

# 基于舆情的大类资产配置模型

## 互联网大数据挖掘系列专题之（十一）

### 报告摘要:

#### ● 利用舆情数据进行资产配置

投资者对资产的关注度舆情指数是一种群体智慧，携带了市场关注热点的信息。投资者情绪随着资产价格变化而波动，同时也反过来影响资产价格。如当市场出现某一热点时，投资者关注度上升，情绪的上升带动资产价格上涨。上涨的价格引来更多的关注度，形成一种正反馈机制。而当投资者热度上升到一个高点时，无法吸引更多的投资者，之前的投资者开始逐渐减少关注，资产价格开始下跌。因此通过对于舆情因子变化趋势的分析，可以去探究投资者关注度的变化情况，从而对资产收益率的变化进行预判。

#### ● 资产选择及舆情因子构建测试

在大类资产选择上，本篇专题选择了权益类、债券类、商品类和现金类四种资产，具体分别是沪深 300 指数、中证 500 指数、中证国债指数、贵金属指数、南华农产品指数、标普 500 指数、万得货币基金指数。采用相应资产的搜索舆情指数作为对应的舆情数据，构造舆情数据因子、结合舆情数据的动量因子以及作为补充的传统因子三类。用因子 IC 均值、IR 值筛选有效因子，通过多元线性回归的方法得到资产预期收益率。

#### ● 基于舆情的大类资产配置模型及历史回测结果

经典 BL 模型对投资者要求较高，投资者错误的观点会使得配置结果更不准确。我们将互联网舆情数据与传统 BL 模型相结合，通过分析互联网舆情数据与资产价格变化的关系，试图得到更加准确的资产预期收益率，提升资产配置的效果。在回测区间内，模型取得了 16.18% 的年化收益率，16.02% 的年化波动率，14.28% 的最大回撤，夏普比率为 1.02。

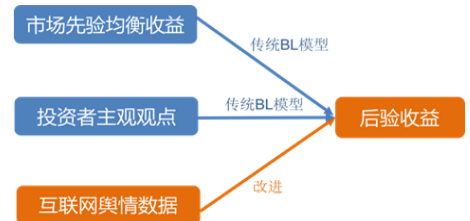
#### ● 模型最新资产配置建议

根据最新资产表现及资产对应的搜索舆情数据表现，模型给出 2017 年 6 月最新的资产配置建议，其中沪深 300、中证 500、中证国债、贵金属、农产品、标普 500、货币基金配置比例分别为：40%、5%、5%、5%、5%、35%、5%。

#### ● 核心假设风险:

本专题内容基于资产舆情数据测算完成，资产舆情是影响资产表现的一类因素，资产表现受其他诸多因素影响，注意模型失效风险。

### 基于舆情的大类资产配置模型



### 模型 2017 年 6 月配置建议

资产	配置比例
沪深300	40%
中证500	5%
中证国债	5%
贵金属	5%
农产品	5%
标普500	35%
货币基金	5%

分析师: 史庆盛 S0260513070004



02087577060

sqs@gf.com.cn

### 相关研究:

- 《基于大数据挖掘的行业轮动策略-互联网大数据挖掘系列专题之（十）》 2017-03-15
- 《基于大数据挖掘的概念轮动策略-互联网大数据挖掘系列专题之（九）》 2016-08-22
- 《多维数据系的大数据择时策略研究——互联网大数据挖掘系列专题之（八）》 2016-05-17

联系人: 陈原文 0755-82797057  
chenyuanwen@gf.com.cn

## 目录索引

一、网络舆情数据介绍 .....	4
1.1 大类资产配置与互联网舆情数据 .....	4
1.2 舆情数据可预测性分析 .....	5
1.3 现有相关研究 .....	5
1.4 大类资产舆情数据 .....	6
二、再谈大类资产配置模型 .....	8
2.1 BL 模型介绍 .....	8
2.2 传统 BL 模型的局限及改进 .....	10
三、舆情因子测试 .....	11
3.1 舆情因子构建及筛选 .....	11
3.2 数据标准化 .....	12
3.3 舆情因子测试 .....	12
四、实证分析 .....	17
4.1 实证分析 .....	17
4.2 回测结果 .....	18
4.3 模型最新配置建议 .....	19
五、总结 .....	20
风险提示 .....	20

## 图表索引

图 1: 百度指数: 代表了某段时间内关键词的搜索热度	4
图 2: 沪深 300 指数月行情与月舆情值关系	7
图 3: 中证 500 指数月行情与月舆情值关系	7
图 4: 中证国债指数月行情与月舆情值关系	7
图 5: 万得贵金属指数月行情与月舆情值关系	7
图 6: 南华农产品指数月行情与月舆情值关系	8
图 7: 标普 500 指数月行情与月舆情值关系	8
图 8: 货币基金指数月行情与月舆情值关系	8
图 9: Black-Litterman 模型主要思想	9
图 10: Black-Litterman 模型配置方法	10
图 11: 改进后 BL 模型	11
图 12: 舆情因子测试流程	13
图 13: 舆情因子测试 - 沪深 300	14
图 14: 舆情因子测试 - 中证 500	14
图 15: 舆情因子测试 - 中证国债	15
图 16: 舆情因子测试 - 贵金属	15
图 17: 舆情因子测试 - 农产品	16
图 18: 舆情因子测试 - 标普 500	16
图 19: 舆情因子测试 - 货币基金	17
图 20: 实证分析示意图	18
图 21: 回测净值结果	19
图 22: 模型最新配置建议	20
表 1: 舆情数据来源一览	5
表 2: 大类资产舆情数据示例	6
表 3: 各类资产月舆情值与月行情	8
表 4: 舆情因子和作为补充的其它因子	11
表 5: 舆情因子测试 - 沪深 300	14
表 6: 舆情因子测试 - 中证 500	14
表 7: 舆情因子测试 - 中证国债	15
表 8: 舆情因子测试 - 贵金属	15
表 9: 舆情因子测试 - 农产品	16
表 10: 舆情因子测试 - 标普 500	16
表 11: 舆情因子测试 - 货币基金	17
表 12: 相关数据	17
表 13: 回测结果分年度展示	18
表 14: 模型最新配置建议	19

# 一、网络舆情数据介绍

## 1.1 大类资产配置与互联网舆情数据

当前，我们正处于互联网大数据时代，我们所面对数据的维度在不断增加。除了传统新闻报道、日益火热的自媒体外，我们自身在网络的交互活动中也在不停的产生数据，如浏览数据、点击数据、搜索数据、留言数据等。这些原本零散碎片的信息，综合起来就成为一个全新的数据源，等待着人们去挖掘其中的价值。以资产配置问题而言，传统的研究范式都聚焦于基本的宏观数据、财务数据、资产历史行情数据等，研究互联网舆情数据在资产配置问题中的应用相对较为缺乏。本报告中我们试图将舆情与大类资产配置问题相结合，通过对不同资产舆情数据的分析，提升资产配置的效果。

目前以搜索引擎的搜索数据为基础构建的舆情指数主要有百度指数、360指数、谷歌趋势（Google Trends）等。百度指数以全球最大的中文搜索引擎百度的搜索数据为基础构建，360指数以搜索引擎360搜索数据为基础构建，据最新的《中国网民搜索行为调查报告》，百度搜索的常用率遥遥领先（手机端88.6%，PC端84.3%），360搜索次之（手机端2.6%，PC端11.9%），两者占据了国内搜索引擎市场的90%以上；谷歌趋势以全球最大搜索引擎Google的搜索数据为基础构建，但在国内常用率较低。基于用户搜索行为，通过后台数据分析得出的某段时间内关键词的搜索热度，代表着某段时间，人们对某话题的关注程度。在本报告中我们将资产对应的搜索舆情指数作为投资者情绪的直接代理变量，作为一个直接的舆情数据来源。

