

基于网络舆情再探指数轮动策略研究

——互联网大数据挖掘系列研究之(十三)

陈原文 S0260517080003
邮箱：chenyuanwen@gf.com.cn
0755-82797057
广发证券金融工程
2018年06月





01

01

02

03

04

|研究背景|

>

海外大数据相关研究

由于信息的爆炸式增长，对信息获取有效性、针对性的需求孕育而生。亚马逊作为全球最大的电子商务网站，拥有20亿账户的账户数据。通过对用户行为数据的收集，构造用户画像，对用户进行精准的商品推荐，帮助百万级用户发现自己无法找到的东西。



数据显示，亚马逊30%的网页浏览量来自于该智能推荐系统，用户购买率、转换率会达到60%左右，从中亚马逊获取了10%到30%的附加利润。

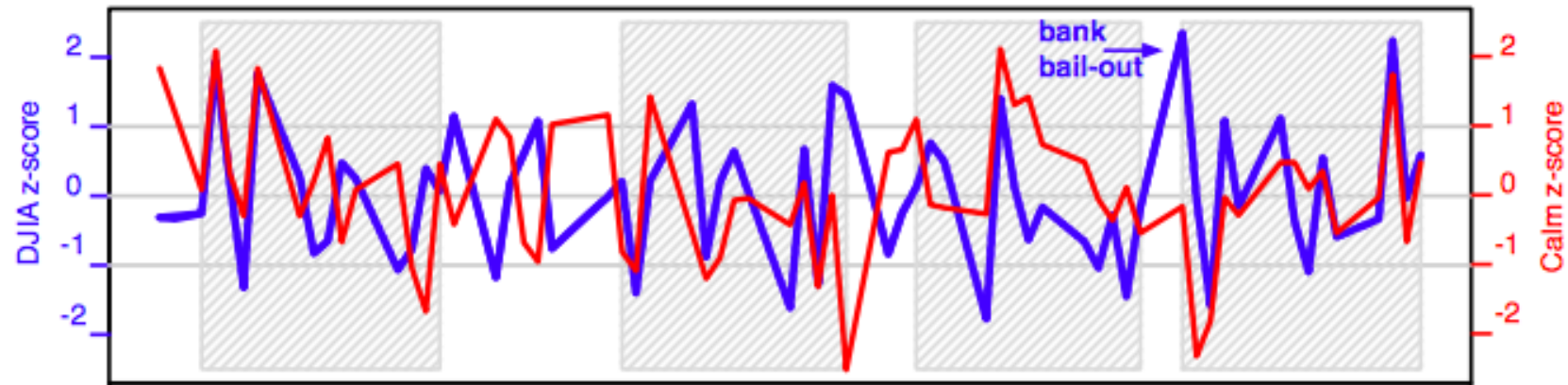
海外大数据舆情相关研究

2010年，美国印第安纳大学的计算机科学家Johan Bollen发现，从海量Twitter信息中提取的情绪指标与道琼斯工业指数存在显著正相关关系，Twitter上用户的情感可以显著的预测道琼斯工业指数在未来3到4天的涨跌情况，预测准确性达87.6%。

STATISTICAL SIGNIFICANCE (P-VALUES) OF BIVARIATE GRANGER-CAUSALITY CORRELATION BETWEEN MOODS AND DJIA IN PERIOD FEBRUARY 28 2008 TO NOVEMBER 3, 2008.

Lag	OF	Calm	Alert	Sure	Vital	Kind	Happy
1 day	0.085*	0.272	0.952	0.648	0.120	0.848	0.388
2 days	0.268	0.013**	0.973	0.811	0.369	0.991	0.7061
3 days	0.436	0.022**	0.981	0.349	0.418	0.991	0.723
4 days	0.218	0.030**	0.998	0.415	0.475	0.989	0.750
5 days	0.300	0.036**	0.989	0.544	0.553	0.996	0.173
6 days	0.446	0.065*	0.996	0.691	0.682	0.994	0.081*
7 days	0.620	0.157	0.999	0.381	0.713	0.999	0.150

(p-value < 0.05: **, p-value < 0.1: *)



数据来源：广发证券发展研究中心

海外大数据舆情相关研究

2015年，欧洲央行发布研究报告，通过分析投资者的情绪变化而找到预测股票市场的办法。通过搜集Twitter和Google中“牛市”和“熊市”两个词出现的次数作为反映投资者情绪的指标，并与股票市场的涨跌情况进行对比，分析二者之间的关系。

Predicting the daily returns of selected stock market indices using Twitter bullishness

Bullishness	DJIA		SP500		Russell1000		Russell2000	
Lag	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value
1	12.56	0.01***	10.98	0.05**	10.72	0.05**	11.02	0.05**
2	2.27	0.67	2.61	0.65	2.46	0.67	2.66	0.65
3	2.18	0.69	3.69	0.53	4.037	0.48	4.58	0.43
4	-7.81	0.15	-8.10	0.16	-9.99	0.08*	-10.28	0.08*
5	-1.12	0.80	-1.28	0.79	-1.35	0.77	-1.37	0.78

数据来源：广发证券发展研究中心

实证结果显示，前一天的Twitter指数对美国几大股票指数——道琼斯(DJIA)、标普500(SP500)、Russell 1000、Russell 2000第二天的走势有显著的正影响。Twitter每日情绪指数(DSI)每上升一个点，可以导致DJIA在次日上升12.56个基点。由此可见，互联网舆情会对股票市场的资产定价产生影响。

海外大数据舆情相关研究

通过对Thomson Reuters网站2003年到2013年超过两千万条新闻数据的收集，提取标题和正文的关键词，通过文本挖掘的方法，构建情绪指标，构建动量策略，做多正面情绪的资产组合，做空负面情绪的资产组合。实证结果显示，该策略累计收益为224%，最大回撤为-10.9%，信息比为0.84，夏普比率为1.47。

