

大数据思维在金融学研究中的运用*

蔡庆丰 郭春松 陈诣之

内容提要:随着大数据时代的到来,大数据思维也日益改变现代金融学的研究范式,拓展新的研究领域。一方面,大数据使得金融学摆脱数据样本的束缚,基于海量数据的大样本甚至全样本研究日益成为西方金融学者的共识,并且数据来源和类型的多样化也促使现代金融学研究日益多元化。另一方面,数据处理技术和运算能力的不断进步,使得金融学者可以在一些非传统领域拓展新的研究视角,例如媒体效应分析、网络数据分析等。此外,在数据样本日益扩大和数据类型不断丰富的背景下,更为精准的量化分析也开始成为西方金融学者关注的重点。本文通过对新近几年来西方金融学者相关研究的梳理与评述,从上述三个角度来总结大数据思维对现代金融学研究的影响与应用。

关键词:大数据 金融学 非结构化数据

尽管目前学界对于“大数据”的定义仍存在诸多争议,但对于大数据对经济社会的深远影响却已是普遍共识:随着大数据时代(The Age of Big Data)的到来,人们的生活方式、工作方式、组织方式与社会形态都将随之发生深刻改变,甚至在某些方面,这种改变可能是颠覆性的。比如说,在传统的研究范式中,学者研究的重点往往集中于因果关系上,而在“大数据”的思维范式里,通过海量数据的处理,研究的重心则将转移到相关关系分析上——人们更多地通过验证相关关系来解决问题或做出决策,从过去研究“为什么”转变成如今研究“是什么”,这一思维的改变不仅对人类生活产生深远影响,对于科学研究也将是一次全新的挑战(Mayer-Schönberger, 2012)。

现代金融学研究自20世纪50年代以来得以迅速发展,在公司金融(Corporate Finance)、金融市场(Financial Market)、资产定价(Asset Pricing)等领域建立起严谨成熟的理论体系。与其他社会科学一样,金融实证研究基本上无法通过特定的实验来进行,这也就决定了在研究方法上主要依靠现实数据收集和计量分析手段。由此,数据采集及其分析对于金融学研究的重要程度不言而喻。在数据类型、来源和特征发生巨变的“大数据时代”,金融学研究

也必然面临诸多挑战和变化。近年来,越来越多西方金融学者将“大数据思维”引入现代金融学研究,相关研究成果也越来越多地被《金融学期刊》(*Journal of Finance*)、《金融经济学期刊》(*Journal of Financial Economics*)、《金融研究评论》(*Review of Financial Studies*)、《美国经济评论》(*American Economic Review*)、《经济学季刊》(*Quarterly Journal of Economics*)等顶级金融学、经济学期刊接受刊发。本文通过对近年来西方金融学、经济学学术文献的梳理,对大数据思维在近几年西方金融学研究中的运用及其影响进行归纳与评述,以期为国内学者在借鉴、运用“大数据思维”进行金融问题研究方面提供参考和启示。我们基于相关文献的梳理后认为,大数据思维在现代金融学研究中的体现主要有以下三个方面:首先,大数据的出现在一定程度上发展了传统的金融学研究模式;其次,大数据技术能够给金融学提供更多的研究领域,丰富金融学的应用范畴;最后,大数据背景下的量化分析方法将在金融学研究中得到更加广泛的应用。

一、大数据丰富了现代金融学的研究内涵

学者通常将大数据的四个显著特征归纳为4V

* 蔡庆丰、陈诣之,厦门大学经济学院、两岸关系和平发展协同创新中心,邮政编码:361005,蔡庆丰电子邮箱:qfcai@xmu.edu.cn;郭春松,中国银监会福建监管局。基金项目:国家自然科学基金项目“资本市场中的社会关系与政治地理中的资本市场:来自中国的理论与实证”(71373219);国家自科基金青年项目“金融部门的利益冲突、自我膨胀及其对实体经济的影响”(71103150);中央高校基本科研业务费立项课题。感谢匿名审稿人提出的修改建议,文责自负。

特征,即大容量(Volume)、高速率(Velocity)、多样性(Variety)、真实性(Veracity)。基于大数据的4V特征,在金融学研究中,使用可获得的最大样本容量,使用更为广泛真实的数据来源进行实证研究和经验分析,往往能突破传统数据分析的各种局限,拓展新的研究思路,呈现新的研究结果。

(一)海量数据使金融学实证研究进入大样本时代

数据容量的几何倍增是大数据的首要特征。据2011年麦肯锡全球研究院(McKinsey Global Institute)的研究报告,当年全球数据增量已达1.8ZB^①。随着近几年计算机存储技术的高速发展和云存储技术的不断突破,可以预见,未来数据容量扩张的速度还将加快。海量数据的出现,意味着数据样本容量的激增。Mayer-Schönberger(2012)指出,大数据能够使未来研究从随机样本变为所有数据。因此在金融学研究中,大样本甚至全样本数据替代原有的随机数据成为可能,实证结论也将产生新的变化。

1. 原有结论的稳健性得到加强。在统计推断中,样本容量的增加会提高实证检验结论的稳健性(Hayashi, 2000)。海量数据的出现,意味着金融学者能够使用更全面、更高频的数据样本对原有金融结论进行检验分析,进而提高研究结论的可靠性与说服力。我们以天气变化对资本市场的影响为例进行评述。

Saunders(1993)在研究天气变化对股市收益的影响时,仅使用了纽约华尔街地区的气温、相对湿度、降水量、风力、日照情况、云量等六项天气指标,以日数据为样本观察天气变化与当日股票价格变化两者之间的相关关系。作者研究发现,华尔街地区多云阴暗的天气变化与纽约证券交易所、美国证券交易所当日股票价格变化呈现显著的负相关关系,但由于数据样本的限制,当时并没有检验全球其他国家和地区的天气状况与股市收益之间的关系,也没有对其作用机制进行更为严谨的分析。10年之后的2003年,遵循类似的研究思路,Hirshleifer & Shumway(2003)运用美国国家气象数据中心(NC-DC)提供的全球天气观测数据,挑选了全球26个国家的主要证券交易所所在地的天气情况作为研究样本,同样证实了一国股市日收益变化与当天该国交易所云层覆盖情况存在显著的负相关性。通过增加数据样本容量,他们的研究证实了天气状况对当地股票市场的影响在不同国家都是显著且一致的,这进一步增强了Saunders(1993)结论的稳健性。

随着数据处理技术的进步,气象研究数据进一步丰富。Goetzman et al(2013)利用美国国家海

洋和大气管理局(NOAA)综合数据库(ISD)所提供的天气数据也进行了相关金融学问题的研究。ISD数据库由全球超过20000个气象站(其中包括美国的7610个气象站)提供,并且能够以小时数据作为频率进行考察,几乎可以涵盖美国所有地区的天气变化情况。他们从中选择了美国所有机构投资者所在地区的天气变化数据,研究机构投资者的投资行为与实时天气变化之间的关系。他们同样也发现,多云、阴暗天气状况会使得机构投资者高估资产价值,而这一分析结果与20年前Saunders(1993)的研究结论保持一致。Goetzman et al(2013)用全样本数据替代既有研究的小样本数据,解决了实证研究中的内生性问题,不仅增强了原有结论的可靠性,而且通过天气变化的高频数据进一步分析了天气变化与资本市场之间可能的影响机制。

