

基于收益预测和 CVaR 约束的投资组合优化

袁靖松

金融数学专题展示

2023.12

Outline

1 研究背景

2 理论模型

3 数据实验

4 结论与不足

改进期望收益

- Mean-Variance 存在缺陷，模型建立在资产收益率服从正态分布等诸多限制性假设之上；均值作为期望收益与实际市场不符；低通滤波；不适用于短期投资；

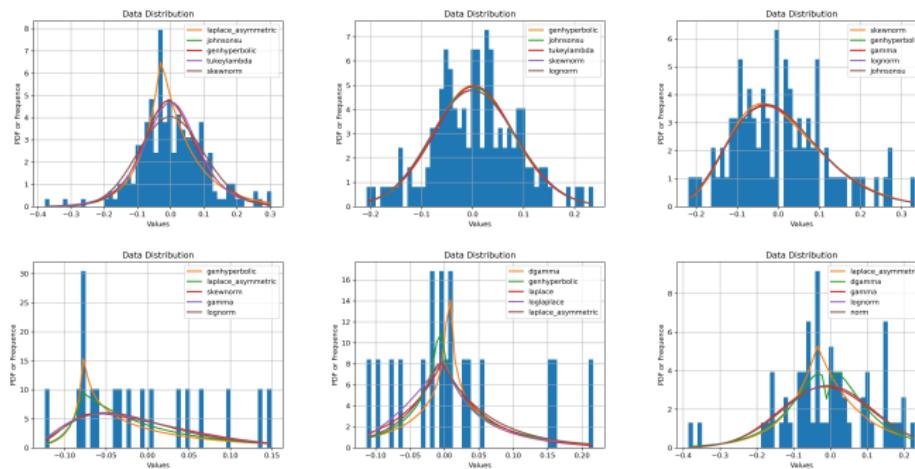


Figure 1: Distribution

改进风险度量

① 使用方差度量风险

- 对异常值敏感，极端风险事件，会显著影响方差计算；
- 只关注收益率分布的前两矩，忽略分布的尾部；

② 使用 CVaR 来度量风险损失

- 对异常值不敏感
- 不依赖分布假设
- 考虑分布的尾部，更好处理异常值

资产及收益率

- N 种资产
- 收益率矩阵 $N \times T$

$$R = \begin{bmatrix} r_1^1 & r_1^2 & r_1^3 & \dots & r_1^T \\ r_2^1 & r_2^2 & r_2^3 & \dots & r_2^T \\ r_3^1 & r_3^2 & r_3^3 & \dots & r_3^T \\ \dots & \dots & \dots & \dots & \dots \\ r_N^1 & r_N^2 & r_N^3 & \dots & r_N^T \end{bmatrix} = [R_1 \quad R_2 \quad R_3 \quad \dots \quad R_T]$$

其中 R_t 为 t 时期 N 种资产的收益率矩阵；

- 权重矩阵 $N \times 1$

$$w = [w_1 \quad w_2 \quad w_3 \quad \dots \quad w_N]^T$$

VaR 及 CVaR

- **价值损失函数**

$$L(w, R) = -w^T R$$

其中， $w \in W \subseteq R^N$ ， W 为可行集； R 为随机变量，密度函数为 $P(R)$ ；

- $L(w, R)$ 是依赖于 w 的随机变量，小于临界值 λ 的概率为：

$$\varphi(w, \lambda) = \int_{L(w, R) \leq \lambda} P(R) dR$$

- **置信水平**: $\alpha \in (0, 1)$

$$\text{VaR} : \lambda_\alpha(w) = \inf\{\lambda \in R; \varphi(w, \lambda) \geq \alpha\}$$

$$\text{CVaR} : \phi_\alpha(w) = E[L(w, R) | L(w, R) \geq \text{VaR}(w)]$$

$$= \frac{1}{1 - \alpha} \int_{L(w, R) \geq \lambda_\alpha(w)} L(w, R) P(R) dR$$

CVaR 计算

- 利用 Rockafellar 和 Uryasev 同时计算 VaR 和 CVaR 的功能函数 $F_\alpha(w, R)$ 有

$$F_\alpha(w, \lambda) = \lambda + \frac{1}{1 - \alpha} \int_{R \in R^N} [L(w, R) - \lambda]^+ P(R) dR$$

