

特別研究報告書

偏相関行列が疎な構造を持つ
資産配分問題について

指導教員

福嶋 雅夫 教授

山下 信雄 助教授

京都大学工学部情報学科

数理工学コース

平成14年4月入学

平成18年3月卒業

關 雄至

平成18年1月31日提出

偏相関行列が疎な構造を持つ資産配分問題について

關 雄至

摘要

近年、経済のグローバル化に伴い株式や債券、通貨など数多くの資産に分散して投資することの重要性が増してきている。計算機に支援された分散投資では、リターンとリスクを定量的に表した数理モデルに基づいて、投資目的に最適な資産の配分割合（ポートフォリオ）を決定する。Markowitz の平均・分散モデルはリターンをポートフォリオの期待収益率で表し、リスクをその分散で表現した数理モデルであり、凸 2 次計画問題として定式化される。これまでに平均・分散モデルに対しては非常に多くの研究が行われ、その改良されたものが実際の投資に用いられている。しかし平均・分散モデルで用いる各資産の収益率に関する共分散行列は一般に密な行列になるため、投資対象の資産数が大きいときには計算容量、計算時間の観点からポートフォリオを求めることが困難となる。また過去の資産価格データから共分散行列を推定する場合、データ数が少ないときには共分散行列が正則にならないので平均・分散モデルは数値的に不安定になる。

本報告書では、偏相関行列 $H = [\rho_{ij}]$ （共分散行列 $V = [\sigma_{ij}]$ の逆行列を正規化した行列）が疎な構造を持つ平均・分散モデルを提案する。偏相関行列の (i, j) 成分 ρ_{ij} は資産 i と j 以外の収益率を固定したときの i と j の収益率の相関係数となり、資産 i と j の収益率が条件付き独立ならば、 $\rho_{ij} = 0$ となる。ここで条件付き独立にならない資産ペアの集合、つまり条件付き従属の集合を E とする。本報告書では $(i, j) \in E$ ならば $\sigma_{ij} = \tilde{\sigma}_{ij}$ であり、 $(i, j) \notin E$ ならば $\rho_{ij} = 0$ となる共分散行列 $V = [\sigma_{ij}]$ を用いるモデルを提案する。ここで $\tilde{\sigma}_{ij}$ は適当な手法で推定された共分散行列の (i, j) 成分である。この行列 V を求める問題は $\sigma_{ij} = \tilde{\sigma}_{ij}, (i, j) \in E$ のもとで行列式を最大にする正定値行列補完を求めることと等価であり、 E がある性質を持つときには容易に解くことができる。また E の要素が少ないときには H は疎な行列となるため、少ない計算容量でこのモデルを実装することができる。本報告書では条件付き従属の集合 E の構成方法をいくつか提案する。また行列 H の疎性を利用した計算容量、計算時間を少なくする実装の工夫についても言及する。さらに実際のヒストリカルデータを用いて数値実験を行い、提案したモデルでも従来の平均・分散モデルと同じようなポートフォリオが得られることを確認した。さらに、ヒストリカルデータの数が少なくても提案した手法で求めた共分散行列 V が数値的に安定であることが分かった。

目次

1	はじめに	1
2	準備	2
2.1	平均・分散モデル	2
2.2	条件付き独立	3
2.3	正定値行列補完	5
3	偏相関行列が疎な構造をもつ平均・分散モデル	6
3.1	Type1:ヒストリカルデータから求めた偏相関係数を用いるモデル	7
3.2	Type2:ヒストリカルデータから求めた相関係数を用いたモデル	7
3.3	Type3:クラス別モデル	7
3.4	Type4:クラス別モデルに偏相関係数が大きくなるクラス間の資産ペアを加えたモデル	8
3.5	Type5: k 重対角行列モデル	8
3.6	Type6:並び替え k 重対角行列モデル	8
4	疎性を利用したアルゴリズム	9
4.1	Goldfarb-Idnani 法	9
4.2	計算時間を減らす方法	10
5	数値実験	12
5.1	共分散行列の正則性の分析	12
5.2	ポートフォリオの分析	13
5.3	パフォーマンスの分析	14
6	まとめ	14

1 はじめに

近年、オンライン取引による手数料の低下や長引く低金利状態などの様々な理由から、株式などの危険資産への投資が機関投資家以外にも幅広く普及してきた。特に、外国債券、投資信託や投資型年金への個人マネーの流入が続いている。そのような大規模な資金を運用する場合には、収益（リターン）を増やすだけでなく、損をする可能性（リスク）を科学的にコントロールする必要がある。そのため、リスクとリターンを定量的に扱う数理モデルの重要性が増してきている。

これまでに様々な目的や用途に合わせた数理モデルが数多く提案されている。例えばシングルファクターモデル [8] や CAPM (Capital Asset Pricing Model) [9]、平均・分散モデル [6] などが有名である。その中でも平均・分散モデルはシンプルかつ拡張性の高さから多くの研究が行われている [5]。平均・分散モデルとは、リターンをポートフォリオの期待収益率で表し、リスクをその分散で表現したモデルである。「収益率の期待値が一定ならば、その分散は小さいほど望ましい」という基準に従うとき、平均・分散モデルは次の凸 2 次計画問題に定式化できる。

$$\begin{aligned} \min \quad & x^T V x \\ \text{s.t.} \quad & r^T x \geq \delta \\ & \sum_{j=1}^n x_j = 1 \\ & x_j \geq 0, j = 1, \dots, n \end{aligned} \tag{1.1}$$

ここで、 n は投資対象となる資産の数、 $x_i (i = 1, \dots, n)$ は資産 i への投資割合（ポートフォリオ）、 $r_i (i = 1, \dots, n)$ は資産 i の期待収益率、 $V = [\sigma_{ij}] \in R^{n \times n}$ は資産 i と j の収益率の共分散 σ_{ij} を要素とする共分散行列（分散共分散行列ともいう）である。また δ は投資家が要求するポートフォリオの期待収益率である。

投資対象として株式、債券、オプションなど金融派生商品など国内外に様々な金融商品が存在する。そのため、それら全てを組み合わせたポートフォリオを求める問題 (1.1) の次元（資産数 n ）は非常に大きくなる。このとき、以下の理由で問題 (1.1) を取り扱うことが困難になる。一般に共分散行列 V は密な行列となるため、その保存には $O(n^2)$ の計算容量（領域計算量ともいう）が必要になる。そのため資産数が数千となるような問題では、市販のコンピュータでは扱うことができない。また 2 次計画問題 (1.1) に対する解法である Goldfarb-Idnani 法 [3] や内点法は、 V が密なときには $O(n^3)$ 以上の計算時間がかかる。そのため n が大きいときは問題 (1.1) を解くコストが莫大なものになる。もう一つの欠点としては、共分散行列 V の推定の難しさがあげられる。一般によく使われる手法では、過去 T 期間の資産価格を用いて共分散行列を計算する (2.1 節参照)。この方法では、期間数 T が資産数 n に比べて小さいときには、共分散行列 V が正則でなくなる。そのため問題 (1.1) の求解が困難になるばかりでなく、共分散行列としての V の妥当性も疑わしくなる。

