### 金融工程



## 最好的金融文本挖掘系统实现开源——光大证券量化投资周报(2014-08-31)

金融工程报告

#### ◆内容提要:

# 我们说到做到!中文云系统现已在 github 实现开源。有志于互联网金融文本挖掘研究的同行,敬请垂询。

光大证券金融工程团队目前持续跟踪的量化投资策略主要包括以下5个:

#### 1、长线情绪择时策略

通过互联网金融文本挖掘算法,在网络大数据中提取投资者情绪表达相关内容, 经过情感识别构建的长线择时模型;上周普通投资者情绪指标大幅回落,下行 通道已经形成。据历史统计规律,建议投资者保持谨慎。

#### 2、短线期指择时策略

通过提取股指期货隔夜持仓数据,经统计处理构建短线择时模型,把握市场知情者动向:

#### 3、投资时钟行业配置策略

光大投资时钟模型是对经典投资时钟模型的细化和改进。在不同的时区各行业有着不同的表现。由月度宏观数据确定时区,参照历史表现给出看多的行业;根据光大投资时钟,目前所处经济时区为 1D 时区,看多的行业包括传媒,计算机,休闲服务,银行,有色金属。

#### 4、多因子时钟策略

该策略是针对多因子模型的因子选择问题的一种新方法。我们运用时区这一外生变量来选择因子。其中时区产生于光大投资时钟模型;根据光大投资时钟,目前所处经济时区为1D时区,选择的因子为.独家:关注度,规模:市值,动量:1个月收益率,动量:2个月收益率,预测:12个月EPS,预测:12个月EPS变动,权重分别为-1,-1,-1,1,1。

#### 5、新闻套利策略

通过互联网金融文本挖掘算法,每日对全部概念板块进行跟踪,全面揭示概念板块关注度走势,构建新闻套利模型。新闻套利模型包括看多概念及其龙头标的。其中龙头标的系通过文本相关性算法得出。

#### 各投资模型样本外结果:

模型名称	时间段	累计收益	年化收益	年化夏普	最大回撤
长线情绪择时	2010 年初至今	1.687	0.247	1.168	0.195
短线期指择时	2010年4月至今	1.092	0.193	0.737	0.211
投资时钟行业配置	2013 年初至今	0.275	0.238	1.076	0.137
多因子时钟	2013 年初至今	0.272	0.189	2.107	0.041
新闻套利	2014 年初至今	0.493	0.894	3.771	0.108

#### 分析师

冯 剑

(执业证书编号: S0930511090001)

021-22169176 fengj@ebscn.com

#### 联系人

袁 婳 021-22169311 yuanhua@ebscn.com

#### 相关研报

每日《大数据操盘手册》

每日《空穴来风概念操作手册》



## 目录

<b>—</b> 、	长线情绪择时模型3	}
	图表 1-1 长线情绪择时策略信号对应指数	3
	图表 1-2 长线情绪择时策略相对大盘表现(2010年至今)	3
	图表 1-3 长线情绪择时策略交易统计(2010年至今)	3
	图表 1-4 长线情绪择时策略近期交易	4
	图表 1-5 长线情绪择时策略相对大盘表现	4
	图表 1-6 长线情绪择时策略月度收益率表现	5
	图表 1-7 长线情绪择时策略年度表现	5
=,	短线期指择时模型6	3
	图表 2-1 短线期指择时策略相对大盘表现(2010年至今)	6
	图表 2-2 短线期指择时策略交易统计(2010年至今)	
	图表 2-3 短线期指择时策略近期交易	6
	图表 2-4 短线期指择时策略相对大盘表现	7
	图表 2-5 短线期指择时策略月度收益率表现	7
	图表 2-6 长线情绪择时策略年度表现	8
三、	投资时钟行业配置模型9	)
	1、行业配置对应 30 只股票	9
	图表 3-1 投资时钟策略相对大盘表现(2013年至今)	9
	图表 3-2 本期投资时钟股票组合收益曲线	
	图表 3-3 投资时钟本期量化投资组合(30 只)	
	图表 3-4 投资时钟行业配置策略相对大盘表现	. 10
	图表 2-5 投资时钟行业配置策略月度收益率表现	. 11
	图表 3-6 投资时钟行业配置策略年度表现	. 11
	2、行业配置对应行业指数	. 12
	图表 3-7 投资时钟策略(对应行业指数)相对大盘表现(2013年至今)	. 12
	图表 3-8 投资时钟行业配置策略(对应行业指数)相对大盘表现	. 12
	图表 3-9 投资时钟行业配置策略月度收益率表现(对应行业指数)	. 13
	图表 3-10 投资时钟行业配置策略年度表现(对应行业指数)	. 13
四、	多因子时钟模型14	1
	图表 4-1 因子策略相对大盘表现(样本外, 2013 年至今)	. 14
	图表 4-2 因子策略相对大盘表现	. 14
	图表 4-3 因子策略月度收益率表现	. 15
	图表 4-4 因子策略年度表现	. 15
五、	新闻套利模型16	;
	图表 5-1 新闻套利策略相对大盘表现(2014年至今)	. 16
	图表 5-2 新闻套利近期多头持仓	. 16
	图表 5-3 新闻套利策略相对大盘表现	. 17
	图表 5-4 新闻套利策略月度收益率表现	. 17



