

行业资产的 Black-Litterman 模型配置研究^①

——基于社交网络情绪文本挖掘算法

孟 勇¹ 任 梦² 赵 心³

(1. 山西财经大学统计学院; 2. 山西财经大学财政与公共经济学院;
3. 山西财经大学金融学院)

研究目标: 构建反映行业股价走势的基于社交网络文本挖掘算法的行业投资者情绪指标, 并改善嵌入行业投资者情绪指标的 Black-Litterman 模型对资产的配置结果。研究方法: 基于社交网络文本挖掘算法度量投资者情绪, 运用主成分分析法构建行业投资者情绪指标, 并嵌入 Black-Litterman 模型中构建投资者观点矩阵, 确定行业资产配置比。研究发现: 基于行业投资者情绪的 BL 模型有效提高了资产配置的日均收益率和夏普比率。实证结果在样本外验证(除受新冠疫情影响阶段)、暴涨暴跌阶段以及经过允许卖空和交易成本调整后仍稳健, 进而证实了投资者情绪对资产组合有显著影响。研究创新: 基于社交网络文本挖掘算法构建投资者情绪指数, 解决了仅依赖于预期收益或历史数据的预测模型无法直观揭示投资者心理认知和行为的局限性问题, 从一个崭新的视角科学地解决 Black-Litterman 模型中投资者观点的生成问题。研究价值: 扩展了 Black-Litterman 模型理论体系研究, 并推动了行为金融理论在资产配置中的应用。

关键词 投资者情绪 Black-Litterman 模型 行业资产配置 文本挖掘

中图分类号 F275.0; F832.51 文献标识码 A

DOI:10.13653/j.cnki.jqte.2022.01.002

引 言

资产组合的选择是指将投资者的资本优化分配给不同的金融资产, 以降低证券组合的非系统性风险, 最大限度地获取预期收益。Markowitz (1952) 开创性地对风险进行了数学化的规范描述, 提出了资产组合选择的均值-方差模型 (MVT), 也就是在投资回报和风险之间进行权衡选择投资组合, 这一模型是现代投资组合理论的基础。但 MVT 设定了较为严苛的假设, 包括市场没有达到强势有效、投资者希望方差越小越好以及只关心资产的期望收益和方差等, 这些假设与实际的金融环境存在差异, 使模型存在不足, 模型对参数的变动较为敏感。VaR (Jorion, 1996) 和 CVaR (Rockafellar 等, 2000) 是这一方向风险度量的代表, 学者们提出并分析了不同投资情况下的 VaR 模型 (Berkowitz 等, 2011) 和 CVaR 模型 (Rockafellar 等, 2002)。

随着资产组合模型的发展和实践, 学者们逐渐认识到, 投资者倾向于关注自身认为被低

^① 本文获得国家社会科学基金项目“超高维稀疏网络模型及其组合风险应用研究”(19BTJ026)资助。

估但有发展潜力的股票或资产，也就是依赖自己的心理感知选择资产组合，投资者情绪在金融市场中的潜在作用得到了经济学家的极大关注。大量研究表明投资者在不确定状态下的投资决策是非理性的，非理性行为是投资者认知因素（Ellis 和 Harper，1975）以及文化环境因素等多种因素复杂作用的结果。John Maynard Keynes（1936）在解释股票市场异常时提到了“动物精神”，荣获 1978 年诺贝尔经济学奖的 Simon（1976）提出“有限理性”假定，行为主体的心理因素影响这一决策。1979 年，Kahneman 和 Tversky 发表 *Prospect Theory: An Analysis of Decision under Risk* 一文，处于行为金融理论研究的“里程碑”地位。Kahneman 等（1982）认为行为人是存在“认知偏差”的，将启发式与偏差分类为因果关系和归因、过度自信、保守主义、代表性以及判断偏差等，这显然与传统金融学范式的假定相悖。

Kahneman 作为一名有影响力的心理学家，第一次从理论上将心理学和经济学研究结合在一起，这一突破性的研究使其获得 2002 年诺贝尔经济学奖。2017 年，行为经济学大师 Richard H. Thaler 获得诺贝尔经济学奖，其对于行为金融理论的发展做出了极大贡献，他对噪声交易者模型的许多附加预测做了检验（Lee 等，1991），研究证实封闭式基金是特定类型的噪声交易者情绪的衡量指标。模型证明金融市场存在一种反馈机制，即认知和行为偏差、投资决策偏差、价格偏离市场价值间的反馈。De Long 等（1990）首次引入“非理性噪声交易者”构建资产定价行为模型，正式证明了噪声交易者情绪与资产价格之间的关系。大量投资实践活动也证实投资决策活动是客观环境与主观心理活动共同作用的产物，决策主体的情绪、价值观、个人偏好等非理性因素影响投资决策。“投资者面临的首要问题，乃至投资中最大的敌人就是自己。”华尔街教父 Graham 曾说。无论是投资理论还是实践都表明人的情绪、性格和情感等自身固有的各种认知偏差会对行为人的决策构成重要的影响，研究金融市场的新方法——行为金融学的发展也证实了这一点。

基于这些认识，1992 年，美国华尔街著名的高盛公司以 Markowitz 模型的构建思路为基础，将模型改进，提出 Black-Litterman 模型。该模型的贡献之处在于引入了投资者对某资产的看法（主观观点），采用贝叶斯法则将先验预期收益率与投资者的主观意见结合，计算新的预期收益率，解决了均值一方差模型配置头寸过于集中及对输入参数过于敏感等问题。但是 BL 模型的提出也引发了一个问题——如何更为精确地度量投资者主观观点？金融市场的复杂性和动态性以及投资者的行为偏差等引起了研究者对生成观点矩阵过程的关注，Black 和 Litterman（1991、1992）研究的投资者观点主要指具有专业背景和实战经验的投资专家观点。基于此，研究者强调投资者观点的生成问题，旨在使用预期收益或历史数据，运用不同的预测模型构建更精确或无偏差的投资者观点。Beach 和 Orlov（2007）作为早期的研究者，使用计量方法——GARCH 模型构建投资者观点。学习算法和仿真方法也被用于生成投资者观点。Didenko 和 Demicheva（2013）采用随机森林集成学习算法生成投资者观点，以整合到美国股票市场的不同类型的投资组合优化中。Chiarawongse 等（2012）基于马尔可夫链蒙特卡罗（MCMC）的模拟方法形成对市场的定性观点。Creamer（2015）将投资者根据市场数据和财报数据形成的观点融合到 BL 模型中，并给出数值算例。基于混合方法来预测收益的研究在近几年成为趋势，并证明了其有效性与准确性。Kim 和 Won（2018）将长短期记忆（LSTM）模型与几种类型的 GARCH 模型融合，实证结果表明，混合模型比单个计量经济学模型“显著提高”了预测性能。Kara 等（2019）提出一种混合方法，即采用 GARCH 模型预测股票指标，再通过支持向量回归模型将指标预测转换为收益预测，最终将收益转换为投资者看法向量引入 Black-Litterman 模型中。然而，上述模型在构建投资

者观点时存在局限性——忽略了投资者的认知偏差等因素，仅假定预测模型产生数据。以上的方法只是对预期收益或历史数据预测进而确定投资者观点的改进，但预期收益或历史数据无法直观地揭示投资者的心理认知和行为以及现实情况。

在投资者情绪的研究中，学者对情绪较为统一的定义是：投资者受市场波动、心理等因素影响，对资产定价持有主观偏差的投资者预期或者投资者信念。国内外的文献集中于构建以下两种不同的投资者情绪度量指数。研究初期，学者们构建基于主观调查的直接投资者情绪指数和基于客观市场的间接投资者情绪指数。由于传统情绪指标存在显著的局限性，即调查数据频率较低、样本容量有限或受访者由于受社会道德等的影响隐藏真实看法，Baker 和 Wurgler (2007) 采用主成分因子分析法——自上而下 (Top Down) 的分析法合成一个单一的总体情绪变量。采用自上而下的模式是因为其优点是以一种简单、更为直观方式融入了泡沫、危机、眼花缭乱的股价变化模式和变幻莫测的投资者情绪等。基于主成分分析法构建投资者情绪指数成为了国内外核心金融期刊的主要方法，学者们根据所研究内容的不同，选取不同的符合情绪指数的源指标。但是该方法除反映投资者情绪外，还反映情绪与其他经济因素相互作用后的均衡结果 (Da 等, 2015)，换句话说，人们永远无法确定某个特定的市场指标是否受到投资者关注的驱动。为排除经济因素对情绪构建的影响，以及随着大数据时代的到来，学者们采用爬虫技术从搜索引擎数据、媒体报道和报纸文章数据以及社交平台数据等不同的文本资源中爬取情绪，并构建有效的投资者情绪指标。与调查法和市场交易数据的情绪度量方法相比，基于社交网络的文本挖掘爬取的数据更加透明、更为真实，这一方法能够客观抓取更多的个人有用信息。网络上发布的与股票相关的帖子、微博等包含了投资者对股票市场的分析、投资决策的观点以及对当前市场状态的评论等。数据挖掘是一种具有高速性、多样客观性和规模性的理想情绪度量方法。