使用大样本或者全样本数据研究以加强结论的稳健性和说服力已经成为现代金融学的重要研究思路,而大数据的出现无疑为这一方法提供了数据基础和技术保障。不仅如此,随着数据样本容量的扩张,金融学者能够在传统金融学研究基础上,利用数据内容的不断细化,在原有研究结论的基础上发现更多更为精准、原有研究没能发现的研究成果。例如,在考察金融市场中存在的性别差异问题时,最初金融学者只是依靠一些投资机构提供的管理数据(Administrative Data),笼统地分析性别差异对个人投资规模的影响,但是由于数据样本的限制,内生性问题较严重,研究结论不仅稳健性不高,并且在研究视角上也显得过于单一。Sundén & Surette(1998)利用美联储与美国收入统计局(Statistics of Income)提供的“消费者金融调查”(Surveys of Consumer Finances)数据,研究了工人在选择投资养老金行为中所存在的性别差异现象。此外,他们利用政府部门公布的抽样调查数据替代商业机构发布的特定数据,也验证了性别差异的确存在于金融市场活动中,并且发现性别差异在单身投资者群体中表现得更加显著。

随着数据样本容量的不断增加和丰富,有学者进一步证明了中小投资者在股市投资行为中存在着性别差异现象(Barber & Odean, 2001)。那么,这种性别差异是否在公司金融领域也同样存在呢?受制于数据采集的可获得性,既有的性别差异问题研究较少深入到公司高管这一群体,这也导致这一问题的研究始终难以深入。进入21世纪以来,一方面女性在金融市场的参与度不断提高,另一方面,机构投资者的相关数据库也日益丰富。Adams & Fer-

reira(2009)运用 IRRC(Investor Responsibility Research Center)所提供的美国标普 1500 上市公司董事会数据研究了女性董事在公司金融和公司治理中的作用。通过对美国 1600 余家上市公司近 10 年的董事会发展变化,他们发现了女性董事比例越高会对企业发展产生负面影响,这与主流媒体的宣扬内容相违背。Faccio et al(2014)运用 BvD(Bureau van Dijk)公司提供的泛欧企业财务分析数据库(Amadeus Data Base)研究了整个欧洲地区 15 万余家企业中 CEO 群体的性别差异问题,研究结果显示,在职业经理人这个群体中性别差异同样显著。

2. 原有结论的显著性发生改变。实证研究中,回归系数显著性与样本容量有着密切的关系,样本容量的改变有可能会改变结果的显著性,进而改变研究结论。大样本容量回归运算得出的结论往往可靠性更高(Hayashi, 2000)。随着样本数据的激增和丰富,运用大样本数据替代原有的小样本数据进行研究分析时,可能会改变原有结论的显著性。

Rotton & Rosenberg(1984)在研究股市价格变化时,提出了一个大胆的猜想——月球运动可能会影响到股市的波动。但受限于当时的天文观测条件,他们仅采用简单的月球运动观测数据,运用简单的时间序列分析方法进行实证研究,结果并未发现月球运动与道琼斯指数变化之间存在显著的相关关系。由此他们认定资本市场运行并不会受到月球运动的影响。然而,诸多学者研究发现月球运动确能影响人类的行为模式(Kelly et al, 1998),而投资者的有限理性行为也会影响金融市场的运行(Hirshleifer, 2001; Kahneman & Riepe, 1998)。由此我们可以合理推论,月球运动可能会对金融市场产生影响。在此基础上,越来越多的学者对 Rotton & Rosenberg(1984)之前的结论提出了质疑。

进入 21 世纪,天文观测技术和数据处理分析技术得到了进一步的发展。Dichev & Janes(2003)重新对月球运动规律与股票市场波动之间的关系进行了考察,他们从当时最专业的月球运动网站 www.lunaroutreach.org 采集观测数据,运用被学者普遍接受的“月球会合周期”(synodic lunar cycle)^②划分月球运动的规律,通过将非金融数据与同期美国股票市场的收益数据进行实证分析。他们研究发现,美国过去 100 年间,新月(New Moon)前后 15 天的股票收益大约是满月(Full Moon)前后 15 天收益的 2 倍,股票市场与月球运动之间存在着显著的相关关系,从而推翻了 Rotton & Rosenberg(1984)

股票市场不受月球运动影响的研究结论,同时也为后续研究奠定了基础。此后, Yuan et al(2006)则从美国海军天文台(USNO)网络数据库中获得了更加精确的月球四相位^③运动数据,这一数据不仅捕捉了从美国不同地区观测月球运动的详细数据,并且收集了地球上不同国家和地区的人们观测到的月球变化数据。他们通过将这一天文观测数据引入金融学研究模型,研究了月球运动与美国等 48 个国家股市收益变化之间的关系,结果同样证实了新月前后的股市收益率显著高于满月前后的股市收益率。他们采用全球的股票市场数据进行研究,使得 Dichev & Janes(2003)的结论更加稳健可信。并且,他们还通过高频数据的处理进一步发现,月球运动两相位间的股市收益率之差在世界范围内均保持在 3% 至 5% 之间,进一步丰富了研究成果。尽管目前学者关于月球运动对股票市场的影响机制并没有达成一致,但就月球运动会对股市产生显著影响这一研究成果却被普遍接受。

随着大数据样本的广泛使用,金融学者不仅发现了一些小样本研究未能发现的研究结论,还推翻了忽略小样本内生性而得到的一些研究结论。我们以市场的知情交易问题研究为例。在市场微观结构理论中,交易者按照是否具有信息优势被分为知情交易者和非知情交易者,其中,通过掌握更多私人信息的知情交易者进行的投资交易占有所有交易的比率被定义为知情交易概率 PIN(Probability of information-based trading)。知情交易概率作为测度市场知情交易的指标,自提出以来尽管广为引用,但同时也备受争议。该指标提出之初, Easley et al(1996)、Easley et al(2002)等学者分别运用纽约证券交易所 1 年和 10 年的市场数据实证研究发现,知情交易会显著影响股票收益。但这一研究结论却遭到诸多学者的质疑,尤其是近年来海量高频数据的引入,诸多学者认为这一研究结论并不成立。Duarte & Young(2009)从证券市场研究机构(ISSM, Institute for the Study of Securities Markets)和纽约证券交易所交易与报价(TAQ, Trade and Quote)数据库中收集了 23 年纽约证券交易所的股票交易数据,实证研究发现,知情交易本身并不影响股票收益,股票收益波动只是由信息交易引发的股票交易频率变化而导致的。在最新的研究中, Lai et al(2014)将研究范围扩大到全球主要成熟市场,通过考察 47 个国家的 3 万余只股票的市场数据发现,知情交易概率(PIN)影响股票收益的结论在全球范围内都是不显著的,这对 Eas-

ley et al(2002)的研究发现提出了强烈的质疑。目前,西方金融学界对于知情交易概率(PIN)是否影响资本市场收益的争论仍在继续,但随着大数据的使用,越来越多的金融学者更倾向于支持知情交易并不影响市场收益的观点。可见,随着大数据时代的到来及其广泛运用,金融学者除了拓展新的研究领域,延展新的研究思路,发现新的研究结论外,既有的一些研究成果也面临大数据的质疑和重新验证,这对于完善和发展现代金融学研究无疑是有益的。

