

基于新闻舆情的选股策略研究

——互联网大数据挖掘系列研究之(十四)

罗 军 S0260511010004
020-87579006

邮箱：lj33@gf.com.cn

安宁宁：S0260512020003
0755-23948352

邮箱：ann@gf.com.cn

陈原文 S0260517080003

邮箱：chenyuanwen@gf.com.cn
0755-82797057

广发证券金融工程
2018年11月15日





01

|研究背景|



互联网新闻数据

随着计算机科学技术的快速发展，媒体新闻的传递从过去的只有电视、报纸、杂志几种渠道，发展到今天以互联网媒体为主流媒体进行信息传递。

在金融投资领域，人们投资决策所需的各种资讯很大部分是通过各种互联网媒体获得的，而媒体的新闻和报道往往引导着投资者的注意力。

纵观海内外，学术界关于新闻媒体对资本市场的研究，有着一系列丰富的研究成果。

雪球

聪明的投资者都在这里

和讯
hexun.com

中国财经网络门户

sina 新浪财经
finance.sina.com.cn东方财富网
eastmoney.com

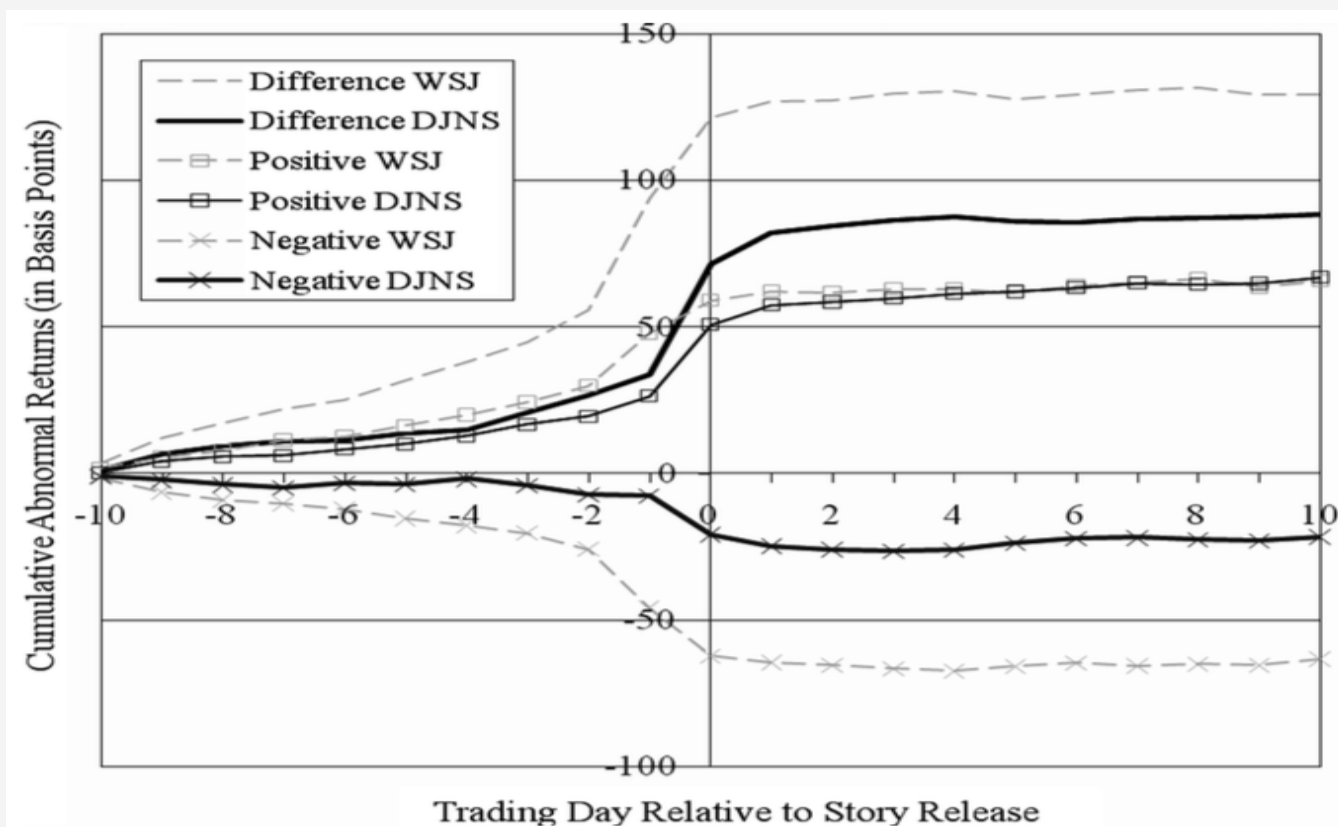
Baidu 新闻

腾讯新闻 事实派

数据来源：广发证券发展研究中心

海外相关研究

2008年，Tetlock首次利用情感分析法对《华尔街日报》、《道琼斯新闻》等媒体新闻内容进行情感分析，判定新闻内容信息为积极还是消极，实证结果表明消极信息会带来向下的价格压力。



数据来源：TETLOCK等《More Than Words: Quantifying Language to Measure Firms' Fundamentals》，广发证券发展研究中心

海外相关研究

2009年，Fang 和 Peress 利用几大主流报纸的文章数量，作为媒体覆盖率的衡量指标。通过分析1993年至2002年期间，NASDAQ 成分股的媒体覆盖率和股票回报率之间的关系，得出结论：在不区分正负面新闻的情况下，受到广泛报道的股票收益显著低于无报道股票。

	Average Monthly Return					Average No. of Stocks		
	Media Coverage			No – High	<i>t</i> -Statistics for No – High	Media Coverage		
	No	Low	High			No	Low	High
All stocks	1.35	1.11	0.96	0.39	2.13	1,430.08	284.82	245.40

数据来源：LILY FANG, JOEL PERESS等《Media Coverage and the Cross-section of Stock Returns》，广发证券发展研究中心

