468。这表明,对于非出口企业而言,数据管理能力得分每提高 1 个标准差(0.1650),企业 *TFP* 将增长 22.22%。进一步地,表4第(3)、(4)列回归结果表明,数据管理能力与出口企业分组的交互项均为负值(-0.3406 和 -0.4038),并在行业、地区和年份固定效应充分控制的前提下,上述系数在 10%显著性水平上拒绝零假设。这表明,对于出口企业而言,在存续年限、所有制类型、行业、地区和年份特征充分控制的前提下,数据管理能力每提高 1 个标准差(0.1333),数据管理能力对于 *TFP* 的影响效应较对照组企业偏低 5.3827 个百分点。因此,异质性检验表明,数据管理能力对于企业生产率的正向效应在非出口企业分组相对较高。

表4第(5)、(6)列回归结果表明,在数据管理能力与存续年限、出口企业分组的交互项同时引入回归方程的前提下,在引入行业、地区和年份固定效应前后,数据管理能力与存续年限较长企业分组的交互项系数均在 1%水平上显著为正,而数据管理能力与出口企业分组的交互项系数均在 5%水平上显著为负。这进一步证明,对于存续年限较长、非出口企业而言,数据管理能力的生产率效应更具经济显著性。异质性检验结果一定程度上与 Pugna et al.(2019)、Ben-David et al.(2013)等文献的理论推断相一致。这或许因为,对于存续年限较长的企业而非出口企业而言,不完全契约所引致的信息不对称问题更为突出,从而造成对于上述企业而言,数据管理能力升级需求更为迫切,数据管理能力对企业生产率的边际效应也相应更高。

3. 稳健性检验

由于 2018 年 CEES 数据中数据管理能力得分、所有制类型、出口等变量的横截性特征,基准回归方程(1)难以运用固定效应模型充分控制企业异质性因素,从而难以有效解决遗漏变量偏误对本文所关注的参数估计值( $\hat{\beta}_1$ )的潜在影响。参照已有研究的设计思路(Bloom et al.,2017;Bloom et al.,

表 4 数据管理能力得分对于企业生产率影响的异质性检验 (交互项回归)

	TFP					
	(1)	(2)	(3)	(4)	(5)	(6)
数据管理能力得分	1.0461*** (0.1416)	0.7755*** (0.1450)	1.5209*** (0.1331)	1.3468*** (0.1351)	1.1669*** (0.1518)	0.9176*** (0.1538)
数据管理能力×存续年限(>11 年)	0.8597*** (0.2225)	1.0369*** (0.2213)			0.9414*** (0.2271)	1.1265*** (0.2254)
数据管理能力×出口企业			-0.3406 (0.2329)	-0.4038* (0.2328)	-0.5146** (0.2386)	-0.5994** (0.2369)
企业存续年限(>11 年)	-0.0244 (0.1154)	-0.1558 (0.1150)	0.3976*** (0.0353)	0.3532*** (0.0355)	-0.0656 (0.1179)	-0.2017* (0.1171)
出口企业	0.4836*** (0.0366)	0.4714*** (0.0380)	0.6515*** (0.1204)	0.6672*** (0.1218)	0.7408*** (0.1238)	0.7719*** (0.1245)
国有企业	0.5524*** (0.0679)	0.5729*** (0.0691)	0.5527*** (0.0684)	0.5761*** (0.0697)	0.5564*** (0.0678)	0.5786*** (0.0690)
港澳台企业	0.3768*** (0.0632)	0.2885*** (0.0695)	0.3789*** (0.0632)	0.2949*** (0.0694)	0.3767*** (0.0631)	0.2901*** (0.0694)
外资企业	0.6964*** (0.0694)	0.4524*** (0.0678)	0.7054*** (0.0692)	0.4636*** (0.0678)	0.6962*** (0.0695)	0.4512*** (0.0679)
行业固定效应	否	是	否	是	否	是
城市固定效应	否	是	否	是	否	是
年份固定效应	否	是	否	是	否	是
样本量	5268	5268	5268	5268	5268	5268
R <sup>2</sup>	0.1596	0.2588	0.1576	0.2561	0.1603	0.2597

2018; Brandt et al., 2012), 本文分别通过引入更多控制变量、子样本回归等方法, 就数据管理能力对于企业生产率的影响效应进行稳健性检验。

(1) 引入管理者能力与公司治理的相关变量后, 数据管理能力对于企业生产率的影响效应仍然显著为正。根据已有研究(Bloom et al., 2017), 管理实践与管理者能力和公司治理等因素存在关联。如果数据管理能力单纯反映企业管理者能力或公司治理水平差异, 那么, 由于遗漏变量问题的存在, 基准回归方程(1)对于数据管理能力生产率效应的估计就存在“谬误回归”的风险。为此, 将全部受访企业按数据管理能力得分中位值(0.4861)划分为高、低两组, 表 5 给出了不同分组企业管理者能力、公司治理等相应指标上的统计差异。研究发现, 对于高数据管理能力企业分组而言, 管理者能力、公司治理均表现更优。具体而言, 对于高数据管理能力企业, CEO 获得大专以上学历的占比为 66.07%, 中高层管理人员的受教育年限平均为 13.1857 年, 分别较对照组企业高出 9.1500 个百分点和 0.3392 年。进一步地, 对于高数据管理能力企业, 有 54.51% 的企业设立了董事会, 有 45.21% 的企业设立了党组织, 上述两个指标较对照组企业分别高出 15.5400 和 13.8200 个百分点; 而在高数据管理能力企业分组中, 仅有 36.85% 为家族企业, 较对照组企业偏低 1.1600 个百分点。

