

# 金融工程

证券研究报告

2019 年 09 月 05 日

## 市场微观结构探析系列之二：订单簿上的 alpha

### 微观结构与高频数据

以往的研究对于 A 股市场的日间低频数据分析已有较为深厚的积累，但对日内微观情境下的市场结构特征我们仍缺乏足够的认知。本文希望借助高频数据对微观交易特征进行初步探索；在此基础上我们希望从高频的市场微观特征中归纳总结出增量的低频 alpha 信息。

### 基于订单簿的 Spread 指标

若短期价格取决于当前多头需求量和空头供给量构建的均衡价格，那么盘口信息将是判断个股短期走势的重要依据。根据每个 Tick 上的买卖挂单强度，我们构建了 Spread 指标以度量当前时间的盘口买卖挂单的强弱差异。从 5 个 Tick 到 200 个 Tick 的预测窗口，我们发现 Spread 指标对股价短期走势具有显著预测能力，在买盘挂单相对越厚的时点股价短期上行概率将越大。

将指标调整为日频后，我们发现 Spread 指标在日频率与 Tick 频率窗口下的含义具有明显差异。在日频率上，盘口委买挂单相对越厚的股票在当日实际承受了越强的主动卖压，二者之间相关性显著，因此这部分股票在截面上的相对收益应该越低。

### 从“高频时序”到“低频 alpha”

我们将高频数据低频化后构建了月度的 Spread 选股因子，从 2010 年至 2019 年 8 月 30 日的回溯区间内，因子表现出显著的选股能力。因子中性化后多空组合年化收益为 14.6%，多空组合 IR 达到 2.01；因子 IC 均值 -4.0%，ICIR 达到 -2.10，IC 胜率为 75.7%。

我们分别在沪深 300 指数、中证 500 指数、中证 800 指数以及中证 1000 指数成分股中回溯了因子的选股能力，各股票池中因子 IC 均值均超过 4.0%，因子选股能力呈现出较好的稳健性。此外，Spread 因子相对于反转、波动、换手、流动性等量价指标具有较高的独立性，截面相关系数均值分别为 13.6%，-10.4%、11.7%、-15.1%，因子能带来较显著的增量信息。

风险提示：因子失效风险，模型失效风险，市场风格变动风险。

### 作者

吴先兴 分析师  
SAC 执业证书编号：S1110516120001  
wuxianxing@tfzq.com  
18616029821

缪铃凯 联系人  
miaolingkai@tfzq.com

### 相关报告

- 1 《金融工程：买卖压力失衡-利用高频数据拓展盘口数据》 2017-08-01
- 2 《金融工程：利用交易型 alpha 捕获低频模型短期收益》 2019-03-18

## 内容目录

1. 微观结构与高频数据 .....	3
2. 时间序列因子 .....	4
3. 基于订单簿的 Spread 指标 .....	5
3.1. 订单簿 .....	5
3.2. Spread 指标构造 .....	5
3.3. Spread 指标的低频特征 .....	7
4. 从“高频时序”到“低频 alpha” .....	8
4.1. 因子构造 .....	8
4.2. 因子绩效 .....	8
4.3. 分域绩效 .....	10
4.4. 因子增量信息 .....	11
5. 总结 .....	12

## 图表目录

图 1: 平安银行 20190701 行情示例 .....	3
图 2: 各类型因子图示 .....	4
图 3: 000062.SZ 日收益累计序列 .....	6
图 4: Spread 与 BuyRate 相关系数 .....	7
图 5: Spread 指标原始值分组收益 .....	9
图 6: Spread 指标原始值 IC 序列 .....	9
图 7: Spread 指标中性化值分组收益 .....	9
图 8: Spread 指标中性化值 IC 序列 .....	9
图 9: Spread 指标分域多空净值 .....	10
图 10: Spread 指标与常见量价因子相关性 .....	11
表 1: 订单簿示例-平安银行-20190701093012 .....	5
表 2: Spread 指标平均绩效 (中证 500 指数成分股) .....	6
表 3: 不同频率的 Spread 指标 .....	8
表 4: Spread 指标原始值分年绩效 .....	9
表 5: Spread 指标中性化值分年绩效 .....	10
表 6: Spread 指标分域绩效 .....	10
表 7: 指标增量信息分析 .....	11

## 1. 微观结构与高频数据

众多的学术文献和研究报告已经对于股票市场在月频、周频和日频的收益异象进行了深入分析和探究，业界通过多因子模型以异象指标为特征进而筛选股票从而获取超越市场指数的超额收益。以往的券商研究对于 A 股市场的低频数据分析可谓高屋建瓴、积累深厚，但对日内等微观情境下的市场结构特征仍旧一知半解。

对于低频的因子模型，我们主要以财务数据、EOD 行情数据等作为输入，输出股票在截面上的收益相对排序值。A 股市场的高频数据主要包括行情快照切片数据、逐笔成交数据、逐笔委托数据等，其中通过对逐笔成交数据降频可得的不同频率的分时 K 线。相较于 Level-1 数据只提供买卖各 5 挡的挂单快照数据，Level-2 数据则提供买卖各 10 挡深度的快照信息。

下图为平安银行在 2019 年 7 月 1 日的行情示例，其中上表为盘口行情数据，其包含股票买卖各 10 挡的挂单金额以及挂单数量、最新价格、累计成交数量、累计成交金额等信息，表中只截取部分行列，数据频率为每 3 秒一次的快照切片。下表为逐笔成交明细数据，逐笔成交数据可以精确到毫秒级别，其记录了每笔成交的价格、数量以及买卖双方的订单编号等信息，虽然该数据可精确到毫秒，但是交易所仍是以 3 秒 1 次的频率推送该数据。