そのような欠点を克服するために、マルチファクターモデル [8] やアセットアロケーションモデル [5] が提案されている。マルチファクターモデルは、各資産の収益率を GNP や為替レートなどのいくつかのファクターとパラメータによって表現したモデルである。ファクターの数を K 個とすると、マルチファクターモデルは、0 でない要素の数が $n + K^2$ 個である行列 $\hat{V} \in R^{(n+K) \times (n+K)}$ を用いた 2 次計画問題に定式化でき、その問題を解くのに必要な計算容量や計算時間は問題 (1.1) に比べて大幅に少なくなる。しかし、適切なファクターの選定が難しく、パラメータの推定に膨大な時間がかかる。また、モデルが単純化されすぎているため、求めたポートフォリオのパフォーマンスが良くないことが報告されている。アセットアロケーションモデルでは、多数の資産をいくつかのクラスに分割し、第 1 段階で各クラスを代表するインデックスに関する平均・分散モデルを解いて各クラスに分配する割合を決めた後に、第 2 段階で各クラス内での平均・分散モデルを解いて全体のポートフォリオを求める。その結果、

各段階で解く 2 次計画問題は小規模なものになるので、全てを一度に扱う問題 (1.1) に比べて、計算容量と計算時間は少なくなる。しかし、クラスやインデックスの構成方法が難しく、上手に構成しないとパフォーマンスの高いポートフォリオを求めることができない。また、このモデルでは異なるクラスに含まれる資産の間の相関は考慮されていない。例えば、同じ企業系列に属する会社 α と β が異なるクラスに含まれている場合、実際には α と β の株価には相関があるはずであるが、この方法では α と β 固有の相関を考慮できない。

本報告書では、多くの資産ペアにおいてそれらの収益率が条件付き独立となることを仮定したモデルを提案する。ここで、資産ペア (i, j) の収益率が条件付き独立となるとは、「ある資産 i と j 以外の収益率を固定したときに資産 i と j の収益率が独立になること」である [7]。為替や GNP などマクロなファンダメンタルスの影響で、ほとんどの資産ペアの間では共分散が 0 とはならないが、条件付き独立となることが予想できる。偏相関行列 $H = (\rho_{ij})$ の成分 ρ_{ij} は、 (i, j) 以外の収益率を固定したときの資産 i と j の収益率の条件付き分布における (i, j) の相関係数である。従って、資産 i と資産 j の収益率が条件付き独立であれば、 $\rho_{ij} = 0$ となる。また、各資産の収益率が多次元正規分布に従うとすると、共分散行列 $V = (\sigma_{ij})$ の逆行列を正規化したものは偏相関行列となる。そのため、条件付き独立な資産ペアの数が多いとき、偏相関行列 (または V^{-1}) は疎な行列となる。しかし、過去の資産価格を用いて共分散行列を推定すると、誤差などの影響により、一般には偏相関行列は密な行列となる。そこで、資産の特徴 (属する業種など) から条件付き独立となる資産ペアをあらかじめ予測し、そのペアに対しては偏相関係数が 0 となる共分散行列 V を求めることを考える。ここで、条件付き独立にならない (条件付き従属) と予測される資産ペアの集合を E とする。本報告書では、 $(i, j) \in E$ ならば $\sigma_{ij} = \tilde{\sigma}_{ij}$ であり $(i, j) \notin E$ ならば $\rho_{ij} = 0$ となる共分散行列 $V = [\sigma_{ij}]$ (およびその偏相関行列 $H = [\rho_{ij}]$) を用いた平均・分散モデルを提案する。ここで $\tilde{\sigma}_{ij}$ は、適当な手法を用いて推定した i と j の共分散である。第 2.3 節で示すように、このような V を求めることは、 $\sigma_{ij} = \tilde{\sigma}_{ij}$ 、 $(i, j) \in E$ のもとで行列式を最大にする正定値行列補完を求めることと等価になり、 E がある性質を持つとき、容易に求めることができる。提案したモデルでは、偏相関行列 H が疎になる性質を利用して、少ない計算容量と計算時間で問題 (1.1) の解を得ることができる。さらに、数多くの資産ペアが条件付き独立な場合、少ないデータから得られた V でも正則となる。

本報告書は以下のように構成されている。まず第 2 節で本報告書で重要となる平均・分散モデル、条件付き独立、正定値行列補完について紹介する。第 3 節では偏相関行列が疎となるモデルを提案し、条件付き従属となる資産ペアの集合 E の構成方法をいくつか提案する。第 4 節では偏相関行列が疎になるときに計算容量、計算時間を減らす計算方法を示す。第 5 節で過去の株式データを用いた数値実験の結果を報告する。

2 準備

2.1 平均・分散モデル

本報告書では、資源配分問題のモデルとして平均・分散モデルを考える。対象とする資産数を n とし、 $x_i (i = 1, \dots, n)$ を資産 i への投資配分とする。ポートフォリオ x の収益率を $R(x)$ とする。ここで、 $R(x)$ は確率変数となることに注意する。ポートフォリオの収益率の期待値 $E[R(x)]$ を一定値 δ 以上にしつつ、分散 $V[R(x)]$ を小さくすることを考えると、平均・分散モ

デルは以下のように定式化できる.

$$\begin{aligned}
 \min \quad & V[R(x)] \\
 \text{s.t.} \quad & E[R(x)] \geq \delta \\
 & \sum_{j=1}^n x_j = 1 \\
 & x_j \geq 0, j = 1, \dots, n
 \end{aligned} \tag{2.1}$$

ここで, 資産 j の収益率 R_j がある確率分布に従う確率変数であるとし, その期待値を r_j , R_j と R_k の共分散を σ_{jk} とする. そのとき, $R(x)$ の平均と分散は次のようになる.

$$\begin{aligned}
 E[R(x)] &= E\left[\sum_{j=1}^n R_j x_j\right] = \sum_{j=1}^n r_j x_j \\
 V[R(x)] &= E[R(x) - E[R(x)]]^2 = E\left[\sum_{j=1}^n R_j x_j - \sum_{j=1}^n r_j x_j\right]^2 \\
 &= E\left[\sum_{j=1}^n \sum_{k=1}^n (R_j - r_j)(R_k - r_k) x_j x_k\right] \\
 &= \sum_{j=1}^n \sum_{k=1}^n \sigma_{jk} x_j x_k
 \end{aligned}$$

r_j および σ_{ij} は, 過去の資産価格 (ヒストリカルデータ) を用いて, 以下のように推定できる. 各資産 j に対して, 過去 T 期間の価格データが与えられたとする. ここで t 時点での資産 j の価格を $P_{jt} (j = 1, \dots, n, t = 1, \dots, T)$ とする. そのとき, t 時点での銘柄 j の収益率 $R_{jt} (t = 1, \dots, T-1)$ は

$$R_{jt} = (P_{jt+1} - P_{jt})/P_{jt} \tag{2.2}$$

となる. そのため, 収益率の期待値 r_j は

$$r_j = \frac{\sum_{t=1}^{T-1} R_{jt}}{T-1} \tag{2.3}$$

と推定できる. また, 共分散 σ_{ij} は

$$\sigma_{jk} = \sum_{t=1}^{T-1} \frac{1}{T-1} (R_{jt} - r_j)(R_{kt} - r_k) \tag{2.4}$$

で求めることができる. ここで, $V = [\sigma_{ij}]$ とすると, V は半正定値行列となる. しかし, $T \leq n$ のときは V は正定値行列とならないので, $x^T V x$ を目的関数とする問題 (2.1) を解くことは容易ではない.

