

公告溢价效应与资产定价：文本机器学习视角*

唐国豪 朱琳 陈世程

摘要：本文在中国股票市场中，针对上市公司公告文本数据，采用文本分析机器学习方法进行信息提取，旨在揭示公告信息与资产预期回报之间的关系以及对资本市场的影响渠道。本文首先依据监督式训练方法构造了基于公告的情感词典，并以此为基础采用机器学习方法对公告效应进行实证分析，最后从多个渠道探究了公告溢价效应的市场反馈机制。本文研究发现，基于机器学习的公告文本情感能显著预测股票收益率，且明确存在正向显著的公告溢价效应。异质性分析发现，公告效应在小规模、成长型公司中溢价显著；与国企相比，民营企业的公告效应更显著。机制分析发现，公告溢价产生的主要原因可能是由于散户投资者的过度关注。对于金融机构关注较多和信息披露质量较高的公司，公告溢价效应较弱。

关键词：公告溢价效应 文本分析 机器学习 市场反馈

一、引言

党的二十大报告指出，要加快发展数字经济，实施创新驱动发展战略。在经济金融领域，创新驱动发展离不开科技与数据要素的加持。2023年中央金融工作会议进一步强调了科技金融与数字金融的重要作用。随着经济数字化转型程度的加深和数字经济发展战略的落地，科技驱动与数据赋能显得尤为关键。因此，在目前资产定价的前沿研究中，金融机器学习等先进技术的应用和非传统大数据信息的深入挖掘，受到了广泛关注。

近年来，Gu et al. (2020)、李斌等(2019)、马甜等(2022)在中美金融市场中应用了各类机器学习方法，验证了诸如人工智能等以技术驱动的金融创新能够有效预测和解释金融市场收益。另一方面，诸如文本、音频、视频等大数据提供了很多传统数据所缺少的信息，大大拓展了研究的边界，甚至给研究范式与投资实践方式带来变革。不少学者指出，金融领域的前沿问题探索应建立在对金融大数据进行智能提取和分析的基础上(洪永淼、汪寿阳, 2021; Goldstein et al., 2021)。金融文本数据区别于传统的交易、财务数据，其规模庞大、词汇极多(符合高维特征)，且结构复杂(叙述结构等)。如何有效应用金融机器学习方法从非结构化的文本数据中挖掘有效信息，探究其与资产预期回报之间的关系，对于提升金融市场的定价效率具有积极意义。

在众多金融文本数据中，上市公司公告具有较高的分析价值。首先，上市公司公告来源于公司自身，由于需要满足合规性要求，其内容具有较高的权威性与可信度；其次，上市公司公告是公司向投资者进行信息传播的重要途径，既能展示公司经营发展状况，又能及时、公平地被投资者获取，从而影响股票收益率的走势。因此，对公告文本进行量化研究分析，对研究股票收益率变动的成因以及理解市场对公告的反馈机制有重要意义。本文旨在深入探究以下问题：第一，如何用科学的方法度量上市公司公告的文本信息；第二，中国股票市场是否存在公告溢价效应；第三，利用上市公司公

*唐国豪、朱琳(通讯作者)、陈世程，湖南大学金融与统计学院，邮政编码：410006，电子邮箱：ghtang@hnu.edu.cn, zhulin1@hnu.edu.cn, chensc1@hnu.edu.cn。基金项目：国家自然科学基金青年项目(72003062)。感谢匿名审稿专家的宝贵建议，文责自负。

告的披露是否能够提高市场定价效率,并探究资本市场对公告信息的反馈机制。

本文使用2000—2020年上市公司公告文本数据,通过有监督式词汇训练方法构造基于公告的动态情感词典,并运用机器学习方法判断公告的情感倾向,在此基础上研究上市公司公告信息与资产预期回报之间的关系。结果表明,第一,不同类别公告产生的溢价效应有显著的差别。利好类公告的股票收益率表现优异,能获得正的显著收益率;利空类公告表现糟糕,收益率显著为负。第二,公告情感词典能显著区分公告效应的方向。在全样本中空投资组合的平均年化收益率为20.04%($t=5.28$),在不确定类公告中正面公告的平均年化收益率为30.98%($t=4.22$);且所有多空投资组合风险调整后的收益率依然显著。第三,在对公告效应的异质性分析中,公告效应在小规模、成长型公司中溢价显著;与国企相比,基于公告的投资组合在民营企业的收益率更高;针对不同行业,公告效应在运输业、制造业和文体服务业更为显著。除此之外,本文对公告溢价效应的市场反馈机制的探究显示,金融机构关注度越高和信息披露质量越高的公司,公告溢价效应较低;而投资者关注度对公告效应的影响效果则相反——散户投资者关注度越高的投资组合公告溢价效应越高。这说明散户投资者倾向于对公告信息进行过度解读,存在过度交易的行为,可能是产生公告溢价效应的主要原因。

本文研究的主要贡献为:(1)提供了研究公告效应的新视角。与以往公告效应的文献中对某一类公告效应使用事件研究法不同,本文对全部公告使用文本分析方法提取公告信息,从信息价值和资产定价视角研究公告效应。(2)本文丰富了中文金融文本分析的研究。本文在对公告情感词典的构建过程中,首次将股票收益率作为权重有监督式地计算词汇的情感方向,使字典与股票收益率之间的联系更加紧密,且能对词典及时进行动态调整。(3)本文提供了量化投资的新思路。本文使用机器学习方法充分挖掘上市公司公告中的定价信息,排除人为定义中潜在的主观性,在获得良好收益同时,排除了前视偏差,对二级市场的投资决策具有重要的参考价值。

本文余下部分的组织架构如下:第二部分进行文献综述;第三部分介绍样本选取与数据处理;第四部分详细描述文本信息的提取与量化;第五部分和第六部分分别是实证分析与稳健性检验;第七部分是对公告溢价效应的市场反馈机制做进一步分析;第八部分是结论与建议。

二、文献综述

关于上市公司公告的披露是否会对股票市场产生显著影响的问题,国内外许多学者进行了大量的实证研究。已有研究表明,公司公告会引发所属公司的股价异常波动,形成“公告效应”。不过,市场对于不同类型公告的反应存在差异,而全面分析所有类型公告产生效应的研究目前稍显不足。关于盈余公告,Ball & Brown(1968)最早提出盈余公告后漂移现象(PEAD),表明市场价格并非立即反应盈余信息,而是在一段时间后才逐渐吸收这些信息(Dellavigna & Pollet, 2009;吴世农等, 2005;孔东民、柯瑞豪, 2007);公开增发公告对市场产生显著的负面效应(Hess & Frost, 1982;Asquith & Mullins, 1986;刘力等, 2003;孔东民、付克华, 2005);而市场对定向增发公告的反应显著为正(Wruck, 1989;Hertzel & Smith, 1993;章卫东、李德忠, 2008)。此外,增减持公告对市场也有显著影响,上市公司大股东的股份增减持行为可能传递出市场价格未能准确反映公司的内在价值(Cheng & Lo, 2006;黄建欢等, 2009)。

在金融文本研究领域,随着大数据技术的发展,日益关注信息价值的增量预测能力。在对美国市场的研究中,学者们针对不同的文本提取情绪指标:Das & Chen(2007)爬取的是股票留言板中的内容,Tetlock(2007)统计了《华尔街日报》金融专栏新闻,Loughran & McDonald(2011)分析的是公司年度报表(10-K filings),Garcia(2013)使用《纽约时报》中的词汇,Huang et al.(2014)解读的是分析师报告中的信号,而Jiang et al.(2019)则关注公司的财务披露。他们构建的情绪指标均发现对股票市场收益有显著的预测作用,说明金融文本数据的情感分析在市场预测中扮演着重要角色。

针对中国市场,游家兴和吴静(2012)研究新闻媒体对金融市场存在的负面效应;张纯和吴明明

(2015)运用LDA(latent dirichlet allocation)主题模型发现媒体跟踪报道的信息广泛地影响了A股市场投资者;谢德仁和林乐(2015)发现年度业绩说明会上的正面语调能显著预测公司未来的业绩;罗炜等(2017)发现财经媒体关于创业企业报道中的正面情绪能够预测创业企业的发展前景;而姜富伟等(2021a)拓宽了研究对象,分析中国人民银行货币政策执行报告,探究货币政策报告的文本信息与宏观经济和股票市场的关系。