图 1：平安银行 20190701 行情示例

SecID	ExTime	LastPx	BidPx[0]	BidPx[1]	BidVol[0]	BidVol[1]	OfferPx[0]	OfferPx[1]	OfferVol[0]	OfferVol[1]
b'000001'	93000000	14.03	14.02	14.01	21200	268144	14.03	14.04	7900	29200
b'000001'	93003000	14.06	14.07	14.06	100000	166100	14.08	14.09	856600	127400
b'000001'	93006000	14.07	14.07	14.06	30400	160600	14.08	14.09	682700	142500
b'000001'	93009000	14.06	14.06	14.05	52600	426400	14.08	14.09	580200	145200
b'000001'	93012000	14.07	14.06	14.05	40100	425200	14.07	14.08	13000	472000
b'000001'	93015000	14.08	14.07	14.06	5600	380	14.08	14.09	430900	160700
b'000001'	93018000	14.06	14.06	14.05	3600	417380	14.07	14.08	31700	405800
b'000001'	93021000	14.07	14.05	14.04	417180	509800	14.07	14.08	80800	393300
b'000001'	93024000	14.07	14.07	14.05	5400	399480	14.08	14.09	348100	164600
b'000001'	93027000	14.08	14.08	14.07	128998	9200	14.09	14.1	165600	1100600
b'000001'	93030000	14.08	14.08	14.07	148598	31000	14.09	14.1	67000	1102600
b'000001'	93033000	14.10	14.09	14.08	3600	92098	14.1	14.11	1087700	807400
b'000001'	93036000	14.09	14.09	14.08	60800	101198	14.1	14.11	866300	807400
b'000001'	93039000	14.09	14.09	14.08	124020	102498	14.1	14.11	620100	807400
b'000001'	93042000	14.09	14.08	14.07	113498	85100	14.09	14.1	1200	456080
b'000001'	93045000	14.10	14.08	14.07	60098	81400	14.1	14.11	415280	807400
b'000001'	93048000	14.09	14.09	14.08	113800	54098	14.1	14.11	297480	808900

SecID	ExTime	ExID	TradeDate	TradeTime	TradePrice	Volume	Turnover	TradeBuyNo	TradeSellNo	TradeFlag
b'000001'	93000010	b'SZ'	20190701	93000010	14.02	3000	42060	336112	78863	b'F'
b'000001'	93000010	b'SZ'	20190701	93000010	14.02	1000	14020	336112	79306	b'F'
b'000001'	93000010	b'SZ'	20190701	93000010	14.02	400	5608	336112	87258	b'F'
b'000001'	93000010	b'SZ'	20190701	93000010	14.02	5000	70100	336112	99073	b'F'
b'000001'	93000010	b'SZ'	20190701	93000010	14.02	5000	70100	336112	103729	b'F'
b'000001'	93000010	b'SZ'	20190701	93000010	14.02	1000	14020	336112	145586	b'F'
b'000001'	93000010	b'SZ'	20190701	93000010	14.02	2000	28040	336112	160598	b'F'
b'000001'	93000010	b'SZ'	20190701	93000010	14.02	1000	14020	336112	251133	b'F'
b'000001'	93000010	b'SZ'	20190701	93000010	14.02	400	5608	336112	276941	b'F'
b'000001'	93000010	b'SZ'	20190701	93000010	14.03	300	4209	336374	8082	b'F'
b'000001'	93000050	b'SZ'	20190701	93000050	14.02	1500	21030	336112	340261	b'F'
b'000001'	93000070	b'SZ'	20190701	93000070	14.02	1300	18226	336112	341506	b'F'
b'000001'	93000070	b'SZ'	20190701	93000070	14.02	18400	257968	336112	342262	b'F'
b'000001'	93000070	b'SZ'	20190701	93000070	14.02	2500	35050	339036	342262	b'F'
b'000001'	93000070	b'SZ'	20190701	93000070	14.02	17000	238340	340205	342262	b'F'

资料来源：Wind，天风证券研究所

日内高频数据反应了市场在微观层面的交易结构特征，多用于程序化交易、高频交易、日内回转等策略模型构建，而以往券商的金工研究对此较少涉及。我们希望能借由该系列报告揭开高频数据的面纱，增进对于市场微观交易结构特征的理解，同时希望能帮助读者在进行相关研究时起到抛砖引玉作用。

作为该系列的第一篇报告，我们希望借助高频数据对日内微观交易结构特征进行初步探索；在此基础上，我们希望从高频的市场微观特征中归纳总结出增量的低频 alpha 信息。

## 2. 时间序列因子

由于 A 股市场 T+1 的交易制度，多因子模型的收益预测区间一般为日频以上，其借助于 alpha 因子给出未来一段时间股票的超额收益预测。无论是日频还是月频的 Alpha 因子，其所给出的信号是对应时间周期下的股票间相对位序关系，因此对于 alpha 因子  $f$  我们常用相关系数 IC 度量指标的预测能力：

$$IC_{f,t} = \text{Corr}(f_s, r_s)$$

其中  $f_s$  为指标  $f$  在股票  $s$  上的取值， $r_s$  为股票  $s$  在截面  $t$  上的区间收益。

相反地，类似于 CTA 择时策略，我们可以将单只股票  $s$  在交易区间  $T$  上的收益切割成  $N$  个区间  $t_i$ 。若时间序列指标  $f$  在每个区间  $t_i$  起始点取值  $f_{t_i}$ ，则我们可以以  $IC_{f,s}$  度量指标  $f$  对于股票  $s$  未来一段区间涨跌幅的预测能力：

$$IC_{f,s} = \text{Corr}(f_{t_i}, r_{t_i})$$

其中  $f_{t_i}$  为信号  $f$  在区间  $t_i$  起始点的取值， $r_{t_i}$  为股票  $s$  在区间  $t_i$  的收益率。若二者呈现出很强的线性相关性，因子取值的两端即可以用于确认个股择时的开平仓点。

