

doi: 10.19920/j.cnki.jmsc.2022.03.004

# 分析师社交圈、利益冲突与研究报告文本信息<sup>①</sup>

刘亚辉, 黄凯, 尹玉刚\*, 汪勇  
(西南财经大学金融学院, 成都 611130)

**摘要:** 通过创新性地量化分析师研究报告中的文本信息含量和分析师身处的金融社交圈, 研究分析师在“圈子”里所面临的利益漩涡对其研报信息含量的影响。发现: 分析师凭借所经营的社交圈在提高其研究报告文本信息以获取市场信任的同时, 也通过更乐观的荐股评级达到职业发展的目的, 且分析师会在“博取市场信任”与“职业生涯发展”之间进行动态权衡。权衡的结果是, 在职业生涯发展目的实现的基础上, 深度社交圈为其提供了更多的信息来源, 反映为更高的研究报告信息含量以进一步博取市场信任, 且其在社交圈中的议价能力亦随之提高。上述结论为评价分析师研究报告信息含量提供了新的视角, 同时为社交圈对分析师行为影响的研究提供了新的证据与补充。

**关键词:** 分析师; 社交圈; 文本分析; 利益冲突

**中图分类号:** F830.9   **文献标识码:** A   **文章编号:** 1007-9807(2022)03-0062-26

## 0 引言

证券分析师在抑制信息不对称、促进价格的信息发现过程中起到了不可忽视的作用<sup>[1-4]</sup>。尤其近年来随着更多高学历、行业经验丰富甚至具有海外经历者的加入, 证券分析师队伍不断成长和壮大。但分析师是否有效发挥了信息中介作用? 现有文献并未得出一致结论。大部分研究认为研究报告(下文简称为“研报”)中存在普遍的乐观偏差, 并将其归因于分析师所面临的利益冲突<sup>②</sup>[5-9]。也有学者认为分析师向市场提供了显著的增量信息, 在抑制信息不对称、改善信息环境

中发挥了不可替代的作用<sup>[10-13]</sup>。

现有文献大多以盈余预测及荐股评级等量化指标来衡量研报的信息价值<sup>[14-16]</sup>。然而, 作为分析师基于公开及私有信息做出的结论性预测, 这些量化指标无法涵盖所有信息, 且不具有及时的可验证性。更为重要的是从分析师解读信息到做出盈余预测和荐股评级的过程是个“黑匣子”, 更容易受制于利益漩涡而失去客观性<sup>③</sup>[15]。仅基于这些量化指标衡量研报信息价值可能有失偏颇。作为研报更重要的组成部分, 文字分析是分析师提供信息的重要载体, 占据了研报的大部分篇幅<sup>④</sup>。相对于其他量化指标, 它更多的是对客观事

① 收稿日期: 2018-04-17; 修订日期: 2020-12-29。

基金项目: 国家自然科学基金资助项目(71903154; 71873266; 71874145)。

通讯作者: 尹玉刚(1988—), 男, 安徽六安人, 博士, 副教授, 博士生导师。Email: yinyg@swufe.edu.cn

② 对分析师的研究集中在受雇于券商的卖方分析师, 券商与承销上市公司、证券投资基金之间的利益关联很容易通过雇佣关系传导给分析师, 对分析师的独立性与客观性形成压力, 形成利益冲突。分析师面临的利益冲突主要有3种, 一是来自所属券商的压力, 不能发表对客户(或投资者)不利的报告; 二是分析师出于对自身职业生涯的考虑; 三是分析师与被研究公司之间存在利益共谋。

③ 在研报发布初投资者是无法验证这些量化指标的客观性的, 只能在一段时间后通过市场表现或公司披露财报进行验证。

④ 研报文字分析主要包括公司最近的业绩表现、发展战略、行业竞争力以及公司治理等方面的内容。

实的描述,并更受投资者关注<sup>⑤</sup>。但是囿于研报文本可得性、文本挖掘技术的发展等因素,分析师研报的相关文献却较少涉及文本信息<sup>[17]</sup>。

中国是个关系型社会<sup>[18]</sup>,在股市这个名利场更是如此。无论研报的量化指标还是文字分析,都不可避免地受到分析师社交圈的影响。在社会关系资本化的资本市场,“圈子”已然成为利益传输的重要桥梁<sup>[6,9]</sup>。对于分析师来说,“圈子”也是其获取私有信息的潜在途径<sup>[19,20]</sup>。“圈子”在资本市场中的影响近年来不断受到学者们的关注。现有对分析师社交圈的研究主要集中在商业关系上,例如券商与上市公司、证券投资基金之间的利益关联通过雇佣关系影响分析师的独立性<sup>[9]</sup>。但是在错综复杂的人际关系中,同窗校友之情、共同调研之谊都是关系的重要内涵,与雇佣关系相比,这种私人关系对分析师来说更加平等,也具有更多的主动权和互动性<sup>[18]</sup>。

因此,本文从增量文本信息和传统量化指标的双重视角较为全面地考察私人“圈子”对研报信息含量的影响。以文本相似度代理研报所提供的增量信息,以量化指标的乐观偏差度量分析师对职业生涯的考量,进而较为全面地研究分析师在利益漩涡中的行为轨迹。以文本相似度衡量增量信息的方法在财经新闻以及财务报告的文本分析中已被较多的学者采用<sup>[21-23]</sup>,将该方法应用在研报中。通过手工搜集研报全文内容,利用余弦相似度计算出每份研报与该公司前3个月所有研报文本相似度的均值作为该研报所提供的增量信息。文本相似度越低,该研报信息含量越高<sup>⑥</sup>。同时,运用社会网络分析的方法,计算出分析师在基金经理中拥有的人脉资源(“圈子”),将“圈子”的范围限制在基金经

理之内,是因为基金经理有《新财富》明星分析师的投票权,而入选《新财富》是分析师职业生涯的终极目标;并且基金经理通常拥有其所持股票的私有信息,分析师可以获取“圈子”内基金经理的私有信息而得到信息优势,进而体现在研报的文本内容中<sup>⑦</sup>。

基于2006年~2016年的样本数据,发现“圈子”能够显著提高分析师研报的文本信息含量,其原因是作为获取私有信息的潜在途径,“圈子”增加了分析师的相对信息优势<sup>[19]</sup>。进一步研究表明,这一优势主要体现在“圈子”内基金经理持股的研报上,证实了分析师通过“圈子”获取私有信息的推论。作为利益冲突形成的土壤,广泛的社交圈在使分析师有获取信息的优势同时,也使其更容易陷入利益漩涡之中。拥有较广社交圈的分析师总体上发布了更高的荐股评级,而这种系统的乐观性是由“圈子”内基金经理持股所致。

高荐股评级与分析师利益如何关联?作为重要的外部激励,入选《新财富》是分析师职业生涯的重大跨越<sup>[6]</sup>。“圈子”能显著提升入选概率,且这一提升作用主要通过分析师对“圈子”内基金经理持股发布更高的荐股评级实现。对研报文本信息而言,虽然对当选概率没有显著影响,但信息含量高的研报能够引起投资者对该股票显著且持续的关注。

入选《新财富》,分析师在财富与声誉(投资者关注)均显著提升之后,分析师对二者的权衡会发生怎样的改变?结果表明,分析师当选后对“圈子”内外荐股评级的差异显著降低。这是因为当选后随着地位提高,他们无需再以高评级取悦基金经理。与“圈子”外相比,当选后“圈子”内股

⑤ 美国《机构投资者》对美国市场的调查显示,1998年以来投资者更注重研报的文本,例如在2010年~2011年,受访者认为文本的重要性排第5,而盈余预测排在第12位。并且尽管很多金融数据库已经包含了研报中量化指标的数据,投资者每年花费在研报全文购买上的资金仍高达数百万美元。

⑥ 之所以用文本相似度度量分析师所提供的关于上市公司的特有信息,是因为除去文本相似度被广泛运用的科学性以外,对于投资者来说,研报有价值的部分在于研报文本中与其他分析师所提供的不一样的基本面等分析。对于文本中与其他分析师所提供的一样的广为人知的共同信息,不但私有信息含量很少,且这一共同信息可能正如董大勇等<sup>[50]</sup>所述那样,仅仅是研报信息的“相互借鉴”,并未提供增量信息。

⑦ 基金经理的私有信息不一定是从上市公司获得的内幕信息,更多的是关于公司更为真实的基本面信息。

票的研报信息含量显著更高,这一方面是因为分析师更加注重自身声誉,倾向于提供更多的特质信息,另一方面是分析师在当选后获取私有信息的能力进一步增强。

本文的主要贡献体现在两个方面。第一,不同于现有文献以盈余预测准确度、荐股评级等较为主观的量化指标衡量研报的信息价值,而以较为客观的研报文本为标的,运用文本挖掘中的相似度算法,为研报信息含量的评估提供新的方法;第二,“圈子”产生的利益冲突对分析师行为的影响仍是学术界较为前沿的话题,仅有的一些研究也局限在分析师通过券商而产生的外部利益关联上,通过研究分析师私人关系,为“圈子”与分析师利益冲突的研究提供了新的证据与补充。

## 1 文献综述与研究假设

### 1.1 研报文本分析回顾

在信息传递中扮演关键角色的分析师受到投资者和学术界的普遍关注<sup>[14,15]</sup>。分析师一方面广泛搜集有价值的公开信息,另一方面通过实地调研等其他途径获得私有信息,并对这些信息进行专业解读,以研报的形式传递到市场<sup>[25-29]</sup>。在我国资本市场处于新兴加转轨、信息环境有待改善且投资者以散户为主的背景下,分析师在抑制信息不对称、促进价格发现方面被寄予更高的期望。作为分析师信息传递的重要载体,研报成为评价其表现的重要依据。个股研报常常对公司近期的业绩表现、发展战略、行业竞争力以及公司治理等方面的信息进行初步分析,进而得出盈余预测、荐股评级和目标价位等量化指标。囿于大量研报全文的不可得性及文本挖掘技术的发展,现有研究多局限于研报的量化指标,对文本内容较少涉及。

然而,正如 Tsao<sup>[30]</sup>所言,“荐股评级和目标价位是分析师研报中无关痛痒的部分,真正有价值的是产生这些结论的逻辑分析与细节”<sup>⑧</sup>。盈余预

测、荐股评级和目标价位等量化指标是分析师基于基本面信息得出的结论性建议,但从信息到结论的推导是个“黑匣子”,难以避免地掺杂着分析师的主观倾向;而文字分析更多的是对客观事实的描述,包含了更多的增量信息。Asquith等<sup>[31]</sup>对1126份研报全文手工构建情绪指标,发现在控制荐股评级后,这一指标对研报发布日未来5天收益率仍有显著解释力,但其样本量过小,回归中用到的观测数仅有193个,结论缺乏普适性。Twedt和Rees<sup>[32]</sup>通过词性标注对2057份深度研报文本构建量化指标,也发现了研报文本对收益率的显著影响,而小样本导致的稳健性问题也削弱了这一研究的代表性。Huang等<sup>[17]</sup>首次在大样本范围内对研报文本进行分析,发现研报文本比盈余预测、荐股评级等包含了更多的信息,且文本信息对公司盈利有更长期的预测力。Huang等<sup>[33]</sup>进一步发现分析师在研报文字中对管理层电话会议内容进行了信息挖掘,而投资者对这部分挖掘的信息更为关注。伊志宏等<sup>[34]</sup>利用文本分析技术,发现研报文本中的负面情感倾向能够降低股价暴跌风险。马黎珺等<sup>[35]</sup>通过机器学习对分析师研报进行了文本分析,发现前瞻性语句的情感倾向与研报发布后的市场反应显著正相关。本文虽然同样关注文本内容,但侧重的是研报信息的“增量”而非“倾向”,是对前述文献的补充与完善。

文本相似度被广泛应用在基于内容的推荐系统、信息检索及论文重复率检测等领域,在财经文本中的应用也初露端倪。其原理是用文本之间的相似程度代理信息差异,相似度越高,则二者的信息差异越小。利用这一思想,Hoberg和Philips<sup>[21]</sup>发现年报中产品描述章节文本相似度高的企业相互并购后获得了更好的协同效应。类似的,以年报中产品描述章节的文本相似度为上市公司重新划分行业,Hoberg和Philips<sup>[22]</sup>发现这种新行业划分能更好地解释公司盈利能力和销售增长等个体

<sup>⑧</sup> 原文是“Stock ratings and price targets are just the skin and bones of analysts’ research. The meat of such reports is in the analysis, detail, and tone”。

特征,新的行业风险因子也具有更强的解释力。Bushman 等<sup>[23]</sup>发现银行年报的文本相似度与其系统风险正相关,并且文本相似度高的银行之间的风险存在较高的联动性。

本文以每份研报与前3个月目标公司所有研报文本相似度的均值作为其信息含量的代理,这一方面为衡量分析师研报信息提供了新的角度,得以更全面地研究分析师信息中介角色的发挥,另一方面扩展了文本挖掘技术在财经文本领域的应用。

### 1.2 “圈子”与研报文本信息

“社会关系资本化,资本市场江湖化”<sup>[17]</sup>,关系在资本市场的影响不容忽视。研究的第一个问题是分析师在基金经理中的“圈子”对研报文本信息和荐股评级的影响。将“圈子”的范围限制在基金经理之内是因为他们有《新财富》明星分析师的投票权,而入选《新财富》在很大程度上是分析师职业生涯成功的标志,并且基金经理通常拥有其所持股票的私有信息,分析师可以获取“圈子”内基金经理的私有信息而得到信息优势,进而体现在研报的文本内容中。“圈子”包括校友关系、调研关系和地域关系3个层次。校友关系是形成社交网络的重要纽带,在资本市场中,校友间的利益共享、信息共享已被学者发现并证实<sup>[18]</sup>。调研亦是分析师获取信息的重要渠道,有两层含义,一是调研者之间很可能已经具有紧密的联系;二是“交个朋友多条路”,调研时新建立的关系也很可能是新合作的开始。而工作地点同城且同一个区,会由于地理的临近而为社交关系的形成提供很大的潜在可能,因此,将地域关系也纳入“圈子”的范畴。借助社会网络分析方法,计算每个分析师在基金经理中“圈子”(Network)的大小。

