

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

内容摘要：数字创新是经济高质量发展的重要推动力量，全国统一大市场会放大数字创新的规模经济和创造性破坏效应，从而成为在科技创新领域被“卡脖子”的困境下引领企业数字创新的关键因素。当前数字创新量化研究凤毛麟角，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 至 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企业数据库进行匹配。然后，本文以全国统一大市场与数字创新之间关键的联系为切入点，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。本文为全国统一大市场的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。此外，本文首次为数字创新提供了创新性的衡量方法，为后续研究奠定了重要微观基础。

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

二十一世纪以来，数字经济蓬勃发展。中国虽然数字经济规模位居全球第二^①，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。产生这种差异的重要原因是在当前中国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，尤其是以美国为首的西方国家对中国的科技封锁，从对中兴、华为等中国科技企业的全面制裁到 2022 年“芯片四方联盟”和芯片法案的通过，无一例外都是遏制中国科技创新能力提升的企图，加之我国部分基础和关键领域数字创新能力不足^②，从而使得我国在数字经济领域遭遇了芯片断供之类诸多“卡脖子”问题^③。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》基本原则第一点就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，该规划中“创新”一词出现了总共 52 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的重要性。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^④。《中华人民共和国国民经济和社会发展第十四个五年规划和 2035 年远景目标纲要》在第五章中首先就

^① 数据来源于中国信息通信研究院 2021 年发布的《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》，中国科技创新能力只能排名第 14，与数字经济规模世界第二的排名存在明显差距。

^③ 在半导体领域被“卡脖子”现象尤为突出：中国缺乏高精尖芯片制造商，只能大量依靠进口芯片。据中国海关数据显示，2021 年中国进口集成电路 6355 亿颗，同比增长 16.9%。

^④ 《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%。

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，从而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，以微观层面的企业为切入点探究如何引领数字创新，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、发展动力和创新激励。随着中美博弈日趋复杂，逆全球化浪潮不断涌现，加之新冠肺炎疫情的冲击，全球市场动荡，需求萎靡。与此同时，随着中国科技水平的提升，通过依靠国际市场的外循环实现“贸易引致学习”进而提升创新能力的途径大打折扣。中国作为拥有巨大且稳定内需市场的大国，全国统一大市场的建设可能成为其实现自主创新能力提升的重要抓手。国务院《“十四五”数字经济发展规划》指出要“发挥我国超大规模市场优势，提高数字技术基础研发能力”。超大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（欧阳峤和汤凌霄，2017）。《中共中央 国务院关于加快建设全国统一大市场的意见》的发展目标中也指出要“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”。当前，我国超大规模市场对技术创新、产业升级的激励效应还不充分。受限于市场分割和区域壁垒，我国市场需求与创新之间存在错配，无法实现“需求引致创新”，创新资源流动受阻，企业创新行为受到抑制（张杰和周晓艳，2011；王磊和邓芳芳，2016；冯艺和连俊华，2020；吕越等，2021；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和合理配置，发挥市场在资源配置中的作用，有利于促进竞争、优化分工、激励创新，增强企业创新能力^②。这也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市场环境对企业创新的引领作用。

相比传统创新，数字创新具有更强的非竞争性从而产生更大的规模经济效应，准入门槛也更低从而具有更高的“创造性破坏”概率（Benkler, 2006; Guellec & Paunov, 2017）。在需求端，全国统一大市场所带来的更大市场规模会放大数字创新的规模经济效应，激励企业进行数字创新；在供给端，全国统一大市场所带来的更多的竞争主体会放大“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、实现技术升级的核心因素，甚至成为打破高端芯片等数字技术领域“卡脖子”困境的关键突破口。市场一体化与全国统一大市场一脉相承。《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，加快建设高效规范、公平竞争、充分开放的全国统一大市场。”从经济学视角来看，这意味着国内各个城市、省份和地区之间的市场一体化程度不断加深^③，市场一体化是建设全国统一大市场的关键。因此，本文在合理地识别和衡量数字创新后，主要以我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，本文的基础性贡献是为数字创新的识别和测度提供创新性和有针对性的方法。“工欲善其事，必先利其器”，只有科学准确地测度数字创新才能更好地为经济学分析服务。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，国内外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; Fichman et al., 2014; Guellec & Paunov, 2017）。

^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。

数字创新不像数字经济可以用计量分析等工具测度并衡量（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成果可以根据其特征和性质做出是与否的判断。本文创新性引入预训练语言模型的机器学习方法，为解读和量化数字创新这一新生事物奠定基础，同时针对性地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，本文将视角聚焦于产品市场一体化，结合数字创新这一全新事物的全新特征，将其纳入到市场一体化的分析框架内，对新古典经济学中市场与创新的关系进行再审视，是对该领域的重要拓展，赋予了传统理论数字经济时代的全新内涵。创新是增长的基础，创新理论最先由熊彼特提出（Schumpeter, 1934），而后逐渐形成创造性破坏的内生化创新的增长模型（Romer, 1990; Aghion & Howitt, 1992）。随着异质性企业理论的提出，创新增长理论也逐渐过渡到企业层面（Klette & Kortum, 2004; Akcigit & Kerr, 2018）。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（Doh & Kim, 2014; Guo et al., 2016; 张杰等, 2015），制度环境（王永进和冯笑, 2018; 张杰等, 2021）等政策因素以及国际贸易（Atkeson & Burstein, 2010）都是企业创新的影响因素。但从企业创新活动的本质来看，市场才是企业创新的关键影响因素（张杰和周晓艳, 2011）^①。从市场的买卖双方两大参与主体来看，供给端和需求端都会对企业创新行为产生影响（Romer, 1990）。需求端市场规模的扩大会促进企业的研发创新（Acemoglu & Linn, 2004; 范红忠, 2007; 钱学锋等, 2021）。供给端市场竞争对企业研发创新的影响尚未有定论，现有文献的结论主要有抑制作用（Schumpeter, 1942; Dorn et al., 2020），促进作用（Correa & Ornaghi, 2014; 张杰等, 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）这两个本质差异以及由此衍生的自生长性、不确定性、无边界性等特点（Corrado et al., 2005, 2009; Benkler, 2006）。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数字经济时代的全新内涵。

第三，本文基于数量和质量双重视角考察市场一体化对企业数字创新的影响。以往利用专利进行创新研究的文献大多侧重于企业创新水平的某一方面。部分文献将创新研究的视角聚焦于专利的数量（寇宗来和刘学悦, 2020; 王永进和刘卉, 2021; 崔静波等, 2021）。对于专利质量的研究也层出不穷，目前主要存在的专利质量衡量方法有专利被引法（Hall et al., 2005, Hsu et al., 2014; 诸竹君等, 2020）、专利宽度法（Akcigit et al., 2016; Aghion et al., 2019）和知识宽度法（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（郝项超等, 2018; 诸竹君等, 2020）。本文立足全国统一大市场，探究市场一体化对企业数字创新数量和质量的影响，为如何推动企业创新“提质增量”相关研究添砖加瓦。

本文剩余部分安排如下：第二部分介绍了数字创新概念、衡量方法、合理性验证和基础特征事实；第三部分详细地进行了理论分析并提出相关研究假说；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具体的影响机理；最后是本文的结论和政策建议。

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

在探究数字创新的衡量方法之前，需要明确的是数字创新的具体概念。数字创新这一概念衍生于信息系统领域对数字技术特性的研究，普遍将其认为是在数字技术应用的加持下传

^① 市场与创新的关系是本文最基本的出发点，详细论述请见第三部分理论分析。

统创新的数字化变革^①（Yoo et al., 2010; Nambisan et al., 2016）。当前国内外对于数字创新的研究呈现多学科交叉的特点，但大多仅仅停留在理论层面，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，必要条件是数字化。Nambisan et al.（2016）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的一种创新。Guellec & Paunov（2020）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，其认为数字创新属于无形资产（intangible）。从企业层面看，Boland et al.（2007）提出数字创新是企业利用数字技术手段来提升创新绩效和改善运营效率；Autio & Thomas（2015）指出制造业企业的数字创新是将新兴数字技术嵌入到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；闫俊周等（2021）认为数字创新可分为数字产品、数字服务、数字过程、数字组织和数字商业模式五类的创新^②。

从以上关于数字创新概念的界定，我们可以发现，从最初的数字产品到如今的数字服务、流程、组织和商业模式等，数字创新的内涵随着数字经济时代的发展逐渐丰富。对于任何新生事物的界定，最本质的出发点是该新生事物所展现的事实。从特征事实来看：数字部件和物理部件进行新的组合如工业机器人、通讯系统属于数字创新，数字部件本身的更迭或开创性创新如软件和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，有形的计算机和智能设备的创新也属于数字创新；数字产品如无人机、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、互联网平台也属于数字创新。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。当前数字技术已经成为一种新的生产要素，数字技术的嵌入改变了原有的组织特征，使得不同形态的边界之间不断交互、更加模糊（Nambisan et al., 2019），从以上特征事实也可以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产品、服务、流程等角度进行数字创新的内涵界定已经不合时宜，也意味着现有研究对于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，新一代数字技术迅速迭代、不断演化，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、流程等的融合发展打破了对于原有创新过程和新生事物的认知，突破了原有的传统创新的理论范式。尽管数字创新在创新过程和产生的新生事物上突破了传统认知，但是归根结底，数字创新的本质是创新，与传统创新的不同则是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^③，本文从一般性的角度给出了数字创新的概念：

数字创新是以数字化思维和想法为导向，利用现有的数字化（主要）和非数字化的知识和物理部件，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而迭代形成的创新；从物理特征来看，数字创新既包含无形的软件、代码和方法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；从创新结果来看，数字创新既包括数字产品创新，也包括数字生产或服务流程的创新；数字创新的主体包括企业、国家、研究机构、个人等；数字创新的目的是在数字产品、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内在动力^④。

^① 具体来讲是传统物理产品的数字化，例如将软件、程序等添加到书籍，服装、汽车等传统物理产品中。

^② 为了更清晰地对比现有研究对数字创新的定义，我们将其汇总成表格形式，详见附录表 1。

^③ Schumpeter（1934）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。”事实上，其还将创新分为五类：新产品、新生产方式、新的销售市场、新的原材料来源、新的组织形式。

^④ 当前大多研究经常性将数字化创新和数字创新在概念上进行混淆，很多时候使用数字化创新来代替数字创新。数字化创新指的是利用数字组件（如数字技术和器件）与传统物理组件进行结合从而创造新生事物的过程，在数字技术的加持下，传统

与传统创新相比,数字创新具有一些全新的特点,这些特点主要来自于数字技术自身的特性(异质性、可编辑性和自参考性等)并在此基础上不断演变而来的。首先,数字创新具有数字非竞争性和更低的进入门槛这两个本质的属性。**第一,数字创新具有数字非竞争性。**数字非竞争性的存在,使得数字创新所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。知识的非竞争性指的是新想法的成本主要源于研发过程,但在应用过程中,增加使用者并不会增加多少成本,也并不妨碍原有的使用者使用知识的效率,即公共物品的特性。随着数字技术的发展,相比传统产品,无形成分如软件在数字创新所产生的新事物的价值上具有更高的比重。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., 2005, 2009; Benkler, 2006)。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,平均生产成本就会越低^①。**第二,数字创新具有更低的进入门槛(成本)。**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活动(例如生物医学制药等实验发明),这种进入壁垒的降低意味着更多潜在创新主体会不断涌入数字创新领域,加快了新数字产品取代现有数字产品的过程,在数字技术开放共享特点的加持下,企业可以低成本地将数字技术应用于创新过程中,“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,风险水平比过去更高。传统创新所形成的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,全新的具有引领性的数字创新可能会占领更大的市场份额(例如自媒体平台的出现极大侵占了传统媒体平台的份额),甚至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。在数字非竞争性和更低的进入门槛两个基本特征的衍生下,数字创新还具有自生长性、不确定性、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性和共享性,加之数字技术具有可重复编程的特征,从而使得数字创新可以自我进化迭代并再次重塑。**不确定性**指的是数字创新更低的进入成本从而导致的更高的“创造性破坏”概率使得新兴数字技术的出现往往会对传统技术形成颠覆与替代,市场前景的不确定性大大提升(蔡莉等, 2019)。**无边界性**指的是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

