

MD&A 文本实据性与企业权益资本成本： 基于文本挖掘的方法

师展,樊重俊

(上海理工大学管理学院,上海 200082)

摘要:我国已全面实施以信息披露为核心的注册制改革,实据性作为信息真实性和完整性的体现,有助于提高企业信息披露质量,从而提高资本市场资源配置效率。以2016—2021年我国沪深A股上市公司为样本,采用文本挖掘方法对管理层讨论与分析(MD&A)文本实据性进行度量,探究MD&A文本实据性对权益资本成本的影响。研究发现:MD&A文本实据性程度越高,权益资本成本越低,且两者关系在文本复杂性较低、信息中介跟踪较低以及外部治理环境较好的样本组中更加显著。研究结论丰富和拓展了权益资本成本的影响因素及文本信息增量价值的相关研究,为监管部门构建系统化的文本信息披露制度提供了新的方向。

关键词:管理层讨论与分析(MD&A);文本实据性;权益资本成本;文本挖掘

中图分类号:F272.3 **文献标识码:**A **文章编号:**1005-0566(2024)07-0169-10

MD&A textual evidentiality and the cost of equity capital: A text mining based approach

SHI Zhan, FAN Chongjun

(Management School, University of Shanghai for Science and Technology, Shanghai 200082, China)

Abstract: China has fully implemented the registration system reform centred on information disclosure, and textual evidentiality, as an embodiment of information authenticity and completeness, helps to improve the quality of corporate information disclosure, thus enhancing the efficiency of resource allocation in the capital market. Taking China's A-share listed companies in Shanghai and Shenzhen from 2016 to 2021 as a sample, text mining method is used to measure the textual evidentiality of management discussion and analysis (MD&A), and the impact of MD&A textual evidentiality on the cost of equity capital is explored. The study finds that the higher the degree of MD&A textual evidentiality, the lower the cost of equity capital, and the relationship is more significant in the sample group with lower textual complexity, lower information intermediary tracking, and better external governance environment. The findings enrich and expand the research related to the influencing factors of the cost of equity capital and the incremental value of textual information, and provide a new direction for regulators to build a systematic textual information disclosure system.

Key words: management discussion and analysis(MD&A); textual evidentiality; cost of equity capital; text mining

收稿日期:2024-04-05 修回日期:2024-05-19

基金项目:2020 教育部哲学社会科学重大课题攻关项目“工业互联网发展中新型生产关系研究”(20JZD010);国家自然科学基金青年项目“基于在线评论情感分析的社交媒体用户推荐”(71601119)。

作者简介:师展(1989—),女,山东德州人,上海理工大学管理学院博士生,研究方向为文本分析、信息披露。通信作者:樊重俊。

在面对经济发展受到需求萎缩、供给受挫、预期低迷等多重压力的情况下,企业所面临的不确定性不断上升,从而进一步加剧了企业融资的难度,如何解决企业融资难和融资贵的问题从而降低企业运营成本,成为推动企业高质量发展的重要策略。股权融资是企业重要的融资方式之一,“十四五”规划提出“提高直接融资,特别是股权融资比重”,而权益资本成本的高低是衡量市场资源配置效率的重要指标,其高低受多方面因素的影响,由于我国股市系统性风险和市场波动性的存在,投资者更加敏感于企业对外披露的信息质量和透明度的高低。已有研究表明,信息披露质量可以影响投资者决策,减轻其对不确定性的担忧,从而提高资本市场运行效率^[1-2]。

管理层讨论与分析文本(Management discussion and analysis, MD&A)是上市公司年报的重要组成部分,包括了对公司过去业绩的回顾和对公司未来发展的展望,这使得 MD&A 成为理解和评估上市公司综合经营状况的关键文档,对于投资者决策具有重要的指导作用。以往研究多聚焦在 MD&A 提供的数字信息对投资者决策的影响上,随着自然语言处理技术的发展,学者发现 MD&A 文本信息包含丰富的增量信息,可以帮助投资者预测企业未来业绩和股票市场表现^[3]。已有文献主要从文本可读性和语调等文本语言特征出发研究其对企业资本市场表现的影响,但除了语调和可读性外,MD&A 文本实据性特征,即说话者在陈述一个命题时,说明信息来源并承诺其真实性,也是反映信息披露质量的重要特征之一。那么,这一特征会不会对企业融资成本产生影响呢?基于此,本文从 MD&A 文本信息披露内容角度出发,重点研究 MD&A 文本实据性特征对企业权益资本成本的影响及其作用机制,以期对资本市场稳健运行提供相应的政策建议。

本文可能的贡献包括如下几点。第一,丰富和拓展了企业权益资本成本的影响因素研究。近年来,随着自然语言处理技术的发展,学者们开始关注企业对外披露报告的文本信息对权益资本成本的影响,但多聚焦在文本语言特征即可

读性和语调方面,而并未关注文本内容特征如实据性方面。第二,丰富了文本信息增量价值的相关研究。本文通过研究年报文本实据性特征对权益资本成本的影响,验证了 MD&A 文本实据性特征具有信息“沟通效应”,为后来学者对于文本信息披露质量的研究提供了新的方向。第三,构建了基于年报 MD&A 文本的实据性度量模型。借鉴段钊等^[4]和郭亚军等^[5]的研究,利用词性标注和非线性拉开档次法构建了年报 MD&A 文本实据性模型。

