

2023 年 06 月 18 日

## 股东户数因子的深挖和改进

——开源量化评论（76）

### 金融工程研究团队

**魏建榕（首席分析师）**

证书编号：S0790519120001

**张翔（分析师）**

证书编号：S0790520110001

**傅开波（分析师）**

证书编号：S0790520090003

**高鹏（分析师）**

证书编号：S0790520090002

**苏俊豪（分析师）**

证书编号：S0790522020001

**胡亮勇（分析师）**

证书编号：S0790522030001

**王志豪（分析师）**

证书编号：S0790522070003

**盛少成（分析师）**

证书编号：S0790523060003

**苏良（分析师）**

证书编号：S0790523060004

**何申昊（研究员）**

证书编号：S0790122080094

### 相关研究报告

《因子切割论》-2020.9.17

《高频股东数据的隐含信息量——开源量化评论（13）》-2020.12.31

《扎堆效应的识别：以股东户数变动为例——开源量化评论（66）》-2022.11.22

**魏建榕（分析师）**

weijianrong@kysec.cn

证书编号：S0790519120001

**苏俊豪（分析师）**

sujunhao@kysec.cn

证书编号：S0790522020001

### ● 股东户数因子回归股本与股价后效果显著

我们使用向后填充，按月抽样的方法，得到月度的股东户数因子（HolderNumber，以下简称 HN），原始的 HN 因子 RankIC 为-1.97%，这表明股东户数越多，股票未来预期收益越低，但 HN 因子的表现十分不稳定。

股东户数与股票流通市值存在较高的正相关性，两者的截面秩相关系数均值高达 0.55，市值行业中性化后，HN 因子的 RankIC 下降至-1.48%，RankICIR 为-0.54，稳定性依然欠佳。

股票的流通市值 = 流通股本 × 股价（未复权），如上所述，股东户数与流通市值强正相关，但流通股本和股价与股东户数的关系却截然相反：流通股本与股东户数的秩相关系数均值高达 0.76，比流通市值更显著，而股价与股东户数的相关性却显著为负，秩相关系数均值为-0.50。

和市值行业中性化相比，对股本、股价分别或合并作回归剔除，得到的残差因子在显著性与稳定性上都要更优。

### ● 区间股价振幅越高，股东户数变化率因子效果越好

我们构造最近两期财报间股东户数的季度变化率因子 HNpct，HNpct 因子 RankIC 为-2.02%，RankICIR 为-1.28，RankIC 胜率为 63%。多空收益方面，五分组下 HNpct 因子的年化多空收益为 5.55%，收益波动比为 1.01。HNpct 因子的多头端区分效果不明显，空头端的负向收益更为显著。

股东户数的季度变化率的绝对幅度 abs(HNpct)（即 HNpct 因子的绝对值）同样具有一定负向选股能力，abs(HNpct)因子 RankIC 为-2.15%，RankICIR 为-0.98，RankIC 胜率为 56%。多空收益方面，五分组下 abs(HNpct)因子的年化多空收益为 4.37%，收益波动比仅为 0.53。

我们基于股价区间振幅，对不同振幅下 HNpct 与 abs(HNpct)因子的选股能力进行测试。结果显示，振幅越大，对应的 HNpct 因子与 abs(HNpct)因子的表现越好，其中，最高振幅分组下，HNpct 与 abs(HNpct)因子的 RankIC 分别为-2.64%与-2.39%。

基于以上测试结果，我们进一步构造了 HNpct 与 abs(HNpct)的精选与标准化因子，和对照基准相比，精选因子的 RankIC、多空收益与多头收益变化不大，但 RankICIR、多空收益波动比、多空收益最大回撤均有较好改善；而标准化因子在 RankIC、多空收益与多头收益上有所下降，但因子稳定性与多空收益稳定性提升则更为显著。

