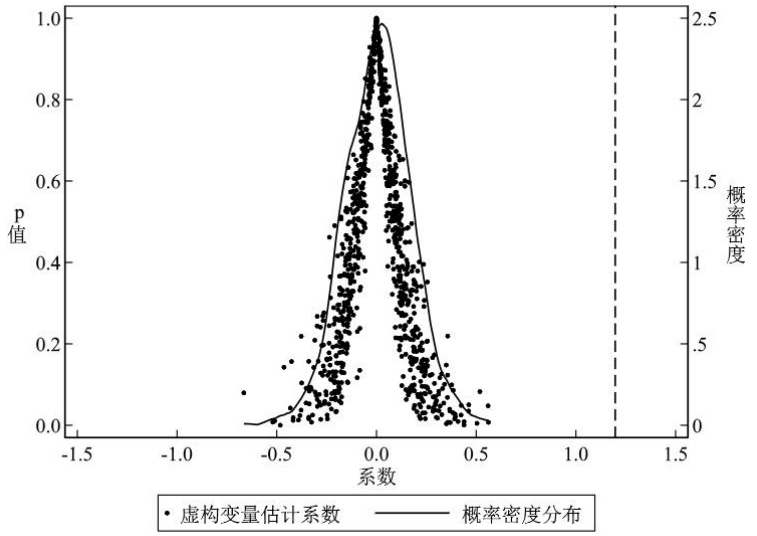
《数据共享与企业创新》附录

附录 1 安慰剂检验结果

本文通过置换检验排除不可观测因素对于基准结果的干扰，具体的做法是根据处理组数量随机抽取相应的样本并构造虚拟解释变量，附图 1 展示了重复 1000 次随机过程的估计系数与对应 *p* 值。可以看出，虚构变量的估计系数均远小于虚线代表的真实处理效应。



附图 1 安慰剂检验

附录 2 内生性处理结果

考虑到供应商特征可能与客户企业的创新水平相关，导致遗漏变量、反向因果等问题，本文通过控制供应商特征、工具变量回归进行处理。附表 1 第（1）列结果显示，控制供应商基本特征后，供应商数据共享的影响仍然显著为正。附表 1 第（2）、（3）列显示，工具变量（*sharing_rate*）一阶段估计结果符合预期，二阶段 *datasharing* 的估计系数仍然显著为正。同时，K-P *F* 统计量大于弱工具变量检验的 10%临界值 16.38，显著拒绝弱工具变量的原假设。以上结果表明在处理潜在的内生性问题后，基准结果保持稳健。

附表 1 内生性处理结果

变量	(1) *innovation*	(2) *datasharing*	(3) *innovation*
datasharing	1.1910**		1.2346**
	(2.0258)		(2.0325)
sharing_rate		1.2414***	
		(4.6942)	
控制变量	Yes	Yes	Yes
企业固定效应	Yes	Yes	Yes
行业—年份固定效应	Yes	Yes	Yes
样本量	6246	6405	6405
Adj. R^2	0.8773		
K-P *F* 统计量		22.040	

注：*、**和***分别表示 10%、5%和 1%的显著性水平，括号内为 *t* 值。下表同。

附录 3 数据特征的影响结果

附表 2 各列结果显示，一方面，数据共享与供应商数据量、传输速度的交互项均显著为正，表明用户数字足迹体量越大、传输速度越快，越有利于提升客户企业的创新数量。另一方面，数据多样性的交互项并不显著，供应商的数据多样性未能转化为客户企业的创新优势。

附表 2 数据特征的影响

变量	(1)	(2)	(3)
		innovation	
volume×*datasharing*	0.3134***		
	(3.0007)		
variety×*datasharing*		0.0638	
		(1.2276)	
velocity×*datasharing*			0.2242***
			(5.8189)
控制变量	Yes	Yes	Yes
企业固定效应	Yes	Yes	Yes
行业—年份固定效应	Yes	Yes	Yes
样本量	6405	6405	6405
Adj. R^2	0.8850	0.8793	0.9224