## 一、长线情绪择时模型

情绪表达判断,判断决定操作。从某种意义上理解,情绪本身就代表 着投资者对大盘走势的一种预期,它包含着投资者对各种影响因素的反应, 因此是一种较为综合的择时方法。

我们通过互联网金融文本挖掘算法,在网络大数据中提取投资者情绪 表达相关内容,经过情感识别构建长线情绪择时模型。模型的核心指标是 情绪曲线。如图 1-1 所示, 我们把情绪曲线画在上半部分图形中。情绪曲 线具有上轨和下轨,类似布林通道。我们根据情绪曲线与其上下轨的穿越 情况给出择时信号,上穿下轨为买入,下穿上轨为卖出。

上周普通投资者情绪指标大幅回落,下行通道已经形成。据历史统计 规律,建议投资者保持谨慎。

-20131101 -20131008

20131202 20140102

1.25 8. 0.75 32000.50

图表 1-1 长线情绪择时策略信号对应指数

2000

2010120

图表 1-2 长线情绪择时策略相对大盘表现(2010年至今)

2011040 20110503 20110601 20110701 20110801 20110901 20120201

20120301 20120405

20111201

评价指标	累计收益	年化收益	年化标准差	年化夏普	最大回撤
择时策略	1.687	0.247	0.185	1.168	0.195
指数表现	-0.324	-0.084	0.186	-0.574	0.406

20120702

20120502 20120601 20120801 20120903

-20121101 -20121008

20130104 20130301

20121203

20130502

20130401

20130603

20130701 20130801 20130902

图表 1-3 长线情绪择时策略交易统计(2010年至今)

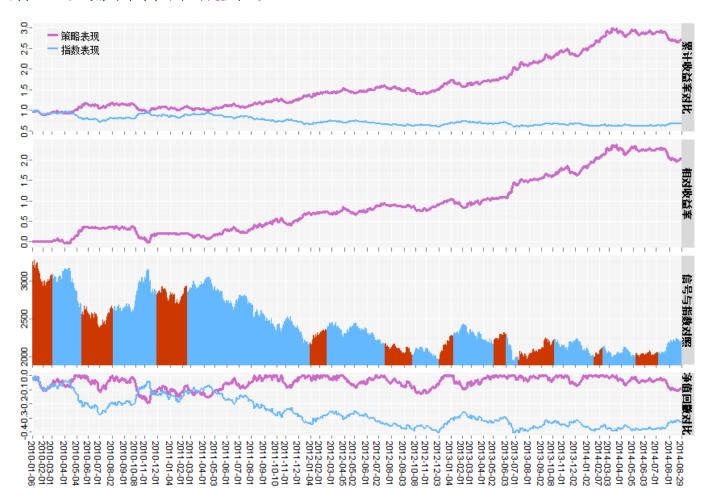
信号	发出次数	盈利次数	胜率	平均收益率	最低收益率	最高收益率	平均持仓天数	最低持仓天数	最高持仓天数	被动胜率	超额胜率
1	10	8	0.8	0.04	-0.07	0.14	38.4	15	63	0.5	0.3
-1	10	8	8.0	0.07	-0.08	0.23	74.4	21	213	0.5	0.3



