

doi: 10.19920/j.cnki.jmsc.2022.08.003

科创板注册制下的审核问询与 IPO 信息披露^①

——基于 LDA 主题模型的文本分析

俞红海, 范思妤, 吴良钰, 马质斌

(南京大学工程管理学院, 南京 210093)

摘要: “以信息披露为核心”是 IPO 注册制的本质要求。在科创板试点注册制背景下, 发行上市审核问询能否完善 IPO 信息披露内容、提升 IPO 信息披露质量, 是市场能否在资源配置中发挥决定性作用的关键。以 2019 年 7 月~2021 年 9 月科创板 341 家 IPO 公司为样本, 本文研究了注册制审核问询对 IPO 信息披露的作用。研究表明: 1) 在上市审核阶段, 信息披露程度较低、技术信息披露不够充分的招股书, 会收到更多的审核问询函。2) 注册制审核问询能提升招股书信息披露的程度, 问询前后两版招股说明书在文本长度、文本句长、文本可理解性、和文本定量信息上显著增加, 总体来说审核问询函对特定主题关注会导致招股书在对应主题信息披露的增加。3) 进一步分析表明, 审核问询函能降低机构投资者报价分歧度。文章研究结论有助于理解注册制改革背景下交易所审核问询发挥的监管作用, 同时也丰富了 IPO 信息披露影响因素相关研究。

关键词: 注册制; 审核问询; 信息披露; LDA 主题模型

中图分类号: F830 **文献标识码:** A **文章编号:** 1007-9807(2022)08-0045-18

0 引言

2018 年 11 月 5 日, 在首届中国国际进口博览会开幕式主旨演讲中, 习近平总书记表示, 将在上海证券交易所设立科创板并试点注册制。随后习总书记在中央政治局会议中提出: 科创板要真正落实以信息披露为核心的证券发行注册制。“以信息披露为核心”的要求, 是由注册制本质决定的。注册制下, 监管部门仅作基于发行人文本的形式审查而不做价值判断, 把企业合规性问题、成长性问题、未来前景问题等条件转化为严格的信息披露要求, 通过充分信息披露和风险揭示, 交给市场参与主体自行判断。既然要让市场自行作出判断和决策, 要让市场力量均衡博弈以充分发挥市场化定价作用, 就要确保发行人在发行上市审核的过程中真实、充分、准确、及时地将企业重要

事项向外界披露。

然而, 科创板上市审核信息披露现状仍与注册制“以信息披露为核心”的要求存在较大差距。上交所在就首轮问询及回复情况答记者问中, 对招股书信息披露质量用“五个不够”进行了总结, 认为目前科创板股票发行存在以下信息披露质量问题: 对科技创新相关事项披露不够充分; 对企业业务模式披露不够清晰; 对企业生产经营和技术风险揭示不够到位; 信息披露语言表述不够友好; 文件格式和内容安排不够规范。以上情况表明, 交易所在发行上市审核问询的过程中不仅关注招股书整体信息披露是否符合规范, 还会重点关注招股书在特定主题上的信息披露是否充分。那么, 发行人哪些信息披露质量问题会引发交易所问询? 交易所在审核问询中更加关注哪些主题的信息披

^① 收稿日期: 2021-06-28; 修订日期: 2022-01-04。

基金项目: 国家社会科学基金资助重大项目(19ZDA105)。

作者简介: 俞红海(1978—), 男, 浙江嘉兴人, 博士, 教授, 博士生导师。Email: hhyu@nju.edu.cn

露?此外,在科创板试点注册制的制度背景下,原本核准制下的上市实质性阻碍在注册制试点中得到逐步放宽,此时审核问询发挥的监管作用可能更为重要,交易所审核问询究竟能否有效发挥监管作用,提升IPO信息披露,从而助力于注册制下资源市场化配置?以上这一系列问题的研究,不仅有助于考察交易所在发行上市审核问询过程中的监管效果,理解“以信息披露为核心”在实现资源市场化配置中的关键作用,更能为持续深入推进注册制改革实践提供有益参考。

目前关于审核问询函影响的研究较少,从现有的研究来看, Li 和 Liu^[1]发现交易所审核问询会影响发行人IPO定价,当发行人收到更多的审核问询函时,从IPO申请日到最终发行日的价格降低幅度更大。Schuldt 和 Vega^[2]表明发行人的盈余管理行为会导致审核问询函中包含更多与收入确认相关的问题,审核问询函能约束公司上市后的盈余管理行为。Lowry 等^[3]对美国证监会披露的问询函与招股书进行了文本分析,研究表明来自美国证券交易委员会的审核问询会影响企业信息披露的程度和申请注册总时长,并发现“收入确认”是美国证券交易委员会最为关注的主题。综合来看,审核问询函具有一定的监管效力,能约束企业盈余操纵等机会主义行为,提升发行人信息披露程度。注册制在美国等市场化程度较高的股票市场已是一项普遍且成熟的发行制度,但我国资本市场发展程度和成熟资本市场仍存在一定差距,相较于欧美发达资本市场以机构投资者为主体的投资者结构,中国股票市场是以散户为主,而散户的风险识别能力和价值评估能力远远低于专业机构投资者。我国注册制试点从科创板开板以来至今才2年有余,仍处于一个探索阶段,科创板注册制诸多改革实践的成效仍有待检验。目前国内仅胡志强和王雅格^[4]研究了审核问询函对企业信息披露的影响,但他们对文本的分析仅基于特定字典展开,未能准确捕捉文字背后的语义,导致对招股书文本信息的衡量相对粗糙。审核问询函本文和招股书文本作为注册制信息披露的重要载体,其篇幅之庞大,主题内容之繁多,用简单的词典法进行文本分析并不能准确识别其中潜在信息、凝练其中主题内容。

本文基于 Blei 等^[5]提出的 LDA 主题分类模型展开文本分析,使用机器学习方法将审核问询函文本依据内容语义划分成12个主题,并用主题长度来刻画审核问询函对不同主题的关注程度。从以下几个层面展开研究:首先,检验交易所是否能识别发行人招股书中存在的信息披露质量问题。研究结果发现,在上市审核阶段,信息披露程度较低、技术信息不够充分的招股书申报稿,会收到更多的审核问询函。其次,通过对比招股书申报稿和招股书注册稿这两版文件的变化,检验注册制审核问询能否提升发行人IPO信息披露。结果表明注册制审核问询能提升招股书信息披露的程度:问询前后两版招股说明书的文本长度、文本句长、文本可理解性和文本定量信息显著增加。总体而言审核问询函对特定主题关注会导致招股书在对应主题信息披露的增加。进一步研究发现,审核问询函充分降低了信息不对称性,给予了投资者发挥鉴别能力和定价能力的基础,使得机构投资者报价分歧度降低。

相较于已有研究,本文的主要贡献体现在以下三个方面。第一,创新性地使用了 LDA 主题模型,来科学评估中国科创板注册制改革实施对IPO信息披露的影响。现有文献主要集中于研究企业披露的文本特征^[6],如使用文本可读性、文本不确定性、语调或情绪等进行描述^[7-11],但基于特定词典展开的文本分析并不能准确捕捉文本语义。近期也有少量文献如 Lowry 等^[3]开始基于 LDA 模型研究美国市场 IPO 注册制对信息披露的影响。但由于制度背景和投资者结构的显著差异,Lowry 等^[3]关于美国市场的研究并不能回答“中国市场实施注册制改革是否有利于提升 IPO 信息披露水平?”这一问题,而这一问题的回答又具有非常重要的理论价值和实践意义。本文基于 LDA 主题模型对科创板市场审核问询函文本以及招股书文本进行了主题分类,准确识别海量异构文本中的潜在信息,刻画分析了监管者进行审核问询的范围以及关注重点,并在此基础上回答了上述问题。此外,Lowry 等^[3]在研究过程中缺乏对内生性问题的充分考虑,而本文利用中国市场制度变革提供的研究机会,较好地控制了潜在的内生性问题。

第二,从文本主题的角度分析和研究了注册制审核问询对 IPO 信息披露的影响路径和影响机制。相较于胡志强和王雅格^[4],本文的研究不仅证实了审核问询的有效性,更重要地,从文本主题层面证实了多轮审核问询的针对性,实证结果表明,交易所对于特定主题的关注会显著提升发行人招股书对应主题的信息披露程度。这为理解注册制改革中监管机构发挥的重要作用提供了证据支持,有利于总结我国现阶段注册制的经验。

第三,从机构投资者报价行为视角揭示了注册制审核问询对科创板 IPO 定价机制的影响。中国科创板注册制询价机制改革强调机构投资者在询价、定价、配售等环节的主导作用,而充分的 IPO 信息披露是机构投资者在资源配置发挥作用的基础。本文检验了科创板注册制下的审核问询对机构投资者报价行为的影响。研究表明,交易所审核问询能降低机构投资者报价分歧度。这一结论有助于更好地理解“以信息披露为核心”在实现资源市场化配置中的关键作用,同时也是对 Lowry 等^[3]研究工作的重要补充。