总而言之，alpha 指标给出未来一段时间所以股票间的相对强弱预测，而时间序列指标给出单只股票在未来一段时间绝对涨跌幅的预测。

多因子模型一般基于低频数据获取日间的选股 alpha 收益，而对于日内高频股票行情指标，在假设持有底仓情况下我们给出获取日内收益的两种思路：

- 通过高频指标构建日内 alpha 因子，获取多空组合在日内特定区间的收益差；
- 通过高频指标构建时间序列因子，通过个股择时以高抛低吸方式获取收益；

对于方法 a，我们可以将每日的交易时间拆分成多个区间段，若日内 alpha 指标对于底仓中的股票在未来一段区间（例如 30 分钟）具有显著选股能力，可通过买入多头组合同时卖出空头组合，并在区间末反向平仓获取区间上的多空组合收益差。

方法 b 通过构建时间序列指标对个股在日内进行高抛低吸。如指标在时间序列上分布是平稳的，我们可以根据指标取值确定信号的开平仓点，在开仓信号触发后买入或卖出股票并在平仓信号触发后反向平仓获取个股的“T0 收益”。

图 2：各类型因子图示



资料来源：天风证券研究所

日内方法 b 相较于方法 a 显然更加灵活，因此在实际中我们常通过 alpha 模型获取日间超额收益，并通过日内高抛低吸方式获取日内收益。本文将基于高频数据探索具有日内择时能力的时间序列因子，根据高频数据低频化思想，我们将尝试从时间序列因子所揭示的交易规律中归纳总结出低频的 alpha 特征。

### 3. 基于订单簿的 Spread 指标

#### 3.1. 订单簿

高频数据主要包含逐笔数据和行情快照两大类：逐笔成交数据包含“量”和“价”两类信息，通过降频逐笔数据可以得到分时 k 线以及日频 k 线等信息，常见的波动、反转、换手等技术指标本质上是基于降频后的成交“量”和“价”信息衍生而得；而行情快照则是对订单簿的截面切片，其反应了切片时点市场买卖盘口不同档位的价格与挂单数量分布，该信息在多因子模型中较少被利用。

以平安银行为例，其在 2019 年 7 月 1 日 9 时 30 分 12 秒的订单簿快照如下图所示，Level-2 行情包含了买卖挂单各 10 档深度的价格与数量明细。由于行情快照信息较为独立，我们将报告的切入点聚焦于股票订单簿切片快照中所反映的市场的微观结构特征。

表 1：订单簿示例-平安银行-20190701093012

买单档位	买单数量(股)	买单价格(元)	卖单价格(元)	卖单数量(股)	卖单档位
Bid1	40100	14.06	14.07	13000	Ask1
Bid2	425200	14.05	14.08	472000	Ask2
Bid3	629800	14.04	14.09	147700	Ask3
Bid4	204900	14.03	14.10	1097800	Ask4
Bid5	1196500	14.02	14.11	807400	Ask5
Bid6	533132	14.01	14.12	87000	Ask6
Bid7	1039800	14.00	14.13	39200	Ask7
Bid8	22500	13.99	14.14	45600	Ask8
Bid9	32500	13.98	14.15	59900	Ask9
Bid10	25200	13.97	14.16	42100	Ask10

资料来源：Wind，天风证券研究所

#### 3.2. Spread 指标构造

盘口信息是判断个股短期走势的重要依据，短期的市场价格是取决于当前买盘需求量和卖盘供给量构建的均衡价格。因此，我们判断盘口买卖挂单的相对强弱对于股票价格的短期走势具有统计意义上显著的预判作用。当需求远大于供给时，均衡价格将上移；相反地，当供给远大于需求时，均衡价格将下降。

我们分别定义指标 BID、ASK 度量买、卖盘口所提供的流动性强弱，考虑价格可能出现跳档，我们在此用挂单金额而非挂单数量：

$$\begin{cases} \text{BID} = \sum_{i=1}^{10} \text{bidPx}_i * \text{bidVol}_i * w_i \\ \text{ASK} = \sum_{i=1}^{10} \text{askPx}_i * \text{askVol}_i * w_i \end{cases}$$

其中  $\text{bidPx}_i$ 、 $\text{bidVol}_i$  分别为买盘第  $i$  档挂单的价格和数量， $\text{askPx}_i$ 、 $\text{askVol}_i$  分别为卖盘第  $i$  档挂单的价格和数量； $w_i$  为不同档位权重，考虑价格的优先次序，我们令  $w_i = 1 - \frac{i-1}{10}$ ，赋予靠前的档位以更高的权重。

区别于通用意义下的盘口 Bid/Ask Spread<sup>1</sup> 定义，我们定义本文的 Spread\_Tick 指标以度量每个 Tick 时间点盘口买卖挂单的强弱差异：

<sup>1</sup>  $\text{Bid/Ask Spread} = \text{AskPrice}_1 - \text{BidPrice}_1$



$$Spread\_Tick = \frac{BID - ASK}{BID + ASK}$$

指标值越大意味着当前买盘挂单强度越大，股价短期上涨的概率越大；指标值越小则意味着当前盘口卖压越大，价格短期下行的概率越大。

按照前文所述，我们测试时间序列指标  $Spread\_Tick$  对于个股短期收益是否具有预测能力。对于任意一只股票  $s$ ，我们分别以 5、10、20、40、100、200 个 Tick 为预测周期，每日计算指标值  $f$  与之后  $N$  个 Tick 之后的收益  $r$  的相关系数  $IC_{f,s}$

$$IC_{f,s,t} = Corr(f_{t_i}, r_{t_i}), t_i \in t$$

$IC_{f,s,t}$  度量了对于股票  $s$  在  $t$  日信号值  $f$  大小与未来一段区间收益的线性相关程度。此外，由于我们希望通过信号获取短期内做多或者做空的机会，我们利用盘口均价计算股票的短期收益变化，从而度量信号多空两端捕获绝对收益的能力。