中美市场的研究均表明文本情感分析在市场预测中具有显著价值,然而,针对上市公司公告文本,现有研究仍存在如下拓展空间:首先,缺乏专门针对公告文本的词典,其专业性和特殊用语导致目前词典的覆盖率极低,难以完成对公告文本的深入分析;其次,现有词典未考虑股票收益对情感色彩的影响,对于资产定价研究而言,缺少关键的监督变量;最后,由于文本数据量级较大,难以收集与处理,现有文献中构建的文本情绪时间跨度均较短,并且没有考虑市场时变性的影响,这些因素都对研究公告文本情感与金融市场的关系产生较大制约。

近年来,金融学开始探讨机器学习对现有研究范式的价值和意义,美国市场的大量研究尝试使用包括机器学习方法在内的大数据方法构建预测和定价模型,来解决高维数据给传统研究方法带来的挑战。例如Light et al.(2017)采用“偏最小二乘”(partial least squares, PLS)来检验公司特征对预期收益率的预测能力;Kozak et al.(2020)和Kelly et al.(2019)分别运用主成分分析(principal components analysis, PCA)和增量主成分分析(instrumented PCA, IPCA)方法,研究发现仅依赖于少数主成分的模型就可以较好地预测截面收益;Gu et al.(2020)对比了包括神经网络在内的各类机器学习模型在美国股票市场的应用,结果发现非线性机器学习模型的预测表现可以有效地超越传统线性模型。

不少机器学习研究也着眼于中国市场。苏治等(2017)对近年来在金融实证里应用的深度学习模型进行了总结归纳,并对未来中国金融市场中机器学习的应用给出了评述;李斌等(2019)运用12种机器学习算法构建股票收益率预测模型,发现非线性方法能获得较高的投资绩效;马甜等(2022)研究表明深度学习(生成式对抗网络)方法在收益率预测精度和因子投资绩效上均有很大提升。本文紧跟前沿,在判断公告情感倾向时使用岭回归分类器,排除人为因素对公告信息判断的干扰,从动态视角探索公告信息的时变性,在实证上丰富了中国市场股票截面收益影响因素的研究。

综上所述,目前学者对公告效应的实证研究多采用事件研究法,但该方法不能全面提取公司公告中的信息。同时,现有中文文本词典对上市公司公告的分析效果有限。因此,有必要借助前沿的机器学习和文本分析等技术,对公司公告文本信息进行综合全面地分析,深入研究其与资产预期回报之间的关系以及对资本市场的影响。

三、数据处理

(一)样本选取及数据来源

本文的研究对象是中国股票市场的上市公司,主要包括主板上市以及在创业板上市的所有公司。本文选取的样本区间为2000年1月1日至2020年12月31日,数据的频率为日度。

首先,上市公司公告文本数据来源于Wind数据库。然而,在处理过程中发现Wind数据库缺失了2011—2012年间的大部分数据。为弥补这一缺失,本文采用国泰安数据库(CSMAR)中的公告数据作为补充,以确保样本的完整性。

其次,为研究公司股票的收益率和公司特征情况,本文从Wind数据库获得诸如日度股票收益率数据、公司市值数据、特别处理(special treatment, ST)与特别转让(particular transfer, PT)股票数据等数据。另外,有关因子模型的数据,如资本资产定价(CAPM)模型、Fama-French三因子、五因子(FF3、FF5)模型等,均来源于国泰安数据库。

最后,进行稳健性检验和进一步分析时所用到的公司股权性质文件、行业分类、被研报关注度、

被分析师关注度以及上市公司信息披露考评文件等数据,均来自于国泰安(CSMAR)数据库。

本文按照资产定价领域的主流研究范式(Fama & French, 1993),在选择样本研究区间时考虑了几个关键因素。首先,在20世纪90年代中国股票市场建立初期,由于上市公司数量较少,可获得的样本量有限,为避免样本偏差对研究结果的影响;其次,1996年底上交所和深交所实行10%涨跌幅限制的交易机制,这一交易机制的改变对上市公司的股票收益产生了较大影响,为避免机制变化对研究结果的干扰;最后,我国在2000年左右加入了世界贸易组织(WTO),证券市场进行了一系列的改革,市场透明度与会计数据的质量显著提高。因此,本文将研究的起点选在2000年1月1日。

(二)数据预处理

为确保投资组合的可行性和实证分析结果的严谨性,需要对数据进行预处理。首先,排除了被标记为ST(特别处理)和PT(特别转让)状态的公司,因为这些公司可能面临监管困境和流动性不足,且存在较高的退市风险。其次,剔除了上市未满半年的股票数据,以应对IPO抑价效应,因为公司上市初期股票价格波动较大。

初始公告数据包含4600009条记录,在对数据预处理后,研究样本期间(2000年1月1日至2020年12月31日)的有效样本数量为4345741条。图1展示了该期间月度公告数量的变化。总体来看,公告数量呈逐年上升趋势,每年4月份呈现高峰,8月份为小高峰,形成了锯齿状上升的曲线。

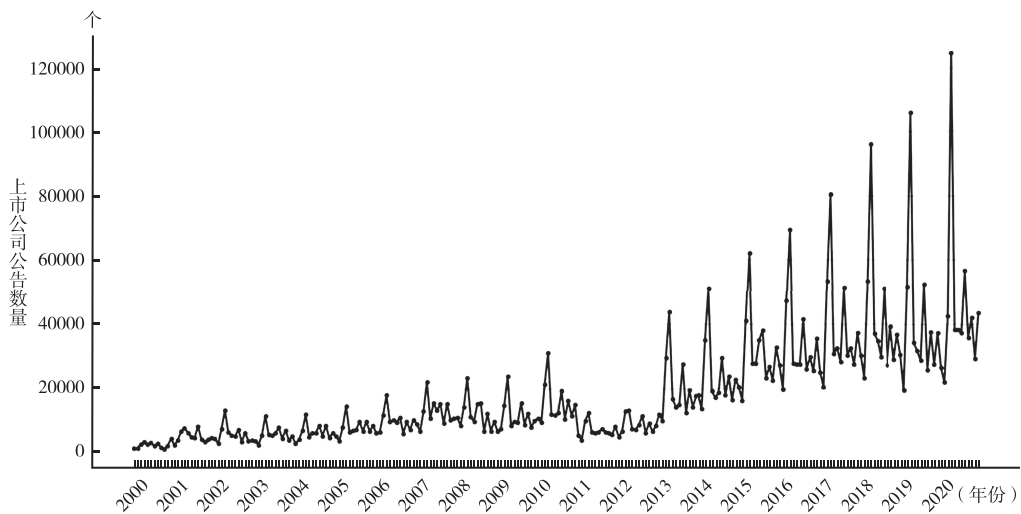


图1 2000年1月至2020年12月月度公告数量

四、文本信息的提取与量化

(一)文本分词方法

文本分析的基础是对公告文本进行分词处理,将文本准确的分成具有中文经济含义的单词是研究的重点。本文对公告标题使用的分词方法是Python中的jieba分词,但是jieba中自带的通用词库无法准确地识别出金融文本中的专有名词。根据目前在中国金融市场中文本分析研究(姜富伟等, 2021b)的经验,我们选择搜狗词库中下载量较大的金融、经济、财会词库,将这些专业术语添加到Python中开源的jieba模块词库中,以提高算法在金融文本中的分词效果。同时,针对公告标题专业术语多、语言简练概括等特点,我们还加入上市公司缩写词库,上市公司全称词库以及金融机构名称库,以保证机构名词的分词准确性。这一系列词库的加入提高了分词模型对专业词汇的准确识别,分词效果显著提升。

(二)公告情感词典

目前关于金融文本的研究中有两个使用率和认可度较高的情感词典。Loughran & McDonald (2011)在研究中指出,与通用文本相比经济金融领域的文本用词和情感倾向明显不同,基于通用文本研发的情感词典难以有效适用于金融领域文本分析。为此,两位学者研发了LM词典,该词典在目前英文经济金融文本分析中占有重要地位。而针对中文财经新闻文本,姜富伟等(2021b)开发了适用于中文经济金融文本的情感词典。较之于常用的知网的情感词典,该词典在中文的经济金融文本上展示出了更显著的适用性,为后续的中文金融文本情感分析打下了良好基础。