社会关系对分析师具有双重影响。一方面它是资本市场信息扩散的重要渠道之一,分析师可以籍此取得信息优势。Horton 和 Serafeim<sup>[20]</sup>用社会网络分析法对10 508名分析师和42 376名公司高管构建校友关系网络,发现网络中心度高的分析师盈余预测更准确,并且失业率更低。Cohen 等<sup>[19]</sup>发现校友关系是分析师获取私有信息的重

要渠道,与上市公司高管有校友关系的分析师取得了显著更高的荐股超额收益。程博和潘飞<sup>[36]</sup>也发现分析师与CEO之间的校友关系增加了分析师的信息获取优势。Cohen 等<sup>[37]</sup>对校友网络的进一步研究发现,基金经理的持仓集中于与高管有校友关系的上市公司股票,且有关系的持仓组合收益比无关系的持仓组合收益率每年高出7.8%,这部分超额收益主要来自于上市公司重大事项公布前后,间接证实了基金经理通过校友关系获得了信息优势。

本文则用文本相似度量研报信息含量,研究“圈子”的影响,并提出如下假设:

**H1a** “圈子”越广,分析师研报文本信息含量越高。

**H1b** 分析师对“圈子”内基金经理持股的研报文本信息含量较“圈子”外研报更高。

另一方面,“圈子”在赋予分析师信息优势的同时,也影响着研报的独立性与客观性。“圈子”内私有信息的获取不是无偿的,分析师向基金经理付出的成本是更乐观的荐股评级。Tan 等<sup>[8]</sup>发现对于基金经理重仓的股票,分析师公开发布的研报乐观偏差较大,但未公开的研报不存在乐观偏差,间接证明了分析师与基金经理之间可能存在的利益共谋。李志生等<sup>[38]</sup>研究发现证券分析师在媒体上发布的荐股行为中存在内幕交易和利益输送,使得财富向机构投资者转移。Yin<sup>[9]</sup>通过将受雇于基金公司控股股东或发起人的分析师与这些基金公司进行配对,直接研究了分析师与关联基金经理之间的利益互动,发现分析师倾向于对关联基金公司持有的股票发布较高的荐股评级,而基金经理则在乐观评级发布后显著减少了持股数量。上述研究集中在雇佣引起的商业关系上,证实了分析师对与其具有商业关联的股票发布了较高的荐股评级,这一逻辑同样适用于私人关系,即分析师与基金经理的“圈子”。基金经理对其持股的信息优势和《新财富》明星分析师投票权,有足够的诱惑力吸引分析师用较高的荐股评级与基金经理达成协议。据此本文提出如下假设:

**H2a** “圈子”越广,分析师荐股评级越高.

**H2b** 分析师对“圈子”内基金经理持股的荐股评级较“圈子”外高.

入选《新财富》能极大促进分析师职业生涯的发展及个人财富的积累,因此竞争异常激烈.基金经理投票的评选机制是“圈子”效应发挥作用的温床. Emery 和 Li<sup>[39]</sup>把竞选明星分析师称为“人气竞赛”,他们发现市场认可度是当选最重要的因素. 吴偲立等<sup>[40]</sup>对我国《新财富》明星分析师评选进行研究,发现获奖前明星分析师与普通分析师的研报信息含量无显著差异,影响获奖概率的不是提供信息的多寡而是其曝光率及所属券商的市场地位. 这些研究都表明能否当选更重要的是基金经理对分析师的认可度. 在同等条件下,分析师更容易获得“圈子”内基金经理的选票,“圈子”越广的分析师越容易当选. 然而,分析师获得基金经理的选票和私有信息不是无偿的,他们需要为“圈子”内基金经理持股提供更高的荐股评级. 同时,校友关系、调研关系和地域关系只是名义上的,实质利益同盟的形成离不开对“圈子”的维护. 只有当分析师确实为基金经理提供了较高的荐股评级时,才能得到基金经理的投票回报. 从分析师对“圈子”内、外基金经理持股的荐股评级差异和跟踪偏好两个方面衡量分析师对“圈子”的维护. 因此,本文提出假设:

**H3** “圈子”会提升分析师入选《新财富》的概率,这种提升效应形成的途径是分析师对“圈子”内、外基金经理持股的荐股评级差异和跟踪偏好.

在普遍认识到研报量化指标存在乐观偏差的情况下,投资者更关注文本信息. 在我国股票市场也存在类似的情况,一是 Wind、东方财富、国泰安和中国研究数据服务平台等金融数据库投入大量资金从券商研究所购买卖方分析师的研报全文,并将其以收费的方式向用户限量提供;二是虽

然中美股市在投资者结构、市场交易制度安排方面有明显差异,但投资者对信息获取的诉求是—致的. 在分析师研报的量化指标存在普遍乐观偏差的环境下,投资者更看重研报的文本信息. 在与中国的机构投资者和中小投资者的交流中,他们也表达了对研报文本信息的重视. 据此本文提出假设:

**H4** 研报文本信息含量越高,投资者对目标公司的关注度越高.

当选明星分析师后,其利益和地位得到了极大提高<sup>[6]</sup>,分析师与“圈子”内基金经理的交流也随之改变. 一方面,分析师当选后无须再一味地以高级别取悦基金经理;另一方面,明星分析师获取信息的能力增强,且更注重文本信息含量对自身声誉的影响. 因此,分析师当选后对利益与自身声誉的动态权衡必然发生改变<sup>⑨</sup>. 据此提出假设:

**H5** 明星分析师当选后对“圈子”内、外股票的荐股评级差异降低,“圈子”内股票研报信息含量提高.

## 2 研究设计

### 2.1 研究样本与数据来源

数据有3个来源. 在构建分析师社交圈时,分析师和基金经理的学历、毕业院校、所属券商等个人特质数据从中国研究数据服务平台(CNRDS)获得,由于样本期内(2006年~2016年)部分院校出现更名、合并等现象,统一将院校名称改为现用名称,并以此构建校友圈;分析师和基金经理调研数据亦来自 CNRDS. 分析师所属券商的办公地址从全国企业信用信息公示系统手动搜集,并以此构建地域圈. 研报全文数据来自和讯财经,时间跨度为2006-01~2016-12,共168 184份原始研报;当研报为团队撰写时,将该研报视为每一团队成员的单独成果,分离

<sup>⑨</sup> 高质量的研报文字能够提高投资者对分析师的认可度,进而影响到对研报量化指标的置信程度,并且无论是从分析师职业操守的角度,还是从监管层对研报信息的质量要求,分析师都有动力提供较高质量的文字分析.

后共有 193 097 个分析师—研报数据, 将这些研报按标的公司、分析师姓名、所属券商和发布日期与国泰安数据库(CSMAR)中的研报量化指标匹配后, 最终得到 64 431 个研究样本。其他控制变量数据来自 CSMAR 数据库。

## 2.2 关键指标构建

分析师社交圈为分析师在基金经理中拥有的人脉资源。与申宇等<sup>[18]</sup>不同, 对社交圈从深度和广度两个层面进行了扩展, 在深度上除了考虑直接关系, 还加入了间接关系; 在广度上把“圈子”的范畴在校友圈之外加入了调研圈和地域圈。

“圈子”的计算依赖于社会网络分析技术<sup>[41]</sup>, 连接矩阵(adjacent matrix)是其常用的数据形式(见表1)。

表1 一个连接矩阵

Table 1 An adjacent matrix

$$\begin{pmatrix} - & A & B & C & D \\ A & - & 1 & 0 & 0 \\ B & 1 & - & 1 & 0 \\ C & 0 & 1 & - & 1 \\ D & 0 & 0 & 1 & - \end{pmatrix}$$

如果 A 和 B 曾经是校友, 则赋值为 1, 否则为 0。如果 A 和 B 在同一时间调研了同一家上市公司, 则认为二者之间存在调研关系; 如果 A 和 B 所属券商的办公地在同一个城市的同一个区, 则二者存在地域临近关系<sup>⑩</sup>。这样就构建了 3 个仅由 0 和 1 组成的对称矩阵, 三者加总, 得到初始矩阵(有权)。

“朋友的朋友亦是朋友”。间接关系也属于“圈子”的范畴, 并且与直接关系相比, 其强度势必减弱。为将间接关系和直接关系强度统一加总计算, 采用 Yin 和 Liu<sup>[42]</sup>提出的指数衰减法对间接关系进行处理。具体来说, 假设 A 和 B 没有直接关系(仅考虑直接关系时二者的权重为 0), 但 A 和 B 同时与 C 有直接关系, 即 A 和 B 通过 C 建

立了间接关系, C 是 A 和 B 的中介<sup>⑪</sup>。将 A 和 B 的间接关系进行指数衰减(具体算法见附录 A)。

连接矩阵中包括分析师和基金经理两类节点, 而《新财富》明星分析师评选时只有基金经理才有投票权, 二者不能等同而论, 因此对传统的社会网络分析方法进行改进。具体方法是对所有的节点按照类型进行排序, 例如矩阵共有  $n = n_1 + n_2$  个节点,  $n_1$  个节点为分析师,  $n_2$  个节点为基金经理。这样排序之后可以根据每个连线的坐标确定其关系类型, 例如某个点的坐标为  $(x, y)$ ,  $x > n_1, y < n_1$ , 则此连线表示分析师与基金经理之间的关系。

基于综合考虑了直接和间接关系的连接矩阵, 用社会网络分析中的点度中心性来衡量分析师“圈子”的大小。在对矩阵根据分析师和基金经理进行排序后, 分析师  $i$  ( $i < n_1$ ) 的中心度为

$$Network_i = \sum_j^n E(i, j), (i \leq n_1; j > n_1) \quad (1)$$

式中  $E(i, j)$  为分析师  $i$  和基金经理  $j$  的关系值, 即排序后连接矩阵中元素  $(i, j)$  的值。  $Network_i$  越大, 表示该分析师在基金经理中拥有越广的人脉。

尽管分析师的校友关系、地域关系在不同年份保持稳定, 但其调研圈会有较大波动, 同时每年均有离职和新进的分析师和基金经理, 这将导致分析师的“圈子”大小有较大的年度差异。因此在样本期内的每一年度, 都按照上述方法计算分析师“圈子”的大小。

文本相似度为研报文本余弦相似度。两个文本之间的余弦相似度计算主要分为分词、构建词向量、计算余弦夹角 3 个步骤。

分词是把研报文本拆分成词语, 以进行更深

⑩ 部分券商存在双总部或多总部的情形, 这可能导致基于注册地计算的地域圈不精确, 在稳健性检验中详细考察并排除了这种可能性对本文结论的影响。

⑪ 分析师和基金经理的本科、硕士、博士阶段可能不在同一学校, 因此存在这种交叉校友的现象。

入的分析。相比英文分词,中文的分词更加复杂<sup>⑫</sup>。Jieba是中文分词中最常用的工具,其自带的通用词汇库已较为完善<sup>⑬</sup>。考虑到研报属于财经文本,有较多的财经专业术语,将搜狗输入法官方推荐的“财经金融词汇大全”和“财务会计词汇大全”导入Jieba中,以进一步提高分词的准确性。分词中另一个重要问题是对停用词的处理。如果将停用词认为是噪音,则在分词时剔除之;如果认为停用词也是种信息,则保留之<sup>⑭</sup>。考虑到过多的停用词意味着研报没有提供太多有价值的信息,所以将保留了停用词的文本相似度作为稳健性检验。

研报文本分词后要将每份研报向量化<sup>⑮</sup>。量化的核心问题是每个词语对应的数值应该是多少?采用最常用的TF-IDF方法,其中TF(text frequency)指某一词语在这份研报中出现的次数, IDF指基础文库中包含该词语的文档数<sup>⑯</sup>。TF越大,说明该词语在这份研报中越重要;IDF越大,说明在基础文库中,该词语是个较为通用的词语,没有特质信息含量,因此权重越小。对于一篇研报中的某个词语,其TF-IDF最终的计算公式为

$$TF - IDF = \frac{f_{t,d}}{n_d} \ln\left(\frac{N}{1 + n_t}\right) \quad (2)$$

式中 $f_{t,d}$ 是该词语在研报d中出现的频率; $n_d$ 是该研报包含的词语总数; $N$ 是基础文库包含的研报份数; $n_t$ 是基础文库中包含该词语的研报份数。较低的 $n_t$ 表明包含词语t研报份数较少,意味着该

词代表了较高的特质信息。

向量化后的研报可以计算两两之间的相似度,即余弦夹角,计算公式为

$$Sim(A, B) = \cos(\theta) = \frac{A \times B}{|A| |B|} = \frac{\sum_{k=1}^n A_k B_k}{\sqrt{\sum_{k=1}^n A_k^2} \sqrt{\sum_{k=1}^n B_k^2}} \quad (3)$$

式中 $A, B$ 分别表示向量化之后的研报; $\cos(\theta)$ 表示 $A, B$ 的余弦夹角; $A_k$ 表示向量 $A$ 的第 $k$ 个元素; $B_k$ 表示向量 $B$ 的第 $k$ 个元素; $n$ 为向量 $A, B$ 的维度。研报的余弦值越大,说明二者的夹角越小,两篇文档的相似度越高。

在计算出研报之间两两相似度之后,计算每一份研报与对应目标公司前3个月所有研报的相似度的均值,作为这篇研报的文本相似度。如果在分词时剔除了停用词,变量为 $SimWu$ ,如果保留了停用词,变量为 $SimYou$ 。文本相似度越高,则该研报提供的新信息越少。

市场关注度采用百度指数作为度量<sup>[43-45]</sup>。百度是国内占有量最大的搜索引擎,自2010年3月谷歌宣布退出中国大陆市场后,2010年~2016年间百度占搜索引擎市场份额的70.17%,而即使在谷歌退出前的2009年~2010年也达到59.51%。因此,选取百度搜索指数作为市场关注的代理具有较好的代表性。百度指数值越大,说明该股票受到投资者的关注度越高。其他变量构造见变量定义表2。

