

分析师:

郑兆磊

zhengzhaolei@xyzq.com.cn

S0190520080006

猎金系列之三十三： 网络关注度因子真的有选股 Alpha 吗

2021 年 12 月 6 日

报告关键点

基于互联网文本的研究一般有三个方向：文本情绪挖掘、文本主题聚焦、关注度/热度指标构建。尤其是关注度指标的构建，大家关注颇多。那么基于互联网的**关注度是否真的就有 Alpha 选股能力呢？**这是本文关注的焦点。

投资要点

- 基于互联网文本的研究一般有三个方向：文本情绪挖掘、文本主题聚焦、关注度/热度指标构建。尤其是关注度指标的构建，大家关注颇多。那么**基于互联网的**关注度是否真的就有 Alpha 选股能力呢？****这是本文关注的焦点。
- 本文以百度开发的**搜索指数**为基础，获取个股舆情信息，挖掘并构建**关注度因子**。百度搜索指数是以海量网民行为数据为基础开发的数据分享平台，可以用来观测市场上相关关键词的搜索指数趋势，是一个非常好的代表互联网文本关注度/热度的数据指标。
- 根据频度的不同，我们构建了**周度及月度的个股关注度因子**。IC 及分位数检验结果显示，关注度因子有着较好的选股能力。行业市值中性化后：周度关注度因子 IC 为 0.062、T 值达到 17；月度关注度因子 IC 达到 0.074，T 值达到 8。分位数各组别严格单调；周度因子多空收益率高达 44.8%，夏普比达到 3.95，月度因子多空收益达到 20.1%，夏普比达到 1.53。即便在沪深 300 以及中证 500 内因子依然具有非常好的选股效果。
- **特异性方面：**我们通过相关性以及 Fama-MacBeth 回归测试以验证该因子与 Barra 十个风格因子的特异性。从测试结果来看，该因子与 Barra 风格因子相关性低，同时 Fama-MacBeth 回归测试也验证了周度以及月度关注因子的特异性。
- 但通过暴力的方式测试关注度因子与兴证金工底层所有因子（兴证金工底层共计 165 个因子）的相关性，结果发现关注度指标与部分因子相关性相对较高。这 5 个因子均是常见的衡量市场流动性和市场结构的指标。**如果将这些因子和 Barra 十个风格因子放到一起进行 Fama-MacBeth 测试，结果发现其特异性明显减弱，对收益率的解释力度所剩无几。**

风险提示：本报告模型及结论全部基于对历史数据的分析，当市场环境变化时，存在模型失效风险。

目录

1、基于网络数据衍生的关注度日益成为焦点.....	3 -
2、百度指数简介	3 -
2.1、网络热度的最佳体现——搜索引擎	3 -
2.2、百度指数介绍	4 -
2.3、百度搜索指数获取思路	6 -
3、百度搜索指数数据统计分析	7 -
3.1、百度搜索指数数据基本概况	7 -
3.2、百度搜索指数对 A 股的覆盖度.....	7 -
4、关注度因子构建及有效性特异性分析	8 -
4.1、关注度因子逻辑阐述	8 -
4.2、关注度因子效性测试	9 -
4.3、关注度因子在宽基指数成分股内有效性分析.....	12 -
4.4、关注度因子真的有效吗？	14 -
5、结语	16 -
图表 1、基于百度另类数据构建的大数据基金	4 -
图表 2、百度搜索指数趋势界面展示	5 -
图表 3、百度资讯指数趋势界面展示	5 -
图表 4、股票简称与股票代码（数字部分）搜索指数对比——贵州茅台为例..	7 -
图表 5、百度搜索总指数与万得全 A 趋势对比.....	7 -
图表 6、百度搜索指数交易日覆盖度统计	8 -
图表 7、股价的过度反映与基本价值偏离	9 -
图表 8、不同周期下构建的关注度因子的 IC 检验.....	10 -
图表 9、周度关注度因子分位数组合测试结果	10 -
图表 10、月度关注度因子分位数组合测试结果.....	10 -
图表 11、行业市值中性化前后因子的 IC 表现.....	10 -
图表 12、行业市值中性化后周度关注度因子分位数组合测试结果.....	11 -
图表 13、行业市值中性化后月度关注度因子分位数组合测试结果.....	11 -
图表 14、行业市值中性化周度关注度因子多空组合净值及收益率表现.....	11 -
图表 15、行业市值中性化月度关注度因子多空组合净值及收益率表现.....	12 -
图表 16、宽基成分股内因子的 IC 表现	12 -
图表 17、周度关注度因子在宽基指数下分位数组合测试结果.....	13 -
图表 18、月度关注度因子在宽基指数下分位数组合测试结果.....	13 -
图表 19、宽基指数下周度关注度的平均水平.....	14 -
图表 20、宽基指数下月度关注度的平均水平.....	14 -
图表 21、关注度因子与传统因子相关性表现.....	15 -
图表 22、剥离流动性后的因子表现	15 -
图表 23、月度关注度因子与 Barra 十个风格因子 Fama-MacBeth 回测结果	15 -
图表 24、月度关注度因子与兴证金工底层所有因子的相关性分析.....	16 -
图表 25、月度关注度因子与其他因子的 Fama-MacBeth 回测结果.....	16 -

报告正文

1、基于网络数据衍生的关注度日益成为焦点

量化投资是一种借助数学、统计学、信息技术等知识对海量的历史大数据进行挖掘，综合归纳出有效的因子或者模型程序来代替人为的主观判断的一种投资方式。显然，挖掘方法以及数据都是量化投资不可或缺的核心要素。

在过去，量化投资者更多的依靠大型金融资讯平台提供的传统金融数据进行研究分析。这导致数据同质化严重，侵蚀了传统因子的有效性，从而稀释了量化投资的利润空间。

大数据时代的到来给量化投资带来一片新的沃土，信息技术的进步让更多的另类数据留存变为可能。每时每刻，互联网上大量的个体活动、商业活动都在产生新的数据并被记录。量化投资者不必再局限于传统的量价数据中厮杀，许多量化投资者开始尝试从千奇百怪的另类数据中提取交易信号。这些另类数据可能是个人在搜索引擎中的搜索数据、社交媒体评论，也可能是公司的交易数据、新闻公告数据，还可能是政府网站的公开信息披露，这些微妙而复杂的信息实际上可能反映着市场上瞬息万变的情绪和供求关系。这些另类数据正在成为传统数据的重要补充，帮助量化投资者打开了新的视角，在挖掘独特的超额收益因子上发挥着巨大价值。大数据洪流的冲击，使得量化投资正在演变为一场数据驱动的信息化战争。

而在所有的另类数据中，基于互联网产生的信息便是一座隐藏起来的巨大金矿。互联网的存在，让广大群众足不出户就能够知晓天下的新闻与趣事，同时来自广大个体的分散声音，也能够传递并被记录下来。而搜索引擎作为互联网生活不可或缺的一个组成部分，帮助用户从眼花缭乱的信息资源中快速准确地找到自己所需的信息，每时每刻产生着数以万计的搜索数据。