一、理论基础与研究假设

我国资本市场目前正在经历一场深刻的改革,这场改革的核心是建立以信息披露为基础的注册制度^[6]。这一改革意味着强调更加透明和真实的信息披露,这不仅要求上市公司更加主动地披露财务信息,而且需要重点提高信息披露的全面性和准确性。管理层讨论与分析文本实据性反应了企业披露其经营情况的有关证据,因此本文认为是具有“沟通效应”的,详实的信息披露不仅是对信息披露全面性要求的回应,也是对企业在市场中更加负责任和透明经营理念的倡导,会对投资者决策及风险感知产生重要影响,从而影响企业权益资本成本。文本信息实据性特征对企业权益资本成本的影响主要通过如下 3 个机制实现。

首先,信息实据性会影响投资者有限注意力。根据 Broadbent^[7]和 Kahneman^[8]的理论观点,投资者的有限注意力在很大程度上塑造了他们的注意力分配方式,直接影响了他们对信息的感知结果。Broadbent^[7]提出的选择性注意力理论强调,由于认知资源的有限性,在处理信息时,个体往往表现出一种选择性关注的倾向,这种选择性关注特定信息的行为会影响个体的决策。Kahneman^[8]的“注意力经济学”理论则进一步揭示了投资者在信息处理中的局限性,指出由于认知过程的有限性,人们在决策中会依赖启发式和简化及明确的信息。信息披露的属性,尤其是详实具体的信息,可能通过塑造投资者有限注意力的分配方式,从而显著改变他们对信息的理解和应用。在有限注意

力的框架下,信息的实据性可以被视为一种筛选机制,引导投资者将更多关注力集中在对其认为最有价值的信息上,这会影响投资者对市场状况的解读及风险感知,从而影响投资者决策及企业在资本市场上的定价。详实具体的信息会在投资者的有限注意力分配中占据主导地位,当企业能够提供实据性较高的信息披露时,投资者更容易理解企业的经营状况、风险因素以及未来的盈利潜力,从而有助于降低投资者的风险感知,因此会对企业的权益资本成本产生影响。

其次,信息实据性可以降低信息不对称程度,提高投资者对信息提取的效率。美国证监会(SEC)要求公司的信息披露应具体到公司的独特情况,避免过于泛泛。换句话说,企业披露的信息越详实可证,投资者感知到的信息真实性越高。如“我们与腾讯合作了一个耗资2亿的游戏项目”比“我们与相关互联网企业合作了游戏项目”更具披露公司相关性,前者提供了明确的、特定的信息,使读者能够清楚的了解信息的详细内容,有利于读者快速提取有用信息,而后者过于泛泛,于读者来说没有太大的信息价值。缺乏实据性的信息可能导致投资者在分析企业状况、制定投资策略时面临更大的困扰,从而不利于投资者进行理性决策。Heinle 等^[9]提出了风险披露定价效应模型,这一模型的前提是投资者对公司未来的现金流大小是不确定的,而公司通过披露信息来向市场传递一个关于其现金流方差的信号,但这个信号被认为是“不完美”的,研究发现,市场对于公司现金流方差的不确定性进行了定价,但具有更多实体元素的信号被赋予了更大的权重,这意味着市场更倾向于重视那些信息披露实据性较高的公司,因为这有助于降低投资者对公司未来现金流方差的不确定性,从而使他们更有信心作出投资决策。

最后,实据性越高,越能体现信息的真实性。DePaulo^[10]总结了自我表征观点(self-presentational perspective)的5类欺诈线索,认为相对于说真话的人,说谎者的言语表达会尽量减少对具体细节和信息的描述,当人们说谎时,通常会选择避免过多

涉及具体细节,以降低被揭穿的风险,他们可能会使用模糊的措辞、概括性的陈述,或者故意回避相关具体细节以模糊事实,从而保持言辞的一致性,这种言辞模式旨在混淆听众,使得检验真实性变得更加困难。相较之下,说真话的人往往更愿意提供更为详实和具体的描述,因为他们的目标是准确真实的传达信息,而非隐藏真相。真实的故事陈述往往包含丰富的实体信息^[11],这些信息以详实的描绘为基础,能够提高读者的信息可靠性感知。实体信息的涵盖范围广泛,既包括具象的人物、处所、地点等,也涉及抽象的情感、思想和理念等,当企业在其年报中提供实据性较高的信息时,投资者通常能够更全面地了解公司的运营状况、财务状况以及未来发展计划,这种信息透明度有助于投资者更准确地评估企业的健康状况和风险因素,从而有利于提高投资者风险评估的准确性,降低投资者寻求价格补偿的空间,减少公司的权益融资成本。根据以上分析,本文提出假设:MD&A 文本信息实据性程度越高,企业权益资本成本越低。

二、研究设计

(一)样本选择与数据来源

本文通过 python 网络爬虫技术,在中国证监会指定信息披露网站:巨潮资讯网上爬取 2016—2021 年我国沪深 A 股上市公司的年度财务报告为研究样本,公司相关财务数据来源于 CSMAR 数据库。为保证数据的合理性,我们剔除了 ST 公司、金融类公司,剔除了有缺失值的样本公司,同时对所有变量进行了缩尾处理。经过以上操作,最终我们获得了 6 142 个公司的年度财务报告样本。

(二)变量定义

1. 解释变量:MD&A 文本实据性(*Evidentiality score, ES*)。

段钊等^[4]结合以往研究,并根据文本描述的“5W1H”原则,即何时(when)、何地(where)、何人(who)、何事(what)、何故(why)、如何(how),提出了用于衡量文本实据性的时间、地点、主语、行为动词和数量词五要素,且经专家评估后,提出了包括五要素组合、四要素组合和三要素组合的 16 个有