图 1：百度指数：代表了某段时间内关键词的搜索热度



数据来源：百度指数网站、广发证券发展研究中心

表1: 舆情数据来源一览

指数	含义	功能模块	数据区间	上线时间
百度指数	以网民在百度的搜索量为数据基础,以关键词为统计对象,科学分析并计算出各个关键词在百度网页搜索中搜索频次的加权和。	趋势研究、需求图谱、 舆情洞察、人群画像	2011 至今	2007.11
360 指数	以网民在 360 搜索的搜索量为数据基础,计算出各个关键词在 360 搜索网页搜索中搜索频次的加权和。	走势分析、需求图谱、 人群特征	2013 至今	2014.5
谷歌趋势	基于搜索日志分析的应用产品,通过分析 Google 全球数以十亿计的搜索结果,给出某一搜索关键词各个时期下在 Google 被搜索的频率和相关统计数据。	热度变化趋势、区域搜索 热度、相关主题/查询	2004 至今	2006.5

数据来源: 广发证券发展研究中心

## 1.2 舆情数据可预测性分析

金融市场上有许多传统金融理论所无法解释的异常现象,比如特殊时间段出现超额收益的日历效应、小市值股票更容易有超额收益的规模效应等。学术上将其归类于行为金融学范畴,从人们的投资心理、行为特征等角度进行分析。如认知心理学认为投资者容易过于自信,盲目冲动,表现出“追逐热点的羊群效应”、“追涨杀跌”等现象,而这些都与经典经济学理论中的“理性人假设”相矛盾。而要分析投资者的情绪变化,一般的基本面数据、宏观数据、资产行情数据等都无能为力。

互联网舆情数据则恰巧不同,通过文本数据挖掘、情感分析、用户行为分析等得到的舆情数据可以从一个侧面及时的描述投资者的情绪面。舆情数据范围十分广泛,比如投资者对不同话题的关注度、投资者对不同事件的情感倾向、不同事件之间的相互关联等。这些舆情数据都刻画了投资者情绪的不同方面,本篇专题报告中重点讨论了由资产对应的搜索舆情数据所代表市场关注度数据对资产收益率的预测能力。

市场关注度舆情指数本身就是一种群体智慧,携带了市场关注热点的信息。投资者情绪随着资产价格变化而波动,同时也反过来影响资产价格。如当市场出现某一热点时,投资者关注度上升,上升的热度带动资产价格上涨。上涨的价格引来更多的关注度,形成一种正反馈机制。而当投资者热度上升到一个高点时,无法吸引更多的投资者,之前的投资者逐渐降低关注,随后资产价格开始下跌。因此通过对于舆情因子变化趋势的分析,可以去探究投资者关注度的变化情况,从而对资产收益率的变化进行预判。

## 1.3 现有相关研究

现有相关研究主要着眼于投资者情绪对股票市场的影响,如: Baker and Wurgler (2006) 第一次讨论了投资者情绪和股票截面收益率间的联系,发现由换手率、新股回报率、股息率等六个因子构成的投资者情绪指数与股票截面收益率间的反转效应。PC Tetlock (2007) 则讨论了媒体新闻和股票收益率间的关系,研究不同悲观程度的新闻报导对相应股票收益率的影响。Bollen, Mao and Zeng (2010) 研究



了社交媒体 Twitter 上评论信息的情感倾向与道琼斯股指的关系，通过文本情感分析方法得到市场某时间段情感指数与股指的变化关系。Azar and Lo (2017) 着眼于特定事件，在 FOMC 会议期间分析 Twitter 上的相关评论数据对股指的影响。

而将投资者情绪与资产配置问题相结合的研究目前还相对较少，所以本报告将互联网关注度数据作为投资者情绪的直接代理变量，试图分析舆情数据在资产配置问题中的应用。

## 1.4 大类资产舆情数据

在大类资产配置的资产中，本篇专题报告构建了以权益类、债券类、商品类、货币类四类资产组成的资产池，具体组成为：

- 1) 沪深300指数：代表权益类国内大市值股票资产。
- 2) 中证500指数：代表权益类国内小市值股票资产。
- 3) 标普500指数：代表权益类海外股票资产。
- 4) 中证国债指数：代表债券类资产。
- 5) 南华农产品指数：代表商品类农产品资产。
- 6) 万得贵金属指数：代表商品类贵金属资产。
- 7) 万得货币基金指数：代表现金类资产。

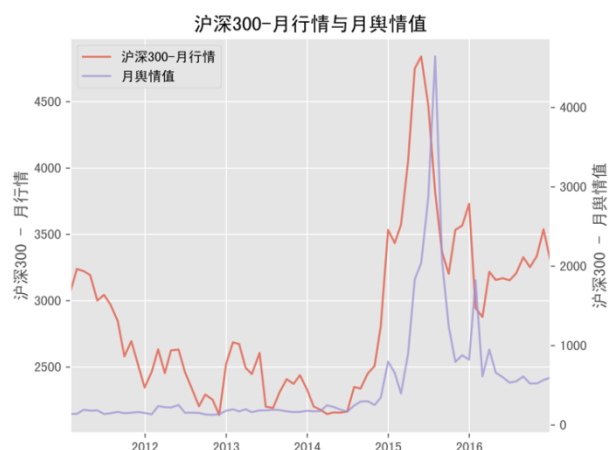
本篇专题报告同时抓取了2011 ~ 2017 时间段七种资产的对应的搜索舆情数据，将资产该月的对应的搜索舆情数据累计值作为月舆情值，观察不同资产行情数据与月舆情值间关系。

表 2：大类资产舆情数据示例

日期	沪深 300	中证 500	中证国债	贵金属	农产品	标普 500	货币基金
2015-01-01	1129	258	1259	1188	917	254	1514
2015-01-02	935	238	1544	1457	853	218	1232
2015-01-03	1024	295	1444	1179	933	205	1675
2015-01-04	1695	313	1705	1539	1405	229	1886
2015-01-05	4061	699	1736	1819	2076	290	2101
2015-01-06	3377	700	1712	1755	1878	309	1960
2015-01-07	3159	576	1723	1798	1842	342	1925
2015-01-08	2793	566	1706	1754	1752	264	1894
2015-01-09	2872	477	1625	1744	1575	250	1845
2015-01-10	1268	270	1471	1051	795	235	1611

