

一、正文的补充说明

(一) 带控制变量的基准回归结果

附表 1 大数据试验区与企业数字创新质量			
	(1) <i>DigPatCite</i>	(2) <i>DigPatCite</i>	(3) <i>DigPatCite</i>
<i>BigData</i>	0.1355*** (0.0417)	0.1021*** (0.0385)	0.1001*** (0.0383)
<i>Age</i>		1.1273*** (0.1612)	1.1161*** (0.1608)
<i>Size</i>		0.0876*** (0.0183)	0.0855*** (0.0182)
<i>Lev</i>		0.2268*** (0.0597)	0.2175*** (0.0598)
<i>ROA</i>		0.0371 (0.0722)	0.0488 (0.0715)
<i>Growth</i>		0.0117 (0.0081)	0.0135* (0.0081)
<i>Fixratio</i>		0.2453*** (0.0781)	0.2429*** (0.0781)
<i>TOP10</i>		-0.0099*** (0.0009)	-0.0099*** (0.0009)
<i>Indenp</i>		-0.0001 (0.0017)	-0.0002 (0.0016)
<i>Mshare</i>		-0.0035*** (0.0010)	-0.0035*** (0.0010)
<i>SOE</i>		0.0628 (0.0441)	0.0593 (0.0443)
<i>GDP</i>			0.0176 (0.0436)
<i>Ind</i>			-0.0008 (0.0020)
<i>Fiscal</i>			0.0939*** (0.0225)
企业固定效应	Yes	Yes	Yes
年份固定效应	Yes	Yes	Yes
观测值	32184	32184	32184
调整 R ²	0.9227	0.9294	0.9297

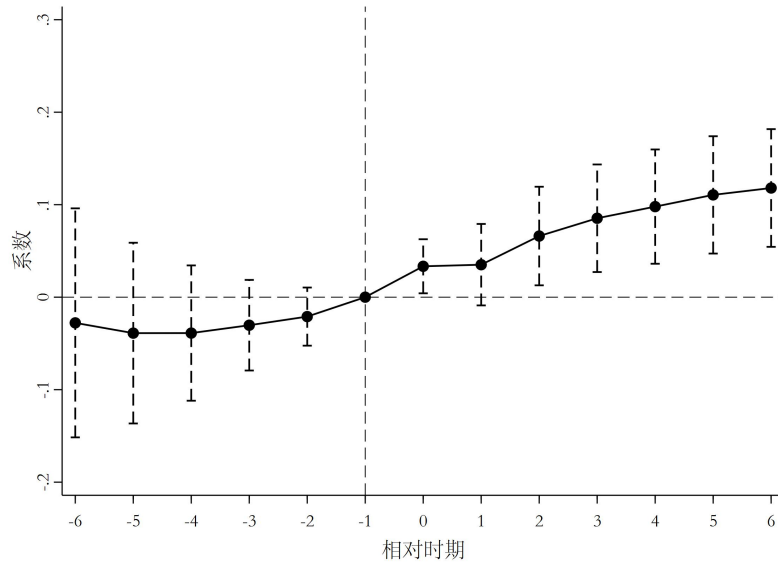
注：括号内为聚类稳健标准误；*、**和***分别表示在 10%、5%和 1%的统计水平上显著。以下各表同。

(二) 异质性分析回归结果

附表 2 异质性检验						
被解释变量 <i>DigPatCite</i>	行业竞争程度		数字化转型程度		所有权性质	
	(1) 高	(2) 低	(3) 高	(4) 低	(5) 国有	(6) 非国有
<i>BigData</i>	0.1390*** (0.0514)	0.0333 (0.0485)	0.0551 (0.0361)	0.1055*** (0.0334)	-0.0268 (0.0570)	0.1967*** (0.0520)
控制变量	Yes	Yes	Yes	Yes	Yes	Yes
企业固定效应	Yes	Yes	Yes	Yes	Yes	Yes
年份固定效应	Yes	Yes	Yes	Yes	Yes	Yes
观测值	15809	15995	12712	14409	11634	20350
调整 R ²	0.9554	0.9131	0.8972	0.9731	0.9392	0.9273

(三) 稳健性检验

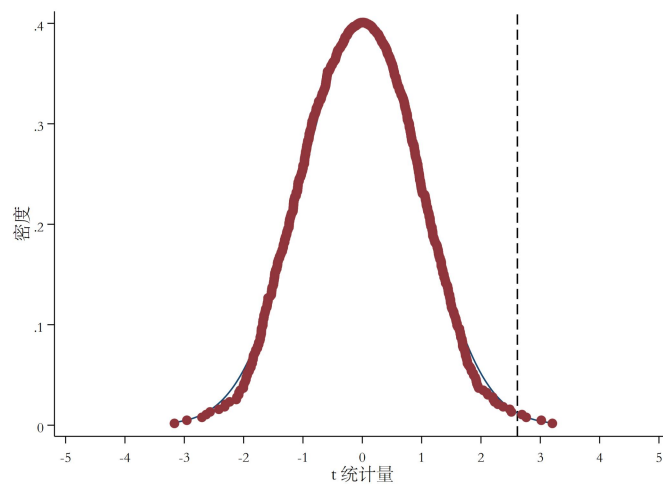
1. 动态效应检验。



附图 1 大数据试验区设立对企业数字创新质量影响的动态趋势图

2. 安慰剂检验。

前文虽然从双重差分法应用前提假设的角度论证了本文基本结论的稳健性,但还无法排除是否存在一些偶然随机因素驱动了处理组和对照组在数字创新质量上的系统性差异。为此,本文按照大数据试验区的设立时间和城市数量,采用随机数生成相对应的虚拟大数据试验区变量,并代入模型(1)中替换真实的大数据试验区变量进行回归分析。本文利用自助法(bootstrap)对上述步骤重复 1000 次,并绘制了伪大数据试验区变量回归系数 β_1 的 t 值核密度图(见附图 2),其中右边的虚线为基准回归表 2 第 2 列中 *BigData* 系数的 t 值。本文发现模拟所得绝大多数回归系数的 t 值都分布在零值附近且不显著,只有少量的系数值分布在虚线右侧。由此可推测,本文基本结果确实是设立大数据试验区所产生的,而非其他偶然因素。



