

Alpha

aglv

July

1 General

1.1 价格序列处理

尝试使用对数收益率来作为收益率值 $\rightarrow CVaR_{neg}$ 在 230101-240101 上表现不佳

1.2 标准化

z-score 方法：标准化后因子回测效果明显变优，收益序列平稳

对同一股票的前一段时间数据作为样本 or 某个时间截面针对所有股票计算均值标准差

按照 axis=1 求标准化，求得效果与原数据相同

取 rank ?

1.3 离群值处理

1.4 缺失值处理

2 波动与收益率

《收益率的非对称分布与尾部蕴含的 Alpha》东方证券

《收益和波动共舞：非对称性理论蕴含的 alpha》华安证券

2.1 非对称性的衡量因子

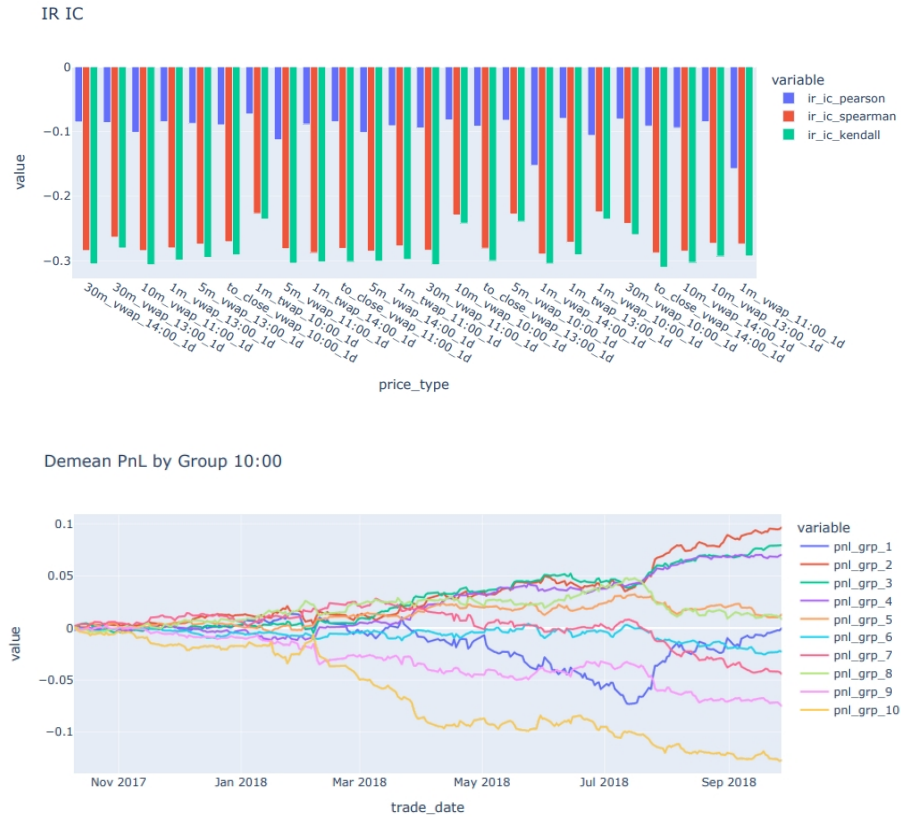
偏度 (skewness): z-score 后的三阶矩

$$Skewness = E\left[\left(\frac{X - \mu}{\sigma}\right)^3\right]$$

remark:

偏度大小很受离群值的影响，三次方加大了偏差，实际上关系到了尾部分布

偏度因子计算结果：



改进？

特征偏度： ϵ_i 表示 Fama-French 三因子中性化之后的收益率

日内特质波动率：

$$iDVol_t = \sqrt{\sum_{i=1}^N \epsilon_{t,i}^2}$$

特质偏度：

$$iDSkew_t = \frac{\sqrt{N} \sum_{i=1}^N \epsilon_{t,i}^3}{iDVol^{3/2}}$$

E_ϕ : 左右尾概率作差表示分布的非对称性

$$E_\phi = P(x \geq k) - P(x \leq -k)$$

S_ϕ : 基于熵构造原理

其中

$$h = 1.06\sigma n^{-\frac{1}{5}}$$

简化求 f:

$$\frac{\partial k(\frac{r_i - x}{h})}{\partial h} = -(\frac{r_i - x}{h})k(\frac{r_i - x}{h})$$

$$\frac{\partial^2 k(\frac{r_i - x}{h})}{\partial h^2} = (\frac{r_i - x}{h})^2 k(\frac{r_i - x}{h}) + \frac{r_i - x}{h} k(\frac{r_i - x}{h})$$

$$Skew_{i,t} = \frac{1}{n} \sum_{d \in S(t)} \left(\frac{R_{i,d} - \bar{R}_i}{\hat{\sigma}_i} \right)^3$$

where $S(t)$ and n are the set of trading days and the number of trading days, \bar{R}_i and $\hat{\sigma}_i$ are the sample mean and the sample standard deviation of stock i .