● **风险提示：**模型基于历史数据测试，市场未来可能发生重大改变。

## 目 录

1、 股东户数因子：回归股本与股价后效果显著 .....	3
2、 股东户数变化率因子的能力受股价振幅影响 .....	5
3、 风险提示 .....	8

## 图表目录

图 1： 股东户数与股票流通市值强正相关 .....	3
图 2： 回归市值前后 HN 因子表现均不佳 .....	3
图 3： 股本和股价与股东户数的相关性相反 .....	4
图 4： 市值的两个拆分因素对股东户数的影响方向相反 .....	4
图 5： 股东户数在不同中性化条件下的表现：股本、股价、行业中性化时表现最好 .....	4
图 6： HN_price_ind 因子的分组收益单调性较好，年化多空收益为 16.1% .....	5
图 7： HNpct 因子的 RankIC 显著为负 .....	5
图 8： HNpct 因子的空头端负向收益更为显著 .....	5
图 9： abs(HNpct) 同样具备一定的负向选股能力 .....	6
图 10： abs(HNpct) 的多空收益稳定性欠佳 .....	6
图 11： 振幅越高，HNpct_q 因子效果越好 .....	7
图 12： 振幅越高，abs(HNpct_q) 因子效果越好 .....	7
表 1： 股东户数在不同中性化条件下的表现：股本、股价、行业中性化时表现最好 .....	4
表 2： 计算基于股价振幅排序的 HNpct_q、abs(HNpct_q) 因子所需数据的示例 .....	6
表 3： 计算基于股价振幅排序的 HNpct_q、abs(HNpct_q) 因子的步骤 .....	7
表 4： 基于因子切割论的思想构造精选因子与标准化因子 .....	7
表 5： 因子表现：标准化因子 HNpct_diff4、abs_HNpct_diff6 稳定性较高 .....	8

上市公司的股东户数随财报公布，股东户数可以有效反映投资者尤其是个人投资者对公司的关注程度。在个人投资者行为较为活跃的 A 股市场中，股东户数的相关信息可以帮助我们挖掘与理解个人投资者的交易行为与习惯，是不可多得独特 alpha 来源。

在开源金工 2020 年发布的报告《高频股东数据的隐含信息量》中，我们发现，交易所投资者互动平台上披露的高频股东数据是有效的补充信息；而在 2022 年发布的报告《扎堆效应的识别：以股东户数变化为例》中，我们构造了股东户数变动因子 SNC 与人均持股占比变动因子 PCRC，在中证 1000 指数中，因子具有优异的增强表现。本文我们将就股东户数这一数据，展开进一步讨论并构造相应的因子。

本文包含两个相对独立的结论：

- (1) 股东户数因子回归股本与股价后效果显著。
- (2) 区间股价振幅越高，股东户数变化率因子效果越好。

## 1、股东户数因子：回归股本与股价后效果显著

我们使用向后填充，按月抽样的方法，得到月度的股东户数因子 (HolderNumber, 以下简称 HN)，原始的 HN 因子 RankIC 为-1.97%，这表明股东户数越多，股票未来预期收益越低，但 HN 因子的表现十分不稳定，RankICIR 仅为-0.48。

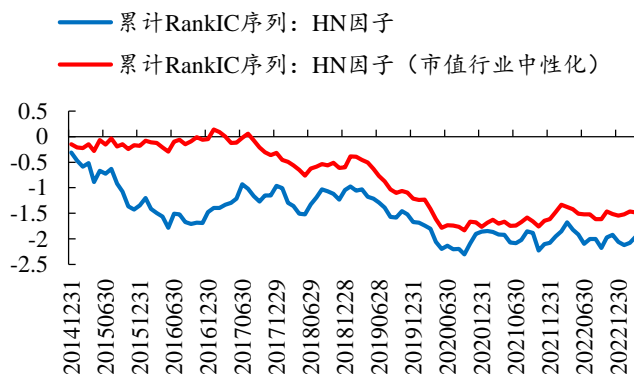
股东户数与股票流通市值存在较高的正相关性，两者的截面秩相关系数均值高达 0.55 (图 1)，市值行业中性化后，HN 因子的 RankIC 下降至-1.48%，RankICIR 为-0.54，稳定性依然欠佳 (图 2)。

