ファクター・リスク寄与度に基づく
ポートフォリオの構築について

東京理科大学工学部経営工学科　黒木 裕鷹
東京理科大学工学部経営工学科　小坪 琢人
東京理科大学工学部経営工学科　永井 直哉

平成 29 年 7 月 29 日

目 次

第 1 章　はじめに **1**

1.1　Bloomberg 投資コンテスト 1
1.2　価格変動の源泉 2
1.3　資本資産価格モデル (CAPM) 2
1.4　戦略の決定 3
1.5　レポート構成 3

第 2 章　ファクターモデルに基づくポートフォリオ選択 **4**

2.1　シングル・ファクターモデルとしての CAPM 4
2.2　CAPM に対する批判 5
2.3　マルチ・ファクターモデル 6
2.4　APT 6
2.4.1　仮定 6
2.4.2　APT の主定理 7
2.4.3　ファクターのリスク・プレミアムの推定 7

第 3 章　データ解析 **9**

3.1　使用したデータ 9
3.2　データのクリーニング・概観 10
3.3　ファクターモデルの推定 14
3.4　リスク・プレミアム 15
3.5　様々なポートフォリオ 16
3.6　ポートフォリオ選択 17

第 4 章　途中経過 **19**

4.1　提出したポートフォリオ 19
4.2　7 月中のパフォーマンス 19
4.3　リバランス 21

第 5 章　今後の課題 **22**

概 要

運用期間が数カ月程度の短期間で投資パフォーマンスを得るために，近年盛んに研究されているマルチ・ファクターモデルを用いたポートフォリオを構築した．Fama-French の 3 ファクターモデルに代表されるマルチ・ファクターモデルには，ファクターの選択方法や学習データに対して過学習してしまう問題が指摘．このように，適切なファクター選択やファクター間の相関がもたらす交絡などの影響は，マルチ・ファクターモデルの重要な課題である．

株価の、時系列データという性質上，たに観測されるデータは常に将来のものであり，これはいわゆる外挿の一種である．そのため，時系列に関するモデリングではモデルの複雑化にともなうオーバーフィッティングの影響を非常に受けやすい．この問題を解決するため，モデル推定では近年パターン認識の分野でしきりに利用されている lasso 回帰を用いた．また，共通要因が株価に与える影響は時系列的に変動するため，最適なモデルの選択には交差検証法を用いた．推定したマルチ・ファクターモデルからファクター・リターンに対する銘柄の影響度を推定し，特異なパフォーマンスを示す銘柄を投資対象としてポートフォリオに組み込む手法を提案した．

バックテストでは，提案したマルチ・ファクターモデルと手法を用いて構築したポートフォリオが長期的に安定してベンチマーク（日経 225 や TOPIX）をアウトパフォームする結果となった．これはファクターのリスク・プレミアムを上手く抽出できた結果であると考えられ，同様の戦略によるポートフォリオを Bloomberg Global Investment Contest において提出した．

第1章　はじめに

金融資産の運用では，どのように各資産の将来価格を予測するのかが重要な課題となる．株価の将来価格の予測は，1) ファンダメンタルズ分析に基づく手法、2) 株価変動の時系列特性から変動パターンを抽出し，将来価格の予測に用いるテクニカル分析、3) 統計モデルによる手法、を用いて行われるのが一般的である．このようにアプローチの仕方は様々存在するが，企業や投資家にとって，将来の株価予測モデルの構築は金融資産運用を行う上で重要な課題である．本レポートでは Bloomberg 投資コンテストに向け無裁定価格理論 (APT) とマルチ・ファクターモデルによる株価収益率モデルを構築し，ポートフォリオ運用に応用する．

1.1　Bloomberg 投資コンテスト

本コンテストで行うシミュレーションは一般的に行われている金融取引とは異なる．最終的な目標は一定のルールのもとでより高い収益を獲得することであるため，投資の戦略もルールに基づいたものでなくてはならない．ルールを概観し，以下にまとめた．

1. ルール
   (a) ポートフォリオ登録時点で時価総額が 1 億円以内になるように登録
   (b) 10 銘柄以上，30 銘柄以下でのポートフォリオを構成
   (c) 2017 年 7 月 31 日までであれば 1 度だけ銘柄の入れ替え (リバランス) が可能
   (d) ロングポジションのみ（空売り禁止）
   (e) 手数料は考慮しない
2. パフォーマンス計測期間
   (a) 2017 年 7 月 3 日 ～ 2017 年 8 月 31 日
3. パフォーマンス測定方法
   (a) ポートフォリオ機能「トータルリターン (%)」を頻度日次・円建てで計測
   (b) ポートフォリオ登録は登録日より 2 営業日前の終値をコスト価格として登録
   (c) ポートフォリオ登録後，6 月中の価格変動による時価総額の増減を含め 7 月 3 日から 8 月 31 日の間でパフォーマンスを計測

以上のルールの中で，時間的な制約は 1-c, 2-a である．これらにより 1 か月単位のバイアンドホールドを強いられ，さらにルール 3-b により終値での取引に限定され，鞘取りを行うことは困難である．つまり，長期的な視点から価格変動のメカニズムを考察し，どの銘柄をポートフォリオに組み込むのか、という問題を考えなければならない．次節では，価格変動の源泉について考察した．

1.2 価格変動の源泉

金融資産は一般的な商品と異なり，投資目的やリスクヘッジ目的で購入されることがほとんどである．そのため資産価格は需要供給の関係だけでなく価格変動の予想にも影響される．株式市場が効率的であるという効率的市場仮説のもとでは，あらゆる「情報」が株価形成に瞬時に反映されているはずであるため，それらの「情報」を用いた将来価格の予測は不可能ということになる．一方，テクニカル分析に基づく株価予測モデルは，株価の将来価格は過去の株価の変動パターンから予測可能であるという仮定のもとに用いられる．一般の投資家は株式市場で適切な価格形成が行われているか十分な情報を得ることが出来ないため，高度に習熟したプロフェッショナル（証券アナリストや裁定取引を行う投資家など）による株価予測が有益な情報となる．一方アナリストによる予測は広く投資家間で共有され，株価上昇が見込まれる銘柄に買いが集まるかもしれない．そこで本研究では，アナリストやテクニカル指標による株価予測の情報は，1) 検証することが困難、2) バイアンドホールドによる運用には向かない、などの理由から考察の対象としない．本研究では、株価は資本資産価格モデル (CAPM) により決定されると仮定した統計モデルによる予測を試みる。

1.3 資本資産価格モデル (CAPM)

株式市場に上場している様々な企業は、それぞれが異なった価格変動の性質を持っている．企業の経営方針や持っている技術，リリースしている商品，組織体系などが多種多様であるためだ．しかし，複数の企業に共通している性質も考えられる．例えば所属国や業種，企業の規模などだ．このような共通の要因（ファクター）のうち，株価の変動に関係を持つものが存在する可能性は十分にある．ファクターモデルとは、株価変動のある部分は共通要因の変動によって説明可能である、とするモデルである。

資産価格変動の共通要因を考えた理論として最も広く知られているものの一つは，ウィリアム・シャープらによる資本資産価格モデル (Capital Asset Pricing Model : CAPM) だろう．市場ポートフォリオを唯一の共通要因とした資産価格の評価手法である．CAPM は 1960 年代より不動の地位を築き，その計算の簡便さもあり現在でも広く用いられている．しかし 1970 年代以降，CAPM に対する様々な批判や問題点が提起され，代わりとなる新たな理論が提唱されてきた。株価変動の共通要因は複数存在すると仮定したマルチ・ファクターモデルである。

最も代表的なマルチ・ファクターモデルは Fama-French の 3 ファクターモデル [5] だろう．このモデルを構築したファーマとフレンチは米市場における実証分析を行い，企業規模 (Size) や簿価比時価率 (Value) が株価の収益構造に関与していると結論付けた．企業規模の小さい銘柄や，時価総額が総資産額に比べて割安な銘柄は平均的に高い収益を見せることを示したのである．この現象はそれぞれ小型株効果，バリュー株効果と呼ばれ，CAPM における代表的なアノマリー（説明できない事象）として認識されている．ファーマとフレンチはその功績により 2013 年にノーベル経済学賞を受賞している．実際に似たような性質 (業種や企業規模など) を持つ銘柄は似たような価格変動を見せることが多く，ファーマとフレンチが対象とした米市場に限らず，金融資産の価格変動が市場ポートフォリオ以外のファクターにも影響されているという主張は自然なものだと考えられる．実際、多くの先進国市場を対象にした Fama-French の 3 ファクターモデルの有用性が報告されている。