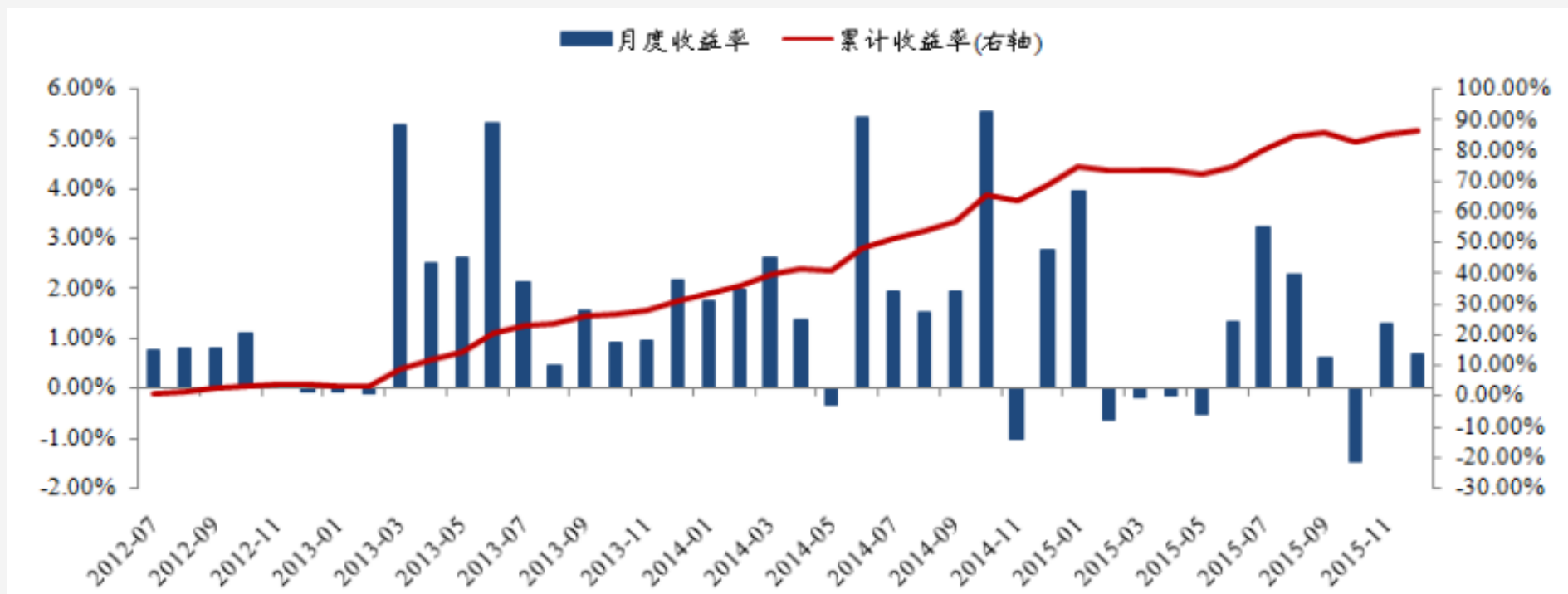
Years	MSCI World Equity	Bonds	Benchmark	Portfolio	Switches
2004	11.3%	2.1%	6.6%	8.5%	4
2005	15.8%	0.9%	8.1%	6.6%	11
2006	15.6%	3.5%	9.4%	4.5%	12
2007	4.7%	9.8%	7.2%	13.5%	10
2008	-38.7%	12.3%	-17.0%	10.3%	6
2009	25.7%	-0.7%	11.7%	53.8%	6
2010	10.0%	5.7%	7.8%	11.4%	6
2011	-5.5%	6.2%	0.2%	8.4%	5
2012	15.7%	1.6%	8.4%	4.5%	11
2013	28.9%	-0.8%	13.1%	10.7%	11

	MSCI World Equity	Bonds	Benchmark	Portfolio
Total Return	86%	48%	66%	224%
Mean Return p.a.	6.4%	4.0%	5.2%	12.4%
Volatility p.a.	16.3%	3.6%	7.7%	8.5%
Maximum Drawdown	-54.5%	-4.6%	-27.4%	-10.9%
Sharpe Ratio	0.39	1.09	0.68	1.47
Alpha p.a.				7.3%
Tracking Error p.a.				8.7%
Information Ratio				0.84
Switches p.a.				8.2

数据来源：广发证券发展研究中心

海外大数据舆情相关研究

2012年，CAYMAN ATLANTIC公司成立了一只基于网络社交媒体Twitter、搜索引擎Google及其他新闻媒体数据指导交易的对冲基金。自2012年7月至2015年12月，CAYMAN ATLANTIC的累计收益为86.15%，年化收益率达19.43%，在42个月中有32个月取得正收益，月度胜率达76.19%，最大回撤仅为1.49%，该大数据基金在可控的风险下创造了较高的收益。



