

大数据技术赋能下软硬信息对 中小微企业融资的差异性研究

池仁勇^{1, 2} 石雨露²¹

(1 浙江工业大学 中国中小企业研究院, 杭州 310023;

2 浙江工业大学 管理学院, 杭州 310023)

【摘要】: 以浙江省 816 家中小微企业为样本, 利用分层线性回归模型分析大数据技术赋能视角下软硬信息对中小微企业融资的差异性影响。研究表明, 非结构化软信息能够增强中小微企业的信用评级从而缓解融资约束的作用效果强于结构化硬信息对其影响, 而囿于技术桎梏无法充分利用软信息的问题在大数据技术的出现之后也得到了充分解决, 并且大数据技术发展越成熟, 非结构化软信息对信用评级的影响作用越能得到充分发挥。

【关键词】: 大数据技术 企业信用 融资约束 软信息 硬信息

0 引言

中小微企业融资难、融资贵已成为普遍现象, 如何解决该问题成为社会关注焦点。大量研究表明, 中小微企业不仅缺乏规模效应, 更缺少固定资产抵押、齐全的财务数据等结构化硬信息, 但中小微企业却具有成长性、创业创新精神、对市场和新技术的敏感性等丰富的非结构化软信用信息(即软信息)。事实上, 银行在审核贷款时仍然更加注重企业是否有可抵押的固定资产等, 这使得软信息的作用被极大的掩盖。究其原因非结构化的软信息难以被传统商业银行甄别和利用, 这不仅是由于软信息调查、甄别成本大, 传统商业银行风险收益难以对等, 还因为软信息具有变化幅度大、时效短等特性, 极大增加了商业银行监控难度。但是, 随着互联网快速发展而兴起的大数据征信模式不仅能够提高硬信息的保存年限, 更能有效解决软信息收集难度大、成本高的技术性难题, 故其可以将两种信息纳入同一评价体系中, 因此, 本文将从理论层面探讨大数据征信对中小微企业信用评价的影响机理, 并以浙江省 816 家中小微企业为研究样本, 验证大数据征信对提高中小微企业信用刻画准确度从而缓解企业融资约束的效果, 以揭示大数据征信解决中小微企业融资难的内在机理。

1 文献综述与研究假设

1.1 软信息和硬信息的作用机制

非结构化软信息包括了意见、个体性格、行为方式、管理规章制度、市场声望等, 其收集方式主要依赖于人工, 依靠具有

作者简介: 池仁勇, 管理学博士, 浙江工业大学中国中小企业研究院院长, 浙江工业大学管理学院教授、博士生导师, 研究方向: 技术经济、中小企业创新管理;

石雨露, 浙江工业大学管理学院硕士研究生, 研究方向: 技术经济、中小企业创新管理。

基金项目: 国家社会科学基金重大项目——“大数据背景下中小微企业征信体系的理论与方法”(项目编号: 17ZDA088;项目负责人: 池仁勇)成果之一

特殊经验的专业人员根据特定的目的进行收集。软信息具有较强的灵活性，中小微企业因规模较小、成立时间较短、组织层级较少，所以信息传递会更快更有效率，企业可以凭借软信息向外界传递经营发展质量优劣的信号，丰富企业信用等级的评估，进而缓解中小微企业的融资约束。软信息还具有粘性，借贷双方可以通过多次交易建立专属客户通道，能将以往的信息沉淀下来，有效降低双方信息不对称程度。

硬信息具有结构化、不依赖特定情境的特点，凭借其具有的特定结构，收集和处理工作可以委派给技术能力较低的工人或者计算机，从而降低交易成本，企业可以将更多的精力和资金用于提高其经营能力，进而改善企业的信用评价。不仅如此，硬信息还可编码，在传递的过程中不会造成信息失真，能应用在更广的区域范围，提升企业信用评价的可信度，助力企业获得资金支持。

软信息和硬信息通过不同的机制影响中小微企业的信用评价进而影响其融资约束，虽然传统的信用评价更注重硬信息的分析，但是对于中小微企业而言，软信息发挥的作用要强于硬信息。一方面，中小微企业具有规模较小、内部控制不规范、财务报告不健全等特征，其本身的硬信息表现较差，难以凭借有限的硬信息获得贷款支持；另一方面，中小微企业具有高成长性和灵活性的特征，其软信息包含了更加丰富的内容，能更全面真实的反映企业的信用水平。池仁勇等在使用软信息和硬信息对中小企业信用风险进行预测时，发现软信息的预测能力要明显高于硬信息的预测能力^[1]。