また，ファクターモデルと密接に関わる理論としてステファン・ロスの無裁定価格理論 (APT)[7] がある．APT は CAPM と異なり，全資産の収益率の同時分布が正規分布である

ことを必要しない．APT が必要とする仮定は,「投資家はただ飯をいくらでも食べたがる」という，ほぼ自明の行動原理のみである [1]．そのため APT は CAPM よりも柔軟な理論として知られている．

1.4 戦略の決定

市場には無数の裁定取引を行う投資家が存在すること，さらに本コンテストのルールにより，鞘取りに関しては全く行うことが出来ない．このことは市場に裁定機会が存在しないことを主な仮定とする APT と非常に相性の良いものであり，APT は本コンテストのルールの下で収益を出すために十分有用であると考えられる．独自のマルチ・ファクターモデルを構築し,「直近で平均的に勝てている投資スタイル」を見つけ出すことが出来れば2ヶ月間の収益を競う本コンテストにおいても成果を上げることが出来るのではないかと考えた．マルチ・ファクターモデルと APT により，CAPM よりも確からしい資産の価格変動の構造を考え，確かな運用手法を提案することを目標にする．また，本コンテストのパフォーマンス測定は 2 カ月間のみの短いものであり，偶発的な金融危機については考えないこととする．とはいえ収益に見合わないリスクを持つポートフォリオを構成することには利点がないため，特に理由がない限り構成銘柄数は最大の 30 銘柄を考える．

マルチ・ファクターモデルにより各銘柄の収益構造を推定した後の問題はポートフォリオに組み込む資産をどのように選択するかである．投資には必ずリスクが伴い，投資家はそのリスクを代償にリターンを求める．そこでここでは，それぞれのファクターの持つリスクに晒される「価値」を考えることとした．この「価値」は通常リスク・プレミアムと呼ばれる．高いリスク・プレミアムを持つファクターに対してリスクを取り，低いリスク・プレミアムを持つファクターに対して分散化することを考えていく．

1.5 レポート構成

本レポートの 2 章ではポートフォリオ構築に使用した諸理論や既存のモデルを体系だてて述べる．3 章では実際に行ったデータ解析の手順を述べ，バックテストの結果と共に提案する手法を示す．4 章では提出したポートフォリオの途中経過を示し，行ったリバランスについても触れる．5 章では提案手法に対する課題の提示と，まとめを行う．

第2章　ファクターモデルに基づくポートフォリオ選択

ポートフォリオ構築の基盤とした理論や，参考にした既存のモデルについて述べる．なお，金融工学に基づく基本的な理論は文献 [1][2] を参考にした．

2.1　シングル・ファクターモデルとしての CAPM

市場を構成する資産数を N とする．CAPM の枠組みにおいて金融資産 i の収益率 r_i は，唯一の共通要因（シングル・ファクター）である市場ポートフォリオの収益率 R_M に依存する変動と，資産特有の変動 ε_i に分けられ以下の式 (2.1) ように表される．また，資産特有の変動 ε_i は資産間で相互に無相関であると仮定される．r_f は無リスク利子率を表す定数であり，信用の高い長期国債の年利が一般的に用いられる．

$$\begin{aligned}
&r_i - r_f = \alpha_i + \beta_i(R_M - r_f) + \varepsilon_i, \qquad (i = 1, \cdots, N) \\
&\varepsilon_i \sim i.i.d.N(0, \sigma_i^2), \\
&Cov(R_M, \varepsilon_i) = 0, \\
&Cov(r_i, r_j) = 0, \qquad (i \neq j)
\end{aligned} \tag{2.1}$$

ここで，β_i は金融資産 i の市場ポートフォリオへの感度を表し，α_i は市場ポートフォリオに対する期待超過収益率を表している．式 (2.1) より r_i の分散を求めると

$$Var(r_i) = \beta_i^2 Var(R_M) + Var(\varepsilon_i) \qquad (\because Cov(R_M, \varepsilon_i) = 0) \tag{2.2}$$

となる．この右辺第 1 項はシステマティック・リスクと呼ばれ市場ポートフォリオの動きにより説明可能な部分である．また右辺第 2 項は企業固有のアンシステマティック・リスクと呼ばれる．

これらのリスクのうち，アンシステマティック・リスクはポートフォリオ選択によって除去可能であるとしばしば言われる．このことを見るために，式 (2.1) により記述される資産を用いた空売り禁止のポートフォリオを考える．資産 $i (i = 1, \cdots, N)$ の保有比率が w_i であるようなポートフォリオ P の収益率 R_P は以下で与えられることになる．ただし，$\sum_{i=1}^N w_i = 1, w_i \geq 0$ である．

$$\begin{aligned}
R_P &= \alpha_P + \beta_P R_M + \varepsilon_P \\
&= \sum_{i=1}^N w_i\alpha_i + \left(\sum_{i=1}^N w_i\beta_i\right) R_M + \sum_{i=1}^N w_i\varepsilon_i
\end{aligned} \tag{2.3}$$

また，このポートフォリオの分散は以下で与えられる．

$$\begin{aligned} Var(R_P) &= \beta_P^2 Var(R_M) + Var(\varepsilon_P) \\ &= \left(\sum_{i=1}^{N} w_i \beta_i \right)^2 Var(R_M) + \sum_{i=1}^{N} w_i^2 Var(\varepsilon_i) \end{aligned} \tag{2.4}$$

この β_P がポートフォリオの市場ポートフォリオに対する感度になり，α_P がポートフォリオの市場ポートフォリオに対する期待超過収益率となる．ここで，式 (2.4) の右辺第 2 項がポートフォリオのアンシステマティック・リスクである．$0 \leq w_i \leq 1$ であることより，以下の式が成立する．

$$\sum_{i=1}^{N} w_i^2 Var(\varepsilon_i) \leq \sum_{i=1}^{N} w_i Var(\varepsilon_i) \tag{2.5}$$

つまり，アンシステマティック・リスク $Var(\varepsilon_i)$ が銘柄ごとに大差ない場合，十分分散化されたポートフォリオのアンシステマティック・リスクは，各資産のアンシステマティック・リスクに比べ非常に小さくなる．

2.2 CAPM に対する批判

1.3 節で述べたように，CAPM に対しては様々な批判が為され，実証分析が行われてきた．以下にその代表的なものを挙げる．

- Basu(1977)[3]
  ニューヨーク証券取引所において，株価収益率 (PER) の低い銘柄が高い銘柄に比べて良いパフォーマンスを見せることを実証した．

- Banz(1981)[4]
  企業規模の小さい銘柄が大きい銘柄に比べて良いパフォーマンスを見せることを実証した．

- Jagadeesh and Titman(1993)[6]
  モメンタムの強い銘柄が市場をけん引し続ける傾向があることを実証した．

ファクターモデルによるアンシステマティック・リスクはポートフォリオ選択によって除去可能であると言われてきたが，これが不可能であることをいずれの研究も示している．企業特有の撹乱項が企業ごとに独立ではなく，共分散が影響しているということである．マーケット以外の共通要因も存在しうることを示唆している．CAPM の枠組みでは，市場ポートフォリオによって説明できる部分 (市場ベータ) と出来ない部分 (アルファ) に分けられ，超過収益であるアルファはポートフォリオ・マネージャーの手腕によるものたと解釈されてきた．しかし次節で述べるマルチ・ファクターモデルの発見により，マネージャーの手腕によると言える部分は徐々に減り，代わりにファクターのリスク・プレミアムによる部分が増えることとなった．

