

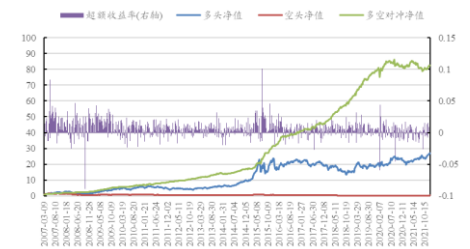
再谈信息不对称理论下的因子研究

高频数据因子研究系列七

报告摘要:

- **因子开发迭代更新越来越重要。**近几年来，随着传统多因子模型在 market 的应用逐渐广泛，因子的波动特征逐渐加大，因子拥挤等原因造成了因子的收益逐渐下降。为了能够寻找更好的 Alpha 收益，在多因子模型框架中，因子作为底层基础，因子的开发、迭代、更新就显得越来越重要。低频相关的数据的因子开发目前边际上的贡献已越来越小，高频数据在用于量化投资中存在一定优势。本篇专题探讨日内高频数据在因子选股中的运用。
- **信息不对称理论。**在个股的交易中，基于市场的非充分有效特征，存在拥有信息优势的交易者，在市场的微观结构中可能更加如此。微观结构理论将市场参与者划分为拥有信息优势的交易者与没有信息优势的交易者两类，认为信息对资产的定价具有较大影响。基于信息不对称理论下的市场微观结构构建 VPIN 因子，研究该因子在选股中的应用。
- **VPIN 因子实证分析。**从因子分档结果来看，VPIN 因子在全市场、中证 800 以及中证 500 范围内的分档效果明显。从因子 IC 值统计结果来看，VPIN 因子的 IC 方向为负向，全市场范围内选股，因子 IC 均值为-0.0426，负 IC 占比 74%，在回测期内，全市场选股多头相对中证 800 指数年化超额收益率为 18.67%，信息比为 1.28。VPIN 因子与传统 BARRA 相关的因子之间的相关性较低，与流动性类因子相关性较高，相关系数为 0.18。
- **风险提示。**策略模型并非百分百有效，市场结构及交易行为的改变以及类似交易参与者的增多有可能使得策略失效。

图：VPIN 因子全市场表现



数据来源：Wind，广发证券发展研究中心

分析师：陈原文



SAC 执证号：S0260517080003



0755-82797057



chenyuanwen@gf.com.cn

分析师：罗军



SAC 执证号：S0260511010004



020-66335128



luojun@gf.com.cn

分析师：安宁宁



SAC 执证号：S0260512020003



SFC CE No. BNW179



0755-23948352



anningning@gf.com.cn

请注意，陈原文、罗军并非香港证券及期货事务监察委员会的注册持牌人，不可在香港从事受监管活动。

相关研究：

行业聚类方法探讨：——重构 2021-11-17

行业轮动框架之十

海量技术指标掘金 Alpha 因子：多因子 Alpha 系列报告之

(四十二)

目录索引

一、高频因子思考：从低频信息到高频信息.....	5
二、信息不对称理论策略回顾.....	6
（一）PIN 模型.....	6
（二）交易量实时更新的(VPIN)模型.....	8
（三）交易量加权的 (VWPIN)模型.....	9
三、交易量实时更新的 VPIN 因子构造.....	9
（一）VPIN 因子构造.....	9
（二）VPIN 因子计算步骤.....	9
（三）VPIN 因子计算举例.....	10
四、实证分析.....	11
（一）数据说明.....	11
（二）因子分档表现.....	12
（三）VPIN 因子实证结果.....	13
五、VPIN 因子与 BARRA 因子相关性分析.....	34
（一）BARRA 因子说明.....	35
（二）VPIN 与 BARRA 相关性分析.....	35
六、VPIN 因子敏感性分析.....	36
（一）敏感性分析-VPIN 因子.....	36
七、总结.....	37
八、风险提示.....	38

图表索引

图 1: 买卖订单流二叉树.....	7
图 2: 高频数据示例图	10
图 3: 拆分交易篮子操作示例图.....	10
图 4: 买单量与卖单量计算示例图-1.....	11
图 5: 买单量与卖单量计算示例图-2.....	11
图 6: 不平衡订单量计算示例图.....	11
图 7: 全市场 VPIN 因子五档-周度调仓	12
图 8: 中证 1000 指数内 VPIN 因子五档-周度调仓	12
图 9: 中证 800 指数内 VPIN 因子五档-周度调仓	12
图 10: 中证 500 指数内 VPIN 因子五档-周度调仓	12
图 11: 沪深 300 指数内 VPIN 因子五档-周度调仓	13
图 12: 创业板指数内 VPIN 因子五档-周度调仓	13
图 13: 全市场选股 VPIN 因子 IC 值及 IC 累计值.....	14
图 14: 全市场选股-多-空策略净值走势-周度调仓	14
图 15: 全市场选股-多-中证 800 策略净值走势-周度调仓	15
图 16: 中证 1000 VPIN 因子 IC 值及 IC 累计值	17
图 17: 中证 1000 选股-多-空策略净值走势-周度调仓	18
图 18: 中证 1000 选股-多-中证 1000 策略净值走势-周度调仓	19
图 19: 中证 800 VPIN 因子 IC 值及 IC 累计值	21
图 20: 中证 800 选股-多-空策略净值走势-周度调仓	21
图 21: 中证 800 选股-多-中证 800 策略净值走势-周度调仓	22
图 22: 中证 500 VPIN 因子 IC 值及 IC 累计值	24
图 23: 中证 500 选股-多-空策略净值走势-周度调仓	25
图 24: 中证 500 选股-多-中证 500 策略净值走势-周度调仓	26
图 25: 沪深 300 VPIN 因子 IC 值及 IC 累计值	28
图 26: 沪深 300 选股-多-空策略净值走势-周度调仓	29
图 27: 沪深 300 选股-多-沪深 300 策略净值走势-周度调仓	30
图 28: 创业板 VPIN 因子 IC 值及 IC 累计值.....	32
图 29: 创业板选股-多-空策略净值走势-周度调仓	32
图 30: 创业板选股-多-创业板指策略净值走势-周度调仓	33
图 31: 全市场 VPIN 手续费敏感性测试	36
图 32: 中证 1000 VPIN 因子手续费敏感性测试	36
图 33: 中证 800 VPIN 因子手续费敏感性测试	37
图 34: 中证 500 VPIN 因子手续费敏感性测试	37
图 35: 沪深 300VPIN 因子手续费敏感性测试	37
图 36: 创业板 VPIN 因子手续费敏感性测试.....	37
表 1: 广发金工高频数据因子挖掘系列报告一览	6
表 2: 全市场选股-整体与分年度 IC 表现.....	13

