

新闻媒体、资讯特征与资本市场信息效率

杨玉龙¹, 吴文², 高永靖³, 张倩男¹

(1. 浙江工商大学 财务与会计学院, 浙江 杭州 310018; 2. 清华大学 五道口金融学院, 北京 100083;
3. 上海财经大学 会计学院, 上海 200433)

摘要:我国媒体产业呈现双轨制特色, 产业中同时存在政策导向媒体与市场导向媒体。两类媒体的功能定位与激励机制迥然不同, 使得两者对资本市场的影响存在差异。文章发现, 在资本市场上, 媒体主要通过负面报道, 促进公司特质性信息反映在股价中, 而正面报道则无此作用。在区分政策导向媒体和市场导向媒体后, 文章进一步发现, 媒体对资本市场信息整合的影响主要体现在政策导向媒体上, 市场导向媒体则无显著影响。为了考察这种差异产生的原因, 文章对两类媒体在资讯报道方面的文本特征(信息挖掘与信息传递)进行了比较, 发现政策导向媒体相对于市场导向媒体进行了更深层次的信息挖掘与更精准的信息传递, 这是它对资本市场信息整合更具影响力的真正原因。

关键词: 政策导向媒体; 市场导向媒体; 股价同步性; 知情交易概率

中图分类号: F235.92 **文献标识码:** A **文章编号:** 1001-9952(2018)06-0109-17

DOI: 10.16538/j.cnki.jfe.2018.06.008

一、引言

我国的媒体最具特色之处是其双重性, 以及由此衍生的媒体产业的双轨制运营(李良荣, 2003)。改革开放以后, 我国经济开始逐步转型, 媒体产业结构也开始出现变化。特别是1992年之后, 媒体企业化经营取得突破, 部分媒体由政府拨款转向自负盈亏, 经济上逐渐独立(田中初, 2005)。同时, 媒体数量迅速增长, 管理机构开始对媒体产业采取“抓大放小”的分层管理策略。于是, 媒体出现分化, 一部分媒体继续维持传统运营模式, 资讯内容偏重在政策方面; 另一部分媒体如都市商业类报刊则逐步获得相对独立的编辑权, 并开始注重市场对于资讯的诉求(李良荣, 2003)。我们将前者称为“政策导向媒体”, 将后者称为“市场导向媒体”, 这便形成了我国媒体产业双轨制运营的特色。

通过对政策导向媒体和市场导向媒体正负面词汇使用的横向与纵向比较, 我们可以看到媒体双轨制运营对资讯内容的鲜明影响。图1绘制了巨灵财经数据库2000—2013年所收录的政策导向媒体与市场导向媒体(相关定义与分类见下文)的正负面报道比例情况。图中有三点值得注意: 一是媒体的正面报道比例远高于负面报道; 二是市场导向媒体的正面报道要少于政策导向

收稿日期: 2017-11-08

基金项目: 浙江省自然科学基金项目(LQ15G020002, LY17G020006); 国家自然科学基金项目(71032005, 71272013, 71502166, 71602112, 71702165); 教育部人文社科重大项目(15JJD630008)

作者简介: 杨玉龙(1986—)(通讯作者), 男, 山东枣庄人, 浙江工商大学财务与会计学院副教授;

吴文(1989—), 男, 山东单县人, 清华大学五道口金融学院博士后;

高永靖(1992—), 男, 河南兰考人, 上海财经大学会计学院博士研究生;

张倩男(1993—), 女, 安徽淮北人, 浙江工商大学财务与会计学院研究生。

媒体,而负面报道要多于政策导向媒体;三是从时间趋势来看,正面报道比例在逐步减少,负面报道比例则逐渐增加,而且这一趋势在市场导向媒体中更加明显。

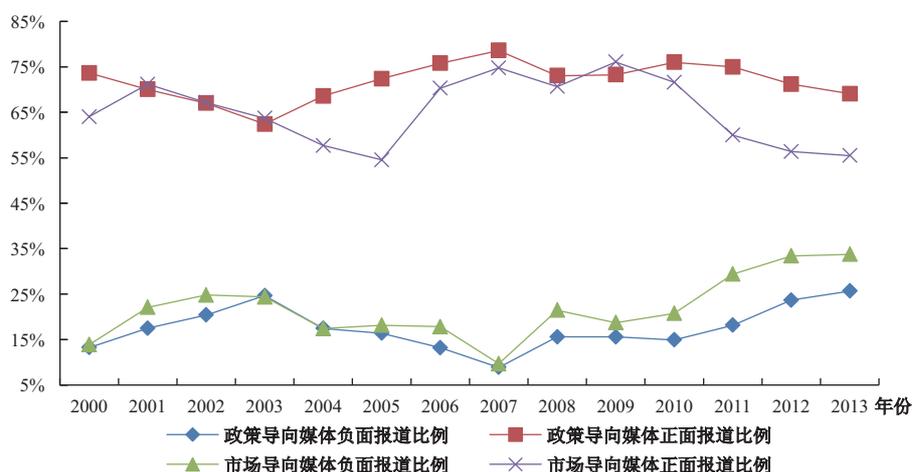


图1 2000—2013年政策导向媒体与市场导向媒体历年的正负面报道比例

资讯文本是信息的载体,不同类型媒体在资讯文本方面的差异必然导致在信息传递效果上有所不同。以往的文献从整体上探讨了媒体在优化公司治理、提升资本市场信息效率方面的作用,并给予了正面评价。更深入的研究还需要对媒体类型和资讯文本特征同时进行识别和分析,从而考察不同类型媒体的差异化表现。毕竟,对资本市场上的投资者而言,他们更感兴趣的问题是选取哪一类媒体、阅读哪一类资讯来支持其决策。换言之,他们需要识别对资本市场信息整合有实质性影响的媒体与资讯。因此,本文将从我国媒体的双轨制运营出发,探讨政策导向媒体与市场导向媒体在影响资本市场信息整合方面的差异,并分析导致这种差异的原因。

本文发现,媒体做出的负面报道能够促进特质性信息反映在个股股价中,正面报道反而不利于特质性信息整合。在将媒体区分为政策导向媒体和市场导向媒体后,本文发现媒体对资本市场的影响主要由政策导向媒体促成,市场导向媒体则无显著影响。随后,本文对媒体的资讯内容进行了文本分析,并计量了资讯内容的两项特征:信息传递和信息挖掘。分析结果表明,政策导向媒体的报道在这两项特征上显著不同于市场导向媒体。引入调节变量模型后,我们发现政策导向媒体主要凭借更深层次的信息挖掘与更精准的信息传递,更有力地在影响资本市场的信息整合。

本文可能存在以下两点贡献:第一,从我国媒体产业双轨制运营的特色入手,探讨了政策导向媒体与市场导向媒体在影响资本市场信息整合效率方面的差异。我们使用股价同步性和知情交易概率两个指标来衡量媒体信息整合效率,发现政策导向媒体可以更好地提升资本市场信息效率。另外,本文还发现媒体提升资本市场信息效率的关键在于信息挖掘的深入程度和信息传递的精准程度。这有助于监管部门评价和选择信息发布的媒体平台,并在资讯内容方面对媒体报道给予指引与建议。

第二,在技术方面,本文依托中国知网“中文信息结构库”提供的正负面词库,结合 Loughran 和 McDonald(2011)提供的英文正负面词汇清单,通过文本挖掘得到了适用于我国媒体报道的正负面词库。同时,我们使用 Python 编程语言,以巨灵财经报刊数据库为基础,构建了时间跨度超过十年(2000—2013年)的媒体正负面报道数据库。此外,我们还利用潜在语义主题分配模型(Latent Dirichlet Allocation,简称 LDA 主题分析)度量了不同类型媒体对上市公司公告跟踪报道