图1：股东户数与股票流通市值强正相关



数据来源：Wind、开源证券研究所（数据区间 201412~202303）

图2：回归市值前后 HN 因子表现均不佳



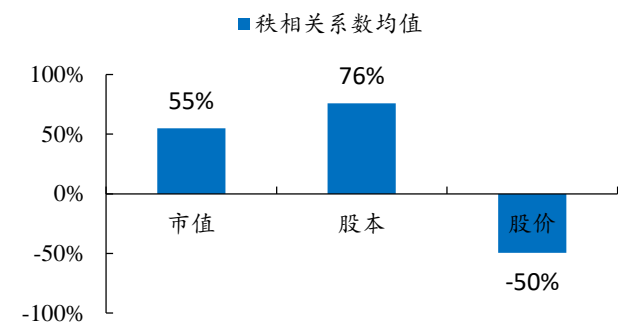
数据来源：Wind、开源证券研究所（数据区间 201412~202303）

股票的流通市值 = 流通股本 × 股价（未复权），如上所述，股东户数与流通市值强正相关，但流通股本和股价与股东户数的关系却截然相反：如图 3 所示，流通股本与股东户数的秩相关系数均值高达 0.76，比流通市值更显著，而股价与股东户数的相关性却显著为负，秩相关系数均值为-0.50。

流通股本越多，可分配给到的股东数量自然也越多；而未复权股价则可视作该股票的投资门槛，门槛越低，资金量较小的个人投资者投资难度越低，则股东户数

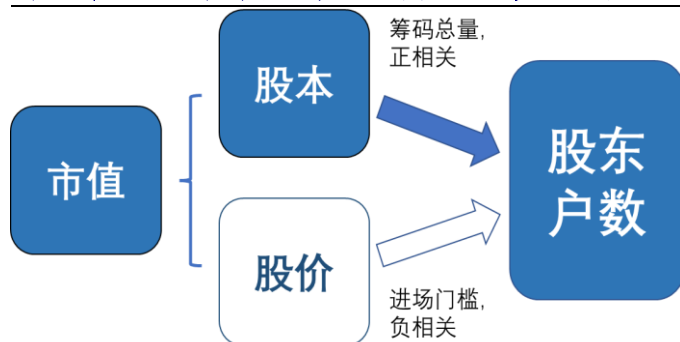
也更多，股本与股价对股东户数的影响如图 4 所示。

图3：股本和股价与股东户数的相关性相反



数据来源：Wind、开源证券研究所

图4：市值的两个拆分因素对股东户数的影响方向相反



资料来源：开源证券研究所

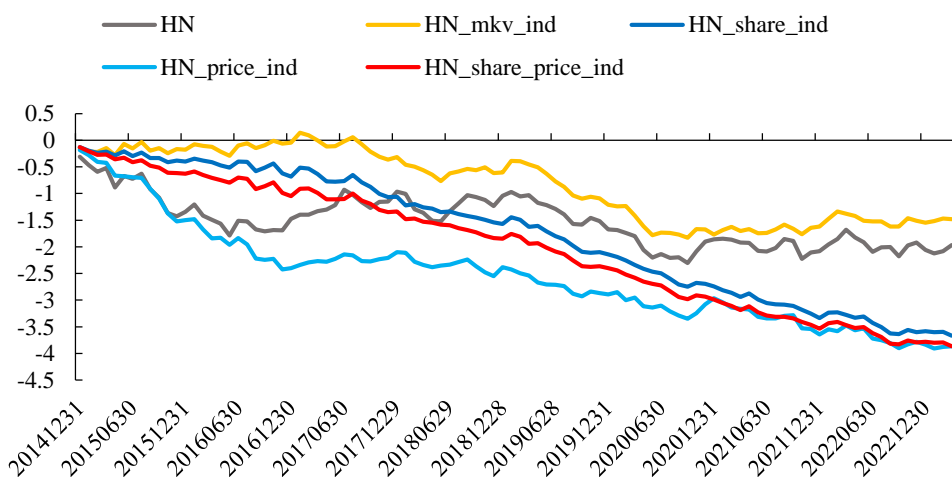
基于以上分析，我们认为应当把股本与股价对股东户数的影响分开看待，作为对照，我们测试了 HN 因子在不同中性化条件下的表现，如表 1、图 5 所示。在中性化前后，HN 因子都表现出较为显著的负向选股能力，这表明，股东户数越多，散户拥挤度越高，股票预期收益越差。