表 6 第(1)—(4)列报告了将管理者能力、公司治理相关变量引入回归方程(1)之后的估计结果。表 6 第(2)、(3)两列估计结果表明, 在将管理者能力、公司治理相关指标分别引入回归方程之后, 数据管理能力对于 TFP 的半弹性系数估值从第(1)列的 1.0234 分别下降为 0.8913 和 0.6540, 但均在 1% 显著性水平上拒绝零假设。这表明, 虽然管理者能力、公司治理能够部分地解释数据管理能力对于 TFP 的影响效应, 但在分别控制上述因素的前提下, 数据管理能力对于企业 TFP 仍有显著为正的边际效应。进一步地, 将管理者能力、公司治理相关指标全部引入回归方程之后, 数据管理能

表 5 2015—2017 年不同分组企业管理者能力与公司治理指标的统计差异

变量名称	全部企业		高数据管理能力		低数据管理能力	
	样本量	平均值 (标准差)	样本量	平均值 (标准差)	样本量	平均值 (标准差)
管理者能力						
CEO 受教育程度(虚拟变量, 大专及以上学历)	5772	0.6143 (0.4868)	2847	0.6607 (0.4736)	2925	0.5692 (0.4953)
CEO 受教育程度(虚拟变量, 高中及以下学历)	5772	0.3857 (0.4868)	2847	0.3393 (0.4736)	2925	0.4308 (0.4953)
中高层管理员工受教育年限(年)	4938	13.0157 (2.4748)	2463	13.1857 (2.4090)	2475	12.8465 (2.5278)
公司治理						
董事会(虚拟变量)	5739	0.4663 (0.4989)	2829	0.5451 (0.4981)	2910	0.3897 (0.4878)
家族企业(虚拟变量)	5826	0.3744 (0.4840)	2874	0.3685 (0.4825)	2952	0.3801 (0.4855)
党组织(虚拟变量)	5706	0.3822 (0.4860)	2820	0.4521 (0.4978)	2886	0.3139 (0.4642)

力对于企业 *TFP* 的半弹性系数估值从第(1)列的 1.0234 下降为 0.5820,但仍在 1%水平上显著为正<sup>①</sup>。这表明,即使充分考虑管理者能力、公司治理对于数据管理能力和 *TFP* 的影响,数据管理能力对于企业 *TFP* 的边际效应仍有 56.87%无法被管理者能力、公司治理等变量所解释。

(2)剔除小微企业样本之后,数据管理能力对于企业生产率的影响效应仍然显著为正。根据已有文献(Syverson,2011; Brandt et al.,2012),小微企业不仅在投入—产出行为与生产率分布上与其他企业分组相比存在较大差异,而且组织管理、数据驱动决策等管理实践对于管理层级较为扁平的小微企业可能并不适用(Bloom et al.,2017)。因此,由于未能剔除上述样本,基准回归结果有可能就数据管理能力对于企业生产率的边际效应产生估计偏误。为此,参照现有研究 (Brandt et al., 2012)的通常做法,表 6 第(5)、(6)列将 2015—2017 年员工人数小于 20 人的企业样本予以剔除,在此基础上就数据管理能力对于企业 *TFP* 的影响效应进行稳健性检验。结果表明,在引入管理者能力、公司治理等变量前后,数据管理能力对于企业 *TFP* 的半弹性系数符号均为正,并始终在 1%统计水平上显著。进一步地,表 6 第(6)列回归结果表明,在基准回归控制向量组、管理者能力和公司治理等因素充分控制的前提下,数据管理能力每提高 1 个标准差(0.1448),企业 *TFP* 将平均提升 7.24%(0.5003×0.1448),仍具有较强的经济显著性。

综上,稳健性检验表明,在考虑更多影响因素的前提下,数据管理能力对于企业生产率仍然存在稳健的促进效应。在基准回归基础上,稳健性检验进一步证实了数据管理能力生产率效应的存在性,也使本文研究假说 1 得到更为细致的验证。

五、进一步讨论

1. 内在机制检验

研究假说 2 表明,在企业生产运营中,数据管理能力的提高将通过提高企业家风险预测能力、

① 参照现有研究(张三峰和魏下海,2019)的做法,本文对基准回归模型及稳健性检验模型均进行了方差膨胀因子(VIF)检验。结果表明,本文的核心解释变量及所有控制变量的 VIF 均小于多重共线临界标准 10。