Idiosyncratic skewness is calculated following Boyer et al. (2010) shown in Eq. (8) to daily stock return data observed during the pre-crisis period. Idiosyncratic skewness for stock i at the end of month t is calculated as follows:

$$IdioSkew_{i,t} = \frac{1}{n-2} \times \frac{\sum_{d \in S(t)} \varepsilon_{i,d}^3}{IdioVol_{i,t}^3}$$

where $IdioVol_{i,t}$ is calculated following Eq. (7).¹¹

Finally, to calculate coskewness, we follow the approach discussed in the literature:

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_{i,m}(R_{m,t} - R_{f,t}) + \varepsilon_{i,t}$$

where $R_{i,t}$, $R_{m,t}$, and $R_{f,t}$ are the monthly return of stock i , the monthly return of the market, and the monthly risk-free rate at the end of month t , respectively. Then, for stock i in month t , denoted $CoSkew_{i,t}$, is estimated as follows:

$$CoSkew_{i,t} = \frac{E[\varepsilon_{i,t}(R_{m,t} - R_{f,t})^2]}{\sqrt{E[\varepsilon_{i,t}^2]E[(R_{m,t} - R_{f,t})^2]}}$$

where $\varepsilon_{i,t}$ is the regression residual calculated from Eq. (11) above.

以用于衡量一个分布是否对称。但由于单纯的距离无法体现分布非对称性的方向，即无法区分分布是左偏还是右偏，所以 Jiang 等人（2020）在距离前乘以 E_φ 因子的符号 $Sign(E_\varphi)$ ，使得 S_φ 加入了非对称性方向的考量。

$$S_\varphi = Sign(E_\varphi) * \frac{1}{2} \left\{ \int_{-\infty}^{-k} (f_1^{\frac{1}{2}} - f_2^{\frac{1}{2}})^2 dx + \int_k^{+\infty} (f_1^{\frac{1}{2}} - f_2^{\frac{1}{2}})^2 dx \right\}$$

S_φ 定义中的 x 和 k 的含义与 E_φ 相同。由于 $f_1(x)$ 、 $f_2(x)$ 的真实分布未知，所以对二者的估计采用非参数估计中的核密度估计法，核密度函数选取高斯核函数。

$$\widehat{f}(x) = \frac{1}{nh} \sum_{i=1}^n \kappa\left(\frac{r_i - x}{h}\right)$$

$$\kappa(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

化简计算：

$$X_2 = -X + 2\mu_X$$

$$\begin{aligned} P(X_2 < x) &= P(-X + 2\mu_X < x) \\ &= P(X > 2\mu_X - x) \\ &= 1 - F(2\mu_X - x) \end{aligned}$$

$$f_2(x) = f(2\mu_X - x)$$

记 S_ϕ 的被积函数为 g

$$\begin{aligned} g &= (f_1^{\frac{1}{2}} - f_2^{\frac{1}{2}})^2 \\ &= f(x) + f(2\mu_X - x) - 2[f(x)f(2\mu_X - x)]^{\frac{1}{2}} \end{aligned}$$

则

$$S_\phi = \frac{1}{2} * \text{sign} * \left(\int_{-\infty}^{-k} g + \int_k^{\infty} g \right) \\ = \text{sign} * \left(2 - \left(\int_{-\infty}^{-k} [f(x)f(2\mu_X - x)]^{\frac{1}{2}} dx + \int_k^{\infty} [f(x)f(2\mu_X - x)]^{\frac{1}{2}} dx \right) \right)$$

计算复杂度优化:

考虑到计算积分的时间成本, 将

$$\text{sign}(E_\phi) \rightarrow \text{sign}_{skew}$$

$$\text{被积差值函数的积分} \rightarrow E[i|i > series_k \text{ for } i \text{ in } diffseries]$$

$Asym_p$: 在分布接近对称分布时表现更好

$$Asym_p = \begin{cases} -corr(f, F) & \text{if } 0 < var(f) \\ 0 & \text{if } var(f) = 0 \end{cases}$$

remark:

1、怎么计算两个函数之间的相关关系?

? \rightarrow 使用 f 和 F 作用于实际收益率的分点上的两个序列的相关系数

2、怎么进行回测:

? 择时 or 选股, 原论文中使用 t 个月的窗口期

3、计算复杂度过大, 尝试借用中心极限定理对问题进化简, 假设每日的收益率独立同分布

2.2 尾部蕴含 alpha

$$CVaR_- = E[X|X < VaR], P(X < VaR) = 5\%$$

$$CVaR_+ = E[X|X > VaR], P(X > VaR) = 5\%$$

(p.s. 是否能考虑将算术求和的线性期望改变成其他形式的期望来增加对极端情况的厌恶

Tail Beta:

$$R_j^e = \beta_j^T R_m^e + \varepsilon_j, R_m^e < -VaR_m(\bar{p})$$

2.3 idea

1、简化计算 \rightarrow 中心极限定理、分布函数按照统计分数值计算

2、因子值做标准化, 策略更加稳定 \rightarrow 数据做标准化的方法? z-score/Max-Min/ L_1, L_2

3、测试因子 \rightarrow 修改函数参数、修改时序窗口期长度、因子值做比率

$$\widehat{\beta}_j^T := \tau_j(\widehat{k/n})^{1/\widehat{\alpha}_m} \frac{VaR_j(\widehat{k/n})}{VaR_m(\widehat{k/n})}$$

