基于大数据挖掘的行业轮动策略研究

史庆盛 S0260513070004
广发证券金融工程
2017年2月

02

03

04

05

01

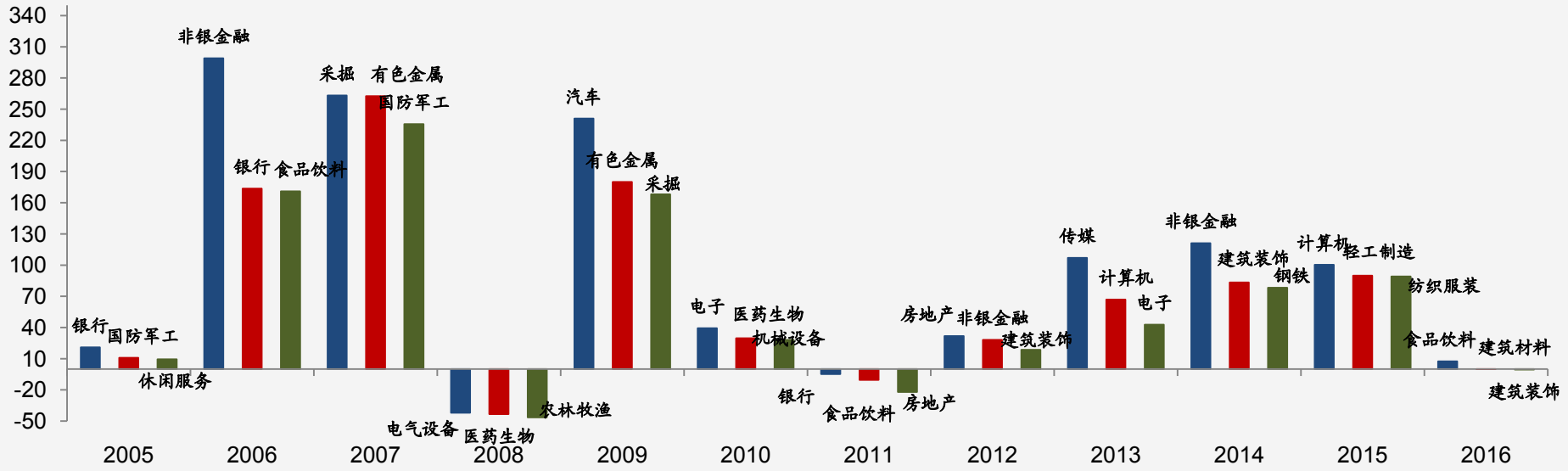
|研究目的及背景|

>

一、研究目的

行业表现一览-年度表现

历史年度行业涨幅前三分布



数据来源：广发证券发展研究中心

行业板块业绩分化严重：

2015年计算机行业涨幅最高约为**100.29%**，垫底行业为非银金融，跌幅**-16.90%**。

2016年食品饮料行业涨幅最高约为**7.43%**，而传媒行业涨幅约为**-32.39%**。

如何把握行业轮动成权益配置的关键！

一、研究目的

行业表现一览-月度表现

	银行	非银金融	食品饮料	房地产	休闲服务	国防军工	有色金属	计算机	采掘	传媒	医药生物	通信	家用电器	综合
单月前三频率	34	28	25	25	23	23	22	22	20	19	18	18	16	15
连续两个月前三频率	14	5	5	1	3	4	7	6	4	6	5	1	1	2
	农林牧渔	钢铁	电子	建筑装饰	电气设备	建筑材料	汽车	机械设备	商业贸易	纺织服装	交通运输	公用事业	化工	轻工制造
单月前三频率	14	14	13	13	13	11	11	8	7	6	6	4	3	1
连续两个月前三频率	4	2	2	3	2	2	1	0	1	2	1	0	0	0

数据来源：广发证券发展研究中心

行业板块业绩分化严重：

分月度看，从05年到16年，单月涨幅前三的行业中，银行、非银、食品饮料频率最多，而化工、轻工制造、公用事业最少，连续两个月涨幅前三的行业中，银行、有色金属、计算机频率最多，机械设备、化工、轻工制造最少，行业轮动切换快。

如何把握行业轮动成权益配置的关键！

一、研究背景

计算机技术的不断发展推动了互联网的迅速普及和发展，互联网上沉淀的数据规模已呈指数型速度增长，大数据随之产生。

传统的量化投资研究面临瓶颈，而互联网大数据由于具有数量大、类型繁多、价值密度大、时效性高的特征，为量化投资提供了新的数据来源。

纵观海内外，大数据相关的量化投资策略的研究近几年兴起，关于舆情大数据与金融市场涨跌之间的关系已然成为量化新的研究方向和研究领域。

我国A股市场是散户为主的市场，众多中小投资者对A股的关注和热情是股市上涨的重要推动力，而投资者对市场的热情往往体现在热门财经网站的新闻阅读量、股吧关注人数和搜索网站的搜索量上，对这些“网络热度”即时的分析和利用，可以从新的视角上了解投资者情绪，指导投资。