虽然研究方法上不断提升，但研究内容上，学者们更多从金融市场整体层面构建投资者情绪指数。随着金融市场不确定性的增加，投资组合相关研究逐渐重视行业效应的影响，Salhin 等 (2016) 从行业层面检验市场情绪的作用，从行业层面构建投资者情绪指数具有重要的投资现实意义 (向诚和陆静, 2018)。各行业在不同信息的冲击下个人投资者情绪表现是否一致？是否会影响投资决策？这些问题还没有得到很好的解释。国内只有极少数学者对具体的行业投资者情绪进行了部分研究，卢米雪和朱喜安 (2014) 将封闭式基金折价率、交易量、波动率、行业 beta 值和每股流通市值等基本面作为情绪代理变量，运用主成分分析法分别构建 23 个行业的投资者情绪指数，并分析了各行业的投资者情绪对收益的影响效应。结果表明，投资者情绪对当期收益和预期收益的影响具有显著的行业差异性。

基于文本挖掘技术的行业投资者情绪指标是否更好地反映行业股价走势？投资者情绪是否影响股市的资产配置？配置结果是否得到明显改善？为解决以上问题，本文依据已有的理论和现实背景，基于数据挖掘技术，爬取文本数据，构建各行业文本情绪指标。按行业对资产进行分类以简化投资决策，将得到的行业投资者情绪作为 BL 模型的观点矩阵，最终得出行业间的最优资产配置比。本文深入更为细分的行业层面，将构建的我国行业投资者情绪指标引入 Black-Litterman 模型，生成投资者观点矩阵。这一改进为构建 BL 模型的主观观点开辟了一条新的路径，以情绪指标更为真实地揭示投资者对股票市场的观点，方法更为科学，从而提升 BL 模型资产配置的收益率，同时也对分析师主观态度难于量化的问题进行了补充。本文的主要创新体现在：第一，构建了加入文本情绪指标的行业投资者情绪指标，发现绝大部分的行业情绪指标能够反映该行业股价走势的动态。第二，将投资者情绪指标和

Black-Litterman 资产配置模型结合起来,以行业投资者情绪指标创建 BL 模型中的投资者主观观点矩阵,从具体配置结果来看,改进的 BL 模型比 Markowitz 模型在日预期收益率、夏普比率及资产配置比维度上得到明显改善。上述发现在考虑了允许卖空、交易成本和股市暴涨暴跌阶段后仍然稳健。第三,本文细分到中国各行业股票市场,给出了新兴股票市场投资者情绪在资产配置中具有重要影响的经验数据。

一、行业投资者情绪指标构建

1. 行业投资者情绪指标构建的研究设计

本文构建行业投资者情绪指标采用主成分分析法。主成分分析法是一种降维方法,即把原先多个特征用更少的特征数取代表征绝大部分的信息。本文行业投资者情绪指标的代理变量包括成交量 CJL 、成交额 CJE 、换手率 HSL 、涨跌幅 ZDF 、市净率 PB 、市盈率 PE 和文本情绪指标 $TSENT$ 。通过代理变量与收益率 Granger 因果关系检验保留或剔除变量。

第一主成分、累积方差贡献率 80% 以及 eigenvalue 大于 1 准则作为提取主成分个数的依据。Baker 和 Wurgler (2007) 只选取第一主成分的因子组成来构建情绪指标。易志高和茅宁 (2009) 构建情绪指标同样采用主成分法,但依据准则同 BW 不同,采用累积方差贡献率 85%,对主成分进行特征值的加权平均,所得结果为投资者情绪指标。本文采用累积方差贡献率 85% 准则进行主成分分析。

Baker 和 Wurgler (2007) 的研究认为,投资者情绪在引导其本人决策时,也会受到宏观经济因素的影响,即理性情绪。通过多元回归将这一理性预期成分剥离出来,仅保留非理性情绪。但由于中国股市存在的特有性,本文认为,投资者情绪指标构建中不应剔除宏观经济因素。《上海证券交易所统计年鉴 (2018~2019 年)》数据显示,在中国股市中,中国股市自然人投资者持股账户数与机构投资者比例为 99.78 : 0.12,呈现出自然人投资者占比大而机构投资者占比少的结构,而散户一般对风险的认识不足,无法像机构投资者充分利用信息来进行交易,容易受到宏观经济因素的影响,从而导致股市出现系统性偏差。因此,受宏观经济基本面因素影响的理性情绪也是投资者情绪的一部分,因此不剔除各投资者情绪代理变量中所包含的宏观因素。

2. 情绪代理变量的解释

代理变量的合理选取是能否合理构建投资者情绪指标的基本前提。但投资者情绪会受到社交网络传导的影响,进一步影响股市的各基本面。本文基于学者们对投资者情绪代理变量选取的研究 (Baker 和 Wurgler, 2007; Ho 和 Hung, 2009; Kurov, 2010; Baker 等, 2012; Stambaugh 等, 2012; Corredor 等, 2013),并结合中国股票市场及各行业的发展现状和数据的可获得性,选取成交量 CJL 、成交额 CJE 、换手率 HSL 、涨跌幅 ZDF 、市净率 PB 、市盈率 PE 和文本情绪指标 $TSENT$ 等 7 个投资者情绪代理变量,以上代理变量间接包含了投资者情绪的某些成分。从交易行为角度来看,股票的成交量 CJL 、成交额 CJE 、换手率 HSL 等指标直接反映了投资者的交易热度,是投资者市场观点较为有效的反映。涨跌幅 ZDF 这一指标是一种市场表现,情绪波动激发投资者的交易行为,最终反映在股市中,因此,涨跌幅可以反向追踪投资者情绪。市盈率 PE 和市净率 PB 是较为常见的评估股价指标,高 PE 低 PB 的股价趋势,间接捕捉到投资者高涨的情绪。前 6 个指标是测度投资者情绪的客观情绪指标,文本情绪指标 $TSENT$ 是唯一的主观情绪指标,将更好地捕捉即时的投资者情绪。行业投资者情绪将在 7 个代理变量中有所体现。

(1) 行业日度股票成交量 CJL 及日度股票成交额 CJE 。成交量在某种程度上可以代表股市的流动性,往往随着股价的变化而变化,该值反映投资意愿落实到交易行动上的量值,即投资者的参与程度。成交额同样反映出投资者的交投热度。当投资者情绪高涨或对市场越乐观时,提高了对股票投资的活跃度,成交额也增加。

(2) 行业日度平均换手率 HSL 。换手率意味着市场中股票流通性和投资者交投的热情。当投资者对股市未来预期较为悲观时,其往往更愿意持有资金,参与投资的情绪低落,股票换手率降低。反之,投资者较为乐观时,情绪更为高涨,股票换手率大大增加。

(3) 行业日度涨跌幅 ZDF 。涨跌幅为正时,说明市场结构越倾向于看好后市,投资者倾向于重复买入动作,情绪越高涨;反之,则越低迷。涨跌幅反映了投资者情绪的变化。

(4) 行业日度每股市净率 PB 。市净率反映出某一股票是否能以较少的投入获得较高的产出。市净率较低,意味着这一股票投资风险较低,投资价值较高,此时,投资者会更加关注这类股票。反之,市净率较高的股票投资者关注度降低。

(5) 行业日度每股市盈率 PE 。同市净率相反,市盈率较高的股票投资价值被高估,投资者的关注度相应提高;反之,投资者的关注度下降。

(6) 文本情绪指标 $TSENT$ 。用户的搜索行为和网络论坛的评论内容能直接反映用户自身的择股决策以及对相关政策的评价或市场状态的解读,这些文本易获得、实效性高,为构建有效的不同层面或频次的投资者情绪度量指标提供了数据源。计算贴吧文本中每日正向、负向情绪词语的总频数,构建文本情绪指标 $TSENT$,如式(1)所示。

$$TSENT_t = \frac{f_{pos,t} - f_{neg,t}}{f_{pos,t} + f_{neg,t}} \quad (1)$$

其中,第 t 日贴吧文本中正向情绪词的总数由 $f_{pos,t}$ 表示,第 t 日贴吧文本中反向情绪词的总数由 $f_{neg,t}$ 表示。文本情绪得分体现了投资者当日对各行业股票表达的积极或消极情绪,得分正(负)绝对值越高,表示投资者对股市的情绪越乐观(悲观)。

7个投资者情绪代理变量的选取存在主观性,没有对所选的变量的合理性进行筛选。据此,基于Wang等(2006)提出的投资者情绪代理变量筛选的优化程序,首先确定影响情绪的代理变量的范围,确定过程由于存在主观性问题进而需要进行合理性检验,即从相关性层面和Granger因果关系判断情绪代理变量选取的合理性,筛选出更适合反映行业投资者情绪的代理变量。

3. 数据来源

本文行业投资者情绪指标构建需要两类变量,一类是客观情绪代理变量,另一类是主观情绪代理变量。

客观情绪代理变量——成交量 CJL 、成交额 CJE 、换手率 HSL 、涨跌幅 ZDF 、市净率 PB 和市盈率 PE 数据源于Wind金融数据库,行业的分类标准依据Wind中的申银万国一级行业^①分类。选取能涵盖我国股票市场中价格波动较大和较小的2019年4月15日至2019年6月21日的更高频的日度数据作为样本期。为评价构建的投资者情绪指标解释能力的稳健性,本文选取2019年6月24日至9月27日、10月8日至12月31日和2020年1月2日至3月27

^① 申银万国一级行业包括采掘、化工、钢铁、有色金属、建筑材料、建筑装饰、电气设备、机械设备、国防军工、汽车、家用电器、纺织服装、轻工制造、商业贸易、农林牧渔、食品饮料、休闲服务、医药生物、公用事业、交通运输、房地产、电子、计算机、传媒、通信、银行、非银金融和综合,共28个行业。