2.2 条件付き独立

いま n 個の確率変数 R_1, R_2, \dots, R_n があるとする. R_i と R_j が条件付き独立であるとは, 「 R_i と R_j 以外の変数を固定したときに R_i と R_j が独立になること」である. ここで $R = (R_1, \dots, R_n)^T$ は多変量正規分布に従う n 次元確率ベクトルだとする. 多変量正規分布とは, 平均ベクトルと共分散行列により完全に規定される分布であり, R の確率密度関数は

$$f(R) = \frac{1}{\sqrt{(2\pi)^n}} \exp -\frac{1}{2}(R-r)^T V^{-1}(R-r)$$

となる. ここで $r = (r_1, \dots, r_n)^T \in R^n$ は R の平均ベクトル, $V = (\sigma_{ij}) \in R^{n \times n}$ は R の共分散行列である.

このとき, n 次元確率ベクトル $R = (R_1, R_2, \dots, R_n)$ をベクトル $R^{(1)} = (R_1, R_2)^T \in R^2$ とベクトル $R^{(2)} = (R_3, \dots, R_n)^T \in R^{n-2}$ に分解すると, R は

$$R = \begin{pmatrix} R^{(1)} \\ R^{(2)} \end{pmatrix}$$

と書ける. これに対応して平均ベクトル, 共分散行列を

$$r = \begin{pmatrix} r^{(1)} \\ r^{(2)} \end{pmatrix}, V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}$$

と分割する. ここで $r^{(1)} \in R^2$, $r^{(2)} \in R^{n-2}$, $V_{11} \in R^{2 \times 2}$, $V_{12} \in R^{2 \times (n-2)}$, $V_{21} \in R^{(n-2) \times 2}$, $V_{22} \in R^{(n-2) \times (n-2)}$ である.

このとき, $R^{(2)} = r^{(2)}$ を与えたもとの $R^{(1)}$ の条件付き分布は, 平均ベクトルが

$$r^{(1)} + V_{12}V_{22}^{-1}(R^{(2)} - r^{(2)})$$

共分散行列が

$$V_{11} - V_{12}V_{22}^{-1}V_{21} \quad (2.5)$$

の 2 次元正規分布に従う.

次に, $V = (\sigma_{ij})$ の逆行列を $Q := V^{-1} = (q_{ij})$ と記し, 上と同様に Q を

$$Q^{-1} = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix}$$

と分割する. すると

$$(Q_{11})^{-1} = V_{11} - V_{12}V_{22}^{-1}V_{21} \quad (2.6)$$

という関係を得る. これは (2.5) と一致する. 一方,

$$(Q_{11})^{-1} = \frac{1}{q_{11}q_{22} - q_{12}q_{21}} \begin{pmatrix} q_{22} & -q_{12} \\ -q_{21} & q_{11} \end{pmatrix} \quad (2.7)$$

である. (2.5) と (2.6) より, (2.7) が R_3, \dots, R_n を与えたときの R_1 と R_2 の条件付き分布の共分散行列となる. この行列は正規化した行列 (相関行列) を偏相関行列となり, その成分は偏相関係数となる. R_1 と R_2 の偏相関係数 ρ_{12} は,

$$\rho_{12} = \frac{\frac{-q_{12}}{q_{11}q_{22} - q_{12}q_{21}}}{\sqrt{\frac{q_{11}}{q_{11}q_{22} - q_{12}q_{21}}} \sqrt{\frac{q_{22}}{q_{11}q_{22} - q_{12}q_{21}}}} = \frac{-q_{12}}{\sqrt{q_{11}}\sqrt{q_{22}}}$$

である. 以上の議論は, 任意の $1 \leq i < j \leq n$ においても変数を適当に並び替えることで一般化できる. つまり, R_i と R_j の偏相関係数は

$$\rho_{ij} = \frac{-q_{ij}}{\sqrt{q_{ii}}\sqrt{q_{jj}}} \quad (2.8)$$

与えられる. したがって, R_i と R_j がその他変数を固定したときに条件付き独立になる必要十分条件は, 共分散行列 V の逆行列 V^{-1} の (i, j) 要素が 0 となること, すなわち偏相関係数 $\rho_{ij} = 0$ になることである.

2.3 正定値行列補完

条件付き従属の集合 E と、それに対応して $\tilde{\sigma}_{ij}, (i, j) \in E$ が与えられているとする。このとき本報告書では

$$\begin{aligned}\sigma_{ij} &= \tilde{\sigma}_{ij}, (i, j) \in E \\ q_{ij} &= 0, (i, j) \notin E\end{aligned}\tag{2.9}$$

を満たすような行列 $V = [\sigma_{ij}]$ を求める必要がある。ここで、 $V^{-1} = [q_{ij}]$ である。(2.9) の条件を満たす V を求めることは、行列式を最大化する正定値行列補完を求めることと密接な関係がある。ここで、 $\sigma_{ij} = \tilde{\sigma}_{ij}, (i, j) \in E$ となるような正定値行列 $V = [\sigma_{ij}]$ を、 $\tilde{\sigma}_{ij} \in E$ の正定値行列補完という。正定値行列補完に対しては、次の定理が知られている。

定理 1 $\tilde{\sigma}_{ij}, (i, j) \in E$ に対する正定値行列補完が存在するとする。このとき、次の最大化問題

$$\begin{aligned}\max \quad & \det(V) \\ \text{s.t.} \quad & \sigma_{ij} = \tilde{\sigma}_{ij}, (i, j) \in E \\ & V \succeq 0\end{aligned}\tag{2.10}$$

の解は唯一存在する。さらにその解 V の逆行列を $Q = [q_{ij}]$ とすると、 $q_{ij} = 0, (i, j) \notin E$ となる。

ヒストリカルデータを用いて推定された $\tilde{V} = [\tilde{\sigma}_{ij}]$ は一般には正定値行列となることから、定理 1 の仮定は満たされる。よって、この $\tilde{\sigma}_{ij}$ を用いた問題 (2.10) の解は (2.9) を満たすことがわかる。問題 (2.10) は半正定値計画問題であるため、一般には凸 2 次計画問題よりも難しい問題であるが、 E にコーダルグラフに関係した性質があれば、その解を陽に示すことが可能である。

ここで、その性質を紹介するために E に関係したグラフを定義する。頂点の集合を $\aleph = \{1, 2, \dots, n\}$ とし、辺の集合を \bar{E} とするグラフを $G(\aleph, \bar{E})$ とする。ただし、 $\bar{E} = E \setminus \{(k, k) | k = 1, \dots, n\}$ である。グラフに関して次の用語を定義する。

- 完全グラフ：任意の 2 つの頂点の間に辺を持つグラフ
- クリーク：あるグラフの部分グラフで完全なもの
- 極大クリーク：他のクリークの真の部分グラフにならないクリーク
- サイクル: 辺 $(k_1, k_2), (k_2, k_3), \dots, (k_{c-1}, k_c), (k_c, k_1)$ が E に含まれる頂点の組 (k_1, k_2, \dots, k_c) をサイクルの長さと呼ぶ
- コード (弦): サイクル中でつながっていない 2 点を結ぶ辺
- コーダルグラフ：長さ 4 以上のすべてのサイクルがコードを持つグラフ

