参赛队员姓名： John Wang

中学： 上海中学国际部

省份： 上海

国家、地区： 中国

指导老师姓名： 樊海潮

指导老师单位： 复旦大学经济学院

论文题目： 数字化与制造业企业创新—基于中国 A 股上市制造业企业的研究

数字化与制造业企业创新

——基于中国 A 股上市制造业企业的研究

作者: John Wang

论文摘要:

实践证明，特别是经历了全球范围的重大疫情考验：作为实体经济的基础，制造业再次毫无疑问确认是国民经济不可替代的重要支柱，对于稳定经济增长、扩大居民就业、繁荣商品市场、推动创新方面发挥着极其重要和长远的作用。同时，制造业也逐渐成为数字化的主战场。当前世界正处于新工业革命的窗口期，中国制造业企业能否借力数字化实现创新发展、高质量发展，事关中国经济能否成功实现新旧动能转换、抢占数字时代的发展先机。

然而，相关领域的文献依然存在部分缺陷。第一，现有文献未对数字化能否促进企业业绩增长达成共识。第二，文献中普遍使用的度量企业数字化水平的方式仍存在改进空间。第三，未有文献从微观企业的角度研究数字化对创新的作用，缺乏微观基础。第四，数字化对企业产生的异质性作用未被深入研究。

本文通过对 2013-2017 年中国 A 股上市公司企业年报文本进行挖掘，构建了制造业企业的数字化指数，并将其与企业财务指标和专利数据匹配，构建平衡面板数据，考察制造业企业数字化能否促进企业创新。回归结果显示，制造业企业数字化水平的提高推动了企业的创新产出。本研究分别从更换 TOBIT 回归模型、使用工具变量防止潜在内生性、使用 LASSO 回归进行特征筛选三方面证明这一结论的稳健性。进一步地，本研究发现，研发投入是数字化促进企业创新的重要中介。最后，本文研究了数字化促进企业创新的异质性作用，发现资金约束较宽松的企业、规模较小的企业、国有企业和传统行业企业更加受益于数字化过程。

本文从数据上证明了数字化对企业创新的重要促进作用，数字化程度更高的制造业企业在新工业革命时期将会有更强大的创新力和竞争力。本研究的实证结果为中国制造业企业的转型升级提供了一定的政策参考：政府应当积极推进数字化与实体制造业相结合，促进高端制造的发展，同时通过一些制度层面的措施，切实缓解企业数字化过程中存在的融资难问题。

关键词: 数字化；企业创新；制造业转型；文本挖掘

目录

一、引言 **4**

二、研究假说 **6**

（一）数字化促进企业创新的作用机理 6

（二）数字化促进制造业企业创新的异质性作用 7

三、研究设计 **9**

（一）数据来源 9

（二）变量设定 9

1、词典的构造 9

2、文本挖掘与核心解释变量构建 11

3、被解释变量与中介变量 11

4、其他控制变量 11

（三）模型设定 12

1、主模型 12

2、中介效应 13

3、异质性分析 13

四、实证结果 **13**

（一）描述性分析 13

（二）数字化对制造业企业创新的影响 15

（三）稳健性检验 15

（四）进一步研究：数字化、研发投入与企业创新 17

（五）异质性分析 18

五、政策建议 **20**

参考文献 **22**

致谢 **25**

老师简历： **26**

作者简历： **32**

一、引言

作为实体经济的基础，制造业再次毫无疑问确认是国民经济不可替代的重要支柱，对于稳定经济增长、扩大就业市场、繁荣市场经济、促进创新方面发挥着极其重要和长远的作用。中国政府高度重视制造业的发展，并积极推动制造业的转型和升级，国家最高领导人习近平总书记明确指出“制造业高质量发展是中国经济高质量发展的重中之重”。目前世界正处于第四次工业革命的窗口期，汇集了产业链和价值链的制造业，必然将会成为全球经济竞争的主战场，因此，如果中国制造业企业无法从简单的原料加工、功能外观仿制、低科技含量向高附加值、有自主知识产权、自主创新和自有品牌的方向转型，中国工业经济将无法实现进一步的转型和发展。但当前中国制造业仍有大而不强、低端产能过剩、资源利用效率低、创新能力不足等问题，与发达国家的高端制造模式仍有一定差距。

随着互联网基础设施的完善，作为通用技术（general purpose technology）的大数据技术和作为补充技术（complementary technology）的云技术和区块链技术实现快速发展，数字产业加速扩张，已经成为中国新的经济增长点（邱子迅和周亚虹，2021）。全球各主要工业化国家相继推出大数据政策，以推动数字经济的发展。美国最早于 2011 年制定《大数据研究与发展计划》，标志着美国已经将发展大数据升高到国家战略；澳大利亚政府于 2013 年颁布《公共服务大数据战略》，通过数字技术推动公共部门改革（张勇进和王璟璇，2014）。中国政府同样重视数字化的应用，特别是制造业领域，强力推进与数字产业的融合发展。2016 年，国务院颁布《关于深化制造业与互联网融合发展的指导意见》，该文件特别强调了深化制造业与互联网融合发展的重要性。特别地，该《意见》明确指出了制造业在数字化时代中的重要位置，是我国实施“互联网+”行动的主战场。因此制造业能否借力数字化实现转型发展，事关中国能否抢占数字时代的发展先机。

企业的数字化进程能否促进企业的高质量发展引起了国内外经济学家们的广泛关注，但是学者们获得的结论却不尽相同。部分学者发现，虽然大部分公司在数字化方面进行了大量的投资，但是收效甚微。如 Solow（1987）曾宣称，“计算机的作用随处可见，但不包括对生产率的提升”，这就是著名的“信息技术生产率悖论”。Bharadwaj et al.（1999）研究发现，虽然公司进行了大量对网络和数字化的投资，但是公司之间的投资绩效却存在巨大差异。Hajli et al.（2015）同样发现，并不是所有企业在数字化的进程中受益，获得改进的只是小部分企业。Lane et al.（2001）发现，与发达国家相比，发展中国家企业数字化的效率明显较弱，

因为发展中国家在全球产业链中较弱的技术位置会使得自身知识吸收能力和组织学习能力较弱。因此，特别是在发展中国家，数字化与公司绩效之间错综复杂的关系使得投资人对企业的数字化项目投资表现出越发谨慎的态度（Lee et al.，2005）。

也有部分学者发现，近期企业的数字化进程提升了企业的绩效。Brynjolfsson 和 Hitt（2000）认为企业的数字化有助于帮助企业合理计划生产流程、快速应对消费者需求的变化、增加组织机动性进而提高企业绩效。杨德明和刘泳文（2018）通过对中国上市公司进行详细研究，发现传统企业与互联网的融合（即“互联网+”）有助于公司业绩的提升，这其中的机制是企业的互联网化促进了企业差异化的形成，但是并没有提升企业的成本领先水平。