日三个季度的日度数据以及 2015 年 3 月 16 日至 9 月 25 日股市大盘暴涨暴跌阶段进行检验。

主观情绪代理变量——选取东方财富网股吧作为投资者情绪的主观情绪代理变量的文本来源。东方财富网股吧是一种以话题为导向的社交平台，网站信息更具有权威性、全面性和准确性以及文本内容的可获得性，因此提高了信息挖掘的可操作性。在 R 语言中读取网址，选取同基本面相对应的每一个股中发布的贴吧作为金融文本数据。各行业在统计时间内的总贴吧量基本均在万条以上，将每一行业的总贴吧文本按日期归类，同一交易日的文本放入同一个 txt 文档中，非交易日文本放入前一交易日的 txt 文档中。然后对文本进行清洗去噪，清洗的目标是将数字、停用词、后缀、标点符号、空白符和特殊字符等一并去除，并进行词干化、词形还原、n-gram 分析，文本清洗有助于进行分词及情感分析。基于分词结果和课题《金融市场情绪指标构建》中通过人工标注及筛选建立的自定义情绪词典进行匹配，倾向性为正向的词语权重为 1，负向的词语权重为 -1。

4. 合理性检验

合理性检验的目的是从数据特征验证所选 7 个变量的合理性，并剔除这些代理变量所带有的主观成分，因为情绪的代理变量和收益率之间可能存在前者引起后者变化，或后者引起前者变化，或两者间相互影响，以往的研究大部分没有考虑这一问题。若代理变量变动是行业流动的原因，则保留该变量来构建情绪指数，反之则剔除。本文采用 Granger 因果关系检验进行合理性检验。

(1) 指标描述性分析。除选取代理变量外，在 Wind 数据库中提取统计时间内的申银万国一级行业第 t 日的收盘价 P_t ，并根据公式 $R_t = \ln(P_t/P_{t-1})$ 对应求各行业的对数收益率时间序列。限于篇幅，28 个行业的指标描述性统计结果无法全部列出，本文只列出采掘和化工两个行业，表 1 和表 2 列明采掘和化工两个行业投资者情绪代理变量的基本统计信息。

从表 1、表 2 的统计结果可以看出，由于量纲不同，各变量的均值和标准差差异较大，无法比较各变量的离散程度，使用两者的比值来对比分析。投资者情绪代理变量中涨跌幅 ZDF 的变异系数较大，波动幅度明显高于其他代理变量。文本情绪指标 $TSENT$ 的变异系数也相对较大。由 J-B 统计量的 p 值判断各变量是否满足正态分布，大部分变量不满足正态分布。ADF 单位根检验中，两个行业除市净率 PB 和市盈率 PE 外，其余代理变量均不存在较强的自相关性，对 PB 和 PE 变量选择一阶差分后的序列检验是否存在单位根，结果显示，均在 5% 的显著性水平上通过了 ADF 检验。

表 1 投资者情绪代理变量描述性统计（采掘）

变量	均值	标准差	偏度	峰度	J-B 统计量 Prop	ADF 检验	Prop
P_t	6.63	0.33	1.15	-0.17	0.0045	-1.4530	0.1398
R_t	0.00	0.02	-0.87	1.78	0.0010	-5.3936	0.0000
CJL_t	3527.77	1545.29	1.11	0.75	0.0028	-1.2459	0.0311
CJE_t	24483.58	11456.05	1.18	0.89	0.0011	-1.3500	0.0362
HSL_t	0.66	0.39	1.30	0.69	0.0005	-1.7872	0.0022
ZDF_t	-0.16	1.71	-0.81	1.47	0.0047	-5.1633	0.0000
PB_t	1.01	0.06	1.13	-0.17	0.0055	-1.8242	0.1351
PE_t	46.45	17.14	1.48	0.22	0.0001	-1.7556	0.2076
$TSENT_t$	0.24	0.25	-0.58	-0.65	0.1880	-2.1475	0.0193

表2 投资者情绪代理变量描述性统计 (化工)

变量	均值	标准差	偏度	峰度	J-B 统计量 Prop	ADF 检验	Prop
P_t	11.06	0.78	0.81	-0.83	0.0395	-1.4195	0.1922
R_t	0.00	0.02	-0.69	1.28	0.0180	-5.0705	0.0000
CJL_t	1804.90	657.51	0.69	-0.01	0.1426	-1.0397	0.0199
CJE_t	13491.88	5390.83	0.59	-0.53	0.2041	-1.1375	0.1122
HSL_t	2.36	0.95	0.75	0.13	0.0956	-0.8870	0.5866
ZDF_t	-0.17	1.97	-0.79	1.66	0.0027	-4.5268	0.0000
PB_t	1.83	0.12	0.83	-0.74	0.0405	-1.3489	0.1920
PE_t	43.14	5.40	1.39	0.48	0.0003	-1.2138	0.1948
$TSENT_t$	0.03	0.20	-0.15	-1.11	0.3327	-3.8839	0.0000

(2) 各行业投资者情绪代理变量的筛选及 KMO 检验和 Bartlett's 球形检验。平稳时间序列是格兰杰因果关系检验的必要条件。由本文的投资者情绪代理变量 ADF 检验结果显示原变量或一阶差分后的序列均为平稳时间序列, 因此使用 Granger 因果关系检验分别检验各行业代理变量与其对数收益率 (反映行业流动) 间的 Granger 因果关系, 部分行业结果由表 3 前 8 列所示, 未列明的行业 7 个投资者情绪代理变量均进行保留。

表3 部分行业投资者情绪代理变量筛选结果及 KMO 检验和 Bartlett's 球形检验结果

行业	CJL_t	CJE_t	HSL_t	ZDF_t	PB_t	PE_t	$TSENT_t$	KMO 检验	Bartlett's 球形检验
采掘	+	+	+	+	-	+	+	0.67	****
化工	+	+	+	-	-	+	+	0.68	****
有色金属	+	+	+	-	-	+	+	0.76	****
建筑装饰	+	+	+	+	-	+	+	0.58	****
电气设备	-	-	-	+	-	+	+	0.42	****
机械设备	+	+	+	-	+	+	-	0.77	****
汽车	+	+	+	+	-	+	+	0.74	****
家用电器	+	-	+	+	+	+	+	0.73	****
纺织服装	-	-	+	+	-	+	+	0.52	****
轻工制造	-	-	-	+	+	+	+	0.63	****
商业贸易	+	+	+	+	-	-	+	0.72	****
公用事业	+	+	+	+	-	+	+	0.71	****
房地产	+	-	+	+	-	+	-	0.67	****
电子	+	+	+	+	-	-	+	0.54	****
传媒	-	-	+	+	-	+	+	0.52	****
通信	+	+	+	-	-	+	+	0.59	****
非银金融	+	+	+	+	-	-	+	0.48	****
综合	+	+	+	+	-	+	+	0.67	****

注: “+”表示选择, “-”表示未选择, “****”表示小于 5×10^{-5} 。

基于表 3 的筛选结果可以看出, 机械设备和房地产行业的 Granger 因果关系检验结果显示, 文本情绪指标 $TSENT$ 不是这两个行业流通指数波动率的格兰杰原因, 进行剔除。其他行业根据 Granger 因果关系检验结果, 剔除不是行业流通收益率格兰杰原因的代理变量。

数据的标准化处理是进行数据分析的前提, 去除各代理变量的单位限制, 从而解决数据的可比性。代理变量很大程度不满足正态分布, 因此, 选择 0~1 标准化, 对标准化后的投资者情绪代理变量做 KMO 检验和 Bartlett's 球形检验。表 3 最后两列报告了部分行业的 KMO 检验和 Bartlett's 球形检验结果, 结果显示, 标准化后的数据 KMO 值符合 KMO 度量标准。Bartlett's 球形检验的伴随概率均小于 0.05, 适合做主成分分析。

5. 行业投资者情绪指标构建

本文将构建两种情绪指标, 一种为未加文本情绪指标的行业投资者情绪指标; 另一种为加入文本情绪指标的行业投资者情绪指标。

分别对标准化后的未加文本情绪指标和加入文本情绪指标的各行业情绪代理变量进行主成分分析, 提取累积方差贡献率 85% 以上的一个或多个主成分。限于篇幅, 本文只列出采掘和化工行业的分析结果如表 4 所示。

表 4 构建行业投资者情绪指标的主成分分析结果

行业	未加文本情绪主成分分析			加文本情绪主成分分析		
	主成分	特征值	累积方差贡献率	主成分	特征值	累积方差贡献率
采掘	Comp 1	3.5915	0.7183	Comp 1	3.6024	0.6004
	Comp 2	1.0005	0.9184	Comp 2	1.3487	0.8252
	—	—	—	Comp 3	0.6560	0.9345
化工	Comp 1	2.1451	0.7150	Comp 1	2.2859	0.5715
	Comp 2	0.7883	0.9778	Comp 2	0.9781	0.8160
	—	—	—	Comp 3	0.6822	0.9865

根据主成分载荷矩阵可以得出由预处理后的代理变量表示的主成分方程式, 并将各主成分的特征值作为权重计算平均数, 构造出各行业未加文本情绪的投资者情绪综合指标 $PSENT$ 和加入文本情绪的投资者情绪综合指标 $SENT$ 。所得的每一行业综合情绪指标 $PSENT$ 和 $SENT$ 为日度数据, 表示每日投资者的情绪量化结果。

