

市场一体化对企业数字创新的影响^{*}

——兼论数字创新衡量方法

胡增玺 马述忠

内容提要:数字创新是经济高质量发展的重要推动力量,全国统一大市场建设会放大数字创新的规模经济和创造性破坏效应,成为在科技创新领域被“卡脖子”困境下引领企业数字创新的关键。当前数字创新量化研究较少,本文创新性使用自然语言处理与机器学习交叉范畴的预训练语言模型,基于 1985—2017 年中国专利全文数据库和数字经济核心产业分类描述文本进行数字创新识别工作,并将识别到的数字创新与天眼查数据库匹配。之后,本文以全国统一大市场与数字创新之间的关键联系为切入点,基于数量和质量双重视角,探究市场一体化对企业数字创新的影响及其渠道。结果发现:市场一体化水平的提高可以显著促进企业数字创新“提质增量”,市场一体化带来的市场规模的扩大和市场竞争的加剧是影响企业数字创新的重要渠道,市场规模的扩大促进企业数字创新,而市场竞争对企业数字创新的影响是非线性的倒 U 型。本文首次为数字创新提供了创新的衡量方法,为后续研究奠定了微观基础。本文还为全国统一大市场建设以及技术创新市场导向机制的完善提供了重要的理论和经验证据。

关键词:数字创新 自然语言处理 市场一体化 市场规模 市场竞争

一、引言

21 世纪以来,数字经济蓬勃发展。中国数字经济规模居全球第二,但与部分发达国家尤其是美国还存在一定差距:美国是消费互联网和产业互联网“双腿跑”;中国则是消费互联网“单脚跳”,产业互联网起步较晚,存在明显短板。^①产生这种差异的重要原因是当前中国处在由消费互联网向产业互联网迭代的关键阶段,中美博弈日趋复杂,以美国为首的西方国家对科技封锁,从华为等企业的全面制裁到 2022 年芯片四方联盟的成立,无一例外都是遏制中国科技创新能力提升的企图。加之中国部分基础和关键领域数字创新能力不足,从而使得中国在数字经济领域遭遇了芯片断供诸多“卡脖子”问题。^②源源不断的数字创新是数字经济高质量发展的动力来源。国务院《“十四五”数字经济发展规划》基本原则第一点提出了要坚持创新引领,并给出到 2025 年数字化创新引领发展能力大幅提升的目标,“创新”一词在该规划中共出现了 51 次,可见推动数字创新,强化关键技术创新能力对中国数字经济发展的必要性。同时,作为经济高质量发展的微观基础,企业是创新的主体,是推动创新创造的主力军。^③《国民经济和社会发展第十四个五年规划和 2035 年

* 胡增玺、马述忠(通讯作者),浙江大学中国数字贸易研究院,邮政编码:310058,电子信箱:zengxi_hu@163.com, mashuzhong@zju.edu.cn。本研究得到国家自然科学基金面上项目(71973120)的资助。作者感谢匿名审稿专家的宝贵意见,当然文责自负。

① 中国信息通信研究院,2021:《全球数字经济白皮书》和《中国数字经济发展白皮书》。

② 根据中国科学技术发展战略研究院 2021 年发布的《国家创新指数报告 2020》,中国科技创新能力排名第 14,与数字经济规模世界第二的排名存在明显差距。

③ 《2020 年全国科技经费投入统计公报》显示,2020 年企业研发经费占全国的比重达 76.6%,对全国的贡献达 77.9%。

远景目标纲要》第五章指出要“强化企业创新主体地位”。供给端的企业可以通过增强数字技术创新能力并将其运用到生产流程上,数字赋能传统产业转型升级,为经济高质量发展提供重要推动力。^①因此,在当前中美博弈尤其是美国对中国科技封锁的背景下,以企业为切入点探究数字创新,对化解“卡脖子”难题极具现实意义。

科技创新最大的激励来自于制度健全的市场环境,这能够为其提供足够的市场预期与发展动力。随着中美博弈日趋复杂,加之新冠肺炎疫情冲击,全球市场动荡、需求萎靡。中国作为拥有巨大内需市场的大国,全国统一大市场的建设可能成为提升自主创新能力的关键抓手。国务院《“十四五”数字经济发展规划》指出要“发挥中国超大规模市场优势,提高数字技术基础研发能力”。超大规模市场具有规模巨大的技术需求,推动技术创新,形成创新的比较优势(欧阳晓和汤凌霄,2017)。《中共中央国务院关于加快建设全国统一大市场的意见》的发展目标也指出,要“发挥超大规模市场具有丰富应用场景和放大创新收益的优势,通过市场需求引导创新资源有效配置,促进创新要素有序流动和合理配置”。但中国超大规模市场对技术创新的激励效应还不充分。受限于区域壁垒,中国市场需求与创新之间存在错配,无法实现需求引致创新,要素流动受阻,企业创新行为受到抑制(张杰和周晓艳,2011;卿陶和黄先海,2021)。建设全国统一大市场可以充分促进创新要素流动,发挥市场在资源配置中的作用,有利于促进竞争、优化分工,增强企业创新能力。^②这意味着只有形成破除分割、打通堵点的全国统一大市场才能充分发挥其对创新的引领作用。

相比于传统创新,数字创新具有更强的非竞争性,能够产生更大的规模经济效应,准入门槛更低从而具有更高的“创造性破坏”概率(Guellec & Paunov,2017)。在需求端,全国统一大市场会放大规模经济效应,激励企业数字创新;在供给端,全国统一大市场会放大“创造性破坏”效应,促进技术更迭。这意味着全国统一大市场可能成为驱动企业数字创新的核心因素,甚至成为打破关键领域“卡脖子”困境的突破口。《中共中央国务院关于加快建设全国统一大市场的意见》提出了建设全国统一大市场的整体要求,即“加快建立全国统一的市场制度规则,打破地方保护和市场分割,打通制约经济循环的关键堵点,促进商品要素资源在更大范围内畅通流动”。从经济学视角看,这意味着各地区之间的一体化程度不断加深,市场一体化是建设全国统一大市场的关键。^③本文在合理识别和衡量数字创新后,主要以中国产品市场一体化为切入点,从理论层面探究影响效应和渠道,并进行实证分析。相较以往文献,本文可能的边际贡献如下:

第一,为数字创新的测度提供创新性和有针对性的方法。对于传统创新,现有文献往往使用R&D投入、专利或者全要素生产率来衡量创新水平(张杰等,2015;寇宗来和刘学悦,2020;安同良和姜妍,2021)。而对于数字创新这一新生事物,目前的研究大都仅仅停留在理论层面,对数字创新做出定性表述(Yoo et al.,2010)。数字创新不像数字经济可以用计量工具测度并衡量(蔡跃洲和牛新星,2021)。从全要素生产率视角看,数字创新的贡献很难分解;从R&D投入视角看,总研发投入中用于数字创新的比重难以准确衡量;而从专利视角看,任何一个专利发明是否属于数字创新可以根据其特征和性质做出判断。本文创新性引入预训练语言模型的机器学习方法,为量化数字创新这一新生事物奠定基础,同时探究了数字创新基础的特征事实,为后续数字创新的研究奠定了重要基础。

第二,本文结合数字创新的全新特征,将其纳入市场一体化的分析框架,弥补了新古典经济学

① 陶长琪、袁琦璟,2021:《“数字+”创新赋能经济高质量发展》,《中国社会科学报》。

② 《国家统计局新闻发言人就2022年一季度国民经济运行情况答记者问》,http://www.stats.gov.cn/xxgk/jd/sjjd2020/202204/t20220418_1829758.html。

③ 市场一体化主要包括产品、要素、服务等一体化,需要说明的是本文关注的是产品市场一体化,关于市场一体化的论述详见第三部分理论分析。

中市场与创新研究的不足,赋予了传统理论数字经济时代的内涵。创新理论最先由熊彼特提出,而后逐渐形成创造性破坏的内生增长模型(Aghion & Howitt,1992)。随着异质性企业理论的提出,创新理论逐渐过渡到企业层面(Klette & Kortum,2004)。创新理论的另一大研究领域是创新的影响因素。税收(Akcigit et al.,2022)、补贴(张杰等,2015)、制度(张杰等,2021)等政策因素以及国际贸易(Atkeson & Burstein,2010)都是创新的影响因素。但从创新活动的本质来看,市场才是其关键影响因素(张杰和周晓艳,2011)。从买卖双方两大参与主体看:需求端市场规模的扩大会促进企业创新(Acemoglu & Linn,2004);供给端市场竞争对创新的影响尚未有定论,现有文献的结论主要有抑制(Dorn et al.,2020)、促进(Correa & Ornaghi,2014)和倒U型关系(Aghion et al.,2005)。而数字创新相比传统创新具有数字非竞争性、更低的进入门槛(成本)等特点(Yoo et al.,2010;Guellec & Paunov,2017)。这些差异可能会突破传统创新的范式,为市场与创新理论增添全新内涵。

