

# 大数据技术应用能抑制企业避税吗

## ——基于国家级大数据综合试验区的分析

何涌 吴腾

(湖南工业大学经济与管理学院, 湖南株洲 412007)

**摘要:**基于国家级大数据综合试验区试点政策,利用2013—2022年沪深A股上市公司数据,采用多期DID模型,研究了大数据技术应用对企业避税的影响。研究表明,大数据技术应用显著抑制了企业避税,该结论在经过PSM-DID、替换被解释变量、排除政策干扰等一系列稳健性检验后依然成立;机制分析发现,大数据技术应用有助于提升信息透明度和企业生产效率进而抑制了企业避税。异质性分析表明,大数据技术应用对东部地区和非高科技企业避税抑制效果更为显著。本文的研究结论不仅丰富了企业避税影响因素的相关研究,而且揭示了国家级大数据综合试验区政策所产生的积极治理效应,为后续持续推进大数据技术应用提供理论支持与经验证据。

**关键词:**大数据技术;企业避税;信息透明度;企业生产效率;多期DID

**中图分类号:**F273 **文献标识码:**A **文章编号:**1003-5230(2024)05-0057-13

### 一、引言

随着大数据、人工智能和区块链等新兴技术的迅速发展,数据已经成为数字经济时代的基础性和战略性资源。企业避税问题一直都是政府治理的难题,避税行为不仅影响了公共财政的正常运转,也破坏了公平竞争的市场环境。传统税收征管模式存在税收边界难以识别、信息不对称以及课税对象不明确等问题<sup>[1]</sup>,然而,在数字经济时代,数据要素的非竞争性、规模经济和低复用成本,使得企业和税务部门能够通过大数据技术实现对经济活动的系统精准监控。在此背景下,探讨大数据技术应用对于我国税收治理的影响具有重要的理论和现实意义。

为全面推进我国大数据的发展与应用,2015年9月国务院颁布了《促进大数据发展行动纲要》,率先在贵州等地区开展大数据相关试点工作,探索数据要素与实体产业融合的有效路径。2016年10月,国家发展和改革委员会(以下简称“国家发展改革委”)将京津冀、广东、河南、上海、内蒙古、重庆及沈阳纳入第二批国家级大数据综合试验区建设名单。国家级大数据综合试验区(简称“大数据综合试

收稿日期:2024-04-14

基金项目:国家社会科学基金一般项目“金融科技的创新扩散、风险溢出与包容性监管研究”(22BJY109)

作者简介:何涌(1974—),男,湖南宁乡人,湖南工业大学经济与管理学院教授;

吴腾(1996—),男,安徽池州人,湖南工业大学经济与管理学院博士生。

点”)的设立对我国大数据技术应用具有重要意义,标志着大数据技术从政策制定向实际应用的跨越。该试点政策的主要任务包括大数据基础设施建设、推动大数据在企业研发和经营各环节的应用、重塑商业模式及产业链体系等。具体而言,试点政策通过统筹规划数字基础设施建设推动数字技术与企业生产经营的深度融合,强化大数据创新应用能力<sup>[2]</sup>。综上所述,大数据试点政策作为大数据技术应用的重要体现,有利于进一步提升我国数字经济发展水平。

现有研究大多从经济增长、绿色发展和企业成长等角度研究大数据综合试点的政策影响。从经济增长的角度来看,大数据试点政策提升了区域全要素生产率<sup>[3]</sup>,通过生产效率的提升显著推动经济高质量发展<sup>[4]</sup>。在绿色发展方面,大数据试点政策通过技术创新、产业升级、资源配置等渠道促进低碳经济转型<sup>[5]</sup>。在企业成长方面,相关文献则较多关注于大数据试点对企业全要素生产率<sup>[6]</sup>的影响。然而,关于大数据综合试点对企业避税的影响尚无清晰回答。在数字经济不断发展、数字技术水平不断提升的背景下,大数据综合试点的设立将如何影响企业避税行为?这种关系是否会因企业特性差异而变化?为此,本文基于国家级大数据综合试验区的外生冲击作为准自然试验,采用双重差分法探讨大数据技术应用对企业避税的影响及其理论机制,以期回答上述问题。

## 二、文献综述

### (一)大数据的经济后果

伴随着数字技术的迅速发展,学术界对大数据展开了广泛研究,多数文献肯定了大数据带来的积极影响。从现有文献来看,关于大数据经济后果方面的研究大致可归纳为两个方面。一是对宏观经济的影响。大数据的发展对宏观经济具有显著的促进作用。徐林、侯林岐等发现,国家级大数据综合试验区政策显著提升了城市的创新能力<sup>[2]</sup>;邱子迅和周亚虹基于国家级大数据综合试验区的相关数据,验证了数字经济发展对地区全要素生产率的促进作用<sup>[3]</sup>。这些研究表明,大数据的发展不仅有助于提高整体经济效率,还能促进创新驱动的发展模式。二是对微观企业经济活动的影响。Gao等研究发现,大数据显著降低了企业对金融资产配置的依赖性,减轻了企业的金融化倾向<sup>[7]</sup>。李增福等指出,大数据税收征管对企业盈余管理具有显著的治理效应<sup>[8]</sup>。史丹和孙光林的研究表明,大数据的发展有助于提升制造业企业的全要素生产率<sup>[9]</sup>。此外,艾永芳和孔涛研究了区域大数据发展对企业绿色创新的影响,结果显示大数据技术主要通过影响资金投向来促进企业的绿色创新,而非资金来源<sup>[10]</sup>。

### (二)企业避税的影响因素

近年来,企业避税行为受到学界的广泛关注。企业避税的影响因素相关研究可以从管理者、企业以及宏观层面等进行阐述。从管理者层面来看,管理者的特征对企业避税决策有显著影响。例如,过度自信的高管更倾向于实施企业避税<sup>[11]</sup>。此外,CFO兼任董事能够发挥良好的治理效应,从而抑制企业的过度避税行为<sup>[12]</sup>。从企业层面来看,融资约束是影响避税行为的重要因素。企业面临较强的融资约束时,可能会出现激进的避税行为,但随着企业内部控制的改善,这种避税动机将会受到抑制<sup>[13]</sup>。除融资约束外,信息透明度的提高能够显著降低企业避税程度<sup>[14]</sup>。从宏观层面来看,法律制度对企业避税的治理效果是学术界关注的重点之一。许红梅和李春涛研究发现,社会保险法的实施通过加强社保费的征收力度,抑制了企业避税行为<sup>[15]</sup>。曹越等研究指出,公平竞争审查制度能够显著降低高行政垄断地区企业的避税程度<sup>[16]</sup>。

