



CWL

量化投资与机器学习

1月

January



CML

量化投资与机器学习

Jan.

腊月初二

星期三

1月 01

价格波动率

期货-波动率因子

利用近月合约 F_{1t} 和次近月合约 F_{2t} 的价格来估计现货 spot 价格 P_t ：

$$P_t = F_{1t} \left(\frac{F_{1t}}{F_{2t}} \right)^{\frac{n_{0t}}{n_1}}$$

计算“有效的”每日对数价格变化（经过标准差的调整）：

$$r_\tau = \frac{\log P_\tau - \log P_{\tau-n}}{\hat{s}_n / \hat{s}_1}$$

利用当前周及其前四周的每日对数价格变化数据序列计算波动率：

$$\hat{\sigma}_t = \sqrt{\frac{1}{N-1} \sum_{\tau=1}^N (r_{t\tau} - \bar{r}_t)^2}$$

其中

- ① n_{0t} 为从时间 t 到近月合约到期的天数；
- ② n_1 为近月合约到期至次月合约到期的天数；
- ③ \hat{s}_n 和 \hat{s}_1 分别为间隔 n 天何间隔 1 天的价格变化的标准差；间隔天数 n 是指自上次交易日以来的天数，如周一的间隔天数通常为 3 天（中间有周末两天非交易日）；对于每个间隔类别 n 的标准差，通过计算整个样本期内对数价格变化的标准差得到；
- ④ N 是五周区间内（前周及其前四周）有效交易日的天数。

说明

波动率衡量了商品价格的变动程度，波动率的变化会影响现货价格、期货价格和库存水平。

【参考文献】

Pindyck, R., 2004, Volatility and Commodity Price Dynamics, Journal of Futures Markets, 29, 1029-1047.



Jan.

腊月初三

星期四

1月

02

基差因子

期货-期限结构因子

$$-\frac{1}{K} \sum_{t=T-K+1}^T \frac{P_{main,t} - P_{spot,t}}{P_{spot,t}} * \frac{365}{M_{main,t}}$$

其中

- ① $P_{spot,t}$ 、 $P_{main,t}$ 分别为现货调整后价格和主力合约的 t 日结算价；
- ② $M_{main,t}$ 分别为主力合约距到期日天数。

说明

基差因子为主力合约年化基差在过去 K 个交易日的均值；做多贴水高的品种，做空贴水低的品种，通过基差收敛获得展期收益。

【参考文献】

王冬黎, 常海晴, 2022, 商品多因子模型框架再探究, 东证期货.



Jan.

腊月初四
星期五

1月 03

价值盲区掘金因子

基本面因子改进

基于过去 2 年 (q-1 期至 q-8 期) 的单季度归母净利润计算业绩超预期维度代表因子：

$$SUE_t = \frac{X_q - E(X_q)}{\sigma_q} = \frac{X_q - [X_{q-4} + u_q]}{\sqrt{\frac{(X_{q-1}-X_{q-5}-u_q)^2+\dots+(X_{q-4}-X_{q-8}-u_q)^2}{4}}}$$

$$u_q = \frac{\sum_{n=1}^4 (X_{q-n} - X_{q-n-4})}{4}$$

计算反映市场修正错误定价充分程度的 3 个月低换手波动率：

$$\text{低换手波动率}_{3m} = std(\text{最近3个月日度换手率序列})$$

计算盈利预期重估因子：

$$EP_FY3_qoq = \frac{\text{一致预期净利润 FY3}}{\text{总市值}}$$

将上述 3 个因子进行行业市值中性化后，求 rank 值等权求和，即得到价值盲区掘金因子：

$$F_{value} = Rank_{SUE1-\text{归母净利润}_MRQ} + Rank_{\text{低换手波动率}_{3m}} + Rank_{EP_FY3_qoq}$$

说明

若基本面业绩未超预期或市场已经对超预期定价，盈余公告后的价格偏移效应可能就会失效；价值盲区掘金因子从业绩超预期、市场交易充分度、估值预期重估强度三个维度综合评估了市场对基本面存在低估个股的定价修正效率，用以寻找市场中那些业绩超预期但市场定价尚未修正的个股。

[参考文献]

郭策, 2023, 价值盲区掘金因子：纳入市场对预期的修正效率, 中银证券.



Jan.