从利用的方式来看基于互联网的文本数据主要包括文本情绪判断、热点主题提取、热度指标计算。其中热度指标是大家经常会聚焦的一个方向。我们可以基于雪球、股吧、各大新闻网站等计算个股帖子的数目，以生成个股关注度/热度。很多第三方数据库也在提供类似的指标供机构投资者使用。这些指标计算方式相同，反应问题相同。但基于互联网产生的热度指标是否真的具有 Alpha 呢？这是本报告聚焦的问题。

我们选择百度搜索引擎作为我们的切入点，以个股在搜索引擎上的频次作为选股因子，研究其选股有效性和特异性，以充分论证热度指标是否就有 Alpha 能力。

2、百度指数简介

2.1、网络热度的最佳体现——搜索引擎

“百度一下，你就知道”，作为全球最大的中文搜索引擎，百度是网民获取中文信息和服务的最主要入口，“有问题，问百度”的思维惯性已深深地植入到中国

大多数网民的心里。百度每天响应来自 100 余个国家和地区的数十亿次搜索请求，服务 10 亿互联网用户，覆盖了用户上至国家大事、下至生活常识的信息需求，是个天然的另类大数据提供平台。

那么百度搜索引擎所产生的另类数据究竟能对量化投资带来什么帮助呢？百度日均搜索数据达 100 亿，这其中包括对个股、主题词汇、公司主营业务等方面信息的搜索。这些数据可以反映散户对证券市场、个股、投资主体关注程度，可以用于判断热点板块、进行事件驱动型投资、预测经济景气程度。百度也充分发挥自己的数据优势，积极的布局量化投资领域。早在 2014 年 7 月 8 日，百度便与中证指数、广发基金跨界合作，推出国内首只反映市场情绪的互联网大数据金融市场指数——“中证百度百发策略 100 指数”。

截止今日，已经有 2 家公募机构选择与百度合作，基于百度提供的数据构建大数据基金，具体参见图表 1。

图表 1、基于百度另类数据构建的大数据基金

名称	代码	管理人	上市时间
广发百发 100 指数 A	000826.OF	广发基金	2014-10-30
广发百发 100 指数 E	000827.OF	广发基金	2014-10-30
易方达易百智能量化策略 A	005437.OF	易方达基金	2018-1-24
易方达易百智能量化策略 C	005438.OF	易方达基金	2018-1-24

资料来源：wind，兴业证券经济与金融研究院整理

2.2、百度指数介绍

百度基于自身优势，以百度海量网民行为数据为基础开发了数据分享平台——百度指数。提供关键词搜索指数服务，帮助用户研究关键词搜索趋势、洞察网民需求变化、监测媒体舆情趋势、定位数字消费者特征等。

基于统计样本的类型不同，百度指数可以分为搜索指数以及资讯指数：

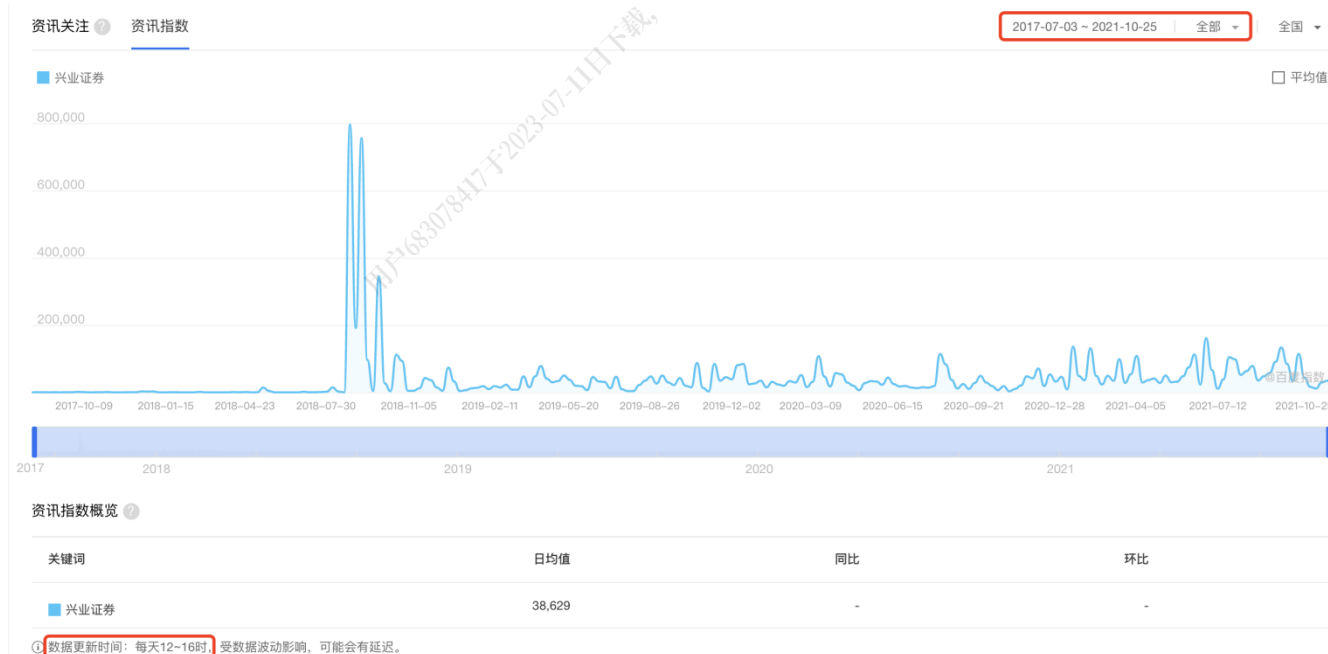
- 1、搜索指数主要展示的是**互联网用户**对关键词搜索关注程度及后续的持续变化情况。**百度搜索指数算法说明**：以网民在百度的搜索量为统计对象，科学分析并计算出各个关键词在百度网页搜索中搜索频次的加权。根据数据来源的不同，搜索指数分为 PC 搜索指数、移动搜索指数及 PC+移动搜索指数。
- 2、资讯指数反映的是**新闻资讯**在互联网上对特定关键词的关注和报道程度及其持续变化。**百度资讯指数算法说明**：以百度智能分发和推荐内容数据为基础，将网民的阅读、评论、转发、点赞、不喜欢等行为的数量加权求和得出资讯指数。

图表 2、百度搜索指数趋势界面展示



资料来源：百度指数，兴业证券经济与金融研究院整理

图表 3、百度资讯指数趋势界面展示



资料来源：百度指数，兴业证券经济与金融研究院整理

百度指数反映网民的主动搜索需求，所有影响网民搜索行为的活动都可能影响百度指数。为了保证指数的可信度，百度指数在防作弊(反作弊)方面做了大量工作，研究并建立了一整套相应的防作弊(反作弊)技术体系，力求将作弊现象减

请务必阅读正文之后的信息披露和重要声明

少到最低限度。用户发现任何作弊情况(嫌疑)可随时进入“意见反馈”举报,对于已经有明确证据证明使用作弊手段来提高排名的关键词,百度指数会对其数据进行很大幅度的降权处理。