第三,基于数量和质量双重视角考察市场一体化对数字创新的影响。以往利用专利进行创新研究的文献大多侧重于创新的某一方面,如专利数量(寇宗来和刘学悦,2020)。对于专利质量,目前的衡量方法有专利被引法、专利和知识宽度法(Akcigit et al.,2016;张杰和郑文平,2018)。一些文献尝试从数量和质量双重视角探究企业创新水平的提升(诸竹君等,2020)。本文立足全国统一大市场,探究市场一体化对数字创新数量和质量的影响,丰富了企业创新“提质增量”的相关研究。

本文其余部分安排如下:第二部分介绍了数字创新概念、识别方法、合理性验证和特征事实;第三部分进行理论分析并提出假说;第四部分延续理论分析中市场一体化与数字创新的关系并实证探究影响机理;最后是结论和政策建议。

二、数字创新的概念、衡量方法及特征事实

(一) 数字创新的概念

数字创新的概念衍生于信息系统领域对数字技术的研究,普遍认为是数字技术加持下传统创新的数字化变革(Yoo et al.,2010;Nambisan et al.,2017)。当前对于数字创新的研究呈现多学科交叉的特点,但大多仅停留在理论层面,对数字创新做出定性表述。Yoo et al.(2010)将数字创新定义为数字部件和物理部件进行新的组合,以生产新的产品,其关注点为产品创新。Nambisan et al.(2017)认为数字创新是利用数字技术改善现有产品、流程和商业模式或开发新产品。Guellec & Paunov(2017)认为数字创新是IT行业内外基于软件代码和数据的新产品和新流程,属于无形资产。从企业层面看,Autio & Thomas(2020)指出制造业企业的数字创新是将新兴数字技术嵌入现有的传统非数字化产品和服务中,赋予这些传统产品和服务新的数字技术特性。从表现形式来看,Fichman et al.(2014)认为数字创新为新产品、新工艺或新商业模式的出现。

对于任何新生事物的界定,最本质的出发点是其展现的特征事实:数字和物理部件进行新的组合如工业机器人等属于数字创新,数字部件本身的迭代或开创性突破如通讯方法也属于数字创新;无形的软件和技术创新属于数字创新,有形的智能设备的创新也属于数字创新;数字产品如半导体部件的创新属于数字创新,数字生产和服务流程应用的如工控系统也属于数字创新。尽管事实例证中的数字创新多种多样,但是可以确定的是数字化的思维贯穿始终。当前,数字技术的嵌入使得不同形态的边界不断交互、越发模糊(Nambisan et al.,2019),从以上特征事实也可看出事物之间本身就存在不同程度的交叠。因此,再简单地从数字技术在传统领域的渗透或从产品、服务等角度对数字创新进行界定已经不合时宜,也意味着现有数字创新的定义存在一定局限性。

数字创新作为数字经济时代独有的概念,与数字经济的发展紧密相连。随着数字经济的发展,新一代数字技术改变了产品与服务形态、企业组织形式以及产业结构(余江等,2017)。不断推陈出新的数字组件以及其与产品、服务、流程等的融合打破了对原有创新过程和新生事物的认知,突

破了传统创新的范式。尽管数字创新突破了传统认知,但归根结底,其本质是创新,与传统创新的不同则源于数字经济这一全新纪元。立足于数字经济时代,基于创新的概念,本文从一般的角度给出了数字创新的内涵:^①

数字创新是以数字化思维和想法为导向,利用现有的以数字化为主的知识 and 物理部件,通过改进或创造新的事物,进而取得创新成果的行为。

从组成来看,数字创新既包括数字组件自身开创性的创新,也包括数字组件和传统组件进行重组并由数字组件驱动形成的创新;从物理特征看,数字创新既包含无形的软件等数字技术和想法的创新,也包含有形的数字和物理部件结合的创新;从结果来看,数字创新包括产品、服务、商业模式等方面;数字创新的主体包括企业、研究机构、个人等;其目的是在数字产品、技术、服务、基础设施以及解决方案等方面为数字经济发展提供动力。

与传统创新相比,数字创新具有一些全新特征:第一,数字非竞争性。数字非竞争性的存在,使得数字创新形成的产品具有比传统创新更低的生产和传播成本,从而产生更大的规模经济效应。随着数字技术的发展,无形成分如软件在数字创新产生的新事物中占据更高的比重。更高无形化的组成使得数字创新形成的产品的边际生产成本不断降低,甚至对于完全无形化产品(例如,游戏软件),生产和传播成本无限接近于零(Corrado et al.,2009;Yoo et al.,2010)。这就产生了巨大的规模回报,只要该成果被应用的次数越多,平均生产成本就越低。第二,更低的进入门槛。以软件、编程为代表的数字技术是数字创新的核心,其资本投入要求远远低于传统创新,意味着更多潜在创新主体会不断涌入数字创新领域。而数字技术的开放共享特点意味着企业可以低成本地将数字技术应用于创新过程,“创造性破坏”概率大大提升(Guellec & Paunov,2017)。同时,在“创造性破坏”概率更高和市场进入机会更多的地方,风险也更高。传统创新成果的生产和应用可能会减少相关领域其他产品的市场份额,而数字创新可能会导致赢者通吃的结局,在位者可能面临失去整个市场的风险(例如,自媒体平台与传统媒体平台)。此外,数字创新还具有自生长性、无边界性等特点。自生长性指数字创新具有动态迭代性和可再塑造性(Yoo et al.,2010),数字创新的非竞争性意味着数字技术的开放和共享,加之数字技术可重复编程的特征,使得数字创新可以自我进化。无边界性指数字创新打破了组织边界和产业边界,不再局限于组织内部,组织外的个体也可以参与到数字创新过程中,主体更加多元化。

(二)数字创新的识别策略

从准确性和全面性来看,衡量创新水平最好的方法是创新的实际经济价值。目前仅有极少数研究基于部分企业的调研数据采用这一方法分析企业的创新行为(崔静波等,2021)。对于更全面的企业面板数据(如工业企业),则很难识别创新的经济效应。专利是创新行为最终形成的具体事物,体现了创新的成果,是衡量创新的重要方面。根据上文数字创新的界定,任何一个专利发明是否属于数字创新可以根据其特征和性质做出判断。因此,只需要找到一个尽可能全面包含数字创新形成的具体事物的范本,并与专利一一对应,即可实现对数字创新的识别。本文发现国家统计局公布的《数字经济及其核心产业统计分类(2021)》定义的数字经济核心产业的描述涵盖的事物可以非常全面地囊括数字创新的成果。因此,本文以数字经济核心产业描述为基础,搭建了其与专利描述之间的桥梁,进行数字创新的识别。

1. 数据介绍

(1) 中国专利数据库

本文在进行数字创新识别时主要使用来源于中国国家知识产权局的专利数据库,涵盖了自

^① 熊彼特(Schumpeter)对创新的定义是:“创新是引入新的产品、服务或改进提供的产品、服务的想法的现实实现。”

1985年《专利法》实施以来至2017年在国家知识产权局申请并公开的所有专利,数据总量超过2400万条。该数据包含专利的基本信息如申请日、专利类型、分类号、公开公告日、申请人、主权项和摘要等。为了后续工作的顺利开展,本文对专利数据进行了一些预处理:第一,删除授权专利重复公开信息;第二,依据专利的地址信息进行城市信息识别;第三,删除不具备数字创新的内涵且摘要缺失严重的外观专利。

(2) 数字经济核心产业分类文本

《数字经济及其核心产业统计分类(2021)》借鉴国内外相关数字经济分类方法建立了相关数字经济行业分类标准。该分类不仅给予数字经济严格的定义,更将其分为五类:数字产品制造业、数字产品服务业、数字技术应用业、数字要素驱动业和数字化效率提升业,前四类被定义为数字经济核心产业。数字经济核心产业不仅代表着数字经济发展基础的数字产业化,也是为产业数字化提供数字技术、产品、服务、基础设施和解决方案的经济活动。综合考虑类别涵盖范畴和分类数量,本文选择数字经济核心产业中类。为了得到语义清晰、表述通顺的行业上中类上下文描述及所涵盖的对象,本文将每个行业中类的行业小类描述合并,并删除“指”等相关连接词、“制造”等动词和存在语义冲突的表述。本文还去掉了没有实际描述性意义的“0205 其他数字产品服务”。最终得到了22个类别及其描述性文本(下文统一使用行业代码指代相关类别)。^①