其中， λ 即 VaR 值， $[0, x]^+ = \max(0, x)$

- Rockafellar 和 Uryasev 提出，当 $P(R)$ 的解析式未知时，可以使用基于历史数据的情景分析法，产生情景矩阵，将多重积分转化为求和运算；
- 假设有 m 种情况，可以取 N 种债券 m 个时期的收益率

$$\widehat{F}_\alpha(w, R) = \lambda + \frac{1}{m(1 - \alpha)} \sum_{t=1}^m [L(w, R_t) - \lambda]$$

优化问题建模

- 目标函数
 $\min_w \lambda + \frac{1}{m(1-\alpha)} \sum_{t=1}^m [-w^T R_t - \lambda]$
 - 决策变量: w
 - 约束条件
 $s.t. w^T I = 1$

数据实验

① 历史数据

- 沪深 300 名单内股票及指数月度收益数据
- 时间范围 2005.04(month85) 至 2021.12(month285)

② 部分因子

- 财务因子**: 利润能力、偿债能力、资产周转率、现金流;
- 成长因子**: 营收增长、利润增长、现金流增长;
- 市场因子**: 股价相关、市值因子、动量;
- 其他因子**: 风险、交易量、评级、股东持股变化、技术指标;

财务因子	成长因子	市场因子	风险因子	技术指标	其他因子
ROE_q	Sales_G_q	return_1m	beta	macd	LN_capital
ROE_ttm	Profit_G_q	return_3m	std_1m	dea	HAlpha
ROA_q	OCF_G_q	return_6m	std_3m	dif	grossprofitmargin_q

Table 1: Factors

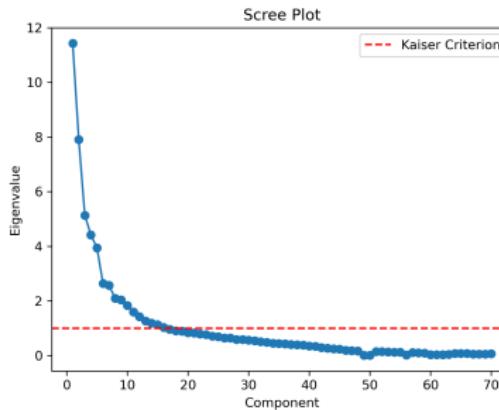
选股及预测

① 打分法选股

- 线性回归模型拟合训练集数据
- 分别使用回归系数、信息系数、因子协方差法为因子分配权重
- 根据不同股票的因子，得到分数较高的股票

② 预测测试集收益率

- 使用 PCA 进行因子合成 (Number of components based on Kaiser Criterion: 16)



优化问题重构

- ① 决策变量为 *weights*
- ② 根据历史数据计算 VaR
 - 历史组合收益 $\text{Portfolio return} = \text{Weights}.T \times R$
 - 置信水平为 $\alpha = 99\%$
 - 拟合组合收益的分布，取 1% 处分位数为 *VaR* 值 = λ
- ③ 加入预测数据计算 CVaR(目标函数)
 - 合并过去 40 个月历史数据，及预测的下 1 个月数据，形成新的数据集
 - 代入 $m, \alpha, L(w, R_t) = -\text{Weights}.T \times R_t, \lambda$ 到之前的解析式中

$$\widehat{F}_\alpha(w, R) = \lambda + \frac{1}{m(1-\alpha)} \sum_{t=1}^m [L(w, R_t) - \lambda]$$

Rolling Window

- ① 投资目标是短期投资
 - ② 使用滚动训练的方法调整回归模型
 - 训练集最少使用 60 个月的数据做训练
 - 测试集是训练集之后的 3 个月
 - 每 3 个月更新一次训练集及测试集
 - ③ 调整每次选出来的股票
 - 测试集中每个月都更新选股