## 图表 1-4 长线情绪择时策略近期交易

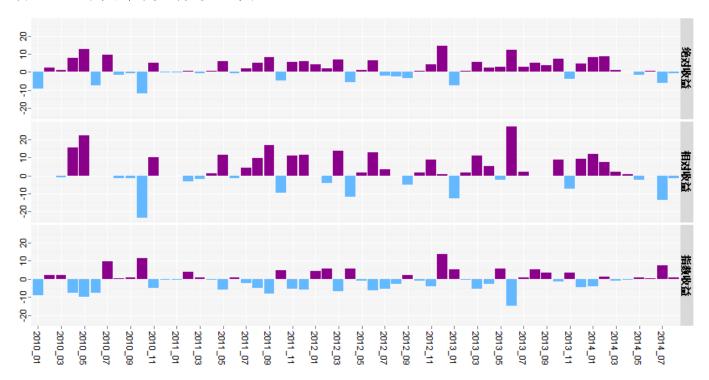
信号	建仓日	平仓日	建仓日指数	平仓日指数	收益率	盈亏	持仓交易日数
-1	2013-10-14	2014-01-20	2237.773	1991.253	0.11	盈	69
1	2014-01-20	2014-02-17	1991.253	2135.415	0.07	盈	15
-1	2014-02-17	2014-05-13	2135.415	2050.728	0.04	盈	58
1	2014-05-13	2014-07-06	2050.728	2059.375	0	盈	38
-1	2014-07-06	2014-08-29	2059.375	2217.2	-0.08	亏	40

图表 1-5 长线情绪择时策略相对大盘表现

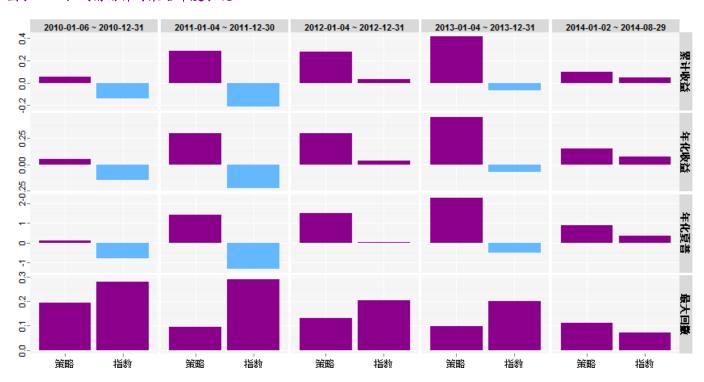




## 图表 1-6 长线情绪择时策略月度收益率表现



图表 1-7 长线情绪择时策略年度表现





## 二、短线期指择时模型

短线择时方面,我们采集股指期货信息,建立了期指持仓集中度择时模型。该模型通过分析股指期货隔夜持仓数据,把握市场知情者动向。在算法上,该模型选取多空 HHI 系数比和多空持仓比作为基础指标,运用统计手段消除套保持仓影响,挖掘股指期货隔夜持仓与次日大盘涨跌之间的相关关系,进而对大盘短线走势作出预测。除了短线交易者外,该模型还可以为调仓时机把握和套利交易的参数设置提供一定的参考。

### 图表 2-1 短线期指择时策略相对大盘表现(2010年至今)

评价指标	累计收益	年化收益	年化标准差	年化夏普	最大回撤
择时策略	1.092	0.193	0.222	0.737	0.211
指数表现	-0.248	-0.066	0.222	-0.400	0.414

#### 图表 2-2 短线期指择时策略交易统计(2010年至今)

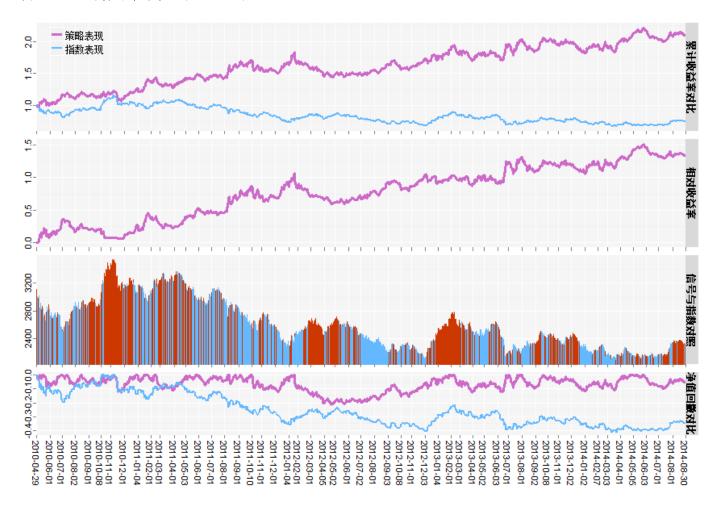
信号	发出次数	盈利次数	胜率	平均收益率	最低收益率	最高收益率	平均持仓天数	最低持仓天数	最高持仓天数	被动胜率	超额胜率
1	203	111	0.55	0	-0.08	0.08	2.36	1	42	0.5	0.05
-1	203	111	0.55	0	-0.05	0.11	2.82	1	46	0.5	0.04