附录 4 考虑处理组过少问题的结果

（一）处理组样本特征

本文统计了 14 家曾接受数据共享的客户企业的行业分布，以及 23 条处理观测值的年份分布，结果如附表 3 所示。首先，处理组企业分布于制造业；批发和零售业；交通运输、仓储和邮政业；信息传输、软件和信息技术服务业；房地产业等多个行业门类，其中多数行业也高度依赖用户数据支撑营销决策与产品迭代。其次，处理观测值的数量在 2012—2020 年呈现先上升后回落企稳的变化趋势，这也与 2012 年印发的《“十二五”国家战略性新兴产业发展规划》提出对大数据的支持、2017 年正式施行的《中华人民共和国网络安全法》强调对网络数据的保护等时点较为吻合。最后，本文还计算得到处理组企业在数据共享发生前的 *innovation* 均值为 0.9824，略低于对照组企业的均值 1.1494，组间差异经 *t* 检验不存在统计显著性。上述结果表明，处理组的行业与年份分布相对合理，一定程度上缓解了结论外推性不足的担忧，且基准结果并非源自处理组初始具有较高的创新水平。

附表 3 处理组样本的行业、年份分布

行业名称	客户企业数	年份	观测值数
通用设备制造业	1	2012	1
计算机、通信和其他电子设备制造业	1	2013	1
批发业	3	2014	0
零售业	1	2015	2
水上运输业	2	2016	7
装卸搬运和仓储业	1	2017	4
电信、广播电视和卫星传输服务	3	2018	3

互联网和相关服务	1	2019	2
房地产业	1	2020	3

（二）倾向得分匹配与粗化精确匹配

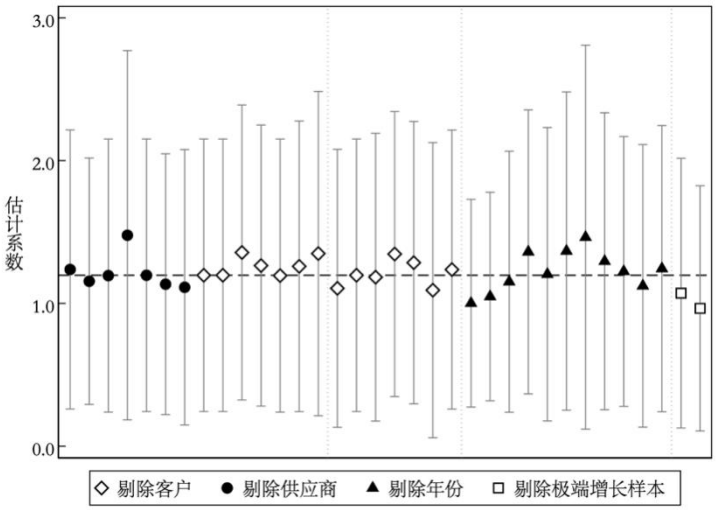
供应商在制定数据共享的决策时可能将客户企业的部分特征纳入考虑，这会使得样本中的处理组与对照组存在显著差异。为此，本文采用匹配方法构造可比的对照组。具体地，本文以基准回归中的全部控制变量为协变量，分别采用基于倾向得分匹配（PSM）的 1:1 近邻匹配、1:3 近邻匹配，以及粗化精确匹配（CEM）。三种方法下，处理组与对照组的协变量均不存在显著的组间差异，附表 4 结果显示基于上述匹配样本的回归结果保持稳健。

附表 4 基于匹配样本的回归结果

变量	(1) PSM 1:1	(2) PSM 1:3	(3) CEM
datasharing	0.8015*	0.7676*	1.2156**
	(1.7287)	(1.7284)	(2.3901)
控制变量	Yes	Yes	Yes
企业固定效应	Yes	Yes	Yes
行业—年份固定效应	Yes	Yes	Yes
样本量	523	935	3330
Adj. R^2	0.9231	0.9241	0.8969

