

Contest

東京理科大学工学部経営工学研究科

新田 翔

東京理科大学工学部第一部経営工学科

黒木 裕鷹

平成 29 年 7 月 19 日

概 要

Bloomberg Global Investment Contest においてルールの下で結果を出すため、近年盛んに研究されてきたマルチ・ファクターモデルを用いた。しかし既存のマルチ・ファクターモデルには数理統計学の観点での問題点がしばしばあった。ファクターの数を増やすことによる多重共線性の問題やオーバーフィッティングの問題である。時系列データの性質上、新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、近年パターン認識の分野でしきりに利用されている lasso 回帰と交差検証法 (cross validation) を用いることとした。また、定性的な議論により独自のファクターも併せて取り入れた。

コンテストと同様のルール下で行ったバックテストでは、独自のマルチ・ファクターモデル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、同様の戦略によりポートフォリオを提出した。

第1章 ルールに基づいた戦略の決定

1.1 ルールの概観

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か月単位のパイアシンドホールドを強いられ，さらにルール3-bにより裁定機会が失われることになる．つまり，金融資産の価格変動は何を根拠として起こっているのか，という視点で考える必要があり，それに基づいて価格が上昇する銘柄を追い求めていくことになる．次の1.2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から，その価格は需要供給の関係だけでなく将来などの情報にも影響される．つまり，「情報」が価格に織り込まれるのである．すると，未だ織り込まれていない「情報」を見つけ出して特定の銘柄を購入し，織り込まれた時点で売却する，という戦略が考えられる．この手法は一見有効であるように思えるが，ここには一つ盲点ある．それは，証券アナ

リストや裁定取引を行う投資家をはじめとする高度に習熟したプロフェッショナルの存在である。習熟したプロ達はいち早く織り込まれていない情報を日々見つけ出し、ポートフォリオに組み込むことを目標にしている。アナリスト達が織り込んでいない情報を見つけ出す手段として、アナリスト達が注目していない、またはポートフォリオに組み込めないような市場や超小型銘柄に焦点を当てる方向が考えられる。しかしルール 1-c, 2-a, 3-b の存在により、この手法も困難であることが分かる。織り込まれていない情報が織り込まれたとして、8 月 31 日までその価格が継続する可能性や測定期間内にその情報が織り込まれる可能性が高くないからだ。またリバランスを行うことにより利益を確定させても、その先 1 か月以上で同様のリスクを背負わなくてはならない。

織り込まれていない「情報」に着目するような投資戦略を避けるとすれば、他にどのようなアプローチが考えられるだろうか。次の 1.3 節では別の見方から価格変動を眺める。

1.3 価格変動の共通要因

世界には無数の企業が存在し、それぞれが異なった性質を持っている。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のうち、株価の変動に関係を持つものも当然存在するだろう。Fama と French 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。というのも、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ということだ。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、代表的なアノマリーとして認識されている。先ほどのモデルは Fama-French の 3 ファクターモデルと呼ばれ、両者はその功績により 2013 年にノーベル経済学賞を受賞した。実際に似たような性質 (業種や企業規模など) の銘柄は似たような価格変動を見せることが多く、Fama と French が対象とした米市場に限らず、金融資産の価格変動が共通の要因に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。

このような価格変動の共通要因の存在は、平均的にリターンを得る投資戦略の足掛かりとなるのではないだろうか。

1.4 様々な理論

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のはウィリアム・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続けてきた。2 章で触れるが、CAPM が必要とする仮定が非常に限定的であり、到底成立するものではない、という主張である。そこで、本コンテストではより確からしい共通要因を考えるため、独自のマルチ・ファクターモデルを構築することとした。

詳細については後に触れるが、APT は CAPM と異なり、全資産の収益の同時分布が正規分布であることを必要しないのである。また、これも後述するが、APT はファクターモデルの基礎と捉えることが出来る。

