金融工程



金融文本挖掘前述----系统、数据、指标

金融工程

目前为止,我们所进行策略开发利用的数据基本上都是那些以数字形式存放在标准数据库中的结构化数据,而这部分数据仅仅占了所有金融信息中的一小部分,而金融信息中绝大部分的数据均是以文本的形式存在的一种非结构化的数据。

众所周知,财经新闻、股票论坛、股票研究报告等媒介信息能够迅速地反应 投资者的投资情绪,同时也会对投资者的投资策略起到潜移默化的影响。这 些文本信息中蕴含了大量未被利用的信息,如果能够合理地利用这些信息, 我们相信这将为我们理解市场提供一个全新的角度。

基于光大金工小组在金融文本挖掘中多年的研究与探索,我们将文本挖掘与金融量化完美的联系起来,为金融量化研究开拓了一个全新的研究方向。同时依托于光大金工强大的中文处理平台和丰富准确的历史数据积累,我们开发了了众多基于文本挖掘的指标和因子产品。同时,基于这些数据,我们开发的光大普通投资者情绪指标择时模型、关注度选股模型、概念板块事件套利模型等均取得了优秀的表现。 这也进一步过来说明了文本挖掘技术确实能够成为金融量化的一个有力工具。

在该篇报告中,我们将首先介绍光大中文云文本挖掘系统的整体架构,然后我们介绍基于该中文云系统所开发的四个数据引擎:光大关键词指数、光大情感指数、光大关键词网络图、光大关注度因子。在各个数据引擎中我们也将分别详细的介绍基于该数据引擎所开发的一些量化指标,如光大宏观系列指标、光大概念板块热度指标、光大情感指数系列、关注度因子、概念板块龙头股等的具体实现与编写细节。

分析师

倪蕴韬 (执业证书编号: S0930512070002) 021-22169338 niyt@ebscn.com

张斯会 (执业证书编号: S0930512020001) 021-22169104 zhangsihui@ebscn.com

联系人

吴晶 021-22167203 wujing@ebscn.com

相关研报

互联网数据挖掘系统,行为金融新领域 2010-09 量化择时之三——情绪拐点择时

2012-02



图表目录

图表	1:	光大中文云文本挖掘系统架构	. 3
		在财经新闻中搜索与 A 股相关的词组统计构成 A 股关注指数与上证指数	
图表	3:	光大经济担忧指数	. 5
图表	4 커	允大资金面担忧指数	. 5
		光大通胀担忧指数	
图表	6:	光大加息关注指数	. 5
图表	7:	基于研究报告的"石墨烯"概念热度与石墨烯板块指数	. 6
图表	8:	基于股票论坛的"石墨烯"概念热度与石墨烯板块指数	. 6
图表	9	基于财经新闻的"石墨烯"概念热度与石墨烯板块指数	. 6
图表	10	光大普通投资者多空指数	. 7
图表	11	光大机构多空指数	. 7
		基于论坛帖子文本的'页岩气'网络图	
图表	13:	基于论坛帖子文本的'抗癌'网络图	. 8
图表	14	基于股票论坛文本数据过去两年各概念板块 top10 龙头股	. 8
图表	15:	'000001.SS'关注度因子与股价	11