在清晰界定了数字创新的内涵和外延之后,如何构建衡量企业数字创新水平的指标是接下来需要重点考虑的。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,即将创新这一行为与其成果所应用于的产品或服务进行关联。然而目前仅有极少数的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,如崔静波等(2021)采用北京中关村自主创新企业的面板数据并采用了新产品收入作为衡量创新产出的一个方面。然而对于大部分更全面的企业面板数据(如工业企业和上市公司数据),则很难识别创新的经济效应,目前大部分的研究往往使用 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)。此外,当前数字技术的发展已经改变了企业原有的资源组态,使企业的组织边界更加模糊(Nambisan et al., 2019)。对于数字创新这一全新事物,从全要素生产率视角看,其贡献很难分解;从 R&D 投入视角看,总研发投入中用于其的比重难以准确衡量。而专利是创新行为最终形成的具体事物,体现的

物理组件被赋予了数字化属性(Yoo et al., 2010; 陈剑等, 2020)。数字化创新内含于数字创新,主要指的是数字创新中数字组件和传统组件进行重组并由数字组件驱动而迭代形成的创新。此外,数字创新还包括开创性的数字组件本身的创新,这种创新可能是天马行空且不依附于任何数字技术的。例如,组成第一台计算机的真空电子管,则是开创性的数字创新。尽管数字化创新所代表的数字组件与传统组件的结合占据了数字创新的很大比重,但是由于无法预知未来会有怎样的开创性的数字组件创新,考虑到数字创新的一般性定义,本文还是将数字化创新与数字创新进行了简单区分。

^① 数字创新的不断增加的规模回报本质上与其成果的无形成分有关,有形部分可能会产生规模经济,但其程度与无形部分不同,因为它们的可变成本不是零。

是创新行为的成果，成果也是用来衡量创新行为的重要方面。根据上文界定的数字创新的内涵和外延，数字创新所形成的具体事物既包含数字产品，也包含应用在生产和服务流程上的数字技术、方法、基础设施等相关成果；在组成上既包含了数字和物理部件结合形成的成果，也包含了数字部件自身迭代或开创性的成果；在物理特征上具有有形和无形并存的属性。这意味着任何一个专利发明是否属于数字创新成果可以根据其特征和性质做出是与否的判断。因此，只需要找到一个尽可能全面包含数字创新所形成的具体事物的范本，并将专利一一对应上去，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）》所定义的数字经济核心产业的描述涵盖的所有对象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，从而进行数字创新的识别工作^①。

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的是中国专利数据库，该数据来源于中国国家知识产权局，涵盖了自1985年《专利法》实施以来至2017年微观层面在国家知识产权局申请并公开的所有专利数据^②。数据总量为24472883条，按照专利类型划分来看，发明专利共8479217件，发明授权共3350988件，实用新型共7910960件，外观设计共4731718件。该数据包含专利的基本信息和申请发明信息，具体包含了申请日、专利类型、分类号、主分类号、公开公告日、申请人、参考文献、代理人、代码、地址、分案原申请号、发明设计人、国别、国际公布、国际申请、公开公告号、名称、审查员、申请国代码、申请号、申请来源、邮编、页数、优先权、专利代理机构、专利号、主权项和摘要等基础信息。

为了后续识别工作的顺利开展，我们对专利数据进行了一些预处理工作。第一，授权的发明专利会公开第二次，这会导致发明专利的重复。因此，对于授权的发明专利，我们删掉了其在授权前公开的信息。第二，由于专利数据缺乏能直接识别所在城市的信息，我们依据专利信息中的地址信息进行了识别：第一步，对于能够直接识别所在省市的专利，我们进行省市（如浙江省杭州市）字段匹配；第二步，对于无法直接识别地级市的专利（如地址直接写的县级市或者区）我们进行了人工的字段识别和匹配。第三，考虑到外观专利仅是更改物品的外表，并不具备数字创新的内涵；同时，后续的识别工作主要依据专利的摘要信息，而多数外观专利的摘要变量缺失严重。因此，我们将外观专利删除，仅保留发明专利和实用新型专利。

（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，对照《国民经济行业分类》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，更是将其划分为五类：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，其中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方面，也是为产业数字化发展提供数字技术、产品、服务、基础设施和解决方案的经济活动。该分类编码有六位，前两位代表行业大类，中

^① 本文主要采用的数据为所有进行过数字创新的企业，来源于天眼查企业数据库（详情见数据来源），该数据库企业非常全面，但是企业日常经营的很多细致的指标是无法得到的。在数据可得性的限制下，本文创新性地对专利进行识别，不失为不得已条件下的一种有效手段。

^② 本文的专利数据获取于2020年，来源是国家知识产权局，由于考虑到专利授权的时间因素的影响，我们将样本时间截止至2017年（包括）。《中华人民共和国专利法》第四章规定，在专利审查和批准过程中，发明专利会区分“公布”和“授权”两个过程。国务院专利行政部门收到发明专利申请后，经初步审查认为符合要求的，自申请日起满十八个月，即行公布。对于发明专利，自申请日起三年内，国务院专利行政部门对发明专利申请进行实质审查，给出修改、驳回或授权决定。发明专利权自授权公告之日起生效。而对于实用新型和外观设计专利，经初步审查（无须实质性审查）没有发现驳回理由的，由国务院专利行政部门作出授予实用新型专利权或者外观设计专利权的决定。实用新型专利权和外观设计专利权自公告之日起生效。这也意味着，对于发明专利，为了获取其最终更新的信息，一般要与数据获取时间保持在三年以上。因此，我们使用截至2017年的专利数据。

间两位代表行业中类，最后两位代表行业小类。

综合考虑类别涵盖范畴和分类数量，我们将研究视角聚焦于数字经济核心产业中类范畴。为了得到语义清晰、表述通顺的行业中类上下文描述及其所涵盖的对象，我们将每个行业中类的几个行业小类描述合并，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。我们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^①及其描述性文本（下文统一使用行业代码指代相关类别）。

2. 识别方法

本文的基础性贡献是提供一套开创性的数字创新识别方法，为后续分析提供高质量的保障。识别思想非常简单且清晰：每一个专利都有摘要，该段文字描述了此专利的名称、所属技术领域、需要解决的技术问题、主要技术特征以及用途等方面，通常不超过 300 字，非常简明扼要。而数字经济核心产业分类的文本详细记录了所涵盖的数字技术、产品、服务、基础设施和解决方案等的具体事物及其描述，如工业机器人类别的描述是“指用于工业自动化领域的工业机器人，如焊接专用机器人、喷涂机器人、工厂用物流机器人、机械式遥控操作装置（遥控机械手）等”。这些类别的描述详细介绍了该类别的核心特征以及哪些具体事物对象属于该类别，并囊括了很多举例，非常全面和具体，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/多个类别所涵盖的对象上，那么将专利进行分类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？我们使用机器学习方法范畴的预训练语言模型 SBERT，基于 Python 进行专利的摘要文本和每个类别描述之间的文本相似度的计算。

（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（NLP，Natural Language Processing）的范畴，NLP 的核心是将人类的语言表示成更易于计算机理解的方式。随着大规模语料库的建设和各种语言知识库的出现，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 研究院提出的一种预训练语言模型，其基础建立在 transformer 之上，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模型成就。BERT 提出了一种全新的框架结构，采用 masked language model 对双向 transformer 进行训练，生成深层的双向语言表征；此外，在预训练后，对输出层进行一个额外的微调，各种各样的下游任务就可以取得优异的表现（Devlin et al., 2018）。Reimers & Gurevych（2020）基于 BERT 网络做了修改，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的句子 embedding 向量，语义相近的句子其 embedding 向量距离就比较近，从而可以用来进行相似度计算。同时，该方法可以在保证精度不变的情况下大幅缩短计算时间，具有较强的现实应用性。该方法提供了训练超过十亿次，包含中文的多语言的预训练语言模型^②，我们基于 SBERT 进行文本余弦相似度的计算。

（2）数字创新识别

经过第一步的计算，我们得到了对于任一个类别，每个专利都与其有一个相似度的数值（范围在 0 和 1 之间）。考虑每个类别的专利相似度分布是有差异的，直接使用常见的 0.5 为阈值是不严谨的，因此本文将采取严格的人工识别策略和计算分段准确率的方法来对每一个类别划分阈值。具体而言，识别的逻辑是：对于任一类别，相似度越高的专利被归到该类

^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，具体使用方法等详见 <https://www.sbert.net/index.html>。

的概率也越高，相似度最高的专利往往可以 100%确定属于该类别，如果按照相似度从高到低排序，专利被归类的概率存在一个大的下降幅度，这就说明对于该类别的归类存在一个分水岭。由于本身无法直接计算专利被归到该类别的概率，综合考虑人工成本和相似度分布范围，我们采用每 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。

在具体的操作上，对于每一个类别，按照相似度从高到低对专利进行排序，并以每 0.05 对 0 和 1 之间的范围进行分割，从上到下分别对于每一个临界值点向下抽取 100 个专利，并进行严格的识别方法^①来鉴别该专利是否属于该类别，并计算 100 个专利中的严格准确率^②。并将每一段的严格准确率从上到下进行排列，当识别到某一临界值其准确率（如 60%）和下一临界值（如 30%）存在较大差异时^③，我们会将该临界值（如 60%）作为该类别数字创新识别的阈值。此外，我们还计算了每个类别相似度排名前 100 的专利的严格准确率来检查使用 SBERT 计算文本相似度是否准确。综合所有类别前 100 的准确率可以达到 91%，可以认为使用 SBERT 计算文本语义相似度是非常可靠和稳健的。阈值平均严格准确率可以达到 64%，而阈值下一临界准确率仅为 36%，存在较大差异，可以认为识别过程是非常稳健的^④。保留每个类别高于阈值的专利，总共得到 1023661 条数字创新专利。

（三）数字创新识别合理性检验

经过以上工作，本文已经得到了识别出的属于数字创新的专利成果。接下来的首要任务是评估本文识别的数字创新是否合理，或者需要为识别的数字创新找到一个可靠的比较基准。尽管从本质上来说数字创新与传统创新的差异主要体现在非竞争性和成本方面，但是由于数据可得性所限，无法从本质特征上进行数字创新的合理性检验。此外，尽管数字创新与传统创新存在诸多本质差异，这些差异所抽象形成的表征则很大程度上依赖于数字经济的发展，即数字化的特征。因此，本文利用这一抽象化特征，将识别到的数字创新与数字经济发展进行对比，确保本文的识别方法是合理可信的。

考虑到目前没有权威机构公布数字创新相关统计，同时我们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^⑤。比对结果如图 1 所示，图 1 左是数字创新总量与数字产业化增加值的趋势图，图 1 右是数字创新占比与数字产业化占 GDP 的比重对比图。我们发现：在总量上数字创新和数字产业化的增长趋势是大致相同的，都呈现出增长速度不断增加的态势；在占比上数字创新占比与数字产业化占比的趋势大致相同。总体来说，不论是从总量还是占比方面都在一定程度上支持本文识别的数字创新是合理的。

^① 严格的识别方法的思路是：对于有详细且精确描述类别的专利，专利的识别严格按照描述来进行，需要符合描述的核心特征；对于只有简要且宽泛描述类别的专利，主要根据这个专利的用途和性质，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。具体的一些举例详见附录。

^② 考虑到存在上万条专利识别信息，我们招募了一批志愿者进行辅助识别，只有当超过半数识别人员按照严格识别方法认为该专利属于该行业时，我们才会将该专利归到该行业中。

^③ 较大的差异我们一般认为至少大于 20%。

^④ 详见附录表 2。

^⑤ 目前有不少的研究进行了数字经济的测度分析，《“十四五”数字经济发展规划》指出我国的数字经济核心产业增加值占国内生产总值（GDP）比重达到 7.8%，综合来看蔡跃洲和牛新星（2021）的测度最接近。