总的来说,百度指数为我们提供了一个十分便利的数据接口,去观测市场上相关关键词的搜索指数趋势。如果将这些关键词限定在金融领域内,就能获得对我们投资有所帮助的舆情信息。

2.3、百度搜索指数获取思路

首先,在搜索指数的选择上,我们选取的是**百度搜索指数**而不是百度资讯指数,有以下两个原因:

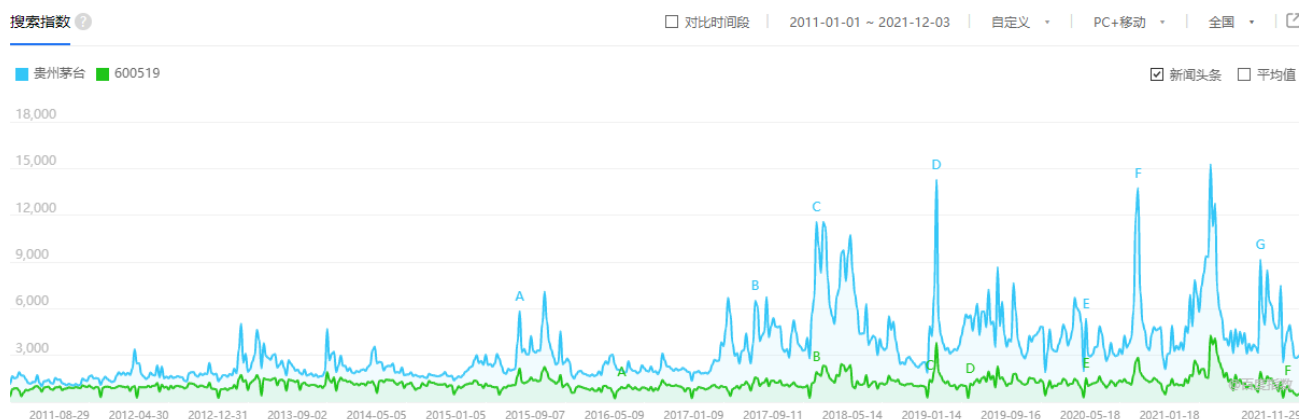
1. 我们关注的是市场上对于每只股票的关注度,而百度资讯指数是基于百度自身的智能分发,带有百度的主观色彩,可能会偏离互联网用户的主观搜索意愿;
2. 百度指数提供的搜索指数开始时间为 2011 年 1 月 1 日,数据长度超 10 年,数据量上更加充足;资讯指数开始时间为 2017 年 7 月 3 日,数据长度较短,不足以提供充足的样本。

其次,在搜索方式的选择上,我们使用的是**股票代码(数字部分)**的方式来获取该个股的关注度。我们认为,在搜索引擎上搜索一只个股时,常用且快捷的方式可能有两种——**股票简称与股票代码(数字部分)**(注:若无特殊说明,本文所指股票代码皆为**股票代码的数字部分**)。而根据我们的测试,使用股票代码构建的因子在选股方面的效果远远强于股票简称:考虑到用户搜索习惯的不同,我们认为会使用股票代码搜索的网民,代表更纯粹的股市参与者,他们在搜索引擎上获取的信息会切实影响他们接下来的买进或卖出行为,并传导至股市中。而使用股票简称搜索的网民当中,可能包含了仅仅只是想要了解该公司的用户,这些关注行为并没有反映到市场的交易中,从而给数据带来了噪音,影响因子的选股效果。

此外,根据百度指数的算法规则,达到收录标准的关键词会自动收录,未达到收录标准的关键词需要用户购买加词权限,将未收录的关键词加入百度指数,百度指数将为该词提供 1 年的数据计算分析服务。而根据我们的测试,并不是每个股票代码都有相应的搜索指数提供,根据其收录标准我们可以理解为该股票在搜索引擎上的关注度极低,相对其他股票可以忽略不计。

最终,我们对每只个股采集了其对应**股票代码的搜索指数**,并将未收录的字段的关注度标注为空,最终获得百度指数目前关于**全 A 市场上仍存续的 3922 个上市公司**的所有可得的**历史搜索指数序列**。

图表 4、股票简称与股票代码（数字部分）搜索指数对比——贵州茅台为例



资料来源：百度指数，兴业证券经济与金融研究院整理

3、百度搜索指数数据统计分析

3.1、百度搜索指数数据基本概况

我们得到的百度搜索指数数据的时间窗口为 2011 年 1 月 1 日至 2021 年 9 月 30 日，共计 3926 天。在此期间共获得 3922 只个股关于股票代码的历史搜索指数。

从百度搜索总指数的趋势来看，其与市场的走势存在非常密切的关系，特别是 15 年股灾前后指数骤升骤降也与搜索指数的变化如出一辙。因此我们认为其可能与市场存在着密切联系，可以用来构建因子，帮助我们进行选股。

图表 5、百度搜索总指数与万得全 A 趋势对比



资料来源：百度指数，Wind，兴业证券经济与金融研究院整理

3.2、百度搜索指数对 A 股的覆盖度

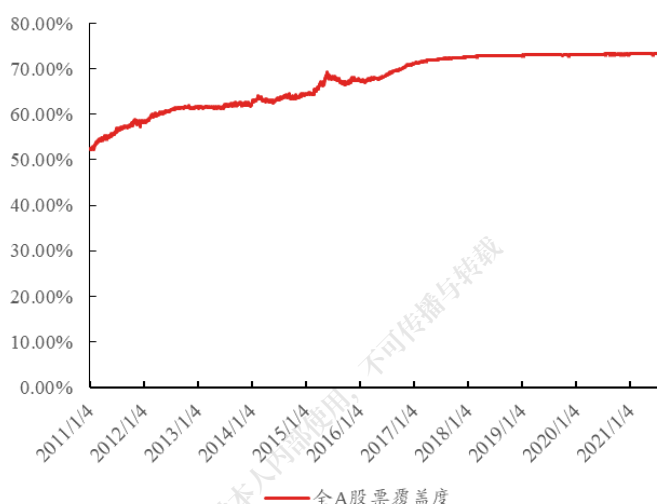
考虑到百度指数并不会收录每一个关键词，仅达到收录标准的关键词会被自动收录，因此我们认为未达到收录标准的关注度极低，可以认为是 0。但如果市场上有太多的 A 股并没有被百度指数收录，那么我们一杆子地将它们的关注度

都定为空，这些股票之间便没有区分度，可能会影响到最终的结果。因此我们统计了在时间窗口 2011 年到 2021 年 9 月的交易日股票覆盖度情况。

自 2011 年起，百度指数对于 A 股市场的覆盖度就不断提升。2011 年的股票覆盖度仅 52.50%，至 2021 年 9 月已经趋于 73%。目前基本趋于稳定，从 2017 年至今便维持在 70%-73% 的水平。

