

# 建投金工深度专题104：高频订单失衡及价差因子

主要观点

本文概述

本文利用高频的逻辑挖掘出盘口数据中有价值的信息，并将其处理得到两大类共14个高频因子（逐档订单失衡率类因子、中间价变化率类因子），最后降为月频的低频选股因子，在单因子回测中取得优秀的选股效果。其中MPC5\_neut因子IC均值-7.26%，年化IR为-3.09，年化多空收益达到30.63%，夏普比率高达2.88，总体选股效果是所有因子里最好的。

高频量价选股因子定义

我们通过高频数据构造出一些高频量价选股因子，第一类因子叫逐档订单失衡率（SOIR类因子），为订单簿各档委托买入和委托卖出量之间的差异占总量的比例。第二类因子叫中间价变化率因子（MPC类因子），为中间价的短期百分比变化率的均值、日频最大值、日频偏度。

SOIR和MPC因子在低频化后出现逻辑反转

我们看到SOIR类因子在高频上与收益率正相关，MPC1因子与收益率正相关，MPC5因子与收益率负相关。然而将高频量价因子降频后，SOIR类因子与MPC类因子均与收益率负相关。我们从以下两点原因来解释。从散户来看，在短期内散户容易存在追高杀跌行为。短期追高，价格上涨，但随着时间的累积，价格会逐渐处于高位，长期来看价格会回落。从主力的角度，主力对市场的短时操纵造成了价格的涨跌。强的买卖压力一般是大单交易造成的，大单交易很可能是主力的“对倒”行为，其目的主要是吸引散户，此时高频因子与收益率呈正相关。但从长期来看，市场价格则会回落，因此造成了低频上因子与收益率呈反向关系。

MPC5\_neut因子IC均值-7.26%，年化多空收益高达30.63%

最后我们对SOIR和MPC两大类共14个高频量价因子进行单因子分析。具体回测时间为最近10年（2010年1月-2020年7月），样本池为全市场，月频调仓。SOIR类因子不做中性化处理，MPC类因子分别做和不做中性化处理进行对比。其中表现较好的因子有：SOIR3因子IC均值-4.86%，年化IR-1.63，年化多空收益20.46%，夏普比率1.94。SOIR4因子IC均值-5.13%，年化IR-1.95，年化多空收益21.81%，夏普比率2.28。SOIR5因子IC均值-5.37%，年化IR-2.29，年化多空收益21.32%，夏普比率2.41。MPC1\_neut因子IC均值-6.81%，年化IR-2.87，年化多空收益26.99%，夏普比率2.66。MPC5\_neut因子IC均值-7.26%，年化IR-3.09，年化多空收益30.63%，夏普比率2.88，总体选股效果是所有因子里最好的，在传统量价因子里面也是罕见的。

报告主体内容

一、高频订单失衡及价差因子定义和投资逻辑

1.1

高频订单失衡及价差因子研究引言

日内盘口数据主要包括委托买卖价格以及委托交易量，其反应了当前时刻的市场供给和需求强度。传统的学术研究和高频因子计算中通常只利用到了买一档和卖一档的盘口信息，但实际上，盘口中第一档以外的数据也包含了许多有价值的信息。本文主要利用各档盘口数据来构造订单失衡率因子，比较盘口各档信息对未来股票价格的预测能力，并通过加权合成为一个包含5档信息的订单失衡率因子。

另外日内盘口中间价变动的分布中蕴含着对未来价格的预测信息，学术研究表明，市场投资者们可能对盘口中间价的极端变动情况也比较感兴趣，偏度可以反应数据分布偏离均值的程度，而最大值则可以直接刻画数据的极端值，因此本文除了研究盘口中间价的百分比变动的均值，还研究了盘口中间价百分比变动的偏度和最大值因子。