(二)数据多样性拓展金融学研究视角

大数据之“大”不仅指海量数据的激增,还特指数据类型从单一的结构化数据发展为多元化的非结构化数据。未来数据不再仅仅局限于以结构化数据的形式出现,还包括文本、图像、音频、视频等各种非结构化数据。数据类型的这种多样性有时被称之为混杂性。Mayer-Schönberger(2012)认为未来的科学研究将由精确性转化为混杂性,这表明在大数据时代,学者们将从单因素分析进入多元影响分析,这也说明大数据时代的数据分析与处理能够同时从多个角度进行。数据多样性还体现在数据来源和用途上,互联网平台的建立和云处理技术的不断突破,使得学者对数据的获取方式更加多元化,现代金融学研究思路与领域也因此得以丰富。总之,运用大数据处理技术,我们可以将非结构化数据引入现代金融学研究,拓展研究思路,丰富研究领域。

1. 广泛的数据来源丰富了传统金融学的研究内容。在传统的金融学研究中,研究样本往往限于金融、经济、财务等结构化数据。而近几十年来,这种基于结构化数据的研究几近处于“过度研究”的状态,很多研究甚至有“数据挖掘”之嫌。随着数据来源和类型的不断增加,越来越多的金融学者们开始将各种非结构化数据引入现代金融学研究。

例如,在公司金融领域中,CEO的个人特征和特质对上市公司的影响一直以来都是这个领域的研究热点。但既有的研究往往都是从CEO的年龄、学历、工作经验等出发,利用上市公司所披露的公开信息来考察CEO个人特征和特质对上市公司的影响(Bertrand & Schoar, 2003; Adams et al, 2005)。然而,金融学者利用大数据思维对CEO的考察已经深入到各个方面,比如,感情状况(Mayew & Venkatachalam, 2012)、心理特征(Graham et al, 2013)、早年经历(Malmendier et al, 2011)、管理和领导才能(Kaplan et al, 2012; Bolton et al, 2008)、参军经历(Malmendier et al, 2011; Benmelech & Frydman, 2014)、外貌形象(Gra-

ham et al, 2010; Halford & Hsu, 2014)等。

上述研究的开展得益于大数据时代背景下的更多非财务数据的可得性,以及非结构化数据的可量化性和可分析性。Benmelech & Frydman(2014)在研究美国上市企业CEO早年从军经历对企业发展的影响时,从美国工商业名人录(Who's Who in Industry and Commerce)和盖尔研究组(Gale Research Group)以及各类新闻报道、人物传记中,搜集了美国自二战以来至2006年间所有美国上市公司CEO的详细军方资料(包括其是否具有作战经历、军衔级别、服役部队资料等非结构化数据信息)。他们研究发现,具有军队履历的CEO相比于没有军队履历的CEO具有更低的投资倾向和R&D投入,并且其管理的企业发生欺骗行为的概率也会更低,同时早年从军的CEO也能够使其领导的公司在行业不景气时表现得更为出色。

Gompers et al(2012)在研究不同投资者之间能否成功合作的关键因素时,在各大金融网站、美国证券交易委员会(SEC)文件、新闻报道、个人自传等各类信息网站和数据库中手动搜集到关于风险投资家的工作经历、学历、性别等文本信息,还通过信息挖掘等技术获得了相关网站和数据库没有公布的信息。他们利用所收集到的大量非结构化数据研究发现,具有相同特征的风险投资者会更加愿意一起合作。例如,双方都是拥有名校学位的风险投资家,共同合作一次以上的概率比平均水平高出8.5%;如果双方从同一大学相同专业毕业,合作概率可以高出20.5%;如果是相同的少数族裔,则合作概率会高出22.8%。

可以预见,随着大数据时代的到来,数据类型和来源将更加丰富多元,已经不局限于传统的结构化数据库,中国的金融学者在研究时应该打破固有的结构化数据思维模式,尝试将更多半结构化数据和非结构化数据引入金融学研究,拓展研究领域,丰富研究思路。

2. 数据多样化使金融学与其他学科间的联系加强。不同学科之间的研究交流对于各学科的发展进步均有意义。而大数据时代,多元数据的运用与分析能够为不同学科之间更为紧密的联系和互动提供基础。金融学作为一门从经济学中分化出来的应用学科,利用数据多样化的特征,引入其他学科中的研究成果和数据内容,加强与不同学科的交叉影响,可以进一步扩大自身的研究边界。

心理学研究已经证实了球迷的心态会随着球队表现不同而呈现出强烈的波动(Wann et al, 1994)。根据这一理论基础,Edmans et al(2007)研究了足球

赛事结果对股票市场的影响。根据行为金融学的研究成果,股市波动显然会受到投资者情绪的直接影响,而足球赛事结果又会对投资者情绪产生直接影响,由此可以推断,足球赛事结果会影响股票市场。为了验证这一想法,他们从 www.rdasilva.demon.cn.uk 网站中获得了从 1973 年 1 月至 2004 年 12 月期间所有的国际足球比赛结果,其中包括世界杯、欧洲杯、美洲杯等国际重要赛事以及各国之间友谊赛的结果。研究结果符合作者的预期,输掉一场足球淘汰赛会导致该国股票市场在第二天产生平均 49 个基点的收益损失,并且这一损失随着比赛重要程度的提升和本国资本市场规模的降低而显著增加,由此证实体育比赛^⑥能影响投资者心态,进而影响股票市场波动,这与心理学研究成果保持一致。

除了社会科学之外,西方金融学者还通过大数据处理方法将自然科学的相关研究成果引入金融学研究。Frydman et al(2014)在研究投资者交易行为时,将神经学数据(Neural Data)引入到行为金融学研究。他们利用功能性磁共振成像(FMRI, Functional Magnetic Resonance Imaging)的方法对特定实验者的大脑神经元活动进行了观测,观察实验者在特定实验中某种特定投资行为前后的神经元数据变化。结果显示,实验中的“投资者”大脑神经元变化能有效地解释其投资行为的前后差异,尤其是投资收益发生较大波动时,神经元数据的变化更加明显。他们利用神经元数据的变化对行为金融学中普遍存在的处置效应(Disposition Effect)与实现效用模型(Realization Utility Model)进行了验证。从这一点可以看出,运用大数据思维,导入神经学科的研究数据和研究方法,对现代金融学中的投资者行为和不确定性问题进行研究,无疑丰富了现代金融学的研究思路。类似地,Sornette(2014)总结了物理学理论在金融学中的应用,他认为现代物理学的建模方法能够帮助解决金融学模型中的发展瓶颈。可以预见,在大数据时代运用大数据思维,对多元化数据进行分析和处理,可以有效加强各学科之间的关联和互动,延展丰富金融学研究思路和研究领域。