的信息挖掘深度和信息传递精度,并以此为基础解释了政策导向媒体和市场导向媒体在影响资本市场信息整合效率方面的差异。

二、理论分析与研究假说

(一)正负面报道与资本市场信息效率

公司股票价格的形成和变动有赖于三个层面的信息,即宏观信息、行业信息以及公司特质性信息(Piotroski 和 Roulstone, 2004)。其中,宏观信息和行业信息会引起资本市场上所有公司或特定行业所有公司的股价波动,只有与公司基本面相关的特质性信息才会引起单个公司股价产生独立于市场或行业趋势的特异性波动。宏观和行业信息所引发的市场上多只股票的同向价格波动导致了股市的“同涨同跌”现象,即股价同步性。因此,股价同步性可以作为个股股价中特质性信息含量的反向测度(Morck 等, 2000)。本文主要使用股价同步性作为资本市场信息效率的测度指标。之所以选取这个指标,主要基于以下两点考虑:第一,股价同步性指标可以测度个股交易中整合的公司特质性信息(Morck 等, 2000),这便于我们直接探讨哪种类型的媒体能够挖掘更多的公司特质性信息;第二,媒体跟踪上市公司时覆盖的时间跨度往往很长,能够涵括这一时间跨度影响的市场类指标以股价同步性为佳,因为它可以精确地截取研究所需的市场交易时段进行计算。

在资本市场中,不同类型信息的扩散模式存在差异。这种信息传播差异可能来自两个方面:一方面,读者有关注负面消息的倾向。相对于好消息,人们更倾向于传播坏消息,所谓“好事不出门,坏事行千里”。在演化过程中,人类继承了原始社会狩猎环境下的相关习性,出于生存需要对负面与危险信号极为关注(Liu 等, 2014),这一遗传习性直接影响了我们对于信息资讯的传递。个体在获知好消息后,平均而言会向 6 个人传递;而在获知坏消息后,平均而言会向 23 个人传递(赵亚东和高研, 2012)。Heath(1996)的实验研究也表明,在社会氛围普遍消极的情境下,人们更乐意传递坏消息。另一方面,个体有累积坏消息并打包释放的倾向。个体在面临多项损失或多项收益时,为了得到最大的效用,通常将损失合并在一起而将收益逐一分开(Kahneman 和 Tversky, 1979)。这导致上市公司管理层有动机暂时隐藏坏消息,而等待特定时机一起公布(Kothari 等, 2009)。

综上所述,当媒体大规模披露坏消息时,坏消息应当比好消息传递得更快,^①而且管理层可能暂时隐藏坏消息而等待时机一起释放,导致坏消息披露时包含的公司特质性信息更多。因此,媒体针对上市公司的负面报道比例越高,上市公司股价同步性越低。而正面信息在传递中则相对缓慢且分散,使得上市公司股价在整合正面资讯时不如负面资讯及时,股价更多地随宏观或行业层面信息波动,从而股价同步性较高。据此,本文提出以下假说:

假说 1: 对上市公司的报道中,负面报道比例越高,该上市公司股价同步性越低;正面报道比例越高,该上市公司股价同步性越高。

(二)媒体双轨制运营与资本市场信息整合

政策导向媒体与市场导向媒体的资金来源不同。政策导向媒体享有公费订阅的支持,并在税收上得到减免等优惠;市场导向媒体则被推向市场,自负盈亏,易为广告费用等商业利益所俘获,从而在针对上市公司的报道中表现出较弱的独立性。

^①Hong 等(2000)认为“Bad news travels slowly”,这看似与本文矛盾。但他们强调的是管理层会有意识地暂时隐藏公司的负面信息,使坏消息的释放与好消息相比会出现延迟。本文试图论证的是公司的坏消息被媒体挖掘(释放)出来后,在资本市场上传递的速度要快于好消息。这也与生活常识相一致,所谓“好事不出门,坏事传千里”。因此,本文的观点与其并不矛盾。

媒体的利益诉求会影响其报道行为,其中以商业利益最为典型。媒体挖掘和传播有关上市公司的私有信息可以促进资本市场的信息整合,但在挖掘出上市公司私有信息之后,并不必然通过发布报道来扩大报刊销量获益,也可通过威胁上市公司高管索取封口费来谋利,这便是“有偿沉默”。方军雄(2014)发现,如果公司在IPO期间向媒体支付费用以换取沉默,那么这类公司在IPO之后的盈余质量更差,也更可能出现业绩反转。这表明媒体的有偿沉默纵容了内部人的私有信息交易,牺牲其他投资者利益而谋取了私利。

与“有偿沉默”相对应的是“有偿报道”,既然媒体会屈从商业利益,自然有公司或机构会加以利用,使媒体发布对自身有利的信息。逯东等(2016)认为,机构投资者存在主动管理媒体的行为,即利用媒体制造信息噪音来引导市场热点的短期炒作行为;实证研究发现,为了获取超额收益,机构投资者会通过主动管理媒体来提高其重仓股票的媒体关注度。尽管这种“主动媒体管理”能在短期内推动股价上涨,但同时也会导致坏消息被藏匿,从而恶化资本市场的信息环境,增加未来的股价崩盘风险(逯东等,2016)。

媒体因商业利益而做出的有偏或虚假报道,会因投资者的羊群效应而将不良影响放大。媒体报道中充斥着大量所谓股评家的专业建议,普通投资者极易受其引导,争相购入特定股票。热点资讯总是稀缺资源,媒体通常会针对某一热点事件进行多次报道,而且不同媒体也会针对同一热点事件进行重复报道。这很可能会凸显本来微不足道的事件的重要性,对投资者预期产生正反馈作用,进而增强原有事件对投资者预期和决策的冲击。游家兴和吴静(2012)研究发现,股票价格对基本面的背离程度受媒体报道影响,特别地,媒体报道情绪越极端,股价对基本面的背离程度就越大。股价原本是公司价值的良好信号,但媒体的有偏或虚假报道会扭曲这一信号,加剧资本市场上的信息不对称。

政策导向媒体的收入来源主要是财政支持,因而更注重信息公平因素。^①市场导向媒体的收入来源主要是广告收入等市场业务,因而更容易被商业利益所俘获,其“有偿沉默”与“有偿报道”等违规行为与政策导向媒体相比要更为严重。此外,资讯发布有着严格的纪律要求,低级别媒体通常需要等待高级别媒体的指示或讯号。政策导向媒体在行政级别上往往高于市场导向媒体,意味着其报道不仅独立于商业利益,而且更加及时,这强化了其对资本市场的影响。因此,在资本市场信息整合方面,我国媒体的影响主要体现在政策导向媒体上,而较少体现在市场导向媒体上。据此,本文提出以下假说:

假说2:我国媒体对资本市场信息效率的影响主要体现在政策导向媒体上,而市场导向媒体则无显著影响。

三、研究设计

(一)数据来源与样本选取

1. 媒体类型与媒体报道数据。本文以巨灵财经报刊数据库为基础,该数据库涵括了2000—2013年的媒体报道。本文使用Python编程语言,对数据库提供的上市公司相关报道进行统计,得到媒体报道相关变量。通过对数据库所覆盖的137家媒体进行追根溯源,我们可以准确

^① 容易引发争议的是,有相反观点认为,政策导向媒体既然受财政支持,则有可能受到行政干预或管制,所报道的资讯更多的是宏观或行业层面信息,导致特质性信息更少。然而,这一观点在财经资讯领域并无可靠的证据支持。以《中国证券报》为例,它是典型的政策导向媒体,由于不可能每天都有政策变动或宏观事件发生,该报刊每天都有大量的个股信息。从大样本实证考察来看,中央媒体(主要由政策导向媒体构成)的确发掘了大量特质性信息(杨玉龙等,2016)。

地甄别和区分各个媒体,挖掘其主办单位及行政级别、注册地等信息,以便于探讨不同类型的媒体在资本市场信息整合方面的差异化表现。上市公司公告虽然借助媒体发布,但这是上市公司的主动披露行为,并不反映媒体的态度,不符合本文研究意图,故予以剔除。

2. 上市公司数据来源与样本选取。本文以 2003—2013 年我国 A 股上市公司为样本,并做了如下筛选:(1)剔除金融保险类公司;(2)剔除在公司治理和股权性质等变量存在数据缺失的公司。样本期之所以从 2003 年开始,是因为 CSMAR 数据库从这一年开始提供公司治理和股权性质数据。本文最终得到了 18 419 个样本。媒体的主办单位、行政级别、注册地信息等数据均手工搜集得到,其他数据来自 CSMAR 数据库和 Wind 数据库。本文对所有连续变量进行了上下 1% 的缩尾处理。

(二)主要变量定义。本文主要涉及三类变量:媒体类型与媒体报道相关变量、上市公司市场交易相关变量以及公司财务与治理相关变量。

1. 媒体类型。本文将媒体区分为政策导向媒体和市场导向媒体两种类型。我们在分类时主要依据媒体主办单位的行政级别,并兼顾该媒体所在城市影响力。本文对政策导向媒体和市场导向媒体的划分参考了杨玉龙等(2017)等文献。