然而,以上两个词典均不能满足本文研究的需求。一方面,本文研究的上市公司公告文本与其他金融文本存在显著差异。公告属于公布性公文,公告文本(特别是公告标题)的用语特殊,相较于其他金融文本(如财经新闻),具有模式化、^①专业性强、严肃准确、简洁精炼等特点,由此导致一般词典中的单词在公告文本中的覆盖率极低,无法完成对公告文本的分析。另一方面,本研究中收益率信息具有重要的指导作用,仅从文本中无监督地提取信息的手段可能忽视关键信息。因此,本文针对公告文本信息,参考Ke et al.(2019)和Garcia et al.(2023)提出的有监督的文本挖掘方法,构造了一个全新的动态文本情感词典,我们将其称之为“公告情感词典”,专门适用于公告文本分析。

该公告情感词典的构建过程包括三个主要步骤。首先,准备构造词典的语料。词典的语料来源于公告标题本身,准备分为两步:(1)将每条公告标题与其次日的收益率匹配,生成公告标题对应次日收益率变动表;(2)将jieba分词后的单词与收益率变动表结合,统计每个单词每年的累计次日收益率和频率。

其次,对于分词完成后的词典进行单词的筛选。筛选过程是剔除无情感意义的单词:(1)根据单词的词性,删除数词、简称词、形容词、助词、连词、介词以及地名词;(2)删除在一年中仅出现一次或者出现次数异常高的单词。前者是因为分词得到了关于某公司相对专有的单词,这对于我们的投资组合构建不具有参考意义;后者则是因为分词得到了大量公司在公告标题中都会包含的单词,例如“公告”“有关”等,均不包含情感色彩。

最后,本文采用有监督方式定义词典中每个词的情感,排除人为主观因素对定义的潜在影响。传统的主题分类模型(Blei et al., 2003)相当于对文本单词进行无监督的降维,不设定标签。本文采用有监督的文本挖掘方法,将单词与股票收益率相关联,用每个单词一年内的收益率作为情感倾向标签,从而排除短期波动影响。记公告发布日为 t , r_{t+1} 为公告对应公司下一个交易日的收益率。 r_u 是上涨(up)词汇 u 的年平均收益率, r_d 是下跌(down)词 d 的年平均收益率。

$$r_u = \frac{\sum_{i=0}^{N_i} r_{i,t+1}}{N_i}, r_d = \frac{\sum_{j=0}^{N_j} r_{j,t+1}}{N_j} \quad (1)$$

其中, N_i 为包含上涨词 i 且 $r_{i,t+1} > 0$ 的公告数量, N_j 为包含所有下跌词 j 且 $r_{j,t+1} \leq 0$ 的公告数量。基于计算得到的 r_u 和 r_d ,相应单词被分类到正面词袋和负面词袋。本研究根据上述步骤逐年动态更新,构建了动态公告情感词典。

本文构造的公告词典有三个方面的创新:首先,针对公告用语的特殊性采取了具有针对性的处理,提高词典信噪比;其次,词典构造过程中考虑了股票收益率,以年度平均收益率作为权重赋予单词情感倾向,有效排除了非系统性噪音,加强了词典与股票收益率的关联;最后,本词典是逐年动态更新的构建方式,考虑了时间维度上的词汇变化,增强了词典的时效性和适用性。

^①公告标题具有固定的模式,部分公告类型的标题是由发文单位、事由、文种名称构成,而事由由部分多使用介词“关于”,例如“董事会关于募集资金2019年度存放与使用情况的专项报告”“关于公司可转换公司债券评级调整的公告”“关于对外投资发起设立产业基金的公告”,等等。

(三)机器学习分类模型

本小节,我们将利用已构建的公告情感词典对文本进行情感分析。考虑到词典构成的稀疏矩阵(sparse matrices)特征,本研究采用岭回归分类器进行建模,以应对稀疏数据进行分类的处理需求。

与各种分类算法相比,岭回归分类器在准确性和训练(预测)时间之间有很好的权衡。一方面,岭回归分类器分类准确度高,由于其采用的是惩罚最小二乘损失,根据不同的计算性能需求选择适合的数值求解器,从而优化模型区分能力;另一方面,岭回归分类器运算速度快、训练时间短。相较于随机森林等非线性算法,具有更高的运算速度和更低的预测成本。特别是在处理高维数据时,线性模型因其适应性强,在高维特征空间中常展现出更佳的性能。

考虑一个由 n 个公告标题组成的集合,和一个由 m 个单词组成的情感词典。我们将使用向量 $d_i \in \mathbb{R}_+^m$ 表示标题 i , $d_{i,j}$ 是出现在标题 i 中情感词 j 的情感倾向收益率。在矩阵形式中,这是一个 $n \times m$ 的文档矩阵 $D, D = [d_1, d_2, \dots, d_n]'$ 。

若回归分析中因变量 Y 为定性变量,则可通过哑变量(dummy variable)进行编码。在每个交易日 t 对市场上所有股票的收益率进行排序,将每日收益率排名前 30% 的股票标记为正面类别 1,后 30% 的股票标记为负面类别 -1,剩余样本标记为中性类别 0。具体如下所示:

$$y_{i,t} = \begin{cases} 1, & \text{正面} \\ 0, & \text{中性} \\ -1, & \text{负面} \end{cases} \quad (2)$$

对于多分类问题,岭回归分类器视为多输出回归,预测类对应的输出值最大。岭回归算法是在多元线性回归基础上对系数大小施加惩罚的线性模型,其回归模型与多元线性回归模型无异。岭回归的模型为:

$$y_{i,t} = \beta_t d_{i,t-1} \quad (3)$$

其中, $d_{i,t-1}$ 是公告 i 的信息特征。岭回归的损失函数则是在最小化残差平方和的基础上加正则化项。目标方程可以写为:

$$\mathcal{L}(\beta; \cdot) = \frac{1}{NT} \sum_{i=1}^N (y_i - f(d_{i,t-1}; \beta_i))^2 + \phi(\beta; \cdot) \quad (4)$$

$$\phi(\beta; \lambda) = \lambda \sum_{j=1}^N \beta_j^2 \quad (5)$$

其中, $\mathcal{L}(\beta; \cdot)$ 为岭回归的损失函数, $\phi(\beta; \cdot)$ 为正则化项, λ 为正则化系数。通过求解最小化 $\min_{\beta} \mathcal{L}(\beta; \cdot)$ 得到回归的估计值 $\hat{\beta}_t$ 。

岭回归算法为实现三分类任务,构造三个独立的岭回归模型 $Ridge(\hat{y}_{i,t} = k) = \hat{\beta}_t D, k = -1, 0, 1$ 。 k 代表不同的情感倾向类别。模型最终判定的情感倾向是基于三个模型估计值中的最大值对应的类别。^①

(四)公告文本量化指标

基于公告情感词典和岭回归分类器,本研究量化了公告标题的情感倾向,分别为正面情感公告($\hat{y}_{i,t} = 1$)、负面情感公告($\hat{y}_{i,t} = -1$)以及中性情感公告($\hat{y}_{i,t} = 0$)。为避免出现前视偏误,情感倾向性的评估基于该时间点之前的情感词典。本文统计了正面情感公告和负面情感公告的数量,结果如图 2 所示,其中,正面情感公告数量为图中的实线所示,负面情感为虚线所示。

^①为确保机器学习算法的有效性,文本以滚动窗口的方式划分训练、验证和测试数据集。具体而言,我们选取了连续 3 年的数据作为训练集,紧接着的 1 年数据作为验证集,以及之后的 1 年数据作为测试集,逐年滚动地对模型调参。

