

基于舆情的大类资产配置

史庆盛 S0260513070004
广发证券金融工程
邮箱：sqs@gf.com.cn
2017年6月

01

I

网络舆情
数据介绍

>

02

II

再谈大类资
产配置模型

>

03

III

舆情因子
测试

>

04

IV

实证分析

>

05

IV

结论

>



01

02

03

04

01

|网络舆情介绍|



1.1 互联网舆情数据

身处大数据时代，我们所面对的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，我们自身在与网络的交互活动，也在不停的产生数据。

浏览数据、点击数据、搜索数据、留言数据.....这些原本零散的碎片数据，综合起来就成为全新的数据源，等待人们去挖掘其中的价值。

一个典型就是互联网舆情数据，本报告尝试将互联网舆情数据与大类资产配置问题联系起来，提升传统模型的准确性。



1.2 舆情指数

百度作为最大的中文搜索引擎，在国内有着巨大的用户基数，是数亿网民上网的流量入口。

而百度指数是百度公司基于用户行为，通过数据分析得出的某关键词搜索热度规模。可以代表某段时间内，人们对某话题的关注程度，作为一个直接的舆情指标。

本报告以**资产的搜索舆情指数**作为**舆情指标**，探究舆情数据在大类资产配置中的应用。



1.3 舆情数据可预测性分析

行为金融学 & 非理性行为：

金融市场上有许多异常现象是传统金融理论所无法解释，比如日历效应、规模效应等。学术上将其分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。

从认知心理学分析，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。这些都与经典经济学理论中的“理性人假设”矛盾冲突。

舆情因子：

舆情搜索指数代表了人们对关键词的关注度情况，我们将这种关注度作为投资者情绪的**直接代理变量**（direct proxy）。投资者对热点的追逐首先就表现在对热点关注度的上升。这种关注度的变化与资产收益率之间的关系正是我们研究的焦点所在。



1.3 舆情数据可预测性分析

舆情数据的背后：

1. 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，同时也反过来影响了资产价格。
2. 当市场出现某一热点时，投资者关注度上升，带来资产价格上涨。上升的价格引来更多的关注度，形成一种**正反馈机制**。
3. 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，之前的投资者开始撤出，带来资产价格下降。
4. 所以通过舆情因子变化趋势的分析，我们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。



1.4 大类舆情资产数据

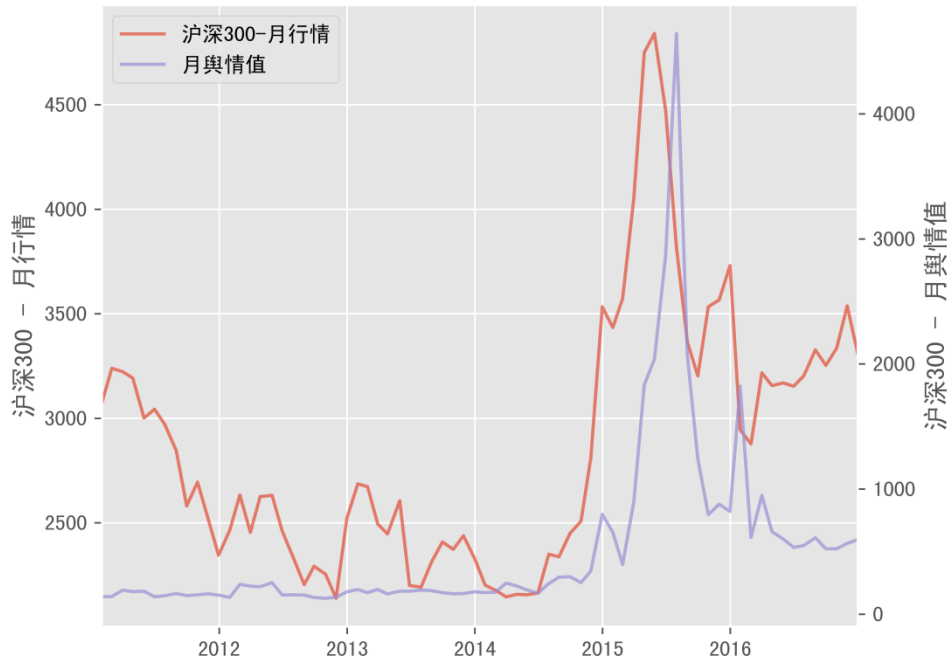
我们抓取了 2011-01-01 ~ 2017-02-28 时间段 7 种资产（沪深300、中证500、中证国债、贵金属、农产品、标普500、货币基金）的搜索指数舆情数据，示例如下：

Date	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	1444	1457	853	218	32
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

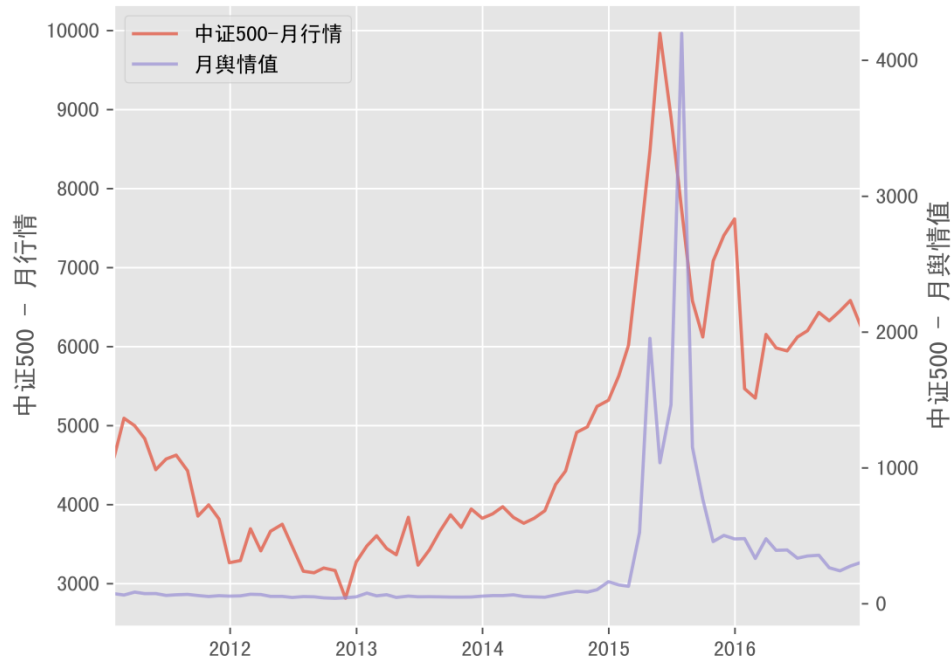
1.4 大类舆情资产数据

在 2011 ~ 2016 时间段，把资产在该月的舆情指数总和作为月舆情值，简单观察不同资产的月行情与月舆情值间关系。可以发现权益类、债券类、大宗商品、类现金等不同资产与对应的搜索舆情指数数据的关系不径相同。