#### 图表 2-3 短线期指择时策略近期交易

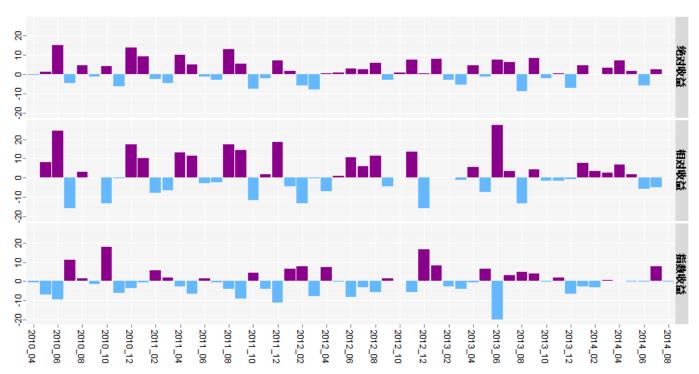
信号	建仓日	平仓日	建仓日指数	平仓日指数	收益率	盈亏	持仓交易日数
-1	2014-08-05	2014-08-06	2376.797	2358.206	0.01	盈	1
1	2014-08-06	2014-08-13	2358.206	2358.459	0	盈	5
-1	2014-08-13	2014-08-14	2358.459	2357.198	0	盈	1
1	2014-08-14	2014-08-29	2357.198	2317.304	-0.02	亏	11
-1	2014-08-29	2014-08-30	2317.304	2317.304	0	亏	1



#### 图表 2-4 短线期指择时策略相对大盘表现

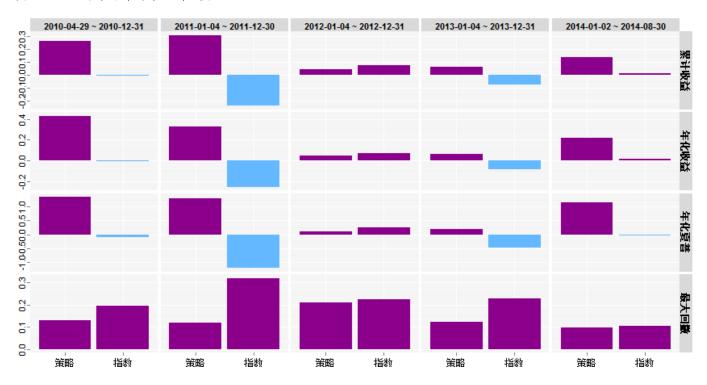


图表 2-5 短线期指择时策略月度收益率表现





## 图表 2-6 长线情绪择时策略年度表现





## 三、投资时钟行业配置模型

根据光大投资时钟,目前所处经济时区为 1D 时区,看多的行业包括传媒,计算机,休闲服务,银行,有色金属。

投资时钟行业配置模型是对经典投资时钟模型的细化和改进。前期研究表现,经典投资时钟模型仅通过考察通胀和产出的升降趋势来判断所处的经济周期阶段,并不能完整揭示中国市场的周期性特征。

在经典投资时钟模型的基础上,我们进一步考虑了通胀和产出相对位置,从而将经典投资时钟模型的时区加以细化,建立了更为适合中国情况的投资时钟行业配置模型。

模型的数据源和经典研究一致,用 CPI 同比增长率代表通胀状况,工业产出缺口(GAP)代表经济产出状况。在上述宏观数据披露后,我们会对时区相应作出调整。在不同的时区,我们参照历史表现给出看多的行业。

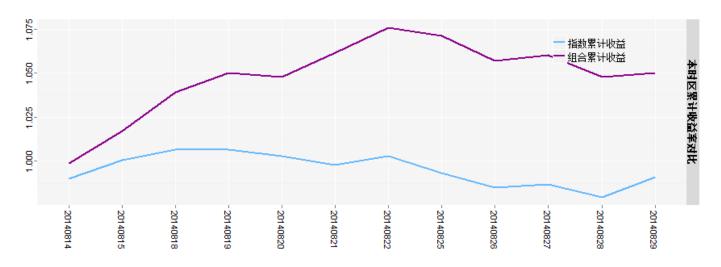
投资时钟行业配置模型表现分为两部分。在第一部分中,最终投资标的为看多行业中的 30 只股票,股票由多因子模型选出;第二部分,投资标的即为行业指数。

#### 1、行业配置对应30只股票

图表 3-1 投资时钟策略相对大盘表现 (2013年至今)

评价指标	累计收益	年化收益	年化标准差	年化夏普	最大回撤
投资时钟策略	0.464	0.398	0.187	1.940	0.124
指数表现	0.056	0.049	0.180	0.130	0.168

