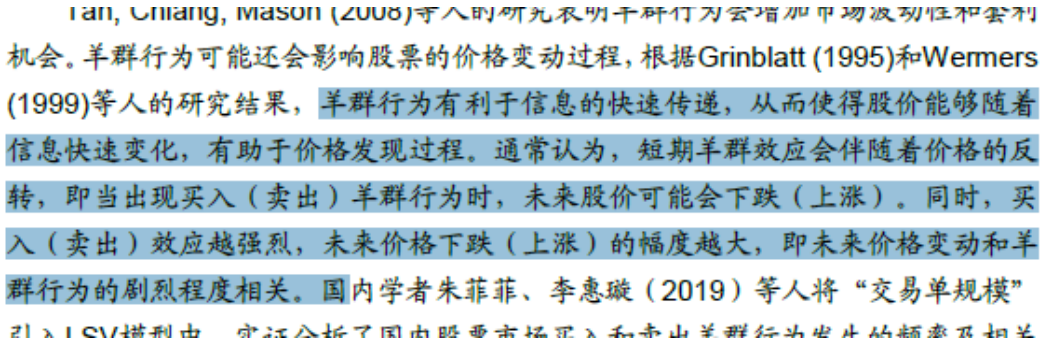
1.Factors Based on the herd effect

Factor Logic

When buying (selling) herd behavior occurs, stock prices may fall (rise) in the future. At the same time, the stronger the buying (selling) effect is, the greater the range of future price decline (rise) is, that is, the future price change is related to the intensity of herd behavior.


Tan, Chiang, Mason (2008)等人的研究表明羊群行为会增加市场波动性和套利机会。羊群行为可能还会影响股票的价格变动过程，根据Grinblatt (1995)和Wermers (1999)等人的研究结果，羊群行为有利于信息的快速传递，从而使得股价能够随着信息快速变化，有助于价格发现过程。通常认为，短期羊群效应会伴随着价格的反转，即当出现买入（卖出）羊群行为时，未来股价可能会下跌（上涨）。同时，买入（卖出）效应越强烈，未来价格下跌（上涨）的幅度越大，即未来价格变动和羊群行为的剧烈程度相关。国内学者朱菲菲、李惠璇（2019）等人将“交易单规模”引入LSV模型中，实证分析了国内股票市场买入和卖出羊群行为发生的频率及相关

Factor Construction

According to the minute-level stock transaction price data, we judge whether the order is driven by the buyer or the order is driven by the seller. Then, we construct the daily herd behavior measure of individual stocks according to the above expression and obtain the factor.

根据Lakonishok和Christoffersen等人的研究，羊群行为度量为

$$H(i,T) = \left|\frac{B(i,T)}{B(i,T)+S(i,T)} - P_T\right| - AF(t,T)$$

$$AF(i,T) = \sum_{k=0}^{N_{i,T}} \binom{N_{i,T}}{k} p_T^k (1-p_T)^{N_{i,T}-} \left|\frac{k}{N_{i,T}} - p_T\right|$$

$$HB(i,T) = -H(i,T) \quad if \ \frac{B(i,T)}{B(i,T)+S(i,T)} > P_T$$

$$HS(i,T) = H(i,T) \quad if \ \frac{B(i,T)}{B(i,T)+S(i,T)} < P_T$$

其中B(i,T)代表在特定时间时间段（小时，天，周）内对于某只股票的买方驱动单的数量，S(i,T)代表特定时间段卖方驱动单的数量。P_T为在该特定时间段内，所有股票的买单占其交易单比例的在横截面上平均值，即绝对值中的项代表了特定股票买单比例的相对水平。其中AF(i,T)是调整项，其表达式为

$$AF(i,T) = \sum_{k=0}^{N_{i,T}} \binom{N_{i,T}}{k} p_T^k (1-p_T)^{N_{i,T}-} \left| \frac{k}{N_{i,T}} - p_T \right|$$

（1）、如果当前时刻股票成交价大于之前一笔（5s或3s前）的买一价，则被定义为买方驱动单;

（2）、如果当前时刻股票成交价小于之前一笔的卖一价则被定义为卖方驱动单;

（3）、如果当前时刻股票成交价等于之前一笔的成交价，则被分为第三类驱动单。当统计在一段时间内的买方和卖方驱动单时，买方B(i,T)=买方驱动单+第三类驱动单/2，卖方S(i,T)=卖方驱动单+第三类驱动单/2。

Factor Processing

Based on the original factor, we will further improve the factor, including using MAD method to eliminate extreme value, neutralization and standardization.

