大数据应用对中国企业市场价值的影响

——来自中国上市公司年报文本分析的证据

张叶青 陆瑶 李乐芸

经济研究, 2021.12

吕漫妮 2022.5.18

目录

- 引言
- 研究背景和动机
- 研究问题和结论
- 贡献
- 大数据与实体企业融合的作用机理
- 研究设计与实证结果
- 结论

研究背景和研究动机

- 自2013年,众多全球知名企业开始迅速收购大数据上、下游厂商,布局大数据战略;
- 我国2020年**数字经济增加值**占GDP38.6%, **数字经济增速**为GDP增速的3.2倍;
- 党的十八届五中全会正式将大数据发展上升为国家战略;
- 党的十九大报告进一步明确将"推动互联网、大数据、人工智能和实体经济深度融合" 列为深化供给侧结构性改革和建设现代化经济体系的重要举措。
- 疫情期间经济活动非聚集性的需要进一步激发了实体经济对数字化、智能化技术的应用需求。
- 面对大数据蓬勃发展的趋势和疫情冲击导致的数字化需求,我们亟需思考在数字化技术与传统产业深度融合过程中,大数据应用对微观企业的影响及其作用机理,从而为后疫情时期的高质量经济发展奠定基础。
- 然而,截至目前,探索大数据应用及其对公司实际影响的研究非常匮乏。

研究问题与结论

- 大数据在中国上市公司中的应用普遍吗?哪些因素影响了公司应用大数据的决策?大数据应用如何影响公司的市场价值?其背后的影响机制是什么?上述影响在不同的公司、行业间是否存在着异质性?
- > 基于年报得到的大数据应用指标可以衡量大数据在中国上市公司中的应用程度
- 规模较大、有形资产比例较低、盈利能力较强,以及所在地区市场化程度较高的公司更可能在生产经营过程中应用大数据
- > 大数据应用能够显著提高公司的市场价值
- 大数据的应用显著提高了公司的生产效率和研发投入,但现实中的技术和人才 供给不足则会限制大数据应用对企业的积极作用
- 大数据应用对公司市场价值的促进作用在小规模公司、非国有公司和所在行业 竞争较为激烈的公司中更为显著

贡献

- 利用非结构化的文本数据构建了公司层面的大数据应用程度的**衡量指标**,更为直接、有效、准确,覆盖面广
- 实证检验了大数据应用对公司市场价值的积极意义,并深入探索了其影响机制、 影响的异质性,为大数据应用对于公司竞争力提高、宏观经济增长的意义构建 了理论基础,提供了实证依据,为政府制定大数据相关政策提供了参考

大数据与实体企业融合的作用机理

- 企业应用的大数据:企业收集、处理与利用的海量、高速、多样化的数据要素或资产
- 在大数据与实体经济深度融合的进程中,大数据的应用影响企业生产经营和最终市场价值的作用机理是什么?

大数据应用与生产效率:大数据应用可以通过**信息渠道**来影响企业的生产效率;大数据 **降低企业的劳动力成本**,从而提高企业的生产效率;大数据可以从**组织管理**的角度降低 企业生产运营的成本

大数据应用与研发创新:大数据应用有利于企业更精准地把握消费者的需求,明确当前生产过程的不足和技术创新的方向,进而提高研发投入的需求;大数据应用降低了企业研发过程的不确定性和成本,增强了研发的动机;大数据应用为研发过程积累了丰富的资源和信息,增强了企业研发效率

大数据应用过程中的摩擦:大数据与实体经济的融合受到**相关人才、技术**等方面匮乏的限制。从人才的角度,企业应用大数据所需的人才或来源于内部资源,或来源于外部市场。从技术支持的角度,所在地区的数据类服务供给和数据中心等新型基础设施建设是企业利用大数据实现增值的重要保障

数据来源及变量描述

• 样本选择

2006—2017 年中国A股所有上市公司,剔除(1)ST、*ST 和PT公司;(2)IPO当年的观测值和已退市的公司;(3)净资产为负的观测值;(4)主要变量缺失的公司;(5)**信息传输、软件和信息技术服务业的公司**

获得2501家上市公司共20623个样本

数据来源

大数据应用的相关变量来自于对公司年报的文本分析,市场交易和财务数据来自于国泰安数据库,CPI 数据来自于国家统计局,对连续变量进行上下1%的缩尾

数据来源及变量描述

• 大数据应用指标 - InBigdata

公司年报中提及表中大数据相关关键词的次数加一后取对数,按照"公司—年份"的观测值确定每年缩尾上下极值各1%

其中,关键词的选取借鉴了以往文献(Chen et al., 2012; McAfee & Brynjolfsson, 2012; Farboodi et al., 2019)、政府文件以及业界报告等,**紧扣定义**,按照**普适性原则**进行筛选