表 3: 全市场选股-多-空对冲分年度策略表现-周度调仓	15
表 4: 全市场选股-多-中证 800 指数对冲分年度策略表现-周度调仓	16
表 5: 全市场选股-分年度换手率	16
表 6: 中证 1000 选股-整体与分年度 IC 表现	17
表 7: 中证 1000 选股-多-空对冲分年度策略表现-周度调仓	18
表 8: 中证 1000 选股-多-中证 1000 指数对冲分年度策略表现-周度调仓	19
表 9: 中证 1000 选股-分年度换手率	20
表 10: 中证 800 选股-整体与分年度 IC 表现	20
表 11: 中证 800 选股-多-空对冲分年度策略表现-周度调仓	22
表 12: 中证 800 选股-多-中证 800 指数对冲分年度策略表现-周度调仓	23
表 13: 中证 800 选股-分年度换手率	23
表 14: 中证 500 选股-整体与分年度 IC 表现	24
表 15: 中证 500 选股-多-空对冲分年度策略表现-周度调仓	25
表 16: 中证 500 选股-多-中证 500 指数对冲分年度策略表现-周度调仓	26
表 17: 中证 500 选股-分年度换手率	27
表 18: 沪深 300 选股-整体与分年度 IC 表现	28
表 19: 沪深 300 选股-多-空对冲分年度策略表现-周度调仓	29
表 20: 沪深 300 选股-多-沪深 300 指数对冲分年度策略表现-周度调仓	30
表 21: 沪深 300 选股-分年度换手率	31
表 22: 创业板选股-整体与分年度 IC 表现	31
表 23: 创业板选股-多-空对冲分年度策略表现-周度调仓	33
表 24: 创业板选股-多-创业板指数对冲分年度策略表现-周度调仓	34
表 25: 创业板选股-分年度换手率	34
表 26: BARRA 因子说明	35
表 27: VPIN 因子与 BARRA 相关性表	36

一、高频因子思考：从低频信息到高频信息

近年来，A股市场机构化趋势明显，量化私募机构的管理规模也迅速扩大，产生了一批管理规模超过百亿的量化私募机构。与此同时，传统的风格因子波动增大，从市场获取超额收益的难度在增加。

因子拥挤是因子收益下降的原因之一。因子代表着市场某方面的非有效性、或者是一段时期内的定价失效。当某类因子收益高的时候，会吸引更多的资金进入，从而出现因子拥挤，降低因子的预期收益。一旦新的因子被公开，套利资金的介入会使得错误定价收窄，因子收益也会跟着下降。因此，在多因子选股模型中，因子的开发和更新迭代变得越来越重要。

以传统日频价量和更低频财务数据为基础的因子开发是一种研究途径。由于基础因子广为人知，在此基础上进行因子挖掘的收益提升空间相对有限。而且日频数据由于本身的数据量和信息量有限，过度挖掘会增大过拟合的风险。

以高频价量数据为基础的因子开发在当下具有更大的收益提升空间。与低频因子相比，高频数据在用于量化投资中存在一定优势。

首先，高频价量数据的体量明显大于低频数据。以分钟行情为例，用压缩效果较好的mat格式存储2020年全市场股票的分钟行情数据（包括分钟频的开高低收价格数据、买卖盘挂单数据等），约为12GB。如果是快照行情（目前上交所和深交所都是3秒一笔）或者level 2行情，数据量要大很多。因此，高频数据因子挖掘对信息处理能力和处理效率的要求较高。而且，日内数据，尤其是level 2数据，一般要额外付费，甚至需要自行下载存储实时行情，在此基础上构建的因子拥挤度较低。

其次，高频价量数据一般是多维的时间序列数据，数据中噪声比例较高，而且与ROE、PE这类低频指标本身就具有选股能力不同的是，原始的高频行情数据一般不能直接用作选股因子，而要通过信号变换、时间序列分析、机器学习等方法从高频数据中构建特征，才能作为选股因子。此类因子与低频信号的相关性较低，而且由于因子开发流程相对复杂，不同投资者构建的因子更具有多样性。

此外，高频数据开发的因子一般调仓周期较短，意味着在检验因子有效性的时候，同一段测试期具有更多的独立样本。例如，在一年的测试期内，只有12个独立的样本段用于检验月频调仓的因子，与之相比，有约50个独立的时段用于检验周频调仓因子，有超过240个独立的时段用于检验日频调仓的因子。独立样本的增多有助于检验高频因子的有效性。

高频数据挖掘因子的难点在于数据维度大、噪声高。凭借专业投资者的经验或者是参阅已发表的文献，可以从高频数据中提炼出一部分有选股能力的特征。此外，机器学习方法擅长从数据中寻找规律和特征，是高频数据因子挖掘的有力工具。本篇专题报告通过学术上关于高频相关的研究结果借鉴，从高频价量数据中提炼选股因子。

本篇报告从高频因子出发，聚焦市场微观结构中的信息不对称。本篇报告是基于信息不对称理论相关报告的第二篇。

表 1: 广发金工高频数据因子挖掘系列报告一览

信息不对称理论下的因子研究-高频数据因子研究系列六
高频价量数据的因子化方法-多因子 Alpha 系列报告之（四十一）
深度学习框架下高频数据因子挖掘-深度学习研究报告之七
基于个股羊群效应的选股因子研究-高频数据因子研究系列三
基于日内高频数据的短周期选股因子研究-高频数据因子研究系列二
基于日内高频数据的短周期选股因子研究-高频数据因子研究系列一

数据来源：Wind，广发证券发展研究中心

二、信息不对称理论策略回顾

传统的有效市场假说认为，在完全有效的市场状态下，价格能够充分反映资产的所有公开信息以及私有信息。但是在非强有效的市场状态下，参与交易的人掌握的信息可能并不相同，往往掌握更多信息的市场参与者相对于信息匮乏的市场参与者而言，处于市场的有利地位。

信息不对称理论将市场上的交易者分为两类：信息优势的交易者与没信息优势的交易者，有信息优势的交易者往往会根据自身掌握的信息优势做出对自身投资决策有利的决策。如在有利好消息信息优势时，会买入股票；在有利空消息信息优势时，会卖出股票。一般而言，拥有信息优势的交易者越多，市场上的信息不对称程度就越强；拥有信息优势的交易者越少，市场上的信息不对称程度就越低。