2. 媒体正负面报道。我们参照汪昌云和武佳薇(2015)及张纯和吴明明(2015)提供的计量方案来识别媒体的正面报道与负面报道。具体而言,我们根据每篇媒体报道中正负面词汇数的相对大小来判断报道是正面的还是负面的:如果报道中的正面词汇数超过负面词汇数,则定义为正面报道;如果报道中的负面词汇数超过正面词汇数,则定义为负面报道;其他则视为中立报道。进一步地,对每一家公司的所有媒体报道进行描述性统计,得到该上市公司的负面报道比例和正面报道比例。在此基础上,我们进一步区分不同媒体类型,得到政策导向媒体的负面报道比例和正面报道比例,以及市场导向媒体的负面报道比例和正面报道比例。

3. 股价同步性。本文参考杨玉龙等(2016)的计量方法,使用实证 CAPM 模型的拟合系数 R^2 来衡量股价同步性:

$$R_{i,t} = \alpha_i + \beta_{m,i} R_{m,t} + \varepsilon_{i,t} \quad (1)$$

模型的拟合系数 R^2 越大,表明股价的波动包含越多的市场风险因素,即公司特质风险对股价的解释力越差。为了满足 OLS 的正态分布条件,我们对拟合系数 R^2 进行了对数变换,得到股价同步性指标 RSQ_m 。

$$RSQ_m = \text{Log}(R^2/(1 - R^2)) \quad (2)$$

本文变量定义见表 1。

表 1 变量定义

变量名称	变量符号	变量定义
因变量		
股价同步性	RSQ_m	见上文
解释变量		
媒体负面报道比例	Neg_News_Ratio	当年所有媒体对上市公司负面报道占报道总数的比例
媒体正面报道比例	Pos_News_Ratio	当年所有媒体对上市公司正面报道占报道总数的比例
政策导向媒体负面报道比例	$Politic_Neg_Ratio$	当年政策导向媒体对上市公司负面报道占报道总数的比例
政策导向媒体正面报道比例	$Politic_Pos_Ratio$	当年政策导向媒体对上市公司正面报道占报道总数的比例
市场导向媒体负面报道比例	$Market_Neg_Ratio$	当年市场导向媒体对上市公司负面报道占报道总数的比例
市场导向媒体正面报道比例	$Market_Pos_Ratio$	当年市场导向媒体对上市公司正面报道占报道总数的比例

续表 1 变量定义

变量名称	变量符号	变量定义
控制变量		
换手率	<i>Tvr</i>	个股周换手率的年均值
政治关联	<i>Pc</i>	虚拟变量,若公司董事长或总经理现任或曾任政府(或军队)官员、人大代表或政协委员,则取 1, 否则取 0
独立董事比例	<i>Indep_dir</i>	公司董事会中独立董事的比例
第一大股东持股	<i>Lshr</i>	第一大股东持股比例
股权制衡	<i>Othshr</i>	第二到第五大股东持股比例之和
机构持股比例	<i>Inst</i>	机构投资者持股比例之和
高管持股比例	<i>Excutive</i>	高管持股数额/公司流通股数
四大所	<i>Big4</i>	若审计师来自四大会计师事务所,则取 1, 否则取 0
操纵性应计项目	<i>Da</i>	当年操纵性应计数,采用修正的 <i>Jones</i> 模型,分行业、分年度测算
盈利相对波动	<i>Droe</i>	当年净资产收益率变动占所有上市公司净资产收益率变动均值的比例
个股相对波动	<i>Stdp</i>	个股回报扣除市场回报后的标准差
公司规模	<i>Size</i>	企业总资产的自然对数
销售增长率	<i>Sgrowth</i>	(当年营业总收入-上年营业总收入)/上年营业总收入
公司上市年限	<i>Age</i>	公司上市后运营年数

(三)模型设定。本文采用模型(3)至模型(5)来研究媒体的正负面报道对股价同步性的影响。模型(3)等式右侧前两项是关键解释变量负面报道比例(*Neg_News_Ratio*)和正面报道比例(*Pos_News_Ratio*),模型(4)等式右侧前两项分别为政策导向媒体负面报道比例(*Politic_Neg_Ratio*)与正面报道比例(*Politic_Pos_Ratio*),模型(5)等式右侧前两项分别为市场导向媒体负面报道比例(*Market_Neg_Ratio*)与正面报道比例(*Market_Pos_Ratio*)。

$Control_n$ 表示控制变量,包括操控性应计项目 *Da*、上市公司政治关联 *Pc*、独立董事比例 *Indep_dir*、机构持股比例 *Inst*、是否由四大审计 *Big4*、第一大股东持股 *Lshr*、股权制衡 *Othshr*、高管持股比例 *Excutive*、销售增长率 *Sgrowth*、盈利相对波动 *Droe*、个股股价波动 *Stdp*、个股换手率 *Tvr*、公司规模 *Size*、公司上市年限 *Age* 以及行业虚拟变量 *Ind* 和年度虚拟变量 *Year*。

$$RSQ_m = \alpha + \beta_1 Neg_News_Ratio + \beta_2 Pos_News_Ratio + \sum \gamma Control_n + \sum \delta Ind + \sum \eta Year + \varepsilon \quad (3)$$

$$RSQ_m = \alpha + \beta_1 Politic_Neg_Ratio + \beta_2 Politic_Pos_Ratio + \sum \gamma Control_n + \sum \delta Ind + \sum \eta Year + \varepsilon \quad (4)$$

$$RSQ_m = \alpha + \beta_1 Market_Neg_Ratio + \beta_2 Market_Pos_Ratio + \sum \gamma Control_n + \sum \delta Ind + \sum \eta Year + \varepsilon \quad (5)$$

四、实证结果分析

(一)描述性统计

1. 媒体类型与正负面报道。表 2 列示了我国媒体的正负面报道比例,以及每篇报道的正负面词汇使用情况。可以看到,无论是市场导向媒体还是政策导向媒体,正面报道比例远远大于负面报道比例(3 倍以上),而且每篇报道中使用的正面词汇平均数量也远远超过负面词汇(约为 2 倍)。此外,相对于政策导向媒体,市场导向媒体的负面报道比例显著较高,而正面报道比例则显著较低,而且每篇报道中使用的负面词汇数量也显著较多,正面词汇数量显著较少。因此,政策导向媒体与市场导向媒体因其功能定位差异,的确在资讯内容方面表现不同。

表 2 媒体类型与正负面报道

	市场导向媒体		政策导向媒体		均值差异	T 值
	观测数	均值	观测数	均值		
负面报道比例	186 811	0.190	337 435	0.132	0.0582***	56.226
正面报道比例	186 811	0.663	337 435	0.715	-0.0524***	-39.594
每篇报道中负面词汇数量	186 811	8.273	337 435	7.054	1.2182***	27.567
每篇报道中正面词汇数量	186 811	13.489	337 435	13.826	-0.3365***	-5.378
每篇报道中词汇总量	186 811	500.759	337 435	628.829	-128.070***	-36.268

注: **、*和*分别表示在 1%、5% 和 10% 的水平上显著, 下表同。

2. 变量描述性统计。从表 3 中可以看出, 无论是政策导向媒体还是市场导向媒体, 负面报道比例都小于正面报道比例, 说明媒体对不同上市公司也以正面报道为主。需要指出的是, 表 2 是针对媒体所有报道所做的统计分析, 表 3 则是针对上市公司的媒体报道所做的统计分析。