沪深300-月行情与月舆情值



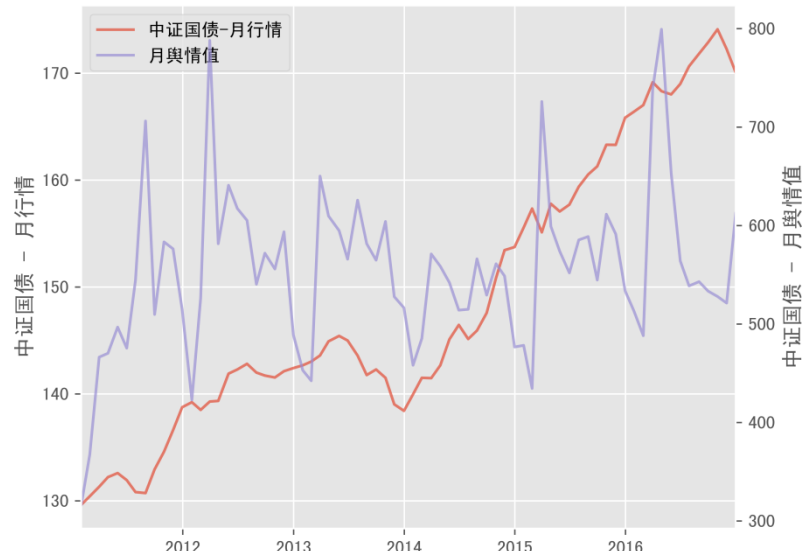
中证500-月行情与月舆情值



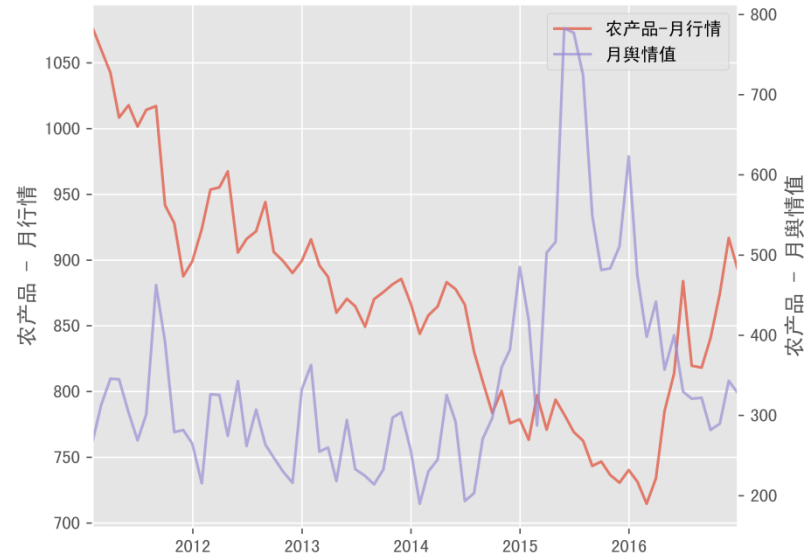
数据来源：广发证券发展研究中心

1.4 大类舆情资产数据

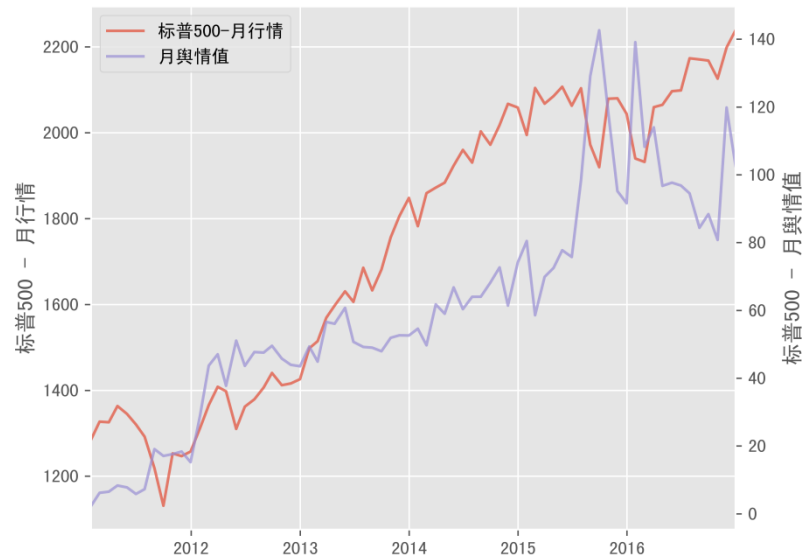
中证国债-月行情与月舆情值



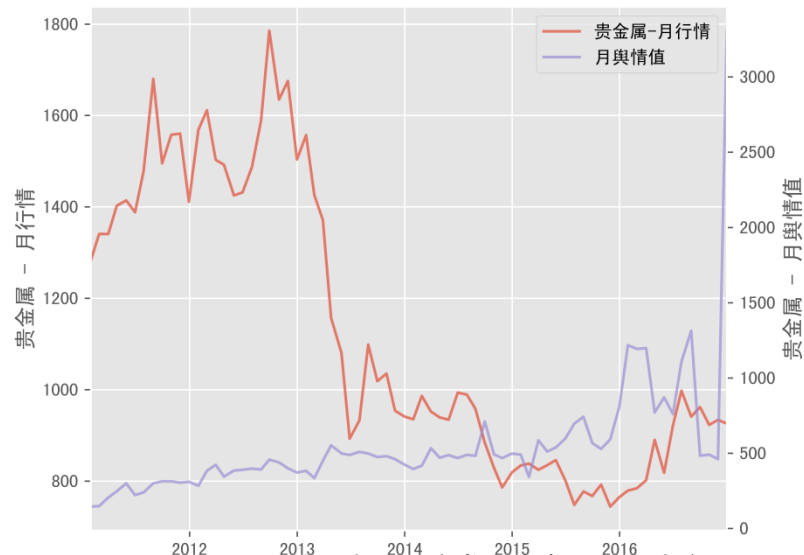
农产品-月行情与月舆情值



标普500-月行情与月舆情值

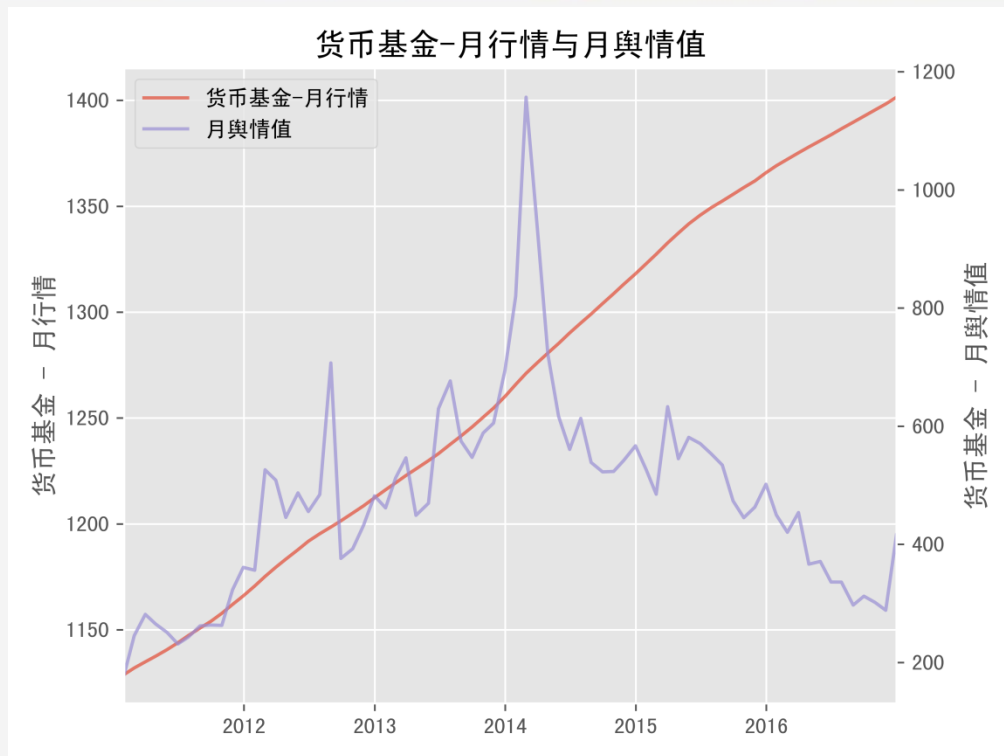


贵金属-月行情与月舆情值



数据来源：广发证券发展研究中心

1.4 大类舆情资产数据



资产	月行情-月舆情值相关系数
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：广发证券发展研究中心

从上面的结果看出，搜索舆情数据与各类资产行情数据的相关性不同。搜索舆情数据与权益类资产如沪深300、中证500、标普500成正相关、债券类、货币类的相关性较弱、商品类相关性为负相关。