针对“信息技术生产率”悖论形成的原因和研究中出现的分歧，Brynjolfsson（1993）给出了自己的解释：首先，矛盾可能来源于企业数字化指标的测量失真；其次，企业数字化的进程并非一蹴而就的，其对企业绩效产生作用会有一个时滞；第三，企业内部自身的管理漏洞而引起的问题。以上文献的研究大都是使用信息通讯设备（软件、固话、电脑等）的总价值或数量，或是专家打分作为企业数字化的度量。前者无法实现对企业数字技术应用和企业数字战略的测度，后者则存在较大的主观性，容易受到打分专家的认知和情感倾向的影响。此外，虽然有学者发现企业之间的数字化效果存在差异，但是并没有从企业管理和组织结构的角度考察数字化对企业绩效提升的异质性作用。除了现有研究存在较大分歧外，作者发现大部分文献考察的对象都聚焦于企业的业绩，即以企业的产出或企业的资产收益率作为研究对象。然而企业的产出和企业的资产收益率是一种企业发展的“量”的指标，并不能很好地表示企业发展的“质”。同时，现有研究互联网化、数字化与经济增长质量的文章大都以地区的生产率为研究对象（郭家堂和骆品亮，2016；邱子迅和周亚虹，2021），尚未有文献从微观企业的角度研究数字化对创新的作用，缺乏微观基础。本文的目的是通过实证研究弥补相关文献的不足，试图回答：中国制造业企业的数字化进程是否会推动企业的创新行为？在不同组织架构的企业之间，这种推动作用是否会表现出异质性？

本研究通过对中国 A 股上市制造业企业 2013-2017 年的年报进行文本挖掘，构造了制造业企业数字化指数，并将该指数与企业财务指标匹配，构建了五年的平衡面板数据。通过建立回归模型，本文发现制造业企业数字化显著促进了企业的创新产出，且本研究从更换回归模型、使用工具变量、特征筛选三个角度对主要结论进行稳健性检验。在研究数字化促进企业创新的同时，本文同样将研发投入纳入研究范畴，具体考察了研发投入在这个过程中起到的中介作用。结合现有考察企业创新产出影响因素的文献和数字经济时代的新特征，本文

进一步研究了哪些特征的企业更加受益于数字化进程。最后本文基于实证分析中获得的结论提出相应的政策建议。

本文余下的结构安排如下：第二部分介绍分析框架与作用机制，第三部分说明本文的指标构建方法、使用的数据和计量模型，第四部分介绍本文的回归分析结果，第五部分对本文的核心结论进行总结，并据此提出政策建议。

二、研究假说

（一）数字化促进企业创新的作用机理

制造业企业数字化指企业通过引进先进的数字技术，在经营过程中，实现生产、运营、管理和销售等各个层面的信息化、数据化和智能化，其本质是不断增强企业自身竞争力以实现效益和利润增值的战略行为（戚聿东和蔡呈伟，2020）。企业的战略行为是一个事关企业全局的具有整体性和长期性的决策，这与企业所面临的竞争环境密不可分。因此，在分析企业数字化如何促进企业创新之前，本研究首先要分析其在数字经济时代所呈现的新特征。

数字经济的重要基础设施是互联网，与互联网相关的领域普遍存在两个特征：一是固定成本很高但是边际成本极低，二是具有极强的网络外部性（谢平等，2015）。前者意味着，企业进行数字化前期需要投入大量的设计、开发和调试等费用支出，但是由于互联网属性的存在，信息的传递成本接近于零，每增加一单位的产出或增加一个产品的用户的边际成本极低。后者意味着，使用该网络的成员越多，信息交流越充分，每一个成员参与得到的正效用就越大，因此网络产生的价值也就更大，从而进一步促进更多的用户接入该网络，在此基础上创造更高的网络价值。。

基于数字经济的特点，本研究从企业效率、互联网思维和市场环境变化三个角度分析数字化对企业创新的作用。

云计算、区块链等大数据技术的补充技术使海量数据的存储和计算成为可能，降低了个体利用和处理信息的成本，社会整体的运行效率大幅提高。制造业企业更加追求生产效率，数字技术的应用可以使企业生产流程数据化，有助于发现生产过程中的缺陷，优化生产流程。因此企业家往往会大胆尝试将数字化技术与传统生产结合，这种基于数据的互联网生产模式本身即是一种创新，特别是大数据、云计算技术、人工智能及机器人，近年来在制造业中得到了广泛的应用。如：以云计算为基础构建的应用更容易帮助管理者深入了解供应链，可以

随时查看库存、生产水平、可用容量、订单状态等信息；对数控机床和 3D 打印进行基于云计算的管理，可以消除复杂的工艺流程；借助云计算部署的应用更容易进行虚拟仿真工具和建模，从而可以更容易进行试生产并减少浪费；生产者利用云计算技术，通过大数据更加精确有效地判断消费者的消费习惯和消费特征，个体的异质性需求得到了满足，生产实现精细化、特异化。近几年随着云技术研发的不断深入，相关专利数量从 2014 年的 2262 项迅速增长到 2017 年的 6618 项。

数字经济的发展在提高企业效率的同时，改变了制造业企业面临的市场环境，加剧了行业内的竞争。首先，由于数字技术尚处于技术引入的阶段，目前缺乏相对成熟的与传统行业相融合的应用，因此对于企业而言，特别是小型初创企业，会为了争夺市场份额获取行业领先地位，在巨大潜在利益的推动下，进行多元化和数字化的探索，推进数字化与传统行业相融合（谢伏瞻，2019），进而催生了诸如智慧安防、自动驾驶、新材料、物联网等新兴产业。其次，跨界竞争成为极其常见的现象，在网络外部性的作用下，互联网巨头很容易进入传统行业，并击败行业原来的巨头，加剧了传统制造业企业内部的竞争，迫使行业内的企业加快数字化转型进程，从而倒逼企业创新。最后，数字化促进了信息在企业之间的流通，在这种环境下，企业需要在动态竞争中保持高度的竞争积极性和警觉性，从而加快创新的步伐。

数字化除了通过将高效的技术渗透进企业的生产流程推进企业创新外，其带来的以“开放”、“平等”、“协作”、“共享”为主要内涵互联网思维（李海舰等，2014）同样有助于提高企业的创新水平。具体而言，生产由企业生产向社会生产转变，即企业在自身优势领域集中资源突破，在自身的非优势领域选择与外部企业合作，从原先的经营短板补缺点向经营长板增优点转变，有助于企业创新水平的提高。如，小米集团将部件生产和组装委托给质量好成本低的工厂，自身的业务全部放在新产品研发和用户沟通上，大幅提升了公司的创新水平。小米在 2012 年之前仅申请了 257 项专利，但是在 2015 年其专利申请数迅速增长至 1300 项。