1.2

## 高频量价因子1（逐档订单失衡率Step Order Imbalance Ratio）定义和投资逻辑

在第一篇高频量价选股因子研究中（具体参考《因子深度研究系列：高频量价选股因子初探》），我们提出了一个衡量订单不平衡性质的OIR因子，在本文中，为了分别研究盘口各档的订单不平衡因子的选股能力，我们提出了SOIR<sub>i</sub>(i=1,2,3,4,5)这一组逐档订单失衡率因子，其中每个因子仅包含第*i*档的信息。最后，为了充分利用盘口各档数据信息，我们采用衰减加权的方法对SOIR<sub>i</sub>加权得到SOIR因子，根据对买卖压力的影响力的不同将不同档位赋予相应权重，赋予靠近交易价格档位更高的权重。具体公式如下：

由于订单簿上的委托量反应了交易者们对于股票未来价格的预期，当交易者预期股票未来价格上升，他们将通过下买单持有更多的股票多头头寸，这将导致买盘的委托量增加，反之亦然。因此，买卖盘委托量的不平衡可以反应市场的总体情绪和方向。

SOIR<sub>i</sub>衡量了买卖委托量不均衡程度在其总量中的占比，同OIR一样，SOIR<sub>i</sub>也采用比率的相对强弱形式，能够区分买卖委托量绝对值差异大但比率小的情况。SOIR为SOIR<sub>i</sub>的衰减加权合成因子，根据学术研究和普遍市场认知，靠近交易价格档位的SOIR<sub>i</sub>对后续的收益率的短期解释力应该更强，因此在因子合成时我们赋予靠近交易价格档位的SOIR<sub>i</sub>更高权重。另外对比OIR，SOIR采用各档订单失衡率的加权，而不是各档加权委托量的失衡率，可以避免某一档订单量过大对总体比率的影响，是对OIR的一种改进尝试。SOIR反应了盘口各档综合的买卖委托量不均衡程度，SOIR为正说明市场买压大于卖压，未来价格趋向上涨，且SOIR的值越大，上涨的概率越高，反之亦然。

图1为2020年7月1日9时57分股票A的盘口情况，根据定义计算出SOIR1 = 0.86, SOIR2 = 0.93, SOIR3 = 0.68, SOIR4 = -0.31, SOIR5 = 0.20, SOIR = 0.64。可以看出，SOIR1到SOIR3显著为正，释放了较强的买入信号，而SOIR4和SOIR5买入信号转弱，且SOIR4释放了卖出信号，而最终加权后的SOIR因子也是有较强的买入信号，该股票在下一分钟的涨幅高达3.57%。这个例子说明越靠近交易价格档位的影响力更大，在因子合成时应该赋予靠近交易价格的档位更高的权重。

另外从全天的分钟频数据来看（图2），SOIR分钟因子在当天9时57分达到高点0.64，股票A在下一分钟上涨了3.57%，并一路拉升，9点59分时达到涨停。因此这个例子说明了SOIR因子短期内对股价有非常显著的正向影响。

### 1.3

## 高频量价因子2（中间价变化率Midpoint Price Change）定义和投资逻辑

在限价订单簿市场中，交易者的三种行为可能会引起市场中间价的变动：（1）交易者发起了主动成交，且交易量大于第一档的委托量，使第一档的价格发生改变；（2）交易者撤销了第一档上的全部委托订单；（3）交易者在买一价和卖一价之间新增了新的限价委托订单。

MPC因子衡量的是市场中间价的短期变动趋势，该趋势刻画了市场交易者的最新交易和挂单撤单行为，反应了市场交易者对股票价格未来走势的最新预期。例如，当市场中间价出现上升时，此时可能是发生了大量的主动买入交易，使得卖一价上升，或是有交易者以高于当前买一价的价格下了限价订单，使得买一价上升，这两种情况都反应了市场交易者对未来股价走势的乐观预期。因此，MPC为正说明股票未来短期价格趋向上涨，且MPC的值越大，其上涨的概率越高，反之亦然。具体公式如下：

图3分别为2020年7月1日9时51分、9时52分股票B的盘口情况，根据定义计算，9时52分的MPC1=2.23%，从图4的成交量数据中可以看出，这段时间内盘口中间价的大幅变动是由成交量突增（经过检验是主动买入量突增引起的）造成的。该股票在下一分钟的收益率高达1.34%。