1.5 戦略の決定

1.2 節で述べたように、これから価格に織り込まれる「情報」を追うのではなく違う視点から戦略を決定する必要がある。また 1.3 節で述べた、Fama-French の 3 ファクターモデルをはじめとするマルチ・ファクターモデルにより、CAPM よりも確からしい資産の価格変動の構造を考えることができる。そこで独自のマルチ・ファクターモデルを構築し、「直近で平均的に勝っ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。なお、本コンテストは 2 カ月間のみのものであり、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何の利点も考えられないため、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。

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。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、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。

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。

第2章 理論

2.1 シングル・ファクターモデルとしてのCAPM

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、以下のように表される。また、 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られる。

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

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(??)より r_i の分散を求めると

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

となり、この右辺第1項はシステマティック・リスクと呼ばれ、シングル・ファクターであるマーケットの動きにより説明可能な部分である。また右辺第2項はマーケットに依存しない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。

次に式(??)により記述される資産で構成したポートフォリオを考える。市場を構成する資産数を N とすると、資産 i ($i = 1, \dots, N$)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。

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

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。

2.2 CAPM に対する批判

1.4 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)
ニューヨーク証券取引所において、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。
- Rolf Banz(1981)
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, Titman(1993)
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。

式 (??) の第 2 項であるアンシステムティック・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 ε_i が企業ごとに独立ではなく、共分散が存在しているということである。つまりこのことは、マーケットではない共通要因も存在し、それらがアノマリーの原因となってい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 (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と解釈された。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??にそのイメージ図を示した。

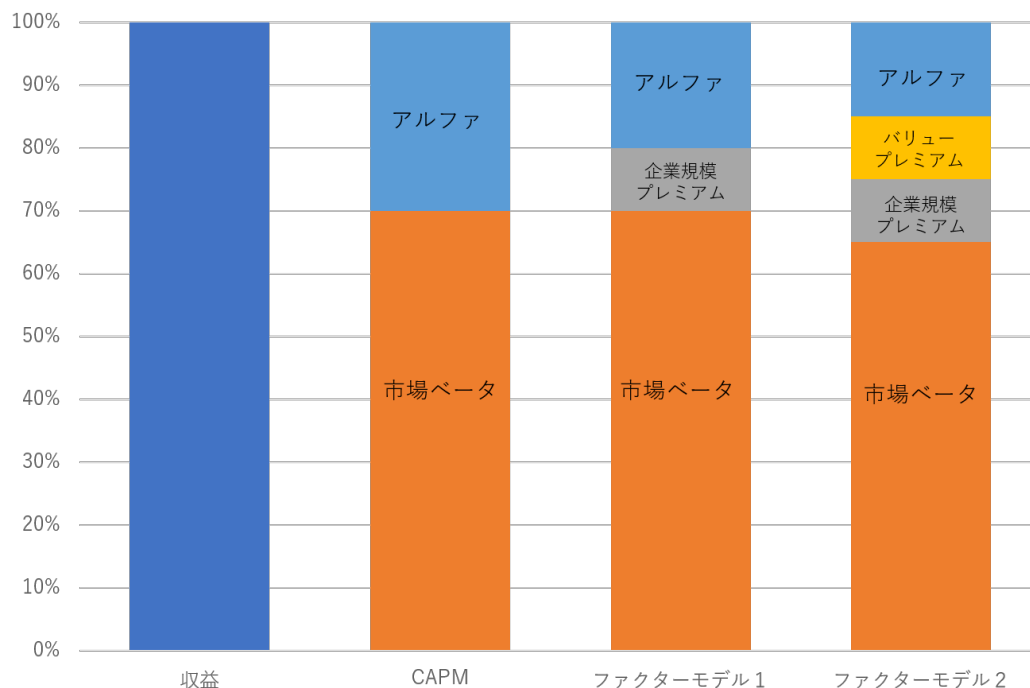


図 2.1: ベータの移り変わり

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 $F_k (k = 1, \dots, m)$ 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 は以下のよう表される。

$$r_i = \alpha_i + \beta_{i,1}F_1 + \beta_{i,2}F_2 + \dots + \beta_{i,k}F_k + \dots + \beta_{i,N}F_m + \varepsilon_i \quad (2.5)$$