综上所述,一方面,学者们关于大数据技术应用的研究正在迅速推进,相关研究均验证了其产生的积极影响。但是,已有研究大多基于赋能视角考察大数据技术产生的影响,鲜有文献深入探讨大数据技术所具备的治理效应。另一方面,已有关于企业避税影响因素研究成果较为丰富,但较少学者关注到数字技术对于企业避税的影响。因此,本文从企业避税视角考察大数据技术的治理效果,可能在以下方面有所创新。(1)以国家级大数据综合试验区为准自然试验,从微观主体角度探讨大数据技术应用对企业避税的影响,丰富了大数据技术应用对于微观主体影响的相关研究。(2)从信息透

明度与企业生产效率的角度出发,探讨大数据技术应用对企业避税的潜在机制,为大数据综合试点效果提供新的分析视角。(3)进一步从大数据综合试点地区的技术溢出、人才集聚现象等方面考虑大数据综合试点对企业避税的异质性影响,最终为政策制定者和企业管理者提供有价值的理论指导和实践参考。

### 三、理论分析与研究假设

#### (一)大数据技术应用与企业避税

避税是企业节约税负支出,将国家利益转化为公司利益的行为<sup>[17]</sup>。本文从信息不对称及委托代理两个视角分析企业避税动机。一方面,信息不对称是引发企业融资约束的主要原因。已有研究表明,我国企业普遍存在融资约束问题,在此情境下,企业为了增加内源融资,有动机通过实施某些避税行为来减少税负支出,增加现金持有<sup>[18]</sup>。另一方面,由于信息不对称,现代公司制度的两权分离使得公司所有者和管理者之间存在不同程度的委托代理问题。由于薪酬合同和业绩参照的存在,或是出于其他自利目的,管理层有动机通过避税实现收入操纵和资源转移<sup>[19]</sup>。

大数据技术应用将数据要素融入经济社会各个领域,为解决这些问题提供了新思路。在内部信息处理方面,隐瞒收入、夸大支出和转移定价是企业税收规避的常用手段<sup>[20]</sup>,大数据技术应用能够提升企业内部的前沿数字技术应用水平,增强企业信息收集能力,使得企业能够准确及时地分析处理各类经营活动和财务信息<sup>[21]</sup>,这无疑使管理层更难进行盈余操纵和关联交易等行为。在外部信息沟通方面,大数据技术应用拓宽了数字技术在社会生产、交易等环节的应用广度与深度,有效改善企业内外部信息的流转状况,提高了企业信息透明度<sup>[22]</sup>。企业对外界隐匿内部信息是企业遮掩避税活动的主要方式<sup>[23]</sup>,企业信息透明度的提升有利于利益关联方进行外部监督并及时参与到企业活动中实施避税监督。此外,大数据技术应用能够帮助金融机构与政府精准识别,有效缓解资金错配现象<sup>[2]</sup>。通过财政金融支持措施,为企业营造一个宽松的金融生态和完善的制度保障,缓解了中小企业“融资难、融资慢”的难题,很大程度上抑制了企业通过税收规避储备现金流的避税动机<sup>[24]</sup>。综上,大数据技术应用降低了信息不对称并缓解了委托代理问题,通过增强企业内部治理水平,抑制了企业避税行为。基于此,本文提出研究假设1。

假设1:大数据技术应用能够抑制企业避税。

#### (二)大数据技术应用对企业避税的影响机制

##### 1.大数据技术应用、信息透明度与企业避税

大数据技术应用提升了企业信息透明度。一方面,大数据技术应用带来的新型涉税监管技术,如金税四期极大增强了税务部门的监察能力<sup>[25]</sup>。受制于涉税监管技术的进步,企业通过盈余操纵调整利润的难度增加,同时管理层出于自利动机而披露虚假信息的机会主义成本也将明显上升<sup>[26]</sup>,避税风险及违规成本的增加促使企业核算信息准确性、披露真实财务信息。另一方面,大数据技术的应用增强了分析师的关注及其盈余预测的准确性。应用大数据技术的企业会引起以分析师为代表的外界广泛关注,相比于外部投资者,分析师能够精准分析向外界传达更可信的企业信息<sup>[27]</sup>。在大数据监测的背景下,尽管企业有意隐瞒内部不利信息,分析师依然能够更加及时、精准地掌握企业真实经营状况,从侧面提升了企业信息透明度。

信息透明度的提升有效抑制了企业的避税行为。复杂且不透明的经营管理活动为企业隐藏信息和实施避税行为提供了便利<sup>[28]</sup>,但随着信息透明度的提升增加了企业的非税收成本,压缩避税空间。一方面,企业信息透明度的提高将降低外部监督的难度和成本,使得税务部门和其他利益相关者能够更容易甄别企业的避税行为<sup>[29]</sup>。另一方面,信息透明度的提高迫使管理层经营行为接受股东和监管者更为深入的观察和更为严厉的控制,从而加强了对企业各项经营决策的监督,抑制企业避税动机<sup>[30]</sup>。基于此,本文提出研究假设2。

假设 2: 大数据技术应用能够提升信息透明度来抑制企业避税。

## 2. 大数据技术应用、企业生产效率与企业避税

大数据技术应用能够提升企业全要素生产率。一方面, 大数据技术应用能够优化人力资本结构, 提升企业创新能力。大数据技术应用产生劳动替代效应增加了对于数字型劳动者的需求, 部分提供隐性知识服务的高技能人才通过数字平台促进信息共享, 增强企业人力资本竞争优势<sup>[31]</sup>。高质量劳动力和知识资本融入生产经营过程后, 产生的扩散效应将促进新知识的产生, 使新产品新技术迅速向周边地区溢出, 由此产生的“羊群效应”将会吸引更多高技能人才, 形成人力资本结构与创新能力循环提升的逻辑闭环, 推动全要素生产率攀升<sup>[32]</sup>。另一方面, 大数据技术应用能够使企业充分利用数据资源, 拓展自身业务范围。企业利用数字资源获取数据要素, 通过“乘数效应”推动新兴技术融合, 提升大数据技术赋能生产经营业务程度, 为资本、劳动和技术等传统生产要素提供更多的配置途径<sup>[33]</sup>。进一步地, 企业对于数据要素的应用能扩大市场匹配的范围, 实现对不同市场的有机整合, 推动市场需求的倍增与市场匹配范围的扩大, 从而提升企业全要素生产率<sup>[9]</sup>。