表1：股东户数在不同中性化条件下的表现：股本、股价、行业中性化时表现最好

因子简称	处理方法	RankIC	RankICIR
HN	-	-1.97%	-0.48
HN_mkv_ind	市值、行业中性化	-1.48%	-0.54
HN_share_ind	股本、行业中性化	-3.67%	-1.87
HN_price_ind	股价、行业中性化	-3.87%	-1.40
HN_share_price_ind	股本、股价、行业中性化	-3.86%	-2.12

数据来源：Wind、开源证券研究所（数据区间 201412~202303）

图5：股东户数在不同中性化条件下的表现：股本、股价、行业中性化时表现最好

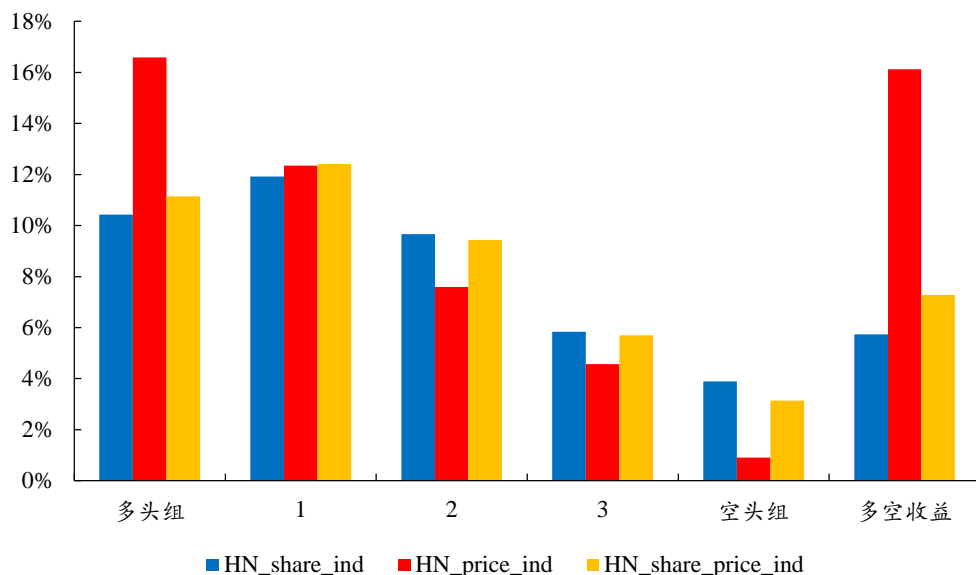


数据来源：Wind、开源证券研究所（数据区间 201412~202303）

和市值行业中性化相比，对股本、股价分别或合并作回归剔除，得到的残差因子在显著性与稳定性上都要更优，不过，HN\_share\_ind、HN\_price\_ind、HN\_share\_price\_ind 三个因子中，仅 HN\_price\_ind 因子的分组收益单调性较好，其多空

收益也更高（图6）。

**图6: HN\_price\_ind 因子的分组收益单调性较好，年化多空收益为 16.1%**



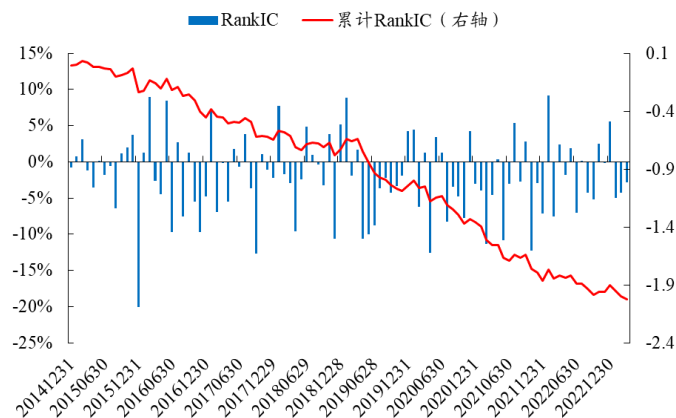
数据来源：Wind、开源证券研究所（数据区间 201412~202303）

## 2、股东户数变化率因子的能力受股价振幅影响

本节我们将基于开源金工“因子切割论”（详见开源金工报告《因子切割论》）的思想，探究股东户数的变化率因子在不同股价振幅下的表现差异，并据此构造表现更稳健的股东户数变化率因子。