附图 2 安慰剂检验

3.控制样本选择性偏误。

考虑到大数据试验区政策并非严格意义上的自然实验,因此本文的研究样本可能存在选择性偏误(Selection Bias)的问题。为此,本文使用倾向得分匹配—双重差分模型(PSM-DID)模型解决这一问题。具体步骤如下:①将企业层面的控制变量设定为匹配变量。②按照 1:1 近邻匹配原则,将企业样本进行逐年匹配,然后按照年份排序进行纵向合并,得到回归需要的面板数据。③对得到的面板数据进行平衡性检验后发现,重新匹配后的处理组和对照组之间变量的均值差异都在 10%的水平上不显著,表明本文的匹配结果是平衡的。④重新使用 DID 方法再次估计大数据试验区政策对企业数字创新质量的影响。基于 PSM 方法的回归结果如附表 3 列(1)所示,变量 *BigData* 的系数仍然在 1%的水平上显著,且系数值相比基本结果并未有明显差异,说明本文的样本受选择性偏误的影响较小。

4.剔除数字经济企业样本。

从行业角度看,数字经济行业的企业可能对数字创新的重视程度更高,且自身的数字创新能力也更强。因此,在回归中包含这部分企业样本可能会高估本文的估计结果。为此,本文将行业类别属于信息传输或软件和信息技术服务业或计算机、通信和其他电子设备制造业认定为数字经济企业,并将这部分企业样本删除。附表 3 列(2)的结果显示,在剔除数字经济企业样本后,变量 *BigData* 的系数为 0.0697,相比基准结果有所下降,但仍然在 1%的统计水平上显著,验证了本文基准结果的稳健性。

5.更换企业数字创新质量的测度方法。

本文重新采用如下指标来测度企业数字创新质量:①专利范围(Patent Scope),指企业当年专利国际分类号中所有二级专利分类号的数量。如果企业某项数字专利的二级专利分类号越多,说明企业在数字专利创造的过程中进入了多个知识领域(Ahuja 和 Lampert, 2001),并创造了新的知识,创新质量较高。②参考张杰和郑文平(2018)的方法,使用专利知识宽度法测度数字创新质量。专利知识宽度能够反映出专利所涉及的识复杂性和广泛性,该指标越大,意味着企业数字专利涉及的识领域越广,创新质量越高。附表 3 列(4)~(5)分别报告了大数据试验区对企业数字专利范围和专利知识宽度的回归结果,可以看出,不论使用何种指标替换被解释变量,*BigData* 的估计系数都至少在 5%的水平上显著为正,说明大数据试验区确实有助于提高企业的数字创新质量,本文基本结论十分稳健。

6.排除其他竞争性假说。

在本文的样本期内,国家为推动数字经济发展,出台了许多的试点政策,如“宽带中国”和“智慧城市”,甚至各地政府也陆续出台了“数字经济发展规划”,并将支持企业数字技术应用和创新发展纳入规划中。为排除上述事件冲击对本文基本结论的影响,本文分别生成了“宽带中国”、“智慧城市”和“重点产业”三个虚拟变量。若企业所在城市在第 t 年为宽带中国、智慧城市试点城市或在产业政策本文中明确数字产业为重点支持产业,则“宽带中国”、“智慧城市”和“重点产业”分别赋值为 1,否则为 0。随后将上述变量代入模型(1)中予以控制,回归结果如附表 3 列(6)所示。在考虑上述政策干扰的情况下,*BigData* 的回归系数依旧在 1%的水平上显著为正,且系数值相比本文的基准结果并无显著差异。这些结果表明其它与数字经济相关的政策冲击不会使本文基本结论产生本质上的变化。

附表 3		其他稳健性检验			
	PSM-DID	剔除数字经济 行业样本	数字专利 范围	数字专利 知识宽度	排除其他政策 干扰
	(1) <i>DigPatCite</i>	(2) <i>DigPatCite</i>	(4) <i>DigPatCite</i>	(5) <i>DigPatCite</i>	(6) <i>DigPatCite</i>
<i>BigData</i>	0.1002*** (0.0384)	0.0697*** (0.0160)	0.0341** (0.0159)	0.0165** (0.0070)	0.0975*** (0.0378)
智慧城市					0.0563 (0.0344)
宽带中国					0.1035** (0.0435)
是否为重点 支持产业					0.0270 (0.0930)
控制变量	Yes	Yes	Yes	Yes	Yes
企业固定效应	Yes	Yes	Yes	Yes	Yes
年份固定效应	Yes	Yes	Yes	Yes	Yes
观测值	32083	30045	32184	32184	32184
调整 R ²	0.9297	0.9282	0.7178	0.6740	0.9367

参考文献

[1] 张杰,郑文平.创新追赶战略抑制了中国专利质量么?[J].经济研究,2018,53(05):28-41.

[2] Ahuja, G., and C. M. Lampert. Entrepreneurship in the Large Corporation: A Longitudinal Study of How Established Firms Create Breakthrough Inventions [J] . Strategic Management Journal, 2001, 21(6-7): 267-294.