对于信息优势占比程度的估计，有以下三种模型：

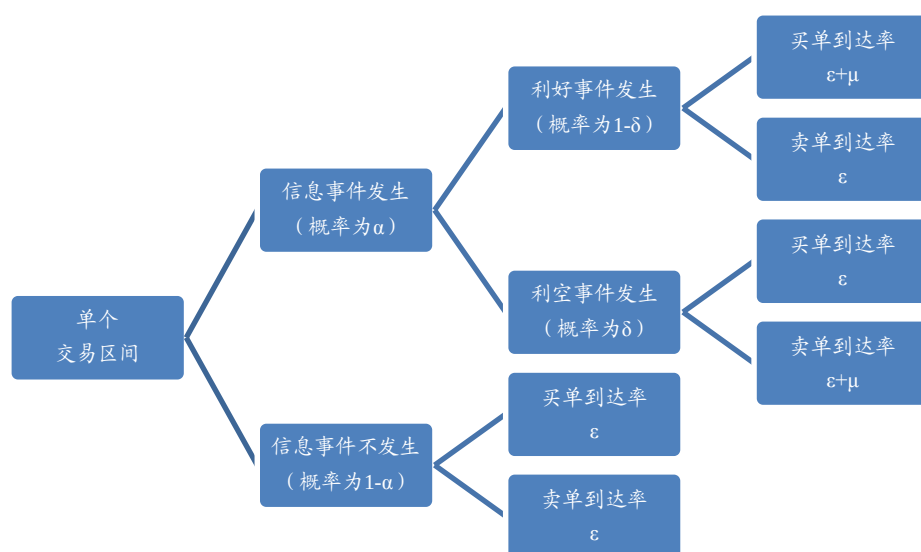
（一）PIN 模型

1. 模型推导

$$PIN = \frac{\alpha\mu}{\alpha\mu + 2\varepsilon}$$

没有好/坏消息时，信息优势交易者不参与交易，因此交易双方的委托单到达率均为 ε 。当信息事件发生时，有以下两种情况。当利好消息发生，信息优势的交易者将事先得知该消息并买入股票，非信息优势交易者可能买入股票也可能卖出股票，此时买方与卖方委托单到达率分别为 $\varepsilon + \mu$ 、 ε 。同理，当利空消息发生，信息优势交易者将利用该消息并卖出股票，非信息优势交易者可能买入或卖出股票，此时买方与卖方委托单到达率分别为 ε 、 $\varepsilon + \mu$ 。具体买卖订单流二叉树如下图所示：

图 1: 买卖订单流二叉树



数据来源: Wind, 广发证券发展研究中心

$$\text{因此 } PIN = \frac{\alpha\mu}{\alpha(\mu+2\varepsilon)+2\varepsilon(1-\alpha)} = \frac{\alpha\mu}{\alpha\mu+2\varepsilon}$$

2. 极大似然估计法参数

PIN模型假设买卖订单服从泊松分布, 其实求解PIN的重点在于估计四个变量 ($\alpha, \mu, \delta, \varepsilon$)

没有消息的时候, 有B笔买单S笔卖单的概率为:

$$(1-\alpha)e^{-\varepsilon T} \frac{(\varepsilon T)^B}{B!} e^{-\varepsilon T} \frac{(\varepsilon T)^S}{S!}$$

发生利好消息的时候, 有B笔买单S笔卖单的概率为:

$$(1-\delta)e^{-(\mu+\varepsilon)T} \frac{[(\mu+\varepsilon)T]^B}{B!} e^{-\varepsilon T} \frac{(\varepsilon T)^S}{S!}$$

发生利空消息的时候, 有B笔买单S笔卖单的概率为:

$$\alpha\delta e^{-\varepsilon T} \frac{(\varepsilon T)^B}{B!} e^{-(\mu+\varepsilon)T} \frac{[(\mu+\varepsilon)T]^S}{S!}$$

$$\begin{aligned} L((B,S)|\theta) &= (1-\alpha)e^{-\varepsilon T} \frac{(\varepsilon T)^B}{B!} e^{-\varepsilon T} \frac{(\varepsilon T)^S}{S!} + \alpha\delta e^{-\varepsilon T} \frac{(\varepsilon T)^B}{B!} e^{-(\mu+\varepsilon)T} \frac{[(\mu+\varepsilon)T]^S}{S!} \\ &\quad + \alpha(1-\delta)e^{-(\mu+\varepsilon)T} \frac{[(\mu+\varepsilon)T]^B}{B!} e^{-\varepsilon T} \frac{(\varepsilon T)^S}{S!} \end{aligned}$$

假设观察到的第一天的数据: $M = (B_i, S_i)_{i=1}^I$

那么 (B,S) 的似然函数可写为: $L(M|\theta) = \prod_{i=1}^I L(\theta|(B_i, S_i))$, 求解即可解出参数值。

3. 缺点

计算量过大，当买卖订单数量庞大的时候，计算量复杂度较高；在很多情况下无法估计出PIN值；PIN值估计量可能会偏离真实值，准确性不高。

（二）交易量实时更新的(VPIN)模型

交易量实时更新的(Volume Synchronized PIN)模型，简称为VPIN模型，是一种非参数估计模型。与PIN模型相比，它不需要进行复杂的极大似然估计获得信息优势交易占比概率的隐含参数，而是通过计算等交易量区间的交易量不平衡性来获得信息优势交易程度的测度。该模型的内在逻辑是：在高频交易的现实世界中，当信息连续地到达市场，投资者也会连续的对信息进行分析并做出交易决策。由于投资交易并不服从均匀分布，而是具有一定的不规则性，那么同一信息引起的交易行为将以不同速度到达市场。基于上述思路，通过度量单位时间内的交易量的不平衡性所构建的PIN模型则与现实情况存在偏差，因此VPIN模型对此进行优化。该模型以相同的交易量划分交易区间，认为只有当交易量达到设定的阈值时，信息才完全融入了市场中。具体的VPIN模型构建步骤如下：

1. 划分等交易量的交易区间

将按时间排列的交易数据分成等交易量的交易篮子，每个交易篮子所包含的交易量定为 V ，如果上一笔交易装满了一个篮子且还有剩余，那么余下的部分将会被分配给下一个篮子，令 $\tau = 1, 2 \dots n$ 为所有等交易量的交易篮子。

2. 基于正态分布假设对交易方向进行判断

为刻画交易量的不平衡性，需要先将交易方向进行划分，得到卖方交易量 V^S 与买方交易量 V^B 。假设 $\tau = 1, 2 \dots n$ 为交易篮子的下标，那么单个交易篮子的总交易量为 $V_\tau = V_\tau^B + V_\tau^S$ 。可以得到如下交易量计算公式：

$$V_\tau^B = \sum_{i=t(\tau-1)+1}^{t(\tau)} V_i * Z\left(\frac{P_i - P_{i-1}}{\sigma_{VP}}\right)$$

$$V_\tau^S = \sum_{i=t(\tau-1)+1}^{t(\tau)} V_i \left[1 - Z\left(\frac{P_i - P_{i-1}}{\sigma_{VP}}\right)\right] = V_\tau - V_\tau^B$$

$$OI_\tau = |V_\tau^S - V_\tau^B|$$

其中， $t(\tau)$ 为第 τ 个交易篮子(第 τ 个等交易量区间)， Z 表示正态分布下的累计密度函数。如果等交易量区间内价格没有发生变化，则认为该交易时段内的买方交易量与卖方交易量相同。如果等交易量区间内价格上升，则买方交易量的比重将会提升。

最后将每个交易篮子中的交易不平衡量除以篮子大小，再求平均值，即可得到VPIN值。

$$VPIN = \frac{\alpha\mu}{\alpha\mu + 2\epsilon} = \frac{\alpha\mu}{V} \approx \frac{\sum_{\tau=1}^n |V_\tau^S - V_\tau^B|}{nV}$$

3. 优缺点及改进

不论是计算复杂程度还是构建思路，VPIN模型相较于PIN模型而言，都更加贴合实际情况的需要，但是VPIN模型仍然具有改进的空间。一方面，VPIN模型没有考虑到信息优势交易者隐藏其意图而拆小单进行交易的情况，其计算公式只反映了交易量的不平衡性，未能反映订单数量的不平衡性；另一方面，该模型需要提前利用长期历史数据确定交易篮子的总交易量V，只能对历史数据进行回测。