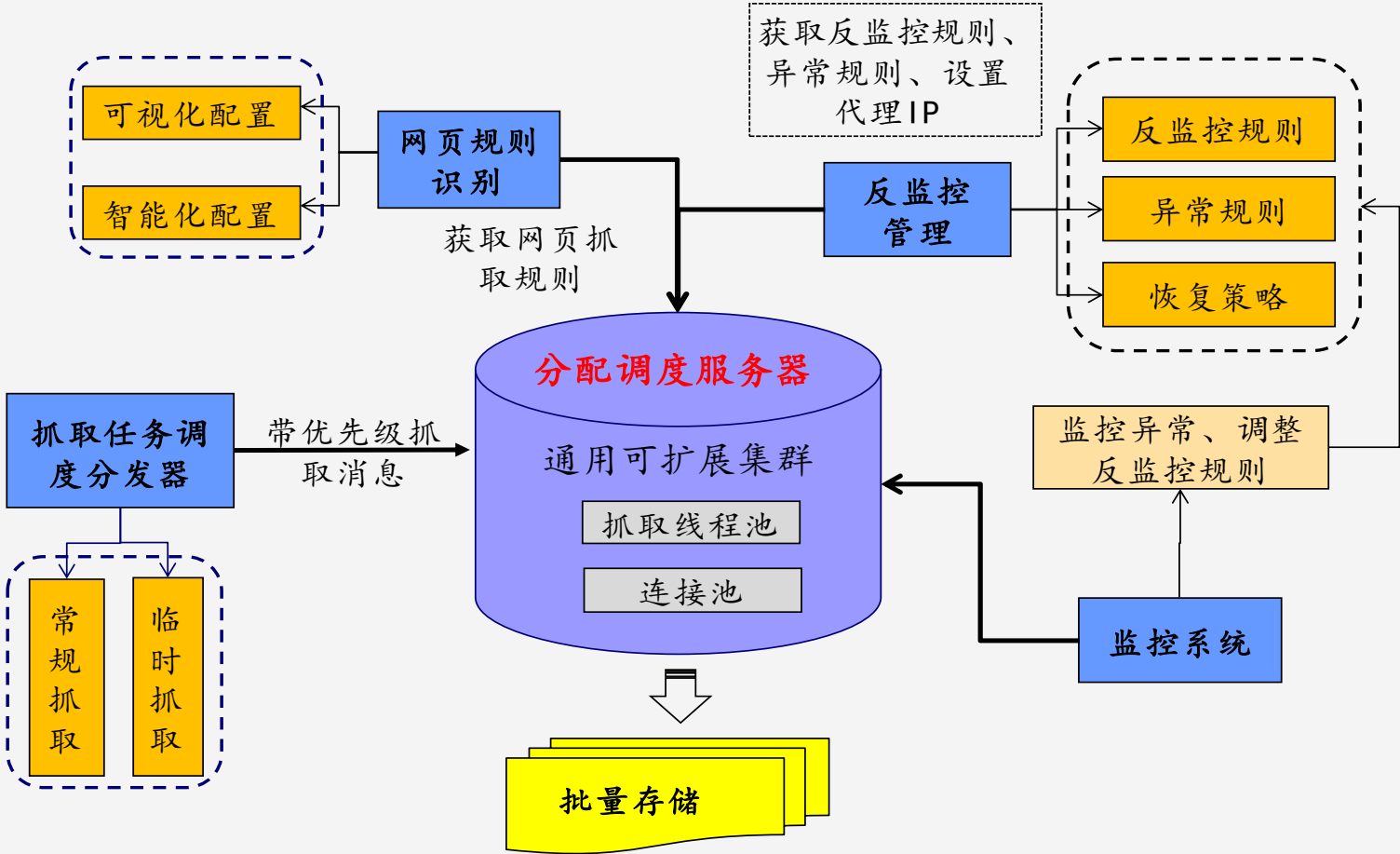
数据来源：广发证券发展研究中心

国内部分互联网大数据基金产品

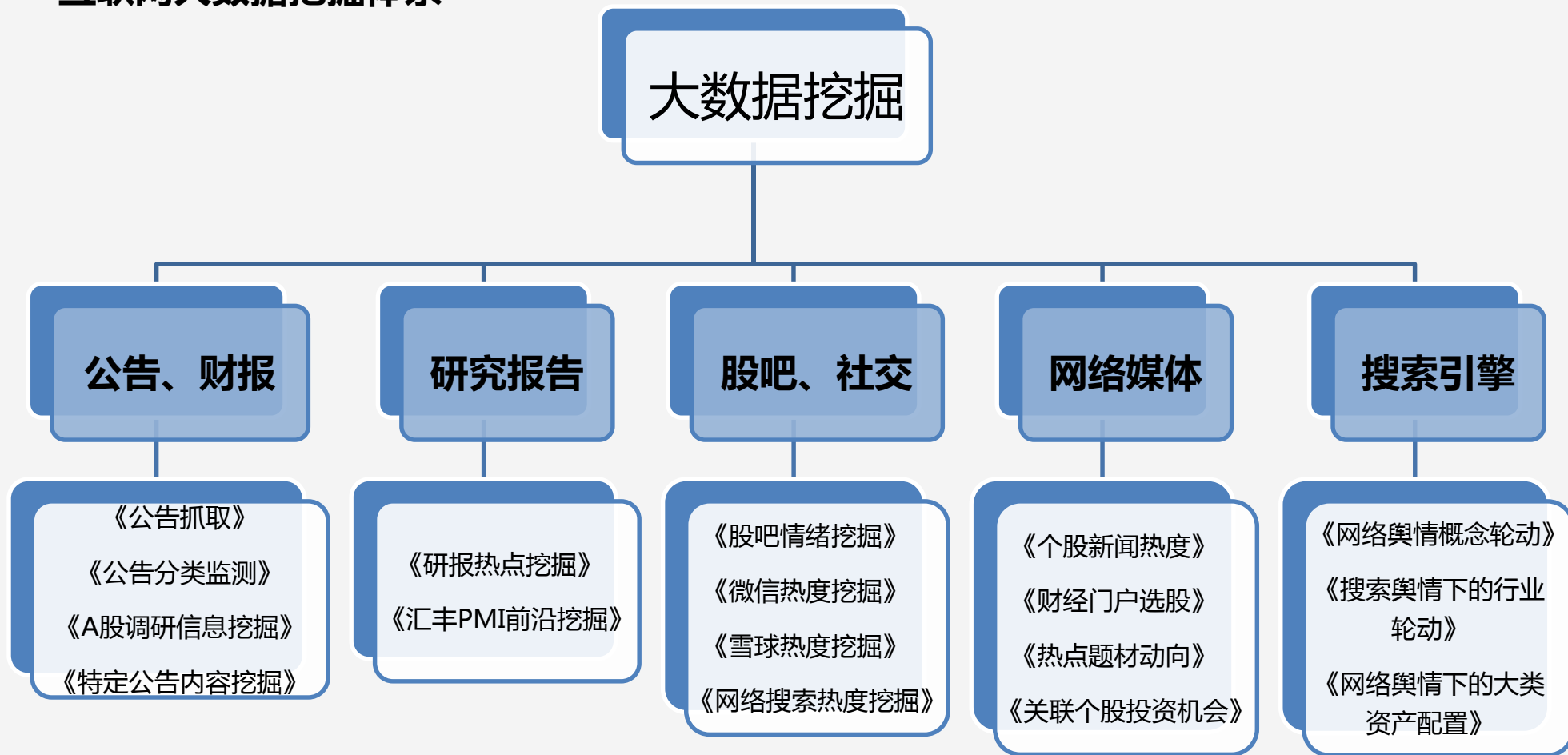
名称	代码	跟踪指数	上市时间	公司	开放/封闭	主动/被动	类型	成立以来表现
博时中证银联智惠大数据100指数型证券投资基金	002588.OF	中证银联智惠大数据100	2016/5/20	博时基金	开放式	被动	指数型基金	40.36%
广发中证百度百发策略100A指数型证券投资基金	000826.OF	中证百度百发策略100指数	2014/10/30	广发基金	开放式	被动	指数型基金	26.13%
广发中证百度百发策略100E指数型证券投资基金	000827.OF	中证百度百发策略100指数	2014/10/30	广发基金	开放式	被动	指数型基金	25.90%
泰达宏利同顺大数据量化优选灵活配置混合型证券投资基金	002263.OF		2016/2/23	泰达宏利基金	开放式	主动	灵活配置型基金	21.00%
南方大数据300A指数证券投资基金	001420.OF	大数据300	2015/6/24	南方基金	开放式	被动	指数型基金	8.94%
东兴证券众智优选基金	002465.OF		2016/6/13	东兴证券股份有限公司	开放式	主动	灵活配置型基金	5.30%
大成中证360互联网+大数据100指数型证券投资基金	002236.OF	中证360互联网+大数据100指数	2016/2/3	大成基金	开放式	被动	指数型基金	4.30%
广发百发大数据策略精选灵活配置混合型证券投资基金	001741.OF		2015/9/14	广发基金	开放式	主动	灵活配置型基金	0.90%
广发百发大数据策略精选灵活配置混合型证券投资基金	001742.OF		2015/9/14	广发基金	开放式	主动	灵活配置型基金	0.90%

数据来源:公司官网、广发证券发展研究中心, 数据截止至2018年5月23日

互联网大数据抓取体系



互联网大数据挖掘体系



大数据研究现状

专题策略报告

《基于网络新闻热度的择时策略-互联网大数据挖掘系列专题(一)》
《那些年一起追过的财经小编选股策略-互联网财经频道文本挖掘策略》
《基于互联网挖掘的热点选股策略-互联网大数据挖掘系列专题之(五)》
《基于大数据挖掘的关联个股投资机会-互联网大数据挖掘系列专题之(六)》
《基于大数据挖掘的Smart Beta策略-互联网大数据挖掘系列专题之(七)》
《多维数据下的大数据择时策略研究-互联网大数据挖掘系列专题之(八)》
《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)》
《基于舆情的大类资产配置策略-互联网大数据挖掘系列专题之(十)》
《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(十一)》
《基于网络舆情的指数轮动策略研究-互联网大数据挖掘系列专题之(十二)》