（三）逐一剔除处理组

为排除基准结果由特定处理组样本所主导的可能性，本文借鉴李书娟等（2021）的做法，逐一剔除处理组客户企业、进行数据共享的供应商以及各年份样本。此外，本文还计算了处理组客户企业在首次接受数据共享前后的专利申请数差值，分别剔除增长值、增长率高于均值的极端增长样本。附图 2 呈现了逐一剔除样本后 *datasharing* 的估计系数及 90%置信区间。可以看出，核心解释变量的估计系数始终显著为正，且均分布在虚线代表的基准估计值附近，表明本文的基准结论并非由少数供应商、客户、年份或极端增长样本所驱动。



附图 2 逐一剔除处理组的估计结果

（四）野聚类自助法

本文采用基于 Webb 六点分布权重的野聚类自助法 (Cameron et al., 2008; Webb, 2023)，得到核心解释变量 *datasharing* 基准估计系数的稳健 *p* 值为 0.033，IV 估计系数的稳健 *p* 值为 0.049，统计显著性均未发生实质变化。

附录 5 排除替代性解释的结果

附表 5 第（1）列结果显示 *datasharing* 的估计系数仍然显著为正，表明即便排除供应商是否拥有 APP 这一差异的影响，数据共享对客户企业创新的促进作用依然成立。附表 5 第（2）列以客户企业是否至少存在一家收集用户数字足迹的供应商（*datacollection*）作为解释变量，其估计系数并不显著。附表 5 第（3）列显示，在控制供应商能力特征后，*datasharing* 的估计系数在 5%水平上显著为正。

附表 5 排除替代性解释的结果

变量	(1) 剔除供应商未开发 APP 样本	(2) 供应商收集数据	(3) 控制供应商能力
datasharing	1.0860*		1.2201**
	(1.8431)		(2.0991)
datacollection		0.2671	
		(1.6353)	
控制变量	Yes	Yes	Yes
企业固定效应	Yes	Yes	Yes
行业—年份固定效应	Yes	Yes	Yes
样本量	979	6405	6405
Adj. R^2	0.8934	0.8784	0.8794

附录 6 其他稳健性检验的结果

附表 6 第（1）列为采用泊松伪极大似然回归重新估计供应商数据共享对客户企业创新绝对数量（*innovation_count*）的影响的结果。第（2）列为剔除了客户与至少一家供应商位于同一地级市的样本的结果。第（3）列为基准回归纳入逆米尔斯比（*IMR*）的结果。第（4）列为剔除客户企业所在行业为数字经济核心产业样本的结果。第（5）列为控制政策虚拟变量的结果。可以看出，各列中基准结果保持稳健。

附表 6 其他稳健性检验结果

变量	(1) *innovation_count*	(2) 剔除同城样本	(3) Heckman	(4) 剔除数字企业	(5) 排除同期政策
datasharing	0.8992***	1.2469**	1.1985**	1.4709*	1.1425**
	(3.5136)	(2.0418)	(2.0684)	(1.7440)	(2.0279)
IMR			-0.0151		
			(-0.4755)		
控制变量	Yes	Yes	Yes	Yes	Yes
企业固定效应	Yes	Yes	Yes	Yes	Yes
行业—年份固定效应	Yes	Yes	Yes	Yes	Yes

样本量	2976	4754	6405	5690	6405
Adj. R^2	—	0.8932	0.8793	0.8731	0.8798

参考文献

[1] 李书娟、陈邱惠、徐现祥，2021，《不利冲击下经济增长恢复的经验——基于中国经济目标管理实践》，《经济研究》第 7 期，第 59~77 页。

[2] Cameron, A. C., J. B. Gelbach, and D. L. Miller, 2008, "Bootstrap-based Improvements for Inference with Clustered Errors," *The Review of Economics and Statistics*, 90(3), pp. 414~427.

[3] Webb, M. D., 2023, "Reworking Wild Bootstrap - based Inference for Clustered Errors," *Canadian Journal of Economics*, 56(3), pp. 839～858.