综上，本研究认为，制造业企业的数字化会显著促进企业创新。

（二）数字化促进制造业企业创新的异质性作用

考虑到现阶段虽然很多制造业企业进行了数字化探索，但是不同企业之间数字化的效果却存在巨大差异。数字化与传统制造相结合本身是一种基于新通用技术的创新行为，因此一些影响企业创新的因素可能会对数字化促进企业创新的效果产生调节作用。但是同时，数字化时代下创新过程又表现出了一些不同的特点，这些新特点同样是本研究需要关注的。

首先，企业创新行为是一个具有典型投入沉没性、过程不可逆性以及产出不确定性的长

周期高风险活动，其持续性需要稳定、充足的金融资源作为保障（Hottenrott 和 Peters，2012；王玉泽等，2019）。企业如果面临较强的资金约束，其创新行为的持续性将会受到严重影响。制造业企业的数字化创新同样满足以上特征，且其过程的不确定性更高，研发周期更长。企业在引进数字技术、将数字技术与传统生产流程相结合、培训员工适应并熟练新技术等过程都中需要投入大量的支出，因此数字化转型的效果一定程度上取决于企业当前的资金能力，资金约束越强的企业在数字化创新的过程中受到的限制越多，因此这些企业的创新产出也相对较低。

其次，新技术和新产业创造的价值在不同企业中的分配是不平衡的。行业领先的企业希望通过加快新技术在实际生产中的应用巩固甚至进一步强化其在行业中的优势地位，同时已经具备一定数字化基础和技术能力的后发企业也希望借助数字化的机遇进行追赶，因此行业内的竞争将会更加激烈。行业领先的企业可能在数字化创新的过程中失败，而技术跟随者可能利用其基础设施优势或工艺优势，成为数字化时代制造业竞争的赢家。在数字化的进程中，企业的规模优势可能不再是决定企业竞争力的核心因素，一些初创企业，特别是一些先导行业的企业，完全可以借力数字化，实现企业创新水平的飞跃，进而在竞争中获得优势地位。综上，我们无法确定在数字化与制造业相融合的时期，企业规模如何影响企业数字化创新的效果，而是需要实证给出检验。

最后，现有多数文献表明，由于国有企业所有者空置，高管监督和激励不足的问题在国有企业中普遍存在（Shleifer，1998），不利于企业创新（王宇等，2020）。同时，制造业数字化本质是将数字技术与传统技术融合，是一种补充性技术的应用，满足分散式创新的特征（Wu et al.，2019），这种创新的效果往往受制于企业的组织架构，特别是企业高管的权利（刘政等，2020），国有企业权力集中的特点可能会削弱企业数字化创新的效果。但是另一方面，现有研究发现，相比于其他所有制的企业，国有企业在融资上面临较小的约束（余明桂等，2019），往往能够获得大量的银行贷款和政府补贴，从而有利于增强它们数字化的创新能力。综合来看，企业所有制在数字化促进企业创新中起到的作用是不确定的，同样需要我们通过实证进行检验。

三、研究设计

（一）数据来源

本文以 2013-2017 年中国 A 股的上市制造业企业作为研究对象，将通过文本挖掘获得的企业数字化指数、企业专利申请和企业财务指标进行匹配，剔除样本时间段内新上市、退市和重组的企业，删除年报文本无法挖掘和具有缺失值的样本，最终获得 920 个制造业企业在 2013 年至 2017 年五年内的平衡面板数据（即，总样本量为 4600）。本文进行挖掘的年报文本来源于巨潮信息网，企业专利申请数据来源于 WIND 数据库，其他财务指标数据来源于 CSMAR 数据库。下面首先说明本文的核心自变量，企业数字化指数的构造方式。

（二）变量设定

研究数字化对企业创新的影响的最大难点在于如何量化企业的数字化程度。除了借助对公司高管对数字化的态度或对公司信息通讯设备存量的调查外（Brynjolfsson 和 McElheran，2016；张三峰和魏下海，2019），目前主流的方法是通过公司年报的内容测试企业数字化的程度（杨德明和陆明，2017），因为公司的年报往往会对公司发展近况进行披露。如，杨德明和刘泳文（2018）通过阅读年报中的董事会报告部分，统计其中关于“互联网+”的相关描述，进行打分，构造企业的“互联网+”指数。这种方法的一个明显不足在于词汇的选取和打分的方法较为主观，容易受到打分专家认知程度的影响。戚聿东和蔡呈伟（2020）首次通过使用文本挖掘的方法构造企业数字化指标，认为与数字化相关的关键词出现的次数较多，说明该公司对这一关键词所代表的事物就越发重视，并用该指标研究数字化对企业资产收益率的影响。邱子迅和周亚虹（2021）在前者的基础上，使用中文 wiki 训练 Word2Vec 模型，构造人工智能词典，并对企业年报进行文本挖掘，用词典内关键词占企业年报总词数的比值衡量企业人工智能化水平，考察国家级大数据试验区的建立是否促进了区域内企业智能化意识的提高，进而增加企业的研发投入。本文参考后者的方式，使用文本挖掘的方法构造本文的核心解释变量——制造业企业数字化指数。

1、词典的构造

在进行文本挖掘之前，我们首先使用 Word2Vec 技术构造与数字化相关的词典。一方面，人的认知有限，无法通过手动整理的方式获得更多更准确的与数字化相关的目标词汇；另一方面，手动整理词汇的工作量非常庞大，且耗费时间较长。Mikolov et al.（2013）提出

的 Word2Vec 技术完美地弥补了手动整理词典的不足，其主要思想是通过向量代表各个词语，然后通过神经网络模型，在大量文本语料的基础上对向量参数进行训练（沈艳等，2019）。训练后的模型可以根据不同词汇的向量距离度量词汇之间的相似度。考虑到中文 wiki 内容量庞大且具有较强的专业性，因此本文使用中文 wiki 作为语料训练 Word2Vec 模型，进而保证词典构造的准确性。模型训练完毕后，分别获得与“数字化”和“人工智能”词向量相关系数最高的 50 个词汇，将两组词汇去重后得到本文的数字化词典。