⑫ 英文以空格作为分词标记即可,但是中文分词需要完善的词汇库,词汇库应该尽可能的包含目标文本中所有的词汇。  
 ⑬ Jieba分词是国内团队开发的主要针对中文分词的开源项目,是目前最为流行的分词组件之一。Jieba可跨语言调用,本文用到的是其R语言版本,该项目托管在github上,网址https://github.com/qinwf/jiebaR。  
 ⑭ 停用词是指“的”、“了”、“吗”等没有实际意义的结构助词或语气词以及特殊符号,本文用到的是哈工大停用词表,网址为https://github.com/chdd/weibo/blob/master/stopwords/哈工大停用词表.txt。  
 ⑮ 基础文库是计算IDF必须用到的概念。对某个词语, IDF指这个词语在该研报所属特定领域内的通用程度,基础文库就是这个“特定领域”内所有研报的集合。本文中研报的基础文库是研报目标公司所属行业的所有研报集合。  
 ⑯ 在附录B中,从两个方面验证了以文本相似度代理增量信息的有效性。一是根据已有研究,给出负面荐股等级的研报中包含了更多的新信息。与此一致,本文发现卖空或减持等级的研报文本相似度显著更低。二是现有文献发现,当股票收益率出现跳跃时,说明股票受到了价值相关的新信息冲击。以月度收益率向下跳跃代理负面信息冲击,发现与同一年度其他月份的均值相比,当月研报的文本相似度显著更低;而盈余预测准确度无显著差异;荐股评级反而更高。进一步地,在以文本相似度为被解释变量的回归中,代理收益率是否出现向下跳跃的哑变量系数显著为负;而以盈余预测准确度和荐股评级为被解释变量时,哑变量的系数并不显著(或为正),表明本文的研报文本相似度能够更好地捕捉这一信息冲击。

表 2 主要变量定义

Table 2 Definition of main variables

变量名称	变量符号	变量定义
社交圈	<i>Network</i>	分析师在基金经理中拥有的人脉资源, 详见正文变量设计部分
“圈子”内股票	<i>NetDummy</i>	哑变量, 分析师“圈子”内基金经理持有研报标的股票, 则为 1, 否则为 0
文本相似度	<i>SimYou</i>	保留停用词的研报余弦相似度, 该数值越高, 文本相似度越高, 信息量越少, 详见变量设计
	<i>SimWu</i>	剔除停用词的研报余弦相似度, 该数值越高, 文本相似度越高, 信息量越少, 详见变量设计
盈余预测偏差	<i>RawBias</i>	(预测盈余 - 实际盈余) / 实际盈余
	<i>RelativeBias</i>	在公司两个季报发布期间, 分析师对某公司的 <i>RawBias</i> / 所有分析师对该公司的 <i>RawBias</i> 均值
荐股评级	<i>Rank</i>	研报中给出的评级, -2 表示卖出, -1 表示减持, 0 表示中性, 1 表示增持, 2 表示买入
荐股评级差异	<i>RankDiff</i>	分析师对“圈子”内股票荐股评级均值 - “圈子”外荐股评级均值
跟踪偏好差异	<i>NumDiff</i>	分析师在某一年度对“圈子”内基金经理持股发布的研报数量占该年度该分析师发布研报的总量
是否明星分析师	<i>Star</i>	哑变量, 发布时分析师是明星分析师, 则为 1, 否则为 0
当选次数	<i>Times</i>	分析师在本年度之前当选为明星分析师累计次数
学历	<i>Edu</i>	分析师的最高学历, 博士为 3, 硕士为 2, 本科为 1, 其他为 0
从业时间	<i>Experience</i>	截止研报发布日分析师从业的季度数
券商规模	<i>Asset</i>	分析师所属券商当年总资产的自然对数
券商所属分析师数量	<i>AnaNum</i>	分析师所属券商旗下当年的分析师数量
是否名校毕业	<i>EliteSchool</i>	哑变量, 毕业于 211 或 985 为 1, 否则为 0
百度指数	<i>Baidu</i> (-1, 1)	后 1 天与前 1 天百度搜索指数的差值
	<i>Baidu</i> (-3, 3)	后 3 天与前 3 天百度搜索指数的差值
	<i>Baidu</i> (-5, 5)	后 5 天与前 5 天百度搜索指数的差值
重仓股比例	<i>ZcRatio</i>	按年度将各行业的股票按照基金持股比例中位数为高、低两组, <i>ZcRatio</i> 为某一分析师跟踪高组的研报数量 / 跟踪低组的研报数量
买入评级比例	<i>BuyRatio</i>	分析师买入评级数量 / 所有荐股评级数量
基金持股比例	<i>Fund</i>	研报发布时上市公司该季度的基金持股比例
杠杆率	<i>Lev</i>	总负债 / 总资产
市净率	<i>Mb</i>	总市值 / 净资产
企业性质	<i>Nature</i>	哑变量, 国企为 1, 民企为 0
企业规模	<i>Size</i>	上市公司流通市值的自然对数
个股收益率	<i>Ret</i>	当天股票的收益率
波动率	<i>Sigma</i>	前 30 个交易日个股收益率的波动率
市场收益率	<i>MktRet</i>	发布当天沪深 300 指数收益率
是否可卖空	<i>Short</i>	哑变量, 是融券标的为 1, 非融券标的为 0

2.3 实证模型

社交圈对分析师研报信息含量具有如下影响: 一方面“圈子”(*Network*)拓展了私有信息的获取途径, 分析师籍此能够向市场提供更多的增量文本信息(*SimWu*); 另一方面“圈子”使分析师更

容易陷入其独立性与职业发展追求的利益旋涡, 降低其研报量化指标的客观性。因此, 以 *Network* 为核心解释变量, 并综合现有文献中研报信息的其他相关因素, 利用模型(4)来检验“圈子”对研报文本信息、盈余预测偏差以及荐股评级的影响

$$Quality_{j,t} = \beta_0 + \beta_1 Network_{i,t} + \beta_2 Lag\_Ind_{i,j,t} + \beta_3 Edu_{i,t} + \beta_4 Experience_{i,t} + \beta_5 Asset_{i,t} + \beta_6 AnaNum_{i,t} + \beta_7 EliteSchool_{i,t} + \sum_{k=7}^{17} \beta_k Control_{k,j,t} + \varepsilon_{i,j,t} \quad (4)$$



式中  $Quality_{i,j,t}$  是分析师  $i$  在第  $t$  年发布的研报  $j$  的信息,包括文本相似度 ( $SimWu_{i,j,t}$ )、盈余预测偏差 ( $RelativeBias_{i,j,t}$ ) 以及荐股评级 ( $Rank_{i,j,t}$ );  $Lag. Ind_{i,j,t}$  分别为 3 个被解释变量的一阶滞后项. 参考尹玉刚<sup>[46]</sup>, 分析师学历 ( $Edu_{i,t}$ )、从业时间 ( $Experience_{i,t}$ )、所属券商规模 ( $Asset_{i,t}$ )、券商所属分析师数量 ( $AnaNum_{i,t}$ ) 以及是否名校毕业 ( $EliteSchool_{i,t}$ ) 对研报质量也有显著影响,在模型中均加以控制. 其他控制变量的选取参考现有文献通用的做法<sup>[43-46]</sup>, 在  $\sum_{k=7}^{17} \beta_k Control_{k,j,t}$  中,  $k$  表示第  $k$  个控制变量,包括企业规模 ( $Size$ )、杠

杆率 ( $Lev$ )、市净率 ( $Mb$ )、企业性质 ( $Nature$ )、个股收益率 ( $Ret$ )、波动率 ( $Sigma$ )、市场收益率 ( $MktRet$ )、基金持股比例 ( $Fund$ ) 和是否可卖空 ( $Short$ ) 等. 在回归中同时控制了年度和行业固定效应.

国内外学者对当选明星分析师的影响因素进行了较为深入的研究<sup>[39,40]</sup>. 但是尚未有文献将“圈子”纳入模型,本文将  $Network$  作为核心解释变量引入如下 logistic 模型

$$P(Star_{i,t} = 1) = \frac{\exp(X\beta_{i,t})}{1 + \exp(X\beta_{i,t})} \quad (5)$$

建立如下模型

$$X\beta_{i,t} = \beta_0 + \beta_1 Network_{i,t} + \beta_2 SimWu_{i,t} + \beta_3 RelativeBias_{i,t} + \beta_4 Rank_{i,t} + \beta_5 Times_{i,t} + \beta_6 ZcRatio_{i,t} + \beta_7 BuyRatio_{i,t} + \beta_8 Edu_{i,t} + \beta_9 Experience_{i,t} + \beta_{10} Asset_{i,t} + \beta_{11} AnaNum_{i,t} + \beta_{12} EliteSchool_{i,t} + \varepsilon_{i,t} \quad (6)$$

利用模型 (6) 检验“圈子”以及文本相似度、荐股评级对当选明星分析师概率的影响. 除核心变量  $Network$  和研报质量 ( $SimWu$ 、 $RelativeBias$ 、 $Rank$ ) 之外,参考吴俣立等<sup>[43]</sup>,同时控制如下变量: 评选前分析师累计获奖次数 ( $Times$ ); 跟踪机构重仓股比例

( $ZcRatio$ )、买入评级占比 ( $BuyRatio$ )、分析师学历 ( $Edu$ ) 和是否名校毕业 ( $EliteSchool$ )、从业时间 ( $Experience$ )、所属券商规模 ( $Asset$ ) 以及券商所属分析师数量 ( $AnaNum$ ). 上述变量的计算方法见表 2.

建立如下模型

$$X\beta_{i,t} = \beta_0 + \beta_1 Network_{i,t} + \beta_2 RankDiff_{i,t} + \beta_3 NumDiff_{i,t} + \beta_4 Rank_{i,t} + \beta_5 Times_{i,t} + \beta_6 ZcRatio_{i,t} + \beta_7 BuyRatio_{i,t} + \beta_8 Edu_{i,t} + \beta_9 Experience_{i,t} + \beta_{10} Asset_{i,t} + \beta_{11} AnaNum_{i,t} + \beta_{12} EliteSchool_{i,t} + \varepsilon_{i,t} \quad (7)$$

利用模型 (7) 检验“圈子”以及荐股评级差异 ( $RankDiff$ )、跟踪偏好差异 ( $NumDiff$ ) 对当选概率的影响,之所以将  $RankDiff$  和  $NumDiff$  加入模型是因为  $Network$  仅是分析师潜在的关系,其在评选中作用的发挥依赖于分析师对该关系的维护,而“圈子”内外荐股评级差异和跟踪偏好差异是维护关系的重要手段. 其他控制变量的选取与模型 (6) 一致.

在分析师荐股评级等量化指标存在普遍乐观偏差的先验认知下,投资者更关注研报文本信息. 尹玉刚<sup>[46]</sup>发现投资者更关注明星分析师,而对分析师教育背景、经验、努力程度等影响研报质量的分析师特质信息关注度较低. 在控制上述已有文献中的变量后,  $Network$ 、 $SimWu$  对投资者关注度是否有显著影响? 采用和模型 (4) 一样的控制变量,用如下模型进行考察

$$Baidu_{i,j,t} = \beta_0 + \beta_1 Network_{i,t} + \beta_2 SimWu_{i,j,t} + \beta_3 Rank_{i,j,t} + \beta_4 Times_{i,t} + \sum_{k=5}^{15} \beta_k Control_{k,j,t} + \varepsilon_{i,j,t} \quad (8)$$

式中  $Baidu_{i,j,t}$  是分析师  $i$  跟踪的上市公司  $j$  在研报发布窗口期内的搜索热度指标<sup>[42-44]</sup>. 同时,以异常换手率 ( $Turnover$ ) 作为稳健性检验.

曾当选明星分析师和从未当选明星分析师对不同分析师而言是质的区别,一朝当选,名气与利益随之而来. 他们的研究报告信息有何不同? 用如下模型来检验这一差异

$$Quality_{i,j,t} = \beta_0 + \beta_1 EverStar_{i,t} + \beta_2 NetDummy_{i,j,t} + \beta_3 EverStar_{i,t} \times NetDummy_{i,j,t} + \beta_4 Times_{i,t} + \beta_5 Edu_{i,t} + \beta_6 Experience_{i,t} + \beta_7 Asset_{i,t} + \beta_8 AnaNum_{i,t} + \beta_9 EliteSchool_{i,t} + \sum_{k=10}^{20} \beta_k Control_{k,j,t} + \varepsilon_{i,j,t} \quad (9)$$

式中  $EverStar_{i,t}$  为在年度  $t$  之前分析师  $i$  是否曾当选明星分析师的虚拟变量,其他变量选取与模型(4)相同。

当选明星分析师后,其收入、地位以及人脉资源均得到巨大提升<sup>[6]</sup>,此时分析师对利益与声誉的权衡有何不同?用如下模型来研究这一动态变化

$$Quality_{i,j,t} = \beta_0 + \beta_1 AfterStar_{i,t} + \beta_2 NetDummy_{i,j,t} + \beta_3 AfterStar_{i,t} \times NetDummy_{i,j,t} + \beta_4 Times_{i,t} + \beta_5 Edu_{i,t} + \beta_6 Experience_{i,t} + \beta_7 Asset_{i,t} + \beta_8 AnaNum_{i,t} + \beta_9 EliteSchool_{i,t} + \sum_{k=10}^{20} \beta_k Control_{k,j,t} + \varepsilon_{i,j,t} \quad (10)$$

式中  $AfterStar_{i,t}$  为分析师  $i$  当选明星分析师前/后的虚拟变量,其他变量选取与模型(4)相同。

有64 431个文本相似度的有效观测值,其中  $SimWu$  均值为 0.087,标准差为 0.043,中位数为 0.080,离散程度和偏度较小。 $Rank$  的特征与前人研究一致,体现出较高的乐观偏差(均值为 1.313),在 75% 分位数仍为买入评级,且标准差较小(0.616);  $BuyRatio$  均值达到 0.370,呈左偏态(中位数 0.300)。从跟踪上市公司的基金持股特征来看,  $ZcRatio$  均值为 0.795,且分布离散程度低(标准差为 0.254),中位数为 0.878,呈现右偏态,说明分析师倾向于对基金重仓股发布研报。 $RankDiff$  的均值为 0.102,说明分析师对“圈子”内持股普遍发布了较高的荐股评级。

### 3 实证结果及分析

#### 3.1 描述统计分析

表 3 列示了主要变量的描述统计结果,从表中可以看出,在剔除个人信息(学历、经验等)缺失的数据后共有 7 052 个有效的分析师数据,其  $Network$  均值为 0.133,标准差为 0.572,中位数为 0.005,说明分析师在基金经理中的人脉资源差异较大。剔除研报作者信息缺失的数据后共