效的实据性模式,详情如表 1 所示。其中,“时间”和“地点”分别代表事件发生的时空信息;“主语”代表本事件的施行者,即行为主体。“行为动词”一般指事件主体的具体行动或动作;“数量词”指量词如“10 个”“30 万元”“百分之几”等。

表 1 有效实据性模式表

序号	评价指标	有效的实据性模式	模式
1	五要素组合	时间地点主语动词数量词	$(tdg) + (ns) + (nt) + (v) + (mx \cdot mf \cdot mg)$
2	四要素组合	时间地点主语动词	$(tdg) + (ns) + (nt) + (v)$
3		时间地点主语数量词	$(tdg) + (ns) + (nt) + (mx \cdot mf \cdot mg)$
4		时间地点动词数量词	$(tdg) + (ns) + (v) + (mx \cdot mf \cdot mg)$
5		时间主语动词数量词	$(tdg) + (nt) + (v) + (mx \cdot mf \cdot mg)$
6		地点主语动词数量词	$(ns) + (nt) + (v) + (mx \cdot mf \cdot mg)$
7	三要素组合	时间地点主语	$(dg) + (ns) + (nt)$
8		时间地点动词	$(tdg) + (ns) + (v)$
9		时间地点数量词	$(tdg) + (ns) + (mx \cdot mf \cdot mg)$
10		时间主语动词	$(tdg) + (nt) + (v)$
11		时间主语数量词	$(tdg) + (nt) + (mx \cdot mf \cdot mg)$
12		时间动词数量词	$(tdg) + (v) + (mx \cdot mf \cdot mg)$
13		地点主语动词	$(ns) + (nt) + (v)$
14		地点主语数量词	$(ns) + (nt) + (mx \cdot mf \cdot mg)$
15		地点动词数量词	$(ns) + (v) + (mx \cdot mf \cdot mg)$
16		主语动词数量词	$(nt) + (v) + (mx \cdot mf \cdot mg)$

注:表格来源:段钊,涂秋阳,胡春等.上市公司社会责任信息披露中的信任修复[J].经济管理,2023,45(8):145-166.

段钊等^[4]的研究对象为企业社会责任报告,本文认为年报文本 MD&A 与企业社会责任报告均为金融类文本,且在编写过程中,管理层均具有较大的自由裁量权,因此本文认为将段钊等^[4]衡量社会责任报告实据性的方法用于 MD&A 文本实据性模型的构建和评估是合理的。对于不同实据性模式要素的提取,本文使用 python 语言借助结巴词性标注功能对每个句子的词进行词性标注,然后计算每个年报 MD&A 文本中 3 种要素模式组合所占的频率。具体实施步骤如图 1 所示。

(1) 数据预处理

选择巨潮资讯网作为文本爬虫的接口,从巨潮资讯网爬取 2016—2021 年度上市公司财务报告 21 603 份。首先,对爬取的年报文本进行格式转换,将 pdf 格式的年报文本转换为 txt 格式;然后,采用正则匹配方法从年报中将 MD&A 部分的文本信息提取出来;最后,进行 MD&A 文本预处理,去除文本中的空白、乱码、空格等,仅保留文字信息,以句号为界划分句子,将每个句子分行存储,并利用结巴分词模块对每个句子进行分词,利用停用词表过滤句子中无意义的“的”“了”“呢”等停用词。

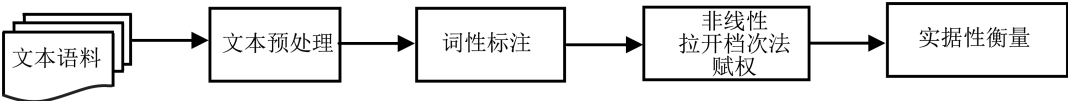


图 1 MD&A 文本实据性衡量步骤

(2) 基于统计模型的词性标注

词性(par-of-speech, Pos)是语言分析和理解的基础,它帮助我们了解单词在句子中扮演的角色以及词与词之间的关系。不同的词在不同的语境下具有不同的词性,词性标注即需要给不同语境句子下的词标注上词性。目前常用的词性标注方法有基于规则的统计方法、基于统计模型的标注方法以及两种方法相结合的词性标注方法。结巴采用统计模型的标注方法,词性标注集使用的是北大计算所词性标注集,从而实现词性标注,结巴的词性标注表部分示例如表 2 所示。

(3) 计算每种模式的句子频率。根据段钊等^[4]构建的有效实据性模式,使用 Python 程序编

表 2 结巴词性标注表示例(部分)

词性代码	词性	举例	数量
a	形容词	高明 尖 诚 粗陋 冗杂 丰盛 顽皮 很贵 挺好用...	4 306
ad	副形词	努目 完全 努力 切面 严实 慌忙 明确 仓惶 详细...	110
ag	形语素	详 笃 睦 奇 洋 裸 渺 忤 虐 黷 怠 峻 恣 鄙 秀...	46
an	名形词	麻生 猥琐 腐生 困苦 危难 负疚 刚愎 危险 悲苦...	40
b	区别词	劣等 洲际性 超常规 同一性 年级 非农业 二合一...	1 363
c	连词	再者说 倘 只此 或曰 以外 换句话说 虽是 除非...	504
d	副词	幸免 四顾 绝对 急速 特约 从早 务须 逐行 挨边...	2 422
df	不要	不要	1
dg	副语素	俱 辄	2
e	叹词	好哟 嘎 天呀 哎 哇呀 啊哈 嗟 诶 喏 呜呼 哇塞...	34