以下の図 2.1 にそのイメージを示した．左から 1 番目の棒グラフはポートフォリオ収益率全体を表している．2 番目の棒グラフは CAPM の $R_P = \alpha_P + \beta_P R_M$，3 番目は $R_P = \alpha_P + \beta_{market} F_{market} + \beta_{size} F_{size}$，4 番目は $R_P = \alpha_P + \beta_{market} F_{market} + \beta_{size} F_{size} + \beta_{value} F_{value}$ とモデル化したときの収益構造のイメージ図である．

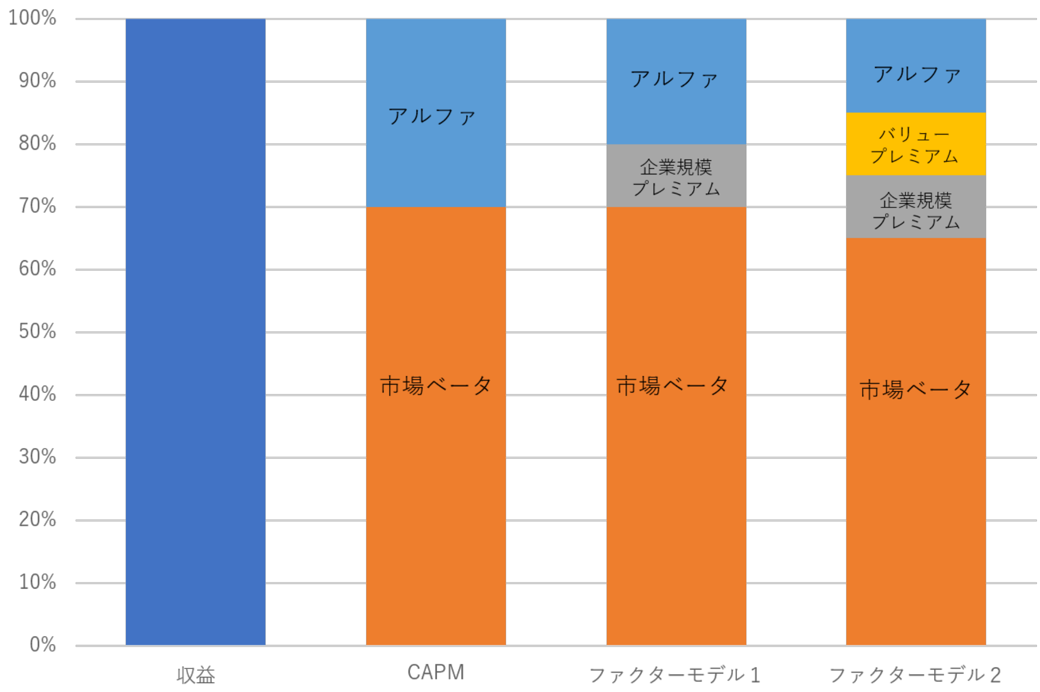

図 2.1: モデル別にみた収益構造

2.3 マルチ・ファクターモデル

ここでは，各資産の収益率が m 個の共通要因に依存するマルチ・ファクターモデルを考える．金融資産 i の収益率 r_i は企業特有の部分 ε_i と共通要因であるファクター $F_k(k=1,\cdots,m)$ によって決定される．また，ε_i は資産間で相互に無相関であると仮定すると，r_i は以下のように表される．

$$r_i = \alpha_i + \beta_{i,1}F_1 + \beta_{i,2}F_2 + \cdots + \beta_{i,k}F_k + \cdots + \beta_{i,m}F_m + \varepsilon_i \tag{2.6}$$

ここで，2.1 節と同様，資産 $i(i=1,\cdots,N)$ の保有比率が w_i であるようなポートフォリオ P の収益率 R_P を考えると以下のようになる．

$$\begin{aligned} R_P &= \alpha_P + \beta_{P,1}F_1 + \beta_{P,2}F_2 + \cdots + \beta_{P,m}F_m + \varepsilon_P \\ &= \sum_{i=1}^{N}\alpha_i + F_1\sum_{i=1}^{N}w_i\beta_{i,1} + \cdots + F_m\sum_{i=1}^{N}w_i\beta_{i,m} + \sum_{i=1}^{N}w_i\varepsilon_i \\ &= \sum_{i=1}^{N}\alpha_i + \sum_{k=1}^{m}\sum_{i=1}^{N}w_i\beta_{i,k}F_k + \sum_{i=1}^{N}w_i\varepsilon_i \end{aligned} \tag{2.7}$$

2.4 APT

2.4.1 仮定

APT では，金融資産の価格構造がファクターモデルで表されることを仮定する．ここでは一般形として，任意の資産 i の収益率は式 (2.6) のように表され，ε_i は資産間で相互で無相

関であるとする．またAPTでは追加として，各ファクターの期待値が0でなければならない．これは，あらかじめ使用するファクターを正規化することにより解決する．以上APTの仮定についてまとめたものを以下に示した．

APTの仮定

各資産 i の収益率は以下のように表される．

$$r_i = \alpha_i + \beta_{i,1}F_1 + \cdots + \beta_{i,k}F_k + \cdots + \beta_{i,m}F_m + \varepsilon_i \tag{2.8}$$

$$E(\varepsilon_i) = 0 \qquad (i = 1, 2, \cdots, N) \tag{2.9}$$

$$E(F_k) = 0 \qquad (k = 1, 2, \cdots, m) \tag{2.10}$$

$$Cov(\varepsilon_i, \varepsilon_j) = 0 \qquad (i \neq j) \tag{2.11}$$

ここで，式(2.8)について期待値を取ると

$$E(r_i) = \alpha_i \tag{2.12}$$

となる．つまり，式(2.8)における定数項は資産 i の期待収益率に相当する．ファクターモデルを仮定しているので，複数の資産で構成されるポートフォリオの収益率は式(2.7)と同様に記述される．ここで，式(2.11)により，ポートフォリオが十分な数の資産で構成されるとき，ε_P は無視しても構わない．つまり，十分に分散化されたポートフォリオの収益は，ファクターに対する感度のベクトル $(\beta_{P,1}, \cdots, \beta_{P,m})$ によってのみ特徴づけられる．

2.4.2 APTの主定理

APTの発想は，市場に裁定機会がないというものであった．無数の市場参加者が存在することによって裁定取引は常に行われ，適正でない金融資産の価格は裁定取引のできない水準に収束することになる．これはノー・フリーランチの原理と呼ばれる．この原理を価格理論に応用することにより，APTの主定理が導かれる．十分に分散化されていれば，ファクターへの感度が全く同じになっているポートフォリオ同士のリターンは同じでなければならない，ということである．

APTの主定理

証券の収益率がマルチ・ファクターモデルで表されるとする．このとき，任意の資産 i の各ファクターへの感度を $(\beta_{i,1}, \beta_{i,2}, \cdots, \beta_{i,m})$ とすると，資産 i のリスク・プレミアムは

$$\lambda_1\beta_{i,1} + \lambda_2\beta_{i,2} + \cdots + \lambda_m\beta_{i,m} \tag{2.13}$$

で近似的に表される．ただし，$(\lambda_1, \lambda_2, \cdots, \lambda_m)$ は各ファクターのリスク・プレミアムである．

2.4.3 ファクターのリスク・プレミアムの推定

APTの主定理で用いるファクターのリスク・プレミアムの推定について触れる．資産を組み合わせて，特定のファクターへの感度のみが1で，その他のファクターへの感度が0に

なるポートフォリオ（ファクター・ポートフォリオ）を考える．例えば，第 1 ファクターのみへの感度が 1 になるようなポートフォリオを P_1 とすると，その収益率は

$$R_{P_1} = \alpha_{P_1} + 1 \times F_1 + 0 \times F_2 + \cdots + 0 \times F_m + \varepsilon_{P_1} \tag{2.14}$$

と表せる．このとき，α_{P_1} が第 1 ファクターの期待収益率であり，リスク・プレミアムとなる．

一般に，m 個のファクターを想定すると，単一ファクター・ポートフォリオの収益率ベクトルは

$$\begin{pmatrix} R_{P1} \\ R_{P2} \\ \vdots \\ R_{Pm} \end{pmatrix} = \begin{pmatrix} \alpha_{P1} \\ \alpha_{P2} \\ \vdots \\ \alpha_{Pm} \end{pmatrix} + \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{pmatrix} + \begin{pmatrix} \varepsilon_{P1} \\ \varepsilon_{P2} \\ \vdots \\ \varepsilon_{Pm} \end{pmatrix} \tag{2.15}$$

