銘柄数制約付き決定係数最大化ポートフォリオ構築問題の効率的解法

05000167 慶應義塾大学大学院理工学研究科 *田中 克弘 TANAKA Katsuhiro
02702020 慶應義塾大学理工学部 山本 零 YAMAMOTO Rei

1. 概要

決定係数最大化ポートフォリオ (MPP) は，資産運用モデルとして有効性が主張されている．その資産運用分野では取引コスト抑制の観点から銘柄数制約の解決は一つのテーマである．一方 MPP でそのような制約を組み入れたモデルは見かけない．

そこで，MPP に銘柄数制約を導入した問題の効率的解法ならびにその制約を課したことによる投資パフォーマンスの優位性を示す．

2. 問題設定

2.1. 決定係数最大化ポートフォリオ

$n \in \mathcal{N} := \{1,\dots,N\}$，$k \in \mathcal{K} := \{1,\dots,K\}$，$t \in \mathcal{T} := \{1,\dots,T\}$ とする．このときある資産の収益率 $R_{n,t}, (n \in \mathcal{N},\ t \in \mathcal{T})$ をマルチファクターモデルにより次のとおり記載する．

$$R_{n,t} = \beta_{n,0} + \sum_{k=1}^{K} \beta_{n,k} f_{k,t-1} + \varepsilon_{n,t-1}, (n \in \mathcal{N}, t \in \mathcal{T}).$$

ここで $f_{k,t-1}(k \in \mathcal{K},\ t \in \mathcal{T})$ はファクター，$\beta_{n,0}(n \in \mathcal{N})$ は切片，$\beta_{n,k}(k \in \mathcal{K}; n \in \mathcal{N})$ はファクターローディング，$\varepsilon_{n,t-1}(n \in \mathcal{N}; t \in \mathcal{T})$ は残差である．$\beta_{n,0}(n \in \mathcal{N})$，$\beta_{n,k}(n \in \mathcal{N}; k \in \mathcal{K})$，$\varepsilon_{n,t-1}(n \in \mathcal{N}; t \in \mathcal{T})$ は最小二乗法で計算できる．

また $R_{n,t}$，$\varepsilon_{n,t-1}(n \in \mathcal{N}; t \in \mathcal{T})$ を中心化した行列を $\boldsymbol{R}^\top := (R_{n,t} - \sum_{t=1}^{T} R_{n,t}/T)$，$\boldsymbol{E}^\top := \varepsilon_{n,t-1} \in \mathrm{R}^{\mathrm{N}\times\mathrm{T}}$ とする．この時収益率と残差の分散共分散行列 $\boldsymbol{P} := \frac{\boldsymbol{R}^\top \boldsymbol{R}}{T}$，$\boldsymbol{Q} := \frac{\boldsymbol{E}^\top \boldsymbol{E}}{T} \in \mathrm{R}^{\mathrm{N}\times\mathrm{N}}$ を定義する．

構成するポートフォリオ $\boldsymbol{x} \in \mathrm{R}^{\mathrm{N}}$ を導入することで，投資可能集合 $\mathcal{X} := \{\boldsymbol{x} \in \mathrm{R}^{\mathrm{N}} \mid \sum_{n=1}^{N} r_n x_n \geqq \rho, \sum_{n=1}^{N} x_n = 1, 0 \leqq x_n \leqq \alpha, (n \in \mathcal{N})\}$ を定義する．ここで ρ と α はユーザが定義するパラメータで，それぞれ期待収益率と投資比率上限である．

以上から，決定係数 $R^2(\boldsymbol{x}) := 1 - \boldsymbol{x}^\top \boldsymbol{Q} \boldsymbol{x} / \boldsymbol{x}^\top \boldsymbol{P} \boldsymbol{x}$ より MPP は以下の通り定式化できる:

$$\left|\begin{array}{ll} \min\limits_{\boldsymbol{x}} & \dfrac{\boldsymbol{x}^\top \boldsymbol{Q} \boldsymbol{x}}{\boldsymbol{x}^\top \boldsymbol{P} \boldsymbol{x}} \\ \text{s.t.} & \boldsymbol{x} \in \mathcal{X}. \end{array}\right. \qquad \text{(MPP)}$$