## 二、高频转低频的方法和逻辑

### 2.1

## 高频量价因子转低频的构造方法

### 2.2

#### 高频量价因子高频和低频IC对比

下面我们看下各因子的分钟IC均值和月频IC均值。由下表可以看出，高频SOIRi的IC值均显著为正，且随着档数的升高IC绝对值降低，这一结果符合我们的预期以及学术研究的结论，这也是为什么我们在做因子加权时给SOIR1最高的权重，SOIR最低的权重。另一方面，低频SOIRi的IC值方向出现了反转，这一现象与我们第一篇高频量价选股因子研究提出的OIR因子相同，在之前研究中，我们也指出这是由散户存在追涨杀跌以及主力的短时市场操纵引起的，特别是用委托量构造的高频因子均会出现这种反转逻辑。

由上表可以看出，在高频（分钟）级别上，MCI\_B和MCI\_A因子方向和前面逻辑完全一致，MCI\_A和未来短期收益为负向关系，MCI\_B和未来短期收益为正向关系。

在低频（月频）级别上，MCI\_B和MCI\_A因子方向均显著为正，主要是流动性因子在长期来看具有风险溢价，因而与未来收益为正向关系。

### 三、高频订单失衡及价差因子和常用因子的相关性

下面我们看下SOIR和MPC两大类共14个高频量价因子和常用选股因子的IC相关性。

经过统计，我们发现这些因子和自由流通市值（LnFloatCap）的相关性较低，因此对于这些因子可以不做市值中性的处理。

另外，SOIR类因子和BP\_LR、SP\_TTM等估值因子与Momentum\_6m、Momentum\_12m等动量反转因子的相关性较高。

最后MPC1、MPC5因子和Momentum\_3m、Momentum\_6m等动量反转因子相关性较高，MPC1\_max、MPC5\_max因子和BP\_LR、SP\_TTM等估值因子、TurnoverAvg1M、Volatility1M等技术因子的相关性较高，而MPC1\_skew、MPC5\_skew因子和常用因子的相关性均不高。