图表 3-2 本期投资时钟股票组合收益曲线





## 图表 3-3 投资时钟本期量化投资组合(30尺)

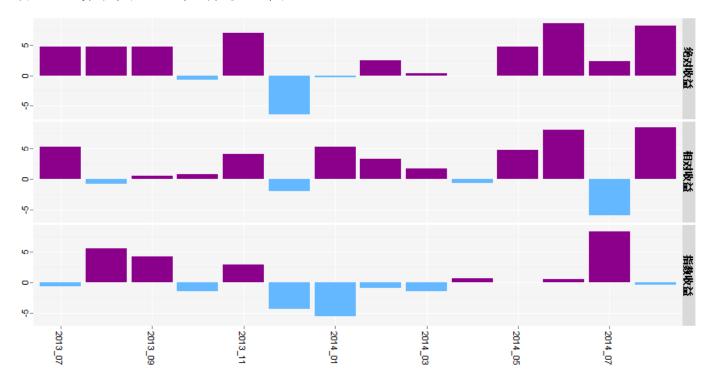
行业	股票代码	股票简称	累计收益率.	. 行业	股票代码	股票简称	累计收益率
传媒	000719.SZ	大地传媒	1.05	休闲服务	002159.SZ	三特索道	0.98
传媒	300059.SZ	东方财富	1.13	休闲服务	002558.SZ	世纪游轮	1.03
传媒	300148.SZ	天舟文化	1.12	银行	002142.SZ	宁波银行	0.98
传媒	300291.SZ	华录百纳	1.01	银行	600015.SH	华夏银行	0.98
传媒	300383.SZ	光环新网	1.02	银行	600016.SH	民生银行	0.98
传媒	601999.SH	出版传媒	1.26	银行	600036.SH	招商银行	0.98
计算机	002421.SZ	达实智能	1.08	银行	601169.SH	北京银行	1.03
计算机	300096.SZ	易联众	1.11	银行	601288.SH	农业银行	0.97
计算机	300167.SZ	迪威视讯	1.09	有色金属	000795.SZ	太原刚玉	1.05
计算机	300202.SZ	聚龙股份	1.03	有色金属	002295.SZ	精艺股份	1.06
计算机	300368.SZ	汇金股份	1.22	有色金属	300127.SZ	银河磁体	0.94
休闲服务	000524.SZ	东方宾馆	1.16	有色金属	600114.SH	东睦股份	0.96
休闲服务	000613.SZ	大东海 A	1.11	有色金属	600255.SH	鑫科材料	1.19
休闲服务	000888.SZ	峨眉山A	0.99	有色金属	600673.SH	东阳光科	0.86
休闲服务	002059.SZ	云南旅游	1.08	NA	NA	NA	NA

图表 3-4 投资时钟行业配置策略相对大盘表现





## 图表 2-5 投资时钟行业配置策略月度收益率表现



图表 3-6 投资时钟行业配置策略年度表现





## 2、行业配置对应行业指数

图表 3-7 投资时钟策略(对应行业指数)相对大盘表现(2013年至今)

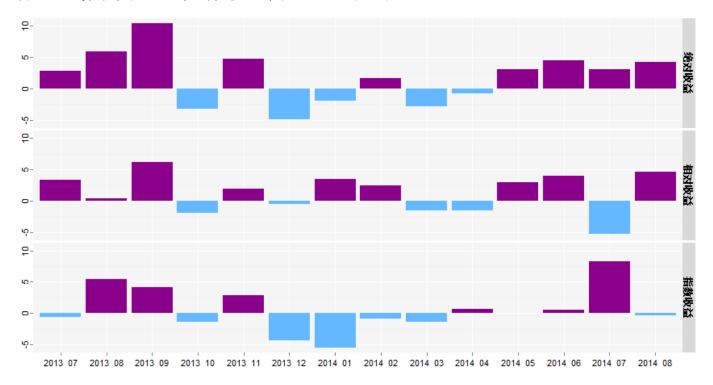
评价指标	累计收益	年化收益	年化标准差	年化夏普	最大回撤
投资时钟策略	0.275	0.238	0.192	1.076	0.137
指数表现	0.056	0.049	0.180	0.130	0.168

图表 3-8 投资时钟行业配置策略(对应行业指数)相对大盘表现





## 图表 3-9 投资时钟行业配置策略月度收益率表现 (对应行业指数)



图表 3-10 投资时钟行业配置策略年度表现 (对应行业指数)





## 四、多因子时钟模型

根据光大投资时钟,目前所处经济时区为 1D 时区,选择的因子为.独家:关注度,规模:市值,动量:1个月收益率,动量:2个月收益率,预测:12个月 EPS 变动,权重分别为-1,-1,-1,1,1。