2.2. 銘柄数制約

この節では銘柄数制約を (MPP) に適用する．まず $\boldsymbol{z} \in \{0,1\}^{\mathrm{N}}$ を導入することで集合 $\mathcal{Z}$ を定義する．

$$\mathcal{Z} := \{\boldsymbol{z} \in \{0,1\}^{\mathrm{N}} \mid \sum_{n=1}^{N} z_n \leqq S\}.$$

S は銘柄数上限である．また $\boldsymbol{x}$ と $\boldsymbol{z}$ に関する SOS1 制約 (SOS1$(1-z_n, x_n)$) を設定する．

$$\begin{cases} 0 \leqq x_n \leqq \alpha, & \text{if } z_n = 1, \\ x_n = 0 & \text{otherwise,} \end{cases} \qquad (n \in \mathcal{N}).$$

これは Gurobi のような商用ソルバーで表記できる．

よって提案する問題は以下のとおりである:

$$\left|\begin{array}{ll} \min\limits_{\boldsymbol{x},\boldsymbol{z}} & \dfrac{\boldsymbol{x}^\top \boldsymbol{Q} \boldsymbol{x}}{\boldsymbol{x}^\top \boldsymbol{P} \boldsymbol{x}} \\ \text{s.t.} & \boldsymbol{x} \in \mathcal{X}, \boldsymbol{z} \in \mathcal{Z}, \\ & \text{SOS1}(1-z_n, x_n), (n \in \mathcal{N}). \end{array}\right. \qquad (\mathbb{MPP})$$

3. 解法

3.1. 既存解法

Gotoh and Fujisawa[1] が考案した手法を説明する．

$(\mathbb{MPP})$ の目的関数は $\frac{\boldsymbol{x}^\top \boldsymbol{E}^\top \boldsymbol{E} \boldsymbol{x}}{\boldsymbol{x}^\top \boldsymbol{R}^\top \boldsymbol{R} \boldsymbol{x}}$ に書き直せる．また $\eta := 1/\sqrt{\boldsymbol{x}^\top \boldsymbol{R}^\top \boldsymbol{R} \boldsymbol{x}} (> 0)$ とする．このとき $\boldsymbol{y} := \boldsymbol{x}\eta$ を定義することで, 集合 $\mathcal{Y} := \{(\boldsymbol{y}, \eta) \mid \sum_{n=1}^{N} r_n y_n \geqq \rho\eta, \sum_{n=1}^{N} y_n = \eta, 0 \leqq y_n \leqq \alpha\eta, (n \in \mathcal{N})\}$ を定義する．ここで $(\mathbb{MPP})$ を次のとおり書き直す:

$$\left|\begin{array}{ll} \min\limits_{\boldsymbol{y},\eta,\boldsymbol{z}} & \boldsymbol{y}^\top \boldsymbol{E}^\top \boldsymbol{E} \boldsymbol{y} \\ \text{s.t.} & \boldsymbol{y}^\top \boldsymbol{R}^\top \boldsymbol{R} \boldsymbol{y} = 1, (\boldsymbol{y}, \eta) \in \mathcal{Y}, \boldsymbol{z} \in \mathcal{Z}, \\ & \text{SOS1}(1-z_n, y_n), (n \in \mathcal{N}). \end{array}\right.$$

更にベクトル $\boldsymbol{u}, \boldsymbol{v} \in \mathrm{R}^{\mathrm{T}}$ を導入し書き直す:

$$\left|\begin{array}{ll} \min\limits_{\boldsymbol{y},\eta,\boldsymbol{z},\boldsymbol{u},\boldsymbol{v}} & \boldsymbol{v}^\top \boldsymbol{v} \\ \text{s.t.} & \boldsymbol{v} = \boldsymbol{E}\boldsymbol{y}, \boldsymbol{u} = \boldsymbol{R}\boldsymbol{y}, \boldsymbol{u}^\top \boldsymbol{u} = 1, \\ & (\boldsymbol{y}, \eta) \in \mathcal{Y}, \boldsymbol{z} \in \mathcal{Z}, \\ & \text{SOS1}(1-z_n, y_n), (n \in \mathcal{N}). \end{array}\right.$$