一、研究背景——舆情指数

趋势研究

以“银行” 为例，银行相关政策出台或新闻事件发生，引起人们广泛关注，舆情指数随即升高。

数据来源：广发证券发展研究中心

一、研究背景——舆情指数

舆情洞察

舆情洞察清晰显示了网络的舆情变动，并给出了对应热点新闻事件，反映了广大网民对银行业利率、贷款业务变动等的关注。

数据来源：广发证券发展研究中心

01

03

04

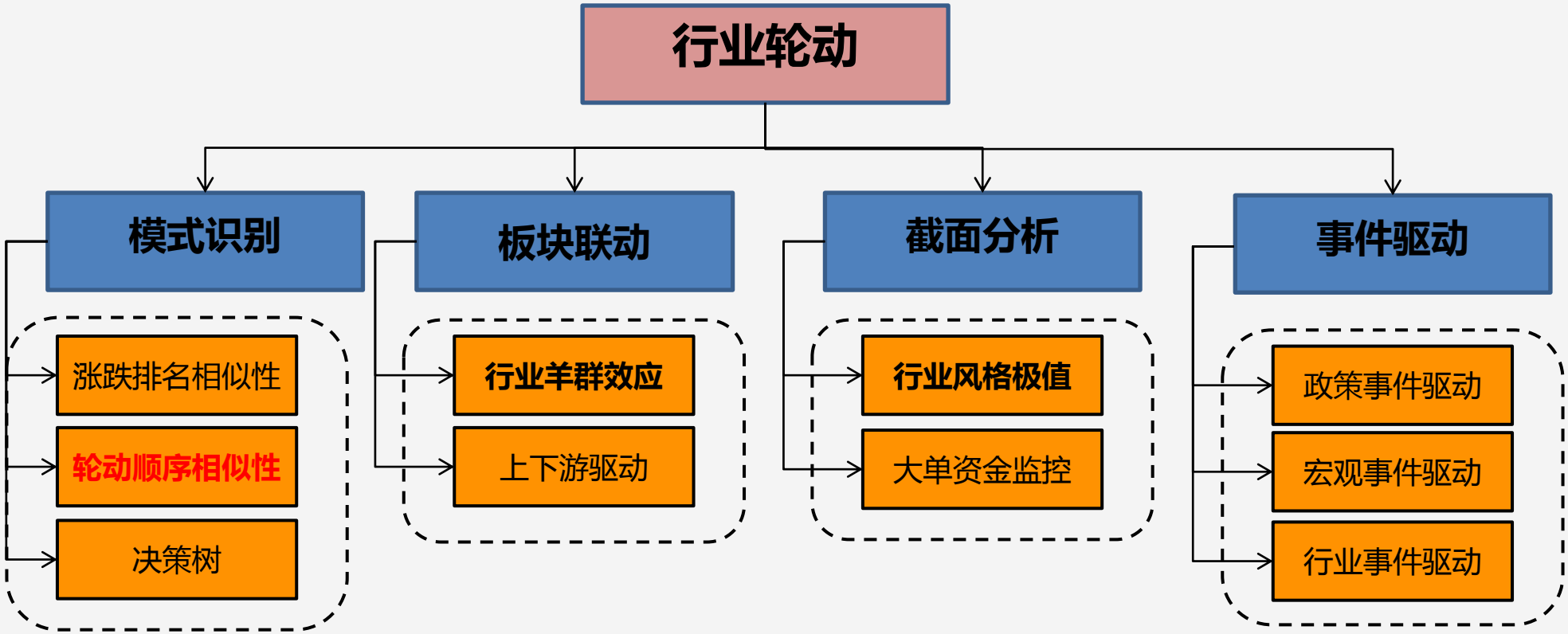
05

02

|研究现状|

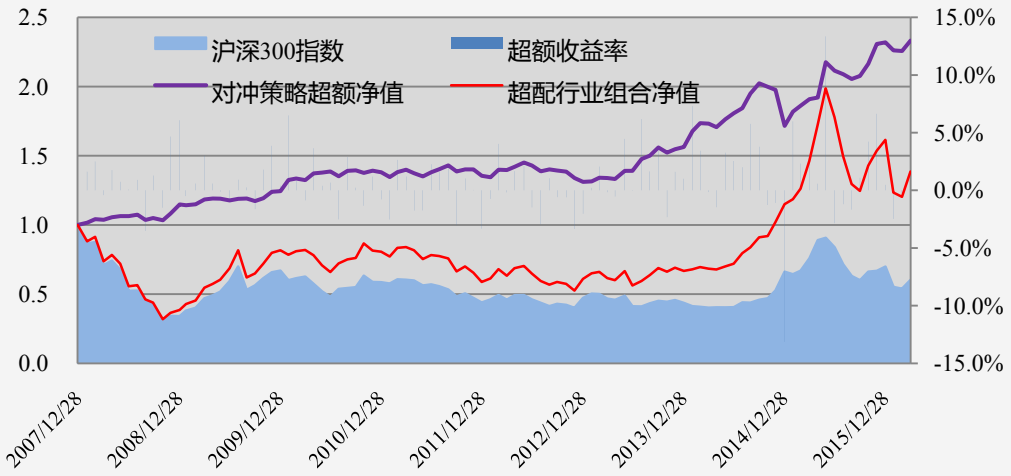
>

二、行业轮动策略研究现状



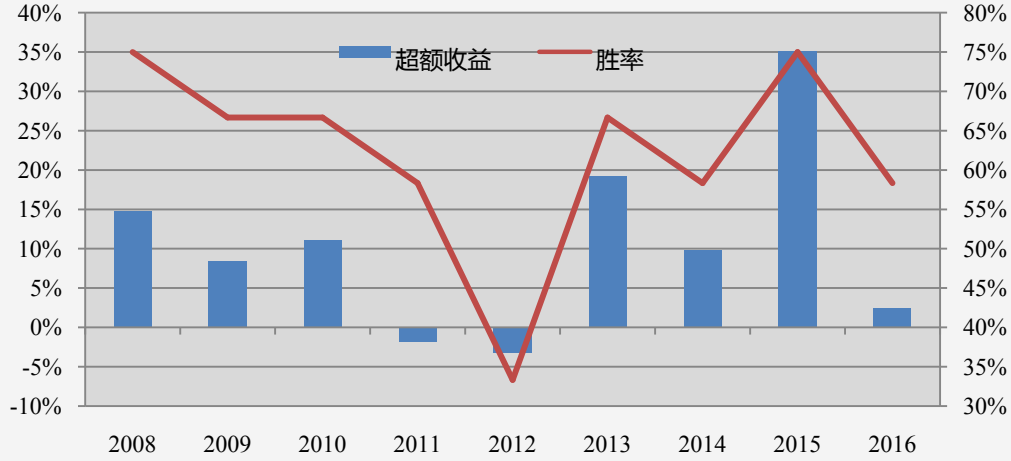
二、行业轮动策略研究现状——相似性匹配行业轮动策略

相似性匹配行业轮动策略历史回测结果



数据来源：广发证券发展研究中心

相似性匹配行业轮动策略年度超额收益



数据来源：广发证券发展研究中心

策略原理：

观察行业启动序列，与历史样本相匹配

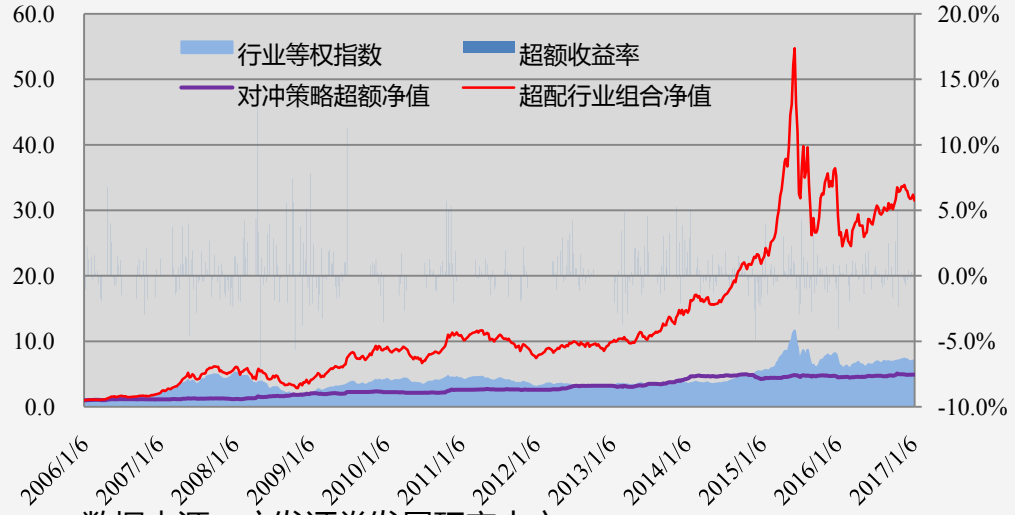
相似性匹配行业轮动策略分年度表现

时间	超额收益率	胜率	最大回撤
全样本	137.63%	62.04%	15.18%
2008	14.75%	75.00%	3.75%
2009	8.39%	66.67%	1.63%
2010	11.02%	66.67%	2.52%
2011	-1.88%	58.33%	5.24%
2012	-3.23%	33.33%	9.59%
2013	19.19%	66.67%	2.32%
2014	9.81%	58.33%	15.18%
2015	35.11%	75.00%	5.62%
2016（截止12.31）	2.50%	58.33%	2.69%