限于篇幅，表 1 只展示 “数字化”和“人工智能”词向量模型返回的前十五大近义词和相关系数。从中我们可以发现，输出的词汇中不仅包括诸如“数位化”、“人机交互”、“系统集成”等难以靠人工获得的近义词，更不乏具有极强专业色彩的词汇。“地理信息系统”指利用数字化技术对整个或部分地球空间中的地理数据进行采集和分析的系统，同样是数字化的重要体现。“专家系统”是一个能够模拟人类决策的过程的系统，该系统具有大量专门知识与经验，它通过人工智能技术，根据现有知识进行推理和判断，是人工智能技术重要的应用领域。“虚拟现实”是一种利用现实生活数据通过计算机技术使其与输出设备结合，并使用户沉浸到模拟环境中的仿真技术。这些词汇应该是其他研究者难以预先加入词典的。

表 1 “数字化”与“人工智能”的前十五大近义词

“数字化”前十五大近义词		“人工智能”前十五大近义词	
词汇	相关系数	词汇	相关系数
数位化	0.704	人工智慧	0.807
数码化	0.701	AI	0.762
网络化	0.683	机器人学	0.676
电子化	0.681	专家系统	0.659
计算机化	0.668	人机交互	0.647
数据处理	0.660	智能	0.640
网络系统	0.654	计算能力	0.635
智能化	0.654	计算机	0.634
计算机技术	0.647	虚拟现实	0.625
数码影像	0.646	编程	0.623
软硬件	0.646	电脑程式	0.617
地理信息系统	0.640	人工神经网络	0.612

“数字化”前十五大近义词		“人工智能”前十五大近义词	
词汇	相关系数	词汇	相关系数
系统集成	0.638	计算机系统	0.610
图像处理	0.637	计算机程序	0.608
交互式	0.629	自然语言	0.608

2、文本挖掘与核心解释变量构建

通过以上方法获得词典后，本文参考王克敏等（2018）的方法对上市制造业公司 2013-2017 年年报进行清洗，然后使用 Python 软件包“jieba”进行分词，对每年每个公司年报中词典词汇出现的频次和总词汇数进行统计，定义本文核心解释变量数字化指数ai为公司年报每千词中数字化相关词汇出现的次数，即：

$$\mathrm{ai}_{\mathrm{it}} = \frac{\text{数字化词典中的词汇在 i 公司的 t 年年报中出现的次数}}{\text{i 公司的 t 年年报的总词汇数}} \times 1000$$

3、被解释变量与中介变量

以往的研究大多以企业的业绩指标作为企业绩效的衡量，如企业产出、净资产收益率、总资产收益率等（杨德明和刘泳文，2018；戚聿东和蔡呈伟，2020），但是这些指标只能衡量企业的产出效果，并不能代表企业的创新水平。同样有很多研究将企业的研发投入或研发投入占比作为企业创新的代理变量（邱子迅和周亚虹，2021），但实际的创新活动往往具有高风险特征，研发投入转化为创新产出存在较大难度，因此使用研发投入作为核心被解释变量有可能高估企业的创新能力。因此，本文参考唐松等（2020）的因变量选择，使用企业专利申请数的对数（lnpat）作为被解释变量。但这并不意味着本研究忽视了研发投入的作用。大量文献已经证明研发投入对专利产出有重要影响（Pakes 和 Griliches，1980），因此本文在考察数字化能否促进企业创新的同时，还侧重于考察不同特征企业在推进数字化的过程中其研发投入的转化作用。根据研究假说，本文认为研发投入起到重要的中介作用，故选取企业研发投入的对数（lnrd）作为中介变量。

4、其他控制变量

为了尽可能克服遗漏变量带来的影响，参考现有文献，本文将多个企业层面的特征纳入控制变量：企业规模（size），即企业总资产数额的对数；高管工资（lnwage），即董事、监事及高管前三名薪酬总额的对数；资金约束（kz），即 kz-score，计算参考李茫茫和黎文靖

（2017），该指数越高意味着企业面临的资金约束越强；成长性（grow），即企业营收增长率；两职合一（two），即董事长与总经理兼任，如果兼任则赋值为 1，否则为 0；政府补贴（sub），即企业收到政府补贴金额的对数；股权性质（state），即如果公司的股权性质为国有，则赋值为 1，否则为 0。详细描述性统计参见表 2。此外，为了研究数字化对企业创新影响的异质性，我们还纳入了非时变的高科技行业变量（hightech），即如果制造业企业的子行业位于《国家重点支持的高新技术领域》名录中，则赋值为 1，否则为 0。

表 2 变量的描述性统计

符号	赋值	观测数	均值	标准差
lnpat	企业申请专利数的对数	4600	3.261	1.424
lnrd	企业研发投入的对数	4600	17.944	1.286
ai	企业数字化指数，构建方法如前文所述	4600	0.660	0.966
size	企业总资产的对数	4600	8.248	1.065
lnwage	董事、监事及高管前三名薪酬总额的对数	4600	14.333	0.844
kz	kz-score，计算参考黎文靖和李茫茫（2017）	4600	1.915	5.186
grow	企业营收增长率	4600	0.292	1.078
two	董事长与总经理兼任则赋值为 1，否则为 0	4600	0.283	0.451
sub	企业收到政府补贴金额的对数	4600	0.013	0.028
state	股权性质为国有则赋值为 1，否则为 0	4600	0.312	0.463
hightech	所属行业为高科技行业则赋值为 1，否则为 0	4600	0.409	0.492

（三）模型设定

1、主模型

考虑到样本具有面板数据结构，本文构建如下实证模型：

$$\mathrm{lnpat_{ikt}} = \beta_0 + \beta_1 ai_{ikt} + \beta_2 othercontrols_{ikt} + \alpha_i + \delta_t + \gamma_{kt} + \epsilon_{ikt} \quad (1)$$

其中，下标i代表制造业企业，下标k代表制造业内具体细分行业，下标t代表时间。被解释变量为企业专利申请lnpat，核心解释变量为企业的数字化指数ai，β_1为我们感兴趣的系数，othercontrols则包含了前述控制变量。由于企业的专利申请会受到某些不随时间变化但是又可能与企业数字化相关的不可观测特征，如果直接使用 OLS 对系数进行估计，会导致系数出现偏误，因此，本文利用面板数据结构，控制了企业固定效应α_i。为了剔除时间趋势和行

业发展的影响，我们控制了时间固定效应δ_t和行业-时间效应γ_{kt}。ϵ_{ikt}为残差项。在对以上方程进行估计时，所有标准误均采用聚类标准误，聚类到企业层面。