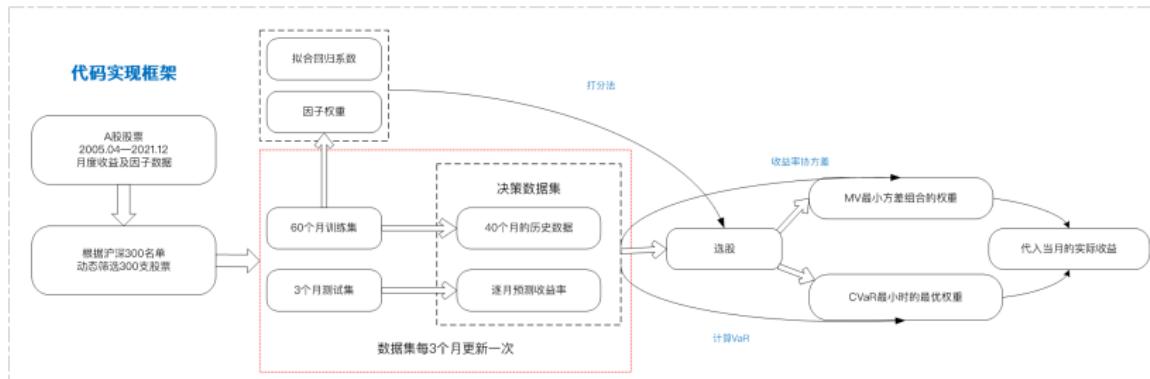


Figure 3: Framework

回测思路

① 三种因子分配权重方法

- 按回归系数归一化配比
- 按信息系数归一化配比
- 按协方差矩阵的逆做配比

② 两种最优权重求解方法

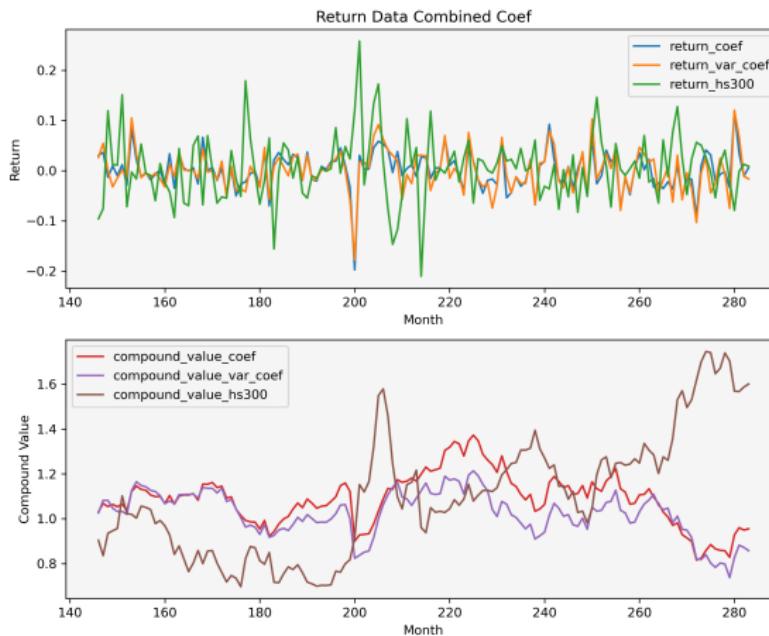
- Mean-Variance 模型最小方差组合的权重
- 使 CVaR 值最小的权重