3. 非结构化数据分析进入金融学研究领域。数据可以分为结构化数据(Structured Data)和非结构化数据(Unstructured Data),前者是指能够用二进制编码进行存储管理的数据,后者则是无法用已经设定好的数据模型管理的信息(Wikipedia),这也意味着传统的数据处理方法无法直接应用在文本、图片、音频、视频等非结构化数据的处理上。进入大数据时代,数

据挖掘(Data Mining)、文本分析(Text Analytics)^⑦、自然语言处理(Natural Language Processing)^⑧等方法开始运用于非结构化数据处理,非结构化数据的可量化分析使得金融学者能够将其与传统的公司财务数据、市场交易数据等结构化数据相结合进行金融学研究。

近几年西方的金融学研究中,文本分析受到越来越多的关注。通过对上市公司在互联网上所公开披露的各种文件、报告进行文本分析,我们能够把文件中透露出的语气、态度进行量化分析(Demers & Vega, 2011; Davis & Tama-Sweet, 2012),这是大数据时代非结构化数据的典型处理模式之一。Loughrann & McDonald(2013)将 1997 至 2010 年间美国拟上市公司向 SEC 提交 IPO 的申请文件(S-1 文件)进行了计算机自动文本分析,计算申请文件中在介绍公司商业策略、经营管理及未来预期等方面时不确定性语气单词(uncertain, weak modal, negative, positive, legal, strong modal 等)的使用频率,从而考察文件内容的不确定性水平是否会影响投资者对 IPO 估值的能力。他们的研究表明,S-1 文件内容不确定性水平越高,公司 IPO 首日收益越高,绝对报价修改情况越大,上市后的波动率也越大。同期,Jegadeesh & Wu(2013)也运用非结构化数据处理技术对上市公司发布的 10-K 文件进行文本分析处理,对文件中的积极词汇和消极词汇的词频进行量化测度,并赋予相应权重,由此计算出 10-K 文件整体语气倾向,在此基础上探讨 10-K 文件整体语气倾向与公司股价走势之间的相关关系。实证结果显示,两者之间存在着显著的相关关系。

利用文本分析,金融学者不仅能够对上市公司的信息披露文件进行整体语气分析,判断上市公司的整体发展情况,还能够更深入地挖掘出某些“潜藏”的企业特征。Zhou(2014)收集了 2003 至 2010 年间近 9 万份上市公司的季度报告,利用文本分析方法,研究企业高管是否习惯性地企业业绩不佳的原因归结于企业外部因素的变化而非自身原因。Zhou(2014)将报告分解成为各个语句,如果句子中有提到宏观经济词汇(例如“Economy”、“Economic Growth”等),该语句便属于“归咎企业外部因素”的语句,同时他还利用 Loughrann & McDonald(2013)的文本分析方法将季度报告中的语句分为“积极语句”和“消极语句”。通过计算机文本分析,Zhou(2014)发现在“归咎于企业外部因素”的语句中消极词频显著多于积极词频,并且内容主要涉及

企业不良的财务状况等。通过文本分析技术,Zhou (2014)挖掘出潜藏在上市公司季度公告中的高管态度,并在此基础上进一步实证研究发现,上市公司高管的这种归咎思维(Blame Game)和消极态度并不利于企业的后续发展。

近年来,随着人工智能技术的发展,衍生出更多更新的非结构化数据处理技术。西方金融学者也都积极尝试着将其引入学术研究。Gurun & Butler (2012)在研究企业价值与当地媒体关系时,就运用了TNSMI(TNS Media Intelligence)数据库。该数据库收集了各种形式的广告数据,包括报刊、卫星电视、有线电视、广播、杂志等媒体广告。他们研究发现,相比于外地企业,本地媒体在报道本地企业时较少使用负面性描述,并且,本地媒体对于本地企业的偏袒程度往往与该企业的市值密切相关。Peress & Schmidt (2014)在研究中小投资者的投资行为时,采用“新闻压力”(News Pressure)这一指标来测度中小投资者的关注度转移。他们从David Stroemberg's website中收集了美国各大主要新闻广播电台1991至1996年间的新闻内容,将每次新闻播报中头条三条新闻的播报时间加总起来作为“新闻压力”指标,认为排在前三条的新闻播报往往是当地最重要的新闻事件,播报时间越长说明受到关注将越高,因此中小投资者的注意力也更有可能会被转移。通过量化新闻音频内容,他们研究发现,当投资者注意力更多地被转移时,尽管交易规模没有发生改变,但交易频率会有一定程度的降低。

二、大数据分析拓展了现代金融学的研究思路

大数据时代除了数据量的激增之外,大数据的处理技术也在不断进步,发展可谓日新月异。根据摩尔定律(Moore's Law),计算机的信息处理速度每隔18个月就能提高一倍。信息处理能力的提升意味着学者们可以进入之前较少关注的研究领域。近年来,新闻媒体内容开始被金融学者广泛收集用作研究素材,同时越来越多的金融学者将互联网上的各种数据和信息引入金融学研究,发现了不少有意思的研究结论。

(一)金融市场中的媒体效应研究

金融市场中的信息传递一直是现代金融学研究的热点。媒体作为信息传播的主要媒介,必然会深刻影响着投资者行为。但在过去相当长一段时间里,由于难以定量测度媒体对各类信息的传播能力,学者们在研究媒体效应时往往只能进行定性分析。随着大数据相关处理技术的进步,各种媒体上的各类非结构化数据和信息的收集整理及量化分析也有

了可能,有些金融学者开始对金融市场中的媒体效应进行实证分析。

媒体传递信息能够有效降低金融市场的信息不对称。Tetlock(2010)对上述命题进行了深入研究,并根据过去29年道琼斯新闻存档(DJ News Archive)中美国所有上市公司的新闻资料构建了样本,包含了大约220万个新闻事件。作者通过数据挖掘、文本分析等数据处理技术研究发现,媒体报道的确能够在一定程度上解决上市公司与投资者之间的信息不对称,尤其是对于那些能够从媒体新闻中发现信息并利用信息改变投资行为的投资者。Tetlock(2010)的研究肯定了媒体对于解决金融市场信息不对称问题的作用,但媒体报道是否能够即时准确地反映市场中的真实信息呢?对于这一问题,Engelberg & Parsons(2011)在ProQuest新闻数据库中采集了美国18个主要城市的主流报刊和2份全国性媒体的所有新闻报道,通过比较不同地区媒体在报道同一新闻事件时各地投资者的交易行为差异,证实媒体报道的口径和倾向性会显著影响投资者的交易行为,这也说明媒体对于金融市场而言,并不仅仅只是简单纯粹的信息传递者,同样也是重要的市场参与者,并可能还带有某种偏见。

