基于簇信息的分布式鲁棒优化的投资组合研究

Robust Stochastic Portfolio Optimization: A Data-driven Clustering Approach

王天宇

清华大学 经管学院

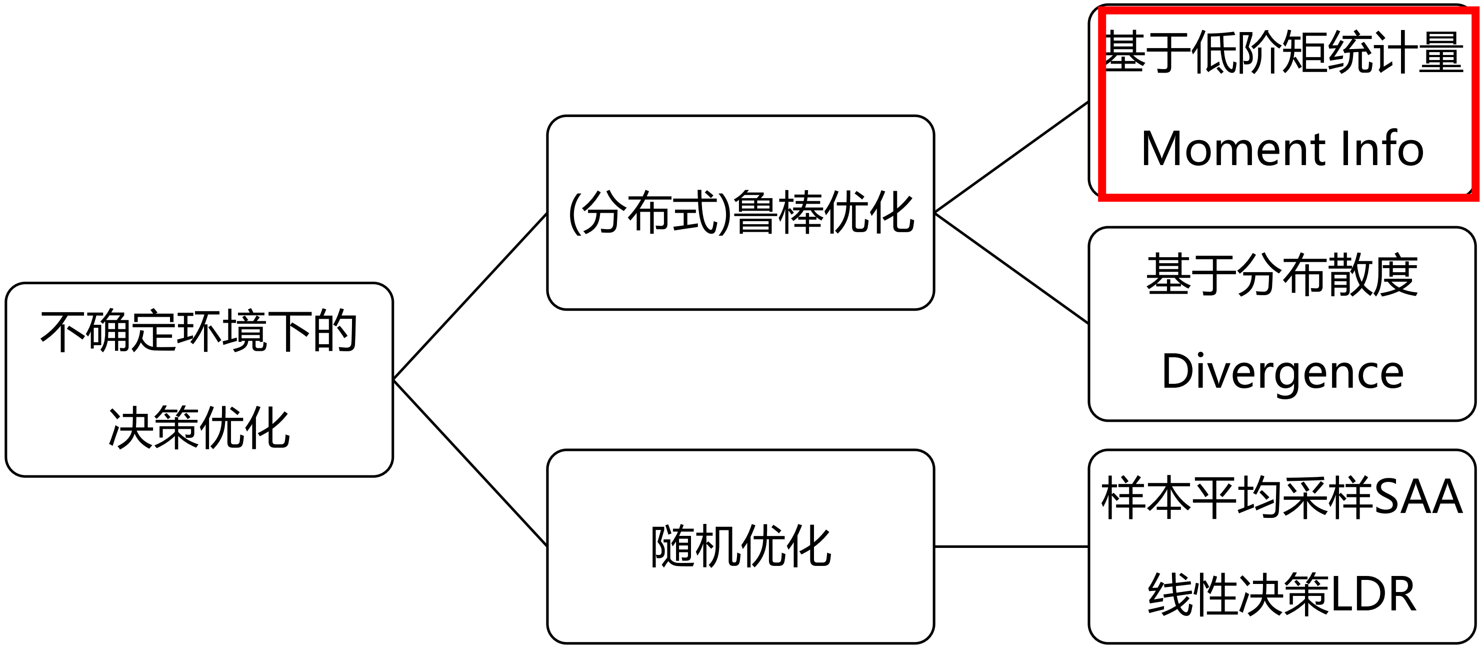
2019.12.14

整体框架

- 问题背景：数据不确定决策 / 投资风险概念 / 策略模型
- 模型分析：模型提出 / 理论分析 / 实际应用
- 实验方法：对比策略 / 数据集实验
- 实验结果：实验结果 / 扩展方法 / 鲁棒性检验
- 结论扩展：结果阐释 / 未来方向

问题背景

- 不确定环境下的决策



- 鲁棒优化与随机优化的区别：

➢ 是否给定先验的分布进行参数决策

➢ 鲁棒优化在决策中更加稳健，但是通常结果过于保守。

投资组合的选取

整体思路：最大化收益 / 最小化风险

- 收益评价
  - Markowitz (1952)：提出定量分析投资组合 $\sup_x \left\{ \mu' x - \frac{\gamma}{2} x' \Sigma x \right\}$
  - DeMiguel et al. (2009): 14种不同策略在实际数据集的表现比较