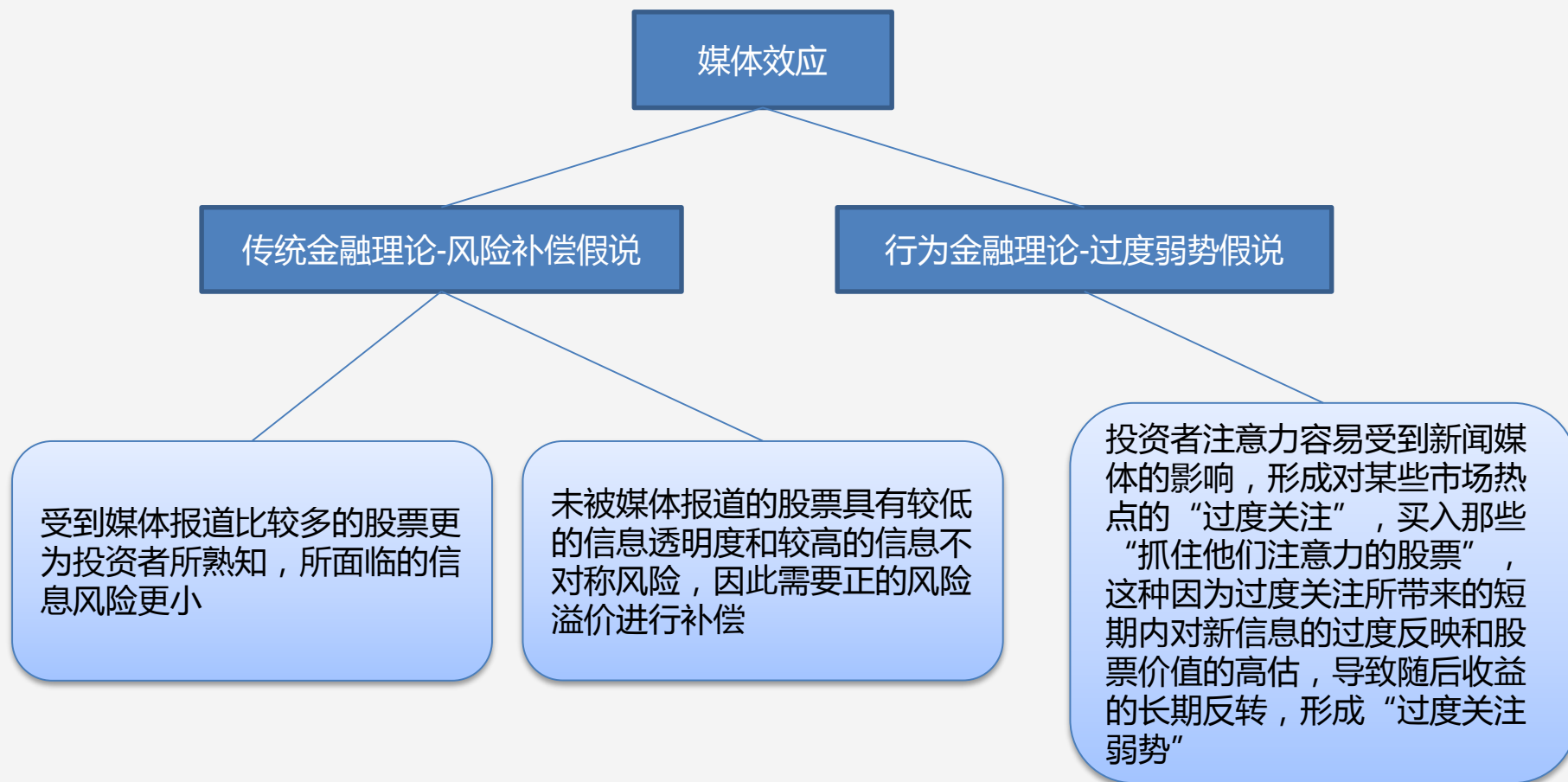
海外相关研究

2011年，Joseph等利用谷歌提供的谷歌搜索量指数（SVI）来衡量投资者关注度并对Russell 3000 指数样本股票在2004年至2008年的数据进行研究，发现谷歌搜索量指数与投资者关注度有正相关性，可以作为个人投资者关注度的衡量。该指数的上升能够预测股票收益在两周内的上涨和一年内价格的反转。

	Week 1 (1)	Week 2 (2)	Week 3 (3)	Week 4 (4)	Week 5-52 (5)
ASVI	18.742*** (7.000)	14.904** (7.561)	3.850 (6.284)	-1.608 (6.903)	-28.912 (17.162)
Log Market Cap × ASVI	-21.182*** (6.508)	-15.647** (6.768)	-4.710 (6.516)	4.290 (6.398)	16.834 (88.624)
Log Market Cap	2.653 (3.023)	3.858 (3.160)	3.144 (3.063)	3.575 (3.186)	-39.229 (67.405)
Percent Dash-5 Volume × ASVI	3.552** (1.639)	1.904 (1.522)	1.687 (1.612)	-2.744 (1.717)	16.258 (23.822)
Percent Dash-5 Volume	1.607 (1.644)	1.351 (1.652)	1.486 (1.659)	0.364 (1.711)	119.901*** (31.765)
APSVI	-2.532*** (0.930)	-1.379 (0.990)	-0.701 (0.808)	-0.704 (0.639)	2.286 (9.909)
Absolute Abn Ret	1.314 (1.879)	-2.389 (1.979)	-1.128 (1.563)	-0.463 (1.405)	-1.510 (28.505)