企业生产效率的提升一定程度上削弱了避税行为带来的预期收益, 从而抑制了企业避税。首先, 生产效率的提高意味着企业能够通过正规渠道获取更多的超额收益<sup>[34]</sup>, 这使得避税行为带来的额外收益相对减少。由于避税行为通常伴随着一定的风险和成本, 决策的关键在于避税收益与成本之间的权衡, 避税边际收益下降将抑制企业避税动机。其次, 企业生产效率提升减轻了企业的税费负担<sup>[35]</sup>, 这意味着企业有足够的资金承担正常的税务负担。相较于避税, 依赖高效生产所获得的收益不仅更加稳定和可持续, 还能降低企业在税务审计中暴露的风险。最后, 生产效率的提升意味着其市场竞争力增强<sup>[36]</sup>, 企业能够通过创新和优化供应链等手段扩大市场份额和盈利空间。在此情况下, 企业更多地依赖核心业务的增长来获得利润, 而避税行为反而可能影响企业的声誉和市场地位<sup>[37]</sup>。因此, 随着避税收益的下降, 企业会更倾向于通过进一步提高生产效率来实现长期发展目标, 抑制了企业避税行为。基于此, 本文提出假设 3:

假设 3: 大数据技术应用能够提升企业生产效率来抑制企业避税。

## 四、研究设计

### (一) 样本选取与数据来源

本文数据主要来源于企业年报、Wind 数据库、国泰安数据库、中国统计年鉴、国家统计局和 CNRDS 数据库等。选取 2013—2022 年中国沪深 A 股上市公司为初始样本, 结合国家级大数据综合试验区政策对企业避税的影响及作用机制展开研究, 最终获得 3387 家企业 29276 个观察值。在此基础上对研究样本进行如下处理: 剔除金融类、房地产类企业; 剔除 ST、\*ST 类企业; 剔除数据少于 3 年的企业; 剔除主要变量缺失的企业。最后保留 2719 家企业共 22084 个观察值, 为避免极端值影响, 本文还对连续变量进行了 1%~99% 的缩尾处理。

### (二) 模型设定

基于理论分析, 本文采用多期双重差分法分析大数据综合试点对企业避税程度的影响及作用渠道。具体模型设定为:

$$BTD_{i,t} = \alpha_0 + \alpha_1 Bigdata_{i,p,t} + \alpha_2 Controls + \mu_i + \theta_t + \epsilon_{i,p,t} \quad (1)$$

式(1)中,  $i, p, t$  分别表示上市公司、地区及时间,  $\mu_i$  代表个体固定效应,  $\theta_t$  代表时间固定效应,  $\epsilon_{i,p,t}$  是随机扰动项。被解释变量  $BTD_{i,t}$  代表企业避税程度, 解释变量  $Bigdata_{i,p,t}$  代表上市公司  $i$  所属地区  $p$  在  $t$  年是否成为大数据试点的虚拟变量, 其余变量代表控制变量, 包括无形资产密集度、固定资产密集度、企业规模、企业年龄、经营费用率、第一大股东持股比例、产权性质和是否为四大审计。

### (三) 变量选取

#### 1. 被解释变量: 企业避税(BTD)

综合已有文献, 参考刘行和叶康涛的研究方法<sup>[38]</sup>, 使用异常税会差异(BTD)来衡量企业避税。通常来说, BTD(税收差异)越大, 企业利用 BTD 进行避税的可能性越高。BTD = (税前会计利润 -

应纳税所得额)/期末总资产。应纳税所得额=(所得税费用-递延所得税费用)/名义所得税率。此外,参照 Desai 等的研究方法,进一步使用扣除应计利润影响后的税会差异(DD-BTD)作为被解释变量的稳健性检验<sup>[19]</sup>。

### 2.解释变量:大数据综合试点(Bigdata)

基于《促进大数据发展行动纲要》政策,国务院分别于 2015 年与 2016 年分两批设立 10 个试点。本文据此设置了大数据综合试验区政策变量,若地区 p 于 2015 年成为试点,则该地区 2015 年及以后年份的 Bigdata 取值为 1;若地区 p 于 2016 年成为试点城市,则该地区 2016 年及以后年份的 Bigdata 取值为 1,其余为 0。

### 3.中介变量

(1)信息透明度(Absacc)。会计盈余是投资者和分析师评价企业的重要指标,但管理层常为追求自身利益而干预会计盈余。盈余操纵降低了财务报告的可靠性,误导信息使用者,损害了公众利益。因此,本文采用 Dechow 等提出的修正 Jones 模型计算得到的应计盈余管理来衡量企业信息透明度<sup>[39]</sup>。具体模型如下:

$$\frac{TA_{i,t}}{A_{i,t-1}} = \alpha_0 \frac{1}{A_{i,t-1}} + \alpha_1 \frac{\Delta REV_{i,t} - \Delta REC_{i,t}}{A_{i,t-1}} + \alpha_2 \left( \frac{PPE_{i,t}}{A_{i,t-1}} \right) + \epsilon_{i,t} \quad (2)$$

式(2)中,TA<sub>i,t</sub>代表企业当年总应计项目,等于营业利润减去经营活动产生的现金净流量;A<sub>i,t-1</sub>代表上一年度总资产,ΔREV<sub>i,t</sub>代表本年营业收入变动额,ΔREC<sub>i,t</sub>代表本年应收账款变动额,PPE<sub>i,t</sub>为本年固定资产。将式(2)分行业分年度回归,回归结果中残差的绝对值即为应计盈余管理程度,残差越小则企业会计信息透明度越高。

(2)企业生产效率(PE)。本文参考宋佳、张金昌等的研究,基于生产力二要素理论,采用熵值法衡量企业生产效率<sup>[40]</sup>。与此同时,结合赵国庆和李俊廷对于企业生产效率的衡量,将硬科技部分的指标替换为研发费用占营业收入比例、数字资产占无形资产比例、企业人工智能专利申请数以及企业绿色技术专利申请数,通过熵值法合成企业生产效率作为替换指标来提升稳健性<sup>[41]</sup>。硬科技部分的指标选取依据源于“十四五”规划中对硬科技代表性领域的界定,主要包括人工智能、信息技术、新能源以及智能制造等,上述四个指标能够较好地反映企业硬科技水平。指标体系说明如表 1 所示,并利用熵值法确定相应权重,计算出企业生产效率指标。同时,将两个指标扩大 1000 倍以增强模型解释力。