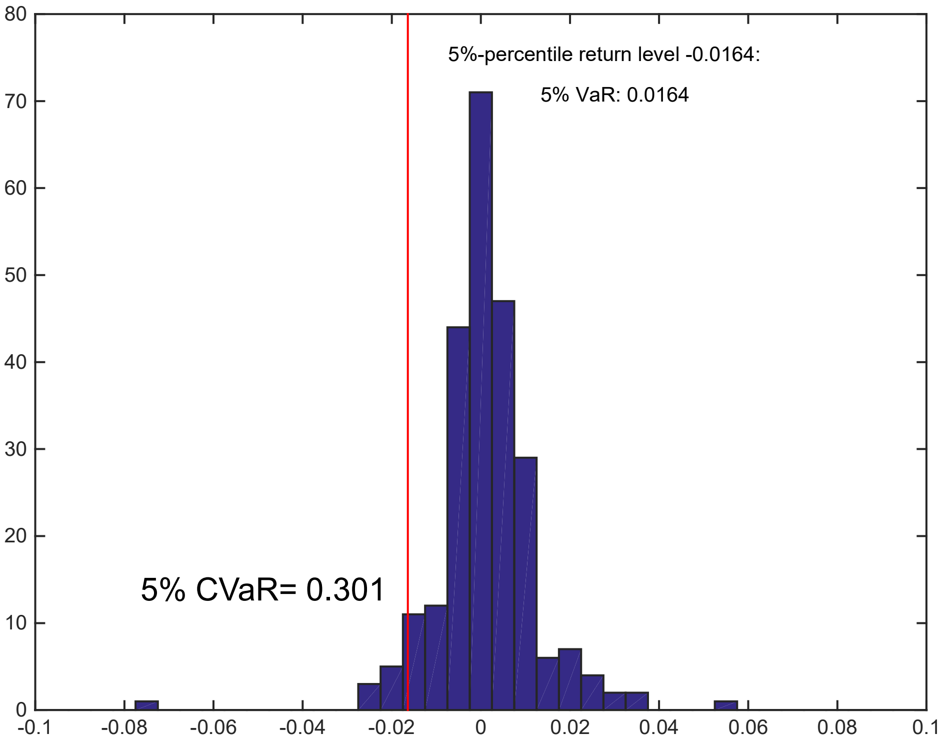
- 风险评价
  - 风险价值 (VaR):

$$\mathrm{VaR}_\epsilon[\tilde{r}] \triangleq \inf\left\{ v \in \mathbb{R} \mid \mathbb{P}\left[\tilde{r} + v \geq 0\right] \geq 1 - \epsilon \right\}$$

  - 条件风险价值 (CVaR):

$$\mathrm{CVaR}^*_\epsilon[\tilde{r}] \triangleq \mathbb{E}_{\mathbb{P}}\left[-\tilde{r} \mid -\tilde{r} \geq \mathrm{VaR}_\epsilon(\tilde{r})\right]$$