6. 各行业情绪指标与其收盘价的相关性分析

为了观察行业情绪指标的波动对其收益变动的解释能力, 分别对未加文本情绪的情绪指标 $PSENT$ 和加入文本情绪的情绪指标 $SENT$ 与股票流动指数 P_t 进行 Pearson 相关性检验, 并绘制投资者情绪综合指标与行业股价趋势图, 图 1(a) 和图 1(b) 为采掘行业未加和加入文本情绪的情绪指标与其股价走势对比图; 图 2(a) 和图 2(b) 为化工行业未加和加入文本情绪的情绪指标与其股价走势对比图。

通过图 1 和图 2 两个行业未加和加入文本情绪的情绪指标与股价走势的对比可以看出, 采掘、化工和家用电器行业的两类情绪指标在一定程度上反映了行业股票流通指数收盘价的变动, 走势基本一致。2019 年 4 月 15 日至 5 月 17 日先涨后跌, 波动较大, 采掘行业情绪序列与行业股指序列的峰值相近; 5 月 21 日之后变得较为平稳。而且, 从对比图中也可以看出将文本情绪作为情绪代理变量加入主成分分析构建的情绪指标 $SENT$ 与行业股指 P_t 的趋势更为相近。

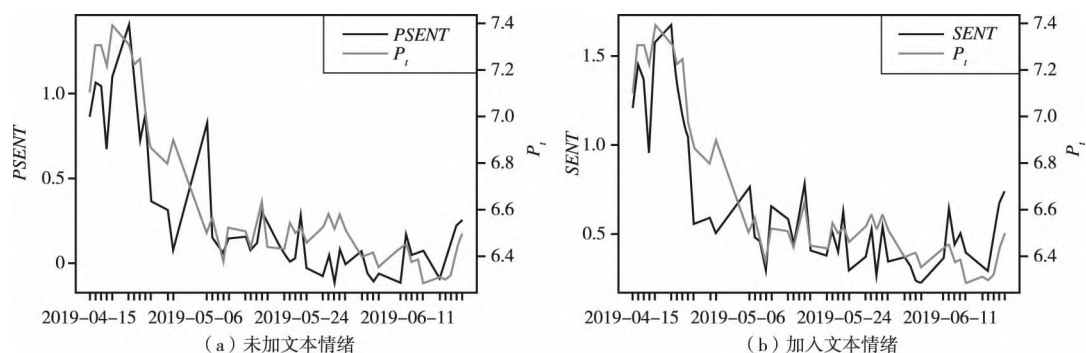


图1 情绪指标与采掘行业股价走势对比

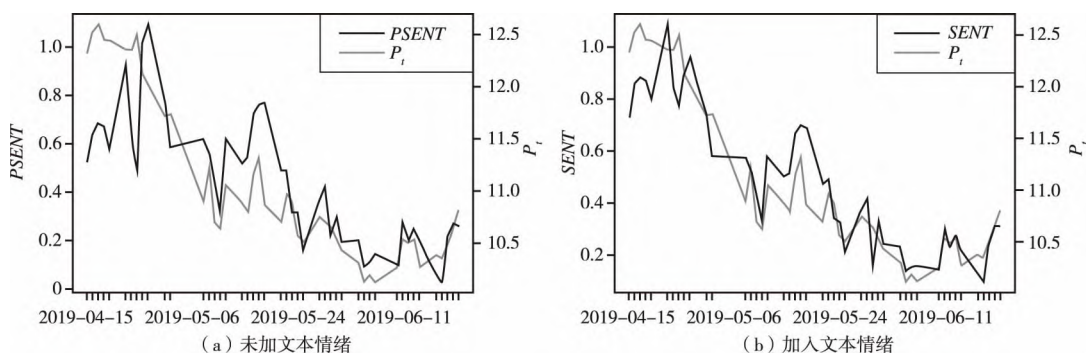


图2 情绪指标与化工行业股价走势对比

限于篇幅,表5列明了部分行业 $PSENT$ 和 $SENT$ 同相应行业股票流通指数 P_t 的相关性分析结果。

表5 部分行业两类投资者情绪指标与行业股票流动指数相关性分析结果

行业		$PSENT$ 与 P_t 的相关系数	$SENT$ 与 P_t 的相关系数
采掘		0.6691***	0.7516***
钢铁	4.15~6.21	0.6421***	0.6173***
	4.15~5.17	-0.0192	-0.2813
	5.20~6.21	0.4748*	0.2158
建筑材料	4.15~6.21	0.7178***	0.7058***
	4.15~5.17	0.6533**	0.6126**
	5.20~6.21	0.5236**	0.5149*
电气设备		0.0708	-0.1034
机械设备		0.8628***	—
国防军工		0.0505	0.0415
轻工制造		-0.6841***	-0.5464***
医药生物	4.15~6.21	0.5930***	0.5371***
	4.15~5.17	—	-0.6601**
	5.20~6.21	0.2313	0.2357

(续)

行业		$PSENT$ 与 P_i 的相关系数	$SENT$ 与 P_i 的相关系数
房地产		0.8590 ***	—
通信	4.15~6.21	0.8364 ***	0.7884 ***
	4.15~5.17	0.9142 ***	0.9176 ***
	5.20~6.21	0.8128 ***	0.8267 ***

注：*、**、*** 分别表示在 10%、5%、1% 水平上显著。

由表 5 以及各行业相关性分析结果可知, 采掘、化工、有色金属、建筑装饰、家用电器、纺织服装、商业贸易、农林牧渔、休闲服务、公用事业、交通运输、电子、传媒和综合等 14 个行业中两类投资者情绪指标与行业股指均具有一定正相关性, 并在 5% 显著性水平下均显著。相比较而言, 加文本情绪指标的投资者情绪指标 $SENT$ 与行业股票流动指数 P_i 相关性较大, 表明在以上 14 个行业中 $SENT$ 更能反映股票收盘价的动态。机械设备和房地产行业由于在变量筛选过程中文本情绪指标未通过检验, 因此只构建了未加文本情绪指标的投资者情绪指标 $PSENT$, 得到与 P_i 的相关系数分别为 0.8628 和 0.8590, 均在 5% 的显著性水平上显著。相反, 轻工制造行业的情绪指标与股票流动指数显著负相关, 从图 3 轻工制造行业未加和加入文本情绪的情绪指标与行业股价走势对比可以看出, 负相关的原因主要存在于“熊市”时期, 即股价呈下跌趋势, 投资者的情绪反而不断高涨。钢铁、建筑材料、汽车、食品饮料、医药生物、计算机、通信、银行和非银金融等 9 个行业中未加文本情绪指标的投资者情绪指标与行业股指相关性高于加入文本的投资者情绪指标。是不是受到了股价波动的影响呢? 将样本期分为两阶段, 4 月 15 日至 5 月 17 日为第一阶段, 股价波动相对较大; 5 月 20 日至 6 月 21 日为第二阶段, 股价波动相对平稳。每一阶段对两类情绪指标与股票流动指数进行相关性检验。由检验结果可知, 在第一阶段中, 建筑材料、汽车、食品饮料、计算机、银行和非银金融行业的未加文本的投资者情绪指标与股指的相关系数显著地大于加入文本的投资者情绪指标 (食品饮料的相关系数为负), 这一阶段的投资者情绪很大程度影响到整个研究期内的检验结果。通信行业两阶段中加入文本情绪指标的投资者情绪指标均显著地略大于未加文本的投资者情绪指标, 因此最终选择分阶段构建投资者情绪综合指标。钢铁行业的阶段检验结果中只有平稳时期的未加文本的投资者情绪指标与行业股指的相关性显著, 而在整个研究期的检验结果是非常显著的。医药生物行业在平稳阶段两类情绪指标与行业股指的相关性不显著, 在波动较大阶段除文本情绪指标外的其他代理变量 KMO 检

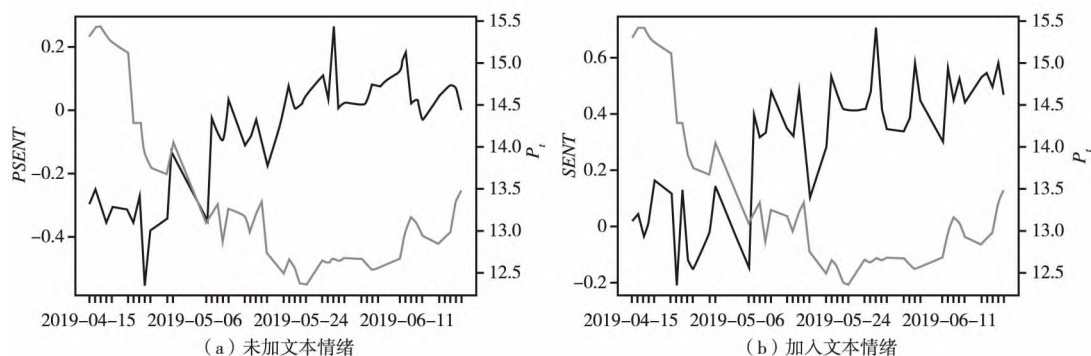


图 3 情绪指标与轻工制造行业股价走势对比

验值为 0.48, 依据度量标准无法做主成分分析; 加入文本情绪指标的投资者情绪指标显著与行业股指负相关。电气设备和国防军工行业的情绪指标与股票流动指数相关性很小, 且不显著, 因此不再对两行业构建投资者综合情绪指标。