因而基于该维度构建的周度或者月度关注因子的覆盖度相对较高，支持进一步选股的研究。

图表 6、百度搜索指数交易日覆盖度统计



资料来源：百度指数，Wind，聚源，兴业证券经济与金融研究院整理

4、关注度因子构建及有效性特异性分析

4.1、关注度因子逻辑阐述

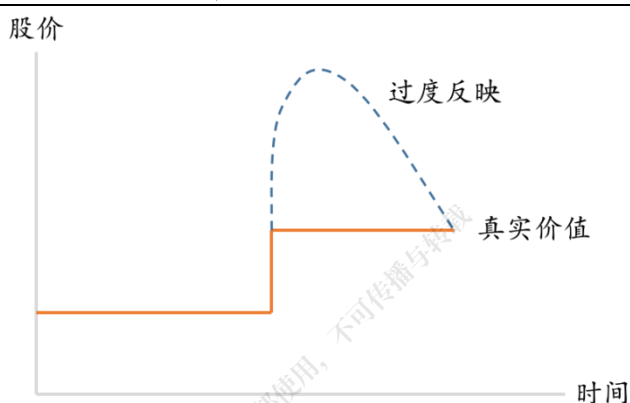
“盛极必衰，否极泰来”。世界的普遍规律告诉我们：凡事到达了一个极端，就会朝另一个方向发展。股票市场似乎也存在着相应的规律，总在牛市与熊市之间轮动。当市场情绪高涨，股市空前繁荣，股票市场的持续火热便会不断地吸引新的投资者进入股票市场，继而推动股市的增长，造成市场“虚假繁荣”的假象。当越来越多的股民们蜂拥进入股市，一旦利空消息来临，浇灭了市场的情绪，价格又会迅速崩盘，熊市骤然而至。

经典的经济学和金融学理论经的假设中，总是将每个人都看做是理性的投资者，认为个体在进行投资决策时会进行理智的分析，当股票价格低于上市公司的内在价值时，投资者开始买入股票；而当股票价格高于上市公司的内在价值时，开始卖出股票。但事实上，人非完人，大多数人都做不到完全的理性，往往被情绪牵动着行为。受到市场情绪的影响，投资者往往会做出非理性的投资行为，形成了市场的暴涨和崩盘现象。此外，中国的股市投资者组成中，有大量的散户投资者，相较机构投资者其行为更容易因为市场情绪而波动，造成了市场上“追涨杀跌”的现象。

请务必阅读正文之后的信息披露和重要声明

细化到个股的投资上，沾上了热点的个股股价往往反映惊人，蜂拥而入的资金推动股价迅速的提升，甚至偏离其真实价值。同时这样的股票也会受到市场更多的关注，反映在互联网上则表现为对该只股票的搜索行为不断增加。当股价开始下跌，回归其真实价值，市场情绪又会变得悲观，大量抛售行为使得股票进一步下跌乃至无人问津，反映在互联网上即不断减少的搜索行为。这也就意味着，个股的关注度与股价的走势有一定的负相关性。基于这个假设，我们根据百度指数提供的每只个股的搜索指数，构建关注度因子，验证股票当期关注度与下期股价收益率的关系。

图表 7、股价的过度反映与基本价值偏离



资料来源：兴业证券经济与金融研究院整理

由于关注度对股价的传导具有一定的滞后性，因此关注度因子定义为某只股票在一个时期内所有搜索指数之和，从而构建出不同频度的关注度因子。以周度关注度因子为例：平安银行在 2011 年 1 月 10 日至 2011 年 1 月 14 日共 5 个交易日的搜索指数分别为 11147、10218、9674、9610、9416，那么平安银行在这段时间内的周度关注度因子即为所有搜索指数之和 50065。

关注度因子定义：对于某只股票，在一个时间段 T 内，存在 t0 到 tn 共 n 个日度时点搜索指数的值，那么这个时间段 T 内对该只股票的关注度大小为 $\sum_{t=t_0}^{t_n} Searching_index_t$ 。

个股数据：全市场 A 股(剔除未被百度指数收录的股票)

行业分类：中信一级行业

数据区间：2011 年 1 月至 2021 年 9 月

特殊处理：如果某个关键词未被收录，即对其的关注度尚未达到收录标准，那么我们将其剔除。

4.2、关注度因子效性测试

首先我们基于全市场对周度、月度关注度因子进行 IC 检验。根据前面关注度因子构建的逻辑，我们认为关注度越低的股票会有更高的超额收益。

总的来看，这两个因子的都取得了较好的 IC 均值结果，周度和月度分别为 6.51% 和 8.57%，且统计上也十分显著。

图表 8、不同周期下构建的关注度因子的 IC 检验

关注度因子分类	平均值	标准差	最小值	最大值	IC_IR	t统计量
周度关注度因子	6.51%	12.58%	-47.37%	38.23%	0.52	12.14
月度关注度因子	8.57%	14.67%	-43.61%	42.91%	0.58	6.61

资料来源：百度指数，Wind，聚源，兴业证券经济与金融研究院整理

分位数组测试结果显示进一步论证了因子的选股有效性：周度关注度因子各组别严格单调，多头收益达到 24.7%，多空收益率高达 42.9%，夏普比达到 2.51；月度关注度因子多头收益达到 21.6%，多空收益率高达 27.7%，夏普比达到 1.32。

图表 9、周度关注度因子分位数组测试结果

组别	年化收益率	波动率	Sharpe比率	平均换手率	最大回撤率	年化超额收益率	信息比率
top	24.7%	29.1%	0.85	0.51	64.2%	17.6%	2.14
1	20.7%	28.2%	0.73	1.01	55.8%	13.8%	2.62
2	14.9%	28.1%	0.53	1.16	57.1%	8.3%	1.95
3	13.0%	28.2%	0.46	1.22	60.6%	6.6%	1.89
4	9.0%	28.2%	0.32	1.23	66.4%	2.9%	0.93
5	5.2%	27.9%	0.19	1.21	67.2%	-0.8%	(0.29)
6	0.7%	27.6%	0.03	1.15	72.8%	-5.1%	(1.68)
7	-2.5%	27.7%	(0.09)	1.04	72.1%	-8.1%	(2.05)
8	-5.8%	27.1%	(0.22)	0.84	76.4%	-11.5%	(2.07)
bottom	-13.5%	26.7%	(0.51)	0.43	82.4%	-19.2%	(1.84)
市场	6.2%	27.3%	0.23		65.6%		
L-S	42.9%	17.1%	2.51		34.9%		

资料来源：百度指数，Wind，聚源，兴业证券经济与金融研究院整理

图表 10、月度关注度因子分位数组测试结果

组别	年化收益率	波动率	Sharpe比率	平均换手率	最大回撤率	年化超额收益率	信息比率
top	21.6%	32.0%	0.68	0.62	42.3%	12.9%	1.34
1	17.7%	30.2%	0.59	1.16	39.5%	8.9%	1.28
2	16.0%	29.8%	0.54	1.31	42.7%	7.4%	1.45
3	12.5%	29.3%	0.43	1.37	49.1%	4.1%	1.23
4	9.9%	28.9%	0.34	1.39	52.8%	1.6%	0.59
5	6.1%	28.4%	0.22	1.39	57.0%	-2.0%	(0.75)
6	4.9%	28.5%	0.17	1.34	56.9%	-3.1%	(0.96)
7	1.5%	28.3%	0.05	1.24	62.4%	-6.4%	(1.55)
8	-0.6%	28.1%	(0.02)	1.04	64.9%	-8.5%	(1.32)
bottom	-5.8%	27.3%	(0.21)	0.54	70.8%	-14.0%	(1.13)
市场	8.3%	28.4%	0.29		53.1%		
L-S	27.7%	21.1%	1.32		27.7%		