表 1 企业生产效率指标

因素	子因素	指标	衡量方式
劳动力	活劳动	研发人员薪资占比	(研发费用-工资薪酬)/营业收入
		研发人员占比	研发人员数量/员工总人数
		高学历人员占比	本科以上员工人数/员工总人数
	劳动对象	固定资产占比	固定资产/资产总额
生产工具	硬科技	制造费用占比	(经营活动现金流出+固定资产折旧+无形资产摊销+减值准备-购买商品接受劳务支付的现金-支付给职工以及为职工支付的现金)/(经营活动现金流出+固定资产折旧+无形资产摊销+减值准备)
		研发费用占比	研发费用/营业收入
	软科技	数字资产占比	数字资产/无形资产
		人工智能创新产出	企业申请的人工智能专利数量
	软科技	绿色技术创新产出	企业申请的绿色技术专利数量
		总资产周转率	营业收入/平均资产总额
		权益乘数倒数	所有者权益/资产总额

### 4.控制变量

本文参照刘行和叶康涛的研究<sup>[38]</sup>,本文选取无形资产密集度(Lntang)、固定资产密集度(PPE)、企业规模(Asset)、企业年龄(AGE)、经营费用率(AGC)、第一大股东持股比例(TOP1)、产权性质(SOE)和是否为四大审计(BIG4)作为控制变量。具体定义详见表 2。

表 2

变量定义及描述性统计

变量	变量定义	观测值	均值	标准差	最小值	最大值
BTD	(税前会计利润-应纳税所得额)/期末总资产	22084	0.003	0.027	-0.077	0.136
Bigdata	详见上文	22084	0.303	0.459	0.000	1.000
Absacc	公式(2)中残差的绝对值	22084	0.060	0.064	0.000	0.551
PE1	参考宋佳,张金昌等(2024)的方法进行计算	22084	5.296	2.430	0.787	16.747
PE2	替换硬科技指标作为 NPRO 的代替指标	22084	4.622	2.399	40.777	10.519
Lntang	期末无形资产净额与期末总资产之比	22084	0.048	0.053	0.000	0.417
PPE	期末固定资产净额与期末总资产之比	22084	0.213	0.153	0.003	0.715
Asset	期末总资产的自然对数	22084	22.342	1.274	19.863	26.456
AGE	企业上市年份	22084	2.175	0.768	0.693	3.401
AGC	管理费用和销售费用之和与主营业务收入之比	22084	0.152	0.121	0.000	0.695
TOP1	第一大股东持有的股份与公司总股份之比	22084	34.130	14.639	7.950	75.460
SOE	是否为国有企业	22084	0.337	0.473	0.000	1.000
BIG4	是否被国际四大事务所审计	22084	0.062	0.241	0.000	1.000

## 五、实证结果分析

### (一)描述性统计

由表 2 可知,企业避税(BTD)标准差为 0.027,均值为 0.003,说明平均而言,会计利润要高于应纳税所得额。信息披露质量(Absacc)最大值为 0.551,最小值为 0,表明我国不同企业之间信息透明度的差距较大。企业生产效率(PE1)最大值和最小值分别为 16.747 和 0.787,均值为 5.296,不难看出,企业之间生产效率水平存在较大差异。控制变量描述性统计结果与前人研究无实质性差异,不再赘述。

### (二)基准回归

表 3 为大数据综合试点影响企业避税程度的基准回归结果。为保证结果可靠性,列(1)没有加入控制变量,大数据综合试点的回归系数为-0.002 且在 5%水平上显著;列(2)在列(1)的基础上加入控制变量重新进行回归,大数据综合试点的回归系数为-0.003 且在 1%水平上显著,表明大数据综合试点能够显著抑制企业避税,假设 1 得到验证。

表 3 基准回归结果

变量	(1) BTD	(2) BTD	变量	(1) BTD	(2) BTD
Bigdata	-0.002** (-2.564)	-0.003*** (-2.798)	TOP1		-0.000 (-0.211)
Lntang		-0.042*** (-4.413)	SOE		-0.003** (-2.096)
PPE		-0.008** (-2.145)	BIG4		0.003 (1.418)
Asset		-0.002** (-2.063)	常数项	-0.003*** (-5.769)	0.035** (2.085)
AGE		0.003*** (2.814)	年份固定效应	控制	控制
AGC		-0.024*** (-4.068)	个体固定效应	控制	控制
			观测值	22084	22084
			R <sup>2</sup>	0.041	0.048

注: \*、\*\* 和 \*\*\* 分别表示在 10%、5% 和 1% 的显著性水平,括号内是经过稳健标准误调整的 t 值,下同。

### (三)假定识别检验

基准回归结果显示,大数据综合试点能够显著抑制企业避税行为的发生,但考虑遗漏变量的存在,需要进一步展开识别检验。

## 1. 平行趋势检验

双重差分法要求在实施大数据综合试点之前,实验组和控制组企业的避税程度具有相同的增长趋势。本文采用动态效应分析图的方法,对平行趋势假设进行检验,估计模型如式(3)所示。

$$BTD_{i,t} = \gamma_0 + \sum_{s=-6}^{-2} \gamma_s \text{pre}_{i,p,t} + \gamma_1 \text{cru}_{i,p,t} + \sum_{t=1}^6 \gamma_t \text{post}_{i,p,t} + \gamma_2 \text{Controls} + \mu_i + \theta_t + \varepsilon_{i,p,t} \quad (3)$$

式(3)中, $\text{pre}_{i,p,t}$ 表示试点设立前的虚拟变量, $\gamma_s$ 反映试点设立前2~6期的政策效应; $\text{cru}_{i,p,t}$ 表示试点设立当期的虚拟变量, $\gamma_1$ 反映试点设立当期的政策效应; $\text{post}_{i,p,t}$ 表示试点设立之后的虚拟变量, $\gamma_t$ 反映试点设立后1~6期的政策效应,其他变量定义与基准回归模型(1)保持一致。本文以设立大数据综合试点前2~6年与后6年的时间,并将试点设立前1年作为基准年份。图1显示了2013—2022年的回归系数变化及其95%的置信区间。从图1中的前2~6期可以看出, $\gamma_s$ 估计系数并不显著,表明处理组和对照组的企业避税程度在大数据综合试点设立之前的变化趋势基本相同,平行趋势假设成立。而在设立大数据综合试点的后6期, $\gamma_t$ 估计系数显著为负,表明处理组和对照组的企业避税程度在设立大数据综合试点之后有所不同,验证了大数据综合试点的动态效果。