综上所述, 采掘、化工、有色金属、建筑装饰、家用电器、纺织服装、商业贸易、农林牧渔、休闲服务、公用事业、交通运输、电子、传媒、通信(分阶段)和综合行业选取加入文本情绪指标构建投资者情绪指标; 钢铁、建筑材料、机械设备、汽车、轻工制造、食品饮料、医药生物、房地产、计算机、银行和非银金融行业选取未加文本情绪指标构建投资者情绪指标。由于我国对不同行业建立不同的政策, 投资者的信心水平受到不同程度的影响, 进而导致投资者在投资过程中所形成的情绪存在明显的差异, 加之各时期内各行业的投资价值不同, 使得行业间投资者情绪指标间存在明显差异性。构建的各行业投资者情绪指标与相应行业的股票流动指数的走势也存在一定的差异, 进一步验证了单一投资者情绪无法揭示各行业投资者综合情绪。

二、Black-Litterman 模型及观点矩阵的构建

1. Black-Litterman 模型的基本假设及算法

Markowitz 模型依据历史信息(先验预期收益率)计算投资组合的资产配置, 但在真实的金融市场中不确定地会出现某些信号, 使得投资者对市场产生主观观点, 即调整资产配置。Black 和 Litterman (1992) 基于传统 Markowitz 模型融入投资者观点, 得到后验资产收益率, 构建 Black-Litterman 模型(以下简称 BL 模型)。该模型的求解过程为:

假设市场中存在 N 种资产, 其收益率为 $R = \{R_1, R_2, \dots, R_N\}$, 服从多元正态分布, 即 $R \sim N(\mu, \Sigma)$, 其中 μ 和 Σ 分别为各资产预期收益率的期望值和协方差的估计值。期望收益 μ 是先验预期收益率期望值 π 加扰动项 ε_1 , $\pi = \lambda \Sigma \omega_{mkt}$, λ 为风险规避系数, 反映市场风险趋避的程度, ω_{mkt} 为市场组合各资产的投资配置; ε_1 是随机向量, $\varepsilon_1 \sim N(0, \tau \Sigma)$, τ 为规模参数, 反映主观观点相对于先验信息的比重。但学术界对于 τ 参数取值存在差异, He 和 Litterman (1999) 设定一个 τ 值, 使 $\omega/\tau = P \Sigma P'$; Satchell 和 Scowcroft (2000) 设定 τ 值为 1; Lee (2000) 设定取值范围为 $[0.01, 0.05]$; Blamont 和 Firoozye (2003) 设定 τ 近似于 1/观测值数量。综合以上研究结果, 本文将 τ 值设定为 0.05。

令 K 为观点量, 用各资产收益率的线性方程组来代表主观观点, 可表述为:

$$P\mu = Q + \varepsilon_2 \quad (2)$$

其中 P 是 $K \times N$ 维矩阵, 即对 N 个资产有 K 个观点, 每一行为权重和, 绝对观点的权重和为 1, 相对观点的权重和为 0; Q 为 $K \times 1$ 列向量, 即看法向量, 每一观点中绝对或相对预期收益率; ε_2 是观点的误差项, $\varepsilon_2 \sim N(0, \Omega)$, Ω 是 $K \times K$ 的对角矩阵, 表示投资者的观点与真实情况有差别, 受信息水平的影响。

根据贝叶斯法则, 结合先验信息与投资者观点的各资产后验预期收益率分布为 $R | Q, \Omega \sim N(\mu_{BL}, \Sigma_{BL})$, 其中

$$\mu_{BL} = [(\tau \Sigma)^{-1} + P' \Omega^{-1} P]^{-1} [(\tau \Sigma)^{-1} \pi + P' \Omega^{-1} Q] \quad (3)$$

$$\Sigma_{BL} = \Sigma + [(\tau \Sigma)^{-1} + P' \Omega^{-1} P]^{-1} \quad (4)$$

得到预期收益率的期望值与方差, 再根据风险最小化或收益率最大化输出最优投资组合权重

和预期收益。

2. 观点矩阵的构建

本文将行业投资者情绪指标量化为投资者观点矩阵，即通过构建的情绪指标转化得到 P 、 Q 、 Ω 。在使用构建的 26 个行业的投资者情绪指标（除电气设备和国防军工行业外）构造 P 、 Q 、 Ω 时，需要对 P 、 Q 、 Ω 的含义及计算方法做相应推广。

在本文应用背景下， $K=N=26$ ，即 26 个行业类别中的观点个数均为一个。 P 为 $N \times N$ 维矩阵，且只含有绝对收益，是单位矩阵。 Q 为每一观点中的预期收益率，即在先验预期收益率的基础上用投资者情绪指数调整得到后验预期收益率。本文采用 GARCH 模型对各行业收益率与相应行业投资者情绪指标进行回归，用估计的回归系数乘以各行业投资者情绪指数的平均数表示投资者的观点误差，即投资者看涨或看跌或不涨不跌的情绪调整。 Ω 为投资者的观点与真实情况的差别，受信心水平向量 confidences ($K \times 1$ 维向量) 的影响，本文采用各行业情绪指标的均值表示信心水平。至此，得到了 BL 模型的输入矩阵 P 、 Q 、 Ω 。

3. 实证研究

使用历史数据估计预期收益率 π 和协方差矩阵 Σ ，并融合投资者的个人观点，即确定 P 、 Q 、 Ω 后，运用 R 语言中的 BLCOP 包的 BLView() 函数构造观点，然后调用 posteriorEst() 函数对后验预期收益率的期望值和方差进行计算经观点调整后的日预期收益率。再调用 optimalPortfolios fPort() 函数计算各行业的投资比例，并进行卖空限制以及不考虑交易成本，选择效率前沿上风险最小的投资组合，图 4 为基于行业投资者情绪的主观观点的 BL 模型下的效率前沿，红色圆点代表最小风险水平下投资组合对应的风险和收益率，根据后验分布得到行业资产配置（如图 5 所示）。

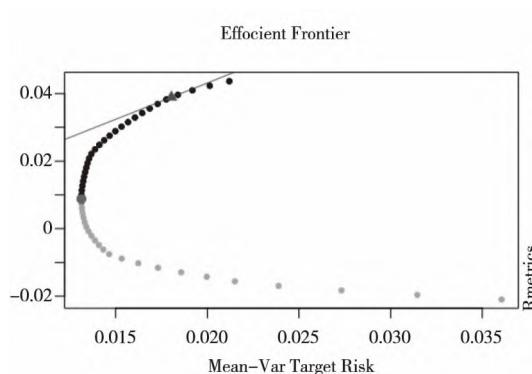


图 4 基于行业情绪指标的 BL 模型的效率前沿

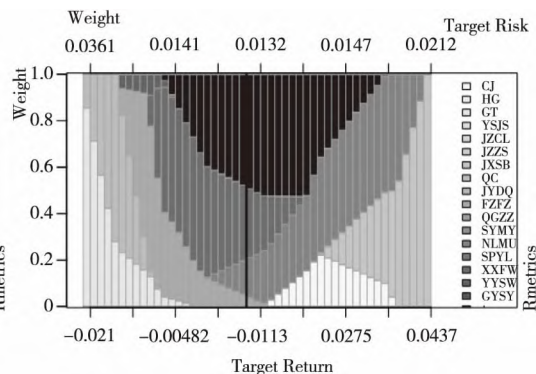


图 5 基于行业情绪指标的 BL 模型的行业资产配置

由图 5 可知，最小风险水平下的资产配置集中于纺织服装、农林牧渔、休闲服务和银行等 4 个行业，比例分别为 4.68%、14.36%、33.43% 和 47.53%，4 个行业经观点调整后的日预期收益率分别为 -0.0064、0.0319、-0.0055 和 0.0129，由此可得样本期内的日收益率均值为 0.0086。

基于行业投资者情绪的 BL 模型的相对表现如何？本文设置传统 Markowitz 均值一方差模型的投资组合作为对照组。图 6 为均值一方差模型下的效率前沿，最小风险水平下投资组合权重比例为纺织服装 5.2%、农林牧渔 13.81%、休闲服务 33.77% 以及银行 47.22%（如图 7 所示），日收益率均值为 -0.0008。

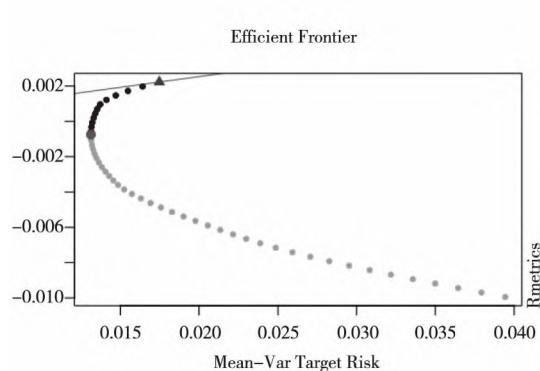


图6 Markowitz 均值-方差模型的效率前沿

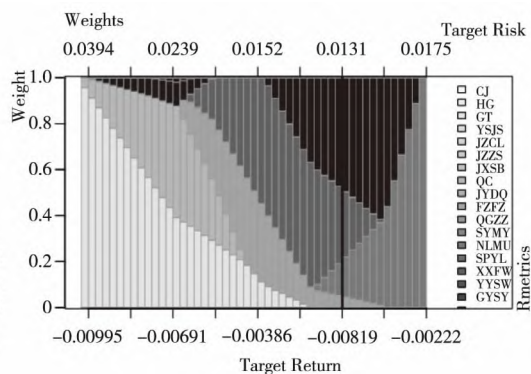


图7 Markowitz 均值-方差模型的行业资产配置

表6列出了两种模型的实证结果,可见,在样本期内,最小风险水平下的资产配置均集中于纺织服装、农林牧渔、休闲服务和银行4个行业,配置比例变化不大,对预期收益率为负的行业配置比例略微减少,预期收益率为正的行业配置比例增加,但基于本文构建的行业投资者情绪的BL模型投资组合的日预期收益率明显高于其他配置模型,由亏损变盈利。表中还列明了各配置模型的夏普比率,即风险调整后的收益率,Markowitz模型的日收益率为负,并且小于无风险日利率(0.01%),因此求得夏普比率为负。结果可知,本文提出的基于行业投资者情绪的BL模型在行业配置中具有一定的竞争优势。