腊月初五
星期六

1月 04

地理动量溢出效应

另类因子-动量溢出

$$RET_{it}^{GEO} = geographic_weight_{ij,t} \times ReT_{jt}$$

$$geographic_weight_{ij,t} = \frac{GMV_{j,t}}{\sum_{j=1 \text{ and } j \neq i}^N GMV_{j,t}}$$

其中

- ① ReT_{jt} 为股票 j 在 t 时刻的月度收益率；
- ② $GMV_{j,t}$ 为 t 时期同行业内公司 j 的市值。

说明

地理动量溢出效应因子刻画了与单只股票同位于相同地理其余股票涨跌对该股票涨跌的联动效应（领先滞后效应）。

[参考文献]

Usman Ali, David Hirshleifer, 2020, Shared analyst coverage: Unflying momentum spillover effects, Journal of Financial Economics, 6(136), 649-675.



Jan.

腊月初六

星期日

1月

05

散户羊群效应

高频因子-资金流类

$$RankCorr(R_t, S_{t+1})$$

其中

- ① R_t 为过去 20 个交易日的日度收益率序列，用于表示市场涨跌；
- ② S_{t+1} 为小单 (<4万元) 过去 20 个交易日同期的资金净流入序列（前置 1 期），
用于表示散户行为；
- ③ 使用 $close_t/open_t - 1$ 日内收益，使用小单非主动净流入（开盘至 10 点之间），
可以提升原始散户羊群效应的表现。

说明

衡量了散户追涨杀跌的程度，与股票未来收益负相关，即股票的散户羊群效应越高，未来预期收益越差。

[参考文献]

- ① 魏述榕, 盛少成, 2022, 新型因子：资金流动力学与散户羊群效应, 开源证券.
- ② 魏述榕, 盛少成, 2022, 大小单资金流 alpha 探究2.0：变鼠精师与高频测算, 开源证券.



Jan.

腊月初七
星期一

1月 06

客户加权毛利率

另类因子-因子改进

$$c_gross_margin_i = \sum_{j=1}^{N_i} w_{ij} * gross_margin_j$$

其中

- ① $gross_margin_j$ 为股票 i 的客户 j 的毛利率；
- ② w_{ij} 为股票 i 对客户 j 的加权权重，可以等权、销售占比加权等；
- ③ 除了加权客户的毛利率，还可以加权 ROE、ROA、市盈率倒数等基本面因子。

说明

客户毛利率因子对股票未来收益有负向预测能力，即客户毛利率越高，供应商公司未来股价收益越低；客户毛利率越高，说明客户相对供应商有更高的定价权，进而一定程度上压低了供应商的收入。

【参考文献】

Jussa et al., 2015, The Logistics of Supply Chain Alpha, Deutsche Bank Markets Research.



Jan.

腊月初八

星期二

1月

07

极端收益反转

高频因子-动量反转类

基于股票 i 的分钟行情数据，每日寻找出日内收益最极端的那根 bar，即 S 最大的那根 bar：

$$S = |r_{i,t} - \text{median}(\{r_{i,t}\})|$$

每月底，将过去 20 天的每天 S 最高那根 bar 收益率求均值得到最极端收益；同理，每月底，将过去 20 天的每天 S 最高那根 bar 前一分钟的收益率求均值得到前一分钟收益；将两者截面排序值等权求和，即得极端收益率反转因子：

$$\text{rank}\left(\frac{1}{20} \sum r_{i,t_{\max(S)}}\right) + \text{rank}\left(\frac{1}{20} \sum r_{i,t_{\max(S)}-1}\right)$$

说明

日内，最极端收益前呈反转特性，最极端收益后呈动量特性；切割出的极端收益反转因子具有强反转效应。

[参考文献]

魏庭榕, 盛少成, 苏良, 2022, 日内极端收益前后的反转特性与因子构建, 开源证券.



Jan.

腊月初九

星期三

1月

08

最大异常日收益率

行为金融因子-投资者注意力

$$ABNRETD = \max_t |r_{i,t} - \bar{r}_t|$$

其中

① $r_{i,t}$ 为股票 i 在 t 交易日的收益率；

② \bar{r}_t 为交易日 t 的全市场收益率；

③ m 为交易日窗口，一般取为月度。

说明

最大异常日收益率 ABNRETD 为每月日频收益率与市场收益率绝对差值的最大值，属于极端日收益类因子；极端收益（过高收益或过低收益）的股票相较其他股票更能吸引投资者的注意，因子值越高，投资者对该股票的关注度越高；而有限注意力导致投资者更倾向于交易引起他们关注的股票，投资者关注度越高的股票通常意味着越多的投资购买；但 A 股市场的做多与做空不对称，使得投资者非理性买入高于非理性卖出，进而导致关注度高的股票由于投资者净买入而存在溢价，未来也更可能出现反转。

[参考文献]

Jussa et al., 2015, The Logistics of Supply Chain Alpha, Deutsche Bank Markets Research.