表 3 主要变量描述统计

Table 3 Descriptive statistics of main variables

变量	均值	标准差	25%分位数	中位数	75%分位数	样本数
<i>Network</i>	0.133	0.572	-0.291	0.005	0.423	7 052
<i>Edu</i>	2.023	0.443	2.000	2.000	2.000	7 052
<i>Experience</i>	9.752	8.595	3.730	6.880	13.312	7 052
<i>Asset</i>	17.444	1.421	16.511	17.509	18.403	7 052
<i>AnaNum</i>	43.704	25.603	22.000	42.000	60.000	7 052
<i>EliteSchool</i>	0.897	0.304	1.000	1.000	1.000	7 052
<i>Star</i>	0.189	0.391	0.000	0.000	0.000	5 198
<i>Times</i>	0.532	1.254	0.000	0.000	0.000	7 029
<i>SimWu</i>	0.087	0.043	0.063	0.080	0.102	64 431
<i>SimYou</i>	0.091	0.042	0.066	0.084	0.105	64 431
<i>RelativeBias</i>	0.939	2.292	-0.025	0.083	0.412	228 179
<i>Rank</i>	1.313	0.616	1.000	1.000	2.000	218 882
<i>ZcRatio</i>	0.795	0.254	0.692	0.878	1.000	6 006
<i>BuyRatio</i>	0.370	0.358	0.000	0.300	0.667	6 187
<i>RankDiff</i>	0.102	0.313	-0.031	0.052	0.244	48 709
<i>NumDiff</i>	0.043	0.035	0.018	0.032	0.058	48 977
<i>Baidu(-1,1)</i>	0.417	5.083	-0.342	0.032	0.899	6 149
<i>Baidu(-3,3)</i>	0.268	4.203	-0.371	0.118	1.006	6 139
<i>Baidu(-5,5)</i>	0.178	2.125	-0.178	0.031	0.390	6 127

### 3.2 分析师社交圈、利益冲突与研报信息

表4利用模型(4)分析了社交圈对分析师研报文本信息含量、相对预测偏差以及荐股评级的影响,其中  $Edu_{i,j}$ 、 $EliteSchool_{i,j}$ 、 $Experience_{i,j}$ 、 $Asset_{i,j}$ 以及  $AnaNum_{i,j}$ 等分析师特征的频率为年度,其他控制变量则在每次研报发布日进行更新。注意到模型中解释变量之间的量纲不同,以原始数据直接回归将导致系数对量纲过于敏感,且不利于不同变量系数之间的对比。为消除量纲差异以及系数的可比,常用的做法是Greenland等<sup>[47]</sup>提出的标准化回归方法。其思路是先将被解释变

量和所有自变量进行标准化,用标准化后的数据进行多元回归。其回归的系数 $\beta_i$ 代表变量*i*变动1个标准差,被解释变量变动 $\beta_i$ 个标准差。需要注意的是,虽然在模型中加入了前人发现的控制变量,仍无法穷尽被解释变量的所有影响因素,但可以合理地推断这些因素中相当的一部分不会随着时间推移而迅速消失。因此,通过将被解释变量的一阶滞后项加入模型,能够从一定程度上更全面地控制这些因素对被解释变量的影响。下文所有固定效应模型的实证分析均遵循此范式。

表4 分析师社交圈与研报信息含量关系分析

Table 4 Analysis on the relationship between analysts' social ties and information content in research reports

解释变量	被解释变量								
	SimWu			Rank			RelativeBias		
	1	2	3	4	5	6	7	8	9
Network	-0.014 *** (-4.54)	-0.021 *** (-6.37)	-0.020 *** (-5.79)	0.044 *** (20.67)	0.046 *** (20.66)	0.050 *** (20.74)	0.007 *** (2.86)	0.008 *** (3.10)	0.005 ** (1.98)
Lag. Ind	0.625 *** (22.24)	0.625 *** (22.21)	0.633 *** (20.10)	0.545 *** (26.87)	0.544 *** (26.88)	0.544 *** (24.04)	0.001 (0.69)	0.001 (0.73)	0.003 (1.25)
Edu		-0.001 (-0.29)	-0.003 (-0.87)		0.032 *** (17.17)	0.033 *** (16.38)		-0.004* (-1.89)	-0.004* (-1.67)
Experience		-0.006 ** (-2.13)	-0.006 ** (-2.20)		-0.021 *** (-11.21)	-0.025 *** (-12.54)		0.004* (1.93)	0.005 ** (2.12)
Asset		0.009 *** (2.78)	0.010 *** (2.85)		0.027 *** (12.01)	0.036 *** (14.40)		0.002 (0.91)	0.001 (0.33)
AnaNum		0.012 *** (3.68)	0.013 *** (3.69)		-0.053 *** (-22.73)	-0.053 *** (-20.88)		0.001 (0.54)	0.002 (0.79)
EliteSchool		0.009 *** (3.11)	0.011 *** (3.56)		0.002 (0.94)	0.002 (1.00)		-0.003 (-1.44)	-0.003 (-1.53)
Size			-0.015 *** (-3.16)			0.017 *** (5.41)			0.035 *** (10.18)
Lev			0.008* (1.88)			0.003 (1.19)			0.033 *** (10.89)
Mb			0.002 (0.53)			0.005 ** (2.45)			0.001 (0.38)
Nature			-0.006 (-1.58)			-0.008 *** (-3.07)			-0.013 *** (-4.46)
Ret			-0.036 *** (-12.31)			0.022 *** (8.31)			-0.001 (-0.43)
Sigma			-0.065 *** (-15.49)			-0.010 *** (-3.49)			0.002 (0.81)
MktRet			0.001 (0.17)			-0.012 *** (-5.54)			-0.001 (-0.34)
Fund			-0.004 (-1.25)			0.003 (1.19)			0.006 ** (2.57)
Short			0.004 (0.92)			-0.008 *** (-2.85)			-0.039 *** (-12.74)
行业固定效应	控制	控制	控制	控制	控制	控制	控制	控制	控制
年度固定效应	控制	控制	控制	控制	控制	控制	控制	控制	控制
样本数	85 849	85 849	72 136	215 769	215 769	183 258	214 988	214 988	182 578
调整 R <sup>2</sup>	0.396	0.396	0.414	0.300	0.303	0.309	0.129	0.129	0.145

注: 括号内为*t*统计量(采用Newey-West稳健标准误进行调整); \*\*\*, \*\*, \* 分别表示在1%、5%、10%水平上显著;采用Greenland等<sup>[47]</sup>的标准化回归方法,其常数项为0,故未列示。

分析师的信息来源有两部分,上市公司披露的公开信息和分析师获得的私有信息<sup>[26]</sup>。私有信息包括厂房设备维护、日常运营状况、公司管理水平等无须依法披露但与公司价值息息相关的信息。Cheng 等<sup>[28]</sup>发现分析师通过调研活动获取了有价值的私有信息,并有更好的盈余预测及荐股表现,这一现象在重资产的行业(如传统制造业)中更为明显。分析师获得私有信息的另一重要渠道为其“圈子”,即本文定义的社交圈。“圈子”广的分析师能够获得更多的私有信息,因此,其文本信息含量更高,体现为更低的文本相似度(表 4 第 3 列 *Network* 系数为 -0.020, *t* 值为 -5.79),验证了上文提出的 H1a。

分析师一方面通过“圈子”获得了更多的私有信息,另一方面也需要发布更乐观的荐股评级作为补偿。表 4 第 6 列和第 9 列表明,“圈子”显著增加了分析师研报的荐股评级(*Network* 系数为

0.050, *t* 值为 20.74)和盈余预测乐观偏差(*Network* 系数为 0.005, *t* 值为 1.98),这与 H2a 一致。表 4 整体上看,在控制了年度和行业固定效应后,无论是仅以“圈子”作为独立解释变量的单变量分析,还是逐渐加入其他控制变量后的多元回归分析,“圈子”对研报信息含量、盈余预测偏差和荐股评级的系数都符合前文假设。

表 5 和表 6 展示了 logistic 模型(模型(5))的实证结果。出于自身职业生涯的考虑,分析师具有发布乐观荐股评级的倾向。Emery 和 Li<sup>[39]</sup>和吴偲立等<sup>[40]</sup>的研究都表明,高荐股评级能够帮助分析师当选。与前人研究一致,表 5 第 4 列说明较高的荐股评级能够增加当选为明星分析师的概率(*Rank* 系数为 0.928, *z* 值为 3.71)。但是分析师研报的文本信息含量和盈余预测准确度对当选概率无显著影响(第 2 列 *SimWu* 系数为 0.096;第 3 列 *RelativeBias* 系数为 -0.039,均不显著)。

表 5 分析师社交圈、荐股评级与当选明星分析师概率

Table 5 Analysts' social ties, stock recommendations and probability of being a star analyst

解释变量	被解释变量 <i>Star</i>				
	1	2	3	4	5
<i>Network</i>	0.182* (1.74)				0.353** (2.30)
<i>SimWu</i>		0.096 (0.61)			0.091 (0.56)
<i>RelativeBias</i>			-0.039 (-0.32)		-0.101 (-0.38)
<i>Rank</i>				0.928*** (3.711)	0.390 (1.06)
<i>Times</i>	1.859*** (18.44)	1.895*** (13.88)	1.858*** (18.41)	1.854*** (18.38)	1.892*** (13.87)
<i>ZcRatio</i>	0.571*** (4.96)	0.341** (2.2)	0.557*** (4.84)	0.569*** (4.95)	0.352** (2.09)
<i>BuyRatio</i>	0.195* (1.89)	0.133 (0.85)	0.235** (2.33)	-0.523** (-2.31)	-0.258 (-0.75)
<i>Edu</i>	0.214** (2.06)	-0.197 (-1.23)	0.212** (2.04)	0.196* (1.88)	-0.199 (-1.20)
<i>Experience</i>	0.116 (1.07)	0.317** (2.0)	0.099 (0.91)	0.094 (0.87)	0.355** (2.25)
<i>Asset</i>	0.828*** (6.08)	1.062*** (4.82)	0.842*** (6.17)	0.802*** (5.90)	0.975*** (4.44)
<i>AnaNum</i>	1.376*** (10.57)	1.291*** (6.70)	1.421*** (11.14)	1.402*** (11.05)	1.251*** (6.47)
<i>EliteSchool</i>	0.156 (1.37)	0.050 (0.30)	0.186* (1.64)	0.193* (1.71)	-0.026 (-0.15)
样本数	5 923	3 290	5 917	5 917	3 278
对数概率	-2 150.00	-1 098.00	-2 150.704	-2 143.00	-1 092.168
赤池信息准则	4 322.00	2 217.00	4 323.408	4 309.00	2 212.335

注: 括号内为 *z* 统计量; \*\*\*, \*\*, \* 分别表示在 1%、5%、10% 水平上显著; 采用 Greenland 等<sup>[47]</sup>的标准化回归方法, 其常数项为 0 故未列示。

分析师社交圈能够显著增加其当选为明星分析师的概率。“圈子”文化在中国由来已久,其在金融领域的作用逐渐受到学者的广泛关注<sup>[18-20]</sup>。在基金经理投票的评选机制下,明星分析师评选更成了“圈子”发挥作用的温床:分析师更容易获得“圈子”内基金经理的选票,并为基金经理的持股发布高荐股评级。表 5 第 1 列从实证上支持了这一推论。控制其他控制变量后,“圈子”对分析师当选的影响系数为 0.182,并且在 10% 的置信水平上显著。在控制了文本相似度、盈余预测偏差以及荐股评级后(第 5 列),“圈子”的系数增加到 0.353,统计量增加到 2.30,并且荐股评级在加入“圈子”后不再显著,这说明相对于荐股评级,“圈子”是影响分析师当选更重要的因素。

“圈子”为分析师与基金经理的利益交换提供了潜在途径,但其发挥作用还依赖于分析师对关系的维护。从两个角度衡量分析师对“圈子”内基金经理关系的维护,一是分析师在某一年度对“圈子”内基金经理持股荐股评级的均值减去对“圈子”外基金经理持股荐股评级的均值(*Rank-Diff*),二是分析师在某一年度对“圈子”内基金经理持股发布的研报数量占该年度该分析师发布研报的总量,即跟踪偏好(*NumDiff*)。Gu 等<sup>[48]</sup>发现分析师倾向于跟踪与其有校友关系的基金经理持有的股票。

从表 6 看出,*RankDiff*(表 6 第 1 列系数为 0.249, *z* 值为 2.43)和 *NumDiff*(第 2 列系数为 0.242, *z* 值为 2.49)对分析师当选都有显著影响,同时加入这两个变量后各自依然显著(第 3 列 *RankDiff* 系数为 0.193, *z* 值为 1.84; *NumDiff* 系数为 0.197, *z* 值为 1.95)。并且加入这两个变量中的任何一个后“圈子”不再显著(第 1 列、第 2 列和第 3 列 *Network* 的系数分别为 0.137、0.122 和 0.122; *z* 值分别为 1.29、1.16 和 1.15)。从第 3 列进一步发现荐股评级差异(系数 0.193)和跟踪偏好(系数 0.197)对当选的影响大于“圈子”(系数为 0.122)。这说明“圈子”效应的发挥依赖于分析师对基金经理的利益输送。表 5 与表 6 共同验证了 H3。尽管更高的荐股评级有助于分析师当选,进而实现其利益,但文本信息含量却可能影响分析师的市场影响力和公信力。

表 6 分析师跟踪偏好、乐观偏差与当选明星分析师概率  
Table 6 Following preference, optimistic bias and the probability of being stars

解释变量	被解释变量 <i>Star</i>		
	1	2	3
<i>RankDiff</i>	0.249** (2.43)		0.193* (1.84)
<i>NumDiff</i>		0.242** (2.49)	0.197* (1.95)
<i>Network</i>	0.137 (1.29)	0.122 (1.16)	0.122 (1.15)
<i>Times</i>	1.849*** (18.33)	1.852*** (18.36)	1.848*** (18.32)
<i>Rank</i>	0.937*** (3.71)	0.911*** (3.62)	0.940*** (3.71)
<i>ZcRatio</i>	0.570*** (4.95)	0.542*** (4.69)	0.543*** (4.70)
<i>BuyRatio</i>	-0.567** (-2.49)	-0.533** (-2.35)	-0.564** (-2.47)
<i>Edu</i>	0.201* (1.93)	0.202* (1.40)	0.204* (1.95)
<i>Experience</i>	0.108 (0.99)	0.108 (0.10)	0.109 (1.00)
<i>Asset</i>	0.790*** (5.81)	0.802*** (5.89)	0.800*** (5.86)
<i>AnaNum</i>	1.383*** (10.64)	1.376*** (10.58)	1.388*** (10.65)
<i>EliteSchool</i>	0.152 (1.33)	0.151 (1.31)	0.141 (1.22)
样本数	5 917	5 917	5 917
对数概率	-2 140.00	-2 139.00	-2 138.00
赤池信息准则	4 305.00	4 305.00	4 304.00