1 制度背景、理论分析与研究假说

1.1 制度背景

严格的信息披露是注册制改革最重要的支撑,而交易所审核问询是确保信息披露“真实、准确、完整”的主要监管手段。图 1 说明了科创板首次公开发行股票注册流程以及每一阶段对应的信息披露。发行人通过保荐人以电子文档形式向上交所提交发行上市申请文件,所提交的第一稿招

股说明书即为申报稿,最后一稿招股说明书为注册稿。上交所需要在 5 个工作日内作出是否予以受理的决定。上交所决定受理后,发行人应于当日在交易所指定渠道立即披露招股说明书及相关文件。随后,发行人进入发行上市审核阶段,这一阶段又可细分为上交所的上市审核和证监会的注册审核两个环节。交易所通过设立科创板上市审核中心,负责上市审核工作。审核员、审核小组、审核部门会对招股说明书申报稿进行层层审核^②。并按照发行人提出审核问询、发行人回答问题的互动式流程开展审核工作。发行人及其保荐人、证券服务机构应当按照交易所发行上市审核机构审核问询要求及时、逐项回复审核问询函,补充说明或者修改招股说明书,并进行必要的补充调查和核查,审核问题和回复函、修改后的招股说明书都会及时同步向市场公开。审核部门问询通过后,相关文件会被递交到上交所审核中心进一步进行审核问询,最终相关文件会被递交至科创板股票上市委员会,上市委员会将视情况进行问询并发布上市委审议结果,发行人至此完成从申报到上会的全过程。过会后,上交所会将审核意见及发行人注册申请文件报送证监会履行注册程序,证监会同意则注册生效,发行人应当按照规定在注册决定有效期内发行股票。

综上所述,交易所通过设立多层级的审核问询体系,对发行人 IPO 信息披露进行层层把关,审核问询中的一问、二问乃至多轮问询,是交易所不断聚焦问题、突出重点的过程,是交易所督促发行人履行信息披露义务、震慑欺诈发行的重要监管手段。

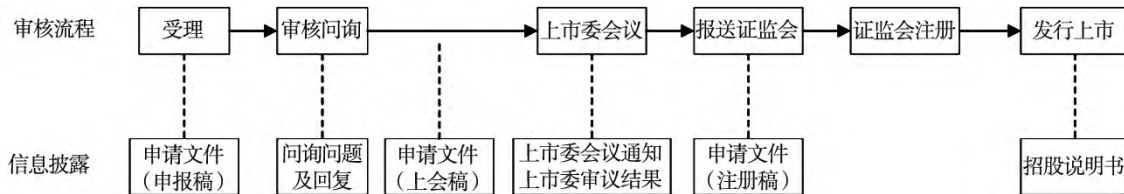


图 1 科创板 IPO 流程和对应信息披露

Fig. 1 IPO registration process of STAR market and information disclosure

② 在审核中,交易所重点关注以下方面:1) 重点关注发行人的信息披露是否达到真实、准确、完整的要求,是否符合招股说明书内容与格式准则的要求。2) 重点关注发行上市申请文件及信息披露内容是否包含对投资者作出投资决策有重大影响的信息,披露程度是否达到投资者作出投资决策所必需的水平。3) 发行上市申请文件及信息披露内容是否一致、合理和具有内在逻辑性。4) 发行上市申请文件披露的内容是否简明易懂,是否便于一般投资者阅读和理解。

1.2 理论分析与研究假说的提出

问询函监管是证券交易所为了规范公司信息披露、保护投资者合法权益而对公司实施审核的一种监管手段。问询函按照不同的被问询事项可以分为财报问询函、并购重组问询函、上市发行审核问询函等。以前文献关于监管问询函的研究大部分集中于对财报问询函的讨论,研究发现财报问询函能发挥治理作用^[12,13],改善上市公司信息披露行为,并对同行业公司的信息披露具有“溢出作用”^[14],在我国财报问询函是资本市场重要的监管手段之一,在信息披露监管方面发挥了重要作用^[15],能够降低上市公司股价崩盘风险^[16]。除了针对财报进行监管问询,交易所还会针对并购重组事项进行问询,并购重组问询函可以降低并购重组信息不对称进而提升并购绩效^[17],同时自愿披露也会提升并购重组的效率^[18]。综上所述,问询函监管具有一定的信息含量,能发挥监管作用,从而改善企业信息披露行为。尽管学术界已经就其他类型问询函对信息披露的影响进行了相对充分的研究,却鲜有文献聚焦于上市发行审核问询函。那么,审核问询函这一监管手段,能否丰富完善IPO信息披露内容、IPO提高信息披露质量?本文认为,审核问询函能从以下三个方面对IPO信息披露产生影响。

第一,审核问询函具有监管效力,专业化的审核问询函能切中发行人信息披露问题的要害,提升IPO信息披露。同发达国家资本市场相比,中国市场投资者理性程度、发行主体、中介机构和询价对象定价自主性与定价能力还存在较大不足;同时,科创板公司招股说明书具有一定的专业性,不仅有动辄几百页的篇幅,而且涉及到诸如股权结构、会计财务、研发活动、核心技术等专业性很强的主题内容。综上所述,缺乏相关专业背景和行业知识的中小投资者很难通过自己阅读找到招股书的症结所在。而交易所审核问询程序由具有相关知识背景的专业审核人员实施,交易所通过发送审核问询函的监管方式督促发行人履行信息披露义务,并就招股书中的股权结构、主营业务、核心技术、财务状况等多方面信息的展开专业询问。这样一种高度专业化的监管模式,能有发挥审核问询函的监管效力,从而提升IPO信息披露。

第二,审核问询函具有震慑作用,审核问询函由交易所发出,若发行人未能按照审核问询函内容履行信息披露义务,则发行人需要面对交易所终止发行上市审核决定的风险。交易所主要就发行人信息披露不足或瑕疵之处进行审核问询,首轮问询后,若发行人对交易所提出的审核问询问题不予理睬或未能提供让交易所满意的信息披露,则交易所会采取再次发函的方式进行“刨根问底”式问询,甚至终止发行上市审核。其次,公司信息披露存在重大疑问时,交易所将启动对保荐机构的现场督导,导致发行人及其保荐机构面临更为严苛的层层检视。当公司未能在发行上市环节落实以信息披露为中心的监管要求带来的风险大于收益时,出于自身收益最大化的考虑,公司有动机改善IPO信息披露行为以通过交易所上市审核。

第三,审核问询函能引发市场治理效应。随着交易所及时向外界公开审核问询和回复内容,发行人对关键审核问询的回复也会立即被外部市场感知,内部问题被置于外部市场监督之下,形成有效地市场约束。如果发行人不能对审核问询函中的问题逐条给予解释、补充披露或进行更正,则第三方机构以及投资者也会意识其中可能存在的潜在风险,进而加大了发行人的发行成本、市场交易成本。此外,信息不确定性和信息不对称会影响投资者对投资标的的估值判断^[19,20],从而导致柠檬市场下的价值低估。而更多的IPO信息披露会带来更准确的市场定价,更低的抑价水平^[21],并有效对冲诉讼风险^[22]。因此,发行人有动机完善IPO信息披露程度,提升IPO信息披露质量。

综合以上分析,提出本文的研究假说:科创板市场注册制实行过程中,交易所审核问询函能提升IPO信息披露水平。

2 研究设计

2.1 样本选择和数据来源

以2019年7月~2021年9月间在上海证券交易所科创板成功上市的341家公司为样本。利用爬虫技术,从上海证券交易所科创板股票审核

网站^③中获取招股说明书申报稿、招股说明书注册稿、回复函共三类文本原始数据^④。IPO 申报财务数据、中介机构业务排名、IPO 机构初步询价明细等数据来自 Wind 数据库,子公司个数数据从发行人招股说明书中手工收集。

2.2 基于 LDA 主题模型的文本分析

2.2.1 LDA 主题模型

LDA (latent Dirichlet allocation, LDA) 隐含狄利克雷分布,是一种概率主题模型,它由 Blei 等^[5]首次提出,可以依据语义将文档划分成不同的主题。主题模型的本质是模拟从主题(隐含)到词项(可见)的生产过程。基于提前预设的最优主题数, LDA 模型可以将每篇文档表示为在不同主题上的概率分布,而其中每一主题又可以看作词语的概率分布。近年来, LDA 主题分类模型在金融经济领域的研究中逐步得到了应用: Bao 和 Datta^[23]利用主题模型研究了 10-K 报表的风险披露如何影响投资者的风险认知。Dyer 等^[24]利用 LDA 主题模型对财报文本进行了分析,发现“公允价值”、“内部控制”和“风险因素”三个主题的信息披露增加是财报长度增加的最重要原因。Huang 等^[25]将盈余电话会议后迅速发布的大量分析师报告的主题内容与电话会议本身的主题内容进行比较,从而探究分析师信息发现角色和信息解释角色。而 Bellstam 等^[26]使用 LDA 的方法提取分析师报告中与企业创新相关的主题,并以此度量企业创新活动,不再局限于专利数量等传统指标。在提取审核问询函、招股书文本内容时,相较于传统的文本分析方法, LDA 主题模型具有以下优点: 第一, LDA 可以从海量的异构文本数据中间快速挖掘其潜在语义信息,非常适用于诸如招股说明书之类的篇幅巨大、主题内容繁多的文本体裁。第二,相较于使用人工阅读、词典法等方式来判别文本内容, LDA 主题模型具有客观性、可复制性、高效性。第三, LDA 主题模型对文本的分析不受章节、段落的限制,区别于按章节总结整