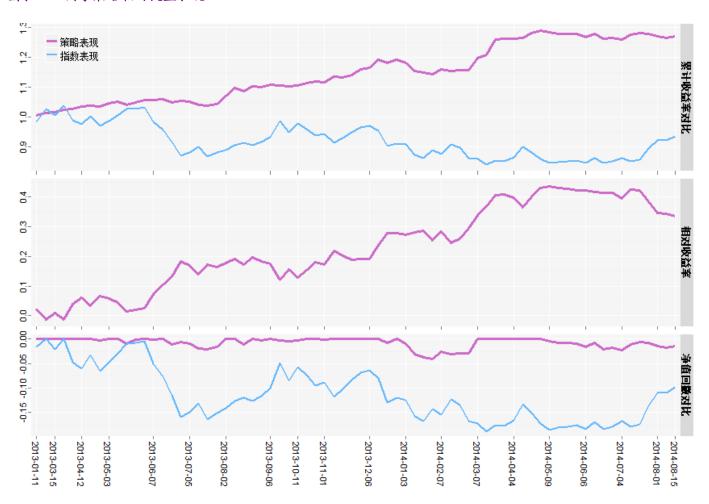
多因子时钟模型,是对多因子模型的因子选择问题的一种新方法。我 们运用时区这一外生变量来选择因子。其中时区产生于光大投资时钟模型。

在研究方法上, 我们根据 2003-2012 年的因子收益率数据, 通过综合 考虑 13 个因子在各时区的平均收益率、胜率、最大回撤、收益率分布等指标, 对时区的最佳因子进行了样本内的筛选。得出各时区对应的最佳因子选择之后, 我们以最佳因子等权的方式, 进行了样本外回测。表现如下:

图表 4-1 因子策略相对大盘表现(样本外, 2013年至今)

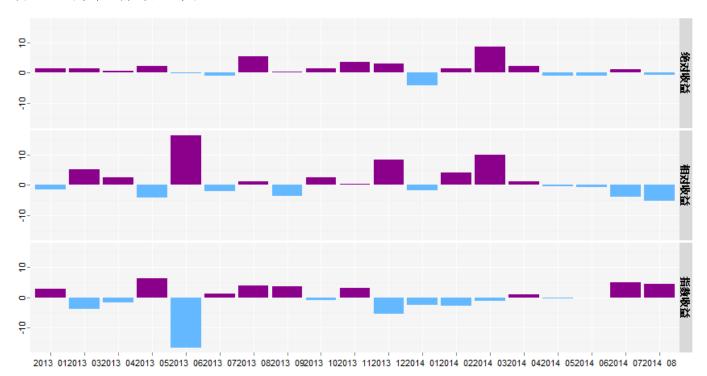
评价指标	累计收益	年化收益	年化标准差	年化夏普	最大回撤
因子策略	0.272	0.189	0.076	2.107	0.041
指数表现	-0.065	-0.047	0.182	-0.388	0.189

图表 4-2 因子策略相对大盘表现





## 图表 4-3 因子策略月度收益率表现



## 图表 4-4 因子策略年度表现





## 五、新闻套利模型

市场热点是每一位投资者都密切关注的内容,也是证券研究的永恒主题。我们通过互联网金融文本挖掘算法,每日对全部概念板块进行跟踪,全面揭示概念板块关注度走势。根据关注度走势的相对变动,我们构建了新闻套利模型。

新闻套利模型包括看多概念及其龙头标的。其中龙头标的是通过文本 相关性算法得出,对于概念和股票之间的真实联系我们无法做出保证,特 此声明。

### 图表 5-1 新闻套利策略相对大盘表现(2014年至今)

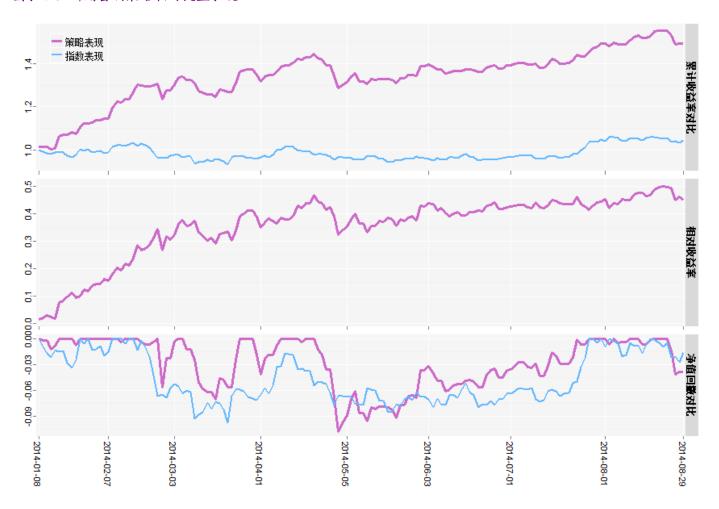
评价指标	累计收益	年化收益	年化标准差	年化夏普	最大回撤
新闻套利策略	0.493	0.894	0.225	3.771	0.108
指数表现	0.045	0.072	0.159	0.289	0.097