资料来源：百度指数，Wind，聚源，兴业证券经济与金融研究院整理

进一步，我们尝试对该因子进行行业市值中性化处理。相比未做任何处理的关注度因子，行业市值中性化后关注度因子 IC 均值都有所下降，但因子的整体稳健性得到提升，风险调整 IC 以及 T 值都有所改进。

图表 11、行业市值中性化前后因子的 IC 表现

	平均值	标准差	ICIR	T 统计量
中性化周度关注度因子	6.23%	8.62%	0.72	16.96
周度关注度因子	6.51%	12.58%	0.52	12.14
中性化月度关注度因子	7.38%	10.62%	0.70	7.86
月度关注度因子	8.57%	14.67%	0.58	6.61

资料来源：百度指数，Wind，聚源，兴业证券经济与金融研究院整理

从分位数组测试的结果来看，行业市值中性化后因子的稳健性更高：周度因子多空组合年化收益率为 44.8%，波动率为 11.3%，Sharpe 比率达到 3.95，最大回撤率控制在 13.3%；月度因子的多空组合表现稍逊一筹，年化收益率为 20.1%，

波动率为 13.1%，Sharpe 比率达到 1.53，最大回撤率控制在 18.2%。

后续如无特殊强调，我们使用的均是行业市值中性化后的因子。

图表 12、行业市值中性化后周度关注度因子分位数组测试结果

组别	年化收益率	波动率	Sharpe比率	平均换手率	最大回撤率	年化超额收益率	信息比率
top	19.2%	24.7%	0.78	0.50	53.9%	11.3%	1.96
1	15.4%	26.4%	0.58	1.04	55.8%	8.4%	2.30
2	16.2%	27.1%	0.60	1.22	56.1%	9.3%	2.98
3	13.4%	27.1%	0.49	1.30	57.9%	6.7%	2.47
4	12.0%	27.4%	0.44	1.33	54.8%	5.4%	1.96
5	8.8%	28.1%	0.31	1.33	63.3%	2.6%	0.94
6	6.0%	27.9%	0.22	1.29	64.9%	-0.1%	(0.02)
7	0.8%	28.4%	0.03	1.21	73.7%	-4.8%	(1.58)
8	-5.0%	28.7%	(0.17)	1.03	78.4%	-10.2%	(2.73)
bottom	-19.5%	30.1%	(0.65)	0.60	92.0%	-23.8%	(3.40)
市场	6.2%	27.3%	0.23		65.6%		
L-S	44.8%	11.3%	3.95		13.3%		

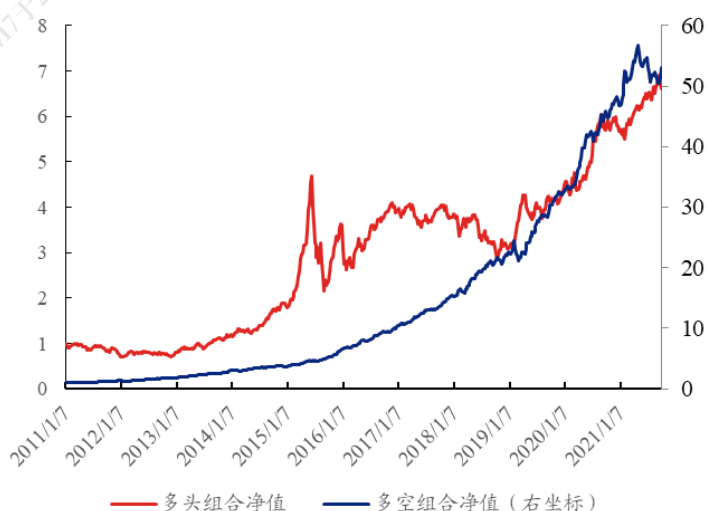
资料来源：百度指数，Wind，聚源，兴业证券经济与金融研究院整理

图表 13、行业市值中性化后月度关注度因子分位数组测试结果

组别	年化收益率	波动率	Sharpe比率	平均换手率	最大回撤率	年化超额收益率	信息比率
top	14.6%	25.5%	0.58	0.53	39.1%	4.8%	0.69
1	13.7%	27.6%	0.50	1.10	42.4%	4.7%	1.13
2	13.5%	28.3%	0.48	1.31	44.1%	4.7%	1.40
3	12.9%	28.4%	0.45	1.40	43.5%	4.2%	1.52
4	11.5%	28.7%	0.40	1.45	48.5%	3.1%	1.25
5	10.5%	29.3%	0.36	1.47	49.9%	2.2%	0.78
6	7.2%	29.3%	0.25	1.44	56.5%	-0.8%	(0.27)
7	4.8%	29.3%	0.16	1.40	58.3%	-3.0%	(0.92)
8	2.1%	29.3%	0.07	1.24	63.2%	-5.6%	(1.42)
bottom	-7.0%	31.1%	(0.22)	0.78	77.3%	-13.7%	(1.83)
市场	8.3%	28.4%	0.29		53.1%		
L-S	20.1%	13.1%	1.53		18.2%		

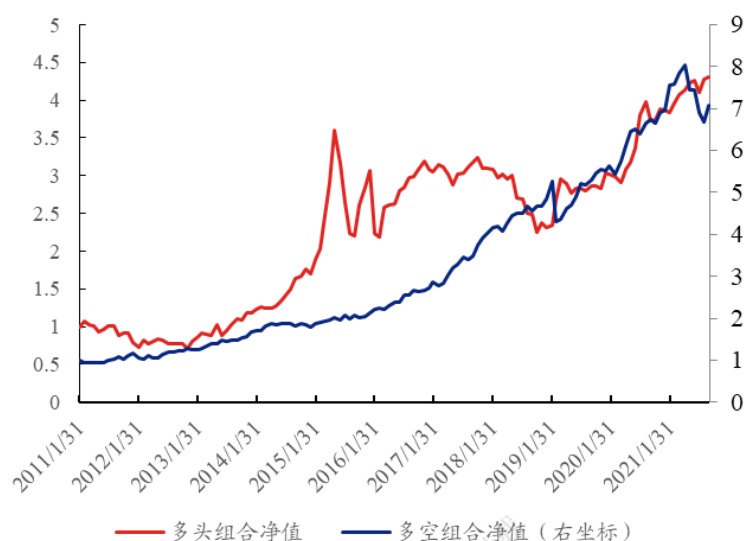
资料来源：百度指数，Wind，聚源，兴业证券经济与金融研究院整理

图表 14、行业市值中性化周度关注度因子多空组合净值及收益率表现



资料来源：百度指数，Wind，聚源，兴业证券经济与金融研究院整理

图表 15、行业市值中性化月度关注度因子多空组合净值及收益率表现



资料来源：百度指数，Wind，聚源，兴业证券经济与金融研究院整理

4.3、关注度因子在宽基指数成分股内有效性分析

既然关注度因子在全市场的表现如此优异，那么将相同的选股策略用于宽基指数下是否仍旧有如此优秀的选股能力呢？我们通过 IC 以及分位数组合测试验证关注度因子在沪深 300、中证 500 指数内的选股能力。