Jan.

腊月初十
星期四

1月 09

截面动量

期货-动量趋势因子

$$XSMom_t = \frac{P_t - P_{t-J}}{P_{t-J}}$$

其中

- ① P_t 为期货品种在 t 日的收盘价；
- ② 期货截面动量策略会涉及三个独立变量： J (回看期)、 K (持有期)， L (滞后期)，一般取较长回看期（如 $J=20$ ），因子信号滞后执行（如 $L=3$ ）等操作能规避短期的反转效应，提升 策略表现。

说明

与股市一样，期货市场同样存在“前者恒强，弱者恒弱”的动量效应。截面动量效应是指在相同时间点上做多涨幅高的品种，做空涨幅少的品种，可以获得持续稳定的收益。

[参考文献]

张革, 2022, 动量及高阶矩因子在商品期货截面上的运用, 中信期货.



Jan.

腊月十一
星期五

1月

10

供应商加权毛利率

另类因子-因子改进

$$s_gross_margin_i = \sum_{j=1}^{N_i} w_{ij} * gross_margin_j$$

其中

- ① $gross_margin_j$ 为股票 i 的供应商 j 的毛利率；
- ② w_{ij} 为股票 i 对供应商 j 的加权权重，可以等权、采购占比加权等；
- ③ 除了加权供应商的毛利率，还可以加权 ROE、ROA、市盈率倒数等基本面因子。

说明

供应商毛利率因子与客户毛利率因子计算逻辑类似，刻画的是供应商基本面信息对客户股价变动的预测能力。

[参考文献]

Jussa et al., 2015, The Logistics of Supply Chain Alpha, Deutsche Bank Markets Research.



Jan.

腊月十二

星期六

1月 11

乖离率

高频技术因子

计算每天 1 分钟级平均收盘价：

$$Close_{ma,t} = \frac{\sum_{i=t-239}^t Close_{i,t}}{240}$$

计算每日收盘价相对当日 1 分钟级移动平均收盘价的偏差：

$$BIAS_{i,t} = \frac{Close_{i,t} - Close_{ma,i,t}}{Close_{ma,i,t}} * 100$$

说明

将过去 20 个交易日的日度高频因子进行时间衰减加权即可降频为月度因子：

$$\widehat{\text{Factor}}_j = \frac{1}{\sum_{j=1}^{20} \frac{j}{20}} \times \sum_{i=1}^{20} \widehat{\text{Factor}}_{j,k} \times \frac{j}{20}$$

说明

BIAS 乖离率指标通过测算股价在波动过程中与移动平均线出现偏离的程度，从而得出股价在剧烈波动时因偏离移动平均趋势而造成的可能回档或反弹。

[参考文献]

姚紫璇, 陈升锐, 2024, 技术指标因子高能化, 中信建投.





Amihud 非流动性比例

期货-流动性因子

$$ILLIQ = \frac{1}{R} \sum_{i=1}^R \frac{|r_i|}{Amount_i}$$

其中

- ① r_i 为期货合约在第 i 个交易日的收益率；
- ② $Amount_i$ 为期货合约在第 i 个交易日的成交额；
- ③ R 为回看期，回看期 $R \leq 40$ 个交易日时，因子表现较好。

说明

因子取值越高，期货品种的流动性越差；做多流动性差的品种，做空流动性高的品种，可获得非流动性补偿。

[参考文献]

- ① Amihud, Y., 2002, Illiquidity and stock returns: cross-section and time-series effects, *Journal of financial markets*, 5(1), 31-56.
- ② 冯佳睿,姚石, 2017, 商丘期货因子挖掘与组合构建再探究竟, 海通证券.



Jan.