表 3：高频订单失衡及价差因子和常用因子的相关性<sup>②</sup>

	SOIR1	SOIR2	SOIR3	SOIR4	SOIR5	SOIR	MPC1	MPC5	MPC1_max	MPC5_max	MPC1_skew	MPC5_skew
EP_TTM	-0.28	-0.19	-0.11	-0.02	0.11	-0.14	-0.05	-0.04	-0.70	-0.62	-0.28	-0.13
BP_LR	<b>-0.72</b>	<b>-0.66</b>	<b>-0.51</b>	-0.29	0.01	<b>-0.56</b>	-0.08	-0.08	<b>-0.64</b>	<b>-0.58</b>	-0.27	0.09
SP_TTM	<b>-0.65</b>	<b>-0.60</b>	<b>-0.47</b>	-0.27	0.00	<b>-0.51</b>	-0.08	-0.08	<b>-0.71</b>	<b>-0.64</b>	-0.27	0.08
Earnings_SQ_YoY	0.31	0.34	0.33	0.28	0.19	0.34	0.08	0.09	-0.22	-0.21	-0.17	-0.29
Sales_SQ_YoY	<b>0.51</b>	<b>0.49</b>	0.41	0.28	0.08	0.44	0.04	0.05	0.15	0.15	0.05	-0.20
ROE_SQ_YoY	0.35	0.44	0.43	0.36	0.25	0.43	0.21	0.22	-0.11	-0.13	-0.08	-0.24
ROE_TTM	0.18	0.25	0.23	0.18	0.12	0.23	-0.01	0.00	-0.38	-0.33	-0.14	-0.22
ROA_TTM	0.45	<b>0.49</b>	0.42	0.29	0.11	0.44	0.03	0.03	-0.14	-0.13	-0.05	-0.28
ROIC_TTM	0.37	0.42	0.37	0.26	0.11	0.38	0.02	0.03	-0.21	-0.19	-0.08	-0.27
Momentum_1m	0.14	0.40	0.48	<b>0.51</b>	0.50	0.42	0.83	0.84	0.25	0.25	0.37	0.37
Momentum_3m	0.24	0.45	<b>0.50</b>	0.49	0.43	<b>0.46</b>	<b>0.58</b>	<b>0.59</b>	0.20	0.21	0.17	0.13
Momentum_6m	0.46	<b>0.59</b>	0.60	0.52	0.38	<b>0.58</b>	<b>0.50</b>	<b>0.51</b>	0.22	0.23	0.14	0.00
Momentum_12m	<b>0.57</b>	0.67	0.64	<b>0.53</b>	0.35	<b>0.64</b>	0.42	0.43	0.25	0.25	0.12	-0.05
Momentum_24m	<b>0.66</b>	<b>0.69</b>	0.62	0.48	0.27	<b>0.65</b>	0.28	0.27	0.29	0.29	0.12	-0.11
LnFloatCap	-0.07	0.01	0.04	0.05	0.08	0.02	0.03	0.06	-0.39	-0.27	-0.09	-0.01
AmountAvg_1M	0.16	0.13	0.04	-0.07	-0.19	0.07	-0.08	-0.04	0.13	0.30	0.12	-0.02
TurnoverAvg1M	0.41	0.27	0.12	-0.06	-0.27	0.18	-0.05	-0.04	0.80	<b>0.85</b>	0.33	0.02
TurnoverAvg3M	0.42	0.25	0.09	-0.09	-0.30	0.15	-0.13	-0.12	0.77	<b>0.81</b>	0.31	0.02
TurnoverAvg6M	0.42	0.24	0.07	-0.11	-0.32	0.15	-0.15	-0.14	0.74	<b>0.77</b>	0.31	0.02
Volatility1M	0.44	0.32	0.17	-0.01	-0.22	0.23	-0.08	-0.07	0.82	0.88	0.32	-0.01
Volatility3M	0.43	0.27	0.11	-0.07	-0.28	0.18	-0.17	-0.16	0.79	<b>0.85</b>	0.30	0.01
Volatility6M	0.44	0.24	0.07	-0.10	-0.33	0.15	-0.23	-0.22	0.74	<b>0.79</b>	0.28	-0.02
Beta_100W	-0.36	<b>-0.50</b>	<b>-0.52</b>	-0.47	-0.36	<b>-0.50</b>	-0.30	-0.29	-0.02	0.10	0.08	0.14
SOIR1	1.00	<b>0.91</b>	<b>0.79</b>	<b>0.62</b>	0.33	<b>0.86</b>	0.14	0.13	0.39	0.39	0.17	-0.15
SOIR2	<b>0.91</b>	1.00	0.96	<b>0.84</b>	0.60	<b>0.98</b>	0.42	0.41	0.41	0.39	0.25	0.00

SOIR3	0.79	0.96	1.00	0.95	0.78	0.99	0.53	0.52	0.32	0.29	0.23	0.07
SOIR4	0.62	0.84	0.95	1.00	0.93	0.92	0.59	0.57	0.19	0.15	0.18	0.14
SOIR5	0.33	0.60	0.78	0.93	1.00	0.72	0.60	0.57	0.00	-0.04	0.10	0.21
SOIR	0.86	0.98	0.99	0.92	0.72	1.00	0.46	0.45	0.34	0.32	0.22	0.02
MPC1	0.14	0.42	0.53	0.59	0.60	0.46	1.00	0.99	0.25	0.23	0.48	0.51
MPC5	0.13	0.41	0.52	0.57	0.57	0.45	0.99	1.00	0.26	0.24	0.49	0.51
MPC1_max	0.39	0.41	0.32	0.19	0.00	0.34	0.25	0.26	1.00	0.97	0.50	0.22
MPC5_max	0.39	0.39	0.29	0.15	-0.04	0.32	0.23	0.24	0.97	1.00	0.53	0.28
MPC1_skew	0.17	0.25	0.23	0.18	0.10	0.22	0.48	0.49	0.50	0.53	1.00	0.81
MPC5_skew	-0.15	0.00	0.07	0.14	0.21	0.02	0.51	0.51	0.22	0.28	0.81	1.00