## 2. 安慰剂检验

为进一步缓解遗漏变量问题对估计结果产生的影响,本文通过随机化大数据综合试点设立的时间,并重复这个随机过程500次,来进行安慰剂检验。图2展示了大数据综合试点对企业避税程度的安慰剂检验结果,可以发现处理效应的绝对值都集中分布在0的两侧,远离基准回归的估计结果(-0.003)。同时,大部分回归系数的p值都大于0.1。因此,可以排除基准回归结果是由不可观测因素导致的。

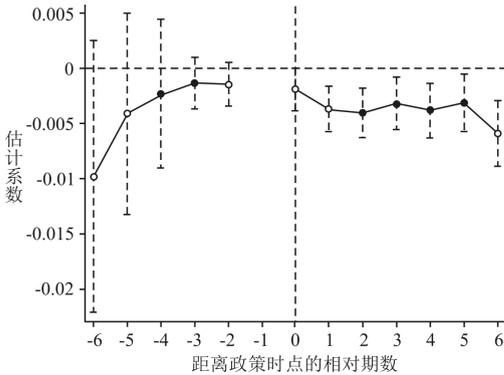


图1 平行趋势检验

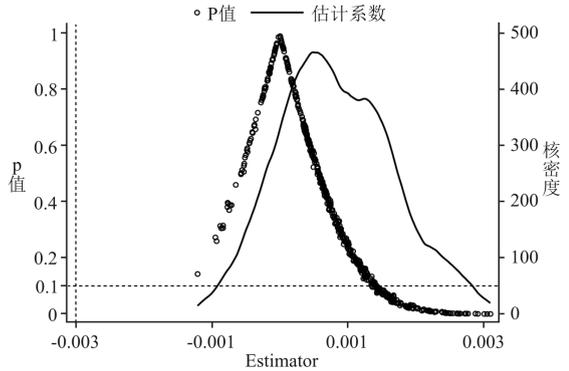


图2 安慰剂检验

## (四) 稳健性检验

考虑到大数据综合试点与企业避税之间存在双向因果、遗漏变量及被解释变量自身存在序列相关特性等问题,本文采用PSM-DID、替换被解释变量、系统GMM、排除其他政策的影响及剔除异常年份等方法对以上问题进行检验以增加前文基准回归结果的稳健性。

### 1. PSM-DID 实证结果分析

为解决样本选择偏差,本文采用倾向得分匹配—双重差分法(PSM-DID)来检验大数据综合试点对企业避税的效果。利用前文的控制变量中的无形资产密集度、固定资产密集度、企业规模、经营费用率、第一大股东持股比例和是否为四大审计作为匹配的协变量,通过Logit模型测算出每个样本的倾向得分,再使用卡尺1:1近邻匹配给实施大数据综合试点的企业逐年匹配对照组,进而缓解样本选择偏差带来的内生性问题。匹配结果如表4所示,匹配后的变量的标准化均值偏差均在10%以内,满足平衡性检验的要求,与匹配前相比,处理组与控制组在协变量上存在显著差异,t检验所对应的p值在1%水平下显著;经过匹配,协变量的标准偏差基本上大幅度下降,这也使得t检验所对应的p值在10%的显著性水平下不显著,表明控制组与处理组之间存在较小选择性偏差,匹配后的处理组与控制组数据满足条件独立同分布假设,匹配效果较好。

变量	匹配前 匹配后	均值		偏差%	t 检验	
		控制组	处理组		t 值	P 值
Lntang	U	0.046	0.048	-5.0	-3.27	0.001
	M	0.046	0.046	1.1	0.61	0.541
PPE	U	0.181	0.224	-29.2	-19.01	0.000
	M	0.182	0.178	2.4	1.50	0.133
Asset	U	22.542	22.301	18.8	12.56	0.000
	M	22.508	22.494	1.1	0.64	0.521
AGC	U	0.154	0.145	7.7	5.03	0.000
	M	0.154	0.156	-1.5	-0.81	0.417
TOP1	U	34.173	33.38	5.4	3.58	0.000
	M	33.988	33.936	0.4	0.20	0.838
BIG4	U	0.094	0.036	23.6	16.41	0.000
	M	0.084	0.072	4.9	2.57	0.010

匹配后的回归结果如表 5 所示。列(1)为经过逐年 PSM 之后的回归结果,与匹配之前的估计结果相一致,且相较于基准回归的回归系数有所提升,回归结果的稳健性得到了验证,说明大数据综合试点能够降低企业避税程度。

## 2. 更换被解释变量

本文将被解释变量替换为异常税收差异(DDBTD)重新进行回归,由表 5 列(2)的回归结果可知,大数据综合试点仍对降低企业避税程度有正向影响,与基本回归结论保持一致。

## 3. 系统 GMM

差分 GMM 估计与系统 GMM 估计可以被用来解决动态面板数据中存在的内生性问题。然而,差分 GMM 估计容易受到弱工具变量而产生有偏估计结果,Blundell 等为了解决此问题提出了系统 GMM 估计,并通过验证得出系统 GMM 估计比差分 GMM 估计的偏差更小<sup>[42]</sup>。因此,本文采用系统 GMM 估计来处理内生性问题,从表 5 列(3)可以看出,模型中的 AR(1)检验的 P 值均在 1%水平上显著;AR(2)检验的 P 值均不具有显著性,说明模型存在一阶序列自相关且不存在二阶序列自相关。同时 Hansen 检验的 P 值分别为 0.359,不具有显著性,说明不存在工具变量的过度识别。系统 GMM 回归中包含被解释变量滞后一期和核心解释变量的回归结果。列(3)结果显示,被解释变量滞后一期(L.BTD)的系数估计值为 0.391,在 1%水平上显著,表明上一期企业避税行为每增加 1 个百分点,可以导致当期企业避税行为增加 0.391 个百分点。

表 5 稳健性检验结果(1)

变量	(1)	(2)	(3)
	PSM-DID BTD	更换被解释变量 DDBTD	系统 GMM BTD
L.BTD			0.391 *** (13.341)
Bigdata	-0.004 *** (-3.170)	-0.002 ** (-2.319)	-0.010 ** (-2.363)
控制变量	控制	控制	控制
年份固定效应	控制	控制	控制
个体固定效应	控制	控制	控制
观测值	18003	22084	13155
R <sup>2</sup>	0.041	0.044	
AR(1)			0.000
AR(2)			0.559
Hansen 检验			0.359