理信息的方法^[17], LDA 主题模型可以推断同一文档、同一段落属于多个主题的概率分布。这对招股说明书或审核问询函之类的文本尤其重要,因为此类文本中同一主题可能在全文上下多次穿插出现。

2.2.2 文本语料的获取和处理

在利用爬虫技术从相关网站获取原始文档 pdf 的基础上,进行如下操作: 第一,文本预处理。利用 Python 将 pdf 格式的原始文档转换为文本格式,同时剔除标点符号、页眉页脚; 第二,中文分词。选取停用词表为哈尔滨工业大学、四川大学编制的停用词表,并从原始文本中手工收集的公司名称、地理名词、会计金融术语、专业术语等,并构建用户词典。采用 Python 的 jieba 分词包对审核问询函和前后两版招股说明书的文本进行中文分词处理。第三,进一步处理文本语料。诸如“公司”、“发行人”这类在所有主题内都频繁出现的词会影响对于主题的解读,为了确保 LDA 主题模型输出的主题分布不被高频词占据从而最终影响主题质量,参考 Dyer 等^[24],本文人工审核了问询函(招股书)出现频率最高的 100(1 000) 个词、所有文本中覆盖率最高的 100(1 000) 个词,将其中无意义的词语加入停用词表。此外,进一步删除了英文、数字、特殊字符、单字与词频数为 1 的词,最终形成用于主题分类的审核问询函文本语料和招股说明书文本语料。

2.2.3 LDA 主题模型训练与模型提取

作为一种机器学习方法, LDA 主题模型本质上是对已有的文本数据做主题分布的拟合,因此 LDA 主题模型学习的结果和训练所用文本语料高度相关。为避免产生样本不平衡的问题,在使用时应尽量控制训练文本语料中的样本类别。交易所问询函和上市公司招股说明书虽然在关注的主题上存在相似,但在文本长度、格式上均存在显著差异。因此,如果将审核问询函与招股说明书语料混合进行训练,会出现样本类别的严重不平衡,此

^③ 网址: <http://keb.sse.com.cn/>

^④ 交易所向发行人发送的审核问询函并不会在科创板股票审核网站上单独披露,但会随着发行人及保荐机构关于审核问询函回复函的发布向外界公开。在回复内容中,发行人会先列明审核问询函问题,再就每一审核问询函所列问题进行回复。因此,实际上是从发行人及保荐机构的回函中手工提取审核问询函的内容。本文关注的是上交所发出的首封审核问询函文本,并对该文本语料进行 LDA 主题建模。

时 LDA 主题模型的训练结果会显著偏向文本语料中比例占优的招股说明书文本语料,而无法有效提取审核问询函语料中的主题. 因此,参照 Lowry 等^[3]在审核问询函和招股说明书文本语料上分别训练 LDA 主题模型,从而得到审核问询函与招股说明书各自的主题分布情况. 此外, LDA 模型需要人为地给定一个提前预设的主题数,主题数的设置会影响主题的生成和模型结果的可解释性. 将主题数设置得太低会导致主题过于宽泛和模糊,而主题数设置过高可能会引入经济上没有意义的话题. 参照 Bao 和 Datta^[23]的做法,将语料划分为 80% 的训练集和 20% 的测试集,并使用困惑度得分 (Perplexity) 作为评估模型的指标,困惑度由 Blei 等^[5]提出,用以评估语言模型优劣,较小的困惑度得分意味着模型对新文本有较好的预测作用. 其具体计算详见式(1). 其中 D 是测试数据中的文本信息集合, M 是其中包含的文本数量,在 M 文本中, N_m 是文本中包含的单词数量, $p(w_m)$ 表示在第 m 个文本中单词 w_m 的概率. 本文在审核问询函、招股说明书对应的训练集语料上,设定不同的主题数 K 进行训练,并使用困惑

度判断筛选出最适合的主题数 K 对应的模型

$$Perplexity(D) = \exp\left(-\frac{\sum_{m=1}^M \ln_D p(w_m)}{\sum_{m=1}^M N_m}\right) \quad (1)$$

问询函和招股书的预设主题数量与模型困惑度得分之间的关系如图 2a、图 2b 所示. 虽然困惑度得分随着主题数增加持续递减,但较多的主题数也可能会引入经济上没有意义的话题. 因此,虽然要选择困惑度较小的模型,但同时应该考虑主题数增大带来的相关问题. 综上,在使用困惑度得分评估模型的基础上,进一步地根据主题的解释能力来调整主题个数. 由图 2a 可知,对于问询函 LDA 主题模型,在主题数 $K > 10$ 之后,随着主题数的进一步增加,困惑度的下降速度趋于平缓,进一步增加主题数带来的边际效用不大. 本文将审核问询函最优主题数量固定在 10 左右区间. 进一步地,根据预设不同主题数下 LDA 模型输出主题分布的可解释性,调整主题个数,最终,选择将审核问询函主题数预设设为 $K = 12$,并训练出相应模型作为问询函 LDA 主题模型. 同理,对于招股书,本文选择将招股书主题数预设设为 $K = 30$,并训练出相应模型作为招股书 LDA 主题模型.

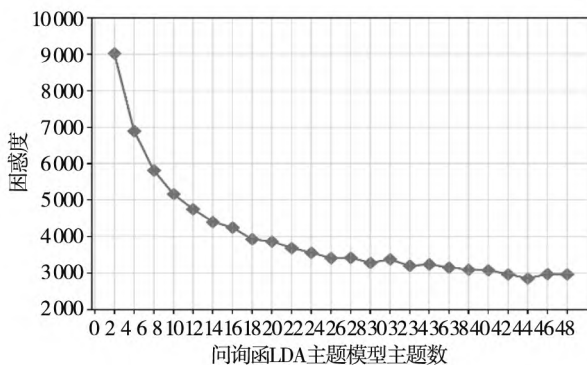


图 2a 问询函困惑度 - 主题数变化图
Fig. 2a SEC letter perplexity and number of topics

问询函主题描述性统计如表 1 所示. LDA 主题模型将问询函语料按照预设最优主题数划分为 12 个主题,按次序对其人工标签为“核心技术”、“损益情况”、“股权结构”、“行业竞争”、“资产”、“生产经营”、“应收款项”、“董监高”、“风险披露”、“客户及供应商”、“收入确认”、“生产产能”. 从审核问询函主题长度平均数来看,上交所的提问重点关注了发行人股东“股权结构”方面

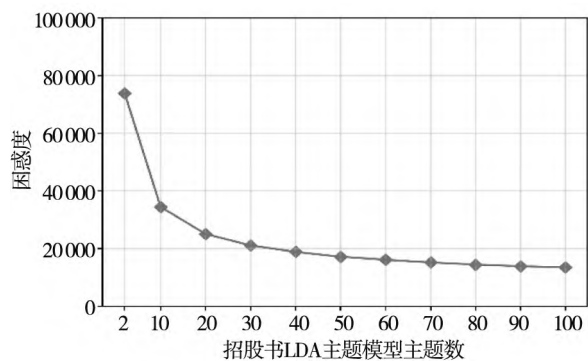


图 2b 招股书困惑度 - 主题数变化图
Fig. 2b Prospectus perplexity and number of topics

的问题,这突出了发行人股权结构问题的重要性——所持股份权属是否清晰、股权是否稳定性是科创企业长期发展的前提. 再者,企业的业务模式也是上交所审核问询重点内容之一. 发行人的主营业务、产品及服务、组织生产和销售的方式,都是有关企业是否持续经营能力的重要信息,而上交所从“生产产能”、“生产经营”、“客户及供应商”、“行业竞争”等主题展开了相关审核问询. 最

后,核心技术、研发活动、研发人员、发明专利等事项是科创企业竞争力所在,是判断发行人科创属性的重要依据。因此,“核心技术”主题也是上交所审核问询监管的关注重点。同时,值得注意的是,审核问询中有关“风险披露”的内容仍相对较少,这可能表明上交所在督促发行人充分进行风险揭示方面做得仍不够到位。对比申报稿匹配主题平均字符数和注册稿匹配主题平均字符数可知,注册稿在各个主题上的对应字符数均有所增加。

2.2.4 检验 LDA 主题模型输出的主题分布

为进一步验证以上 LDA 主题模型训练与模型选取是否合理、可靠、有效,本文对 LDA 主题模型输出的主题分布进行了验证。基于训练好的问询函 LDA 主题模型,可以得到每一篇审核问询函文本属于不同主题的概率。用审核问询函文本属于某一主题的概率来刻画上交所对于该主题的关注程度。本文以新冠疫情爆发作为外生冲击,检验

审核问询函中“风险披露”主题概率随审核时间的变化情况。如图 3 所示,在新冠疫情爆发后,审核问询函对“风险披露”主题的关注出现了显著的上升,并在疫情得到有效控制后逐渐下降。综上,检验结果在一定程度上支持 LDA 主题模型的有效性。