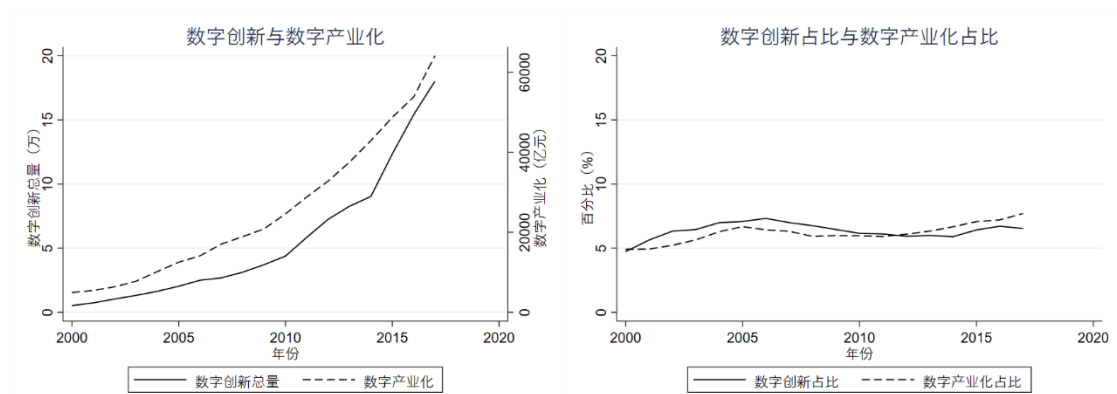


图 1 数字创新与数字经济核心产业对比

经过与数字产业化增加值进行比对，数字创新识别的合理性得到了初步的印证。基于数字经济越发达的地区数字创新可能越多这一思想，接下来本文将在地区层面进一步印证数字创新识别的合理性^①。通过利用专利详细信息中的地址信息，本文将数字创新专利汇总到省和地级市层面。在省份层面，本文借鉴韩兆安等（2021）测算的省际数字经济规模指标，分别对比了 2012-2017 年各省数字创新总量和数字经济规模之间的相关系数（表 1 第 2 列），大部分的相关系数都在 0.9 以上且在 1%水平上显著，充分印证了本文数字创新识别的合理性。在城市层面，本文借鉴赵涛等（2020）提出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（表 1 第 3 列），大部分的相关系数都在 0.7 以上且在 1%水平上显著，充分印证了本文数字创新识别的合理性。

表 1 地区层面数字创新与数字经济相关系数比较

年份	省份数字经济增加值相关系数	地级市数字经济指数相关系数
2011		0.6837***
2012	0.9327***	0.7231***
2013	0.8931***	0.7004***
2014	0.8823***	0.7076***
2015	0.9113***	0.7391***
2016	0.9534***	0.7327***
2017	0.9744***	0.7759***
以上所有年份	0.9107***	0.7050***

注：***为在 1%水平上显著。

（四）数字创新特征事实

经过了数字创新的合理性检验，本文也在地区、企业性质和规模等方面探究了数字创新特征事实^②，并将其展现的特质与传统创新进行比对，以期实证分析提供启发和思路：

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；同时，与传统创新不同的是，数字创新也呈现集聚效应，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。

对于传统创新而言，寇宗来和刘学悦（2020）指出中国企业的创新行为呈现区域不平衡特征。我们根据国家统计局东中西和东北地区的划分来探究各区域数字创新发展的趋势。通过统计发现 2000 以来东部地区数字创新占比 71.6%。相比其他地区，东部地区不论是在规

^① 本文也曾尝试在行业内层面进行细化的合理性检验，但是当前行业层面关于数字经济的衡量指标尚有缺失，因此本文只能在地区层面进行进一步的检验。

^② 对于企业层面特征事实的探究主要依据下文详细介绍的数字创新与天眼查企业匹配数据库。

模还是增速上都远远领先。这表明我国数字创新也呈现出了很大的区域不平衡性^①。为了更进一步佐证，参考寇宗来和刘学悦（2020）计算创新基尼系数的方法，我们将数字创新分别汇总到省份和城市层面并计算数字创新基尼系数^②。在省份层面，创新基尼系数在 0.6 左右徘徊，说明数字创新的省份分布是不平衡的；在城市层面，基尼系数大多数年份高于 0.8，说明城市层面的不平衡性是高于省份层面的；值得注意的是，2006 年之后，数字创新的基尼系数有一定的下降，说明不均衡性在逐渐好转。

此外，我们还分别计算了以北京为核心的环渤海地区、以上海为核心的长三角地区和以深圳为核心的珠三角地区的数字创新占比，总的来说三大创新区的数字创新总量占比超过了 60%，呈现出了数字创新的集聚效应。我们也分别计算了北京、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^③。与传统创新不同的是，本文发现随着数字经济的发展，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，上海对于周边的辐射效应是最强的。

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，我们按照企业所有制将企业划分为国有企业、民营企业和外资及港澳台企业三种类型^④。相比于国有企业和外资、港澳台企业，民营企业在数字创新专利的申请上一骑绝尘，无论是在增速上还是规模上都明显更高^⑤。这意味着民营企业是我国数字创新的主体，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。

参考寇宗来和刘学悦（2020）的方法，本文以企业注册资本衡量企业规模，在企业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^⑥。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平衡性，尤其是在 2006 年之后，规模上的不平衡性进一步下降，2017 年的基尼系数在 0.3 左右。这一现象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，我国不断涌现数字技术尤其是互联网企业。不断有个体和初创企业凭借自身独特的数字技术创新成为细分行业的领头羊。2014 年达沃斯论坛上，李克强总理提出“大众创业、万众创新”这一理念，进一步推动了个人和中小企业创新创业的行为。不论企业规模，任何主体都可以参与到数字创新当中，百花齐放^⑦。这也意味着中小企业是我国数字创新的中坚力量，是数字经济发展的推动力。

综上所述，数字创新是传统创新在数字经济时代的延伸。一方面，数字创新展现出更的集聚效应和民营企业是创新主体的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，比如数字创新的溢出效应更强，数字创新在企业规模维度并不存在较大的不平衡性等。

三、理论框架与研究假说

本文最基本的出发点是市场与创新的关系。一方面，企业创新受到企业内部和外部多种因素的驱动，但市场始终处于决定地位。从外部因素来看，融资约束（张璇等，2017）、企业社会网络（刘善仕等，2017）、需求拉动因素（刘小玄和吴延兵，2009）、产业集群和知识溢出（朱秀梅，2008）以及政府政策与制度（王永进和冯笑，2018；张杰等，2021）等是驱

^① 见附图图 1。

^② 见附图图 1。

^③ 见附图图 2。

^④ 天眼查企业数据提供了企业性质变量，我们对该变量所含标签进行了人工识别对应。

^⑤ 见附图图 3。

^⑥ 见附图图 3。

^⑦ 据央视财经报道，十年来，平均每年新增 1.7 万家高新技术企业，我国高新技术企业已经达到 27.5 万家，科技型中小企业有 22.3 万家。<https://finance.sina.com.cn/7x24/2021-08-01/doc-ikqciyzk8930336.shtml>

动创新的重要因素；从内部因素来看，市场营销能力（陈晓红和于涛，2013）、高管行为（李春涛和宋敏，2010）、人力资源管理（刘善仕等，2007）以及市场势力和企业规模（周黎安和罗凯，2005）等都是驱动创新的重要因素。尽管影响因素多种多样，但是从本质来看，除了政府政策因素外，其他的影响因素可以说都是企业的市场行为或者属于市场因素这一框架。另一方面，从企业创新本身或者动机上来说，这一行为本质上就是由市场来驱动的。驱动企业创新最关键的前提是创新成本与收益的比较，只有当创新收益高于创新成本时，企业创新才能得到激励（张杰和周晓艳，2011）。而企业创新的收益则是通过市场来实现的，只有市场上存在对该创新产品和服务的有效需求且支付意愿足以支撑其市场化推广，才能使得企业的创新投入转化为市场收益，进而从根本上支撑企业创新（康志勇和张杰，2008）。从以上两方面可以看出，市场是创新的关键决定因素。此外，《中华人民共和国国民经济和社会发展规划第十四个五年规划和 2035 年远景目标纲要》指出要完善技术创新市场导向机制，国务院《“十四五”数字经济发展规划》指出要发挥我国超大规模市场优势，提高数字技术基础研发能力。因此，本文以市场与创新的理论框架为基础，立足于全国统一大市场，探究市场一体化对企业数字创新影响背后的理论逻辑。

市场一体化与市场分割是同一个硬币的两面（陈甬军和丛子薇，2017）。区域市场一体化指的是不同区域之间的“经济边界”逐步消失的过程（国务院发展研究中心课题组，2005）。市场一体化主要涵盖产品市场一体化、要素市场一体化和服务市场一体化三个方面（李雪松和孙博文，2015）。需要重点强调的是由于多种原因所限，本文主要关注的是产品市场一体化。产品市场一体化指的是产品在不同地点之间的自由流动程度，这种自由流动程度越高市场一体化程度也越高（Donaldson, 2015）。产品市场一体化的本质是限制性贸易壁垒的下降，从理论上讲这一壁垒通常指代的是“冰山成本”，不论市场分割是由于自然、技术乃至制度等因素导致的，这一切都反映在该成本上。这也是为什么采用价格法^①来衡量市场分割是更具有现实应用性和有效性的，因为价格反映了产品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上，区域贸易和国际贸易的模型是一致的，尽管国际贸易和区域贸易可能在消费者偏好、技术以及要素禀赋的设定上可能存在不同，但是标准工具都是引力模型分析框架，本文也将借鉴经典引力模型框架分析市场一体化产生影响的渠道。

从产品市场一体化的定义出发，市场一体化的提高意味着限制性贸易壁垒的降低，即“冰山成本” τ_{ij} 会降低^②。一方面，在 EK 模型中，通过对引力模型的推导^③，随着*i*地和*m*地之间的贸易壁垒的降低（即 d_{mi} 下降），*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，即*i*地所面临的目的地*m*的市场规模会增加（Eaton & Kortum, 2002）。从现实上也可以解释该现象：考虑一个极端情况，当某地区和周围地区存在完全贸易壁垒（即 τ_{ij} 无限大）时，该地区企业的任何产品无法被周围地区的消费者消费，只能由本地消费者消费；而当完全不存在贸易壁垒（即 $\tau_{ij}=1$ ）时，该地区企业所有产品都可以被周围地区消费，同时也由本地消费者消费。这意味着该地区企业的产品在市场一体化逐渐提高的过程中可以被更多的消费者消费（不论这个企业是否开拓外地市场，所面临的需求扩张是事实），即面临的市场规模是扩大的。另一方面，在异质性企业模型开放经济中，当 τ_{ij} 下降时，零利润（ZCP）条件不断右移，内生决定的市场中的企业数目是增加的^④（Melitz, 2003），这意味着市场竞争的加剧（Melitz & Ottaviano, 2008）。从现实上也可以解释该现象：考虑同上的极端情况，当某地区和周围地区存在完全贸易壁垒（即 τ_{ij} 无限大）

^① 自从 Anderson & van Wincoop（2004）对于贸易成本的研究发表以来，大多研究在理解空间价格差距如何被用来估计贸易成本方面取得了大量进展（Donaldson, 2015）。

^② 即从*i*地运输到*j*地的成本，包括运输成本、交流成本、税收、制度性限制措施等等一切可以反应在价格上的因素。一般由 τ_{ij} 表示，但由于习惯不同，在 Eaton & Kortum（2002）中使用 d_{ni} 。

^③ 详见 Eaton & Kortum（2002）的式（11）及之前的推导流程，由于篇幅所限，本文无法将全部推导过程放上来。

^④ 详见 Melitz（2003）的式（20）和（21）。

时,周围地区的企业无法进入该地区,该地区企业的竞争对手仅有当地厂商;而当完全不存在贸易壁垒(即 $\tau_{ij}=1$)时,周围地区企业可以自由进入该地区市场,该地区企业不仅面对当地厂商的竞争,还面临周围地区企业的竞争,由此面临更大的市场竞争。综上所述,从基本引力模型框架可以看出,市场一体化的影响主要是通过需求端市场规模扩大和供给端市场竞争的加剧两个渠道来实现的。