ここで非凸制約 $\boldsymbol{u}^\top \boldsymbol{u} = 1$ に対し，$\boldsymbol{u}^{k-1}$ を入力値とし $(\boldsymbol{u}^\top)\boldsymbol{u}^{k-1}$ とすることで，次の混合整数変数二次計画問題を書く:

$$\left|\begin{array}{ll} \min\limits_{\boldsymbol{y},\eta,\boldsymbol{z},\boldsymbol{u},\boldsymbol{v}} & \boldsymbol{v}^\top \boldsymbol{v} \\ \text{s.t.} & (\boldsymbol{u}^\top)\boldsymbol{u}^{k-1} = 1, \\ & \boldsymbol{v} = \boldsymbol{E}\boldsymbol{y}, \boldsymbol{u} = \boldsymbol{R}\boldsymbol{y}, \\ & (\boldsymbol{y}, \eta) \in \mathcal{Y}, \boldsymbol{z} \in \mathcal{Z}, \\ & \text{SOS1}(1-z_n, y_n), \\ & (n \in \{\mathcal{N}\}). \end{array}\right. \qquad (\mathbb{MPP}(\boldsymbol{u}^{k-1}))$$

これを繰り返し解き，有限回の反復で局所解に収束する．

以上より既存アルゴリズムは次のとおりである．

アルゴリズム 1: Normalized linearization (NL)

Step1 $k \leftarrow 1$. $\epsilon(> 0)$ を設定し，初期値 $(\boldsymbol{y}^0, \boldsymbol{u}^0, \boldsymbol{v}^0)$ を $(\boldsymbol{u}^0)^\top \boldsymbol{u}^0 = 1$ を満たすよう設定する．

Step2 $(\mathbb{MPP}(\boldsymbol{u}^{k-1}))$ を解き，$(\hat{\boldsymbol{y}}^k, \hat{\boldsymbol{u}}^k, \hat{\boldsymbol{v}}^k)$ を得る．

Step3 $(\boldsymbol{y}^k, \boldsymbol{u}^k, \boldsymbol{v}^k) \leftarrow \frac{(\hat{\boldsymbol{y}}^k, \hat{\boldsymbol{u}}^k, \hat{\boldsymbol{v}}^k)}{\sqrt{(\hat{\boldsymbol{u}}^k)^\top \hat{\boldsymbol{u}}^k}}$ と基準化する．もし $\sqrt{(\hat{\boldsymbol{u}}^k)^\top \hat{\boldsymbol{u}}^k} \geqq 1+\epsilon$ なら，$k \leftarrow k+1$ とし，Step2 へ．

ここで Step1 の初期値 $(y_n^0, (n \in \mathcal{N}))$ は Gotoh and Fujisawa[1] と同様，一律に $1/\sqrt{N}$ とする．

3.2. 二段階アプローチ

NL を離散変数を含む $(\mathbb{MPP})$ に適用する際，課題が二つある．それは 1.「得た局所解の質の悪さ」，2.「計算負荷」である．そこで解消するアプローチを考える．

3.2.1. 初期値の校正

$(\mathbb{MPP})$ は，問題の非凸性が増加するため，解の精度は初期値に大きく依存するものと想定する．初期値が $1/\sqrt{N}$ だと，全銘柄を保有することを意味しており，最適解から乖離すると予想される．そこで初期値は (MPP) から得た解とする．(MPP) と $(\mathbb{MPP})$ の違いは銘柄数制約の有無のみで，(MPP) は $(\mathbb{MPP})$ の緩和問題である．そのため双方の最適解は近しいので, 初期値として有効と予想される．

よって (MPP) と $(\mathbb{MPP})$ を続けて解く二段階アプローチを取る:

アルゴリズム 2: 二段階アプローチ (TS)

Step1 NL で (MPP) を解き，最適解は $(\boldsymbol{y}^*, \boldsymbol{u}^*, \boldsymbol{v}^*)$.

Step2 $(\boldsymbol{y}^0, \boldsymbol{u}^0, \boldsymbol{v}^0) \leftarrow (\boldsymbol{y}^*, \boldsymbol{u}^*, \boldsymbol{v}^*)$. NL で $(\mathbb{MPP}(\boldsymbol{u}^{k-1}))$ を反復し $(\mathbb{MPP})$ を解く．

