

量化研究与策略

iLampard

2016 年 12 月 26 日

目录

1	量化选股基础	2
1.1	因子的预处理	2
1.1.1	中性化	2
1.1.2	去极值化	2
1.1.3	标准化	2
1.2	策略表现的指标计算	3
1.2.1	对冲组合的净值计算方法	3
2	量化选股策略	4

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， $f = (f_1^1, \dots, f_i^j \dots f_k^m)^T, i = 1, 2 \dots m, j = 1, 2, \dots k$ ，对应的 $(k \times 1)$ 回归系数向量 $\beta = (\beta_1, \dots, \beta_j \dots \beta_k)^T$ 。

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

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

公式(1)中的残差向量 $(\varepsilon_1^1, \dots, \varepsilon_k^m)$ 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。

]

风格中性处理的一般方法则是将标准化的因子对市值对数和beta进行回归，去回归的残差作为因子值。

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。对于去极值化后的因子序列 $\{f_i, i = 1, \dots, N\}$ ，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

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

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

用 $\{r_t^s\}$ 和 $\{r_t^f\}$ 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

$$NPV_1 = NPV_0 \left(1 + (1 - p)[(1 + r_1^s) - (1 + r_1^f)] \right)$$

[

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

$$\begin{aligned} NPV_1 &= NPV_0 \left(1 + \frac{V_0^s(1 + r_1^s) - V_0^f(1 + r_1^f)}{V_0} \right) \\ &= NPV_0 \left(1 + \frac{V_0(1 - p)(1 + r_1^s) - V_0(1 - p)(1 + r_1^f)}{V_0} \right) \\ &= NPV_0 \left(1 + (1 - p)[(1 + r_1^s) - (1 + r_1^f)] \right) \end{aligned}$$

□

]

- 在下一个调仓日之前(包括调仓日)的任意 $t = n$ 时刻，净值为

$$NPV_n = NPV_0 \left(1 + (1 - p) \left[\prod_{i=1}^n (1 + r_i^s) - \prod_{i=1}^n (1 + r_i^f) \right] \right) \quad (2)$$

- 在第一个调仓日 t_1^* 之后， $t = t_1^* + 1$ 时刻，净值为

$$NPV_{t_1^*+1} = NPV_{t_1^*} \left(1 + (1 - p)[(1 + r_{t_1^*+1}^s) - (1 + r_{t_1^*+1}^f)] \right) \quad (3)$$

从公式(2)和公式(3)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

$$NPV_{t_i^*+n} = NPV_{t_i^*} \left(1 + (1 - p) \left[\prod_{i=1}^n (1 + r_i^s) - \prod_{i=1}^n (1 + r_i^f) \right] \right) \quad (4)$$

(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

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

(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

$$\text{NPV}_{n+1} = \text{NPV}_n \left(1 + [(1 + r_{n+1}^s) - (1 + r_{n+1}^f)] \right) \quad (6)$$

2 量化选股策略