数据来源：JOSEPH等《In Search of Attention》，广发证券发展研究中心

媒体效应的解释



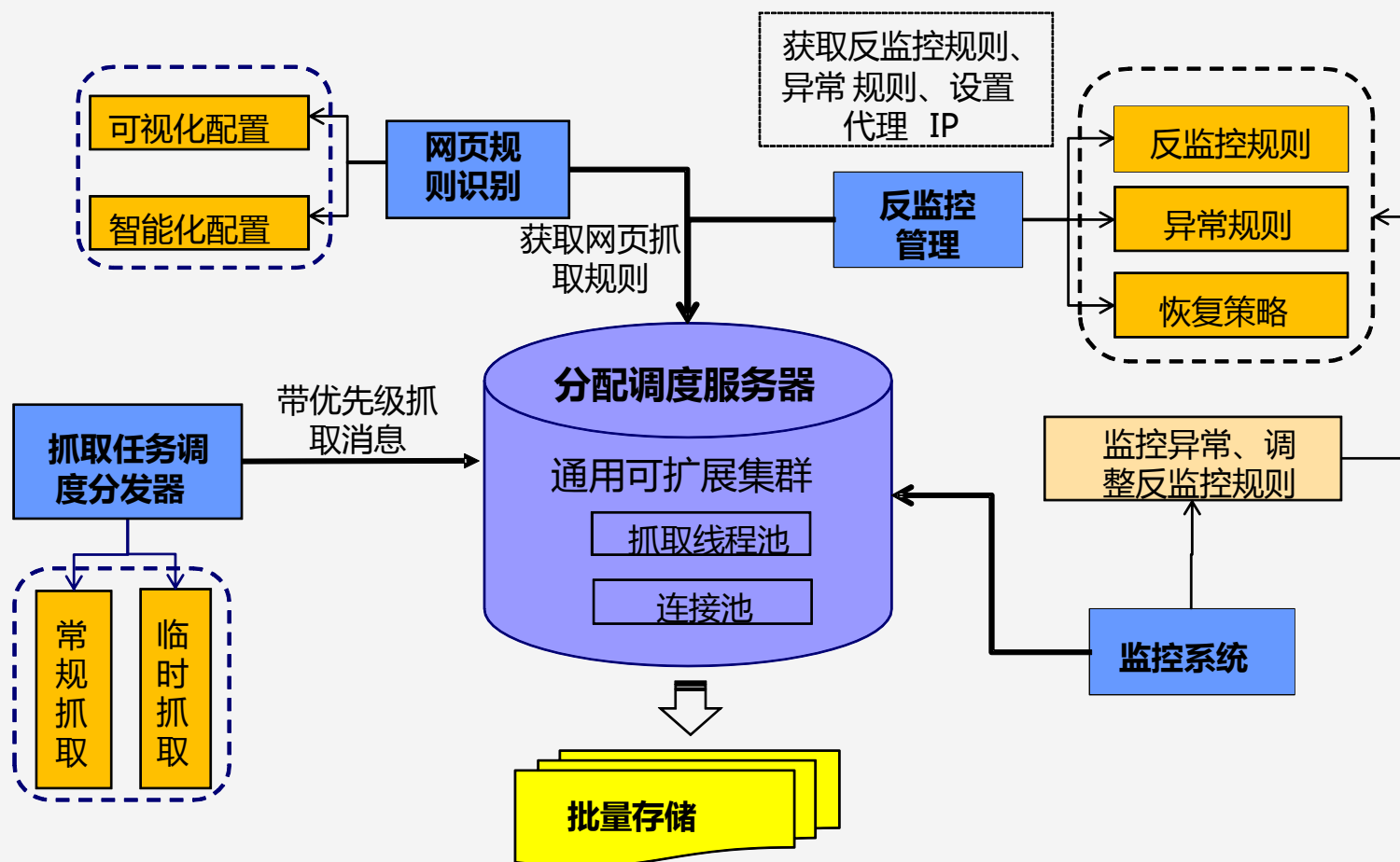
数据来源：广发证券发展研究中心

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
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
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
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
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
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
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
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
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
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

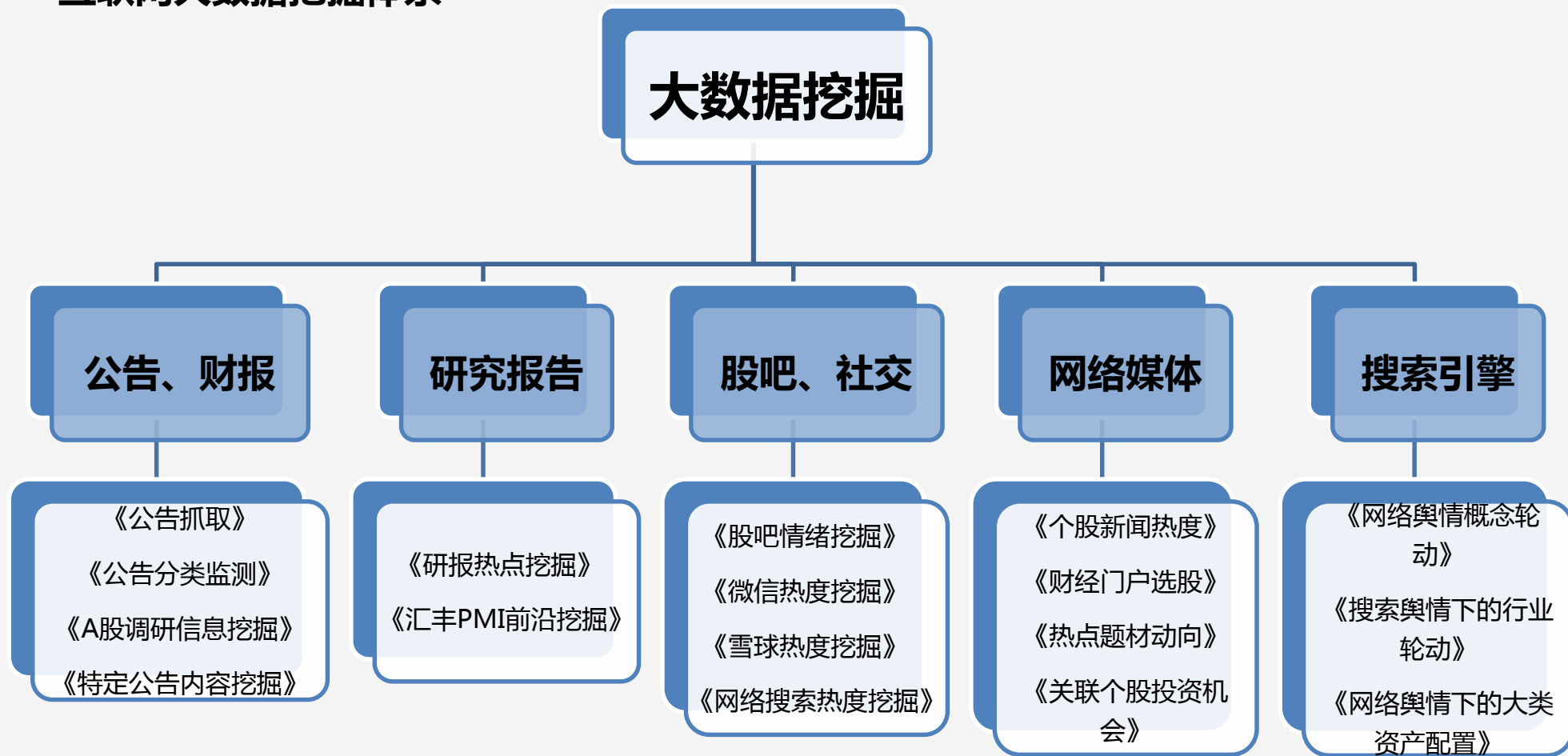
数据来源:公司官网、广发证券发展研究中心, 数据截止至2018年11月14日

互联网大数据抓取体系



数据来源：广发证券发展研究中心

互联网大数据挖掘体系



数据来源：广发证券发展研究中心

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）》
《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）》