腊月十四
星期一

1月 13

计入预期后的新闻情绪

另类因子-文本分析

$$senti_i^t = \frac{1}{N} \sum_{j=1}^N (pos_{i,j}^t * 1 + neg_{i,j}^t * (-1) + neu_{i,j}^t * 0)$$

$$senti_improved_i^t = \frac{senti_std_i^t}{(1 + re_i^t)}$$

其中

- ① 从 [t-1 日 14:55, t 日 14:55] 区间内所有新闻中筛选出与公司相关性大于等于 60% 且新闻等级为 2 级以上的共计 N 条新闻；
- ② $pos_{i,j}^t$ 、 $neg_{i,j}^t$ 、 $neu_{i,j}^t$ 分别股票 i 的第 j 条新闻中属于正面、负面、中性情绪的概率；
- ③ $senti_improved_i^t$ 为 $senti_i^t$ 标准化至 0-1 的情绪值，值越大越偏向正面；
- ④ re_i^t 为股票 i 在 t 时刻过去一个月的涨跌幅。

说明

计入前期涨跌情况有助于提升原始情绪因子的表现；对于负向新闻，若前期下跌，表明利空出尽股价将改善，若前期上涨，负向新闻会促进股价下跌；对于正向新闻，若前期下跌，正向新闻将有助于抬升股价，若前期上涨，股价可能已经利好出尽将会下跌。

[参考文献]

古翔, 郑文才, 陈宣筠, 2023, 如何优化新闻文本因子, 中金公司.



Jan.

腊月十五

星期二

1月

14

加权预期收益率

分析师、机构持仓

$$WTR = \sum \frac{P_i^e}{P_i^0} = \frac{\sum P_i^e \times P / P_i^0}{P} = \sum \frac{P_i^e}{P} \omega_i = TR \times W$$

其中

- ① P_i^e 为第 i 家机构发布的股票目标价格；
- ② P_i^0 表示在第 i 家机构发布价格预测前一个交易日的收盘价；
- ③ P 为月末计算因子时的股票收盘价；
- ④ N 为发布分析师预期目标价格的机构数量。

说明

加权后的目标收益率，加入了利用后续行情验证目标价格的权重 ω_i ，如果股票价格走势能够验证分析师的目标收益率 ($\omega_i > 1$)，则应给予目标价格更大的权重；反之，给予更小的权重。

[参考文献]

魏建榕, 2021, 分析师目标价的 Alpha 信息, 开源证券.



Jan.

腊月十六
星期三

1月 15

错误定价因子 MISV_TTM

基本面因子改进

基于一阶段剩余价值增长模型进行如下横截面回归预测得到股票的合理估值 $\frac{\widehat{V}_{it}}{B_{it}}$:

$$\frac{V_{it}}{B_{it}} - 1 = \gamma_0 \frac{NI_{it}}{B_{it}} + \gamma_1 \frac{B_{it-1}}{B_{it}} + \varepsilon_{it}$$

计算股票合理估值与当前估值 PB 的差异比即为错误定价因子：

$$MISV_TTM_{it} = \frac{\widehat{V}_{it}}{B_{it}} / PB_{it} - 1$$

其中

- ① V_{it} 为股票 i 在 t 时刻的总市值；
- ② B_{it} 为股票 i 在 t 时刻的净资产（内在价值）；
- ③ NI_{it} 为股票 i 在 t 时刻的归母净利润 TTM，计算时可以纳入业绩预告和业绩快报的信息，其中业绩预报取预测值上下限的均值；
- ④ PB_{it} 为股票 i 在 t 时刻的市净率估值因子。

说明

错误定价因子衡量了股票内在的合理估值与账面估值 PB 之间的差异，错误定义因子值越高，说明账面估值 PB 相对股票内在合理估值偏低，股票低估越严重。

[参考文献]

郭策, 2024, 价值掘金因子, 中银证券.



Jan.

腊月十七

星期四

1月

16

对新闻的反应波动性

另类因子-文本分析

对过去 7 个交易日内与股票 i 相关的每条新闻（共计 N 篇新闻），计算新闻发布后 15 分钟内股票分钟收益率序列的均值，并对所有新闻对应的均值等权聚合到股票维度：

$$ret_std_{i,n} = std(\text{股票 } i \text{ 在新闻 } n \text{ 发布后 15 分钟内 1 分钟收益率序列})$$

$$news_reaction_std_{i,t} = \frac{1}{N} \sum_{n=1}^N ret_std_{i,n}$$

其中

- ① 若新闻发布时间为非交易时段，则取新闻发布后第一个交易日开盘后首个 15 分钟的收益率序列计算因子；
- ② 可以对因子进行截面均值距离化处理来提高因子的单调性。

说明

该因子统计了新闻发布后 15 分钟内的股价波动率，用于衡量股价在新闻发布后短期内的反应一致性，波动率越大代表市场分歧越大，反之代表市场分歧越小；因子取值越高，未来收益越低，说明短期内存在市场对信息反应的不确定和投资者预期分歧。

[参考文献]

薛春晓, 陈泽鹏, 2024, 新闻中的有限注意力和“凸显效应”因子构建, 方正证券.