注:括号内为 *z* 统计量;\*\*\*、\*\*、\* 分别表示在 1%、5%、10% 水平上显著;采用 Greenland 等<sup>[47]</sup>的标准化回归方法,其常数项为 0,故未列示。

利用模型(8)检验了文本信息含量对市场关注度的影响,结果见表 7。实证发现,研报发布后的第 1 天(*Baidu*(-1,1))投资者更关注分析师自身的“圈子”(表 7 第 3 列的系数为 0.012, *t* 值为 1.97)以及当选次数(表 7 第 1 列系数为 0.007, *t* 值为 2.05;而在第 3 列加入 *SimWu* 后 *Times* 不显著, *t* 值为 0.733)。但第 3 天(*Baidu*(-3,3))和第 5 天(*Baidu*(-5,5))投资者更关注研报的信息含量,第 6 列和第 9 列 *SimWu* 系数分别为 -0.018 和 -0.010, *t* 值分别为 -3.09 和 -1.76,而分析师的“圈子”和当选次数对市场关注没有显著影响。验证了 H4。尹玉刚<sup>[46]</sup>将投资者关注明星分析师而忽视分析师特质的现象称为

认知偏差,因为这些分析师特质更能影响研报质量。值得注意的是,从表 7 看出投资者对新发布的研报在第 1 天 ( $Baidu(-1,1)$ ) 出现了关注偏

差,但是随着时间推移,这种偏差在研报发布后第 3 天和第 5 天 ( $Baidu(-3,3)$  和  $Baidu(-5,5)$ ) 消失,市场关注重新回归富含文本信息的研报。

表 7 分析师“圈子”与市场关注(百度指数)

Table 7 Analysts' social ties and market attention ( Baidu Index)

解释变量	被解释变量								
	$Baidu(-1,1)$			$Baidu(-3,3)$			$Baidu(-5,5)$		
	1	2	3	4	5	6	7	8	9
<i>Network</i>	0.009** (2.38)		0.012** (1.97)	0.008** (1.99)		0.009 (1.44)	0.0004 (0.11)		0.007 (1.13)
<i>SimWu</i>		-0.008 (-1.34)	-0.008 (-1.31)		-0.018*** (-3.11)	-0.018*** (-3.09)		-0.010* (-1.78)	-0.010* (-1.76)
<i>Times</i>	0.007** (2.05)	0.004 (0.79)	0.004 (0.73)	-0.002 (-0.57)	-0.004 (-0.70)	-0.004 (-0.74)	-0.004 (-1.02)	-0.005 (-0.92)	-0.005 (-0.96)
<i>Size</i>	-0.032*** (-5.50)	-0.018** (-2.00)	-0.019** (-2.08)	-0.018*** (-3.07)	-0.014 (-1.54)	-0.014 (-1.60)	0.030*** (5.13)	0.021** (2.38)	0.021** (2.33)
<i>Lev</i>	0.002 (0.38)	-0.006 (-0.4)	-0.006 (-0.70)	0.009* (1.76)	0.001 (0.08)	0.001 (0.12)	0.009* (1.66)	0.006 (0.73)	0.006 (0.75)
<i>Mb</i>	0.002 (0.59)	0.003 (0.54)	0.003 (0.54)	0.002 (0.59)	0.0003 (0.05)	0.0003 (0.06)	-0.001 (-0.35)	-0.005 (-0.87)	-0.005 (-0.87)
<i>Nature</i>	0.021*** (4.56)	0.020*** (2.82)	0.020*** (2.83)	0.011** (2.42)	0.018** (2.47)	0.018** (2.48)	-0.011** (-2.39)	-0.006 (-0.77)	(0.005) (-0.76)
<i>Ret</i>	0.041*** (8.87)	0.021*** (3.68)	0.021*** (3.68)	0.041*** (8.97)	0.020*** (3.52)	0.020*** (3.52)	0.055*** (11.92)	0.036*** (6.49)	0.036*** (6.49)
<i>Sigma</i>	-0.007 (-1.56)	0.017** (2.11)	0.017** (2.12)	-0.010** (-2.10)	0.027*** (3.31)	0.027*** (3.32)	-0.014*** (-3.02)	0.049*** (5.93)	0.049*** (5.94)
<i>MktRet</i>	-0.010*** (-2.63)	0.005 (0.81)	0.005 (0.85)	-0.004 (-1.15)	0.033*** (5.08)	0.033*** (5.11)	0.008** (2.04)	0.036*** (5.58)	0.036*** (5.60)
<i>Fund</i>	-0.019*** (-4.93)	-0.029*** (-4.78)	-0.029*** (-4.80)	-0.025*** (-6.53)	-0.030*** (-5.07)	-0.030*** (-5.08)	-0.010*** (-2.61)	-0.011* (-1.81)	-0.011* (-1.82)
<i>Short</i>	0.007 (1.46)	0.008 (0.98)	0.008 (1.01)	0.005 (1.04)	0.005 (0.60)	0.005 (0.63)	-0.008 (-1.61)	-0.003 (-0.41)	-0.003 (-0.39)
行业固定效应	控制	控制	控制	控制	控制	控制	控制	控制	控制
年度固定效应	控制	控制	控制	控制	控制	控制	控制	控制	控制
样本数	79 257	33 727	33 727	79 257	33 727	33 727	79 212	33 701	33 701
调整 $R^2$	0.003	0.001	0.001	0.02	0.01	0.01	0.009	0.008	0.008

注: 括号内为  $t$  统计量(采用 Newey - West 稳健标准误进行调整); \*\*\*, \*\*, \* 分别表示在 1%、5%、10% 水平上显著; 采用 Greenland 等<sup>[47]</sup> 的标准化回归方法, 其常数项为 0 故未列示。

### 3.3 进一步研究

上文研究为分析师社交圈与研报文本信息含量和荐股评级的关系提供了直接证据,并证明了荐股评级及其在“圈子”内外的差异对分析师职业生涯的重要影响,以及研报文本信息含量对市场关注的作用。下面则进一步研究不同分析师群体对其“圈子”内/外基金经理持股研报的差异,以及明星分析师在当选前后研报各指标(文本信息含量和量化指标)的变化,从而

考察“圈子”导致的利益冲突在分析师当选前后的动态过程,结果如表 8 所示。

表 8 表明,从研报文本增量信息的角度,曾当选的明星分析师从“圈子”内获取信息的能力更强,在面板 A 中体现为  $DiffStar$  的绝对值(0.011)大于  $DiffNonStar$  的绝对值(0.009);从动态变化看,明星分析师当选后从“圈子”内获取信息的能力进一步增强,  $DiffAfter$  的绝对值(0.012)大于当选前的  $DiffBefore$  的绝对值(0.005)。

表8 不同分组下“圈子”内外研报信息差异

Table 8 Difference of research report information between inside and outside social ties under different groups

面板 A: 明星/非明星分析师研报信息差异				
是否明星分析师	基金经理持股	<i>SimWu</i>	<i>Rank</i>	<i>RelativeBias</i>
		1	2	3
明星	“圈子”内	0.079 (0.00)	1.421 (0.00)	1.407 (0.00)
	“圈子”外	0.091 (0.00)	1.350 (0.00)	0.629 (0.00)
<i>DiffStar</i>		-0.011 <sup>***</sup> (0.00)	0.071 <sup>***</sup> (0.00)	0.778 <sup>***</sup> (0.00)
非明星	“圈子”内	0.08 (0.00)	1.292 (0.00)	0.978 (0.00)
	“圈子”外	0.089 (0.00)	1.255 (0.00)	0.712 (0.00)
<i>DiffNonStar</i>		-0.009 <sup>***</sup> (0.00)	0.037 <sup>***</sup> (0.00)	0.266 <sup>***</sup> (0.01)
面板 B: 明星分析师当选前后研报信息差异				
当选前后	基金经理持股	<i>SimWu</i>	<i>Rank</i>	<i>RelativeBias</i>
		1	2	3
当选前	“圈子”内	0.081 (0.00)	1.445 (0.00)	1.044 (0.00)
	“圈子”外	0.086 (0.00)	1.249 (0.00)	0.440 (0.00)
<i>DiffBefore</i>		-0.005 <sup>***</sup> (-0.01)	0.196 <sup>***</sup> (0.00)	0.603 <sup>***</sup> (0.00)
当选后	“圈子”内	0.079 (0.00)	1.415 (0.00)	1.467 (0.00)
	“圈子”外	0.091 (0.00)	1.367 (0.00)	0.676 (0.00)
<i>DiffAfter</i>		-0.012 <sup>***</sup> (0.00)	0.048 <sup>***</sup> (0.00)	0.791 <sup>***</sup> (0.00)

注: 括号内为信息差异的  $p$  值; \*\*\* 表示在 1% 水平上显著。

这一证据与上文的分析一致, 分析师当选后从“圈子”内获取信息的能力增强。其次, 在前文论述中, 将荐股评级作为分析师获取信息的成本, 那么这一成本的动态变化有什么特征? 从明星分析师与非明星分析师的静态对比看, 明星分析师整体上对“圈子”内基金经理的持股给予了更高的荐股评级(面板 A 第 2 列中 *DiffStar* 的值为 0.071, 而 *DiffNonStar* 的值为 0.037) 这也是明星分析师能获取更多增量信息的原因。尤其是明星分析师当选前, 对“圈子”内基金经理持股的荐股评级显著更高(面板 B *DiffBefore* 的值为 0.196)。有趣的是, 在当选明星分析师后随着分析师自身财富和市

场地位的迅速提高, 他们出于对声誉的考量, 给出了相对较低的荐股评级(*DiffAfter* 的值为 0.048)。

最后, 曾当选明星的分析师对“圈子”内基金经理持股给予了更乐观的盈余预测(面板 A *DiffStar* 的值为 0.778), 且“圈子”内外的乐观程度差异更大(面板 A *DiffNonStar* 的值为 0.226)。这也表明明星分析师群体给出的盈余预测整体上比非明星分析师更乐观。仅从明星分析师群体看, 当选后他们对“圈子”内的乐观程度更高(面板 B *DiffBefore* 的值为 0.603, 而 *DiffAfter* 的值为 0.791)。结合荐股评级的动态变化, 可以看出明星分析师当选后虽然给予了更客观的荐股评级, 但盈余预测中的

乐观偏差却在增加。

在加入控制变量后,用模型(9)更细致地研究了明星/非明星分析师研报信息的差异,结果见表9。表9第1列表明明星分析师和非明星分析师的文本相似度整体上没有差异(*EverStar*系数为0.006, *t*值为0.96),但他们呈现出共同特征——对“圈子”内股票研报的文本相似度都更低(表9第2列中*NetDummy*的系数为-0.009,且在5%水平上显著)。然而在第3列中加入*NetDummy*和*EverStar*的交叉项后发现其系数为-0.013,且在5%水平上显著,这说明明星分析师对“圈子内外文本相似度的差异要显著大于非明星分析师,即明星分析师更能获得“圈子”内基金经理的私有信息,并体现在文本信息中,也验证了H1b。高的荐股评级(第6列*EverStar*系数为0.094, *t*值为28.40);无论是明星还是非明星分析师,都会对“圈子”内的股票给出更高的荐股评级(第5列*NetDummy*系数为0.038, *t*值为

16.31);与表9结论相对应,发现明星分析师更注重给“圈子”内基金经理持股发布较高的荐股评级(第6列交叉项*EverStar* × *NetDummy*系数为0.008, *t*值为2.17),验证了H2b。对于盈余预测偏差,第9列*EverStar*系数为0.003, *t*值为0.90,这是由于盈余预测偏差的来源有两个:一是分析师的乐观偏差,曾当选的分析师面临更严重的利益冲突;其次,盈余预测偏差也包含了信息,明星分析师有更广泛的信息搜集和处理能力,预测更准确。两个相反的效应相互抵消,导致其系数不显著。同样的,交叉项也没有显著的解释力(第9列*EverStar* × *NetDummy*系数为-0.003, *t*值为-0.81)。首次当选后分析师的知名度大大提高,同时也获得极大经济利益。这对分析师有两方面的重要影响,一是其更加注重自己的市场影响力及公信力;二是随着其地位的提高,他们通过“圈子”内基金经理获取私有信息的能力也得到增强。

表9 明星/非明星分析师研报信息差异

Table 9 Difference of information content between star and non-star analysts

解释变量	被解释变量								
	<i>SimWu</i>			<i>Rank</i>			<i>RelativeBias</i>		
	1	2	3	4	5	6	7	8	9
<i>EverStar</i>	0.006 (0.96)		0.009 (1.44)	0.096*** (29.67)		0.094*** (28.40)	0.002 (0.67)		0.003 (0.90)
<i>NetDummy</i>		-0.009** (-2.41)	0.002 (0.03)		0.038*** (16.31)	0.031*** (8.75)		-0.010*** (-4.28)	-0.008** (-2.15)
<i>EverStar</i> × <i>NetDummy</i>			-0.013** (-2.44)			0.008** (2.17)			-0.003 (-0.81)
控制变量	控制	控制	控制	控制	控制	控制	控制	控制	控制
行业固定效应	控制	控制	控制	控制	控制	控制	控制	控制	控制
年度固定效应	控制	控制	控制	控制	控制	控制	控制	控制	控制
样本数	72 649	72 649	72 649	183 059	183 059	183 059	182 379	182 379	182 379
调整 R <sup>2</sup>	0.080	0.080	0.080	0.096	0.093	0.097	0.146	0.146	0.146