2. 识别方法

本文的基础性贡献是提供一套数字创新识别方法,为后续分析提供高质量的保障。具体的识别思想是:专利摘要描述了专利的名称、主要技术特征及用途等,通常不超过300字。数字经济核心产业分类详细记录了涵盖的数字技术、产品、服务和基础设施等具体事物及其描述。这些描述全面且具体地介绍了该类别的核心特征以及哪些具体事物属于该类别,并囊括了很多举例,几乎涵盖了所有数字经济的事物对象。如果可以通过专利的描述和数字经济核心产业类别之间的语义关系将专利归类到某一/多个类别对应的对象上,那么数字创新识别的第一步就迎刃而解。如何同时对多达1600余万的专利数据进行识别?本文使用机器学习方法范畴的预训练语言模型SBERT,利用Python进行专利摘要文本和每个类别描述之间的文本相似度计算。

(1) 文本相似度的计算

文本相似度的计算属于自然语言处理(natural language processing, NLP)范畴。随着大规模语料库的出现,基于语料库的统计机器学习方法进入自然语言处理的视野。2018年底BERT的面世,NLP进入了预训练模型的时代。BERT是谷歌AI研究院提出的建立在transformer基础上的一种预训练语言模型,拥有强大的语言表征能力和特征提取能力,打破了11项NLP基准测试任务,成为NLP发展史上的里程碑式的成就。BERT提出了一种全新的框架结构,在预训练后,对输出层进行一个额外的微调,各种下游任务就可以取得优异的表现。Reimers & Gurevych (2019)基于BERT做了修改,提出了Sentence-BERT(SBERT)网络结构,该网络结构利用孪生(siamese)和三级(triplet)网络结构生成具有语义意义的定长句子embedding向量,语义相近句子的embedding向量距离比较近,从而可以用来进行相似度计算。该方法提供了训练超十亿次,包含中文的多语言预训练语言模型,本文基于SBERT进行文本余弦相似度的计算。^②

① 具体类别为:0101 计算机、0102 通讯及雷达设备、0103 数字媒体设备、0104 智能设备、0105 电子元器件及设备、0106 其他数字产品、0201 数字产品批发、0202 数字产品零售、0203 数字产品租赁、0204 数字产品维修、0301 软件开发、0302 电信、广播电视和卫星传输服务、0303 互联网相关服务、0304 信息技术服务、0305 其他数字技术、0401 互联网平台、0402 互联网批发零售、0403 互联网金融、0404 数字内容与媒体、0405 信息基础设施、0406 数据资源与产权交易和0407 其他数字要素。

② 具体来说,我们使用了SBERT提供的paraphrase-multilingual-MiniLM-L12-v2模型进行相似度计算,具体方法等详见<https://www.sbert.net/index.html>。

(2) 数字创新识别

经过计算,本文得到了任一类别与任一专利之间的相似度数值(范围在 0—1 之间)。考虑每个类别相似度分布的差异性,直接使用常见的 0.5 为阈值是不严谨的,因此本文采取严格的人工识别策略和计算分段准确率的方法对每一类别划分阈值。^① 具体而言,对于任一类别,相似度越高的专利被归到该类的概率越高,如果按照相似度从高到低排序,专利被归类的概率存在大幅下降,说明该类别的相似度存在一个分水岭。由于本身无法直接计算专利被归类的概率,综合考虑人工成本和相似度分布范围,本文采用每 0.05 为一分界线,抽取一批分界线上的专利并计算严格准确率,将其作为该相似度临界值上专利属于该类别的概率。

在具体操作上,对于每一类别,从上到下分别对每一分界线向下抽取 100 个专利计算准确率。然后将每一段的准确率从上到下排列,当某一临界值的准确率和下一临界值存在较大差异时,本文会将该临界值作为该类别数字创新识别的阈值。此外,本文还计算了每个类别相似度排名前 100 的专利的准确率,检查使用 SBERT 计算文本相似度是否准确。所有类别前 100 的平均准确率为 91%,可以认为使用 SBERT 计算文本语义相似度是非常稳健的。阈值平均准确率可以达到 64%,而阈值下一临界准确率仅为 36%,存在较大差异,可以认为识别过程比较稳健。^② 保留每个类别高于阈值的专利,共得到 1023661 条数字创新专利。

(三) 数字创新识别合理性检验

考虑到目前没有权威机构公布数字创新相关统计,同时本文使用的数字创新类别对照是数字经济核心产业,即数字产业化部门。因此,本文首先将识别出的数字创新专利在年度层面汇总并与蔡跃洲和牛新星(2021)测度的数字产业化增加值进行比对。结果如图 1 所示,图 1 左侧是数字创新总量与数字产业化增加值的趋势,图 1 右侧是数字创新占比与数字产业化占 GDP 比重的对比。可以发现:数字创新在总量上和数字产业化的增长趋势大致相同,在占比上与数字产业化占比的趋势大致相同,在一定程度上支持本文识别的数字创新是合理的。

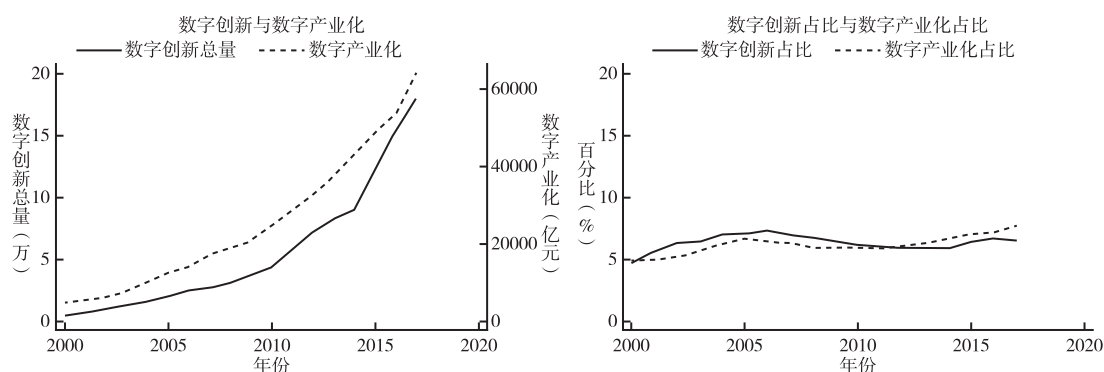


图 1 数字创新与数字经济核心产业对比

基于数字经济越发达的地区数字创新可能越多的前提,本文在地区层面进一步印证识别的合理性。通过专利的地址信息,将数字创新汇总到省和地级市层面。在省份层面,本文借鉴韩兆安等(2021)的省级数字经济规模指标,对比了 2012—2017 年各省数字创新总量和数字经济规模之间

^① 严格的识别方法思路是:对于有详细且精确描述的种类,专利的识别严格按照描述进行,需要符合描述的核心特征;对于只有简要且宽泛描述的种类,主要根据专利的用途和性质,与描述相符的专利可以被划分进去。对于一些具有高度专业词汇描述和不常见不了解的设备均使用搜索引擎和专利检索系统检索出详情和图片来进一步判断。考虑到存在上万条专利识别信息,我们招募了志愿者进行辅助识别。

^② 限于篇幅,每个类别对应的阈值标准未在正文报告,留存备案。

的相关系数,大部分的相关系数都在0.9以上且在1%水平上显著。在城市层面,借鉴赵涛等(2020)的数字经济发展水平的指数体系,构建了地级市数字经济发展指数,并分别对比了2011—2017年各地级市数字创新总量和数字经济发展指数之间的相关系数,大部分的相关系数都在0.7以上且在1%水平上显著。这充分印证了本文数字创新识别的合理性。^①

(四)数字创新特征事实^②

本文也在地区和企业异质性等方面探究了数字创新的特征事实,以期实证分析提供启发:

特征事实一:与传统创新一致,中国数字创新发展呈现区域不平衡性,东部地区贡献了大部分的数字创新;与传统创新不同的是,数字创新呈现更多的集聚效应,在环渤海、长三角和珠三角地区,随着数字经济的发展,北京、上海和深圳都呈现了创新的溢出效应。

中国企业的传统创新呈现区域不平衡特征(寇宗来和刘学悦,2020)。本文根据国家统计局的地区划分,探究各区域数字创新发展的趋势后发现:2000年以来东部地区数字创新不论是规模还是增速上都远远领先于其他地区。更进一步,本文参考寇宗来和刘学悦(2020)将数字创新分别汇总到省份和城市层面,计算数字创新基尼系数。在省份层面,创新基尼系数在0.6左右徘徊;在城市层面,基尼系数大多数年份高于0.8。数字创新呈现出了区域不平衡性。

本文分别计算了以北京、上海和深圳为核心的环渤海地区、长三角地区和珠三角地区的数字创新占比。三大创新区的数字创新总量占比超过了60%,呈现出集聚效应。本文还计算了北京、上海和深圳在各自创新区中的占比来反映“中心—外围”理论中的“虹吸”和“辐射”效应。与传统创新不同的是,北京、上海和深圳在各自区域内都呈现了数字创新的“辐射”效应。

