量化研究与策略

iLampard

2016 年 12 月 30 日

目录

1 量化选股基础

1.1 因子的预处理

从各个数据源得到的因子数据称为原始因子。在构建Alpha策略前，要对截面原始因子进行一定的处理。

假设有k个行业（一般使用申万23个一级行业），每个行业均有m个股票。用$f_{i,t}^{j}$表示t时刻第i个行业的j支股票的因子。因为是针对截面数据做处理(t作为一个固定值)，下面的表示将隐去t角标，因子就以f_i^j表示。

1.1.1 中性化

一般来说，因子需要对行业进行中性化处理，以去除行业对因子的影响。方法是将原始因子对行业虚拟变量进行回归，取得回归的残差作为因子值。

根据假设，将这些股票的因子合并成一个$(km)\times 1$的向量，$f=(f_1^1,...f_i^j...f_k^m)^T, i=1,2...m, j=1,2,..k$，对应的$(k\times 1)$回归系数向量$\beta=(\beta_1,...\beta_j...\beta_k)^T$。

回归变量可以表示为$(km)\times k$的矩阵$X=\left(u_1^1,...u_i^j...u_k^m\right)^T$，其中$u_i^j=(0,..,1,0,...0)$ 为第i个元素为1总长度为k的单位行向量。

$$\begin{pmatrix} f_1^1 \\ f_1^2 \\ ... \\ f_i^j \\ ... \\ f_k^m \end{pmatrix}_{(km)\times 1} = \begin{pmatrix} 1 & 0 & ... & 0 \\ 1 & 0 & ... & 0 \\ ... & & & \\ 0 & 0 & 1_{j*i,i} & 0 \\ ... & & & \\ 0 & ... & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ ... \\ \beta_i \\ ... \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1^1 \\ \varepsilon_1^2 \\ ... \\ \varepsilon_i^j \\ ... \\ \varepsilon_k^m \end{pmatrix} \tag{1}$$

公式(1)中的残差向量$(\varepsilon_1^1,...\varepsilon_k^m)$就是行业中性化后的因子

[

公式(1)的含义是，行业中性化处理等价于每个股票的因子减去行业内股票该类因子的平均值。因为对指示函数$\mathbf{1}_{industry}$回归，那么回归系数β等于所有应变量的均值。

]

还有一种中性化的方法是同时对行业以及风格（如市值）进行中性化，用$(km)\times 1$的向量$\left(M_1^1, M_1^2...M_k^m\right)^T$ 表示所有股票在该截面时刻的市值，原始因子对行业虚拟矩阵和市值向量同时做回归可得

$$\begin{pmatrix} f_1^1 \\ f_1^2 \\ ... \\ f_i^j \\ ... \\ f_k^m \end{pmatrix}_{(km)\times 1} = \begin{pmatrix} 1 & 0 & ... & 0 \\ 1 & 0 & ... & 0 \\ ... & & & \\ 0 & 0 & 1_{j*i,i} & 0 \\ ... & & & \\ 0 & ... & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ ... \\ \beta_i \\ ... \\ \beta_k \end{pmatrix} + \gamma \begin{pmatrix} M_1^1 \\ M_1^2 \\ ... \\ M_i^j \\ ... \\ M_k^m \end{pmatrix} + \begin{pmatrix} \varepsilon_1^1 \\ \varepsilon_1^2 \\ ... \\ \varepsilon_i^j \\ ... \\ \varepsilon_k^m \end{pmatrix} \tag{2}$$

公式(2)中的残差向量$(\varepsilon_1^1,...\varepsilon_k^m)$就是行业中性化以及市值中性化后的因子。

1.1.2 去极值化

因子序列中要去除某些极值（异常值）的影响，一般做法是将原始因子值调整到3倍绝对偏离中位数（或者是均值）的范围内。

1.1.3 标准化

不同类型的因子（财务，市值等）需要标准化之后才能互相比较。对于原始因子序列$\{f_i, i = 1, ...N\}$, 用μ, σ表示这一序列的均值和标准差, 则标准化后的因子序列为

$$\{f_i^s = \frac{f_i - \mu}{\sigma}, i = 1, ...N\}$$

1.2 策略表现的指标计算

1.2.1 对冲组合的净值计算方法

一个对冲组合的净值计算主要涉及到调仓的设定。此处调仓是指使得组合的股票多头和对应期货（以指数代替）的空头市值相同，即完全对冲。在下一个调仓日之前，组合的股票多头和期货空头数量不再调整，随着市场的波动，组合会出现多空头寸的暴露。调仓的频率假设有三种，为年，月和日。

假设有$t = 0, 1, 2, ...N$共N个交易日，调仓日为$\{t_i^*, i = 1, 2...k\}$共k个。调仓时间日为调仓日的收盘前。

用$\{r_t^s\}$和$\{r_t^M\}$表示策略和标的指数的日简单收益序列。用V_t, V_t^s, V_t^f表示t时刻的组合价值，组合中股票账户市值，以及期货账户市值（以指数代替）。

假设组合的收益为累计复利计算方式，另外假设组合中期货账户占用资金比例为p（用于保证金账户）,那么组合中$1 - p$的部分可投资于股票账户。对冲组合的净值NPV_t变化可以演示为

