

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；

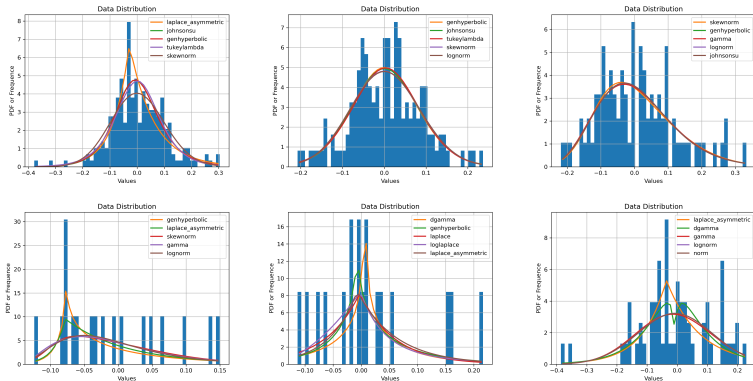


Figure 1: Distribution

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
- ③ 结合机器学习方法与现代投资组合优化理论 [1]

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 $F_\alpha(w, \lambda)$ 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, $[x]^+ = \max(0, x)$

- 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 [3]

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

优化问题建模

1 CVaR 最小化

- 目标函数 (决策变量: w)

$$\min_w \lambda + \frac{1}{m(1-\alpha)} \sum_{t=1}^m [-w^T R_t - \lambda]^+$$

- 约束条件

$$\text{s.t. } w^T I = 1$$

$$w \geq 0$$

2 方差最小化

- 目标函数 (决策变量: w)

$$\min_w w^T \Sigma w$$

- 约束条件

$$\text{s.t. } w^T I = 1$$

$$w \geq 0$$

数据及因子

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	HAalpha
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
- 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)

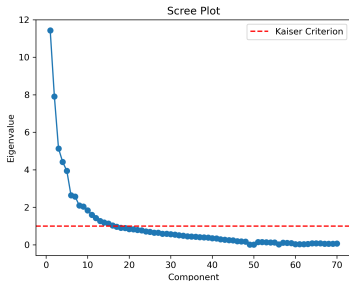


Figure 2: Scree Plot

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 $Portfolio\ return = Weights.T \times R$
 - 置信水平为 $\alpha = 99\%$
 - 拟合组合收益的分布，取 1% 处分位数为 $VaR\ 值 = \lambda$
- 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 $m, \alpha, L(w, R_t) = -Weights.T \times R_t$, λ 到之前的解析式中

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

Rolling Window

- 1 投资目标是短期投资
- 2 使用滚动训练的方法调整
 - 训练集最少使用 60 个月的数据，最多不超过 84 个月
 - 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
 - 测试集每个月轮动一次，持仓时间实际上是 1 个月
 - 测试集使用上一个月的因子数据，结合打分法选出下一个月的股票 (避免未来函数)

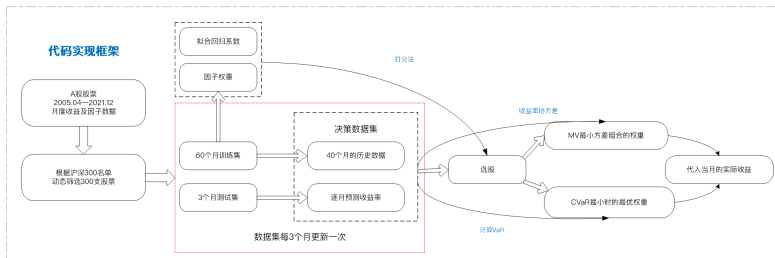


Figure 3: Framework

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
 - 将最优权重分别代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-Coef