针对上述两个问题，李平等人在2020年在论文《知情交易概率于风险定价——基于不同PIN测度方法的比较研究》中，提出了VWPIN模型。

（三）交易量加权的 (VWPIN)模型

基于物理时间和交易量加权的知情交易概率(volume-weighted probability of informed trading)模型，简称VWPIN模型，继承了VPIN模型的基本思路，认为不同的交易量反映的信息量也不同，同时基于知情交易者可能存在大单交易，或为隐藏知情交易意图而拆小单进行交易的实际情况，将订单数量的不平衡性加入模型的考虑范围。具体而言，VWPIN模型计算了固定时间范围内，交易量加权的订单数量不平衡程度。具体计算步骤如下：

$$VWPIN = \sum_{i=1}^n w_i Pin_i = \sum_{i=1}^n w_i \frac{|S_i - B_i|}{S_i + B_i}$$

$$w_i = \frac{TradVol_i}{\sum_{i=1}^n TradVol_i}$$

其中，n为固定时间范围内划分的交易区间个数， Pin_i 第i个区间的知情交易概率， S_i 为第i个交易区间的卖单数量， B_i 为第i个交易区间的买单数量， w_i 为第i个交易区间的交易量加权系数， $TradVol_i$ 为第i个交易区间的交易量。

相比之前的模型，VWPIN模型具有计算简便、综合考虑订单数量和交易量、不受估计区间限制等优点。

三、交易量实时更新的 VPIN 因子构造

（一）VPIN 因子构造

根据信息不对称理论及VPIN模型的定义，本篇专题报告构建VPIN因子，用来衡量市场中个股反映信息不对称程度的刻画。

（二）VPIN 因子计算步骤

VPIN模型的计算是利用个股的区间i的交易量 V_i 、价格 P_i 来实现的。具体计算如下：

首先，将按时间排列的交易数据分成等交易量的交易篮子，每个交易篮子所包含的交易量定为 V_t （本文定为前三个月的日均成交量的五十分之一），如果上一笔交易装满了篮子且还有剩余，那么余下的部分将会被分配给下一个篮子，令 $\tau = 1, 2 \dots n$ 为所有等交易量的交易篮子

其次，根据第 t 天个股的区间交易量、区间价格，计算当天每个交易时段的价格变动（ ΔP ），计算当天所有交易时段价格变动的标准差（ $\sigma_{\Delta P}$ ），并计算每个交易时段的价格变动的正态分布分位数 $Z(\frac{P_i - P_{i-1}}{\sigma_{\Delta P}})$ 。

再次，将每个交易时段的成交量与价格变动的正态分布分位数相乘，计算第 τ 个交易篮子的买单交易量 $V_{\tau}^B = \sum_{i=t(\tau-1)+1}^{t(\tau)} V_i * Z(\frac{P_i - P_{i-1}}{\sigma_{\Delta P}})$ ，计算卖单交易量 $V_{\tau}^S = \sum_{i=t(\tau-1)+1}^{t(\tau)} V_i * [1 - Z(\frac{P_i - P_{i-1}}{\sigma_{\Delta P}})]$ ，再计算交易不平衡量 $OI_{\tau} = |V_{\tau}^S - V_{\tau}^B|$ 。

最后，将个股在第 t 天所有交易篮子内的交易不平衡量 OI 取平均，得到第 t 天的VPIN值，对个股在日历周内的VPIN值数据求均值，得到VPIN因子。

（三）VPIN 因子计算举例

假设当前交易日为 t ，每1分钟作为一个交易区间，则交易区间为 $i = 1, 2 \dots 240$ ，那么VPIN因子计算举例如下图所示。

首先提取高频数据，相关变量与字段如下图所示。

图 2：高频数据示例图

指标	时间	股票代码	收盘价	成交量	当日累计成交量
变量字符	Date	Stockid	Price	Vol	Section_vol

数据来源：Wind，广发证券发展研究中心

假设提取到的数据如图3所示，此时我们假设过去三个月日均成交量为5000，那么一个交易篮子的量为 $5000/50=100$ ，当某一个交易时段的数据填满桶还有剩余，那么将一部分成交量填满上一个交易篮子，剩余部分成交量分给下一个交易篮子。计算每个时间区间的价格变动（ ΔP ），根据价格变动计算所有样本价格变动的标准差（ $\sigma_{\Delta P}$ ），以及价格变动的正态分布分位数 $Z(\frac{P_i - P_{i-1}}{\sigma_{\Delta P}})$ 。

图 3：拆分交易篮子操作示例图

时间	股票代码	成交量	时点当日累积成交量	Bucket	ΔP	$\sigma_{\Delta P}$	$Z(\frac{P_i - P_{i-1}}{\sigma_{\Delta P}})$
9:31	SZ300750	20	20	1	0	3.61881	0.5
9:32	SZ300750	40	60	1	1	3.61881	0.6238
9:33	SZ300750	40	100	1	0	3.61881	0.5
9:33	SZ300750	40	140	2	0	3.61881	0.5
9:34	SZ300750	30	170	2	-1	3.61881	0.3762
9:35	SZ300750	10	180	2	0	3.61881	0.5
9:36	SZ300750	20	200	2	1	3.61881	0.6238
9:36	SZ300750	20	220	3	1	3.61881	0.6238

数据来源：Wind，广发证券发展研究中心

处理数据可得买单量、卖单量和交易不平衡量。

图 4: 买单量与卖单量计算示例图-1

时间	股票代码	Bucket	$Z(\frac{P_i - P_{i-1}}{\sigma_{\Delta P}})$	成交量 V_i	$V_i * Z(\frac{P_i - P_{i-1}}{\sigma_{\Delta P}})$	$V_i * [1 - Z(\frac{P_i - P_{i-1}}{\sigma_{\Delta P}})]$
9:31	SZ300750	1	0.5	20	10	10
9:32	SZ300750	1	0.6238	40	24.952	15.048
9:33	SZ300750	2	0.5	40	20	20
9:34	SZ300750	2	0.3762	30	11.286	18.714
9:35	SZ300750	2	0.5	10	5	5
9:36	SZ300750	2	0.6238	20	12.476	7.524

数据来源: Wind, 广发证券发展研究中心

$$\text{买单交易量: } V_{\tau}^B = \sum_{i=t(\tau-1)+1}^{t(\tau)} V_i * Z(\frac{P_i - P_{i-1}}{\sigma_{\Delta P}})$$

$$\text{卖单交易量: } V_{\tau}^S = \sum_{i=t(\tau-1)+1}^{t(\tau)} V_i * [1 - Z(\frac{P_i - P_{i-1}}{\sigma_{\Delta P}})]$$

图 5: 买单量与卖单量计算示例图-2

$V_i * Z(\frac{P_i - P_{i-1}}{\sigma_{\Delta P}})$	$V_i * [1 - Z(\frac{P_i - P_{i-1}}{\sigma_{\Delta P}})]$	V_{τ}^B	V_{τ}^S
10	10		
24.952	15.048		
20	20	54.952	45.078
11.286	18.714		
5	5		
12.476	7.524	48.762	51.238