一、光大中文云文本挖掘系统

在光大中文云系统中, 我们主要是分为三个具体模块, 分别是数据获取与 清洗模块、数据引擎与指标构建模块、策略开发模块。

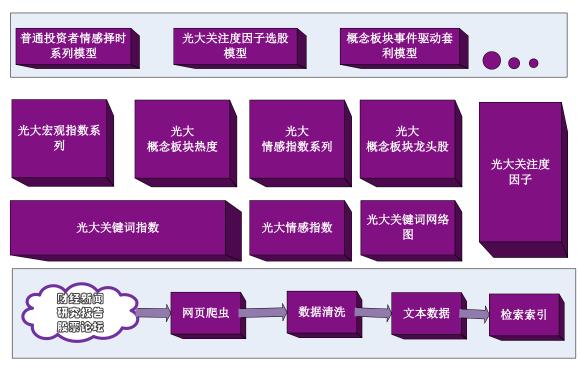
处于最底层的是数据获取与清洗模块。在该模块中,我们及时地从相应的 网站获取所需的所有文本信息。然后对获取得到的数据进行清洗,清除掉一些 重复和无意义的文本,然后将其存储于相应的文件库内。目前我们获取的文本 资料主要由三个来源,一是热门财经新闻,其次是所有分析师的研究报告,最 后是一些热门股票论坛上的所有帖子。从 2010 年 6 月到现在,我们的系统上 较为完整的积累了近 36 万篇研究报告、300 万篇财经新闻、2500 万条论坛帖子。 而考虑到文本数量的巨大以及为了维持较高的性能,我们为这些文本资料建立 了相应的全文索引,通过该全文索引,我们可以迅速地在海量的文本资料中获 取所需的具体资料,目前从 2500 万条论坛帖子中搜寻出某个关键词所在的所 有位置仅需要 1 秒左右。

系统的中间部分是数据引擎与指标构建模块。针对于底层的全文索引,为了方便我们日常的管理与引用,我们开发了各种数据引擎,这些数据引擎在一定程度上隔绝了应用和底层的文本资料,保持了各自的独立性,同时也可以让使用者较为方便地进行调用。这些数据引擎主要包括了光大关键词指数、光大情感指数、光大关键词网络图以及光大关注度因子。在该模块中,我们还完成了基于各数据引擎,构造我们的因子与指标产品的工作。目前我们的因子和指标产品主要包括有光大宏观系列指数、光大情感指数系列、光大概念板块热度、光大概念板块龙头股、光大关注度因子等。

而在文本挖掘系统最上层的模块就是基于中间模块所产生的各指标和因子数据,我们所开发的各类投资模型,目前主要有普通投资者情感择时系列模型、光大关注度因子选股模型、概念板块事件驱动的套利模型等。

在数据的维护及更新上,我们编写好所有的自动执行的程序,在相应的时间自行地启动和执行,这样可以方便我们及时获取最新的数据,并且减轻了我们进行系统维护的成本。

图表 1: 光大中文云文本挖掘系统架构



数据来源:光大金工中文云 www.chinesecloud.net



二、光大关键词指数

在关键词指数数据引擎中,我们可以从选定的文本资料源中搜索任意的关键词或者词语组合,提取包含有该关键词或者关键词组合的文本资料。然后按周统计搜索得到的文本资料数目,将其除以本周总的文本资料数目得到各周关键词指数。通过该数据引擎,我们可以对市场上对于该方面的关注的值进行量化,同时通过该数据引擎,我们也可以迅速的了解历史上市场对于该方面的关注程度值及目前所处的历史位置。

基于该数据引擎以及我们自定义的一些关键词组合,我们构造了光大宏观系列指数、光大概念板块热度等指标。

MA(20) 0.05 7天平均词频数 0.03 0.02 0.01 0.00 2010 "000001" 股价走势 3400 MAC201 3200 MA160 3000 \$ 2800 열 2600 2400 2200 2000 2011 Mar

图表 2: 在财经新闻中搜索与 A 股相关的词组统计构成 A 股关注指数与上证指数

数据来源:光大金工中文云 www.chinesecloud.net

2.1 光大宏观系列指数

针对不同的宏观元素,我们首先构造了表现特定宏观现象的关键词组合,然后应用关键词指数数据引擎在所有的宏观研究报告中搜索该关键词组合,则我们分别得到了光大宏观系列指数。例如当我们在构造光大经济担忧指数时,我们首先设计了一个如"经济 AND (下降 OR 下滑 OR 放缓)"类似的关键词组合,然后通过我们的关键词指数数据引擎在宏观研究报告中搜索含有该关键词组合的文本资料,然后得到一个经济担忧指数的时间序列,该时间序列即为光大经济担忧指数。通过这种方式,我们可以对市场长期关心的因素编制相应指标,以反映专业分析师团体对该因素的预期及整体关注程度。