注: 括号内为 *t* 统计量(采用 Newey - West 稳健标准误进行调整); \*\*\*, \*\* 分别表示在 1%、5% 水平上显著; 采用 Greenland 等<sup>[47]</sup> 的标准化回归方法, 其常数项为 0, 故未列示; 为节约篇幅, 其他控制变量也未在表中显示。

利用模型(10)动态研究了分析师入选《新财富》后对“圈子”内外评级差异的动态变化,结果见表10。从第1列可以看出,尽管当选后明星分析师研报的文本相似度整体有所上升(*AfterStar*系数为0.012, *t*值为1.81),但是他们当选后对“圈子”内股票研报的文本相似度显著降低(第3列*AfterStar* × *NetDummy*系数为-0.041, *t*值为-3.18),这一方面是因为他们

更容易获取“圈子”内基金经理的私有信息,另一方面是因为他们更注重自身的市场号召力,有动机提供更有信息的研报。前文提到分析师当选后市场知名度及地位会有显著提高,这很可能会影响其与“圈子”内基金经理的“议价”能力:他们一方面能够更容易获得“圈子”内基金经理所拥有的私有信息,另一方面无需一味地以高荐股评级取悦基金经理。第6列验证了这



一推论,明星分析师当选后对“圈子”内股票的荐股评级依然显著( $AfterStar \times NetDummy$  系数为  $-0.021$ ,  $t$  值为  $-3.40$ )。与表 9 的结论类

似,盈余预测偏差则在明星分析师当选前后没有显著差异(第 9 列第 3 行系数不显著)。表 10 的实证结果验证了 H5。

表 10 明星分析师当选前后研报信息差异

Table 10 Difference of information content between before and after being star analysts

解释变量	被解释变量								
	SimWu			Rank			RelativeBias		
	1	2	3	4	5	6	7	8	9
<i>AfterStar</i>	0.012* (1.81)		0.019*** (2.69)	-0.001 (-0.33)		0.001 (0.29)	0.008*** (2.58)		0.010*** (2.81)
<i>NetDummy</i>		-0.020*** (-3.68)	0.017 (1.33)		0.049*** (16.02)	0.068*** (10.08)		-0.011** (-1.93)	-0.006 (-0.85)
<i>AfterStar</i> × <i>NetDummy</i>			-0.041*** (-3.18)			-0.021*** (-3.40)			-0.006 (-1.00)
控制变量	控制	控制	控制	控制	控制	控制	控制	控制	控制
行业固定效应	控制	控制	控制	控制	控制	控制	控制	控制	控制
年度固定效应	控制	控制	控制	控制	控制	控制	控制	控制	控制
样本数	34 701	34 701	34 701	103 427	103 427	103 427	103 119	103 119	103 119
调整 $R^2$	0.100	0.100	0.100	0.120	0.120	0.120	0.205	0.205	0.205

注: 括号内为  $t$  统计量(采用 Newey-West 稳健标准误进行调整); \*\*\*, \*\*, \* 分别表示在 1%、5%、10% 水平上显著; 为消除量纲差异的影响, 采用 Greenland 等<sup>[47]</sup> 的标准化回归方法, 其常数项为 0, 故未列示; 为节约篇幅, 其他控制变量也未在表中显示。

3.4 稳健性检验

实证分析中, 用剔除了停用词的文本相似度度量研报增量文本信息, 然而作为没有实际意义的结构助词或语气助词以及特殊符号, 停用词从某种程度上代表了特殊的“信息”——无用信息。为保证结论的可靠性, 使用保留了停用词的文本相似度(*SimYou*) 进行稳健性检验, 结果见表 11。表 11 第 3 列加入所有控制变量后, 结论不变(*Network* 系数为  $-0.032$ ,  $t$  值为  $-7.50$ )。

尽管以百度指数作为市场对股票的关注度是学者常用的做法<sup>[43-45]</sup>, 但也有采用换手率作为市场关注度的度量<sup>[49, 50]</sup>。研报文本相似度对市场关注度的影响在替换市场关注度指标后是否依然存在? 本文构建异常换手率(*Turnover*) 指标来检验结果的稳健性。异常换手率是研报发布后  $i$  日与发布前  $i$  日换手率的均值差异( $i$  分别取值 1, 3, 5)。检验结果展示在表 12 中。从中看出研报文本相似度对股票换手率的影响与百度指数具有同样的模式, 结论稳健。

表 11 保留停用词的稳健性检验

Table 11 Robustness test of remaining stop words

解释变量	被解释变量 <i>SimYou</i>		
	1	2	3
<i>Network</i>	-0.029*** (-7.43)	-0.040*** (-9.60)	-0.032*** (-7.50)
<i>Edu</i>		0.0005 (0.13)	-0.009** (-2.33)
<i>Experience</i>		-0.015*** (-4.54)	-0.011*** (-3.03)
<i>Asset</i>		0.021*** (5.23)	0.019*** (4.36)
<i>AnaNum</i>		0.014*** (3.49)	0.022*** (5.023)
<i>EliteSchool</i>		0.015*** (4.23)	0.021*** (5.66)
控制变量	不控制	不控制	控制
行业固定效应	控制	控制	控制
年度固定效应	控制	控制	控制
样本数	87 162	87 162	72 769
调整 $R^2$	0.040	0.050	0.080

注: 括号内为  $t$  统计量(采用 Newey-West 稳健标准误进行调整); \*\*\*, \*\* 分别表示在 1%、5% 水平上显著; 采用 Greenland 等<sup>[47]</sup> 的标准化回归方法, 其常数项为 0, 故未列示; 为节约篇幅, 其他控制变量也未在表中显示。

表 12 分析师“圈子”与市场关注度(异常换手率)  
Table 12 Analysts' social ties and market attention (Turnover)

解释变量	被解释变量		
	Turnover(-1,1)	Turnover(-3,3)	Turnover(-5,5)
SimWu	-0.002 (-0.66)	-0.011*** (-3.07)	-0.016*** (-4.23)
Times	0.010*** (2.68)	0.005 (1.32)	0.003 (0.76)
控制变量	控制	控制	控制
行业固定效应	控制	控制	控制
年度固定效应	控制	控制	控制
样本数	72 641	72 641	72 641
调整 R <sup>2</sup>	0.07	0.07	0.07

注: 括号内为 *t* 统计量(采用 Newey-West 稳健标准误进行调整); \*\*\* 表示在 1% 水平上显著; 采用 Greenland 等<sup>[47]</sup> 的标准化回归方法, 其常数项为 0 故未列示; 为节约篇幅, 其他控制变量也未在表中显示。

在前文中, 用模型(10)研究了分析师当选明星后股票研报质量(SimWu 和 Rank)的变化, 发现明星分析师更加注重研报的文本信息含量, 而降低了“圈

子”内股票荐股评级。为验证这一结论的可靠性, 直接以 RankDiff 及 NumDiff 为被解释变量, 采用如下模型检验分析师当选明星后的动态变化

$$Diff_{i,t} = \beta_0 + \beta_1 AfterStar_{i,t} + \beta_2 EverStar_{i,t} + \beta_3 Network_{i,t} + \beta_4 Edu_{i,t} + \beta_5 Experience_{i,t} + \beta_6 Asset_{i,t} + \beta_7 AnaNum_{i,t} + \beta_8 EliteSchool_{i,t} + \sum_{k=9}^{19} \beta_k Control_{k,j,t} + \varepsilon_{i,j,t} \quad (11)$$

式中  $Diff_{i,t}$  代表 RankDiff 和 NumDiff, 控制变量的选取与模型(4)相同, 回归结果展示在表 13 中。

表 13 明星/非明星分析师研报差异稳健性检验

Table 13 Robustness test of difference of information content between star and non-star analysts

解释变量	被解释变量											
	RankDiff			NumDiff			RankDiff			NumDiff		
	1	2	3	4	5	6	7	8	9	10	11	12
EverStar	0.060*** (13.32)	0.052*** (10.39)	0.050*** (9.00)	0.052*** (26.34)	0.027*** (12.61)	0.018*** (7.70)						
AfterStar							-0.013** (-2.25)	-0.015** (-2.40)	-0.012** (-2.30)	0.015*** (5.33)	0.021*** (7.66)	0.024*** (8.10)
Network		0.048*** (6.84)	0.035*** (4.49)		0.261*** (109.20)	0.263*** (101.00)		-0.003 (-0.29)	-0.011 (-1.16)		0.306*** (100.20)	0.312*** (93.37)
控制变量	不控制	不控制	控制	不控制	不控制	控制	不控制	不控制	控制	不控制	不控制	控制
行业固定效应	控制	控制	控制	控制	控制	控制	控制	控制	控制	控制	控制	控制
年度固定效应	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	47 558	47 558	38 684	223 999	223 999	185 730	29 644	29 644	23 948	126 179	126 179	104 638
调整 R <sup>2</sup>	0.070	0.070	0.070	0.130	0.180	0.180	0.080	0.100	0.100	0.160	0.230	0.230

注: 括号内为 *t* 统计量(采用 Newey-West 稳健标准误进行调整); \*\*\*、\*\* 分别表示在 1%、5% 水平上显著; 采用 Greenland 等<sup>[47]</sup> 的标准化回归方法, 其常数项为 0 故未列示; 为节约篇幅, 其他控制变量也未在表中显示。

与表 9 和表 10 的结论吻合, 发现相对于非明星分析师, 明星分析师对“圈子”内股票的荐股评级更高, 也更倾向于对“圈子”内的股票发布

研报。另外, 我国部分券商存在双总部或多总部的情况, 这可能会导致正文中基于注册地计算的分

析师“地域圈”有偏差。实际上,虽然很多券商存在双总部或多总部的情况,但大部分分析师主要在注册地的研究所办公,其它总部的研究所主要是为在当地开展业务的分析师设置的临时办公场所。但是依然检验了双总部或多总部情形对分析师“地域圈”构建造成的可能影响。从各地方证监局网站统计了各自辖区内券商的情况,得到每家券商的总部数量与所在地,发现样本所包含的133家券商中,有72家券商只有1家总部,双总部或多总部的有61家。对于多总部或双总部的券商,无法确切地知道某个分析师所属的确切总部进而计算“地域圈”,因此表14基于72家单总部券商旗下的分析师样本重复了表4的实证模型。与表4相比,表14中主要解释变量的系数与显著性保持一致。同样的,社交圈对分析师预测偏差的影响也保持稳健,这些结论进一步表明多总部券商的情况并不影响本文主要结论。

表14 基于单总部券商的子样本分析

Table 14 Subsample analysis based on single-headquarter brokerage

解释变量	被解释变量		
	<i>SimWu</i>	<i>Rank</i>	<i>RelativeBias</i>
<i>Network</i>	-0.023 *** (-4.04)	0.049 *** (11.54)	0.011 ** (2.25)
<i>Lag. Ind</i>	0.606 *** (12.38)	0.519 *** (13.72)	0.133 *** (3.83)
控制变量	控制	控制	控制
行业固定效应	控制	控制	控制
年度固定效应	控制	控制	控制
样本数	30 759	65 066	64 779
调整 $R^2$	0.381	0.311	0.020

注:括号内为  $t$  统计量(采用 Newey-West 稳健标准误进行调整);\*\*\*、\*\* 分别表示在 1%、5% 水平上显著;采用 Greenland 等<sup>[47]</sup>的标准化回归方法,其常数项为 0,故未列示。

### 4 结束语

研究报告是分析师向市场传递信息的重要载体,以往研究多以盈余预测偏差和荐股评级等量化指标评价研报的信息价值,但这些量化指标本质上是分析师基于其公共信息和私有信息的结论

性预测,而非信息本身。分析师从解读信息到作出预测的过程是个“黑匣子”,具有较强的主观性且易受到外部利益冲突的干扰,从投资者角度来说这些指标也不具有及时的可验证性,以其作为评价分析师信息中介功能的依据可能片面且不客观。因此,本文创新性地以文本相似度衡量研报信息含量,并通过量化分析师社交圈的大小,研究“圈子”对分析师独立性的影响及其在研报信息含量中的体现,运用逻辑演绎并实证得出一些有价值的发现。

“圈子”是分析师获取私有信息的重要途径,人脉广泛的分析师凭借信息优势能够向市场提供具有显著增量信息的研究报告。实证发现“圈子”广的分析师研报文本信息含量更高,这一现象在“圈子”内基金经理持股的研报中更为显著。同时在“社会关系资本化,资本市场江湖化”的股票市场,“圈子”增加了分析师陷入利益漩涡的概率,分析师可能以高级别的研报谋求私人利益。对该问题的实证分析证实了这一推论,“圈子”广的分析师整体上荐股评级更高,并且其对“圈子”内基金经理持股的荐股评级显著高于“圈子”外。

本文进一步分析“圈子”效应在分析师研报中发挥作用的机制。作为重要的外部激励,当选明星分析师是分析师职业生涯的终极追求,而基金经理手握明星分析师评选的投票权。实证发现对“圈子”内基金经理持股发布高级别能显著增加分析师入选《新财富》的概率,从某种程度上讲,高荐股评级是分析师获取基金经理选票的筹码。尽管文本信息含量对当选概率没有显著影响,但是高信息含量的研报文本能够引起更大的市场反应,标的股票获得了更广泛的市场关注。

作为分析师职业发展的重要节点,入选《新财富》后分析师的声誉与利益发生了显著改变,此时“圈子”效应发生了何种变化?对这一动态问题的研究发现当选明星后分析师更注重文本信息含量,而对“圈子”内股票的荐股评级有所下降。其原因是分析师的职业生涯利益得到实现后更注重声誉,并且明星分析师获取私有信息的途径更加多元化,而以高级别获取选票的动机减弱。

本文的研究价值主要体现在两方面。一是为全面评价分析师研报信息含量提供了新的方法。不同于以往局限于研报量化指标的研究, 本文将其范围拓展到了更为客观和重要的研报文本, 并且相对于量化指标, 投资者更关注研报的文本信息。二是为“圈子”在资本市场的作用提供了新

的补充, 发现“圈子”对分析师具有双重影响: 一方面为分析师发现私有信息提供了便利, 有利于其信息中介作用的发挥; 另一方面影响分析师的独立性, 提高了荐股评级的乐观偏差。这为投资者如何充分利用研报信息以及规范分析师行为提供了新的思考与启示。