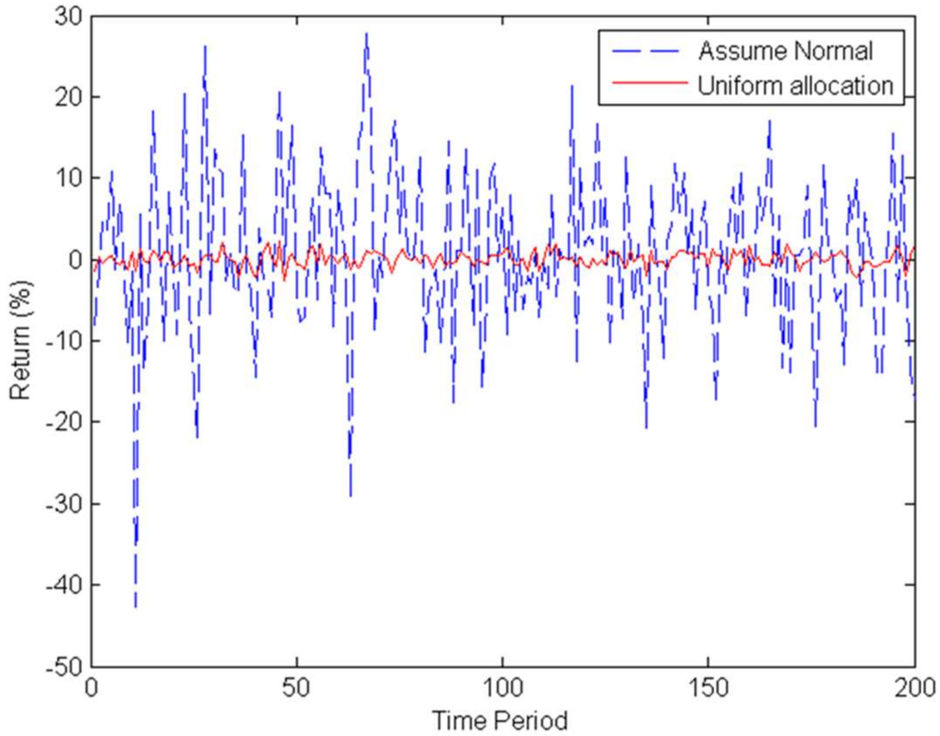
$$\mathrm{CVaR}^*_\epsilon[\tilde{r}] \geq \mathrm{VaR}_\epsilon(\tilde{r})$$



投资组合的选取

- 直接进行如随机采样(SAA)或线性决策(LDR)进行优化

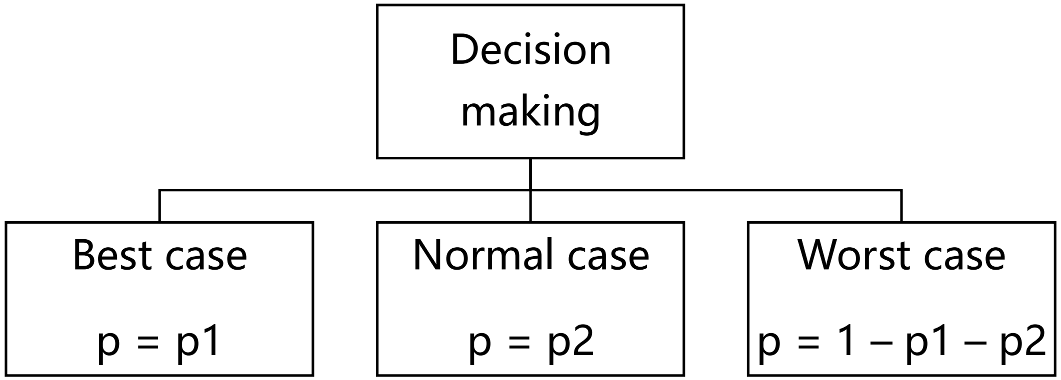
➢ 结果波动性较大 (i.e. 过拟合)



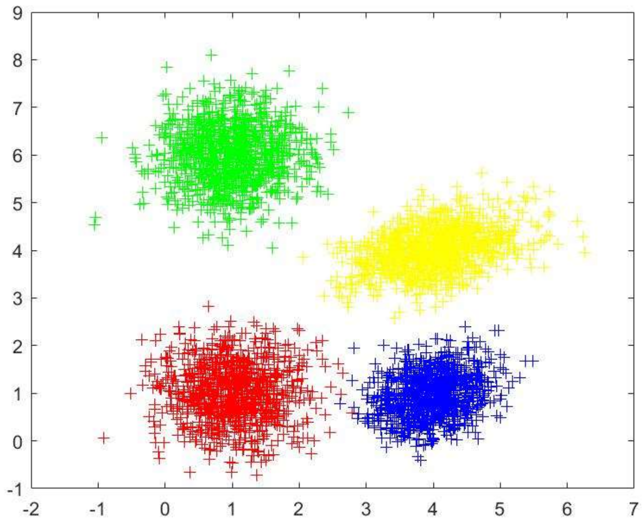
思考延伸

- Chen et al. (2019): 将分布式鲁棒优化的思想应用到**多种情形。**

➤实际生活的决策树
不同可能情形下的期望进行决策



➤统计的聚类方法:
如K-Means, K-medoids



问题形成

- 符号表示 Notations

➢ 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 为不同簇，(μ_s, Σ_s)为相应资产的均值-协方差矩(估计)信息。