2、中介效应

参考温忠麟和叶宝娟（2014）对中介效应的相关介绍，本文建立以下模型研究研发投入在数字化促进制造业创新中承担的作用，以进一步探索数字化创新的作用机制。

$$\mathrm{lnrd_{ikt}} = \mu_0 + \mu_1 ai_{ikt} + \mu_2 othercontrols_{ikt} + \alpha_i + \delta_t + \gamma_{kt} + \epsilon_{ikt} \quad (2)$$

$$\mathrm{lnpat_{ikt}} = \theta_0 + \theta_1 ai_{ikt} + \theta_2 lnrd_{ikt} + \theta_3 othercontrols_{ikt} + \alpha_i + \delta_t + \gamma_{kt} + \epsilon_{ikt} \quad (3)$$

本研究在获得主模型，即式（1）的估计结果后，对方程（2）和（3）进行估计。如果β_1, μ_1, θ_2同时显著大于零，则说明数字化进程显著提高了企业的研发投入，进而显著增加了企业的专利产出。进一步地，如果θ_1显著大于零，则说明研发投入是数字化促进企业创新的部分中介，否则研发投入是一个完全中介。

3、异质性分析

以上模型均只研究了解释变量对被解释变量的平均作用，忽略了不同特征的个体可能存在相异的作用。本研究建立以下模型考察数字化促进制造业企业创新的异质性作用。

$$\mathrm{lnpat_{ikt}} = \beta_0 + \beta_1 ai_{ikt} + \beta_2 co_{ikt} + \beta_3 (ai \times co)_{\mathrm{ikt}} + \beta_4 othercontrols_{ikt} + \alpha_i + \delta_t + \gamma_{kt} + \epsilon_{ikt} \quad (4)$$

其中，co为我们感兴趣的可能影响到数字化对企业创新促进作用的企业特征，即协变量。核心解释变量数字化对被解释变量的偏效应为

$$\frac{\partial \mathrm{lnpat}}{\partial \mathrm{ai}} = \beta_1 + \beta_3 co$$

因此，在异质性分析中，我们感兴趣的参数为β_3，即协变量对解释变量对被解释变量偏效应的影响。

四、实证结果

（一）描述性分析

图 1 制造业企业数字化指数随时间的变化趋势

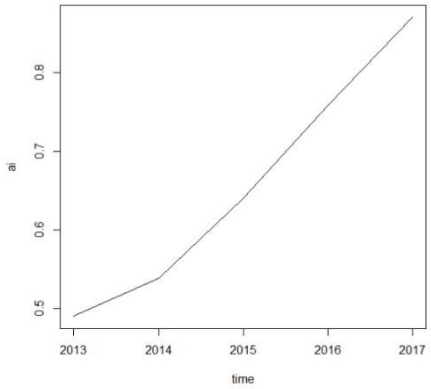

图 1 展示了制造业企业的数字化指数的平均水平随时间的变化趋势。可以发现，随着时间的推移，制造业企业的平均数字化水平在逐年增加。2013 年时，全制造业的平均数字化水平只有 0.4906，到 2018 年，该指数就增长到了 0.8702，由此可见中国制造业企业的数字化水平在迅速提高。2015 年 7 月国务院颁布了相关文件，目的是推动互联网与传统行业相互融合，利用互联网技术所具备的优势和特点，创造新的经济发展机会，推动产业经济的转型和发展。图 1 也显示，在 2015 年之后，中国制造业企业的数字化指数迅速升高，企业在主动利用数字技术改善自身的经营，力求实现创新发展。

图 2 制造业企业数字化与企业专利产出

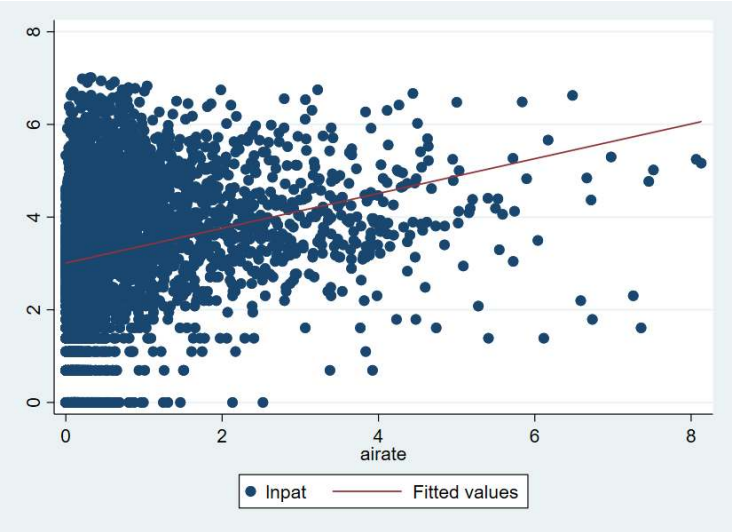

图 2 直观展示了制造业企业数字化水平和企业专利产出的关系。横坐标为本文构造的我国制造业企业的数字化指数，纵坐标为企业专利产出的对数。可以发现，二者呈现出一个显著的正相关关系，即数字化水平越高的企业，其专利产出越高。但是从图中我们仅能够获得一个相关关系，为了识别数字化和企业创新之间的因果关系，我们需要通过回归给出答案。

（二）数字化对制造业企业创新的影响

表 3 针对数字化与制造业企业创新的基准关系进行了实证检验。在模型（1）中，本文仅控制核心解释变量企业数字化和各种固定效应，结果发现，制造业企业的数字化水平的回归系数为正，且通过了 1%的统计显著性检验，说明制造业企业数字化水平的提高显著推动了企业的创新产出。在控制了相关控制变量后，如表 3 列（2）所示，上述结论依旧维持不变。这说明，在数字化水平提高后，企业的核心创新竞争力得以逐步加强。

表 3 数字化对制造业企业创新的影响

被解释变量	(1) lnpat	(2) lnpat
ai	0.0903*** (0.0320)	0.0625** (0.0280)
state		-0.1947 (0.2529)
size		0.5437*** (0.0651)
lnwage		0.0561** (0.0285)
kz		-0.00102 (0.000885)
grow		0.0193** (0.00777)
two		0.0771 (0.0546)
sub		-0.320 (0.931)
固定效应	企业、时间、行业-时间效应	企业、时间、行业-时间效应
常数项	3.201*** (0.0211)	-2.030*** (0.696)
NT	4600	4600

注：上标三颗星、两颗星、一颗星分别代表在 1%、5%、10%的显著性水平下显著，括号内为聚类到企业层面的聚类标准误。下文皆同，不再赘述。

（三）稳健性检验

本文从三个主要方面对主要结论的稳健性进行检验：一是回归模型的更换，二是采用工具变量的方法，试图解决可能存在的内生性问题，三是使用机器学习的方法对特征进行筛选，

以检验本文的核心自变量是否对特征筛选过程稳健。

首先，本研究的被解释变量为专利产出的对数，整体为零值堆积与正值连续分布共存的混合特征，因此本文采用了更适合这种数据结构的 TOBIT 模型进行检验。本文在进行 TOBIT 回归时，加入时间虚拟变量和行业虚拟变量力求对固定效应的控制，回归结果如表 4 模型（1）所示，数字化水平的提高有助于企业创新力的提升这一结论依然成立。