表 3 变量描述性统计

变量	样本数	均值	中位数	标准差	最小值	最大值
<i>RSQ_m</i>	18 419	-0.417	-0.381	0.712	-2.446	1.111
<i>Neg_News_Ratio</i>	18 419	0.213	0.143	0.249	0	1
<i>Pos_News_Ratio</i>	18 419	0.462	0.556	0.378	0	1
<i>Politic_Neg_Ratio</i>	18 419	0.112	0.021	0.160	0	0.725
<i>Politic_Pos_Ratio</i>	18 419	0.288	0.261	0.282	0	1
<i>Market_Neg_Ratio</i>	18 419	0.097	0	0.155	0	0.750
<i>Market_Pos_Ratio</i>	18 419	0.174	0.083	0.213	0	1
<i>Tvr</i>	18 419	0.011	0.008	0.009	0.001	0.050
<i>Pc</i>	18 419	0.405	0	0.491	0	1
<i>Indep_dir</i>	18 419	0.363	0.333	0.051	0.250	0.556
<i>Lshr</i>	18 419	0.169	0.090	0.181	0.001	0.658
<i>Othshr</i>	18 419	0.076	0.051	0.078	0.003	0.362
<i>Inst</i>	18 419	0.039	0.008	0.065	0	0.310
<i>Excutive</i>	18 419	0.041	0	0.117	0	0.563
<i>Big4</i>	18 419	0.194	0	0.395	0	1
<i>Da</i>	18 419	-0.002	-0.004	0.083	-0.252	0.282
<i>Droe</i>	18 419	0.096	0.100	0.238	-1.113	0.685
<i>Stdp</i>	18 419	0.031	0.028	0.010	0.015	0.058
<i>Size</i>	18 419	21.683	21.559	1.189	19.115	25.356
<i>Sgrowth</i>	18 419	0.173	0.108	0.512	-0.812	3.573
<i>Age</i>	18 419	9.257	9	4.282	3	23

(二) 回归分析

1. 正负面报道与资本市场信息效率。本文采用模型(3)至模型(5)来研究媒体的正负面报道对股价同步性的影响, 以普通最小二乘法(OLS)进行估计。从表 4 中可以看到, 在对上市公司的报道中, 负面报道比例越高, 股价同步性越低, 但不显著; 正面报道比例越高, 股价同步性越高。这一结果部分支持了假说 1, 我们将在下文考察不同类型媒体报道对股价同步性的影响。此外, 公司信息不够透明(*Da* 较高)或者具有政治关联, 均会导致较高的股价同步性。

2. 媒体双轨制与资本市场信息整合。我们进一步将媒体区分为政策导向媒体和市场导向媒体,以便考察我国媒体的双轨制运营对资本市场信息整合功能的影响。从表4中可以看到,促成资本市场信息整合的媒体报道主要是政策导向媒体的负面报道,市场导向媒体的负面报道并没有显著传递有关上市公司的特质性信息。值得注意的是,市场导向媒体的正面报道也会提高上市公司的股价同步性。本文认为,市场导向媒体的正面资讯进入股价的速度缓慢,而且在报道方面通常紧跟政策导向媒体,造成信息冗余重复而特质性信息不足,这也是其导致股价同步性上升的重要原因。

表4 媒体正负面报道与股价同步性

	全部媒体			政策导向媒体			市场导向媒体	
	系数	T值		系数	T值		系数	T值
<i>Intercept</i>	1.3351***	11.247	<i>Intercept</i>	1.292***	11.187	<i>Intercept</i>	1.2009***	10.354
<i>Neg_News_Ratio</i>	-0.009	-0.412	<i>Politic_Neg_Ratio</i>	-0.069**	-2.496	<i>Market_Neg_Ratio</i>	-0.025	-0.81
<i>Pos_News_Ratio</i>	0.1762***	12.806	<i>Politic_Pos_Ratio</i>	0.219***	12.46	<i>Market_Pos_Ratio</i>	0.1299***	5.2292
<i>Da</i>	0.0675	1.498	<i>Da</i>	0.1655***	3.3989	<i>Da</i>	0.1749***	3.5786
<i>Pc</i>	0.0329***	4.1505	<i>Pc</i>	0.0177**	2.0733	<i>Pc</i>	0.0203**	2.3742
<i>Indep_dir</i>	-0.392***	-4.877	<i>Indep_dir</i>	-0.407***	-5.055	<i>Indep_dir</i>	-0.399***	-4.938
<i>Inst</i>	-0.993***	-12.44	<i>Inst</i>	-0.935***	-11.72	<i>Inst</i>	-0.966***	-12.01
<i>Big4</i>	-0.046***	-4.418	<i>Big4</i>	-0.045***	-4.333	<i>Big4</i>	-0.044***	-4.212
<i>Lshr</i>	-0.09***	-2.994	<i>Lshr</i>	-0.09***	-2.992	<i>Lshr</i>	-0.093***	-3.052
<i>Othshr</i>	-0.514***	-8.636	<i>Othshr</i>	-0.51***	-8.56	<i>Othshr</i>	-0.529***	-8.838
<i>Excutive</i>	0.0617	1.317	<i>Excutive</i>	0.0645	1.3764	<i>Excutive</i>	0.068	1.4463
<i>Sgrowth</i>	-0.063***	-7.762	<i>Sgrowth</i>	-0.064***	-7.815	<i>Sgrowth</i>	-0.064***	-7.785
<i>Droe</i>	0.2765***	14.131	<i>Droe</i>	0.2897***	14.891	<i>Droe</i>	0.3169***	16.337
<i>Stdp</i>	-12.99***	-15.32	<i>Stdp</i>	-12.65***	-14.92	<i>Stdp</i>	-12.72***	-14.92
<i>Tvr</i>	-18.3***	-27.91	<i>Tvr</i>	-18.12***	-27.61	<i>Tvr</i>	-18.29***	-27.76
<i>Size</i>	0.0177***	4.2426	<i>Size</i>	0.0199***	4.7954	<i>Size</i>	0.0244***	5.8644
<i>Age</i>	-0.016***	-12.28	<i>Age</i>	-0.016***	-12.49	<i>Age</i>	-0.016***	-12.67
<i>Year</i>	控制		<i>Year</i>	控制		<i>Year</i>	控制	
<i>Ind</i>	控制		<i>Ind</i>	控制		<i>Ind</i>	控制	
<i>F</i> 值	297.04	<i>Pr</i> <0.001	<i>F</i> 值	290.96	<i>Pr</i> <0.001	<i>F</i> 值	294.9	<i>Pr</i> <0.001
样本数	18 419		样本数	18 419		样本数	18 419	
<i>Adj. R</i> ²	0.4211		<i>Adj. R</i> ²	0.4161		<i>Adj. R</i> ²	0.4194	

注:回归分析考虑了异方差,并做了公司层面的 cluster 调整,下表同。

五、稳健性测试与拓展性分析

(一)稳健性测试。以股价同步性来反映股价对公司特质性信息的整合是有争议的。West(1988)研究发现,基于噪声信息的交易降低了公司的股价同步性。Kelly(2014)认为,公司股票的 R^2 越低,其信息环境越差,表明 R^2 并非度量信息效率的有效指标。为解决这一问题,我们将使用直接反映特质性信息整合的指标知情交易概率进行稳健性测试。

知情交易概率(*Possibility of Informed Trading, PIN*)是指个股交易中知情交易的比重。其基本计量思路是将给定区间内针对个股的交易分为两类:一类是由私有信息促成的知情交易;另一类是由随机扰动因素驱动的噪声交易。促使股价向真实价值回归的动力主要来自知情交易,因

此个股交易中知情交易的比重越高, 股价对于信息的整合越有效。我们沿用已有研究中估计 PIN 指数的经典方法 (Easley 和 O'hara, 1987, 1992), 计量方法详见朱伟骅 (2008) 及孔东民等 (2013)。

表 5 列示了 PIN 对媒体正负面报道的回归结果。可以看到, 负面报道可促使知情交易比例上升, 而正面报道则会减少知情交易, 这与上文所论证的“负面报道促进特质性信息整合, 正面报道不利于特质性信息整合”相一致。同时, 我们可以看到媒体报道对知情交易产生影响主要体现在政策导向媒体上, 市场导向媒体对知情交易概率则无显著影响。