Jan.

腊月十八

星期五

1月 17

市价偏离度

高频因子-流动性类

$$MPB_t = \overline{TP}_t - \frac{M_t + M_{t-1}}{2} = \overline{TP}_t - \overline{MP}_t$$

其中

① $\overline{TP}_t = \begin{cases} \frac{T_t}{V_t} & V_t \neq 0 \\ \overline{TP}_{t-1} & V_t = 0 \end{cases}$ 为平均交易价格；

② $M_t = \frac{P_t^B + P_t^A}{2}$ 为买一价和卖一价的均值；

③ P_t^B 和 P_t^A 为 t 时刻的买一价和卖一价； T_t 和 V_t 为 t 时刻成交额和成交量。

说明

市场偏离度为平均交易价格和平均委托挂单价格的差值；当市场偏离度为正，交易均价更接近卖一价，为卖方发起的交易，卖压大，未来价格更趋向于下行，即市场交易均价将向市场平均中间价回归，反之亦然。

【参考文献】

丁鲁明, 陈升锐, 2020, 高频盈价选股因子初探, 中信建投证券.



Jan.

腊月十九

星期六

1月 18

CSSD 模型

行为金融因子-羊群效应

利用个股收益 r_i 和市场收益 R_m 计算股票收益的横截面标准差：

$$CSSD = \sqrt{\frac{\sum_{i=1}^N (r_i - \bar{r}_m)^2}{N - 1}}$$

以 CSSD 为 y , 以 D_t^u (市场收益高于计算收益阈值是取值为 1, 否则取值为 0) 和 D_t^d (市场收益低于计算收益阈值是取值为 1, 否则取值为 0) 为 x , 进行如下时序回归:

$$CSSD = \alpha + \beta_1 D_t^u + \beta_2 D_t^d + \varepsilon$$

说明

CSSD 模型对羊群效应的检验主要基于时序回归上市场极端收益项前回归系数正负及显著性；当 β_1 或 β_2 显著小于 0 时，说明市场极端上涨或下跌时，CSSD 取值明显下降，投资者的投资行为越来越一致，市场上羊群效应越明显。

[参考文献]

Christie, W. G., and Huang, R. D, 1995, Following the Pied Piper: Do Individual Returns Herd around the Market? *Financial Analysts Journal*, 51(4), 31-37.



Jan.

腊月二十

星期日

1月

19

Flesch-Kincaid 年级水平

另类因子-文本分析

$$FKGL = (0.39 * \text{words_per_sentence}) + (11.8 * \text{syllables_per_word}) - 15.59$$

其中

- ① *words_per_sentence* 为文档中单词总数/文档中句子总数；
- ② *syllables_per_word* 为文档中总音节数/总词数；
- ③ 分析的文档为年度 10-K 报告，借助自然语言处理技术得到文档单词总数、文档句子总数、文档音节总数等统计指标。

说明

Flesch-Kincaid 年级水平表示需要多少年的教育才能理解该文本，也可用于衡量年度财务报告的语言复杂度，得分越低，文档越容易阅读，复杂度越低；年报可读性与公司的业绩表现和盈利持续性存在一定关系，可用来衡量公司管理和信息披露质量，较低可读性可能意味着公司在试图掩饰不利信息，较高可读性可能意味着公司对自身业绩表现充满信心，愿意更透明化的披露信息；但财报可读性与未来股票收益之间并未发现显著关系。

【参考文献】

Li, F. 2009, "Annual report readability, current earnings, and earnings persistence", *Journal of Accounting and Economics*, 45(2-3), 221-247.



Jan.