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;

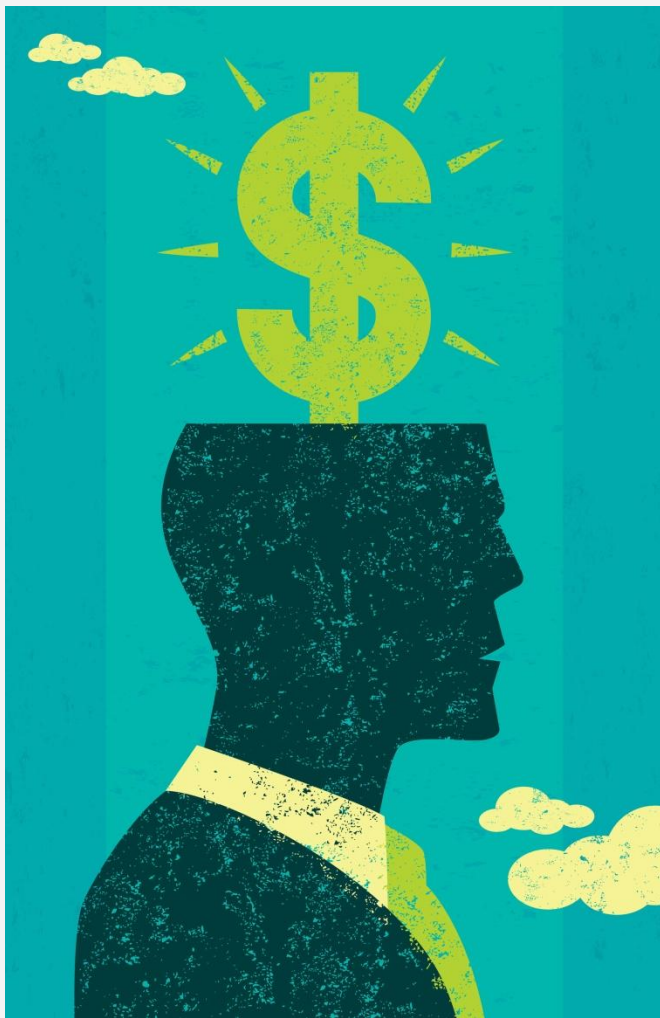


02

|策略构建|



互联网新闻数据可预测性分析



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和模型所无法解释，比如规模效应、日历效应等。行为金融学从投资者非理性心理和行为等方面对这些现象给予一定的理论支撑。

有限关注一直是行为金融学中研究的热门话题，当投资者面对诸多信息时，极易受到外界信息的干扰，进而导致个性化的选择注意差别，最终反映在股票价格上面。

有限关注的存在在一定程度上限制了投资者学习和决策的过程。我国大部分投资者为散户投资者，由于缺乏专业的知识，容易受到外界信息的干扰，从而造成投资损失。

投资者关注度的衡量

目前学术上一般用引起投资者注意的事件或信息作为投资者关注度间接的代理变量。传统关注度的代理变量有超额收益、异常交易量、换手率、涨跌停板等，但这些指标都是金融资产本身的交易特性和价格行为，并不能直接反映投资者对股票市场的关注程度。

在现实股票市场中，互联网成为广大股民获取信息最主要的渠道。互联网搜索引擎提供的相关数据，准确刻画投资者的心理活动。因此，我们可以使用**个股的新闻数量**作为投资者关注度的衡量指标，探究投资者关注度和股票收益之间的关系。



投资者关注度的衡量

为了尽可能地获取个股在所有热门新闻网站上的新闻数量，有以下两种思路：

1. 分别获取各个热门财经新闻网站有关个股的新闻信息，这种方法的缺点是效率极低，需对不同网站的网页源码进行解析，导致抓取的工作量大，耗费时间长。
2. 借助搜索引擎的相关新闻搜索功能，通过搜索引擎我们即可获得个股在各个热门财经网站上的新闻信息。这既增加了效率也增加了数据量。目前具有新闻搜索功能的搜索引擎主要有新浪财经新闻搜索、百度新闻搜索、360新闻搜索等。

本文采用**第二种搜索引擎采集新闻**来构建投资者关注度指标。

个股新闻数量描述性统计

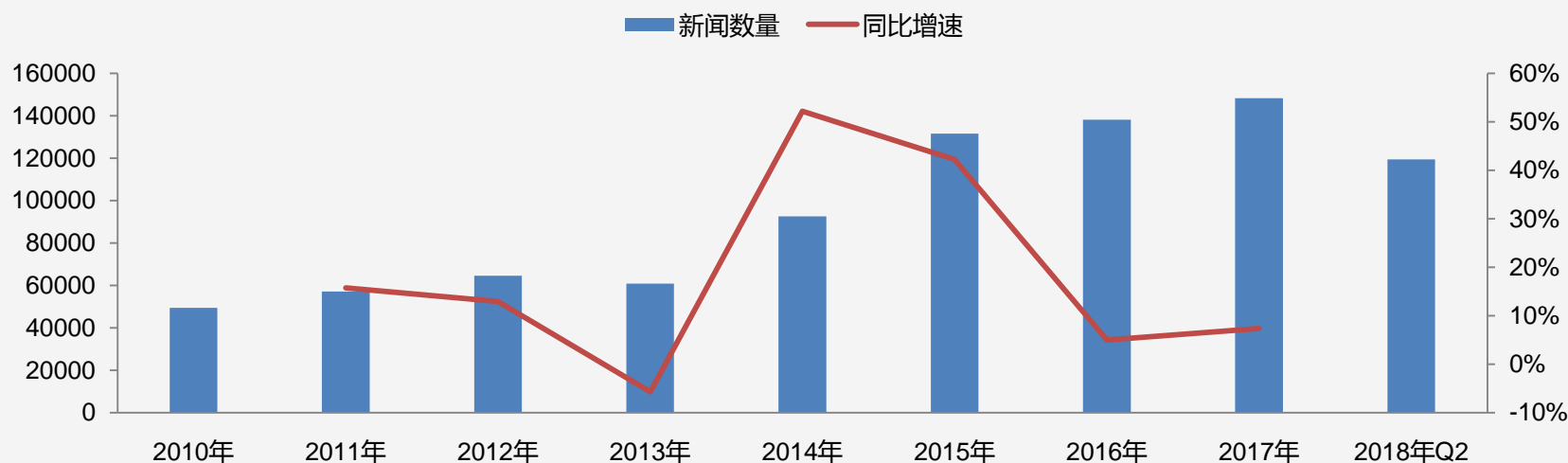
指标	数值（单位：条）
股票个数	1117
平均数	706.56
标准差	790.76
最小值	34
25%分位数	364
50%分位数	518
75%分位数	736
最大值	8832