去极值

采用MAD法对异常的数据进行剔除。MAD法是针对均值标准差方法的改进，样本均值用样本中位数代替，样本标准差用样本MAD（Median Absolute Deviation）代替。

$$md = median\{x_i, i = 1,2,3 \dots, n\}$$

$$MAD = median(|x_i - md|), i = 1,2,3, \dots, n$$

$$MAD_e = 1.483 \times MAD$$

采用与3σ法等价的方法，将数据中偏离中位数超过3倍MAD_e的值，替换为对应的值，这样去除了极值的同时也将极值的股票之间保序。

中性化

将原始因子值和股票的市值以及行业进行回归，取残差项作为新因子

$$H(i,T) = \beta_0 + \beta_1 size(i,T) + \beta_2 industry(i,T) + \varepsilon(i,T)$$

其中市值自变量使用对数市值，行业自变量为0或者1的哑变量

标准化

使用因子的Z-Score值作为新的因子值

$$H_{new}(i,T) = \frac{H(i,T) - mean(H(i,T))}{std(H(i,T))}$$

将原始因子值经过上述处理之后的结果如下所示。

Factor Expectation

We expect the treated factors performed better, and factor IC value was around 0.05.

未经处理过的因子指标在全市场中选股，从2007年至今，IC均值为0.047，正IC占比为69.90%，多头组合在回测期内年化收益率为34.02%，信息比率为1.05；多头组合相对中证500指数年化超额收益率为24.61%，信息比率为2.59；

经过处理后的因子指标全市场中选股，从2007年至今，IC均值为0.051，正IC占比为73.77%，多头组合在回测期内年化收益率为34.62%，信息比率为1.09；多头组合相对中证500指数年化超额收益率为24.98%，信息比率为2.84。

2.Return Distribution Factor

Factor logic

There is a correlation between high moments and future returns of stocks.

Factor Construction

Three indexes of realized variance, skewness and kurtosis are constructed by using intra-day minute level data of stock. The factor value is the index mean of the past 20 days.

高频已实现方差：$RVar_i = \sum_{j=1}^{N} r_{ij}^2$

高频已实现偏度：$RSkew_i = \frac{\sqrt{N}\sum_{j=1}^{N} r_{ij}^3}{RVar_i^{3/2}}$

高频已实现峰度：$RKurtosis_i = \frac{N\sum_{j=1}^{N} r_{ij}^4}{RVar_i^2}$

其中r_{ij}为股票 i 的日内 1 分钟对数收益序列$\{j = 1, \cdots, N\}$，因子值为过去 20 日的指标均值。

Factor Expectation

The high frequency skewness factor has significant stock selection ability, while variance and kurtosis have no stock selection ability.

3.Proportion of high frequency upstream and downstream fluctuation factor (高频上下行波动占比因子)

Factor Logic

Feunou B et al.(2015)3 separated volatility into upward volatility and downward volatility, and found that investors paid more attention to downward volatility and required higher risk compensation.

Factor Construction

Where rij is the daily 1-minute logarithmic return series of stock i, and the factor value of stock is the index mean value of the past 20 days. $I_{r_{ij}}$ denotes the indicator of wheather the return is positive or negative.

$$高频下行波动占比 = \frac{\sum_{j=1}^{N} r_{ij}^2 \cdot I_{r_{ij}<0}}{\sum_{j=1}^{N} r_{ij}^2}$$

Factor Expectation

IC is around 0.06 and long short portfolio monthly average yield difference is around 1.5%.

4.Intraday volume distribution Factor

Factor Logic

Generally speaking, intraday trading volume of stocks presents a "U" or "W" pattern, The volume distribution at each point in time can reflect the behavior characteristics of investors, containing additional information.

Factor Construction

We divided the daily minute-level data into eight intervals at half an hour intervals, calculated the volume proportion of each interval, and used the average value of the stock index in the past 20 days as the factor value.

我们以半小时为间隔划分成八个区间，计算每个区间成交量占比$\boldsymbol{VolumeRatio}$，并使用股票过去 20 日指标均值作为因子值。

$$\boldsymbol{VolumeRatio_t} = \frac{\boldsymbol{Volume_t}}{\boldsymbol{Volume_{total}}}$$

Factor Expectation

At the beginning of trading, the factor is positively correlated with earnings, while in the middle of trading, the factor is negatively correlated with earnings.