新闻媒体对上市公司的直接报道对于资本市场的影响更为直接(Tetlock, 2007; Tetlock et al, 2007; Peress, 2008)。随着大数据处理技术的进步, Hillert et al(2014)对1989至2010年间共22年的来自NYSE、AMEX、NASDAQ的所有普通股股票的收益数据与同一时期《纽约时报》(New York Times)、《华盛顿邮报》(Washington Post)、《今日美国》(USA Today)、《华尔街日报》(Wall Street Journal)四大报刊和41种地方性报刊共220万篇相关报道之间的相关关系进行了研究,并运用文本分析法计算出每篇新闻报道中所含负面词汇的比例,在此基础上研究了媒体报道的倾向性对股票市场收益的影响。他们的研究结论证实了上市公司的相关媒体报道、市场短期动量效应以及长期反转效应之间存在着系统性关联。他们的研究样本几乎涵盖了美国所有主流财经媒体,这相当于运用全样本数据容量进行研究,从而有力地证实了媒体报道对于股票市场的直接影响。

Dang et al(2014)则使用更大范围的样本数据,从国别数据角度对金融市场的媒体效应进行了研究。他们从RavenPack数据库中获得了41个国家2000至2009年间所有媒体关于上市公司的实时新闻报

道,RavenPack 数据库不仅提供新闻报道,还运用专业的评测手段将新闻报道进行评级分类。所有新闻报道根据一定的标准划分为积极的(Positive)与消极的(Negative)两大类。他们借助于该数据库研究发现,在机构投资者占比较低的股票市场上,新闻报道引发的上市公司股价波动会更加显著。此外,如果该公司有更高的新闻通用性(News commonality),新闻报道会引发公司更高的涨幅和更多的交易量,并且这一效应在机构投资者占比较高的成熟市场更加显著。

随着非结构化数据在经济学、金融学中的广泛运用,越来越多的西方金融学者从不同角度对媒体效应展开研究。Peress(2014)研究了1989至2010年间OECD国家爆发的全国性新闻媒体罢工事件对该国股票市场的影响。他们通过查询事件报道和官方记录,获得了每次罢工事件的历时与性质等特征,并在此基础上考察了52起罢工事件对期间88个交易日的影响效应。研究发现,在全国性新闻媒体罢工当日,该国股票市场的交易量、波动率均有显著下降,这从另一角度证实了新闻媒体对金融市场在信息传播过程中的重要作用。

(二)使用互联网数据进行金融学研究

进入21世纪以来,越来越多的人参与到社交网络平台 and 互联网交易平台中。互联网的一个重要特征就是每个参与者在互联网空间中的行为都会留痕,并被存储形成互联网数据。互联网空间中每时每刻都在产生的海量数据是最为典型的大数据,这些数据往往蕴涵着丰富的研究价值有待挖掘。

Bollen et al(2011)收集了每日全球知名社交网站Twitter上数以亿计的美国用户所发布的情绪词汇,通过加权汇总处理形成情绪指标。他们认为国民投资情绪可以通过该指标进行量化表示,并且通过实证研究发现,这一指标与道琼斯指数存在着显著的相关关系。沿着这一研究思路,Karabulut(2013)以另一个同样广泛使用的社交网络平台Facebook上的互联网数据作为研究样本,采集了Facebook自2004年12月创立至2012年4月期间,其社交平台上1.6亿美国用户在更新“个人状态”(Status)时所使用的“情绪”(emotion)词汇,并通过汇总每天所有更新“个人状态”用户的情绪词汇,构建了国民幸福指数GNH(Gross National Happiness)变量。作者认为,GNH指数在一定程度上可以反映出美国国民的情绪状况,其中也包括参与金融市场的投资者情绪,并且所构造的GNH指标能够有效预测美国股市的未来走势。

随着社交网络平台的发展,越来越多的投资者喜

欢聚集在一些专业的网络平台交流投资经验,分享投资建议。Chen et al(2014)从美国最大的投资者社交网站SA(Seeking Alpha)下载了其中“只关注某一只股票走势”的分析文章以及相应评论,时间跨度从2005年到2012年,共97070篇文章以及在文章发表两天内所发表的相关评论共计459679条。他们借鉴了Da et al(2011)的文本分析方法,将文章及评论内容进行了量化分析,结果发现,文章和评论能够有效预测股票的未來走势和收益,并且这一影响在控制了财经主流媒体的新闻报道后依然是显著的。他们通过海量互联网数据的量化分析,证实了群体建议效应(Peer-based advice)存在于金融市场的投资决策中。

Heimer & Simon(2012)收集了美国某个电子商务平台上所有详细的注册账户资料、交易信息、客户和客服之间聊天记录等数据内容。他们通过利用所收集的超过100万条的交易结果以及交易双方的聊天记录等数据,证实了Hirshleifer(2010)的结论,即投资者的社交行为会增加其交易的频率。他们的研究首次采用完全真实的全样本互联网交易数据进行投资者行为的研究,从而更加贴近于投资者的真实情况。在互联网时代,金融体系中的各类投资者都会在各种互联网空间上留下痕迹,形成各种结构化和非结构化数据,并且在很多时候,投资者在互联网空间上更会真实地表达自己的情绪和观点。更为重要的是,这些映射到互联网空间的各种数据、信息、图像和文笔等海量样本,可以通过计算机技术直接进行收集与分析。金融学者利用互联网上所形成的海量数据,可以从虚拟的互联网空间去捕捉、刻画现实世界中的投资者个性特征和行为特点,并纳入到现代金融学各个领域的研究中,丰富研究思路。

三、非结构化数据的量化分析成为金融学研究新热点

量化分析(Quantitative analysis)又称定量分析,是指人们可以使用具体的数据表示一些模糊、抽象的因素,以便用来进行分析或运算。在大数据时代,一切皆可量化,这是大数据思维最核心的理念之一。更大范围更广思路的量化分析也成为西方金融学研究的新趋势。

(一)投资者情绪量化分析

投资者情绪(Investor Sentiment)是有限套利环境下影响资产价格的系统性风险,是影响金融市场波动的重要因素之一。但对于如何测度投资者情绪,学术界一直存在着争议。在较早的研究中,对于