写代码,计算3种模式的句子在全文总句中所占的频率,即 X_n = 不同要素模式组合的句数/MD&A 文本总句数($n=3,4,5$)。

(4)对3种模式进行赋权。段钊等^[4]在对社会责任报告实据性模型进行不同要素模式间的赋权时,采用了郭亚军等^[5]提出的非线性拉开档次法。本文预借鉴这一方法对年报文本的3种要素模式进行赋权,运用python软件进行编码计算,得出五要素组合、四要素组合和三要素组合的权重系数分别为0.229,0.321和0.450,其中非线性因子为 X_3 。继而计算得到各上市公司年报 MD&A 文本的实据性程度。

2. 被解释变量:权益资本成本(*cost of equity capital, CEC*)

已有研究发现,事前权益资本成本模型相对于事后模型在预测准确性和稳定性方面具有优势,在事前权益资本成本模型中,PEG模型和MPEG模型综合考虑了市场估值、盈利增长和企业特定风险等因素,提供了更全面和准确的资本成本估计^[12],而OJN模型在假设方面更加严谨和科学^[13]。因此,综合以上3种模型的优势,本文借鉴冯来强等^[14]的研究,采用PEG、MPEG和OJN3种模型计算结果的均值来代表企业权益资本成本CEC。3种模型的计算公式如下:

$$CEC_{PEG} = \sqrt{\frac{EPS_{t+2} - EPS_{t+1}}{P_0}} \quad (1)$$

$$CEC_{MPEG} = \sqrt{\left(\frac{DPS_{t+1}}{P_0}\right)^2 + \frac{EPS_{t+2} - EPS_{t+1}}{P_0}} \quad (2)$$

$$CEC_{OJN} = A + \sqrt{A^2 + \frac{EPS_{t+1} \times (g - (r - 1))}{P_0}} \quad (3)$$

式(1)、式(2)和式(3)分别为PEG、MPEG和OJN模型的计算公式,其中 P_0 是年初的股票收盘价, EPS_{t+2} 和 EPS_{t+1} 分别为分析师预测的 $t+1$ 期和 $t+2$ 期的每股收益, $DPS_{t+1} = a \times EPS_{t+1}$ 为 $t+1$ 期的预测股利, DPS_{t+1} = 过去三年的平均股利支付率 $\times EPS_{t+1}$, $A = \frac{1}{2} \left((r - 1) + \frac{DPS_{t+1}}{P_0} \right)$, g =

$$\frac{(EPS_{t+2} - EPS_{t+1})}{P_0}, g \text{ 和 } r - 1 \text{ 分别表示企业的短期盈余增长率和长期增长率。}$$

用以上3种模型计算所得结果的均值来衡量企业权益资本成本,即

$$CEC = \frac{CEC_{PEG} + CEC_{MPEG} + CEC_{OJN}}{3}。$$

3. 控制变量

考虑到权益资本成本的影响因素,控制变量包括:公司规模(*size*)、财务杠杆(*lev*)、总资产净利润率(*ROA*)、营业收入增长率(*growth*)、现金流比率(*cashflow*)、董事会规模(*board*)、账面市值比(*Btm*)、系统性风险(*Beta*)、总资产周转率(*ATO*)。同时控制了行业和年份。

(三)模型构建

本文构建如下OLS模型,考虑到企业上一年的文本实据性会对企业下一年的权益资本成本产生影响,因此,对权益资本成本做滞后一期处理。

$$CEC_{i,t} = \beta_0 + \beta_1 \times ES_{i,t-1} + \sum \beta_j controls_{i,t-1} + \sum Year + \sum Industry + \varepsilon \quad (4)$$

式(4)中,CEC为权益资本成本,ES为MD&A文本实据性程度,Control为控制变量;Year和Industry分别为年份和行业,均为虚拟变量。

三、实证检验及分析

(一)描述性统计分析

表3展示了主要变量的描述性统计结果。解释变量实据性程度(ES)的最大值和最小值分别为0.397和0.302,均值为0.347,表明研究样本的实据性程度普遍不高。被解释变量企业权益资本成

表3 描述性统计结果

变量	N	Mean	SD	Min	Median	Max
CEC	6 142	0.126	0.042	0.05	0.12	0.277
ES	6 142	0.347	0.02	0.302	0.346	0.397
LIQ	6 142	-0.026	0.023	-0.12	-0.019	-0.001
Size	6 142	22.801	1.338	20.493	22.587	26.847
Lev	6 142	0.423	0.188	0.072	0.418	0.845
ROA	6 142	0.065	0.049	-0.047	0.055	0.232
Growth	6 142	0.236	0.377	-0.361	0.156	2.446
Cashflow	6 142	0.066	0.065	-0.108	0.063	0.249
Board	6 142	2.129	0.197	1.609	2.197	2.708
BM	6 142	1.128	1.367	0.098	0.669	8.082
BETA	6 142	1.212	0.318	0.422	1.204	2.117
ATO	6 142	0.679	0.414	0.122	0.585	2.48

本(CEC)的最大值和最小值分别为 0.277 和 0.05,均值为 0.126,标准差为 0.042,说明我国企业之间权益资本成本存在较大差异。各个控制变量的最大值和最小值相差较大,说明上市公司之间财务状况、治理水平和经营风险等存在较大的差异。以上变量的描述性统计与现有研究基本保持一致。

(二)基准回归结果分析

表 4 为实据性对企业权益资本成本的实证结果。其中,表 4 第(1)列仅加入了自变量和因变量进行回归,结果显示,MD&A 文本实据性和企业权益资本成本的相关系数为 -0.284,在 1% 的水平下显著负相关,本文假设 H1 得到验证;进一步地,在第(2)列控制了年份和行业,结果如表 4 所示,实据性与企业权益资本成本依然在 1% 的水平下显著负相关;第(3)列加入控制变量,结果仍然不变,且模型的 R^2 显著增加。以上均证实了本文提出的假设。