数据来源：广发证券发展研究中心

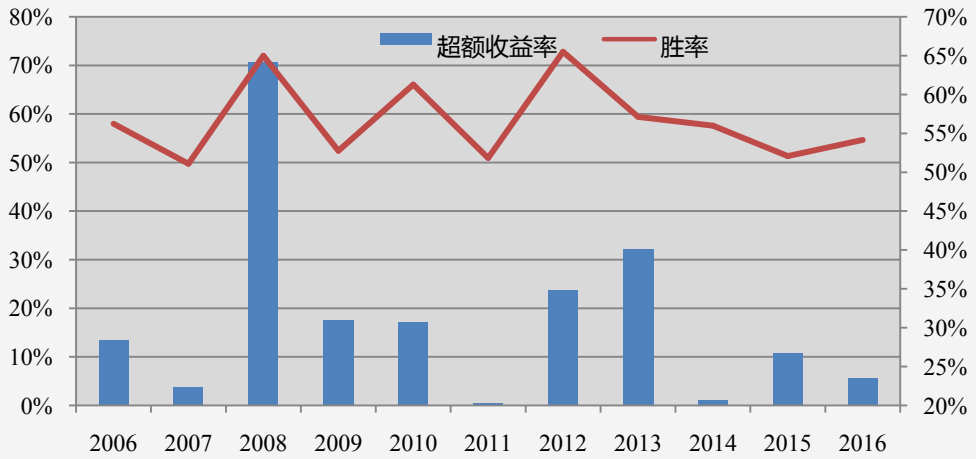
二、行业轮动策略研究现状——羊群效应行业轮动策略

羊群效应行业轮动策略历史回测结果



数据来源：广发证券发展研究中心

羊群效应行业轮动策略年度超额收益



数据来源：广发证券发展研究中心

策略思想：

根据行业在过去一周是否存在羊群效应，并且根据龙头股数量是否适宜来选择行业进行轮动。

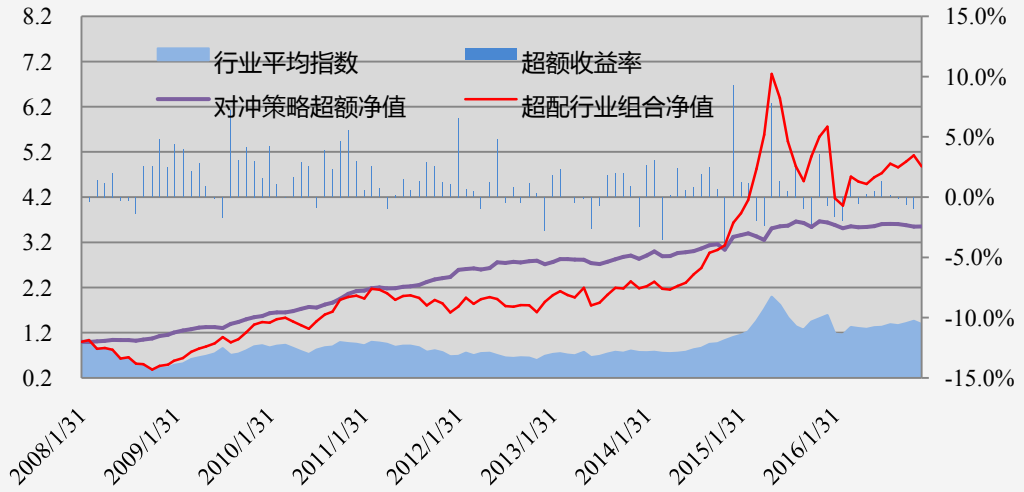
羊群效应行业行业策略分年度表现

时间	超额收益率	胜率	最大回撤
全样本	392.00%	56.30%	14.70%
2006	13.40%	56.30%	4.50%
2007	3.90%	51.10%	10.80%
2008	70.80%	65.00%	7.60%
2009	17.70%	52.80%	6.30%
2010	17.30%	61.30%	4.10%
2011	0.40%	51.90%	6.00%
2012	23.80%	65.50%	1.90%
2013	32.20%	57.10%	5.50%
2014	1.10%	56.00%	14.50%
2015	10.80%	52.10%	6.00%
2016(截止12.31)	5.7%	54.2%	4.3%

数据来源：广发证券发展研究中心

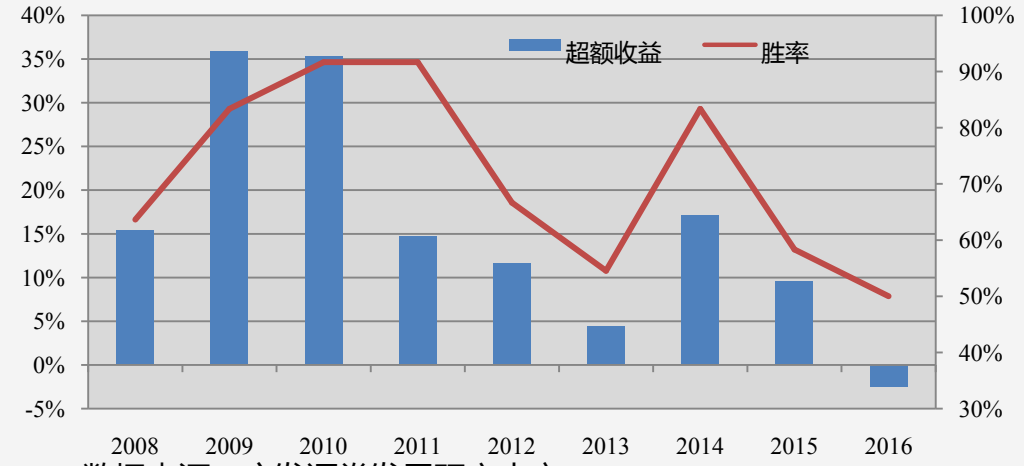
二、行业轮动策略研究现状——因子极值行业轮动策略

因子极值行业轮动策略历史回测结果



数据来源：广发证券发展研究中心

因子极值行业轮动策略年度超额收益



数据来源：广发证券发展研究中心

策略思想：

根据行业内个股的各因子是否达到创新高或创新低，计算各行业极值比例，来挖掘行业板块机会。

因子极值行业轮动策略分年度表现

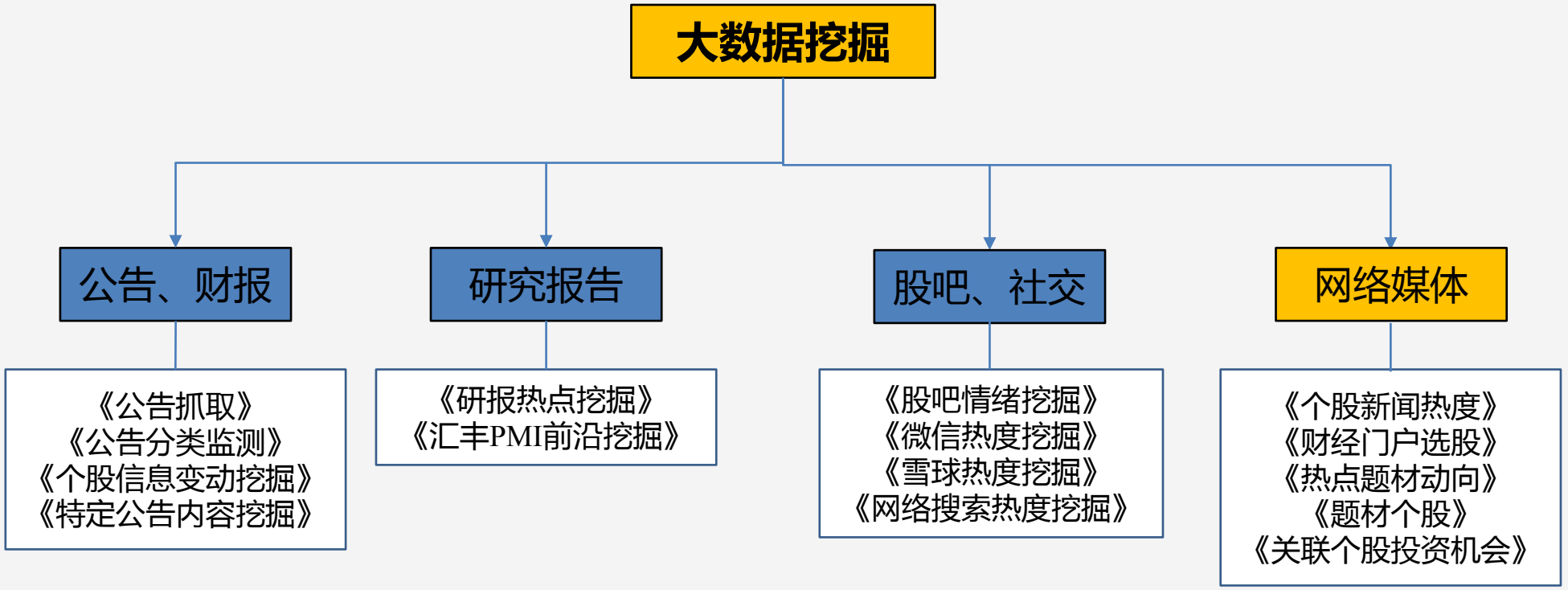
时间	超额收益率	胜率	最大回撤
全样本	255.11%	71.70%	4.26%
2008	15.40%	63.64%	1.91%
2009	35.85%	83.33%	1.81%
2010	35.31%	91.67%	0.89%
2011	14.74%	91.67%	0.95%
2012	11.61%	66.67%	2.77%
2013	4.41%	54.55%	3.83%
2014	17.06%	83.33%	3.76%
2015	9.61%	58.33%	4.26%
2016(截止12.31)	-2.43%	50.00%	3.49%