投资者情绪的测度往往采用间接法,例如使用封闭式基金折价(CEFD)来表示投资者情绪(Lee et al, 1991; Neal & Wheatley, 1998)。近年来广泛使用的是 Baker & Wurgler(2006)基于封闭式基金折价、交易量、IPO 数量及上市首日收益率、股利收益和股票发行/证券发行比例共 6 个单项情绪指标,运用主成分分析构造了度量投资者情绪的复合指数(后文简称 BW 投资者情绪指数)。他们也指出,这一领域更为严谨深入的研究关键在于寻找更全面、更即时的指标更好地测度市场上的投资者情绪。有学者基于互联网空间的海量数据,开始尝试使用各种更为直接的测度指标来表达投资者情绪;有学者尝试利用媒体新闻内容捕捉市场上的投资者情绪。Tetlock(2007)对《华尔街日报》各专栏的新闻报道进行文本分析,同样运用主成分分析法构建了一项指标来衡量媒体报道内容的消极程度,并以该指标表示投资者情绪。作者在此基础上研究发现,新闻报道的内容越消极,股票市场价格下行的压力会越大,但是公司股价随后发生反转的可能性也越大。同时,当新闻报道消极程度出现异常时,市场交易量也会随之大幅放大。随后的文献中, Garcia(2012)用《纽约时报》中“金融市场”(Financial Markets)版块和“华尔街话题”(Topics in Wall Street)版块中的积极词汇数量和消极词汇数量来构建投资者情绪指标,实证研究发现,这一指标在经济衰退时能够更好地预测股票市场的走势。

搜索引擎可谓网民获取信息的首要方式。投资者同样会通过搜索引擎搜寻相关的市场信息和公司信息,因此搜索引擎的搜索量也能在一定程度上反映投资者情绪的变化。为了验证这一猜想, Da et al(2015)根据 Google 搜索主要词汇的结果,构建了 FEARS (Financial and Economic Attitudes Revealed by Search)指标来表示国民的投资情绪。他们通过分析 2004 至 2010 年间的相关数据研究发现, FEARS 指标能很好地预测股票市场波动,这表明 FEARS 指标很好地量化了投资者情绪。近年来,随着投资者中网民比例的提升,投资者的个性特征、情绪态度和交易行为越来越多地真实且即时地映射到互联网空间,形成大量的结构化和非结构化的海量数据。大数据处理技术又使得这些信息和数据的处理成了可能,这也就使得越来越多的西方金融学者使用互联网数据来直接量化投资者情绪(Karabulut, 2013; Bollen et al, 2011)。

(二)外貌形象量化分析

自 Hamermesh & Biddle(1994)研究发现长相

更出众的员工往往能够获得更高的收入以来,这一领域一直吸引着不少学者的研究兴趣。然而,个人形象往往是图片格式的非结构化数据,难以批量处理和量化处理,只能依托人工评判和手工处理,除了标准难以统一之外,处理速度和处理能力上也受到很大的局限。

在较早的研究中,对于个人形象这种非结构化数据,通常是根据心理学(Cunningham et al, 1990)的方法,随机选择志愿者对照片进行人工打分。对于这一种传统方法, Pareek & Zuckerman(2013)可谓做到了极致。他们将基金经理的照片都进行了统一处理,选择相同类型照片(包括背景相同、表情类似等),并统一转换为黑白照和相同的像素水平,然后交由耶鲁大学电子实验室 25 位随机挑选的志愿者对照片上人的长相进行独立打分。他们在此基础上研究发现,长相得分高的基金经理往往能够获得更多的资金净流入,也更容易在激烈的市场竞争中存活下来,但这并不意味着这些基金经理拥有更高的投资水平。但这样的相貌随机量化打分法还是受制于 25 个志愿者的审美偏好。为了扩大评分人的群体数量, Duarte et al(2012)编写特殊的计算机程序(Amazon's Mechanical Turk)来对借款人的形象进行量化打分。通过该网络程序,打分人可以是任意一位美国公民,并且遍布各个阶层与族裔。这一网络公众随机打分方法将评判人数扩大数十倍,从而将个人的主观审美偏好降到了最低。他们研究发现,长得更帅或更漂亮的借款人能够更容易获得银行贷款,并且所要求的利息也会更低。尽管这一方法尽量降低了个人审美偏好的影响,但人工打分始终无法完全避免主观因素的影响,需要更为客观的方法进行量化。

随着计算机模式识别技术的发展,参照生物统计学家对人体外貌分析的研究结果,我们完全可以使用计算机程序等对个人照片进行自动打分(Schmid et al, 2008)。这种全自动化、纯标准化打分模式可以完全避免个人偏好的影响。在最新的研究中,有些西方学者也将这一技术应用于金融学问题的研究。Halford & Hsu(2014)在分析高管外貌形象对公司价值的影响时,运用“Google 图片”搜集到 2000 年以来标准普尔 500 强上市公司的 677 位 CEO 的个人照片,并将所有照片上传至 anaface.com 这一专业网站,通过计算机模式识别技术统一给这些 CEO 的个人照片进行打分。他们根据该网站的程序计算结果构建了 CEO 面部吸引力指数(Facial Attractiveness Index)进行实证分析。他们研究发现,形象更好即面部吸引力

四、结束语

指数更高的 CEO 在上任当天,其所在公司的股价上涨更多,收益率更高;当该企业宣布进行并购时往往也能带来更高的收益。他们还结合新闻媒体数据分析得出更精确的结论,形象好的 CEO 出现在电视新闻中会比其出现在报刊新闻中给任职企业带来更多的收益,这也证明了美丽溢价(Beauty Premiums)的存在性。尽管在上述研究中,采用人工智能自动打分的与研究结果与志愿者人工打分的结果基本保持一致,但前者的可靠性程度显然会更高些。

(三)其他量化分析

Da et al(2011)运用 2004 至 2008 年间 Google 搜索引擎上上市公司相关词汇搜索频率(SVI)的变化来衡量投资者关注点的变化。这种测量方法能够更加及时、有效地捕捉到投资者关注点的变化,尤其是散户投资者。他们研究发现,某上市公司的 SVI 增加预示着未来 2 周内该公司股价将会上升,但一年内也会出现最终的价格反转。所构造的 SVI 这一指标也有助于解释上市公司 IPO 当天的较高溢价和随后的股价逆转,长期表现不佳的问题。他们认为 SVI 指标与过去的投资者关注度代理指标相比,两者既相似又有所区别——SVI 指标能够在时间趋势上更加精准地测度投资者关注度,对中小投资者关注度的量化测度也会更加精确。沿着这一思路,Chi & Shanthikumar(2014)将地理区域的概念融入到了 SVI 指标中,通过统计 Google 搜索的 IP 地址确定该投资者的所在州(State),并根据每个州的地理中心点测量投资者距离。他们研究发现,“本地偏好”(Local Bias)现象在投资者对上市公司的互联网搜索行为同样普遍存在,并且名气越大的公司这一现象越明显。