测试结果显示：相比全市场，关注度因子在宽基指数范围内表现有所下降。但无论是沪深 300 内亦或是中证 500 内，关注度因子都呈现出较强的选股能力。

图表 16、宽基成分股内因子的 IC 表现

	股票池	平均值	标准差	ICIR	T 统计量
周度关注度因子	沪深 300	2.85%	10.20%	0.28	6.55
周度关注度因子	中证 500	4.65%	11.05%	0.42	9.86
周度关注度因子	全市场	6.23%	8.62%	0.72	16.96
月度关注度因子	沪深 300	4.38%	10.17%	0.43	4.88
月度关注度因子	中证 500	5.64%	11.89%	0.47	5.37
月度关注度因子	全市场	7.38%	10.62%	0.70	7.86

资料来源：百度指数，Wind，聚源，兴业证券经济与金融研究院整理

图表 17、周度关注度因子在宽基指数下分位数组合测试结果

组别	年化收益率	波动率	Sharpe比率	平均换手率	最大回撤率	年化超额收益率	信息比率	宽基指数
top	11.3%	21.1%	0.54	0.38	37.7%	6.8%	0.86	沪深300
1	8.8%	21.9%	0.40	0.82	38.8%	4.7%	0.74	沪深300
2	6.6%	21.9%	0.30	1.00	42.9%	2.7%	0.48	沪深300
3	3.9%	22.4%	0.18	1.10	53.9%	0.2%	0.03	沪深300
4	2.9%	23.8%	0.12	1.14	63.3%	-0.5%	(0.09)	沪深300
5	5.8%	23.5%	0.25	1.14	44.3%	2.2%	0.41	沪深300
6	-0.5%	23.8%	(0.02)	1.09	55.4%	-3.8%	(0.66)	沪深300
7	-0.6%	24.5%	(0.02)	1.00	60.4%	-3.7%	(0.60)	沪深300
8	-2.4%	24.9%	(0.09)	0.79	65.9%	-5.4%	(0.80)	沪深300
bottom	-1.7%	25.5%	(0.07)	0.39	68.6%	-4.7%	(0.60)	沪深300
市场	3.5%	22.5%	0.16		53.0%			沪深300
L-S	11.1%	13.5%	0.82		21.0%			沪深300
top	14.3%	24.3%	0.59	0.43	50.2%	9.9%	1.04	中证500
1	9.7%	25.7%	0.38	0.93	54.9%	6.0%	0.80	中证500
2	7.9%	25.7%	0.31	1.13	59.3%	4.3%	0.59	中证500
3	8.2%	26.0%	0.32	1.23	56.6%	4.7%	0.71	中证500
4	4.7%	27.4%	0.17	1.28	66.0%	1.8%	0.34	中证500
5	7.3%	26.9%	0.27	1.28	60.4%	4.0%	0.56	中证500
6	5.0%	27.4%	0.18	1.25	62.6%	1.9%	0.27	中证500
7	3.2%	27.8%	0.12	1.18	67.7%	0.3%	0.04	中证500
8	-1.6%	28.6%	(0.06)	1.01	74.7%	-4.2%	(0.52)	中证500
bottom	-13.4%	30.2%	(0.45)	0.56	84.4%	-15.5%	(1.55)	中证500
市场	2.9%	26.8%	0.11		69.4%			中证500
L-S	28.5%	14.6%	1.95		23.1%			中证500

资料来源：百度指数，Wind，聚源，兴业证券经济与金融研究院整理

图表 18、月度关注度因子在宽基指数下分位数组合测试结果

组别	年化收益率	波动率	Sharpe比率	平均换手率	最大回撤率	年化超额收益率	信息比率	宽基指数
top	12.4%	21.0%	0.59	0.50	32.2%	6.6%	0.81	沪深300
1	4.6%	21.1%	0.22	0.97	34.6%	-0.7%	(0.12)	沪深300
2	8.1%	22.9%	0.35	1.17	41.2%	3.0%	0.50	沪深300
3	5.8%	23.9%	0.24	1.25	41.7%	1.2%	0.20	沪深300
4	4.3%	24.8%	0.17	1.30	54.5%	0.0%	(0.01)	沪深300
5	3.2%	24.5%	0.13	1.31	43.7%	-1.2%	(0.22)	沪深300
6	1.7%	26.5%	0.06	1.27	56.5%	-2.2%	(0.34)	沪深300
7	0.6%	25.0%	0.02	1.19	52.0%	-3.6%	(0.64)	沪深300
8	2.4%	25.9%	0.09	0.97	46.1%	-1.7%	(0.27)	沪深300
bottom	0.9%	27.1%	0.03	0.48	62.4%	-3.0%	(0.34)	沪深300
市场	4.6%	23.5%	0.19		44.8%			沪深300
L-S	8.7%	14.6%	0.59		23.6%			沪深300
top	12.3%	24.8%	0.50	0.54	43.5%	4.9%	0.56	中证500
1	9.6%	26.3%	0.36	1.07	43.0%	2.9%	0.47	中证500
2	8.8%	26.2%	0.33	1.26	48.3%	2.2%	0.46	中证500
3	6.3%	26.9%	0.23	1.36	53.0%	0.1%	0.02	中证500
4	8.8%	27.0%	0.33	1.42	45.2%	2.5%	0.52	中证500
5	6.0%	27.5%	0.22	1.42	52.7%	0.0%	(0.01)	中证500
6	6.1%	27.1%	0.22	1.42	52.2%	-0.1%	(0.02)	中证500
7	6.0%	28.3%	0.21	1.35	54.5%	0.2%	0.03	中证500
8	0.7%	29.3%	0.02	1.20	65.5%	-4.7%	(0.75)	中证500
bottom	-4.2%	30.5%	(0.14)	0.72	74.2%	-9.1%	(1.11)	中证500
市场	6.1%	26.8%	0.23		52.9%			中证500
L-S	14.1%	15.0%	0.94		23.3%			中证500