表 1 "大数据"相关关键词及定义

关键词	定义
大数据	企业收集、处理与利用的海量、高速、多样化的数据要素或资产。
海量数据	根据高德纳公司对大数据的定义,海量规模是大数据的重要特征之一。
数据中心	安置计算机系统及相关部件的设施,用于在网络基础设施上传递、加速、展示、计算、存储数据信息。信息时代下,大数据需要安全可靠、高效率的数据中心进行存储、计算和交换。
信息资产	指由企业拥有或者控制的能够为企业带来未来经济利益的信息资源。根据高德纳公司的报告,大数据本质上是一种信息资产。
数据化	将均匀、连续的数字比特结构化和颗粒化,形成标准化的、开放的、非线性的、通用的数据对象,并 基于不同形态与类别的数据对象,实现大数据的应用。
算力	也称哈希率,指比特币网络处理能力的度量单位,也是计算机计算哈希函数时输出的速度。

数据来源及变量描述

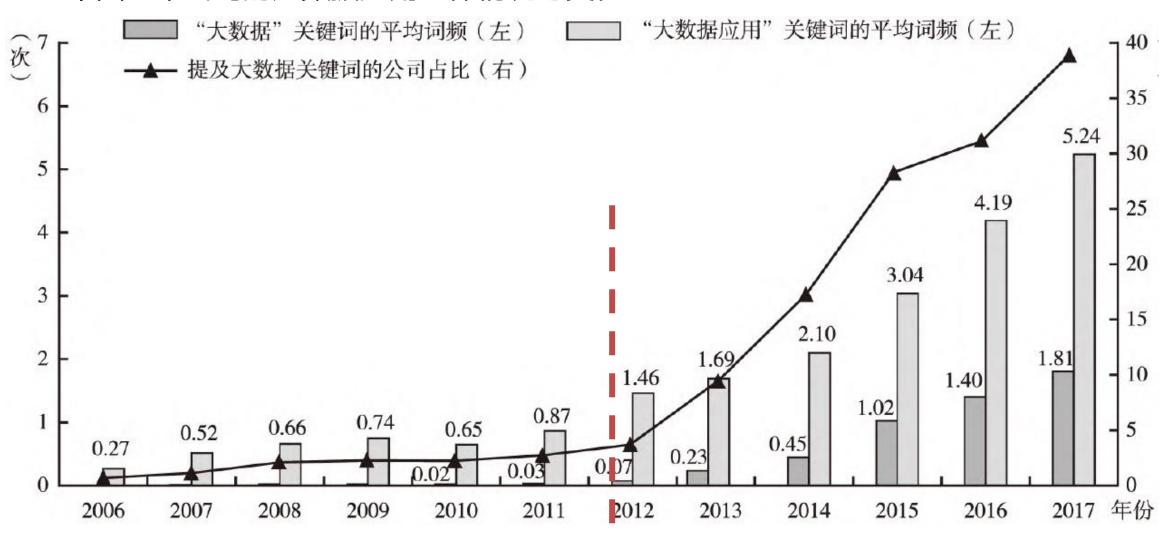
• 其他变量的定义

被解释变量:公司的估值指标 **托宾Q值**——公司总市值与总负债之和除以公司总资产控制变量包括:公司规模(InAssets)、公司杠杆率(Lev)、固定资产比率(PPE_TA)、公司年龄(InAge)、国有性质的虚拟变量(SOE)、销售收入增长率(SalesGrowth)和总资产收益率(ROA)。其中公司规模和销售收入增长率经过CPI调整

变量	变量定义	均值	标准差	中位数	观测数
ln <i>Bigdata</i>	表1中大数据相关关键词在年报中出现的次数加一后取对数	0.180	0.508	0.000	20623
Tobin's Q	公司总市值与总负债之和除以公司总资产	2. 485	1. 689	1. 946	20623
lnAssets	总资产取自然对数	21. 778	1. 241	21. 609	20623
Lev	总资产除以股东权益	0. 457	0. 204	0. 461	20623
PPE_TA	固定资产除以总资产	0. 246	0. 173	0. 213	20623
lnAge	当年年份减去上市年份加1,再取自然对数	2. 183	0. 694	2. 303	20623
SOE	按公司实际控制人性质确定	0. 468	0. 499	0. 000	20623
SalesGrowth	当年营业收入除以上一年营业收入后减一	0. 194	0. 442	0. 119	20623
ROA	净利润除以总资产	0. 0371	0. 0519	0.0336	20623