表 4 MD&A 文本实据性对企业权益资本成本的实证结果

变量	(1)	(2)	(3)
ES	-0.284 *** (-10.335)	-0.244 *** (-9.385)	-0.214 *** (-8.313)
Size	-	-	-0.003 *** (-5.006)
Lev	-	-	0.014 *** (3.522)
ROA	-	-	0.051 *** (3.679)
Growth	-	-	0.005 *** (3.827)
Cashflow	-	-	-0.033 *** (-3.548)
Board	-	-	-0.008 *** (-3.165)
BM	-	-	0.008 *** (12.547)
BETA	-	-	-0.003 (-1.433)
ATO	-	-	0.006 *** (3.994)
常数项	0.224 *** (23.520)	0.192 *** (17.727)	0.256 *** (15.281)
控制年份	否	是	是
控制行业	否	是	是
样本量	6 142	6 142	6 142
调整后的 R^2	0.017	0.144	0.183

注:括号内为 t 检验值。*、**、*** 分别表示在 $p < 0.01$ 、 $p < 0.05$ 、 $p < 0.10$ 有统计学意义。下同。

(三)稳健性与内生性检验

第一,替换因变量。分别采用 PEG 模型、MPEG 模型和 OJN 模型计算权益资本成本,并分别对 MD&A 文本实据性进行回归。实验结果如

表 5 Panel A 所示,用 3 种方式计算的企业权益资本成本的数据对 ES 的回归结果均在 1% 的水平下显著负相关,原假设依然成立。

第二,工具变量回归。使用工具变量法进行两阶段最小二乘法回归(2SLS)解决互为因果的内生性问题。使用同年度同行业其他公司 MD&A 文本的实据性得分均值(ES_mean)作为工具变量。实验结果如表 5 Panel B 所示,第一阶段回归结果如列(1)所示,显示工具变量与实据性的相关性在 1% 的水平下显著相关,证明了工具变量的有效性,weak IV 检验说明不存在弱工具变量风险。第二阶段回归结果如列(2)所示,回归系数依然在 1% 的水平下显著为负,研究结论稳健。

第三,Heckman 两阶段法。使用 Heckman 两阶段法解决样本选择偏误,通过考虑潜在样本的选择行为来修正 OLS 估计结果的偏误。构建 ES 的虚拟变量 ES_dum ,并将第一阶段估计的 IMR 放入第二阶段作为控制变量进行回归,结果如表 5 Panel C 列(2)所示,MD&A 文本实据性与企业权益资本成本的回归系数仍然在 1% 的水平下显著为负,原假设依然成立。

第四,PSM 倾向得分匹配。采用倾向得分匹配法(PSM)解决自选择问题。按照实据性中位数构造实验组和对照组,选取可能影响实据性的变量:公司规模($size$)、财务杠杆(lev)、净资产收益率(ROE)、现金流比率($cashflow$)、两职合一($dual$)、第一大股东持股比例($Top1$)、账面市值比(Btm)、托宾 Q($TobinQ$)、上市年限($listage$)、是否四大($Big4$)作为控制变量构建 Logit 模型,采用 1:1 最近邻法进行配对,最终匹配成功的样本有 5 395 个,匹配后的 p 值均大于 0.1,使用成功匹配的样本重新进行回归,实验结果如表 5 Panel D 所示,回归系数依然显著为负。

四、进一步研究

(一)机制检验:基于缓解信息不对称程度的视角

根据优序融资理论,信息不对称会导致投资者面临较大的不确定性^[15],从而企业融资成本增加。如果上市公司能够披露实据性较高的文本信

表 5 稳健性和内生性检验

Panel A: 替换因变量			
变量	(1) <i>CEC_{MPEG}</i>	(2) <i>CEC_{OJN}</i>	(3) <i>CEC_{PEG}</i>
<i>ES</i>	-0.234 *** (-5.905)	-0.180 *** (-5.858)	-0.134 *** (-5.129)
控制变量	是	是	是
常数项	0.244 *** (8.730)	0.277 *** (12.875)	0.221 *** (11.902)
控制年份	是	是	是
控制行业	是	是	是
样本量	6 142	6 142	6 142
调整后的 <i>R</i> ²	0.210	0.172	0.230
Panel B: 工具变量回归			
变量	(1) 第一阶段	(2) 第二阶段	
	<i>ES</i>	<i>CEC</i>	
<i>IV</i>	0.963 *** (18.020)	-	
<i>ES</i>	-	-0.505 *** (-3.750)	
控制变量	是	是	
常数项	-0.020 (-0.964)	0.348 *** (7.295)	
控制年份	是	是	
控制行业	是	是	
样本量	6 142	6 142	
调整后的 <i>R</i> ²	0.104	0.162	
弱工具变量检验	324.703	-	
Panel C: Heckman 两阶段法			
变量	(1) 第一阶段	(2) 第二阶段	
	<i>ES_dum</i>	<i>CEC</i>	
<i>ES</i>	-	-0.231 *** (-7.475)	
<i>IMR</i>	-	-0.100 *** (-6.708)	
控制变量	是	是	
<i>Indep</i>	-0.193 (-0.409)	-	
<i>Top1</i>	-0.474 ** (-2.398)	-	
<i>Balance1</i>	-0.014 (-0.142)	-	
<i>Big4</i>	-0.042 (-0.499)	-	
常数项	-1.608 ** (-2.215)	0.446 *** (12.215)	
控制年份	是	是	
控制行业	是	是	
样本量	6 142	6 142	
调整后的 <i>R</i> ²	0.016	0.189	
Panel D: PSM 倾向得分匹配			
变量	<i>CEC</i>		
<i>ES</i>	-0.203 *** (-6.247)		
控制变量	是		
常数项	0.245 *** (10.539)		
控制年份	是		
控制行业	是		
样本量	5 395		
调整后的 <i>R</i> ²	0.183		