特征事实二:与传统创新一致,民营企业贡献了大部分的数字创新。

不失一般性,本文按照所有制将企业划分为国有、民营、外资与港澳台企业三种类型。相比于其他企业,无论是增速还是规模,民营企业的数字创新专利申请一骑绝尘。这意味着民营企业是中国数字创新的主体,是数字创新发展的动力源泉。

特征事实三:相比传统创新,数字创新的不平衡性在企业规模上明显降低。

本文以注册资本衡量企业规模,在规模维度计算并绘制了创新基尼系数以及注册资本前10%和20%企业的数字创新占比。令人意外的是,数字创新在企业规模维度上没有表现出非常明显的的不平衡性,2017年的基尼系数在0.3左右。这可能与数字创新本身对于资本需求远远低于传统创新活动有关(Guellec & Paunov, 2017)。在过去的十几年中,借助新一代信息技术的发展,中国不断涌现数字技术尤其是互联网企业。不断有初创企业凭借自身独特的技术创新成为细分行业的领头羊。

综上,数字创新是传统创新在数字经济时代的延伸。一方面,数字创新展现出的集聚效应和民营企业主体现象与传统创新一致;另一方面,数字创新表现出异于传统创新的特征,比如数字创新溢出效应更强,数字创新在企业规模维度不存在较大的不平衡性等。

三、理论框架与研究假说

产品市场一体化是指产品在不同地点之间的自由流动程度(Donaldson, 2015),其本质是限制性贸易壁垒的下降。理论上这一壁垒通常指“冰山成本”,不论是自然、技术乃至制度因素导致的市场分割都反映在该成本上。这也是为什么采用价格法来衡量市场分割是现实和有效的,因为价格反映了由限制性壁垒导致的成本加成(吕冰洋和贺颖,2020)。在市场一体化的分析框架中,尽管国际贸易和区域贸易可能在消费者偏好及要素禀赋的设定上存在不同,但标准分析工具都是引

① 限于篇幅,省份和城市层面的对比结果未在正文报告,留存备案。

② 限于篇幅,数字创新特征事实未在正文报告,留存备案。

力模型,本文也借鉴经典引力模型框架分析市场一体化的影响渠道。

产品市场一体化的提高意味着限制性贸易壁垒降低,即“冰山成本” τ_{ij} 降低。^①一方面,在EK模型中,对引力模型推导可知, i 地感知到的目的地 m 的市场规模 $(d_{mi}/p_m)^{-\theta}X_m$ 在其他因素不变的情况下,会随着 d_{mi} 的降低而增加(Eaton & Kortum,2002)。^②另一方面,在异质性企业模型中,当冰山成本下降时,零利润条件不断右移,内生决定的市场中的企业数目增加(Melitz,2003)。^③这意味着市场竞争加剧(Melitz & Ottaviano,2008)。从基本引力模型框架可以看出,市场一体化的影响主要通过需求端市场规模的扩大和供给端市场竞争的加剧两个渠道来实现。

相比传统创新,数字创新具有的更强的非竞争性和更高的无形化特征,意味着与其相关的产品的边际生产成本更低(Corrado et al.,2009;Yoo et al.,2010);数字技术的可分配性赋予了数字创新不再受限于专有资源和创新中介的特性,数字创新一旦完成则可以形成终端产品或应用到终端服务上,加速了产品和服务的市场化进程(Autio et al.,2018)。以上特点意味着数字创新具有更强的规模经济效应。而市场一体化可以扩大市场规模,企业可以销售给更多消费者。面对市场规模的扩张,企业通过加大研发投入,利用数字技术完善产品生产流程,生产数字产品,占领更多的市场份额,从而在规模经济下获得更大收益,弥补了创新投入的沉没成本(Aghion et al.,2022);甚至可能出现“赢者通吃”的格局,进一步激励企业研发投入(Hall & Harhoff,2012),促使企业产出更多更高质量的数字创新成果。

创新是在不断变化的市场竞争环境中可持续发展的核心,经营过程中通过创新来提供与竞争对手不同的产品及服务,才能够取得优势。数字创新主要依赖无形的数字技术,所要求的资本投入相比传统创新大大降低,大量由于门槛所限无法进入创新领域的个体也可以参与到数字创新过程中,提高了“创造性破坏”的概率(Guellec & Paunov,2017)。这种“创造性破坏”体现在数字创新所导致的产品、服务、商业模式乃至组织结构发生重大变化:传统行业中通过应用数字技术并形成以服务为导向的新型产品可以迅速开拓市场,数字创新的赋能效应能够重塑企业重大业务的流程(Fitzgerald et al.,2014),乃至改变价值创造方式实现高质量发展(何帆和刘红霞,2019)。这意味着,企业进行数字创新的不确定性更大,且一旦竞争对手先一步成功进行了数字创新,企业可能血本无归。此外,数字创新还具有更强的规模经济效应和自生长性,自我更迭和传播速度远远高于传统创新。在应对竞争时,企业采用数字创新会极大缩短研发和市场推广周期,且一旦进行了成功的被市场广泛接受的数字创新,依靠数字创新更强的规模经济效应,企业会更迅速地占领市场,赢得竞争。

从上述理论分析看,市场竞争对数字创新的影响比传统创新更加扑朔迷离。当然,如果立足于数字创新的本质特征,并结合市场竞争理论框架,还是能够清晰把握其中脉络。数字创新的进入门槛大大降低意味着大企业才能承担创新成本,市场竞争不利于数字创新,理论可能无法立足。此外,为了从激烈的竞争中脱颖而出,企业很可能借助于数字创新,在产品质量或者生产效率上战胜竞争对手(Correa & Ornaghi,2014)。尽管创新是一个长期过程,但数字创新依靠具有自我迭代性特点的数字技术可以加速产品和服务的应用,意味着应对激烈竞争所采取的数字创新战略见效是迅速的。数字创新更大的规模经济也会加速这一过程,一旦企业进行的数字创新成功迎合市场,其将会迅速占领市场,打败竞争对手。需要注意的是,除了开创性的数字创新外,大部分数字创新是在已有数字技术的基础上迭代演化的,即数字创新符合技术落后者必须赶上前沿技术,才能在未来

① 即从地运输到地的成本,包括运输、交流、税收、制度性限制措施等一切可以反映在价格上的因素。一般由 τ_{ij} 表示,由于习惯不同,在Eaton & Kortum(2002)中使用 d_{it} 。

② 详见Eaton & Kortum(2002)的(11)式及之前的推导,由于篇幅所限,本文没有包含全部推导过程。

③ 详见Melitz(2003)的(20)式、(21)式。

争夺技术领先地位的逐步创新内生增长框架。这意味着市场竞争对企业数字创新的影响可能是倒 U 型的 (Aghion et al., 2005)。

综上所述,市场规模扩大和市场竞争加剧是市场一体化影响企业数字创新的两大渠道。有研究指出,需求端拉动了 60%—80% 的创新 (Utterback, 1974), 意味着不论市场竞争的影响如何,都无法掩盖市场规模对数字创新的拉动作用。同时本文发现市场一体化与数字创新之间存在明显的正相关关系。^① 这说明市场一体化对企业数字创新的总体影响可能是正向的。

基于以上理论分析,本文提出如下三个研究假说:

假说 1: 市场一体化程度的提高对企业数字创新水平的影响是正向的。

假说 2: 市场一体化程度的提高会扩大企业面临的市场规模,从而促进企业数字创新。

假说 3: 市场一体化程度的提高会加剧企业面临的市场竞争,且如果数字创新符合逐步创新模型的设定,市场竞争对其的影响是非线性的倒 U 型。

四、实证分析

(一) 数据说明与实证设计

1. 企业数据匹配

本文需要为识别到的数字创新进行企业层面的匹配。考虑到工业企业数据库样本时间存在限制,以及上市公司数据缺乏小微企业。为了使匹配的企业数据更全面,本文借助于天眼查平台。天眼查企业基本信息数据库涵盖了中国几乎所有企业,每个企业的变量包括公司名称、统一社会信用代码、参保人数、注册资本、实缴资本、成立日期、营业期限、所属省市区、公司类型、所属行业等。

本文通过清理专利申请人与企业基本信息进行全面匹配。在去掉个人申请人后的 972001 个样本观测值中,共匹配上 798790 条数据,匹配成功率高达 82%; 其中有 192572 家企业,匹配上 164100 家,匹配成功率高达 85%。根据每个企业成立时间和退出时间识别企业存续区间,从而构建面板数据。综合考虑市场一体和数字创新专利数据的时间维度,面板数据的时间跨度为 2001—2017 年。此外,本文还借助数据中的经营状态变量去除了迁入迁出的企业,从而保证样本中不存在企业所在地的变动。

