高频因子

模型研报来自于海通证券：《选股因子系列研究（七十六）——基于深度学习的高频因子挖掘》
“本文基于分钟 K 线数据、盘口委托挂单数据、逐笔成交数据构建得到了 164 个 30”

其中高频因子构建的研报汇总在《选股因子系列研究（六十四）——基于直观逻辑和机器学习的高频数据低频化应用》表1
分钟频率的指标序列，并将其作为模型的输入。”

高频因子可以分为收益率分布、成交量分布，量价复合、资金流和日内动机。 cr -《高频量价因子在股票与期货中的表现》
以下整理的只是来自研报的内容，部分公式可以对其他数据互相套用。比如收益率和成交量

分钟级别数据

注：本节选取的高频因子都是基于分钟级别的数据(除非有注明)。来自的研报(海通证券)如下:
1、《选股因子系列研究（十九）——高频因子之股票收益分布特征》
2、《选股因子系列研究（二十五）——高频因子之已实现波动分解》
3、《高频量价因子在股票与期货中的表现》
4、《选股因子系列研究（四十六）——日内分时成交中的玄机》

1、
(1) 高频收益方差： (其中r_{ij}为股票i在第j分钟的收益)

$$\mathrm{RVar}_i = \sum_{j=1}^{N} r_{ij}^2$$

(2) 高频收益偏度:

$$R_{\mathrm{Rexw}i} = \frac{\sqrt{N}\sum_{j=1}^{N} r_{ij}^2}{\mathrm{RVar}_i^{3/2}}$$

(3) 高频收益峰度:

$$\mathrm{RKurt}_i = \frac{N\sum_{j=1}^{N} r_{ij}^4}{RVar_i^2}$$

2、
按“系统波动+特质波动”的拆分，通过Fama-French回归将股票收益分解为系统收益和特质收益。

$$r_i = \alpha + \beta_{MKT} MKT + \beta_{SMB} SMB + \beta_{HML} HML + \varepsilon_i$$

其中，r_i为股票收益，MKT为市场收益，SMB为市场溢价，HML为估值溢价，回归残差为股票特质收益。对高频的横截面作回归即可。

(4) 高频波动： (和(1)的高频收益方差一样，其中r_i^t是股票i在t时刻的收益)

$$\text{高频波动} = \left(\sum_t \left(r_i^t\right)^2\right)^{\frac{1}{2}}$$

(5) 高频特质波动:

$$\text{高频特质波动} = \left(\sum_t \left(\varepsilon_i^t \right)^2 \right)^{\frac{1}{2}}$$

(6) 高频系统波动：

$$\text{高频系统波动} = \left(\sum_t \left(r_i^t - \varepsilon_i^t \right)^2 \right)^{\frac{1}{2}}$$

(7) 高频特异度：

$$\text{高频特异度} = \frac{\sum_t \left(\varepsilon_i^t \right)^2}{\sum_t \left(r_i^t \right)^2}$$

(8) 高频上行波动：

$$\text{高频上行波动} = \left(\sum_t \left(r_i^t I_{\{r_i^t > 0\}} \right)^2 \right)^{\frac{1}{2}}$$

(9) 高频下行波动：

$$\text{高频下行波动} = \left(\sum_t \left(r_i^t I_{\{r_i^t < 0\}} \right)^2 \right)^{\frac{1}{2}}$$

(10) 高频上行波动占比：

$$\text{高频上行波动占比} = \frac{\sum_t \left(r_i^t I_{\{r_i^t > 0\}} \right)^2}{\sum_t \left(r_i^t \right)^2}$$

(11) 高频下行波动占比:

$$\text{高频下行波动占比} = \frac{\sum_t \left(r_i^t I_{\{r_i^t < 0\}} \right)^2}{\sum_t \left(r_i^t \right)^2}$$

3、

(12) 日内成交量占比：(半小时划分区间，一天八个区间)：

$$\text{VolumeRatio}_t = \frac{\text{Volume}_t}{\text{Volume}_{\text{total}}}$$

(13) 高频量价相关性：(其中P_t为价格序列，V_t为成交量序列，分钟级别)

$$\rho = \text{corr}(P_t, V_t)$$

(14) 高频仓价相关性：(期货，价格和持仓量之间的相关性)

$$\rho = \text{corr}(P_t, V_t)$$

(15) 资金流因子：

<1>

$$\text{flowInRatio} = \sum_i \sum_j \text{Volume}_{ij} \cdot \text{Close}_{ij} \cdot \frac{\text{Close}_{ij} - \text{Close}_{ij-1}}{|\text{Close}_{ij} - \text{Close}_{ij-1}|} \Big) / \sum_i \text{Amount}_{i,\text{total}}$$

*Amount*为成交额

<2>

$$\text{flowInRatio} = \sum_i \sum_j |OI_{ij} - OI_{ij-1}| \cdot \text{Close}_{ij} \cdot \frac{\text{Close}_{ij} - \text{Close}_{ij-1}}{|\text{Close}_{ij} - \text{Close}_{ij-1}|} / (OI_{t-R} \cdot \text{Settle}_{t-R})$$