对于任意一只股票我们可逐日计算 Spread 指标值与未来  $N$  个 Tick 收益间的相关系数以及 Spread 指标多头和空头的平均收益与胜率。

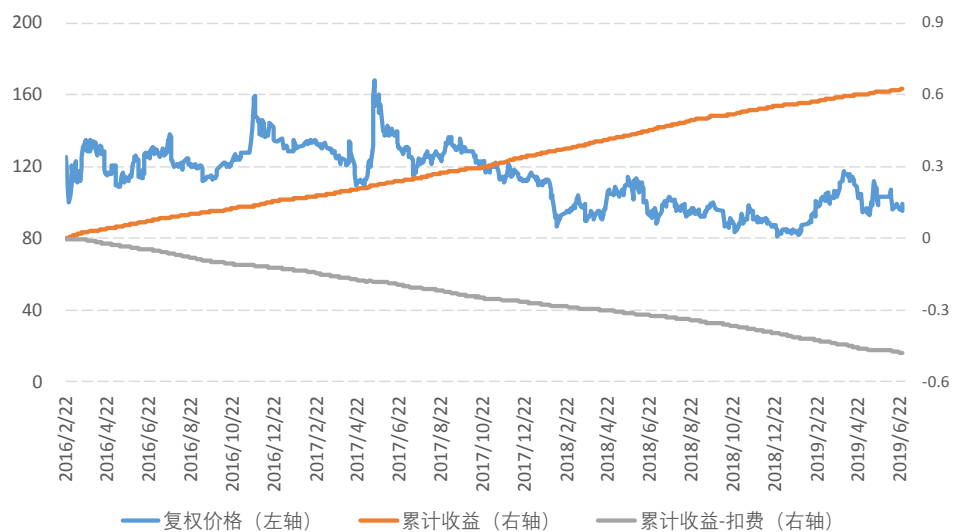
表 2: Spread 指标平均绩效 (中证 500 指数成分股)

Tick 数量	IC 均值	空头收益	空头胜率	多头收益	多头胜率
5	16.1%	-0.033%	57.8%	0.035%	56.2%
10	18.7%	-0.048%	61.7%	0.049%	60.5%
20	20.1%	-0.063%	64.3%	0.067%	63.3%
40	20.0%	-0.078%	65.7%	0.085%	64.7%
100	19.4%	-0.098%	66.5%	0.111%	65.0%
200	19.2%	-0.120%	66.1%	0.135%	63.9%

资料来源: Wind, 天风证券研究所

2016 年以来 Spread 指标在中证 500 指数成分股内股票的平均预测表现如下表所示。指标表现出较强的预测能力，从 5 个 tick (15 秒) 到 200 个 tick (10 分钟)，IC 均值均超过 15%，多头组合具有显著的正收益，空头组合具有显著的负收益，指标对于日内短期收益具有较好的择时能力。

图 3: 000062.SZ 日收益累计序列



资料来源: Wind, 天风证券研究所

高频交易天生的敌人在于交易成本，以 000062.SZ 为例，我们逐日根据 Spread 指标滚动历史分布以及当前 Tick 指标取值确定开平仓点，平均每日约触发 30 次信号，每次信号

平均持仓周期约 27 个 Tick，每日逐笔收益均值的时间序列累计曲线如上图所示。Spread 指标对于股价的短期走势具有显著判断能力，但日内高换手策略产生的交易成本将相当巨大，要获取超额收益需要更多独立信号的组合。

### 3.3. Spread 指标的低频特征

为在低频多因子模型中引入盘口信息，我们需要将对 Spread 指标进行降频处理。为分析指标在低频结构下的特征，我们对每只股票的每日盘口相对强弱 Spread 取平均值以得到当日的整体的盘口买卖相对强弱度量值。

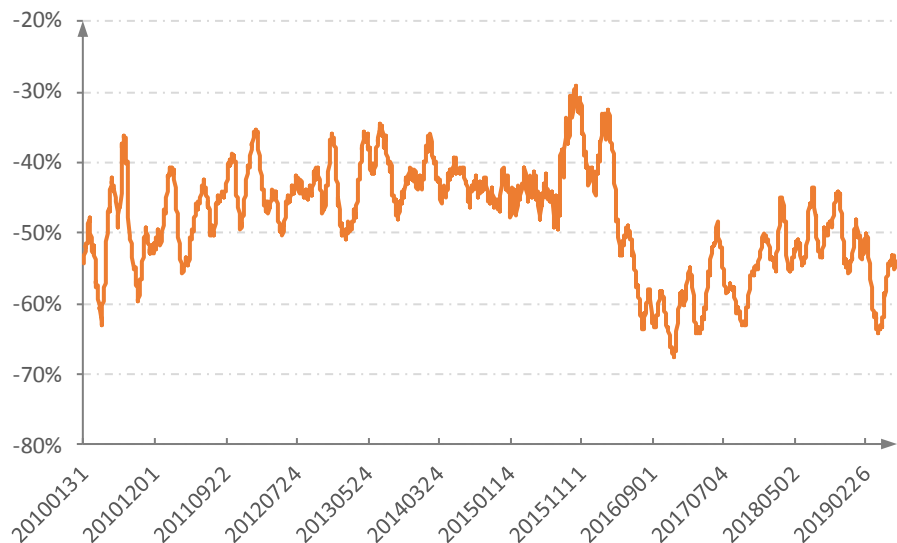
$$Spread\_Date_t = \frac{1}{N} * \sum_i Spread\_Tick_{t,i}$$