综上，本文提出假设：

H1a:软信息和硬信息对中小微企业的信用评价具有正向影响。

H1b:软信息对中小微企业的信用评价的影响高于硬信息。

1.2 企业规模的调节作用

根据资源基础观，大型企业和中小微企业所具有的信息资源存在差异。一方面，大型企业的规模较大，经营年限相对来说较长，财务制度会更加完善，较高的信息透明度使得金融机构和银行等贷款机构在进行投资决策时更加倾向于选择大型企业，缓解企业的融资约束，故在大型企业中，硬信息发挥的作用更大。另一方面，对于中小微企业而言，由于其规模较小，内部控制不够完善，财务制度也不够规范，此时，企业家个人的诚信水平、参与的社会公益活动、创新活力等非结构化的软信息更能反映中小微企业的真实信用水平。因此，本文提出假设：

H2a:在大型企业中，硬信息发挥的作用更强。

H2b:在中小微企业中，软信息发挥的作用更强。

1.3 大数据征信的调节作用

大数据技术是数据对象、技术与应用三者的有机统一，将大数据技术应用到征信领域进而兴起的大数据征信模式逐渐得到学者的青睐，表1总结了当前学者大数据征信的研究观点。

大数据征信一定程度上破除了企业信用数据的部门所有，打破了信息孤岛和信息鸿沟，使跨部门数据查询和收集成为可能。可以对全网获取的硬信息进行交叉检验，提高其精确度，为企业的战略决策提供有价值的历史信息。此外，大数据征信拥有的独特模型算法可以从海量软信息中挖掘出有价值的内容，并将其应用在企业的信用评价中，显著提高软信息的利用率。

表 1 大数据技术对企业信用评级的影响研究观点

观点	描述
降低信息处理成本	大数据技术自动筛选海量数据，减少冗余，降低了第三方机构收集软信息的成本 ^[2] ，有利于提升中小微企业的信用评级，缓解其融资约束。
提高信息获取及时性	大数据背景下，使数据 T+0 传递成为可能，中小微企业的信用信息、交易记录可以被实时动态的监控到 ^[3-4] 。
丰富信用数据内容	大数据技术打破了数据部门分割的现象 ^[5] ，使数据共享、非结构化数据收集处理成为可能，政府与市场、公司数据传递更加畅通，可以更加全面的还原企业的信用能力 ^[6] 。
破除信用信息时空阻隔	大数据技术实现线上数据实时动态分析，打破了时空阻隔，提高了征信的时效性和精准性 ^[7-8] 。

虽然大数据征信模式能极大的发挥软信息和硬信息对企业信用评价的影响，但是对硬信息的影响要弱于软信息，这是因为硬信息的数据结构相对而言比较规范，其作用在传统征信模式下已经得到极大的发挥，大数据征信本质上并不能改变硬信息。而囿于技术不足，软信息作用在传统征信模式下难以得到有效体现，而这种技术痛点却能在大数据技术和征信体系结合的模式下得到迎刃而解，其能够对这些软信息进行线上自动处理，利用机器学习算法，将不具有规范结构的信息内容进行文本分析，形成规模效应，极大限度地挖掘软信息的价值，完善中小微企业的信用评价。

因此，本文提出以下假设：

H3a: 大数据征信正向调节软信息和硬信息对企业的信用评价。

H3b: 大数据征信在软信息对企业信用评价影响中发挥的调节作用强于硬信息。

2 数据和变量选择

2.1 样本选取和研究方法

为了论证上述假设，本文采取企业和区域两个层面的数据，利用 HLM 软件构建分层线性回归模型进行研究分析。企业层数据来源于对浙江省中小微企业高层管理人员发放的调查问卷，区域层样本数据是浙江省经济和信息化厅发布的各地级市两化融合指数。

2.2 变量选取

企业层面和区域层面的核心变量和控制变量定义如表 2 所示，为了排除各个区域的宏观经济情况和企业自身情况对模型的影响，本文采用企业年龄和企业所处行业以及区域 GDP 增长率和人均 GDP 作为控制变量。