数据来源: Wind, 广发证券发展研究中心

$$\text{计算不平衡订单量 } OI_{\tau} = |V_{\tau}^S - V_{\tau}^B|$$

图 6: 不平衡订单量计算示例图

股票代码	Bucket	V_{τ}^B	V_{τ}^S	OI	开始时间	结束时间
SZ300750	1	54.952	45.078	9.874	9:31	9:33
SZ300750	2	48.762	51.238	2.476	9:33	9:36

数据来源: Wind, 广发证券发展研究中心

日度VPIN值:

$$VPIN = \frac{\alpha \mu}{\alpha \mu + 2\varepsilon} = \frac{\sum_{\tau=1}^n |V_{\tau}^S - V_{\tau}^B|}{nV} = \frac{\sum_{\tau=1}^n |OI_{\tau}|}{nV}$$

计算VPIN的周度平均值, 计为VPIN因子值。

四、实证分析

(一) 数据说明

选股范围: 全市场、中证500、中证800、创业板指、沪深300等

股票预处理: 剔除ST/*ST、涨跌停板、上市未满1年股票, 计算前三个月的日

均成交量（前三个日历月至少要有40个交易日）

因子预处理：MAD去极值、Z-Score标准化、行业市值中性化

回测区间：2007.03.09–2021.12.31

分档方式：根据当期股票的因子值，从小到大分为五档

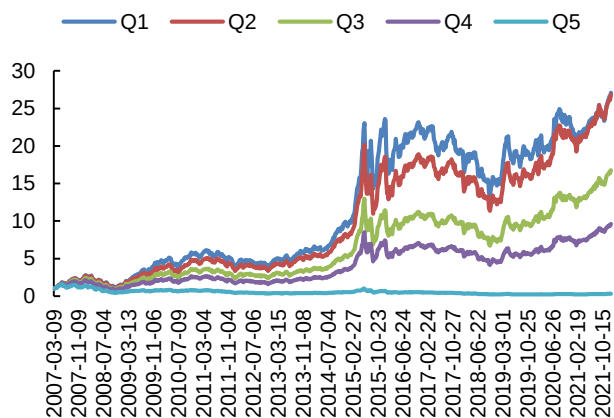
调仓周期：每周最后一个交易日以收盘价调仓

交易费用：千分之三（卖出时收取）

（二）因子分档表现

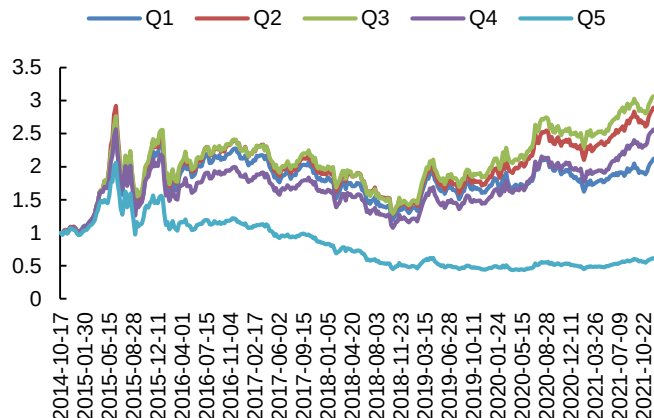
1. 因子分档表现-VPIN因子

图 7：全市场 VPIN因子五档-周度调仓



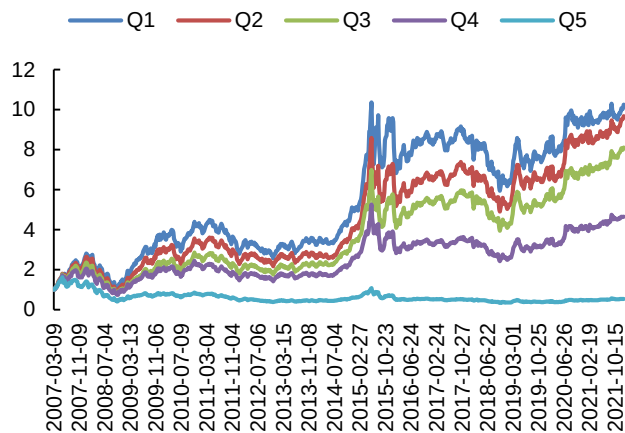
数据来源：Wind，广发证券发展研究中心

图 8：中证1000指数内VPIN因子五档-周度调仓



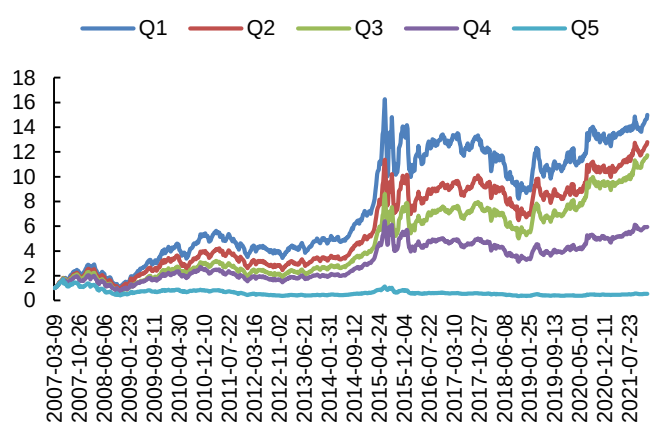
数据来源：Wind，广发证券发展研究中心

图 9：中证800指数内VPIN因子五档-周度调仓



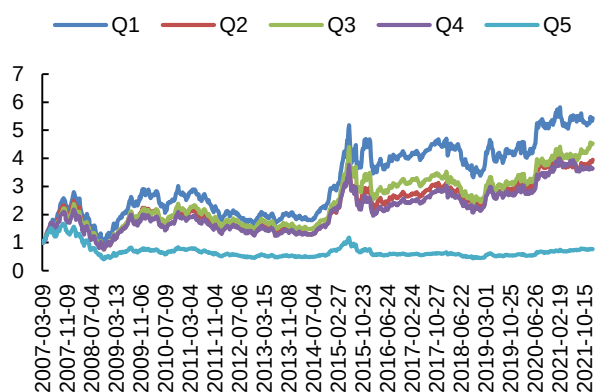
数据来源：Wind，广发证券发展研究中心

图 10：中证500指数内VPIN因子五档-周度调仓



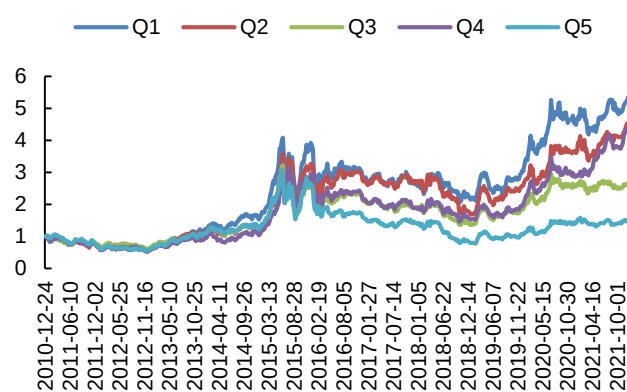
数据来源：Wind，广发证券发展研究中心

图 11: 沪深300指数内VPIN因子五档-周度调仓



数据来源: Wind, 广发证券发展研究中心

图 12: 创业板指数内VPIN因子五档-周度调仓



数据来源: Wind, 广发证券发展研究中心

(三) VPIN 因子实证结果

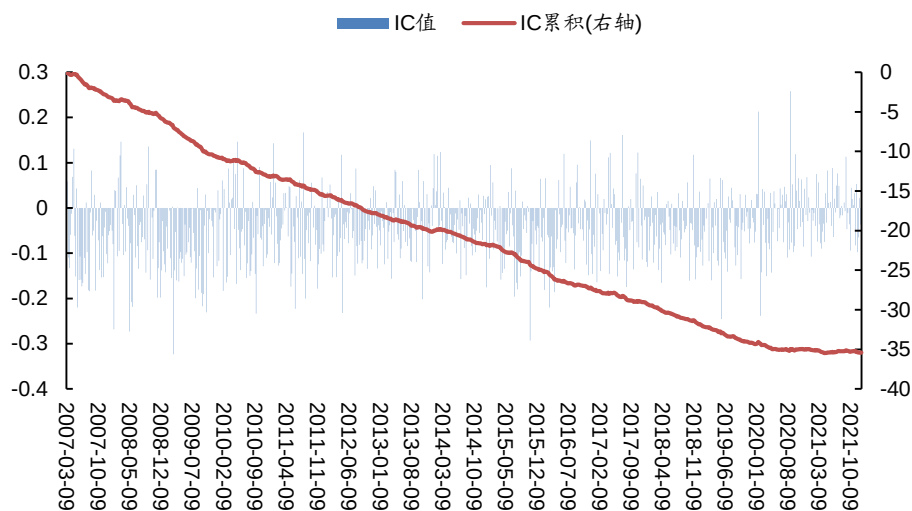
1. 全市场选股

表 2: 全市场选股-整体与分年度IC表现

范围	IC 均值	IC 最大值	IC 最小值	负 IC 所占比例	IC 累计值
2007	-7.53%	13.09%	-22.02%	80.95%	-3.16