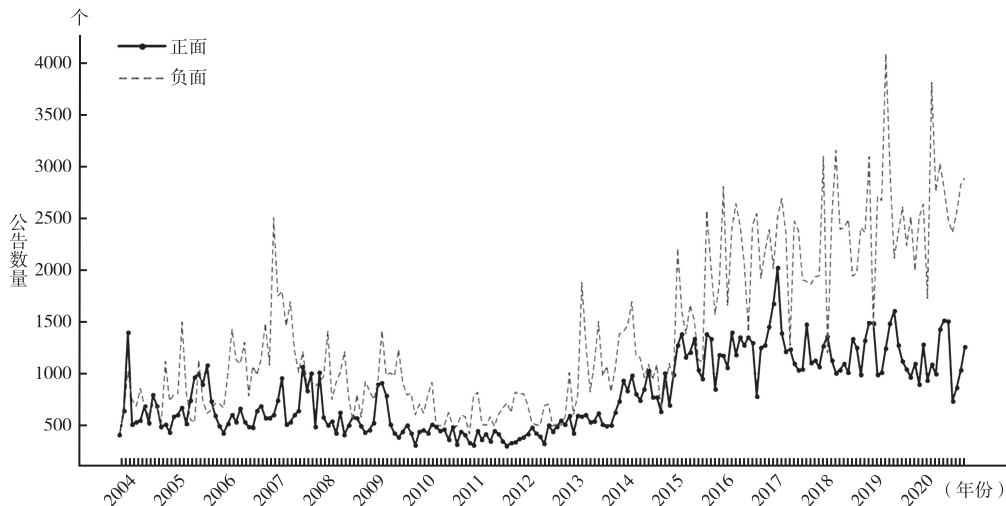


图2 正面情感公告与负面情感公告的数量

五、实证分析

为验证中国股票市场上公告溢价效应的存在性,本文将从三个维度开展实证分析。首先,分析不同类别公告所产生收益的差异性;其次,评估基于公告情感词典构建的投资组合能否实现显著的溢价效应;最后,针对不确定性类别的公告,探讨基于情感词典构建的投资组合是否同样能够带来显著的溢价效应。通过这三个层面的逐步深入,旨在全面验证公告溢价效应的存在性。

(一)基于公告类别的溢价分析

本文依据 Wind 数据库中公司公告的分类标准,将所有公告标题归类为 61 个不同类别。其中,股东大会类的公告数量最多,占比达 7.96%,董事会公告、中介公告、资金投向公告、关联交易、收购兼并等类别的公告也较为常见,各自占总公告数量的 3% 以上。为避免因样本量过小而导致的实证结果偏差,本研究剔除了占比不足 1% 的公告类别,包括配股上市、招股说明书、上市公告书、申报稿和发行结果五个类别。

基于公告类别,本文构建了相应的投资组合。令公告发布日为 T 日,在下一个交易日(即 $T+1$ 日)构造投资组合,具体为 $T+1$ 日开盘时买入对应类别的所有股票,并持有 一定天数,最终得到各类公告在不同持有期下构造的投资组合的收益情况,计算收益率、夏普比率等指标评估各投资组合的表现。^①

表 1 展示了基于公告类别构建的投资组合的绩效评估结果。其中,机构调研公告的表现最好,夏普比率高达 1.95,年化收益率为 62.65%;员工持股公告子样本的夏普比率为 1.38,年化收益率能达到 49.47%。相反,风险提示、特别处理及终止上市公告的投资组合表现不佳,年化收益率为负,分别为 -23.66%、-23.75% 以及 -27.13%。该结果验证了不同类别公告所产生的收益有显著的差异,且与经济学直觉相契合。机构调研公告通常被视为公司积极信号的释放,吸引投资者关注高质量及具有研究价值的企业;员工持股公告则体现了管理层对公司未来的信心,增强了投资者对企业价值的正面预期。反之,风险提示、特别处理和终止上市等公告揭示了公司的潜在经营问题,通过公告向交易市场释放的消极信号会导致股票价格的下跌。

(二)基于公告情感词典的溢价分析

本文利用机器学习技术(岭回归分类器)预测的公告情感倾向(正面、中性、负面)来构建投资组

^①由于篇幅限制,文中仅展示了具有代表性的几类公告的绩效表现,全部公告类别持有不同天数的投资组合的表现,留存备案。

合,深入分析公告情感倾向对股票市场反应的影响,验证公告溢价效应的存在性问题。具体做法是,在公告发布的次日构造投资组合,并分析通过买入正面情感公告股票同时卖空负面情感公告股票所构建的多空对冲投资组合的收益表现。

表1 不同类别公告构造的投资组合绩效分析

| 公告类别 | 夏普比率 | 年化收益率% | 年化波动率% | 较基准胜率 ^① |
|------|-------|--------|--------|--------------------|
| 机构调研 | 1.95 | 62.65 | 32.10 | 0.58 |
| 员工持股 | 1.38 | 49.47 | 35.98 | 0.54 |
| 股权激励 | 0.96 | 28.73 | 30.04 | 0.53 |
| 业绩快报 | 0.79 | 19.4 | 24.49 | 0.52 |
| 增发预案 | 0.69 | 25.51 | 36.98 | 0.49 |
| 风险提示 | -0.67 | -23.66 | 35.16 | 0.47 |
| 特别处理 | -0.92 | -23.75 | 25.94 | 0.40 |
| 终止上市 | -0.95 | -27.13 | 28.44 | 0.44 |

注:表中投资组合采用市值加权,年化收益率与年化波动率均以百分比形式汇报。

表2详细展示了基于公告情感倾向构建的投资组合在不同持有期的平均年化收益率情况。分析结果显示,基于公告词典和机器学习方法,不同情感分类的投资组合收益率具有明显的区分度,且能实现较高的年化收益率。例如,仅持有1天的正面情感公告的投资组合的年化收益率为24.04%($t=3.00$),而买入正面卖空负面公告的对冲投资组合的年化收益率为20.04%($t=5.28$),结果说明基于机器学习构造的公告情感词典对公告的有效性,能够在实际的投资中创造较高回报。此外,表2还展示了投资组合在不同持有期的收益率变化情况。持有2天的年化收益率为18.35%($t=2.24$),延长至30天的年化收益率为20.54%($t=2.74$),表明正面情感的投资组合均能获得显著的正收益率,具有较强的稳健性。进一步分析,正面公告的发布不仅在短期内迅速反映到股价中,还会对股票价格有持续的影响,显示出一定的上涨惯性。对于负面情感公告的投资组合,持有2~30天时收益率虽然呈现上升趋势,分别在7.92%~12.38%之间,但是 t 值并不显著,这可能与负面信息迅速被市场吸收以及市场整体呈上涨趋势有关。

表2 基于公告情感词典构造的投资组合收益表现

| 投资组合 | 持有天数 | | | | | |
|-------|--------------------|--------------------|-------------------|-------------------|-------------------|--------------------|
| | 1 | 2 | 5 | 10 | 20 | 30 |
| 正面 | 24.04*** [3.00] | 18.35** [2.42] | 17.69** [2.34] | 17.90** [2.38] | 18.49** [2.45] | 20.54*** [2.74] |
| 中性 | 16.12* [1.83] | 13.68* [1.77] | 13.67* [1.79] | 14.32* [1.87] | 10.38 [1.35] | 11.70 [1.53] |
| 负面 | 3.99 [0.51] | 7.92 [1.05] | 9.52 [1.26] | 12.38* [1.65] | 11.46 [1.52] | 11.53 [1.54] |
| 正面-负面 | 20.04*** [5.28] | 10.42*** [5.28] | 8.17*** [4.05] | 5.52*** [2.73] | 7.02*** [3.26] | 9.01*** [4.32] |

注:表中投资组合采用市值加权,年化收益率以百分比形式汇报,方括号内为各投资组合的 t 值。***、**、*分别表示在1%、5%、10%的水平上显著,下同。

(三)公告情感词典在不同公告分类中的表现

在前文探究公告类别对股价影响的研究中,我们发现部分公告类别因其积极的信号而带来明确可观的回报,这类公告被归纳为“利好类公告”,包括机构调研、员工持股、股权激励;相反,有的类别

^①较基准胜率表示投资期内投资组合收益率大于基准收益率的次数占比。此处的基准为上证综合指数。

公告一经披露必能引起股价的下跌,我们将之总结为“利空类公告”,分别是特别处理、终止上市、风险提示。除上述两类外,大多数公告的投资回报不稳定,其收益方向难以预判,我们将这类公告称为“不确定类公告”。本文将聚焦于这三大类公告,尤其是“不确定类公告”,利用公告情感词典对它们进行进一步的情感分类分析,探索在不同类别公告中,尤其是在不确定性较高的公告中,基于公告情感词典构建的投资组合表现,这一分析有助于深化我们对公告情感倾向与股市表现关系的理解。