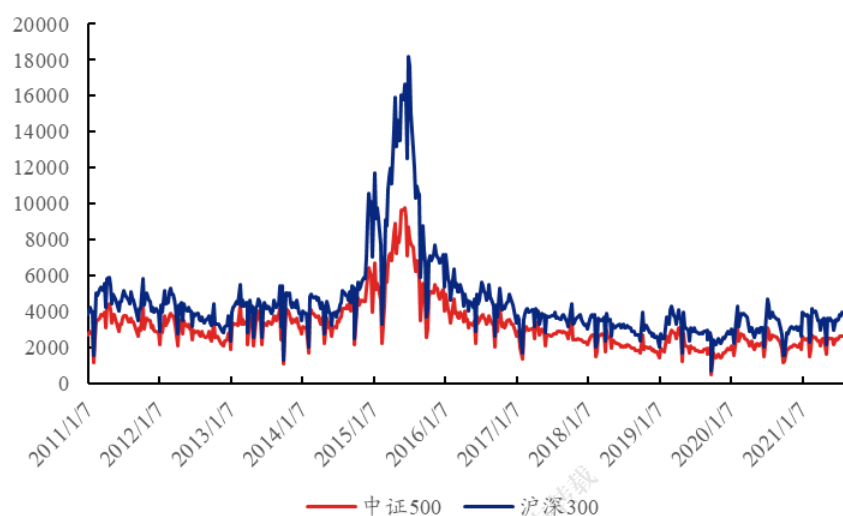
资料来源：百度指数，Wind，聚源，兴业证券经济与金融研究院整理

显然，在宽基指数内关注度因子的表现并没有在全市场那么优异。原因可能是因为：沪深 300 和中证 500 的成分股本身就是市场上规模前 300 以及规模 300-800 之间的股票，相应地也受到市场参与者更多的关注。这些股票的信息流动

请务必阅读正文之后的信息披露和重要声明

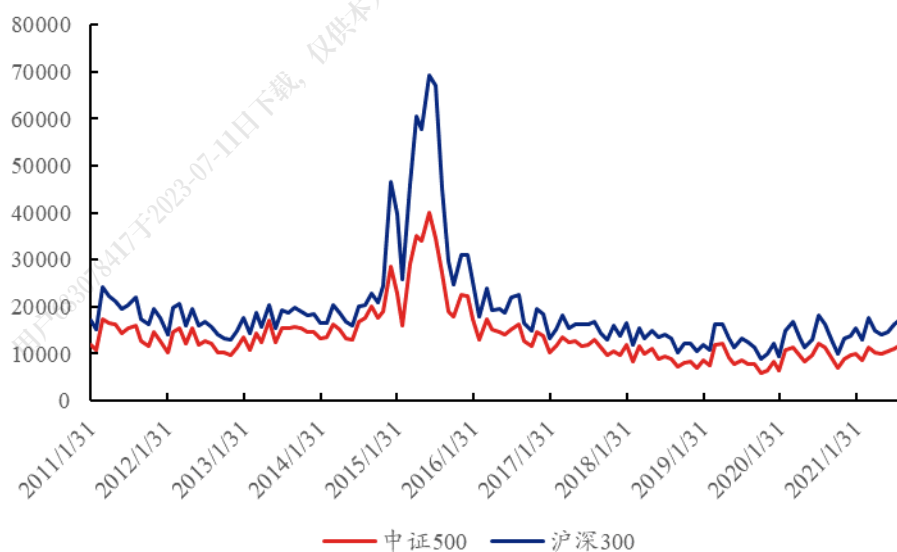
更快，整体 Alpha 效果会更弱。

图表 19、宽基指数下周度关注度的平均水平



资料来源：百度指数，Wind，聚源，兴业证券经济与金融研究院整理

图表 20、宽基指数下月度关注度的平均水平



资料来源：百度指数，Wind，聚源，兴业证券经济与金融研究院整理

4.4、关注度因子真的有效吗？

前面通过大量的篇幅论证了基于网络热度构建的关注度因子的选股有效性。但该因子真的具有较好的选股效果吗？我们通过相关性分析以及 Fama-Mac 测试来进一步论证。

与 Barra 十个因子的相关性分析显示：除 Barra 流动性因子外，月度关注度因

请务必阅读正文之后的信息披露和重要声明

子以及周度关注度因子与 Barra 其他风格因子相关性很低。而剥离了流动性之后的周度以及月度关注度因子的选股效果并没有降低。因此这种相关性并不影响关注度因子对收益率解释的特异性。

图表 21、关注度因子与传统因子相关性表现

	月度关注度因子	周度关注度因子
市值	0.04	0.04
贝塔	0.14	0.13
动量	-0.01	-0.03
残差波动	0.25	0.23
非线性市值	0.01	0.01
账面市值比	-0.02	-0.02
流动性	0.50	0.47
盈利预期	0.08	0.08
成长	0.04	0.04
杠杆	0.14	0.13

资料来源：百度指数，Wind，聚源，兴业证券经济与金融研究院整理

图表 22、剥离流动性后的因子表现

	平均值	标准差	最小值	最大值	ICIR	T 统计量
月度关注度因子	0.074	0.09	-0.31	0.27	0.78	8.82
周度关注度因子	0.063	0.08	-0.25	0.28	0.75	17.57

资料来源：百度指数，Wind，聚源，兴业证券经济与金融研究院整理

进一步 Fama-Mac 回归测试也验证关注度因子对收益率解释的显著力度。

图表 23、月度关注度因子与 Barra 十个风格因子 Fama-MacBeth 回测结果

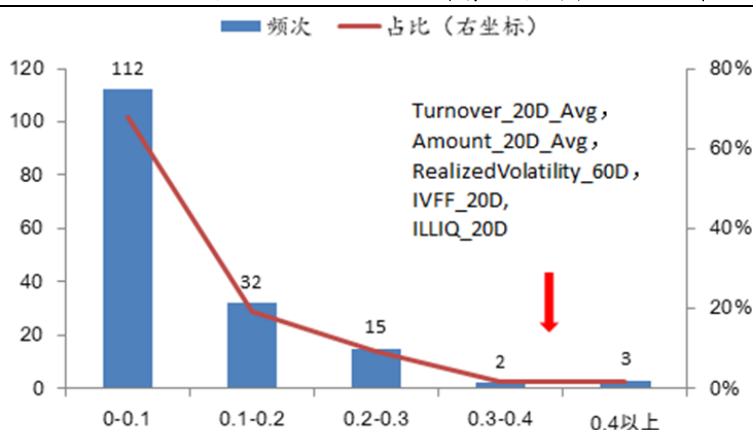
因子	T 统计量-综合回归	T 统计量-单独回归
月度关注度	-3.00	-5.46
Size	-2.12	-2.04
Beta	1.95	-0.25
Momentum	2.49	-0.37
Residual Volatility	-1.63	-2.13
Nonlinear Size	-6.49	-5.45
BookToPrice	0.74	0.80
Liquidity	-2.19	-2.88
Earnings Yield	3.11	1.12
Growth	1.55	-0.08
Leverage	-1.73	-0.89

资料来源：百度指数，Wind，聚源，兴业证券经济与金融研究院整理

但通过暴力的方式测试关注度因子与兴证金工底层所有因子（兴证金工底层共计 165 个因子）的相关性，结果发现关注度指标与部分因子相关性相对较高。图表 24 展示了以月度关注度为例进行的测试结果，不难发现其与 5 个因子相关性达到 0.3 以上。这 5 个因子均是常见的衡量市场流动性和市场结构的指标。