$G(\aleph, \bar{E})$ がコーダルグラフならば、次の性質を満たすような極大クリーク族 $\{C_r | r = 1, \dots, l\}$ が存在する。

性質 1: $E = \bigcup_{r=1}^l C_r \times C_r$

性質 2: $r = 1, 2, \dots, l-1$ に対して $s \geq r+1: C_r \cap (C_{r+1} \cup C_{r+2} \cup \dots \cup C_l) \subset C_s$ を満たす。

この極大クリーク族は $O(n + |\bar{E}|)$ で求めることができる。次に二つの集合族 $\{S_r\}, U_r$ を以下のように定義する。

$$S_r = C_r \setminus (C_{r+1} \cup C_{r+2} \cup \dots \cup C_l), r = 1, \dots, l$$

$$U_r = C_r \cap (C_{r+1} \cup C_{r+2} \cup \dots \cup C_l), r = 1, \dots, l$$

ここで, $\aleph = \cup_r S_r$ かつ $S_i \cap S_j = \emptyset$ であることに注意する. S_1, S_2, \dots, S_l の順にそれぞれの要素を取り出して並べたときに, $1, 2, \dots, n$ となるように頂点の番号をつけかえる. そのつけかえの置換行列を P とする. このとき問題 (2.10) の解は

$$V = P^T L_1^T L_2^T \dots L_{l-1}^T D L_{l-1} \dots L_2 L_1 P \quad (2.11)$$

となる. ただし,

$$(L_r)_{ij} = \begin{cases} 1 & i = j \\ (\tilde{V}_{U_r U_r} \tilde{V}_{U_r S_r})_{ij} & (i, j) \in U_r \times S_r \\ 0 & \text{otherwise} \end{cases}$$

$$D = \begin{pmatrix} D_{S_1 S_1} & & & \\ & D_{S_2 S_2} & & \\ & & \ddots & \\ & & & D_{S_l S_l} \end{pmatrix}$$

$$D_{S_r S_r} = \begin{cases} \tilde{V}_{S_r S_r} - \tilde{V}_{S_r U_r} \tilde{V}_{U_r U_r}^{-1} \tilde{V}_{U_r S_r} & r \leq l-1 \\ \tilde{V}_{S_r S_r} & r = l \end{cases}$$

ここで, E の要素が少ないとき, U_r, S_r は小さくなることに注意すると, L_r, D は疎な行列となることが分かる.

この正定値行列補完を求めるためには, E に対して $G(\aleph, \bar{E})$ がコーダルグラフである必要がある. 任意の集合 E に辺を付け足して最小となるコーダルグラフを求める問題は NP 完全である. しかし, 最小次数法などの効率の良いヒューリスティックなアルゴリズムを用いれば, 高速にコーダルグラフとなる E の拡張を計算できる [1].

3 偏相関行列が疎な構造をもつ平均・分散モデル

適当な方法で得られた共分散行列 $\tilde{V} = [\tilde{\sigma}_{ij}]$ と, 条件付き従属になると予想される資産ペア (i, j) の集合を E が与えられているとする. このとき, 一般には $|E| \ll n^2$ となることが予想される. そこで,

$$\sigma_{ij} = \tilde{\sigma}_{ij}, (i, j) \in E$$

$$q_{ij} = 0, (i, j) \notin E$$

となる行列 $V = [\sigma_{ij}]$ と $V^{-1} = [q_{ij}]$ を共分散行列として用いることを提案する. この共分散行列を用いるメリットとしては以下のものがあげられる.

- () 条件付き独立の性質を考慮している.
- () V^{-1} が疎になるため, 計算容量が少なくなる.
- () \tilde{V} が正則でなくても, V は正則になりやすい.

本報告書で提案するポートフォリオを決定する手順は以下のとおりである.

手順 1: 適当な手法で収益率が条件付き従属となる資産ペアの集合 E を求める.

手順 2: グラフ $G(N, \bar{E})$ がコーダグラフとなるように E を拡張する .

手順 3: ヒストリカルデータから得られる共分散行列を $\tilde{V} = [\tilde{\sigma}_{ij}]$ とする . (2.11) 式によって

$$V = P^T L_1^T L_2^T \dots L_{l-1}^T D L_{l-1} \dots L_2 L_1 P$$

を求める .

手順 4: $V = (\sigma_{ij})$ を用いた凸 2 次計画問題 (1.1) を解いてポフォリオ x を求める .

手順 4 において, V^{-1} の疎性を利用して 2 次計画問題を効率よく解く方法は第 4 節で説明する . この節の残りでは, 提案モデルにおいて最も重要となる, 手順 1 における疎な集合 E の構成方法を提案する .

3.1 Type1: ヒストリカルデータから求めた偏相関係数を用いるモデル

まずヒストリカルデータから作る共分散行列を \tilde{V} とする . その \tilde{V} に対する偏相関係数を求める . そしてその値がある閾値 ϵ_1 以上となる資産ペア (i, j) は条件付き従属であると考え, 集合 E に加える . つまり, 全ての資産ペア (i, j) に対して

$$\rho_{ij} = \frac{-\tilde{q}_{ij}}{\sqrt{\tilde{q}_{ii}}\sqrt{\tilde{q}_{jj}}} \quad (3.1)$$

を計算し

$$E = \{(i, j) \mid |\rho_{ij}| \geq \epsilon_1\} \quad (3.2)$$

とする . ここで \tilde{q}_{ij} は \tilde{V} の逆行列 \tilde{V}^{-1} の (i, j) 成分である .

この方法では, ヒストリカルデータによって条件付き独立に近いと考えられるものを条件付き独立とみなすので, 得られる V はヒストリカルデータによる \tilde{V} に近くなると予想される .

しかし偏相関係行列を計算するためには V^{-1} の計算が必要となるため n が大きいときは $O(n^3)$ の計算時間がかかってしまう . また, \tilde{V} が正則でない場合には V を計算することができない .

3.2 Type2: ヒストリカルデータから求めた相関係数を用いたモデル

Type1 と違い, ヒストリカルデータから作る相関係数 $\tilde{\sigma}_{ij}$ の大きさによって E を求めるモデルである . 具体的には,

$$E = \{(i, j) \mid |\tilde{\sigma}_{ij}| \geq \epsilon_2\} \quad (3.3)$$

とする . ここで ϵ_2 は閾値である .

この方法では \tilde{V} が正則でない場合でも, V を求めることができる . 一方, 相関係数が大きい (小さい) 資産ペアでも, 条件付き独立 (従属) になることもある . そのため, 得られた V は現実の情報を反映していないことがある .