2008	-5.51%	14.68%	-27.34%	72.55%	-2.81
2009	-9.31%	4.43%	-32.32%	92.16%	-4.75
2010	-4.76%	14.66%	-23.36%	73.08%	-2.47
2011	-4.73%	16.73%	-22.26%	70.59%	-2.41
2012	-4.74%	11.77%	-23.21%	80.00%	-2.37
2013	-4.26%	8.46%	-20.16%	78.43%	-2.17
2014	-2.94%	12.38%	-15.71%	71.15%	-1.53
2015	-6.34%	9.52%	-29.29%	80.77%	-3.30
2016	-4.91%	14.94%	-21.97%	78.43%	-2.50
2017	-3.59%	16.18%	-17.45%	70.59%	-1.83
2018	-5.05%	11.79%	-16.58%	80.39%	-2.58
2019	-4.72%	6.66%	-24.57%	72.55%	-2.41
2020	-1.36%	25.84%	-23.82%	54.90%	-0.69
2021	-0.83%	11.32%	-10.74%	57.69%	-0.43
All	-4.67%	25.84%	-32.32%	74.18%	-35.42

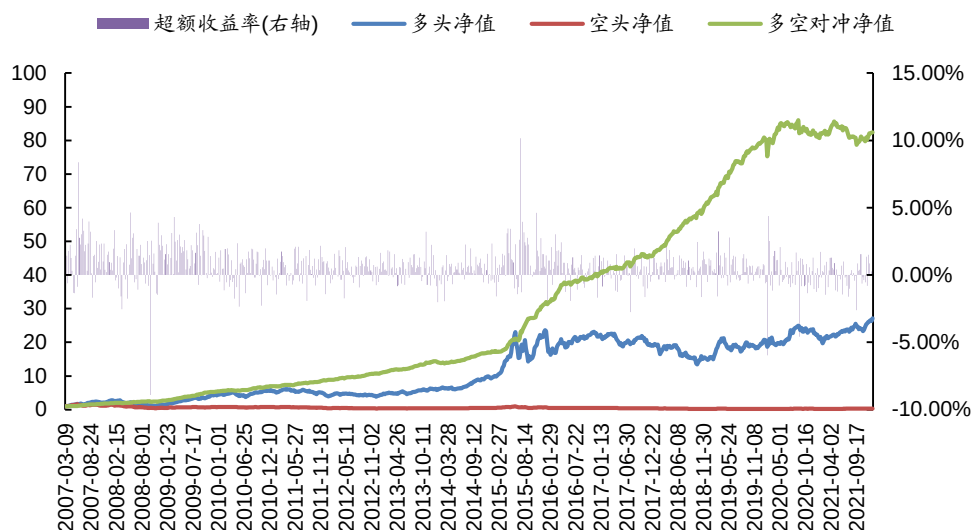
数据来源: Wind, 广发证券发展研究中心

图 13: 全市场选股 VPIN因子IC值及IC累计值



数据来源: Wind, 广发证券发展研究中心

图 14: 全市场选股-多-空策略净值走势-周度调仓



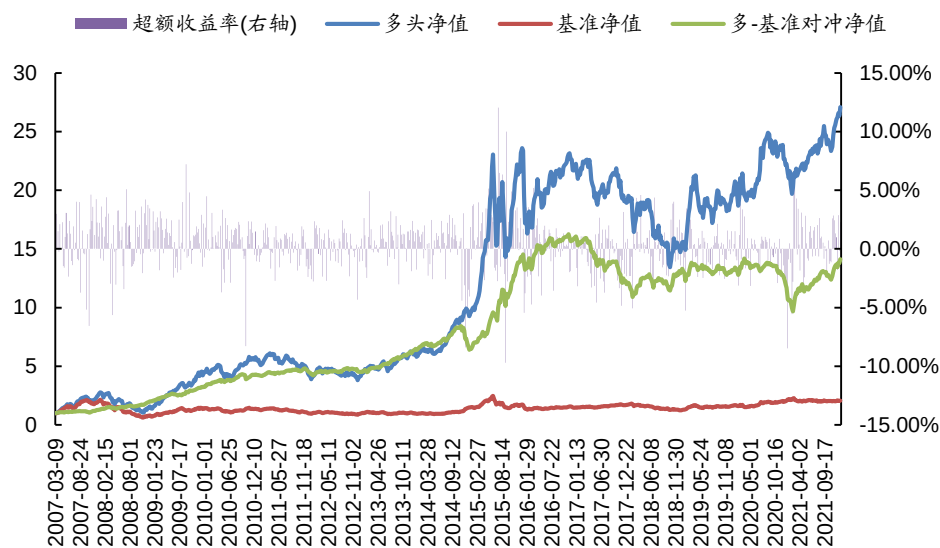
数据来源: Wind, 广发证券发展研究中心

表 3: 全市场选股-多-空对冲分年度策略表现-周度调仓

范围	累积收益率	最大回撤	年化波动率	年化收益率	信息比率
2007	86.67%	2.67%	12.76%	110.24%	8.96
2008	42.21%	8.92%	14.64%	41.23%	2.89
2009	103.22%	0.74%	8.41%	100.41%	12.37
2010	28.61%	3.14%	7.99%	27.37%	3.51
2011	26.83%	2.06%	6.71%	26.24%	4.01
2012	23.73%	1.73%	5.34%	23.73%	4.55
2013	31.80%	1.58%	5.85%	31.09%	5.45
2014	18.42%	4.54%	5.62%	17.65%	3.21
2015	85.38%	2.61%	13.57%	81.03%	6.17
2016	28.45%	1.92%	7.34%	27.82%	3.89
2017	14.41%	2.76%	6.33%	14.11%	2.28
2018	32.19%	1.63%	5.39%	31.47%	5.99
2019	30.16%	1.59%	6.10%	29.49%	4.95
2020	1.94%	7.02%	11.00%	1.90%	0.18
2021	1.52%	8.18%	6.31%	1.43%	0.23
All	8142.89%	8.92%	9.14%	33.68%	3.78