数据来源：广发证券发展研究中心

对个股从2010年至2018年6月的新闻量数据进行统计。其中，新闻数量最多的来自华夏幸福（8832条）、中国动力（8259条）、科大讯飞（5714条）；新闻数量最少的来自中迪投资（43条）、圣济堂（41条）、华创阳安（34条）。

新闻数量分年度统计

新闻数量分年度统计(单位：条)

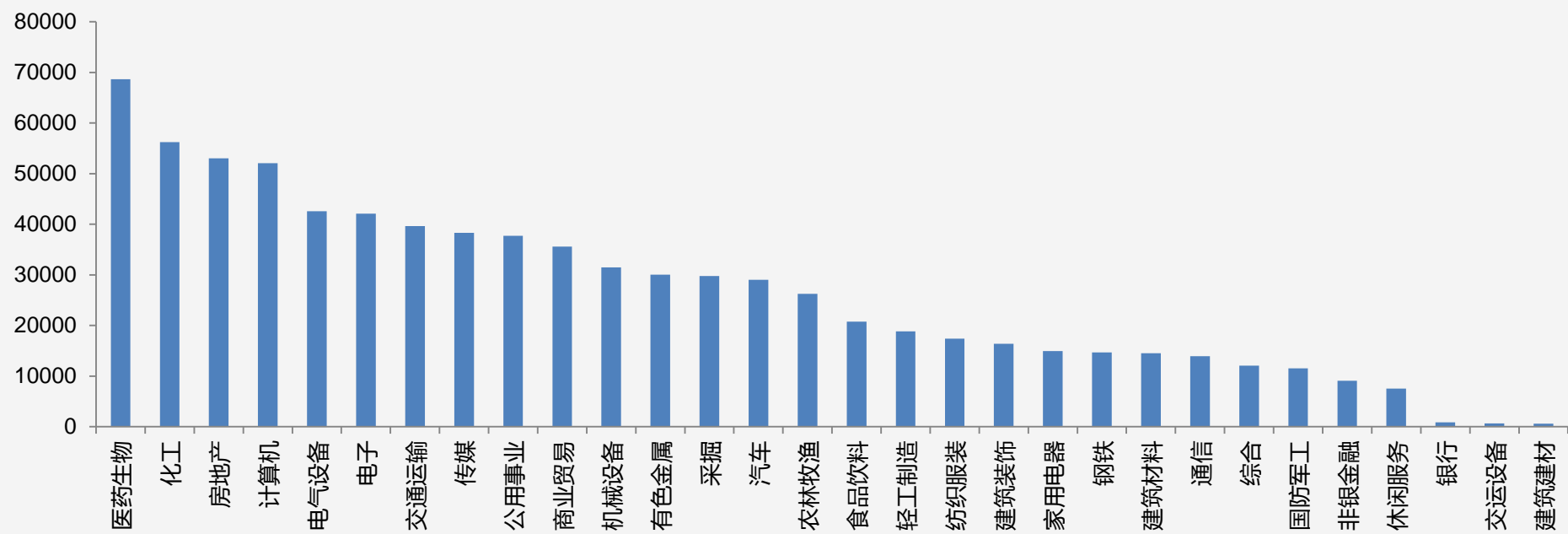


数据来源：广发证券发展研究中心

随着移动互联网的普及和快速发展，关于个股的财经新闻数量显著增长，2018年上半年个股新闻数据总量高达12万条，占2017年年度新闻总量的80%。

新闻数量分行业统计

行业新闻数量一览(单位：条)