数据来源：广发证券发展研究中心

总结

当前行业轮动的主要方法是从因子价量数据及指数的价量数据出发寻找可能的行业轮动规律；在当前大数据时代下，行业的舆情数据提供了新的研究视角，这些舆情数据能够更为及时地反应当前市场上投资者情绪，对投资者的投资决策起到潜移默化的作用，我们可以从舆情数据出发来研究行业轮动的规律。

二、大数据量化研究现状


大数据挖掘
公告、财报
研究报告
股吧、社交
网络媒体
《公告抓取》
《公告分类监测》
《个股信息变动挖掘》
《特定公告内容挖掘》
《研报热点挖掘》
《汇丰PMI前沿挖掘》
《股吧情绪挖掘》
《微信热度挖掘》
《雪球热度挖掘》
《网络搜索热度挖掘》
《个股新闻热度》
《财经门户选股》
《热点题材动向》
《题材个股》
《关联个股投资机会》

二、大数据量化研究现状

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）》

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、舆情指数搜索工具；

01

02

04

05

03

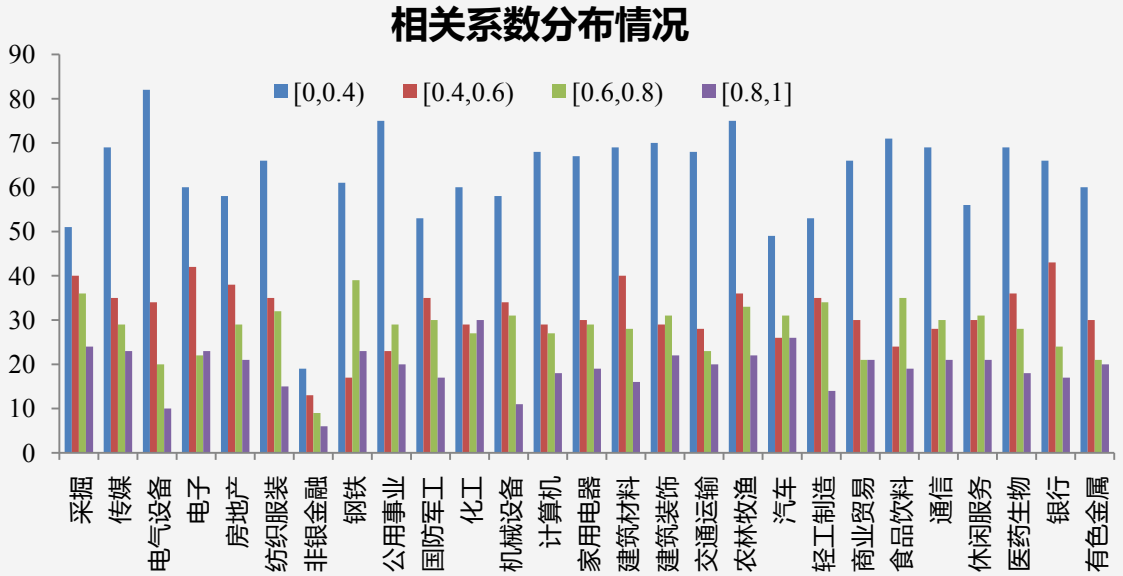
|行业轮动策略构建|

>

三、行业轮动策略构建——策略介绍

策略原理

我们发现各行业的滞后5阶舆情数据和行情数据有很强的正相关关系，当投资者对于某个行业的关注度急剧上升时，说明该行业可能是近期投资热点。

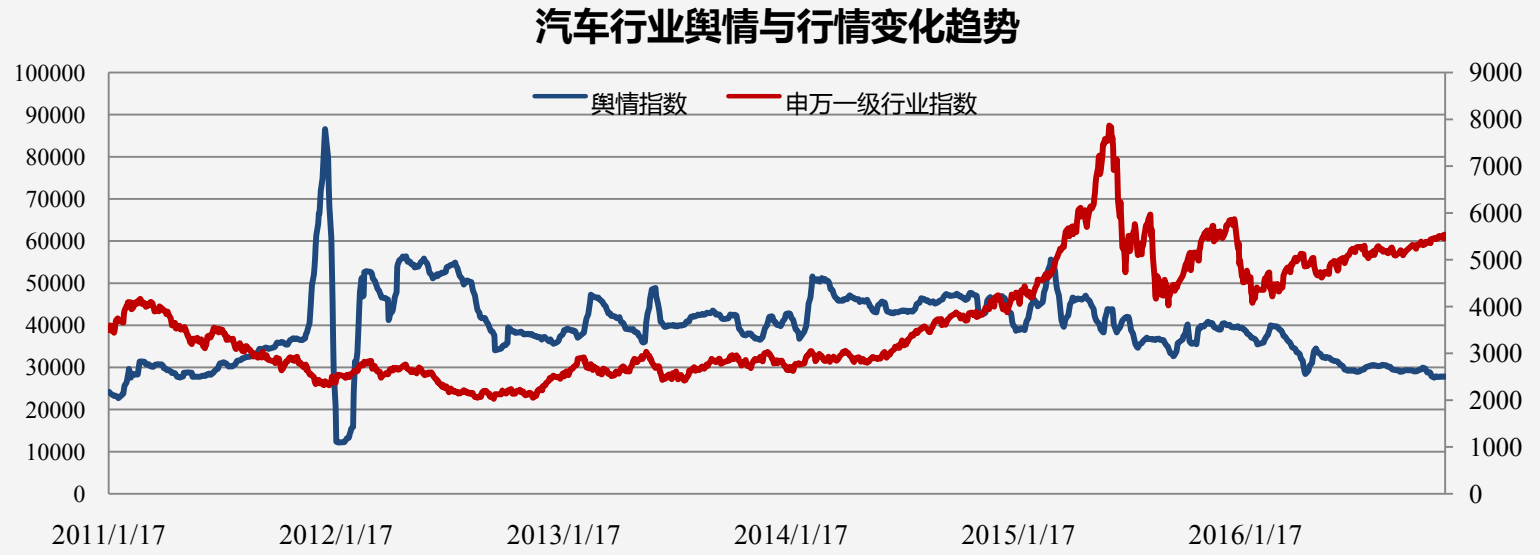


数据来源：广发证券发展研究中心

行业	采掘	传媒	电气设备	电子	房地产	纺织服装	非银金融	钢铁	公用事业	国防军工	化工	机械设备	计算机	家用电器
正相关个数	151	156	146	147	146	148	47	140	147	135	146	134	142	145
正相关均值	0.51	0.46	0.38	0.46	0.47	0.45	0.45	0.48	0.44	0.48	0.49	0.45	0.44	0.46
行业	建筑材料	建筑装饰	交通运输	农林牧渔	汽车	轻工制造	商业贸易	食品饮料	通信	休闲服务	医药生物	银行	有色金属	
正相关个数	153	152	139	166	132	136	138	149	148	138	151	150	131	
正相关均值	0.46	0.46	0.44	0.46	0.51	0.47	0.45	0.46	0.46	0.5	0.44	0.45	0.46	