2. 变量说明^②

(1) 企业数字创新

对于企业数字创新的数量,本文使用企业每年被识别为数字创新的专利数加 1 取对数来刻画。对于数字创新质量,采用专利被引刻画。具体而言,借鉴 Akcigit et al. (2016) 的方法,采用每个企业数字创新专利平均被引次数以及最大被引次数,并加 1 取对数衡量企业数字创新的质量。考虑到存在右侧“断尾”问题,本文仅采用授权后 5 年作为窗口期计算被引。由于中国国家知识产权局数据库无法准确提供专利被引用信息,本文基于申请号对 Google Patent 进行了爬取。考虑到 5 年窗口期,企业数字创新质量指标截至 2016 年而非 2017 年。

(2) 市场一体化

本文借鉴吕冰洋和贺颖 (2020) 采用价格法计算城市与其所有相邻城市的产品市场分割指数。价格数据是 2001—2017 年的居民消费价格指数,考虑到 2016 年该指数类别发生了细微调整,将 2016 年后的类别与之前的对应,最终得到八类产品部门的价格指数。数据主要来源于各省的统计年鉴、各个城市的统计年鉴和统计公报。为了更细致地考察交通基础设施的作用,本文使用高德地图,用城市对之间驾车距离的倒数作为权重进行加权,得到一个城市与其相邻城市之间的市场分割

① 限于篇幅,未在正文报告,留存备案。

② 限于篇幅,变量的描述性统计未在文中报告,留存备案。

指数 $Markseg_{ct}$ ，其中 c 代表城市， t 代表时间。对于市场一体化的指标，本文参考盛斌和毛其淋 (2011)，根据公式 $MarketInt_{ct} = \sqrt{1/Markseg_{ct}}$ 取对数计算市场一体化程度。

(3) 其他变量

为了获取企业层面的控制变量，本文对其他变量进行了清理：根据企业成立年份，将年份减去该成立年份并加 1 取对数作为企业年龄变量 (Age)。对企业资产变量，天眼查数据中可以使用的是企业注册资本。由于该变量包含不同国家的货币，本文按照注册年份各国货币对人民币汇率将其转换成人民币。对于企业人数变量，清理了参保人数指标。

3. 基准回归模型

基准回归模型为：

$$DigInov_{fct} = \beta_0 + \beta_1 MarketInt_{ct} + \beta_2 Control_{fct} + \eta_f + \eta_t + \varepsilon_{fct} \quad (1)$$

其中，被解释变量 $DigInov_{fct}$ 为年份 t 、所在城市 c 和行业 i 的企业 f 的数字创新衡量指标（数量和质量）。核心解释变量 $MarketInt_{ct}$ 为城市 c 在年份 t 的市场一体化水平。 $Control_{fct}$ 为控制变量，为了在有限可用变量的条件下尽可能控制更多企业特征，本文将企业层面不随时间变化的变量与年份虚拟变量交互，控制更多的企业特征。控制变量包括企业年龄、注册资本与年份虚拟变量的交互、参保人数与年份虚拟变量的交互。 η_f 为企业固定效应， η_t 为时间固定效应。 ε_{fct} 为随机误差项。

(二) 基准结果

表 1 报告了市场一体化对企业数字创新的基准回归结果。考虑到城市行政级别差异会影响地区间的可比性，后文实证分析都使用剔除四个直辖市的市場一体化指标 $MarketInt$ 。第 (1) 列和第 (4) 列是市场一体化对企业数字创新数量 ($Quantity$) 的回归。第 (2) 列和第 (5) 列使用平均被引 (Ave_cite) 作为创新质量衡量指标，第 (3) 列和第 (6) 列使用最大被引 (Max_cite) 作为创新质量衡量指标。结果显示市场一体化可以显著提升数字创新数量和质量。假说 1 得证。

表 1 基准回归结果

变量	(1)	(2)	(3)	(4)	(5)	(6)
	<i>Quantity</i>	<i>Ave_cite</i>	<i>Max_cite</i>	<i>Quantity</i>	<i>Ave_cite</i>	<i>Max_cite</i>
<i>MarketInt_muni</i>	0.0801 *** (0.0135)	0.0499 *** (0.0131)	0.0610 *** (0.0154)			
<i>MarketInt</i>				0.0788 *** (0.0155)	0.0403 *** (0.0149)	0.0505 *** (0.0175)
<i>Age</i>	0.0369 *** (0.001)	0.0167 *** (0.000922)	0.0184 *** (0.00109)	0.0397 *** (0.00104)	0.0168 *** (0.00100)	0.0183 *** (0.00119)
<i>_cons</i>	0.0596 *** (0.00603)	0.0609 *** (0.00582)	0.0770 *** (0.00683)	0.0528 *** (0.00695)	0.0587 *** (0.00667)	0.0742 *** (0.00783)
N	1386631	1236271	1236271	1087497	968456	968456
<i>Capital × Year</i>	√	√	√	√	√	√
<i>Emp × Year</i>	√	√	√	√	√	√
Year Fixed Effect	√	√	√	√	√	√
Firm Fixed Effect	√	√	√	√	√	√
R ² _Adj	0.19	0.12	0.16	0.18	0.12	0.15

注： $Quantity$ 为企业数字创新数量的衡量指标， Ave_cite 为利用平均被引计算企业数字创新质量的衡量指标， Max_cite 为利用最大被引计算企业数字创新质量的衡量指标。 $Capital \times Year$ 为注册资本和年份虚拟变量的交互， $Emp \times Year$ 为参保人数与年份虚拟变量的交互。回归中控制了年份固定效应和企业固定效应。***、** 和 * 分别表示 1%、5% 和 10% 的水平上显著；括号内为稳健标准误。下同。

(三) 异质性分析^①

第一, 本文将数字创新分为发明专利和实用新型专利, 回归结果如表 2 第 (1) — (4) 列所示。结果表明, 市场一体化对数字创新发明专利的质量和数量以及实用新型专利的数量都有显著促进作用。本文还将数字创新分为数字产品和数字生产/服务流程创新。^② 回归结果如表 2 第 (5) — (8) 列所示。市场一体化显著促进企业进行更多更高质量的数字产品和数字生产/服务流程类型创新。

表 2 分类回归

变量	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	发明		实用新型		数字产品		数字生产和服务	
	Quantity	Ave_cite	Quantity	Ave_cite	Quantity	Ave_cite	Quantity	Ave_cite
MarketInt	0.0491 *** (0.0101)	0.0361 *** (0.0128)	0.0320 *** (0.0122)	0.00922 (0.0112)	0.0698 *** (0.0145)	0.0351 ** (0.0141)	0.0354 *** (0.00824)	0.0301 *** (0.00930)
_cons	0.0364 *** (0.00455)	0.0457 *** (0.00575)	0.0347 *** (0.00549)	0.0339 *** (0.00502)	0.0466 *** (0.00652)	0.0518 *** (0.00630)	0.0172 *** (0.00370)	0.0193 *** (0.00417)
N	1087497	968456	1087497	968456	1087497	968456	1087497	968456
其他控制变量	✓	✓	✓	✓	✓	✓	✓	✓
Capital × Year	✓	✓	✓	✓	✓	✓	✓	✓
Emp × Year	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓	✓	✓	✓	✓
R ² _Adj	0.25	0.14	0.13	0.08	0.17	0.11	0.24	0.15

第二, 为检验市场一体化对数字创新的促进效应是否存在地区异质性, 本文分东部和其他地区样本回归进行探究。结果显示市场一体化对东部地区的数字创新的促进作用显著, 对中西部 and 东北地区企业数字创新的作用是不显著的。当前城市群成为区域经济发展的新型空间组织和重要战略单元, 相比于东部地区, 中西部地区城市群和经济带的建设存在滞后性, 对周边城市的溢出效应不明显, 可能在一定程度上解释东部和其他地区效果的差异。

第三, 为检验市场一体化对数字创新的促进作用是否存在企业所有制的异质性, 本文就国有企业、外资及港澳台企业和民营企业进行分样本回归, 发现市场一体化水平的提高对民营企业数字创新促进作用显著, 对国有企业和外资企业不显著。

第四, 为检验市场一体化对数字创新的促进作用是否存在企业规模的异质性, 本文使用注册资本和参保人数作为企业规模的衡量标准, 通过加入企业规模的交互项探究企业规模的异质性。市场一体化与企业规模的交互项系数都为负且在统计上显著, 说明市场一体化对数字创新的推动作用在中小规模的企业更大。

(四) 稳健性检验^③

第一, 考虑到交通基础设施可能包括铁路、航运等除公路外的重要设施, 本文使用城市间直线

① 限于篇幅, 部分异质性分析表格未在文中报告, 留存备案。

② 根据数字经济核心产业统计分类的定义, 原则上可以将数字产品制造业的数字创新归类到数字产品创新上来, 将数字产品服务业、数字技术应用业和数字要素驱动业的数字创新归到数字生产/服务流程创新上。