#### 参考文献:

- [1]朱红军, 何贤杰, 陶林. 中国的证券分析师能够提高资本市场的效率吗——基于股价同步性和股价信息含量的经验证据[J]. 金融研究, 2007, 29(2): 110-121.  
Zhu Hongjun, He Xianjie, Tao Lin. Can securities analysts improve the efficiency of capital market in China[J]. Journal of Financial Research, 2007, 29(2): 110-121. (in Chinese)
- [2]Altunkılıç O, Hansen R S. On the information role of stock recommendation revisions[J]. Journal of Accounting and Economics, 2009, 48(1): 17-36.
- [3]Barber B M, Lehavy R, Trueman B. Ratings changes, ratings levels, and the predictive value of analysts' recommendations[J]. Financial Management, 2010, 39(2): 533-553.
- [4]周铭山, 林靖, 许年行. 分析师跟踪与股价同步性——基于过度反应视角的证据[J]. 管理科学学报, 2016, 19(6): 49-73.  
Zhou Mingshan, Lin Jing, Xu Nianhang. Star analyst coverage and stock price synchronicity: Empirical evidence based on market overreaction[J]. Journal of Management Sciences in China, 2016, 19(6): 49-73. (in Chinese)
- [5]Chen P F, Novoselov K E, Wang Y. Regulatory effects on analysts' conflicts of interest in corporate financing activities: Evidence from NASD Rule 2711[J]. Journal of Corporate Finance, 2018, 48: 658-679.
- [6]Hong H, Kubik J D. Analyzing the analysts: Career concerns and biased earnings forecasts[J]. The Journal of Finance, 2003, 58(1): 313-351.
- [7]许年行, 江轩宇, 伊志宏. 分析师利益冲突, 乐观偏差与股价崩盘风险[J]. 经济研究, 2012, 47(7): 127-140.  
Xu Nianhang, Jiang Xuanyu, Yi Zhihong. Conflicts of interest, analyst optimism and stock price crash risk[J]. Economic Research Journal, 2012, 47(7): 127-140. (in Chinese)
- [8]Tan H T, Duh R R, Chiu S B, et al. Does an asset management firm's stock holding made in response to buy-side analysts' prior recommendations induce subsequent forecast optimism? [J]. Behavioral Research in Accounting, 2016, 28(2): 55-68.
- [9]Yin Y. Fund portfolio holdings and affiliated analyst's objectivity: Do equity and employment relationship matter? [J]. Journal of Behavioral Finance, 2018, 19(1): 16-29.
- [10]储一昀, 仓勇涛. 财务分析师预测的价格可信吗? ——来自中国证券市场的经验证据[J]. 管理世界, 2008, 24(3): 58-69.  
Chu Yiyun, Cang Yongtao. Are the prices predicted by financial analysts reliable? Empirical evidence from Chinese securities market[J]. Management World, 2008, 24(3): 58-69. (in Chinese)
- [11]潘越, 戴亦一, 林超群. 信息不透明, 分析师关注与个股暴跌风险[J]. 金融研究, 2011, 33(9): 138-151.  
Pan Yue, Dai Yiyi, Lin Chaoqun. Information opacity, analyst attention and stock price crash risk[J]. Journal of Financial Research, 2011, 33(9): 138-151. (in Chinese)
- [12]姜超. 证券分析师, 内幕消息与资本市场效率[J]. 经济学(季刊), 2013, 12(2): 429-452.

- Jiang Chao. Securities analyst, financial analysts, insider information and efficiency of capital market [J]. *China Economic Quarterly*, 2013, 12(2): 429–452. (in Chinese)
- [13] 于季胜, 王成龙, 王艳艳. 分析师社交媒体在信息传播效率中的作用——基于分析师微博的研究 [J]. *管理科学学报*, 2019, 22(7): 107–126.
- Yu Lisheng, Wang Chenglong, Wang Yanyan. The role of analysts' social media in the information dissemination: Evidence from analysts' microblogs [J]. *Journal of Management Sciences in China*, 2019, 22(7): 107–126. (in Chinese)
- [14] Ramnath S, Rock S, Shane P. The financial analyst forecasting literature: A taxonomy with suggestions for further research [J]. *International Journal of Forecasting*, 2008, 24(1): 34–75.
- [15] 张雪兰, 何德旭. 证券分析师利益冲突影响投资者利益吗——一个经验研究评述 (1995–2007) [J]. *金融研究*, 2008, 30(7): 170–183.
- Zhang Xuelan, He Dexu. Are investors harmed by financial analysts' conflicts of interest a literature review on empirical study between 1995–2007 [J]. *Journal of Financial Research*, 2008, 30(7): 170–183. (in Chinese)
- [16] Bradshaw M T. Analyst information processing, financial regulation, and academic research [J]. *The Accounting Review*, 2009, 84(4): 1073–1083.
- [17] Huang A H, Zang A Y, Zheng R. Evidence on the information content of text in analyst reports [J]. *The Accounting Review*, 2014, 89(6): 2151–2180.
- [18] 申宇, 赵静梅, 何欣. 校友关系网络, 基金投资业绩与“小圈子”效应 [J]. *经济学(季刊)*, 2015, 14(1): 403–428.
- Shen Yu, Zhao Jingmei, He Xin. Alumni networks, funds' performance and “Small World” effect [J]. *China Economic Quarterly*, 2015, 14(1): 403–428. (in Chinese)
- [19] Cohen L, Frazzini A, Malloy C. Sell-side school ties [J]. *The Journal of Finance*, 2010, 65(4): 1409–1437.
- [20] Horton J, Serafeim G. Security Analyst Networks, Performance and Career Outcomes [R]. SSRN, 2009. [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1522313](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1522313)
- [21] Hoberg G, Phillips G. Product market synergies and competition in mergers and acquisitions: A text-based analysis [J]. *The Review of Financial Studies*, 2010, 23(10): 3773–3811.
- [22] Hoberg G, Phillips G. Text-based network industries and endogenous product differentiation [J]. *Journal of Political Economy*, 2016, 124(5): 1423–1465.
- [23] Bushman R M, Chen J V, Williams C D. Informativeness and Timeliness of 10–K Text Similarity for Predicting Tail-Risk Comovement [R]. SSRN, 2017.
- [24] 董大勇, 张尉, 赖晓东, 等. 谁领先发布: 中国证券分析师领先——跟随影响因素的实证研究 [J]. *南开管理评论*, 2012, 15(5): 56–63.
- Dong Dayong, Zhang Wei, Lai Xiaodong, et al. Who is leading to release? The empirical study about influencing factors of the security analysts leading-following forecast in China [J]. *Nankai Business Review*, 2012, 15(5): 56–63. (in Chinese)
- [25] 贾琬娇, 洪剑峭, 徐媛媛. 我国证券分析师实地调研有价值吗? ——基于盈余预测准确性的一项实证研究 [J]. *投资研究*, 2015, (4): 96–113.
- Jia Wanjiao, Hong Jianqiao, Xu Yuanyuan. Are analysts' site visits valuable in China? An empirical study based on earnings forecast accuracy [J]. *Investment Research*, 2015, (4): 96–113. (in Chinese)
- [26] 谭跃, 钟子英, 管总平. 公平信息披露规则能缓解证券分析师的利益冲突吗 [J]. *南开管理评论*, 2013, 16(4): 43–54.
- Tan Yue, Zhong Ziyong, Guan Zongping. Can regulation of fair disclosure alleviate analysts' conflicts of interest [J].

- Nankai Business Review ,2013 ,16( 4) : 43 - 54. ( in Chinese)
- [27]谭松涛,崔小勇. 上市公司调研能否提高分析师预测精度[J]. 世界经济,2015,(4): 126 - 145.  
Tan Songtao , Cui Xiaoyong. Can analysts' site visits improve earnings forecast accuracy [J]. The Journal of World Economy ,2015 ,( 4) : 126 - 145. ( in Chinese)
- [28]Cheng Q , Du F , Wang X , et al. Seeing is believing: Analysts' corporate site visits [J]. Review of Accounting Studies , 2016 ,21( 4) : 1245 - 1286.
- [29]肖斌卿,彭毅,方立兵,等. 上市公司调研对投资决策有用吗——基于分析师调研报告的实证研究[J]. 南开管理评论,2017,20(1): 119 - 131.  
Xiao Binqing , Peng Yi , Fang Libing , et al. Is survey on listed company valuable to investment decision? An empirical study based on analysts' investigation report [J]. Nankai Business Review ,2017 ,20( 1) : 119 - 131. ( in Chinese)
- [30]Tsao A. When a stock's rating and target collide [N]. Business Week Onlin ,2002 -04 -25.
- [31]Asquith P , Mikhail M B , Au A S. Information content of equity analyst reports [J]. Journal of Financial Economics ,2005 , 75( 2) : 245 - 282.
- [32]Twedt B , Rees L. Reading between the lines: An empirical examination of qualitative attributes of financial analysts' reports [J]. Journal of Accounting and Public Policy ,2012 ,31( 1) : 1 - 21.
- [33]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.
- [34]伊志宏,朱琳,陈钦源. 分析师研究报告负面信息披露与股价暴跌风险[J]. 南开管理评论,2019,22(5): 192 - 206.  
Yi Zhihong , Zhu Lin , Chen Qinyuan. Negative information disclosure of analysts' reports and stock price crash risk [J]. Nankai Business Review ,2019 ,22( 5) : 192 - 206. ( in Chinese)
- [35]马黎珺,伊志宏,张澈. 廉价交谈还是言之有据? ——分析师报告文本的信息含量研究[J]. 管理世界,2019,35(7): 182 - 200.  
Ma Lijun , Yi Zhihong , Zhang Che. Evidence-based statements or cheap talks? The information content of text in analyst reports [J]. Management World ,2019 ,35( 7) : 182 - 200. ( in Chinese)
- [36]程博,潘飞. 语言多样性、信息获取与分析师盈余预测质量[J]. 管理科学学报,2017,20(4): 50 - 70.  
Cheng Bo , Pan Fei. Linguistic diversity , information acquisition and the quality of analyst's earnings forecast [J]. Journal of Management Sciences in China ,2017 ,20( 4) : 50 - 70. ( in Chinese)
- [37]Cohen L , Frazzini A , Malloy C. The small world of investing: Board connections and mutual fund returns [J]. Journal of Political Economy ,2008 ,116( 5) : 951 - 979.
- [38]李志生,李好,刘淳,等. 天使还是魔鬼? ——分析师媒体荐股的市场效应[J]. 管理科学学报,2017,20(5): 66 - 81.  
Li Zhisheng , Li Hao , Liu Chun , et al. Angel or devil? Market effects of media recommendations [J]. Journal of Management Sciences in China ,2017 ,20( 5) : 66 - 81. ( in Chinese)
- [39]Emery D R , Li X. Are the Wall Street analyst rankings popularity contests? [J]. Journal of Financial and Quantitative Analysis ,2009 ,44( 2) : 411 - 437.
- [40]吴俣立,张峥,乔坤元. 信息质量,市场评价与激励有效性——基于《新财富》最佳分析师评选的证据[J]. 经济学(季刊),2016,15(1): 723 - 744.  
Wu Weili , Zhang Zheng , Qiao Kunyuan. Information quality , market evaluation and incentive effectiveness: Evidence from the New Fortune's analysts ranking [J]. China Economic Quarterly ,2016 ,15( 1) : 723 - 744. ( in Chinese)
- [41]Scott J. Social network analysis [J]. Sociology ,1988 ,22( 1) : 109 - 127.

- [42] Yin Y , Liu Y. Quasi-Closeness: A toolkit for social network applications involving indirect connections [J]. *Mathematical Problems in Engineering* , 2017 , 7420281.
- [43] 宋双杰 , 曹 晖 , 杨 坤. 投资者关注与 IPO 异象——来自网络搜索量的经验证据 [J]. *经济研究* , 2011 , 46( S1) : 145 - 155.  
Song Shuangjie , Cao Hui , Yang Kun. Investor attention and IPO anomalies: Evidence from google trend volume [J]. *Economic Research Journal* , 2011 , 46( S1) : 145 - 155. ( in Chinese)
- [44] 刘 锋 , 叶 强 , 李一军. 媒体关注与投资者关注对股票收益的交互作用: 基于中国金融股的实证研究 [J]. *管理科学学报* , 2014 , 17( 1) : 72 - 85.  
Liu Feng , Ye Qiang , Li Yijun. Impacts of interactions between news attention and investor attention on stock returns: Empirical investigation on financial shares in China [J]. *Journal of Management Sciences in China* , 2014 , 17( 1) : 72 - 85. ( in Chinese)
- [45] 李俊生 , 姚东旻. 互联网搜索服务的性质与其市场供给方式初探——基于新市场财政学的分析 [J]. *管理世界* , 2016 , 32( 8) : 1 - 15.  
Li Junsheng , Yao Dongmin. The nature of internet search service and its market supply mode: Based on the analysis of new market public finance [J]. *Management World* , 2016 , 32( 8) : 1 - 15. ( in Chinese)
- [46] 尹玉刚. 分析师教育 , 经验和努力: 投资者关注与认知偏差 [J]. *投资研究* , 2014 , 33( 12) : 126 - 141.  
Yin Yugang. Analysts' education background , experience and efforts: Investor attention and cognitive bias [J]. *Investment Research* , 2014 , 33( 12) : 126 - 141. ( in Chinese)
- [47] Greenland S , Schlesselman J J , Criqui M H. The fallacy of employing standardized regression coefficients and correlations as measures of effect [J]. *American Journal of Epidemiology* , 1986 , 123( 2) : 203 - 208.
- [48] Gu Z , Li Z , Yang Y G , et al. Friends in need are friends indeed: An analysis of social ties between financial analysts and mutual fund managers [J]. *The Accounting Review* , 2019 , 94( 1) : 153 - 181.
- [49] Chemmanur T J , Yan A. Advertising , attention , and stock returns [J]. *Quarterly Journal of Finance* , 2019 , 9( 3) : 1 - 5.
- [50] 权小锋 , 吴世农. 投资者关注 , 盈余公告效应与管理层公告择机 [J]. *金融研究* , 2010 , 32( 11) : 90 - 107.  
Quan Xiaofeng , Wu Shinong. Investor attention , earnings announcement effect and announcement timing of management [J]. *Journal of Financial Research* , 2010 , 32( 11) : 90 - 107. ( in Chinese)
- [51] Jiang G J , Zhu K X. Information shocks and short-term market underreaction [J]. *Journal of Financial Economics* , 2017 , 124( 1) : 43 - 64.
- [52] Barndorff-Nielsen O E , Shephard N. Econometrics of testing for jumps in financial economics using bipower variation [J]. *Journal of Financial Econometrics* , 2006 , 4( 1) : 1 - 30.