腊月廿一
星期一

1月 20

b 型成交量分布

高频因子-成交分布类

- ① 计算同价成交量：将日内相同分钟收盘价的成交量累加至一起，得到在当前价格上的成交量总和；
- ② 将成交量累计最大的价格定义为该股当日的成交量支撑点 Volume Support Price，包含该支撑点的附近区域定义为成交量支撑区域 Volume Support Area；
- ③ 逐步计算 VSP 周围的成交量累计值（价格由近及远，成交量由大及小），该累计值与全天成交量总和的比值超 50% 的最小区域，即为成交量支撑区域 VSA，该区域上限价格为 VSA_High，下限价格为 VSA_Low；
- ④ 计算上限价格 VSA_High 与当日最低价之间的差异即得到 vsa_high2min。

说明

若上限价格与当日最低价越接近，说明成交量支撑区域越接近日内的低价区域，则该股日内的成交量分布越接近于 b 型成交量分布，预期上涨；当个股价格出现急剧下跌然后盘整，或者短期急剧上涨但整体处于低位时，成交量分布将出现 b 型曲线。

[参考文献]

郑兆磊, 2022, 成交量分布中的 Alpha, 兴业证券.



Jan.

腊月廿二

星期二

1月 21

买入意愿因子

期货-高频因子

$$buy_sell = \frac{count_{buy>sell}}{count_{all}}$$

其中

- ① $count_{buy>sell}$ 为某个期货品种在一个交易日内快照瞬间买一大于卖一的总快照数量；
- ② $count_{all}$ 为某个期货品种在一个交易日内所有快照数量。

说明

买入意愿因子代表了一个交易日内买单多于卖单的时间所占的比例，比例越大，说明市场对该品种的买入意愿越强烈，该品种在后市表现相对更优的可能性更大。

[参考文献]

赵妍, 高智威, 邬子峰, 2024, 基于日内高频博弈信息的商品 cta 策略, 国金证券.



Jan.

腊月廿三
星期三

1月 22

新闻网络节点度

另类因子-网络结构

① 以标题中股票为 lead , 以同时出现在新闻正文中的其他股票为 follower, 确定领边 $l_{ij,T}$, 并构建股票共现图的邻接矩阵 \mathcal{W}_T :

$$\begin{aligned} l_{ij,T} &\stackrel{\text{def}}{=} \bigcup_{m_d \in D_T} \{(i, j)_{m_d} \mid j \text{ in } m_d \text{ title}, i \text{ in } m_d \text{ headline}, i \neq j\} \mathcal{W}_T \\ &= [\omega_{ij,T}]_{n \times n} = \left[\frac{l_{ij,T}}{\sum_{j=1}^n l_{ij,T}} \right]_{n \times n} \end{aligned}$$

② 基于股票共现图邻接矩阵 \mathcal{W}_T 计算节点度 degree:

$$\begin{aligned} \mathcal{A}_{i,t}(\omega_{ij,t-l:t}) &= \#\{j : \omega_{ij,t-l:t} \neq 0\} + \#\{j : \omega_{ji,t-l:t} \neq 0\}, j \\ &= 1, 2, \dots, n, \end{aligned}$$

说明

新闻节点度为新闻共现网络中与节点相连的邻居节点的数量，节点度较高，说明该公司在新闻报道中频繁与其他多个公司共同提及，有较高的媒体关注度；媒体在报道时更愿意报道积极的、乐观的消息，有较高关注度的公司表现会优于那些受到较少媒体关注的公司。

[参考文献]

Hu Junjie, Härdle Wolfgang, 2021, Networks of news and cross-sectional returns, IRTG 1792 Discussion Papers 2021-023, Humboldt University of Berlin, International Research Training Group 1792 "High Dimensional Nonstationary Time Series".



Jan.

腊月廿四
星期四

1月

23

时序回归因子

期货-动量趋势因子

$$close_t = a + b * t + c * t^2$$

$$\text{时序回归因子} = b * R^2$$

其中

- ① $close_t$ 为某商品期货过去 J 日收盘价序列；
- ② t 为该收盘价序列对应的时间序列： $t = 1, 2, 3, 4, \dots, J$ ，对应长期趋势；
- ③ R^2 为上述时序回归拟合优度；
- ④ t^2 刻画了趋势的非线性部分，对应短期加速涨跌的情况。

说明

主次价差因子为次主力合约较主力合约年化升贴水率在过去 K 个交易日的均值；
做多贴水高的品种，做空贴水低的品种，通过基差收敛获得展期收益。

[参考文献]

周道, 2022, 商品期货Alpha因子拾遗, 中信期货.