因此在之后应用搜索舆情数据进行资产配置时，针对不同的资产我们采取不同的分析策略。



02

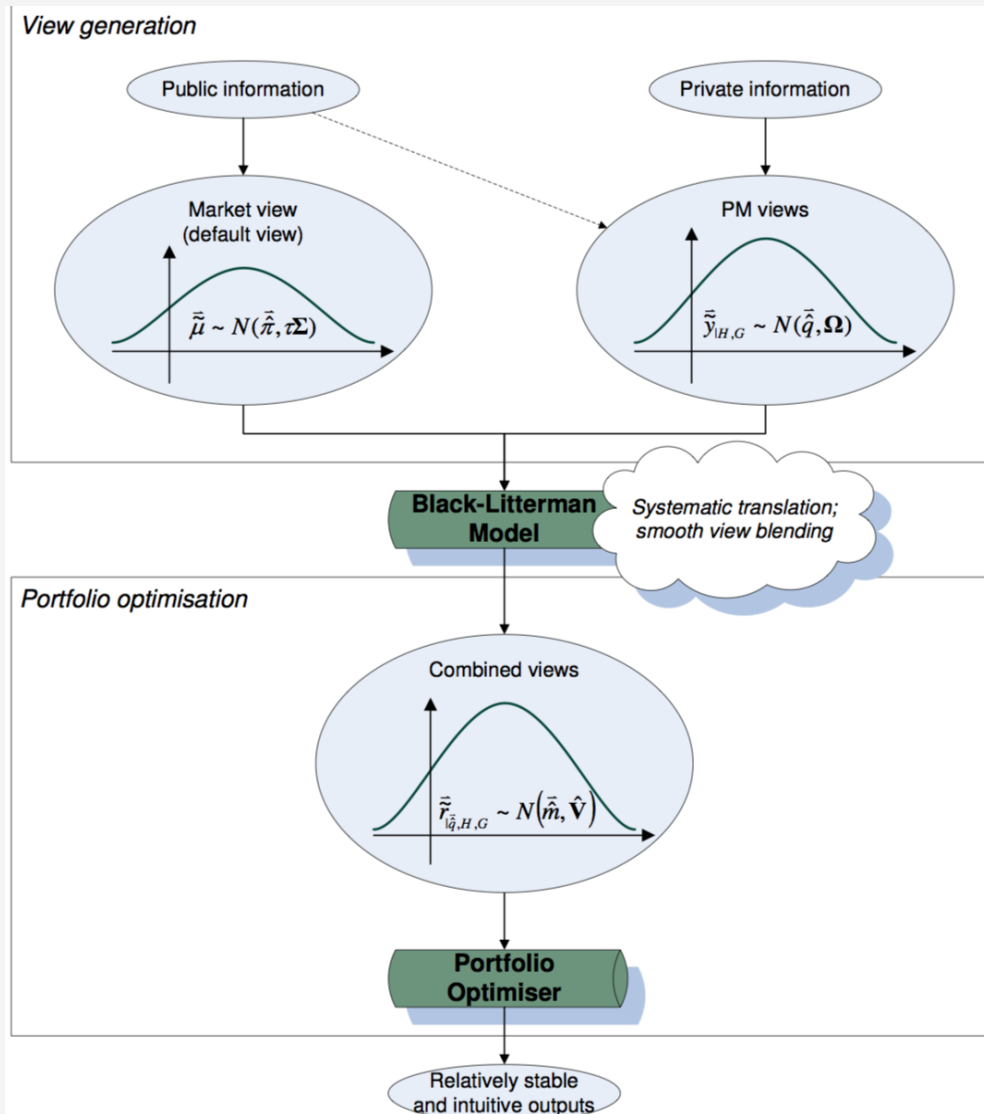
|再谈资产大类配置模型|



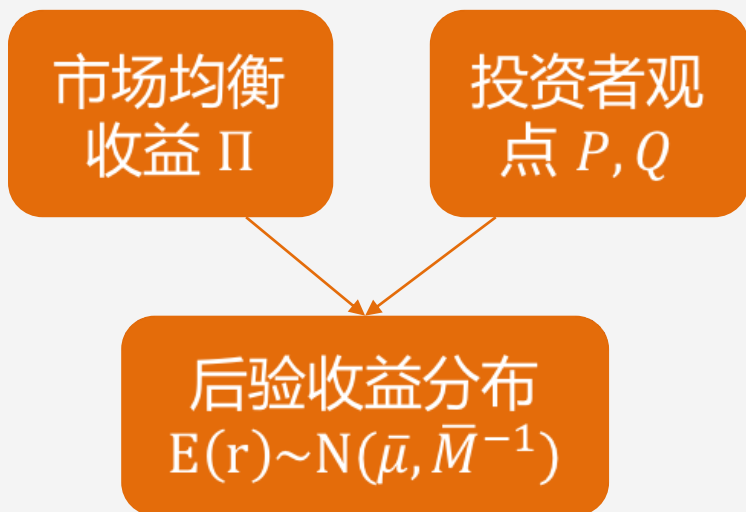
2.1 Black-Litterman 模型

- 由高盛的 Fisher Black 和 Robert Litterman 在 1992年提出的资产配置模型。
- 以贝叶斯方法，将投资者主观观点与市场先验均衡收益相结合，得到对市场的后验期望收益。
- 再根据后验期望收益，在约束条件下得到各类资产的配置比例。

数据来源：广发证券研究中心、Cheung W. The Black-Litterman model explained[J]. Journal of Asset Management, 2010, 11(4):229-243.

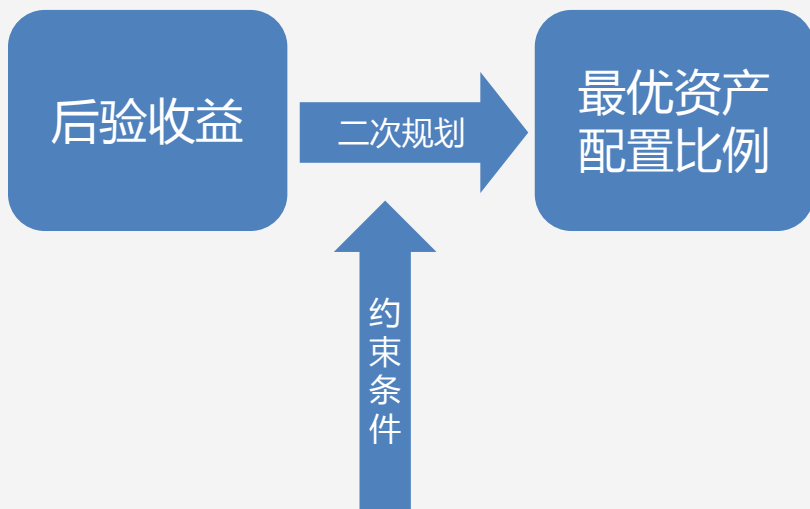


2.1 Black-Litterman 模型



风险厌恶系数 τ , 观点不确定度 Ω , 资产价格协方差 Σ , 投资者观点 P , 投资者预期收益 Q :

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

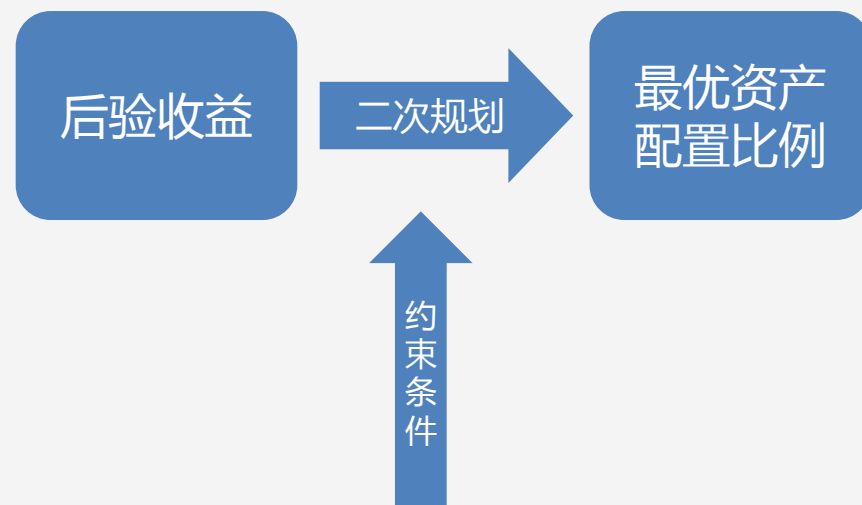


最优化时, 可根据实际情况, 加入相应的约束条件: 权重和为1、单品种持仓上限、持仓下限等.....