- 没有稳定明显地跑赢指数,CVaR 策略小幅度优于 MV 策略
- 交易因子和财务因子, 占较大权重

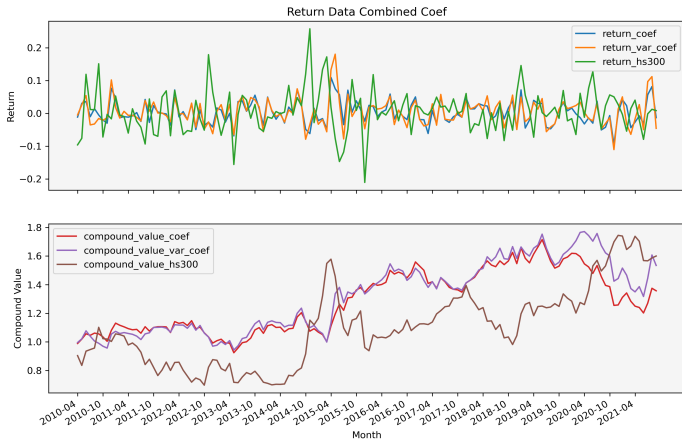


Figure 4: Return and Compound Value

策略回撤-Coef

- 两种策略最大回撤在 20% – 30%，总体上回撤幅度低于指数
- 大幅回撤发生在 2020 年 5 月份之后

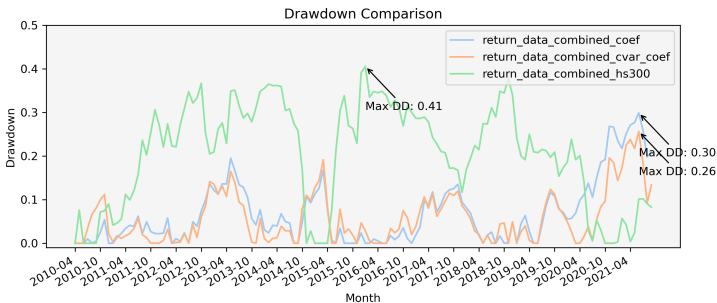


Figure 5: Drawdown Comparison Coef

策略收益表现-IC

- 总体上完全可以跑赢指数，CVaR 策略完全优于 MV 策略；
- 2015 年股市异常波动，高杠杆场外配资是股市异常波动主要原因；2020 年年初，选股策略失效，可能是受疫情影响；

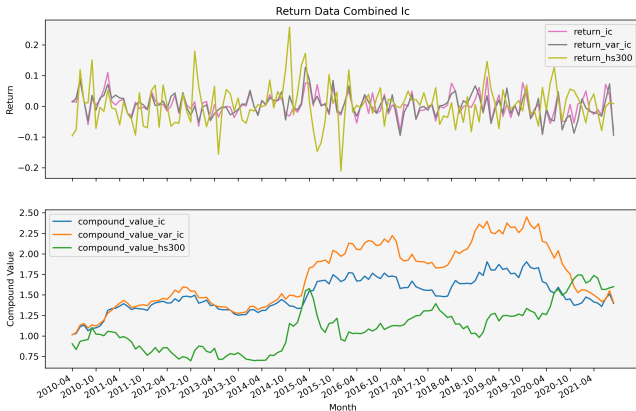


Figure 6: Return and Compound Value

策略回撤-IC

- 在 2019 年末及 2020 年初，回撤幅度大幅增加
- 财务因子及股价因子，占较大权重

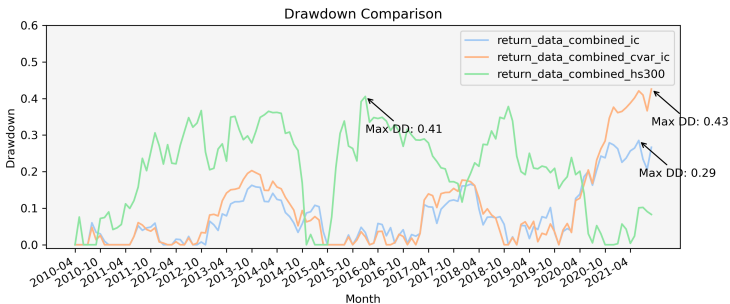


Figure 7: Drawdown Comparison IC

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；
- 2015 年异常波动；2019 年年初异常波动，处于 18 年以来衰退末期，风险偏好回升，估值修复；

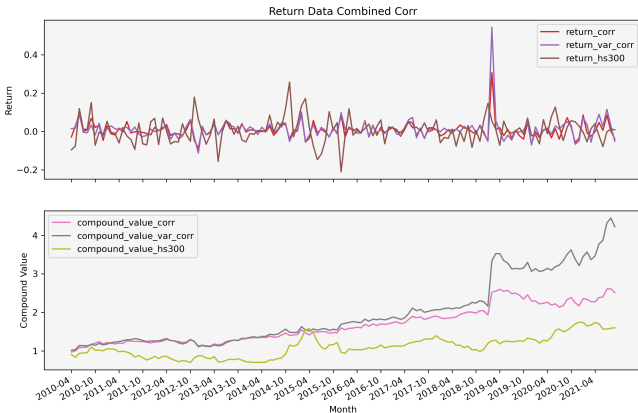


Figure 8: Return and Compound Value

策略回撤-Corr

- 总体上回撤幅度较小
- 股价及增长率因子占比为主，重视技术及情绪因子
- 疫情之后，传统的成长，动量等因子逐渐失效，估值，盈利，杠杆，非线性市值等因子表现突出