市场一体化对数字创新的影响,从需求端市场规模渠道来看:相比传统创新,数字创新具有更强的非竞争性,在组成上具有更高的无形化特征,这些特点意味着数字创新产生的产品的边际生产成本更低(Corrado et al., 2005, 2009; Benkler, 2006),甚至有些数字软件的生产 and 传播成本接近于零;同时,数字技术的可分配性特点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者应用到终端服务上,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相比传统创新具有更强的规模经济效应。而市场一体化可以扩大一个地区中企业所面临的市场规模,企业可以销售给更多的消费者,使得企业不仅可以支配本地市场,也可以支配外地市场。面对市场规模的扩张,企业通过创新的研发投入,利用数字技术完善产品生产流程,生产数字产品,薄利多销,占领更多的市场份额,带来更大的收益,从而在规模经济的条件下获得更大的利润,弥补了创新投入的沉没成本(Bustos, 2011; Aghion et al., 2020)。这也意味着市场一体化带来的更大市场规模会进一步放大数字创新的规模经济效应,甚至可能出现“赢者通吃”的格局,攫取大部分的市场利润,从而激励企业研发投入(Hall & Harhoff, 2012),促使企业产出更多更高质量的数字创新成果。

市场一体化对数字创新的影响,从供给侧市场竞争渠道来看:市场一体化意味着企业会面临其他地区同行业企业的竞争,市场竞争更加激烈。市场竞争对创新的影响属于市场结构与创新的研究分支,学界尚未达成一致,主要有三个观点:第一,具有市场势力的企业在创新实力与风险承担方面占据优势,市场竞争不利于创新(Schumpeter, 1942)。第二,竞争有利于创新企业降低成本,而垄断则会导致企业缺乏创新动力从而不利于创新绩效的提升(Arrow, 1962)。第三,在完全垄断的市场条件下,企业缺乏竞争对手,对研发的动力和热情不够高,不利于创新;在完全竞争的市场条件下,企业规模较小难以承担研发费用和 risk,不利于创新;而处于两者之间的市场结构下,企业既能持有上述优点,又能规避不足之处,从而最大程度有利于创新(Aghion et al., 2001, 2005)。创新则是在不断变化的市场竞争环境中可持续发展的核心,在经营过程中发现运用创新提供与竞争对手不同的产品及服务,常常能够取得竞争优势。而数字创新主要依赖无形的数字技术,所要求的资本投入要比传统创新大大降低,由此降低了准入门槛,打破了组织边界和产业边界,大量由于门槛所限无法进入创新领域的个体和个人也可以参与到数字创新过程中,提高了“创造性破坏”概率(Guellec & Paunov, 2017)。这种“创造性破坏”体现在数字创新导致组织的产品、服务、商业模式乃至组织结构发生重大变化:在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开拓全新市场(Gobble, 2016),数字创新所产生的数字赋能效应使企业能够优化客户体验和创建新的商业运营模式,重塑企业重大业务的过程(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)。这意味着在此过程中,企业本身数字创新的不确定性更大,且一旦竞争对手先一步进行了成功的数字创新,企业则可能血本无归。此外,相比传统创新,数字创新所产生的产品和服务还具有更强的规模经济效应,同时数字创新所依赖的数字技术具有自生长性,自我更迭和传播速度远远高于传统创新。这意味着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,企业则会更加迅速地占领市场,赢者通吃赢得竞争。

从对市场竞争渠道的理论分析来看,市场竞争对数字创新的影响相比传统创新可能更加

扑朔迷离。但如果立足于数字创新的特征,并结合市场竞争理论框架分析,还是能够清晰把握影响脉络。由于数字创新所需的成本和进入门槛大大降低,不具备市场势力的大量创新主体可以参与到数字创新过程中来,这就意味着规模大的企业才能承担创新成本,市场竞争不利于数字创新的理论可能无法立足。激烈的竞争使得企业有被淘汰的可能性,为了从激烈的竞争中脱颖而出,企业很可能借助于所需固定资本投入更低数字创新,在产品质量或者生产和服务效率上战胜竞争对手(Correa & Ornaghi, 2014)。尽管创新是一个长期的过程,但是对于数字创新而言,依靠具有自我迭代性特点的数字技术,使得此次科技革命比以往的工业革命要快的多(Schwab, 2016),这意味着应对激烈的竞争采取的数字创新战略见效可能是迅速的。同时,数字创新更大的规模经济效应也会加速这一过程,一旦企业进行的数字创新成功迎合市场,那么其将会迅速占领市场,打败竞争对手。此外,需要注意的是除了开创性的数字创新外,大部分的数字创新是在已有数字组件的基础上迭代演化的,这意味着数字创新符合技术落后者必须首先赶上前沿技术,才能在未来争夺技术领先地位的逐步创新的内生增长框架。那么市场竞争对企业数字创新的影响可能是倒 U 型的:当两家企业在技术水平上并驾齐驱时,双方都可能有强大的创新动力来逃离竞争;如果某企业远远落后于技术前沿,产品市场竞争的加剧可能会降低创新的增量回报,从而产生抑制创新的效应(Aghion et al., 2001, 2005)。

综上所述,市场规模的扩大和市场竞争的加剧是市场一体化影响企业数字创新的两大重要渠道。有研究指出需求端拉动 60%至 80%的创新(Utterback, 1974),这意味着不论供给端市场竞争的影响是促进抑或非线性,可能都无法掩盖需求端市场规模对于数字创新的拉动作用。同时本文也进行了特征事实上的探究,发现市场一体化程度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: 市场规模的扩大是重要的影响机制之一,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: 市场竞争的加剧是重要的影响机制之一,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

上文已经对所有的专利数据进行了数字创新的识别和合理性检验,构建了本文数字创新专利基础数据库。为了搭建与微观企业的桥梁,我们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(官方来源截至 2013 年),以及上市公司数据聚焦于上市企业而非全样本,尤其缺乏小微企业。为了使得与数字创新专利进行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信息(包括工业企业和上市公司),每个企业的变量主要包括公司名称、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(所有制)、所属行业等。

天眼查提供的批量查询功能使得本文可以通过清理专利申请人进而与该平台上企业基

^① 基于下文计算的城市层面的市场一体化程度,本文在时间和城市维度上探究了其数字创新之间的关系。详见附录图 4。

本信息进行全面匹配。具体而言：第一步将具有多位共同申请人的专利对申请人进行拆分，第二步去掉个人申请人，第三步通过天眼查的批量查询功能进行企业匹配。在去掉个人申请并拆分后的 972001 个样本中，共匹配上 798790 条数据，匹配成功率高达 82%；其中共有 192572 家企业，匹配上的有 164100 家，匹配成功率高达 85%。为了获得面板数据，基于匹配上的企业数据，本文根据每个企业成立时间和退出时间，识别企业存续区间，从而构建本文实证分析所使用的企业面板数据。综合考虑市场一体（2001 年后）和数字创新专利数据（2017 年前）的时间维度，企业面板数据的时间跨度为 2001 至 2017 年。此外，考虑到企业可能存在迁入迁出现象，本文借助企业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

（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数量加 1 取对数来刻画。此外，专利可以分为发明专利和实用新型专利；数字创新可以根据结果的不同分为数字产品和数字生产/服务流程上的创新。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/服务流程创新上来。因此，本文还分别对数字创新中的发明专利和实用新型专利、数字产品和数字生产/服务流程上的创新进行分类汇总，并加 1 取对数作为异质性分析的指标。

对于企业数字创新的质量，本文采用专利被引次数来刻画。具体而言，本文借鉴 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（稳健性），并加 1 取对数来测度企业数字创新的质量。考虑到存在右侧“断尾”问题，本文仅采用专利申请至授权后 5 年作为窗口期计算专利被引量。由于中国国家知识产权局的企业专利数据库无法准确有效地提供企业专利被引用次数信息，本文基于申请号对 Google Patent 进行了爬取工作。考虑到 5 年窗口期，企业数字创新质量指标截止到 2016 年而非 2017 年。

（2）市场一体化

现有文献关于国内市场一体化的测度有很多方法：生产法（Young, 2000；白重恩, 2004）、贸易法（Naughton, 1999；Poncet, 2003）和价格法（桂琦寒等, 2006；吕冰洋和贺颖, 2020）等。考虑到中国省际贸易数据难以获得，城市层面的更难以获得；地级市在经济、政治和发展上承担的承上启下作用，同时城市群在推动区域经济一体化和发展上起到的战略性作用。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。计算该指数的价格数据是 2001 至 2017 年的居民消费价格指数，考虑到 2016 年该价格指数类目发生了细微的调整，因此本文将 2016 年之后的类别与之前的进行对应，最终得到食品、烟酒、衣着、居住、娱乐教育文化用品及服务、医疗保健和个人用品、交通和通信、家庭设备用品及维修服务八类产品部门的价格指数。数据主要来源于各个省份的统计年鉴，各个城市的统计年鉴和统计公报等。为了更为细致地考察交通基础设施扮演的作用，本文使用高德地图城市对之间的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指数 $MarkSeg_{ct}$ ，其中 c 代表城市， t 代表时间。

为了得到反向衡量市场一体化的指标，本文参考盛斌和毛其淋（2011）的方法根据公式 $MarketInt_{ct} = \sqrt{1/MarkSeg_{ct}}$ 取对数计算市场一体化程度，该指标越大代表城市与其相邻城市的市场一体化程度越高。

（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：本文首先根据企业成立年份，将每年减去该成立年份并加 1 取对数作为企业年龄变量（Age）。对于企业资产变量，天眼查数据中可以使用的是企业注册资本指标（实缴资本变量缺失过多，不具备使用条件）。

由于该变量包含不同国家的货币（人民币为主），本文对其进行了清理，并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，本文清理了参保人数指标。主要变量的统计性描述如表 2 所示：

表 2 描述性统计

变量名	Obs	Mean	SD	Min	Median	Max
MarketInt	1208755	0.43	0.032	.2882322	.4328212	.5631669
MarketInt_muni	1539662	0.43	0.032	.2876783	.429931	.5644777
Quantity	1583153	0.16	0.434	0	0	8.22121
Ave_cite	1424253	0.11	0.381	0	0	5.899898
Max_cite	1424253	0.13	0.460	0	0	5.899898
Capital	1542572	7.44	2.027	2.397895	7.409827	12.73095
Emp	1459819	3.10	2.297	0	3.218876	8.325791
Age	1583153	1.80	0.879	0	1.94591	6.926577

注：MarketInt 剔除直辖市样本计算的市场一体化指标，MarketInt_muni 为不剔除四大直辖市样本来计算城市对的价格变动进而计算得到的市场一体化指标。Quantity 为数字创新数量，Ave_cite 为利用平均被引计算的数字创新质量，Max_cite 为利用最大被引计算的数字创新质量，Capital 为注册资本，Emp 为参保人数，Age 为企业年龄，本文对以上解释变量加 1 取对数。

3.基准回归模型

本文的基准回归模型为：

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

其中，被解释变量 $DigInov_{fcit}$ 为年份 t 、所在城市 c 和行业 i 的企业 f 的数字创新（数量和质量）衡量指标。核心解释变量 $MarketInt_{ct}$ 为城市 c 在年份 t 的市场一体化水平。 $Control_{fcit}$ 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进行交互以控制更多的企业特征。控制变量具体包括企业年龄，企业注册资本与年份虚拟变量的交互，企业参保人数与年份虚拟变量的交互。 η_f 为企业固定效应，用来控制企业不随时间变化的特征，值得注意的是，样本中的企业所在地和行业并不随时间变化，因此控制企业固定效应可以有效控制城市和行业层面特征； η_t 为时间固定效应。 ε_{fcit} 为随机误差项。

（二）基准结果

表 3 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，因此后文的实证分析使用的都是剔除四个直辖市的市場一体化指标 $MarketInt$ 。（1）和（4）列是市场一体化对企业数字创新数量的回归。结果显示不论是否剔除直辖市，市场一体化的系数都是正显著的。这说明在其他条件不变时，市场一体化可以显著促进企业产出更多的数字创新成果。（2）和（5）列为基于平均被引作为创新质量衡量指标的回归结果，（3）和（6）为基于最大被引作为创新质量衡量指标的回归结果。结果显示在其他条件不变时，市场一体化可以显著提升企业数字创新质量。由此，假说 1 得证。此外，企业年龄的系数为正且显著，这说明成立时间越久的企业数字创新越多质量越高。