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

2.1 Black-Litterman 模型

- 1) 从资产配置问题的效用函数 $U = w^T E(r) - \frac{\lambda}{2} w^T \Sigma w$ 可以容易看出，优化问题的关键在于资产的预期收益率 $E(r)$ 和资产超额收益率间的协方差 Σ 。
- 2) 协方差 Σ 一般都是从资产的历史价格得到，或者从别的方法得到一个预期的协方差，我们在这里简单采用历史协方差。
- 3) 预期收益率 $E(r)$ 采用 Black-Litterman 模型中得到的后验收益，所以**预期收益率 $E(r)$** 预测的结果是整个资产配置问题的关键。



$$\left\{ \begin{array}{l} \text{Max}(w^T E(r) - \frac{\lambda}{2} w^T \Sigma w) \\ \text{s. t. : } \Sigma w_i = 1, \\ w_i \leq \text{max}w, \\ w_i \geq \text{min}w \end{array} \right.$$

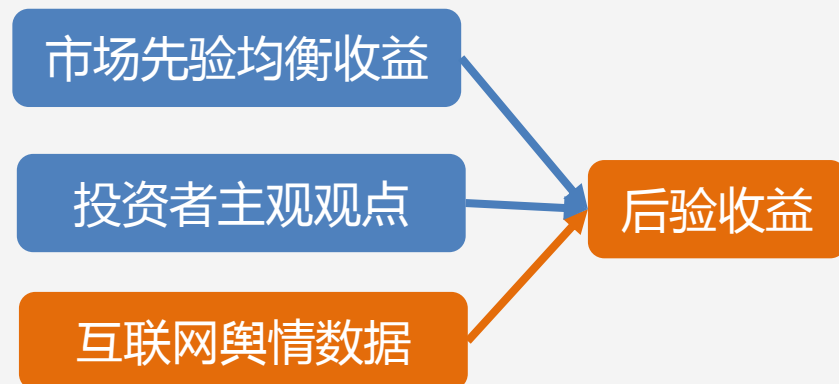
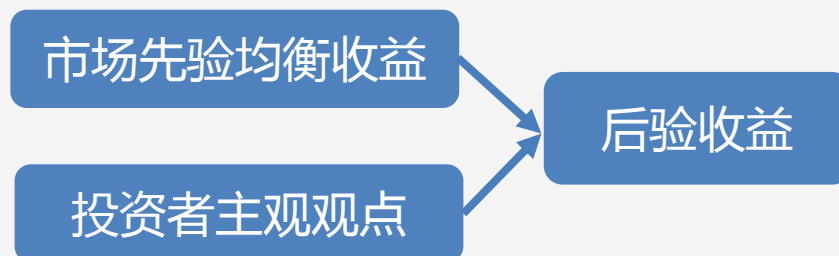
2.2 传统Black-Litterman 模型的局限

【局限】

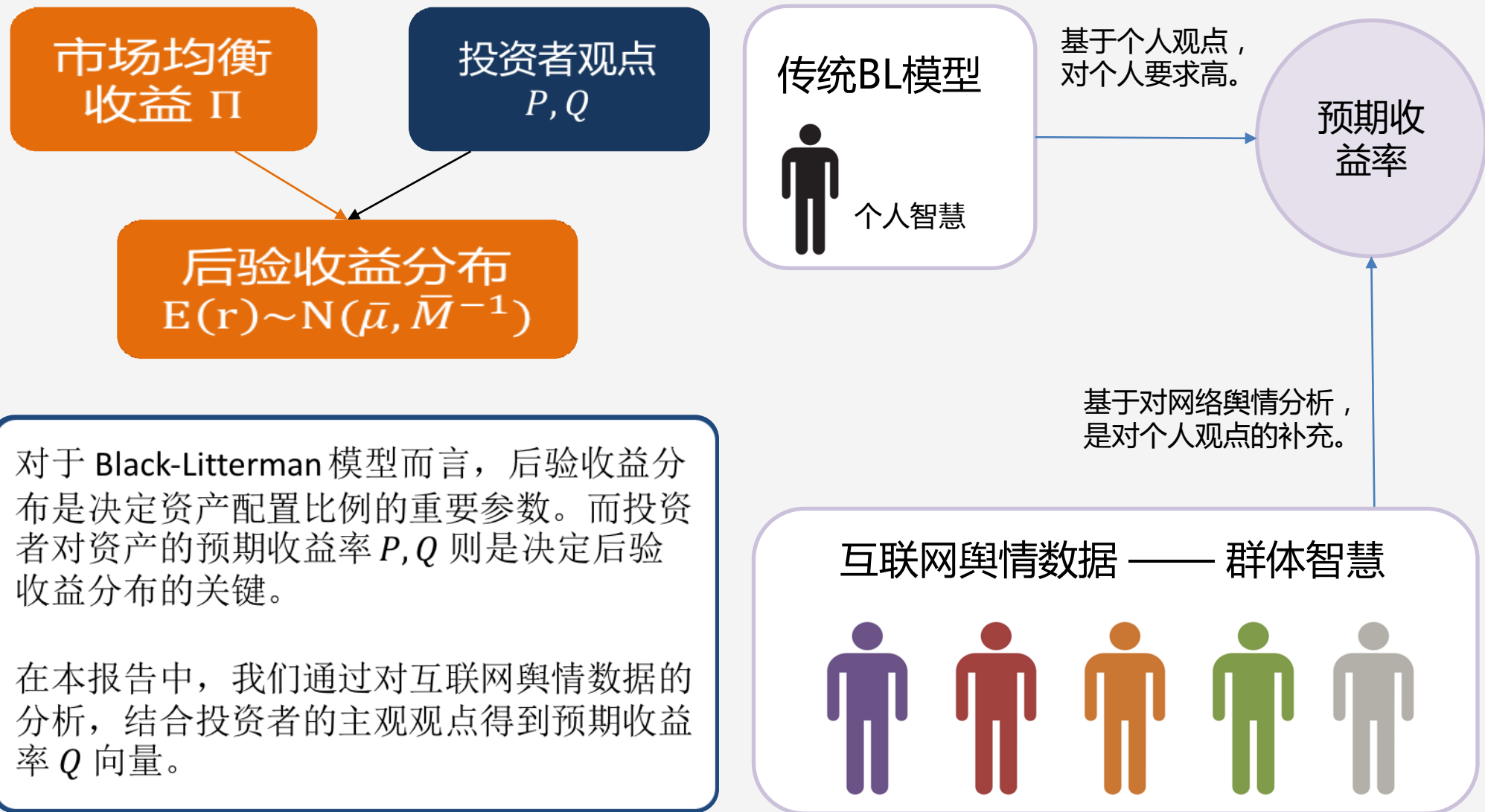
- 传统的 Black-Litterman 模型中需要投资者输入自己对不同资产的主观观点。
- 但实践中投资者很难对不同资产都有比较准确的观点。有时提出的错误观点会使得模型给出更不准确的结果。

【改进】

- 我们试图在传统 Black-Litterman 模型基础上，与互联网舆情数据相结合。通过分析舆情数据，以追求得到更准确的预期收益率，从而实现更好的资产配置效果。



2.3 对 Black-Litterman 模型的改进





03

|舆情因子测试|



3.1 舆情因子

为了更好的预测资产收益率，我们对之前抓取得到的舆情指数数据进行初步处理，分成 3 大类：

1. 基于舆情数据的舆情因子
2. 结合了舆情数据的动量因子
3. 作为补充的传统因子



3.2 舆情因子的构建

因子分类	因子指标	因子描述
舆情因子	近1月舆情值	<ul style="list-style-type: none"> 相应时间周期内，对每日舆情值的加总。 描述了该时间段内舆情累积值，代表对资产关注度的累积量。 分 1、2、3 月三个不同的时间周期看时间长度的影响效果。
	近2月舆情值	
	近3月舆情值	
	近1月舆情值波动率	<ul style="list-style-type: none"> 相应时间周期内，对每日舆情值的波动率。 描述了该时间段内日舆情值的波动情况，代表了对资产关注度的变化程度。 分 1、2、3 月三个不同的时间周期看时间长度的影响效果。
	近2月舆情值波动率	
	近3月舆情值波动率	
	近1月舆情值增长率	<ul style="list-style-type: none"> 相应时间周期内，舆情值总量的环比变化率。 描述了该时间段内舆情值总量的变化情况，代表了对资产关注度的变化程度。 分 1、2、3 月三个不同的时间周期看时间长度的影响效果。
	近2月舆情值增长率	
	近3月舆情值增长率	
	近1月舆情值增长率变化率	<ul style="list-style-type: none"> 相应时间周期内，舆情值增长率的变化率。 描述了该时间段内舆情值变化率的变化情况，代表了对资产关注度的二阶变化程度。 分 1、2、3 月三个不同的时间周期看时间长度的影响效果。
	近2月舆情值增长率变化率	
	近2月舆情值增长率变化率	