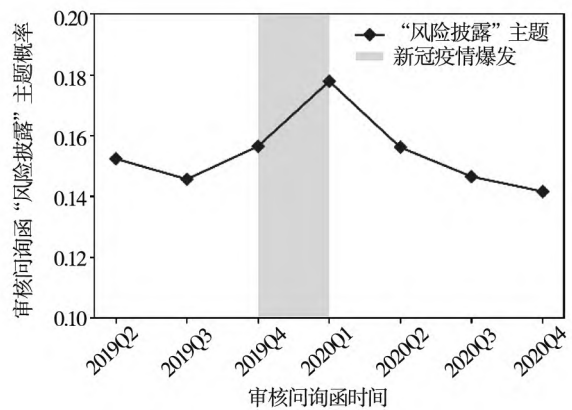


图 3 风险披露主题随时间的变化图

Fig. 3 Risk topic length changes over time

表 1 问询函主题描述性统计

Table 1 Descriptive statistics of SEC letter topics

	主题 1	主题 2	主题 3	主题 4	主题 5	主题 6	主题 7	主题 8	主题 9	主题 10	主题 11	主题 12
主题标签	核心技术	损益状况	股权结构	行业竞争	资产	生产经营	应收款项	董监高	风险披露	客户及供应商	收入确认	生产产能
主题字符中位数	1 734.65	1 439.23	1 875.13	1 577.34	1 208.57	1 024.62	1 684.47	1 054.47	561.38	1 774.09	1 148.66	1 344.93
主题字符平均数	1 820.15	1 546.21	2 073.15	1 725.67	1 267.2	1 124.36	1 766.91	1 115.38	594.25	1 829.56	1 269.35	1 406.68
前十主题词	研发技术 核心技术 人员 专利 纠纷 应用 参与 生产 知识产权	研发费用 毛利率 费用 成本 差异 会计师 同行业 构成 人员 各期	股东 股份 股权 持股 实际控制人 转让 持有 投资 协议 纠纷	市场 行业 经销商 国内 领域 竞争 风险 发展 募投项目 实现	存货 期末 计提 会计师 固定资产 余额 年末 比例 无形资产 现金	经营 生产 风险 资质 租赁 环保 考务 规定 授予 到期	应收账款 收入 模式 各期 比例 差异 主要客户 政策 会计师 经销	董事 人员 信息 管理人员 划分 控制 关联方 资金 高级 财务会计	风险 准则 要求 信息披露 提示 规定 政府补助 问答 风险因素 公开发行	采购 供应商 价格 原材料 子公司 交易 科技 公允 收购 关联方	收入 合同 服务 收入确认 销售收入 约定 会计准则 会计师 规定 订单	生产 设备 数据 产能 系统 授权 使用 亿元 制造 产量
申报稿匹配主题平均字符	9 045.84	11 170.88	8 753.06	30 419.5	13 858.1	10 273.49	13 858.1	10 684.31	6 177.49	6 472.39	14 686.22	14 686.22
注册稿匹配主题平均字符	12 306.73	22 220.44	11 742.41	37 803.94	21 142.26	12 232.15	21 142.26	12 364.84	7 489.59	7 663.22	20 083.68	20 083.68

注: 使用主题中的词项对问询函主题进行描述,按照每个主题词项在该主题中的概率降序排序,得到该主题前十主题词。

2.3 变量定义

2.3.1 审核问询函变量

从以下三个层面刻画科创板注册制下的审核问询:

1) 审核问询函数量 ($SLNum$): 从交易所受理发行人上市申请文件开始到发行人注册生效为止, 发行人收到交易所问询函件的总数.

2) 审核问询函主题 ($SLTopic$): 借鉴 Lowry 等^[3], 对每一篇问询函文本, 运用 LDA 主题模型提取交易所问询函 12 类主题, 并得到对应主题的概率. 通过将主题 i 对应主题概率乘以该文档的总字符数来构造主题 i 的伪长度, 以度量审核问询函中交易所对于特定主题的关注程度. 数值越大, 则问询函文本对该主题讨论的内容篇幅越多, 交易所对该主题的关注程度越大.

$$SLTopic = \ln(\text{主题 } i \text{ 概率 } \% \times \text{审核问询函总字符数}) \quad (2)$$

3) 审核问询函语气 ($SLTone$): 采用 Python 的 jieba 分词包对审核问询函本文语料进行中文分词处理, 并采用中国台湾大学《中文情感极性词典》判断词语的情感属性是正面或负面, 进行正面语气词汇和负面语气词汇的词频统计. 并参照式 (3) 的方式构建审核问询函语气变量, 其中 POS 代表正面语气词汇的个数, NEG 代表负面语气词汇的个数. 该变量数值越小, 则审核问询函语气越负面、情感越消极

$$SLTone = \frac{(POS - NEG)}{(POS + NEG)} \quad (3)$$

2.3.2 招股书文本特征变量

本文从以下五个方面刻画招股书文本特征, 以从不同的方面度量招股书信息披露内容.

1) 文本长度 ($Length$). 招股书文本的总字符数取自然对数.

2) 文本句长 ($Sentence$). 以招股书文本中平均每一句子中含有的词汇总数来衡量文本句长.

3) 文本可理解性 ($Understand$). 参照陈霄等^[27], 使用由国家教育部发布的《现代汉语语料

库字(词)频表》衡量文本可理解性. 《现代汉语语料库词频表》的语料规模为 2 000 万汉字, 其根据出现的次数, 囊括了 14 629 个汉语常用词, 使用该词频表构建文本可理解性指标. 以招股书文本平均每一句子含有的现代汉语常用词数来度量文本可理解性, 该项指标越大, 文本常用词越多, 可理解性越强.

4) 文本定量信息 ($Quantity$). 借鉴 Huang 等^[25], 以文本平均每一句子中包含“%”、“¥”、“\$”、“人民币”、“美元”、以及阿拉伯数字字符个数来衡量招股书文本定量信息含量.

5) 文本技术信息 ($Technology$). 采取手工收集的方式, 将发行人招股说明书第一章中列示的专业词语释义表中的词汇整理成词典. 该词典涵盖了与发行人核心技术及研发活动相关的专业术语. 以招股书文本中平均每一句子包含的专业词语数来衡量文本技术信息含量, 以反映招股书对科技创新、核心技术相关事项的披露程度.

2.3.3 招股说明书在主题 i 上的信息披露变化程度

对于招股说明书文本, 使用 LDA 模型将招股说明书分类为 30 个主题, 并基于余弦相似度将招股书主题与问询函主题进行一一匹配. 与问询函主题 i 最相似的招股书主题, 也将被标签为主题 i ^⑤. 例如, 基于余弦相似度找到与问询函“股权结构”主题最为相似的某一招股书主题, 此时该招股书主题也将被标签为“股权结构”主题. 审核问询前后两版招股书在主题 i 上的信息披露变化 ($Disclosure\ Change\ in\ Topic\ i\ DC_i$) 用式 (4) 进行度量

$$DC_i = \ln \left(\frac{\text{主题 } i \text{ 概率 } \% \times \text{招股书注册稿总字符数}}{\text{主题 } i \text{ 概率 } \% \times \text{招股书申报稿总字符数}} \right) \quad (4)$$

2.3.4 控制变量

参考 Lowry 等^[3], 选取如下变量作为控制变量, 控制变量的具体定义见表 2.

⑤ 此处涉及到如何将问询函 12 类主题与招股书 30 类主题相配, 具体地, 对于语料中每一个主题, 可以根据语料中每一个词语在该主题上的出现概率, 构建一个表示该主题的向量, 向量的维数等于语料中总词语数, 向量的第 n 项等于第 n 个词语在该主题上的出现概率. 对于问询函的每一个主题向量, 本文使用余弦相似度计算其与招股书每一个主题向量的相似度, 将相似度最高的招股书主题确定为该问询函主题对应的招股书主题, 这样就完成了每个问询函主题与招股书主题的匹配.