表 2 变量的定义

变量名称	符号	变量定义
自变量		
软信息	SOFTINFOR	企业家的兴趣爱好，社交情况等多项综合加权得分
硬信息	HARDINFOR	企业收入增长率和税前利润增长率的综合指数
控制变量		
企业年龄	AGE	企业成立至调查日的时间
所属行业	INDU	虚拟变量，工业制造业为 1, 其他为 0
GDP 增长率	GROGDP	GDP 增长额/上年 GDP
个人 GDP	PERGDP	人均国内生产总值，人民生活水平的一个标准
调节变量		
企业规模	SIZE	虚拟变量，中小微企业为 1, 大型企业为 0
基础环境	BD1	基础环境指数
工业应用	BD2	工业应用指数
应用效益	BD3	应用效益指数
大数据技术	BIGDATA	浙江省信息化和工业化融合指数
因变量		
企业信用	CREDIT	企业在供应商，客户，消费者处评价的综合得分

2.3 同源偏差

为避免同源偏差对实验结果产生干扰，本文采用 Harman 单因子法对回收的问卷进行检验，将企业层所有变量的题项进行探索性因子分析，发现单因子解释的变异都小于 40%，认为同源偏差不严重，数据具有可信度，可以进一步分析。

2.4 信效度分析

本文对变量信息进行信效度分析，变量对应的 KMO 值为 0.790，Bartlett 球形检验的显著性均为 0.000，问卷的信度检验较好，采用 SmartPLS 对变量进行进一步检验，结果见表 3 和表 4。

表 3 变量信度与效度

变量	Alpha	组合信度	平均抽取变量量
SOFTINFOR	0.711	0.837	0.632

HARDINFOR	0.827	0.917	0.847
CREDIT	0.872	0.913	0.724

表 4 区别效度-HTMT

变量	SOFTINFOR	HARDINFOR	CREDIT
SOFTINFOR			
HARDINFOR	0.203		
CREDIT	0.445	0.069	

3 实证结果和分析

3.1 描述性统计

采用 SPSS26 对调研的 816 家企业的样本数据进行处理，表 5 代表了企业层面和区域层面各个变量的描述性统计。

表 5 变量的描述性统计

变量	最小值	最大值	平均值	标准差
AGE	0.000	36.000	8.990	6.800
INDU	0.000	1.000	0.480	0.500
SIZE	0.000	1.000	0.060	0.250
SOFTINFOR	1.000	5.000	3.400	0.720
HARDINFOR	1.000	6.000	4.640	0.970
CREDIT	1.000	5.000	3.900	0.830
BIGDATA	60.550	103.050	84.290	11.980
GROGDP	0.070	0.110	0.090	0.020
PERGDP	11.060	11.850	11.390	0.290

3.2 变量的相关性检验

企业层面和区域层面的变量是连续型变量，因此选择皮尔逊相关性系数进行相关性检验。企业层和区域层的变量相关系数

的绝对值都小于 0.600, 且该相关性并不显著, 说明变量之间的独立性较好, 可以进行进一步分析。

3.3 模型设定与检验

3.3.1 软信息和硬信息的主效应

利用上述样本数据, 本文分析软信息和硬信息对企业信用评价的影响如表 6 所示。结果表明, 软信息和硬信息对企业的信用评价有显著的正向影响, 并且软信息对企业信用评价的影响要高于硬信息, 假设 H1a 和 H1b 得到验证。

表 6 模型参数估计和检验结果

变量	模型 1	模型 2	模型 3
AGE	0.004	0.004	0.004
	(0.004)	(0.004)	(0.004)
INDU	0.239***	0.238***	0.239***
	(0.049)	(0.049)	(0.049)
SIZE	0.069	1.181***	-0.079
	(0.107)	(0.374)	(0.467)
SOFTINFOR	0.713***	0.710***	0.710***
	(0.035)	(0.035)	(0.036)
HARDINFOR	0.175***	0.197***	0.175***
	(0.028)	(0.028)	(0.028)
SIZE×SOFTINFOR		0.323***	
		(0.104)	
SIZE×HARDINFOR			0.048
			(0.146)
常数	0.502**	0.406*	0.510**
	(0.207)	(0.208)	(0.208)
Pseudo-R ²	0.366	0.373	0.365

3.3.2 企业规模的调节作用

企业规模在软信息和企业信用评价之间的影响具有显著的正向调节作用, 如表 6 中的模型 2、模型 3。在中小微企业中, 软

信息发挥的作用要强于硬信息，但是在大型企业中，硬信息强于软信息的作用并不显著，这是因为对于大型企业而言，其经营年限较久，拥有较丰富的社会关系，其软信息也比较丰富，因此假设 H2a 得到验证，但假设 H2b 未得到验证。