表 5 媒体正负面报道与知情交易概率

	全部媒体			政策导向媒体			市场导向媒体	
	系数	T 值		系数	T 值		系数	T 值
<i>Intercept</i>	0.4099***	26.933	<i>Intercept</i>	0.4114***	27.156	<i>Intercept</i>	0.407***	29.368
<i>Neg_News_Ratio</i>	0.0022***	2.924	<i>Politic_Neg_Ratio</i>	0.003***	3.3712	<i>Market_Neg_Ratio</i>	0.0008	1.0122
<i>Pos_News_Ratio</i>	-0.004***	-6.433	<i>Politic_Pos_Ratio</i>	-0.004***	-6.544	<i>Market_Pos_Ratio</i>	-0.004***	-5.807
<i>Da</i>	-0.003	-0.516	<i>Da</i>	-0.003	-0.506	<i>Da</i>	-0.003	-0.55
<i>Pc</i>	-0.002*	-1.886	<i>Pc</i>	-0.002*	-1.912	<i>Pc</i>	-0.002**	-2.056
<i>Indep_dir</i>	-0.004	-0.357	<i>Indep_dir</i>	-0.003	-0.317	<i>Indep_dir</i>	-0.004	-0.399
<i>Inst</i>	0.1059***	10.474	<i>Inst</i>	0.1045***	10.36	<i>Inst</i>	0.1056***	10.386
<i>Big4</i>	0.004***	3.0572	<i>Big4</i>	0.004***	3.0431	<i>Big4</i>	0.0039***	2.9648
<i>Lshr</i>	0.0177***	4.6563	<i>Lshr</i>	0.0177***	4.6564	<i>Lshr</i>	0.0179***	4.7067
<i>Othshr</i>	0.0281***	3.7438	<i>Othshr</i>	0.0278***	3.709	<i>Othshr</i>	0.0283***	3.7695
<i>Excutive</i>	-0.006	-1.083	<i>Excutive</i>	-0.006	-1.095	<i>Excutive</i>	-0.006	-1.092
<i>Sgrowth</i>	0.0032***	3.0711	<i>Sgrowth</i>	0.0032***	3.0995	<i>Sgrowth</i>	0.0032***	3.1234
<i>Droe</i>	-0.009***	-3.734	<i>Droe</i>	-0.009***	-3.861	<i>Droe</i>	-0.01***	-4.157
<i>Stdp</i>	-0.556***	-5.176	<i>Stdp</i>	-0.567***	-5.282	<i>Stdp</i>	-0.577***	-5.355
<i>Tvr</i>	2.281***	27.573	<i>Tvr</i>	2.2746***	27.51	<i>Tvr</i>	2.2793***	27.499
<i>Size</i>	-0.012***	-21.1	<i>Size</i>	-0.012***	-21.22	<i>Size</i>	-0.012***	-21.65
<i>Age</i>	-0.0045**	-2.504	<i>Age</i>	-0.0041**	-2.488	<i>Age</i>	-0.0039**	-2.414
<i>Year</i>	控制		<i>Year</i>	控制		<i>Year</i>	控制	
<i>Ind</i>	控制		<i>Ind</i>	控制		<i>Ind</i>	控制	
<i>F 值</i>	88.694	<i>Pr</i> <0.001	<i>F 值</i>	88.655	<i>Pr</i> <0.001	<i>F 值</i>	88.274	<i>Pr</i> <0.001
样本数	18 419		样本数	18 419		样本数	18 419	
<i>Adj. R</i> ²	0.1785		<i>Adj. R</i> ²	0.1784		<i>Adj. R</i> ²	0.1778	

(二)拓展性分析。本文的主要实证结果表明, 在促进资本市场信息整合方面, 政策导向媒体显示出了强于市场导向媒体的影响力。我们将从资讯特征的角度, 探讨造成这一差异的原因。本文主要考察两项资讯特征: 信息挖掘和信息传递。所谓信息挖掘, 是指媒体通过公开信息渠道或者调查暗访等手段搜集各类信息, 并进行分析比对, 形成独立的观点和结论。所谓信息传递, 是指媒体针对上市公司已经发布的信息, 通过对语句进行重新组织、概括、演绎等, 以投资者更容易理解的方式再次呈现出来。

以文本分析来考察媒体的资讯特征, 目前只有很少的研究关注了媒体报道更深层次的信息角色。Drake 等 (2014) 将媒体对上市公司盈余公告的跟踪报道区分为被上市公司雇用以传播信息的“信息传递”类报道 (*Presswire Services*) 和带有作者“信息创造”的报道 (*Newswire Services*), 发现媒体对上市公司年度盈余公告的报道能够减少现金流误定价, 但不能降低应计误定价, 且

这种降低会计误定价的作用主要是由媒体“信息传递”而非“信息创造”功能实现的。他们对于信息传递和信息挖掘的度量,得益于特殊的数据库(News Wire),很难被复制。

在分析师跟踪的相关研究中,有文献探讨了分析师报告的不同“信息角色”。这类研究主要通过市场反应来度量分析师报告的“信息解释”与“信息挖掘”两类信息角色。Huang等(2014)认为用市场反应来间接度量分析师报告的“信息角色”有两点局限:一是以市场反应来度量信息挖掘和信息解释,其前提条件是投资者能够充分理解分析师报告中的信息,且信息能迅速融在价格中,但现有研究表明这一前提条件并不成立;二是上市公司披露信息和分析师跟踪发布报告在时间上比较接近,学者难以区分资本市场反应是针对上市公司信息披露还是分析师跟踪报告。因此,他们主张用潜在语义主题分配模型(Latent Dirichlet Allocation,简称LDA主题分析)来直接度量分析师报告的信息挖掘和信息解释程度。本文也使用LDA主题分析来直接度量我国媒体报道的“信息角色”,并区分政策导向媒体和市场导向媒体,探讨两者之间的差异。

1. 信息挖掘与信息传递的度量。以LDA模型来度量媒体信息挖掘与信息传递的基本思想是,比较上市公司公告与随后的媒体跟踪报道之间的差异,根据这一差异来衡量分析师报告信息挖掘或传递的程度。附录A描述了我们对公司公告与随后的媒体跟踪报道进行匹配的具体过程。为避免出现噪音,本文仅选择公司当天有且仅有一条信息公告的观测值,从而可以对上市公司公告与媒体报道进行精确匹配。这会使样本量略有变化,我们最终得到18108个公司年份观测值。

本文参考Huang等(2014)及张纯和吴明明(2015)的方法,以公司公告和媒体报道的文本内容为对象进行LDA主题建模。这一建模的基本思路是,所有公告或媒体报道的内容尽管千差万别,但是主题(话题)却很有限,无非是产业支持政策、信息披露、股权发行、风险和披露、债务和担保、法律诉讼等,我们发现这些公告和资讯大致可以归结为40个有意义的主题(建模过程详见附录B)。而后我们可以测算两项内容:一是每篇上市公司公告与随后的跟踪报道所涵括的主题数量;二是每篇上市公司公告与随后的跟踪报道在描述同一主题时的用词分布。

在此基础上,我们采用媒体跟踪报道与对应的上市公司公告之间的主题分布差异来度量媒体跟踪报道的信息挖掘程度。这样处理背后的逻辑是,如果媒体对上市公司公告进行了深入分析,则其报道的隐含主题集合与公司公告的隐含主题集合之间会存在差异。因此,公司公告和媒体报道的主题分布差异越大,则媒体对公司信息进行了越深入的解读和挖掘。

同时,我们统计了媒体报道与公司公告在描述同一主题时的用词差异,反映媒体的信息传递水平。这样处理背后的逻辑是,如果媒体对上市公司公告进行了有效的信息传递,那么其报道的用词和表述应当与上市公司公告高度一致。因此,在描述同一主题时的用词分布差异越小,则信息传递的扭曲越小。

媒体的信息挖掘水平和信息传递水平分别记为Discover和Dissemination,本文对这两个变量的计量主要参考了Huang等(2014)及张纯和吴明明等(2015)的研究,略有调整,详见附录C。

2. 媒体类型与资讯特征。表6列示了政策导向媒体和市场导向媒体在资讯特征上的差异化表现。可以看到,政策导向媒体在信息挖掘方面的表现要显著优于市场导向媒体(0.387>0.341);

表6 媒体类型与资讯特征

资讯特征	市场导向媒体		政策导向媒体		均值差异	T值
	观测数	均值	观测数	均值		
信息挖掘(Discover)	5707	0.341	12401	0.387	0.0464***	12.108
信息传递(Dissemination)	5707	0.269	12401	0.325	0.0560***	18.168

同时,政策导向媒体的信息传递水平平均值为0.325,大于市场导向媒体的信息传递水平平均值0.269。可见,政策导向媒体在信息挖掘和信息传递方面显著优于市场导向媒体,这与上文所述政策导向媒体具有更强的经济独立性与及时性相一致。