图表 3: 光大经济担忧指数



数据来源:光大金工中文云 www.chinesecloud.net

图表 5: 光大通胀担忧指数



数据来源:光大金工中文云 www.chinesecloud.net

图表 4 光大资金面担忧指数



数据来源:光大金工中文云 www.chinesecloud.net

图表 6: 光大加息关注指数



数据来源:光大金工中文云 www.chinesecloud.net

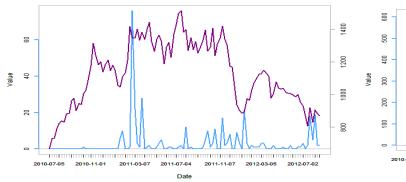
2.2 概念板块热度

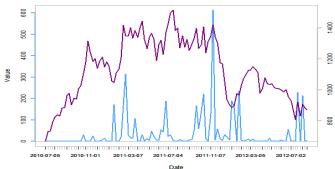
在光大概念板块热度中,其构造方式基本与宏观系列指数的构造方式一致,首先我们针对于每一个概念板块,制定了与该板块相关联的一系列关键词组合。例如针对于页岩气概念板块,我们构建了如"页岩气 OR 页岩 OR 煤层气"类似的关键词组合,然后应用关键词指数数据引擎分别在所有财经新闻、论坛帖子、研究报告中搜索该关键词组合,则我们分别得到了基于财经新闻、股票论坛、研究报告的各概念板块热度。

通过该概念板块热度,我们可以对于各概念在某段时间内的所受到的关注 程度进行量化以及迅速的了解目前该概念板块所受到的关注程度所在的历史 相对位置。

基于概念板块热度, 我们开发了概念板块事件驱动套利策略系列, 取得了 良好的成果, 具体的可以参考我们后续的报告。

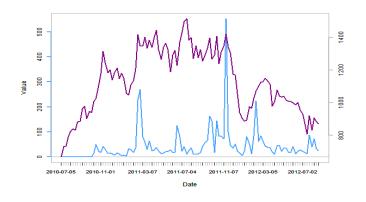
图表 7: 基于研究报告的"石墨烯"概念热度与石墨烯板块指数 图表 8: 基于股票论坛的"石墨烯"概念热度与石墨烯板块指数





数据来源:WIND 资讯,光大金工中文云 www.chinesecloud.net 图表 9基于财经新闻的"石墨烯"概念热度与石墨烯板块指数

数据来源:WIND资讯,光大金工中文云 www.chinesecloud.net



数据来源:WIND 资讯,光大金工中文云 www.chinesecloud.net

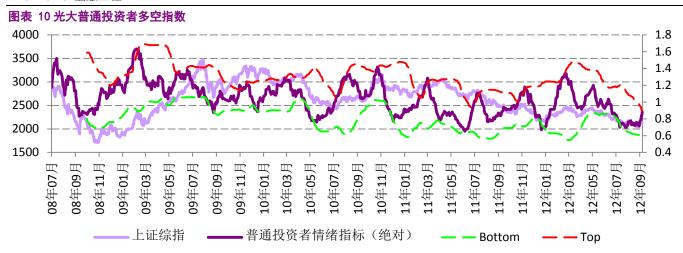
三、光大情感指数

在该情感指数数据引擎中,我们从特定的文本资料中,通过关键词组合搜索提取相应的文本资料,然后借助于我们的合作单位武汉大学 ROST 团队开发的中文情感识别模型与程序,对文本资料进行情感判别并打分,其中情感判别分为两类:乐观与悲观,正的分值对应乐观情绪,分值越高情绪越乐观;负的分值对应悲观情绪,分值越低情绪越悲观。然后将乐观的得分除以悲观的得分得到改关键词组合的情感指数。