Mironov(2014)在研究公司高管腐败问题时构建了“腐败倾向”(PTC, Propensity To Corrupt)指数来表示高管发生腐败问题的可能性,并从俄罗斯中央银行(Russian Central Bank)和俄国相关政府机构泄露的数据中收集了首都莫斯科市所有市民的银行交易数据、个人收入数据、交通违章数据等,其中包括 1997 至 2007 年间在莫斯科市发生的近 680 万次交通违章和近 16 万次交通事故数据。随后作者通过模型设定对莫斯科市区近 275 万驾驶员的违章倾向进行了统计,并筛选出企业高管的相关数据,这包括莫斯科市所有 58157 家私人持股公司的管理人员。作者研究发现,腐败倾向(PTC)更高的高管收入水平往往也更高,但不可思议的是,其所任职公司的经营业绩也会比同行业企业更好。

本文通过梳理近年来西方金融学界文献,总结了大数据思维对金融学研究产生的影响及应用。总体来看,影响主要体现在:研究的样本容量在不断扩大,从过去的随机样本或小样本研究转变成成为大样本或全样本研究,这与大数据思维中数据的大容量特征相吻合;同时受到大数据的多样性特点影响,金融学研究正逐渐走向多元化,研究的内容也将持续扩大。

大数据处理技术也为金融学研究提供了更多的应用领域。媒体效应分析、网络数据处理作为金融学研究领域近年来兴起的热点,已经取得了丰硕的研究成果,这无疑得益于数据类型的丰富和处理技术的进步。可以确定的是,伴随着大数据思维的推广与运用,未来将有更多的应用领域等待学者们的探索。除了研究领域扩大,大数据思维还有助于提升研究方法的应用。量化分析一直是金融学界与金融业界的重要分析方法,在大数据环境下,这一思路得到了前所未有的发展。海量、多样化、高实效性的数据,让更大范围更广思路的量化分析在金融学研究的应用中不断扩展,已经成为研究的一项基本思路。

大数据思维对金融学研究的影响通常是多方面、多层次的,而非单一的从某一个方面去影响研究过程。无论是研究内容的确定、数据样本的选择还是模型回归的方法,从近两年西方金融、经济顶级期刊发表的最新文章来看,已经有很大一部分金融学者将大数据的思维方式渗透在整个研究中(Karabulut, 2013; Halford & Hsu, 2014; Dang et al, 2014)。

借助于大数据的数据特征和相关处理技术,金融学研究能够更多地使用实证模型去验证理论假设,严谨程度得到进一步提升。目前,大数据思维在现代金融研究中的运用,在西方金融学界可谓如火如荼,日新月异,国内金融学界也应跟上大数据时代的脚步,主动开启大数据思维。

注:

- ① $1\text{ZB}=2^{10}\text{EB}=2^{20}\text{PB}=2^{30}\text{TB}=2^{40}\text{GB}=2^{50}\text{MB}=2^{60}\text{KB}=2^{70}\text{Bytes}$,即 2 的 70 次方字节。
- ② 月球会合周期即“朔望月”,又称“太阴月”,为月球绕地球公转相对于太阳的平均周期,平均时间为 29.53059 天。其中从地球上观测月球最暗的一天称为“新月”(我国农历初一日),此时月球正好处于地球与太阳之间,三者成一条直线;而地球上观测到的月球最亮一天称为“满月”(我国农历十五或十六),此时地球正好处于月球与太阳中间,三者成一条直线。
- ③ 月球四相位包括新月、上弦月、满月、下弦月。
- ④ Edmans et al(2007)还发现板球(cricket)、橄榄球(Rug-

by)、冰球(Ice Hockey)、篮球(Basketball)的比赛结果均能够对股市产生影响。

- ⑤文本分析是指通过对文本内容进行挖掘和数据分析,以推断文本作者的目的和意图。由于文本属于非结构化数据,文本分析必须将它们从一个无结构的原始文本转化为结构化的计算机可以识别处理的信息,即对文本进行科学的抽象,建立它的数学模型,用以描述和代替原始文本。
- ⑥自然语言处理是指研究能实现人与计算机之间用自然语言进行有效通信的各种理论和方法。自然语言处理并不是一般地研究自然语言,而在于研制能有效地实现自然语言通信的计算机系统,特别是其中的软件系统。因而它是计算机科学的一部分。

参考文献:

- Adams, R. B. & D. Ferreira (2009), "Women in the board room and their impact on governance and performance", *Journal of Financial Economics* 94(2):291-309.
- Baker, M. & J. Wurgler (2006), "Investor sentiment and the cross-section of stock returns", *Journal of Finance* 61(4):1645-1680.
- Adams, R. B. et al (2005), "Powerful CEOs and their impact on corporate performance", *Review of Financial Studies* 18(4):1403-1432.
- Barber, M. B. & T. Odean (2001), "Boys will be boys: Gender, overconfidence, and common stock investment", *Quarterly Journal of Economics* 116(1):261-292.
- Benmelech, E. & C. Frydman (2014), "Military CEOs", *Journal of Financial Economics*, forthcoming.
- Bernstein, P. (1992), *Capital Ideas: The Improbable Origins of Modern Wall Street*, Free Press.
- Bertrand, M. & A. Schoar (2003), "Managing with style: The effect of managers on firm policies", *Quarterly Journal of Economics* 118(4):1169-1208.
- Bollen, J. et al (2011), "Twitter mood predicts the stock market", *Journal of Computational Science* 2(1):1-8.
- Bolton, P. et al (2008), "Leadership, coordination and mission driven management", *Working Paper*.
- Chen, H. et al (2014), "Wisdom of crowds: The value of stock opinions transmitted through social media", *Review of Financial Studies* 27(5):1367-1403.
- Chi, S. & D. Shanthikumar (2014), "The geographic dispersion of Google search and the market reaction to earnings announcements", *Working Paper*.
- Cunningham, M. R. et al (1990), "What do women want? Facial metric assessment of multiple motives in the perception of male facial physical attractiveness", *Journal of Personality and Social Psychology* 59(1):61-72.
- Da, Z. et al (2011), "In search of attention", *Journal of Finance* 66(5):1461-1499.
- Da, Z. et al (2015), "The sum of all fears: Investor sentiment and asset prices", *Reviews of Financial Studies* 28(1):1-32.
- Dang, L. T. et al (2014), "Commonality in news around the world", *Journal of Financial Economics*, forthcoming.
- Davis, A. K. & I. Tama-Sweet (2012), "Managers' use of language across alternative disclosure outlets", *Contemporary Accounting Research* 29(3):804-837.
- Demers, E. & C. Vega (2011), "Linguistic tone in earnings press releases: News or noise", *Working Paper*.
- Dichev, I. D. & T. D. Janes (2003), "Lunar cycle effects in stock returns", *Journal of Private Equity* 6(4):8-29.
- Duarte, J. & L. Young (2009), "Why is PIN priced?", *Journal of Financial Economics* 91(2):119-138.
- Duarte, J. et al (2012), "Trust and credit: The role of appearance in peer-to-peer lending", *Review of Financial Studies* 25(8):2455-2484.
- Easley, D. et al (1996), "Liquidity, information, and infrequently traded stocks", *Journal of Finance* 51(4):1405-1436.
- Easley, D. et al (2002), "Is information risk a determinant of asset returns?", *Journal of Finance* 57(5):2185-2221.
- Edmans, A. et al (2007), "Sports sentiment and stock returns", *Journal of Finance* 62(4):1967-1998.
- Engelberg, J. & C. Parsons (2011), "The causal impact of media in financial markets", *Journal of Finance* 66(1):67-97.
- Faccio, M. et al (2014), "CEO gender and corporate risk-taking", *Working Paper*.
- Frydman, C. et al (2014), "Using neural data to test a theory of investor behavior: An application to realization utility", *Journal of Finance* 69(2):907-946.
- Garcia, D. (2012), "Sentiment during recessions", *Journal of Finance* 68(3):1267-1300.
- Goetzmann, W. N. et al (2015), "Weather-induced mood, institutional investors, and stock returns", *Review of Financial Studies* 28(1):73-111.
- Gompers, P. A. et al (2012), "The cost of friendship", *Working Paper*.
- Graham, J. R. et al (2010), "A corporate beauty contest", *Working Paper*.
- Graham, J. R. et al (2013), "Managerial attitudes and corporate actions", *Journal of Financial Economics* 109(1):103-121.
- Grossman, S. & J. Stiglitz (1980), "On the impossibility of informationally efficient markets", *American Economic Review* 70(3):393-408.
- Gurun, U. G. & A. W. Butler (2012), "Don't believe the hype: Local media slant, local advertising, and firm value", *Journal of Finance* 67(2):561-597.
- Halford, J. T. & H. Hsu (2014), "Beauty is wealth: CEO appearance and shareholder value", *Working Paper*.