资讯特征必然会对媒体报道的效果产生影响,本文使用调节效应模型来探讨这种影响。我们仍以股价同步性来衡量资本市场信息效率,区分政策导向媒体和市场导向媒体后,在模型(4)和模型(5)的基础上,加入不同类型媒体的正面报道比例、负面报道比例与资讯特征的交叉项,构建了模型(6)至模型(9)。

$$RSQ_m = \alpha + \beta_1 Market_Neg_Ratio \times Discover + \beta_2 Market_Neg_Ratio + \beta_3 Market_Pos_Ratio \times Discover + \beta_4 Market_Pos_Ratio + \beta_5 Discover + \sum \gamma Control_n + \sum \delta Ind + \sum \eta Year + \varepsilon \quad (6)$$

$$RSQ_m = \alpha + \beta_1 Politic_Neg_Ratio \times Discover + \beta_2 Politic_Neg_Ratio + \beta_3 Politic_Pos_Ratio \times Discover + \beta_4 Politic_Pos_Ratio + \beta_5 Discover + \sum \gamma Control_n + \sum \delta Ind + \sum \eta Year + \varepsilon \quad (7)$$

$$RSQ_m = \alpha + \beta_1 Market_Neg_Ratio \times Dissemination + \beta_2 Market_Neg_Ratio + \beta_3 Market_Pos_Ratio \times Dissemination + \beta_4 Market_Pos_Ratio + \beta_5 Dissemination + \sum \gamma Control_n + \sum \delta Ind + \sum \eta Year + \varepsilon \quad (8)$$

$$RSQ_m = \alpha + \beta_1 Politic_Neg_Ratio \times Dissemination + \beta_2 Politic_Neg_Ratio + \beta_3 Politic_Pos_Ratio \times Dissemination + \beta_4 Politic_Pos_Ratio + \beta_5 Dissemination + \sum \gamma Control_n + \sum \delta Ind + \sum \eta Year + \varepsilon \quad (9)$$

表7列示了回归结果。可以看到,政策导向媒体通过更深入的信息挖掘与更精准的信息传递,不仅可以强化负面报道的信息整合作用,还可以削弱正面报道对资本市场信息整合的不利影响。唯一的例外是,政策导向媒体对负面报道的信息传递在进一步降低股价同步性方面不显著。市场导向媒体的信息挖掘与信息传递则基本上没有展示出增强媒体报道效力的作用。基于以上证据,我们认为政策导向媒体在信息挖掘方面更加深入,在信息传递方面更加精准,使其整合资本市场信息的作用强于市场导向媒体。

表7 资讯特征、媒体报道与股价同步性

Panel A: 信息挖掘的调节作用					
	市场导向媒体			政策导向媒体	
	系数	T值		系数	T值
<i>Intercept</i>	0.6972***	5.6558	<i>Intercept</i>	-0.143	-0.852
<i>Market_Neg_Ratio</i> × <i>Discover</i>	0.0485	0.3284	<i>Politic_Neg_Ratio</i> × <i>Discover</i>	-0.771***	-3.657
<i>Market_Neg_Ratio</i>	-0.172**	-2.534	<i>Politic_Neg_Ratio</i>	-0.0859	-0.8276
<i>Market_Pos_Ratio</i> × <i>Discover</i>	0.1985*	1.8232	<i>Politic_Pos_Ratio</i> × <i>Discover</i>	-0.304*	-1.753
<i>Market_Pos_Ratio</i>	0.162***	3.2597	<i>Politic_Pos_Ratio</i>	0.0694	0.7705
<i>Discover</i>	-0.161***	-2.587	<i>Discover</i>	-0.1261**	-2.0095
其他控制变量	控制		其他控制变量	控制	
年份	控制		年份	控制	
行业	控制		行业	控制	
F值	97.834	Pr<0.001	F值	188.75	Pr<0.001
样本数	5 707		样本数	12 401	
Adj. R ²	0.4483		Adj. R ²	0.4184	

续表7 资讯特征、媒体报道与股价同步性

Panel B: 信息传递的调节作用					
	市场导向媒体			政策导向媒体	
	系数	T值		系数	T值
<i>Intercept</i>	0.6552***	5.3375	<i>Intercept</i>	-0.131	-0.774
<i>Market_Neg_Ratio</i> × <i>Dissemination</i>	-0.09	-0.489	<i>Politic_Neg_Ratio</i> × <i>Dissemination</i>	-0.315	-1.21
<i>Market_Neg_Ratio</i>	-0.125*	-1.842	<i>Politic_Neg_Ratio</i>	-0.082	-0.726
<i>Market_Pos_Ratio</i> × <i>Dissemination</i>	-0.024	-0.173	<i>Politic_Pos_Ratio</i> × <i>Dissemination</i>	-0.366*	-1.663
<i>Market_Pos_Ratio</i>	0.2304***	4.6714	<i>Politic_Pos_Ratio</i>	0.0651	0.7085
<i>Dissemination</i>	-0.103	-1.33	<i>Dissemination</i>	0.0489	0.6128
其他控制变量	控制		其他控制变量	控制	
年份	控制		年份	控制	
行业	控制		行业	控制	
F值	97.68	<i>P</i> <0.001	F值	189.28	<i>P</i> <0.001
样本数	5 707		样本数	12 401	
Adj. R ²	0.4479		Adj. R ²	0.4180	

六、结 论

改革开放以后,我国媒体产业发展迅速,媒体形式也逐渐多样化。然而,媒体数量上的增长带来了诸多问题,如经济利益对媒体独立性的绑架、媒体从业者水平有待提高等,使得读者在选取资讯来源时犹豫不决,并对资讯内容产生了诸多疑问。因此,需要科学的量化研究来探讨媒体的资讯特征,为读者遴选媒体与资讯提供指引。资本市场为我们理解我国媒体的影响力提供了良好的量化场景,尽管这一量化仅限于金融资源配置这一狭小领域,但是仍可以为我们理解媒体的社会功能提供参考。

本文从我国媒体产业的双轨制特色入手,借助文本分析技术,考察了政策导向媒体和市场导向媒体的正负面报道影响资本市场信息整合的效力。本文发现,媒体的负面报道对资本市场信息整合有促进作用,而正面报道则不利于特质性信息反映在股价中,这意味着负面报道对投资者而言更具决策价值。特别地,媒体报道对资本市场的影响主要集中在政策导向媒体上,这进一步缩小了投资者决策时的媒体参考范围。本文进一步探察了政策导向媒体与市场导向媒体出现上述差异的原因。本文利用LDA主题分析模型,测量了两者在信息挖掘深度与信息传递精度上的表现差异。本文发现,政策导向媒体在信息挖掘方面更加深入,在信息传递方面更加精准,使其整合资本市场信息的功能强于市场导向媒体。这一研究结果具有重要的现实意义,有助于投资者重新审视不同的资讯来源,并给予其一定的启发与指引。

参考文献:

- [1]方军雄. 信息公开、治理环境与媒体异化——基于IPO有偿沉默的初步发现[J]. 管理世界, 2014, (11): 95-104.
- [2]孔东民, 刘莎莎, 应千伟. 公司行为中的媒体角色: 激浊扬清还是推波助澜?[J]. 管理世界, 2013, (7): 145-162.
- [3]李良荣. 论中国新闻媒体的双轨制——再论中国新闻媒体的双重性[J]. 现代传播, 2003, (4): 1-4.
- [4]逯东, 付鹏, 杨丹. 机构投资者会主动管理媒体报道吗?[J]. 财经研究, 2016, (2): 73-84.
- [5]田中初. 新闻实践与政治控制[M]. 济南: 山东人民出版社, 2005.
- [6]汪昌云, 武佳薇. 媒体语气、投资者情绪与IPO定价[J]. 金融研究, 2015, (9): 174-189.