令 $X_t^{(m)} = -R_{m,t}^e, t = 1, 2 \dots n$, n 为计算期间内交易日天数。然后将市场超额收益的相反数 $X_t^{(m)}$ 从小到大排序, 得到 $X_{n,1}^{(m)} \leq X_{n,2}^{(m)} \leq \dots \leq X_{n,n}^{(m)}$ 。 $X_t^{(j)}$ 、 $X_{n,i}^{(j)}$ 定义方式同上。取 k 为 n 个交易日中 R_m^e 损失超出其 VaR 值的天数 ($k \approx 0.05n$), 则

$$\frac{1}{\widehat{\alpha}_m} = \frac{1}{k} \sum_{i=1}^k \log X_{n,n-i+1}^{(m)} - \log X_{n,n-k}^{(m)}$$

$$\tau_j(\widehat{k/n}) = \frac{1}{k} \sum_{i=1}^k 1_{\{X_t^{(j)} > X_{n,n-k}^{(j)} \text{ and } X_t^{(m)} > X_{n,n-k}^{(m)}\}}$$

$VaR_j(\widehat{k/n})$ 、 $VaR_m(\widehat{k/n})$ 分别表示股票 j 和市场第 $k+1$ 大损失。尾部 β 因子表示当市场出现极端下跌行情时, 个股收益对市场收益的敏感性。若尾部 β 越大, 表示个股收益对市场的极端负收益越敏感, 使得收益率低于尾部 β 偏小的股票。

图表 5 因子构造关键点总览表: 收益-波动先后关系和计算频率问题

波动滞后因子				
变量	时间先后顺序	日间	日间2	日内高频
收益	收益在前	日涨跌幅	日涨跌幅	5min涨跌幅
波动	波动在后	日波动率	日内波动率	5min波动率
相关系数	/	滚动过去20日	滚动过去20日	日内计算
波动超前因子				
变量	时间先后顺序	日间	日间2	日内高频
收益	收益在后	日涨跌幅	日涨跌幅	5min涨跌幅
波动	波动在前	10日波动率	10日日内波动率	60min波动率
相关系数	/	滚动过去20日	滚动过去20日	日内计算
同期效应因子				
变量	时间先后顺序	日间	日间2	日内高频
收益	同期	日涨跌幅	日涨跌幅	5min涨跌幅
波动	同期	日波动率	日内波动率	5min波动率
相关系数	/	滚动过去20日	滚动过去20日	日内计算

2.4 波动和收益率的同期跨期统计分析

remark:

- 1、日内分钟数据的使用对因子的预测效果提升显著, 高频波动因子效果更好
- 2、从 IC 均值看, 因子收益率一般但是表现稳定

2.5 Idea

- 1、利用波动率和收益率的相关关系: 同期呈正相关关系、未来呈现负相关关系
- 2、非对称性: 同等幅度的上涨或者下跌带来的波动率水平是不一样的
- 3、

3 情绪面指标测试

《A 股择时之情绪面指标测试》华泰证券

3.1 引言

两个维度：‘量价数据’、‘后市预期’

量化数据 → 换手率、成交量表征交易热度，表征市场强弱状态，综合反映投资者参与当下市场投资的意愿

后市预期 → 融资融券、股指期货、期权，表征投资者对未来市场走势的预期。

上述指标中，换手率、成交量等表征全市场交易热度类指标以及基于细分资产强弱势状态构建的市场涨跌类指标反映了投资者参与当下市场交易的意愿，大多与市场走势同步，可作为市场情绪的“温度计”，而不具备显著的择时能力。融资融券行为、期货、期权市场量价和仓位数据反映投资者对后市的预期，指标走势中蕴含投资者观点，具备一定的增量信息，在牛熊切换的拐点领先于市场，基于这类指标构建的策略往往好于基于市场指数自身构建的择时策略

4 A 股市场如何构建动量因子

《A 股市场如何构建动量因子》开源证券

问题 1：A 股市场呈现强反转效应，涨跌幅因子时间维度的切割效果不佳

想法 1：分析反转效应原因（从交易行为入手）→ 反转与投资者的过度思考和资金的过度涌入流出有关，→ 不同交易活跃程度下，动量和反转效应是否存在差异性？

交易行为维度：换手率，单笔成交金额，日度振幅

5 上下影线与威廉指标

基于蜡烛图的择时策略

1、分别计算上下影线，并且进行五日标准化

$$\text{标准化上影线} = \frac{\text{当日上影线}}{\text{上影线.rolling(5).mean()}}$$

用标准化的数据计算 mean() 和 std()

2、主要是源于上下影线和威廉指标体现的多空双方的力量趋势

6 高频因子

《兴业证券》

四类高频指标信息：分布信息，时序信息，关联信息和另类信息

非有效性市场下，过去一段时间的量价信息并不完全反映所有的过往信息