Jan.

腊月廿五

星期五

1月

24

基于异常收益的注意力捕捉

行为金融因子-投资者注意力

对每个行业内分别统计每日行业成分股收益率绝对值高于 7% 的股票数量占比作为市场关注代理指标：

$$indus_abn_d = \frac{n_{>=top7\%,d}}{N_d}$$

将市场关注代理指标与股票日收益进行月度时序回归，代理指标前的系数绝对值 $|\beta_{i,t}|$ 即为股票对市场关注的敏感程度：

$$r_{i,d} = u_{i,t} + \beta_{i,t} indus_abn_d + \rho_{1,i,t} Mkt_d + \rho_{2,i,t} SMB_d + \rho_{3,i,t} HML_d + \rho_{4,i,t} UMD_d + \varepsilon_{i,t}$$

其中

- ① $n_{upper,d}$ 、 $n_{lower,d}$ 、 N_d 分别为 d 日涨停股票数量、跌停股票数量、当日交易股票总数；
- ② $r_{i,d}$ 为个股日度收益；
- ③ Mkt_d 、 SMB_d 、 HML_d 为经典的 Fama-French 三因子；
- ④ UMD_d 为动量因子收益，即全市场过去 11 个月累积收益前后 30% 的组合在 d 日的收益差。

说明

不同股票对同一市场关注度事件的反映程度往往是不同的，可借助个股收益对市场关注度事件的敏感程度来衡量股票对该事件的注意力捕捉效应；捕捉能力越强（敏感程度越高），表明个股更能吸引市场注意，从而导致买盘压力增大，股价高估；上述因子以异常收益占比作为市场关注度事件，极端日收益是衡量关注度的一个重要指标，出现较高或较低收益都代表关注度高。

[参考文献]

- ① Fengjiao LIN, Zhigang QIU, Zheng W, 2023, Cranes among chickens: The general-attention - grabbing effect of daily price limits in China's stock market, Journal of Banking & Finance, 150, 106818.
- ② 陈升锐, 2024, 投资者有限关注及注意力捕捉与溢出, 中信建投.



量化投资与机器学习



Jan.

腊月廿六

星期六

1月

25

客户重要性

另类因子-网络结构

利用 PageRank 算法计算客户重要性：

$$p(i) = \frac{q}{V} + (1 - q) \sum_{j \in \mathcal{V}: j \rightarrow i} \frac{p(j)}{k_j^{out}}$$

其中

- ① $p(i)$ 为客户 i 的 PageRank 得分；
- ② k_j^{out} 是供应商 j 的出度，即从节点 j 发出的链接数量；
- ③ q 为阻尼因子，通常取值为 0.85；
- ④ \mathcal{V} 为指向公司 i 的所有供应商节点的集合；
- ⑤ V 为图中节点的总数量。

说明

衡量客户重要性的 PageRank 得分相比客户节点度（入度），除了考虑指向客户的供应商数量外，还考虑了相连供应商本身的重要性，即使两个客户的入度相同，拥有更多“重要”供应商的客户也会有更高的 PageRank 得分。

[参考文献]

Jussa et al., 2015, The Logistics of Supply Chain Alpha, Deutsche Bank Markets Research.



Jan.

腊月廿七

星期日

1月

26

高频技术因子

商品通道

$$HLC_{i,t} = \frac{High_{i,t} + Low_{i,t} + Close_{i,t}}{3}$$

$$HLC_{ma,i,t} = \frac{\sum_{t=239}^t HLC_{i,t}}{240}$$

$$AVEDEV_{HLC,i,t} = \frac{\sum_{i=t-239}^t |HLC_{i,t} - HLC_{ma,i,t}|}{240}$$

$$CCI_{i,t} = \frac{HLC_{i,t} - HLC_{ma,i,t}}{0.015 * AVEDEV_{HLC,i,t}}$$

说明

将过去 20 个交易日的日度高频因子进行时间衰减加权即可降频为月度因子。

说明

CCI 指标衡量了一段时间内价格与股价平均区间的偏离程度。

[参考文献]

姚紫薇, 陈升锐, 2024, 技术指标因子高频化, 中信建投.



Jan.

