

【云通整理】基于组合权重优化的风格中性多因子选股策略

前言：

任意股票都在同一时刻暴露于多种不同的风险因子下,它们之间的共同作用影响了股票价格的波动。通过对不同风险因子的梳理研究,可实现对股票收益来源的分解剥离,从而定量地研究股价波动的成因。多因子选股作为量化投资研究领域的经典模型,在海内外各类投资机构均受到了广泛研究和实践应用。本文从结构化多因子风险模型入手,探究 A 股市场的整体风险结构。

马科维茨均方差模型：

该模型认为投资者可以用预期收益率 $E[R_p]$, 以及收益率的标准差 σ_p 来完全构建和衡量一个投资组合。依据该观点, 对于一个资产数目为 N , 且各资产头寸相同的投资组合, 如果已知每一个资产收益率的方差 $Var(R_i)$ 和资产两两之间的协方差 $Cov(R_i, R_j)$, 那么我们可以计算这个投资组合的方差：

$$Var(R_p) = Var(R_1 + R_2 + \dots + R_N) = \sum_{i=1}^N Var(R_i) + \sum_{i=j=1}^N Cov(R_i, R_j)$$

对于包含 N 个资产的投资组合, 我们需要计算 $N(N-1)$ 个协方差, 通常以协方差矩阵来表示：

$$\begin{pmatrix} Var(R_1) & Cov(R_1 R_2) & \dots & Cov(R_1 R_N) \\ Cov(R_2 R_1) & Var(R_2) & \dots & Cov(R_2 R_N) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(R_N R_1) & Cov(R_N R_2) & \dots & Var(R_N) \end{pmatrix}$$

此时, 资产的协方差矩阵包含了我们投资组合的一切风险信息。在实际计算中,

我们需要通过历史数据来计算经验协方差矩阵(Empirical Covariance Matrix), 作为协方差矩阵的估计。然而, 使用经验协方差矩阵存在以下问题:

- (1) 数据量要求大。要对协方差矩阵实现较为准确的估计, 需要保证观测值数目大于矩阵的维数。考虑以沪深 300 的 300 个成分股作为投资组合, 以月度数据计算经验协方差矩阵, 则需要至少 $300/12=25$ 年的数据, 因此缺乏现实可行性;
- (2) 依据历史数据进行协方差估计无法反映投资组合中资产的结构变化 (例如并购) ;
- (3) 大量资产两两之间的协方差计算, 容易出现多重比较谬误 (multiple comparison fallacy) 的问题, 因而引起资产之间相关性的错误判断。
- (4) 历史数据中包含大量的噪音, 因此简单使用资产的协方差矩阵进行预测会造成较大的偏差。

为了克服与改进上述模型的缺陷, 结构化多因子风险模型应运而生。

结构化风险因子模型:

结构化风险因子模型利用一组共同因子和一个仅与该股票有关的特殊因子解释股票的收益率, 并利用共同因子和特殊因子的波动来解释股票收益率的波动。结构化多因子风险模型的优势在于, 通过识别重要的因子, 可以降低问题的规模, 只要因子个数不变, 即使股票组合的数量发生变化, 处理问题的复杂度也不会发生变化。

结构化多因子风险模型首先对收益率进行简单的线性分解, 分解方程中包含四个组成部分: 股票收益率、因子暴露、因子收益率和特殊因子收益率。那么, 第 j 只股

票的线性分解如下所示:

$$r_j = x_1 f_1 + x_2 f_2 + x_3 f_3 + x_4 f_4 \dots x_K f_K + u_j$$

其中, r_j 表示第 j 只股票的收益率; x_k 表示第 j 只股票在第 k 个因子上的暴露(也称为因子载荷); f_k 表示第 j 只股票第 k 个因子的因子收益率(即每单位因子暴露所承载的收益率); u_j 表示第 j 只股票的特殊因子收益率。