其中，OI_{ij}表示第i日第j分钟的持仓量，$Settle_{t-R}$为结算价

(16) 趋势强度:

$$\text{trendStrength} = \frac{P_n - P_1}{\sum_{i=2}^{n} \text{abs}(P_i - P_{i-1})_t}$$

其中P_t为价格序列

(17) (改进)反转因子*：（反转指将隔夜和开盘后半小时的涨幅剔除，分钟级不需要剔除，其中w_i为时刻i的权重)

$$\text{Rev}_{\text{vol}} = \sum_{i=1}^{\text{period}} w_i \log \frac{\text{Close}_{t-i+1}}{\text{Close}_{t-i}}, w_i \propto \text{volume}_i$$

4、

(18) 平均单笔成交金额：(AMT_{ij}是成交金额序列，$TrdNum_{ij}$是成交笔数序列)

$$\text{AmtPerTrd}_i = \sum_{j=1}^{N} \text{Amt}_{ij} / \sum_{j=1}^{N} \text{Tr}\, dNum_{ij}$$

(19) 平均单笔流入金额：

$$\text{AmtPerTrd\_inFlow}_i = \frac{\sum_{j=1}^{N} Amt_{ij} \cdot I_{r_{ij}} > 0}{\sum_{j=1}^{N} \text{TrdNum}_{\text{ij}} \cdot \text{I}_{\text{r}_{\text{ij}}} > 0}$$

(20) 平均单笔流出金额：

$$\text{AmtPerTrd\_outFlow}_i = \frac{\sum_{j=1}^{N} Amt_{ij} \cdot I_{r_{ij}} < 0}{\sum_{j=1}^{N} \text{TrdNum}_{\text{ij}} \cdot \text{I}_{\text{r}_{\text{ij}}} < 0}$$

(21) 平均单笔流入金额占比：

$$\text{ApT\_inFlow\_ratio}_i = \frac{\text{AmtPerTrd\_inFlow}_i}{\text{AmtPerTrd}_i}$$

(22) 平均单笔流出金额占比：

$$\text{ApT\_outFlow\_ratio}_i = \frac{\text{AmtPerTrd\_outFlow}_i}{\text{AmtPerTrd}_i}$$

(23) 平均单笔流入流出金额之比：

$$\text{ApT\_netInFlow\_ratio}_i = \frac{\text{ApT\_inFlow\_ratio}_i}{\text{Apt\_outFlow\_ratio}_i}$$

(24) 大单资金净流入金额：(将分钟K线按AmtPerTrd_{ij}从高到低排序，选择前N(N=10%, 20%, 30%)的K线)

$$\text{Amt\_netInFlow\_bigOrder}_i = \sum_{j=1}^{N} Amt_{ij} \cdot I_{\{r_{ij}>0, j \epsilon IdxSet\}} - \sum_{j=1}^{N} Amt_{ij} \cdot I_{\{r_{ij}<0, j \epsilon IdxSet\}}$$

(25) 大单资金净流入率：

$$\text{Amt\_netInFlow\_bigOrder\_ratio}_i = \text{Amt\_netInFlow\_bigOrder} / \sum_{j=1}^{N} \text{Amt}_{ij}$$

(26) 大单驱动涨幅：

$$\text{Mom\_bigOrder}_i = \text{prod}\left(1 + r_{ij} \cdot I_{\{j \in IdxSet\}}\right)$$

Tick级盘口委托

也可以理解为分钟级。3s一次。

注：来自的研报(海通证券)如下:

1、 《选股因子系列研究（四十七）——捕捉投资者的交易意愿》

2、 《选股因子系列研究（四十九）——当下跌遇到托底

1、

若假定委买量的增加代表了 投资者买入意愿的增强，而委卖量的增加代表了投资者卖出意愿的增强，那么可以认为 净委买变化量体现了投资者买入意愿的变化。考虑到委托挂单的变化与股票本身股本有 一定的关联，因此本文将净委买变化量除以股票流通股本，得到净委买变化率。

(1) 净委买变化率：（其中，$净委买变化率_{k,t}^{T}$为T日t到t+1时刻间，使用前K档数据计算得到的）

$$净委买变化率_{k,t}^{T} = \frac{净委买变化量_{k,t}^{T}}{流通股本_{T}}$$

(2) 净委买变化量：

$$净委买变化量_{k,t}^{T} = \sum_{j=1}^{k} 委买变化量_{j,t}^{T} - \sum_{j=1}^{k} 委卖变化量_{j,t}^{T}$$

(3) 平均净委买变化率：

$$平均净委买变化率_{k}^{T} = \mathrm{mean}\left(净委买变化率_{k,t}^{T} \right)$$

(4) 净委买变化率波动率：

$$净委买变化率波动率_{k}^{T} = std\left(净委头变化率_{k,t}^{T} \right)$$

(5) 平均净委买变化率偏度：

$$平均净委买变化率偏度_{k}^{T} = \mathrm{skewness}\left(净委买变化率_{k,t}^{T} \right)$$

2、

(6) 委托成交相关性:($r_{T,t}^{i}$为股票i在T日的高频收益序列，$netBid_{T,t}^{i}$为股票i在T日使用前1档委托挂单数据计算的净委买变化率序列)

$$委托成交相关性_{T}^{i} = \mathrm{corr}\left(r_{T,t}^{i}, netBid_{T,t}^{i} \right)$$

逐笔数据

注：来自的研报(海通证券)如下:

1、《选股因子系列研究（五十六）——买卖单数据中的 Alpha》
2、《选股因子系列研究（五十七）——基于主动买入行为的选股因子》
3、《选股因子系列研究（五十八）——知情交易与主买主卖》

1、

本文使用了“N 倍标准差”的方式，在每个交易日对于每个股票单独设定大单筛选阈值。

(1) 大卖成交金额占比：

$$大卖成交金额占比_{i,t} = \frac{大卖成交金额_{i,t}}{总成交金额_{i,t}}$$

(2) 大买成交金额占比：

$$大买成交金额占比_{i,t} = \frac{大买成交金额_{i,t}}{总成交金额_{i,t}}$$

(3) 大买大卖成交金额占比差值：

$$大买大卖成交金额占比差值_{i,t} = \frac{大买成交金额_{i,t}}{总成交金额_{i,t}} - \frac{大卖成交金额_{i,t}}{总成交金额_{i,t}}$$

(4) 大单成交金额占比：

$$大单成交金额占比_{i,t} = \frac{大买成交金额_{i,t}}{总成交金额_{i,t}} + \frac{大卖成交金额_{i,t}}{总成交金额_{i,t}}$$

(5) 卖单集中度：（卖单成交金额$_{i,t,k}$为股票i在交易日t的第k个买单的成交金额）

$$卖单集中度_{i,t} = \frac{\sum_{k=1}^{N_{i,t}} 卖单成交金额_{i,t,k}^2}{总成交金额_{i,t}^2}$$

(6) 买单集中度：

$$买单集中度_{i,t} = \frac{\sum_{k=1}^{N_{i,t}} 卖单成交金额_{i,t,k}^2}{总成交金额_{i,t}^2}$$

(7) 买卖单集中度差值：

$$买卖单集中度差值_{i,t} = \frac{\sum_{k=1}^{N_{i,t}} 买单成交金额_{i,t,k}^2}{总成交金额_{i,t}^2} - \frac{\sum_{k=1}^{N_{i,t}} 卖单成交金额_{i,t,k}^2}{总成交金额_{i,t}^2}$$

(8) 买卖单集中度之和：

$$买卖单集中度差值_{i,t} = \frac{\sum_{k=1}^{N_{i,t}} 买单成交金额_{i,t,k}^2}{总成交金额_{i,t}^2} + \frac{\sum_{k=1}^{N_{i,t}} 卖单成交金额_{i,t,k}^2}{总成交金额_{i,t}^2}$$

2、

本文着眼于逐 笔数据中的 BS 标志。该字段对于每笔成交的主动成交方向进行了界定，B 为主动买入， 也即，卖出方先挂单，买入方主动触碰卖单并成交。S 为主动卖出，也即，买入方先挂 单，卖出方主动触碰买单并成交。

(9) 主买占比：

$$主买占比\ (占全天成交) = \frac{主动买入金额}{当日总成交金额}$$

(10) 主买强度：

$$主买占比\ (占同时段成交) = \frac{主动买入金额}{同时段总成交金额}$$

(11) 日内主买强度：

$$日内主买强度 = \frac{\text{mean}(\text{ 主动买入金额})}{\text{std}(\text{ 主动买入金额 })}$$

(12) 日内净主买强度：

$$日内净主买强度 = \frac{\text{mean}(\text{ 主动买入金额-主动卖出金额})}{\text{std}(\text{ 主动买入金额-主动卖出金额})}$$

3、

基于股票过去一个月的日内分钟收益序列，可构建以下回归模型：

$$R_{i,T,j} = \gamma_0 + \sum_{k=1}^{4} \gamma_{1,k} D_{T,k,j}^{\text{weekday}} + \sum_{k=1}^{3} \gamma_{2,k} D_{T,k,j}^{\text{Period}} + \gamma_{3,1} R_{i,T,j-1} + \varepsilon_{i,j}$$

其中 $R_{i,T,j}$ 为股票i在T日第j分钟的收益， $D_{T,k,j}^{\text{weekday}}$ 为虚拟变量，表示周一到周四， $D_{T,k,j}^{\text{Period}}$ 为时间虚拟变量，表示开盘后30min，盘中时段以及收盘前30min。得到残差序列，作为预期外收益。在预期外收益为正时，投资者的主动 卖出行为可被认为是知情主卖，而预期外收益为负时，投资者的主动买入行为可被认为 是知情主买。

(13) 知情主卖占比（占全天成交额）：
(14) 知情主卖占比（占同时段成交额）：
(15) 知情主卖占比（占同时段主卖）：
(16) 知情主买占比（占全天成交额）：
(17) 知情主买占比（占同时段成交额）：
(18) 知情主买占比（占同时段主买）：

(19) 知情主买占比（占全天成交额）：
(20) 知情主买占比（占同时段成交额）：
(21) 知情主买占比（占同时段净主买）：