数据来源：广发证券发展研究中心

图 2：沪深300指数月行情与月舆情值关系



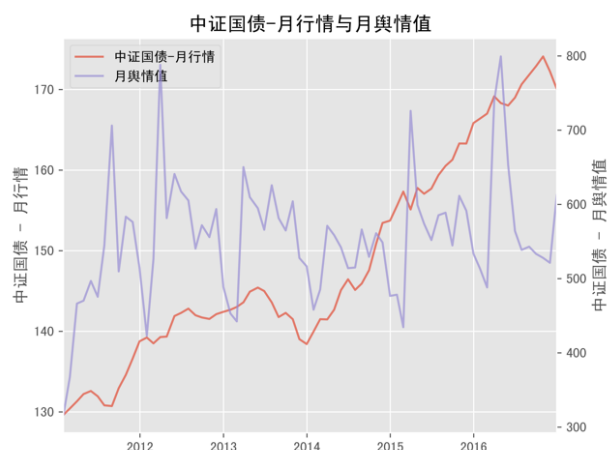
数据来源：Wind、广发证券发展研究中心

图 3：中证500指数月行情与月舆情值关系



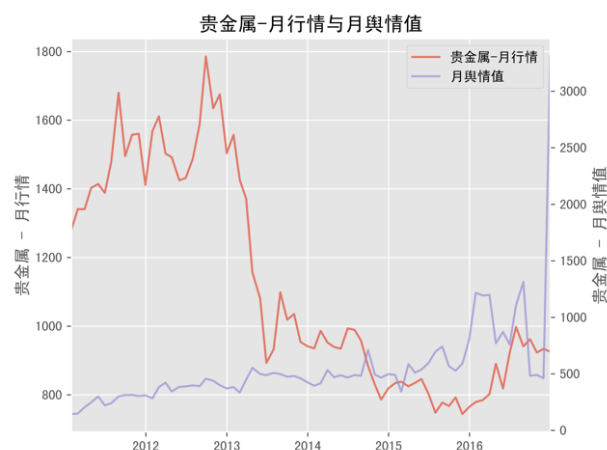
数据来源：Wind、广发证券发展研究中心

图 4：中证国债指数月行情与月舆情值关系



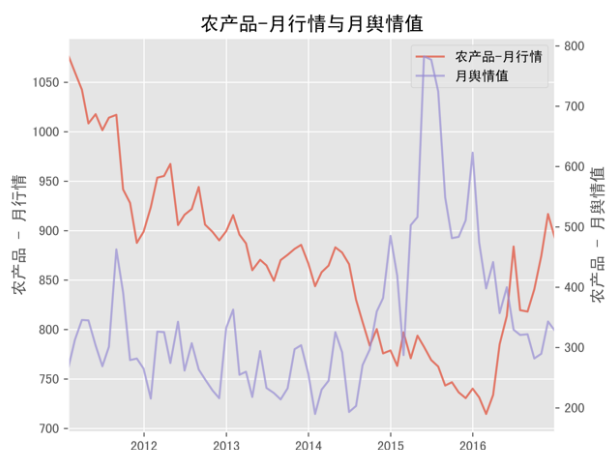
数据来源：Wind、广发证券发展研究中心

图 5：万得贵金属指数月行情与月舆情值关系



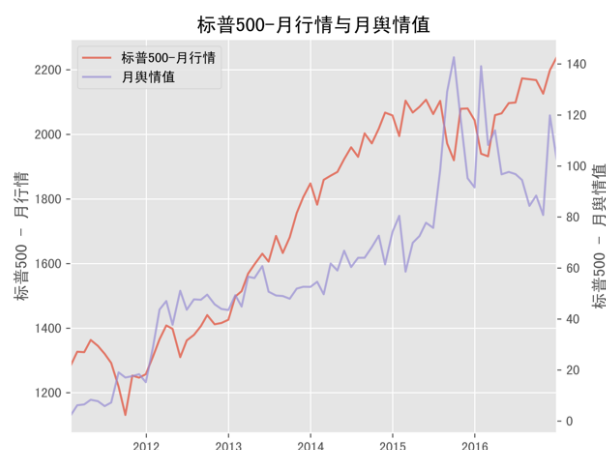
数据来源：Wind、广发证券发展研究中心

图 6：南华农产品指数月行情与月舆情值关系



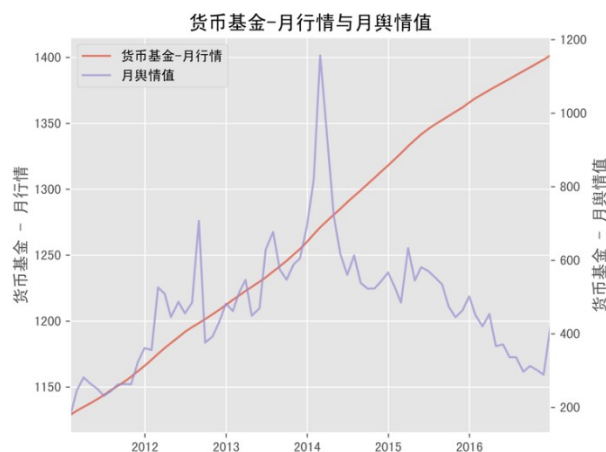
数据来源：Wind、广发证券发展研究中心

图 7：标普500指数月行情与月舆情值关系



数据来源：Wind、广发证券发展研究中心

图 8：货币基金指数月行情与月舆情值关系



数据来源：Wind、广发证券发展研究中心

表 3：各类资产月舆情值与月行情

资产	月舆情与月行情相关系数
沪深300	0.538
中证500	0.555
中证国债	0.217
贵金属	-0.402
农产品	-0.524
标普500	0.826
货币基金	0.185

数据来源：Wind、广发证券发展研究中心

从上面的结果可以看出，资产对应的搜索舆情数据与各类资产行情数据的相关性有明显差异。资产对应的搜索舆情数据与权益类资产如沪深300指数、中证500指数、标普500指数呈现正相关关系，而与债券类、货币类资产的相关性较弱，与商品类相关性为负相关。因此在利用舆情数据进行资产配置时，针对不同的资产本篇专题报告采用不同的策略。

## 二、再谈大类资产配置模型

### 2.1 BL 模型介绍

BL模型是由高盛的 Fisher Black 和 Robert Litterman 在1992年提出的资产配置模型。与传统均值-方差模型不同，BL模型采用贝叶斯方法，将市场均衡收益率作为先验分布，投资者对资产的预期收益率作为似然函数，得到结合了投资者的主



观观点的资产后验收益率分布。并根据得到的资产后验收益率，在约束条件下得到不同资产的配置比例。

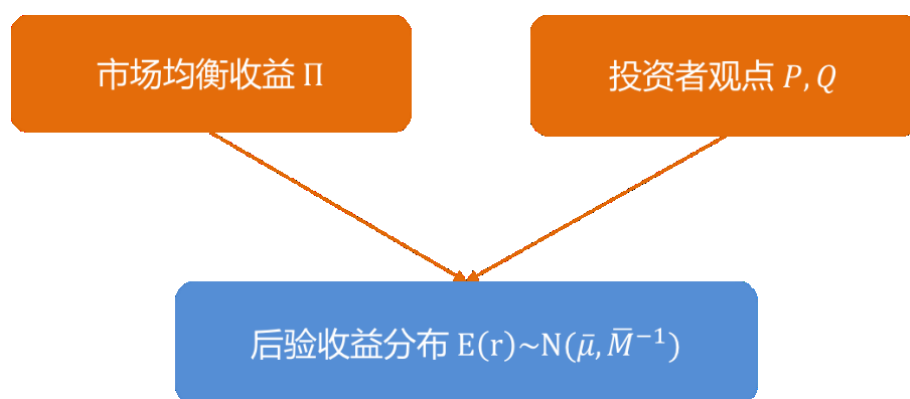
模型一般使用 CAPM 模型的结果  $\Pi = \lambda \Sigma w_{eq}$  作为市场均衡收益  $\Pi$ ，其中  $\lambda$  为风险厌恶系数， $\Sigma$  为资产超额收益率的协方差矩阵， $w_{eq}$  为资产市值比例向量。投资者观点矩阵  $P$  和投资者预期收益率  $Q$  向量可以是对某单一资产的预期收益率，也可以是预期的两种资产收益率的差。模型允许投资者输入对于不同观点的置信度，以提升模型对于输入参数的鲁棒性。如果投资者对某类资产收益率并无特别的预期，可以用市场均衡收益代替。BL模型中资产后验收益率服从以  $\bar{\mu}$  为均值、 $\bar{M}^{-1}$  为方差的正态分布，二者的表达式为：

$$\begin{cases} \bar{\mu} = [(\tau \Sigma)^{-1} + P^T \Omega^{-1} P]^{-1} [(\tau \Sigma)^{-1} \Pi + P^T \Omega^{-1} Q] \\ \bar{M}^{-1} = [(\tau \Sigma)^{-1} + P^T \Omega^{-1} P]^{-1} \end{cases}$$