表 3 基准回归结果

	(1)	(2)	(3)	(4)	(5)	(6)
	Quantity	Ave_cite	Max_cite	Quantity	Ave_cite	Max_cite
MarketInt_muni	0.080*** (0.013)	0.050*** (0.013)	0.061*** (0.015)			
MarketInt				0.079*** (0.015)	0.040*** (0.015)	0.050*** (0.017)

Age	0.037*** (0.001)	0.017*** (0.001)	0.018*** (0.001)	0.040*** (0.001)	0.017*** (0.001)	0.018*** (0.001)
_cons	0.060*** (0.006)	0.061*** (0.006)	0.077*** (0.007)	0.053*** (0.007)	0.059*** (0.007)	0.074*** (0.008)
N	1386631	1236271	1236271	1087497	968456	968456
CapitalxYear	√	√	√	√	√	√
EmpxYear	√	√	√	√	√	√
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 × Year 为注册资本和年份虚拟变量的交互，Emp × 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。***，**，* 分别表示 1%、5% 和 10% 的水平上显著；括号内的标准误为稳健标准误。

（三）异质性分析

第一，本文将数字创新分为发明专利的数字创新和实用新型专利的数字创新，探究市场一体化对这两类数字创新（数量和质量）的影响。回归结果如表 4（1）至（4）列所示。结果表明，市场一体化对企业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的促进作用。此外，根据数字经济核心产业大类的分类，分别对数字产品和数字生产/服务流程上的创新的回归结果如表 4（5）至（9）所示。结果表明，市场一体化能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 4 分类回归

	(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	0.049*** (0.010)	0.036*** (0.013)	0.032*** (0.012)	0.009 (0.011)	0.070*** (0.015)	0.035** (0.014)	0.035*** (0.008)	0.030*** (0.009)
Age	0.007*** (0.001)	0.003*** (0.001)	0.033*** (0.001)	0.016*** (0.001)	0.035*** (0.001)	0.015*** (0.001)	0.008*** (0.001)	0.003*** (0.001)
_cons	0.036*** (0.005)	0.046*** (0.006)	0.035*** (0.005)	0.034*** (0.005)	0.047*** (0.007)	0.052*** (0.006)	0.017*** (0.004)	0.019*** (0.004)
N	1087497	968456	1087497	968456	1087497	968456	1087497	968456
CapitalxYear	√	√	√	√	√	√	√	√
EmpxYear	√	√	√	√	√	√	√	√
Year Fixed Effect	√	√	√	√	√	√	√	√
Firm Fixed Effect	√	√	√	√	√	√	√	√
R ² _Adj	0.25	0.14	0.13	0.08	0.17	0.11	0.24	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。各分类的专利质量指标由分别对各个分类汇总计算而得。Capital × Year 为注册资本和年份虚拟变量的交互，Emp × 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。***，**，* 分别表示 1%、5% 和 10% 的水平上显著；括号内的标准误为稳健标准误。

第二，特征事实一指出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。市场一体化对企业数字创新水平的促进效应在东部和其他地区是否存在异质性？本文分别分东部和其他地区样本进行回归来探究地区的异质性^①。结果显示市场一体化对东部地区企业的数字创新（数量和质量两方面）的促进作用是显著的，对中西部和东北地区企业数字创新（数量和质量两方面）的作用是不显著的。当前城市群成为了区域经济发展的新型空间组织和重要战略单元，从特征事实来看，数字创新的集聚效应和溢出效应都很强。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^②，这可能在一定程度上解释东部和其他地区作用效果的差异。

第三，特征事实三指出民营企业贡献了大部分的数字创新，市场一体化对不同所有制企业是否存在异质性效应？本文分别就国有企业、外资及港澳台企业和民营企业进行分样本回归^③。结果发现市场一体化水平的提高对民营企业数字创新（数量和质量两方面）促进作用显著，对国有企业和外资企业不显著。

第四，特征事实四指出数字创新的不平衡性在企业规模维度上明显降低，市场一体化对数字创新的促进作用是否会对中小企业更大呢？本文分别使用企业的注册资本和参保人数作为企业规模的衡量标准，并通过加入企业规模的交互项探究企业规模的异质性^④。场一体化与企业规模的交互项都为负且在统计上显著，这说明，市场一体化对数字创新（数量和质量两方面）的推动作用在中小规模的企业上会更大。这一结果可能源于不同规模企业对数字创新投入的未来预期不同（Lileeva & Trefler, 2010），相比于大企业，中小企业的市场表现相对较差，也因此对创新投入的未来收益提高预期。

（四）稳健性检验

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的交通设施，因此最直接的城市间距离可能会综合性反映所有的交通难度。因此，本文借鉴吕冰洋和贺颖（2020）使用城市间直线距离的倒数将城市对的价格变动在相邻城市中进行加权平均，并计算市场一体化指数，将其作为稳健性指标进行回归^⑤，回归结果与基准回归中的结果一致。

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，本文首先通过控制行业一时间固定效应，消除随行业变动的遗漏政策等因素的影响^⑥。

第三，为了进一步排除内生性的影响，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进行回归（吕越等，2018；卞元超和白俊红，2021）。海拔高度和坡度反映的是该地区与外界之间产生流动的地理条件，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因素范畴，故满足排他性约束。考虑到海拔和坡度几乎不随时间变化，对于每个城市本文采用剔除该城市后其他所有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使用的工具变量。表 5（1）报告了第一阶段的结果，（2）到（4）列报告了二阶段回归结果。工具变量回归结果显示，市场一体化对企业数字创新（数量和质量两方面）的促进作用是稳健的。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 χ^2 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 5 内生性检验

(1)	(2)	(3)	(4)
MarketInt	Quantity	Ave_cite	Max_cite

① 结果详见附录表 3。
 ② 我们也计算了成渝城市群中心城市的作用发现并不存在溢出效应甚至具有一定的虹吸效应。
 ③ 结果详见附录表 4。
 ④ 结果详见附录表 5。
 ⑤ 结果详见附录表 6。
 ⑥ 结果详见附录表 7 和表 8。

IV1_altitude	0.012*** (0.000)			
IV2_gradient	0.009*** (0.001)			
MarketInt		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 χ^2		0.248	0.561	0.556
其他控制变量	√	√	√	√
Ind-Year Fixed Effect	√	√	√	√
Firm Fixed Effect	√	√	√	√

注：IV1_altitude 为构建的海拔高度工具变量，IV2_gradient 为构建的坡度工具变量。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。其他控制变量包括：企业年龄、注册资本和年份虚拟变量的交互、参保人数与年份虚拟变量的交互。回归中控制了行业年份固定效应和企业固定效应。***，**，*分别表示 1%、5%和 10%的水平上显著；括号内的标准误为稳健标准误。

（五）机制分析

前文理论分析已经阐明了市场规模和市场竞争对企业数字创新存在直接影响。因此，本文借鉴江艇（2022）对渠道检验的论述，主要利用上文的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）的因果关系来探究市场一体化对企业数字创新的影响渠道。值得注意的是，考察机制变量对被解释变量影响的方向是有益且重要的，而理论分析中市场竞争对企业数字创新的影响的方向尚存在不明确之处。因此，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

准确衡量企业面临的市场规模变动是探究该影响渠道的重要前提。由于无法获得企业销售尤其是内销数据，也无法获得城市间贸易流数据和不同产业的消费数据，综合考虑数据可得性条件下，本文使用城市社会消费品零售总额并借鉴 Aghion et al.（2020）的方法衡量一个城市及其所有相邻城市的需求变动：

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

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。该公式从构成上可以被解读为城市 c 所在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被其自身在该小城市群中重要性加权得到。

表 6 报告了市场规模渠道的回归结果。其中（1）列为在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变动的影响，控制了城市和行业时间的固定效应。（2）列是市场规模对企业数字创新数量的回归结果，（3）和（4）列为市场规模对企业数字创新质量的回归结果。考虑到市场规模的变动是在城市时间层面，因此回归中控制了最严格的企业和行业时间固定效应。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的回归结果显著为正，意味着市场规模的扩大可以显著促进企业数字创新。由此，假说 2 得以验证。此外，通过划分数字产品和数字生产/服务流程，我们发现市场规模对数字产品创新（数量和质量）正向促进作用更显著^①。这一结果也符合市场规模效应的经济直觉，市场规模扩

^① 结果见附录表 9。

大时企业面对的消费者会更多，企业往往会将数字创新的重点放在产品上，生产出更先进的产品从而迅速占领市场。

表 6 市场规模机制

	(1)	(2)	(3)	(4)
	$\Delta Demand$	Quantity	Ave_cite	Max_cite
MarketInt	0.013*** (0.004)			
$\Delta Demand$		0.076*** (0.008)	0.019** (0.008)	0.028*** (0.009)
_cons		0.081*** (0.002)	0.079*** (0.002)	0.099*** (0.002)
N	4797	1423077	1272760	1272760
其他控制变量		√	√	√
固定效应	city, ind#year		firm, ind#year	
Wald F	1642.574			
Hansen J χ^2	0.006			
R ² _Adj		0.19	0.12	0.16

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。其他控制变量包括：企业年龄、注册资本和年份虚拟变量的交互、参保人数与年份虚拟变量的交互。***，**，* 分别表示 1%、5% 和 10% 的水平上显著；括号内的标准误为稳健标准误。

2. 市场竞争

在进行市场竞争的机制检验前需要明确如何衡量市场竞争。考虑到无法获取企业销售数据，因此本文采用常见的企业竞争对手数量来衡量市场竞争，一般而言，竞争对手越多市场竞争越激烈。因此，本文利用天眼查企业数据计算了每个城市每个行业每年存续的企业数，并将城市 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 的企业的竞争对手数量来衡量企业面临的市场竞争。

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，在进行具体的市场竞争机制实证分析前，我们需要检验数字创新是否符合逐步创新模型的设定。在此，我们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。对于任意时期，我们计算了每个企业该期之前累积的数字创新数量 $DigInov_Past$ ，将其作为解释变量加入本文的基准回归模型中。我们发现，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）的影响是显著为正的^①，这说明了企业的技术积累对未来数字创新具有重要的作用，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。

表 7（1）是基于工具变量回归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，证实了市场竞争是渠道之一。表 7（2）至（4）是在控制了企业、行业-时间和城市-时间固定效应下市场竞争对企业数字创新数量和质量的结果，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）。但是，当数字创新遵循经典内生增长模型的逐步创新假定时（正如上文验证的那样），市场竞争对创新的作用可能是微妙的，而且很可能是非单调的（Aghion et al., 2001, 2005），线性的结果可能掩盖了非线性事实。

因此，本文加入市场竞争的二次项来考察是否存在市场竞争与数字创新是否存在非线性

^① 详见附录表 10。

关系，表 7（5）至（7）列报告了该回归结果。结果显示，市场竞争对数字创新（数量和质量两方面）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 型关系中点位置之间的联系：市场竞争（竞争对手的对数形式）的最小值为 4.2，最大值为 16.8，均值为 12.9，而倒 U 型最高点（中点）约为 12；可以发现市场竞争的取值范围涵盖了倒 U 型最高点所在位置且该位置比较靠近分布的中心。其次，我们对倒 U 型关系进行了二次回归^①，具体做法是：首先确定倒 U 型关系的最高点所在位置，然后在该断点处对之前的市场竞争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 型关系说明了随着市场竞争的提高，其对数字创新（数量和质量）的促进作用先增大后减小，最后过度的市场竞争甚至可能会抑制企业数字创新。这意味着，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。我们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；但是对于数字生产/服务流程的影响总体来看是正向的。

表 7 市场竞争机制

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	MarketCom	Quantity	Ave_cite	Max_cite	Quantity	Ave_cite	Max_cite
MarketInt	0.009*** (0.003)						
MarketCom		-0.010** (0.005)	-0.007 (0.005)	-0.009 (0.006)	0.135*** (0.016)	0.077*** (0.015)	0.115*** (0.018)
MarketCom ²					-0.006*** (0.001)	-0.003*** (0.001)	-0.005*** (0.001)
_cons		0.221*** (0.061)	0.170*** (0.059)	0.218*** (0.071)	-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		√	√	√	√	√	√
固定效应	ind#city, ind#year			firm, ind#year, city#year			
Wald F	37828.354						
Hansen J χ^2	8.569						
R ² _Adj		0.20	0.12	0.16	0.20	0.12	0.16