数据来源：广发证券发展研究中心

三、行业轮动策略构建——策略介绍



数据来源：广发证券发展研究中心

舆情信息的变化要先于申万一级行业指数的变化，可以通过各行业舆情数据和行情数据的关系，来判断行业指数是否在接下来一段时间内出现上涨。

策略原理

行业的舆情趋势与行情趋势有较强的正相关关系，且舆情变化在短期内会先于行情变化，当某个行业指数的舆情涨幅较大，同时对应市场行情还未上涨或者涨幅很小，接下来市场行情上涨的可能性较大。可以对涨幅设置一个阈值，当同时满足高于或低于某个阈值时发出看多看空信号。

设置参数：舆情涨幅阈值 A；

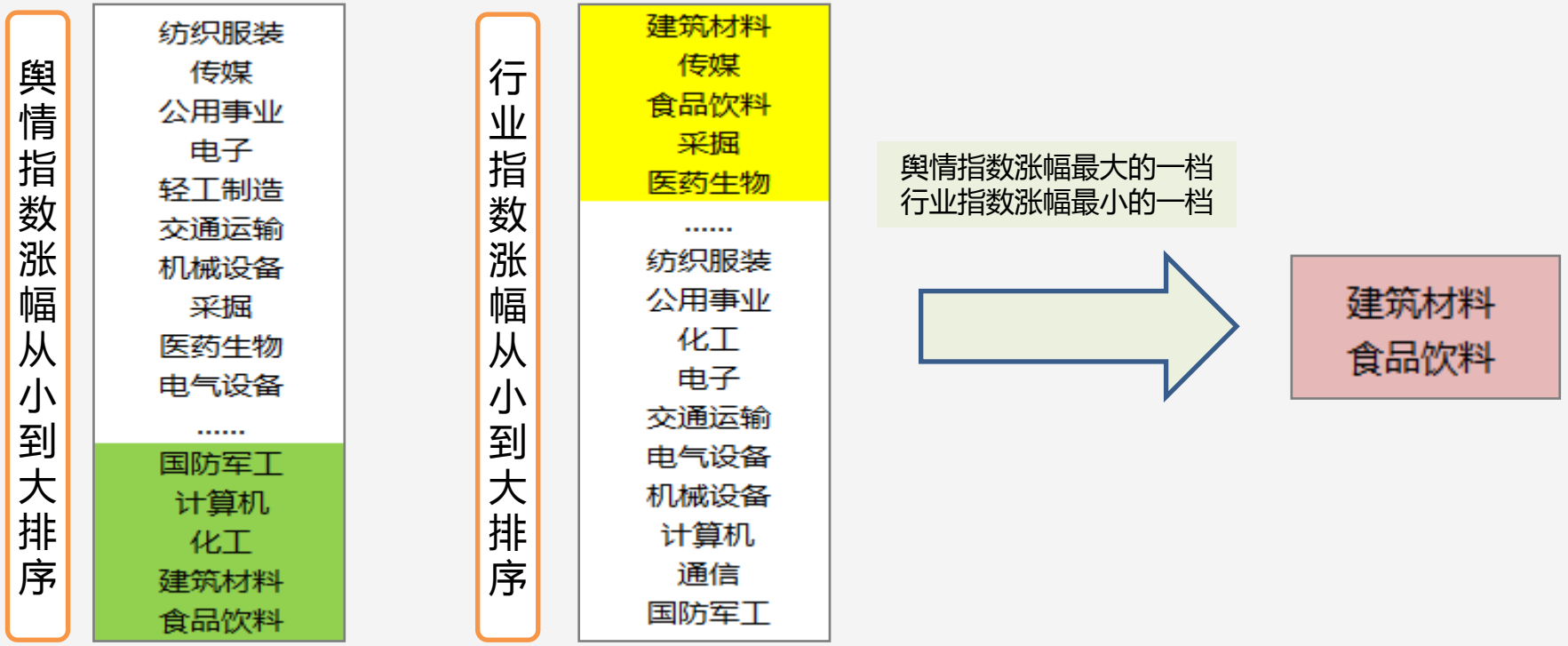
申万一级行业指数涨幅阈值 B；

申万一级行业指数涨幅阈值 C。

择时策略：根据过去一周舆情以及对应的行业指数的涨跌幅，当某个行业满足舆情涨幅大于阈值A，对应的行业指数涨幅大于阈值B且小于阈值C时，发出买入信号，在本周的第一个交易日买入，固定持有一段时间后平仓。

策略原理

如果在一个周内选出的热点行业数目太多，对选出的行业做进一步筛选，将这些行业按舆情涨幅排序分成三档，同时按对应的行业指数涨幅分成三档，选择同时满足舆情涨幅处于最大的一档，对应行业指数涨幅处于最小的一档。如果筛选不出热点行业，放宽对行业指数涨幅的限制，直至筛选出热点行业，停止筛选过程。



以2014年11月17日—11月24日为例对策略原理进行具体说明：

	时间	传媒	电气设备	房地产	国防军工	化工	家用电器	建筑装饰	汽车	轻工制造	食品饮料	通信	医药生物
舆情指数	2014/11/14	937	268	2387	500	601	568	231	36366	196	147	560	740
	2014/11/15	953	246	2095	568	477	578	197	40311	165	148	483	642
	2014/11/16	865	219	2094	686	495	586	171	41202	180	131	424	631
	2014/11/17	1029	340	2494	582	628	676	256	40429	186	176	592	778
	2014/11/18	1079	298	2992	1042	592	635	288	43834	198	156	636	863
	2014/11/19	1115	353	3136	1200	676	701	269	44563	182	169	611	800
	2014/11/20	1047	288	3223	1335	625	728	246	47222	184	189	558	770
	2014/11/21	1119	292	3283	1127	593	611	255	47486	205	151	563	696
	2014/11/22	981	222	3648	1139	505	575	182	47925	186	140	482	656
	2014/11/23	1041	255	3474	1225	553	690	224	49411	201	153	485	722
	2014/11/24	1524	318	3729	1675	698	718	252	48383	226	201	591	886
申万一级行业指数	2014/11/14	954.79	4345.45	3159.53	1282.71	2199.22	3273.6	1910.91	3791.73	2120.6	4932.62	1884.03	5795.44
	2014/11/17	971.18	4406.39	3160.09	1298.87	2219.08	3265.9	1916.59	3842.36	2139.45	4947.26	1904.8	5857.67
	2014/11/18	978.46	4407.3	3150.98	1296.09	2225.56	3240.94	1915.09	3846.86	2147.83	4865.07	1906.44	5853.99
	2014/11/19	997.91	4410.28	3148.97	1298.87	2238.34	3246.35	1905.37	3834.75	2166.67	4840.12	1926.69	5900.03
	2014/11/20	994.79	4418.26	3143.76	1278.2	2240.27	3238.77	1907.37	3827.9	2169.66	4824.83	1918.97	5860.34
	2014/11/21	1012.41	4450.93	3188.7	1287.42	2258.52	3270.74	1935.43	3868.96	2184.77	4885.53	1931.33	5895.89
	2014/11/24	1027.75	4488.91	3348.02	1311.03	2279.23	3325.21	2011.81	3922.2	2201.02	4912.23	1948.75	5902.11

数据来源：广发证券发展研究中心

11月16日和11月23日是判断舆情指数涨跌幅的起止时间，11月14日和11月21日是判断行业指数涨跌幅的起止时间。根据舆情涨幅大于阈值A，对应行业指数涨幅大于阈值B且小于阈值C的选择标准，初步选出了传媒、电气设备等12个行业。