设一共有  $n$  种资产，投资者有  $k$  种观点，上式中各变量的含义分别为：

- 1)  $P$  表示投资者观点矩阵，是  $k \times n$  维矩阵，表示  $k$  种对不同资产预期收益率的观点。
- 2)  $Q$  表示投资者预期收益率，是  $k$  维向量，代表了每种观点中的预期收益率大小。
- 3)  $\Sigma$  表示资产价格协方差矩阵，是  $n \times n$  维矩阵，一般使用各类资产的历史超额收益率协方差矩阵。
- 4)  $\tau$  表示先验观点的不确定度，是个标量参数。
- 5)  $\Omega$  表示投资者主观观点的不确定度，是  $k$  维对角阵，表示对每种观点的信心水平。
- 6)  $\Pi$  表示作为先验的市场均衡收益率，一般用 CAPM 均衡收益  $\Pi = \lambda \Sigma w_{eq}$ ，其中  $w_{eq}$  是资产的市值比例， $\lambda$  是投资者风险厌恶系数。
- 7)  $\bar{\mu}$  表示作为后验资产预期收益率的期望，是  $n$  维向量。
- 8)  $\bar{M}^{-1}$  表示作为后验资产预期收益率的协方差矩阵，是  $n \times n$  维矩阵。

图 9: Black-Litterman 模型主要思想

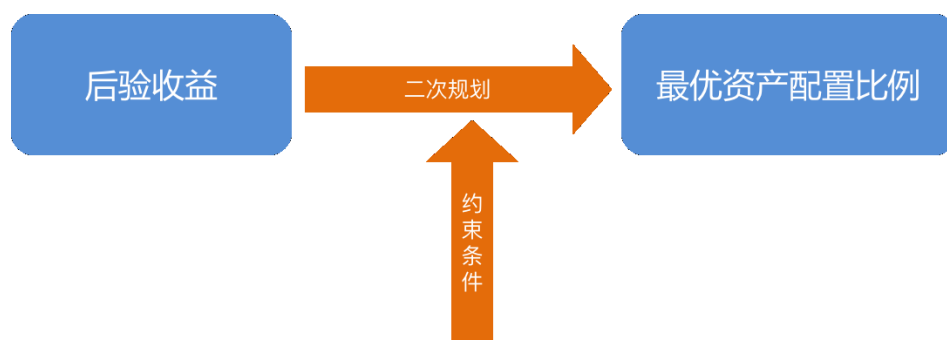


数据来源：广发证券发展研究中心

得到资产的后验收益率求最优化时，可以根据实际情况添加约束条件，如：非卖空条件、各资产持仓比例和为1、单资产持仓下限、单资产持仓上限、组合方差上限等。于是将资产配置问题转为二次规划问题：

$$\begin{cases} \text{Max} (w^T \bar{\mu} - \frac{\lambda}{2} w^T \bar{M}^{-1} w) \\ \text{s.t. : } \Sigma w = 1 \\ w_i \leq \text{max} w \\ w_i \geq \text{min} w \\ w^T \bar{M}^{-1} w \leq \text{max} \sigma \end{cases}$$

图 10: Black-Litterman模型配置方法



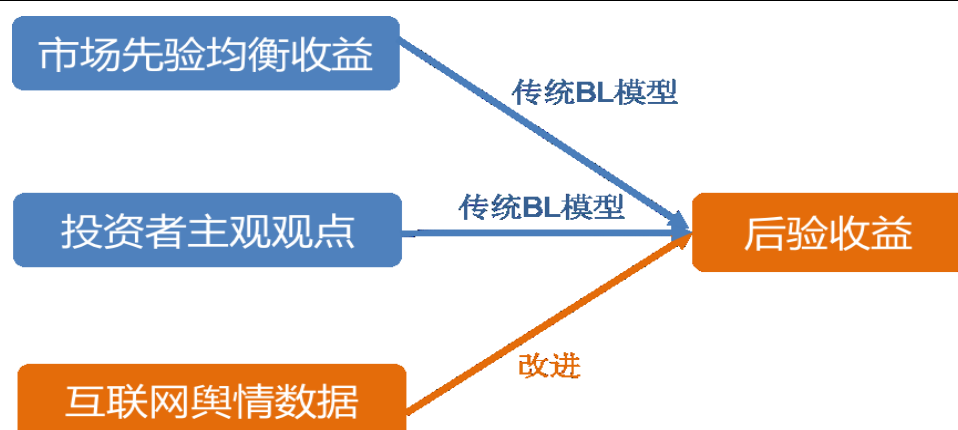
数据来源：广发证券发展研究中心

## 2.2 传统 BL 模型的局限及改进

传统BL模型中需要投资者输入自身对不同资产收益率的主观观点，因此对投资者自身提出了一个较高要求。因为投资者往往由于自身背景知识、熟悉领域有限等局限，很难对所有资产都给出较准的预期收益率，而当投资者提出错误的资产预期收益率观点时往往会使得模型给出更不准确的结果。使得传统 BL 模型在实践中可操作性不高。

本报告中我们试图在传统BL模型基础上，结合互联网舆情数据。通过分析互联网舆情数据，捕捉投资者情绪的变动，希望得到更准确的资产预期收益率，从而实现更好的资产配置效果。对BL模型而言，资产的后验收益率分布是决定资产配置比例的重要参数。而投资者的主观观点P，对资产的预期收益率Q是决定后验收益的关键变量。本报告中，我们通过对资产所对应的搜索舆情数据的分析，结合投资者主观观点得到预期收益率Q向量。

图 11: 改进后BL模型



数据来源：广发证券发展研究中心

对于  $n$  种资产而言，如果每种资产给出一个预期收益率，则后验预期收益率  $Q$  向量就是  $n$  维向量。对于确定的资产收益率如现金存款等，可以直接给出确定的预期收益率，而对收益率不确定的资产，则将改进后BL模型得出的结果作为资产的预期收益率。

### 三、舆情因子测试

#### 3.1 舆情因子构建及筛选

为了更好的预测资产收益率，本篇专题报告对资产对应的搜索舆情数据进行初步处理，从增加因子的多样性角度考虑，把因子分成 3 大类：

- 1) 基于各大类资产对应的搜索舆情数据构造的纯舆情因子。
- 2) 结合了舆情数据的动量因子。
- 3) 作为补充的传统因子。

报告中采用多元线性回归的方法，以 12 月作为滑动窗口期，每期根据因子的 IC 值、IR 值筛选当期选用因子进行多元线性回归，得到资产下一月的预期收益率。

表 4: 舆情因子和作为补充的其它因子

因子分类	因子指标	因子描述
舆情因子	近1月舆情值	当月前1月的舆情累积值
	近2月舆情值	当月前2月的舆情累积值
	近3月舆情值	当月前3月的舆情累积值
	近1月舆情值波动率	当月前1月的日舆情值波动率
	近2月舆情值波动率	当月前2月的日舆情值波动率
	近3月舆情值波动率	当月前3月的日舆情值波动率
	近1月舆情值增长率	当月前1月的舆情累积值增长率
	近2月舆情值增长率	当月前2月的舆情累积值增长率
	近3月舆情值增长率	当月前3月的舆情累积值增长率
	近1月舆情值增长率变化率	当月前1月的舆情累积值增长率的变化率