3.3 Type3: クラス別モデル

同一クラスに含まれる資産の間には相関があり, 違うクラスに含まれる資産の間には相関がないと考える . ここで, n 個の資産を L 個のクラス $J_k (k = 1, \dots, L)$ に分類する . そして

$$E = \{(i, j) \in J_k \times J_k \mid k = 1, \dots, L\} \quad (3.4)$$

とする．

このとき， $G(\mathbb{N}, E)$ はコーダルグラフになっている．従って，手順 2 においてコーダルグラフに拡張する必要がない．しかし，このモデルはアセットアロケーションモデルと本質的に同じであり，そのためアセットアロケーションモデルと同様の問題点を持つ．

3.4 Type4:クラス別モデルに偏相関係数が大きくなるクラス間の資産ペアを加えたモデル

Type3 で求めた集合 E に，相関の強いクラスの組み合わせに含まれる資産ペアを加えることを考える．まず過去のデータから各クラスの平均収益率を求める．さらにクラス間の偏相関係数 ρ'_{kh} を求め，その値がある閾値 ϵ_3 以上になるいくつかのクラスの組 $G = \{(k, h) \mid \rho'_{kh} > \epsilon_3\}$ を求める．そして

$$E = \{(i, j) \mid (i, j) \in J_k \times J_h, (k, h) \in G\} \quad (3.5)$$

とする．例えば，業種 J_1 と業種 J_2 の相関が大きいとき， $i \in J_1, j \in J_2$ となる資産ペア (i, j) も E に含まれる．

Type3 のモデルよりさらに多くの情報を含んでいるため，より現実に近い V が得られると期待できる．その分 Type3 のモデルより集合 E の要素数が増え，計算時間がかかる．また相関が非常に大きな資産ペアでもクラス間に全く関係がなければ無視してしまうことがある．

3.5 Type5: k 重対角行列モデル

$|i - j| \leq k$ となる資産ペア (i, j) が集合 E に含まれるようなグラフを考える．すなわち

$$E = \{(i, j) \mid |i - j| \leq k\} \quad (3.6)$$

とする．このとき，グラフ $G(\mathbb{N}, E)$ はコーダルグラフとなる．また V^{-1} は k 重対角行列となる．

E の構造がシンプルなので閾値の設定などの作業が必要ない．しかし， $|i - j|$ が大きい資産ペア (i, j) では，たとえ条件付き独立でなくても条件付き独立とみなされるため，情報が欠落することがあると予想される．

3.6 Type6:並び替え k 重対角行列モデル

同じクラスに含まれる資産が近くなるように並び替えた k 重対角行列モデル．まず，隣り合う資産が同じクラスに含まれるように資産を並び替える． $(1, 2, \dots, n)$ から $(1', 2', \dots, n')$ と並び替えたとして，

$$E = \{(i', j') \mid |i' - j'| \leq k\} \quad (3.7)$$

E は Type5 と同じ要素数になる．さらに，同じクラスに含まれる資産同士の $|i - j|$ が小さくなるため，同じクラス内であるにもかかわらず条件付き独立だとみなされる資産ペアの数が，Type5 よりも少なくなる．

ただし，最も相関が大きい資産どうしが隣り合うように並び替えたわけではないので，依然欠落する情報はある．

4 疎性を利用したアルゴリズム

ここではまず, 凸 2 次計画問題を解く一般的なアルゴリズムとして Goldfarb-Idnani 法 (GI 法) を紹介する [3]. その後, 逆行列 \tilde{V}^{-1} が疎になることを用いて計算量を減らす方法を与える.

4.1 Goldfarb-Idnani 法

次の 2 次計画問題 (QP) の解法である Goldfarb-Idnani 法を紹介する.

$$\begin{aligned} \min \quad & f(x) = \frac{1}{2}x^T V x + c^T x \\ \text{s.t.} \quad & a_i^T x \leq b_i, i = 1, \dots, m \end{aligned} \quad (4.1)$$

ここで $x \in R^n$, $V \in R^{n \times n}$, $c \in R^n$, $a_i \in R^n$, $b_i \in R$ である. V は正定値行列であるとする. さらに, k 回目の反復で, 等式が成り立つと考えている制約条件の添字集合を $W_k = \{k_1, \dots, k_{n_k}\}$ で表し, $N_k = [a_{k_1} \dots a_{k_{n_k}}]^T \in R^{n \times n_k}$, $b_k = [b_{k_1} \dots b_{k_{n_k}}]^T \in R^{n_k}$ と書くことにする.

Step0: $x_0 = -V^{-1}c$, $W_0 = \emptyset$, $n_0 = 0$ とする. $k = 0$ とする.

Step1: x_k が実行可能解ならば x_k が最適解なので終了. さもなければ Step2 へ.

Step2: x_k が満たしていない制約 $p \notin W_k$ を一つ選ぶ. $n_k = 0$ なら $\hat{u}_k = 0$, $n_k > 0$ なら $\hat{u}_k = \begin{pmatrix} u_k \\ 0 \end{pmatrix}$ とする.

Step3: $z_k, t_k^{(P)}$ を以下のように求める.

$$z_k = -V^{-1}(I - N_k N_k^*) a_p \quad (4.2)$$

ただし

$$N_k^* = (N_k^T V^{-1} N_k)^{-1} N_k^T V^{-1} \quad (4.3)$$

である.

$$t_k^{(P)} = \begin{cases} \frac{b_p - a_p^T x_k}{a_p^T z_k} & z_k \neq 0 \\ \infty & z_k = 0 \end{cases} \quad (4.4)$$

$r_k, t_k^{(D)}$ を以下のように求める.

$$r_k = -N_k^* a_p \quad (4.5)$$

$$t_k^{(D)} = \min \left\{ -\frac{(\hat{u}_k)_j}{(r_k)_j} \mid (r_k)_j < 0, j = 1, 2, \dots, n_k \right\} = -\frac{(\hat{u}_k)_l}{(r_k)_l} \quad (4.6)$$

ただし, このような l が存在しないときには $t_k^{(D)} = +\infty$ とする.

Step4: $t_k = \min(t_k^{(P)}, t_k^{(D)})$ とする. もし, $t_k = \infty$ なら QP に実行可能解がないので終了. そうでないとき, $u_k := u_k + t_k$ とおく. さらに $t_k^{(P)} = \infty$ のときは (a), そうでないときは (b) を実行

(a) 添字集合 W_k から l に対応する制約条件の番号を除いてあらためて W_k とおく. 行列 N_k , ベクトル b_k, u_k から第 l 番目の列ベクトルや要素を取り除き, それぞれあらためてさらに $n_k := n_k - 1$ とおいて Step3 へ

(b) まず $x_{k+1} = x_k + t_k z_k$ とおく. さらに $t_k^{(P)} \leq t_k^{(D)}$ ならば (b-1), $t_k^{(P)} > t_k^{(D)}$ ならば (b-2) を実行.

- (b-1) $t_k^{(P)} \leq t_k^{(D)}$ ならば, $W_{k+1} = W_k \cup \{p\}$, $u_{k+1} = \hat{u}_k$, $n_{k+1} = n_k + 1$ とおく.
 行列 N_k とベクトル b_k に第 p 番目の制約条件のベクトル a_p や右辺 b_p を追加し, それぞれ N_{k+1} , b_{k+1} とおく. $k = k + 1$ として Step1 へ.
- (b-2) $t_k^{(P)} > t_k^{(D)}$ ならば, $W_{k+1} = W_k \setminus \{l\}$ とおき, 行列 N_k とベクトル b_k , \hat{u}_k から第 l 番目のものを除いてそれぞれ N_{k+1} , b_{k+1} , \hat{u}_{k+1} とおく. $n_{k+1} = n_k - 1$ とおく. $k = k + 1$ として Step3 へいく.