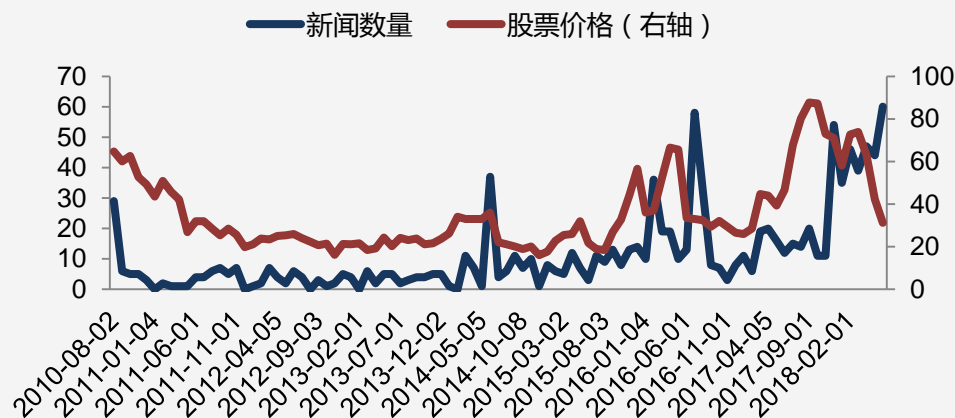


数据来源：广发证券发展研究中心

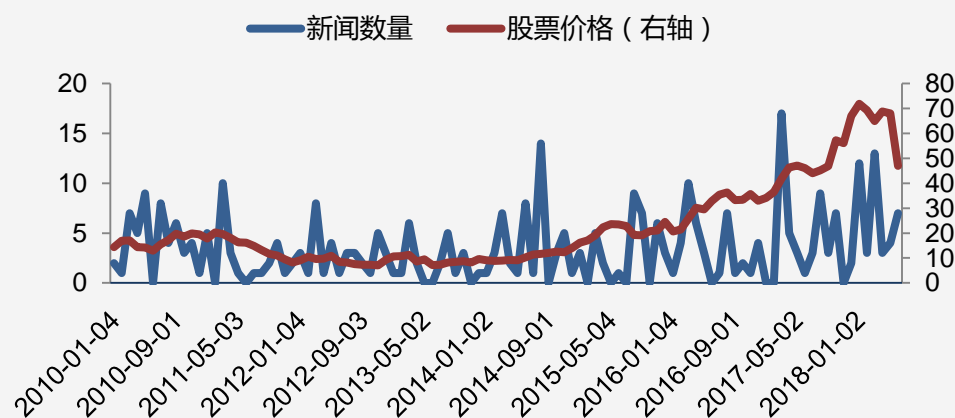
行业分布上，新闻量多的行业主要集中在医药生物、化工、房地产、计算机等行业。

个股新闻数量与股票价格走势

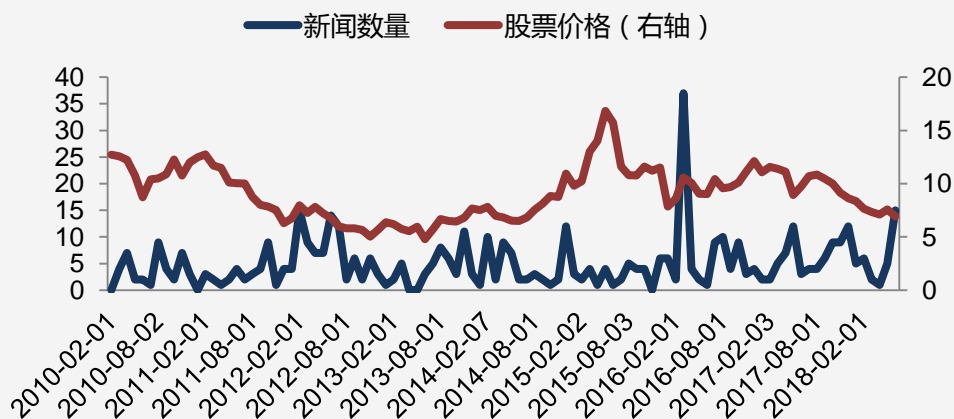
赣锋锂业新闻数量(单位：条)与股票价格(单位：元)走势图



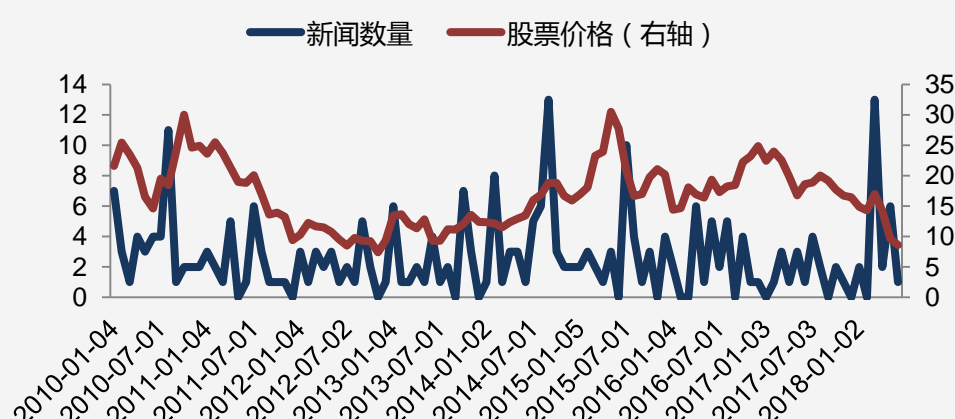
小天鹅A新闻数量(单位：条)与股票价格(单位：元)走势图



中化国际新闻数量(单位：条)与股票价格(单位：元)走势图



贵航股份新闻数量(单位：条)与股票价格(单位：元)走势图

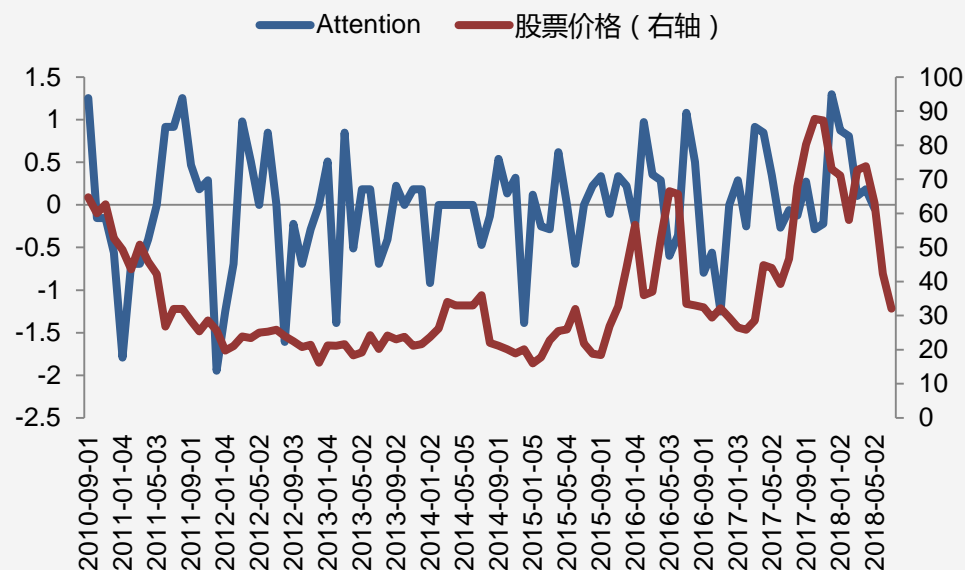


策略构建

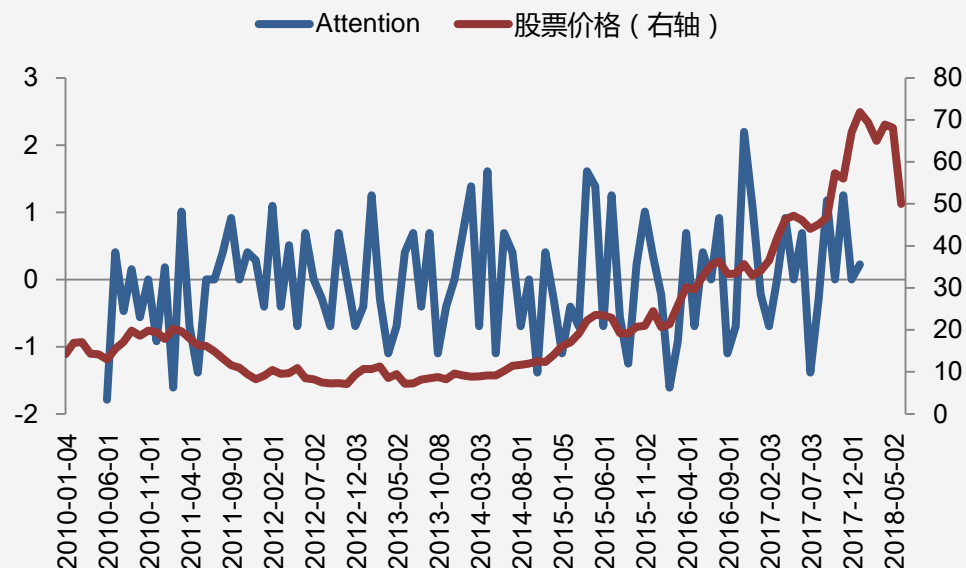
- 数据：中证500指数成分股月度的新闻数量；
中证500指数日频数据
- 股票池：N只股票
- 回测区间：2010年01月01日 – 2018年07月02日
- 策略参数：周期 M
- 根据个股当月的新闻数量 $News_{i,t}$ 构造**投资者关注度指标**：
$$Attention_{i,t} = \ln(News_{i,t}) - \ln[MEDIAN(News_{i,t-1}, News_{i,t-2}, \dots, News_{i,t-M})]$$
- 在每个月的第一个交易日进行调仓，做多上个月关注度较低的股票

投资者关注度与股票价格走势

赣锋锂业投资者关注度与股票价格(单位:元)走势图



小天鹅A投资者关注度与股票价格(单位:元)走势图



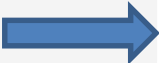
数据来源：Wind，广发证券发展研究中心

可以看出，投资者关注度和股票价格在一定程度上存在负向关系。

投资者关注度策略构建 — 举例

000418.SZ

月份	新闻数量	中位数	Attention
...
2010-06	0	5	-1.79176
2010-07	8	5	0.405465
2010-08	4	7	-0.47
2010-09	6	5	0.154151
2010-10	3	6	-0.55962



升序排列



2010-06

股票代码	Attention
002332.SZ	-2.30259
002353.SZ	-2.19722
300058.SZ	-2.19722
002244.SZ	-2.07944
000988.SZ	-1.94591
...	...
600664.SH	2.079442
000903.SZ	2.351375
600624.SH	2.484907
600151.SH	2.525729
002399.SZ	2.538974



多头组合

股票代码	Attention
002332.SZ	-2.30259
002353.SZ	-2.19722
...	...
600850.SH	-1.50408
002362.SZ	-1.47331