と表される。行列ベクトル表記すれば次式のようになる。

$$\boldsymbol{R}_P = \boldsymbol{\alpha}_P + \boldsymbol{I}_m \boldsymbol{F} + \boldsymbol{\varepsilon}_P \tag{2.16}$$

となる，この単位行列 $\boldsymbol{I}_P$ はファクターモデルの係数行列 $\boldsymbol{B}$ と未知の荷重行列 $\boldsymbol{W}$ の積で表現できると仮定する．

第 $k(k = 1, 2, \cdots, m)$ ファクターに対する感度のみが 1 であるポートフォリオを第 k ファクター・ポートフォリオ P_k とし，P_k における資産 i のウェートを $w_{i,k}$ とする．さらに，資産 i の第 k ファクターへの感度を $\beta_{i,k}$ とすると，以下の式 (2.17) を $\boldsymbol{w}$ について解くことによりファクター・ポートフォリオを構築できる．

$$\boldsymbol{W}^{\mathrm{T}} \boldsymbol{B} = \boldsymbol{I_m} \tag{2.17}$$

式 (2.17) 左辺の $\beta_{i,k}$ による感度行列は式 (2.7) により推定されるので，$w_{i,k}$ によるウェート行列が求める対象である．感度行列の逆行列を左辺の右からかければ良いことになるが，感度行列の型は $n \times m$ であり，正方行列ではない．しかしこれは一般化逆行列 $\mathbf{B}^{\dagger}$ を用いることにより解決する．ただし、$\boldsymbol{B}^{\dagger} = (\boldsymbol{B}^{\mathrm{T}} \boldsymbol{B}^{-1}) \boldsymbol{B}^{\mathrm{T}}$ である。

第3章　データ解析

3.1　使用したデータ

2015年1月から2017年5月にかけて，東証一部上場企業の株価と各指数を日次終値を取得した．なお，データの取得にはBloomberg端末とMicrosoft ExcelのBloombergアドインを使用し，データ分析はR言語により行った．

- 投資対象市場の選択について
  東証一部上場企業の株式の価格データを使用した．
  数々の先行研究により，CAPMだけでなくマルチ・ファクターモデルにおいても市場ポートフォリオが第1の共通要因になることに疑いの余地はない．そこで分析を行いやすくするため対象ユニバースは一つに絞ることを考え，馴染みの深い日本の東京証券取引所の株式を選択した．また東証二部や東証マザーズに関して，流動性の不十分性よりマルチ・ファクターモデルが十分な説明力を発揮しない可能性があるため，除外することとした．また，マルチ・ファクターモデルを最大限活かすために投資先を株式のみとし，コモディティや債券，デリバティブ，ETFなどは利用しなかった．

- ファクターについて
  マルチ・ファクターモデルを構成する際，以下のファクターを使用した．
  - マーケット・ファクター
    東証一部上場銘柄を対象としているため，マーケットファクターとしては東証一部上場銘柄の時価総額を反映しているTOPIXを使用した．マーケットファクターの収益率には対数収益率を用いた．
  - サイズ・ファクター
    Russell/Nomuraの提供する日本株インデックスで代用した．具体的には，配当を含めないSmall Cap Indexから配当を含めないLarge Cap Indexを引くことにより算出した．さらにその変動をみるために差分を取ったものを使用した．
  - バリュー・ファクター
    Russell/Nomuraの提供する日本株インデックスで代用した．具体的には，配当を含めないTotal Value Indexから配当を含めないTotal Growth Indexを引くことにより算出した．さらにその変動をみるために差分を取ったものを使用した．
  - 為替ファクター
    日本円(JPY)の市場価格(対米ドル)を為替ファクターとして使用した．さらにその変動をみるために対数差分を取ったものを使用した．
  - 市場ボラティリティ・ファクター
    S&PがVIXと同様の手法で算出している日本市場の30日インプライド・ボラティリティを使用した．

上記のファクターはそれぞれ単位もスケールも異なるので，正規化 (平均 0 分散 1 に標準化) した後に使用した．

サイズ・ファクターとバリュー・ファクターについては先行研究が豊富であり，その存在はほぼ確実とされている．次に為替ファクターであるが，これは日本市場が主にアメリカ市場から影響を受けることに起因する．マーケット・ファクターが為替ファクターを包含しているのではないか，という疑問は持たれるだろうが，業種や海外進出の有無などによりその感度は銘柄ごとに異なるだろう．そのため円ドル相場をファクターとして取り入れることとした．最後にボラティリティ・ファクターについてであるが，これは低ボラティリティ・ファクターとは異なるものである．というのも投資家たちは市場ボラティリティの大きさに従い，リスクを調整するためにリバランスを行うが，その際に取引が集中する銘柄を検出しようと狙ったものである．このとき，資産価格はボラティリティの変動というよりもその大きさに影響されると考えられるため，差分を取らずに日経 VIX 系列を正規化した．

3.2　データのクリーニング・概観

分析対象期間の 2015 年 1 月から 2017 年 5 月において，何らかの理由（未上場等）により継続的にデータが取得できない企業を除外した．分析対象の 1948 企業に対し，さらに対数差分をとることによって各営業日の日次対数収益率を算出した．

対象銘柄すべての価格推移をプロットすることは紙面の都合上冗長であるため割愛した．ファクターに関しては，前節のように各指数を加工した後に使用する．原系列である各指数について確認するため，Bloomberg 端末にて該当期間のラインチャートを取得し，エクスポートしたものを以下の図 3.1 から 3.4 に示した．図 3.1 は TOPIX の原系列である．ファクターには対数差分を正規化したものを使用した．

図 3.1: TOPIX の時系列プロット

図 3.2 の上段が日本円レート（対米ドル），下段が S&P が提供する日経 VIX の原系列である．日経 VIX は正規化したものを，日本円レートは対数差分を正規化したものをファクターとして使用した．

図 3.2: 日本円レートと日経 VIX の時系列プロット

図 3.3 の上段が Russell Nomura の提供する日本小型株インデックスと日本大型株インデックス，下段が日本小型株インデックスから日本大型株インデックスを減算したものである．ファクターには下段の差分を正規化したものを使用した．

図 3.3: Russell Nomura 日本株インデックス (Size)

図 3.4 の上段が Russell Nomura の提供する日本バリュー株インデックスと日本グロース株インデックス，下段が日本バリュー株インデックスから日本グロース株インデックスを減算したものである．ファクターには下段の差分を正規化したものを使用した．

図 3.4: Russell Nomura 日本株インデックス (Value)

次に，実際にファクターとして使用する時系列を図 3.5 に示した．また，多重共線性などの問題が無いことを確認するため，散布図とヒストグラム，相関行列を図 3.6 にまとめた．図 3.6 の対角線上にある図が各ファクターのヒストグラムとカーネル密度関数，下三角の図がファクター間の散布図，上三角の図がファクター間の相関係数に 100 を掛けた数値を表している．さらに，ファクターの基本統計量を表 3.1 に示した．