3.1 普通投资者情绪指标

在构建普通投资者多空指数时,我们搜索所有的论坛帖子,对每一条帖子进行情感识别得到其情感分值后,我们按交易日将所有的帖子分类汇总,将所有乐观帖子和悲观帖子的分值分别加总,得到每一个交易日的乐观总分(定义为看多值)和悲观总分(定义为看空值)。将看多值除以看空值,然后按20个交易日平滑得到光大普通投资者多空指数(普通投资者情绪指标)。





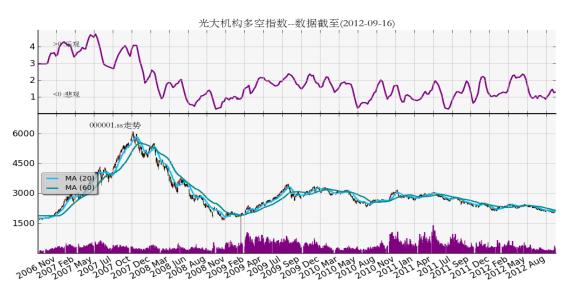
数据来源:光大金工中文云 www.chinesecloud.net

基于普通投资者情绪指标,我们开发出了系列大盘择时策略,具体的请参 考我们的报告《量化择时之三——情绪拐点择时》、《情绪指标多轨线择时策略》。

3.2 机构多空指标

具体的,当我们在构建机构多空指数时,我们搜索所有的研究报告,对每篇研究报告进行情感识别得到每篇报告的情感分值后,同样的,我们按交易日将所有的报告分类汇总,将所有乐观报告和悲观报告的分值分别加总,得到每一个交易日的乐观总分(定义为看空值)。将看多值除以看空值,然后对其求对数值,将结果按四周平滑得到光大机构多空指数。

图表 11 光大机构多空指数



数据来源:光大金工中文云 www.chinesecloud.net

四、光大关键词网络图

在关键词网络图中,我们从选定的文本资料中,通过关键词组合搜索 提取相应的文本资料,然后按照该文本资料进行分词,这里在分词时,我们需 要一个自定义金融字典,否则就会将一些我们所需要的词语拆分开,如将"平 安银行"拆分为"平安"和"银行"两个词语。之后在分词的结果中过滤去除 一些意义不大或与金融无关的词组,计算剩下的每一个词组在该文本资料内的

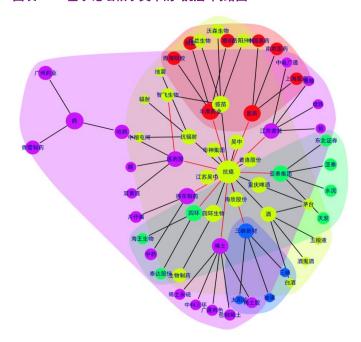


与概念词相关性的指标。在该指标的设定中,我们遵循的标准即为:某个词在特定文档中出现的次数越多,则该词的重要性也就越高;若某个词在其他所有文档中出现的次数越少,则该词的重要性也就越高。通过该原则而设定的相关性指标,选取相关性高的词组做为与该关键词最相关的词组,然后我们对查找出来的与该关键词最相关的词组分别重复上述过程,得到这些词组最相关的词组,构成最初关键词的二级相关词组,最终结果以网络图的形式进行展示。

通过该应用可以迅速地对一些新的或陌生的词组,有一个最全面和直接的 了解,同时也方便于迅速地寻找词组所对应的热门投资标的。

图表 12: 基于论坛帖子文本的'页岩气'网络图

图表 13: 基于论坛帖子文本的'抗癌'网络图



数据来源:光大金工中文云 www.chinesecloud.net

数据来源:光大金工中文云 www.chinesecloud.net

4.1、概念板块龙头股

在关键词网络图中,当我们的输入为与概念板块相关的系列词组时,并且在输出的一级相关词组中过滤掉非股票名称词语,则我们得到各个概念板块的龙头股组合。当然这里我们可以设文本资料的时间段,则我们可以得到各个时间段内各概念板块的龙头股。如下图为过去的两年中,各概念板块的 top10 龙头股。