#### 四、高频订单失衡及价差因子测试结果

最后我们对SOIR和MPC两大类共14个高频量价因子进行单因子分析（包括IC分析和多空收益分析）。具体回测时间为最近10年（2010年1月-2020年7月），样本池为全市场，每月底剔除停牌、一字板、上市未满半年和ST股票，月频调仓。因子做了极值处理（剔除3倍标准差之外的样本）和缺失值处理（直接剔除）。SOIR类因子不做中性化处理，MPC类因子分别做和不做中性化处理进行对比。组合的多空收益分位数用10分位。

##### 4.1

##### SOIR1因子选股效果

首先是SOIR1因子的效果，因子IC均值-2.68%，年化IR-0.94，年化多空收益8.63%，夏普比率0.88，总体选股效果较为一般。

##### 4.2

##### SOIR2因子选股效果

接着我们看下SOIR2因子的选股效果，因子IC均值-4.18%，年化IR-1.27，年化多空收益17.89%，夏普比率1.52，总体选股效果相比SOIR1显著提升，选股效果非常显著。

##### 4.3

##### SOIR3因子选股效果

然后是SOIR3因子，因子IC均值-4.86%，年化IR-1.63，年化多空收益20.46%，夏普比率1.94，选股效果相比SOIR1和SOIR2因子进一步提升。

##### 4.4

##### SOIR4因子选股效果

对于SOIR4因子，因子IC均值-5.13%，年化IR-1.95，年化多空收益21.81%，夏普比率2.28，选股效果优于SOIR1到SOIR3。

##### 4.7

##### MPC1因子选股效果

对于中间价变化率因子MPC，首先是MPC1因子的选股效果，因子IC均值-5.36%，年化IR-1.54，年化多空收益19.52%，夏普比率1.42，选股效果显著。

## 高频订单失衡及价差因子在指定样本池内的测试结果

我们检测了SOIR和MPC两大类共14个高频量价因子在指定样本池的测试效果，首先在中证500指数样本内做测试，下表是各因子的多空收益分析。

我们看到几乎所有因子在中证500样本内的年化多空收益都在10%以上，部分因子在15%以上，表现最好的仍然是MPC5和MPC5\_neut因子，年化多空收益接近18%，因此SOIR和MPC类因子在中证500样本内的选股效果仍然非常优秀。

然后我们检测SOIR和MPC两大类共14个高频量价因子在沪深300指数样本内的效果（多空收益分析），结果如下表。

我们看到大部分因子在沪深300样本内的年化多空收益都接近10%或在10%以上，表现最好的是MPC5\_skew因子，年化多空收益达12.35%，因此SOIR和MPC类因子在沪深300样本内的选股效果也是非常不错的。

## 五、总结和思考

高频数据中蕴含了丰富的市场交易信息，它能带我们通过数据窥探知情交易者的隐藏信息，也让我们更近距离地感受市场交易者的情绪，从而帮助我们更准确地拿捏市场股票价格的走势。本文利用高频的逻辑挖掘出盘口数据中有价值的信息，并将其处理得到高频因子，最后降为月频的低频选股因子，在后续的因子回测中取得良好的选股效果。

第一部分主要通过高频分钟数据构造出一些高频量价选股因子，第一类因子叫逐档订单失衡率（SOIR类因子），订单簿上的委托量反应了交易者们对于股票未来价格的预期，当交易者预期股票未来价格上升，他们将通过买单持有股票的多头头寸，这将导致买盘的委托量增加，反之亦然。SOIR类因子衡量了买卖委托量不均衡程度在其总量中的占比，反应了市场的总体情绪和方向。第二类因子叫中间价变化率因子（MPC类因子），为中间价的短期百分比变化率的均值、日频最大值、日频偏度，MPC因子衡量的是市场中间价的短期变动趋势，该趋势刻画了市场交易者的最新交易和挂撤单行为，反应了市场交易者对股票价格未来走势的最新预期。

第二部分我们采用具体流程把高频因子转为我们常用的月度低频选股因子。首先因为股票的盘口挂单强弱受到市场总体走势的影响，因此我们需要对各股票进行截面标准化以剔除市场对个股的影响。然后我们把标准化后的分钟因子转换成日因子，我们采用了等权的方法。最后我们把日因子转换成月因子，我们按距离每月最后一个交易日（假设为组合调仓日）的时间远近进行加权，考虑到信息的时效性，距离调仓日越远其信息的有效性越弱，因此用衰减加权的方法对日因子加权。