③ 与沪深 300 指数的收益对比

- 分别将两个最优权重，带入当月的实际收益率，计算组合收益
- 将组合的收益与沪深 300 指数的收益对比

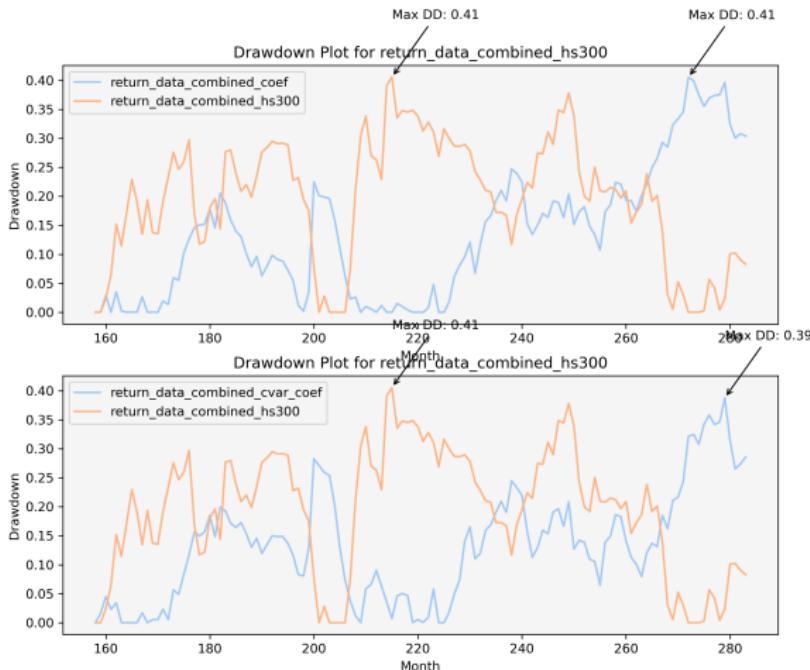
策略收益表现-Coeff

- 10 年期收益及累计收益；总体暂时没有稳定跑赢指数；
- MV 策略总体优于 CVaR



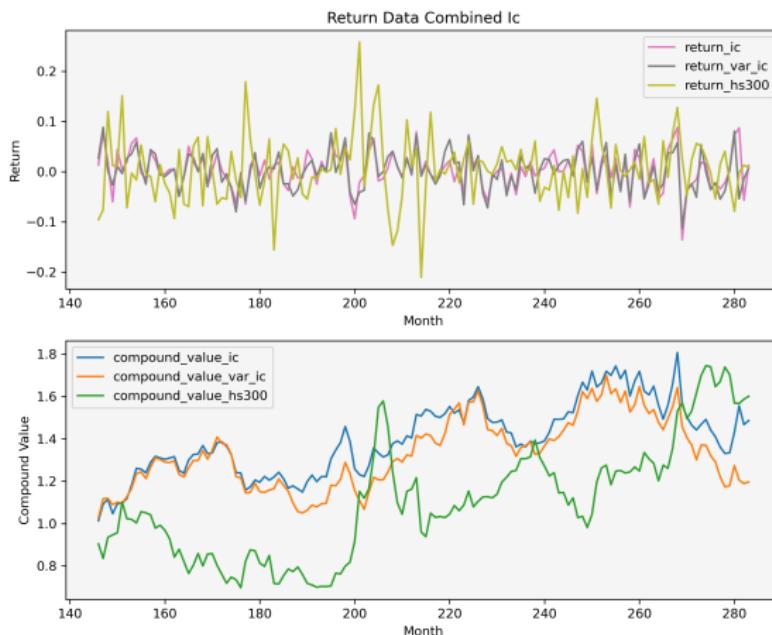
策略回撤-Coef

- 10 年期收益及累计收益；总体暂时没有稳定跑赢指数；
- MV 策略总体优于 CVaR



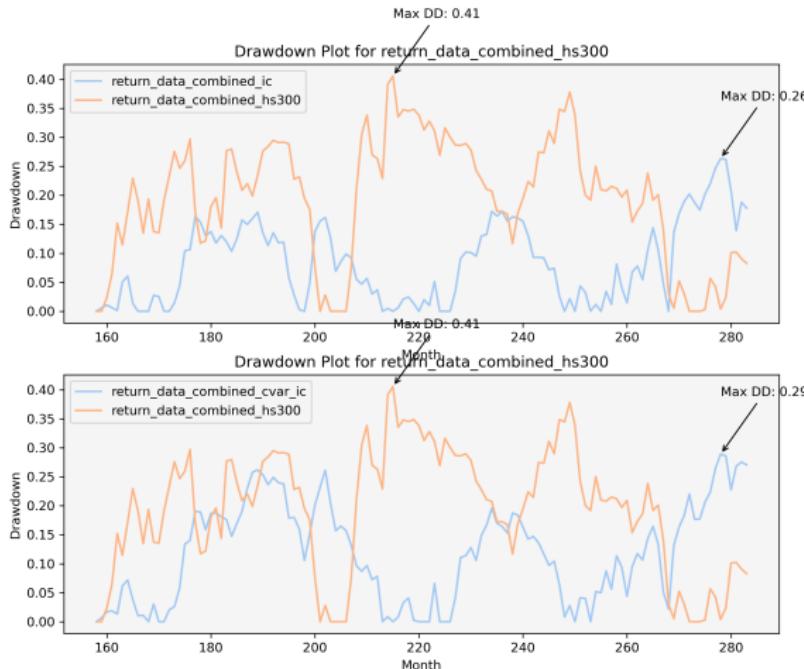
策略收益表现-IC

- 总体上基本可以跑赢指数；
- MV 策略与 CVaR 策略接近



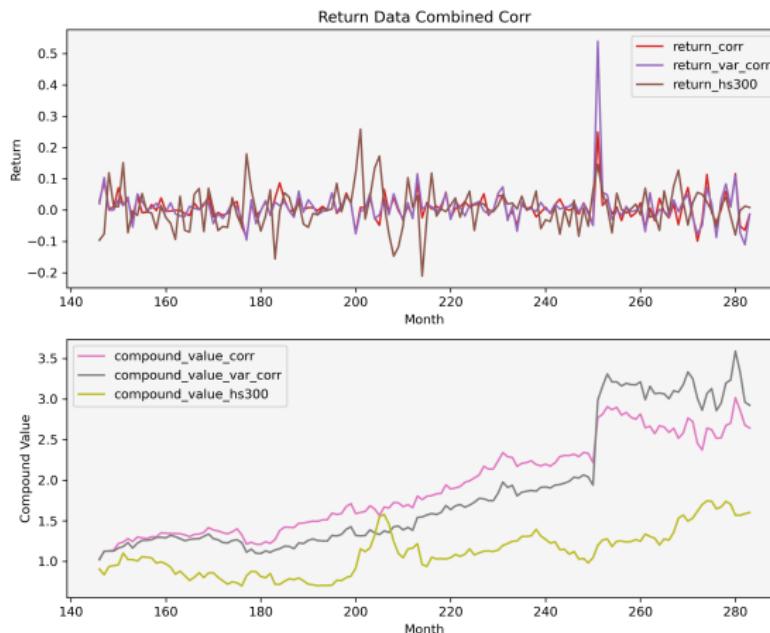
策略回撤-IC

- 10 年期收益及累计收益；总体暂时没有稳定跑赢指数；
- MV 策略总体优于 CVaR