3.2 舆情因子的构建

因子分类	因子指标	因子描述
结合舆情因子的动量因子	近1月舆情值加权日均收益率	<ul style="list-style-type: none"> 相应时间周期内，以每日舆情值加权得到的日收益率的均值。 结合舆情数据后对的资产收益率。 分 1、2、3 月三个不同的时间周期看时间长度的影响效果。
	近2月舆情值加权日均收益率	
	近3月舆情值加权日均收益率	
	近1月舆情值加权日收益率波动率	<ul style="list-style-type: none"> 相应时间周期内，以每日舆情值加权得到的日收益率的标准差。 结合舆情数据后对的资产收益率波动率。 分 1、2、3 月三个不同的时间周期看时间长度的影响效果。
	近2月舆情值加权日收益率波动率	
	近3月舆情值加权日收益率波动率	
作为补充的传统因子	对数市值	<ul style="list-style-type: none"> 资产市值的对数
	近1月收益率	<ul style="list-style-type: none"> 相应时间周期内，以资产的收益率数据。
	近2月收益率	
	近3月收益率	
	近1月日收益率波动率	<ul style="list-style-type: none"> 相应时间周期内，以资产的收益率波动率数据。
	近2月日收益率波动率	
	近3月日收益率波动率	

3.3 数据标准化

由于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：先去极值，后标准化。

1. 去极值：

σ 是序列 x 的标准差，以 2.5 倍标准差为阈值，对数据进行处理，避免极端值对结果的影响。

$$x_{new} = \begin{cases} x + 2.5\sigma, & \text{如果 } x > x + 2.5\sigma \\ x - 2.5\sigma, & \text{如果 } x < x - 2.5\sigma \\ x, & \text{其他} \end{cases}$$

2. 标准化：

其中 \bar{x} 是序列 x 的均值， $\sigma(x)$ 是序列 x 的标准差。将序列变成均值为 0，方差为 1 的新序列。

$$x_{new} = \frac{x - \bar{x}}{\sigma(x)}$$

3. 舆情因子测试

因为资产池中一共只有 7 种大类资产，且大类资产间差异性较大。所以我们在检测因子的时候，对每种资产单独进行分析。

对资产收益率时间序列数据进行回归，检验因子的有效性，挑选因子预测未来资产的收益率情况。

每种资产单独分析

通过回归方式检测因子

挑选因子预测收益率

3.4 舆情因子测试

舆情因子测试流程：

因子值向量 $F = (f_1, f_2, \dots, f_N)$, 资产下月超额收益 $R = (R_1, R_2, \dots, R_N)$, 定义 IC 值为 F 和 R 的相关系数。取窗口期为 12 个月, 所以 F 和 R 都是 12 维的向量。

f_i 是资产第 i 月末的因子值

R_i 是资产第 $i+1$ 月的超额收益率

$$\text{信息系数 } IC = \rho(F, R) = \frac{\text{cov}(F, R)}{\sigma(F)\sigma(R)}$$

定义信息比率 IR 为因子 IC 的均值和因子 IC 的标准差之比, IR 绝对值越高代表因子预测能力越稳定。

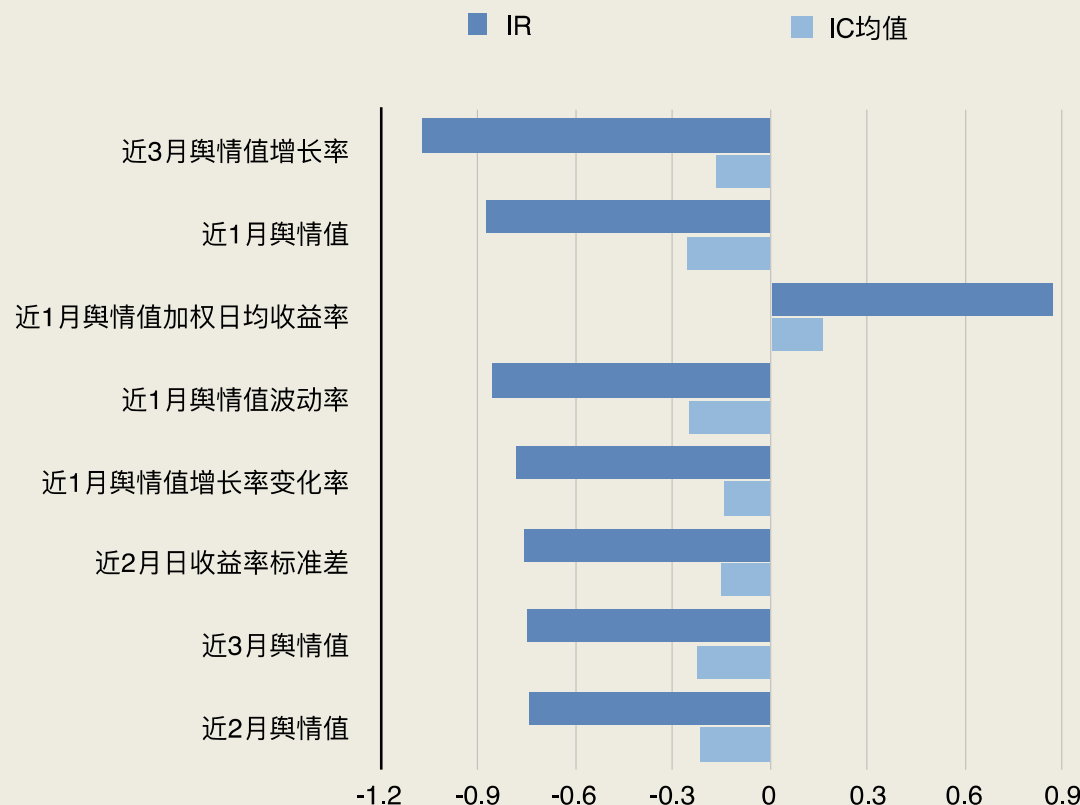
$$\text{信息比率 } IR = \frac{\overline{IC}}{\sigma(IC)}$$