表 2 控制变量定义

Table 2 Definition of control variables

变量名称	变量定义
第一大股东持股比例	上市当年最大的股东持有的股票数量占总股本的比例
净资产收益率	上市当年净利润与净资产的比值
资产负债率	上市当年总负债与总资产的比值
公司规模	上市当年总资产的自然对数
控股子公司数	上市当年控股子公司数量的自然对数
公司年龄	公司成立年龄的自然对数
承销商声誉	承销商为业务排名前五的券商承销时为取值 1, 否则为 0
会计师事务所声誉	会计师事务所为业务排名前五的会计师事务所时为取值 1, 否则为 0
律师事务所声誉	律师事务所为业务排名前五的律师事务所时为取值 1, 否则为 0
风险投资	有风险投资取值为 1, 否则为 0

考虑到数据异常值可能影响实证结果,因此本文对所有连续变量进行 1% 分位双边缩尾处理。经过如上处理后,各变量的描述性统计结果如表 3 所示。由表 3 可知,科创板公司在审核问询全过程中平均会收到 4.29 封问询函。申报

稿文本长度取对数后的均值为 12.68, 其对应字符数大约为 321 258 字符,可见发行人招股说明书确实篇幅庞大、信息丰富,仅凭人工阅读或者使用词典法进行文本分析难以挖掘其中蕴含的有用信息。

表 3 主要变量描述性统计

Table 3 Descriptive statistics of major variables

变量名称	均值	标准差	25%	50%	75%
审核问询函数量	4.290	1.146	3	4	5
审核问询函语气	0.481	0.0820	0.426	0.485	0.541
机构报价分歧度	2.728	3.372	0.823	1.680	3.118
申报稿文本长度	12.68	0.210	12.54	12.66	12.79
申报稿文本可理解性	33.19	3.113	31.23	33.08	34.59
申报稿文本句长	86.98	14.27	80.03	85.22	91.05
申报稿文本定量信息	0.126	0.0260	0.109	0.119	0.136
申报稿文本技术信息	5.094	1.046	4.429	5.028	5.584

2.4 模型设定

本文首先检验交易所是否能识别发行人招股说明书文本中存在的信息披露问题并实施相应的审核问询监管。建立回归模型(5),并进行多元回归分析。此外,为了克服潜在小样本偏误对结论的影响,采用自抽样法(Bootstrap)对样本进行 1 000 次自抽样,计算得到 Bootstrap 标准误

$$SLNum = a + \beta_1 Length/Sentence/Understand \\ Quantity/Technology + Controls + \\ Year + Ind + \varepsilon \quad (5)$$

其中被解释变量 $SLNum$ 为交易所审核问询过程中发布的问询函数量,解释变量为一系列招股说明书信息披露特征的代理变量。 $Controls$ 为控制变量, $Year$ 表示控制年度固定效应, $Industry$ 表示控制行业固定效应, ε 为随机干扰项。

在验证审核问询函可以识别招股说明书信息披露问题的基础上,本文进而考察审核问询函对 IPO 信息披露的影响。为了检验审核问询函和文本信息披露变化之间的关系,设定如下回归模型(6),并采用自抽样法(Bootstrap)解决小样本偏

误问题

$$\Delta TextInformation = \alpha + \beta_1 SLNum + Controls + Year + Ind + \varepsilon \quad (6)$$

其中被解释变量 *TextInformation* 表示前后招股书申报稿和招股书注册稿之间各信息披露指标的变化程度,具体包括前后两版招股书之间文本长度、文本长句、文本可理解性、文本定量信息、文本技术信息的变化程度。其数值等于招股说明书注册稿的文本特征变量与申报稿对应的文本特征变量之比。解释变量 *SLNum* 为交易所问询审核过程中发布的问询函数量。

进一步地,在主题层面上,为考察审核问询函对某一特定主题关注是否会导致发行人招股书对应主题信息披露的变化,设定如下模型(7),并采用自抽样法(Bootstrap)解决小样本偏误问题

$$DC_i = \alpha + \beta_1 SLTopic + Controls + Year + Ind + \varepsilon \quad (7)$$

其中被解释变量 *DC_i* 为发行人前后两本招股书在主题 *i* 上的信息披露变化程度,解释变量 *SLTopic* 为问询函在对应主题的长度,代表问询函对该主题的关注程度。

3 实证结果与分析

3.1 上市公司收到审核问询函影响因素的检验

为了探究交易所是否能识别发行人招股说明书文本中存在的信息披露问题并实施相应的审核问询监管,依照模型(4)进行多元回归,实证结果如表4所示。交易所审核问询函数量与申报稿文本总长度和申报稿技术信息呈显著负相关性。即发行人申报稿的文本总长度越小、技术信息披露越少,收到的问询函就相应越多。这说明,在诸多度量招股书信息披露的文本特征中,上交所更加注重招股书申报稿本身的篇幅,更加注重对科技创新相关事项的披露。这与科创板本身的战略定位是高度契合的。同时,也观察到交易所并未对其他招股书文本特征做出反应,无论是文本可理解

性还是文本平均句长,都与问询函数量没有显著相关性。申报稿文本中的定量信息含量也不会显著影响交易所审核问询函数量,这可能因为招股书中定量信息多集中在会计财务信息方面,而交易所应当更加关注企业财务定量信息的质量而不仅是数量,因此定量信息披露程度和发行人收到的问询函数量间没有显著关系。

3.2 审核问询函与 IPO 信息披露的检验

在验证了交易所审核问询函能识别招股书申报稿信息披露问题的基础上,本文进一步考察审核问询函对 IPO 信息披露的影响。

首先,文本探究了审核问询函数量与招股书文本特征变化之间的关系。如表5所示,交易所审核问询函的数量与发行人招股说明书前后文本长度变化程度、文本句子平均长度变化程度、文本可理解性变化程度、文本定量信息变化程度成正相关。具体地,当审核问询数量增加1封时,注册稿的文本长度变化程度会增加7.9%,文本句长变化程度会增加1.2%。同时每增加一封问询函,文本可理解性变化程度平均会增加0.8%,文本定量信息变化程度平均会增加2.9%。此外,值得注意的是,表4结论表明,交易所会因为招股书技术信息的披露不足进行更多问询,但表5结果显示,发行人后续并没有因问询在技术信息上而进行更详细的披露。本文对此的解释为,“以信息披露为核心”的审核理念的根本是以投资者需求为导向,交易所就招股书技术信息披露问题向发行人进行问询的目的在于,通过充分信息披露和风险揭示,让投资者自主判断发行人的科技创新能力。而考虑到招股书阅读的主体是广大投资者,交易所在审核问询函中会提醒发行人落实信息披露的简明性和可理解性,力求披露内容简明易懂、客观平实,减少使用专业术语、广告语言^⑥。而发行人也会遵守交易所的要求,在回函中尽量避免做到“用专业术语解释专业术语”,而是努力提升招股书在科技创新事项方面的可理解性。以上现实

^⑥ 以上交所向龙软科技发放的审核问询函为例,上交所曾明确要求发行人“补充披露以产品线及应用领域代替核心技术披露的原因及适当性,并以简明易懂的语言逐项披露所列技术中的自有技术和行业通用技术,充分披露发行人的核心技术及所填补的技术空白。”

背景与表 5 实证结果相契合,即审核问询函能显著提升招股书文本的可理解性,但并不会显著提升文本技术信息的篇幅.此外,本文发现审核问询

函正面语气减弱了交易所问询数量对于文本长度等定量信息的影响.这表明当问询函语气更加负面时,发行人会更多地提升招股书信息披露程度.

表 4 申报稿信息披露和交易所审核问询

Table 4 Draft prospectus information disclosure quality and SEC inquiry

变量	(1)	(2)	(3)	(4)
	<i>SLNum</i>	$\ln(SLNum + 1)$	<i>SLNum</i>	$\ln(SLNum + 1)$
申报稿文本总长度	-0.247*** (-2.71)	-0.193** (-2.47)	-0.247*** (-2.61)	-0.193** (-2.50)
申报稿文本句长	0.000 (0.20)	0.000 (0.11)	0.000 (0.18)	0.000 (0.11)
申报稿文本可理解性	0.011 (1.29)	0.009 (1.27)	0.011 (1.26)	0.009 (1.28)
申报稿定量信息	0.477 (0.86)	0.371 (0.83)	0.477 (0.80)	0.371 (0.80)
申报稿技术信息	-0.049** (-1.98)	-0.039* (-1.96)	-0.049* (-1.87)	-0.039* (-1.92)
截距项	3.688*** (3.45)	3.337*** (3.64)	3.688*** (3.35)	3.337*** (3.72)
控制变量	控制	控制	控制	控制
行业效应	控制	控制	控制	控制
时间效应	控制	控制	控制	控制
样本数	341	341	341	341
伪 R^2 / 调整 R^2	0.0219	0.204	0.0219	0.204

注:第(1)列、第(3)列为泊松回归,第(2)列、第(4)列为普通最小二乘法,第(1)列、第(2)列括号中报告经过公司层面聚类调整的 t 值,第(3)列、第(4)列括号中报告了由 Bootstrap 自抽样 1 000 次得到的 z 值,***、**、* 分别表示估计参数在 1%、5% 与 10% 水平上显著.

表 5 交易所审核问询函对发行人信息披露程度的影响

Table 5 SEC letter and issuer information disclosure

变量	被解释变量:注册稿信息披露文本特征/申报稿信息披露文本特征				
	(1)	(4)	(3)	(3)	(5)
	Δ 文本长度	Δ 文本句长	Δ 文本可理解性	Δ 文本定量信息	Δ 文本技术信息
交易所审核问询函数量	0.079*** (7.62)	0.012*** (2.87)	0.008*** (4.11)	0.029*** (3.59)	-0.003 (-0.96)
交易所审核问询函语气	-0.283*** (-2.77)	-0.114** (-2.38)	-0.005** (-0.18)	-0.262** (-2.29)	-0.013 (-0.38)
截距项	1.128*** (5.04)	1.438*** (12.83)	1.209*** (22.14)	1.495*** (6.61)	1.211*** (14.06)
控制变量	Yes	Yes	Yes	Yes	Yes
行业效应	Yes	Yes	Yes	Yes	Yes
时间效应	Yes	Yes	Yes	Yes	Yes
样本数	341	341	341	341	341
调整后 R^2	0.305	0.086	0.093	0.087	0.050

注:括号中报告了由 Bootstrap 自抽样 1 000 次得到的 z 值,***、**、* 分别表示估计参数在 1%、5% 与 10% 水平上显著.