表3 公告情感词典在不同分类中的投资组合表现

| 投资组合 | 利好类公告 | | 不确定类公告 | | 利空类公告 | |
|-------|--------------------|--------------------|--------------------|--------------------|------------------|------------------|
| | 持有1天 | 持有5天 | 持有1天 | 持有5天 | 持有1天 | 持有5天 |
| 正面 | 33.68*** [4.21] | 25.12*** [3.82] | 30.98*** [4.22] | 20.51*** [3.41] | 4.68 [0.75] | 3.87 [0.85] |
| 负面 | 22.11*** [2.82] | 15.36** [2.35] | 15.06** [1.97] | 9.97 [1.64] | -9.70 [-1.26] | -3.26 [-0.59] |
| 正面-负面 | 11.57 [1.33] | 9.76* [1.68] | 15.52* [1.82] | 10.05* [1.76] | 14.93 [1.56] | 7.76 [1.17] |

注:表中投资组合采用市值加权,年化收益率以百分比形式汇报,方括号内为各投资组合的t值。

表3展示了在利好类、不确定性类和利空类公告样本中基于公告情感字典构造投资组合的收益表现。表3左侧的结果表明,公告情感词典在利好类公告中的投资组合收益的区分度不大,即使被分为负面公告的投资组合也会有显著的年化收益率22.11%($t=2.82$);表3右侧的结果显示,在利空类公告中,尽管情感词典的区分度略有提高,但收益差异并不显著。这两部分结果表明,无论是利好还是利空类公告,其表现似乎并不受公告用词情感的显著影响,显示出一定的收益稳健性。这可能是因为上市公司在公告中使用的语言经过精心挑选,旨在树立良好的企业形象,而投资者能够透过表面保持自己的辨别能力。

针对“不确定类公告”,运用情感词典进行再分类,以检验基于情感词典构建投资组合的稳健性。如表3中间结果所示,购买并持有1天正面词汇的公告的收益率为30.98%($t=4.22$),正面负面多空组合的收益率为15.52%($t=1.82$),结果表明,公告情感词典能在不确定收益方向的样本中有效地区分出有收益潜力的股票,从而为投资者带来较高的收益。

六、稳健性检验

前文已经论证了本文构造的投资组合能够获得显著收益率并且有高夏普比率,接下来将计算风险调整的收益率以及进行异质性分析,从多角度研究公告溢价效应的稳健性。

(一) 风险调整后的收益表现

为了评估基于公告信息构造的投资组合收益的稳健性,特别是考虑到优异的业绩表现可能意味着有更多的风险承担,本文将采用不同的资产定价模型进一步检验风险调整后的收益表现(risk-adjusted performance),即超额收益率。具体而言,分别计算CAPM- α 、FF3- α 、FF5- α 来评估投资组合的超额收益率,这些模型分别是包含市场溢价的CAPM模型,包含市场溢价、规模因子和账面市值比因子的Fama-French三因子(FF3)模型和在三因子(FF3)的基础上外加盈利因子和投资因子的Fama-French五因子(FF5)模型。

表4 Panel A展示的是前文所有投资组合中不同类别的公告投资组合的风险调整后的年化收益率以及各组t值(方括号中)。结果显示,“利好类公告”风险调整后的收益率依然显著,如机构调研公告的CAPM- α 为43.67%($t=6.11$),FF3- α 为39.97%($t=7.67$),FF5- α 为38.16%($t=7.58$),而“利空类公告”依然亏损,属于其中风险提示公告的CAPM- α 、FF3- α 和FF5- α 分别为-20.06%($t=-2.31$)、-21.44%($t=-2.52$)和-21.50%($t=-2.54$),均在1%水平上显著。

表4 Panel B展示的是基于公告情感字典构造的买入正面公告卖空负面公告的多空投资组合的

风险调整的年化收益率。结果表明,首先,在全样本中,多空投资组合的风险调整收益率依然表现稳健,CAPM- α 为20.20%($t=4.94$),FF3- α 为20.49%($t=5.03$),FF5- α 为20.25%($t=4.99$)。其次,对不确定类公告运用情感字典,公告情感的区分度依然显著,多空组合的CAPM- α 为15.55%($t=1.81$),FF3- α 为16.03%($t=1.86$),FF5- α 为15.06%($t=1.75$)。最后,在利好类公告和利空类公告的情感词典的投资组合,虽然能获得收益但均不显著,结果和前文保持一致。这些发现不仅再次证实了公告溢价效应的存在性,还验证了公告情感词典在构建投资组合时的有效性和稳健性。

表4 风险调整后的收益表现

| Panel A: 基于公告分类的风险调整收益率 | | | | | | | |
|---------------------------|--------------------|--------------------|--------------------|-------|----------------------|----------------------|----------------------|
| 公告类别 | CAPM- α | FF3- α | FF5- α | 公告类别 | CAPM- α | FF3- α | FF5- α |
| 机构调研 | 43.67*** [6.11] | 39.97*** [7.67] | 38.16*** [7.58] | 风险提示 | -20.06** [-2.31] | -21.44** [-2.52] | -21.50** [-2.54] |
| 员工持股 | 33.85*** [2.95] | 33.79*** [3.23] | 31.12*** [3.01] | 特别处理 | -17.47*** [-2.75] | -18.04*** [-2.87] | -17.72*** [-2.83] |
| 股权激励 | 17.03** [2.24] | 12.69* [1.84] | 11.88* [1.75] | 终止上市 | -9.49 [-1.37] | -10.33 [-1.52] | -9.44 [-1.38] |
| Panel B: 基于公告情感字典的风险调整收益率 | | | | | | | |
| 正面-负面 | CAPM- α | FF3- α | FF5- α | 正面-负面 | CAPM- α | FF3- α | FF5- α |
| 全样本 | 20.20*** [4.94] | 20.49*** [5.03] | 20.25*** [4.99] | 利好类公告 | 11.33 [1.27] | 11.24 [1.26] | 10.89 [1.22] |
| 不确定类公告 | 15.55* [1.81] | 16.03* [1.86] | 15.06* [1.75] | 利空类公告 | 16.04 [1.51] | 16.95 [1.59] | 16.10 [1.51] |

注:表中CAPM- α ,FF3- α 和FF5- α 均以百分比形式汇报,方括号内为对应的t值。

(二)公司特征异质性分析

本文将深入探讨公告溢价效应的异质性,此分析不仅有助于确认公告效应的稳健性,还能进一步理解公告效应在不同公司特征之间的差异化表现。

为探究公告溢价受账面市值比、公司规模、公司盈利能力的影响程度,本文基于不同公司特征进行双变量排序投资组合分析。具体而言,每日初,将上市公司分别按照上一交易日的账面市值比、规模、公司盈利能力(股权收益率)(Jiang et al., 2018)进行从高到低排序,分为三组,同时独立地将上市公司按照发出公告的情感倾向性分为正面、中性和负面三组,得到共9(3×3)个投资组合,计算市值加权的投资组合的年平均收益率与t值,结果如表5所示。

表5 公告溢价效应在不同公司特征下的表现

| 投资组合 | 规模 | | 账面市值比 | | 盈利能力 | |
|-------|--------------------|------------------|--------------------|--------------------|--------------------|-------------------|
| | 小 | 大 | 成长 | 价值 | 低 | 高 |
| 正面 | 41.25*** [4.25] | -0.36 [-0.04] | 17.02* [1.74] | 28.41*** [3.23] | 18.94* [1.81] | 20.29** [2.30] |
| 负面 | 11.75 [1.25] | -5.02 [-0.60] | -17.13* [-1.83] | 14.92* [1.73] | -4.01 [-0.49] | 8.11 [0.89] |
| 正面-负面 | 29.50*** [4.56] | 4.65 [0.84] | 34.15*** [4.85] | 13.49** [2.56] | 22.95*** [3.28] | 12.18** [2.10] |