表 6 稳健性检验

	TFP					
	引入管理者能力和公司治理				子样本回归	
	(1)	(2)	(3)	(4)	(5)	(6)
数据管理能力得分	1.0234*** (0.1264)	0.8913*** (0.1260)	0.6540*** (0.1255)	0.5820*** (0.1254)	0.8951*** (0.1390)	0.5003*** (0.1382)
CEO 大专以上学历		0.2058*** (0.0423)		0.1345*** (0.0420)		0.1566*** (0.0441)
中高层管理人员受教育年限 (自然对数值)		0.6007*** (0.1085)		0.4731*** (0.1092)		0.6590*** (0.1134)
董事会			0.3019*** (0.0400)	0.2855*** (0.0401)		0.2880*** (0.0412)
家族企业			-0.0498 (0.0421)	-0.0479 (0.0419)		-0.0504 (0.0454)
党组织			0.5245*** (0.0425)	0.4833*** (0.0429)		0.4291*** (0.0441)
其他控制变量	是	是	是	是	是	是
行业固定效应	是	是	是	是	是	是
城市固定效应	是	是	是	是	是	是
年份固定效应	是	是	是	是	是	是
样本量	4446	4446	4446	4446	3959	3959
R <sup>2</sup>	0.2482	0.2631	0.2915	0.2994	0.2201	0.2736

注:其他控制变量为企业存续年限、出口特征、所有制类型。

改善人力资本绩效激励、优化目标规划、促进创新等 4 种影响机制促进生产率增长。本文采用逐步回归法的中介效应模型,对上述 4 种机制进行检验。

将全部受访企业按数据管理能力得分中位值(0.4861)划分为高、低两组,表 7 给出了不同分组企业在风险预测能力、绩效激励、目标规划和企业创新等 4 个影响渠道的统计差异。2018 年 CEES 获取了受访企业家对于 2019 年销售收入从“最低”到“最高”5 种状态下的概率分布及其对应的预期数值,本文构建样本变异系数(标准差/期望收入)作为企业经营风险的不确定性指数。统计表明,高数据管理能力企业 2019 年销售收入的不确定性指数为 0.1421,较对照组企业(0.1513)偏低 6.08%。这表明,随着数据管理能力的提高,企业家风险预测能力将提升,从而降低企业经营风险的不确定性。进一步地,按照已有研究的处理思路(Bloom et al., 2017; Buffington et al., 2017),本文构建了受访企业在绩效激励、目标规划两个方面的管理实践得分指标。测算表明,对于高数据管理能力企业而言,中高层管理人员绩效激励、一线员工绩效激励和目标规划 3 项指标的得分分别为 0.6104、0.5998 和 0.6804,较对照组分别高出 5.48%、6.97%和 11.10%。这表明,对于数据管理能力较高的企业分组而言,人力资本绩效激励、目标规划均表现更优。此外,以 2015—2017 年受访企业至少在全国市场领先的新产品销售额占比<sup>①</sup>作为代理变量,分组统计表明,高数据管理能力企业全国市场领先的新产品销售额占比为 8.20%,较对照组企业(4.74%)高出 3.4600 个百分点,数据管理能力提升对于企业创新或具有一定促进作用。

为了对内在作用机制进行检验,本文借鉴温忠麟等(2004)等文献使用中介效应的方法构建递归方程,具体设定为方程(2)—(4),其中方程(2)与基准回归方程(1)基本相同,但在控制向量组

① 即对于上述调查年份而言,受访企业在全国市场、全球市场领先的新产品销售收入对于全部新产品销售收入的比值。本文简称为全国市场领先的新产品销售额占比。

表 7 2015—2017 年不同分组企业不确定性、绩效激励、目标规划与创新指标的统计差异

变量名称	全部企业		高数据管理能力		低数据管理能力	
	样本量	平均值 (标准差)	样本量	平均值 (标准差)	样本量	平均值 (标准差)
不确定性						
2019 年销售收入的不确定性指数(标准差/期望收入)	5298	0.1468 (0.1603)	2619	0.1421 (0.1625)	2679	0.1513 (0.1581)
绩效激励						
中高层管理人员的绩效激励得分(0—1 比值)	4776	0.5954 (0.1632)	2517	0.6104 (0.1602)	2259	0.5787 (0.1649)
一线员工的绩效激励得分(0—1 比值)	4575	0.5811 (0.1739)	2385	0.5998 (0.1720)	2190	0.5607 (0.1738)
目标规划						
目标规划得分(0—1 比值)	5604	0.6466 (0.1865)	2817	0.6804 (0.1773)	2787	0.6124 (0.1893)
企业创新						
全国市场领先的新产品销售额占比(0—1 比值)	5476	0.0642 (0.1702)	2659	0.0820 (0.1866)	2817	0.0474 (0.1513)

$Z_{ijdt}$  中进一步包括企业管理者能力、公司治理等用于稳健性检验的相关代理变量。

$$\ln y_{ijdt} = c_0 + \alpha_{01} DM_{ijdt} + Z_{ijdt} \lambda + \gamma_j + \gamma_d + \gamma_t + v_{ijdt} \quad (2)$$

$$M_{ijdt} = c_1 + \alpha_{11} DM_{ijdt} + Z_{ijdt} \kappa + \gamma_j + \gamma_d + \gamma_t + \mu_{ijdt} \quad (3)$$

$$\ln y_{ijdt} = c_2 + \alpha_{21} DM_{ijdt} + \alpha_{22} M_{ijdt} + Z_{ijdt} \delta + \gamma_j + \gamma_d + \gamma_t + \omega_{ijdt} \quad (4)$$