互联网文本挖掘工具

- | | |
|----------------|----------------------|
| 1、A股新闻热度搜索工具; | 2、A股上市工具公告抓取工具; |
| 3、上市公司信息变更抓取; | 4、文本信息批量识别及处理; |
| 5、汇丰PMI实时监测工具; | 6、个股研报热点监测工具; |
| 7、特定公告实时监测工具; | 8、财经小编选股工具; |
| 9、百度指数搜索工具; | 10、A股上市公司调研信息文本挖掘工具; |

互联网舆情数据

随着移动互联网的迅速发展，人类进入数据时代。大数据带来的信息风暴正深刻改变人们的生活、工作和思维方式。互联网大数据呈现出海量性、多样性、高速性、易变性的特点，为量化投资提供了新的数据来源。

纵观海内外，对大数据的量化投资策略的研究刚刚兴起，但对互联网大数据相关的量化研究已然成为量化投资新的研究方向和热门研究领域。

本报告尝试将互联网舆情数据的变化与大小盘风格轮动问题联系起来，从大小盘的网络舆情数据中挖掘出大小盘轮动的相关规律。





01

02

03

04

02

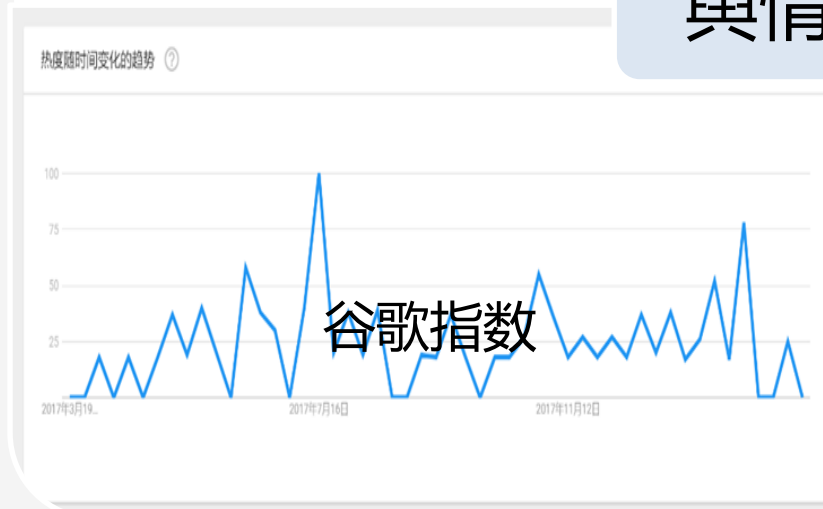
|指数轮动策略构建|

>

舆情数据



舆情数据



数据来源：百度、360、谷歌、微博、广发证券发展研究中心

舆情数据

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间下在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
微指数	新浪微博的数据分析工具，通过关键词的热议度，以及行业/类别的平均影响力，来反映微博舆情或账号的发展走势。	热词趋势、实时趋势、地域解读、属性分析	2016.3至今	2014.9
微信指数	基于微信系统的大数据分析。可以帮助大家分析关键词在微信内的热度情况。数据分析情况限于微信搜索、公众号文章以及朋友圈的数据分析。	热词趋势	最近90天	2017.3

数据来源：百度、360、谷歌、微博、微信、广发证券发展研究中心

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，将这种关注度作为投资者情绪的**直接代理变量**（Direct Proxy）。投资者对不同大小盘风格的追逐变化首先就表现为大小盘风格舆情的相对变化。这种舆情变化的强弱与大小盘风格的轮动之间的关系正是此研究的焦点所在。



舆情数据可预测性分析

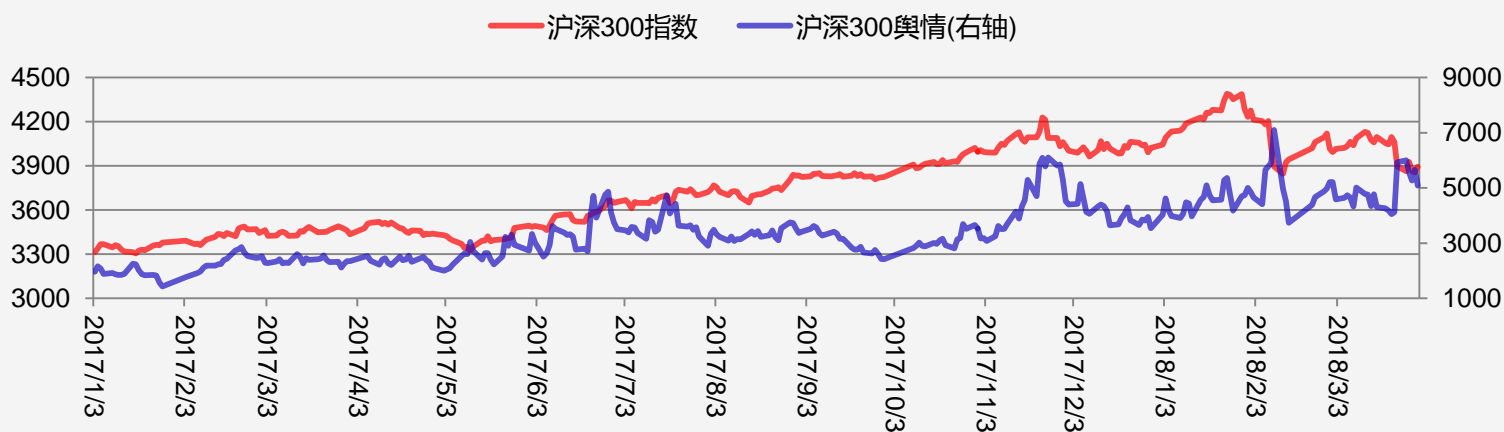
舆情数据的背后：

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。投资者情绪随着大小盘风格的变化而波动，同时也反过来影响了大小盘风格的变化。
2. 当市场出现追逐大盘或者小盘风格时，投资者关注度上升，带来大盘或者小盘风格的上涨。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，强势风格发生转向。
4. 所以通过对大小盘风格舆情相对强弱的变化趋势的分析，可以去探究投资者关注度的变化情况，从而对大小盘风格的变化进行预判。