注:表中为双变量排序结果,投资组合采用市值加权,收益率以百分比形式汇报,方括号内为各组t值。

表5的结果显示,除大规模组以外,基于情感公告构造的多空对冲组合在不同公司特征下的子样本中都产生了显著正收益,这说明公告效应的稳健性。分别来看,投资组合在小规模公司中的股票溢价更高,年化收益率为29.50%($t=4.56$),显著高于大规模公司的4.65%($t=0.84$),这与中国市场中显著存在的规模效应(Liu et al., 2019)有关,小市值公司在规避严格的IPO审核制度的反向并购中视为潜在的壳公司。公告效应在成长型和价值型公司中都具有很高的股票溢价,但在成长型

公司中的溢价能力更高,为34.15%($t=4.85$),高于价值型公司的13.49%($t=2.56$)。而关于公司的盈利能力,正面公告在盈利能力高的公司的收益率为20.29%($t=2.30$),高于低盈利能力的18.94%($t=1.81$),但多空组合的收益率却与此相反。

(三) 股权性质分析

与发达国家市场中私营企业占主导、国企为辅的经济条件不同,国有企业是我国社会主义市场经济体系的经济支柱。在此背景下分析不同股权性质公司对公告溢价效应的影响尤为重要,尤其是探究国有企业和民营企业之间的差异。

本文根据中国上市公司股权性质文件将公司分为国企、民营、外资和其他四个类别,但由于外资和其他两类的样本有限,本文着重研究国企和民营两类公司。具体而言,分别在国企和民营两类企业内,按照公告情感分类,计算各投资组合的年度平均收益率与相对应的 t 值。表6中的左边两列展示了公告情感词典组合在不同股权性质企业中的表现。结果显示,国企和民营企业均能获得显著的超额收益率,但民营的收益率显著高于国企,分别为15.07%($t=2.70$)和20.01%($t=3.40$),公告效应在民营企业中更加显著,可能是因为民营企业经营规模较小、融资约束大等特点所导致的更高的公告溢价效应,这个结果提示了公告效应更可能从错误定价角度而非风险角度进行解释。更进一步,由于国企和民营的控制人不同导致企业内部控制质量存在差异,与国企相比,民营企业产生错误定价的可能性更高。

表6 公告溢价效应在不同股权性质公司和不同行业中的表现

| 投资组合 | 股权性质 | | 行业 | | | |
|-------|--------------------|--------------------|-------------------|-------------------|------------------|-------------------|
| | 国企 | 民营 | 金融科技 | 运输业 | 房地产业 | 文体服务 |
| 正面 | 11.04 [1.25] | 20.61** [2.22] | 31.60** [1.91] | 11.15 [1.06] | 2.68 [0.25] | 11.49 [0.95] |
| 负面 | -4.04 [-0.47] | 0.60 [0.07] | 12.17 [0.69] | -11.58 [-1.04] | -2.90 [-0.26] | -13.59 [-1.08] |
| 正面-负面 | 15.07*** [2.70] | 20.01*** [3.40] | 19.43 [0.94] | 22.73* [1.89] | 5.58 [0.50] | 25.08* [1.80] |

注:表中投资组合采用市值加权,收益率以百分比形式汇报,方括号内为各组 t 值;运输业全称为交通运输、仓储和邮政业。

(四) 行业效应分析

为了进一步分析公告溢价效应在不同行业中的表现,探讨行业间的异质性差异,本文采用证监会行业分类2012版,并加入了新兴的金融科技行业。在每个行业内,计算每种情感公告类别投资组合的市值加权的年度平均收益率与相应的 t 值。考虑到篇幅限制,表6的右边4列仅展示4个具有代表性的行业,分别为金融科技类、运输业、房地产业和文体服务业。

结果显示,公告溢价效应在不同行业内的表现不同,行业效应(industry effect)的异质性表现明显。具体而言,运输业和文体服务业的公告效应较为明显,能获得较高且显著的年度平均收益率,分别为22.73%($t=1.89$)和25.08%($t=1.80$);房地产业的公告效应较弱,无论是在正面、负面还是对冲组合中,投资组合收益均不显著;而金融科技公司虽然多空对冲组合的收益并不显著,但在正面公告的投资组合中能获得年平均收益率31.60%($t=1.91$)。

七、进一步分析

本文尝试从不同角度分析公告溢价效应的市场反馈机制,分别为金融机构关注度、投资者关注度和公司信息披露质量三个视角。

(一) 金融机构关注度

机构投资者在资本市场中扮演着重要的角色,其对信息的搜集和利用能力更具优势,并且在交易过程中会包含所持有的信息(Sias et al., 2006; Yan & Zhang, 2009; Baik et al., 2010),可以促使资

产价格在融合市场信息方面更为迅捷,提高资本市场的定价效率。下面探究机构关注度对公告溢价效应的影响。

本文将金融机构关注度(被研报关注度和被分析师关注度)与公告倾向进行双变量排序,形成投资组合,研究公告溢价效应如何随关注度水平而变化。根据上个交易日末投资者关注度的数值,按中位数将股票分成2组,分别为低关注度组和高关注度组,同时独立地将股票根据上个交易日末的公告情感倾向进行排序,共6(2×3)个投资组合,计算市值加权回报。

表7 金融机构关注度和投资者关注度对公告效应的影响

| 投资组合 | 金融机构关注度 | | | | 投资者关注度 | |
|-------|--------------------|-------------------|--------------------|--------------------|--------------------|--------------------|
| | 被研报关注度 | | 被分析师关注度 | | 低 | 高 |
| | 低 | 高 | 低 | 高 | | |
| 正面 | 24.38** [2.44] | 22.08** [2.45] | 20.10** [2.01] | 26.15*** [2.89] | 29.15*** [2.88] | 19.24* [1.73] |
| 负面 | -4.15 [-0.43] | 12.18 [1.40] | -2.71 [-0.29] | 11.79 [1.35] | 8.87 [0.90] | -5.59 [-0.53] |
| 正面-负面 | 28.53*** [4.10] | 9.90* [1.75] | 22.80*** [3.36] | 14.36** [2.46] | 20.28*** [3.09] | 24.83*** [3.65] |

注:表中为双变量排序结果,投资组合采用市值加权,收益率以百分比形式汇报,方括号内为各组t值。

表7左侧结果显示,与金融机构关注度较高的组相比,机构关注度较低的组产生的公告溢价效应更高,具体来说,被研报关注度低的股票的多空投资组合平均收益率为28.53%($t=4.10$),高于被研报关注度较高组的9.90%($t=1.75$),同样地,分析师关注度低组的公告溢价效应为22.80%($t=3.36$),高于被分析师关注度高组的14.36%($t=2.46$)。机构投资者由于其在投资决策以及信息搜集分析等方面具有专业优势而被视为“成熟投资者”(Jiambalvo et al., 2002),金融机构作为上市公司与投资者之间的信息中介,可以降低企业与投资者之间的信息不对称程度,提升资产的信息效率。也就是说,在金融机构关注度较低的公司,其信息不对称程度较高,错误定价的可能性更高,更容易放大公告溢价效应。

(二)投资者关注度

行为金融理论的“投资者心理偏差观”认为公告效应是一种市场异象,是由于投资者的心理偏差引起的。与机构分析师相比,散户投资者在交易过程中往往更容易出现非理性的关注,如对所获取的信息过度自信,从而引起过度交易(Odean, 1999)。本文使用公告标题进行公告效应的研究也是由于标题更容易吸引散户投资者的关注。由于散户投资者还受到有限关注(Stambaugh et al., 2012)的影响,因此公告其他的细节信息容易被忽略。本小节将从投资者关注度的角度探究公告是否引起了散户投资者的关注,进而产生了公告效应。

首先,我们采用Fama-Macbeth横截面回归检验上市公司发布公告对投资者关注度的影响,我们将投资者关注度(异常交易量, Barber & Odean, 2008)作为被解释变量,上市公司是否发布公告(D , 哑变量)作为主要解释变量,采用股票收益率(Ret)、市值规模($\log(MV)$)作为控制变量,探究在公告发出的当日、次日以及第30日对投资者关注度的影响。