- Hamermesh, D. S. & J. E. Biddle(1994), "Beauty and the labor market", *American Economic Review* 84(5):1174—1194.
- Hayashi, F. (2000), *Econometrics*, Princeton University Press.
- Hillert, A. et al(2014), "Media makes momentum", *Review of Financial Studies* 27(12):3467—3501.
- Hirshleifer, D. & T. Shumway(2003), "Good day sunshine: Stock returns and the weather", *Journal of Finance* 58(3):1009—1032.
- Hirshleifer, D. (2001), "Investor psychology and asset pricing", *Journal of Finance* 56(4):1533—1597.
- Hirshleifer, D. (2010), "Self-enhancing transmission bias and active investing", *Working Paper*.
- Hubbard, D. W. (2011), *How to Measure Anything: Finding the Value of Intangibles in Business*, Wiley.
- Jegadeesh, N. & D. Wu(2013), "Word power: A new approach for content analysis", *Journal of Financial Economics* 110(3):712—729.
- Kahneman, D. & M. W. Riepe(1998), "Aspects of investor psychology", *Journal of Portfolio Management* 24(4):52—65.
- Kaplan, S. N. et al(2012), "Which CEO characteristics and abilities matter?", *Journal of Finance* 67(3):973—1007.
- Karabulut, Y. (2013), "Can Facebook predict stock market activity?", *Working Paper*.
- Kelly, I. W. et al(1998), "The moon was full and nothing happened: A review of studies on the moon and human behavior and human belief", *Skeptical Intelligencer* 2(3/4):22—34.
- Lai, S. et al(2014), "Does PIN affect equity prices around the world?", *Journal of Financial Economics* 114(1):178—195.
- Lee, C. et al(1991), "Investor sentiment and the closed-end fund puzzle", *Journal of Finance* 46(1):75—109.
- Loughran, T. & B. McDonald(2013), "IPO first-day returns, offer price revisions, volatility, and form S-1 language", *Journal of Financial Economics* 109(2):307—326.
- Lour, S. (2012), "The Age of Big Data", *New York Times*.
- Malmendier, Ü. et al(2011), "Overconfidence and early-life experiences: The effect of managerial traits on corporate financial policies", *Journal of Finance* 66(5):1687—1733.
- Mayer-Schönberger, V. (2012), *Big Data: A Revolution that will Transform How We Live, Work, and Think*, John Murray Publishers.
- Mayew, W. J. & M. Venkatachalam(2012), "The power of voice: Managerial affective states and future firm performance", *Journal of Finance* 67(1):1—43.
- McAfee, A. & E. Brynjolfsson(2012), "Big Data: The management revolution", *Harvard Business Review* 90(10):60—6,68,128.
- Mironov, M. (2014), "Should one hire a corrupt CEO in a corrupt country?", *Journal of Financial Economics*, forthcoming.
- Neal, R. & S. M. Wheatley(1998), "Do measures of investor sentiment predict returns?", *Journal of Financial and Quantitative Analysis* 33(4):523—547.
- Pareek, A. & R. Zuckerman(2013), "Trust and investment management: The effects of manager trustworthiness on hedge fund investments", *Working Paper*.
- Peress, J. (2008), "Media coverage and investors' attention to earnings announcements", *Working Paper*.
- Peress, J. (2014), "The media and the diffusion of information in financial markets: Evidence from newspaper strikes", *Journal of Finance* 69(5):2007—2043.
- Peress, J. & D. Schmidt(2014), "Glued to the TV: The trading activity of distracted investors", *Working Paper*.
- Rotton, J. & M. Rosenberg(1984), "Lunar cycles and the stock market: Time-series analysis for environmental psychologists", *Working Paper*.
- Saunders, E. M. (1993), "Stock prices and Wall Street weather", *American Economic Review* 83(5):1337—1345.
- Schmid, K. et al(2008), "Computation of a face attractiveness index based on neoclassical canons, symmetry, and golden ratios", *Pattern Recognition* 41(8):2710—2717.
- Simon, D. & R. Z. Heimer(2012), "Facebook finance: How social interaction propagates active investing", *Working Paper*.
- Sornette, D. (2014), "Physics and financial economics (1776—2014)", *IOP Science* 77(6):1—28.
- Sundén, E. A. & B. J. Surette(1998), "Gender differences in the allocation of assets in retirement savings plans", *American Economic Review* 88(2):207—211.
- Tetlock, P. C. (2007), "Giving content to investor sentiment: The role of media in the stock market", *Journal of Finance* 62(3):1139—1168.
- Tetlock, P. C. et al(2008), "More than words: Quantifying language to measure firms' fundamentals", *Journal of Finance* 63(3):1437—1467.
- Tetlock, P. C. (2010), "Does public financial news resolve asymmetric information?", *Review of Financial Studies* 23(9):3520—3557.
- Wann, D. et al(1994), "Relationships between spectator identification and spectators' perceptions of influence, spectators' emotions, and competition outcome", *Journal of Sport and Exercise Psychology* 16:347—364.
- Yuan, K. et al(2006), "Are investors moonstruck? Lunar phases and stock returns", *Journal of Empirical Finance* 13(1):1—23.
- Zhou, D. (2014), "The blame game", *Working Paper*.

(责任编辑:李仁贵)