结合舆情因子的动量因子	近2月舆情值增长率变化率	当月前2月的舆情累积值增长率的变化率
	近3月舆情值增长率变化率	当月前3月的舆情累积值增长率的变化率
	近1月舆情值加权日收益率	当月前1月以日舆情值为权重得到的加权日收益率
	近2月舆情值加权日收益率	当月前2月以日舆情值为权重得到的加权日收益率
	近3月舆情值加权日收益率	当月前3月以日舆情值为权重得到的加权日收益率
	近1月舆情值加权日收益率波动率	当月前1月以日舆情值为权重得到的加权日收益率波动率
	近2月舆情值加权日收益率波动率	当月前2月以日舆情值为权重得到的加权日收益率波动率
作为补充的其他传统因子	近3月舆情值加权日收益率波动率	当月前3月以日舆情值为权重得到的加权日收益率波动率
	对数市值	大类资产市值的对数
	近1月收益率	资产前1月收益率
	近2月收益率	资产前2月收益率
	近3月收益率	资产前3月收益率
	近1月日收益率波动率	当月前1月日收益率的标准差
	近2月日收益率波动率	当月前2月日收益率的标准差
	近3月日收益率波动率	当月前3月日收益率的标准差

数据来源：广发证券发展研究中心

### 3.2 数据标准化

因为构建的各因子数据量纲不同，为了回归时方便比较，先对数据进行标准化处理。以及为了减少异常值对预测结果带来负面影响，对数据先去极值后进行标准化操作。

1). 去极值处理：以 2.5 倍标准差作为阈值，其中 $\bar{x}$ 为均值， $\sigma$ 为标准差。

$$x = \begin{cases} \bar{x} + 2.5\sigma, & x < \bar{x} + 2.5\sigma \\ \bar{x} - 2.5\sigma, & x < \bar{x} - 2.5\sigma \\ x, & \text{else} \end{cases}$$

2). 标准化处理：将窗口期的因子时间序列数据变换成均值为0，标准差为1的新时间序列：

$$x_{new} = (x_{old} - \bar{x})/\sigma_x$$

### 3.3 舆情因子测试

因为资产池中的七种资产间差异性较大，在进行回归前本篇专题报告逐一对每类资产筛选因子。取窗口期为N月，对资产*i*而言在第t月对因子*f*有N维因子值向量 $F = (f_1, f_2, \dots, f_N)$ ，资产*i*对应的下月超额收益率向量 $R = (R_1, R_2, \dots, R_N)$ 。定义因子IC值为因子值向量*F*和资产下月收益率向量*R*的相关系数：

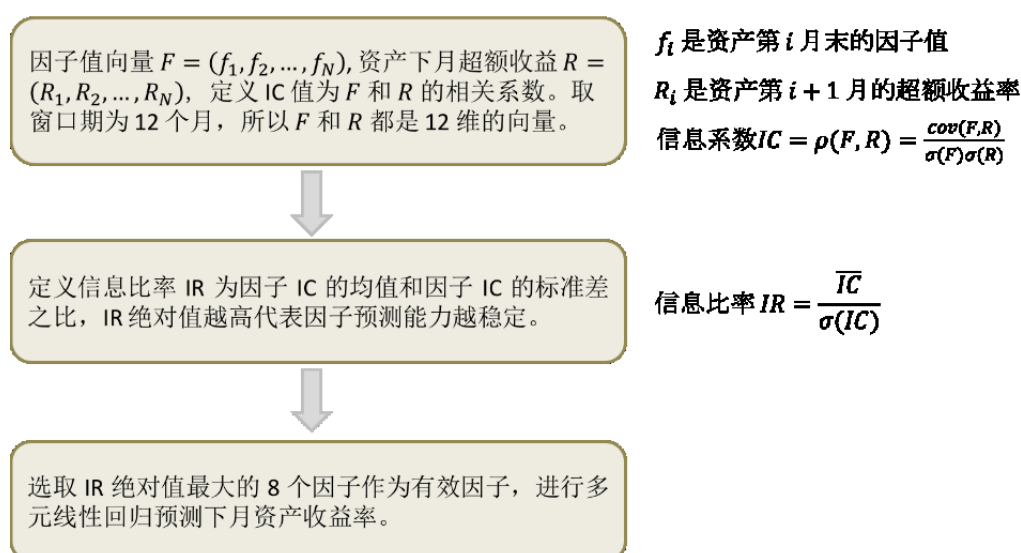
$$IC = \rho_{F,R} = \frac{cov(F,R)}{\sigma_F \sigma_R}$$

同时定义在第  $t$  月，因子信息比率  $IR$  值为前  $N'$  月的因子  $IC$  均值与因子  $IC$  值的标准差之比：

$$IR = \frac{\overline{IC}}{\sigma_{IC}}$$

因子的信息比率  $IR$  值绝对值越高代表在该段时期的因子预测能力越稳定，在每月末计算各因子的信息比率  $IR$  值，取其中绝对值最大的 8 个因子作为当期筛选出的有效因子，利用有效因子进行多元线性回归求出资产的预期收益率作为 BL 模型中的投资者主观收益率。

图 12：舆情因子测试流程



数据来源：广发证券发展研究中心

在第 2017-01-31 期截面对不同种类资产按照上述方法筛选出的有效因子展示如下：



表 5: 舆情因子测试 - 沪深300

因子	IR值	IC均值
近3月舆情值增长率	-1.066	-0.169
近1月舆情值	-0.874	-0.255
近1月舆情值加权日均收益率	0.868	0.161
近1月舆情值波动率	-0.853	-0.249
近1月舆情值增长率变化率	-0.778	-0.139
近2月日收益率标准差	-0.757	-0.145
近3月舆情值	-0.748	-0.224
近2月舆情值	-0.741	-0.218

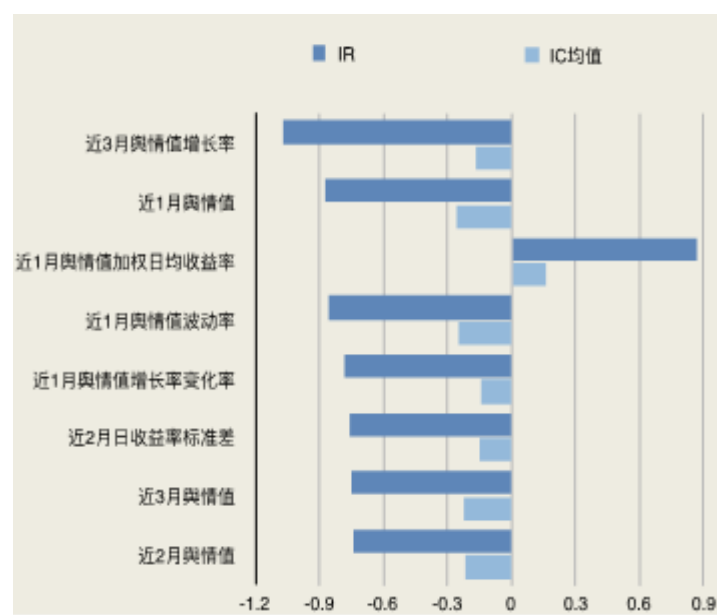
数据来源：广发证券发展研究中心

表 6: 舆情因子测试 - 中证500

因子	IR值	IC均值
近2月舆情值增长率	1.28	0.208
近3月舆情值	-0.729	-0.18
近1月舆情值增长率变化率	-0.631	-0.137
近3月收益率	-0.594	-0.129
近2月舆情值波动率	-0.584	-0.138
近1月舆情值	-0.566	-0.13
近3月舆情值波动率	-0.524	-0.108
近3月舆情值增长率	0.523	0.114