表8 公告发布与投资者关注度的Fama-Macbeth横截面回归

| 变量 | 当日 | | 次日 | | 第30日 | |
|------------|-------------------|-------------------|------------------|------------------|------------------|------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| D | 0.47*** [6.18] | 0.45*** [5.83] | 0.27 [1.14] | 0.28 [1.13] | -0.22 [-0.81] | -0.14 [-0.69] |
| ret | 7.99 [1.24] | 8.00 [1.24] | 1.88** [3.87] | 1.85** [3.91] | 1.26 [1.06] | 1.31 [1.05] |
| $\log(MV)$ | | 0.09* [1.74] | | -0.09 [-0.89] | | -0.13 [-1.24] |

注:表中展示各变量的回归系数,方括号内为经过Newey-West方法调整异方差和自相关后的稳健t值。

表8是检验上市公司公告的发布对投资者关注度影响的回归结果,方括号内是经过Newey-West(1987)方法调整异方差和自相关后的稳健t值。从表8的列(1)和(2)可以看到,在不同的控制变量下,公告(D)都显著为正,且系数较为稳健,说明公告发布当日显著地正向影响投资者关注度。表8的列(3)和(4)显示,公告发布对次日投资者关注度的影响仍为正向,但系数有所降低且t值也不显著。而在30日的时候,列(5)和(6)的结果显示,公告的发布对投资者关注度的影响出现了反转,系数变为负且不显著。

接下来,将投资者关注度与公告效应进行双变量排序,与前文的双变量排序方法一致,将投资者关注度与公告情感倾向进行双变量排序,表7右边两列报告了投资组合年度平均收益率以及对应的t值。结果显示,与金融机构的关注度恰好相反,在散户投资者关注度较高组的投资组合收益率为24.83%(t=3.65),高于低投资者关注度的20.28%(t=3.09)。散户投资者受到有限注意力的制约(权小锋、吴世农,2010;王磊、孔东民,2014),往往更容易出现非理性的关注,对信息的处理和解读出现偏差,引起过度交易从而导致股票错误定价。

(三)公司自身信息披露

信息披露是连接股票市场资金供给方和需求方的重要纽带,企业信息披露质量会影响信息不对称程度和定价效率(曾颖、陆正飞,2006)。本小节将从公司自身信息披露角度探究公告效应的来源。本文选取上市公司信息披露考评,审计师是否来自四大会计师事务所和是否来自境外会计师事务所三个指标评估公司信息披露质量。

表9报告了上市公司信息披露质量与公告情感分类进行双变量排序的投资组合收益率结果,就上市公司信息披露考评而言,评级较好公司公告溢价效应的平均年化收益率为14.68%(t=2.60),低于较差组的24.29%(t=4.22)。通常认为,审计师来自四大会计师事务所和境外会计师事务所的公司,信息披露质量较高,表9结果显示,在审计师是四大会计师事务所的公司中,多空投资组合的年平均收益率为11.64%(t=1.27),低于不是四大会计师事务所的公司组的21.63%(t=4.98),与审计师来自境外会计师事务所的公司结果相一致。结果说明,公司信息披露质量低的上市公司,投资者之间的信息不对称程度较高,产生错误定价的可能性较大,导致公告溢价效应水平较高。

表9 上市公司信息披露与公告溢价效应

| 投资组合 | 上市公司信息披露考评 | | 审计师是否来自四大会计师事务所 | | 审计师是否来自境外会计师事务所 | |
|-------|--------------------|--------------------|-------------------|--------------------|-------------------|--------------------|
| | 较好 | 较差 | 是 | 否 | 是 | 否 |
| 正面 | 21.50** [2.31] | 12.81 [1.45] | 21.14** [2.33] | 21.42** [2.39] | 20.59** [2.03] | 21.04** [2.37] |
| 负面 | 6.82 [0.75] | -11.49 [-1.32] | 9.50 [0.99] | -0.21 [-0.02] | 15.94 [1.52] | -0.05 [-0.01] |
| 正面-负面 | 14.68*** [2.60] | 24.29*** [4.22] | 11.64 [1.27] | 21.63*** [4.98] | 4.65 [0.43] | 21.08*** [5.05] |

注:表中为双变量排序结果,投资组合采用市值加权,收益率以百分比形式汇报,方括号内为各组t值。

八、结论与建议

本文在中国股票市场上,针对上市公司公告的文本数据,采用文本分析机器学习方法,分析了上市公司公告信息与资产预期回报之间的关系,验证了中国市场上公告溢价效应的存在。本文使用2000—2020年上市公司公告文本数据,构建了基于公告的文本情感词典,并运用机器学习方法判断公告的情感倾向(正面/负面),旨在深入理解公告信息对资产预期回报的影响。此外,本文对公告溢价效应进行了稳健性检验和异质性分析,并从不同角度探究了公告溢价效应的市场反馈机制。

本文的研究结果发现:首先,不同类别公告的收益有显著的差别,利好类公告的股票具有较高收

益率,而利空类公告的股票收益率较低。其次,本文构造的公告情感词典能显著区分公告的收益方向。最后,在异质性分析中,公告溢价效应在小规模、成长型公司溢价显著;相较于国企,基于公告的投资组合在民营企业中表现出更高的收益。此外,本文还揭示了公告溢价效应产生的原因是由于散户投资者的过度关注。对于金融机构关注度越高和公司信息披露质量越高的公司,公告溢价效应则越弱。

本文提供了研究公告效应的新视角,通过文本分析和机器学习方法,从信息价值和资产定价视角研究公告效应,拓展了中文金融文本数据在实证资产定价研究中的应用范围。此外,本研究的发现对于上市公司信息披露政策和监管部门的政策制定有参考价值。为了提高市场的信息效率和保护投资者权益,上市公司在信息披露过程中需要更加重视信息的质量和准确性,注重公告中行文和用词的恰当程度,准确反映公司的运营状况和经营水平,避免引起市场偏差或对投资者行为产生误导。监管部门应加强对上市公司信息披露行为的监督和对市场进行有效引导,及时纠正不当的信息披露行为,保护投资者利益,维护资本市场的健康发展。

参考文献:

- 陈卫华 徐国祥,2018:《基于深度学习和股票论坛数据的股市波动率预测精度研究》,《管理世界》第1期。
- 洪永森 汪寿阳,2021:《大数据如何改变经济学研究范式?》,《管理世界》第10期。
- 黄建欢 尹筑嘉,2009:《中国股市限售股解禁的减持效应研究》,《管理科学》第4期。
- 姜富伟 胡逸驰 黄楠,2021a:《央行货币政策报告文本信息、宏观经济与股票市场》,《金融研究》第6期。
- 姜富伟 孟令超 唐国豪,2021b:《媒体文本情绪与股票回报预测》,《经济学(季刊)》第4期。
- 孔东民 付克华,2005:《中国股市增发的市场反应及影响因素分析》,《世界经济》第10期。
- 孔东民 柯瑞豪,2007:《谁驱动了中国股市的PEAD?》,《金融研究》第10期。
- 李斌 邵新月 李玥阳,2019:《机器学习驱动的基本面量化投资研究》,《中国工业经济》第8期。
- 刘力 王汀汀 王震,2003:《中国A股上市公司增发公告的负价格效应及其二元股权结构解释》,《金融研究》第8期。
- 罗炜 何顶 洪莉莎 常国珍,2017:《媒体报道可以预测创业企业的发展前景吗?》,《金融研究》第8期。
- 马甜 姜富伟 唐国豪,2022:《深度学习与中国股票市场因子投资——基于生成式对抗网络方法》,《经济学(季刊)》第3期。
- 权小锋 吴世农,2010:《投资者关注、盈余公告效应与管理层公告择机》,《金融研究》第11期。
- 苏治 卢曼 李德轩,2017:《深度学习的金融实证应用:动态、贡献与展望》,《金融研究》第5期。
- 王磊 孔东民,2014:《盈余信息、个人投资者关注与股票价格》,《财经研究》第11期。
- 吴世农 吴超鹏,2005:《盈余信息度量、市场反应与投资者框架依赖偏差分析》,《经济研究》第2期。
- 谢德仁 林乐,2015:《管理层语调能预示公司未来业绩吗?——基于我国上市公司年度业绩说明会的文本分析》,《会计研究》第2期。
- 游家兴 吴静,2012:《沉默的螺旋:媒体情绪与资产误定价》,《经济研究》第7期。
- 张纯 吴明明,2015:《媒体在资本市场中的角色:信息解释还是信息挖掘?》,《财经研究》第12期。
- 章卫东 李德忠,2008:《定向增发新股折扣率的影响因素及其与公司短期股价关系的实证研究——来自中国上市公司的经验证据》,《会计研究》第9期。
- 曾颖 陆正飞,2006:《信息披露质量与股权融资成本》,《经济研究》第2期。
- Asquith, P. & D.W.Mullins(1986), "Equity issues and offering dilution", *Journal of Financial Economics*, 15(1): 61-89.
- Baik, B. et al.(2010), "Local institutional investors, information asymmetries, and equity returns", *Journal of Financial Economics*, 97(1): 81-106.
- Ball, B. & P.Brown(1968), "An empirical evaluation of accounting income numbers", *Journal of Accounting Research*, 6(2): 159-178.
- Barber, B.M. & T.Odean(2008), "All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors", *Review of Financial Studies*, 21(2): 785-818.
- Blei, D.M. et al.(2003), "Latent Dirichlet Allocation", *Journal of Machine Learning Research*, 3(1): 993-1022.
- Cheng, Q. & K.Lo(2006), "Insider trading and voluntary disclosures", *Journal of Accounting Research*, 44(5): 815-848.
- Das, S.R. & M.Y.Chen(2007), "Yahoo! for Amazon: Sentiment extraction from small talk on the web", *Management*