- 基于簇信息的**模糊分布集 Ambiguity Set**

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

问题转化

- 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)和Chen et al. (2019)类似方法，可将其转化为如下的目标函数:

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

- 而该目标可化为多项式时间的二阶锥优化(SOCP)问题，保证了计算可执行性。

簇的聚类方法

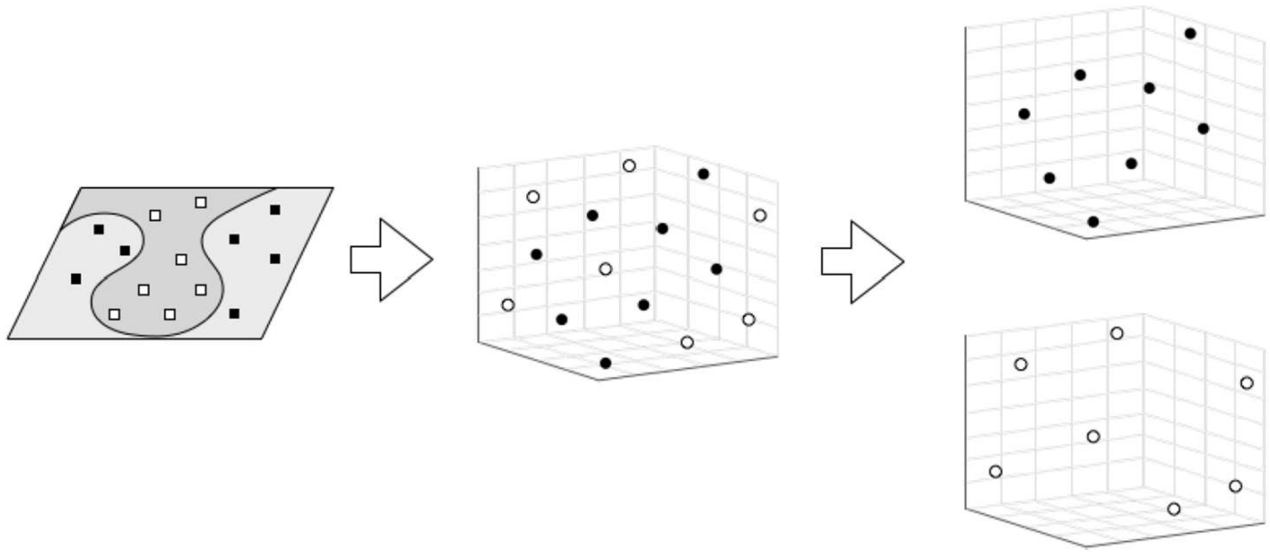
- 基于资产**历史收益** naively clustered by return

- 基于**因子信息**：边际信息 clustered by side information

➢ Fama-French 3-factor model (Fama and French, 1992)

$$E(R_{it}) - R_{ft} = \beta_i[E(R_{mt}) - R_{ft}] + s_i^E(SMB_t) + h_i^E(HML_t)$$

➢ Fama-French 5-factor model (Fama and French, 2013)