息,就可以提高投资者的信息获取能力,从而帮助投资者更准确地评估公司的内在价值和未来成长潜力,提高投资者投资意愿,降低融资成本。据此,借鉴熊家财等^[16]和 Schoenfeld^[17]的做法,采用 *Amihud* 非流动性比率来度量信息不对称程度,计算方法如模型(5)、模型(6)所示:

$$Amihud_{i,t} = \frac{1}{D_{i,t}} \sum_{d=1}^{D_{i,t}} \frac{|R_{i,t,d}|}{V_{i,t,d}} \times 100 \quad (5)$$

式(5)中,*R* 为股票 *i* 在第 *t* 年第 *d* 天的收益率;*V* 为股票 *i* 在第 *t* 年第 *d* 天的交易金额;*D* 为股票 *i* 在 *t* 年的交易天数。由于 *Amihud* 非流动性指标为反向指标,且流动性分布不均匀,偏度较大,因此采用王运陈等^[18]的做法,对其加一后取自然对数,再取相反数,具体公式如(9),*LIQ_{i,t}* 的值越大,股票流动性越好,反之则越差。

$$LIQ_{i,t} = -\ln(1 + Amihud_{i,t}) \quad (6)$$

本文借鉴温忠麟等^[19]的中介效应研究来检验信息不对称程度是否在文本实据性和权益资本成本之间发挥了中介效应。实验结果如表 6 所示,第(1)列显示,MD&A 文本实据性与权益资本成本在 1% 的水平下依然显著负相关。第(2)列显示,MD&A 文本实据性与股票流动性在 5% 的水平下显著正相关,同时第(3)列结果显示股票流动性与权益资本成本在 5% 的水平下显著负相关,且 MD&A 文本实据性与权益资本成本在 1% 的水平下显著负相关,说明股票流动性在 MD&A 文本实据性和权益资本成本之间发挥了部分中介效应。

表 6 中介效应检验

变量	(1) <i>CEC</i>	(2) <i>LIQ</i>	(3) <i>CEC</i>
<i>ES</i>	-0.214 *** (-6.880)	0.033 ** (2.166)	-0.212 *** (-6.801)
<i>LIQ</i>	-	-	-0.082 ** (-2.295)
控制变量	是	是	是
常数项	0.256 *** (11.519)	-0.342 *** (-33.630)	0.227 *** (9.357)
控制年份	是	是	是
控制行业	是	是	是
样本量	6 142	6 142	6 142
调整后的 <i>R</i> ²	0.179	0.465	0.180

(二) 分组检验

1. 文本特征之间的交互作用检验

MD&A 文本复杂性对投资者决策具有重要影

响。首先,复杂的语言和专业术语增加了投资者阅读和理解信息的难度,使得他们在获取关键信息时感到困扰。其次,复杂的文本会导致信息的误解或遗漏,投资者无法准确理解年报中的语言表达,从而影响其投资决策,特别是对于那些缺乏专业知识的散户投资者来说更为困难。因此,对于本文的研究而言,复杂的文本描述很可能会降低 MD&A 文本实据性的信息含量,从而削弱文本实据性对企业权益资本成本的缓解作用。基于此,借鉴迷雾指数对复杂词的衡量方法,以国家汉语水平考试委员会颁布的《汉语水平词汇与汉字等级大纲》中的 3 个及以上字的丙级词和丁级词作为复杂词^[20],并以此计算 MD&A 文本中的复杂词词频。

本文以每年度各行业 MD&A 复杂性的中位数为界限,将样本分为高复杂性和低复杂性两组并分别进行回归。回归结果见表 7 所示,在不同样本组,MD&A 文本实据性的回归结果均显著为负,与基准回归结果一致。交乘项($complex \times ES$)的系数在 1% 的水平下显著为正,表明相较于复杂性较高的样本公司,复杂性较低的样本公司其 MD&A 文本实据性对权益资本成本的影响更加显著。分组检验表明,虽然在文本复杂性较低和较高的两组中,MD&A 文本实据性对企业权益资本成本的影响均显著(分别在 1% 和 10% 的水平下显著负相关),但费舍尔组间差异检验表明两组间的系数存在显著差异($p = 0.000$),进一步证实了在文本复杂性较低组,两者关系更加显著,实验结果与本文的分析一致。

表 7 基于 MD&A 文本复杂性的分组

变量	(1) 文本复杂性低	(2) 文本复杂性高	(3) 全样本
ES	-0.327 *** (-7.772)	-0.078 * (-1.861)	-0.215 *** (-6.851)
Complex × ES	-	-	0.436 *** (3.778)
Complex	-	-	0.006 ** (2.191)
控制变量	是	是	是
常数项	0.284 *** (9.083)	0.223 *** (7.581)	0.236 *** (9.956)
控制年份	是	是	是
控制行业	是	是	是
样本量	3 225	2 917	6 142
调整后的 R ²	0.173	0.203	0.183
p - Value	0.000		-