### 图表 5-2 新闻套利近期多头持仓

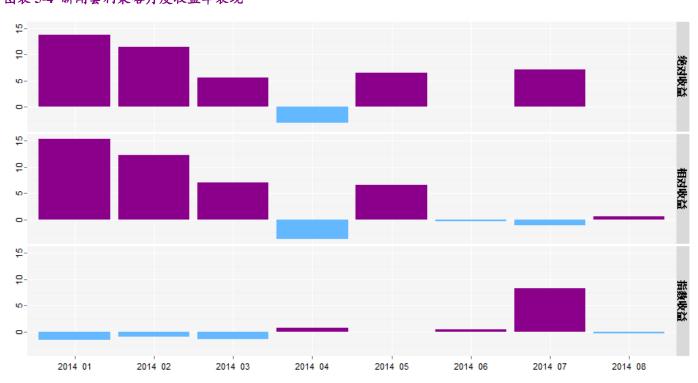
日期	概念名称	股票简称
20140824	土地流转	中联重科, 永安林业, 罗牛山, 新野纺织, 亚盛集团,*ST 大荒, 中华企业, 亚通股份, 东方集团, 海南橡胶
20140824	新三板	宏源证券,长江证券,中关村,云南盐化,东湖高新,苏州高新,辽宁成大,大智慧,东吴证券
20140824	在线教育	西安饮食,陕西金叶,科大讯飞,拓维信息,洪涛股份,达华智能,华平股份,时代出版,新南洋,大连控股
20140824	中蒙	葛洲坝, 哈高科, 锦州港, 亿利能源, 营口港, 杭萧钢构, 晋西车轴, 天津港, 内蒙华电, 中煤能源
20140830	充电桩	许继电气, 科陆电子, 奥特迅, 科士达, 比亚迪, 特锐德, 上海电力, 永泰能源, 动力源, 上海普天
20140830	迪士尼	华纺股份, 新世界, 强生控股, 陆家嘴, 中华企业, 锦江股份, 中路股份, 东方明珠, 申通地铁, 界龙实业
20140830	军工	航天科技,成飞集成,北化股份,中国卫星,洪都航空,抚顺特钢,航天晨光,航天通信,航天电子,中国重工
20140830	无人机	云南盐化, 山河智能, 成飞集成, 高德红外, 山东矿机, 机器人, 洪都航空, 贵航股份, 同济科技, 航天电子



## 图表 5-3 新闻套利策略相对大盘表现



图表 5-4 新闻套利策略月度收益率表现





#### 分析师声明

负责准备本报告以及撰写本报告的所有研究分析师或工作人员在此保证,本研究报告中关于任何发行商或证券所发表的观点均如实反映分析人员的个人观点。负责准备本报告的分析师获取报酬的评判因素包括研究的质量和准确性、客户的反馈、竞争性因素以及光大证券股份有限公司的整体收益。所有研究分析师或工作人员保证他们报酬的任何一部分不曾与,不与,也将不会与本报告中具体的推荐意见或观点有直接或间接的联系。

#### 分析师介绍

冯剑,光大证券研究所首席金融工程分析师,复旦大学经济学博士,8年经济金融研究经验,4年证券研究经验,主要带领团队从事择时策略和量化交易策略研究。

#### 行业及公司评级体系

买入一未来 6-12 个月的投资收益率领先市场基准指数 15%以上:

增持一未来 6-12 个月的投资收益率领先市场基准指数 5%至 15%;

中性一未来 6-12 个月的投资收益率与市场基准指数的变动幅度相差-5%至 5%;

减持一未来 6-12 个月的投资收益率落后市场基准指数 5%至 15%;

卖出一未来 6-12 个月的投资收益率落后市场基准指数 15%以上;