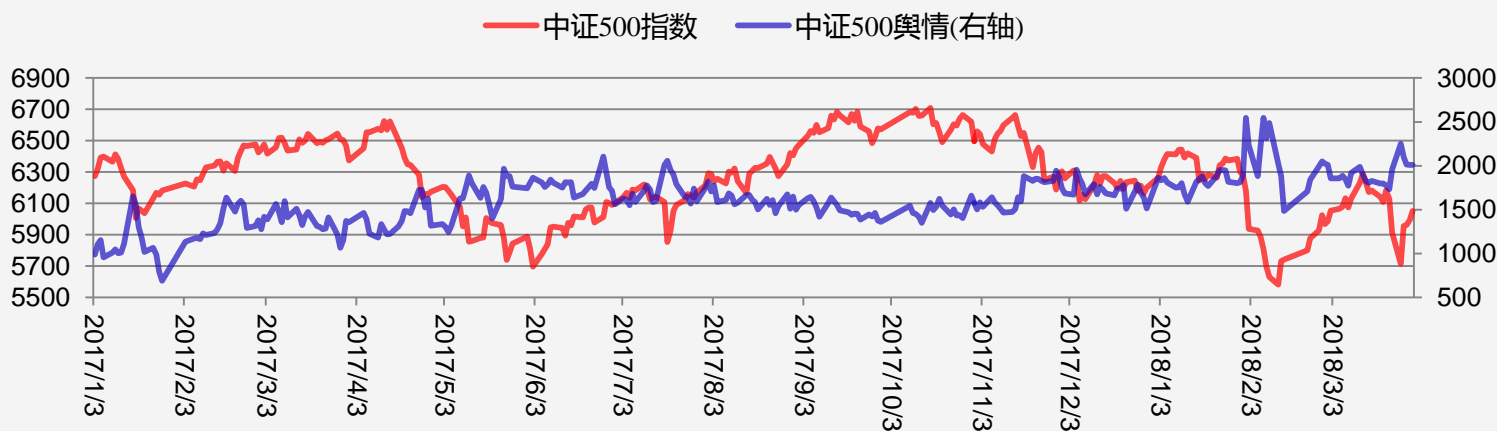


大小盘指数与舆情指数走势图

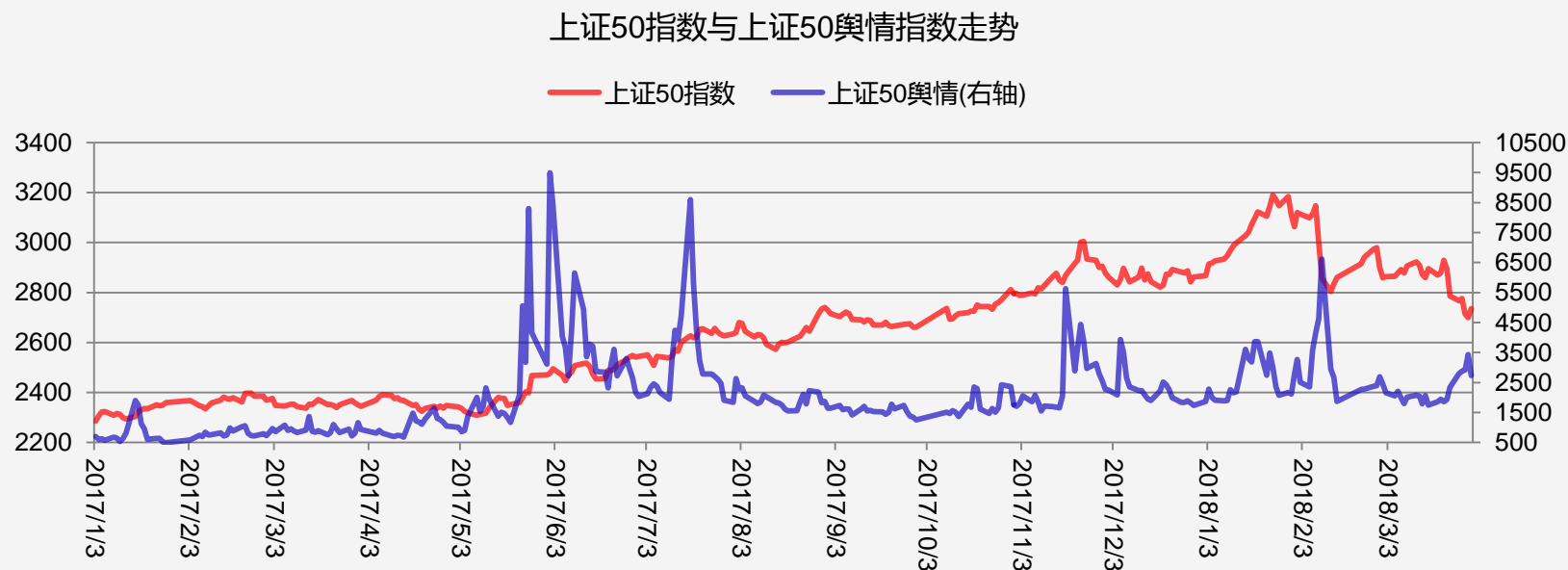
沪深300指数与沪深300舆情指数走势



中证500指数与中证500舆情指数走势



大小盘指数与舆情指数走势图



数据来源：WIND、广发证券发展研究中心

可以看出，指数和对应舆情在一定程度上存在正向相关关系。

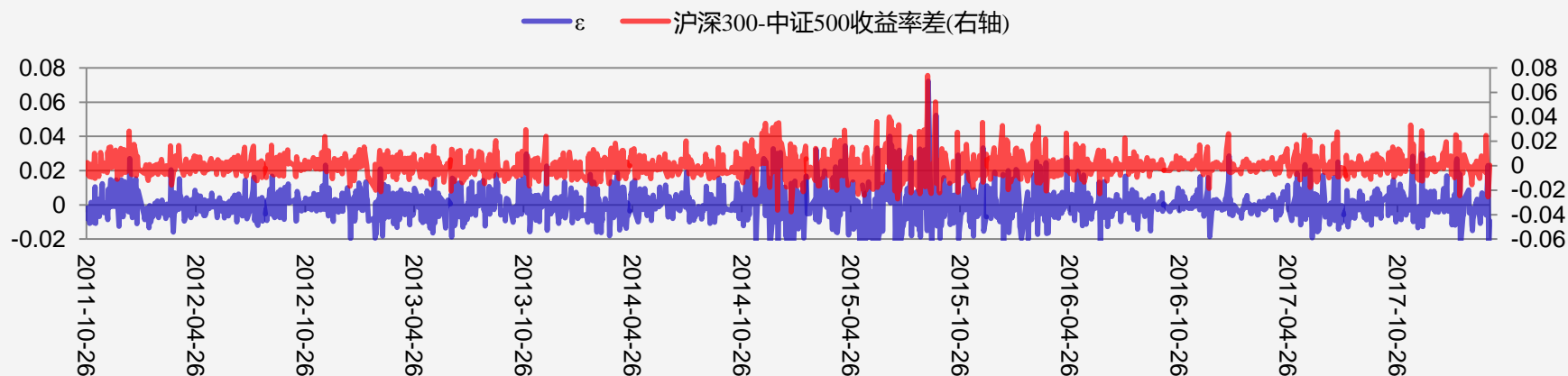
策略构建

- 数据：上证50、沪深300、中证500指数每日开盘价；
上证50、沪深300、中证500指数每日舆情数据；
- 回测区间：2011年01月01日 – 2018年03月30日
- 计算标的股票 (如沪深300与中证500) 之间的日收益率差作为因变量，以过去一段时间的情绪变动率强弱变化作为自变量
- 对 $T = 0$ ，对前 n 日的数据进行回归，得到回归系数及截距项，代入 $T = 0$ 的数据中求得残差 ε ，若 $\varepsilon > 0$ ，则做多沪深300 (上证50)，做空中证500；若 $\varepsilon < 0$ ，则做多中证500，做空沪深300 (上证50)