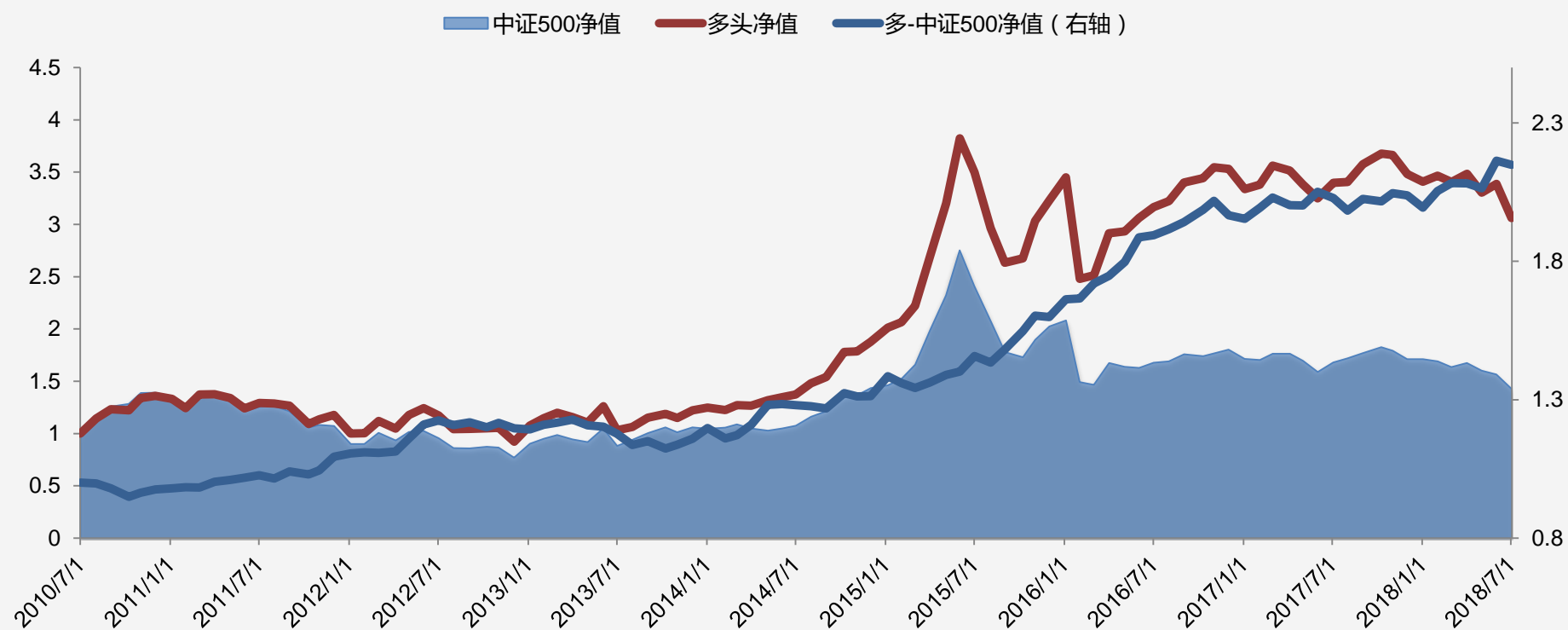
03

|策略实证结果|

>

策略表现---中证500选股（等权）

策略表现



策略分年度对冲表现

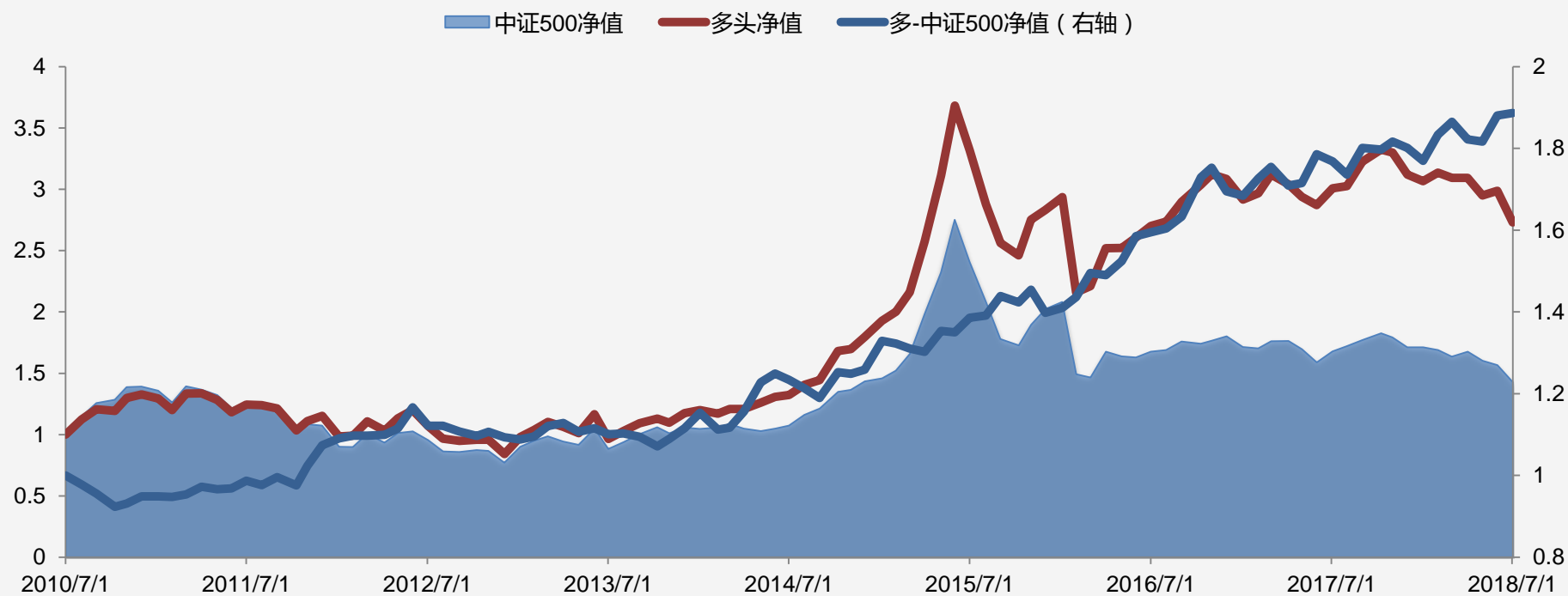
年份	累积对冲收益率	对冲年化收益率	最大回撤	多头累计收益率	基准累计收益率	信息比率
2010	-2.40%	-5.70%	-4.80%	36.00%	39.2%	-0.92
2011	12.10%	12.10%	-1.20%	-13.50%	-22.9%	2.27
2012	9.40%	9.40%	-2.30%	-21.70%	-28.2%	1.455
2013	-3.20%	-3.20%	-8.50%	32.80%	37.4%	-0.542
2014	13.30%	13.30%	-3.10%	53.40%	35.6%	1.552
2015	21.80%	21.80%	-3.00%	72.00%	41.1%	2.591
2016	23.00%	23.00%	-2.60%	9.40%	-10.9%	3.513
2017	3.60%	3.60%	-3.30%	-1.40%	-5.1%	0.685
2018	5.40%	9.50%	-0.90%	-12.10%	-16.7%	1.218

策略分年度换手率统计

年份	平均值	最大值	最小值	标准差
2010	54.00%	74.10%	35.70%	16.00%
2011	61.30%	100.00%	34.50%	21.60%
2012	77.60%	100.00%	51.90%	19.20%
2013	52.50%	87.50%	36.70%	15.20%
2014	62.40%	100.00%	40.00%	21.40%
2015	59.30%	100.00%	36.70%	18.60%
2016	79.70%	100.00%	50.00%	20.90%
2017	69.60%	90.00%	35.70%	19.10%
2018	87.20%	90.00%	83.30%	2.80%
All	66.50%	100.00%	34.50%	20.90%

策略表现---中证500选股（流通市值加权）

策略表现



策略分年度对冲表现

年份	累积对冲收益率	对冲年化收益率	最大回撤	多头累计收益率	基准累计收益率	信息比率
2010	-5.10%	-11.80%	-5.60%	32.70%	39.2%	-1.682
2011	13.20%	13.20%	-2.10%	-13.10%	-22.9%	1.807
2012	1.80%	1.80%	-6.20%	-27.10%	-28.2%	0.271
2013	2.00%	2.00%	-5.10%	39.80%	37.4%	0.394
2014	12.80%	12.80%	-4.80%	52.90%	35.6%	1.27
2015	11.00%	11.00%	-3.90%	57.60%	41.1%	1.254
2016	21.20%	21.20%	-3.30%	8.90%	-10.9%	2.697
2017	6.30%	6.30%	-2.80%	1.10%	-5.1%	0.886
2018	4.80%	8.30%	-2.60%	-12.50%	-16.7%	1.087

策略分年度换手率统计

年份	平均值	最大值	最小值	标准差
2010	54.00%	74.10%	35.70%	16.00%
2011	61.30%	100.00%	34.50%	21.60%
2012	77.60%	100.00%	51.90%	19.20%
2013	52.50%	87.50%	36.70%	15.20%
2014	62.40%	100.00%	40.00%	21.40%
2015	59.30%	100.00%	36.70%	18.60%
2016	79.70%	100.00%	50.00%	20.90%
2017	69.60%	90.00%	35.70%	19.10%
2018	87.20%	90.00%	83.30%	2.80%
All	66.50%	100.00%	34.50%	20.90%



04

|结论|

>

基于新闻数量的选股策略根据新闻数量和股票价格之间存在联动关系构造投资者关注度指标，并通过投资者关注度指标构建策略。实证结果表明，在中证500成分股中，利用个股新闻数量可以取得超额收益。策略要点主要有：

1. 通过网络爬虫抓取中证500从2010年至今的新闻量数据，以该数据构造投资者关注度指标。
2. 策略基于投资者关注度的高低，对中证500成分股进行分档，在每个月第一个交易日做多关注度低的股票，策略表现优异。

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