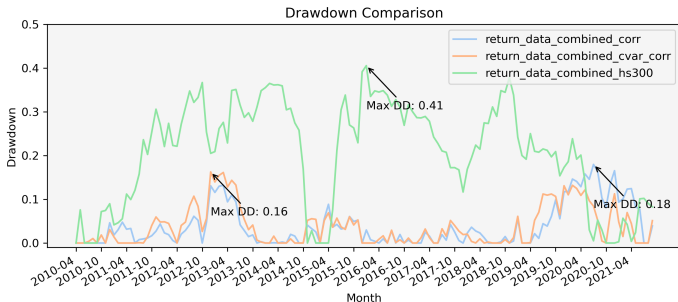


Figure 9: Drawdown Comparison Corr

策略对比

- CVaR 策略基本优于 MV 策略，使用 Corr 方法选股 +CVaR 策略最优
- 10 年期平均年化收益率 15%

策略	最大回撤	夏普比率	信息比率	年化收益率	收益回撤比
Coef	29.901%	1.272	-0.556	3.450%	11.538%
CVaR_coef	26.713%	1.510	-0.309	4.867%	18.218%
lc	28.526%	1.345	-0.526	3.739%	13.108%
CVaR_ic	42.263%	1.395	-0.484	4.025%	9.524%
Corr	17.969%	2.884	0.396	9.346%	52.009%
CVaR_Corr	16.723%	3.256	1.149	15.199%	90.889%
hs300	40.558%	1.336	0.000	6.845%	16.877%

Table 2: Strategy Comparison

小结

- ① CVaR 策略具有一定可行性
 - 投资组合的权重决策，基本优于 MV 最小方差策略
 - CVaR 对损失风险更加敏感，投资组合的波动性更小
 - 一定程度避免了人为模型设定误差和参数估计误差
 - 同时获取超额收益的能力具有一定保障
- ② CVaR 策略的有效性受到诸多条件的影响
 - 数据集划分、轮动周期
 - 因子的数量及有效性
 - 选股方法及选股数量

不足

❶ 数据量小

- 只筛选出了沪深 300 的股票，整体的数据量较低
- 需要较多的历史数据才能拟合的比较精准
- 但历史数据的有效样本略微不足
- 月度数据，策略实效性较低

❷ 历史数据选取时段缺乏标准 (40 个月)

- 过去时段过短，样本量少，反映尾部分布特征的样本较少
- 过去时段过长，样本量过多，收益率分布函数可能变化，无法反映先期收益率的分布特征

❸ 没有考虑投资者观点

- 只考虑了风险指标
- 缺少对资产预期收益率的观点
- 只适合极端风险厌恶的投资者
- 理想的模型应该是权衡收益与风险的

针对 CVaR 策略的改进方向

① 改进目标函数的估计方法

- $F_{\alpha}(w, \lambda) = \lambda + \frac{1}{1-\alpha} \int_{R \in R^N} [L(w, R) - \lambda]^+ P(R) dR$
- 目标函数中包含极大值函数，是非光滑凸规划问题
- 考虑使用非参数核估计方法 [5]

② 改进约束条件

- 考虑交易成本
- 加入投资者观点，如定价误差参数、预期收益率等 $w^T R \geq \pi$
- 梯度计算公式 + 逼近束求解 [2]
- 假设一种分布，如非对称-Laplace 分布 [4]

③ 改进代码实现方法

- 尝试更多 ML/DNN 方法选股及预测
- 改进滚动及调仓周期
- 改进因子配比方法
- 使用更高频数据

因子权重

① 三种方法对应的因子权重

Coef		IC		Corr	
Factors	Weights	Factors	Weights	Factors	Weights
turn_6m	0.11	EP	0.07	EP	0.18
exp_wgt_return_12m	0.06	EPcut	0.06	EPcut	0.14
turn_12m	0.05	ROE_ttm	0.06	BP	0.10
bias_turn_6m	0.04	Profit_G_q	0.06	SP	0.09
exp_wgt_return_1m	0.03	ROE_q	0.05	NCFP	0.07
turn_3m	0.03	OCFP	0.05	OCFP	0.07
exp_wgt_return_6m	0.03	ROE_G_q	0.04	DP	0.06
return_1m	0.03	SP	0.04	G/PE	0.06
std_6m	0.03	DP	0.04	Sales_G_q	0.06
std_1m	0.03	ROA_ttm	0.03	Profit_G_q	0.05

Table 3: Weights of Factors

参考文献 I



Yilin Ma, Ruizhu Han, and Weizhong Wang.

Prediction based portfolio optimization models using deep neural networks.

IEEE Access, 8:115393–115405, 2020.



Li Jia Tong, Jie Shen, and Na Xu.

An infeasible incremental bundle method for nonsmooth optimization problem based on cvar portfolio.

Complexity, 2021.



荣喜民 and 夏江山.

基于 cvar 约束的指数组合优化模型及实证分析.

数理统计与管理, 26(4), 2007.

参考文献 II



黄金波, 吴莉莉, and 尤亦玲.
非对称 laplace 分布下的均值-var 模型.
中国管理科学, 30(5):31–40, 2022.



黄金波, 李仲飞, and 姚海祥.
基于 cvar 核估计量的风险管理.
管理科学学报, 17(3):49–59, 2014.

Thank You!