③ 限于篇幅, 部分稳健性检验表格未在文中报告, 留存备案。

距离的倒数将城市对的价格变动在相邻城市加权平均,并计算市场一体化指数,将其作为稳健性指标进行回归,回归结果与基准回归中一致。

第二,考虑到市场一体化与企业数字创新可能存在遗漏变量或反向因果等导致的内生性问题,本文通过控制行业时间固定效应,消除随行业变动的遗漏政策等的影响。

第三,本文采用城市平均海拔和平均坡度作为市场一体化的工具变量来排除内生性的影响(吕越等,2018;卞元超和白俊红,2021)。海拔高度和坡度反映了该地区与外界之间流动的地理条件,会通过影响地区间贸易成本,进而在一定程度上影响市场一体化。地理条件越复杂、地形起伏越大的地区,与外界的市场一体化程度可能越低,满足相关性要求。同时,海拔高度和坡度属于地理范畴,满足排他性约束。考虑到每个城市海拔和坡度几乎不随时间变化,本文采用剔除该城市后其他城市的市场一体化趋势作为外生时间趋势,与海拔和坡度交互得到最终的工具变量。表3第(1)列报告了第一阶段结果,第(2)一(4)列报告了两阶段回归结果。工具变量回归显示结果是稳健的。Kleibergen-Paap Wald F 等统计量表明不存在弱工具变量问题。过度识别 Sargan 检验的统计量对应的 p 值明显大于 0.1,说明工具变量都是外生的。

表 3 内生性检验

变量	(1)	(2)	(3)	(4)
	<i>MarketInt</i>	<i>Quantity</i>	<i>Ave_cite</i>	<i>Max_cite</i>
<i>IV1_altitude</i>	0.0115 *** (0.000276)			
<i>IV2_gradient</i>	0.00872 *** (0.000745)			
<i>MarketInt</i>		1.895 *** (0.183)	0.526 *** (0.181)	0.929 *** (0.217)
N	1205336	1205336	1075700	1075700
Wald F		2406.972	2054.986	2054.986
Hansen J		0.248	0.561	0.556
其他控制变量	√	√	√	√
Ind-Year Fixed Effect	√	√	√	√
Firm Fixed Effect	√	√	√	√

注:*IV1_altitude* 为海拔高度工具变量,*IV2_gradient* 为坡度工具变量。

(五) 机制分析

本文借鉴江艇(2022)对渠道检验的论述,主要利用工具变量回归考察市场一体化对机制变量(市场规模和市场竞争)的因果关系来探究影响渠道。本文在构建机制变量后,基于最严格的固定效应回归单独检验了机制变量与被解释变量之间的关系。

1. 市场规模

综合考虑数据可得性,本文使用城市社会消费品零售总额,并借鉴 Aghion et al. (2022)的方法衡量一个城市及其所有相邻城市的需求变动:

$$\Delta Demand_{ct} = \sum_z \frac{Cosum_{zt}}{\sum_z Cosum_{zt}} \frac{Cosum_{zt} - Cosum_{zt-1}}{(Cosum_{zt} + Cosum_{zt-1})/2}$$

(2)

其中, $Cosum_{zt}$ 为城市 z 在年份 t 的社会消费品零售总额, z 的取值集合为城市 c 及其所有相邻城市。该公式可以被解读为城市 c 所有企业面临的市场需求变动, 由城市 c 及其所有相邻城市组成的小城市群中的每个城市的需求变动及自身在该城市群中重要性加权得到。

表 4 报告了市场规模渠道的回归结果。其中第(1)列为城市时间层面基于工具变量法探究市场一体化对市场规模的影响, 控制了城市和时间的固定效应。第(2)列是市场规模对企业数字创新数量的回归结果, 第(3)、(4)列为市场规模对企业数字创新质量的回归结果。结果显示市场一体化对市场规模的变动有正向且显著的影响, 证实了市场规模是市场一体化对企业数字创新影响的渠道之一。市场规模对企业数字创新数量和质量回归结果显著为正, 意味着市场规模的扩大可以显著促进企业数字创新, 证实了假说 2。此外, 通过划分数字产品和数字生产/服务流程, 本文发现市场规模对数字产品创新(数量和质量)正向作用更显著。^① 这一结果也符合规模效应的经济直觉, 市场规模扩大时企业会面对更多消费者, 企业往往会将重点放在数字产品创新上, 生产出更先进的产品从而迅速占领市场。

表 4 市场规模机制

变量	(1)	(2)	(3)	(4)
	$\Delta Demand$	$Quantity$	Ave_cite	Max_cite
$MarketInt$	0.0129 *** (0.00423)			
$\Delta Demand$		0.0760 *** (0.00845)	0.0195 ** (0.00784)	0.0278 *** (0.00922)
$_cons$		0.0806 *** (0.00195)	0.0787 *** (0.00187)	0.0986 *** (0.00221)
N	4797	1423077	1272760	1272760
其他控制变量		√	√	√
Wald F	1642.574			
Hansen J	0.0360			
adj_R ²		0.19	0.12	0.16

注: 第(1)列控制了城市、时间固定效应, 第(2)一(4)列控制了企业、行业时间固定效应。

2. 市场竞争

本文采用常见的企业竞争对手数量衡量市场竞争。一般而言, 竞争对手越多市场竞争越激烈。本文利用天眼查数据计算了每个城市、行业每年存续的企业数, 并用城市 c 及其所有周边城市在行业 i 的企业数目, 作为城市 c 行业 i 的企业面临的市场竞争程度。

假说 3 提到如果数字创新符合逐步创新模型的设定, 市场竞争对其的影响是非线性的倒 U 型。因此, 在进行市场竞争机制分析前, 本文需要检验数字创新是否符合逐步创新模型的设定。在此, 主要考察企业过去积累的数字创新是否对当前的数字创新有影响。本文计算了每个企业该期之前累积的数字创新数量, 将其作为解释变量加入基准回归中, 结果如表 5 所示。结果发现, 一个企业过去累积的数字创新对当前的数字创新(数量和质量)的影响显著为正, 说明了数字创新符合逐步创新的设定。那么, 市场竞争对其的影响极有可能是非线性的。

① 限于篇幅未在文中报告, 留存备案。

表 5 技术累积效应的回归结果

变量	(1)	(2)	(3)
	<i>Quantity</i>	<i>Ave_cite</i>	<i>Max_cite</i>
<i>MarketInt</i>	0.0758 *** (0.0155)	0.0392 *** (0.0149)	0.0481 *** (0.0175)
<i>DigitalInnovation_Past</i>	0.0948 *** (0.0183)	0.0341 *** (0.00772)	0.0686 *** (0.0144)
_cons	0.0509 *** (0.00713)	0.0581 *** (0.00667)	0.0730 *** (0.00782)
N	1087497	968456	968456
Capital × Year	√	√	√
Emp × Year	√	√	√
Ind – Year Fixed Effect	√	√	√
Firm Fixed Effect	√	√	√
adj_R ²	0.19	0.12	0.16

注：*DigitalInnovation_Past* 为企业过去累积的数字创新数量（对数形式）。回归控制了行业年份固定效应和企业固定效应。

表 6 第(1)列是基于工具变量考察市场一体化对市场竞争的影响，控制了城市行业和行业时间固定效应，结果表明市场一体化会提高市场竞争程度。表 6 第(2)–(4)列是控制了企业、行业时间和城市时间固定效应下市场竞争对数字创新数量和质量回归结果，市场竞争的系数为负，且仅在对数字创新数量的回归中显著（5%水平上）。然而，上文已验证数字创新遵循经典内生增长模型的逐步创新假定，这可能意味着线性结果掩盖了非线性事实。

因此，本文加入市场竞争的二次项考察市场竞争与数字创新是否存在非线性关系，表 6 第(5)–(7)列报告了回归结果。结果显示，市场竞争对数字创新（数量和质量）的影响是倒 U 型的，即市场竞争对数字创新的促进作用随着市场竞争的提高先扩大再减小。为了验证这种倒 U 型关系是否真的存在，本文首先考察了市场竞争的分布和倒 U 型中点位置之间的联系：市场竞争（竞争企业数的对数形式）的最小值为 4.2，最大值为 16.8，均值为 12.9，而倒 U 型中点约为 12；可以发现市场竞争的取值范围涵盖了倒 U 型最高点位置，且该位置较靠近分布的中心。其次，本文对倒 U 型关系进行了二次回归。^① 具体做法是：首先确定倒 U 型关系的最高点，然后在该断点处对之前和之后的市场竞争分别生成新变量进行断点回归，如果在该点之前的市场竞争的系数显著为正，且在该点之后的市场竞争系数显著为负，则可以证明倒 U 型关系。^② 市场竞争与企业数字创新的倒 U 型关系意味着市场竞争对数字创新的促进作用存在最优水平，只有适度且合理的竞争才可以最大程度促进企业数字创新“提质增量”。