大数据应用的描述与决定因素

• 中国上市公司的大数据应用整体情况与变化



决定因素分析

- probit(模型1) 和OLS(模型2)
- 是否提及大数据相关关键词的虚 拟变量(Bigdata_dummy) (模型 1);年报中大数据关键词的频率 加一后取对数(InBigdata)模型(2)
- 规模较大的公司、有形资产占比低、盈利能力强的公司更可能应用大数据;
- 所在地区市场化度越高,上市公司越可能应用大数据

	probit			OLS
变量		Bigdata	_dummy	lnBigdata
		(1)	(2)	(3)
lnAssets		0. 0831 ***	0.0856***	0. 0947 ***
		(0. 0176)	(0. 0175)	(0.0215)
规模		- 0. 0143 ***	- 0. 0142 ***	- 0. 0456
lnAge		(0. 00312)	(0. 00302)	(0.0716)
Lev		0. 0559	0. 0425	0.00855
Lev		(0.110)	(0.108)	(0.0529)
PPE_TA		-0.656***	- 0. 675 ****	-0.109*
方以次立 FF		(0.112)	(0.111)	(0.0576)
一	L	0. 0422	0. 0478	0.0141
SalesGrown		(0. 0385)	(0. 0382)	(0.0101)
ROA		0. 921 **	0.905**	0. 239*
<u> </u>		(0.406)	(0.403)	(0. 131)
SOE		- 0. 0994 **	-0.0929**	0.00205
SOE		(0. 0418)	(0. 0410)	(0.0282)
Sh1		0. 0446	0. 0555	-0.132
5/11		(0.116)	(0.115)	(0.0759)
Poord Sire		-0.0896	-0.0817	0.0133
Board_Size		(0. 0884)	(0. 0878)	(0.0478)
Education			0.170	
Education			(0.221)	
Fangang_index			0. 0285 ****	
所在地区市場		ν ρ	(0. 0110)	
アハ1工ノビレムコンル GDP_ret	רנע! 	心 反	-1.125	
GDF_fet			(1.275)	
固定效应	7	Year, Industry, Province	Industry, Year	Firm, Province-Year, Industry-Year

大数据应用对公司市场价值的影响

• 基准模型: $Y_{ijpt} = \gamma_0 + \gamma_1 BigData_{ijpt} + \gamma_2 Controls_{ijpt} + \delta_{pt} + \gamma_{jt} + \mu_i + \xi_{ijpt}$

亦具	Tobin's Q					
变量	(1)	(2)	(3)	(4)		
${ m ln} Bigdata$	ln <i>Bigdata</i> 0. 157 **** (0. 0375)		0. 153 *** (0. 0354)	0. 151 **** (0. 0337)		
控制变量	量 Yes Yes		Yes	Yes		
固定效应	Year, Firm	Firm, Industry–Year	Firm, Province-Year	Firm, Industry-Year, Province-Year		
观测值	20623	20623	20623	20623		
adjusted R ²	0. 695	0. 698	0. 699	0. 702		

• 公司应用大数据越多,其在股票市场上的价值越高。

大数据应用对公司市场价值的影响

• 内生性问题的处理:

- 1、工具变量的设计基于**2009 年启动**的"珠峰计划",旨在培养拔尖创新人才,推动创新型国家的建设,首批试点选择了17所高校理工科相关专业,它的实施**增加了试点大学附近理工类人才供给,降低附近公司应用大数据的劳动力成本**
- 2、人力资源成本是大数据决策的关键:人力成本降低,企业更愿意应用大数据→上市公司与试点大学的距离越近,公司受到政策的辐射力度越大,越有可能提高大数据应用程度;所在城市的上市公司数目越多,则上述辐射力度越可能被削弱

$$IV_{ii} = \frac{\ln\left(\sum_{k=1}^{n} \frac{1}{distance_{ik}}\right)}{N_{s}} \times Post_{t}$$

i: 公司 k: 首批试点高校 c: 公司办公地点所在城市

Nc: 2014 年公司i 的办公地点所在城市c 中的上市公司的总数量

Post: 为时间虚拟变量, **2014 年之后**设为1, **2014 年之前**则设为0

• 相关性, 排他性

大数据应用对公司市场价值的影响

・ 工具变量回归结果:

第一阶段回归: IV的系数显著为正,表明在**政策之后,距离 试点高校越近的上市公司对大** 数据应用程度越高

第二阶段回归:大数据应用指标的系数仍然在5%水平上显著为正,即处理了内生性问题后,大数据应用提高公司市场价值的基本结论不变

	第一阶段结果	第二阶段结果	第一阶段结果	第二阶段结果
变量	$\ln\!Bigdata$	Tobin's Q	ln <i>Bigdata</i>	Tobin's Q
	(1)	(2)	(3)	(4)
T 77	0. 076 ***		0.060 ***	
IV	(0.013)		(0.012)	
1 7 7		1. 333 ***		1. 420 ***
lnBigdata		(0.593)		(0.630)
控制变量	Yes	Yes	Yes	Yes
固定效应	Firm, Industry–Year	Firm , Industry–Year	Firm, Industry–Year, Province–Year	Firm, Industry–Year, Province–Year
观测值	17001	17001	17001	17001
弱工具变量检验				
F 统计量 (Cragg-Donald)	106. 734		54. 313	