对于上述方程的时间结构,若我们定义因子暴露是在时刻 t 的结果,那么股票收益率、因子收益率和特殊因子收益率均为 $t+1$ 的结果。在模型中,我们以月频率处理截面数据。

那么对于一个包含 N 只股票的投资组合,假设组合的权重为 $w = (w_1, w_2, \dots, w_N)^T$,那么组合收益率可以表示为:

$$R_p = \sum_{j=1}^N w_n \cdot \left(\sum_{k=1}^K x_{jk} f_{jk} + u_j \right)$$

现在我们假设每只股票的特殊因子收益率与共同因子收益率不相关,并且每只股票的特殊因子收益率也不相关。那么在上述表达式的基础上,可以得到组合的风险结构为:

$$\sigma_p = \sqrt{w^T (X F X^T + \Delta) w}$$

其中, X 表示 N 只个股在 K 个风险因子上的因子载荷矩阵 ($N \times K$) :

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,k} \\ x_{2,1} & x_{2,2} & \dots & x_{2,k} \\ \dots & \dots & \dots & \dots \\ x_{n,1} & x_{n,2} & \dots & x_{n,k} \end{bmatrix}$$

F 表示 K 个因子的因子收益率协方差矩阵 ($K \times K$) :

$$F = \begin{bmatrix} Var(f_1) & Cov(f_1, f_2) & \dots & Cov(f_1, f_k) \\ Cov(f_1, f_2) & Var(f_2) & \dots & Cov(f_2, f_k) \\ \dots & \dots & \dots & \dots \\ Cov(f_k, f_1) & Cov(f_k, f_2) & \dots & Var(f_k) \end{bmatrix}$$

其中,因子收益率的波动率和协方差以因子收益率的日频率数据估算。

Δ 表示 N 只股票的特殊因子收益率协方差矩阵 ($N \times N$) :

$$\Delta = \begin{bmatrix} Var(u_1) & 0 & \dots & 0 \\ 0 & Var(u_2) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & Var(u_k) \end{bmatrix}$$

由于我们假设每只股票的特殊因子收益率相关性为 0,因此 Δ 为对角阵。

结构化多因子风险模型对于投资者的意义并不在于最小化组合风险,而是在给定组合预期收益率的情况下,赋予最优的风险权重。

风险因子筛选:

下面我们来讨论风险模型中的共同因子部分, 在 Barra 的结构化多因子风险模型中, 通常将共同因子分为行业因子和风格因子两部分。

行业因子是风险模型的重要部分,通过对 A 股市场全部个股的行业划分, 反映了

个股所属行业的独有特点，在模型中行业因子以虚拟变量(Dummy Variable)来表示,取值 1 或 0 表示属于/不属于该行业。我们在申万一级 28 个行业分类的基础上,将非银金融行业划分为证券、保险和多元金融 3 个子行业。因此,我们的风险模型包含 30 个行业因子,具体为:

表 1: 行业因子分类					
交通运输	休闲服务	传媒	公用事业	农林牧渔	化工
医药生物	商业贸易	国防军工	家用电器	建筑材料	建筑装饰
房地产	有色金属	机械设备	汽车	电子	电气设备
纺织服装	综合	计算机	轻工制造	通信	采掘
钢铁	银行	证券	保险	多元金融	食品饮料

数据来源：国泰君安证券研究

风格因子是共同因子中另一重要部分,风格因子将在后面我们讨论纯因子组合与股票组合权重优化中起到至关重要的作用。风格因子总共包含 9 大类因子、20 个小因子,其中大类因子包含 Beta、Momentum、Size、Earnings Yield、Volatility、Growth、Value、Leverage 和 Liquidity, 具体为:

表 2: 风格因子分类构建		
大类因子	小类因子	因子计算方式
Beta	BETA	$r_i = \alpha + \beta r_m + e_i$; 利用个股收益率序列和沪深 300 指数收益率序列进行一元线性回归, 收益率序列长度取 250 交易日。股收益率序列和沪深 300 指数收益率序列均以半衰指数加权, 半衰期为 60 日。
Momentum	RSTR	$RSTR = \sum_{t=L}^{T+L} w_t [\ln(1 + r_t)]$; 其中 T=500, L=21, 收益率序列以半衰指数加权, 半衰期为 120 日。
Size	LNCAP	$LNCAP = LN(total_market_capitalization)$; 个股总市值对数值。

	EPIBS	$EPIBS = est_eps / P$; 其中 est_eps 为个股一致预期基本每股收益。
Earnings Yield	ETOP	$ETOP = earnings_ttm / mkt_freeshares$; 历史 EP 值, 利用过去 12 个月个股净利润除以当前市值。
	CETOP	$CETOP = Cash_earnings / P$; 个股现金收益比股票价格。
Volatility	DASTD	$DASTD = (\sum_{t=1}^T w_t \cdot (r_t - \mu(r))^2)^{1/2}$; 其中收益率序列长度取 250 个交易日, 半衰期设定为 40 日。
	CMRA	$CMRA = \ln(1 + \max\{Z(T)\}) - \ln(1 + \min\{Z(T)\})$; 其中 $Z(T) = \sum_{\tau=1}^T [\ln(1 + r_{\tau})]$; r_{τ} 表示个股月收益率, T 代表过去 12 个月。
	HSIGMA	$HSIGMA = std(e_i)$; 其中残差 e_i 为 BETA 计算中所得。
Growth	SGRO	过去 5 年企业营业总收入复合增长率。
	EGRO	过去 5 年企业归属母公司净利润复合增长率。
	EGIB	未来 3 年企业一致预期净利润增长率。
	EGIB_S	未来 1 年企业一致预期净利润增长率。
Value	BTOP	$BTOP = common_equity / current_market_capitalization$; 计算企业总权益值除以当前市值。
	MLEV	$MLEV = (ME + LD) / ME$; 其中 ME 表示企业当前总市值, LD 表示企业长期负债。
Leverage	DTOA	$DTOA = TD / TA$; 其中 TD 表示总负债 TA 表示总资产。
	BLEV	$BLEV = (BE + LD) / BE$; 其中 BE 表示企业账面权益, LD 表示企业长期负债。
Liquidity	STOM	$STOM = \ln(\sum_{i=1}^{21} (V_i / S_i))$; 其中 V_i 表示当日成交量, S_i 表示流通股本。
	STOQ	$STOQ = \ln(\frac{1}{T} \sum_{\tau=1}^T \exp(STOM_{\tau}))$; 其中 T=3。
	STOA	$STOA = \ln(\frac{1}{T} \sum_{\tau=1}^T \exp(STOM_{\tau}))$; 其中 T=12。

数据来源: 国泰君安证券研究

其中,风格因子 Earnings Yield、Volatility、Growth、Leverage 和 Liquidity 均利用若干小类因子组合得到。另外,我们利用样本内单类因子回归最小化残差平方和的方法得到小类因子的组合权重。

总结:

在这一份报告里, 我们简要介绍了 Barra 结构化风险模型, 并对一系列风格因子进行了构建和分析。我们以沪深 300 成分股作为投资组合进行测试。测试结果表明, 大部分因子均较好地反映了沪深 300 指数的变化特征, 以及在不同的市

场环境下投资者的一些投资倾向。在动量因子的分析中，我们发现因子只能反映市场的长期动量变化，而当市场出现大幅震荡的行情时，其变化出现了一定的滞后性。所以在下一步的建模中，我们将进一步添加表征短期和中期动量的细分因子，使其能够捕捉不同情况下的市场变化趋势。

参考文献：

- [1] 国泰君安：基于组合权重优化的风格中性多因子选股策略.
- [2] 私募云通：基金业绩归因系列之 Barra 因子模型
- [3] 私募云通：Barra 对冲基金风险模型（一）