- [7]杨玉龙,孙淑伟,孔祥. 媒体报道能否弥合资本市场上的信息鸿沟?——基于社会关系网络视角的实证考察[J]. 管理世界, 2017, (7): 99–119.
- [8]杨玉龙,吴明明,王璟,等. 异质性媒体与资本市场信息效率[J]. 财经研究, 2016, (3): 83–94.
- [9]游家兴,吴静. 沉默的螺旋: 媒体情绪与资产误定价[J]. 经济研究, 2012, (7): 141–152.
- [10]张纯,吴明明. 媒体在资本市场中的角色: 信息解释还是信息挖掘?[J]. 财经研究, 2015, (12): 72–83.
- [11]赵亚东,高岩. 公安网络舆论的生成机制及其引导功能[J]. 中国人民公安大学学报(社会科学版), 2012, (2): 136–141.
- [12]朱伟骅. 公司治理与内幕交易监管效率研究[J]. 经济学(季刊), 2008, (1): 271–288.
- [13]Drake M S, Guest N M, Twedt B J. The media and mispricing: The role of the business press in the pricing of accounting information[J]. The Accounting Review, 2014, 89(5): 1673–1701.
- [14]Easley D, O'hara M. Price, trade size, and information in securities markets[J]. Journal of Financial Economics, 1987, 19(1): 69–90.
- [15]Easley D, O'hara M. Time and the process of security price adjustment[J]. The Journal of Finance, 1992, 47(2): 577–605.
- [16]Heath C. Do people prefer to pass along good or bad news? Valence and relevance of news as predictors of transmission propensity[J]. Organizational Behavior and Human Decision Processes, 1996, 68(2): 79–94.
- [17]Hong H, Lim T, Stein J C. Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies[J]. The Journal of Finance, 2000, 55(1): 265–295.
- [18]Huang A, Lehavy R, Zang A Y, et al. A thematic analysis of analyst information discovery and information interpretation roles[R]. Working Paper No. 1229, 2014.
- [19]Kahneman D, Tversky A. Prospect theory: An analysis of decision under risk[J]. Econometrica, 1979, 47(2): 263–292.
- [20]Kelly P J. Information efficiency and firm-specific return variation[J]. Quarterly Journal of Finance, 2014, 4(4): 1–44.
- [21]Kothari S P, Shu S, Wysocki P D. Do managers withhold bad news?[J]. Journal of Accounting Research, 2009, 47(1): 241–276.
- [22]Liu G, Xin Z, Lin C. Lax decision criteria lead to negativity bias: Evidence from the emotional stroop task[J]. Psychological Reports, 2014, 114(3): 896–912.
- [23]Loughran T, McDonald B. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks[J]. The Journal of Finance, 2011, 66(1): 35–65.
- [24]Morck R, Yeung B, Yu W. The information content of stock markets: Why do emerging markets have synchronous stock price movements?[J]. Journal of Financial Economics, 2000, 58(1-2): 215–260.
- [25]Piotroski J D, Roulstone D T. The influence of analysts, institutional investors, and insiders on the incorporation of market, industry, and firm-specific information into stock prices[J]. The Accounting Review, 2004, 79(4): 1119–1151.
- [26]West K D. Bubbles, fads and stock price volatility tests: A partial evaluation[J]. The Journal of Finance, 1988, 43(3): 639–656.

附录 A 媒体跟踪报道和上市公司信息披露公告的匹配过程

1. 对于上市公司披露的财务报表, 由于主要内容是表格和数字, 不适用 *LDA* 主题模型进行文本分析, 本文剔除了这类公告。

2. 由于媒体跟踪上市公司公告一般会在新闻内容中出现“**(公司)**(时间)公告称”之类的用语, 本文在新闻全文中首先定位找到“公告”这一关键词, 将其之前直到上一个句终标点(句号、叹号、问号等)或新闻全文字符串的首个字符截取下来, 然后删除其中的空格, 进行时间关键

词扫描。之所以不包括逗号、顿号等非句终标点,是因为存在以下形式的表述:“**(公司)**(时间)进行了**披露,公告称……”。时间关键词包括:

(1)“昨”,代表昨天,可以是“昨天”、“昨日”、“昨晚”等。

(2)“前日”“前天”,代表2天前。

(3)“今日”“今天”“今晚”“今早”,代表当天。

(4)具体的年月日,对于“年”“月”“日”之前的数字,可以是阿拉伯数字和中文数字及相应的全角数字。

(5)如果没有出现以上四类时间关键词,但是出现了“晚间”“上午”“下午”“早间”“一早”“晚上”“傍晚”,则代表当天。

附录 B 对上市公司公告和媒体跟踪报道文本集进行 LDA 主题建模过程

1. LDA 主题模型

LDA 主题模型是在自然语言信息处理中提取大规模语料集的主题结构任务的一类机器学习模型。它可以从文本集中估算提取出主题信息,并把每个文本表示为关于主题的多项式分布,而每个主题被表示为关于语料集中所有不同词语的多项式分布。因此,学者可以利用 LDA 主题模型的输出结果,对不同文本的差异进行比较。

LDA 主题模型用统计过程来模仿文本的生成过程。该算法假设所有语料集的所有文档共享同一组主题,但是主题在不同文档间的分布不同。每个文档被建模为关于主题的概率分布,而每个主题被建模为关于词语的概率分布。该算法假定文档中的每个词语通过以下两个步骤生成:第一步,文档的作者从所有可用的主题分布中选择一个主题;第二步,对于选定的主题,文档的作者从表现该主题的词语分布中选择一个词语。重复这两个步骤,逐词生成整个文档。每篇文档的生成过程都是一样的且相互独立。

在上述假定的基础上,LDA 主题模型估算出最匹配(似然度最大)文本数据的模型参数。参数估算过程依赖于词语在同一篇文档中的出现情况。直观地讲,如果两个词频繁在同一篇文档中出现,那么 LDA 模型把这两个词归类为同一个主题的概率就会较高。LDA 模型的主要输出结果是每个主题的词频分布矩阵(用 *Topic_Word* 表示),这也可以作为 LDA 的建模结果;另外,LDA 模型也会同时输出训练集中每个文档关于主题的频率分布矩阵(用 *Doc_Topic* 表示)。

2. LDA 建模前的文本数据预处理

与一般的中文文本分析准备步骤一样,在 LDA 主题分析之前,首先要对全文进行中文分词。我们采用 NLPPIR 中文分词系统,将每篇上市公司公告和财经媒体跟踪报道分拆成词语列表。然后,为了防止对主题建模的干扰,我们剔除了分词结果中的非中文词语(包括数字、英文等),并根据《哈工大停用词表》剔除了中文停用词(*Stop Words*),如功能词“还有”、“即便”等。最后,中文分词系统会把媒体报道中大量存在的人名、公司名等非专有名词切割成出现频率较高的单个汉字,如“嘉”、“宏”等,这会对主题建模产生干扰,所以我们进一步剔除了分词结果中以单个汉字形式呈现的中文词。

3. 决定主题个数

LDA 主题分析的算法需要人工输入主题个数,而主题个数的选择会影响结果解读。设定太少的主题数会导致最后的主题区分过于宽泛和模糊,而设定太多的主题数又会产生一些没有经济意义的主题。基于 Huang 等(2014)及张纯和吴明明(2015)的研究,我们使用“复杂度”(Perplexity)统计量来确定 LDA 模型的主题数,最后选取 40 作为最优的主题数。

4. LDA 建模结果

由于不同行业的上市公司公告和媒体跟踪报道内容会涉及含有各自行业特点的主题, 我们按照证监会二级行业代码, 剔除“金融、保险业”, 对余下的 21 个行业进行了 LDA 主题分析。表 8 列示了样本中文档数最多的制造业中“机械、设备、仪表”(C7)行业的 LDA 建模结果。其中, “主题的词语分布中前 20 个词语”为 LDA 建模后输出结果, 而“人工拟定主题名称”为我们根据该主题的词语分布中前 20 个词语人工判定的主题语义。从前 10 个主题的前 20 个词语可以看出, LDA 较好地概括了上市公司公告和财经媒体跟踪报道经常涉及的主题, 其中第二个主题“产业政策对公司的影响”、第四个主题“新项目机会”和第八个主题“出口、海外市场”体现了一定的行业特点, 另外六个主题则是上市公司一般都会涉及的主题。

表 8 样本中文档数最多的“机械、设备、仪表”行业 LDA 建模结果 (前 10 个主题的前 20 个词语)