3.3.3 大数据技术的调节作用

零模型。为了判断样本数据是否适用于分层线性模型的方法来估计区域间的差异，在研究之前首先构建了不包含任何变量的零模型。

层-1 结构模型：

$$CREDIT_{ij} = \beta_{0j} + r_{ij}, r_{ij} \sim N(0, \sigma^2) \quad (1)$$

层-2 结构模型：

$$\beta_{0j} = \gamma_{00} + \mu_{0j}, \mu_{0j} \sim N(0, \tau_{00}) \quad (2)$$

其中 i 和 j 分别代表不同的企业和区域， $CREDIT_{ij}$ 代表第 j 个城市的第 i 家企业的信用评价，采用 HLM 软件对零模型进行分析后，分析结果见表 7。表 7 中的模型 4 显示，截距 γ_{00} 的方差成分为 0.004 并且在 5% 的水平下显著，结果表明各城市之间具有显著差异，可以对样本数据进行分层线性模型分析。

随机系数模型。层-1 和层-2 的结构模型如式 (3) 和式 (4) 所示，系数值 β_{pj} 代表企业层面各个解释变量和控制变量对被解释变量的影响效果。

层-1 结构模型：

$$CREDIT_{ij} = \beta_{0j} + \beta_{1j}(AGE)_{ij} + \beta_{2j}(INDU)_{ij} + \beta_{3j}(SIZE)_{ij} + \beta_{4j}(SOFTINFOR)_{ij} + \beta_{5j}(HARDINFOR)_{ij} + r_{ij}, r_{ij} \sim N(0, \sigma^2) \quad (3)$$

层-2 结构模型：

$$\beta_{pj} = \gamma_{p0} + \mu_{pj}, \mu_{pj} \sim N(0, \tau_{p0}), (p=0, 1, 2, 3, 4, 5) \quad (4)$$

跨层次调节模型。综上分析，随机效应变量是企业所属的行业、企业的软信息和硬信息，固定效应变量是企业年龄和企业规模，为了进一步分析区域层变量对企业层的变量的影响情况，在区域层交互项中加入 BIGDATA 核心变量，分析在大数据背景下，软信息和硬信息对企业信用评价的影响，改进模型如下：

层-1 结构模型：

$$CREDIT_{ij} = \beta_{0j} + \beta_{1j}(AGE)_{ij} + \beta_{2j}(INDU)_{ij} + \beta_{3j}(SIZE)_{ij} + \beta_{4j}(SOFTINFOR)_{ij} + \beta_{5j}(HARDINFOR)_{ij} + r_{ij}, r_{ij} \sim N(0, \sigma^2) \quad (5)$$

层-2 结构模型：

$$\beta_{pj} = \gamma_{p0} + \sum_{s=1}^S \gamma_{ps} W_{sj} + \mu_{pj}, \mu_{pj} \sim N(0, \tau_{p0}), (p = 0, 4, 5)$$

$$\beta_{qj} = \gamma_{q0} + \mu_{qj}, \mu_{qj} \sim N(0, \tau_{q0}), (q = 1, 2, 3) \quad (6)$$

其中， W_{ij} 代表区域层面的变量，将上述模型在软件 HLM 进行运算，结果为表 7 中的模型 4 和模型 5。

表 7 分层线性模型参数估计和检验结果

		模型 1	模型 2	模型 3	模型 4	模型 5
固定效应						
截距, β_{0j}	γ_{00}	3.927***	3.952***	3.935***	4.161***	4.062***
BIGDAT 斜率	γ_{01}				0.000	-0.005
GROGD 斜率	γ_{02}				1.090	0.952
PERGDP 斜率	γ_{03}				-0.635	-0.230
AGE 斜率	γ_{10}		0.003	0.002	0.004**	0.002*
INDU 斜率	γ_{20}		0.031	-0.044	0.008	-0.080
SIZE 斜率	γ_{30}		-0.102	-0.230	-0.109	-0.260**
SOFTINFOR 斜率	γ_{40}		0.638***		0.573***	
SOFTINFOR×BIGDATA	γ_{41}				0.004***	
HARDINFOR 斜率	γ_{50}			0.025**		0.070
HARDINFOR×BIGDATA	γ_{51}					0.003
随机效应						
截距, μ_{0j}	τ_{00}	0.004**	0.071**	0.060	0.389***	0.221***
AGE 斜率, μ_{1j}	τ_{10}		0.003	0.002	0.002	0.001
INDU 斜率, μ_{2j}	τ_{20}		0.254***	0.248***	0.262***	0.237***
SIZE 斜率, μ_{3j}	τ_{30}		0.133	0.195	0.125	0.154
SOFTINFOR 斜率, μ_{4j}	τ_{40}		0.008**		0.052**	
HARDINFOR 斜率, μ_{5j}	τ_{50}			0.073*		0.106***