如果将这些因子和 Barra 十个风格因子放到一起进行回归，结果发现其特异性明显减弱，对收益率的解释力度所剩无几。

图表 24、月度关注度因子与兴证金工底层所有因子的相关性分析



资料来源：百度指数，Wind，聚源，兴业证券经济与金融研究院整理

图表 25、月度关注度因子与其他因子的 Fama-MacBeth 回测结果

因子	T 统计量-综合回归	T 统计量-单独回归
月度关注度因子	-1.35	-5.46
Size	-1.22	-2.04
Beta	0.37	-0.25
Momentum	2.97	-0.37
ResidualVolatility	-1.17	-2.13
NonlinearSize	-4.60	-5.45
BookToPrice	0.23	0.80
Liquidity	1.09	-2.88
EarningsYield	3.25	1.12
Growth	1.07	-0.08
Leverage	-1.04	-0.89
Turnover_20D_Avg	-5.00	-3.14
IVFF_20D	-3.75	-4.55
Amount_20D_Avg	-2.48	-3.68
ILLIQ_20D	3.30	2.02
RealizedVolatility_60D	1.97	-1.73

资料来源：百度指数，Wind，聚源，兴业证券经济与金融研究院整理

5、结语

在本篇文章中，我们摒弃传统的量价数据因子挖掘，从另类数据的角度切入，对基于另类数据生成的关注度因子是否具有 Alpha 能力进行了详细的论证和探索。我们用国内市场搜索引擎龙头公司百度提供的另类数据——百度搜索指数作为市场关注度的代理变量，从而构建周度、月度的关注度因子进行选股，并对不同频度的关注度因子的特异性和有效性进行了相应的分析。测试结果显示该因子虽然有效，但特异性较为一般。

风险提示：本报告模型及结论全部基于对历史数据的分析，当市场环境变化时，存在模型失效风险。

分析师声明

本人具有中国证券业协会授予的证券投资咨询执业资格并注册为证券分析师，以勤勉的职业态度，独立、客观地出具本报告。本报告清晰准确地反映了本人的研究观点。本人不曾因，不因，也将不会因本报告中的具体推荐意见或观点而直接或间接收到任何形式的补偿。

投资评级说明

投资建议的评级标准	类别	评级	说明
报告中投资建议所涉及的评级分为股票评级和行业评级(另有说明的除外)。评级标准为报告发布日后的12个月内公司股价(或行业指数)相对同期相关证券市场代表性指数的涨跌幅。其中：A股市场以上证综指或深圳成指为基准，香港市场以恒生指数为基准；美国市场以标普500或纳斯达克综合指数为基准。	股票评级	买入	相对同期相关证券市场代表性指数涨幅大于15%
		审慎增持	相对同期相关证券市场代表性指数涨幅在5%~15%之间
		中性	相对同期相关证券市场代表性指数涨幅在-5%~5%之间
		减持	相对同期相关证券市场代表性指数涨幅小于-5%
		无评级	由于我们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，致使我们无法给出明确的投资评级
	行业评级	推荐	相对表现优于同期相关证券市场代表性指数
		中性	相对表现与同期相关证券市场代表性指数持平
		回避	相对表现弱于同期相关证券市场代表性指数

信息披露

本公司在知晓的范围内履行信息披露义务。客户可登录 www.xyzq.com.cn 内幕交易防控栏内查询静默期安排和关联公司持股情况。

使用本研究报告的风险提示及法律声明

兴业证券股份有限公司经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。

本报告仅供兴业证券股份有限公司（以下简称“本公司”）的客户使用，本公司不会因接收人收到本报告而视其为客户。本报告中的信息、意见等均仅供客户参考，不构成所述证券买卖的出价或征价邀请或要约。该等信息、意见并未考虑到获取本报告人员的具体投资目的、财务状况以及特定需求，在任何时候均不构成对任何人的个人推荐。客户应当对本报告中的信息和意见进行独立评估，并应同时考量各自的投资目的、财务状况和特定需求，必要时就法律、商业、财务、税收等方面咨询专家的意见。对依据或者使用本报告所造成的一切后果，本公司及/或其关联人员均不承担任何法律责任。

本报告所载资料的来源被认为是可靠的，但本公司不保证其准确性或完整性，也不保证所包含的信息和建议不会发生任何变更。本公司并不对使用本报告所包含的材料产生的任何直接或间接损失或与此相关的其他任何损失承担任何责任。

本报告所载的资料、意见及推测仅反映本公司于发布本报告当日的判断，本报告所指的证券或投资标的的价格、价值及投资收入可升可跌，过往表现不应作为日后的表现依据；在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告；本公司不保证本报告所含信息保持在最新状态。同时，本公司对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。

除非另行说明，本报告中所引用的关于业绩的数据代表过往表现。过往的业绩表现亦不应作为日后回报的预示。我们不承诺也不保证，任何所预示的回报会得以实现。分析中所做的回报预测可能是基于相应的假设。任何假设的变化可能会显著地影响所预测的回报。

本公司的销售人员、交易人员以及其他专业人士可能会依据不同假设和标准、采用不同的分析方法而口头或书面发表与本报告意见及建议不一致的市场评论和/或交易观点。本公司没有将此意见及建议向报告所有接收者进行更新的义务。本公司的资产管理部门、自营部门以及其他投资业务部门可能独立做出与本报告中的意见或建议不一致的投资决策。

本报告并非针对或意图发送予或为任何就发送、发布、可得到或使用此报告而使兴业证券股份有限公司及其关联子公司等违反当地的法律或法规或可致使兴业证券股份有限公司受制于相关法律或法规的任何地区、国家或其他管辖区域的公民或居民，包括但不限于美国及美国公民（1934年美国《证券交易所》第15a-6条例定义为本「主要美国机构投资者」除外）。

本报告的版权归本公司所有。本公司对本报告保留一切权利。除非另有书面显示，否则本报告中的所有材料的版权均属本公司。未经本公司事先书面授权，本报告的任何部分均不得以任何方式制作任何形式的拷贝、复印件或复制品，或再次分发给任何其他人，或以任何侵犯本公司版权的其他方式使用。未经授权的转载，本公司不承担任何转载责任。

特别声明

在法律许可的情况下，兴业证券股份有限公司可能会持有本报告中提及公司所发行的证券头寸并进行交易，也可能为这些公司提供或争取提供投资银行业务服务。因此，投资者应当考虑到兴业证券股份有限公司及/或其相关人员可能存在影响本报告观点客观性的潜在利益冲突。投资者请勿将本报告视为投资或其他决定的唯一信赖依据。

兴业证券研究

上 海	北 京	深 圳
地址：上海浦东新区长柳路36号兴业证券大厦15层	地址：北京西城区锦什坊街35号北楼601-605	地址：深圳市福田区皇岗路5001号深业上城T2座52楼
邮编：200135	邮编：100033	邮编：518035
邮箱：research@xyzq.com.cn	邮箱：research@xyzq.com.cn	邮箱：research@xyzq.com.cn