其中,变量  $M_{ijdt}$  表示中介变量,即本文为验证研究假说 2 所使用的不确定性指数、绩效激励、目标规划和企业创新等相关指标。 $v_{ijdt}$ 、 $\mu_{ijdt}$  和  $\omega_{ijdt}$  分别表示所对应回归方程的随机误差项。中介效应检验分为 3 个步骤:①对方程(2)进行回归,判断在不包括中介变量的前提下,数据管理能力对于企业生产率的总效应,如果  $\alpha_{01}$  显著,则表明数据管理能力对于企业生产率的总体效应存在;②对方程(3)进行回归,判断数据管理能力对于中介变量的影响效应( $\alpha_{11}$ );③在方程(2)中引入中介变量  $M_{ijdt}$  构成方程(4)进行回归,检验数据管理能力对于企业生产率的直接效应( $\alpha_{21}$ )和通过中介变量传导的中介效应( $\alpha_{22}$ )。如果  $\alpha_{11}$  和  $\alpha_{22}$  都显著,则表明中介效应存在;在此前提下,如果  $\alpha_{21}$  不显著,表明存在完全中介效应;如果  $\alpha_{21}$  显著,表明存在部分中介效应,中介效应占总效应的比重为  $\frac{\alpha_{11} \alpha_{22}}{\alpha_{21} + \alpha_{11} \alpha_{22}}$ 。进一步地,如果  $\alpha_{11}$  和  $\alpha_{22}$  中至少有一个不显著,则需要使用 Sobel 统计量进行二次检验,Sobel 中介效应检验对应的原假设为“中介效应不显著”。

前文表 6 第(4)列报告了中介效应检验的第 1 步估计结果。测算表明,在全部控制变量充分引入的前提下,数据管理能力对于企业  $TFP$  的半弹性系估值为 0.5820,且在 1%水平上显著为正。这表明,数据管理能力对于企业生产率的总体效应存在。进一步地,表 8 给出了中介效应检验的第 2 步估计结果。回归结果表明,在全部控制变量充分引入的前提下,数据管理能力对于风险预测能力<sup>①</sup>、中高层管理人员绩效激励、一线员工绩效激励、目标规划和全国市场领先的新产品销售额占比

① 为便于计算中介效应占总效应的比重,本文采用“10—不确定性指数”作为企业家风险预测能力的代理指标。该数值越大,代表受访企业未来经营的不确定性程度较低,企业家风险预测能力越强。

表 8 数据管理能力得分对于中介变量的影响效应

	10—不确定性 指数	绩效激励 (中高层)	绩效激励 (一线员工)	目标规划	全国市场领先的新 产品销售额占比
	(1)	(2)	(3)	(4)	(5)
数据管理能力得分	0.0443** (0.0186)	0.0534** (0.0219)	0.0934*** (0.0234)	0.2148*** (0.0268)	0.0589** (0.0236)
其他控制变量	是	是	是	是	是
行业固定效应	是	是	是	是	是
城市固定效应	是	是	是	是	是
年份固定效应	是	是	是	是	是
样本量	2889	2889	2889	2889	2889
R <sup>2</sup>	0.1831	0.1535	0.1621	0.2428	0.1890

注:其他控制变量为企业存续年限、出口特征、所有制类型、中高层管理人员受教育年限、董事会等。以下各表同。

等中介变量的影响效应均在至少 5% 显著性水平上拒绝零假设,并满足内在机制分析的预期符号推断。上述回归结果支持了数据管理能力对于中介变量影响效应的存在性。

表 9 报告了中介效应检验的第 3 步估计结果。一方面,将所有中介变量依次引入回归方程(4)后,估计结果表明,中介变量对于 *TFP* 的影响系数均在至少 10% 水平上显著为正。结合第 2 步估计结果,回归方程(3)和(4)的待估参数  $\alpha_{11}$  和  $\alpha_{22}$  均显著拒绝零假设,支持了上述内在机制中介效应的存在性。进一步地,表 9 的第 3 步估计结果表明,在上述中介变量依次引入回归方程(4)的前提下,数据管理能力对于企业 *TFP* 的直接效应( $\alpha_{21}$ )均在 1% 水平上显著为正,这说明中介效应只能部分地解释总体效应。第 3 步估计结果检验了上述内在机制中介效应占总体效应的比例。测算表明,对于风险预测能力而言,中介效应占总体效应的比例为 4.38%;对于中高层管理人员、一线员工的绩效激励而言,中介效应占比分别为 2.06% 和 3.76%。对于目标规划、企业创新而言,上述效应占比分别为 8.51% 和 5.00%。