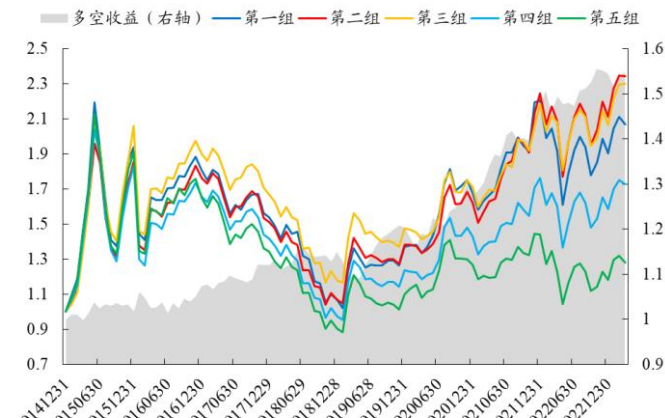
我们构造最近两期财报间股东户数的季度变化率因子 HNpct，HNpct 因子 RankIC 为-2.02%，RankICIR 为-1.28，RankIC 胜率为 63%。多空收益方面，五分组下 HNpct 因子的年化多空收益为 5.55%，收益波动比为 1.01。HNpct 因子的多头端区分效果不明显，空头端的负向收益更为显著（图8）。

**图7: HNpct 因子的 RankIC 显著为负**



数据来源：Wind、开源证券研究所（数据区间 201412~202303）

**图8: HNpct 因子的空头端负向收益更为显著**



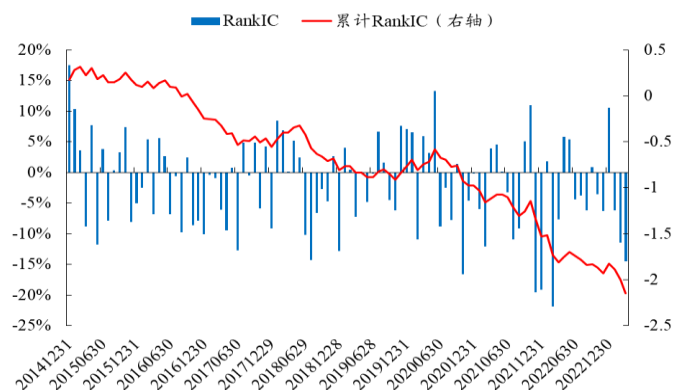
数据来源：Wind、开源证券研究所（数据区间 201412~202303）

股东户数的季度变化率的绝对幅度  $\text{abs}(\text{HNpct})$ （即 HNpct 因子的绝对值）同样



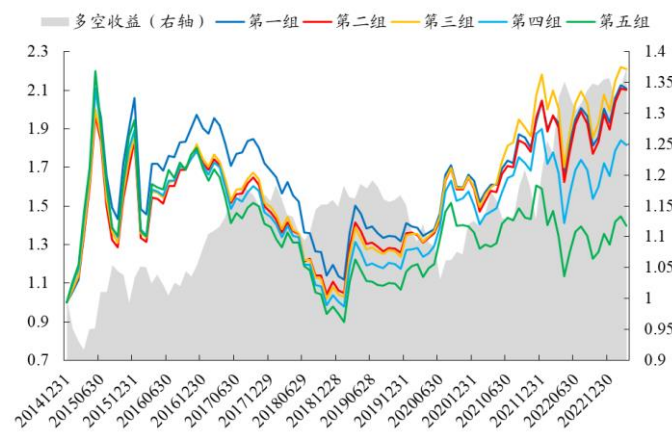
具有一定负向选股能力，abs(HNpct)因子 RankIC 为-2.15%，RankICIR 为-0.98，RankIC 胜率为 56%。多空收益方面，五分组下 abs(HNpct)因子的年化多空收益为 4.37%，收益波动比仅为 0.53。

图9: abs(HNpct)同样具备一定的负向选股能力



数据来源：Wind、开源证券研究所（数据区间 201412~202303）

图10: abs(HNpct)的多空收益稳定性欠佳



数据来源：Wind、开源证券研究所（数据区间 201412~202303）

我们回看过去两年共八个季度的财报股东户数 HNpct 因子与 abs(HNpct)因子，并计算其同期的股价振幅，这里，我们把股价对齐到股东户数的实际截止日而非公告日。如表 2 所示，我们列出了某股票 2023Q1 对应因子值向前回看所需数据，计算得到相应的因子值后，我们再把因子值对齐到实际公告日（即 20230429）。