由于初步筛选出的行业数目较多（>4），对选出的行业做进一步筛选，将这些行业按舆情涨幅排序分成三档，同时按申万一级行业指数涨幅分成三档。

行业	传媒	电气设备	房地产	国防军工	化工	家用电器	建筑装饰	汽车	轻工制造	食品饮料	通信	医药生物
舆情指数	0.20	0.16	0.66	0.79	0.12	0.18	0.31	0.20	0.12	0.17	0.14	0.14
申万一级行业指数	0.06	0.02	0.01	0.00	0.03	0.00	0.01	0.02	0.03	-0.01	0.03	0.02

舆情指数涨幅排序

轻工制造	0.12
化工	0.12
通信	0.14
医药生物	0.14
电气设备	0.16
食品饮料	0.17
家用电器	0.18
汽车	0.20
传媒	0.20
建筑装饰	0.31
房地产	0.66
国防军工	0.79

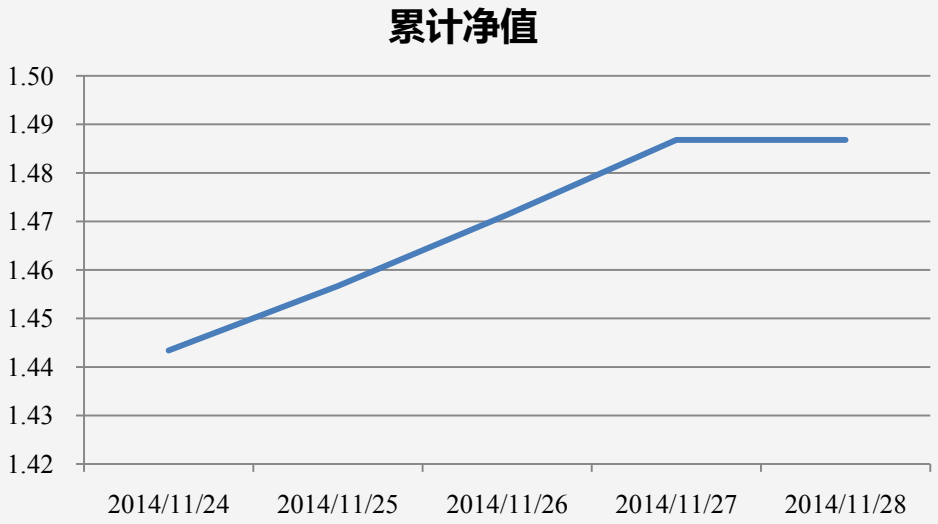
行业指数涨幅排序

食品饮料	-0.01
家用电器	0.00
国防军工	0.00
房地产	0.01
建筑装饰	0.01
医药生物	0.02
汽车	0.02
电气设备	0.02
通信	0.03
化工	0.03
轻工制造	0.03
传媒	0.06

房地产
国防军工

考虑选择满足舆情涨幅处于最大的一档，行业指数涨幅处于最小的一档。成功筛选出热点行业房地产和国防军工，筛选结束。

选择出房地产、国防军工两个行业后，在下一交易日11月24日等权买入，持仓一段时间后平仓。此时等待下一买入信号发出，在买入信号发出前持有资产。本次交易细节如下：



数据来源：广发证券发展研究中心

时间	组合收益率	房地产			国防军工		
		买价	卖价	收益率	买价	卖价	收益率
2014/11/24	1.13%	3307.51	3348.02	1.22%	1297.58	1311.03	1.04%
2014/11/25	2.06%	3307.51	3372.81	1.97%	1297.58	1325.45	2.15%
2014/11/26	3.09%	3307.51	3385.98	2.37%	1297.58	1346.96	3.81%
2014/11/27	4.17%	3307.51	3414.43	3.23%	1297.58	1363.86	5.11%
2014/11/28	0.00%	--	--	--	--	--	--

01

02

03

05

04

|实证分析|

>

回测说明

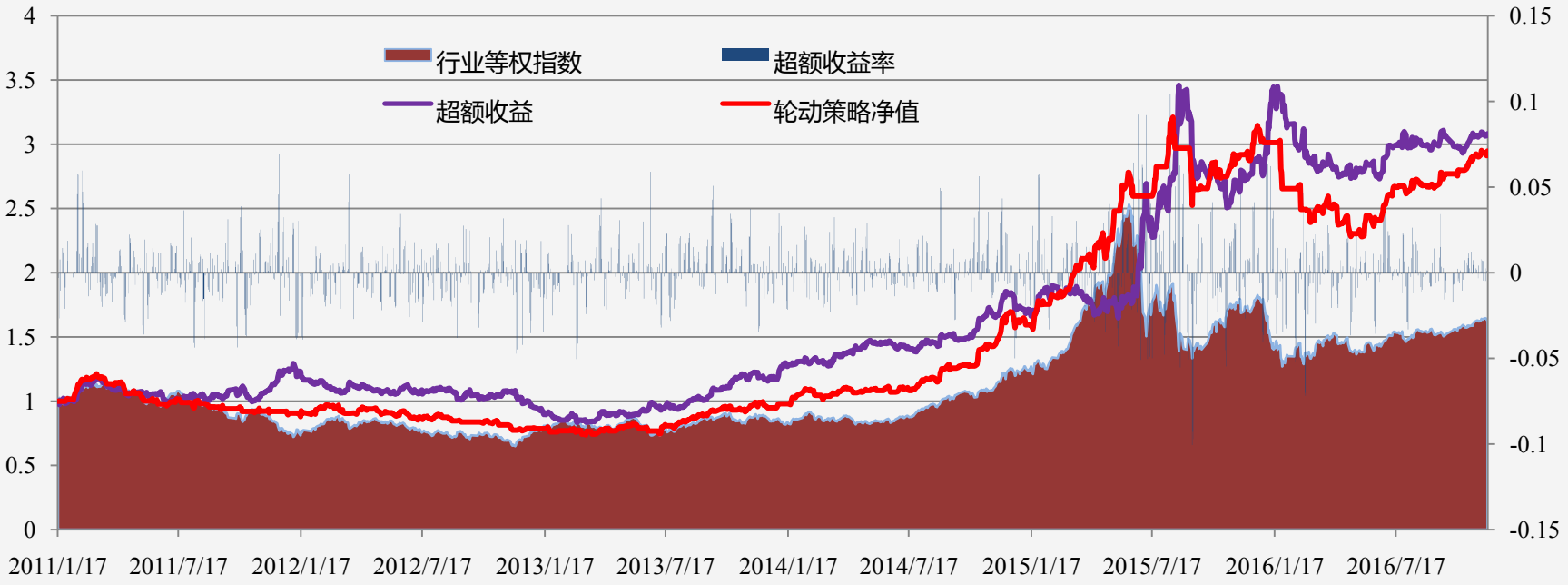
数据： 行业指数历史数据 ：申万一级行业指数开盘价、收盘价;
舆情指数历史数据：对应行业的舆情指数;