图表 14基于股票论坛文本数据过去两年各概念板块 top10 龙头股

概念板块名称	龙头股 1	龙头股 2	龙头股3	龙头股 4	龙头股 5	龙头股 6	龙头股7	龙头股8	龙头股9	龙头股 10
石墨烯	中国宝安	金路集团	中钢吉炭	力合股份	方大炭素	维科精华	天富热电	新华锦	红星发展	华丽家族
锂电池	西藏矿业	钱江摩托	中信国安	天齐锂业	亿纬锂能	赣锋锂业	路翔股份	成飞集成	杉杉股份	佛山照明
智能交通	银江股份	亿阳信通	皖通科技	北京城乡	新大陆	键桥通讯	远望谷	四维图新	启明信息	浙大网新
稀土永磁	包钢稀土	广晟有色	中色股份	中科三环	西藏发展	包钢股份	五矿发展	太原刚玉	鼎立股份	厦门钨业
舟山新区	金鹰股份	中水渔业	宁波港	宁波联合	澳洋顺昌	柘中建设	海越股份	浙江广厦	东百集团	康恩贝
网络游戏	中青宝	博瑞传播	海虹控股	浙大网新	顺网科技	乐视网	中科英华	恒信移动	时代出版	奥飞动漫
新疆区域振兴	国际实业	新疆城建	新疆天业	啤酒花	特变电工	伊力特	北新路桥	天山股份	天富热电	天利高新
高铁	晋亿实业	中国南车	中国北车	太原重工	晋西车轴	中国铁建	中国中铁	特锐德	中铁二局	世纪瑞尔
安防监控	英飞拓	航天长峰	海康威视	大华股份	星网锐捷	世纪瑞尔	鑫茂科技	安居宝	金刚玻璃	远望谷

