人文社科

基于簇信息的分布式鲁棒优化的投资组合研究

清华大学 大学生学术研究推进论坛 Tsinghua Student Research Conference

作者: 王天宇 经管学院 指导教师: 王纯 经管学院 邮箱:wangty.17@sem.tsinghua.edu.cn

关键词：鲁棒优化 投资组合 条件风险价值

问题背景

最小化条件风险价值(CVaR)*是现代**最优投资组合**选取的重要准则，但实际问题仅从历史收益数据中直接估计均值-协方差优化**样本外表现风险较高**。分布式鲁棒优化(DRO)即从数据中最差可能的参数分布出发，期望获得更好的样本外表现。但是这样基于**最差分布**选择的方法，结果过于保守。**本文通过数据驱动下的聚类方法得到基于均值协方差的收益簇信息，由此优化得到的解能够减少原投资组合的保守程度，并在实际数据集上具有较高的可执行性和兼容性。**

模型分析与策略比较

符号表示

x 为决策权重(非负且和为1)；
r 为N种资产收益的随机变量；
s 为不同情形, (μ_s, Σ_s) 为相应的矩信息。

模型提出 簇信息易于从历史数据中计算获取(如K-Means)，形成如下的**簇分布集**:

$$\mathcal{F}(\mu,\Sigma)=\left\{\mathbb{P}\in\mathcal{P}_0\left(\mathbb{R}^{I_r}\times[S]\right)\;\middle|\;\begin{array}{ll}(\tilde{r},\tilde{s})\sim\mathbb{P} & \\ \mathbb{E}_{\mathbb{P}}[\tilde{r_s}]=\mu_s & \forall s\in[S]\\ \mathbb{E}_{\mathbb{P}}\left[(\tilde{r}-\mu_s)(\tilde{r}-\mu_s)^{\top}\right]=\Sigma_s & \forall s\in[S]\\ \mathbb{P}[\tilde{s}=s]=p_s & \forall s\in[S]\end{array}\right\}$$

因为聚类方法在高维数据表现较差， 又选择**Fama-French三因子模型**数据，作为**边际信息**聚类。

模型目标 通过使**最差分布下的F-CVaR最小化**实现最优投资组合，等价于:

$$\inf_{x,v}\left\{v+\frac{1}{\epsilon}\sup_{\mathbb{P}\in\mathbb{F}}\mathbb{E}_{\mathbb{P}}\left[(-\tilde{r}'x-v)^{+}\right]\right\}$$

而使用Popescu(2007)等类似方法，可类似地将其转化为如下的目标函数:

$$\inf_{x,v}\left\{v+\frac{1}{2\epsilon}\sum_{s=1}^{K}p_s(-\mu_s'x-v+\sqrt{x'\Sigma_s x+(\mu_s'x+v)^2})\right\}$$

而该目标**减少**了原问题(Popescu)的**保守程度**，并可化为**多项式时间**的**锥优化问题**，保证了**计算可执行性**。

基准策略 其他与之比较的基准策略和相应的参数有:

Method	Target	Parameter
1/N Policy	$x=(\frac{1}{N},\frac{1}{N},\dots,\frac{1}{N})^T$	NA
Markowitz	$\sup_{x}\{\mu'x-\frac{\gamma}{2}x'\Sigma x\}$	$\gamma=0.5$
CVaR (SAA)	$\inf_{x,v}\left\{v+\frac{1}{\epsilon}\frac{1}{M}\sum_{i\in[M]}(-r_i'x-v)^{+}\right\}$	$\epsilon=0.05$
F-CVaR (Popescu)	$\inf_{x,v}\left\{v+\frac{1}{2\epsilon}[(-\mu'x-v)+\sqrt{x'\Sigma x+(\mu'x+v)^2}]\right\}$	$\epsilon=0.05$

实证结果

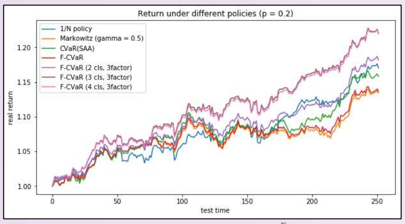
Method	Sharpe ratio	VaR	CVaR
1/N Policy	0.1460	0.5560	0.0481
Markowitz (0.5)	0.1226	0.5614	0.0397
CVaR (SAA)	0.1454	0.6198	0.0397
F-CVaR (Popescu)	0.1304	0.6112	0.0392
F-CVaR (2 cls, 3 factor)	0.1773	0.5306	0.0416
F-CVaR (3 cls, 3 factor)	0.1935	0.4891	0.0461
F-CVaR (4 cls, 3 factor)	0.1953	0.4799	0.0460
F-CVaR (2 cls, return)	0.1692	0.5263	0.0409
F-CVaR (3 cls, return)	0.1589	0.5571	0.0453
F-CVaR (4 cls, return)	0.1531	0.5639	0.0473

(其他年份和种类的数据集获得了相似的实验结果)

数据集*
10 Industry
2012 - 2017
方法
5天滚动窗口
准则
Sharpe ratio*
VaR* / CVaR

在与基准策略的比较中，**基于簇信息尤其是通过三因子数据的方法**表现更好。

研究结论



(p = 0.2 表示每次加入调整的罚因子，即在目标乘以 $1-p\sum_{j=1}^{N}|x_{j,t+1}-x_{j,t}|$ 这一系数)

问题延伸 考虑**交易成本**后，基于簇信息的策略仍然在测试集有较好的收益表现。

研究结论 相比其他DRO方法，**基于簇信息**的方法不仅实证**表现更好易于计算**，而且**聚类方法可作为与多种边际信息**(如CAPM, 4-因子模型)**相兼容的框架。**

分布式鲁棒优化 —收益信息 / 边际信息(因子/市场...)→ 簇信息方法

附录说明

*VaR (风险价值), 定义为在一定概率水平下某投资组合的最大可能损失; CVaR(条件风险价值), 为投资组合的损失超过给定VaR值的条件下，该投资组合损失的期望。本文分析中选取概率水平为ε = 0.05。
*Sharpe ratio (夏普比率), 定义为某投资组合的预期超额收益率与该组合的标准差的比值。
*本工作主要为作者在交换期间，在新加坡国立大学Melvyn Sim教授和清华大学王纯教授的指导下完成。数据集来源为Kenneth R. French – Data Library website. 仿真程序及结果由Python 3.7和Gurobi 8.1.0的接口给出。