第三部分我们分析各因子的分钟IC均值和月频IC均值。我们看到SOIR类因子在高频上与收益率正相关，且随着档数的升高IC绝对值降低，这一结果符合我们的预期以及学术研究的结论。MPC1因子与收益率正相关，MPC5因子则与收益率负相关，这说明中间价的变动在1分钟时为动量效应，5分钟时则出现反转。然而将高频量价因子降频后，SOIR类因子与MPC类因子均与收益率负相关。

我们从以下两点原因来解释。从散户来看，在短期内散户容易存在追高杀跌行为。短期追高，价格上涨，但随着时间的累积，价格会逐渐处于高位，长期来看价格会回落。从主力的角度，主力对市场的短时操纵造成了价格的涨跌。强的买卖压力一般是大单交易造成的，大单交易很可能是主力的“对倒”行为，其目的主要是吸引散户，此时高频因子与收益率呈正相关。

第四部分我们检测了SOIR和MPC两大类共14个高频量价因子和常用选股因子的IC相关性。经过统计发现这些因子和自由流通市值（LnFloatCap）的相关性较低。另外，SOIR类因子和BP\_LR、SP\_TTM等估值因子与Momentum\_6m、Momentum\_12m等动量反转因子的相关性较高。最后我们看到MPC1、MPC5因子和Momentum\_3m、Momentum\_6m等动量反转因子相关性较高，MPC1\_max、MPC5\_max因子和BP\_LR、SP\_TTM等估值因子、TurnoverAvg1M、Volatility1M等技术因子的相关性较高，而MPC1\_skew、MPC5\_skew因子和常用因子的相关性均不高。

第五部分对SOIR和MPC两大类共14个高频量价因子进行单因子分析。具体回测时间为最近10年（2010年1月-2020年7月），样本池为全市场，月频调仓。SOIR类因子不做中性化处理，MPC类因子分别做和不做中性化处理进行对比。其中表现较好的有：SOIR3因子IC均值-4.86%，年化IR-1.63，年化多空收益20.46%，夏普比率1.94。SOIR4因子IC均值-

5.13%，年化IR-1.95，年化多空收益21.81%，夏普比率2.28。SOIR5因子IC均值-5.37%，年化IR-2.29，年化多空收益21.32%，夏普比率2.41。MPC1\_neut因子IC均值-6.81%，年化IR-2.87，年化多空收益26.99%，夏普比率2.66。MPC5\_neut因子IC均值-7.26%，年化IR-3.09，年化多空收益30.63%，夏普比率2.88，总体选股效果是所有因子里最好的。

第六部分测试了这些因子在中证500、沪深300样本池内的表现，我们看到几乎所有因子在中证500样本内的年化多空收益都在10%以上，部分因子在15%以上，表现最好的仍然是MPC5和MPC5\_neut因子，年化多空收益接近18%。大部分因子在沪深300样本内的年化多空收益都接近10%或在10%以上，表现最好的是MPC5\_skew因子，年化多空收益达12.35%，因此SOIR和MPC类因子在中证500、沪深300样本内的选股效果仍然非常不错。

## 参考文献

Jiang Lei. Order Imbalance, Liquidity, and Market Efficiency: Evidence from the Chinese Stock Market. *Managerial and Decision Economics*, 32(7): 469-480, 2011.

Roberto Pascual, Bartolomé Pascual-Fuster. The relative contribution of ask and bid quotes to price discovery. *Journal of Financial Markets*, 20:129–150, 2014.

Turan G.Bali, Nusret Cakici, and Robert F.Whitelaw. Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics*, 99:427-446, 2011.

Diego Amaya, Peter Christoffersen, and Kris Jacobs. Does realized skewness predict the cross-section of equity returns? *Journal of Financial Economics*, 118:135-167, 2015.

详细内容可以参考中信建投金融工程组报告：《高频订单失衡及价差因子》

中信建投金融工程深度专题报告回顾

【资产配置】

【因子选股】

【交易策略与衍生品】

【基金产品研究与FOF】

报告联系人

免责声明：