人工拟定主题名称	主题的词语分布中前 20 个词语
对上市公司消息的市场反应	市场、投资者、行业、增发、问题、亏损、资金、收购、价格、资本、获得、计划、值得、投资、选择、未来、这种、实现、公布、超过
产业政策对公司的影响	机床、收到、资金、补贴、数控、政府、财政、国家、政策、获得、补助、所得税、产业、科技、通知、技术、高新技术、项目、发展、重大
信息披露	披露、信息、媒体、投资者、报道、澄清、交易、风险、重大、传闻、交易所、波动、刊登、指定、价格、深圳、控股、广大、异常、规则
新项目机会	中标、项目、招标、电力、金额、合同、采购、硬盘、国家电网公司、电缆、设备、工程、通知书、变压器、经营、电网、营业、电能、数量、签订
股权发行	基金、投资、管理、责任、配售、资产、信托、计划、发行、账户、申购、债券、资格、国泰、保险、资金、中信、自营、君安、产品
资产转让	转让、资产、评估、交易、持有、关联、净资产、轻骑、价格、经营、审计、价值、债务、负债、出售、完成、协议、总额、产权、全部
风险和披露	产品、问题、管理、会计、资金、披露、理财、风险、财务、要求、措施、发生、生产、信息、使用、制度、整改、损失、工作、处理
出口、海外市场	美国、产品、美元、机械、国际、市场、出口、销售、海外、收购、工程、全球、欧元、欧洲、生产、子公司、倾销、订单、德国、通用
债务和担保	银行、担保、贷款、提供、资金、持有、质押、债券、责任、利息、借款、机械、期限、控股、偿还、子公司、分行、债务、支行、财务
法律诉讼	诉讼、法院、仲裁、人民法院、纠纷、中级、破产、申请、损失、收到、冻结、判决、中华、赔偿、案件、起诉、虚假、民事、债权人、尚未

附录 C 信息挖掘 (Discover) 和信息传递 (Dissemination) 计量步骤

1. 以文档 d 标记上市公司公告与媒体报道的文本, 构建文档 d 中的主题向量 T_d 。在文档分词的基础上, 利用 LDA 建模可以取得各个行业文档集的主题—词汇分布矩阵 $Topic_Word$, 这一矩阵将实现特定主题 (S) 与一组词汇的对应。针对某一文档 d (可以是上市公司公告, 也可以是媒体报道), 以句子为单位, 分析句中的词汇, 若该句所用词汇归属于主题 S 的最多, 则该句被归为主题 S 。以此类推, 我们可以对文档 d 中所有句子进行主题归类。我们再把文档 d 中归属于各个主题的句子数除以文档 d 中句子总数, 形成如下向量:

$$\text{文档 } d \text{ 中的主题向量} = T_d = (S_{d1}, S_{d2}, \dots, S_{d40}) \quad (10)$$

2. 确定文档 d 中的主题后, 针对特定主题 S , 汇总文档中出现的归属于主题 S 的词汇列表, 并统计各个词汇出现的频数, 进而形成每篇文档中每个主题的主题词语向量。

$$\text{上市公司公告中主题 } S \text{ 的主题词语向量} = W_{Ann,S} = (w_{1S}, w_{2S}, \dots, w_{NS}) \quad (11)$$

$$\text{媒体跟踪报道中主题 } S \text{ 的主题词语向量} = W_{Media,S} = (w_{1S}, w_{2S}, \dots, w_{NS}) \quad (12)$$

其中, 元素 w_{iS} 或 w_{iS} 代表所在文档中词语 i 用于描述主题 S 的频数。

3. 度量媒体跟踪的信息挖掘程度。本文计算了上市公司公告与媒体跟踪报道的主题向量余弦值,用1减去该余弦值来衡量两者之间的主题分布差异,其现实意义便是媒体跟踪报道的信息挖掘程度,计算公式如下:

$$Discover = 1 - \frac{\sum_{s=1}^{40} (S_{Media,s} \cdot S_{Ann,s})}{\sqrt{\sum_{s=1}^{40} (S_{Media,s})^2} \cdot \sqrt{\sum_{s=1}^{40} (S_{Ann,s})^2}} \quad (13)$$

其中, $S_{Media,s}$ 和 $S_{Ann,s}$ 分别代表媒体跟踪报道和上市公司公告的主题向量中第 s 个元素。

4. 度量媒体跟踪的信息传递效果。首先针对每个主题 S , 计算该主题的词语向量的余弦,然后将所对应上市公司公告中的前 S_d 个主题下计算的这个值进行算数平均。我们对每篇上市公司公告的主题频率从高到低排序,取累积频率达到 80% 的前 S_d 个主题进行检验,以排除文档中不太重要的主题所带来的干扰。

$$Dissemination = \frac{1}{S_d} \sum_{s=1}^{S_d} \frac{\sum_{j=1}^N (w_{js} \cdot v_{js})}{\sqrt{\sum_{j=1}^N (w_{js})^2} \cdot \sqrt{\sum_{j=1}^N (v_{js})^2}} \quad (14)$$

其中, N 为文档 d 中所有不同的词语个数, v_{js} 和 w_{js} 分别为上市公司公告和媒体跟踪报道中主题 S 的词语向量中第 j 个元素。

Chinese Media, News Characteristics and Information Efficiency of Capital Market

Yang Yulong¹, Wu Wen², Gao Yongjing³, Zhang Qiannan¹

(1. School of Accounting, Zhejiang Gongshang University, Hangzhou 310018, China;

2. PBC School of Finance, Tsinghua University, Beijing 100083, China;

3. School of Accountancy, Shanghai University of Finance and Economics, Shanghai 200433, China)

Summary: Since reform and opening-up, Chinese media industry has undergone profound development. The forms of media have become more diversified. However, the increase in the amount of media is accompanied with some problems, such as the erosion of independence due to business benefits, and the relatively low quality of employees in media industry. Therefore, readers are often confused about how to select reliable media. To address these issues, we need scientific research to provide us with guidance, and help us understand the information characteristics of media.

As a combination of policy-oriented media and market-oriented media, Chinese media industry is unique for its dual-track feature. The policy-oriented media is supported by government finance in some way, and mainly focuses on government policy. In contrast, the market-oriented media is self-financing, and pays more attention to the information demands of capital market. Prior literature finds that media coverage can help improve the quality of corporate governance, and enlarge the information efficiency of capital market. But for investors, they care more about how to select media or news articles to facilitate their decision making. In other words, they want to identify the very media or news article that has a substantial impact on the information integration of capital market. Therefore, further research should take account of media types and characteristics of news articles, in order to figure out the functional distinction between different types of media.

Based on the dual-track feature of Chinese media industry, this paper discusses the differences and reas-

ons of performance in facilitating information integration of capital market between policy-oriented media and market-oriented media. By comparing the use of positive and negative words in news articles, we find that positive reports dominate in all media. But the proportion of positive reports is larger for policy-oriented media than for market-oriented media, while the proportion of negative reports goes opposite. We also find that the proportion of positive reports gradually declines from 2000 to 2013, while that of negative reports gradually increases, and the above trends are more significant in the market-oriented media. Given news articles are carriers of information, so the differences of news articles' features between two types of media must lead to different performances in disseminating information.

Using Python programming language and Genius Finance News Database, we sort and analyze all the news articles related to the listed firms. Consequently, we construct a database of positive and negative media reports, which covers the period from 2003 to 2013, and consists of 137 unique media. With a sample from China A-share listed non-financial firms from 2003 to 2013, we find that the media coverage can improve the information efficiency of capital market through the channel of negative news rather than that of positive news. Moreover, the above effects are mainly driven by the policy-oriented media, while the market-oriented media barely shows no influence. To explore the mechanism behind those relationships, we quantify the depth of information discovery and the accuracy of information dissemination for each news article, using the method of textual analysis. Further empirical evidences indicate that the policy-oriented media discovers information deeper and disseminates information more accurate than market-oriented media, and this is the reason why policy-oriented media performs better than market-oriented media in facilitating the information integration of capital market.

Our results imply that negative news is more valuable for investors' decision-making. Especially, we find that the impact of media on information efficiency of capital market is mainly driven by the policy-oriented media, which is helpful for investors to narrow their scope of media selection. Meanwhile, our results can help regulatory authorities evaluate and select media platform, and assist them to provide media with guidance in respect of news content.

Technically, using the method of textual analysis, we get a database of positive and negative vocabulary which is compatible with Chinese news reports. Based on the vocabulary database, we further analyze the news articles in Genius Finance News Database by Python programming language. Finally, we construct a database of positive and negative news reports, which covers the period from 2000 to 2013. Moreover, we employ the model of Latent Dirichlet Allocation to measure the depth of information discovery and the accuracy of information dissemination, which helps us explore the reason why the policy-oriented media performs better than the market-oriented media in terms of improving information efficiency of capital market.

Key words: policy-oriented media; market-oriented media; stock synchronicity; PIN

(责任编辑 康健)