其次，我们对基准模型的估计可能受到潜在的内生性的影响。如果一个企业希望提升创新能力，或者希望增加其专利产出，则该企业有更强的动机推行数字化，因此可能存在反向因果问题。我们使用同行业其他企业平均的数字化指数作为ai的工具变量。一方面，制造业企业可能会模仿同行业其他企业的数字化进程，因此满足工具变量的相关性；另一方面，同行业其他企业的数字化进程不会直接影响该企业的专利产出，满足工具变量的外生性。工具变量回归结果如表 4 模型（2）所示，一阶段回归结果显示，我们选取的工具变量与变量ai高度相关，且一阶段回归 F 值远大于弱工具变量判定法则的临界值 10，因此认为我们选取的是合格的工具变量。且表 4 中工具变量回归的结果也支持了本文主模型中得到的结论。

最后，本研究使用 LASSO 回归对所有解释变量进行特征筛选，解的路径如表 5 所示。随着惩罚因子由大变小，我们对特征筛选的要求逐渐放松，各解释变量被逐步引入回归方程中。可以发现，继常数项之后，公司规模和数字化变量被引入了模型。只有当惩罚因子大于 3530.3 时，数字化变量才会被特征筛选过程剔除。因此本文认为，数字化是影响企业创新产出的不可忽略的重要因素，其作用也对特征筛选过程稳健。

表 4 稳健性检验

稳健性检验	(1)更换 TOBIT 模型	(2)工具变量回归	
被解释变量	lnpat	lnpat	一阶段工具变量回归系数 0.887*** (0.0201)
ai	0.217*** (0.0210)	0.0625** (0.0280)	
其他控制变量	控制	控制	一阶段回归 F 值 314
固定效应	时间虚拟变量、行业虚拟变量	企业、时间、行业-时间趋势	
NT	4600	4600	

表 5 LASSO 解的路径

节点数	惩罚因子	L1-Norm	EBIC	拟合优度	操作
1	6169.31046	0.00000	3408.85403	0.0000	加入 常数项
2	5621.24592	0.05427	3248.10994	0.0350	加入变量 size
3	3530.30626	0.26621	2707.35459	0.1404	加入变量 ai
4	1677.18149	0.67696	2061.04899	0.2511	加入变量 lnwage
5	344.91444	1.05625	1826.51332	0.2884	加入变量 two
6	216.61632	1.12733	1825.00850	0.2899	加入变量 grow
7	102.91030	1.20071	1827.19175	0.2908	加入变量 kz
8	93.76803	1.20689	1835.28804	0.2909	加入变量 state
9	85.43794	1.22297	1834.3723	0.2910	加入变量 sub

（四）进一步研究：数字化、研发投入与企业创新

本文选取企业的专利产出作为被解释变量，相较于研发投入，专利产出更能代表企业的创新能力和创新产出。但是在考察企业的专利产出时，研发投入也是不可忽视的一个因素。在这部分研究中，本文考察研发投入在数字化促进企业创新中的角色。本研究参考温忠麟和叶宝娟（2014）对中介效应的研究框架，首先对主模型进行回归（如表 6 模型（1）所示，回归结果同表 3 模型（2））。其次，我们以研发投入作为被解释变量进行回归，结果如表 6 模型（2）所示，发现企业数字化显著提升了企业的研发水平。最后，我们将研发投入进行控制，再次考察数字化对企业专利产出的作用，发现数字化和研发投入均对专利产出有显著作用，因此本文认为研发投入是数字化促进企业创新产出的中介。

表 6 研发投入在数字化促进企业创新中的中介作用

被解释变量	(1) lnpat	(2) lnrd	(3) lnpat
ai	0.0625** (0.0280)	0.0344** (0.0167)	0.0582** (0.0278)
lnrd			0.123*** (0.0344)
其他控制变量	控制	控制	控制
固定效应	企业、时间、行业-时间效应	企业、时间、行业-时间效应	企业、时间、行业-时间效应
NT	4600	4600	4600

（五）异质性分析

在前面的分析中，本研究发现，制造业企业数字化水平的提高显著促进了企业的专利产出，且该结论对更换回归模型、使用工具变量、进行特征筛选均表现稳健。具体的机制是，企业推进数字化后，加大了研发投入，进而提高了专利产出。这些都只是在平均意义上获得的结论，我们更需要知道，为什么现实中，不同企业在推进数字化的过程中会产生不同的表现，即我们更需要关注数字化促进企业创新的异质性作用，这对于后疫情时代中国制造业企业转型升级和政策制定具有重要意义。本研究分别从所属行业、高管工资、资金约束、企业规模四个角度来考察数字化促进制造业企业创新的异质性作用。

首先，创新的不确定性强、失败率高、耗费时间长，需要大量资金支持（Hottenrott 和 Peters，2012），因此较强的资金约束不利于企业的创新产出。表 7 中，本研究在基准模型的基础上加入了企业数字化与企业资金约束的代理 kz-score 的交互项（交互项各项均已进行去均值处理），该交互项的系数表示企业数字化对企业创新的促进作用随不同 kz-score 的变化趋势，即 kz-score 对数字化推进企业创新的调节作用。回归结果显示，交互项显著，即较强的资金约束会削弱数字化促进企业创新的作用。

表 7 异质性分析：企业资金约束

被解释变量	(1) lnpat	(2) lnrd	(3) lnpat
ai	0.0583** (0.0274)	0.0344** (0.0164)	0.0541** (0.0273)
kz	-0.000259 (0.000930)	-0.000384 (0.000431)	-0.000212 (0.000905)
ai*kz	-0.0745* (0.0424)	0.000986 (0.0251)	-0.0746* (0.0424)
lnrd			0.123*** (0.0343)
其他控制变量	控制	控制	控制
固定效应	企业、时间、行业-时间效应	企业、时间、行业-时间效应	企业、时间、行业-时间效应
NT	4600	4600	4600

其次，本研究考察企业规模对数字化创新效果的影响。本文同样选择加入数字化与企业规模的交互项进行研究，表 8 展示了回归结果。实证结果证明了，相比于大型企业，小型企业数字化能带来更多的专利产出。本研究的结果符合谢伏瞻（2019）给出的论断：由于新技