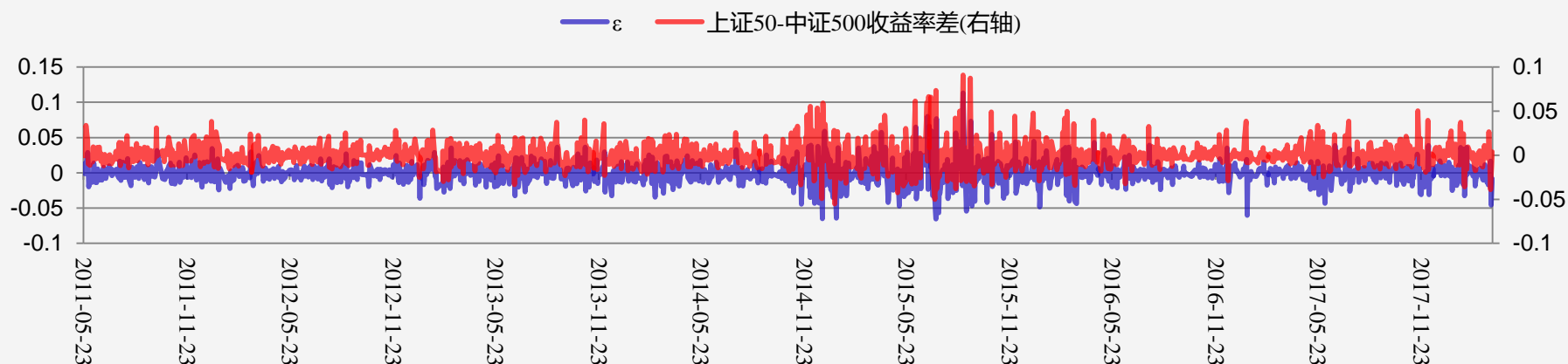
$$Ret_div = \alpha + \beta_i Sentiment_div_i + \varepsilon$$

收益率差与残差走势图

沪深300-中证500收益率差与残差走势



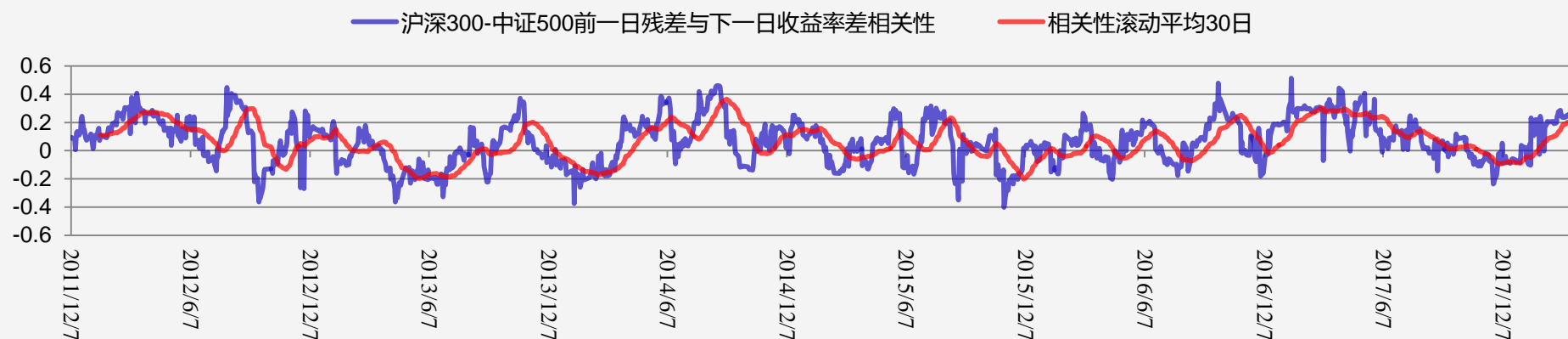
上证50-中证500收益率差与残差走势



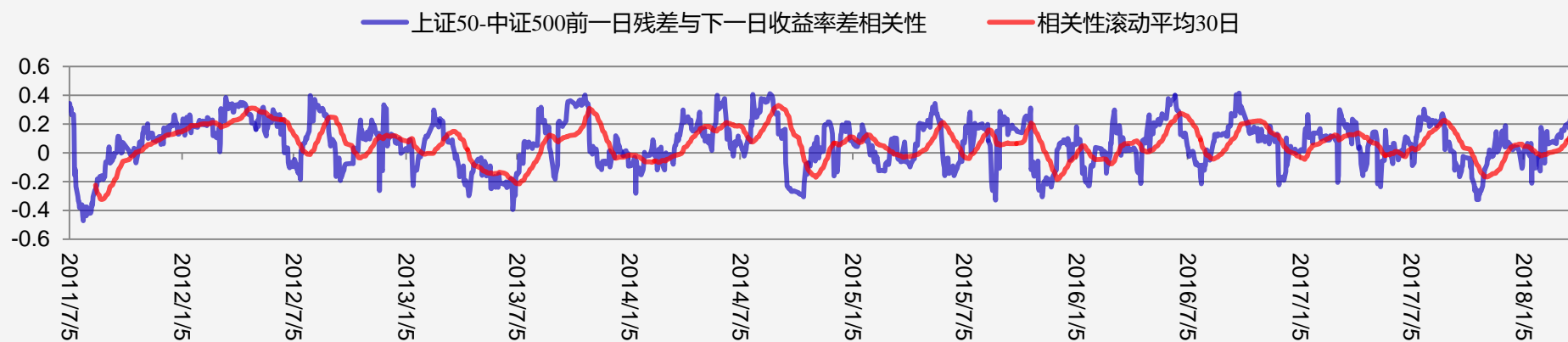
数据来源：WIND、广发证券发展研究中心

残差与收益率差相关性走势图

沪深300-中证500前一日残差与下一日收益率差相关性走势



上证50-中证500前一日残差与下一日收益率差相关性走势



数据来源：WIND、广发证券发展研究中心



01

02

03

04

03

|策略实证结果|

>

策略表现---沪深300与中证500轮动

指标	沪深300-中证500
择时周期	2011-1-1至2018-3-29
累计收益率	249.14%
年化收益率	22.13%
波动率	14.64%
信息比率	1.51
单次平均收益率	0.08%
判断正确率	43.37%
平均盈利率	1.28%
平均亏损率	-0.66%
盈亏比	1.94
最大连续上涨天数	16
最大连续下跌天数	7
单次最大盈利	18.18%
单次最大亏损	-5.21%
最大回撤率	-12.25%
交易次数	725
平均择时天数	3.24