综上,中介效应检验表明,数据管理能力确实通过提高企业家风险预测能力、优化人力资本绩效激励、改进目标规划、促进企业创新等 4 种影响机制促进生产率增长。本文研究假说 2 获得了实证检验的有力支持。在这 4 种内在机制中,目标规划的边际作用更为明显。作为面向数字化商业情境的一项重要管理能力,数据管理能力的提升对于解决企业生产经营信息不对称问题的边际贡献更为突出。这将表现在企业管理层对未来经营目标前瞻性预判能力的提高 (Buffington et al., 2017)。

## 2. 内生性问题的处理

由于核心解释变量的横截性特征,基准回归方程(1)难以引入企业固定效应,充分控制企业异质性因素对于数据管理能力和企业生产率指标的影响。并且,核心解释变量的横截性特征也造成本文难以借鉴 Bloom et al. (2013) 和 Brynjolfsson and McElheran (2016a) 的识别策略,运用数据管理能力的滞后项作为核心解释变量,从而有效规避反向因果所引发的内生性问题。为此,本文将运用工具变量法对内生性问题做出有效处理,从而对数据管理能力对于企业生产率的影响效应进行更为稳健的因果推断。

本文借鉴 Lewbel (1997) 的研究思路,采用数据管理得分与按二位码行业和省份分类的数据管理得分均值差额的三次方作为工具变量 (Lewbel IV)。首先进行数据管理能力对于工具变量的第一



表 9 数据管理能力得分影响企业生产率的机制检验

	TFP				
	(1)	(2)	(3)	(4)	(5)
数据管理能力得分	0.6597*** (0.1609)	0.6757*** (0.1614)	0.6640*** (0.1616)	0.6312*** (0.1636)	0.6554*** (0.1609)
10—不确定性指数	0.6825*** (0.1518)				
绩效激励(中高层)		0.2653* (0.1380)			
绩效激励(一线员工)			0.2775** (0.1276)		
目标规划				0.2732** (0.1291)	
新产品销售额占比(全国领先)					0.5850*** (0.1259)
其他控制变量	是	是	是	是	是
行业固定效应	是	是	是	是	是
城市固定效应	是	是	是	是	是
年份固定效应	是	是	是	是	是
中介效应/总体效应	0.0438	0.0206	0.0376	0.0851	0.0500
样本量	2889	2889	2889	2889	2889
R <sup>2</sup>	0.3361	0.3322	0.3324	0.3324	0.3364

阶段回归,并将其中数据管理能力的拟合值引入回归方程(1)<sup>①</sup>,以此作为第二阶段回归数据管理能力的代理指标,估计其对于企业生产率的半弹性系数。如果第一阶段回归中  $IV$  对于数据管理能力的影响系数具有统计显著性,且  $F$  统计量大于经验规则 10,表明本文选取的工具变量不存在弱工具变量问题。在此基础上,如果第二阶段回归中本文感兴趣的参数估计值( $\hat{\beta}_1$ )显著为正,并且方程(5)的半简化式回归符合排他性约束条件,则表明:本文寻找的工具变量是合理的,而且数据管理能力对于企业生产率具有显著为正的因果效应。

$$\ln y_{ijt} = \pi_0 + \pi_1 DM_{ijt} + \pi_2 IV_{ijt} + Z_{ijt} \eta + \gamma_j + \gamma_d + \gamma_t + \xi_{ijt} \quad (5)$$

值得注意的是,方程(5)的验证思路在于,如果工具变量满足排他性约束要求,工具变量将与基准计量模型的扰动项不相关。因此,将工具变量加入基准计量模型进行回归,其估计系数将会表现为无显著影响(孙圣民和陈强,2017;蒲艳萍和顾冉,2019)。

表 10 报告了工具变量法的实证检验结果。其中,表 10 第(1)列给出了第一阶段回归的估计结果。研究发现, $Lewbel IV$  与数据管理能力得分显著正相关,且在 1%显著性水平上拒绝零假设;并且,第一阶段的  $F$  统计量远大于经验规则 10。这说明,本文所选择的工具变量不存在弱工具变量问题。表 10 第(2)列给出了第二阶段回归的估计结果。测算表明,第二阶段回归中数据管理能力得分对于  $TFP$  的半弹性估计值为 0.7346,在 1%显著性水平上拒绝零假设。这说明,在其他因素不变的前提下,对于受到工具变量有效干预的企业样本而言,数据管理能力每提高 1 个标准差(0.1534),企业  $TFP$  将平均提高 11.27%。表 10 的第(3)列给出了半简化式方程(5)的回归结果。研究发现,将  $Lewbel IV$  直接引入基准回归方程之后,该工具变量对于企业  $TFP$  的影响系数无法拒绝零假设。这说明,工具变量与基准回归方程扰动项不相关,该变量除通过影响数据管理能力而对企业生产率存

① 考虑到稳健性检验对于遗漏变量问题的处理思路,控制向量组  $Z_{ijt}$  除包括所有制类型、存续年限、是否出口等变量外,还包括了企业管理者能力、公司治理等相关代理变量。