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。其他控制变量包括：企业年龄、注册资本和年份虚拟变量的交互、参保人数与年份虚拟变量的交互。***，**，*分别表示 1%、5%和 10%的水平上显著；括号内的标准误为稳健标准误。

综上所述，市场一体化可以促进企业数字创新。这也意味着全国统一大市场的建设能够有效提升企业数字创新水平。在迈向全国统一大市场的进程中，不论是需求端的市场规模，还是供给端的市场竞争，都在引领企业数字创新上有着独特的贡献。

^① Simonsohn 和 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，最高点前的市场竞争系数为 0.017（1%水平上显著），最高点后的市场竞争系数为-0.008（1%水平上显著），证实了倒 U 型关系的存在。

^③ 详见附录表 11。

从需求端的市场规模来看。本文的结果证实了一体化程度更高的市场会带来更大的消费需求，进而显著促进企业数字产品创新，但对于数字生产/服务流程的创新影响不大。全国统一大市场的建设通过畅通循环、提高市场效率，完善标准和计量体系、全面提升服务质量，进一步降低市场交易成本，可以有效弥合市场分割、打通经济循环的堵点，推进商品和服务市场的高水平统一，从而释放更大内需潜力。通过完善的市场信息交互渠道和现代流通网络，企业可以迅速捕捉其他地区的多样化市场需求，突破自然性市场分割的阻碍，使得自己的产品在全国范围的顺畅流动。在面对需求端冲击下转瞬即逝的消费需求，大部分企业往往着眼于提升产品的质量和口碑，以期迅速占领市场，从而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，助力数字创新成果的应用转化，强化创新过程的市场需求导向。

从供给端的市场竞争来看。本文的结果证实了一体化程度更高的市场会带来竞争水平的提升，进而对数字创新（数量和质量）产生倒 U 型的影响，这意味着随着市场竞争的提高，其对数字创新的促进作用先增大后减小，存在最优水平，只有适度且合理的竞争才可以最大程度上促进企业数字创新“提质增量”。全国统一大市场的建设通过深化产权保护、市场准入、公平竞争和社会信用等制度改革，加强反垄断和反不正当竞争，破除地方保护和区域壁垒，积极打造市场化、法治化、国际化营商环境，为各类市场主体投资兴业营造良好生态。企业可以在公平、充分的国内竞争环境下规范经营，与同类型企业进行产品和服务流程等方面的竞争。当市场竞争程度并不激烈时，企业在自身所在的行业领域具有一定的话语权，可以更多地在产品和服务效率等方面进行数字创新，提高产品质量和生产服务效率，从而攫取更大市场份额，弥补前期投入，形成良性循环。而当市场竞争异常激烈时，企业大多采用价格竞争策略，遭遇过度“内卷”困境，加成率降低削弱了产品创新的收益，从而降低企业在产品创新上的投入。另一方面，为了从激烈的竞争中脱颖而出，企业会加大在生产/服务流程上的数字创新投入，不断提升竞争力。

以上发现是在数字经济时代对市场与创新关系进行的再审视，其中所蕴含的丰富政策启示将在下一部分进行详细阐述。

五、结束语

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 至 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，通过计算每个专利的摘要文本与每个数字经济核心行业描述之间的语义相似度，并基于严格的人工识别方法划分阈值来进行数字创新识别工作。在验证了本文识别方法的稳健性后，本文在地区、企业性质和规模等方面探究了数字创新的特征事实，以期实证检验提供启发和思路。之后，本文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作用机制。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作用渠道。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数量和质量提升，对进一步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的作用，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。本文首次为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。

在当前中美博弈日趋复杂，美国对华科技封锁日益激烈的背景下，本文为如何建立全国统一大市场，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。本文的政策启示非常丰富：

(1) 坚持技术创新市场导向机制, 加快我国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“提质增效”, 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。本文的研究也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, 而在数字生产或服务流程创新上并不具备影响效应。此外, 我国当前市场需求萎靡, 作用也有限, 同时经过新冠疫情的打击, 我国居民收入更加雪上加霜。因此, 不仅要把目光放在需求端, 要寻求供给端的替代措施。2022年4月10日发布的《中共中央 国务院关于加快建设全国统一大市场的意见》指出要“进一步规范不当市场竞争和市场干预行为”。本文指出不仅要从供给端去健全公平竞争制度框架, 更要引导合理有序的市场竞争结构, 将我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。当前我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, 无法实现长远发展尤其是进行有效的基础的数字创新。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市场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来细分市场, 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, 提高城市群中心城市的数字创新溢出效应。同时也要推动各地区之间尤其是东部和其他地区之间的科技资源共享, 鼓励不同区域之间科技信息交流互动。例如, 可以利用“东数西算”工程, 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, 要加强对创新型中小企业原始创新和知识产权的保护。中小企业是我国数字创新的主力军, 适当给予部分保护, 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共性数字创新与中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效应, 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。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, 国有企业在全国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, 为我国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。
- 白重恩、杜颖娟、陶志刚、仝月婷, 2004:《地方保护主义及产业地区集中度的决定因素和变动趋势》,《经济研究》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。
- 蔡莉、杨亚倩、卢珊、于海晶, 2019:《数字技术对创业活动影响研究回顾与展望》,《科学学研究》第10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。
- 陈剑、黄朔、刘运辉, 2020:《从赋能到使能——数字化环境下的企业运营管理》,《管理世界》第2期。
- 陈晓红、于涛, 2013:《营销能力对技术创新和市场绩效影响的关系研究——基于我国中小上市企业的实证研究》,《科学学研究》第4期。
- 陈甬军、丛子薇, 2017:《更好发挥政府在区域市场一体化中的作用》,《财贸经济》第2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。

- 范红忠, 2007:《有效需求规模假说、研发投入与国家自主创新能力》,《经济研究》第3期。
- 冯艺、连俊华, 2020:《市场一体化对企业技术创新的影响研究——基于提升创新要素流动性视角的分析》,《价格理论与实践》第9期。
- 桂琦寒、陈敏、陆铭、陈钊, 2006:《中国国内商品市场趋于分割还是整合:基于相对价格法的分析》,《世界经济》第2期。
- 国务院发展研究中心课题组, 2005:《国内市场一体化对中国地区协调发展的影响及其启示》,《中国工商管理研究》12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。
- 郝项超、梁琪、李政, 2018:《融资融券与企业创新:基于数量与质量视角的分析》,《经济研究》第6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。
- 康志勇、张杰, 2008:《有效需求与自主创新能力影响机制研究——来自中国1980—2004年的经验证据》,《财贸研究》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。
- 李春涛、宋敏, 2010:《中国制造业企业的创新活动:所有制和CEO激励的作用》,《经济研究》第5期。
- 李雪松、孙博文, 2015:《密度、距离、分割与区域市场一体化——来自长江经济带的实证》,《宏观经济研究》第6期。
- 刘善仕、刘婷婷、刘向阳, 2007:《人力资源管理系统、创新能力与组织绩效关系——以高新技术企业为例》,《科学学研究》第4期。
- 刘善仕、孙博、葛淳棉、王琪, 2017:《人力资本社会网络与企业创新——基于在线简历数据的实证研究》,《管理世界》第7期。
- 刘小玄、吴延兵, 2009:《企业生产率增长及来源:创新还是需求拉动》,《经济研究》第7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。
- 吕越、田琳、吕云龙, 2021:《市场分割会抑制企业高质量创新吗?》,《宏观质量研究》第1期。
- 欧阳峤、汤凌霄, 2017:《大国创新道路的经济学解析》,《经济研究》第9期。
- 钱学锋、刘钊、陈清日, 2021:《多层次市场需求对制造业企业创新的影响研究》,《经济动态》第5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。
- 王磊、邓芳芳, 2016:《市场分割与资源错配——基于生产率分布视角的理论与实证分析》,《经济理论与经济管理》第11期。
- 王永进、冯笑, 2018:《行政审批制度改革与企业创新》,《中国工业经济》,第2期。
- 王永进、刘卉, 2021:《企业专利申请、出口与生产率动态》,《世界经济》,第6期。
- 闫俊周、姬婉莹、熊壮, 2021:《数字创新研究综述与展望》,《科研管理》第4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、郑文平、翟福昕, 2014:《竞争如何影响创新:中国情景的新检验》,《中国工业经济》第11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。
- 张璇、刘贝贝、汪婷、李春涛, 2017:《信贷寻租、融资约束与企业创新》,《经济研究》第5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。
- 周黎安、罗凯, 2005:《企业规模与创新:来自中国省级水平的经验证据》,《经济学(季刊)》第2期。
- 朱秀梅, 2008:《高技术产业集群创新路径与机理实证研究》,《中国工业经济》第2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”, *The Quarterly Journal of Economics*, 119(3), 1049-1090.
- Aghion, P., U. Akcigit, A. Bergeaud, R. Blundell, and D. Hémous, 2019, “Innovation and Top Income Inequality”, *The Review of Economic Studies*, 86(1), 1-45.
- 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, 2020, "The Heterogeneous Impact of Market Size on Innovation: Evidence from French Firm-Level Exports", NBER Working Paper, 24600.
- 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.
- Aghion, P., C. Harris, P. Howitt, and J. Vickers, 2001, "Competition, Imitation and Growth with Step-by-Step Innovation", *Review of Economic Studies*, 68(3), 467-492.
- Akcigit, U., S. Baslandze, and S. Stantcheva, 2016, "Taxation and The International Mobility of Inventors", *American Economic Review*, 106(10), 2930-81.
- Akcigit, U., J. Grigsby, T. Nicholas, and S. Stantcheva, 2022, "Taxation and Innovation in the Twentieth Century", *The Quarterly Journal of Economics*, 137(1), 329-385.
- Akcigit, U., and W. R. Kerr, 2018, "Growth Through Heterogeneous Innovations", *Journal of Political Economy*, 126(4), 1374-1443.
- Anderson, J. E., and E. Van Wincoop, 2004, "Trade Costs", *Journal of Economic literature*, 42(3), 691-751.
- Arrow, K., 1962, "Economic Welfare and The Allocation of Resources for Invention", *The Rate and Direction of Inventive Activity: Economic and Social Factors* (pp. 609-626). Princeton University Press.
- 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.
- Benkler, Y., 2006, "The Wealth of Networks: How Social Production Transforms Markets and Freedom", *Information Economics & Policy*, 19(2), 278-282.
- Boland Jr, R. J., K. Lyytinen, and Y. Yoo, 2007, "Wakes of Innovation in Project Networks: The Case of Digital 3-D Representations in Architecture, Engineering, and Construction. *Organization Science*", 18(4), 631-647.
- Bustos, P., 2011, "Trade Liberalization, Exports and Technology Upgrading: Evidence on the Impact of MERCOSUR on Argentinean Firms", *American Economic Review*, 101(1), 304-340.
- Corrado, C., C. Hulten, and D. Sichel, 2005, "Measuring Capital and Technology: An Expanded Framework", NBER Chapters, in: *Measuring Capital in the New Economy*, 11-46.
- 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.
- Devlin, J., M. Wei. Chang, K. Lee, and K. Toutanova, 2019, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding", In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171-4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Doh, S., and B. Kim, 2014, "Government Support for SME Innovations in The Regional Industries: The Case of Government Financial Support Program in South Korea", *Research Policy*, 43(9), 1557-1569.
- 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.
- Gobble, M. M., 2016, “News and Analysis of The Global Innovation Scene”, *Research-Technology Management*, 59(4), 2-9.
- Guellec, D., and C. Paunov, 2017, “Digital Innovation and the Distribution of Income”, NBER Working Paper, 23987.
- Guo, D., Y. Guo, and K. Jiang, 2016, “Government-Subsidized R&D and Firm Innovation: Evidence from China”, *Research policy*, 45(6), 1129-1144.
- Hall, B. H., and D. Harhoff, 2012, “Recent Research on The Economics of Patents”, *Annual Review of Economics*, 4(1), 541-565.
- Hall, B. H., A. Jaffe, and M. Trajtenberg, 2005, “Market Value and Patent Citations”, *RAND Journal of Economics*, 16-38.
- Hsu, P. H., X. Tian, and Y. Xu, 2014, “Financial Development and Innovation: Cross-Country Evidence”, *Journal of Financial Economics*, 112(1), 116-135.
- Klette, T. J., and S. Kortum 2004, “Innovating Firms and Aggregate Innovation”, *Journal of Political Economy*, 112(5), 986-1018.
- Lileeva, A. and D. Trefler, 2010, “Improved Access to Foreign Markets Raises Plant-level Productivity for Some Plants”, *Quarterly Journal of Economics*, 125(3), 1051-1099.
- 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”, *The Review of Economic Studies*, 75(1), 295-316.
- Melitz, M. J., and S. J. Redding, 2021, “Trade and Innovation”, CEPR Discussion Papers, 16264.
- 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.
- Naughton, B., 1999, “How Much Can Regional Integration Do to Unify China's Markets?”, Conference for Research on Economic Development and Policy, Stanford University.
- Poncet, S., 2003, “Measuring Chinese Domestic and International Integration”, *China Economic Review*, 14(1), 1-21.
- Reimers, N., and I. Gurevych, 2019, “Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks”, In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 3982–3992, Hong Kong, China. Association for Computational Linguistics.
- Romer, P. M., 1990, “Endogenous Technological Change”, *Journal of Political Economy*, 98(5), 71-102.
- Schumpeter, A. J., 1934, “The Theory of Economic Development: An Inquiry Into Profits, Capital, Credit, Interest, and the Business Cycle”, Harvard University Press.
- Schumpeter, A. J., 1942, “Capitalism, Socialism, and Democracy”, New York: Harper and Brothers.
- Schwab, K., 2016, “The Fourth Industrial Revolution”, *The Cambodian Journal of International Studies*, 65.
- 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.
- Young, A., 2000, “The Razor's Edge: Distributions and Incremental Reform in the People's Republic of China”, *Quarterly Journal of Economics*, 115(4), 1091-1136.