实证区间 ：2011/01/04——2016/12/02

设置参数 ：舆情指数涨幅阈值 A ；
申万一级行业指数涨幅阈值 B ；
申万一级行业指数涨幅阈值 C。

四、行业轮动策略实证分析

回测结果



策略回测结果一览

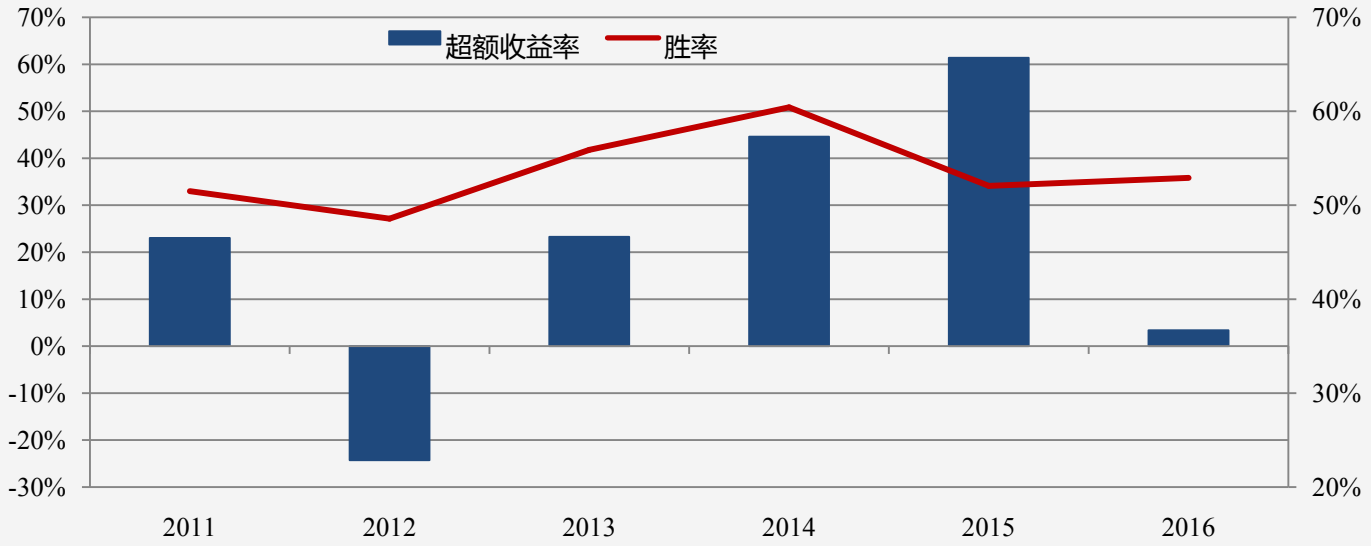
年化收益率	胜率	超额收益率
19.93%	53.57%	20.83%

数据来源：广发证券发展研究中心

四、行业轮动策略实证分析

分年度收益表现

时间	超额收益率	胜率	最大回撤
全样本	208.20%	53.57%	35.61%
2011	23.01%	51.49%	17.02%
2012	-24.30%	48.56%	26.46%
2013	23.25%	55.88%	16.66%
2014	44.60%	60.41%	7.46%
2015	61.39%	52.05%	27.55%
2016	3.35%	52.91%	20.82%



数据来源：广发证券发展研究中心

四、行业轮动策略实证分析

筛选结果

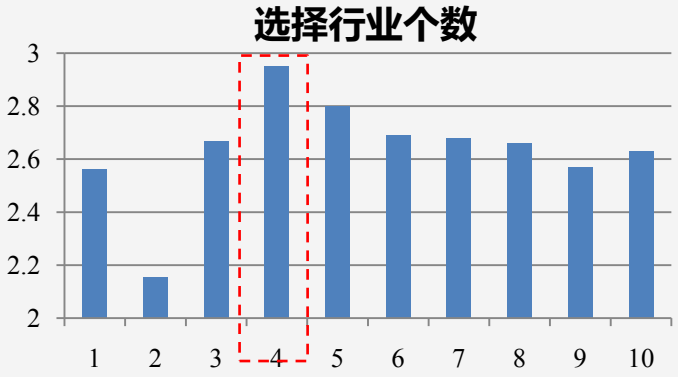
时间	持有行业指数			
2011/1/17				
……	……			
2016/6/13	农林牧渔	商业贸易	医药生物	
2016/6/20				
2016/6/27	轻工制造			
2016/7/4	非银金融	建筑材料		
2016/7/11	纺织服装	非银金融	建筑装饰	
2016/7/18	机械设备	交通运输		
2016/7/25				
2016/8/1	食品饮料			
2016/8/8	有色金属			
2016/8/15	非银金融	公用事业	家用电器	食品饮料
2016/8/22	非银金融	机械设备	建筑材料	食品饮料
2016/8/29	纺织服装			
2016/9/5	房地产			
2016/9/12	纺织服装	计算机	建筑材料	农林牧渔
2016/9/19	房地产			
2016/9/26	纺织服装	商业贸易		
2016/10/10				
2016/10/17	建筑材料	食品饮料		
2016/10/24				
2016/10/31	非银金融	休闲服务	有色金属	
2016/11/7	采掘			
2016/11/14	休闲服务	有色金属		
2016/11/21	化工	商业贸易		
2016/11/28	房地产	非银金融	家用电器	银行

数据来源：广发证券发展研究中心

四、行业轮动策略实证分析

参数敏感性测试

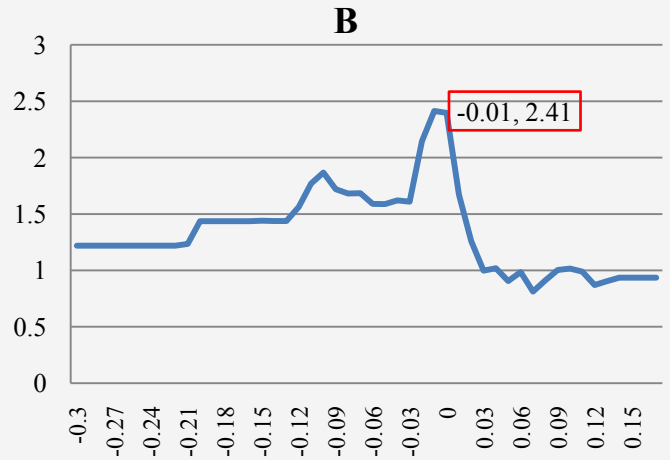
为观察轮动策略的表现对于参数的敏感性，固定其余两个参数，策略累计净值随另一参数的变化而变化。得到测试结果如下（图横轴为参数值，纵轴为累计净值）：

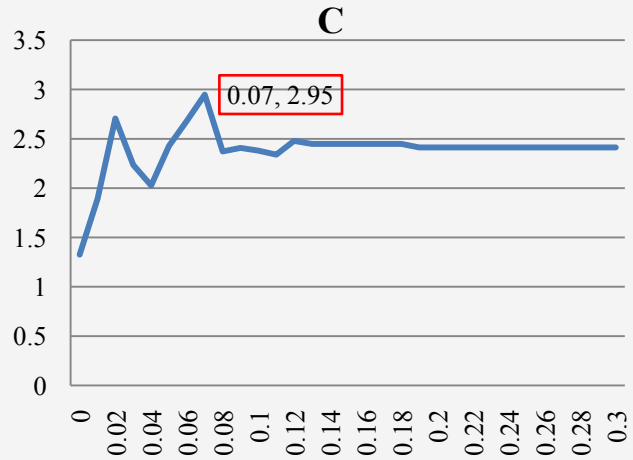


变动参数	固定参数	变化范围	公差	累计净值最大值	累计净值最小值	累计净值变化标准差
阈值A	B，C	0~1	0.01	2.95	0.72	0.47
阈值B	A，C	-0.3~0.17	0.01	2.41	0.81	0.38
阈值C	A，B	0~0.3	0.01	2.95	1.33	0.26

数据来源：广发证券发展研究中心

策略对舆情涨幅阈值A的变动最为敏感，行情涨幅阈值B、C次之。





策略改进

前面的行业轮动策略中根据相关性分析结果（各行业的滞后5阶舆情数据和行情数据有很强的正相关关系），以两者存在很强的正相关关系为前提。这种正相关关系是否在每个交易区间都存在？若不存在，是否会影响策略表现，这里我们对其进行讨论。