制約に等式条件を含む場合は, 最初に等式制約を満たすように p を選び, その後 Step3 で等式制約を取り除かないようにすればよい.

4.2 計算時間を減らす方法

ここで, GI 法のアルゴリズムにおいて, V^{-1} の疎性を利用することで計算時間を減らすことができる工夫を示す.

アルゴリズムにおいて最も計算時間を必要とする箇所は Step3 において, z_k と r_k を求める部分である. z_k, r_k はそれぞれ, 式 (4.2), (4.3), (4.5) より

$$z_k = -V^{-1}(I - N_k N_k^*)a_p \quad (4.7)$$

$$r_k = -N_k^* a_p \quad (4.8)$$

で求まる.

V^{-1} の 0 でない要素数を s とする. 平均・分散モデルでは, $m = n + 2$ であり,

$$(a_1, \dots, a_m)^T = \begin{pmatrix} 1 & 1 & \dots & 1 & 1 \\ r_1 & r_2 & \dots & r_{n-1} & r_n \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix}$$

となる. そのため, N_k の 0 でない要素数は $O(n)$ である.

以下では r_k を求めるときの計算時間を見積もる. z_k においても同様に見積もることができる.

(4.8) 式は

$$r_k = -(N_k^T V^{-1} N_k)^{-1} N_k^T V^{-1} a_p \quad (4.9)$$

と書ける. ここで

$$N' := N_k^T V^{-1} N_k \quad (4.10)$$

$$v' := N_k^T V^{-1} a_p \quad (4.11)$$

とすると, $r_k = (N')^{-1} v'$ となる. 以下の手順で計算を行う.

手順 1: v' を計算する. まず $v_1 := V^{-1} a_p$ を計算し, $v' = N_k v_1$ を計算する.

手順 2: つぎに $(N')^{-1} v'$ を計算する. ここで $n \gg n_k$ のとき手順 2-1 を, $n \approx n_k$ のとき手順 2-2 を実行する.

手順 2-1: コレスキー分解を利用する.

手順 2-2 : 共役勾配法を利用する .

それぞれの手順における計算時間を見積もる .

手順 1 の計算時間は , $v_1 = V^{-1}a_p$ の計算にかかる計算時間は $O(s)$, $v' = N_k^T v_1$ の計算にかかる計算時間は $O(n)$ なので , $O(n+s)$ である .

手順 2-1 の計算時間は , $N_1 = V^{-1}N_k$ の計算にかかる計算時間は $O(sn_k)$, $N' = N_k N_1$ の計算にかかる計算時間は $O(nn_k^2)$, コレスキー分解を用いた $r_k = (N')^{-1}v'$ の計算時間は $O(n_k^3)$, 全て合わせると $O(sn_k + nn_k^2)$ である .

手順 2-2 の計算時間を見積もる前に , 共役勾配法を記述する . いま , 求めたいのは

$$r_k = (N_k^T V^{-1} N_k)^{-1} v'$$

なので , $N_k^T V^{-1} N_k r_k = v'$ となる r_k を求めることと同じである .

ここで , 共役勾配法は与えられた正定値対称行列 M とベクトル b に対して

$$\begin{aligned} \min \quad & f(x) = \frac{1}{2} x^T M x - b^T x \\ \text{s.t.} \quad & x \in R^n \end{aligned} \quad (4.12)$$

の最適解 x^* を求める方法で , 最適解においては $\nabla f(x^*) = Mx^* - b = 0$ を満たす . よって , $M = N_k^T V^{-1} N_k$, $b = v'$ とした (4.12) の最適解が r_k となる . 共役勾配法は以下のように書ける .

Step0: まず適当な初期解 x_0 を決めて , $k = 0$ とする .

Step1: x_k が最適解ならば終了 , さもなければ Step2 へ

Step2: 探索方向 d_k を求める . ただし ,

$$d_k = -\nabla f(x_k) + \frac{\nabla f(x_k)^T \nabla f(x_k)}{\nabla f(x_{k-1})^T \nabla f(x_{k-1})} d_{k-1} \quad (4.13)$$

である .

Step3:

$$t_k = -\frac{d_k^T \nabla f(x_k)}{d_k^T A d_k} \quad (4.14)$$

とし , さらに $x_{k+1} = x_k + t_k d_k$, $k = k + 1$ とする . Step1 に戻る .

Step2 の式 (4.13) において $\nabla f(x_{k-1})$ と d_{k-1} は $k-1$ 回目の反復で既に計算しているので , k 回目の反復で新たに必要なのは $\nabla f(x_k)$ の計算である . いま ,

$$\nabla f(x_k) = -v' + N_k^T V^{-1} N_k x_k$$

である . $N_k^T V^{-1} N_k x_k$ の計算は , $x_1 := N_k x_k$ の計算 , $x_2 := V^{-1} x_1$ の計算 , $x'_k = N_k x_2$ の計算の順に行う . $x_1 = N_k x_k$ の計算にかかる計算時間は $O(n)$, $x_2 = V^{-1} x_1$ の計算にかかる計算時間は $O(s)$, $x'_k = N_k x_2$ の計算にかかる計算時間は $O(n)$, よって , 一回の $\nabla f(x_k)$ の計算にかかる計算時間は $O(n+s)$ となる .

Step3 で最も計算時間が必要なのは式 (4.14) の分母の計算であるが , これは d_k の計算と同じ

要領で計算すればよい．したがって探索方向 d_k と同様にステップ幅 t_k も $O(n+s)$ の計算時間で求めることができる．共役勾配法は多くとも決定変数 x の次元よりも少ない反復回数で解を得ることができるので， $r_k = (N_k^T V^{-1} N_k)^{-1} v'$ を求めるにかかる計算時間は $O(n_k(n+s))$ となる．

以上のことから手順 2 の計算時間は $\min\{O(sn_k + nn_k^2), O(n_k(n+s))\}$ となる．また z_k に関しても同じ議論が展開できる．

V^{-1} が密な行列であるときには，これらの計算には $O(n^3)$ 以上の計算時間が必要であるので，これは， V^{-1} が疎な行列であるとき計算時間が大幅に削減できることを意味している．

5 数値実験

ここでは，提案した共分散行列推定法の有効性を確認するため，実際のデータを用いた数値実験の結果を報告する．ヒストリカルデータとして用いたデータは，以下のとおりである．

D_1 : 2003 年 1 月 6 日から 2005 年 12 月 30 日の東証 1 部企業 200 銘柄の株式価格日足データ
 D_2 : 2004 年 1 月 5 日から 2005 年 12 月 30 日の東証 1 部企業 200 銘柄の株式価格日足データ
 D_3 : 2005 年 1 月 4 日から 2005 年 12 月 30 日の東証 1 部企業 200 銘柄の株式価格日足データ
 D_4 : 2005 年 3 月 10 日から 2005 年 12 月 30 日の東証 1 部企業 200 銘柄の株式価格日足データ
 D_5 : 2003 年 1 月 6 日から 2005 年 12 月 30 日の東証 1 部企業 400 銘柄の株式価格日足データ
 D_6 : 2005 年 1 月 4 日から 2005 年 12 月 30 日の東証 1 部企業 400 銘柄の株式価格日足データ
このとき， D_1 と D_5 では $T = 736$ ， D_2 では $T = 491$ ， D_3 と D_6 では $T = 245$ ， D_4 では $T = 200$ である．