策略净值表现



数据来源：WIND、广发证券发展研究中心

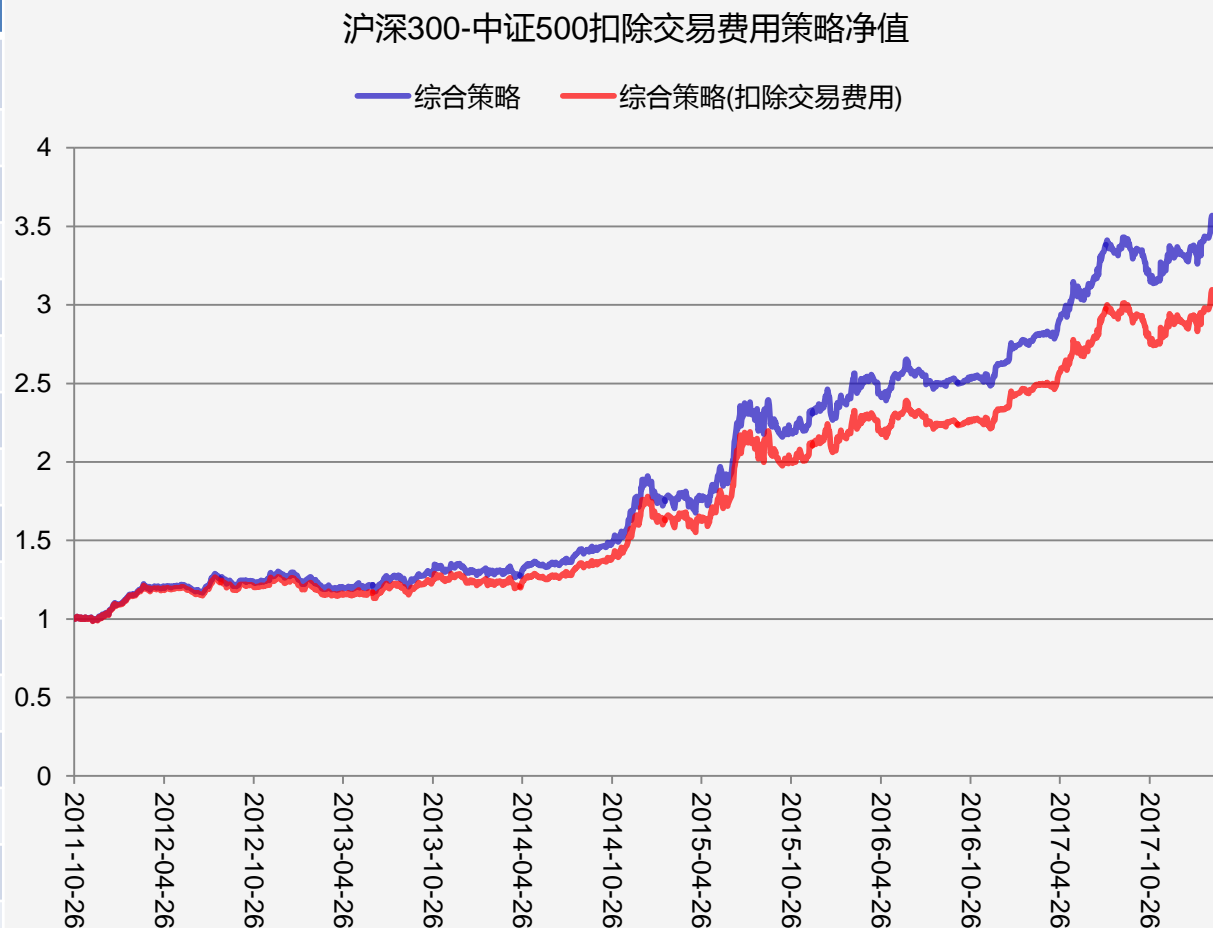
策略分年度表现---沪深300与中证500轮动

	累积收益率	最大回撤	年化波动率	年化收益率	信息比率
2011年	3.37%	-2.83%	10.87%	18.87%	173.50%
2012年	23.32%	-6.30%	9.78%	24.07%	246.03%
2013年	2.39%	-9.39%	12.40%	2.51%	20.25%
2014年	44.36%	-5.27%	14.90%	45.45%	305.08%
2015年	24.49%	-12.25%	22.72%	25.16%	110.75%
2016年	11.85%	-7.97%	12.76%	12.16%	95.33%
2017年	26.53%	-8.57%	11.69%	27.26%	233.27%
2018年	5.17%	-5.06%	16.36%	24.25%	148.23%
All	249.14%	-12.25%	14.64%	22.12%	151.14%

数据来源：WIND、广发证券发展研究中心

策略表现---沪深300与中证500轮动（扣除交易费用）

指标	沪深300-中证500
择时周期	2011-1-1至2018-3-29
累计收益率	202.12%
年化收益率	19.33%
波动率	14.64%
信息比率	1.32
单次平均收益率	0.075%
判断正确率	42.82%
平均盈利率	1.26%
平均亏损率	-0.66%
盈亏比	1.90
最大连续上涨天数	16
最大连续下跌天数	7
单次最大盈利	18.13%
单次最大亏损	-5.23%
最大回撤率	-12.84%
交易次数	725
平均择时天数	3.24



策略分年度表现---沪深300与中证500轮动（扣除交易费用）

	累积收益率	最大回撤	年化波动率	年化收益率	信息比率
2011年	2.94%	-3.05%	10.85%	16.30%	150.29%
2012年	20.50%	-6.64%	9.79%	21.15%	216.06%
2013年	0.12%	-10.37%	12.39%	0.13%	1.03%
2014年	41.43%	-5.36%	14.88%	42.43%	285.20%
2015年	21.64%	-12.84%	22.70%	22.22%	97.90%
2016年	9.20%	-8.06%	12.76%	9.44%	73.97%
2017年	23.88%	-9.01%	11.70%	24.53%	209.76%
2018年	4.54%	-5.14%	16.38%	21.07%	128.64%
All	202.12%	-12.84%	14.63%	19.33%	132.12%

数据来源：WIND、广发证券发展研究中心

策略表现---上证50与中证500轮动

指标	上证50-中证500
择时周期	2011-1-1至2018-3-29
累计收益率	496.80%
年化收益率	30.68%
波动率	21.42%
信息比率	1.43
单次平均收益率	0.12%
判断正确率	44.58%
平均盈利率	1.79%
平均亏损率	-0.99%
盈亏比	1.81
最大连续上涨天数	10
最大连续下跌天数	8
单次最大盈利	17.21%
单次最大亏损	-8.23%
最大回撤率	-17.68%
交易次数	768
平均择时周期	3.26