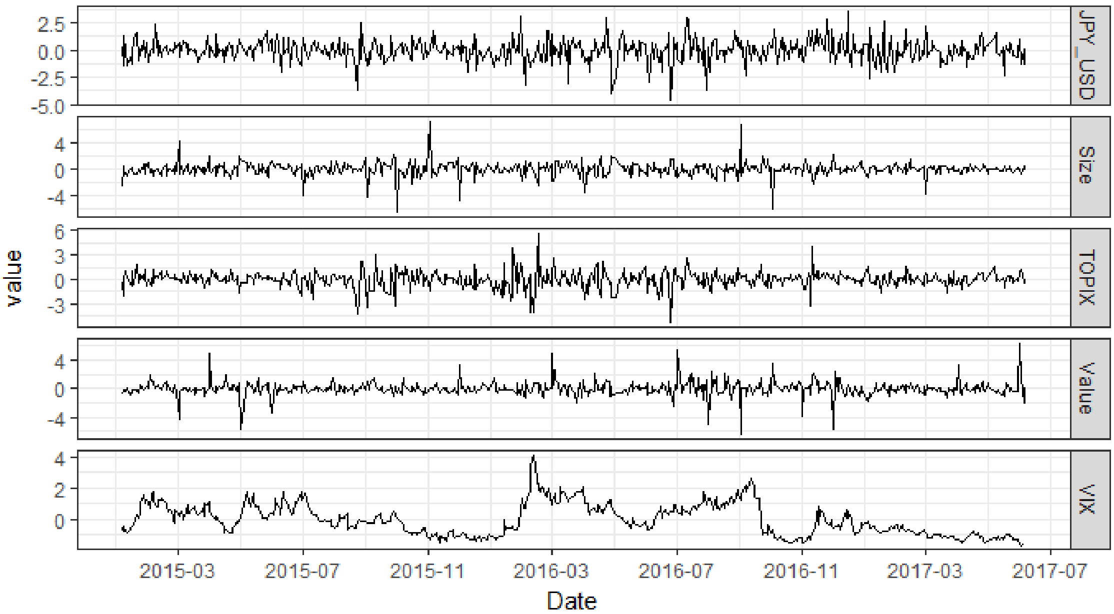

図 3.5: ファクターの時系列プロット

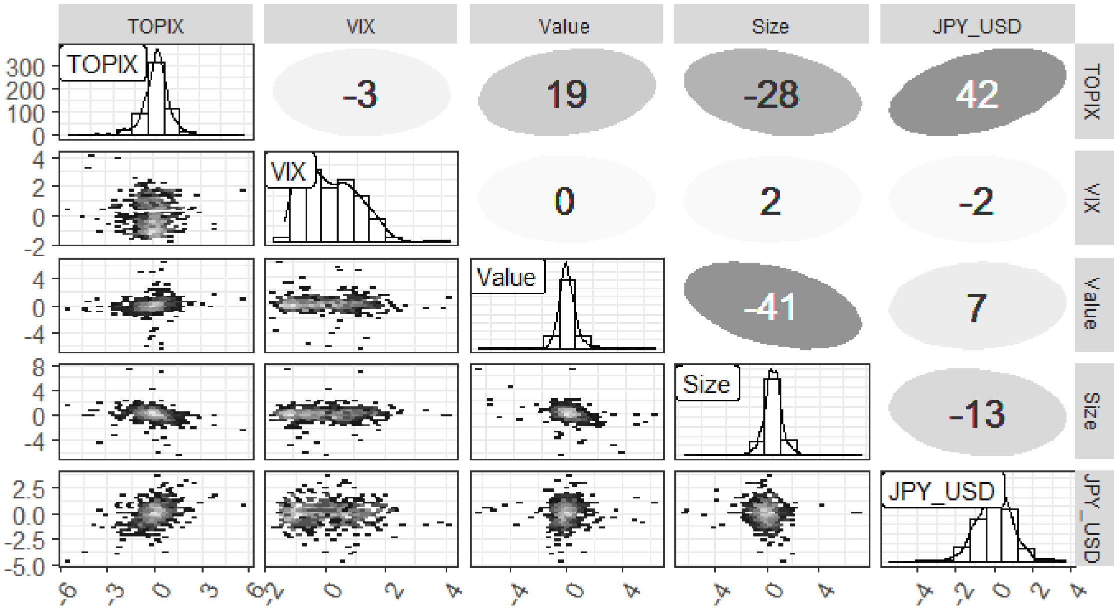

図 3.6: ファクター間の相関関係とヒストグラム

統計量	TOPIX	VIX	Value	Size	JPY_USD
Min	-5.33806	-1.6387	-6.34983	-6.42186	-4.64922
1st Qu	-0.45082	-0.8659	-0.41889	-0.39178	-0.56066
Median	0.05201	-0.1465	-0.06176	0.02299	0.02342
Mean	0.00000	0.0000	0.00000	0.00000	0.00000
3rd Qu	0.49170	0.7329	0.38466	0.46966	0.59423
Max	5.69692	4.1170	6.21356	0.46966	0.59423
sd	1.00000	1.0000	1.00000	1.00000	1.00000

表 3.1: ファクターの基本統計量

図 3.5 より，TOPIX, Value, Size, JPY_USD はどれも定常時系列であることが分かる．対数差分や差分を取っているこれらの系列がこのような動きを示すことは自然なことであるが，中でも Value, Size は時折大きな変動を見せており，収益率の従う確率分布が他とは異なる可能性がある．JPY_USD の差分系列は最も変動が安定しており，TOPIX の対数差分系列がその中間ぐらいであると言える．また VIX に関してであるが，ボラティリティの変化率ではなく，大きさをファクターとして扱うために原系列をそのまま正規化した．そのため明らかに分散が不均一な過程になっている．

次に図 3.6 を確認する．既に述べたが TOPIX, Value, Size は時折大きな変動を見せるため，中心付近で高く，裾の広い分布になっている．図 3.6 右上の，ファクター間の相関係数を 100 倍した数値を確認すると，全体的に低くまとまっているものの，低い相関を持つファクターの組み合わせが存在することが見て取れる．この場合 TOPIX と JPY_USD の 42，Value と Size の-41，TOPIX と Size の-28 であるが，この程度の相関なら多重共線性は考えなくてよい．よって，完全に独立なファクターであるとは言えないまでも，回帰モデルには十分利用

可能なファクターであることが確認できた．モデルのパラメータ推定方法については次節で述べる．

3.3 ファクターモデルの推定

モデル推定のための学習期間を 3 カ月に設定し，推定したモデルでその先 1 カ月の運用パフォーマンス（バックテスト）を計測する．この分析を 1 カ月ごとにローリングしながら推定と計測を繰り返す．そのイメージ図を以下の図 3.7 に示した．

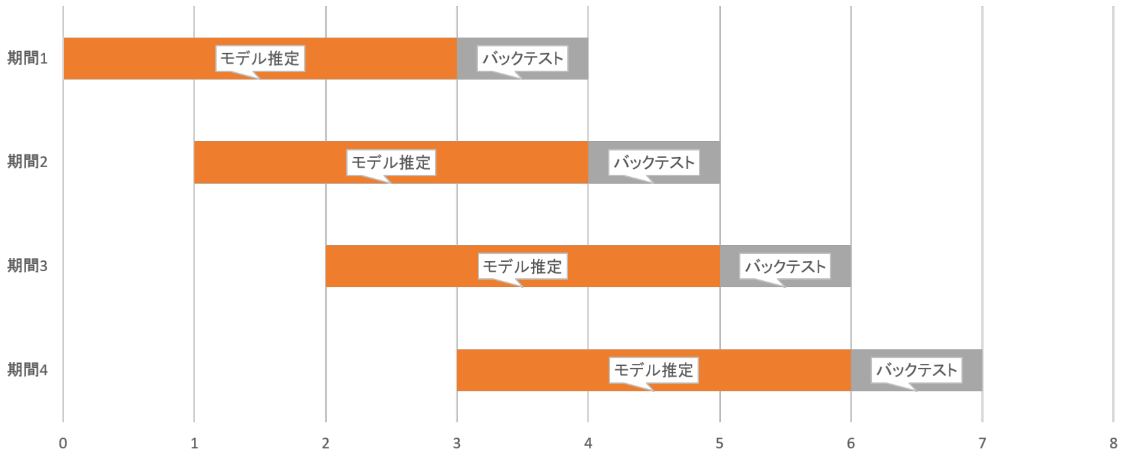

図 3.7: バックテストのイメージ図

マルチ・ファクターモデルを推定する際，Fama-French の 3 ファクターモデルのような線形モデルだけでなく，非線形モデルも考慮に入れることができる．しかしながら，標本の少なさや時系列データにおける外挿の問題により複雑なモデルを考えるメリットが小さく，ここでは線形モデルのみを考えた．