附录

A1 主要现有研究对于数字创新的定义

附录表 1 现有主要研究对于数字创新的定义及涵盖范围

定义	涵盖范围	来源
数字部件和物理部件进行新的组合，以生产新的产品。	数字产品	Yoo et al. (2010); Hanelt et al. (2021)

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的一种创新。	数字产品、数字流程	Nambisan et al. (2016); Guellec & Paunov (2020)
将新兴数字技术嵌入到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。	数字产品、数字服务	Thomas (2015); Khin & Ho (2019)
企业利用数字技术手段来提升创新绩效和改善运营效率。/利用数字技术重塑用户价值、转变运营模式从而实现企业价值。	数字商业模式	Boland et al. (2007); Berman (2012)

A2 严格筛选方法

核心思路是：对于有详细且精确描述的类别，专利的识别严格按照描述来进行，需要符合描述的核心特征；对于只有简要且宽泛描述的类别，主要根据这个专利的用途和特征，与描述相符的专利才可以被划分进去。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。

具体而言，对于 01 数字产品制造业，其中的 0101 计算机制造、0102 通讯及雷达设备制造、0103 数字媒体设备制造、0104 智能设备制造、0105 电子元器件及设备制造和 0106 其他数字产品制造业都具有相当详细且明确的定义，因此可以按照严格识别对照方法来识别。以一些反例为例：《一种自发电灯》的详细描述为“本实用新型公开了一种自发电灯，灯盘上安装有开关，灯盘两端安装有腕带，灯盘中间设置有环形太阳能发电板，太阳能发电板中间安装有灯泡，灯泡内部安装有钢容器和压缩机，压缩机与钢容器相连通，半导体设置在钢容器的下方，太阳能发电板通过开关与压缩机相连接，半导体与灯泡电连接。该自发电灯利用身体温和被压缩的氟利昂之间形成的温差，通过半导体发电点亮灯泡，携带方便，避免了对外接电源的依赖，音乐播放器用于播放音频信息，增加了娱乐性。”虽然该专利含有音乐播放器，单一的音乐播放器可以归类到 0103 数字媒体设备制造的音响设备类别上，但是《一种自发电灯》本质上是电灯，虽然含有音乐播放器但是按照严格定义并不能归类到数字媒体设备中。但是该专利在广义上可以被认为数字媒体设备。又如《机器人自主定位系统》“机器人自主定位系统，属于机器人智能控制装置，解决机器人在地下环境作业的高精度自主定位问题。本发明包括安装平台、设置于安装平台上的传感器子系统、数据处理子系统以及电压转换模块；安装平台上部为承载传感器子系统的转台，安装平台下部的固定基座为密封腔式结构，内部安装数据处理子系统和电机驱动模块；传感器子系统包括惯性传感器模块和三维数字罗盘，数据处理子系统包括中央控制器及装载于其内部的数据预处理模块、数据融合模块和航迹推算模块；本发明可用于在煤矿、隧道、溶洞等地下环境工作的智能机器人，如机器人，盾构机，地铁机车等的自主定位，为智能机器人在地下环境完成勘测、探险和搜索等任务提供有效的位姿信息。”该专利严格来说，只是智能机器人单独的系统和配件，并不能完全构成机器人，因此严格意义上并不能归类到 0104 智能设备制造的工业机器人中来，但是可以归类到应用软件类目中去。

对于 02 数字产品服务业，其下面的四个子类别定义较为宽泛。那么对于这样的宽泛定义的类别，我们将与其相关的专利归到该类别中。以 0201 数字产品批发为例，与数字产品批发过程有关的物流，支付，电商平台，采购销售系统，产品包装，软件传输系统等都可以算入该类别，例如《一种数字出版产品及其权益网上交易方法及系统》、《采用数字广播的电子商务系统和方法》、《用于跨边界采购的装置、方法和系统》、《一种国际物流快速运载信息平台》、《国际贸易快易工作机》等。

对于 03 数字技术应用业和 04 数字要素驱动业，其下面的子类别具有非常详细的定义，我们就严格遵循定义来识别专利是否属于该类别。例如，对于 0303 互联网相关服务，其中描述到“不包括互联网支付、互联网基金销售、互联网保险、互联网信托和互联网消费金融等互联网信息服务”，我们就将任何的互联网支付等相关专利不归类到该类别上。又如，0401 互联网平台的所有描述中都含有“互联网平台”概念，只有一个创新属于互联网“平台”范畴时，才会被归类到该类别上。

A3 每个类别数字创新阈值及其准确率

附录表 2 每个类别数字创新阈值及其准确率

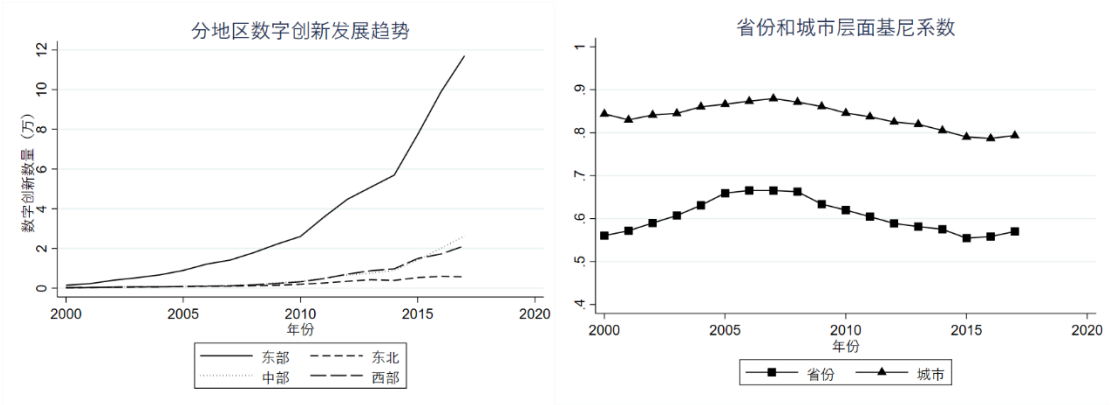
行业	前 100 严格准确率	阈值严格准确率	阈值	下一临界严格准确率
0101	98%	53%	0.55	26%
0102	100%	65%	0.6	41%
0103	93%	69%	0.55	40%
0104	98%	59%	0.55	32%
0105	92%	58%	0.6	37%
0106	92%	55%	0.55	34%
0201	78%	52%	0.55	31%
0202	79%	56%	0.6	32%
0301	100%	67%	0.5	43%
0302	100%	74%	0.5	48%
0303	82%	77%	0.5	43%
0304	100%	68%	0.45	35%
0305	100%	81%	0.65	37%
0401	88%	65%	0.65	32%
0402	78%	59%	0.65	26%
0404	90%	66%	0.6	45%
0405	92%	62%	0.65	30%
0407	80%	59%	0.7	28%
平均	91%	64%	0.58	36%

注：需要说明的是，我们删掉了 0203 数字产品租赁这个类别，因为数字产品租赁的专利非常之少，会非常大影响后续进一步分析的准确率^①；对于 0204 数字产品维修、0403 互联网金融和 0406 数据资源与产权交易类别，考虑到相关专利较少，综合准确率和数量，我们只将前 100 个专利（0204 行业：严格准确率 55%，临界值为 0.6659；0403 行业：严格准确率 59%，临界值为 0.6944；0406 行业：严格准确率 50%，临界值为 0.6642）识别为该类别的创新^②。

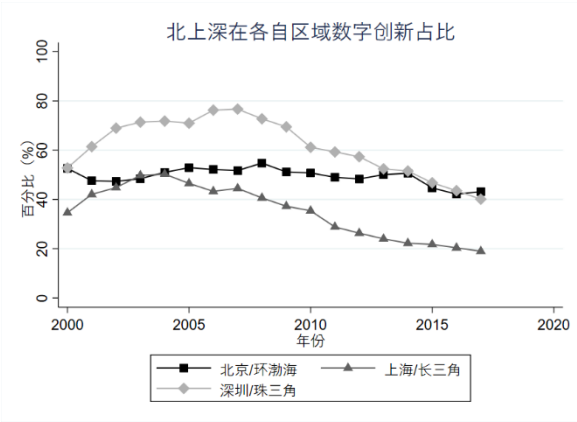
^①通过字段匹配可以发现：同时包含“租”和“电子”的专利只有 1 个属于数字产品租赁，同时包含“租”和“数字”的专利也只有 1 个属于数字产品租赁，同时包含“租”和“影音”/“电脑”/“笔记本”的专利没有数字产品租赁相关专利。

^②对于 0204 数字产品维修，我们通过一些字段匹配如“修”、“手机”、“计算机”等发现属于该行业专利非常少，总共不超过 100 个。这可能是由于行业特性所导致的，计算机等设备的维修具有一般性，维修店铺和维修人员往往不会针对其进行独特的创新并申请专利。对于 0403 互联网金融，由于我们使用的是严格识别策略，该类别下属三个子类定义十分严格，如网络借贷服务的主体是专门从事网络借贷信息中介业务活动的金融信息中介公司、非金融机构支付服务的主体是非金融机构，这样就直接排除了银行、证券、基金等金融机构，涉及到的范围大大缩小。对于 0406 数据资源与产权交易，我们也通过“数据资源”“数据产权”和“交易”等字段匹配发现相关专利很少。