表2: 计算基于股价振幅排序的 HNpct\_q、abs(HNpct\_q)因子所需数据的示例

区间起始日	区间结束日	区间结束日公告时间	区间结束日股东户数	区间股价振幅	HNpct	abs(HNpct)
20210331	20210630	20210831	1410821	22.09%	9.13%	9.13%
20210630	20210930	20211029	1569497	30.02%	11.25%	11.25%
20210930	20211231	20220331	1623582	10.17%	3.45%	3.45%
20211231	20220331	20220428	1608385	25.73%	-0.94%	0.94%
20220331	20220630	20220830	1606920	17.44%	-0.09%	0.09%
20220630	20220930	20221031	1571609	22.22%	-2.20%	2.20%
20220930	20221230	20230404	1512990	12.88%	-3.73%	3.73%
20221230	<b>20230331</b>	<b>20230429</b>	1332076	32.84%	-11.96%	11.96%

数据来源：Wind、开源证券研究所（数据区间 201412~202303）

在获取过去八个季度的股价振幅与相应的 HNpct、abs(HNpct)因子后，对每个股票，我们按照如下方法，计算基于股价振幅排序的 HNpct\_q、abs\_HNpct\_q 因子。这里，HNpct\_q1 因子指的是，过去八个季度中，股价振幅最小的季度对应的股东户数变化率。

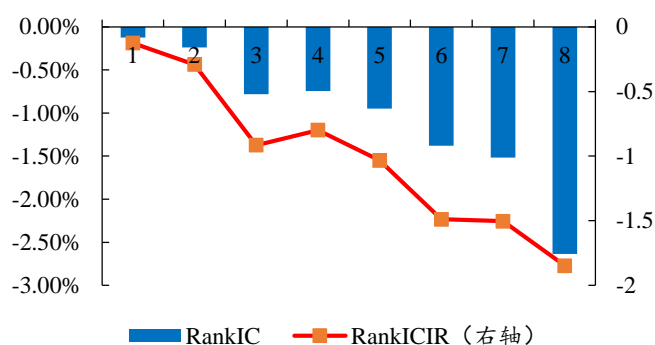
表3：计算基于股价振幅排序的 HNpct\_q、abs(HNpct\_q)因子的步骤

步骤 1	对每个股票，回看其最近过去 8 个季度的股价振幅、HNpct、abs(HNpct) 因子
步骤 2	把 8 个 HNpct、abs(HNpct)因子按照其同期振幅进行排序
步骤 3	按照振幅从小到大，依次记每个季度的 HNpct、abs(HNpct)因子为 HNpct_q1,...,HNpct_q8;abs(HNpct_q1),...,abs(HNpct_q8)

资料来源：开源证券研究所

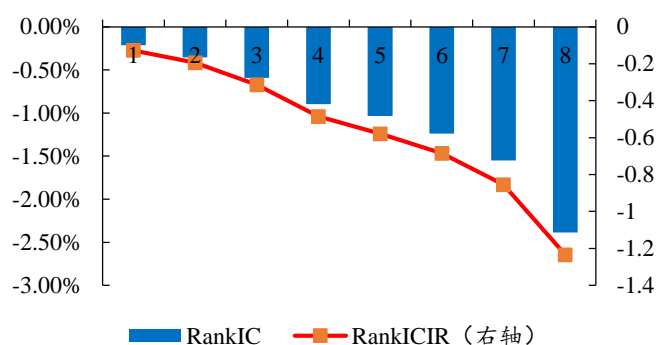
我们测试 HNpct\_q1~ HNpct\_q8、abs(HNpct\_q1)~ abs(HNpct\_q8)因子的表现，结果如图 11、图 12 所示：振幅分组越大，对应的 HNpct 因子与 abs(HNpct)因子的表现越好，其中 HNpct\_q8 与 abs(HNpct\_q8)的 RankIC 分别为-2.64%与-2.39%。