過適合を避けるため，lasso 回帰によるモデル推定と変数選択を行うこととした．正則化項付きのモデル推定は「あえてすべての変数を使う」という立場にある．説明変数を「使う」「使わない」のゼロイチで切り捨てる代わりに，全ての係数を少しずつ抑えるという発想の手法である．具体的には，リッジ回帰では $\sum_i \beta_i^2$ に対して，lasso 回帰では $\sum_i |\beta_i|$ に対して罰則を加えるのである．この操作により回帰係数が過大になることを防ぎ，なおかつ汎化性能を高めることができる．

lasso 回帰で罰則を加える $\sum_i |\beta_i|$ とリッジ回帰で罰則を加える $\sum_i \beta_i^2$ に大差はないように思える．2 乗の β_i^2 を絶対値 $|\beta_i|$ に置き換えただけであるが，こうすることにより興味深い性質が生まれる．任意の説明変数がある程度以上「いらない変数」の場合，その係数が小さくなるだけでなく 0 になってしまうのである．すなわち，係数の過大化を防ぎ汎化性能を高めるだけでなく自動的に「変数選択」を行う能力を持つのである．

以上の理由によりモデル推定には lasso 回帰を用いることとした．$t = 1, \cdots, T$ の期間において，銘柄 i のマルチ・ファクターモデルは以下の式を最小化することにより求まる．α_i

が含まれていないが，これは各推定期間ごとに正規化することによる．

$$\sum_{t=1}^{T}\left(r_{i,t}-\sum_{k=1}^{5}\beta_{i,k}F_{k,t}\right)^2+\lambda_i\sum_{k=1}^{5}|\beta_{i,k}| \tag{3.1}$$

問題は正則化項のパラメータ λ_i に関してであるが，これは交差検証法によって最小二乗誤差が最小の値を使用する．今回の分析では交差検証法の分割を全て 10 とした．

全 26 期間で行った推定のうち，最初と最後の期間に関する結果を表 3.2, 3.3 に示した．

統計量	(Intercept)	TOPIX	VIX	Value	Size	JPY_USD
Min	-0.0143864657	-0.006802000	-0.02629276	-0.01499114	-0.010450605	-0.01182420
1st Qu	-0.0012404059	0.003709114	0.00000000	0.00000000	0.000000000	0.00000000
Median	-0.0001376986	0.007552877	0.00000000	0.00000000	0.000000000	0.00000000
Mean	0.0001227267	0.0074834257	0.0002030726	0.0004132952	0.0011393320	0.0000702478
3rd Qu	0.0012743206	0.010987066	0.00000000	0.00000000	0.001281696	0.00000000
Max	0.0181611082	0.028263950	0.01792821	0.02979067	0.037666321	0.01556331
sd	0.002412701	0.005135135	0.002535910	0.001950954	0.002910319	0.001690928

表 3.2: ファクターモデルの推定結果（訓練期間：2015 年 1 月から 4 月）

統計量	(Intercept)	TOPIX	VIX	Value	Size	JPY_USD
Min	-0.0311943287	0.000000000	-0.027704158	-0.02503910	-0.010465046	-0.008442638
1st Qu	-0.0003752476	0.005102337	0.000000000	0.00000000	0.000000000	0.000000000
Median	0.0009955913	0.008934296	0.000000000	0.00000000	0.000000000	0.000000000
Mean	2.361706e-03	8.874863e-03	1.459822e-03	6.364366e-06	1.335285e-03	1.654883e-04
3rd Qu	0.0036495111	0.012430849	0.001103099	0.00000000	0.002435982	0.000000000
Max	0.0587778636	0.028889523	0.044701605	0.01014139	0.016323614	0.012010414
sd	0.006413166	0.005366411	0.005131025	0.001808388	0.002374741	0.001316902

表 3.3: ファクターモデルの推定結果（訓練期間：2017 年 3 月から 5 月）

3.4 リスク・プレミアム

過去 3ヶ月のデータより推定したマルチ・ファクターモデルに基づいて算出したリスク・プレミアムを以下の図 3.8 に示した．具体的には，2.6 節の式 (2.17) を一般化逆行列を利用することによりファクター・ポートフォリオを推定し，その期待超過収益率をファクターのリスク・プレミアムとした．

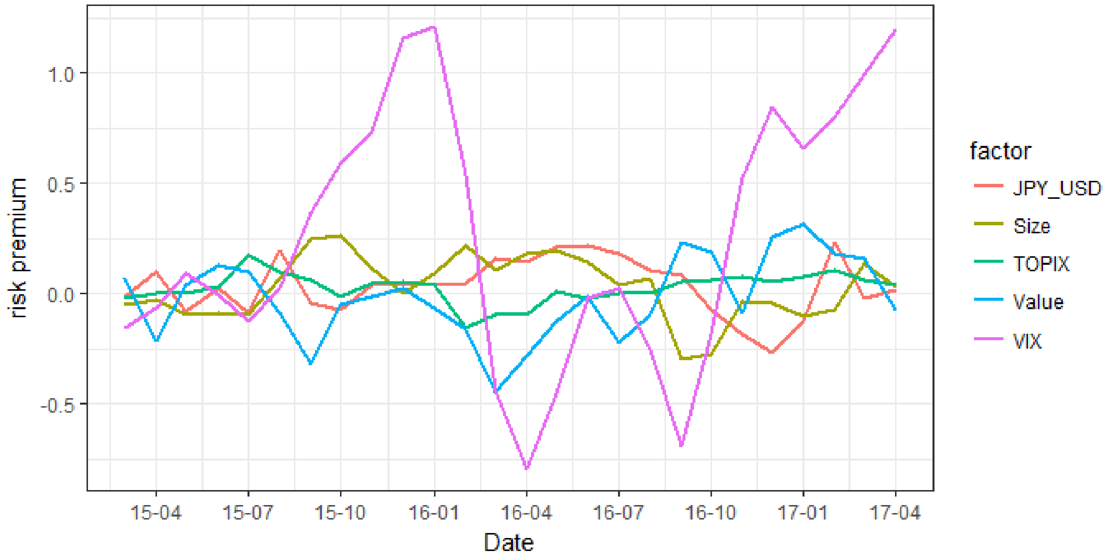

図 3.8: リスク・プレミアムの推移

3.5 様々なポートフォリオ

本レポートで取り入れたファクターが，CAPM におけるアノマリー説明できているかを確かめるため，推定したマルチ・ファクターモデルの各係数 (ベータ) の上位 30 銘柄で等加重ポートフォリオを作成した．また，その累計リターンの推移を以下の図 3.9 に示した．また，市場ポートフォリオに対してのパフォーマンスを確認するために，TOPIX の累計リターンを benchmark として共に示した．

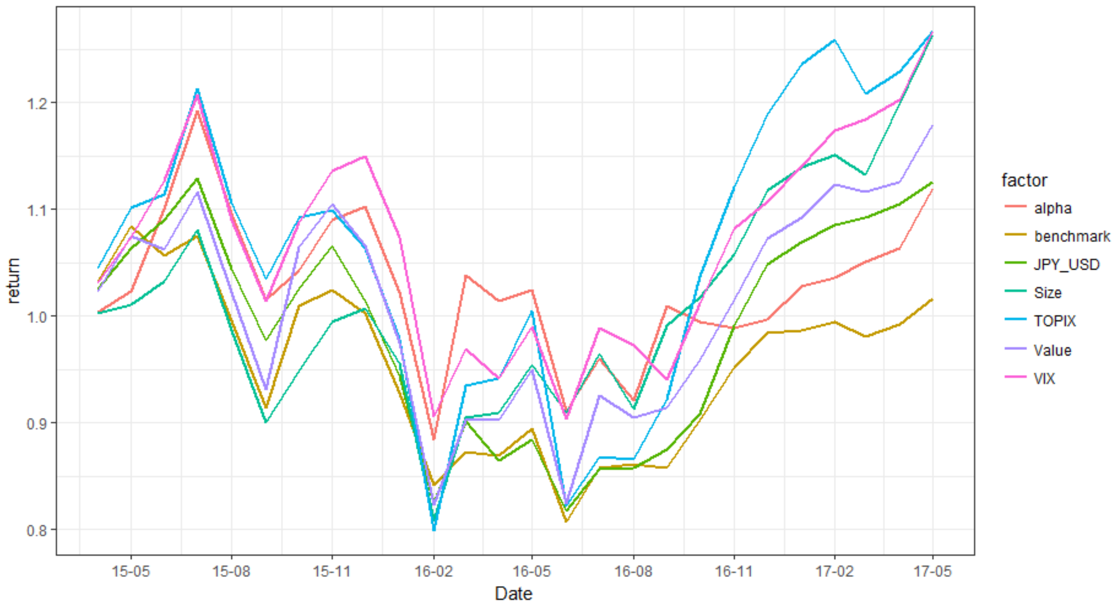

図 3.9: 各ベータ上位 30 銘柄による等加重ポートフォリオ

図 3.9 より，どのポートフォリオも似たような動きをしており，市場ポートフォリオの影響力が非常に大きいことが分かる．しかし全てのファクター・ベータについての上位 30 銘柄のポートフォリオはベンチマークをアウトパフォームしており，市場ポートフォリオだけでは説明のつかないアノマリーの存在をマルチ・ファクターモデルによって抜き出せたことが確認できた．