其他稳健性检验

- **更换大数据衡量指标**:虚拟变量、分类变量、大数据词频密度、更为广义的关键词集合、避免所选关键词的行业特质性的干扰
- 更换研究样本: A 股全部行业的上市公司、剔除金融业公司
- **避免企业因追逐热点而虚假披露大数据信息**: 仅使用大数据发展早期的样本(即2014年及之前的数据)

影响机制的检验——生产效率和研发投入

生产效率 or 研发强度 = $\gamma_0 + \gamma_1 BigData_{ijpt} + \gamma_2 Controls_{ijpt} + \delta_{pt} + \gamma_{jt} + \mu_i + \xi_{ijpt}$

• **生产效率**: 全要素生产率(TFP) (Liu & Lu, 2015; Liu & Qiu, 2016)

InBigdata回归系数显著为正,表明**大数据的应用帮助公司显著提高了自身的生产效率**

• 研发强度: R&D的资金投入和人员投入

大数据应用程度的提高对公司的研发强度有显著的正向影响,从而印证了影响渠道: **大数据应用通过促进公司研发来提高其市场价值**

被解释变量	TFP	R&D_Exp/Assets	R&D workers (%)
	(1)	(2)	(3)
$\ln\!Bigdata$	0. 0245 ** (0. 0108)	1. 503 *** (0. 477)	1. 710 *** (0. 298)
控制变量	Yes	Yes	Yes
固定效应	Yes	Yes	Yes

大数据应用中的摩擦——人才需求和技术支持

公司市场价值 $Y_{ijpt} = \gamma_0 + \gamma_1 BigData_{ijpt} + \gamma_2 Controls_{ijpt} + \delta_{pt} + \gamma_{jt} + \mu_i + \xi_{ijpt}$

• 人才需求角度:公司内部技术人员储备不足,外部劳动力市场的相关人才供给不足

• 技术支持角度:公司所在地区的数据相关技术服务与基础设施建设不足

被解释变量	Tobin's Q					
	内部的技术人员储备		外部的高学历人才供给		数据基础设施建设	
样本	>中值	< 中值	>中值	<中值	国家大数据 中心所在省份	其他省份
	(1)	(2)	(3)	(4)	(5)	(6)
$\ln\!Bigdata$	0. 195 **	0. 0911	0. 185 ***	0. 111 **	0. 204 **	0. 142 ***
	(0.078)	(0.057)	(0.050)	(0.040)	(0. 0661)	(0. 0344)
控制变量	Yes	Yes	Yes	Yes	Yes	Yes
固定效应	Yes	Yes	Yes	Yes	Yes	Yes
difference	0. 104(p < 0. 01)		0. 0741(p < 0. 01)		0. 0618(p < 0. 01)	

内部技术人员储备较为充足的公司能更好地将大数据转化为公司的市场估值;高学历劳动力供给充足的地区的公司能够更好地将大数据转化为公司的市场价值;位于数据基础设施建设好的地区的公司能够更好地实现大数据对公司的增值效果

异质性影响

- 公司规模:大公司的数据更为丰富,因此基于大数据的预测更精确,从而降低生产的不确定性和融资成本。因此,大公司的大数据应用可能对市场价值的提升效果更好。但是,相对于大公司,小公司受到分析师的追踪分析和投资者的关注不足,自身生产经营中和外部资本市场上的信息不对称程度更为严重。因此,大数据的应用对于小公司获得更多信息来提高预测精度的意义更为重大。——小公司更能从大数据应用中实现市场价值的提高
- 国有性质:这两类公司在经营目标、外部经营环境和公司内部治理等方面存在显著差异——大数据应用对非国有公司市场价值的提升作用较大且在统计意义上显著,但是对国有公司的影响很小且为负
- 行业竞争程度: 竞争激烈行业的公司可能更有动力利用大数据来提高生产效率,实现企业增值——大数据应用对公司市场绩效的提升效果在竞争行业(行业主营业务利润率标准差的倒数)中更为显著

结论

本文利用中国A股上市公司披露的年报关键词,构造了公司层面的大数据应用程度的衡量指标,描述了大数据在中国上市公司中的应用情况,探究了公司大数据应用的影响因素,大数据应用对公司市场价值的影响及其作用机制(影响机制和摩擦因素)和异质性。