在作用之外,并不存在影响企业生产率的其他途径。工具变量符合排他性约束条件要求。

本文的内生性处理结果表明,考虑到工具变量的有效性,数据管理能力对于企业生产率存在正向效应。

表 10 数据管理能力得分对于企业生产率影响的工具变量估计

	TFP		
	第一阶段	IV	半简化式
	(1)	(2)	(3)
数据管理能力得分		0.7346*** (0.1547)	0.3828** (0.1954)
Lewbel IV	8.5791*** (0.2889)		3.0183 (2.1154)
其他控制变量	是	是	是
F 统计量	38.1000		22.3100
样本量	4446	4446	4446
R <sup>2</sup>	0.6439	0.2992	0.2997

## 六、结论与政策启示

运用 2018 年 CEES 数据,本文首次根据受访企业在决策数据可得性、决策数据依赖程度、企业数据搜集主体多样性、决策过程数据使用频率、工作过程数据使用频率、使用统计方法预测频率 6 个维度、11 个细分项指标构建了数据管理能力得分,就数据管理能力对于企业生产率的影响效应、作用机理展开了稳健的实证分析。本文主要研究发现如下:

(1)基准回归表明,在控制企业存续年限、出口特征、所有制类型、行业地区和年份固定效应等因素的前提下,数据管理能力对于企业全要素生产率具有显著的促进效应。在其他因素不变的前提下,数据管理能力每提高 1 个标准差(0.1534),企业全要素生产率平均增加 18.15%。稳健性检验表明,无论引入管理者能力和公司治理等更多控制因素,还是剔除存在异方差可能的小微企业样本,数据管理能力对于企业生产率的影响效应始终显著为正。上述回归结果证实了数据管理能力生产率效应的存在性。

(2)异质性检验表明,数据管理能力对于企业生产率的影响效应在存续年限较久企业分组和非出口企业分组相对偏高。考虑到存续年限较久、非出口企业的信息不对称问题可能更为突出,这是造成上述企业数据管理能力对于生产率边际效应偏高的重要原因。进一步地,中介效应检验表明,随着数据管理能力的提高,企业家风险预测能力、人力资本绩效激励、目标规划和企业创新均显著增强,上述因素均为数据管理能力对于企业生产率产生影响的中介渠道。同时,作为面向数字化商业情境的一项重要管理能力,数据管理能力提升对于解决企业生产经营信息不对称的边际作用更为凸显,这表现为目标规划在上述 4 种影响渠道之中存在相对更高的中介效应。

(3)运用工具变量有效解决基准回归的内生性问题,估计结果发现,数据管理能力对于企业生产率仍然具有显著的促进效应。在其他因素不变前提下,对于受到 Lewbel IV 有效干预的企业样本而言,数据管理能力每提高 1 个标准差,企业全要素生产率平均提高 11.27%。结合稳健性检验和工具变量估计结果,数据管理能力对于企业生产率的影响效应满足因果推断的统计要求。

本文的政策启示如下:①随着中国经济进入大数据时代,除互联网、ICT 和人工智能技术对企业生产率具有促进作用外,数据管理的生产率红利也日趋凸显。中国企业应针对数字经济的商业环境变化,使数据管理真正成为企业决策与工作过程的重要环节,增强企业应对不确定性的能力,不断

提高资源配置效率。<sup>②</sup>在数字化转型的经济背景下,中国企业的管理实践创新将发生重要变化。在新经济环境下,管理实践应不再局限于对企业生产经营组织流程的效率挖潜,而是要通过对多源异构数据的有效应用,通过数据驱动决策更为主动地实现需求预测和价值创造。考虑到人力资本、信息通信技术和互联网发展对于数据管理能力的互补性作用,企业在构建契合自身发展实际的数据管理决策模式的同时,应重视通过引进高质量人力资本、加快生产工艺和技术设备的智能化改造优化要素禀赋,使数据管理能力对于企业生产率增长的促进效应得到更大发挥。<sup>③</sup>政府应充分认识到企业家管理创新意识不足是阻碍现阶段中国企业数据管理能力提升的重要原因。应在人工智能扶持基金中列支专项资金,针对企业家群体开展数字化转型与管理实践创新的培训计划,并组织企业家赴“典型企业”深入学习交流,使代表性企业数据管理的成功经验能够在产业链各环节得到有效扩散。通过上述举措,不断激发企业家的数据管理创新意识,助推中国企业加快数字化转型。

值得说明的是,鉴于数据管理对于中国企业数字化转型的重要性,笔者将以数据管理能力与传统组织管理的相互关系、人工智能技术进步对于企业数据管理创新的影响效应和作用机理、企业数据管理能力与人力资本质量的异质性问题作为今后研究方向,围绕中国企业如何通过提升数据管理能力实现生产率增长,开展更为深入的理论与实证研究。