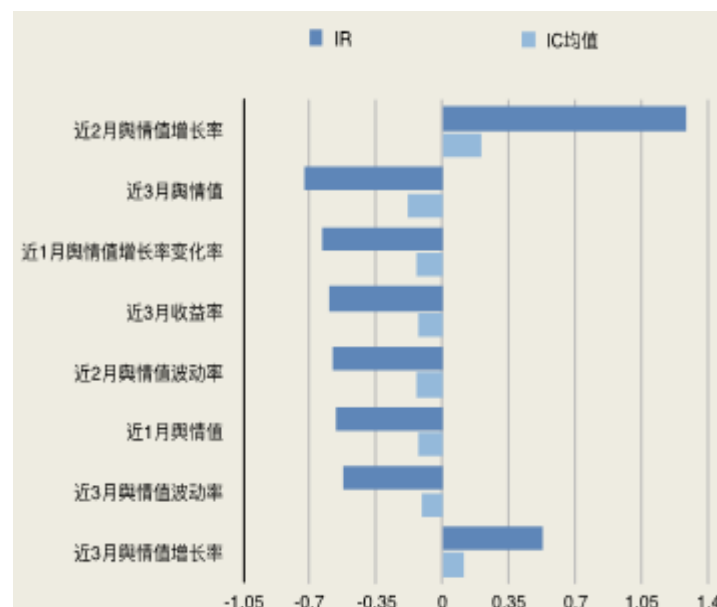
数据来源：广发证券发展研究中心

图 13: 舆情因子测试 - 沪深300



数据来源：广发证券发展研究中心

图 14: 舆情因子测试 - 中证500



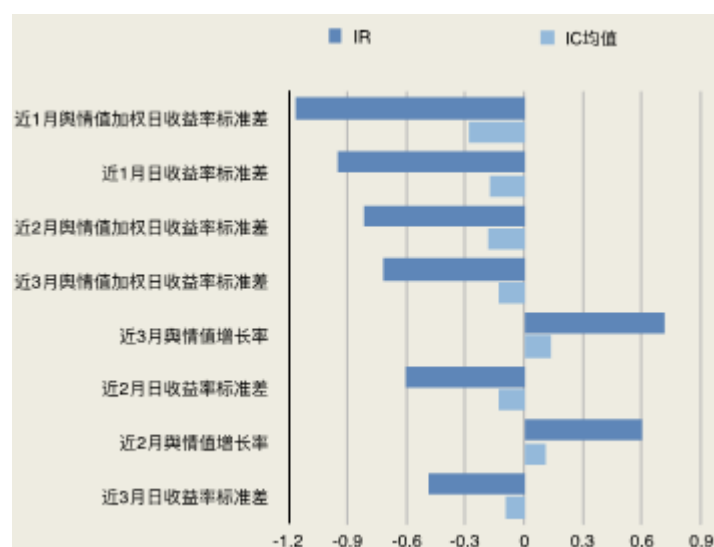
数据来源：广发证券发展研究中心

表 7: 舆情因子测试 - 中证国债

因子	IR值	IC均值
近1月舆情值加权日收益率标准差	-1.162	-0.278
近1月日收益率标准差	-0.952	-0.177
近2月舆情值加权日收益率标准差	-0.819	-0.185
近3月舆情值加权日收益率标准差	-0.72	-0.133
近3月舆情值增长率	0.713	0.137
近2月日收益率标准差	-0.604	-0.127
近2月舆情值增长率	0.6	0.106
近3月日收益率标准差	-0.486	-0.093

数据来源: 广发证券发展研究中心

图 15: 舆情因子测试 - 中证国债



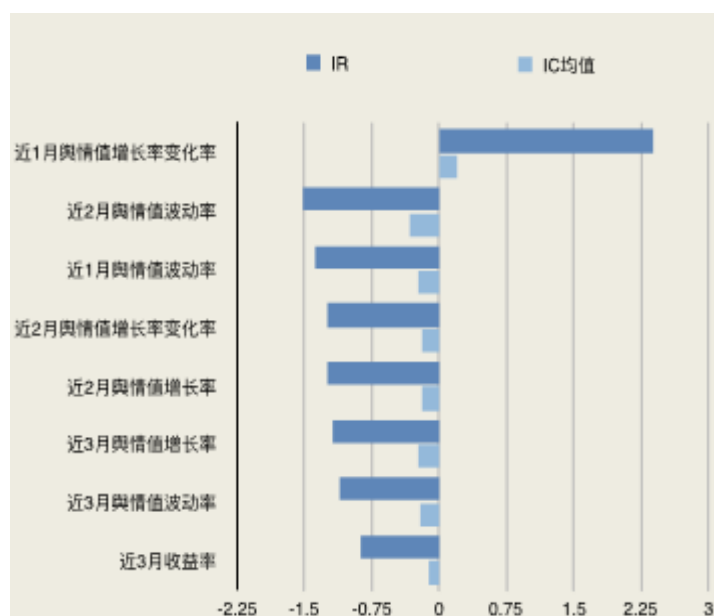
数据来源: 广发证券发展研究中心

表 8: 舆情因子测试 - 贵金属

因子	IR值	IC均值
近1月舆情值增长率变化率	2.39	0.21
近2月舆情值波动率	-1.512	-0.321
近1月舆情值波动率	-1.383	-0.218
近2月舆情值增长率变化率	-1.232	-0.179
近2月舆情值增长率	-1.23	-0.182
近3月舆情值增长率	-1.171	-0.217
近3月舆情值波动率	-1.106	-0.207
近3月收益率	-0.864	-0.118

数据来源: 广发证券发展研究中心

图 16: 舆情因子测试 - 贵金属



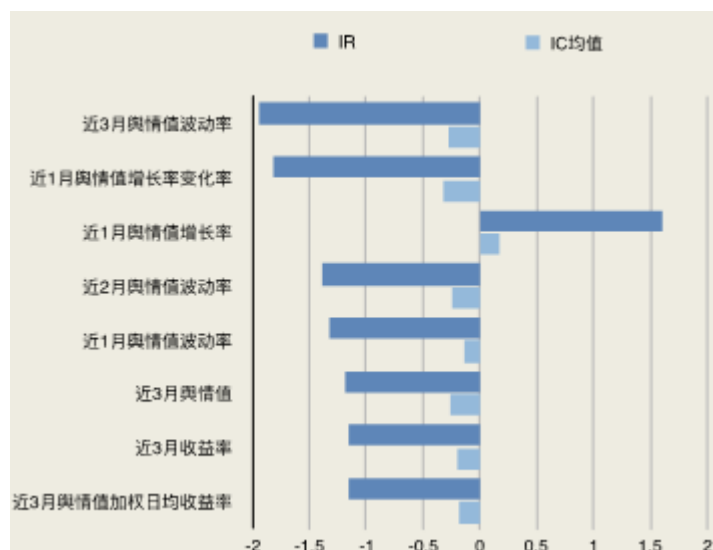
数据来源: 广发证券发展研究中心

表 9: 舆情因子测试 - 农产品

因子	IR值	IC均值
近3月舆情值波动率	-1.938	-0.277
近1月舆情值增长率变化率	-1.81	-0.323
近1月舆情值增长率	1.597	0.17
近2月舆情值波动率	-1.388	-0.238
近1月舆情值波动率	-1.325	-0.14
近3月舆情值	-1.184	-0.262
近3月收益率	-1.161	-0.203
近3月舆情值加权日均收益率	-1.158	-0.19