我们发现个股每日盘口强度与其当日的主动买入成交额占比具有较显著的相关性。对于股票的每日成交额，我们将股票每日成交额拆分成主动买入 *ActiveBuy* 与主动卖出 *ActiveSell* 两部分，并计算主动买入成交额占比 *BuyRate*<sup>2</sup>：

$$BuyRate = \frac{ActiveBuy}{ActiveBuy + ActiveSell}$$

下图展示了 *Spread\_Date* 与当日 *BuyRate* 截面相关系数滚动 20 日的均值，二者之间呈现出显著的负向相关性。

图 4：Spread 与 BuyRate 相关系数



资料来源：Wind，天风证券研究所

*Spread* 与 *BuyRate* 之间的负相关性在逻辑上是显然的。考虑两类极端情况：当股票涨停时盘口只存在买盘挂单和主动卖出；而股票跌停时盘口只存在卖盘挂单和主动买入。因此在日级别的频率上，盘口挂单的强弱。

因此个股盘口挂单强弱与当日的主动买入成交占比实际上可理解为对特定现象在不同维度的刻画，截面上的高相关性验证了这点。在 Tick 频率的微观结构下，盘口委买越厚则其所提供的买盘流动性越强，股票短期获取的绝对收益越高；而在日级别频率上，盘口委买挂单厚的股票实际承受了更强的主动卖压，股票在截面上的相对收益应该越低。

基于此逻辑，我们可以将 Spread 指标从 Tick 频率的时间序列指标低频化加工成日、月频率的 alpha 指标以区分个股在买卖交易盘口的承压情况。

<sup>2</sup> 该指标降频成月度 alpha 因子后表现一般，中性化因子月度 IC 均值为 1.7%。

表 3：不同频率的 Spread 指标

时间频率	因子方向	指标类型	收益类型
Tick 级别	正	时间序列因子	绝对收益
日级别	负	alpha 因子	相对收益

资料来源：天风证券研究所

## 4. 从“高频时序”到“低频 alpha”

### 4.1. 因子构造

基于以上对度量盘口买卖挂单相对强弱的时间序列指标  $Spread\_Tick$  的分析，我们希望将其降频为多因子模型可用的月频 alpha 因子。

根据上文所述我们先将指标降频到日频，但是考虑到交易中成交将对盘口挂单造成冲击，大单的主动买卖时会打薄对应盘口挂单数量，此时订单簿难以反应真实的盘口挂单强弱。

因此，对于每只股票我们每日将其所有的 Tick 行情快照按成交量排序，筛选出成交量最低的 1/2 数据。对于每个 Tick 快照，我们计算其每日盘口买卖挂单的平均相对强弱：

$$Spread\_Date_t = \frac{1}{2N} * \sum_i Spread\_Tick_{t,i} * I_{volume_{t,i} < Median\_Volume_t}$$

其中  $Spread\_Tick_{t,i}$  为交易日  $t$  的 Tick 快照  $i$  上的盘口  $Spread$ ， $volume_{t,i}$  为交易日  $t$  的 Tick 快照  $i$  上的成交额， $Median\_Volume_t$  为交易日  $t$  各个 Tick 成交量的中位数， $N$  为交易日  $t$  的快照总数； $I_A(x)$  为取值 0-1 的示性函数。

此外，个股日度盘口挂单强弱受到市场趋势的影响，在市场大涨或大跌的交易日，全市场股票的买卖挂单强弱会呈现出趋同性的分布。因此，我们在每个交易日对  $Spread\_Date_t$  指标做截面标准化以剔除市场走势的影响。

$$\widehat{Spread\_Date}_t = \frac{Spread\_Date_t - Mean(Spread\_Date_t)}{Std(Spread\_Date_t)}$$

最终将日频  $Spread$  指标降频成月频，我们利用当月所有交易日取值加权平均得到月度的  $Spread$  强度。基于指标的时效性我们赋予临近数据以更高的权重，加权平均得到月度的盘口挂单相对强弱  $Spread\_Month$ ：

$$Spread\_Month_m = \frac{1}{\sum_t w_t} * \sum_t w_t * \widehat{Spread\_Date}_t$$

其中  $\widehat{Spread\_Date}_t$  为月份  $m$  第  $t$  个交易日的盘口  $Spread$ ； $w_t = t/n$ ， $t \leq n$  为各交易日指标值的权重， $n$  为月份  $m$  的交易日数。

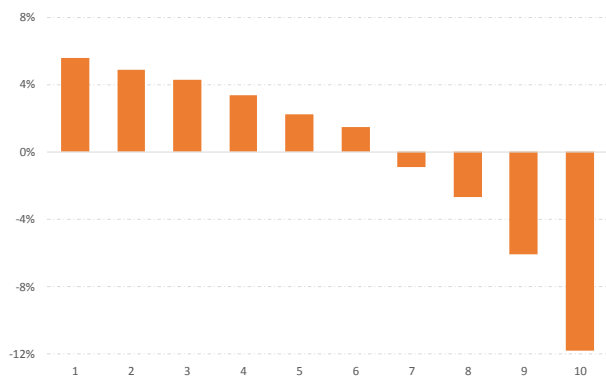
### 4.2. 因子绩效

按照上述方式构建月频的  $Spread$  指标，我们每个月末根据指标值将股票分组以检验  $Spread$  指标的选股能力，其中：

- 股票样本空间为全部 A 股剔除 ST 股票以及新股；
- 回溯时间区间为 2010 年至 2019 年 8 月；

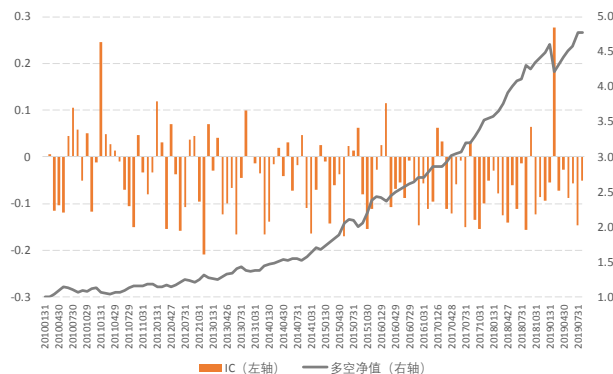


图 5: Spread 指标原始值分组收益



资料来源: Wind, 天风证券研究所

图 6: Spread 指标原始值 IC 序列



资料来源: Wind, 天风证券研究所

Spread 指标原始值从 2010 年至 2019 年 8 月底, 因子 10 分组多空组合年化收益 17.7%, 多空组合年 IR 达到 1.86, 同时因子 IC 均值 -4.5%, ICIR 达到 -1.83, IC 胜率为 71.3%, 因子表现出优秀的选股能力。