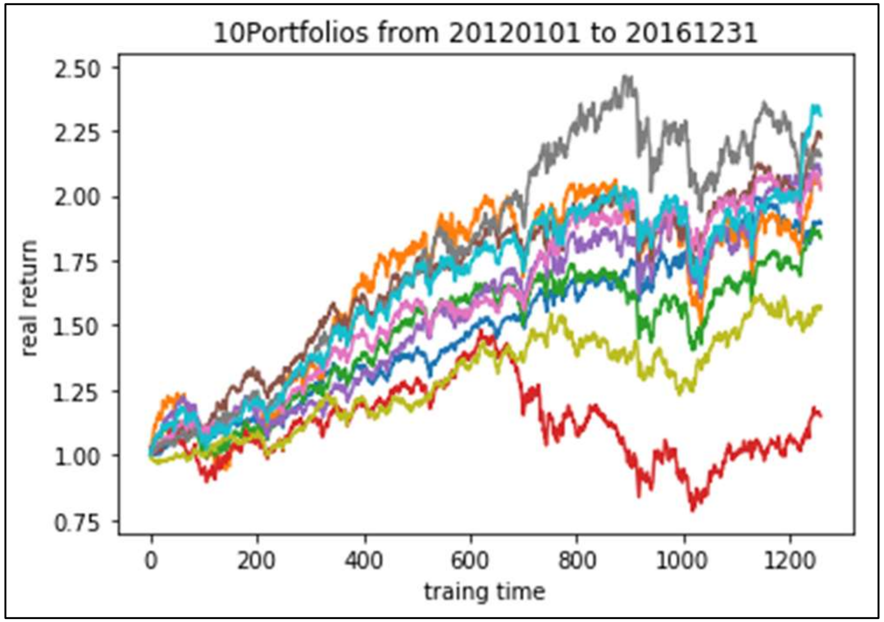


数据集与测试方法

- Base case : 10_Industry_Portfolios_Daily.csv (Daily)

➢训练集(in-sample)： 2012.01.01 – 2016.12.31 (1258 days)

➢测试集(out-of-sample)：2017.01.01 – 2017.12.31 (251 days)



Train set (for calculating moment information)
Rolling_window = 5 yrs

Rebalance period = 5 ds

timeline

- 滚动窗口测试方法 (rolling_window = 5 years)
- 平衡周期 5 trading days

基准策略与评价标准

- 基准策略 benchmark

Method	Target	Parameter
1/N Policy	$x = (\frac{1}{N}, \frac{1}{N}, ..., \frac{1}{N})^T$	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)	$\inf_{x,v} \left\{ v + \frac{1}{2\epsilon} [(-\mu' x - v) + \sqrt{x' \Sigma x + (\mu' x + v)^2}] \right\}$	$\epsilon = 0.05$

- 评价标准：实际测试集的收益

➢ 夏普比率(Sharpe ratio)

$$SharpeRatio = \frac{\mathbb{E}(R_p) - R_f}{\sigma_p}$$

➢ 风险价值(VaR)

$$\mathrm{VaR}_\epsilon[\tilde{r}] \triangleq \inf \{ v \in \mathbb{R} \mid \mathbb{P}[\tilde{r} + v \geq 0] \geq 1 - \epsilon \}$$

➢ 条件风险价值(CVaR)

$$\mathrm{CVaR}_\epsilon^*[\tilde{r}] \triangleq \mathbb{E}_\mathbb{P}[-\tilde{r} \mid -\tilde{r} \geq \mathrm{VaR}_\epsilon(\tilde{r})]$$