3.6 ポートフォリオ選択

ポートフォリオの収益率ファクターに対する感度は 2.2 節の式 (2.7) により与えられていた．今回は 5 ファクターモデルを考えているため，ポートフォリオの収益は次のように与えられる．

$$
\begin{aligned}
R_P &= \alpha_P + \beta_{P,1}F_1 + \beta_{P,2}F_2 + \cdots + \beta_{P,5}F_m + \varepsilon_P \\
&= \sum_{i=1}^{N}\alpha_i + F_1\sum_{i=1}^{N} w_i\beta_{i,1} + \cdots + F_5\sum_{i=1}^{N} w_i\beta_{i,m} + \sum_{i=1}^{N} w_i\varepsilon_i \\
&= \sum_{i=1}^{N}\alpha_i + \sum_{k=1}^{5}\sum_{i=1}^{N} w_i\beta_{i,k}F_k + \sum_{i=1}^{N} w_i\varepsilon_i
\end{aligned}
\tag{3.2}
$$

リスク・プレミアムを算出したことによりファクターごとのリスクを取る価値が定量化できたのであった。ここで，負のリスク・プレミアムを持つに対しては分散化し，より高いリスク・プレミアムを持つファクターに対してはエクスポージャーするようなポートフォリオ構築を目指す．10∼30 銘柄に絞り込まなければならないので，銘柄 i がポートフォリオに組み込まれた場合に 1 を，そうでない場合は 0 を示す変数 x_i を導入することで以下のように定式化できる．

$$
\begin{aligned}
\text{maximize :}\quad & \beta_{P,1}\mu_1 + \beta_{P,2}\mu_2 + \beta_{P,3}\mu_3 + \beta_{P,4}\mu_4 + \beta_{P,5}\mu_5 \\
& = \sum_{k=1}^{5}\beta_{P,k}\mu_k \\
& = \sum_{k=1}^{5}\mu_k\sum_{i=1}^{N} w_i\beta_{i,k} \\
\text{subject to :}\quad & \sum_{i=1}^{N} w_i = 1 \\
& w_i \geq 0 \qquad (i = 1, \cdots, N) \\
& 10 \leq \sum_{i=1}^{N} x_i \leq 30 \\
& \mu_k = \max(0, \lambda_k) \qquad (k = 1, \cdots, 5)
\end{aligned}
\tag{3.3}
$$

このうち，各ファクターのリスク・プレミアム λ_k と各銘柄の回帰係数 $\beta_{i,k}$ は算出済みの定数であるので，変数は w_i, x_i となる．x_i についての制約式が加わったことにより，単体法などの良く知られている手法で厳密解を求めることは非常に難しくなった．このような厳密に解くことが難しい最適化問題にはメタヒューリスティクスによるアプローチが知られてい

るが，計算時間の問題により実行できなかった．そこで，以下の提案手法により近似解を求めることとした．

ポートフォリオ選択の手順

1. 直近3ヶ月のデータより推定したファクターごとのリスク・プレミアムのうち，最もプレミアムの大きいファクターへの感度が高いもの上位100銘柄を選択する．
2. 1で選択した100銘柄のうち，期待収益率が高い30銘柄を選択
3. 2で選択した30銘柄により等加重ポートフォリオを構成

リスク・プレミアムが負であるファクターに対し分散化する目的で，等荷重ポートフォリオを構成した．

この手法を用いて行ったバックテストの結果を以下の図3.10に示した．

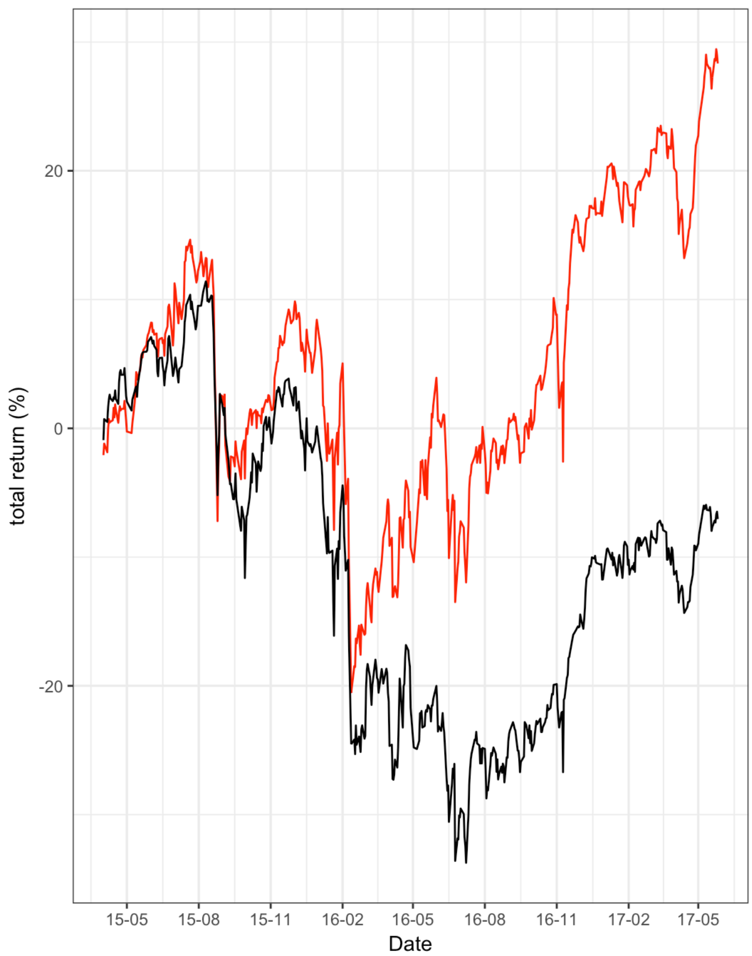

図 3.10: 提案手法によるバックテスト

第 4 章　途中経過

コンテストの測定期間は 2017 年 7 月から 8 月末にかけてであるが，本レポートの提出期限が 2017 年 7 月末であるため，7 月末時点での結果を述べる．提出したポートフォリオの構成とそのリターンを示し，リバランスの内容にも触れる．

4.1　提出したポートフォリオ

3.6 節で提案した手順に基づいて，提出するポートフォリオを決定した．本コンテストのパフォーマンス計測は 2017 年 7 月 3 日より始まるため，2017 年 4 月 3 日から 2017 年 6 月 29 日にかけてデータを取得し，3 章と同様の分析を行った．提案した手法により選択された銘柄とそのウェート，時価総額を表 4.1 に示した．また，ポートフォリオ登録時点では等荷重ポートフォリオを構築したが，7 月 3 日までの価格変動により若干のズレが生じている．

4.2　7 月中のパフォーマンス

提出したポートフォリオの，トータルリターンの推移を図 4.1 に示した．また，東証一部上場銘柄を対象ユニバースとしたため，TOPIX をベンチマークとして合わせて示した．図 4.1 の上段がそれぞれのトータルリターンであり，下段がそのパフォーマンス差である．