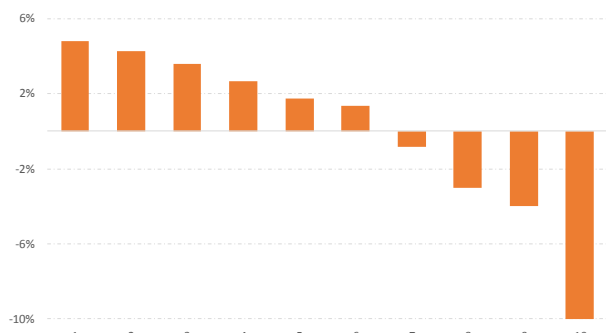
表 4: Spread 指标原始值分年绩效

年份	多空收益	多空 IR	最大回撤	IC 均值	ICIR	IC 胜率	覆盖度
2010	12.9%	1.27	-6.9%	-2.3%	-0.97	54.5%	98.9%
2011	4.8%	0.62	-2.3%	-0.8%	-0.29	58.3%	98.3%
2012	7.5%	0.78	-2.9%	-3.3%	-1.04	50.0%	98.7%
2013	12.9%	1.47	-4.8%	-5.0%	-2.18	75.0%	97.3%
2014	16.8%	2.39	-2.0%	-4.3%	-2.12	66.7%	89.1%
2015	45.1%	3.05	-4.7%	-5.8%	-2.68	75.0%	88.0%
2016	17.7%	2.79	-2.4%	-5.4%	-2.67	83.3%	92.4%
2017	24.0%	3.79	-0.3%	-6.4%	-2.88	75.0%	93.6%
2018	26.2%	4.28	-1.3%	-8.0%	-4.44	91.7%	95.1%
20190830	6.4%	0.74	-8.7%	-2.8%	-0.74	87.5%	99.8%
全样本	17.7%	1.86	-8.7%	-4.5%	-1.83	71.3%	94.9%

资料来源: Wind, 天风证券研究所

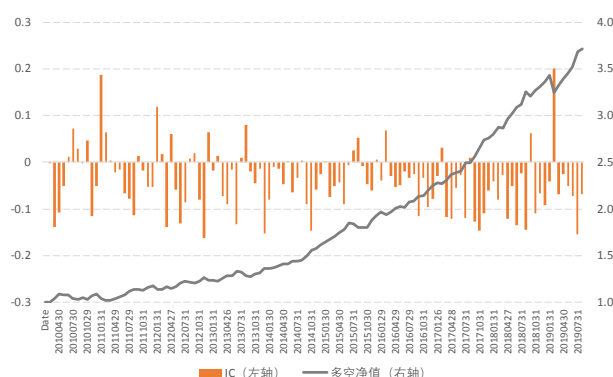
Spread 指标在截面上与行业 and 市值并无显著关系, 其与市值的相关系数均值仅为 3.8%, 但我们习惯于将指标对行业 and 市值正交, 以剔除行业 and 市值风格的影响。Spread 指标中性化后的因子分组收益、多空组合净值以及 IC 时间序列如下图所示, 中性化后因子同样表现出稳定的选股能力。

图 7: Spread 指标中性化值分组收益



资料来源: Wind, 天风证券研究所

图 8: Spread 指标中性化值 IC 序列



资料来源: Wind, 天风证券研究所

Spread 指标行业以及市值中性化后分年绩效表现如下所示,从 2010 年至 20190830 日,多空组合年化收益为 14.6%,多空组合 IR 达到 2.01;因子 IC 均值-4.0%, ICIR 达到-2.10, IC 胜率为 75.7%,因子稳定性相对原始值有较为明显提升。

表 5: Spread 指标中性化值分年绩效

年份	多空收益	多空 IR	最大回撤	IC 均值	ICIR	IC 胜率	覆盖度
2010	10.9%	1.19	-4.8%	-2.8%	-1.37	63.6%	98.7%
2011	6.7%	0.90	-2.2%	-1.2%	-0.55	66.7%	98.3%
2012	5.1%	0.66	-2.3%	-3.1%	-1.19	50.0%	98.7%
2013	9.8%	1.39	-3.6%	-3.8%	-2.03	75.0%	97.3%
2014	15.9%	3.53	-0.2%	-4.7%	-3.69	83.3%	89.1%
2015	18.9%	2.36	-3.9%	-2.5%	-2.00	66.7%	88.0%
2016	15.3%	3.22	-1.2%	-4.2%	-3.21	91.7%	92.4%
2017	25.4%	4.55	-0.2%	-7.3%	-4.29	83.3%	93.6%
2018	22.7%	3.61	-1.6%	-6.9%	-4.11	91.7%	95.1%
20190830	10.6%	1.59	-5.1%	-3.5%	-1.19	87.5%	99.8%
全样本	14.6%	2.01	-6.7%	-4.0%	-2.10	75.7%	94.9%