选取 IR 绝对值最大的 8 个因子作为有效因子, 进行多元线性回归预测下月资产收益率。

3.4 舆情因子测试: 沪深300

因子	IR	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

沪深300因子分析

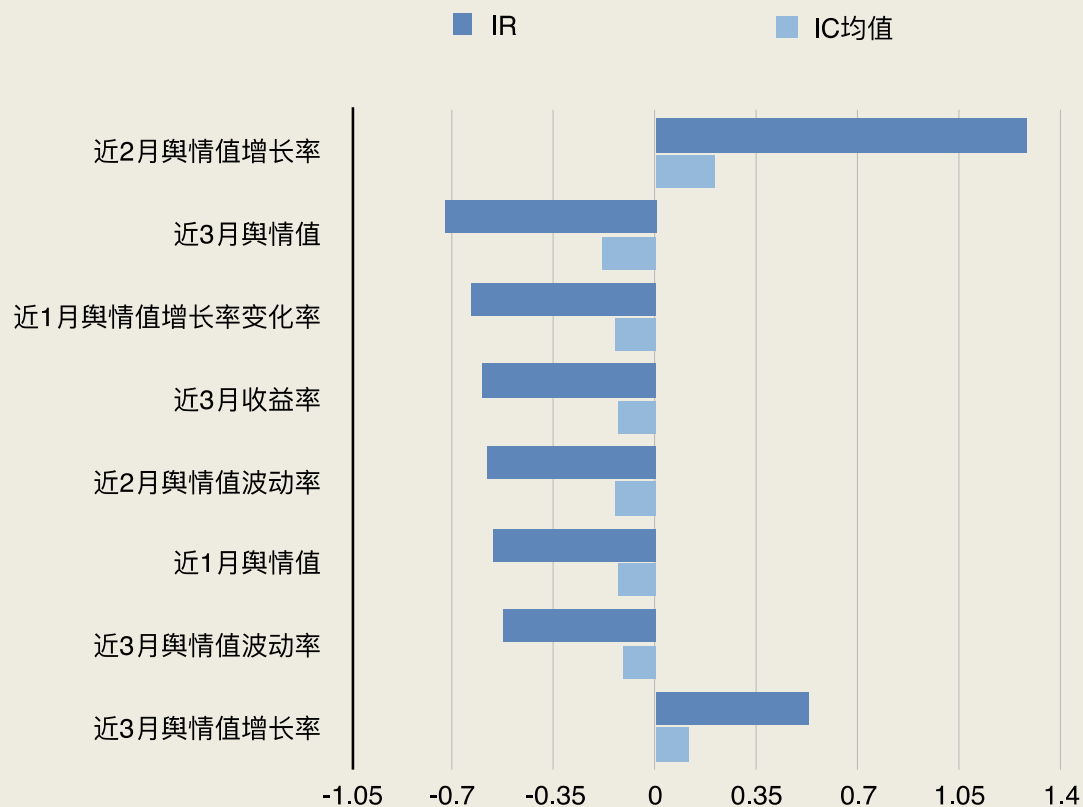


数据来源：广发证券发展研究中心

3.4 舆情因子测试: 中证500

因子	IR	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

中证500因子分析

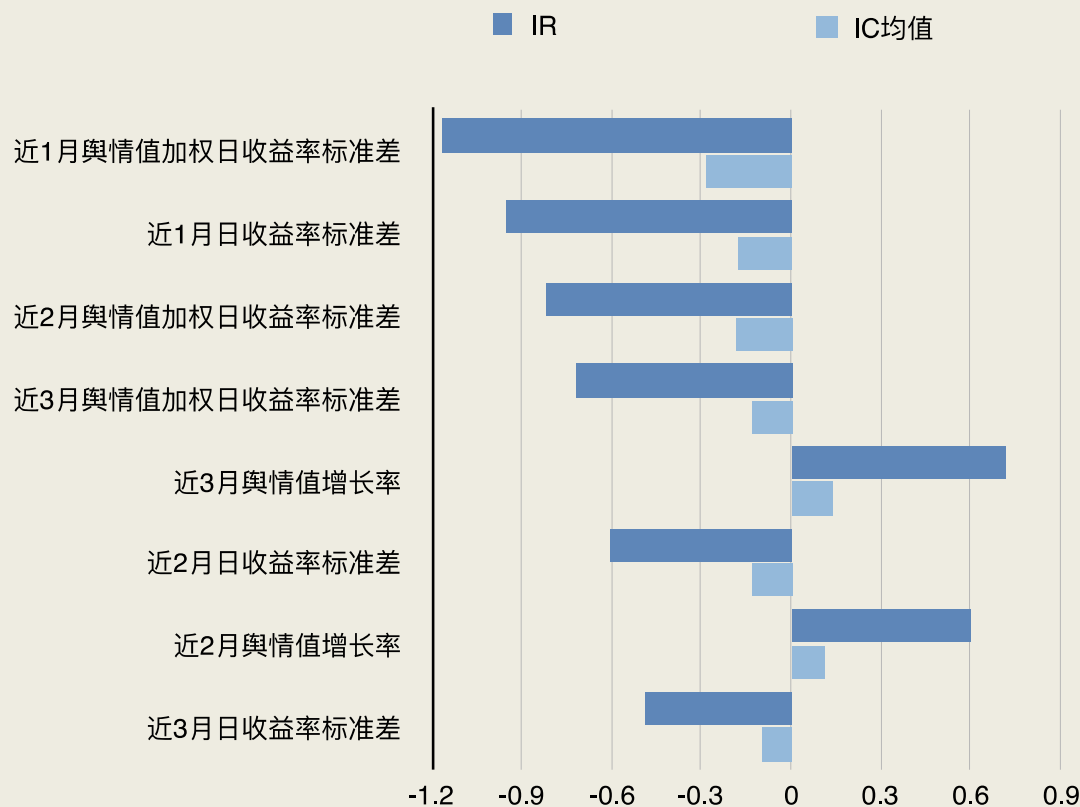


数据来源：广发证券发展研究中心

3.4 舆情因子测试: 中证国债

因子	IR	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

中证国债因子分析

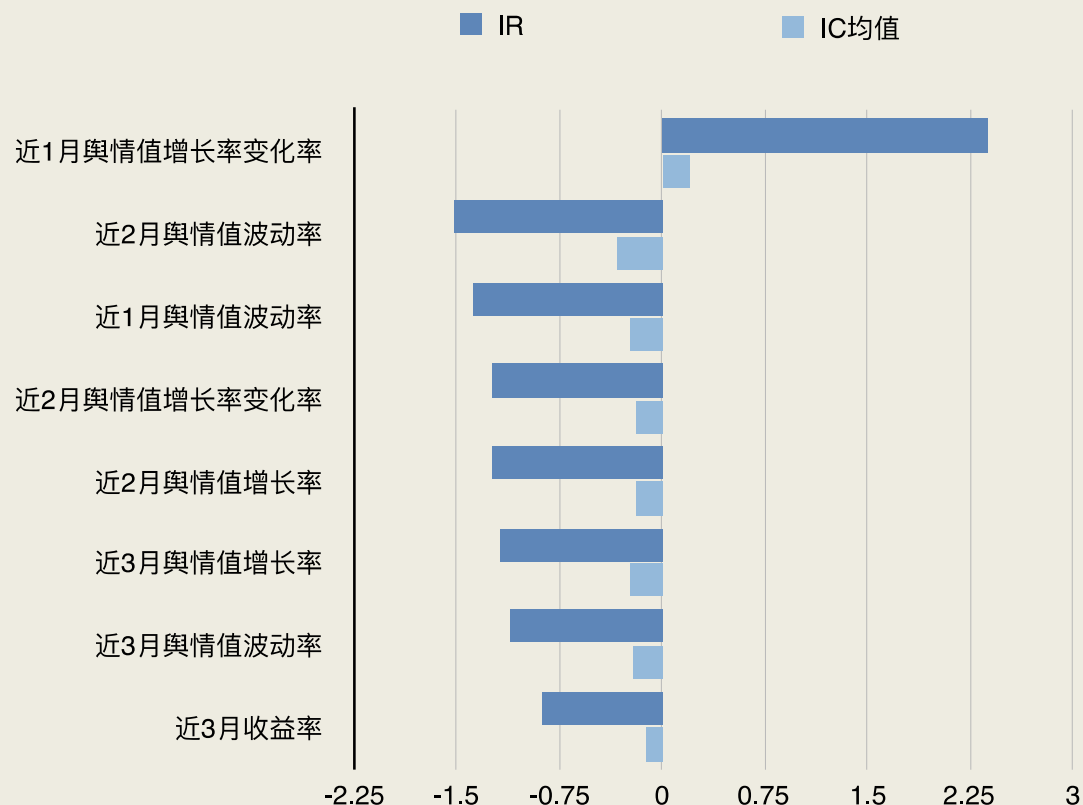


数据来源：广发证券发展研究中心

3.4 舆情因子测试: 贵金属

因子	IR	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

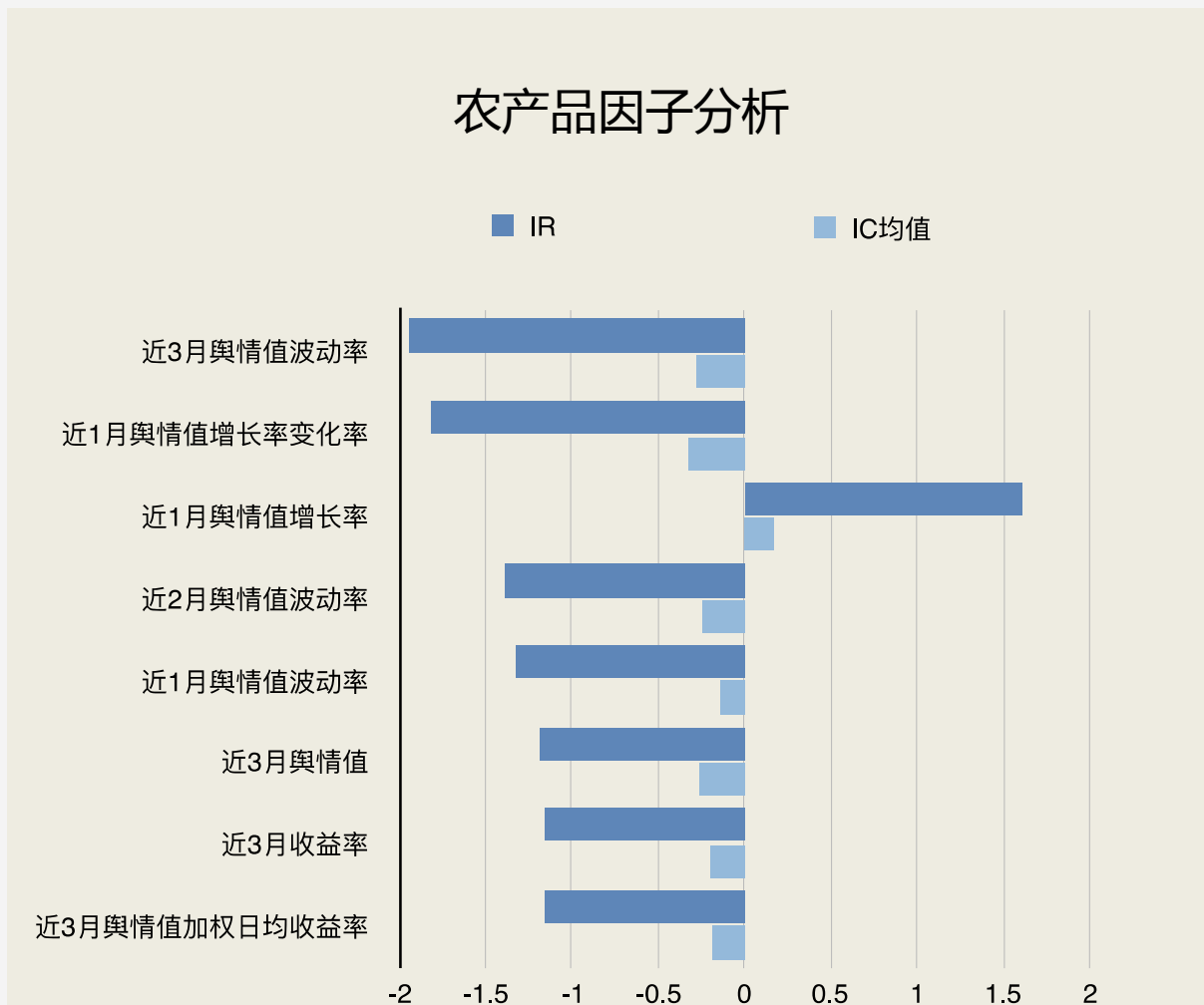
贵金属因子分析



数据来源：广发证券发展研究中心

3.4 舆情因子测试: 农产品

因子	IR	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