实验结果base case

- 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。

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

问题延伸：考虑交易成本

• 为防止平衡时投资组合过于频繁，加入交易成本对收益影响考虑。

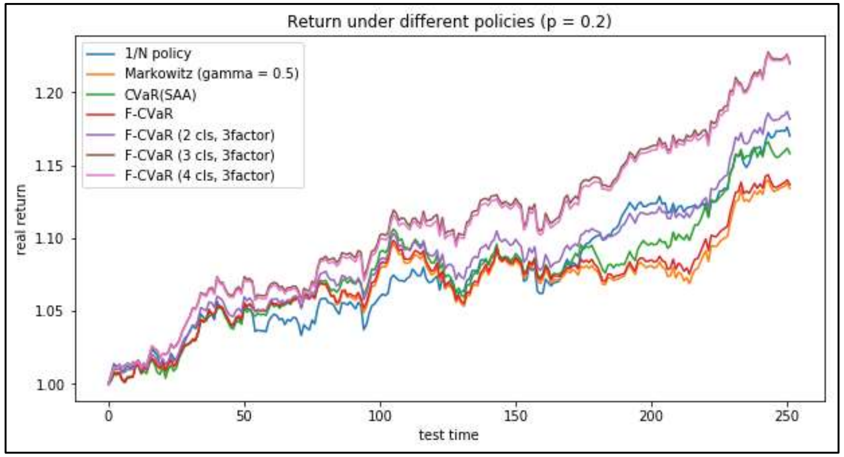
➢ 以F-CVaR (Popescu)方法为例：

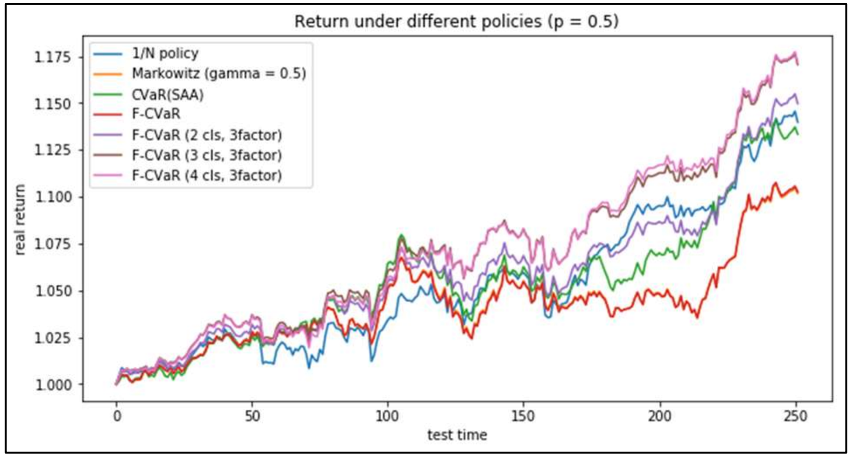
$$target = \inf_{x,v}\left\{v+\frac{1}{\epsilon}\sup_{\mathbb{P}\in\mathbb{F}}\mathbb{E}_{\mathbb{P}}\left[\left(-\tilde{r}'x-p\sum_{j=1}^{N}|x_j-x_j'|-v\right)^+\right]\right\}$$

$$=\inf_{x,v}\left\{v+\frac{1}{2\epsilon}\left[\left(-\mu'x-\boxed{p\sum_{j=1}^{N}|x_j-x_j'|}-v\right)+\sqrt{x'\Sigma x+\left(\mu'x+\boxed{p\sum_{j=1}^{N}|x_j-x_j'|}+v\right)^2}\right]\right\}$$

$$=\inf_{x,v}\left\{v+\frac{1}{2\epsilon}\left[\left(-\mu'x-p\sum_{j=1}^{N}t_j-v\right)+\sqrt{x'\Sigma x+\left(\mu'x+p\sum_{j=1}^{N}t_j+v\right)^2}\right]\right\}$$

(where x_j' is given before, $x_j-x_j'\le t_j$, $x_j'-x_j\le t_j$.)





➢ 即使加入交易成本后，簇信息(尤其是三因子方法)依旧优势显著

鲁棒性检验

- 在其他数据集的实证结果和多种参数的选择中，簇信息的方法均取得了较好的表现。

Method	Sharpe ratio	VaR	CVaR
1/N Policy	-0.0039	1.5202	0.1124
Markowitz (0.5)	0.0168	1.3779	0.0975
CVaR (SAA)	0.0088	1.3981	0.0992
F-CVaR (Popescu)	0.0118	1.4150	0.0979
F-CVaR (2 cls, 3 factor)	0.0226	1.3188	0.0975
F-CVaR (3 cls, 3 factor)	0.0196	1.3565	0.1022
F-CVaR (4 cls, 3 factor)	0.0159	1.3930	0.1052
F-CVaR (2 cls, 5 factor)	0.0229	1.3222	0.0976
F-CVaR (3 cls, 5 factor)	0.0189	1.3637	0.1027
F-CVaR (4 cls, 5 factor)	0.0128	1.4389	0.1057
F-CVaR (2 cls, return)	0.0206	1.3419	0.0968
F-CVaR (3 cls, return)	0.0154	1.3485	0.1016
F-CVaR (4 cls, return)	0.0055	1.4342	0.1052