2. 信息中介的替代效应检验

媒体和分析师扮演着传递和解释金融信息的关键角色。媒体通过报道新闻、分析市场趋势和提供专业评论,将复杂的金融信息转化为更易理解的形式,为投资者提供了决策依据和市场见解。分析师则专注于深入研究公司和行业,发布研究报告、评级和投资建议,帮助投资者评估风险和收益,并做出理性的投资决策。信息中介的作用在于传播信息,这使得投资者有更多的渠道了解到公司的信息,从而有效辅助了自身决策,因此降低了文本实据性对权益资本成本的作用,因此本文猜测,信息中介和文本实据性对企业权益资本成本的交互影响可能存在替代效应。

对于媒体关注度,本文以 $\ln(\text{网络和报刊媒体报道数量} + 1)$ 来进行衡量;分析师关注度使用 $\ln(\text{一年内跟踪分析过某一公司的分析师或者分析师团队数量} + 1)$ 来进行衡量。实验结果如表 8 所示,在不同样本组,MD&A 文本实据性的回归结果均显著为负,与基准回归结果一致。交乘项($media \times ES$ 和 $analyst \times ES$)的系数在 1% 的水平下均显著为正,表明相较于媒体关注度及分析师关注度较高的样本公司,拥有较低关注度的样本公司其 MD&A 文本实据性对权益资本成本的影响更加显著。分组检验表明,虽然在不同分组中,MD&A 文本实据性对企业权益资本成本的影响均显著,但费舍尔组间差异检验表明两组间的系数存在显著差异(p 值均为 0.000),进一步证实了在信息中介关注度较低的组,两者关系更加显著,信息中介和文本实据性对企业权益资本成本的交互影响存在替代效应,实验结果与本文的分析一致。

3. 外部环境因素的影响

外部治理因素会影响企业的信息披露质量^[21]。市场化程度较低的地区缺乏外部监督和透明度,企业有更大的空间进行信息披露管理,如掩盖负面消息,甚至存在不诚信的行为。而在市场化程度较高的地区,企业管理层通常更加重视遵守法律法规和市场规范,因为不良的信息披露会受到严厉的处罚,影响企业的声誉和市场地位^[22]。因此,在市场化程度较高的地区,外部监管对管理

表 8 基于信息中介关注度的分组

变量	(1) 媒体关注低	(2) 媒体关注高	(3) 全样本	(4) 分析师关注低	(5) 分析师关注高	(6) 全样本
<i>ES</i>	-0.295 *** (-7.198)	-0.114 *** (-2.596)	-0.204 *** (-6.558)	-0.309 *** (-7.568)	-0.085 ** (-2.035)	-0.201 *** (-6.620)
<i>Media</i> × <i>ES</i>	-	-	0.161 *** (5.633)	-	-	-
<i>Media</i>	-	-	-0.001 * (-1.780)	-	-	-
<i>Analyst</i> × <i>ES</i>	-	-	-	-	-	0.180 *** (5.257)
<i>Analyst</i>	-	-	-	-	-	-0.001 * (-1.791)
控制变量	是	是	是	是	是	是
常数项	0.232 *** (6.544)	0.253 *** (8.786)	0.246 *** (11.104)	0.244 *** (7.778)	0.228 *** (7.931)	0.240 *** (10.649)
控制年份	是	是	是	是	是	是
控制行业	是	是	是	是	是	是
样本量	3 185	2 957	6 142	3 283	2 859	6 142
调整后的 R^2	0.166	0.205	0.187	0.154	0.243	0.185
p -Value	0.000		-	0.000		-

层的信息操纵行为产生了明显的威慑作用,企业信息更加透明,投资者可以更好的从企业对外披露的报告中获取有价值的信息,从而降低了企业权益资本成本^[21]。本文采用《中国分省份市场化指数报告(2021)》中各省份市场化指数来衡量市场化进程,并依据样本企业所在地区市场化指数的中位数对样本进行分组,大于这个中位数的样本划分为市场化程度较高组,否则为市场化程度较低组。实验结果如表 9 所示,在不同样本组,MD&A 文本实据性的回归结果均显著为负,与基准回归结果一致。交乘项(*market* × *ES*)的系数在 1% 的水平下显著为负,且两者关系在市场化程度较高的样本组中显著,而在市场化程度较低的样本组中不显著,表明相较于市场化程度较低的样本公司,市场化程度较高的样本公司其 MD&A 文

本实据性对权益资本成本的影响更加显著,实验结果与本文的分析一致。

五、结论与建议

本文以我国 2016—2021 年沪深 A 股上市公司 MD&A 文本为研究对象,实证检验了 MD&A 文本实据性与企业权益资本成本之间的关系及其作用机制。研究结果发现:第一,MD&A 文本实据性与企业权益资本成本显著负相关,MD&A 文本实据性越高,企业权益资本成本越低,两者关系在经过一系列稳健性和内生性检验后依然成立;第二,信息不对称程度在 MD&A 文本实据性与企业权益资本成本的关系中发挥了部分中介作用,MD&A 文本实据性可以降低信息不对称程度,从而降低企业权益资本成本;第三,文本特征之间会产生交互作用,较高的文本复杂性会削弱实据性对企业权益资本成本的降低作用,而信息中介和文本实据性对企业权益资本成本的交互影响存在替代效应,且市场环境作为外部治理因素会影响文本实据性与权益资本成本之间的关系,在市场化程度较高组,MD&A 文本实据性对企业权益资本成本的影响更加显著。

基于以上研究结论,本文提出如下政策建议。第一,对于政府和监管部门。监管部门应尽快出台更加明确的 MD&A 披露办法,还包括 MD&A 文本语言特征方面的要求,也应包括 MD&A 文本内容方面的要求,提高 MD&A 文本信息的可靠性和