敬请参阅最后一页特别声明



证券研究报告

2012 3 10 业而4	/ ₁									
核高基	中国软件	欧比特	太极股份	浙大网新	同方股份	浪潮软件	川大智胜	联信永益	华微电子	世纪鼎利
水利水电建设	粤水电	葛洲坝	三峡水利	安徽水利	岷江水电	明星电力	钱江水利	大禹节水	长江电力	新疆天业
苹果	东方通信	莱宝高科	汉王科技	中国联通	超声电子	金利科技	利源铝业	九安医疗	宇顺电子	爱施德
文化传媒	现代投资	天舟文化	中南传媒	中视传媒	中体产业	时代出版	华谊嘉信	珠江实业	博瑞传播	乐视网
IPV6	紫光股份	星网锐捷	力合股份	高鸿股份	中兴通讯	东方通信	道博股份	同方股份	浪潮信息	三五互联
概念板块龙头	三维丝	大禹节水	动力源	九阳股份	奥特迅	中天科技	深长城	工大高新	中联电气	通威股份
云计算	浪潮信息	华胜天成	中电广通	浪潮软件	华东电脑	长城电脑	鹏博士	星网锐捷	新大陆	东软集团
港珠澳大桥	格力地产	东风汽车	紫金矿业	中国建筑	美达股份	华发股份	珠海港	塔牌集团	世荣兆业	天威保变
分拆上市	同方股份	辽宁成大	复星医药	雅戈尔	南京高科	中关村	中国宝安	中青旅	东方市场	鑫茂科技
三网融合	歌华有线	广电网络	数码视讯	东方通信	东方明珠	同洲电子	电广传媒	金亚科技	乐视网	武汉塑料
股权激励	中集集团	兴森科技	人人乐	乐视网	名流置业	湘鄂情	中国建筑	中国宝安	天津港	冠城大通
新材料	大西洋	诚志股份	杭齿前进	方大炭素	中钢吉炭	中国宝安	安泰科技	宝利沥青	皖维高新	东睦股份
宽带提速	中天科技	乐视网	亨通光电	烽火通信	兴森科技	日海通讯	浙大网新	中国联通	达华智能	杰赛科技
中日韩自贸区	新华锦	连云港	青岛双星	大连国际	大连港	青岛金王	新兴铸管	日照港	南宁百货	山东高速
煤层气(页岩气)	江钻股份	海默科技	天科股份	辽宁成大	准油股份	神开股份	杰瑞股份	煤气化	山东墨龙	中海油服
绿色节能照明	飞乐音响	雪莱特	新疆天业	嘉宝集团	三安光电	佛山照明	华微电子	兆驰股份	珠江实业	东方日升
合同能源管理	泰豪科技	九洲电气	智光电气	延华智能	易世达	东方电子	金通灵	方大集团	达实智能	科陆电子
上海本地重组	上港集团	上海机场	嘉宝集团	中纺投资	复旦复华	上海梅林	运盛实业	飞乐音响	泰胜风能	上海贝岭
3G	中国联通	大唐电信	世纪鼎利	中兴通讯	高鸿股份	南京熊猫	爱施德	神州泰岳	中国石油	宏图高科
移动互联网	恒信移动	拓维信息	华胜天成	北纬通信	亿阳信通	卫士通	三元达	神州泰岳	中国联通	三五互联
环保	中原环保	富春环保	三维丝	先河环保	天立环保	霞客环保	现代投资	创业环保	三聚环保	龙净环保
建筑节能	延华智能	泰豪科技	达实智能	方大集团	金晶科技	彩虹精化	嘉寓股份	中航三鑫	鲁阳股份	东百集团
温州金融改革	浙江东日	哈高科	香溢融通	动力源	浙江东方	金丰投资	浙江富润	通程控股	轻纺城	珠江实业
海峡西岸	福建高速	漳州发展	三木集团	厦门港务	冠城大通	建发股份	中国武夷	东百集团	青山纸业	中福实业
移动支付	国民技术	东信和平	恒宝股份	卫士通	证通电子	长电科技	大唐电信	东方通信	南天信息	新大陆
核能核电	中国一重	东方电气	中核科技	大西洋	沃尔核材	上海电气	湖北能源	哈空调	赣粤高速	皖能电力
抗癌	江苏吴中	海欣股份	四环生物	丰原药业	华神集团	普洛股份	九芝堂	亚泰集团	重庆啤酒	江苏索普
物联网	新大陆	东方通信	大唐电信	远望谷	厦门信达	中瑞思创	长电科技	高鸿股份	达华智能	东方电子
皖江城市带	皖通高速	山鹰纸业	皖维高新	国投新集	国风塑业	皖新传媒	金种子酒	恒源煤电	华茂股份	盛运股份
太阳能发电	万家乐	天威保变	利达光电	乾照光电	拓日新能	精功科技	宏润建设	超日太阳	航天机电	东方日升
新能源汽车	比亚迪	福田汽车	长安汽车	安凯客车	东风汽车	中通客车	佛山照明	曙光股份	万向钱潮	路翔股份
节能元器件										
(IGBT)	科达股份	华微电子	长电科技	钱江摩托	东光微电	台基股份	中国南车	中环股份	大恒科技	通富微电
卫星导航	中国卫星	北斗星通	海格通信	华力创通	江苏三友	合众思壮	国腾电子	动力源	北京利尔	四维图新
新三板	苏州高新	中关村	张江高科	动力源	电子城	海泰发展	东湖高新	紫光股份	奥特迅	复旦复华
装饰园林	东方园林	棕榈园林	栖霞建设	泰达股份	世纪鼎利	广田股份	壹桥苗业	国民技术	神州泰岳	金螳螂
外资并购(国际	左吐肌从	二二肌八	联关协机	工业 & 业	阳业肌从	拉比於 #	名从 肌 八	上计肌从		क्रमधी कर ग
<u>板</u>)	东睦股份	三元股份	联美控股	万业企业	阳光股份	杭齿前进	多伦股份	大江股份	珠海港 大法的儿	四川双马
氢能	同济科技	华昌化工	广安爱众	英特集团	金龙汽车	中航地产	北海港	南都电源	东凌粮油	徐工机械
高端装备制造	抗齿前进 	沈阳机床	机器人	西飞国际	中国一重	中集集团	中船股份	中航动控	中国船舶	柳工
钢结构	杭萧钢构	东南网架	法因数控	精工钢构	光正钢构	鸿路钢构	新潮实业	中国中冶	宝钢股份	大金重工