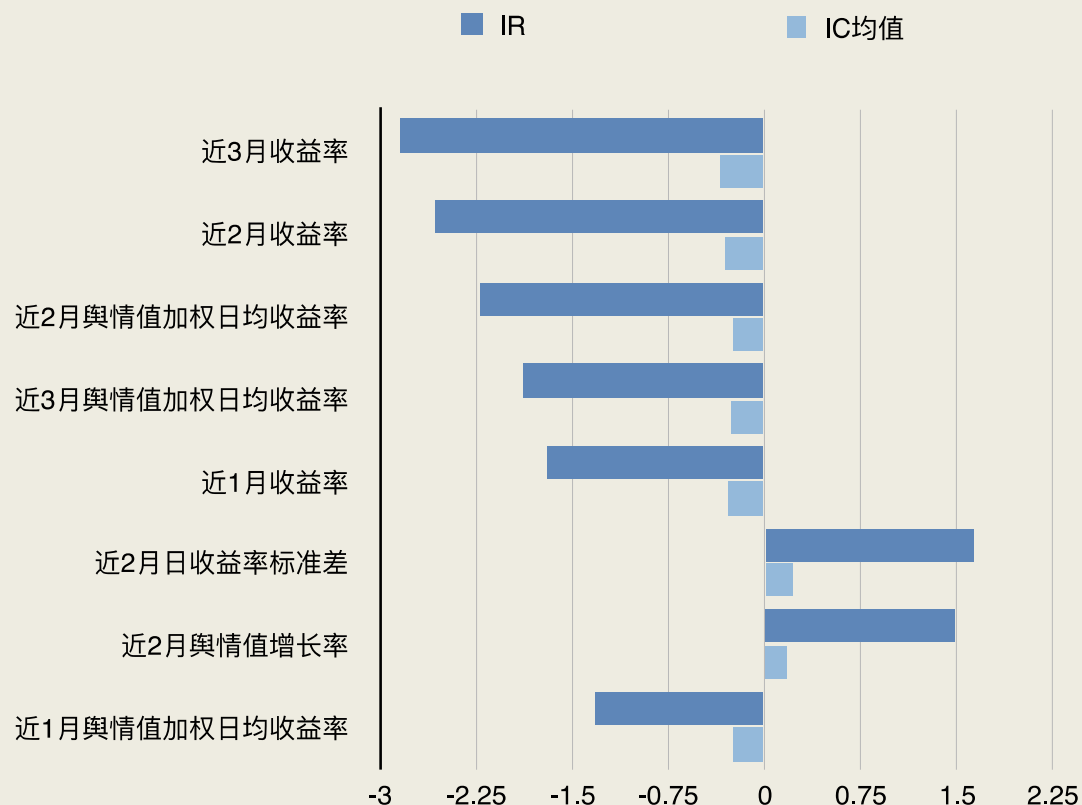


数据来源：广发证券发展研究中心

3.4 舆情因子测试: 标普500

因子	IR	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

标普500因子分析

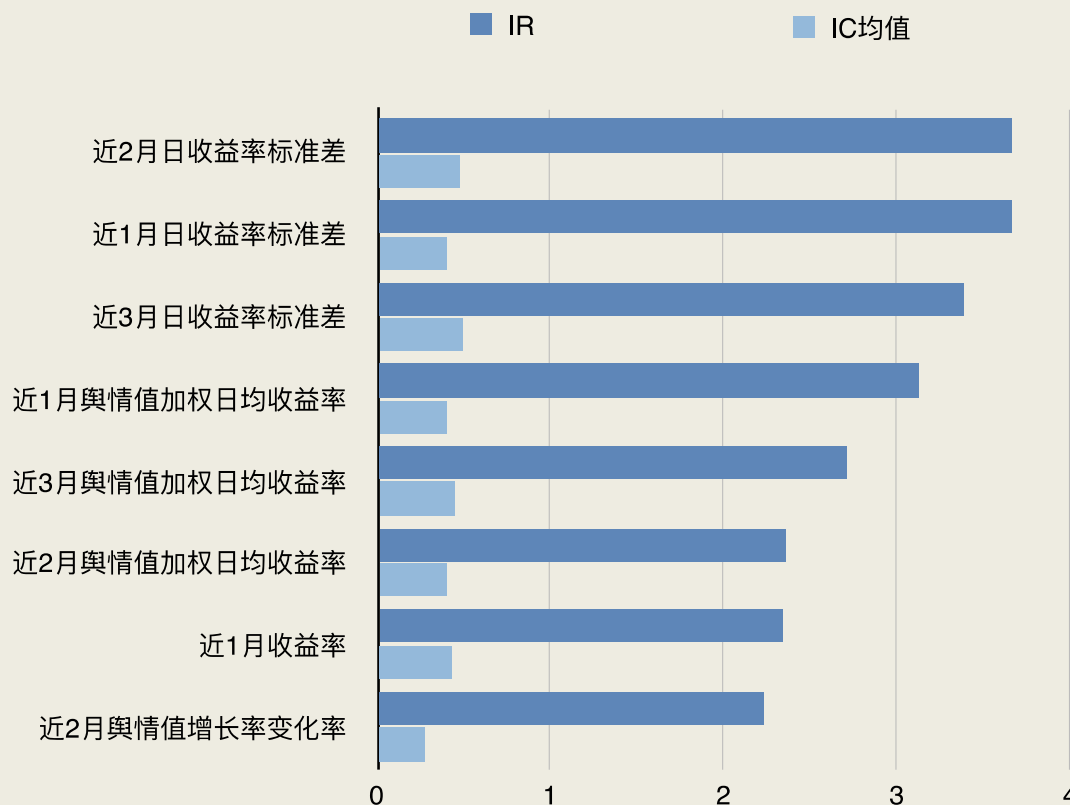


数据来源：广发证券发展研究中心

3.4 舆情因子测试: 货币基金

因子	IR	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

货币基金因子分析



数据来源：广发证券发展研究中心



04

|实证分析|

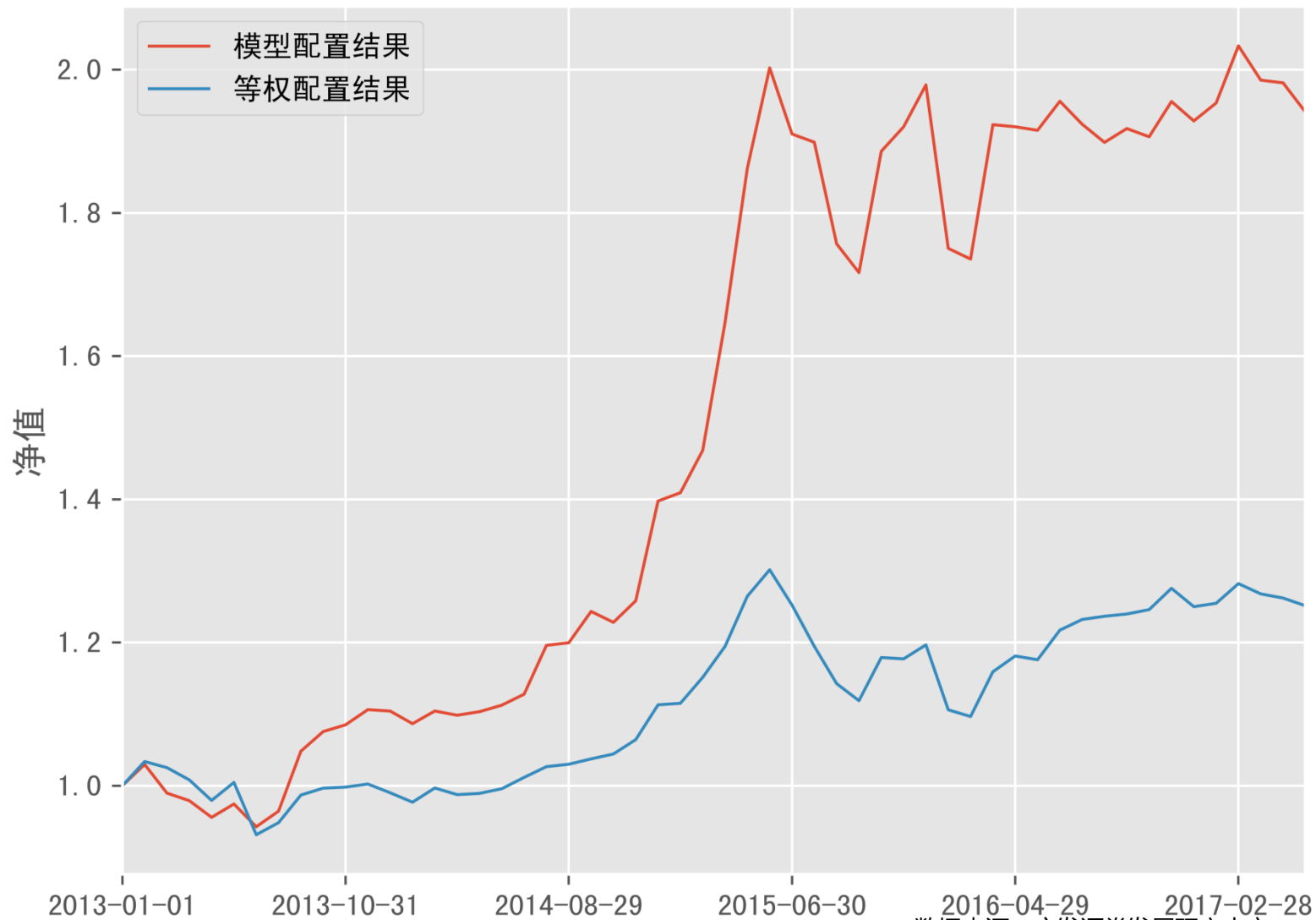


4.2 相关数据

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 ~ 2017-02-28
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
单品种资产持仓限制	最低持仓比例 5%，最高持仓比例40%

4.3 回测结果

基于舆情的BL模型配置结果



4.3 回测结果

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

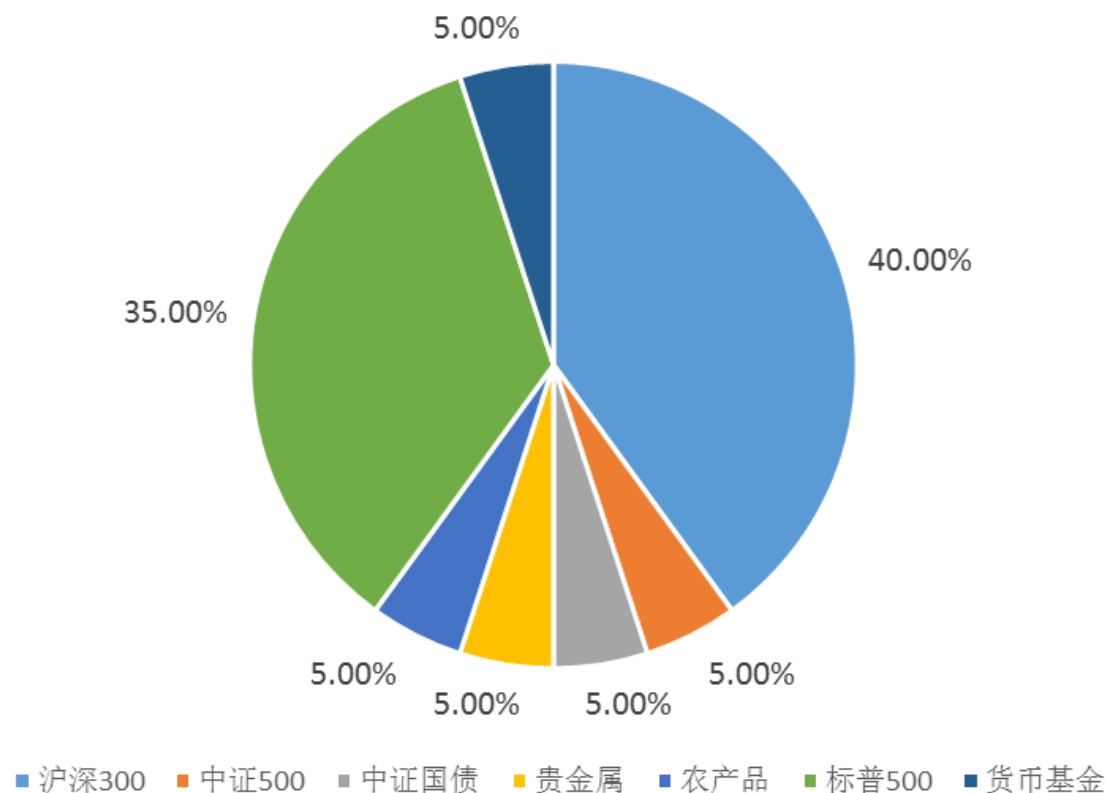
4.4 配置建议

依照上述方法，我们根据资产配置模型进行计算和分析。

给出2017年6月的各类资产配置比例建议：

- a) 沪深300持仓比例：40%
- b) 中证500持仓比例：5%
- c) 中证国债持仓比例：5%
- d) 贵金属持仓比例：5%
- e) 农产品持仓比例：5%
- f) 标普500持仓比例：35%
- g) 货币基金持仓比例：5%

资产配置推荐



数据来源：广发证券发展研究中心



05

|结论|



- 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 Black-Litterman 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。
- 本报告是对于舆情数据在配置资产领域应用中的一个尝试，未来我们可以扩充覆盖更广的舆情数据，通过文本分析、深度学习等方法得到自定义的舆情数据，再进行更加广泛、深入的测试。

本文旨在对所研究问题的主要关注点进行分析，因此对市场及相关交易做了一些合理假设，但这样会导致建立的模型以及基于模型所得出的结论并不能完全准确地刻画现实环境。而且由于分析时采用的相关数据都是过去的时间序列，因此可能会与未来真实的情况出现偏差。本文内容并不是适合所有的投资者，客户在制定投资策略时，必须结合自身的环境和投资理念。

广发证券股份有限公司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。

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。

Thanks !
谢谢