#### 4.是否受其他区位导向性政策干扰

在经济发展的过程中大数据综合试点的设立往往会受到其他区位导向性政策的影响。本文主要聚焦于“智慧城市”政策与“宽带中国”政策,原因在于:“智慧城市”试点政策实施年份为2012—2014年,先于大数据综合试点,且该试点政策包含大数据综合试点中所有省份的部分城市,致力于使用云计算、物联网等智能技术对大数据进行分析整合实现对城市资源的高效配置;“宽带中国”试点政策实施年份为2014—2016年,属于与大数据综合试点的同期政策,且该试点政策包含大数据综合试点中9个省份的部分城市,能够有效提升企业创新水平。因此,选择“智慧城市”政策与“宽带中国”政策进行检验,参考杨仁和李自鑫<sup>[43]</sup>的方法,本文在式(1)的基础上估计如下方程:

$$BTD_{i,t} = \alpha_0 + \alpha_1 Bigdata_{i,p,t} + \alpha_2 did1_{i,p,t} + \alpha_3 did2_{i,p,t} + \alpha_4 Controls + \mu_i + \theta_t + \epsilon_{i,p,t} \quad (4)$$

式(4)中, $did1_{i,p,t}$ 为“智慧城市”政策虚拟变量, $did2_{i,p,t}$ “宽带中国”政策虚拟变量。实证结果如表6列(1)(2)(3)所示,大数据综合试点对所在地区企业避税程度仍存在显著抑制作用。因此,大数据综合试点对所在地区企业避税的抑制作用确实是由大数据综合试点所导致,而非其他政策影响。

#### 5.剔除部分样本分析

在全样本的基础上,通过剔除位于特殊年份样本,确保回归结果免受其他不可观测因素的干扰。具体而言,在样本年限内(2020—2022年)发生了重要的突发公共卫生事件,在很大程度上影响了企业生产经营的决策行为。因此,本文剔除了疫情影响,将样本年限截取为2013—2019年。实证结果如表6列(4)所示,大数据综合试点对所在地区企业避税的影响仍显著。

表6 稳健性检验结果(2)

变量	(1)	(2)	(3)	(4)
	排除政策干扰 BTD	排除政策干扰 BTD	排除政策干扰 BTD	剔除部分样本 BTD
Bigdata	-0.003*** (-2.758)	-0.003*** (-2.738)	-0.003*** (-2.741)	-0.001* (-1.665)
did1	0.002 (0.620)		0.001 (0.537)	
did2		0.000 (0.081)	0.000 (0.137)	
控制变量	控制	控制	控制	控制
年份固定效应	控制	控制	控制	控制
个体固定效应	控制	控制	控制	控制
观测值	21880	21839	21839	14311
R <sup>2</sup>	0.048	0.048	0.048	0.022

#### (五)异质性分析

##### 1.对企业区域差异的异质性分析

数字产业正经历一个明显加速发展的过程,大数据发展水平存在区域间差异,整体上呈现出由东部沿海向西部内陆递减的空间分布特征,其中东部地区与中西部地区企业的融资渠道及避税手段存在较大差异。通过考察东中西部地区的企业,能够更好地体现大数据综合试点对不同区域企业避税治理的差异性,从而提出更具针对性的建议。鉴于此,根据企业所属省份不同划分为东部、中部和西部地区三个子样本分别进行回归。

由表7列(1)(2)(3)可知,大数据综合试点对于东部地区企业避税程度具有显著抑制作用,对于中西部地区企业避税程度抑制效果不明显,这可能与东部地区的经济繁荣度、人才密集度、资本集中度较高有关,而中西部地区可能由于某些经济基础较为薄弱的地区存在产业金融化程度较低、数字基础设施建设不完善从而导致中西部大数据政策试点无法取得预期效果,从而导致其对企业避税程度具有抑制作用但不显著的结果。

## 2.对企业行业差异的异质性分析

数字产业的迅猛发展受到外界投资者的持续关注,高科技企业与非高科技企业采用的避税手段不尽相同,高科技企业掌握更加前沿数字技术,对于企业数据的处理也具有更加丰富的经验,在研究中引入这两类企业有助于更加清晰地展现大数据综合试点对于不同行业的避税治理效果。通过此类分析能够为提出更具针对性的政策提供数据支撑,以提升对于多样化企业的避税治理效果。鉴于此,本文根据企业是否属于高科技行业进行分组回归,当企业属于高科技行业时,取值为1,否则取值为0。按企业技术间差异分组检验的结果如表7列(4)(5)所示,大数据综合试点对高科技企业的避税抑制效果不显著;对于非高科技企业的避税治理产生显著效果。之所以存在上述差异,原因在于高科技企业对于数字技术应用更加广泛,能够通过多样化的渠道获得融资,且其具备的高科技人才资源更加丰富,企业内部的税务处理也会更加规范,从而导致大数据综合试点对高科技企业避税程度并不具有显著影响。相反,非高科技企业面临的融资约束程度大并且数字技术应用的匮乏带来的无效内部治理会增加企业实施避税行为的可能,大数据综合试点的引导与规范表现出明显的避税治理作用。

表7 异质性检验

变量	(1)	(2)	(3)	(4)	(5)
	东部 BTD	中部 BTD	西部 BTD	非高科技企业 BTD	高科技企业 BTD
Bigdata	-0.004*** (-3.465)	0.001 (0.364)	-0.001 (-0.162)	-0.004*** (0.001)	-0.001 (0.001)
控制变量	控制	控制	控制	控制	控制
年份固定效应	控制	控制	控制	控制	控制
个体固定效应	控制	控制	控制	控制	控制
观测值	15731	3865	2488	9668	12416
R <sup>2</sup>	0.049	0.049	0.067	0.023	0.073

## 六、机制检验

大数据综合试点能够通过促进信息透明度和企业生产效率发展来抑制企业避税,为避免中介效应检验的内生性偏误问题<sup>[44]</sup>,本文重点分析大数据综合试点政策对信息透明度和企业生产效率的影响,信息透明度和企业生产效率对企业避税的影响已在上文理论机制中说明,具体模型设定如下。

$$\text{Absacc}_{i,t}/\text{PE}_{i,t} = \alpha_0 + \alpha_1 \text{Bigdata}_{i,p,t} + \alpha_2 \text{Controls} + \mu_i + \theta_t + \epsilon_{i,p,t} \quad (5)$$

式(5)中,Absacc为信息透明度,采用修正Jones模型计算残差的绝对值衡量;PE代表企业生产效率,参考孙雪娇和翟淑萍<sup>[37]</sup>的研究,通过熵值法合成。在此基础上,修改硬科技指标重新衡量企业生产效率作为稳健性检验。其他变量参考基准回归变量解释。