- Science*, 53(9): 1375—1388.
- Dellavigna, S. & J.M.Pollet(2009), “Investor inattention and Friday earnings announcements”, *Journal of Finance*, 64(2): 709—749.
- Fama, E.F. & K.R.French(1993), “Common risk factors in the returns on stocks and bonds”, *Journal of Financial Economics*, 33(1): 3—56.
- Garcia, D.(2013), “Sentiment during recessions”, *Journal of Finance*, 68(3): 1267—1300.
- Goldstein, I. et al.(2021), “Big data in finance”, *Review of Financial Studies*, 34(7): 3213—3225.
- Gu, S. et al.(2020), “Empirical asset pricing via machine learning”, *Review of Financial Studies*, 33(5): 2223—2273.
- Hertzel, M. & R.L.Smith(1993), “Market discounts and shareholder gains for placing equity privately”, *Journal of Finance*, 48(2): 459—485.
- Hess, A.C. & P.A.Frost(1982), “Tests for price effects of new issues of seasoned securities”, *Journal of Finance*, 37(1): 11—25.
- Huang, A. et al.(2014), “Evidence on the information content of text in analyst reports”, *Accounting Review*, 89(6): 2151—2180.
- Jiambalvo, J. et al.(2002), “Institutional ownership and the extent to which stock prices reflect future earnings”, *Contemporary Accounting Research*, 19(1): 117—145.
- Jiang, F. et al.(2018), “Q—theory, mispricing, and profitability premium: Evidence from China”, *Journal of Banking & Finance*, 87(C): 135—149.
- Jiang, F. et al.(2019), “Manager sentiment and stock returns”, *Journal of Financial Economics*, 132(1): 126—149.
- Ke, Z.T. et al.(2019), “Predicting returns with text data”, NBER Working Paper, No. 26186.
- Kelly, B.T. et al.(2019) “Characteristics are covariances: A unified model of risk and return”, *Journal of Financial Economics*, 134(3): 501—524.
- Kozak, S. et al.(2020), “Shrinking the cross—section”, *Journal of Financial Economics*, 135(2): 271—292.
- Light, N. et al.(2017), “Aggregation of information about the cross section of stock returns: A latent variable approach”, *Review of Financial Studies*, 30(4): 1339—1381.
- Liu, J. et al.(2019), “Size and value in China”, *Journal of Financial Economics*, 134(1): 48—69.
- Loughran, T. & B.Medonald (2011), “When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks”, *Journal of Finance*, 66(1): 35—65.
- Newey, W.K. & K.D.West(1987), “A simple, positive semi—definite, heteroskedasticity and autocorrelation consistent covariance matrix”, *Econometrica*, 55(3): 703—708.
- Odean, T.(1999), “Do investors trade too much”, *American Economic Review*, 89(5): 1279—1298.
- Sias, R.W. et al.(2006), “Changes in institutional ownership and stock returns: Assessment and methodology”, *Journal of Business*, 79(6): 2869—2910.
- Stambaugh, R.F. et al.(2012), “The short of it: Investor sentiment and anomalies”, *Journal of Financial Economics*, 104(2): 288—302.
- Tetlock, P.C.(2007), “Giving content to investor sentiment: The role of media in the stock market”, *Journal of Finance*, 62(3): 11—39.
- Wruck, K.H.(1989), “Equity ownership concentration and firm value: Evidence from private equity financings”, *Journal of Financial Economics*, 23(1): 3—28.
- Yan, X. & Z.Zhang(2009), “Institutional investors and equity returns: Are short-term institutions better informed?”, *Review of Financial Studies*, 22(2): 893—924.

Announcement Premium Effect and Asset Pricing: A Textual Machine Learning Perspective

TANG Guohao, ZHU Lin and CHEN Shicheng
(Hunan University, Changsha, China)

Summary: In recent years, the application of machine learning techniques within the financial market has become increasingly prevalent, underscoring the significant potential of these methods in accurately forecasting and explaining the market returns. Concurrently, big data is revolutionizing the finance industry and has the potential to significantly shape future research in finance. Goldstein et al. (2021) noted that new research built on these big data to push the frontier on fundamental issues across areas in finance.

Textual data, with its large size, high dimensionality, and complex structure, epitomizes big data features. Corporate announcements are particularly noteworthy within this domain due to their high-quality information, which not only reveals a company's operational and development status but also communicates critical information that can significantly affect stock price. Consequently, it is valuable to mine effective pricing information from these corporate announcements to examine its predictive impact on returns.

This paper analyzes textual data from corporate announcements spanning from 2000 to 2020. Through textual analysis, this paper constructs the sentiment dictionaries based on these announcements and employs machine learning techniques to assess the sentiment orientation (positive or negative) of each announcement, so as to investigate the influence of announcement content on the anticipated asset returns. Furthermore, this paper undertakes robustness tests and heterogeneity analyses of the announcement premium effect, offering an in-depth examination of the market's reaction mechanisms to this effect from multiple dimensions.

The findings of this paper reveal several key insights. Firstly, there is a significant variance in the returns of stocks based on the category of announcements, with positive news yielding higher returns and negative news leading to lower returns. Secondly, the sentiment dictionary developed in this paper effectively differentiates the direction of stock returns, demonstrating a significantly positive announcement premium effect. Thirdly, in the heterogeneity analysis, the announcement premium effect is notably pronounced in small-scale, growth-oriented firms; compared with state-owned enterprises (SOEs), portfolios based on announcements in private firms exhibit higher returns. Additionally, the paper uncovers that the underlying cause of the announcement premium effect is the excessive attention from retail investors. For companies that are more closely monitored by financial institutions and have higher quality of information disclosure, the announcement premium effect is relatively weaker.

The main contributions of this paper are as follows. Firstly, it offers a novel perspective on analyzing the effects of corporate announcements. Distinct from the conventional event study methods that focus on specific types of announcements, this paper employs textual analysis to extract information from all types of announcements, examining their effects from perspectives of information value and asset pricing. Secondly, this paper broadens the object of Chinese financial text analysis. In constructing the sentiment dictionary for announcements, this paper innovatively incorporates stock returns as weights to calculate the sentiment orientation of words in a supervised manner. This approach strengthens the link between the dictionary and stock returns, making it more relevant for research in asset pricing. Thirdly, this paper opens new avenues for quantitative investment strategies. By harnessing the power of machine learning to extract pricing information from corporate announcements, this method not only enhances the precision of investment analyses but also contributes to the optimization of investment portfolios based on data-driven evidence.

Furthermore, the findings of this paper offer valuable insights for both corporate disclosure policies and regulatory frameworks. It underscores the importance for listed companies to prioritize the quality and accuracy of disclosed information. Regulatory bodies should enhance their oversight and guidance on corporate disclosure practices, promptly addressing inappropriate behaviors to safeguard investor interests and foster the healthy development of the capital market.

Keywords: Announcement Premium Effect; Textual Analysis; Machine Learning; Market Reaction

JEL Classification: G11, G12

(责任编辑:木丰)

(校对:金禾)