## Analysts' social ties , interest conflicts and text information in research reports

LIU Ya-hui , HUANG Kai , YIN Yu-gang\* , WANG Yong

School of Finance , Southwestern University of Finance and Economics , Chengdu 611130 , China

**Abstract:** Using a novel text mining method to quantify the information content in analysts' reports , as well as a social network analysis ( SNA) method to quantify analysts' social ties , the paper investigates how interest

conflicts induced by social ties impact the information content and optimistic bias in analysts' reports. The paper finds that analysts who maintain strong social ties offer more informative reports and achieve significant career success. There is a tradeoff between analysts' reputation (by informative reports) and career concerns (by optimistic bias). The tradeoff is dynamic, which is supported by the empirical evidence that, after being stars, analysts give more importance to report informativeness and show less optimistic bias. Our findings provide a new perspective to appraise analyst reports, as well as additional evidence on the relationship between social ties and analyst behaviors.

**Key words:** analyst; social ties; text analysis; interest conflicts

附录 A

间接关系的计算

社会网络中除了直接关系还存在着广泛的间接关系,忽略这些间接关系将导致信息损失. Yin 和 Liu<sup>[42]</sup>的研究为间接关系和直接关系的统一计算提供了标准. 在一个网络图中,节点  $V_d$  和  $V_s$  (图 A1) 可以通过多条路径建立间接联系,其中包含节点数最少的路径称为测地线(该图中有两条测地线  $V_d-V_2-V_3-V_s$ ,  $V_d-V_6-V_5-V_s$ ),均包含 3 个节点,6 条连线. 在选择最短路径时满足两个条件:一是节点优先,即该路径必须是测地线(包含的节点数最少);二是权重优先,当存在多条测地线时,选其权重最大者为最优路径,在图 A1 中路径  $V_d-V_6-V_5-V_s$  的权重较  $V_d-V_2-V_3-V_s$  权重,因此前者是最优路径.

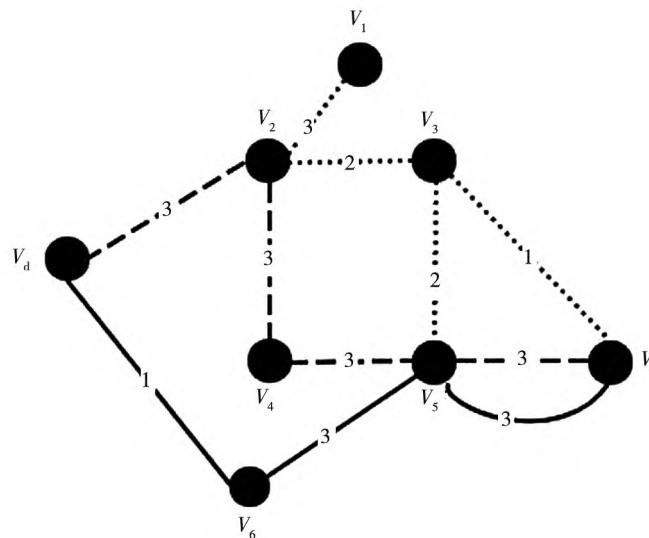


图 A1 最优路径选择

Fig. A1 Optimal path selection

在确定最优路径后, Yin 和 Liu<sup>[42]</sup> 提出对该路径(间接关系)进行指数衰减的两步法. 首先将该路径中所有连线的权重取均值,  $V_d-V_6-V_5-V_s$  的连线均值为  $\frac{1+3+3}{3}$ ; 然后将该均值按连线数进行指数衰减  $\frac{1+3+3}{3} \times 10^{-3}$  作为对间接关系的“惩罚”,并以该数值替换原连接矩阵中的 0(在原矩阵中,无直接关系的节点之间关系权重为 0).

在对分析师构建校友圈、调研圈、地域圈的加总矩阵后参考 Yin 和 Liu<sup>[42]</sup> 的方法替换间接关系权重. 新的连接矩阵包含间接关系,避免了传统社会网络分析方法的信息损失,据此计算的分析师社交圈更为全面.



附录 B

研报文本相似度与增量信息

本文的核心解释变量之一是研报的文本相似度,用于代理该研报所提供增量信息的多寡。本附录则从卖出评级的研报和股票收益率出现跳跃时的研报两个子样本,验证该度量比荐股评级、盈余预测准确度等传统量化指标提供了更为客观有效的信息。

由于分析师中普遍存在着乐观偏差,绝大部分研报给出了买入或卖出的荐股评级。在此背景下,现有研究认为,给出卖出评级的研报包含了更多的信息,基于此,检验此类研报的文本相似度是否更低。根据荐股评级将样本中所有的研报分为卖出、减持、中性、增持和买入5组,发现买入评级的研报有38 828份,占比达45.06%,增持评级占比更高达48.98%,二者(乐观评级)占比之和超过94%。中性评级的研报有4 895份,占比5.68%。反观减持或卖出研报,二者研报总数只有234份,占比仅为0.27%。这表明本文的数据与广泛证实的分析师乐观偏差一致,并且暗示,这些稀有的卖出评级研报包含了更多的信息。

对荐股评级不同的分组内研报的文本相似度进行了均值及差异的T检验。结果如表B1所示。发现卖出评级的研报文本相似度最低(0.076),且其文本相似度显著低于其他分组的研报。在面板B中,bootstrap随机抽样的结论与面板A保持一致,都表明卖出评级的研报中包含了更多的文本信息。

表 B1 基于荐股评级分组的文本相似度差异性检验

Table B1 Text similarity difference test based on recommendation stock rating group

文本相似度	面板 A						面板 B				
	评级	买入	增持	中性	减持	卖出	买入	增持	中性	减持	卖出
SimWu	均值	0.085	0.092	0.094	0.081	0.076	0.091	0.104	0.096	0.085	0.084
	t	4.06	4.34	3.65	3.15	3.23	2.62	2.04	1.85	2.66	1.95
	样本数	38 828	42 204	4 895	39	195					
比较	评级	卖出-买入	卖出-增持	卖出-中性	卖出-减持	卖出-卖出	卖出-增持	卖出-中性	卖出-减持		
	均值	-0.009	-0.016	-0.018	-0.005	-0.007	-0.020	-0.012	-0.001		
	t	-3.09	-5.14	-5.29	-4.31	-3.27	-4.06	-6.36	-8.09		

注: 面板 A 是利用全样本(共 86 161 份研报)计算的均值及 t 统计量; 为避免某一分组内个别研报的文本相似度出现极端值, 从而影响均值, 面板 B 用随机放回抽样的 bootstrap 法重新计算, 每次抽样的样本数量为 20, 抽样次数为 100 000, 并据此计算均值和 t 统计量。

从信息冲击的角度证明文本相似度是研报信息的更好代理变量。Jiang 和 Zhu<sup>[51]</sup>认为, 当股票价格出现大幅上升或下降时, 常常伴随着影响公司估值的有效信息冲击。参考他们的做法, 首先用 Barndorff-Nielsen 和 Shephard<sup>[52]</sup>提出的 t 统计量基于日度收益率数据计算当月收益率是否出现跳跃, 然后分别检验在股价出现跳跃的月份 t 及前后各 1 个月(t-1, t+1)的研报文本相似度、盈余预测和荐股评级是否与当年的其他月份显著不同。由于乐观偏差的存在, 本就乐观的盈余预测和荐股评级在收益率向上跳跃时虽然会更乐观, 但差异可能较小。因此仅考察收益率出现向下跳跃时的情况。页面冲击时研报信息差异的 t 检验结果如表 B2 所示。

表 B2 面板 A 中, 可以看到在负面信息冲击前 1 个月(t-1), 研报文本相似度(SimWu)与当年其他月份并无显著差异, t 统计量仅为 0.91; 而当月(t)的 SimWu 显著低于其他月份, t 统计量为 -1.81; 后 1 个月(t+1)的 SimWu 则更低, 显著性水平也进一步提高。

面板 B 采用双向固定效应模型, 在控制其他变量后进一步考察这一关系是否稳健。

表 B2 负面冲击时研报信息指标的变化

Table B2 Changes of information index is reported under the negative shocks

面板 A: 负面冲击窗口期( $t-1$ 至 $t+1$ 月) 与其他月份的研报信息差异 $t$ 检验									
参数	SimWu			Rank			RelativeBias		
	$y_{t-1}$	$y_t$	$y_{t+1}$	$y_{t-1}$	$y_t$	$y_{t+1}$	$y_{t-1}$	$y_t$	$y_{t+1}$
均值	0.002	-0.003	-0.005	0.039	0.029	0.072	0.086	0.056	-0.130
$t$ 统计	0.91	-1.81	-2.20	1.85	2.41	3.92	1.43	1.66	-2.86
样本数	969	707	580	2 143	1 629	1 618	2 143	1 629	1 618
面板 B: 回归分析									
解释变量	SimWu			Rank			RelativeBias		
	1	2	3	4	5	6	7	8	9
	$y_{t-1}$	$y_t$	$y_{t+1}$	$y_{t-1}$	$y_t$	$y_{t+1}$	$y_{t-1}$	$y_t$	$y_{t+1}$
<i>Jump</i>	0.001 (0.217)	-0.006* (-1.66)	-0.015*** (-4.15)	0.001 (0.36)	0.004* (1.88)	0.007*** (3.08)	0.001 (0.26)	-0.0001 (-0.02)	0.001 (0.29)
<i>Network</i>	-0.038*** (-9.24)	-0.032*** (-7.28)	-0.038*** (-7.29)	0.065*** (26.00)	0.067*** (24.32)	0.067*** (24.34)	0.003 (1.52)	0.004 (1.56)	0.004 (1.58)
<i>Edu</i>	-0.009** (-2.06)	-0.008** (-2.26)	-0.001** (-2.27)	0.040*** (18.34)	0.043*** (18.74)	0.044*** (18.76)	-0.003* (-1.55)	-0.004* (-1.69)	-0.004* (-1.71)
<i>Experience</i>	-0.011*** (-3.03)	-0.011*** (-3.02)	-0.011*** (-3.03)	-0.021*** (-9.85)	-0.029*** (-12.52)	-0.028*** (-12.54)	0.004* (1.86)	0.004* (1.92)	0.005* (1.93)
<i>Asset</i>	0.020*** (4.30)	0.019*** (4.29)	0.019*** (4.25)	0.023*** (8.64)	0.037*** (13.06)	0.039*** (13.08)	0.008 (0.36)	0.0005 (0.17)	0.001 (0.22)
<i>AnaNum</i>	0.023*** (5.12)	0.022*** (5.07)	0.016*** (5.18)	0.005 (1.34)	0.003 (0.99)	0.009 (1.33)	-0.063*** (-21.91)	-0.064*** (-21.98)	-0.064*** (-22.00)
<i>EliteSchool</i>	0.022*** (5.67)	0.021*** (5.55)	0.015*** (4.30)	-0.001 (-0.34)	-0.002 (-0.70)	-0.008 (-0.92)	-0.003 (-1.61)	-0.004 (-1.62)	-0.004 (-1.63)
<i>Size</i>	-0.215*** (-36.69)	-0.215*** (-36.61)	-0.215*** (-36.57)	0.178*** (50.08)	0.177*** (50.05)	0.176*** (50.03)	0.040*** (11.04.)	0.038*** (10.96)	0.039*** (10.98)
控制变量	控制	控制	控制	控制	控制	控制	控制	控制	控制
样本数	72 136	72 136	72 136	183 258	183 258	183 258	183 258	183 258	183 258
调整 $R^2$	0.080	0.085	0.087	0.090	0.091	0.091	0.140	0.138	0.139

注: 基于 BNS 统计量, 当  $t$  月的个股日收益率出现向下跳跃时, 视为该月出现了负面的信息冲击. 本表检验负面冲击的窗口期( $t-1$  至  $t+1$ ) 内每个月的研报信息指标与非窗口期相比的变化, 分别用  $y_{t-1}$ ,  $y_t$ ,  $y_{t+1}$  表示. 面板 A 是单变量的均值差异  $t$  检验. 面板 B 是控制了其他变量的回归结果; 括号内为  $t$  统计量(采用 Newey - West 稳健标准误进行调整); \*\*、\* 分别表示在 1%、5%、10% 水平上显著; 采用 Greenland 等<sup>[47]</sup> 的标准化回归方法, 其常数项为 0, 故未列示.

表中哑变量 *Jump* 是核心解释变量, 当股票  $i$  在  $t$  月收益率出现向下跳跃时, 将  $Jump_{i,t}$  记为 1, 否则为 0. 面板 A 的模式在这里重现,  $t$  月收益率是否出现跳跃与  $t-1$  月的研报相似度并无显著关系(第 1 列); 而第 2 列中, 跳跃当月的 SimWu 系数则显著降低, 为 -0.006,  $t$  值为 -1.66, 在 10% 的置信水平下显著; 在信息冲击事件已经发生后的  $t+1$  月, 研报文本相似度则显著更低, 第 3 列 SimWu 的系数为 -0.015,  $t$  统计量为 -4.15, 在 1% 的置信水平上仍然显著. 研报文本相似度与负面信息冲击之间的关系之所以呈现这种时间上递进的模式, 是因为期望分析师能够提前预测收益率短期跳跃的要求过于严苛, 也不属于本文讨论的范畴. 本文关注的是在信息冲击来临后, 文本相似度能否捕捉分析师对该事件的解读. BNS 检验虽然能够判断某月收益率是否出现跳跃, 但不能精确定位跳跃日. 当收益率在月末出现跳跃时, 该月研报大部分在跳跃前发布. 据此, 随着时间的推移, 研报文本相似度在时间窗口期内应逐步降低, 即所含的信息含量逐步增加. 表 B2 的结论表明, 本文提出的文本相似度能够较好地表征研报中的增量信息. 在股票收益率出现向下跳跃后, 研报的文本信息显著增加, 而盈余预测和荐股评级等传统量化指标无显著变化, 或与预期方向相反, 不能有效、客观的代理研报中的信息.