术的技术范式和技术路径并不清晰，不同类型的创新主体，特别是初创小型企业在新技术可能带来巨大潜在利益的驱动下，通常会积极进行数字化和多元化的商业模式探索。

表 8 异质性分析：企业规模

被解释变量	(1) lnpat	(2) lnrd	(3) lnpat
ai	0.0606** (0.0268)	0.0341** (0.0163)	0.0564** (0.0267)
size	0.540*** (0.0646)	0.769*** (0.0580)	0.446*** (0.0670)
ai*size	-0.150* (0.0766)	-0.0240 (0.0658)	-0.147** (0.0735)
lnrd			0.123*** (0.0340)
其他控制变量	控制	控制	控制
固定效应	企业、时间、行业-时间效应	企业、时间、行业-时间效应	企业、时间、行业-时间效应
NT	4600	4600	4600

再次，本研究通过实证检验企业股权性质对制造业企业数字化创新的影响。表 9 在基准回归的基础上加入了数字化与企业股权性质的交互项，估计结果显示交互项对企业创新产出的回归系数显著大于零，即相比于其他企业，国有企业数字化创新的效果更佳，融资能力的正向调节作用起到主导。

表 9 异质性分析：企业股权性质

被解释变量	(1) lnpat	(2) lnrd	(3) lnpat
ai	0.0524* (0.0272)	0.0290* (0.0158)	0.0489* (0.0272)
state	-0.0538 (0.247)	0.0278 (0.112)	-0.0572 (0.243)
ai*state	0.576*** (0.109)	0.308*** (0.0814)	0.539*** (0.111)
lnrd			0.121*** (0.0338)
其他控制变量	控制	控制	控制
固定效应	企业、时间、行业-时间效应	企业、时间、行业-时间效应	企业、时间、行业-时间效应
NT	4600	4600	4600

最后，企业数字化转型是一个庞大复杂的工程，特别是传统行业企业相比于新兴行业企业，资金投入需求更高。如果传统企业的数字化进程能够有效提升其创新能力，那么传统行业在数字经济时代面临的机遇大于挑战，数字化也会为传统行业孕育新的生产力，成为经济高质量发展的重要驱动力。本研究将所有样本分为高科技行业企业（$\mathrm{hightech} = 1$）和传统行业企业（$\mathrm{hightech} = 0$）两组，分样本进行回归，结果如表 10 所示。本文发现，高科技行业企业的数字化与企业创新产出的关系并不显著，企业创新主要由企业的研发投入主导。值得注意的是，传统行业企业的数字化过程促进了企业的专利产出，且数字化显著提升了企业的研发投入水平。综上，本部分的研究结论说明了数字经济时代，传统企业可以借力数字化发展实现创新水平的提升，挖掘新的生产力。

表 10 异质性分析：高科技制造业企业与传统制造业企业

分组	高科技行业（$\mathrm{hightech} = 1$）			传统行业（$\mathrm{hightech} = 0$）		
被解释变量	(1) lnpat	(2) lnrd	(3) lnpat	(1) lnpat	(2) lnrd	(3) lnpat
ai	0.0395 (0.0373)	0.0214 (0.0213)	0.0352 (0.0371)	0.0976** (0.0412)	0.0518** (0.0257)	0.0924** (0.0410)
lnrd			0.200** (0.0791)			0.101*** (0.0324)
其他控制变量	控制	控制	控制	控制	控制	控制
固定效应	企业、时间、行业-时间效应	企业、时间、行业-时间效应	企业、时间、行业-时间效应	企业、时间、行业-时间效应	企业、时间、行业-时间效应	企业、时间、行业-时间效应
NT	1880	1880	1880	2720	2720	2720

五、政策建议

本文通过对 2013-2017 年中国 A 股上市公司企业年报文本进行挖掘，构建了制造业企业的数字化指数，并将其与企业财务指标和专利数据匹配，构建平衡面板数据，考察制造业企业数字化能否促进企业创新。回归结果显示，制造业企业数字化显著促进了企业创新，且本结论对更换回归模型、使用工具变量、进行特征筛选均稳健。同时，本研究发现，研发投入是数字化促进企业创新的重要中介。最后，本文研究了数字化促进企业创新的异质性作用，发现资金约束较宽松的企业、规模较小的企业、国有企业和传统行业企业更加受益于数字化过程。

基于本文的研究，提出如下建议：

（一）从制造业企业的角度来看，应积极推进数字化过程，特别是传统行业企业，大力采用大数据、云计算、人工智能等新技术，以便更好地实现互联网等高科技技术和实体经济的深度融合。制造业企业应结合自身特征，充分利用数字技术，获取更多有价值的信息，降低生产成本和销售成本，推动自身技术创新。只有这样，企业才能生产出更多满足不同需求的差异化产品，在数字经济时代获得行业内的优势地位。

（二）从政府的角度看，应通过一些制度层面的措施，鼓励制造业采用新技术，降低使用新技术的初期成本和投入，切实缓解制造业企业因对新技术投入而产生的融资难、融资贵的问题。在数字经济时代，研发投入依然是企业提升创新力的重要中介，在当前原材料和融资成本上升的压力下，制造业企业的数字化转型面临巨大的资金困境，政府应适当地给予数字化转型中的制造业企业一定的补贴，并使金融机构和企业更好地对接，缓解企业研发投入不足的问题，促进企业数字化创新。同时，政府应充分利用快速发展的信息通信技术以及不断迭代更新的互联网平台，及区块链和人工智能，让这些高科技技术与传统行业、产业进行深度、紧密融合，特别推动产业互联网的建立和发展，加速制造业的数字化转型，创造新的经济发展生态，将数字科技创新成果不断深度融合于经济发展、社会管理各领域之中，提升整个中国社会的创新力和生产力。

参考文献

【英文文献】:

[1] Bharadwaj, A. S., Bharadwaj, S. G. and Konsynski, B. R., 1999, “Information Technology Effects on Firm Performance as Measured by Tobin’s Q”, *Management Science*, Vol.45.

[2] Brynjolfsson, E. and Lorin Hitt, M., 2000, “Beyond Computation: Information Technology, Organizational Transformation and Business Performance”, *Journal of Economic Perspectives*, Vol.14.

[3] Brynjolfsson, E., McElheran, K., 2016, “Digitization and Innovation: The Rapid Adoption of Data-Driven Decision-Making”, *American Economic Review* ,Vol.106.

[4] Brynjolfsson, E., 1993, “The Productivity Paradox of Information Technology”, *Communications of the ACM*, Vol.36.

[5] Hajli, M., Sims, J. M., Ibragimov, V., 2015, “Information Technology (IT) Productivity Paradox in the 21St Century”, *International Journal of Productivity and Performance Management*, Vol.64.

[6] Hottenrott, H. and Peters, B., 2012, “Innovative Capability and Financing Constraints for Innovation: More Money, More Innovation?”, *Review of Economics and Statistics*, Vol.94.