数据来源: 广发证券发展研究中心

图 17: 舆情因子测试 - 农产品



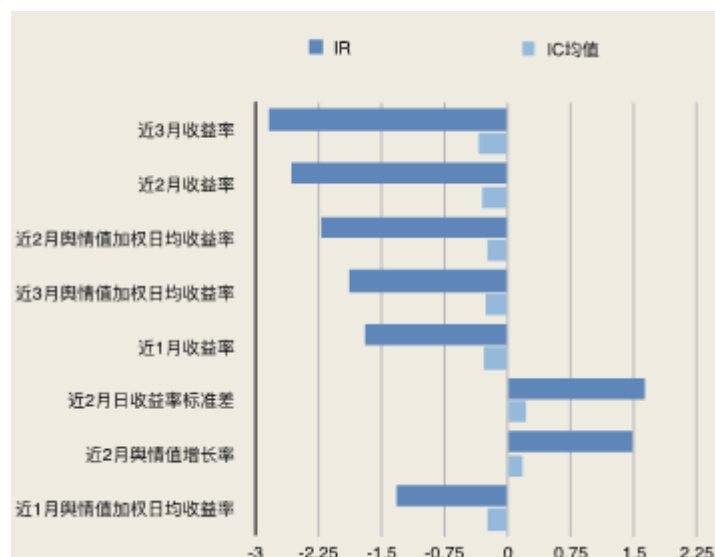
数据来源: 广发证券发展研究中心

表 10: 舆情因子测试 - 标普500

因子	IR值	IC均值
近3月收益率	-2.831	-0.332
近2月收益率	-2.564	-0.306
近2月舆情值加权日均收益率	-2.205	-0.245
近3月舆情值加权日均收益率	-1.884	-0.264
近1月收益率	-1.699	-0.283
近2月日收益率标准差	1.621	0.219
近2月舆情值增长率	1.477	0.173
近1月舆情值加权日均收益率	-1.327	-0.23

数据来源: 广发证券发展研究中心

图 18: 舆情因子测试 - 标普500



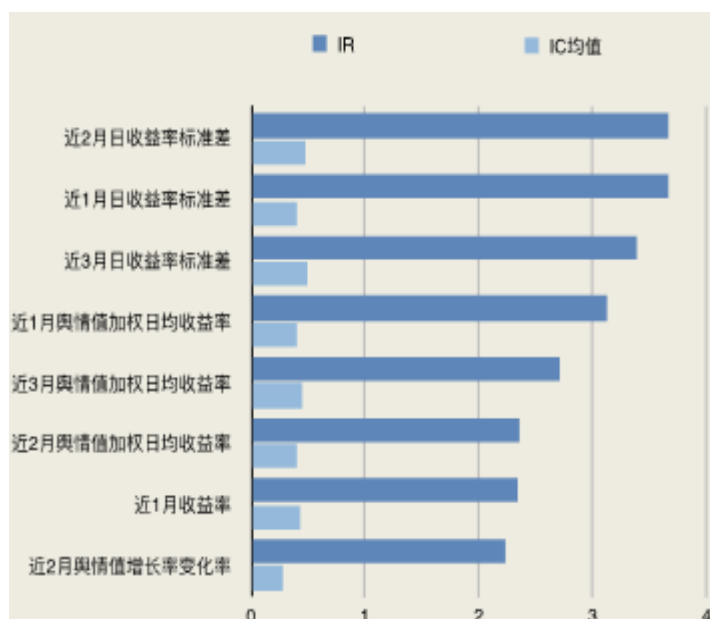
数据来源: 广发证券发展研究中心

表 11: 舆情因子测试 - 货币基金

因子	IR值	IC均值
近2月日收益率标准差	3.67	0.47
近1月日收益率标准差	3.662	0.392
近3月日收益率标准差	3.382	0.492
近1月舆情值加权日均收益率	3.127	0.392
近3月舆情值加权日均收益率	2.71	0.442
近2月舆情值加权日均收益率	2.35	0.403
近1月收益率	2.34	0.425
近2月舆情值增长率变化率	2.239	0.269

数据来源: 广发证券发展研究中心

图 19: 舆情因子测试 - 货币基金



数据来源: 广发证券发展研究中心

## 四、实证分析

### 4.1 实证分析

在实证分析中选取了沪深300指数、中证500指数、标普500指数、中证国债指数、南华农产品指数、贵金属指数、货币基金指数七种资产分别作为权益类、债券类、商品类、现金类资产构成资产池。

资产市值数据中的沪深300指数、中证500指数、标普500指数用当期人民币市值，中证国债指数市值选用银行间国债净值，南华农产品指数、万得贵金属指数市值选用商品期货合约的双边准备金总和、万得货币基金指数市值选用当期货币基金净值总和。无风险利率采用当期银行间一年期存款利率。

回测时间是 2013 ~ 2017 区间，滑动窗口期选为 12 个月，计算信息比率IR时选用IC因子的窗口期为 12 个月。

表 12: 相关数据

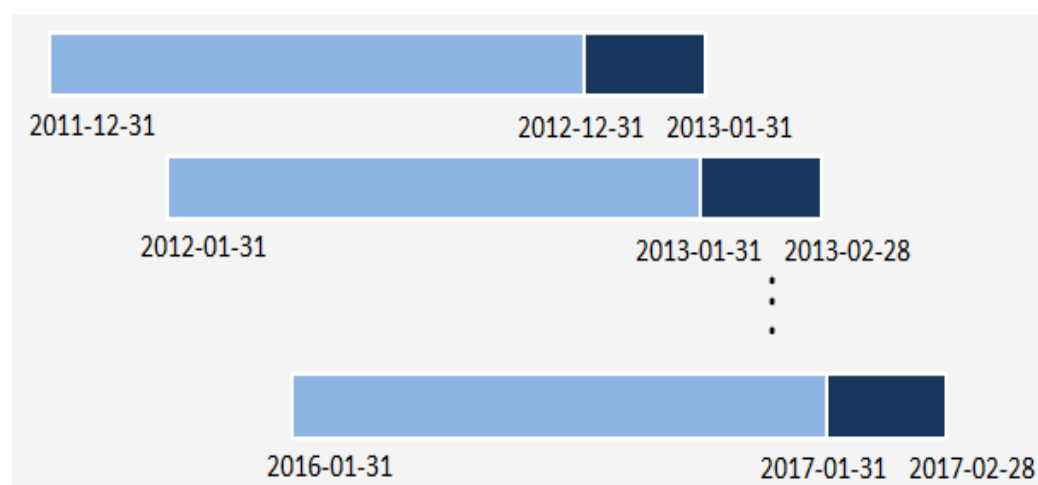
数据参数	具体说明
大类资产类别	沪深300、中证500、中证国债、贵金属、农产品、标普500、货币基金
资产市值	指数的相应市值、银行间国债的净值、商品期货合约的双边准备金总和、货币基金净值总和

回测时间	2013-01-01 ~ 2017-05-28
回测窗口长度	12个月
调仓频率	月度调仓
无风险利率	银行一年期存款利率

数据来源：广发证券发展研究中心

实证分析时以 12 个月为窗口期，对第  $t$  期前 12 个月的数据进行回归，以第  $t$  期的因子数据去预测第  $t+1$  期的收益率。预测的结果作为 Black-Litterman 模型中的资产预期收益率，得出第  $t+1$  期的资产配置比例。

图 20：实证分析示意图



数据来源：广发证券发展研究中心

## 4.2 回测结果

从回测结果看，在 2013~2017 的回测区间内，模型取得了 16.18% 的年化收益率，16.02% 的年化波动率，14.28% 的最大回撤，夏普比率为 1.02。