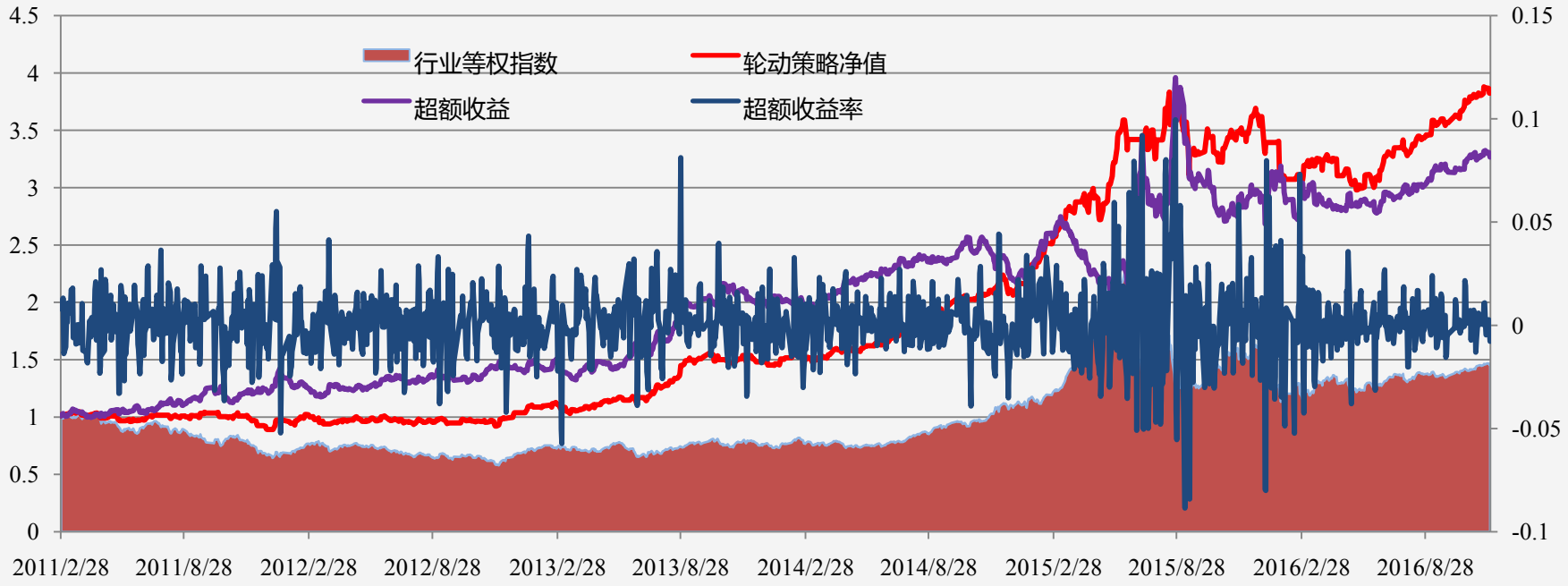
首先对舆情与行情的相关关系进行判断，原有策略的其他设定不变：

设置参数： 舆情指数涨幅阈值 A；
申万一级行业指数涨幅阈值 B；
申万一级行业指数涨幅阈值 C。

择时策略：根据过去一定时期内的舆情指数与对应行情指数的相关系数方向，估计未来一周两者的相关系数方向（正/负）；若为正相关，则根据过去一周舆情以及对应的行业指数的涨跌幅，当某个行业满足舆情涨幅大于阈值A，对应的行业指数涨幅大于阈值B且小于阈值C时，发出买入信号，在本周的第一个交易日买入，固定持有一段时间后平仓。

四、行业轮动策略实证分析

回测结果



策略回测结果一览

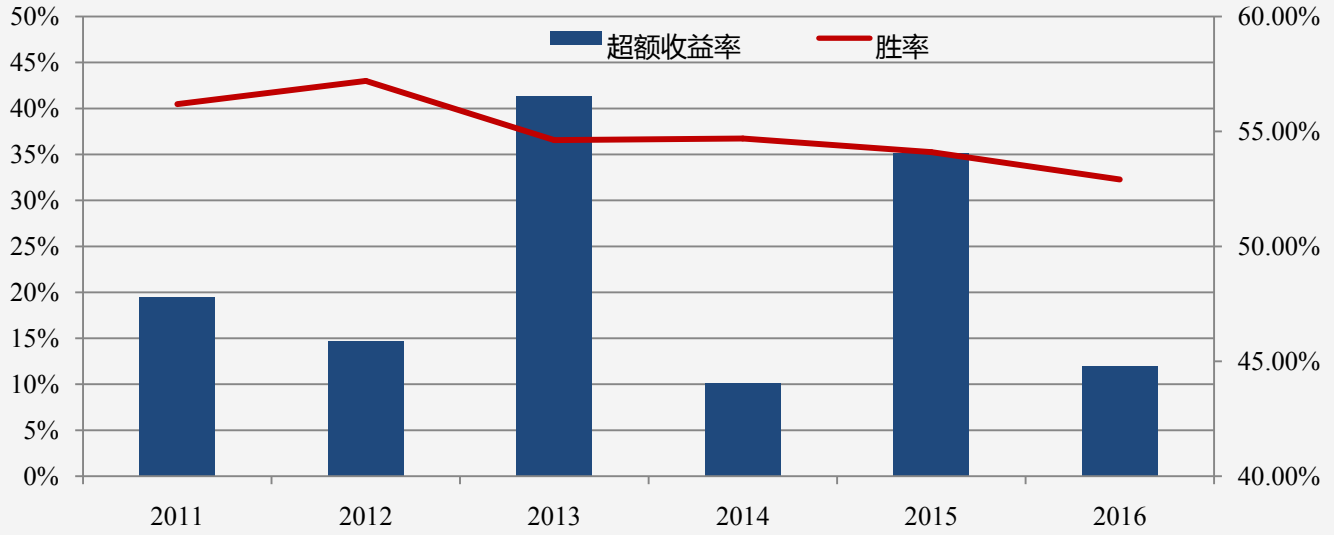
年化收益率	胜率	超额收益率
25.55%	54.95%	22.21%

数据来源：广发证券发展研究中心

四、行业轮动策略实证分析

分年度收益表现

时间	超额收益率	胜率	最大回撤
全样本	226.63%	54.95%	32.23%
2011	19.48%	56.19%	11.05%
2012	14.63%	57.20%	17.35%
2013	41.32%	54.62%	12.06%
2014	10.15%	54.69%	14.63%
2015	35.08%	54.10%	31.61%
2016	12.00%	52.91%	14.64%



数据来源：广发证券发展研究中心

01

02

03

04

05

|研究总结|

>

五、研究总结

基于舆情的行业轮动策略根据舆情变化和行情变化有较强的正相关关系，且舆情变化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、我们利用网络爬虫程序抓取了各行业的舆情指数历史数据，从Wind提取了行业指数的历史数据用于策略构建。

2、策略中我们对各行业的舆情和历史数据涨幅设置阈值，当舆情涨幅和历史数据涨幅突破阈值时，发出看多信号买入行业指数，固定持有一段时间后平仓。对阈值参数组合进行敏感性测试选择出了最优参数组合，同时发现策略对舆情指数涨幅阈值A变动最为敏感，行业指数涨幅阈值B、C次之。

未来的研究方向

- 以一周为时间区间判断舆情和指数行情是否出现上涨，选择出行业后持仓一段时间，导致策略中可能每周会有几天的空仓日期。未来可以改进策略，在空仓日期再做相同判断，选择性买入指数，减少空仓日期。

风险提示

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。

免责声明

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

地址: 广州市天河区林和西路9号耀中广场A座1401 P.C.510075 电话: 020-87576976 WWW.GF.COM.CN