数据来源：WIND、广发证券发展研究中心

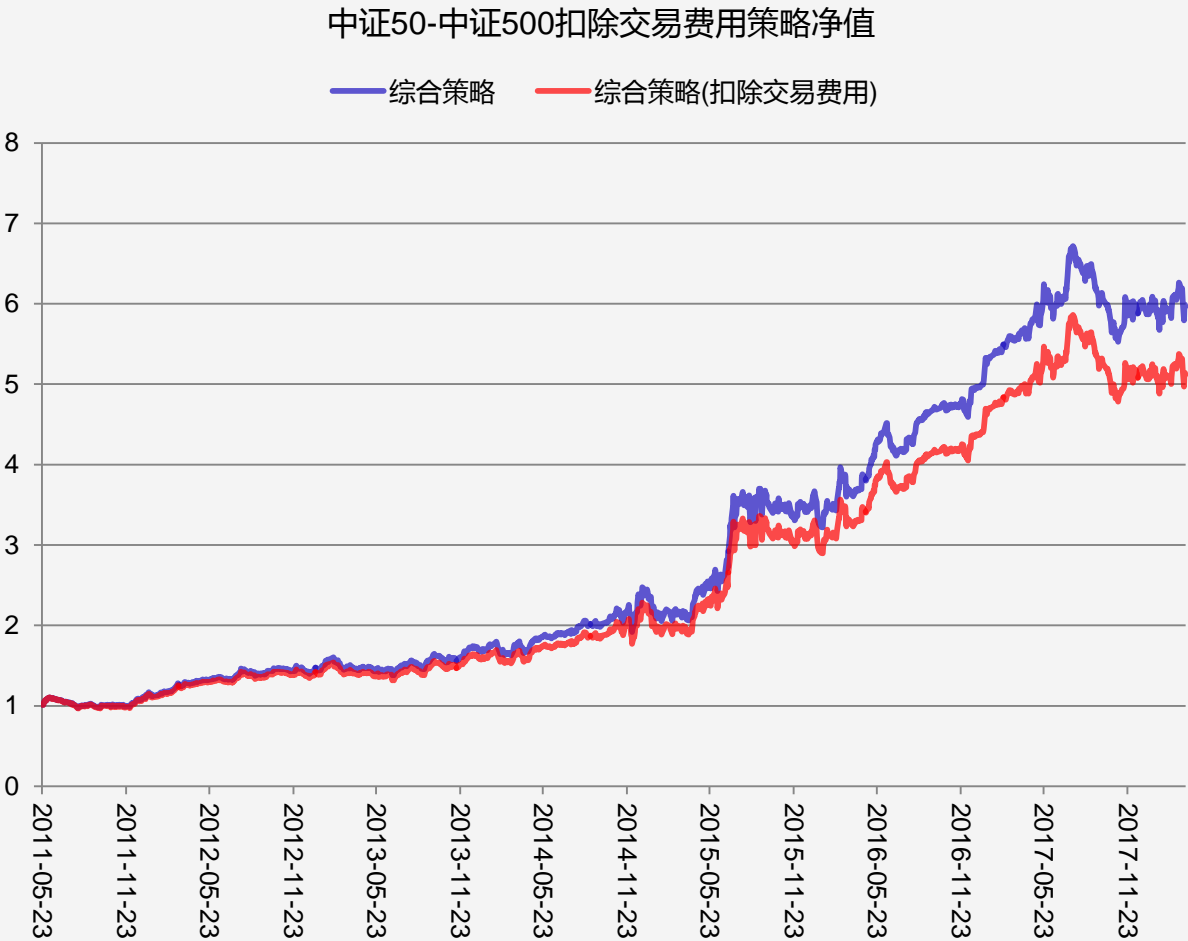
策略分年度表现---上证50与中证500轮动

	累积收益率	最大回撤	年化波动率	年化收益率	信息比率
2011年	11.19%	-11.95%	14.02%	18.92%	134.95%
2012年	27.93%	-6.97%	13.42%	28.84%	214.93%
2013年	20.10%	-13.77%	17.97%	21.22%	118.08%
2014年	41.87%	-15.00%	22.33%	42.88%	192.05%
2015年	42.71%	-16.14%	35.55%	43.96%	123.67%
2016年	43.56%	-12.21%	18.45%	44.85%	243.06%
2017年	19.59%	-17.68%	16.86%	20.11%	119.33%
2018年	0.50%	-7.50%	23.19%	2.18%	9.41%
All	496.79%	-17.68%	21.42%	30.68%	143.26%

数据来源：WIND、广发证券发展研究中心

策略表现---上证50与中证500轮动（扣除交易费用）

指标	上证50-中证500
择时周期	2011-1-1至2018-3-29
累计收益率	412.06%
年化收益率	27.71%
波动率	21.41%
信息比率	1.29
单次平均收益率	0.11%
判断正确率	43.67%
平均盈利率	1.80%
平均亏损率	-0.98%
盈亏比	1.83
最大连续上涨天数	10
最大连续下跌天数	9
单次最大盈利	17.17%
单次最大亏损	-8.27%
最大回撤率	-18.34%
交易次数	768
平均择时周期	3.26



策略分年度表现---上证50与中证500轮动（扣除交易费用）

	累积收益率	最大回撤	年化波动率	年化收益率	信息比率
2011年	9.47%	-12.46%	14.01%	15.93%	113.67%
2012年	25.23%	-7.18%	13.42%	26.04%	194.08%
2013年	17.45%	-14.61%	17.95%	18.40%	102.51%
2014年	38.67%	-15.07%	22.32%	39.60%	177.39%
2015年	39.68%	-16.35%	35.52%	40.83%	114.94%
2016年	40.36%	-12.33%	18.45%	41.54%	225.12%
2017年	17.10%	-18.34%	16.87%	17.56%	104.10%
2018年	-0.10%	-7.51%	23.17%	-0.42%	-1.82%
All	412.06%	-18.34%	21.41%	27.72%	129.48%

数据来源：WIND、广发证券发展研究中心



04

|结论|

>

基于舆情回归的指数轮动策略根据舆情回归残差和行情变化有较强的正相关关系，且舆情回归残差会领先于行情变化，构建了量化择时策略，实证结果表明利用舆情信息作为择时信号源在历史区间内具有良好表现。策略要点：

1、搜索舆情指数基于搜索引擎海量用户基数和使用频数构建，能够代表投资者情绪的直接代理变量。我们利用网络爬虫程序抓取了大小盘风格对应的搜索舆情的历史数据，从Wind提取对应的指数的历史数据用于策略构建。

2、策略中基于大小盘风格的舆情与行情数据变化构建指数轮动策略，观察大小盘风格指数前一日舆情回归的残差以及行情变化的正负关系，在下一个交易日根据这种变化的正负关系分别做多、做空对应的指数。

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（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。

Thanks !
谢谢