无评级—因无法获取必要的资料,或者公司面临无法预见结果的重大不确定性事件,或者其他原因,致使无法给出明确的投资评级。

市场基准指数为沪深300指数。

#### 分析、估值方法的局限性说明

本报告所包含的分析基于各种假设,不同假设可能导致分析结果出现重大不同。本报告采用的各种估值方法及模型均有其局限性,估值结果不保证所涉及证券能够在该价格交易。



### 特别声明

光大证券股份有限公司(以下简称"本公司")创建于1996年,系由中国光大(集团)总公司投资控股的全国性综合类股份制证券公司,是中国证监会批准的首批三家创新试点公司之一。公司经营业务许可证编号: z22831000。

公司经营范围:证券经纪;证券投资咨询;与证券交易、证券投资活动有关的财务顾问;证券承销与保荐;证券自营;为期货公司提供中间介绍业务;证券投资基金代销;融资融券业务;中国证监会批准的其他业务。此外,公司还通过全资或控股子公司开展资产管理、直接投资、期货、基金管理以及香港证券业务。

本证券研究报告由光大证券股份有限公司研究所(以下简称"光大证券研究所")编写,以合法获得的我们相信为可靠、准确、完整的信息为基础,但不保证我们所获得的原始信息以及报告所载信息之准确性和完整性。光大证券研究所可能将不时补充、修订或更新有关信息,但不保证及时发布该等更新。

本报告根据中华人民共和国法律在中华人民共和国境内分发、仅供本公司的客户使用。

本报告中的资料、意见、预测均反映报告初次发布时光大证券研究所的判断,可能需随时进行调整。报告中的信息或所表达的意见不构成任何投资、法律、会计或税务方面的最终操作建议,本公司不就任何人依据报告中的内容而最终操作建议作出任何形式的保证和承诺。

在法律允许的情况下,本公司及其附属机构可能持有报告中提及的公司所发行证券的头寸并进行交易,也可能为这些公司提供或正在争取提供投资银行、财务顾问或金融产品等相关服务。投资者应当充分考虑本公司及本公司附属机构就报告内容可能存在的利益冲突、不应视本报告为作出投资决策的唯一参考因素。

在任何情况下,本报告中的信息或所表达的建议并不构成对任何投资人的投资建议,本公司及其附属机构(包括光大证券研究所)不对投资者买卖有关公司股份而产生的盈亏承担责任。

本公司的销售人员、交易人员和其他专业人员可能会向客户提供与本报告中观点不同的口头或书面评论或交易策略。本公司的资产管理部和投资业务部可能会作出与本报告的推荐不相一致的投资决策。本公司提醒投资者注意并理解投资证券及投资产品存在的风险,在作出投资决策前,建议投资者务必向专业人士咨询并谨慎抉择。

本报告的版权仅归本公司所有,任何机构和个人未经书面许可不得以任何形式翻版、复制、刊登、发表、篡改或者引用。

## 光大证券股份有限公司研究所 销售交易部 机构业务部

上海市新闸路 1508 号静安国际广场 3 楼 邮编 200040 总机: 021-22169999 传真: 021-22169114、22169134

销售交易部 姓名 办公电话 手机. 电子邮件 010-58452028 haohui@ebscn.com 郝耀 北京 010-58452027 黄怡 huangyi@ebscn.com liangchen@ebscn.com 010-58452029 刘公直 liugongzhi@ebscn.com 021-22169086 上海 yanfei@ebscn.com 021-22169087 周薇薇 zhouww1@ebscn.com 021-22169082 徐又丰 xuyf@ebscn.com 021-22169131 李强 liqiang88@ebscn.com 021-22169091 奚亦扬 xiyy@ebscn.com 021-22169083 zhanggong@ebscn.com 021-22169146 罗德锦 luodj@ebscn.com 深圳 黎晓宇 0755-83553559 lixy1@ebscn.com 0755-83553249 huanglh@ebscn.com 黄鹂华 0755-83559378 lixiao1@ebscn.com 张亦潇 0755-23996409 zhangyx@ebscn.com **平渊绕** wangyuanfeng@ebscn.com 机构客户业务部 姓名 办公电话 手机 电子邮件 濮维娜(副总经理) 021-62152373 13611990668 puwn@ebscn.com 张辉 上海 021-22167108 13611990668 zhanghui1@ebscn.com 021-22167101 计爽 18621181721 jishuang@ebscn.com 021-22169129 18918212345 jizhe@ebscn.com 北京 010-59046212 18611386181 zhulin1@ebscn.com 朱林 徐放 010-56513051 18618469955 xufang@ebscn.com 国际业务 陶奕(副总经理) 021-62152393 18018609199 taoyi@ebscn.com

021-22169152

021-22167094

13585893550

18352760578

qidw@ebscn.com

gush@ebscn.com

戚德文

顾胜寒