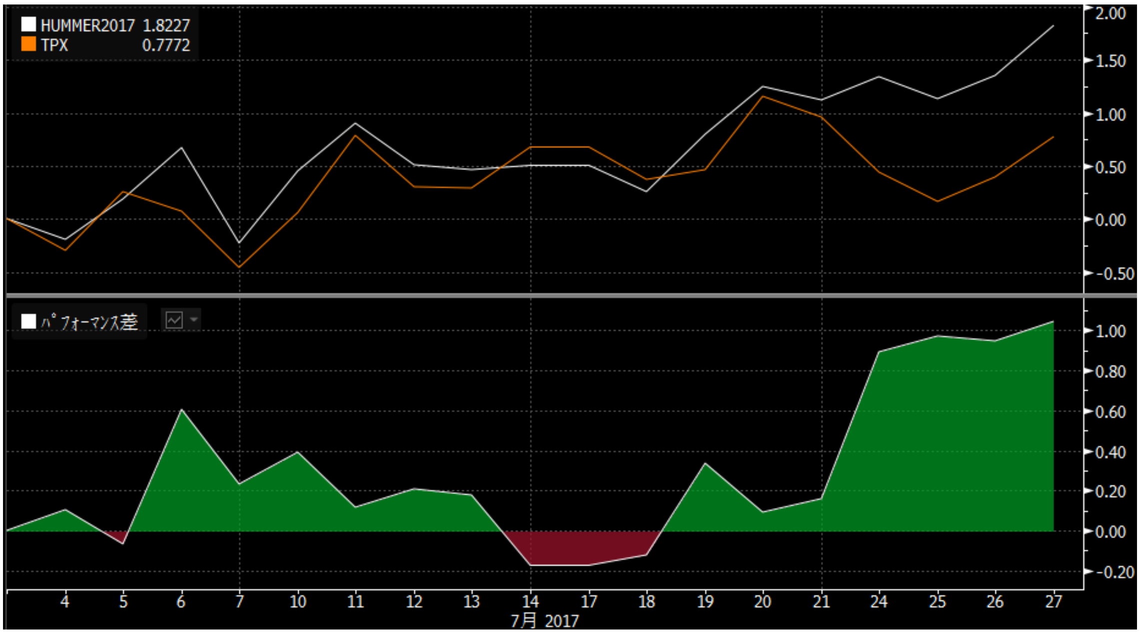

図 4.1: 7 月中のリターン

企業名	ウェート (%)	時価総額 (円, 7 月 3 日時点)
グローブライド	3.34	3,343,642
コメダホールディングス	3.33	3,335,200
ゼンショーホールディングス	3.33	3,339,875
三井ホーム	3.33	3,333,333
井筒屋	3.33	3,341,014
日成ビルド工業	3.35	3,360,302
美津濃	3.28	3,287,108
藤倉ゴム工業	3.36	3,369,011
近鉄百貨店	3.35	3,352,435
ＧＳＩクレオス	3.30	3,309,523
イマジカ・ロボット・ホールディングス	3.34	3,351,877
グランディハウス	3.30	3,309,859
サンフロンティア不動産	3.43	3,432,881
北陸電力	3.32	3,326,772
アイ・オー・データ機器	3.35	3,360,768
アドソル日進	3.36	3,363,838
システムリサーチ	3.36	3,363,148
ルネサスイーストン	3.36	3,367,816
日本電波工業	3.36	3,363,984
ソフトクリエイトホールディングス	3.35	3,361,324
カワチ薬品	3.35	3,353,240
三井製糖	3.39	3,395,734
六甲バター	3.30	3,305,533
森永乳業	3.23	3,234,511
養命酒製造	3.31	3,317,285
ニチハ	3.34	3,350,168
三井住友建設	3.30	3,306,233
太平洋興発	3.33	3,333,333
極東証券	3.28	3,285,697
ＳＯＭＰＯホールディングス	3.35	3,353,441
合計	100.00	100,208,884

表 4.1: 提出したポートフォリオ

2017 年 7 月 3 日の終値を基準とすると，TOPIX のトータルリターンが 0.7772%であるのに対し，提出したポートフォリオのトータルリターンは 1.8227%であった．ベンチマークを 1.0455%アウトパフォームしており，少なくとも 2017 年 7 月では提案した手法が有効であったことが確認できる．

4.3 リバランス

本コンテストのルールにより，リバランスは 7 月中に一度しか行えない．そのため，提案した手法の効果を安定的に得るため，リバランスは 7 月 31 日に行うものとした．1 度目のポートフォリオ提出と同様の手法で銘柄を選択し，等荷重ポートフォリオを再構築した．選択した銘柄とウェート，時価総額を次ページの表に示した．

第5章　今後の課題

1 回目のポートフォリオ提出に間に合わなかった分析や，提出後に浮かんできた疑問点，考えうる批判など数々の課題がある．ここではそれらを列挙し，考察していく．

- データに関する問題

  3.1 節で述べたように，2015 年からの日次データを取得した．この取得を日次ではなく，前場の終値，後場の終値ごとに取得すれば絶対的に標本数が増え，より精度の高い分析が行えるだろう．

- ファクター選択に関する問題

  今回のリサーチでは，3.1 節で述べたように使用するファクターを選択した．その存在が確実といわれるファクターだけでなく，独自のファクターを定性的な理由により取り入れた．しかし 3.2 節で述べたように，正規化したファクター間には弱い相関が存在するものがあった．多重共線性を回避し，汎化性能を高めるためにモデル選択では工夫をしたが，ファクター間に包含関係がある可能性は依然として残ったままである．本来ファクター同士は直交していることが望ましく，統計的な因果推論などを行い適切なファクターの組み合わせ選択を行う必要性があるだろう．

- マルチ・ファクターモデル推定に関する問題

  3.2 節で述べたように，マルチ・ファクターモデルの推定では工夫をした．銘柄の中には，lasso によってマーケット・ファクターやサイズ・ファクターなど，その存在が確実とされているファクターへの感度が 0 と推定されたものがあった．説明力や汎化性能を高めるためとはいえ，このように絶対的なファクターを蔑ろにしてしまうことには議論の余地があるだろう．

  また，モデル推定に用いたデータは直近 3ヶ月間のものであり，1ヶ月単位で (1 月 ∼3 月，2 月 ∼4 月，というような) 推定を行った．例えばこの期間を 2ヶ月間，4ヶ月間，5ヶ月間と変更し同様の作業を行うことで，どのような変化があるのか検証する必要がある．さらに推定を 1ヶ月単位で行うのではなく，1 日単位で行った場合に関しても検証する必要があるだろう．

- リスク・プレミアム推定に関する問題

  上記のように 1 日単位でモデル推定を行うことにより，それに基づく各ファクターのリスク・プレミアムの推移はより滑らかになることが考えられる．たとえそうでなかったとしても，データの数が増えることにより新たな発見につながると予想できる．

  また，ボラティリティ・ファクターのプレミアムの分散は他に比べて非常に大きいものとなっていた．このことは直感に反しており，さらに考察をしていく必要があると感じた．さらに，ファクターのリスク・プレミアム同士がお互いに影響を及ぼしているかについても分析することが出来るだろう．

- ポートフォリオ選択に関して
  計算時間の問題により，式 (3.3) で定式化した最適化問題を実行することが出来なかった．近似として示した提案手法ではなく，多目的最適化の手法を取り入れることにより段階的に最適解探索を考えることが望ましい

関連図書

[1] 小林孝雄，芹田敏夫，日本証券アナリスト協会『新・証券投資理論 I 理論篇』（日本経済新聞出版社，2009）

[2] デービッド・G・ルーエンバーガー『金融工学入門』（日本経済新聞社，2002）

[3] Basu, S. ”Investment Performance of Commin Stocks in Relation to their Price-Earnings Ratios”, Journal of Finance, 663-682(1977)

[4] Banz, R.W. ”The relationship between return and market value of common stock”, Journal of Financial Economics, 3-18(1981)

[5] Fama, E.F. and K.R. French ”Common risk factors in the returns on stocks and bonds”, Journal of Financial Economics, 33, 3-56(1993)

[6] Jagadeesh, N. and Titman, S. ”Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency.”, Journal of Finance 48, 6591(1993)

[7] Ross, S. ”The Arbitrage Theory of Capital Asset Pricing”, Journal of Economic Theory 13, 341-360(1976)