#### 〔参考文献〕

- [1]陈国青,曾大军,卫强,张明月,郭迅华. 大数据环境下的决策范式转变与使能创新[J]. 管理世界, 2020,(2):95-105.
- [2]陈剑,黄朔,刘运辉. 从赋能到使能——数字化环境下的企业运营管理[J]. 管理世界, 2020,(2):117-128.
- [3]陈彦斌,林晨,陈小亮. 人工智能、老龄化与经济增长[J]. 经济研究, 2019,(7):47-63.
- [4]郭海,韩佳平. 数字化情境下开放式创新对新创企业成长的影响:商业模式创新的中介作用[J].管理评论, 2019,(6):186-198.
- [5]郭家堂,骆品亮. 互联网对中国全要素生产率有促进作用吗[J]. 管理世界, 2016,(10):34-49.
- [6]韩先锋,惠宁,宋文飞. 信息化能提高中国工业部门技术创新效率吗[J]. 中国工业经济, 2014,(12):70-82.
- [7]何小钢,梁权熙,王善骞. 信息技术、劳动力结构与企业生产率——破解“信息技术生产率悖论”之谜[J]. 管理世界, 2019,(9):65-80.
- [8]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]. 中国工业经济, 2019,(8):5-23.
- [9]金碚. 工业的使命和价值——中国产业转型升级的理论逻辑[J]. 中国工业经济, 2014,(9):51-64.
- [10]黎波. 大数据环境下的因果学习与稳定预测[A]. 陈剑等. 大数据环境下的运营策略优化与协调研究[C]. 北京:科学出版社, 2020.
- [11]刘业政,孙见山,姜元春,陈夏雨,刘春丽. 大数据的价值发现:4C模型[J]. 管理世界, 2020,(2):129-138.
- [12]蒲艳萍,顾冉. 劳动力工资扭曲如何影响企业创新[J]. 中国工业经济, 2019,(7):137-154.
- [13]施炳展,李建桐. 互联网是否促进了分工:来自中国制造业企业的证据[J]. 管理世界, 2020,(4):130-149.
- [14]孙圣民,陈强. 家庭联产承包责任制与中国农业增长的再考察——来自面板工具变量法的证据[J]. 经济学(季刊), 2017,(2):815-832.
- [15]温忠麟,张雷,侯杰泰,刘红云. 中介效应检验程序及其应用[J]. 心理学报, 2004,(5):614-620.
- [16]肖静华,吴瑶,刘意,谢康. 消费者数据化参与的研发创新——企业与消费者协同演化视角的双案例研究[J]. 管理世界, 2018,(8):154-173.
- [17]杨善林,周开乐. 大数据中的管理问题:基于大数据的资源观[J]. 管理科学学报, 2015,(5):1-8.
- [18]张三峰,魏下海. 信息与通信技术是否降低了企业能源消耗——来自中国制造业企业调查数据的证据[J]. 中国工业经济, 2019,(2):155-173.
- [19]Ben -David, I., J. R. Graham, and C. R. Harvey. Managerial Miscalibration [J]. Quarterly Journal of



- Economics, 2013,128(4):1547–1584.
- [20]Bertsimas, D., and N. Kallus. From Predictive to Prescriptive Analytics[J]. Management Science, 2019,66(3):1–57.
- [21]Bloom, N., and J. Van Reenen. Measuring and Explaining Management Practices across Firms and Countries[J]. Quarterly Journal of Economics, 2007,122(4):1351–1408.
- [22]Bloom, N., E. Brynjolfsson, L. Foster, R. Jarmin, I. Saporta-Eksten, and J. Van Reenen. Management in America[R]. CES Working Paper, 2013.
- [23]Bloom, N., H. Cheng, M. Duggan, H. Li, and F. Qian. Do CEOs Know Best? Evidence from China[R]. NBER Working Paper, 2018.
- [24]Bloom, N., R. Sadun, and J. Van Reenen. Management as a Technology[R]. NBER Working Paper, 2017.
- [25]Blumenstock, J. Don't Forget People in the Use of Big Data for Development [J]. Nature, 2018, 561(7722): 170–172.
- [26]Brandt, L., J. Van Biesebroeck, and Y. Zhang. Creative Accounting or Creative Destruction? Firm-level Productivity Growth in Chinese Manufacturing[J]. Journal of Development Economics, 2012,97(2):339–351.
- [27]Brynjolfsson, E. An Incomplete Contracts Theory of Information, Technology and Organization [R]. MIT Center for Coordination Science, 1991.
- [28]Brynjolfsson, E. ICT, Innovation and the E-Economy[J]. EIB Papers, 2011,16(2):60–76.
- [29]Brynjolfsson, E., and K. McElheran. Data in Action: Data-Driven Decision Making in US Manufacturing[R]. Social Science Research Network, 2016a.
- [30]Brynjolfsson, E., and K. McElheran. The Rapid Adoption of Data-Driven Decision-Making [J]. American Economic Review, 2016b,106(5):133–39.
- [31]Brynjolfsson, E., and L. Hitt. Paradox Lost? Firm-Level Evidence on the Returns to Information Systems Spending[J]. Management Science, 1996, 42(4):541–558.
- [32]Brynjolfsson, E., L. M. Hitt, and H. H. Kim. Strength in Numbers: How Does Data-Driven Decision Making Affect Firm Performance[R]. Social Science Research Network, 2011.
- [33]Buffington, C., L. Foster, R. Jarmin, and S. Ohlmacher. The Management and Organizational Practices Survey (MOPS): An Overview[J]. Journal of Economic and Social Measurement, 2017, 42(1):1–26.
- [34]Cappelli, P. There's No Such Thing as Big Data in HR[J]. Harvard Business Review, 2017, (2):2–4.
- [35]Cheng, H., H. Li, and T. Li. The Performance of State-Owned Enterprises: New Evidence from the China Employer-Employee Survey [EB/OL]. Economic Development and Cultural Change, <https://www.journals.uchicago.edu/doi/abs/10.1086/703100?mobileUi=0>, 2019a.
- [36]Cheng, H., R. Jia, D. Li, and H. Li. The Rise of Robots in China [J]. Journal of Economic Perspectives, 2019b,33(2):71–88.
- [37]Davenport, T. H., and J. G. Harris. Competing on Analytics: The New Science of Winning [M]. Cambridge: Harvard Business Press, 2007.
- [38]He J., X. Fang, H. Liu, and X. Li. Mobile App Recommendation: An Involvement-Enhanced Approach[J]. MIS Quarterly, 2019,43(3):827–849.
- [39]Hilbert, M. The More You Know, the More You Can Grow: An Information Theoretic Approach to Growth in the Information Age[J]. Entropy, 2017,19(2):1–21.
- [40]Huang, T., and J. A. Van Mieghem. Clickstream Data and Inventory Management: Model and Empirical Analysis[J]. Production and Operations Management, 2014,23(3):333–347.
- [41]Khuntia, J., T. Saldanha, S. Mithas, and V. Sambamurthy. Information Technology and Sustainability: Evidence from an Emerging Economy[J]. Production and Operations Management, 2018,27(4):756–773.
- [42]Kusiak, A. Smart Manufacturing Must Embrace Big Data[J]. Nature, 2017, 544(7648):23–25.