本文聚焦于审核问询函主题和对应招股书主题层面,关注交易所问询函对于不同主题的关注程度对发行人在不同主题上披露程度的影响.表6列示回归结果表明,除了“风险披露”和“生产产能”主题之外,审核问询函对特定主题的关注都会导致发行人招股书对应主题的信息披露程度提升,这充分体现了交易所在审核问询中扮演的角色,即通过问询引导发行人披露更多相关信息.就表6中所列示的回归系数而言,“核心技术”主

题对应的系数最大,达到了0.283,并在1%的置信水平上显著.相比之下,损益情况的披露虽然也有大幅增加,但是增幅只有核心技术的42%,这也说明交易所审核问询环节对于公司的科技创新属性的关注度更高,更希望引导公司披露核心技术有关的信息.此外由于风险披露和生产产能是更为格式化的信息,交易所在相关主题上的问询对于篇幅增加的影响并不明显,可能更多的是内容上的修改.

表6 审核问询函主题长度与招股书对应主题变化程度

Table 6 SEC letter topic length and prospectus topic disclosure change

变量	被解释变量: 基于余弦相似度匹配的对应该招股书主题长度的变化程度											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	核心技术	损益情况	股权结构	行业竞争	资产	生产经营	应收款项	董监高	风险披露	客户及供应商	收入确认	生产产能
审核问询函主题 <i>i</i> 长度	0.283*** (5.31)	0.118*** (3.26)	0.106*** (3.83)	0.044* (1.77)	0.085*** (2.62)	0.078*** (2.89)	0.115*** (3.46)	0.060*** (3.41)	0.004 (0.16)	0.108*** (4.76)	0.151*** (4.46)	0.010 (0.28)
审核问询函语气	0.141 (0.58)	-0.788*** (-4.01)	-0.333 (-1.62)	0.009 (0.08)	-0.568*** (-3.27)	-0.136 (-0.80)	-0.538*** (-3.07)	0.038 (0.49)	-0.090 (-0.62)	-0.060 (-0.56)	-0.032 (-0.19)	0.127 (0.75)
截距项	-0.919 (-1.29)	-0.027 (-0.05)	-0.555 (-1.18)	-0.213 (-0.70)	0.385 (0.79)	1.828*** (4.50)	-0.091 (-0.17)	-0.409* (-1.93)	-0.504 (-1.18)	-0.870*** (-2.76)	-0.655 (-1.57)	0.181 (0.39)
控制变量	控制	控制	控制	控制	控制	控制	控制	控制	控制	控制	控制	控制
行业效应	控制	控制	控制	控制	控制	控制	控制	控制	控制	控制	控制	控制
时间效应	控制	控制	控制	控制	控制	控制	控制	控制	控制	控制	控制	控制
样本数	341	341	341	341	341	341	341	341	341	341	341	341
调整后R ²	0.093	0.150	0.117	0.045	0.183	0.039	0.194	0.224	0.055	0.119	0.113	0.047

注: 括号中报告了由 Bootstrap 自抽样 1 000 次得到的 z 值, ***、**、* 分别表示估计参数在 1%、5% 与 10% 水平上显著.

3.3 内生性问题的检验

考虑到科创板发行上市审核流程是一项时间节点清晰、交易所和发行人依照业务流程循序推进活动,首轮审核问询函的发出和发行人就函件完善招股书信息披露之间在时间上有明晰的先后次序,因此两者之间的因果关系比较明确,实证结果受互为因果问题干扰的可能性较小.但考虑到实证回归中仅控制了部分可观测的公司特征变量,本文的主要实证结论仍可能受到遗漏变量的干扰.为了控制可能出现的遗漏变量的问题对实证结果的影响,参照胡志强和王雅格^[4]的做法,利用工具变量法控制可能的内生性问题.

科创板注册制审核规则的变动为构建审核问询相关代理变量的工具变量提供了机会.为进一步落实“以信息披露为核心”的注册制要求,改进发行上市审核方式,切实提高审核效率,优化审核服务,上交所在前期发行上市审核工作的基础上推进“科创板审核 2.0”,更加突出审核问询的重大性、针对性,按照重要性原则和有效性原则,精简优化问询内容,提高信息披露.2020年12月4日,上交所发布《上海证券交易所科创板股票发行上市审核规则(2020年修订)》,其中要求“本所审核和中国证监会注册的时间总计不超过三个月.”^⑦此后,交易所审核问询的节奏加快,更聚焦

⑦ 详情参见《上海证券交易所科创板股票发行上市审核规则(2020年修订)》.

于重点问题,审核问询在总体上减量提质。因此,这一科创板审核规则的变更可能会影响上交所相关部门发放问询函的频次和问询重点,但不会影响发行人对招股书的修改和完善。具体地,构建工具变量 IV_period , 发行人 IPO 受理日期在 2019 年 12 月之前的取值为 0, 反之取值为 1。

工具变量回归结果如表 7 所示。2SLS 第一阶

段的回归结果显示,工具变量与内生解释变量显著负相关,且 F 统计量为 97.337 2, 在 1% 水平上显著,故认为不存在弱工具变量的问题。第二阶段的回归结果显示,第(2)列~第(5)列中,拟合的问询函数量对应的回归系数仍显著为正,这表明本文主要实证结论在控制了内生性问题之后仍保持稳健^⑧。

表 7 工具变量回归

Table 7 Regression of instrumental variables

变量	第一阶段	第二阶段				
	(1)	(2)	(4)	(3)	(5)	(6)
	问询函数量	Δ 文本长度	Δ 文本句长	Δ 文本可理解性	Δ 文本定量信息	Δ 文本技术信息
工具变量 IV_period	-1.532 *** (-9.87)					
问询函数量		0.185 *** (7.96)	0.015* (1.76)	0.021 *** (4.37)	0.047 *** (2.77)	-0.001 (-0.16)
截距项	3.655 ** (2.3)	1.214 *** (3.30)	1.644 *** (12.23)	1.351 *** (17.94)	1.502 *** (5.67)	1.430 *** (12.92)
控制变量	控制	控制	控制	控制	控制	控制
行业效应	控制	控制	控制	控制	控制	控制
时间效应	控制	控制	控制	控制	控制	控制
样本数	341	341	341	341	341	341
F 统计量	97.337 2 ***					
调整后 R^2	0.380 7	0.054	0.119	0.103	0.063	0.087

注: 括号中报告经过公司层面聚类调整的 t 值, **、*、* 分别表示估计参数在 1%、5% 与 10% 水平上显著。

4 进一步研究

4.1 注册制审核问询与机构投资者报价分歧度

新股定价问题一直是中国资本市场的难题, 风险和不确定性是致使投资者对股票估值产生意见分歧的重要原因^[28]。而科创企业本身还具有研发投入高、技术迭代快、业绩波动大等特征, 市场可比公司较少, 估值判断难度进一步增大。以前文献表明, 投资者情绪和意见分歧是我国 IPO 长期弱势的主要因素^[29]。科创板注册制询价机制改革强调机构投资者在询价、定价、配售等环节的主导作用, 而充分的 IPO 信息披露是市场在资源配置

发挥作用的基础。那么, 在发行承销阶段, 审核问询是否充分降低了信息不对称, 给予了投资者发挥鉴别能力和定价能力的基础? 本文认为, 交易所审核问询能督促发行人将企业重要事项真实、充分、准确、及时地向外界披露, 为投资者进行价值判断和决策提供更加有效的参考。充分的信息披露可以使机构投资者对即将发行上市公司具有更为全面清晰的了解, 有助于统一各个投资者对上市公司的意见, 从而降低机构报价意见分歧, 提升定价效率。为了检验审核问询函数量是否可以显著降低机构投资者报价分歧度, 以及审核问询函中哪些主题的信息含量对机构投资者报价分歧度影响最大, 设计了如下回归模型

^⑧ 主题层面的工具变量回归结果与主检验类似, 前文得到的研究结论相对稳健。限于篇幅, 主题层面的工具变量回归结果没有列示。感兴趣的读者可向作者索取。

$$Divergence = \alpha + \beta_1 SLNum/SLTopic + Controls + Year + Ind + \varepsilon \quad (8)$$

其中被解释变量 *Divergence* 表示机构投资者报价分歧度. 参照李冬昕等^[30]、俞红海等^[31] 以初步询价中机构投资者报价的标准差刻画投资者报价分歧度. 投资者报价分歧度衡量了不同机构报价之间的差异, 体现了市场的定价效率. 解释变量分别为审核问询函数量以及审核问询函各个主题的长度.

表8展示了回归结果, 被解释变量反映了科创板市场新股发行时机构之间报价的标准差, 数值越小则表明新股发行时机构的报价越统一, 报价越趋近于内在价值, 意味着市场发行定价效率越高. 在发行承销阶段, 如果充分的信息披露能降低机构报价意见分歧, 提升发行定价的效率, 那么发行人收到的问询函数量以及问询函中关于各个主题的询问长度的回归系数应该显著为负.