数据来源: Wind, 广发证券发展研究中心

图 15: 全市场选股-多-中证800策略净值走势-周度调仓



数据来源: Wind, 广发证券发展研究中心

表 4: 全市场选股-多-中证800指数对冲分年度策略表现-周度调仓

范围	多头累积收益率	累积收益率	最大回撤	年化波动率	年化收益率	信息比率
2007	154.84%	22.42%	12.93%	17.67%	27.23%	1.58
2008	-44.43%	56.64%	9.35%	15.98%	55.27%	3.56
2009	209.52%	62.84%	6.05%	13.08%	61.29%	4.83
2010	24.97%	29.84%	9.78%	14.04%	28.54%	2.08
2011	-24.52%	4.65%	7.62%	8.47%	4.55%	0.55
2012	6.36%	2.58%	7.04%	10.19%	2.58%	0.26
2013	38.77%	40.25%	3.99%	10.27%	39.32%	3.93
2014	56.07%	4.66%	20.52%	15.56%	4.48%	0.29
2015	147.65%	113.94%	12.06%	24.65%	107.77%	4.53
2016	-7.52%	11.65%	9.02%	14.98%	11.40%	0.78
2017	-12.82%	-24.21%	25.37%	13.23%	-23.80%	-1.83
2018	-22.57%	6.91%	10.71%	14.46%	6.77%	0.48
2019	33.01%	1.82%	7.71%	11.20%	1.78%	0.16
2020	13.10%	-8.97%	15.48%	10.65%	-8.80%	-0.84
2021	22.34%	17.71%	19.40%	17.77%	16.63%	0.96
All	2609.62%	1248.88%	40.59%	14.95%	18.67%	1.28

数据来源: Wind, 广发证券发展研究中心

表 5: 全市场选股-分年度换手率

范围	均值	最大值	最小值	累计值
2007	63.53%	77.99%	50.00%	26.68
2008	61.57%	98.81%	50.45%	31.40
2009	63.32%	72.08%	54.68%	32.29
2010	60.14%	69.06%	52.14%	31.27
2011	58.69%	66.32%	50.58%	29.93
2012	57.81%	67.84%	51.03%	28.91
2013	59.45%	65.48%	52.94%	30.32
2014	58.56%	67.46%	48.65%	30.45
2015	65.14%	97.67%	50.14%	33.87
2016	57.92%	66.29%	51.95%	29.54
2017	56.58%	69.20%	47.78%	28.86
2018	57.11%	62.95%	50.59%	29.13
2019	56.99%	62.06%	50.45%	29.06
2020	54.38%	63.95%	47.02%	27.73
2021	53.90%	62.09%	48.32%	28.57
All	58.95%	98.81%	47.02%	448.02

数据来源: Wind, 广发证券发展研究中心

在全市场选股中，VPIN因子表现出显著的选股区分度，因子IC均值为-0.0467，负IC占比74.18%。在多头对冲中证800指数的回测中，策略整体的年化收益率为18.67%，信息比率为1.28。除2017和2020年外，其余年份均可取得超额收益。整体换手率保持在58%左右。

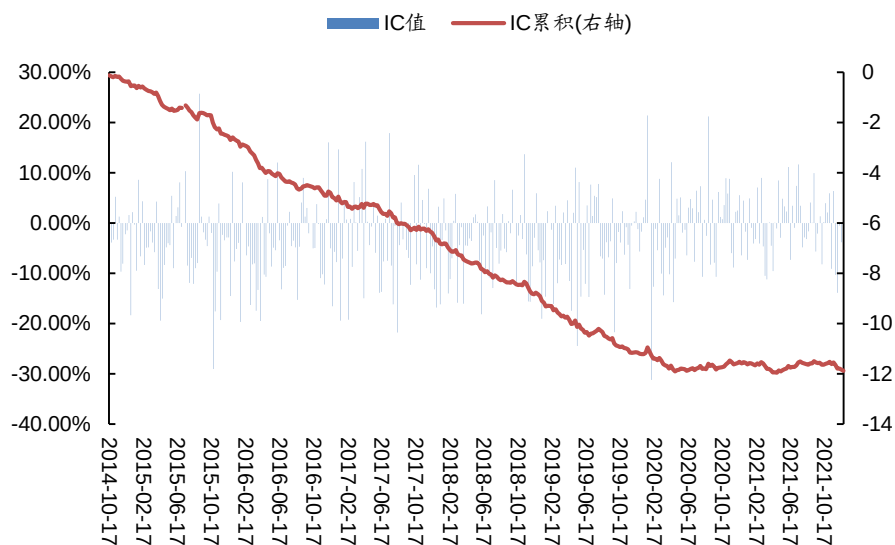
2. 中证1000选股

表 6: 中证1000选股-整体与分年度IC表现

范围	IC 均值	IC 最大值	IC 最小值	负 IC 所占比例	IC 累计值
2014	-3.32%	5.22%	-12.68%	72.73%	-0.37
2015	-4.57%	25.76%	-29.03%	74.51%	-2.33
2016	-4.56%	16.03%	-19.65%	74.51%	-2.32
2017	-3.28%	17.88%	-21.76%	62.75%	-1.67
2018	-4.19%	13.68%	-18.17%	70.59%	-2.14
2019	-4.66%	11.00%	-24.44%	66.67%	-2.38
2020	-0.74%	21.38%	-31.18%	47.06%	-0.38
2021	-0.57%	11.64%	-13.91%	53.85%	-0.30
All	-3.22%	25.76%	-31.18%	64.50%	-11.88