- [43]Lau, R. Y. K., W. Zhang, and W. Xu. Parallel Aspect-oriented Sentiment Analysis for Sales Forecasting with Big Data[J]. *Production and Operations Management*, 2018,27(10):1775-1794.
- [44]Lavalle, S., M. S. Hopkins, E. Lesser, R. Shockley, and N. Kruschwitz. Analytics: The New Path to Value[J]. *MIT Sloan Management Review*, 2010,52(1):1-25.
- [45]Lewbel, A. Constructing Instruments for Regressions with Measurement Error When No Additional Data Are Available, with An Application to Patents and R&D[J]. *Econometrica*, 1997,65(5):1201-1213.
- [46]Milgrom, P., and J. Roberts. The Economics of Modern Manufacturing: Technology, Strategy, and Organization[J]. *American Economic Review*, 1990,80(3):511-528.
- [47]Perrin, A., and J. Jiang. About a Quarter of U.S. Adults Say They Are “Almost Constantly” Online [EB/OL]. <https://www.kyforward.com/pew-research-center-about-a-quarter-of-u-s-adults-say-they-are-almost-constantly-online>, 2018.
- [48]Pugna, I. B., A. Dutescu, and O. G. Stanila. Corporate Attitudes towards Big Data and Its Impact on Performance Management: A Qualitative Study[J]. *Sustainability*, 2019,11(3):1-26.
- [49]Rahwan, I., M. Cebrian, N. Obradovich, J. Bongard, J. F. Bonnefon, C. Breazeal, and M. O. Jackson. Machine Behaviour[J]. *Nature*, 2019,568(7753):477-486.
- [50]Syverson, C. What Determines Productivity[J]. *Journal of Economic Literature*, 2011,49(2):326-65.
- [51]Xu, Z., and A. Dukes. Product Line Design Under Preference Uncertainty Using Aggregate Consumer Data[J]. *Marketing Science*, 2019,38(4):669-689.

## The Effect of Data Management Ability on Firm Productivity ——New Evidence from China Employer-Employee Survey

LI Tang, LI Qing, CHEN Chu-Xia

(Guangdong Institute for International Strategies of GDUFS, Guangzhou 510000, China)

**Abstract:** As the global economy enters the era of big data, the importance of data management ability for risk prediction, business model adjustment and value discovery has become increasingly prominent for China's firms. Drawing on the China employer-employee survey data collected in 2018, this paper first creates the measures of data management and robustly estimates the effect and influence mechanism of data management ability on firm productivity. Compared to the management efficiency in production organization, our statistical analyses reflect the fact that the data management score of Chinese firms is significantly low, which have already restricted the resource allocation and productivity improvement in China. Holding other conditions constant, the basic regression results present that TFP will increase by 18.15% respectively as the data management score increase by one standard deviation. The heterogeneity test shows that the effect of data management ability on firm productivity is relatively positive for older and non-export firms. Mediation test supports hypotheses that data management ability can promote productivity growth by improving risk prediction capabilities of entrepreneurs, performance incentives, goal programming and innovation. Based on robustness check and instrument variable estimations, this paper finds that data management has positive and significant causal effect on firm productivity. These findings provide important implications for Chinese firms to encourage productivity growth by management innovation in this age of digital economy.

**Key Words:** data management ability; productivity; mediation effect; instrumental variable; China employer-employee survey

**JEL Classification:** D21 M11 O14

[责任编辑:李鹏]