### (一)信息透明度机制

首先检验大数据综合试点能否通过提升信息透明度来降低企业避税程度,表8列(2)显示,核心解释变量大数据综合试点(Bigdata)对信息透明度(Absacc)的回归系数为-0.008,在1%水平上显著,说明大数据综合试点能够提升企业信息透明度。信息透明度的提升降低了外部监督的难度,使得管理层行为更容易受到利益相关者的控制,从而降低了避税的动机。可见,提升信息透明度是大数据综合试点抑制企业避税的一个重要渠道。

### (二)企业生产效率机制

接下来检验大数据综合试点能否通过提升企业生产效率来降低企业避税程度,表8列(3)表明,核心解释变量大数据综合试点(Bigdata)对企业生产效率(PE)的回归系数为0.188,且在1%水平上显著,说明大数据综合试点能够提升企业生产效率(PE1)。同时采用更换硬科技指标后的企业生产效率(PE2)作为替代变量重新进行回归以增强稳健性,表8列(4)的结果表明替换变量后依然显著,验证了中介机制的有效性。企业生产效率提升增加了超额收益,使企业有能力负担税费,由于避税相

对收益下降抑制了企业避税动机。综上所述,大数据综合试点能够通过提升信息透明度与企业生产效率发展抑制企业避税行为,假设 2 与假设 3 得证。

表 8 机制分析结果

变量	(1)	(2)	(3)	(4)
	BTD	Absacc	PE1	PE2
Bigdata	-0.003*** (-2.798)	-0.008*** (-3.485)	0.188*** (3.085)	0.462*** (3.137)
控制变量	控制	控制	控制	控制
年份固定效应	控制	控制	控制	控制
个体固定效应	控制	控制	控制	控制
观测值	22084	22084	22084	22084
R <sup>2</sup>	0.048	0.122	0.473	0.314

## 七、结论与启示

本文基于国家级大数据综合试验区的准自然试验,选取 2013—2022 年我国沪深 A 股 2719 家 22084 个观察值的非平衡面板数据,构建多期双重差分模型和中介效应模型分析了大数据技术应用对企业避税的影响及内在机制,得出如下结论。

第一,大数据技术应用在一定程度上抑制了企业避税。实证结果表明,无论是基准回归还是考虑内生性以及进行的一系列稳健性结果,研究结果均表明大数据技术应用能够抑制企业避税行为的发生,因此可认为大数据技术应用显著降低了企业避税程度。第二,将上市企业按照区域及行业间的差异进行分组探究大数据技术应用对于不同企业避税程度影响差异性,结果发现,大数据技术应用对于东部地区、非高科技企业抑制企业避税程度的作用更明显。第三,大数据技术应用主要通过提升企业信息透明度和生产效率,由此能够对企业避税进行有效治理。

本文的政策启示如下。第一,从税收征管角度出发,政府应当鼓励税务部门应用大数据分析和人工智能技术,对企业的财务数据进行深度挖掘和实时监控。通过数据模型分析,识别出异常交易和潜在的避税行为,及时采取监管措施,预防企业利用复杂的跨境交易或其他手段进行避税。第二,鉴于东部地区及非高科技企业对大数据技术应用反应积极,政府可以在东部地区进一步加强大数据基础设施的建设与应用,推广成功经验。针对非高科技企业出台专门的数据治理政策。第三,由于征纳主体双方之间存在信息不对称,税务部分应借助大数据和人工智能技术,建立企业“数字足迹”追踪机制。将企业在经营活动中产生的数据链条化,使税务部门和分析师能够在海量数据中捕捉企业异常活动,从而做出精准预判。企业也应当主动适应大数据技术时代的到来,企业应通过技术创新而不是单纯依靠避税来提高生产效率,与此同时,企业还应加强劳动力数字技能培训,不断强化人力资本水平以提升竞争力。

### 参考文献:

- [1] 黄丽君.数智化税收治理探索:嵌入纳税人自然系统的设计与实现[J].税务与经济,2023(3):24—31.
- [2] 徐林,侯林岐,程广斌.国家级大数据综合试验区创新效应研究[J].科技进步与对策,2022(20):101—111.
- [3] 邱子迅,周亚虹.数字经济发展与地区全要素生产率——基于国家级大数据综合试验区的分析[J].财经研究,2021(7):4—17.
- [4] 苏锦旗,唐诗瑶,张营营.国家级大数据综合试验区能否促进区域经济高质量发展——基于试点区政策的准自然实验[J].现代财经,2023(10):56—73.
- [5] 张修凡,范德成.数字经济发展赋能我国低碳经济转型研究——基于国家级大数据综合试验区的分析[J].科技进步与对策,2023(19):118—128.
- [6] 张益豪,郭晓辉.大数据发展与企业全要素生产率——基于国家级大数据综合试验区的实证分析[J].产业经济研究,2023(2):69—82.

[7] Gao, Y., Wang, Z., Wang, K., Zhang, R., Lu, Y., Effect of Big Data on Enterprise Financialization: Evidence from China's SMEs[J]. *Technology in Society*, 2023; 102351.

[8] 李增福, 骆展聪, 杜玲, 等. “信息机制”还是“成本机制”? ——大数据税收征管何以提高了企业盈余质量[J]. *会计研究*, 2021(7): 56—68.

[9] 史丹, 孙光林. 大数据发展对制造业企业全要素生产率的影响机理研究[J]. *财贸经济*, 2022(9): 85—100.

[10] 艾永芳, 孔涛. 区域大数据发展能促进企业绿色创新吗? [J]. *中南财经政法大学学报*, 2021(6): 116—126.

[11] Hsieh, T. S., Wang, Z., Demirkan S. Overconfidence and Tax Avoidance: The Role of CEO and CFO Interaction[J]. *Journal of Accounting and Public Policy*, 2018, 37(3): 241—253.

[12] 孙雪娇, 翟淑萍, 甦叶. CFO兼任董事能抑制企业极端避税行为吗? [J]. *中南财经政法大学学报*, 2021(1): 67—76.

[13] 陈作华, 方红星. 融资约束、内部控制与企业避税[J]. *管理科学*, 2018(3): 125—139.

[14] Kerr, J.N. Transparency, Information Shocks, and Tax Avoidance[J]. *Contemporary Accounting Research*, 2019, 36(2): 1146—1183.

[15] 许红梅, 李春涛. 社保费征管与企业避税——来自《社会保险法》实施的准自然实验证据[J]. *经济研究*, 2020(6): 122—137.

[16] 曹越, 张玉婷, 周瑞博. 公平竞争审查制度会降低企业避税程度吗? [J]. *中南财经政法大学学报*, 2024(2): 52—66.