- 9 -



垃圾发电	富春环保	深圳能源	泰达股份	东方市场	华电能源	盛运股份	南海发展	凯迪电力	山鹰纸业	美欣达
通用航空	中信海直	海特高新	航天科技	威海广泰	洪都航空	西飞国际	山河智能	航空动力	哈飞股份	贵航股份
涉矿	运盛实业	莱茵置业	华银电力	科学城	大西洋	斯米克	华茂股份	山东高速	华业地产	万泽股份
新能源	宝新能源	比亚迪	福田汽车	西藏矿业	安凯客车	佛山照明	中国宝安	宏润建设	东风汽车	长安汽车
央企重组	东方通信	中电广通	本钢板材	乐凯胶片	葛洲坝	广弘控股	高鸿股份	东百集团	江泉实业	吉电股份
浦东新区	上海金陵	津滨发展	海泰发展	陆家嘴	外高桥	上工申贝	珠海中富	长江投资	上海建工	渝开发
参股金融	现代投资	中原高速	万顺股份	宝通带业	皖维高新	浙江东日	弘业股份	浙江东方	方正电机	金丰投资
成渝特区	渝开发	华智控股	富临运业	中海发展	新希望	西南药业	海信电器	四川路桥	楚天高速	兰生股份
航母	中船股份	中国重工	中国船舶	杭齿前进	中原特钢	抚顺特钢	广船国际	新钢股份	中国一重	大连港
充电桩	动力源	奥特迅	通程控股	科陆电子	东百集团	万马电缆	西部牧业	京能置业	上海普天	国星光电
深圳前海新区	深天健	世纪星源	冠城大通	方大集团	招商地产	广聚能源	罗牛山	华联控股	中国宝安	新都酒店
生物育种	荃银高科	登海种业	隆平高科	丰乐种业	东方海洋	大北农	安徽水利	万向德农	莱茵生物	敦煌种业
保障性住房	栖霞建设	天房发展	中国中冶	中天城投	塔牌集团	保利地产	中国建筑	杭齿前进	渝开发	海螺型材
数字电视	龙元建设	数码视讯	数源科技	同洲电子	歌华有线	东方集团	博闻科技	金亚科技	中信国安	海信电器
触摸屏	华东科技	莱宝高科	超声电子	长信科技	欧菲光	宇顺电子	方兴科技	士兰微	康强电子	长城开发
清洁煤发电	同济科技	华昌化工	广安爱众	英特集团	金龙汽车	中航地产	北海港	南都电源	东凌粮油	徐工机械
风力发电	金风科技	闽东电力	银星能源	申华控股	湘电股份	长征电气	珠海港	九洲电气	上海电气	泰胜风能
智能电网	东方电子	福星晓程	中元华电	特变电工	中天科技	积成电子	华智控股	三维丝	思源电气	国电南瑞
新型煤化工	伊力特	赤天化	中国化学	远兴能源	云维股份	中泰化学	四环生物	中煤能源	兰花科创	山西三维
迪士尼	大西洋	东方明珠	界龙实业	上海建工	陆家嘴	中路股份	斯米克	浦东建设	上海机场	华谊兄弟
黄金珠宝	东百集团	老凤祥	广发证券	浦东金桥	杭齿前进	华泰股份	漫步者	潮宏基	葛洲坝	比亚迪
滨海新区	津滨发展	海泰发展	泰达股份	滨海能源	天津港	天房发展	华北高速	英特集团	天保基建	鑫茂科技
乙醇汽油	渝开发	华资实业	中国石化	兴民钢圈	丰乐种业	天茂集团	敦煌种业	远兴能源	万向德农	恩华药业
海南旅游岛	罗牛山	海南高速	海德股份	海南航空	迪马股份	首创股份	正和股份	中国国旅	海虹控股	中海海盛
北部湾经济区	北海港	两面针	五洲交通	桂冠电力	柳工	南宁百货	中国重工	朗科科技	国恒铁路	中信证券
三沙	海峡股份	罗牛山	海南高速	正和股份	中海海盛	华业地产	中国船舶	中信海直	国联水产	中船股份