资料来源: Wind, 天风证券研究所

### 4.3. 分域绩效

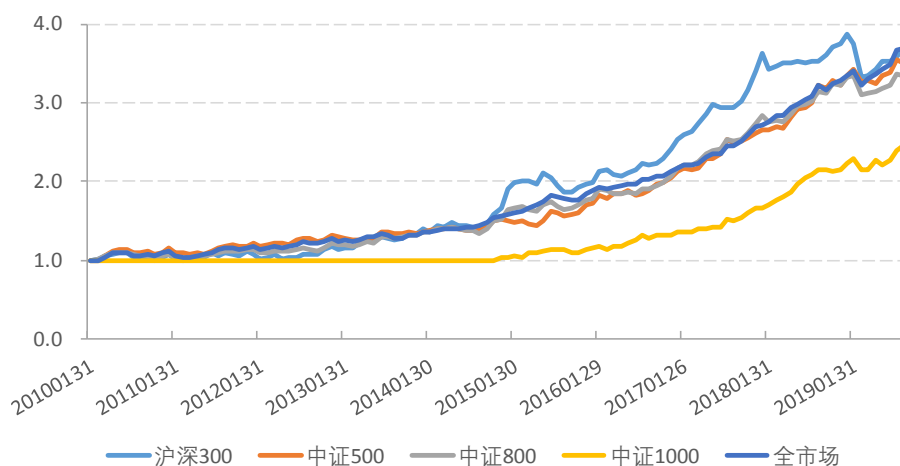
良好的选股因子应该在不同股票池中都具有较佳的选股能力,为分析 Spread 指标的分域选股能力,我们分别在沪深 300 指数、中证 500 指数、中证 800 指数、中证 1000 指数成分股中回溯因子的绩效。2010 年以来,因子绩效如下所示。

表 6: Spread 指标分域绩效

股票池	多空收益	多空 IR	最大回撤	IC 均值	ICIR	IC 胜率	覆盖度
沪深 300	14.5%	1.16	-14.0%	-4.3%	-1.29	67.0%	97.0%
中证 500	14.0%	1.49	-6.4%	-4.1%	-1.54	67.0%	96.1%
中证 800	13.4%	1.48	-8.1%	-4.1%	-1.52	67.0%	96.4%
中证 1000	20.4%	2.35	-6.0%	-4.9%	-2.42	79.3%	92.9%
全样本	14.6%	2.01	-6.7%	-4.0%	-2.10	75.7%	94.9%

资料来源: Wind, 天风证券研究所

图 9: Spread 指标分域多空净值



资料来源: Wind, 天风证券研究所

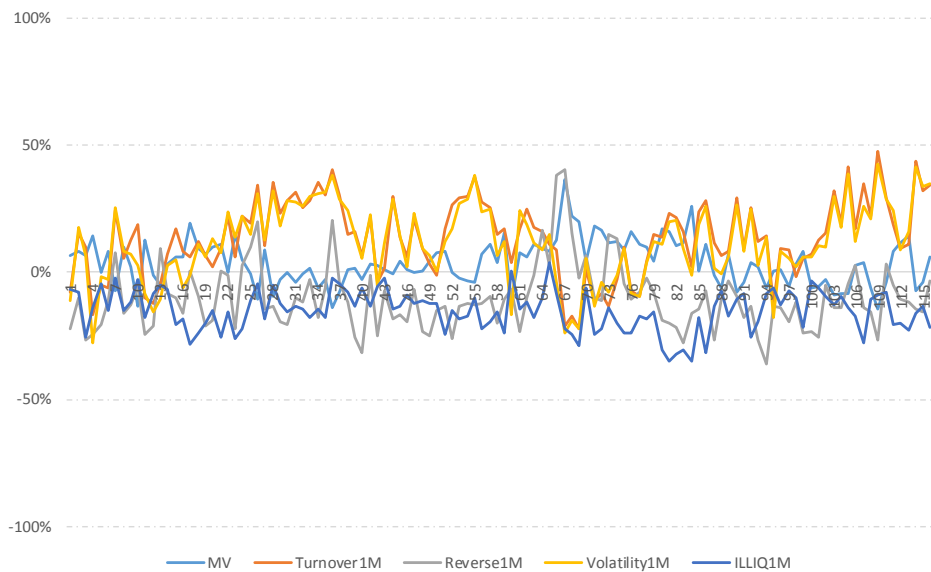
Spread 指标在沪深 300 指数、中证 500 指数、中证 800 指数以及中证 1000 指数成分股中均展现出较佳的选股能力，因子在各股票池中 IC 均值均超过 4%，多空净值曲线稳定向上，因子选股能力表现出较强的稳健性。

#### 4.4. 因子增量信息

量价类因子的挖掘难点在于新指标提供的边际增量信息往往很有限，因子 alpha 来源能被常见的量价类因子所解释。

月频 alpha 模型中涉及到量价类选股指标一般包括反转、波动、换手、流动性等，我们分别计算 Spread 指标与 20 日反转、20 日波动率、20 日换手、20 日非流动性冲击的相关性，如下图所示 Spread 指标与四个常见量价指标具有较低的相关性，截面相关系数均值分别为 13.6%，-10.4%、11.7%、-15.1%，因子表现出有较高的独立性。

图 10: Spread 指标与常见量价因子相关性



资料来源：Wind，天风证券研究所

为进一步分析 Spread 指标所提供的增量 alpha 信息，我们将其与常见的量价类指标复合，对比复合因子在加入 Spread 指标前后选股能力的变化。

表 7: 指标增量信息分析

年份	四因子				五因子			
	多空收益	多空 IR	IC 均值	ICIR	多空收益	多空 IR	IC 均值	ICIR
2011	24.9%	3.49	8.9%	4.01	23.3%	2.98	8.0%	3.88
2012	51.9%	7.19	10.6%	4.88	49.8%	6.65	10.7%	5.04
2013	26.9%	3.42	7.9%	3.61	24.0%	3.86	8.3%	4.79
2014	35.9%	2.95	10.2%	4.46	40.3%	3.36	10.4%	5.46
2015	53.6%	2.36	10.1%	3.39	57.5%	2.97	9.5%	3.80
2016	54.5%	4.78	14.0%	7.81	52.7%	4.50	14.3%	7.82
2017	36.1%	5.68	9.7%	4.66	42.6%	4.72	11.4%	5.38
2018	35.1%	2.61	9.7%	3.44	37.5%	4.11	10.7%	5.19
20190830	25.9%	2.51	10.2%	2.85	24.7%	2.24	9.7%	2.78
全样本	39.6%	3.24	10.1%	4.22	40.4%	3.54	10.4%	4.78