表 9 基于市场化程度的分组

变量	(1) 市场化程度低	(2) 市场化程度高	(3) 全样本
<i>ES</i>	-0.130 (-1.211)	-0.219 *** (-6.768)	-0.217 *** (-6.998)
<i>Market</i> × <i>ES</i>	-	-	-0.044 ** (-2.324)
<i>Market</i>	-	-	-0.001 * (-1.821)
控制变量	是	是	是
常数项	0.177 ** (2.493)	0.260 *** (10.989)	0.262 *** (11.575)
控制年份	是	是	是
控制行业	是	是	是
样本量	578	5564	6 142
调整后的 R^2	0.109	0.188	0.181

详实性,采取“强制性”和“格式化”的披露模式,对“如何披露、披露多少、披露到什么程度”等进行严格细致的规范和指引;设计文本信息披露指标,利用人工智能技术提取相关文本特征,审核文本信息披露的规范性和完整性,并就文本信息披露质量制定相关奖惩措施并付诸实践;政府需进一步优化市场环境,加强信息中介的外部治理作用,提高资本市场资源配置效率。第二,对于企业。应以信息披露真实性和完整性为指导原则,提高MD&A文本实据性,降低企业权益资本成本,提升企业价值和竞争力。同时,由于不同文本特征之间具有交互影响作用,企业应注重加强提升多维度文本特征质量,避免“顾此薄彼”。第三,对于投资者。投资者在解读年报时,可关注文本实据性较高的企业,其提供的报告真实可靠、证据充足,可帮助投资者预测企业未来经营情况,提高自身投资效率。同时,可借助媒体、分析师报道等信息中介渠道对企业进行了解、分析和评价,以弥补专业知识的缺陷,规避投资雷区,保护自身利益。

参考文献:

- [1] CHEN Hanwen, CHEN J Z, LOBO G J, et al. Effects of audit quality on earnings management and cost of equity capital: evidence from China[J]. Contemporary accounting research, 2011, 28(3): 892-925.
- [2] 王俊秋. 政治关联、盈余质量与权益资本成本[J]. 管理评论, 2013, 25(10): 80-90.
- [3] 王嘉鑫,陈今,史亚雅. 年报非财务信息的文本披露语言特征会影响股价崩盘风险吗?[J]. 北京工商大学学报(社会科学版), 2022, 37(3): 98-112.
- [4] 段钊,涂秋阳,胡春,等. 上市公司社会责任信息披露中的信任修复[J]. 经济管理, 2023, 45(8): 145-166.
- [5] 郭亚军,阮泰学,宫诚举. 基于主客观信息综合判断的非线性拉开档次法[J]. 运筹与管理, 2017, 26(6): 149-154.
- [6] 刘一寒,范慧敏,任晨煜. MD&A 纵向文本相似度与分析师盈余预测准确性[J]. 北京工商大学学报(社会科学版), 2024, 39(2): 71-84.
- [7] BROADBENT D E. A mechanical model for human attention and immediate memory[J]. Psychological review, 1957, 64(3): 205-215.
- [8] KAHNEMAN D. Attention and effort[M]. Englewood Cliffs, NJ: Prentice-Hall, 1973.
- [9] HEINLE M S, SMITH K C. A theory of risk disclosure[J]. Review of accounting studies, 2017, 22(4): 1459-1491.
- [10] DEPAULO B M. Nonverbal behavior and self-presentation[J]. Psychological Bulletin, 1992, 111(2): 203-243.
- [11] JOHNSON M K, RAYE C L. Reality monitoring[J]. Psychological review, 1981, 88(1): 67-85.
- [12] 毛新述,叶康涛,张嶝. 上市公司权益资本成本的测度与评价:基于我国证券市场的经验检验[J]. 会计研究, 2012(11): 12-22, 94.
- [13] 刘冰,方政. 公司内部治理机制与股权融资成本:股权性质差异条件下的影响因素分析[J]. 经济管理, 2011, 33(12): 135-140.
- [14] 冯来强,孔祥婷,曹慧娟. 董事高管责任保险与权益资本成本:来自信息质量渠道的实证研究证据[J]. 会计研究, 2017(11): 65-71, 97.
- [15] MYERS S C, MAJLUF N S. Corporate financing and investment decisions when firms have information that investors do not have[J]. Journal of financial economics, 1984, 13(2): 187-221.
- [16] 熊家财,苏冬蔚. 股票流动性与代理成本:基于随机前沿模型的实证研究[J]. 南开管理评论, 2016, 19(1): 84-96.
- [17] SCHOENFELD J. The effect of voluntary disclosure on stock liquidity: new evidence from index funds[J]. Journal of accounting and economics, 2017, 63(1): 51-74.
- [18] 王运陈,贺康,万丽梅,等. 年报可读性与股票流动性研究:基于文本挖掘的视角[J]. 证券市场导报, 2020(7): 61-71.
- [19] 温忠麟,叶宝娟. 中介效应分析:方法和模型发展[J]. 心理科学进展, 2014, 22(5): 731-745.
- [20] 左虹,朱勇. 中级欧美留学生汉语文本可读性公式研究[J]. 世界汉语教学, 2014, 28(2): 263-276.
- [21] 朱光,王纯熙. 年报可读性与内部人交易研究:基于文本挖掘视角[J]. 金融监管研究, 2022(4): 100-114.
- [22] 马嫣然,罗丽. 家族化管理、地区市场化程度与家族企业财务绩效[J]. 西南大学学报(社会科学版), 2022, 48(3): 110-120.

(本文责编:默 黎)