表6 两种模型实证结果比较(风险最小策略下的投资组合/卖空限制/无交易成本)

2019年4月15日~2019年6月21日

行业	行业投资者情绪 BL 模型		Markowitz 模型	
	日预期收益率	资产配置	日预期收益率	资产配置
纺织服装	-0.0064	4.68%	-0.0026	5.2%
农林牧渔	0.0319	14.36%	0.0022	13.81%
休闲服务	-0.0055	33.43%	-0.0029	33.77%
银行	0.0129	47.53%	-0.00004	47.22%
日收益率均值	0.0086		-0.0008	
夏普比率	0.6538		-0.0687	

本文同时对两种模型比较效率前沿上夏普比率最高的投资组合,进行卖空限制以及不考虑交易成本。由表7可见,资产配置行业有了明显的不同,Markowitz模型仅集中配置于农林牧渔一个行业;基于行业投资者情绪的BL模型分散行业配置,日预期收益率明显高于Markowitz模型预期收益率。

表7 两种模型实证结果比较(夏普比率最高策略下的投资组合/卖空限制/无交易成本)

2019年4月15日~2019年6月21日

行业	行业投资者情绪 BL 模型		Markowitz 模型	
	日预期收益率	资产配置	日预期收益率	资产配置
机械设备	0.0437	65.96%		
农林牧渔	0.0319	34.04%	0.0022	100%
日收益率均值	0.0397		0.0022	
夏普比率	2.0204		0.1200	

三、稳健性检验

第二部分的实证部分只选择一季度时间段作为样本期,同时在实证部分求解模型的配置权重时设定了卖空限制和无交易成本的约束,这些约束条件对配置结果有影响。本部分主要研究加入行业情绪构建的 BL 模型相对于时间、约束条件等附件条件的可靠性和非随机性,分阶段、分情况对模型进行稳健性检验。具体包括允许卖空、考虑交易成本、三个季度的样本外测试以及股市暴涨暴跌阶段和较快或较慢上涨/下跌阶段。

1. 允许卖空和考虑交易成本

放松卖空限制,在 optimalPortfolios fPort () 函数中添加 constraints = “ShortOnly”,分别对两种模型求解效率前沿上风险最小和夏普比率最高的投资组合,配置结果同表 6、表 7 的结果。允许卖空后,模型的配置结果不受其影响,本文提出的模型依然具有优势。

考虑交易成本,借鉴谭华清等(2018)的引入方法,交易成本 $t_i = \text{换手率}_i \times \text{单价交易成本}$,经交易成本调整后的收益为:收益 $t_i, \text{adjusted} = \text{收益}_i - \text{交易成本}_i$,本文设定单价交易成本为千分之一。在风险最小策略下,基于行业投资者情绪的 BL 模型资产配置集中于纺织服装、农林牧渔、休闲服务和银行等行业,比例分别为 4.11%、14.53%、32.93%和 48.43%,日预期收益率为 0.0081,夏普比率为 0.7407;Markowitz 模型同样将资产配置于以上 4 个行业,比例分别为 6.43%、12.69%、35.71%和 45.26%,日预期收益率为 -0.0016,夏普比率为 -0.1232,行业情绪的 BL 模型在各维度上都明显高于 Markowitz 模型,由亏转盈。在夏普比率最高策略下,基于行业投资者情绪的 BL 模型资产配置于机械设备和农林牧渔行业,对应配置比例为 66.11%和 33.89%,日预期收益率为 0.0374,夏普比率为 1.9179;Markowitz 模型资产配置于农林牧渔、食品饮料和银行等行业,对应配置比例为 54.12%、2.58%和 43.30%,日预期收益率为 -0.0002,夏普比率为 -0.0216,基于行业投资者情绪的 BL 模型投资组合回报在各维度上都明显高于 Markowitz 模型,竞争优势还是存在的。

2. 样本外测试

本文选取了 2019 年 6 月 24 日至 2019 年 9 月 29 日、2019 年 10 月 8 日至 2019 年 12 月 31 日以及 2020 年 1 月 2 日至 2020 年 3 月 27 日三个季度的数据作为测试集,验证本文提出的基于行业投资者情绪的 BL 模型在训练集外的表现效果,并设定卖空限制和无交易成本的约束。

(1) 2019 年 6 月 24 日~2019 年 9 月 29 日。从股市大盘行情走势图看,2019 年 6 月 24 日至 2019 年 9 月 29 日这一季度上证指数在 2700 到 3000 点之间窄幅震荡,7~8 月延续 4 月份的跌势,但幅度较小,9 月呈现上升走势。在最小风险策略下,基于行业投资者情绪的 BL 模型资产配置于采掘、钢铁、有色金属、农林牧渔、休闲服务和银行等行业,相应的配置比例为 30.74%、3.61%、17.29%、4.01%、19.97%和 24.38%,日预期收益率为 0.0035,夏普比率为 0.4789,均高于 Markowitz 模型的日预期收益率 -0.0046 和夏普比率 -0.3884。在夏普比率最高策略下,本文构建的 BL 模型的日预期收益率为 0.0285,虽小于 Markowitz 模型的日预期收益率 0.0399,但其夏普比率达到了 1.9805,即风险调整后的日收益率明显好于 Markowitz 模型的 1.1083,并且行业配置更为分散,Markowitz 模型只集中于食品饮料和电子行业,行业情绪 BL 模型分散到钢铁、有色金属、机械设备、食品饮料、医药生物和非银金融等 6 个行业,降低了投资风险。

(2) 2019 年 10 月 8 日~2019 年 12 月 31 日。2019 年 10 月 8 日至 2019 年 12 月 31 日这

一季度股市以平稳运行为主, 各行业投资者情绪日度指数整体呈现窄幅震荡走势, 未见明显分歧走势。10月份, 中美经贸磋商取得实质性进展, 投资者情绪得到一定提振, 但国内通胀预期升温, 对经济滞胀的担忧一定程度影响着投资者情绪。在最小风险策略下, 基于行业投资者情绪的 BL 模型的最大回报为 0.0023, 夏普比率为 0.4107, 资产由高到低分别配置于公用事业、银行、食品饮料、休闲服务、化工和纺织服装行业; Markowitz 模型的最大回报为 0.0003, 夏普比率为 0.0313。行业情绪 BL 模型在不同维度上的表现与 Markowitz 模型对比具有明显的资产配置价值。在夏普比率最高策略下, 基于行业投资者情绪的 BL 模型的日收益率为 0.0221, 夏普比率为 2.7569, 同样高于 Markowitz 模型的日预期收益率 0.0039 和夏普比率 0.3451, 资产配置优势在 2019 年第四季度显著。

(3) 2020 年 1 月 2 日~2020 年 3 月 27 日。2020 年 1 月初因央行宣布降准投资者对股市的关注出现高点。而新冠肺炎疫情在 1 月末 2 月初全国范围发生后, 引发了投资者担忧, 行业投资者情绪整体呈现下跌态势。3 月开始国内疫情整体得到控制, 但全球疫情的蔓延, 投资者对全球经济增长前景仍处于悲观情绪状态, 投资者信心继续受到影响, 并且受美国股市在十天内四次熔断的影响, 行业情绪指数延续了下跌趋势。实证结果显示, 资产配置中出现了医药生物行业, 这一行业受疫情影响逐渐向好。但与其他季度的结果不同, 在风险最小策略下, 基于行业投资者情绪的 BL 模型最大日预期收益率为 -0.0027, 在经济数值上低于 Markowitz 模型的日预期收益率 -0.0015。究其原因, 一次爆发性的事件的发生将影响投资者与其感兴趣的话题间的情绪关联状态, 个体非理性情绪在某一事件发生时可能受股市大盘影响较大。在夏普比率最高策略下, 本文提出的基于行业投资者情绪的 BL 模型日预期收益率为 0.0751, 夏普比率为 2.5033, Markowitz 模型日预期收益率为 0.0016, 夏普比率为 0.0533, 本文构建的基于行业投资者情绪的 BL 模型在资产配置上依然具有一定的竞争优势。

(4) 股市暴涨暴跌阶段。2020 年一季度经历了新冠肺炎疫情的重大冲击, 但中国股票市场未出现明显的大幅波动, 然而实证结果与股市平稳季度存在差异。股市巨幅的波动是否对基于行业投资者情绪的 BL 模型的稳健性造成影响? 在短短十年间, 中国股市经历了两次巨幅且频繁的波动, 分别是 2008 年和 2015 年, 出现“股灾”。考虑到东方财富网股吧文本数据的可获取性, 本文选取 2015 年作为股市暴涨暴跌稳健性检验的时间点。2015 年既有“疯牛”又有“股灾”, 涨幅跌幅均达到 30% 以上。牛市起步于 2014 年, 一直到在 2015 年上半年才走完了主升浪。“杠杆”是 2015 年暴涨暴跌的始作俑者, 监管层的强制去杠杆方式让股市在 6 月中旬出现一轮断崖式下跌。本文同样选择一个季度数据进行检验, 设定暴涨阶段为 2015 年 3 月 16 日至 6 月 19 日, 暴跌阶段为 2015 年 6 月 23 日至 9 月 25 日。