资料来源：Wind，天风证券研究所

我们将 20 日反转、20 日波动率、20 日换手、20 日非流动性冲击四个量价类因子进行行业 and 市值中性化并作对称正交处理，以过去 12 个月 IC-IR 为权重复合各指标。四因子的复合指标作为对照组，加入 Spread 指标后的五因子作为实验组。

从上表可知，Spread 指标能提供较好的独立于常见量价类因子的增量信息，复合因子中混入该指标后多空 IR 从 3.24 提高到 3.54，ICIR 从 4.22 提高到 4.78，因子整体选股能力有较明显提升。

## 5. 总结

以往的研究对于 A 股市场的日间低频数据分析已有较为深厚的积累，但对日内微观情境下的市场结构特征我们仍缺乏足够的认知。本文希望借助高频数据对微观交易特征进行初步探索；在此基础上我们希望从高频的市场微观特征中归纳总结出增量的低频 alpha 信息。

若短期价格取决于当前多头需求量和空头供给量构建的均衡价格，那么盘口信息将是判断个股短期走势的重要依据。根据每个 Tick 上的买卖挂单强度，我们构建了 Spread 指标以度量当前时间的盘口买卖挂单的强弱差异。从 5 个 Tick 到 200 个 Tick 的预测窗口，我们发现 Spread 指标对股价短期走势具有显著预测能力，在买盘挂单相对越厚的时点股价短期上行概率将越大。

将数据调整为日频后，我们发现 Spread 指标在日频率与 Tick 频率窗口下的含义具有明显差异。在日频率上，盘口委买挂单相对越厚的股票在当日实际承受了越强的主动卖压，二者之间相关性显著，因此这部分股票在截面上的相对收益应该越低。

我们将高频数据低频化后构建了月度的 Spread 选股因子，从 2010 年至 2019 年 8 月 30 日的回溯区间内，因子表现出显著的选股能力。因子中性化后多空组合年化收益为 14.6%，多空组合 IR 达到 2.01；因子 IC 均值 -4.0%，ICIR 达到 -2.10，IC 胜率为 75.7%。

我们分别在沪深 300 指数、中证 500 指数、中证 800 指数以及中证 1000 指数成分股中回溯了因子的选股能力，各股票池中因子 IC 均值均超过 4.0%，因子选股能力呈现出较强的稳健性。此外，Spread 因子相对于反转、波动、换手、流动性等量价指标具有较高的独立性，截面相关系数均值分别为 13.6%，-10.4%、11.7%、-15.1%，因子能带来较显著的增量信息。

## 分析师声明

本报告署名分析师在此声明：我们具有中国证券业协会授予的证券投资咨询执业资格或相当的专业胜任能力，本报告所表述的所有观点均准确地反映了我们对标的证券和发行人的个人看法。我们所得报酬的任何部分不曾与，不与，也将不会与本报告中的具体投资建议或观点有直接或间接联系。

## 一般声明

除非另有规定，本报告中的所有材料版权均属天风证券股份有限公司（已获中国证监会许可的证券投资咨询业务资格）及其附属机构（以下统称“天风证券”）。未经天风证券事先书面授权，不得以任何方式修改、发送或者复制本报告及其所包含的材料、内容。所有本报告中使用的商标、服务标识及标记均为天风证券的商标、服务标识及标记。

本报告是机密的，仅供我们的客户使用，天风证券不因收件人收到本报告而视其为天风证券的客户。本报告中的信息均来源于我们认为可靠的已公开资料，但天风证券对这些信息的准确性及完整性不作任何保证。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，天风证券及/或其关联人员均不承担任何法律责任。

本报告所载的意见、评估及预测仅为本报告出具日的观点和判断。该等意见、评估及预测无需通知即可随时更改。过往的表现亦不应作为日后表现的预示和担保。在不同时期，天风证券可能会发出与本报告所载意见、评估及预测不一致的研究报告。

天风证券的销售人员、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。天风证券没有将此意见及建议向报告所有接收者进行更新的义务。天风证券的资产管理部门、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。

## 特别声明

在法律许可的情况下，天风证券可能会持有本报告中提及公司所发行的证券并进行交易，也可能为这些公司提供或争取提供投资银行、财务顾问和金融产品等各种金融服务。因此，投资者应当考虑到天风证券及/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一参考依据。

## 投资评级声明

类别	说明	评级	体系
股票投资评级	自报告日后的 6 个月内，相对同期沪深 300 指数的涨跌幅	买入	预期股价相对收益 20%以上
		增持	预期股价相对收益 10%-20%
		持有	预期股价相对收益 -10%-10%
		卖出	预期股价相对收益 -10%以下
行业投资评级	自报告日后的 6 个月内，相对同期沪深 300 指数的涨跌幅	强于大市	预期行业指数涨幅 5%以上
		中性	预期行业指数涨幅 -5%-5%
		弱于大市	预期行业指数涨幅 -5%以下

## 天风证券研究

北京	武汉	上海	深圳
北京市西城区佟麟阁路 36 号	湖北武汉市武昌区中南路 99 号保利广场 A 座 37 楼	上海市浦东新区兰花路 333 号 333 世纪大厦 20 楼	深圳市福田区益田路 5033 号平安金融中心 71 楼
邮编：100031	邮编：430071	邮编：201204	邮编：518000
邮箱：research@tfzq.com	电话：(8627)-87618889	电话：(8621)-68815388	电话：(86755)-23915663
	传真：(8627)-87618863	传真：(8621)-68812910	传真：(86755)-82571995
	邮箱：research@tfzq.com	邮箱：research@tfzq.com	邮箱：research@tfzq.com