此外，本文还分别考察了市场竞争对数字产品创新和数字生产/服务流程创新的影响。^③ 结果发现市场竞争对数字产品创新（数量和质量）的影响也符合倒 U 型关系，经过二次回归的检验，倒 U 型关系仍然稳健；对于数字生产/服务流程的影响总体来看是正向的。

① 参见 Simonsohn & Nelson 提出的对于 U 型的挑战及检验，详见 http://datacolada.org/27#footnote_3_720。

② 例如，市场一体化对企业数字创新数量的结果，最高点前市场竞争的系数为 0.0116（1%水平上显著），最高点后的市场竞争系数为 -0.00580（5%水平上显著），证实了倒 U 型关系的存在。

③ 限于篇幅未在文中报告，留存备案。

表 6 市场竞争机制

变量	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	<i>MarketCom</i>	<i>Quantity</i>	<i>Ave_cite</i>	<i>Max_cite</i>	<i>Quantity</i>	<i>Ave_cite</i>	<i>Max_cite</i>
<i>MarketInt</i>	0.00947 *** (0.00257)						
<i>MarketCom</i>		-0.00961 ** (0.00473)	-0.00662 (0.00456)	-0.00866 (0.00552)	0.135 *** (0.0160)	0.0766 *** (0.0150)	0.115 *** (0.0184)
<i>MarketCom</i> ²					-0.00566 *** (0.000590)	-0.00327 *** (0.000559)	-0.00486 *** (0.000684)
_cons		0.221 *** (0.0612)	0.170 *** (0.0587)	0.218 *** (0.0711)	-0.685 *** (0.115)	-0.351 *** (0.108)	-0.558 *** (0.132)
N	111820	1423562	1273200	1273200	1423562	1273200	1273200
其他控制变量		√	√	√	√	√	√
Wald F	37828.354						
Hansen J	8.569						
adj_R ²		0.20	0.12	0.16	0.20	0.12	0.16

注：第(1)列控制了行业城市、行业时间固定效应，第(2)一(7)列控制了企业、行业时间、城市时间固定效应。

五、结论与政策建议

本文创新性使用自然语言处理与机器学习交叉范畴的预训练语言模型进行数字创新识别工作。在验证了识别方法的稳健性后，本文在地区、企业异质性方面探究了数字创新的特征事实，为实证检验提供了启发和思路。本文还从市场与创新的关系出发，立足于全国统一大市场，从理论上分析了市场一体化对企业数字创新的影响及其机制。在理论分析的启发下，本文实证探究了市场一体化对企业数字创新的影响及其渠道。结果发现：市场一体化水平可以显著促进企业数字创新数量和质量的提升，对细分的发明和实用新型专利、数字产品和数字生产/服务流程创新也有显著作用，且该影响效应在地区、企业性质、企业规模等方面存在异质性。此外，本文还拓展分析了市场一体化对企业数字创新的影响渠道：市场一体化带来的市场规模扩大和市场竞争加剧是影响企业数字创新的重要渠道；市场规模的扩大促进数字创新数量和质量两方面的提升；而市场竞争对企业数字创新的影响是非线性的倒 U 型。本文对数字创新的识别和测度提供科学合理的方法，为后续数字创新的研究奠定了重要基础。

当前中美博弈日趋复杂，美国对华科技封锁日益激烈，本文为如何建设全国统一大市场，完善技术创新市场导向机制，从而突破数字创新“卡脖子”困境提供了重要的理论和实证依据，政策启示非常丰富：

(1) 坚持技术创新市场导向机制，加快国内大循环和全国统一大市场的建设，发挥超大规模市场优势，强化市场竞争的基础地位，以市场为导向提高企业数字创新基础研发能力，从而实现数字创新的“提质增量”，实现经济高质量发展。

(2) 在推动全国统一大市场建设过程中，不仅要把重点放在需求端的市场规模，更要把重点放在供给端的市场竞争。本文的研究指出了把重点放在市场需求端的不足：市场规模仅仅能促进数

字产品创新,而对数字生产或服务流程创新并不具备影响效应。中国当前市场需求不振,作用也有限。经过新冠疫情的打击,短期内居民收入提升困难。因此,不仅要把目光放在需求端,也要寻求供给端的替代措施。本文指出,不仅要健全公平竞争的制度框架,更要引导合理有序的市场竞争结构,将中国集中力量办大事的制度优势和市场机制有机结合,建立公平有序且合理的竞争政策与产业政策联动机制,杜绝恶性竞争,不能依赖于价格战去占领市场,最大化市场竞争对数字创新的推动作用。当前中国制造业低价竞争和过度竞争已经是不可忽视的现象,企业利润率低下,无法实现长远发展。因此,促进竞争不足行业的充分竞争,引导竞争过度行业的合理竞争不失为当前需求不振时促进数字创新发展的发力点,例如,可以利用数字技术释放的个性化需求细分市场,引导企业差异化竞争。只有合理的竞争环境才能最大程度上避免技术创新的重复投入,实现优势科技资源的有效整合,提升基础研究、核心技术、原始创新、协同创新的能力。

(3)需要关注数字创新的地区不平等及其原因。推动城市群一体化建设尤其是成渝、长江中游等中西部城市群,构建创新共享平台,提高中心城市的数字创新溢出效应。同时也要推动各地区尤其是东部和其他地区的科技资源共享,鼓励不同区域之间科技信息交流互动。例如,在“东数西算”工程中,不仅要实现西部的算力资源更充分地支撑东部数据的运算,更需要东部与西部建立科技资源的共享机制,实现东西共创的创新互助机制。

(4)大力推动中小企业数字创新。在全国统一大市场的建设中,加强对创新型中小企业知识产权的保护。中小企业是中国数字创新的主力军,适当给予保护,防止被大企业恶意兼并挤压非常重要。此外,也要发挥大企业和平台作用,引导大企业就基础性数字创新与小企业形成互补联系,完善企业数字创新服务体系。推进高校、科研机构 and 平台在基础性数字技术方面对企业进行开放,形成良性的知识溢出,为企业尤其是中小企业的数字创新技术积累提供帮助。

(5)助力民营企业数字创新。在统一大市场建设中为民营企业营造良好的营商环境,助力民营企业数字创新。此外,国有企业在统一大市场的建设中应该发挥表率作用,深耕数字创新领域,为国家整体的数字创新能力保驾护航。

参考文献

- 安同良、姜妍,2021:《中国特色创新经济学的基本理论问题研究》,《经济学动态》第4期。
- 卞元超、白俊红,2021:《市场分割与中国企业的生存困境》,《财贸经济》第1期。
- 蔡跃洲、牛新星,2021:《中国数字经济增加值规模测算及结构分析》,《中国社会科学》第11期。
- 崔静波、张学立、庄子银、程郁,2021:《企业出口与创新驱动——来自中关村企业自主创新数据的证据》,《管理世界》第1期。
- 韩兆安、赵景峰、吴海珍,2021:《中国省际数字经济规模测算、非均衡性与地区差异研究》,《数量经济技术经济研究》第8期。
- 何帆、刘红霞,2019:《数字经济视角下实体企业数字化变革的业绩提升效应评估》,《改革》第4期。
- 江艇,2022:《因果推断经验研究中的中介效应与调节效应》,《中国工业经济》第5期。
- 寇宗来、刘学悦,2020:《中国企业的专利行为:特征事实以及来自创新政策的影响》,《经济研究》第3期。
- 吕冰洋、贺颖,2020:《迈向统一市场:基于城市数据对中国商品市场分割的测算与分析》,《经济理论与经济管理》第4期。
- 吕越、盛斌、吕云龙,2018:《中国的市场分割会导致企业出口国内附加值率下降吗》,《中国工业经济》第5期。
- 欧阳峤、汤凌霄,2017:《大国创新道路的经济学区解》,《经济研究》第9期。
- 卿陶、黄先海,2021:《国内市场分割、双重市场激励与企业创新》,《中国工业经济》第12期。
- 盛斌、毛其淋,2011:《贸易开放、国内市场一体化与中国省际经济增长:1985—2008年》,《世界经济》第11期。
- 余江、孟庆时、张越、张兮、陈凤,2017:《数字创新:创新研究新视角的探索及启示》,《科学学研究》第7期。
- 张杰、毕钰、金岳,2021:《中国高新区“以升促建”政策对企业创新的激励效应》,《管理世界》第7期。
- 张杰、陈志远、杨连星、新夫,2015:《中国创新补贴政策的绩效评估:理论与证据》,《经济研究》第10期。
- 张杰、郑文平,2018:《创新追赶战略抑制了中国专利质量么?》,《经济研究》第5期。
- 张杰、周晓艳,2011:《中国本土企业为何不创新——基于市场分割视角的一个解读》,《山西财经大学学报》第6期。
- 赵涛、张智、梁上坤,2020:《数字经济、创业活跃度与高质量发展——来自中国城市的经验证据》,《管理世界》第10期。