行为金融学过度反应理论指出, 投资者在面对爆发性或未预期到的事件时, 赋予新信息更高的权重而忽略以往的信息, 从而引起股价的暴涨或暴跌。相反, 反应不足异象同样存在, 投资者的保守主义使其不能及时根据事件变化的情况修正权重, 而对基础信息给予过多权重。此外, 投资者的非理性行为——“羊群行为”使得股市出现“追涨杀跌”现象。自然人投资者由于无法获得全面的信息、为减少自身恐惧以及可能存在缺乏某些知识经验等原因而模仿他人决策或依赖于舆论, 轻视了私人信息。当行情向好时盲目追涨越过价值的限度; 股价下跌幅度较大时, 具有高安全边际投资价值时认赔离场。依据行为金融理论, 股市巨幅波动阶段的投资者情绪值更为极端, 观点误差应高于平均值。前文用回归系数乘以各行业投资者情绪指数的平均数表示投资者的观点误差, 在暴涨暴跌阶段, 本文将观点误差进行翻倍

作为情绪调整值,更为真实地反映各行业收益率的变化。

①暴涨阶段(2015年3月16日~2015年6月19日)。在“疯牛”阶段,最小风险策略下,基于行业投资者情绪的 BL 模型资产配置于有色金属、机械设备、纺织服装、休闲服务、传媒和银行等行业,日预期收益率为 0.0422,夏普比率为 2.2328,在两个维度上均显著高于 Markowitz 模型的日预期收益率 0.0040 和夏普比率 0.2162。夏普比率最高策略下,基于行业投资者情绪的 BL 模型资产配置于采掘和计算机行业,日预期收益率为 0.2422,夏普比率为 9.3876,Markowitz 模型日预期收益率为 0.0081,夏普比率为 0.3506。与股市较为平稳季度的配置结果相比,两种策略下的日预期收益率涨幅至少 5 倍以上,并且本文构建的 BL 模型资产配置优势依然显著。

②暴跌阶段(2015年6月23日~2015年9月25日)。在“股灾”阶段,最小风险策略下,基于行业投资者情绪的 BL 模型资产日预期收益率为 0.0071,夏普比率为 0.2373,在这两个维度上均显著高于 Markowitz 模型的日预期收益率 -0.0047 和夏普比率 -0.1897。夏普比率最高策略下,基于行业投资者情绪的 BL 模型日预期收益率为 0.0126,夏普比率为 0.3788;Markowitz 模型日预期收益率为 -0.0025,夏普比率为 -0.0807。基于行业投资者情绪的 BL 模型将 Markowitz 模型负的日预期收益率调整为正收益率,由亏损变盈利。

(5) 股市较快或较慢上涨/下跌阶段。基于行业投资者情绪的 BL 模型在整个股市暴涨或暴跌阶段竞争优势明显,在较快或较慢上涨、下跌阶段配置效果如何呢?本文分别将选取的暴涨暴跌阶段再细分,根据股市大盘走势,设定 2015 年 3 月 16 日至 2015 年 4 月 30 日为较快上涨阶段,2015 年 5 月 4 日至 2015 年 6 月 19 日为较慢上涨阶段;2015 年 6 月 23 日至 2015 年 8 月 7 日为较快下跌阶段,2015 年 8 月 10 日至 2015 年 9 月 25 日为较慢下跌阶段。

①较快上涨阶段(2015年3月16日~2015年4月30日)。在最小风险策略下,基于行业投资者情绪的 BL 模型的日预期收益率为 0.0530,夏普比率为 4.6814,Markowitz 模型的日预期收益率为 0.0090,夏普比率为 0.7672。在夏普比率最高策略下,基于行业投资者情绪的 BL 模型的日预期收益率为 0.1625,夏普比率为 11.3636,Markowitz 模型的日预期收益率为 0.0132,夏普比率为 0.9357。较快上涨阶段基于行业投资者情绪的 BL 模型配置结果明显高于 Markowitz 模型配置结果,且高于整体暴涨阶段。这一结果符合经济学原理——边际效用递减规律。在股市上涨初期,投资者情绪得到一定的提振,涨幅继续增加时,投资者的预期继续加强,使投资者情绪更为乐观。

②较慢上涨阶段(2015年5月4日~2015年6月19日)。在风险最小策略下,基于行业投资者情绪的 BL 模型的日预期收益率为 0.0228,夏普比率为 1.2667,Markowitz 模型的日预期收益率为 0.0048,夏普比率为 0.2697。这一策略下较慢上涨阶段基于行业投资者情绪的 BL 模型配置结果小于较快上涨阶段的配置结果,由于边际效用递减规律,当大盘继续上涨但涨幅相对较小时,投资者的情绪有所下跌,导致预期减少,资产配置的日预期收益率下降。在夏普比率最高策略下,基于行业投资者情绪的 BL 模型的日预期收益率为 0.3713,夏普比率为 15.0894,Markowitz 模型的日预期收益率为 0.0115,夏普比率为 0.5157。这一策略下较慢上涨阶段基于行业投资者情绪的 BL 模型配置结果略高于较快上涨阶段的配置结果。

③较快下跌阶段(2015年6月23日~2015年8月7日)。较快下跌阶段结果刚好相反,在风险最小策略下,基于行业投资者情绪的 BL 模型资产配置分散于 15 个行业,日预期收益率为 -0.00002,夏普比率为 -0.2906,Markowitz 模型资产集中于 3 个行业,日预期收

益率为-0.0066, 夏普比率为-0.2548。在夏普比率最高策略下, 基于行业投资者情绪的 BL 模型的日预期收益率为 0.0007, 夏普比率为 0.2069, Markowitz 模型的日预期收益率为-0.0037, 夏普比率为-0.1166。这一阶段基于行业投资者情绪的 BL 模型配置的最大收益率高于 Markowitz 模型的结果, 但低于整体暴跌阶段, 为负, 同样可以用边际效用递减来解释。当投资者还沉浸在炒股盈利的喜悦情绪中时, 股市大盘的较快下跌, 使得恐慌、紧张情绪持续扩散。

④较慢下跌阶段(2015年8月10日~2015年9月25日)。在风险最小策略下, 基于行业投资者情绪的 BL 模型的日预期收益率为 0.0183, 夏普比率为 0.7194, Markowitz 模型的日预期收益率为-0.0050, 夏普比率为-0.2056。在夏普比率最高策略下, 基于行业投资者情绪的 BL 模型的日预期收益率为 0.0686, 夏普比率为 1.6348, Markowitz 模型的日预期收益率为-0.0033, 夏普比率为-0.1073。较慢下跌阶段基于行业投资者情绪的 BL 模型配置的最大收益率高于 Markowitz 模型的结果, 同时均高于较快下跌阶段的配置结果。随着股市呈现出震荡走势, 持续一段时间后出现又一次相对较小幅度的下跌, 由于边际效用递减规律, 这一阶段投资者情绪有所缓解, 预期收益率回升。

四、结论

行业对于基金管理者进行产业权重配置起着重要作用, 同时, 个体投资者在挑选和配置个股前, 需要关注个股所在行业的行情, 把握强势行业, 决策投资比例。BL 模型是“科学与艺术”相结合的模型, 既对资产进行量化配置, 又将投资者对股票基本面的主观观点——“艺术”成分嵌入模型中。BL 模型计算的关键问题是投资者观点矩阵的构建, 现有研究更多是将历史数据或预期收益进行预测作为投资者观点矩阵, 但行为金融大量研究已证实投资者的心理偏差极大地会影响投资决策。历史数据无法反映投资者认知和行为偏差, 因此, 在模型应用上必然会产生误差。本实证研究的意义在于将投资者情绪嵌入 BL 模型中, 以投资者情绪指标来更为真实地反映其观点, 使得模型的资产配置结果更加有参考价值。

本文通过文本挖掘技术, 爬取网页内容构建文本情绪指标, 并作为行业情绪指标的代理变量构建各行业的投资者情绪指标, 最终提出一种将行业情绪指标量化为 BL 模型所需观点矩阵的方法。基于行业投资者情绪形成的观点矩阵是否会提高配置的预期收益率以及夏普比率? 通过实证分析表明, 在效率前缘上风险最小和夏普比率最高水平下, 加入行业情绪主观观点的资产组合效率优于 Markowitz 均值一方差模型以及加入股市情绪主观观点的 BL 模型, 改善了行业投资组合的回报以及夏普比率。通过允许卖空、考虑交易成本和样本外测试集进行稳健性检验, 允许卖空和考虑交易成本配置结果没有太大变化。样本外测试集包括了三个季度数据以及股市暴涨暴跌数据, 配置结果表明, 受爆发性事件以及披露的利好或利空消息的影响, 自然人投资者情绪乐观或悲观程度幅度增大, 从而提振或削弱了投资者的预期。股市较大幅度上涨, 基于行业投资情绪的 BL 模型的收益率和夏普比率显著高于 Markowitz 模型的配置结果。除 2020 年第一季度受全球新冠肺炎疫情影响出现基于行业投资情绪的 BL 模型配置结果低于 Markowitz 模型外, 其他阶段均通过了稳定性检验, 本文构建的基于行业投资者情绪的 BL 模型具有一定的竞争优势。