表 7 的研究结果表明, SOFTINFOR 和 BIGDATA 交互项显著, 且系数为正, 表明大数据技术建设越完善, 企业非结构化软信息对企业信用评价的影响越能得到充分发挥。HARDINFOR 和 BIGDATA 交互项为正但并不具有统计学意义, 这表明硬信息和企业信用评价之间的关系不受大数据技术的影响, 假设 H3a 部分通过, H3b 通过。

3.4 稳健性检验

在模型分析中, 解释变量对被解释变量的影响常常存在时间滞后效应, 也即被解释变量会受到解释变量过去值的影响。本研究选取的区域层面的大数据征信变量, 其发展本身需要一定的周期的, 从开始服务企业到对企业产生影响也需要时间, 此外, 由于经济活动的性质, GDP 等宏观经济要素对企业发展的影响往往会延续到下期。为了排除可能的影响, 本文建立滞后 1 期的动态分层分析模型, 滞后一期的结果进一步验证了上述假设, 表明研究结论具有稳健性。

4 研究结论和未来发展展望

4.1 研究结论

本文以浙江省中小微企业为研究样本, 采用调查问卷法, 收集了 816 家中小微企业的软信息和硬信息。研究得到以下结论:

第一, 非结构化软信息和结构化硬信息均对企业的信用评级有显著的正向影响, 且软信息的作用大于硬信息的。同时, 企业规模越小, 软信息对企业信用评价的影响越明显。说明对于中小微企业而言, 软信息更能代表企业真实的信用状况, 也就意味着提高软信息在信用信息评价过程中的作用能为全面刻画企业信用画像提供极大帮助。

第二, 为了探明大数据技术对软信息与硬信息的信用赋能效应, 本文进一步检验大数据技术在软信息和硬信息对企业信用评级中发挥的调节作用。研究表明, 大数据技术建设越完善, 软信息在中小微企业信用评级过程中的重要地位就越突显, 而硬信息在信用评级过程中的权重并未随之发生显著性变化。

4.2 未来发展展望

本文以浙江省 816 家中小微企业为调查样本, 由于受到问卷发放的地理性限制以及对地域性中小微企业发展水平考虑, 因此只对浙江省十个地级城市企业数据进行了分析, 故研究结果可能存在一定的局限性。未来应当进一步扩大样本搜索范围, 从而可以获得更加具有普适性的研究结论。同时, 未来也可以从动态数据着手, 分析大数据技术赋能企业融资发展的动态效应, 从而能进一步洞悉大数据等数字技术影响企业发展的内在机理。

参考文献:

- [1]池仁勇, 朱张帆. 软信息与硬信息孰轻孰重—中小企业授信与信用风险视角[J]. 华东经济管理, 2020, 34(3):112-118.
- [2]赵付春, 汪晓菲. 我国大数据征信商业模式研究—以上海市信用服务企业的创新实践为例[J]. 征信, 2019(11):18-24.
- [3]BOS J W B, MILLONE M. Show Me Yours and I' ll Show You Mine:Sharing Borrower Information in a Competitive Credit Market[J]. Social Science Electronic Publishing, 2016(8):1-58.
- [4]龚柏新. 征信体系对中小企业融资的影响分析[J]. 上海金融, 2008(8):93-96.

-
- [5]BOWMAN A W. Big questions, informative data, excellent science[J]. Stats & Probability Letters, 2018, 136:34-36.
- [6]陈小梅. 基于大数据的小微企业信用评级体系研究[J]. 征信, 2018(10):27-31.
- [7]CROSMAN P. ZestFinance Aims to Fix Underbanked Underwriting[J]. American Banker, 2012, 177(177):7-7.
- [8]SIMSEK Z, VAARA E, PARUCHURI S, et al. New Ways of Seeing Big Data[J]. Academy of Management Journal, 2019, 62(4):971-978.