图11：振幅越高，HNpct\_q 因子效果越好



数据来源：Wind、开源证券研究所（数据区间 201412~202303）

图12：振幅越高，abs(HNpct\_q)因子效果越好



数据来源：Wind、开源证券研究所（数据区间 201412~202303）

基于以上切割的结果，我们按照“精选”与“标准化”的思路构造因子，这里，我们简记：

$$top6(HNpct) = (HNpct\_q8 + HNpct\_q7 + HNpct\_q6 + HNpct\_q5 + HNpct\_q4 + HNpct\_q3)$$

$$bottom4(HNpct) = HNpct\_q1 + HNpct\_q2 + HNpct\_q3 + HNpct\_q4$$

对 abs(HNpct)因子，也有同样的简记方法。作为对照，我们同时还构造了使用两年间全部八个季报数据的 HNpct 均值因子与 abs(HNpct)均值因子，因子构造方法与测试结果如表 4、表 5 所示：

表4：基于因子切割论的思想构造精选因子与标准化因子

	构造公式
HNpct_top4	top4(HNpct)
HNpct_top6	top6(HNpct)
HNpct_diff4	top4(HNpct)-bottom4(HNpct)
HNpct_diff6	top6(HNpct)-bottom2(HNpct)
HNpct_avg	top4(HNpct)+bottom4(HNpct)
abs_HNpct_top4	top4(abs(HNpct))
abs_HNpct_top6	top6(abs(HNpct))
abs_HNpct_diff4	top4(abs(HNpct))-bottom4(abs(HNpct))
abs_HNpct_diff6	top6(abs(HNpct))-bottom2(abs(HNpct))
abs_HNpct_avg	top4(abs(HNpct))+bottom4(abs(HNpct))

资料来源：开源证券研究所

表5：因子表现：标准化因子 HNpct\_diff4、abs\_HNpct\_diff6 稳定性较高

	RankIC	RankICIR	多空收益	多空收益波动比	多空收益最大回撤	多头收益
HNpct_top4	-3.85%	-1.86	11.86%	1.70	6.02%	13.67%
HNpct_top6	<b>-4.08%</b>	-1.82	<b>13.03%</b>	1.73	6.63%	<b>13.97%</b>
HNpct_diff4	-2.79%	<b>-1.89</b>	8.77%	<b>1.76</b>	<b>4.20%</b>	11.91%
HNpct_diff6	-2.82%	-1.72	7.09%	1.25	9.34%	10.69%
HNpct_avg	-4.02%	-1.69	12.53%	1.55	7.02%	13.94%
abs_HNpct_top4	-3.21%	-1.16	7.26%	0.71	15.57%	10.91%
abs_HNpct_top6	<b>-3.21%</b>	-1.07	7.13%	0.63	17.66%	11.05%
abs_HNpct_diff4	-2.69%	-1.80	7.36%	1.46	4.81%	10.42%
abs_HNpct_diff6	-2.98%	<b>-2.04</b>	<b>9.08%</b>	<b>1.80</b>	<b>4.65%</b>	<b>11.37%</b>
abs_HNpct_avg	-3.10%	-1.00	6.65%	0.55	19.28%	11.19%

数据来源：Wind、开源证券研究所（数据区间 201412~202303）

可以看到，和作为对照基准的 HNpct\_avg 因子与 abs\_HNpct\_avg 因子相比，精选因子 HNpct\_top 和 abs\_HNpct\_top 的 RankIC、多空收益与多头收益变化不大，但 RankICIR、多空收益波动比、多空收益最大回撤均有较好改善；而标准化因子 HNpct\_diff、abs\_HNpct\_diff 在 RankIC、多空收益与多头收益上有所下降，但因子稳定性与多空收益稳定性提升则更为显著。

### 3、风险提示

模型基于历史数据测试，市场未来可能发生重大改变。



## 特别声明

《证券期货投资者适当性管理办法》、《证券经营机构投资者适当性管理实施指引（试行）》已于2017年7月1日起正式实施。根据上述规定，开源证券评定此研报的风险等级为R3（中风险），因此通过公共平台推送的研报其适用的投资者类别仅限定为专业投资者及风险承受能力为C3、C4、C5的普通投资者。若您并非专业投资者及风险承受能力为C3、C4、C5的普通投资者，请取消阅读，请勿收藏、接收或使用本研报中的任何信息。因此受限于访问权限的设置，若给您造成不便，烦请见谅！感谢您给予的理解与配合。