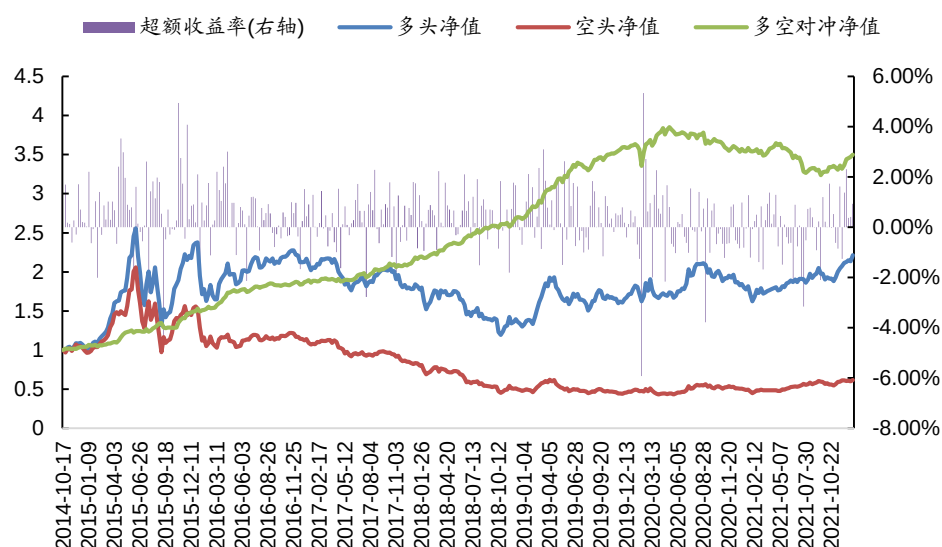
数据来源: Wind, 广发证券发展研究中心

图 16: 中证1000 VPIN因子IC值及IC累计值



数据来源: Wind, 广发证券发展研究中心

图 17：中证1000选股-多-空策略净值走势-周度调仓



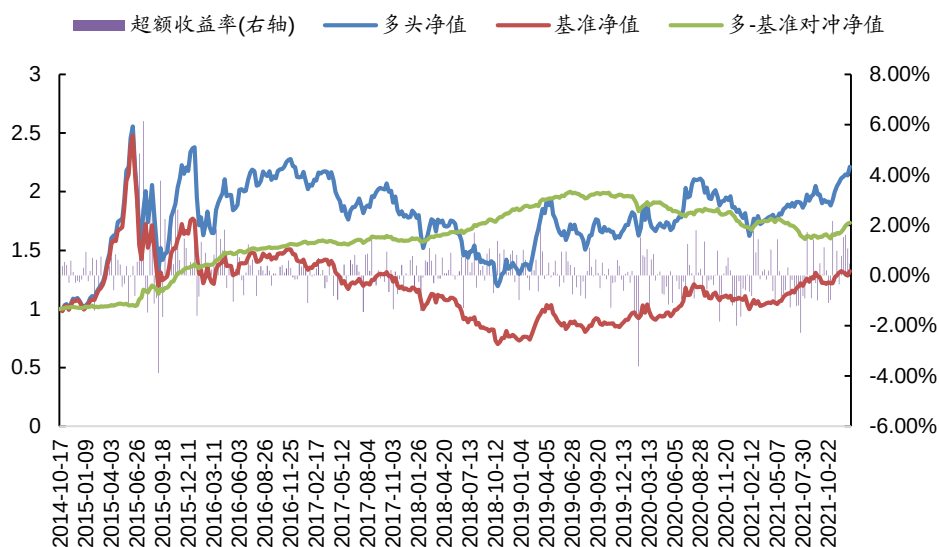
数据来源：Wind，广发证券发展研究中心

表 7：中证1000选股-多-空对冲分年度策略表现-周度调仓

范围	累积收益率	最大回撤	年化波动率	年化收益率	信息比率
2014	4.22%	0.63%	5.17%	20.69%	4.10
2015	43.68%	4.93%	11.40%	41.69%	3.76
2016	24.77%	2.25%	7.57%	24.23%	3.28
2017	14.92%	2.77%	7.14%	14.60%	2.09
2018	24.72%	1.80%	6.47%	24.18%	3.83
2019	34.93%	2.94%	7.35%	34.14%	4.77
2020	-2.09%	8.11%	11.44%	-2.05%	-0.18
2021	-1.03%	11.30%	7.87%	-0.97%	-0.13
All	250.13%	15.94%	8.72%	18.40%	2.16

数据来源：Wind，广发证券发展研究中心

图 18: 中证1000选股-多-中证1000策略净值走势-周度调仓



数据来源: Wind, 广发证券发展研究中心

表 8: 中证1000选股-多-中证1000指数对冲分年度策略表现-周度调仓

范围	多头累积收益率	累积收益率	最大回撤	年化波动率	年化收益率	信息比率
2014	3.69%	3.40%	0.94%	5.80%	16.43%	2.90
2015	128.63%	35.20%	5.82%	11.69%	33.64%	2.95
2016	-10.31%	14.49%	2.43%	4.82%	14.18%	3.01
2017	-16.02%	1.98%	2.59%	4.18%	1.94%	0.47
2018	-27.22%	15.67%	1.91%	4.65%	15.34%	3.38
2019	32.08%	6.68%	2.85%	4.29%	6.54%	1.56
2020	6.49%	-10.80%	10.80%	7.06%	-10.60%	-1.53
2021	21.02%	-0.98%	9.76%	7.07%	-0.92%	-0.13
All	121.21%	77.89%	20.17%	6.88%	8.07%	1.20

数据来源: Wind, 广发证券发展研究中心

表 9: 中证1000选股-分年度换手率

范围	均值	最大值	最小值	累计值
2014	61.52%	67.08%	50.00%	6.77
2015	65.33%	100.00%	51.90%	33.97
2016	57.65%	69.46%	48.78%	29.40
2017	55.33%	65.52%	44.89%	28.22
2018	56.59%	64.57%	47.90%	28.86
2019	56.84%	67.20%	47.45%	28.99
2020	55.06%	66.67%	42.41%	28.08
2021	54.39%	63.02%	45.55%	28.82
All	57.44%	100.00%	42.41%	213.11

数据来源: Wind, 广发证券发展研究中心

在中证1000选股中, VPIN因子IC均值为-0.0322, 正IC占比64.5%。在多头对冲中证1000指数的回测中, 策略整体的年化收益率为8.07%, 信息比率为1.20。整体换手率保持在57%左右。

3. 中证800选股

表 10: 中证800选股-整体与分年度IC表现

范围	IC 均值	IC 最大值	IC 最小值	负 IC 所占比例	IC 累计值
2007	-7.77%	11.31%	-26.37%	80.95%	-3.27
2008	-4.23%	19.84%	-28.46%	66.67%	-2.16
2009	-7.32%	7.07%	-27.26%	80.39%	-3.73
2010	-3.69%	19.45%	-25.13%	71.15%	-1.92
2011	-3.10%	16.09%	-16.33%	64.71%	-1.58
2012	-3.09%	9.78%	-17.55%	66.00%	-1.55
2013	-1.92%	17.10%	-18.36%	64.71%	-0.98
2014	-1.41%	13.64%	-16.96%	57.69%	-0.73
2015	-6.25%	10.14%	-22.63%	80.77%	-3.25
2016	-3.94%	14.87%	-24.30%	64.71%	-2.01
2017	-1.97%	21.14%	-19.17%	56.86%	-1.01
2018	-2.01%	18.47%	-22.92%	54.90%	-1.03
2019	-3.22%	14.88%	-27.99%	66.67%	-1.64
2020	-0.09%	17.95%	-24.95%	43.14%	-0.05
2021	-0.12%	14.06%	-11.78%	53.85%	-0.06
All	-3.29%	21.14%	-28.46%	64.69%	-24.96

数据来源: Wind, 广发证券发展研究中心

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/298051122141006135>