不同时间区间的因子 IC 统计结果如表 6 所示，上午 10:00 之前，下午 14:30 之后的成交量占比因子和股票下月收益负相关；10:00-11:00 的成交量占比因子和股票下月收益显著正相关。

5. High frequency volume-valence correlation

Factor Logic

The correlation coefficient between daily frequency stock price and turnover rate has significant stock selection ability.

Factro Constrction

Take stock for 1 minute within each trading day Price $\boldsymbol{Pt}$ and volume $\boldsymbol{Vt}$ of frequency sequence, calculate Pearson correlation coefficient. The average value of the index in the past 20 days is used as the factor value.

下面我们使用日内数据构建高频量价相关性因子，即每一交易日取股票日内 1 分钟频率的价格$\boldsymbol{P_t}$和成交量$\boldsymbol{V_t}$序列，计算 Pearson 相关系数，

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

并使用过去 20 日指标均值作为因子值。

Factor Expectation

IC is around -0.061 and long short portfolio monthly average yield difference is around 2%.

6. Trend Strength Factor(趋势强度因子)

Factor Logic

It is one of the momentum factors.

Factor Construction

The trend strength can be calculated by using the price series of 1-minute frequency of t day, which can be understood as the ratio of price displacement and distance of the day, which can describe the strength of the trend of the day. We also define the mean of the past R-day index as the trend strength factor.

记 t 日 1 分钟频率的价格序列为$P_t(t = 1, 2, \cdots, n)$，则趋势强度指标定义如下：

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

Factor Expectation

IC is around -0.034 and long short portfolio monthly average yield difference is around 1.18%. But after eliminating the style factor, the performance is poor.

7. Improved reversal factor(改进翻转因子)

Factor Logic

Most companies choose to release important information, such as financial data, after the close, causing stocks to open sharply the next day and fluctuate wildly in the ensuing period. Such opening jumps are the result of the market pricing important data or information properly, rather than mispricing caused by investor behavior.

大多数公司选择在收盘后发布财务数据等重要信息，造成股票次日往往跳空开盘，并在随后一段时间大幅波动。近年来，A 股机构投资者占比提升，市场交易行为在慢慢向美国等发达市场靠拢。这种在美股财报季常见的开盘跳空现象属于市场对重要数据或信息的合理定价，而非投资者行为造成的错误定价。基于以上推断，可以尝试将隔夜和开盘后半小时的涨幅剔除，构建改进反转因子。因子计算公式如下：

Factor Construction

We use daily minute level closing price data to obtain the improved rollover factor.

$$改进反转 = \prod_{n=t}^{n=t-T+1} \frac{Close_{i,j,n}}{Close_{i,10:00,n}} - 1$$

Factor Expectation

IC is around 0.0433 and long short portfolio monthly average yield difference is around 2%.

8. Ratio of average single outflow amount(平均单笔流出金额占比)

Factor Logic

When the stock falls, if the single transaction amount is large, it shows that there is a big buy, which is a bottom-fishing behavior.

Factor Construction

We use the proportion of T's daily average single outflow amount as a factor。

$$平均单笔流出金额占比 = \frac{1}{T}\sum_{n=t}^{n=t-T+1} \frac{\sum_{j=1}^{N} Amt_{i,j,n} \cdot I_{r_{i,j,n}<0} / \sum_{j=1}^{N} TrdNum_{i,j,n} \cdot I_{r_{i,j,n}<0}}{\sum_{j=1}^{N} Amt_{i,j,n} / \sum_{j=1}^{N} TrdNum_{i,j,n}}$$

Factor Expectation

IC is around 0.03 and long short portfolio monthly average yield difference is around 1.5%.

9. Net master buy ratio after opening Factor(开盘后净主买占比因子)

Factor Logic

When the stock falls, if the single transaction amount is large, it shows that there is a big buy, which is a bottom-fishing behavior.

Factor Construction

We use the proportion of T's daily average single outflow amount as a factor。

$$平均单笔流出金额占比 = \frac{1}{T}\sum_{n=t}^{n=t-T+1} \frac{\sum_{j=1}^{N} Amt_{i,j,n} \cdot I_{r_{i,j,n}<0} / \sum_{j=1}^{N} TrdNum_{i,j,n} \cdot I_{r_{i,j,n}<0}}{\sum_{j=1}^{N} Amt_{i,j,n} / \sum_{j=1}^{N} TrdNum_{i,j,n}}$$

Factor Expectation

IC is around 0.03 and long short portfolio monthly average yield difference is around 1.5%.