[7] Lane, P. J., Salk, J. E. and Lyles, M. A., 2001, “Absorptive Capacity, Learning, and Performance in International Joint Ventures”, Strategic Management Journal, Vol.22.

[8] Lee, S. T., Gholami, R. and Tong, T. Y., 2005, “Time Series Analysis in the Assessment of ICT Impact at the Aggregate Level Lessons and Implications for the New Economy”, *Information & Management*, Vol.35.

[9] Mikolov, T., I.Sutskever, K.Chen, G. S. Corrrado, and J. Dean, 2013, “Distributed Representations of Words and Phrases and their Compositionality”, *Advances in Neural Information Processing Systems*, Vol.26.

[10] Pakes, A., and Z., Griliches, 1980, “Patents and R and D at the Firm Level: A First Look”, *NBER working paper*, w0561.

[11] Shleifer, A., 1998, “State versus Private Ownership”, *Journal of Economic Perspectives*, Vol.12.

[12] Solow, R. M., 1987, “We’d better watch out”, *New York Times Book Review*.

[13] Wu, L., Lorin Hitt, Bowen Lou, 2019, “Data Analytics, Innovation, and Firm Productivity”, *Management Science*.

【中文文献】:

[14] 郭家堂、骆品亮，2016:《互联网对中国全要素生产率有促进作用吗？》,《管理世界》, 第 10 期。

[15] 李海舰、田跃新、李文杰，2014:《互联网思维与传统企业再造》,《中国工业经济》，第 10 期。

[16] 李茫茫、黎文靖，2017:《审计具有保险功能吗——基于政府官员变更的自然实验》,《南开管理评论》，第 4 期。

[17] 刘政、姚雨秀、张国胜、匡慧姝，2020:《企业数字化、专用知识与组织授权》,《中国工业经济》，第 9 期。

[18] 戚聿东、蔡呈伟，2020:《数字化对制造业企业绩效的多重影响及其机理研究》,《学习与探索》，第 7 期。

[19] 邱子迅、周亚虹，2021:《数字经济发展与地区全要素生产率——基于国家级大数据综合试验区的分析》,《财经研究》，第 7 期。

[20] 沈艳、陈赟、黄卓，2019:《文本大数据分析在经济学和金融学中的应用：一个文献综述》,《经济学（季刊)》，第 7 期。

[21] 唐松、伍旭川、祝佳，2020:《数字金融与企业技术创新——结构特征、机制识别与金融监管下的效应差异》,《管理世界》，第 5 期。

[22] 王克敏、王华杰、李栋栋、戴杏云，2018:《年报文本信息复杂性与管理者自利——来自中国上市公司的证据》,《管理世界》，第 12 期。

[23] 王宇、王铁男、易希薇，2020:《R&D 投入对 IT 投资的协同效应研究——基于一个内部组织特征的情境视角》,《管理世界》，第 7 期。

[24] 王玉泽、罗能生、刘文彬，2019,《什么样的杠杆率有利于企业创新》,《中国工业经济》, 第 3 期。

[25] 温忠麟、叶宝娟，2014:《中介效应分析:方法和模型发展》,《心理科学进展》，第 5 期。

[26] 谢伏瞻，2019:《论新工业革命加速拓展与全球治理变革方向》,《经济研究》，第 7 期。

[27] 谢平、邹传伟、刘海二，2015:《互联网金融的基础理论》,《金融研究》第 8 期。

[28] 杨德明、刘泳文，2018:《“互联网+”为什么加出了业绩》,《中国工业经济》，第 5 期。

[29] 杨德明、陆明，2018:《互联网商业模式会影响上市公司审计费用么?》,《审计研究》，第 6 期。

[30] 余明桂、钟慧洁、范蕊，2019:《民营化、融资约束与企业创新——来自中国工业企业的证据》,《金融研究》，第 4 期。

[31] 张三峰、魏下海，2019:《信息与通信技术是否降低了企业能源消耗——来自中国制造业企业调查数据的证据》,《中国工业经济》，第 2 期。

[32] 张勇进、王璟璇，2014:《主要发达国家大数据政策比较研究》,《中国行政管理》，第 12 期。

致谢

1. 论文的选题来源，研究背景：

随着互联网基础设施的完善，作为通用技术的大数据技术和作为基础技术的云计算和人工智能及区块链技术实现快速发展，数字产业加速扩张，已经成为中国新的经济增长点。全球各主要工业化国家相继推出大数据等相应政策，以推动数字经济的发展。中国也在推进制造业数字化，开展数字化转型，只有这样才能形成叠加效应、聚合效应、倍增效应，加速新旧发展动能和生产体系的转换。因此，在新一轮的工业革命窗口期，中国制造业企业能否借力数字化实现创新发展、高质量发展，事关中国能否成功实现新旧动能转换、抢占数字时代的发展先机。然而，相关领域的文献依然存在部分缺陷。所以在数字经济的大环境中，我们非常有必要厘清企业数字化和企业创新的关系，并讨论这其中可能存在的提升空间，这对中国的政策推行和企业战略制定具有重要意义。

2. 每一个队员在论文撰写中承担的工作以及贡献：

文章由作者在指导老师樊教授的指导下独立完成。在选题的过程中，作者对企业创新以及数字化的时代背景进行了充分的了解，然后深入阅读了大量与企业绩效、企业创新、地区创新和数字化相关的文献，进一步分析现有文献的不足和缺陷，将文献分类总结，据此发展新的研究话题。在实证过程中，作者收集和整理了中国上市制造业企业公司的各项财务数据和专利数据，并借助爬虫数据进行匹配，构建了平衡面板数据。作者建立实证模型并独立完成了本文的基础回归分析工作，并对主要结论进行了稳健性检验。在此基础上，通过与指导老师的探讨，进一步挖掘具有哪种特征的企业更受益于数字化过程，即进行异质性分析。最后，作者通过实证分析获得的结论，给出了相关的政策建议。

3. 指导老师与学生的关系，在论文写作过程中所起的作用，及指导是否有偿

在本文的研究过程中，作者感谢复旦大学经济学院樊教授无偿的辛苦付出。樊教授对于相关文献和理论的耐心指导使作者受益匪浅，在此特别感谢他提供的文本挖掘方面的技术指导。尤其是从课题的选定，以及对于课题的理解及启发，直至后续深入的研究都给予了专业的指导。在作者撰写论文过程中，提供相关文献使作者更全面的进行详细分析和深度思考，对于遇到的各种问题樊教授都进行了非常详实的解答和诠释。通过作者的努力和樊教授的专业指导，使得本篇论文能够顺利完成。 在此感谢樊教授给予的指导意见。