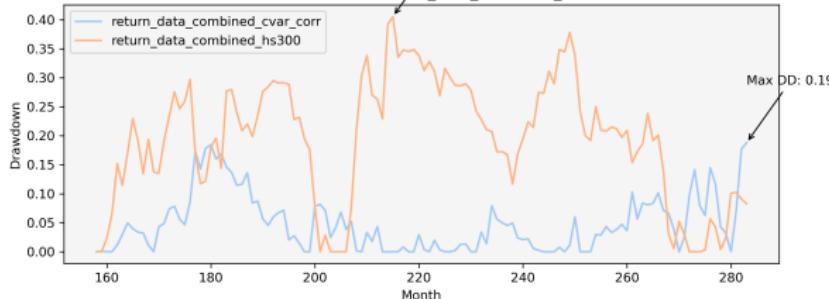
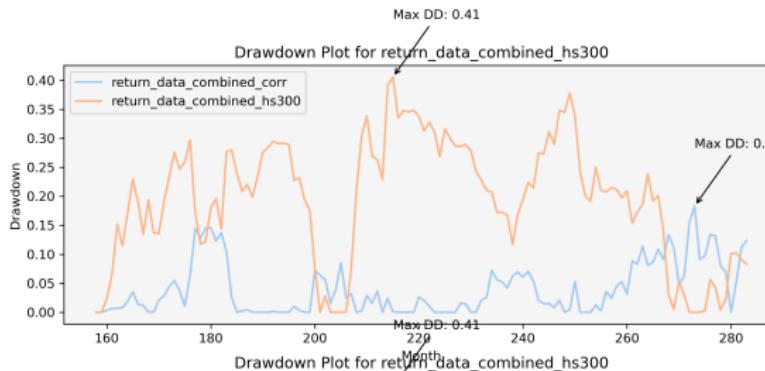
策略收益表现-Corr

- 总体上大幅跑赢指数；
- CVaR 策略逐步优于 MV 策略



策略回撤-Corr

- 10 年期收益及累计收益；总体暂时没有稳定跑赢指数；
- MV 策略总体优于 CVaR



小结

① 理论端

- 考虑交易成本

② 预测端

- 尝试更多 ML/DNN 方法

③ 策略端

- 改进滚动及调仓周期
- 改进因子配比方法

④ 数据端

- 使用更高频数据

⑤ 求解方法

- 松弛线性规划 (LP)
- 次梯度算法 (SA)
- 遗传算法 (GA)
- 光滑化方法 (SM)

不足与改进

① 理论端

- 考虑交易成本

② 预测端

- 尝试更多 ML/DNN 方法

③ 策略端

- 改进滚动及调仓周期
- 改进因子配比方法

④ 数据端

- 使用更高频数据

⑤ 求解方法

- 松弛线性规划 (LP)
- 次梯度算法 (SA)
- 遗传算法 (GA)
- 光滑化方法 (SM)