- 诸竹君、黄先海、王毅, 2020, 《外资进入与中国式创新双低困境破解》, 《经济研究》第 5 期。
- Acemoglu, D., and J. Linn, 2004, “Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry”, *Quarterly Journal of Economics*, 119(3), 1049—1090.
- Aghion, P., and P. Howitt, 1992, “A Model of Growth through Creative Destruction”, *Econometrica*, 60(2), 323—351.
- Aghion, P., A. Bergeaud, M. Lequien, and M. J. Melitz, 2022, “The Heterogeneous Impact of Market Size on Innovation: Evidence from French Firm-Level Exports”, *Review of Economics and Statistics*, 1—56.
- Aghion, P., N. Bloom, R. Blundell, R. Griffith, and P. Howitt, 2005, “Competition and Innovation: An Inverted-U Relationship”, *Quarterly Journal of Economics*, 120(2), 701—728.
- Akcigit, U., S. Baslandze, and S. Stantcheva, 2016, “Taxation and the International Mobility of Inventors”, *American Economic Review*, 106(10), 2930—2981.
- Akcigit, U., J. Grigsby, T. Nicholas, and S. Stantcheva, 2022, “Taxation and Innovation in the Twentieth Century”, *Quarterly Journal of Economics*, 137(1), 329—385.
- Atkeson, A., and A. T. Burstein, 2010, “Innovation, Firm Dynamics, and International Trade”, *Journal of Political Economy*, 118(3), 433—484.
- Autio, E., S. Nambisan, L. D. Thomas, and M. Wright, 2018, “Digital Affordances, Spatial Affordances, and the Genesis of Entrepreneurial Ecosystems”, *Strategic Entrepreneurship Journal*, 12(1), 72—95.
- Autio, E., and L. D. Thomas, 2020, “Value Co-creation in Ecosystems: Insights and Research Promise from Three Disciplinary Perspectives”, *Handbook of Digital Innovation* (pp. 107—132), Edward Elgar Publishing.
- Corrado, C., C. Hulten, and D. Sichel, 2009, “Intangible Capital and US Economic Growth”, *Review of Income and Wealth*, 55(3), 661—685.
- Correa, J. A., and C. Ornaghi, 2014, “Competition & Innovation: Evidence from U. S. Patent and Productivity Data”, *Journal of Industrial Economics*, 62(2), 258—285.
- Donaldson, D., 2015, “The Gains from Market Integration”, *Annual Review of Economics*, 7(1), 619—647.
- Dorn, D., G. H. Hanson, G. Pisano, and P. Shu, 2020, “Foreign Competition and Domestic Innovation: Evidence from US Patents”, *American Economic Review: Insights*, 2(3), 357—74.
- Eaton, J., and S. Kortum, 2002, “Technology, Geography, and Trade”, *Econometrica*, 70(5), 1741—1779.
- Fichman, R. G., B. L. D. Santos, and Z. Zheng, 2014, “Digital Innovation as A Fundamental and Powerful Concept in the Information Systems Curriculum”, *MIS Quarterly*, 38(2), 329—354.
- Fitzgerald, M., N. Kruschwitz, D. Bonnet, and M. Welch, 2014, “Embracing Digital Technology: A New Strategic Imperative”, *MIT Sloan Management Review*, 55(2), 1.
- Guellec, D., and C. Paunov, 2017, “Digital Innovation and the Distribution of Income”, NBER Working Paper, 23987.
- Hall, B. H., and D. Harhoff, 2012, “Recent Research on the Economics of Patents”, *Annual Review of Economics*, 4(1), 541—565.
- Klette, T. J., and S. Kortum, 2004, “Innovating Firms and Aggregate Innovation”, *Journal of Political Economy*, 112(5), 986—1018.
- Melitz, M. J., 2003, “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity”, *Econometrica*, 71(6), 1695—1725.
- Melitz, M. J., and G. I. Ottaviano, 2008, “Market Size, Trade, and Productivity”, *Review of Economic Studies*, 75(1), 295—316.
- Nambisan, S., K. Lyytinen, A. Majchrzak, and M. Song, 2017, “Digital Innovation Management: Reinventing Innovation Management Research in A Digital World”, *MIS Quarterly*, 41(1), 223—238.
- Nambisan, S., M. Wright, and M. Feldman, 2019, “The Digital Transformation of Innovation and Entrepreneurship: Progress, Challenges and Key Themes”, *Research Policy*, 48(8), 103773.
- Reimers, N., and I. Gurevych, 2019, “Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks”, arXiv preprint arXiv:1908.10084.
- Utterback, J. M., 1974, “Innovation in Industry and the Diffusion of Technology”, *Science*, 183(4125), 620—626.
- Yoo, Y., O. Henfridsson, and K. Lyytinen, 2010, “Research Commentary: The New Organizing Logic of Digital Innovation: An Agenda for Information Systems Research”, *Information Systems Research*, 21(4), 724—735.

The Impact of Market Integration on Firm Digital Innovation: An Analysis on Digital Innovation Measurement Methods

HU Zengxi and MA Shuzhong

(China Academy of Digital Trade, Zhejiang University)

Summary: The 21st century has witnessed a thriving digital economy, where continuous digital innovation serves as the primary engine for high-quality development. However, during the crucial transition from consumer Internet to industrial Internet in China, the complexity of China-US competition has been escalating. The technology sanctions on China led by Western countries have resulted in significant challenges such as the cut-off of chip supply due to China's insufficient capacity for digital innovation in key fields. The greatest incentive for technological innovation comes from a well-established market environment. Due to market segmentation, there exists a mismatch between market demand and innovation in China, hindering "demand-induced innovation" and suppressing a firm's innovative behavior. Compared to traditional innovation, digital innovation exhibits more substantial economies of scale and a higher probability of "creative destruction." On the demand side, the larger market scale brought by market integration can amplify the economies of scale, encouraging businesses to pursue digital innovation. On the supply side, the increased number of competitive entities due to market integration can enhance the "creative destruction" effect, promoting technological advancement. This implies that a unified national market might become a critical factor driving a firm's digital innovation.

The purpose of this paper is to provide a quantitative research method for digital innovation and explore the impact of market integration on a firm's digital innovation both theoretically and empirically. In terms of quantifying digital innovation, this paper innovatively employs a pre-trained language model. Based on the China Patent Full-text Database from 1985 to 2017 and the textual descriptions of core digital economy industries, the semantic similarities between each patent's abstract text and each digital economy industry description are calculated. Digital innovation is then identified by setting thresholds through strict manual identification methods. This paper investigates the basic characteristics of digital innovation and analyzes the mechanisms of market integration based on a gravity model framework, pointing out that market scale and competition are important influencing channels. In the empirical analysis, digital innovation firms' panel data are obtained by matching the applicants of digital patents with all firms registered in the industrial and commercial register in China, and a quality indicator is constructed using patent citation data for further analysis.

This paper's findings are as follows: the agglomeration effect and the dominance of private firms in digital innovation are consistent with traditional innovation; digital innovation exhibits unique features compared to traditional innovation, such as stronger spillover effects and less imbalance along the dimension of firm size. The empirical results show that the improvement of market integration level can significantly promote both the quantity and quality of a firm's digital innovation, and this effect varies across regions, the nature of firm, and firm size. The extended analysis of the mechanism reveals that the expansion of market scale and the intensification of market competition brought about by market integration are important channels influencing a firm's digital innovation. The expansion of market scale promotes both the quantity and quality of digital innovation, while the effect of market competition on digital innovation follows an inverse U-shaped curve.

The contributions of this paper are threefold. First, it provides an innovative and targeted method for measuring digital innovation, which has mainly been qualitatively analyzed in existing research. This paper innovatively introduces a machine learning method using a pre-trained language model, laying the foundation for interpreting and quantifying digital innovation and providing an essential basis for future research. Second, this paper focuses on product market integration, incorporating the novel features of digital innovation into the analysis framework of market integration, thereby addressing the gaps in neoclassical economics research on the relationship between the market and innovation. This gives the traditional theory a new connotation in the digital economy era. Third, this paper examines the impact of market integration on firm digital innovation from both quantity and quality perspectives, enriching related research on firm innovation.

Keywords: Digital Innovation; Natural Language Processing; Market Integration; Market Scale; Market Competition

JEL Classification: O12, O31, R12

(责任编辑:王利娜)(校对:王红梅)