数値実験は全て，CPU が 2.53GHz の Pentium4，メモリーが 2GB の計算機上で，MATLAB6.5 を用いて行った．

5.1 共分散行列の正則性の分析

ヒストリカルデータから (2.4) 式を用いて推定した共分散行列 \tilde{V} と，Type1 ~ Type6 で推定した共分散行列 V の正則性を調べた．資産数 n が 200 である $D_1 \sim D_4$ に対して V の最大固有値，最小固有値，条件数，条件付従属な資産ペアの集合 E の密度 ($|E|/n^2$) を表 1 にまとめる．なお，Type1 と Type2 では集合 E の非ゼロ要素の密度が 5%前後になるように，閾値を調節した．また，手順 2 におけるコーダグラフへの拡張には最小次数法を用いた．

Type1 と Type2 に関しては， E の密度を 5%程度になるように閾値 ϵ_1 や ϵ_2 を調節するのが難しいという欠点があった．これは， E をコーダグラフへと拡張する際に，要素数がもとの E に比べ何倍にもなる場合があるためである．そのため，適当に決めた閾値では， V^{-1} が疎でなくなることがあった．

表 1 から，ヒストリカルデータから推定した行列 \tilde{V} はデータの期間数 T が資産数 n に近づくとつれて最小固有値が次第に小さくなり，正則ではなくなっていくことが読み取れる．一方，本報告書で提案した手法では Type1 を除き，全ての方法での条件数はデータ数 T が減少してもそれほど変化しなかった．このことから，提案した手法で推定する \tilde{V} は期間数 T が少なくなってもその影響が小さく，安定して共分散行列を推定できると言える．また，提案した手法で推定した共分散行列 V の条件数は \tilde{V} の条件数に比べて小さいことが分かった．一般に，条件数が小さいほど，凸 2 次計画問題 (1.1) は解きやすい．そのため，従来の手法で推定した共分散行列 \tilde{V} よりも，数値的には解を求めやすい．

表 1: 期間数 T に対する共分散行列の固有値と条件数の変化

データ		従来 \tilde{V}	Type1	Type2	Type3	Type4	Type5	Type6
D_1	最大固有値	275.42	155.86	244.83	49.214	56.456	55.342	53.047
	最小固有値	0.1471	0.2172	0.5237	0.2416	0.2398	0.2398	0.2417
	条件数	1873.1	717.34	467.43	203.62	233.58	230.71	219.45
	密度 [%]	100	7.9	1.8	4.6	6.8	3.5	3.5
D_2	最大固有値	215.86	168.46	137.17	32.367	45.089	48.904	44.586
	最小固有値	0.0977	0.1888	0.0424	0.2126	0.2126	0.2095	0.2098
	条件数	2208.2	891.96	322.84	152.21	212.02	233.37	212.51
	密度 [%]	100	16.3	4.5	4.6	6.8	3.5	3.5
D_3	最大固有値	165.1	57.655	122.79	22.253	35.081	37.112	34.654
	最小固有値	0.0074	0.1951	0.3421	0.1958	0.1958	0.1822	0.1869
	条件数	22050	295.49	358.89	113.63	179.13	203.67	185.33
	密度 [%]	100	3.8	7.3	4.6	6.8	3.5	3.5
D_4	最大固有値	148.08	*	93.578	18.876	31.981	29.583	27.323
	最小固有値	0	*	0.3187	0.1837	0.1837	0.1726	0.1771
	条件数	inf	*	293.57	102.75	174.09	171.37	154.24
	密度 [%]	100	*	4	4.6	6.8	3.5	3.5

5.2 ポートフォリオの分析

データ D_1, D_3 (銘柄数 200) と D_5, D_6 (銘柄数 400) を使って、各手法で推定した共分散行列を用いた問題 (1.1) を解いたポートフォリオ x を比較した。ただし、本実験では投資家の求める収益率 δ は、各データの期間の日経平均株価の平均収益率とした。

データ D_1, D_3 からポートフォリオを求めた結果を表 2 に、データ D_5, D_6 からポートフォリオを求めた結果を表 3 にまとめた。表 2, 3 には、各データの期間におけるポートフォリオ x の分散、 $T = 736$ (データ D_3 と D_6) のデータで推定した共分散行列 \tilde{V}_{736} によるポートフォリオ \tilde{x}^{736} と x の距離、各ポートフォリオで 0 とならない資産数を示す。ここで、距離とはポートフォリオ \tilde{x}^{736} と各モデルで求めたポートフォリオ x に対して $\sqrt{(\tilde{x}^{736} - x)^T \tilde{V}_{736} (\tilde{x}^{736} - x)}$ で計算されるものである。これは、従来の手法による共分散行列 \tilde{V}_{736} を用いたノルムではなかった距離である。なお、Type1 に関して、 $T < n$ となるときには式 (3.1) を計算することができないので実験は行わなかった。

表 2: 資産数 200 の場合

	データ	従来 \tilde{V}	Type1	Type2	Type3	Type4	Type5	Type6	日経平均
分散	D_1	0.3755	0.6262	0.6837	0.7269	0.659	0.4052	0.3965	1.4139
	D_3	0.3621	0.6565	0.6875	0.7368	0.6706	0.4281	0.4112	0.7336
距離	D_1	0	0.2027	0.2415	0.3239	0.2497	0.1751	0.1444	
	D_3	0.0851	0.3112	0.2586	0.3205	0.2536	0.1854	0.1539	
資産数	D_1	29	56	114	166	140	110	109	
	D_3	31	134	122	156	132	106	106	

表 3: 資産数 400 の場合

	データ	従来 of \tilde{V}	Type1	Type2	Type3	Type4	Type5	Type6	日経平均
分散	D_5	0.2169	0.4242	0.4464	0.4128	0.376	0.3992	0.3927	1.4139
	D_6	0.141	*	0.4302	0.4223	0.3885	0.4121	0.3989	0.7336
距離	D_5	0	0.1369	0.1208	0.0971	0.0821	0.1112	0.1026	
	D_6	0.1361	*	0.1048	0.0883	0.0786	0.0968	0.0916	
資産数	D_5	44	141	219	231	208	256	232	
	D_6	41	*	203	202	182	248	217	

提案したモデルで得られたポートフォリオの分散は、日経平均の分散よりは小さくなることを確認できた。ポートフォリオの分散、距離ともに資産数 n が増えるにつれて減少している。また、データ数 T が少なくなっても、提案した手法ではポートフォリオに大きな変化は見られなかった。表 3 より、 $T < n$ の場合 (D_6) では従来手法よりも \tilde{x}^{736} との距離は近くなった。このことから、提案したモデルでも従来手法で求めたポートフォリオに近い結果が得られ、特にデータ数が少ない時に有効であることが確かめられた。また、面白いことに、提案手法によるポートフォリオで選択される銘柄数は従来のモデルよりも多くなった。さらに、 E の要素の密度が高くなるにつれ、提案したモデルでの選択銘柄数は減少していき、従来のモデルの選択銘柄数に近づいていくという性質が見られた。