[17] Kim, J. B., Li, Y., Zhang, L. Corporate Tax Avoidance and Stock Price Crash Risk: Firm—Level Analysis [J]. *Journal of Financial Economics*, 2011, 100(3): 639—662.

[18] 王宇, 赵宇亮, 张东源. 客户集中度对企业避税的影响机制研究——来自中国上市公司的实证分析[J]. *中国软科学*, 2022(7): 118—128.

[19] Desai, M. A., Dharmapala, D. Corporate Tax Avoidance and High—Powered Incentives[J]. *Journal of Financial Economics*, 2006, 79(1): 145—179.

[20] 于明洋, 吕可夫, 阮永平. 文过饰非还是如实反映——企业避税与年报文本复杂性[J]. *经济科学*, 2022(3): 112—126.

[21] 崔建军, 雍雁岚, 王少辉. 国家级大数据综合试验区与企业绿色技术创新[J]. *经济问题探索*, 2023(12): 37—59.

[22] 高玉婷, 李波. 大数据试验区建设助推经济绿色发展的机制与实现路径[J]. *中南民族大学学报(自然科学版)*, 2024(2): 280—288.

[23] 陈凯, 杨亚平. 企业数字化转型缘何增加了征税难度——来自中国上市公司避税活动的证据[J]. *山西财经大学学报*, 2023(12): 111—122.

[24] 黄徐亮, 徐海东, 倪鹏飞. 大数据治理对数字产业发展的影响研究[J]. *经济体制改革*, 2024(1): 97—104.

[25] 张克中, 欧阳洁, 李文健. 缘何“减税难降负”: 信息技术、征税能力与企业逃税[J]. *经济研究*, 2020(3): 116—132.

[26] Rego, S. O., Wilson R. Equity Risk Incentives and Corporate Tax Aggressiveness[J]. *Journal of Accounting Research*, 2012, 50(3): 775—810.

[27] 张宗新, 周嘉嘉. 分析师关注能否提高上市公司信息透明度? ——基于盈余管理的视角[J]. *财经问题研究*, 2019(12): 49—57.

[28] Balakrishnan, K., Blouin, J.L., Guay W R. Tax Aggressiveness and Corporate Transparency[J]. *The Accounting Review*, 2019, 94(1): 45—69.

[29] 支晓强, 王瑶, 侯德帅. 资本市场开放能抑制企业避税吗——基于沪港通的准自然实验[J]. *经济理论与经济管理*, 2021(2): 70—84.

[30] 董竹, 张欣. 会计信息可比性与企业风险承担的关系研究[J]. *外国经济与管理*, 2021(2): 36—50.

[31] 陈岩, 张李叶子, 李飞, 等. 智能服务对数字化时代企业创新的影响[J]. *科研管理*, 2020(9): 51—64.

[32] 欧阳海琴. 数字经济、人才集聚与长三角高质量发展[J]. *南昌大学学报(人文社会科学版)*, 2024(1): 45—55.

[33] 张夏恒, 刘彩霞. 数据要素推进新质生产力实现的内在机制与路径研究[J]. *产业经济评论*, 2024(2): 1—21.

[34] 段俊, 戈亨婷, 张保帅. 数字普惠金融、企业全要素生产率与绿色创新[J]. *哈尔滨商业大学学报(社会科学版)*, 2024(2): 15—30.

[35] 陈冬, 陆佳妮, 周宇洁, 等. 增值税改革与企业所得税避税的联动效应——来自增值税税率简并改革的证据[J]. *财经研究*, 2024(5): 49—63.

- [36] McMillan, J., Woodruff, C. The Central Role of Entrepreneurs in Transition Economies[J]. *Journal of Economic Perspectives*, 2003, 16(3): 153—170.
- [37] 孙雪娇,翟淑萍,于苏.柔性税收征管能否缓解企业融资约束——来自纳税信用评级披露自然实验的证据[J].*中国工业经济*,2019(3):81—99.
- [38] 刘行,叶康涛.企业的避税活动会影响投资效率吗? [J].*会计研究*,2013(6):47—53.
- [39] Dechow,P.M., Sloan,R.G., Sweeney,A.P. Detecting Earnings Management[J]. *Accounting Review*, 1995: 193—225.
- [40] 周名丁,胡春生.数字创新推动制造业企业高质量发展的机理与路径研究[J].*贵州社会科学*,2024(3): 105—123.
- [41] 赵国庆,李俊廷.企业数字化转型是否赋能企业新质生产力发展? ——基于中国上市企业的微观证据[J].*产业经济评论*:2024(3):1—13.
- [42] Blundell, R., Bond, S., Windmeijer, F. Estimation in Dynamic Panel Data Models: Improving on the Performance of the Standard GMM Estimator[J]. *Nonstationary Panels, Panel Cointegration, and Dynamic Panels*, 2001: 53—91.
- [43] 杨仁发,李自鑫.创新型城市试点政策能够促进共同富裕吗? [J].*财经研究*,2023(8):19—33.
- [44] 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.

## Can Big Data Technology Applications Curb Corporate Tax Avoidance: Analysis Based on National-level Big Data Comprehensive Pilot Zones

HE Yong WU Teng

(School of Economics and Management, Hunan University of Technology, Zhuzhou, 412007, China)

**Abstract:** Based on the pilot policy of national-level big data comprehensive experimental zone, using the data of Shanghai and Shenzhen A-share listed companies from 2013 to 2022, a multi-period DID model is used to study the impact of big data technology application on corporate tax avoidance. The result shows that the application of big data technology significantly inhibits corporate tax avoidance, and this conclusion still holds after a series of robustness tests such as PSM-DID, replacing explanatory variables, and eliminating policy interferences; the mechanism analysis shows that the application of big data technology helps to improve the transparency of information and the efficiency of corporate production to inhibit corporate tax avoidance; the heterogeneity analysis shows that the inhibitory effect of the application of big data technology on tax avoidance is more significant for the eastern region and the non-high-tech enterprises. Heterogeneity analysis shows that the application of big data technology has a more significant inhibiting effect on tax avoidance in the eastern region and non-high-tech enterprises. The findings of this paper not only enrich the research on the influence factors of enterprise tax avoidance, but also reveal the positive governance effect of the national-level big data comprehensive pilot zone policy, which can provide theoretical support and empirical evidence for the subsequent continuous promotion of the application of big data technology.

**Key words:** Big Data Technology; Corporate Tax Avoidance; Information Transparency; Corporate Productivity; Multi-period DID

(责任编辑:肖加元)