数据来源:WIND 资讯,光大金工中文云 www.chinesecloud.net

基于该概念板块龙头股,我们对概念板块事件驱动套利策略进行了改进,使得该策略能够更贴近于市场。我们均使用最近三个月的文本数据挖掘出各个概念板块的龙头股,用这些龙头股构成的组合来替代原来所使用的概念板块指数进行策略开发,具体的可以参考我们后续的报告。

五、光大关注度因子

我们按月统计主流股票论坛上每只股票的新发帖总量,将该发贴总量除以该股总市值开三次方得到光大个股关注度因子,同时我们根据个股的新发帖总量相对上一个月的变化率构造关注度变化因子以及我们根据关注度因子在过去一年中最高关注度因子与最低关注度因子间相对位置构造关注度历史相对位置因子。

我们发现无论是关注度因子还是关注度变化因子、关注度历史相对位置因 子与股价表现都呈现明显的负相关性。



图表 15: '000001. SS'关注度因子与股价



数据来源:WIND 资讯,光大金工中文云 www.chinesecloud.net

关于关注度因子, 我们的回测显示, 关注度低的股票在后续的表现中显著的优于关注度高的股票, 具体的可以参考我们的报告《关注度选股因子历史回测》。



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

倪蕴韬,瑞典哥德堡大学金融学硕士,复旦大学世界经济系学士,2010年加入光大证券研究所任金融工程分析师。 主要研究方向:投资时钟量化框架、算法交易、高频策略。

行业及公司评级体系

做多一未来 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--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%;

做空一未来 6-12 个月的投资收益率落后市场基准指数 15%以上。

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

	• • • • • • • • • • • • • • • • • • • •	., ,,		
销售交易团队	姓名	办公电话	手机	电子邮件
北京	王汗青(总经理)	010-56513035	-	wanghq@ebscn.com
	郝辉	010-56513031	-	haohui@ebscn.com
	黄怡	010-56513050	-	huangyi@ebscn.com
	梁晨	010-56513153	-	liangchen@ebscn.com
企业客户	孙威(执行董事)	010-56513038	-	sunwei@ebscn.com
	吴江	010-56513030	-	wujiang@ebscn.com
	杨月	010-56513033	-	yangyue1@ebscn.com
	顾超	021-22169485	-	guchao@ebscn.com
上海	李大志(销售交易部总经理助理)	021-22169128	-	lidz@ebscn.com
	严非(执行董事)	021-22169086	-	yanfei@ebscn.com
	周薇薇	021-22169087	-	zhouww1@ebscn.com
	徐又丰	021-22169082	-	xuyf@ebscn.com
	韩佳	021-22169491	-	hanjia@ebscn.com
	冯诚	021-22169083	-	fengcheng@ebscn.com
深圳	黎晓宇(副总经理)	0755-83024434	-	lixy1@ebscn.com
	黄鹂华(执行董事)	0755-83024396	-	huanglh@ebscn.com
	张晓峰	0755-83024431	-	zhangxf@ebscn.com
	江虹	0755-83024029	-	jianghong1@ebscn.com
	罗德锦	0755-83024064	-	luodj@ebscn.com
富尊财富中心	濮维娜(副总经理)	021-62152373	-	puwn@ebscn.com
	陶奕	021-62152393	-	taoyi@ebscn.com
	戚德文	021-22169152	-	qidw@ebscn.com