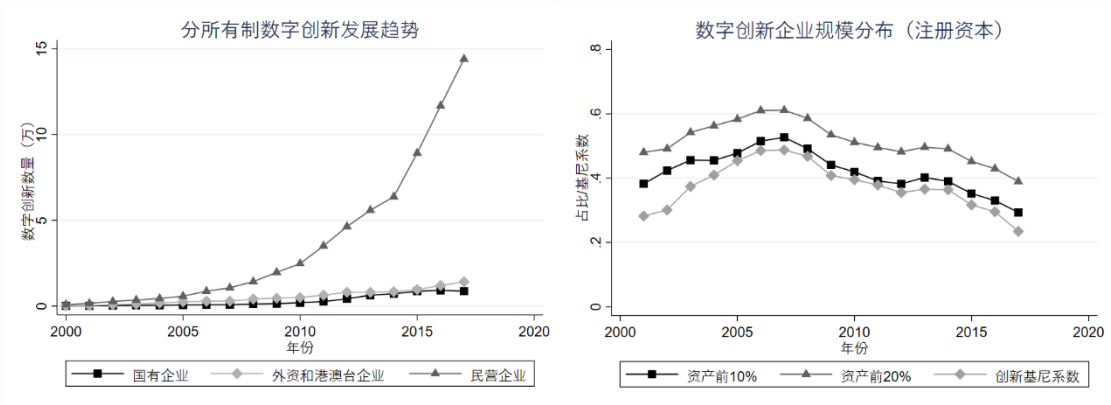
A4 数字创新特征事实



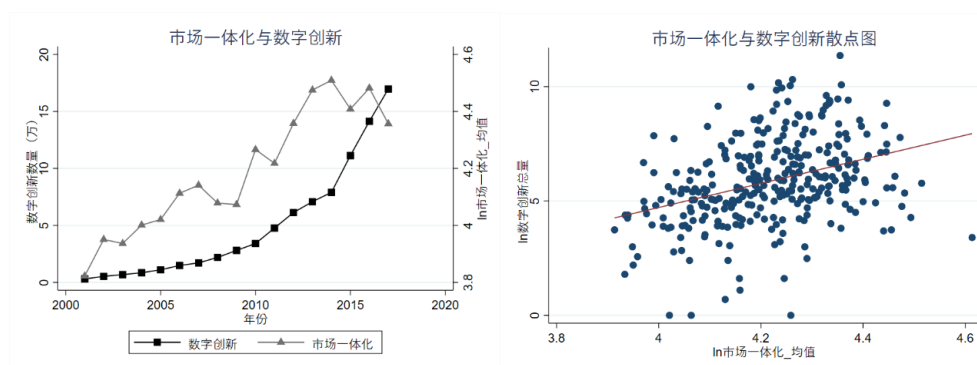
附录图 1 数字创新发展不平衡性



附录图 2 “北上深”在各自区域的数字创新占比



附录图 3 数字创新在企业所有制和企业规模上的分布



附录图 4 市场一体化与数字创新

A5 异质性分析

附录表 3 地区异质性回归结果

	(1)	(2)	(3)	(4)
	东部		非东部	
	Quantity	Ave_cite	Quantity	Ave_cite
MarketInt	0.113*** (0.019)	0.066*** (0.018)	0.024 (0.028)	0.002 (0.026)
Age	0.043*** (0.001)	0.018*** (0.001)	0.036*** (0.002)	0.016*** (0.002)
_cons	0.032*** (0.008)	0.046*** (0.008)	0.089*** (0.013)	0.076*** (0.012)
N	752402	671034	335095	297422
CapitalxYear	√	√	√	√
EmpxYear	√	√	√	√
Year Fixed Effect	√	√	√	√
Firm Fixed Effect	√	√	√	√
R ² _Adj	0.19	0.12	0.17	0.11

注：我们使用 SUEST（基于似无相关模型 SUR 的检验）对组间系数差异进行了检验，数字创新数量和质量的地地区系数差异 χ^2 统计量分别为 370.41 和 25.56，均在 1% 水平下显著，可以拒绝组间系数无差异的原假设，认为组间系数存在显著差异。Quantity 为企业数字创新数量的衡量，Ave_cite 为利用平均被引计算企业数字创新质量的衡量。Capital × Year 为注册资本和年份虚拟变量的交互，Emp × 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。***，**，* 分别表示 1%、5% 和 10% 的水平上显著；括号内的标准误为稳健标准误。

附录表 4 企业所有制异质性回归结果

	(1)	(2)	(3)	(4)	(5)	(6)
	国有企业		外资和港澳台企业		民营企业	
	Quantity	Ave_cite	Quantity	Ave_cite	Quantity	Ave_cite
MarketInt	0.069 (0.115)	-0.004 (0.102)	-0.009 (0.051)	0.045 (0.049)	0.084*** (0.016)	0.037** (0.016)
Age	0.100*** (0.010)	0.033*** (0.008)	0.044*** (0.004)	0.015*** (0.003)	0.040*** (0.001)	0.018*** (0.001)
_cons	-0.071 (0.055)	0.041 (0.048)	0.093*** (0.023)	0.069*** (0.022)	0.049*** (0.007)	0.058*** (0.007)
N	20917	19414	99243	90930	967337	858112

CapitalxYear	√	√	√	√	√	√
EmpxYear	√	√	√	√	√	√
Year Fixed Effect	√	√	√	√	√	√
Firm Fixed Effect	√	√	√	√	√	√
R ² _Adj	0.36	0.20	0.30	0.17	0.16	0.11

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。Capital × Year 为注册资本和年份虚拟变量的交互，Emp × 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。***，**，*分别表示 1%、5%和 10%的水平上显著；括号内的标准误为稳健标准误。

附录表 5 企业规模异质性回归结果

	(1)	(2)	(3)	(4)
	Quantity	Quantity	Ave_cite	Ave_cite
MarketInt	0.315*** (0.068)	0.320*** (0.027)	0.101* (0.061)	0.210*** (0.027)
Age	0.064*** (0.001)	0.064*** (0.001)	0.031*** (0.001)	0.031*** (0.001)
MarketIntxCapital	-0.023** (0.009)		-0.003 (0.008)	
MarketIntxEmp		-0.056*** (0.007)		-0.040*** (0.007)
_cons	0.089*** (0.010)	0.087*** (0.010)	0.088*** (0.008)	0.088*** (0.008)
N	1087497	1087497	968456	968456
CapitalxYear	√	√	√	√
EmpxYear	√	√	√	√
Year Fixed Effect	√	√	√	√
Firm Fixed Effect	√	√	√	√
R ² _Adj	0.19	0.19	0.12	0.12

注：Quantity 为企业数字创新数量的衡量，Ave_cite 为利用平均被引计算企业数字创新质量的衡量。Capital × Year 为注册资本和年份虚拟变量的交互，Emp × 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。***，**，*分别表示 1%、5%和 10%的水平上显著；括号内的标准误为稳健标准误。

A6 稳健性检验

附录表 6 稳健性检验 1

	(1)	(2)	(3)	(4)	(5)	(6)
	Quantity	Ave_cite	Max_cite	Quantity	Ave_cite	Max_cite
MarketInt_muni_direct	0.053*** (0.013)	0.035*** (0.013)	0.045*** (0.015)			
MarketInt_direct				0.052*** (0.015)	0.024* (0.014)	0.032* (0.017)
Age	0.037*** (0.001)	0.017*** (0.001)	0.018*** (0.001)	0.040*** (0.001)	0.017*** (0.001)	0.018*** (0.001)
_cons	0.071*** (0.006)	0.067*** (0.006)	0.084*** (0.007)	0.064*** (0.007)	0.066*** (0.006)	0.082*** (0.008)
N	1368607	1220377	1220377	1069473	952562	952562
CapitalxYear	√	√	√	√	√	√

EmpxYear	√	√	√	√	√	√
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 × Year 为注册资本和年份虚拟变量的交互，Emp × 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。***，**，*分别表示 1%、5%和 10%的水平上显著；括号内的标准误为稳健标准误。

附录表 7 控制更严格的固定效应下对企业数字创新数量的影响

Quantity	(1) 总体	(2) 发明	(3) 实用新型	(4) 数字产品	(5) 数字生产/服务
MarketInt	0.083*** (0.016)	0.050*** (0.010)	0.037*** (0.012)	0.074*** (0.015)	0.035*** (0.008)
Age	0.042*** (0.001)	0.006*** (0.001)	0.036*** (0.001)	0.038*** (0.001)	0.007*** (0.001)
_cons	0.047*** (0.007)	0.037*** (0.005)	0.027*** (0.006)	0.039*** (0.007)	0.020*** (0.004)
N	1087316	1087316	1087316	1087316	1087316
CapitalxYear	√	√	√	√	√
EmpxYear	√	√	√	√	√
Ind-Year Fixed Effect	√	√	√	√	√
Firm Fixed Effect	√	√	√	√	√
R ² _Adj	0.19	0.26	0.13	0.17	0.25

注：本表汇报了在控制行业年份固定效应和企业固定效应下，市场一体化对企业（不同种类）数字创新数量的回归结果。Capital × Year 为注册资本和年份虚拟变量的交互，Emp × Year 为参保人数与年份虚拟变量的交互。回归中控制了行业年份固定效应和企业固定效应。***，**，*分别表示 1%、5%和 10%的水平上显著；括号内的标准误为稳健标准误。

附录表 8 控制更严格的固定效应下对企业数字创新质量的影响

Quality	(1) 总体	(2) 发明	(3) 实用新型	(4) 数字产品	(5) 数字生产/服务
MarketInt	0.040*** (0.015)	0.034*** (0.013)	0.013 (0.011)	0.035** (0.014)	0.029*** (0.009)
Age	0.017*** (0.001)	0.002* (0.001)	0.017*** (0.001)	0.015*** (0.001)	0.002*** (0.001)
_cons	0.059*** (0.007)	0.048*** (0.006)	0.031*** (0.005)	0.051*** (0.006)	0.022*** (0.004)
N	968287	968287	968287	968287	968287
CapitalxYear	√	√	√	√	√
EmpxYear	√	√	√	√	√
Ind-Year Fixed Effect	√	√	√	√	√
Firm Fixed Effect	√	√	√	√	√
R ² _Adj	0.12	0.14	0.08	0.11	0.15

注：本表汇报了在控制行业年份固定效应和企业固定效应下，市场一体化对企业（不同种类）数字创新质量的回归结果。Capital × Year 为注册资本和年份虚拟变量的交互，Emp × Year 为参保人数与年份虚拟变量的交互。回归中控制了行业年份固定效应和企业固定效应。***，**，*分别表示 1%、5%和 10%的水平上显著；括号内的标准误为稳健标准误。

A7 市场规模机制分析

附录表 9 市场规模渠道分类回归结果

	(1)	(2)	(3)	(4)
	数字产品		数字生产和服务	
	Quantity	Ave_cite	Quantity	Ave_cite
ΔDemand	0.076*** (0.008)	0.021*** (0.007)	0.009* (0.005)	-0.003 (0.005)
_cons	0.067*** (0.002)	0.067*** (0.002)	0.038*** (0.001)	0.039*** (0.001)
N	1423077	1272760	1423077	1272760
其他控制变量	√	√	√	√
固定效应	firm, ind#year			
R ² _Adj	0.18	0.11	0.26	0.16

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。其他控制变量包括：企业年龄、注册资本和年份虚拟变量的交互、参保人数与年份虚拟变量的交互。***，**，*分别表示 1%、5%和 10%的水平上显著；括号内的标准误为稳健标准误

A8 市场竞争机制分析

附录表 10 技术的累积效应回归结果

	(1)	(2)
	Quantity	Ave_cite
MarketInt	0.076*** (0.015)	0.039*** (0.015)
Age	0.041*** (0.001)	0.017*** (0.001)
DigitalInnovation_Past	0.095*** (0.018)	0.034*** (0.008)
_cons	0.051*** (0.007)	0.058*** (0.007)
N	1087497	968456
CapitalxYear	√	√
EmpxYear	√	√
Year Fixed Effect	√	√
Firm Fixed Effect	√	√
R ² _Adj	0.19	0.12

注：Quantity 为企业数字创新数量的衡量，Ave_cite 为利用平均被引计算企业数字创新质量的衡量。DigitalInnovation_Past 为企业过去累积的数字创新数量。Capital x Year 为注册资本和年份虚拟变量的交互，Emp x 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。***，**，*分别表示 1%、5%和 10%的水平上显著；括号内的标准误为稳健标准误。

附录表 11 市场竞争渠道分类回归结果

	(1)	(2)	(3)	(4)
	数字产品		数字生产和服务	
	Quantity	Ave_cite	Quantity	Ave_cite
MarketCom	0.136***	0.089***	0.037***	0.020*

	(0.015)	(0.014)	(0.010)	(0.011)
MarketCom ²	-0.006***	-0.004***	-0.001***	-0.001*
	(0.001)	(0.001)	(0.000)	(0.000)
_cons	-0.704***	-0.431***	-0.204***	-0.089
	(0.106)	(0.100)	(0.073)	(0.078)
N	1423562	1273200	1423562	1273200
其他控制变量	√	√	√	√
固定效应		firm, ind#year, city#year		
R ² _Adj	0.18	0.11	0.26	0.16

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。其他控制变量包括：企业年龄、注册资本和年份虚拟变量的交互、参保人数与年份虚拟变量的交互。***，**，*分别表示 1%、5%和 10%的水平上显著；括号内的标准误为稳健标准误。