表8 交易所审核问询与机构投资者报价分歧度

Table 8 SEC inquiry and quotation divergence of institutional investors

变量	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
	机构报价分歧度	核心技术	损益情况	股权结构	行业竞争	资产	生产经营	应收款项	董监高	风险披露	客户及供应商	收入确认	生产产能
问询函数量	-0.449*** (-2.73)												
审核问询函主题 <i>i</i> 长度		-0.546* (-1.67)	-0.668** (-2.04)	-0.100 (-0.34)	0.400 (1.09)	-0.671 (-1.46)	-1.169*** (-2.74)	-0.398 (-0.76)	-0.204 (-0.51)	0.087 (0.26)	-0.674 (-1.52)	-0.246 (-0.80)	0.230 (0.54)
审核问询函语气	3.889* (1.81)	3.537* (1.66)	2.785 (1.29)	3.270 (1.50)	3.411 (1.59)	2.814 (1.30)	2.737 (1.30)	2.994 (1.38)	3.299 (1.53)	3.378 (1.59)	2.837 (1.32)	3.579 (1.62)	3.361 (1.57)
截距项	17.654*** (3.68)	20.895*** (3.52)	21.117*** (3.85)	17.165*** (3.06)	13.426** (2.29)	20.323*** (3.50)	20.830*** (4.00)	19.843*** (3.54)	17.552*** (3.22)	16.132*** (3.15)	20.731*** (3.70)	17.995*** (3.51)	15.254*** (2.75)
控制变量	控制	控制	控制	控制	控制	控制	控制	控制	控制	控制	控制	控制	控制
行业效应	控制	控制	控制	控制	控制	控制	控制	控制	控制	控制	控制	控制	控制
时间效应	控制	控制	控制	控制	控制	控制	控制	控制	控制	控制	控制	控制	控制
样本数	341	341	341	341	341	341	341	341	341	341	341	341	341
调整 <i>R</i> ²	0.199	0.185	0.188	0.180	0.183	0.188	0.208	0.183	0.181	0.180	0.187	0.181	0.181

注: 括号中报告了由 Bootstrap 自抽样 1 000 次得到的 *z* 值, **、*、* 分别表示估计参数在 1%、5% 与 10% 水平上显著.

从回归结果来看, 发行人收到的问询函数量每增加 1 封, 报价分歧度会减少 0.449, 相当于均值的 16.46%. 而主题方面, 几乎所有的解释变量系数都为负, 其中关于核心技术、损益情况和生产经营主题的询问长度的增加都会显著降低机构投资者报价的分歧度, 这说明核心技术、损益情况和生产经营是投资者更为关心的主题, 相关主题问询引导公司的披露会降低机构报价意见分歧, 提升发行定价效率. 其中“损益情况”和“生产经营”主题长度的回归系数分别为 -0.668 和 -1.169, 且在 5% 的水平上显著, 核心技术的问询函主题

长度的回归系数为 -0.546, 在 10% 的水平下显著, 是均值的 20.01%, 这也说明机构投资者对于核心技术的敏感性低于损益情况和生产情况, 即机构投资者在初步询价阶段, 对公司的价值评估分歧和报价水平差距更受企业基本面情况的影响.

4.2 稳健性检验

4.2.1 更换问询函主题与招股书主题的方式

主题之间的匹配计算本质上是词向量之间的相似度, 目前使用比较广泛的计算方法有余弦距

离、KL 距离和 JS 距离。前文使用余弦相似度对问询函主题向量与招股书主题向量进行匹配,从而找到与问询函特定主题最为相似的招股书主题。在稳健性检验中,本文更换问询函主题与招股书主题之间的匹配方式,用 JS 距离进行匹配^⑨,从而为问询函 12 类主题找到与其最为相似的招股书主题,并构建与之相关的实证变量。稳健性检验结果与前文结论相似,前文得到的研究结论依然成立。

4.2.2 更换变量度量方式^⑩

稳健性检验中,用主题 i 在前后两版招股说明书申报稿和招股说明书注册稿之间主题概率变化来衡量招股书在主题 i 上的信息披露变化程度,用主题 i 在问询函文本的主题概率来衡量交易所审核问询对某一主题的关注程度,并进行相关回归分析。稳健性检验结果与前文结论相似,前文得到的研究结论依然成立。

5 结束语

设立科创板并试点注册制是资本市场全面深化改革的全新探索,是近两年来中国资本市场最为关心的重大议题。信息披露直接对接注册制的市场化原则,是确保市场力量均衡博弈以充分发挥市场化定价作用的基础。注册制审核问询则是不断督促发行人及中介机构真实、准确、完整地披露信息的监管手段。本文研究了注册制审核问询对 IPO 信息披露的作用。得到以下结论:1) 在上市审核阶段,信息披露程度较低、技术信息披露不够充分的招股书,会收到更多的审核问询函。2) 注册制审核问询能提升招股书信息披露的程度,问询前后两版招股说明书在文本长度、文本句长、文本可理解性、和文本定量信息上显著增加。总体

而言,审核问询函对特定主题关注会导致招股书在对应主题信息披露的增加。3) 进一步分析表明,审核问询函能降低机构投资者报价分歧度。以上结果表明,注册制改革背景下,交易所切实履行了信息披露的把关责任,提升了 IPO 信息披露水平,为市场在资源配置中发挥决定性作用奠定了基础。

本研究具有重要理论及实践意义。在理论方面,其主要结论证实了注册制审核问询对 IPO 信息披露的提升作用,为理解注册制改革中监管机构发挥的重大作用提供了证据支持,补充了审核问询函经济后果的相关研究,也丰富了 IPO 信息披露影响因素相关文献^[3,4]。此外,本文基于审核问询函文本和招股书文本,运用 LDA 主题模型,将文本中相似的语义内容划分为一类文本主题,更好地去了解监管者进行审核问询的范围和重点。在实践方面,有利于总结我国现阶段注册制的经验,具有一定的政策启示意义。对于监管者主体来说,有利于上交所进一步提升审核内容和审核程序的透明度、规范性,健全优化公司发行上市审核规则体系,未来将科创板注册制实践的经验辐射到整个主板市场;对于发行人主体来说,有利于拟上市公司了解注册制审核的标准和尺度,了解监管者关心的重点主题,按照试点注册制改革理念,在发行上市环节落实以信息披露为中心的监管要求,更好地服务于“以投资者需求为导向”的目的。对于投资者主体来说,核准制下的上市实质性阻碍在注册制试点中得到逐步放宽,这更加强了交易所信息披露方面的监管作用。了解监管者在审核问询的重点问题、重点主题,能够帮助外部投资者更好地发挥价值评估的能力,发挥市场在资源配置中的决定性作用。

^⑨ JS 距离用以衡量两个概率分布间的差异情况,其本质上为另一种衡量概率分布间差异的指标——KL 散度的变种。JS 散度在 KL 散度的基础上,进一步解决了 KL 散度缺乏对称性,不适合用作衡量距离的问题,使得 P 与 Q 的 JS 散度等于 Q 与 P 的 JS 散度,具体定义如下,其中 P 与 Q 分别对应着问询函与招股书的一个主题分布, $JS(P, Q) = \frac{1}{2}KL\left(P, \frac{P+Q}{2}\right) + \frac{1}{2}KL\left(Q, \frac{P+Q}{2}\right)$ 。

^⑩ 限于篇幅,此处不报道稳健性检验具体结果,感兴趣的读者可向作者索取。

参考文献:

- [1] Li B, Liu Z B. The oversight role of regulators: Evidence from SEC comment letters in the IPO process [J]. *Review of Accounting Studies*, 2017, 22(3): 1229 – 1260.
- [2] Schuldt M, Vega J. An examination of SEC revenue recognition comments and IPO earnings management [J]. *Accounting Research Journal*, 2018, 31(3): 371 – 387.
- [3] Lowry M, Michaely R, Volkova E. Information revealed through the regulatory process: Interactions between the SEC and companies ahead of their IPO [J]. *Review of Financial Studies*, 2020, 33(12): 5510 – 5554.
- [4] 胡志强, 王雅格. 审核问询、信息披露更新与 IPO 市场表现——科创板企业招股说明书的文本分析 [J]. *经济管理*, 2021, 43(4): 155 – 172.
Hu Zhiqiang, Wang Yage. Review inquiry, disclosure update, and IPO performance: Textual analysis based on prospectuses of STAR market listed companies [J]. *Business Management Journal*, 2021, 43(4): 155 – 172. (in Chinese)
- [5] Blei D M, Ng A Y, Jordan M I. Latent Dirichlet allocation [J]. *Journal of Machine Learning Research*, 2003, 3(4–5): 993 – 1022.
- [6] 马长峰, 陈志娟, 张顺明. 基于文本大数据分析的会计和金融研究综述 [J]. *管理科学学报*, 2020, 23(9): 19 – 30.
Ma Changfeng, Chen Zhijuan, Zhang Shunming. A survey on accounting and finance research based on textual big data analysis [J]. *Journal of Management Sciences in China*, 2020, 23(9): 19 – 30. (in Chinese)
- [7] Li F. Annual report readability, current earnings, and earnings persistence [J]. *Journal of Accounting & Economics*, 2008, 45(2–3): 221 – 247.
- [8] Loughran T, McDonald B. IPO first-day returns, offer price revisions, volatility, and form S-1 language [J]. *Journal of Financial Economics*, 2013, 109(2): 307 – 326.
- [9] 卞世博, 管之凡, 阎志鹏. 答非所问与市场反应: 基于业绩说明会的研究 [J]. *管理科学学报*, 2021, 24(4): 109 – 126.
Bian Shibo, Guan Zhifan, Yan Zhipeng. Irrelevant answers and market reaction: Evidence from performance briefings [J]. *Journal of Management Sciences in China*, 2021, 24(4): 109 – 126. (in Chinese)
- [10] 姚加权, 冯 绪, 王赞钧, 等. 语调、情绪及市场影响: 基于金融情绪词典 [J]. *管理科学学报*, 2021, 24(5): 26 – 46.
Yao Jiaquan, Feng Xu, Wang Zanjuan, et al. Tone, sentiment and market impacts: The construction of Chinese sentiment dictionary in finance [J]. *Journal of Management Sciences in China*, 2021, 24(5): 26 – 46. (in Chinese)
- [11] 吴武清, 赵 越, 闫嘉文, 等. 分析师文本语调会影响股价同步性吗? ——基于利益相关者行为的中介效应检验 [J]. *管理科学学报*, 2020, 23(9): 108 – 126.
Wu Wuqing, Zhao Yue, Yan Jiawen, et al. Does textual tone in analyst reports affect stock price synchronicity?: An analysis based on mediating effects of stakeholders' behavior [J]. *Journal of Management Sciences in China*, 2020, 23(9): 108 – 126. (in Chinese)
- [12] Kubick T R, Lynch D P, Mayberry M A, et al. The effects of regulatory scrutiny on tax avoidance: An examination of SEC comment letters [J]. *Accounting Review*, 2016, 91(6): 1751 – 1780.
- [13] 陈运森, 邓祎璐, 李 哲. 证券交易所一线监管的有效性研究: 基于财务报告问询函的证据 [J]. *管理世界*, 2019, 35(3): 169 – 185, 208.
Chen Yunsen, Deng Yilu, Li Zhe. Effectiveness of the front-line regulation of the Chinese stock exchanges: Evidence from comment letters [J]. *Management World*, 2019, 35(3): 169 – 185, 208. (in Chinese)

- [14] Brown S V , Tian X L , Tucker J W. The spillover effect of SEC comment letters on qualitative corporate disclosure: Evidence from the risk factor disclosure [J]. *Contemporary Accounting Research* , 2018 , 35(2) : 622 – 656.
- [15] 陈运森, 邓祎璐, 李哲. 非处罚性监管具有信息含量吗? ——基于问询函的证据 [J]. *金融研究* , 2018 , (4) : 155 – 171.
Chen Yunsen , Deng Yilu , Li Zhe. Does the non-penalty regulation have information content? Evidence from inquiry letters [J]. *Journal of Financial Research* , 2018 , (4) : 155 – 171. (in Chinese)
- [16] 张俊生, 汤晓建, 李广众. 预防性监管能够抑制股价崩盘风险吗? ——基于交易所年报问询函的研究 [J]. *管理科学学报* , 2018 , 21(10) : 112 – 126.
Zhang Junsheng , Tang Xiaojian , Li Guangzhong. Does the preventive regulation mitigate stock price crash risk? Evidence from comment letters [J]. *Journal of Management Sciences in China* , 2018 , 21(10) : 112 – 126. (in Chinese)
- [17] 李晓溪, 杨国超, 饶品贵. 交易所问询函有监管作用吗? ——基于并购重组报告书的文本分析 [J]. *经济研究* , 2019 , 54(5) : 181 – 198.
Li Xiaoxi , Yang Guochao , Rao Pingui. The regulatory role of stock exchange comment letters: Evidence from textual analysis of merger and acquisition plans [J]. *Economic Research Journal* , 2019 , 54(5) : 181 – 198. (in Chinese)
- [18] Kimbrough M D , Louis H. Voluntary disclosure to influence investor reactions to merger announcements: An examination of conference calls [J]. *Accounting Review* , 2011 , 86(2) : 637 – 667.
- [19] Rock K. Why new issues are underpriced [J]. *Journal of Financial Economics* , 1986 , 15(1 – 2) : 187 – 212.
- [20] Benveniste L M , Spindt P A. How investment bankers determine the offer price and allocation of new issues [J]. *Journal of Financial Economics* , 1989 , 24(2) : 343 – 361.
- [21] Hanley K W , Hoberg G. The information content of IPO prospectuses [J]. *Review of Financial Studies* , 2010 , 23(7) : 2821 – 2864.
- [22] Hanley K W , Hoberg G. Litigation risk , strategic disclosure and the underpricing of initial public offerings [J]. *Journal of Financial Economics* , 2012 , 103(2) : 235 – 254.
- [23] Bao Y , Datta A. Simultaneously discovering and quantifying risk types from textual risk disclosures [J]. *Management Science* , 2014 , 60(6) : 1371 – 1391.
- [24] Dyer T , Lang M , Stice-Lawrence L. The evolution of 10-K textual disclosure: Evidence from latent Dirichlet allocation [J]. *Journal of Accounting & Economics* , 2017 , 64(2 – 3) : 221 – 245.
- [25] Huang A H , Lehavy R , Zang A Y , et al. Analyst information discovery and interpretation roles: A topic modeling approach [J]. *Management Science* , 2018 , 64(6) : 2833 – 2855.
- [26] Bellstam G , Bhagat S , Cookson J A. A text-based analysis of corporate innovation [J]. *Management Science* , 2021 , 67(7) : 4004 – 4031.
- [27] 陈霄, 叶德珠, 邓洁. 借款描述的可读性能够提高网络借款成功率吗 [J]. *中国工业经济* , 2018 , (3) : 174 – 192.
Chen Xiao , Ye Dezhu , Deng Jie. Can readability of loan description promote lending success rate of online [J]. *China Industrial Economics* , 2018 , (3) : 174 – 192. (in Chinese)
- [28] Miller E M. Risk , uncertainty , and divergence of opinion [J]. *Journal of Finance* , 1977 , 32(4) : 1151 – 1168.
- [29] 邹高峰, 张维, 常中阳. 询价制度下中国 IPO 长期表现 [J]. *管理科学学报* , 2012 , 15(11) : 66 – 75.
Zou Gaofeng , Zhang Wei , Chang Zhongyang. The long-run performance of the initial public offerings after the book-building in Chinese stock markets [J]. *Journal of Management Sciences in China* , 2012 , 15(11) : 66 – 75. (in Chinese)
- [30] 李冬昕, 李心丹, 俞红海, 等. 询价机构报价中的意见分歧与 IPO 定价机制研究 [J]. *经济研究* , 2014 , 49(7) : 151 – 164.
Li Dongxin , Li Xindan , Yu Honghai , et al. Disagreement of institutional investors' bids and IPO pricing mechanism [J].

Economic Research Journal ,2014 ,49(7) : 151 - 164. (in Chinese)

[31]俞红海,李心丹,耿子扬. 投资者情绪、意见分歧与中国股市 IPO 之谜[J]. 管理科学学报,2015 ,18(3) : 78 - 89.

Yu Honghai , Li Xindan , Geng Ziyang. Investor sentiment , disagreement and IPO puzzle in China' s stock market [J]. Journal of Management Sciences in China ,2015 ,18(3) : 78 - 89. (in Chinese)

Registration system review inquiry and IPO information disclosure on STAR market: Textual analysis based on LDA topic model

YU Hong-hai , FAN Si-yu , WU Liang-yu , MA Zhi-bin

School of Management and Engineering , Nanjing University , Nanjing 210093 , China

Abstract “Taking Information disclosure as the core” is the requirement of the registration-based IPO system. In the context of the pilot registration-based IPO system on STAR market , whether the content and quality of IPO information disclosure can be improved is a prerequisite for the market to play a decisive role in resource allocation. Based on a sample of 341 companies successfully listed on the STAR market from July 2019 to September 2021 , this paper examines the relationship between registration system review inquiries and IPO information disclosure. This main results are as follows. (1) At the IPO review stage , prospectuses with a lower degree of information disclosure and insufficient technical information disclosure will receive more review inquiry letters. (2) Registration system review inquiry letters can enhance the degree of prospectus disclosure , sentences' length , textual comprehensibility , and quantitative information; review inquiry letters focusing on a particular topic will lead to an increase in prospectuses' information disclosure on the corresponding topic. (3) Further analysis shows that review inquiry letters reduce the divergence of institutional investors' offering prices. The results can deepen the understanding of the regulatory role played by SSE review inquiries in the context of the pilot registration-based IPO system , and enrich the research related to the impact factors of IPO disclosure.

Key words: registration-based IPO system; review inquiry; information disclosure; LDA topic model