BL 模型中最为重要的是减少投资者观点设定的不确定性, 并合理设定观点的置信度, 因此投资者情绪的度量仍需探索。如何建立更为精准、客观、实时的投资者情绪度量指标是投资者情绪研究课题一直要解决的问题。近五年, 国外学者运用跨学科融合信息处理的方法

开启了投资者情绪研究的新范式，将自然语言处理技术与金融数据特征结合，如提高分类准确度，改进机器学习算法等（Zhao 和 Gui, 2018; Hassan 和 Mahmood, 2018; Zhou 等, 2019），来提取更为可靠的投资者情绪度量指标；国内相关的实证研究才刚刚起步，研究内容还不够系统。而且研究主要将文本内容（书面文本）作为判断情绪极性的唯一信息来源，而忽略了平台服务的其他网络信息，例如有影响力的用户发表的观点或评论对其他投资者情绪极性具有影响力；文本的信息除以文字形式呈现外，音视频、静/动态图像、表情符号等非文本信息越来越频繁使用，占据着社交资源的主导地位，但对图像等进行视觉分析以预测情绪的相关研究较少。在今后的研究中，应寻求更为高效、高质的投资者情绪指标，将基于深度学习技术，从金融贴吧中提取文本和带有信息的图像信息，运用社交网络方法分析用户行为，建立投资者网络信息流量到投资者网络情绪的映射，进一步提高资产配置模型的效果。

参 考 文 献

- [1] Baker M. P., Wurgler J. A., 2007, *Investor Sentiment in the Stock Market* [J], Journal of Economic Perspective, 21 (2), 129~151.
- [2] Baker M., Wurgler J., Yuan Y., 2012, *Global, Local, and Contagious Investor Sentiment* [J], Journal of Financial Economics, 104 (2), 272~287.
- [3] Beach S. L., Orlov A. G., 2007, *An Application of the Black-Litterman Model with EGARCH-M-derived Views for International Portfolio Management* [J], Financial Markets and Portfolio Management, 21 (2), 147~166.
- [4] Black F., Litterman R. B., 1991, *Asset Allocation: Combining Investor Views with Market Equilibrium* [J], Journal of Fixed Income, 1 (2), 7~18.
- [5] Black F., Litterman R. B., 1992, *Global Portfolio Optimization* [J], Financial Analysts Journal, 48 (5), 28~43.
- [6] Blamont D., Firoozy N., 2003, *Asset Allocation Model* [R], Global Markets Research: Fixed Income Research, Deutsche Bank.
- [7] Chiarawongse A., Kiatsupaibul S., Tirapat S., Van Roy B., 2012, *Portfolio Selection with Qualitative Input* [J], Journal of Banking & Finance, 36 (2), 489~496.
- [8] Corredor P., Ferrer E., Santamaria R., 2013, *Investor Sentiment Effect in Stock Markets: Stock Characteristics or Country-Specific Factors?* [J], International Review of Economics and Finance, 27, 572~591.
- [9] Da Z., Engelberg J., Gao P., 2015, *The Sum of all FEARS Investor Sentiment and Asset Prices* [J], Review of Financial Studies, 28 (1), 1~32.
- [10] De Long J. B., Shleifer A., Summers L. H., Waldmann R. J., 1990, *Noise Trader Risk in Financial Markets* [J], Journal of Political Economy, 98 (4), 703~738.
- [11] Didenko A., Demicheva S., 2013, *Application of Ensemble Learning for Views Generation in Meucci Portfolio Optimization Framework* [J], Review of Business and Economic Studies, 1, 100~110.
- [12] Ellis A., Harper R. A., 1975, *A New Guide to Rational Living* [J], Psycritiques, 21 (3), 158~167.
- [13] Creamer G. G., 2015, *Can a Corporate Network and News Sentiment Improve Portfolio Optimization using the Black-Litterman Model?* [J], Quantitative Finance, 15 (8), 1405~1416.
- [14] Hassan A., Mahmood A., 2018, *Convolutional Recurrent Deep Learning Model for Sentence Classification* [J], IEEE Access, 6, 13949~13957.
- [15] He G., Litterman R., 1999, *The Intuition behind Black-Litterman Model Portfolios* [R], Tech-

nical Note, Goldman Sachs Quantitative Resources Group.

[16] Ho C. , Hung C. H. , 2009, *Investor Sentiment as Conditioning Information in Asset Pricing* [J], Journal of Banking & Finance, 33 (5), 892~903.

[17] Kahneman D. , Slovic P. , Tversky A. , 1982, *Judgment under Uncertainty: Heuristics and Biases* [M], Cambridge: Cambridge University Press.

[18] Kahneman D. , Tversky A. , 1979, *Prospect Theory: An Analysis of Decision under Risk* [J], Econometrica, 47 (2), 263~291.

[19] Kara M. , Ulucan A. , Atici K. B. , 2019, *A Hybrid Approach for Generating Investor Views in Black-Litterman Model* [J], Expert Systems with Applications, 128, 256~270.

[20] Kim H. Y. , Won C. H. , 2018, *Forecasting the Volatility of Stock Price Index: A Hybrid Model Integrating LSTM with Multiple GARCH-type Models* [J], Expert Systems with Application, 103 (8), 25~37.

[21] Kurov A. , 2010, *Investor Sentiment and the Stock Market's Reaction to Monetary Policy* [J], Journal of Banking & Finance, 34 (1), 139~149.

[22] Lee C. M. C. , Shleifer A. , Thaler R. H. , 1991, *Investor Sentiment and the Closed End Fund Puzzle* [J], Journal of Finance, 46 (1), 75~109.

[23] Lee W. , 2000, *Advanced Theory and Methodology of Tactical Asset Allocation* [M], New York: John Wiley & Sons.

[24] Markowitz H. M. , 1952, *Portfolio Selection* [J], Journal of Finance, 7 (1), 77~91.

[25] Salhin A. F. S. , Sherif M. , Jones E. , 2016, *Managerial Sentiment, Consumer Confidence and Sector Returns* [J], International Review of Financial Analysis, 47, 24~38.

[26] Satchell S. E. , Scowcroft A. , 2000, *A Demystification of the Black-Litterman Model: Managing Quantitative and Traditional Portfolio Construction* [J], Journal of Asset Management, 1 (2), 138~150.

[27] Simon H. A. , 1976, *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organization* [M], New York: Macmillan Publishing Co. , Inc.

[28] Stambaugh R. F. , Yu J. , Yuan Y. , 2012, *The Short of It: Investor Sentiment and Anomalies* [J], Journal of Financial Economics, 104 (2), 288~302.

[29] Wang Y. H. , Keswani A. , Taylor S. , 2006, *The Relationships between Sentiment, Returns and Volatility* [J], International Journal of Forecasting, 22 (1), 109~123.

[30] Zhao J. Q. , Gui X. L. , 2018, *Deep Convolution Neural Networks for Twitter Sentiment Analysis* [J], IEEE Access, 6, 23253~23260.

[31] Zhou J. , Lu Y. , Dai H. N. , Wang H. , Xiao H. , 2019, *Sentiment Analysis of Chinese Microblog Based on Stacked Bidirectional LSTM* [J], IEEE Access, 7, 38856~38866.

[32] 卢米雪、朱喜安：《行业投资者情绪的测量及其对收益的影响效应研究》[J]，《统计与信息论坛》2014 年第 4 期。

[33] 谭华清、赵学军、黄一黎：《动量思想与大宗商品的战术配置价值——基于 Black-Litterman 模型》[J]，《经济理论与经济管理》2018 年第 10 期。

[34] 向诚、陆静：《投资者有限关注、行业信息扩散与股票定价研究》[J]，《系统工程理论与实践》2018 年第 4 期。

[35] 易志高、茅宁：《中国股市投资者情绪测量研究：CICSI 的构建》[J]，《金融研究》2009 年第 11 期。

[36] 约翰·梅纳德·凯恩斯：《就业、利息和货币通论》[M]，中国社会科学出版社，2009。

Research on Black-Litterman Model Allocation of Industry Asset

Meng Yong¹ Ren Meng² Zhao Xin³

- (1. School of Statistics, Shanxi University of Finance and Economics;
2. School of Public Finance and Economics, Shanxi University of Finance and Economics;
3. School of Finance, Shanxi University of Finance and Economics)

Research Objectives : Construct industry investor sentiment indicators based on social network text mining technology that reflects the trend of industry stock prices, and improve the asset allocation results of the Black-Litterman model embedded in industry investor sentiment indicators. **Research Methods :** Based on social network text mining technology and principal component analysis, this paper constructed investor sentiment indicators of the industry which forms the investor views matrix of the BL model, and determined the asset allocation ratio of the industry. **Research Findings :** The BL model based on industry investor sentiment can effectively improve the average return rate and Sharpe ratio of asset allocation. Such findings are robust to the out-of-sample verification (except for the stage of COVID-19), the stage of sharp rise and fall, short selling and transaction cost adjustments. **Research Innovations :** The construction of investor sentiment index based on social network text mining technology solves the limitation that predictive models that only rely on expected returns or historical data cannot visually reveal the psychological cognition and behavior of investors. Scientifically solve the problem of generating investor's opinion in the Black-Litterman model from a new perspective. **Research Value :** This paper expands the research of Black-Litterman model theory system and promotes the application of behavioral finance theory in asset allocation.

Key Words : Investor Sentiment; Black-Litterman Model; Industry Asset Allocation; Text Mining

JEL Classification : G11, C32, C88

(责任编辑: 白延涛)