调整训练和测试年份至10-14-15

其他不变

Method	Sharpe ratio	VaR	CVaR
1/N Policy	0.1454	0.7178	0.0556
Markowitz (0.5)	0.1559	0.5215	0.0376
CVaR (SAA)	0.1609	0.5678	0.0378
F-CVaR (Popescu)	0.1447	0.5501	0.0377
F-CVaR (2 cls, 3 factor)	0.1832	0.4872	0.0405
F-CVaR (3 cls, 3 factor)	0.1871	0.5031	0.0467
F-CVaR (4 cls, 3 factor)	0.1868	0.4927	0.0462
F-CVaR (2 cls, 5 factor)	0.1832	0.4701	0.0403
F-CVaR (3 cls, 5 factor)	0.1810	0.5152	0.0473
F-CVaR (4 cls, 5 factor)	0.1805	0.4911	0.0465
F-CVaR (2 cls, return)	0.1740	0.5519	0.0412
F-CVaR (3 cls, return)	0.1526	0.6329	0.0485
F-CVaR (4 cls, return)	0.1508	0.6843	0.0515

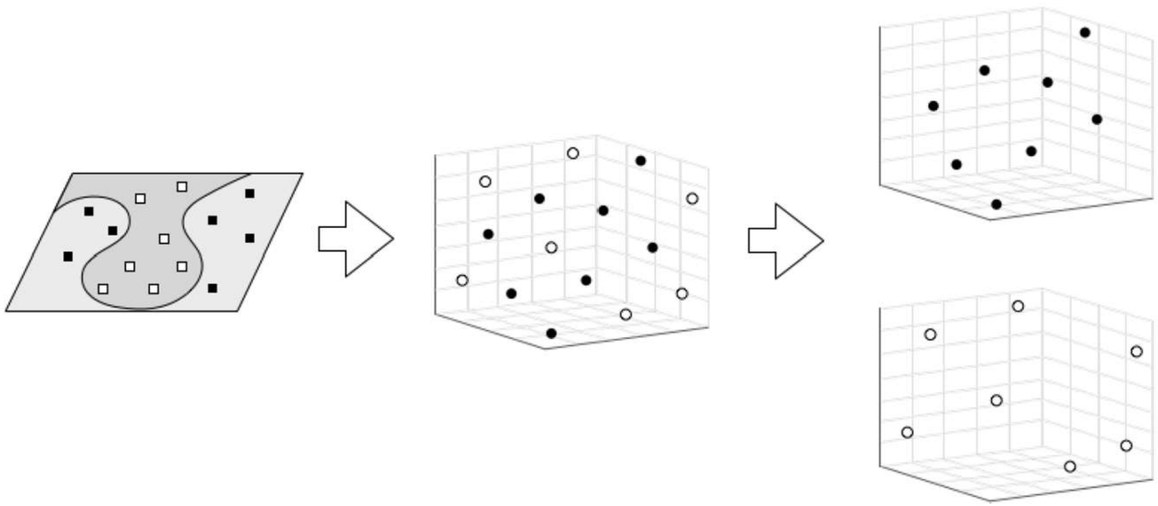
更换数据集为17_Industry_Portfolio

其他不变

结果阐释

因子信息方法的优势

- 克服原来直接通过收益作为**高维**数据的问题：10,17,30 → 3,5；
- **减少**资产收益**非系统波动**，反映市场信息；
- 作为边际信息的**接口，更加具有扩展性**。如之后4-因子，5-因子等其他数据驱动进行簇分类的方法。



方法阐释

基于簇矩信息的DRO方法在**理论**和**实证**上都取得了较好的结果

- 理论的**计算优势**

➢ 计算可执行性 computational tractability 二阶**锥优化**问题 (SOCP)

➢ 在分布式鲁棒优化的保守和随机优化的高收益更好**平衡**

分布式鲁棒优化

收益信息

边际信息(因子/市场...)

簇信息方法

- 实证的**数据兼容性**

➢ 相比其他DRO方法，该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

未来方向

存在不足	未来方向
两阶段调整对模型**假设较多**，未考虑forward-looking	**多阶段调整**的分布式鲁棒优化收益组合模型
实际数据资产收益风险**波动较大**，仍存在误差	统计参数假设分布，**自己生成样本外数据**，更好控制