- 在$t = 0$时刻建立组合，该时刻也是第一个调仓日。此时$V_0^s = V_0^f$, $NPV_0 = 1.0$
- 在$t = 1$时刻，组合的价值变动为$V_0^s(1 + r_1^s) - V_0^f(1 + r_1^f)$, 此时

$$\mathrm{NPV}_1 = NPV_0 \left(1 + (1 - p)[(1 + r_1^s) - (1 + r_1^f)]\right)$$

[

Proof. 根据股票账户以及期货账户的资金比例假设以及$V_0^s = V_0^f$可知

$$\begin{aligned}
\mathrm{NPV}_1 &= \mathrm{NPV}_0 \left(1 + \frac{V_0^s(1 + r_1^s) - V_0^f(1 + r_1^f)}{V_0}\right) \\
&= \mathrm{NPV}_0 \left(1 + \frac{V_0(1 - p)(1 + r_1^s) - V_0(1 - p)(1 + r_1^f)}{V_0}\right) \\
&= \mathrm{NPV}_0 \left(1 + (1 - p)[(1 + r_1^s) - (1 + r_1^f)]\right)
\end{aligned}$$

□

]

– 在下一个调仓日之前(包括调仓日)的任意$t=n$时刻，净值为

$$\mathrm{NPV}_n = \mathrm{NPV}_0 \left(1+(1-p)[\prod_{i=1}^{n}(1+r_i^s)-\prod_{i=1}^{n}(1+r_1^f)]\right) \tag{3}$$

– 在第一个调仓日t_1^*之后，$t=t_1^*+1$ 时刻，净值为

$$\mathrm{NPV}_{t_1^*+1} = NPV_{t_1^*}\left(1+(1-p)[(1+r_{t_1^*+1}^s)-(1+r_{t_1^*+1}^f)]\right) \tag{4}$$

从公式(3)和公式(4)可以得出以下结论

(i) 任意两个调仓日t_i^*, t_{i+1}^*之间，对冲组合的净值为

$$\mathrm{NPV}_{t_1^*+n} = \mathrm{NPV}_{t_1^*} \left(1+(1-p)[\prod_{i=1}^{n}(1+r_i^s)-\prod_{i=1}^{n}(1+r_1^f)]\right) \tag{5}$$

(ii) 期货保证金账户的资金占用降低了组合净值的变化。假设期货保证金占用为零，此时对冲组合的净值为

$$\mathrm{NPV}_{t_1^*+n} = \mathrm{NPV}_{t_1^*} \left(1+[\prod_{i=1}^{n}(1+r_i^s)-\prod_{i=1}^{n}(1+r_1^f)]\right) \tag{6}$$

(iii) 有一种特殊的情况为每日调仓，保证每日结算时多空敞口为零，此种情况下

$$\mathrm{NPV}_{n+1} = \mathrm{NPV}_n \left(1+[(1+r_{n+1}^s)-(1+r_{n+1}^f)]\right) \tag{7}$$

2 量化选股策略

2.1 动态情景多因子策略

[动态情景多因子Alpha模型，东方证券，朱建涛，2016]

2.1.1 概述

传统多因子模型大多是全市场范围对所有股票一视同仁打分评价，而忽视了个股的基本面情况差异以及选股因子对于不同风格股票组合的适用性。本文的动态情景多因子模型(Dynamic Contextual Alpha Model)将全市场的股票按照规模、估值、成长等因子进行了划分，在此基础上对不同分档的股票组合中对个股再赋予一定权重；通过这一方法大幅提升模型对于市场风格切换的适应能力。

2.1.2 算法

假设调仓频率为月度，调仓时点为每个月末。

(i) 在调仓日根据风格因子(可称为分层因子)将股票池分组：方法为因子值从低到高排列，对半分组。每个分层因子对应一个股票池的分组

风格维度	因子名称和含义
规模	总市值
价值	最近财报的净资产/总市值
成长	净资产同比增长率
盈利	净资产收益率
流动性水平	季度日均换手率

表 1: 分层因子表

(ii) 在不分层的情况下，在调仓日对每个α因子进行去极值化和标准化处理, 参见(1.1.2)和(1.1.3)

(iii) 在不分层的情况下，在调仓日对每个α因子和标的阶段收益率R进行行业和市值中性化处理（原文中称为风险调整）。

$f_{i,t}$表示调仓日t时刻的因子i向量，R_t表示调仓日$t-1$到t之间的股票收益向量，M_t为调仓日t时刻的市值向量，X 为行业虚拟矩阵，参见公式(2)。调整后的因子和收益向量为

$$f_{i,t}^a = f - X\beta_1 - \beta_2 * \ln M_t$$
$$R_t^a = R - X\beta_1 - \beta_2 * \ln M_t$$

(iv) 在每个分层空间内，对每个因子计算风险调整IC, 用$IC_{i,t}^a$表示调仓日t时刻的基于因子i的风险调整因子

$$IC_{i,t}^a = \mathrm{Cor}\left(f_{i,t-1}^a, R_t^a\right)$$