3.2.2. 問題削減法

(MPP) の解 $y_n = 0$ となる n において，$(\mathbb{MPP})$ の解も $y_n = 0$ と置く．なぜなら (MPP) は $(\mathbb{MPP})$ の緩和問題なので，その最適解も近いことが予想される．また (MPP) の $y_n > 0, (n \in \mathcal{N})$ を満たす解の個数は，$(\mathbb{MPP})$ のそれ以上であることを前提とすると，(MPP) で $y_n = 0$ となるサンプル n では $(\mathbb{MPP})$ で $y_n > 0$ となる可能性が低いことが想定される．そこで，以下の制約式を付け加える．

$$\begin{cases} z_n = y_n = 0, & \text{if } y_n^0 = 0, \\ \mathrm{SOS1}(1 - z_n, y_n), & \text{otherwise}, \end{cases} \quad (n \in \mathcal{N}). \qquad (*)$$

以上から TS を改良したアプローチは次である:

アルゴリズム 3: 二段階アプローチ+ (TS_+)

Step1 TS の Step1 と同様.

Step2 $(\boldsymbol{y}^0, \boldsymbol{u}^0, \boldsymbol{v}^0) \leftarrow (\boldsymbol{y}^*, \boldsymbol{u}^*, \boldsymbol{v}^*)$. NL で以下の $(\mathbb{MPP}(\boldsymbol{u}^{k-1}, \boldsymbol{y}^0))$ を反復することで $(\mathbb{MPP})$ を解く．

$$\begin{array}{lll} \displaystyle\min_{\boldsymbol{y},\eta,\boldsymbol{z},\boldsymbol{u},\boldsymbol{v}} & \boldsymbol{v}^\top \boldsymbol{v} & \\ \text{s.t.} & (\boldsymbol{u}^\top)\boldsymbol{u}^{k-1} = 1, & \\ & \boldsymbol{v} = \boldsymbol{E}\boldsymbol{y}, \boldsymbol{u} = \boldsymbol{R}\boldsymbol{y}, & (\mathbb{MPP}(\boldsymbol{u}^{k-1}, \boldsymbol{y}^0)) \\ & (\boldsymbol{y}, \eta) \in \mathcal{Y}, \boldsymbol{z} \in \mathcal{Z}, & \\ & (*) & \end{array}$$

4. 数値結果

株価データは日経平均高配当株 50 指数と東証一部上場銘柄から 2008 年 11 月から 2021 年 12 月の月次データおよび日次リターンを獲得できた銘柄を用いた ($N = 45$，1500)．またファクターは 6 つ採用した．

計算機環境は Core i9-9980，32.0 GB，2.4. GHz でソルバーは Gurobi optimizer ver 9.5.2．最大計算時間は 3600 秒 (1 時間)，$\alpha = 1$，$\epsilon = 1.0\mathrm{e}^{-6}$，$\rho = \sum_{n=1}^{N}(r_n/N)$ とする．

$(\mathbb{MPP})$ は以下の 4 つのアプローチで計算する．

- NC: 非凸制約を持つ $(\mathbb{MPP})$ の形式のまま解く
- NL: アルゴリズム 1 を用いて解く
- TS: アルゴリズム 2 を用いて解く
- TS_+: アルゴリズム 3 を用いて解く

なお，非凸制約を含んだ NC は，Gurobi の `NonConvex` の機能を用いることで直接解くことができる．

表 1 は $N = 45$ で解けた回数を比較し，厳密解法の NC の解ける範囲は限定的であることが分かる。

表 1: $N = 45$ で 10 回中解けたインスタンス数

	NC	NL	TS	NC	NL	TS	NC	NL	TS
$S\backslash T$		24			36			48	
3	10	10	10	9	10	10	6	10	10
5	0	10	10	0	10	10	0	10	10
7	0	10	10	0	10	10	0	10	10
10	0	10	10	0	10	10	0	10	10

表 2 は $N = 1500$ で解けた回数を比較し，二段階アプローチは既存解法の NL より多く解けることが分かる．

表 2: $N = 1500$ で 10 回中解けたインスタンス数

	NL	TS	TS_+	NL	TS	TS_+	NL	TS	TS_+
$S\backslash T$		500			750			1000	
3	0	2	10	0	1	10	0	0	10
5	0	2	10	0	2	10	0	1	10
7	0	4	10	0	3	10	0	3	10
10	0	9	10	0	5	10	0	6	10

その他，決定係数の値や事後リターンは当日に発表する．

参考文献

[1] Gotoh, J. and Fujisawa, K. “Convex optimization approaches to maximally predictable portfolio selection”. *Optimization: A Journal of Mathematical Programming and Operations Research*, 63, 11, 1713-1735, (2014).