## 分析师承诺

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、竞争性因素以及开源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的具体的推荐意见或观点有直接或间接的联系。

## 股票投资评级说明

	评级	说明
证券评级	买入（Buy）	预计相对强于市场表现 20%以上；
	增持（outperform）	预计相对强于市场表现 5%~20%；
	中性（Neutral）	预计相对市场表现在-5%~+5%之间波动；
	减持	预计相对弱于市场表现 5%以下。
行业评级	看好（overweight）	预计行业超越整体市场表现；
	中性（Neutral）	预计行业与整体市场表现基本持平；
	看淡	预计行业弱于整体市场表现。

备注：评级标准为以报告日后的 6~12 个月内，证券相对于市场基准指数的涨跌幅表现，其中 A 股基准指数为沪深 300 指数、港股基准指数为恒生指数、新三板基准指数为三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的）、美股基准指数为标普 500 或纳斯达克综合指数。我们在此提醒您，不同证券研究机构采用不同的评级术语及评级标准。我们采用的是相对评级体系，表示投资的相对比重建议；投资者买入或者卖出证券的决定取决于个人的实际情况，比如当前的持仓结构以及其他需要考虑的因素。投资者应阅读整篇报告，以获取比较完整的观点与信息，不应仅仅依靠投资评级来推断结论。

## 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。

## 法律声明

开源证券股份有限公司是经中国证监会批准设立的证券经营机构，已具备证券投资咨询业务资格。

本报告仅供开源证券股份有限公司（以下简称“本公司”）的机构或个人客户（以下简称“客户”）使用。本公司不会因接收人收到本报告而视其为客户。本报告是发送给开源证券客户的，属于商业秘密材料，只有开源证券客户才能参考或使用，如接收人并非开源证券客户，请及时退回并删除。

本报告是基于本公司认为可靠的已公开信息，但本公司不保证该等信息的准确性或完整性。本报告所载的资料、工具、意见及推测只提供给客户作参考之用，并非作为或被视为出售或购买证券或其他金融工具的邀请或向人做出邀请。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、价值及投资收入可能会波动。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。客户应当考虑到本公司可能存在可能影响本报告客观性的利益冲突，不应视本报告为做出投资决策的唯一因素。本报告中所指的投资及服务可能不适合个别客户，不构成客户私人咨询建议。本公司未确保本报告充分考虑到个别客户特殊的投资目标、财务状况或需要。本公司建议客户应考虑本报告的任何意见或建议是否符合其特定状况，以及（若有必要）咨询独立投资顾问。在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议。在任何情况下，本公司不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任。若本报告的接收人非本公司的客户，应在基于本报告做出任何投资决定或就本报告要求任何解释前咨询独立投资顾问。

本报告可能附带其它网站的地址或超级链接，对于可能涉及的开源证券网站以外的地址或超级链接，开源证券不对其内容负责。本报告提供这些地址或超级链接的目的纯粹是为了客户使用方便，链接网站的内容不构成本报告的任何部分，客户需自行承担浏览这些网站的费用或风险。

开源证券在法律允许的情况下可参与、投资或持有本报告涉及的证券或进行证券交易，或向本报告涉及的公司提供或争取提供包括投资银行业务在内的服务或业务支持。开源证券可能与本报告涉及的公司之间存在业务关系，并无需事先或在获得业务关系后通知客户。

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。所有本报告中使用的商标、服务标记及标记均为本公司的商标、服务标记及标记。

## 开源证券研究所

### 上海

地址：上海市浦东新区世纪大道1788号陆家嘴金控广场1号楼10层

邮编：200120

邮箱：research@kysec.cn

### 北京

地址：北京市西城区西直门外大街18号金贸大厦C2座9层

邮编：100044

邮箱：research@kysec.cn

### 深圳

地址：深圳市福田区金田路2030号卓越世纪中心1号楼45层

邮编：518000

邮箱：research@kysec.cn

### 西安

地址：西安市高新区锦业路1号都市之门B座5层

邮编：710065

邮箱：research@kysec.cn