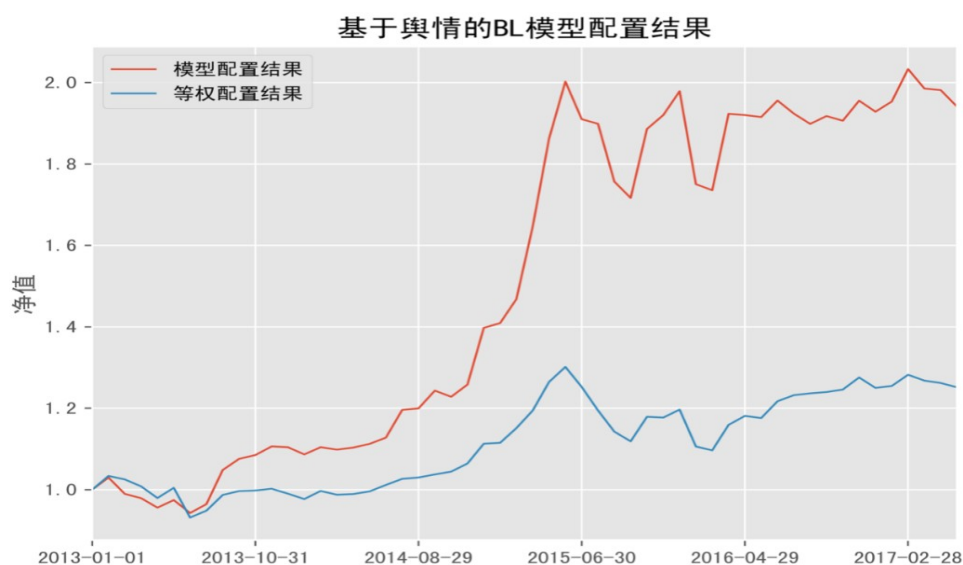
表 13：回测结果分年度展示

	2013年	2014年	2015年	2016年	2017年至今	整体
年化收益率	10.43%	26.57%	41.58%	-2.54%	0.64%	16.18%
最大回撤	8.46%	1.23%	14.28%	2.94%	4.55%	14.28%
年化波动率	11.82%	12.30%	22.63%	17.15%	9.18%	16.02%
夏普比率	0.90	1.99	1.66	-0.07	0.20	1.02

数据来源：广发证券发展研究中心



图 21：回测净值结果



数据来源：广发证券发展研究中心

### 4.3 模型最新配置建议

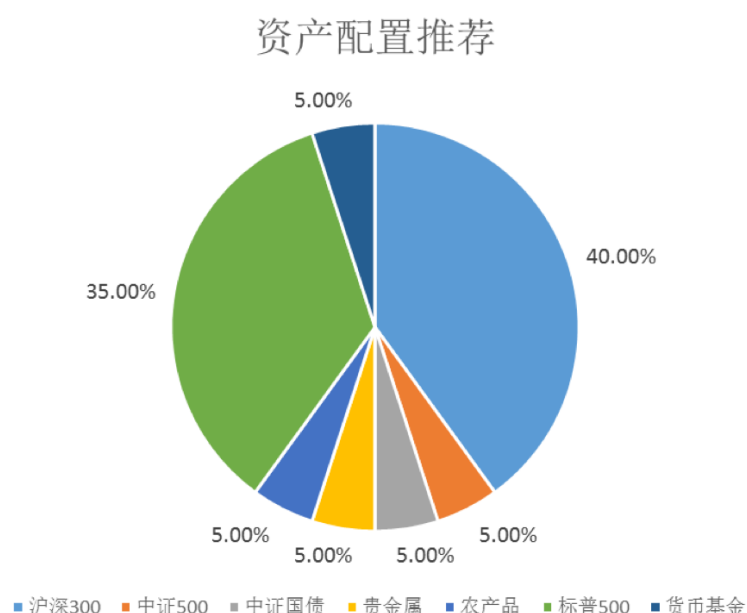
根据最新资产表现及资产对应的搜索舆情数据，模型给出如下的资产配置建议，其中沪深300指数配置比例为40%；中证500指数配置比例为5%；中证国债指数配置比例为5%；万得贵金属指数配置比例为5%；南华农产品指数配置比例为5%；标普500指数配置比例为35%；万得货币基金指数配置比例为5%。

表 14：模型最新配置建议

资产	配置比例
沪深300	40%
中证500	5%
中证国债	5%
贵金属	5%
农产品	5%
标普500	35%
货币基金	5%

数据来源：广发证券发展研究中心

图 22：模型最新配置建议



数据来源：广发证券发展研究中心

## 五、总结

经典的资产配置模型如 Black-Litterman 模型，对投资者本身要求较高，投资者错误的观点会使得配置结果更不准确。本篇专题报告通过对互联网舆情数据进行分析，探寻市场关注度变化与资产收益率间变化关系。构造了多个舆情因子、结合舆情的传统因子并进行测试，以回归的方法预测资产的预期收益率，作为 BL 模型中的投资者主观收益率，结合投资者主观观点得到资产配置比例。经回测得出了较好的资产配置结果。

根据最新资产表现及资产对应的搜索舆情数据，模型给出2017年6月最新的资产配置建议，其中沪深300配置比例为40%；中证500配置比例为5%；中证国债配置比例为5%；贵金属配置比例为5%；农产品配置比例为5%；标普500配置比例为35%；货币基金持仓比率为5%。

## 风险提示

本专题内容基于资产舆情数据测算完成，资产舆情是影响资产表现的一类因素，资产表现受其他诸多因素影响，注意模型失效风险。

## 广发证券—行业投资评级说明

买入： 预期未来 12 个月内，股价表现强于大盘 10%以上。  
持有： 预期未来 12 个月内，股价相对大盘的变动幅度介于-10%~+10%。  
卖出： 预期未来 12 个月内，股价表现弱于大盘 10%以上。

## 广发证券—公司投资评级说明

买入： 预期未来 12 个月内，股价表现强于大盘 15%以上。  
谨慎增持： 预期未来 12 个月内，股价表现强于大盘 5%-15%。  
持有： 预期未来 12 个月内，股价相对大盘的变动幅度介于-5%~+5%。  
卖出： 预期未来 12 个月内，股价表现弱于大盘 5%以上。

## 联系我们

	广州市	深圳市	北京市	上海市
地址	广州市天河区林和西路9号耀中广场A座1401	深圳福田区益田路6001号太平金融大厦31楼	北京市西城区月坛北街2号月坛大厦18层	上海浦东新区世纪大道8号国金中心一期16层
邮政编码	510620	518000	100045	200120
客服邮箱	gfyf@gf.com.cn			
服务热线				

## 免责声明

广发证券股份有限公司（以下简称“广发证券”）具备证券投资咨询业务资格。本报告只发送给广发证券重点客户，不对外公开发布，只有接收客户才可以使用，且对于接收客户而言具有相关保密义务。广发证券并不因相关人员通过其他途径收到或阅读本报告而视其为广发证券的客户。本报告的内容、观点或建议并未考虑个别客户的特定状况，不应被视为对特定客户关于特定证券或金融工具的投资建议。本报告发送给某客户是基于该客户被认为有能力独立评估投资风险、独立行使投资决策并独立承担相应风险。

本报告所载资料的来源及观点的出处皆被广发证券股份有限公司认为可靠，但广发证券不对其准确性或完整性做出任何保证。报告内容仅供参考，报告中的信息或所表达观点不构成所涉证券买卖的出价或询价。广发证券不对因使用本报告的内容而引致的损失承担任何责任，除非法律法规有明确规定。客户不应以本报告取代其独立判断或仅根据本报告做出决策。

广发证券可发出其它与本报告所载信息不一致及有不同结论的报告。本报告反映研究人员的不同观点、见解及分析方法，并不代表广发证券或其附属机构的立场。报告所载资料、意见及推测仅反映研究人员于发出本报告当日的判断，可随时更改且不予通告。

本报告旨在发送给广发证券的特定客户及其它专业人士。未经广发证券事先书面许可，任何机构或个人不得以任何形式翻版、复制、刊登、转载和引用，否则由此造成的一切不良后果及法律责任由私自翻版、复制、刊登、转载和引用者承担。