DRO簇信息的聚类方法在运营管理决策**(供应链，收益管理，服务系统)**其他领域的应用，如 Perakis et al. (2019)等

敬请批评指正！

王天宇

清华大学 经管学院

2019.12.14

说明与补充

- 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. http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html
- 仿真程序及结果由Python 3.7和Gurobi 8.1.0的接口给出。
- 在p8上即使将CVaR作为约束条件进行自适应鲁棒优化，如

$$
\begin{gathered}
\min \epsilon \\
\text{s.t.} 0 \leq \epsilon \leq 1, x \in \mathbb{X} \\
\inf_{x,v} \left\{ v + \frac{1}{\epsilon} \sup_{\mathbb{P} \in \mathbb{F}} \mathbb{E}_{\mathbb{P}} \left[(-\tilde{r}'x - v)^{+} \right] \right\} \leq \tau
\end{gathered}
$$

簇信息方法变换得到的表达式在代数变形后，仍可通过分式线性规划(LFP)和根式的标准方法转化为SOCP多项式时间的优化问题。

附录

- Popescu (2007) 对于基于单均值-协方差的分布式鲁棒优化给出的最差上界：

$$\sup_{\mathbb{P}\in\mathbb{F}} \mathbb{E}_{\mathbb{P}}\left[(-\tilde{r}'x - v)^{+}\right] = \frac{1}{2}[(-\mu'x - v) + \sqrt{x'\Sigma x + (\mu'x + v)^2}]$$

- 多阶段的资产组合优化模型的模糊集框架(可由robust counterpart转化为SDP问题):

$$\mathcal{F} = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^{2I_u+2I_v} \times [S]\right) \;\middle|\; \begin{array}{ll} ((\tilde{\boldsymbol{u}}^1, \tilde{\boldsymbol{u}}^2, \tilde{\boldsymbol{v}}^1, \tilde{\boldsymbol{v}}^2), \tilde{s}) \sim \mathbb{P} & \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{u}}^1 \mid \tilde{s} \in \mathcal{E}_k\right] = \hat{\boldsymbol{\mu}}_k^1 & \forall k \in [K_1] \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{v}}^1 \mid \tilde{s} \in \mathcal{E}_k\right] \leq \hat{\boldsymbol{\sigma}}_k^1 & \forall k \in [K_1] \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{u}}^2 \mid \tilde{s} = s\right] = \hat{\boldsymbol{\mu}}_s^2 & \forall s \in [S] \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{v}}^2 \mid \tilde{s} = s\right] \leq \hat{\boldsymbol{\sigma}}_s^2 & \forall s \in [S] \\ \mathbb{P}\left[(\tilde{\boldsymbol{u}}^1, \tilde{\boldsymbol{u}}^2, \tilde{\boldsymbol{v}}^1, \tilde{\boldsymbol{v}}^2) \in \mathcal{Z}_s \mid \tilde{s} = s\right] = 1 & \forall s \in [S] \\ \mathbb{P}\left[\tilde{s} = s\right] = p_s & \forall s \in [S] \end{array} \right\}$$

参考文献

- Chen, Zhi, Melvyn Sim, Peng Xiong. 2019. Robust Stochastic Optimization. *Working Paper* .
- DeMiguel, Victor, Lorenzo Garlappi, Raman Uppal. 2007. Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The review of Financial studies* 22(5) 1915–1953.
- Fama, Eugene F, Kenneth R French. 1992. The cross-section of expected stock returns. *the Journal of Finance* 47(2) 427–465.
- Eugene F. Fama, & Kenneth R. French. (2014). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1).
- Markowitz, H. M. (1952). Portfolio Selection. Journal of Finance 7:77–91.
- Perakis, Georgia, Melvyn Sim, Qinshen Tang, Peng Xiong. 2018. Joint pricing and production: a fusion of machine learning and robust optimization. Available at SSRN.
- Popescu, Ioana. 2007. Robust mean-covariance solutions for stochastic optimization. *Operations Research* 55(1) 98–112.

参考文献

- Rockafellar, R Tyrrell, Stanislav Uryasev. 2002. Conditional value-at-risk for general loss distributions. Journal of banking & finance 26(7) 1443–1471.
- Shapiro, Alexander, Anton Kleywegt. 2002. Minimax analysis of stochastic problems. Optimization Methods and Software 17(3) 523–542.
- Sharpe, William F. 1966. Mutual fund performance. The Journal of business 39(1) 119–138.