腊月廿八

星期一

1月

27

期货-持仓类因子

套期保值比例

$$HP = \frac{|OI_t - OI_{t-J}|}{\sum_{j=0}^{J-1} Vol_{t-j}}$$

其中

- ① OI_t 、 Vol_t 分别为 t 时刻品种的持仓量和成交量；
- ② 回看期 J 较短时，因子并没有有效的选基能力；J=243 日时，选基能力较显著。

说明

套期保值比例衡量了套期保值者的参与程度；因子值越高，套期保值者参与比例可能更高，投机者更有可能从套期保值者那获得风险补偿。

【参考文献】

- ① 张革, 2022, 商品期货截面风格因子初探, 中信期货.
- ② Fan, J.H. and Zhang, T., 2019, The untold story of commodity futures in China, The Journal of futures markets, 40(4), 671-706.



Jan.

腊月廿九

星期二

1月

28

招聘广告数量

另类因子-招聘信息

方式1：使用当前公司员工数量进行标准化

$$Posting_EMP_xm_{i,t} = \frac{\sum_{j=t-x+1}^t Number_of_job_posting_{i,j}}{Number_of_employee_{i,t}}$$

方式2：使用过去 12 个月的招聘广告数量进行标准化

$$Posting_EMP_xm_{i,t} = \frac{\sum_{j=t-x+1}^t Number_of_job_posting_{i,j}}{Number_of_employee_{i,t}}$$

其中

① $Number_of_job_posting_{i,j}$ 为公司 i 在 t 月的招聘广告总数；

② $Number_of_employee_{i,t}$ 为公司 i 在 t 月的员工总数；

③ x 为回溯窗口长度，可以是 x = 1、3、6、12 个月。

说明

招聘广告数量代表了公司对未来增长或业务扩展计划的信心；短期内，职业发布的增加表明公司的预期增长，与未来股价正相关；但长期内，持续的扩招可能意味着管理层的过度自信，会导致冗余劳动力和低盈利能力，招聘广告的数量与未来股价的关系由正转负。

[参考文献]

Yin, Jenny, and the QES Team, 2024, Alpha Insights From Global Job Positions Data, Wolfe Research.



Jan.

正月初一

星期三

1月 29

时间加权平均的股票相对价格位置

高频因子-收益分布类

计算股票相对价格位置，即股票当期价格相对区间最高最低价的分位数：

$$RPP_{i,t} = \frac{(P_{i,t} - L_i)}{(H_i - L_i)}$$

将股票相对价格位置对时间进行积分，得到时间加权平均的相对价格位置 ARPP：

$$ARPP_i = \int_0^T RPP_{i,t} \cdot dt \Rightarrow ARPP_{i,t} = \frac{\left(\int_0^T P_{i,t} \cdot dt - L_i \right)}{(H_i - L_i)}$$

说明

① $RPP_{i,t}$ 表示股票 i 在某一时间区间内的时刻 t 的相对价格位置；

② H_i 和 L_i 为该时间区间内股票 i 的最高价和最低价；

③ $\int_0^T RPP_{i,t} \cdot dt$ 表示股票 i 在时间区间 0 到 T 内的时间加权平均价格（即 TWAP），可以用分钟线高开低收均值在时序上的均值作为区间的 TWAP。

说明

衡量股票在价格相对高位停留的时间长短，股票在价格相对高位停留的时间越长，因子取值越大。

[参考文献]

朱延波, 2020. 基于时间尺度挖掘的日内买卖压力, 东万证券.



Jan.

正月初二

星期四

1月

30

“国家队”持仓市值变动

分析师、机构持仓

$$\text{国家队持仓市值变动}_{i,t} = \text{国家队持仓市值}_{i,t} - \text{国家队持仓市值}_{i,t-1}$$

其中

“国家队”持仓数据从上市公司的年报、半年报及季报披露的“十大股东”及“十大非限售股东”信息中提取得到，属于“国家队”的机构类型包括：中央汇金、社保基金及养老、国家集成电路、外管局、国新控股及国新资管、证金等。

说明

“国家队”持仓市值变动衡量了“国家队”的买入强度；可以结合“国家队”资金流向对该因子进行择时：在资金入场时以动量思想做多“持仓市值变动”因子，令因子方向为正；在资金出场时，不配置该因子，进而与择时因子一起构建复合因子。

[参考文献]

郭策, 2024, 如何追踪“国家队”构建股票优选组合?(改进版2.0), 中银证券.