5.3 パフォーマンスの分析

データとして D_3 を用いる。まずはじめに 2003 年 1 月から 12 月のデータから共分散行列を推定し、ポートフォリオを求める。2004 年 1 月の初日にその通りに投資を行い、一ヶ月間保持する。ここで総資産は、2004 年 1 月の初日から 2004 年の 2 月の初日のデータによって増減する。次に 2003 年 2 月から 2004 年 1 月までのデータを用いて再度ポートフォリオを計算し、そのポートフォリオを再投資して一ヶ月間保持する。この作業を 2 年間繰り返す。このときの従来手法と Type6 のパフォーマンスを比較した結果を図 1 に示す。ただし、グラフの横軸は経過日数、縦軸は 2004 年 1 月 5 日を 1 としたポートフォリオの収益率である。

図 1 から、提案した手法で求めたポートフォリオの動きは、従来手法で求めたポートフォリオの動きと非常に近い動きとなることが分かった。

6 まとめ

偏相関行列の疎性を利用することで、従来の共分散行列よりも計算容量を抑える共分散行列の推定方法を提案した。また数値実験の結果、提案した手法から求められたポートフォリオが、ヒストリカルデータから推定される \tilde{V} を用いて求めたポートフォリオに近くなることを確認できた。さらに、提案した手法では、データ数が少なくても推定された共分散行列 \tilde{V} が数値的に安定となることを確認できた。

今後の課題として、計算時間の比較を行う必要がある。また、提案した手法における集合 E の要素数と、推定される共分散行列 V やポートフォリオの選択銘柄数との関係を調べる必要がある。

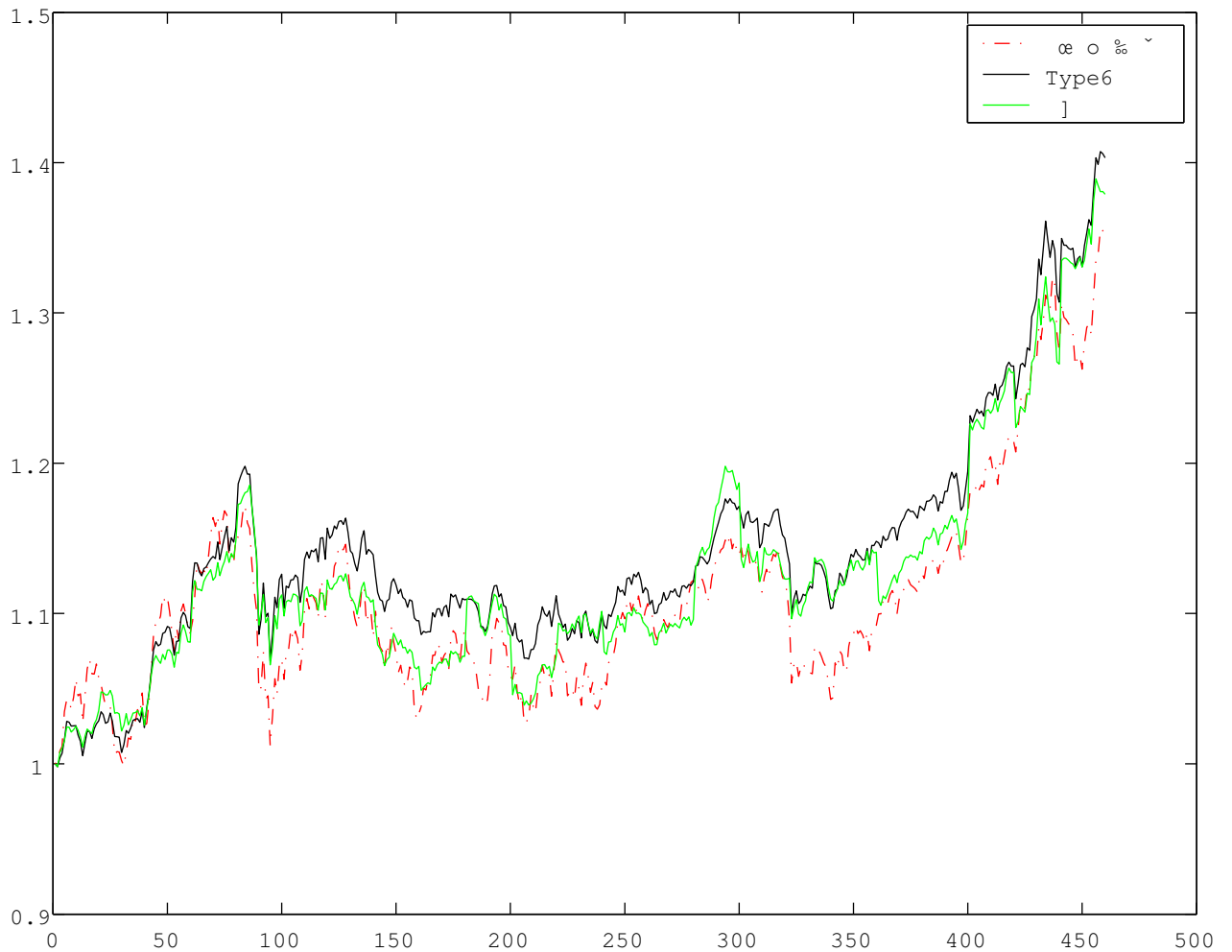


図 1: 収益率

謝辞

日頃からご教授くださり，本研究に対しても熱心なご指導を賜った福嶋雅夫教授，ならびに本報告書の作成にあたって細部に至るまで様々ご指摘と適切にご指導を賜った山下信雄助教授に心より感謝致します．また，日頃からお世話になっている林俊介特任助手をはじめ，福嶋研究室の皆様に厚く御礼申し上げます．

参考文献

- [1] M. Fukuda, M. Kojima, K. Murota, and K. Nakata, Exploiting sparsity in semidefinite programming via matrix completion I: General framework, SIAM Journal on Optimization, 11, pp. 647-674, 2000.
- [2] 福嶋雅夫, 非線形最適化の基礎, 朝倉書店, 2001.
- [3] D. Goldfarb, A. Idnani, A numerically stable dual method for solving strictly convex quadratic programs, Mathematical Programming 27, 1-33, 1983.

- [4] R. Grone, C. R. Johnson, E. M. Sa, and H. Wolkowicz, Positive definite completions of partial hermitian matrices, *Linear Algebra Appl.*, 58, 109-124, 1984.
- [5] 今野浩, 理財工学, 日科技連, 1995.
- [6] H. Markowitz, *Portfolio selection: Efficient diversification of investments*, John Wiley and Sons, New York, 1959
- [7] 宮川雅巳, *グラフィカルモデリング*, 朝倉書店, 1997
- [8] A. Peroid, Large scale portfolio optimization, *Management Science*, 30, 1143-1160, 1984.
- [9] W. F. Sharpe, Capital asset process : A theory of market equilibrium under condition of risk, *J. of Finance*, 19, 425-442, 1964.
- [10] W. F. Sharpe, A simplified model for portfolio anylysis, *Management Science*, 9, 277-293, 1963.
- [11] 矢部博, 八巻直一, *非線形計画法*, 朝倉書店, 1999.
- [12] N. Yamashita, Sparse quasi-Newton updates with positive definite matrix completion, Technical Report 2005-008, *Applied Mathematics and Phisics*, Kyoto University, 2005.