ここで、2.1 節と同様、資産 $i (i = 1, \dots, N)$ 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 を考えると以下ようになる。

$$\begin{aligned} R_P &= \alpha_P + \beta_{P,1}F_1 + \beta_{P,2}F_2 + \dots + \beta_{P,m}F_m + \varepsilon_P \\ &= \sum_{i=1}^N \alpha_i + F_1 \sum_{i=1}^N w_i \beta_{i,1} + \dots + F_m \sum_{i=1}^N w_i \beta_{i,m} + \sum_{i=1}^N w_i \varepsilon_i \\ &= \sum_{i=1}^N \alpha_i + \sum_{k=1}^m \sum_{i=1}^N w_i \beta_{i,k} F_k + \sum_{i=1}^N w_i \varepsilon_i \end{aligned} \quad (2.6)$$

2.4 Fama-French の 3 ファクターモデル

数あるマルチ・ファクターモデルのうち、最も知られているものは Fama-French の 3 ファクターモデルだろう。このモデルは市場ポートフォリオ、時価総額、簿価時価比率の 3 つをファクターとしている。数式としては、以下のように表されることが多い。

$$r_i - r_f = \beta_i^{MKT}(R_M - r_f) + \beta_i^{SMB}SMB + \beta_i^{HML}HML \quad (2.7)$$

ここで、 r_f は無リスク利子率であり、長期国債の年利が用いられることが多い。また SMB は Small Minus Big の略であり、時価総額下位 50%の時価総額加重平均ポートフォリオから上位 50%の時価総額加重平均ポートフォリオを引いたものである。つまり、企業規模の小さい銘柄が大きい銘柄に対して余剰に得ている収益を表すファクターである。次に HML であるが、これは High Minus Low の略であり、簿価時価比率上位 30%の時価総額加重ポートフォリオから下位 30%の時価総額加重ポートフォリオを引いたものである。つまり、総資産額に比べて株価が割安になっている銘柄が割高な銘柄に対して余剰に得ている収益を表すファクターである。

2.5 APT

2.6 ファクターのリスク・プレミアムの算出

資産 i の収益率が式 (??) で表されているとする。このとき資産を組み合わせ、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

$$R_{P_1} = \alpha_{P_1} + 1 \times F_1 + 0 \times F_2 + \dots + 0 \times F_m + \varepsilon_{P_1} \quad (2.8)$$

と表せる．このとき， α_{P_1} が第 1 ファクターのリスク・プレミアムとなる．

次に，各ファクターのリスク・プレミアムを行列計算により求めることを考える．なお 2.3 節と同様， N 資産に対する m ファクターモデルを考える．第 k ($k = 1, 2, \dots, m$) 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 (2.9) のようにまとめることが出来る．

$$\begin{pmatrix} w_{1,1} & w_{1,2} & \dots & w_{1,m} \\ w_{2,1} & w_{2,2} & \dots & w_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n,1} & w_{n,2} & \dots & w_{n,m} \end{pmatrix}^T \begin{pmatrix} \beta_{1,1} & \beta_{1,2} & \dots & \beta_{1,m} \\ \beta_{2,1} & \beta_{2,2} & \dots & \beta_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{n,1} & \beta_{n,2} & \dots & \beta_{n,m} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} \quad (2.9)$$

式 (2.9) 左辺の $\beta_{i,k}$ による感度行列は式 (2.9) 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 B^\dagger を用いることにより解決する．

第3章 実際のデータ分析

3.1 使用したデータ

以下のデータを2015年1月～2017年5月の期間で、日次終値の形で取得した。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。

- 資産価格について

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。

- バリュウ・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、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 系列を正規化するのみに留まった。

3.2 データのクリーニング・概観

対象期間とした 2015 年 1 月～2017 年 5 月において、何らかの理由 (そもそもまだ起業されていない、上場していない、等) により価格データが欠けている企業が見受けられたため、当該企業を除外した。残った 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冗長であるため割愛した。使用するファクターについてのみ時系列プロットを行い、以下の図??に示した。また、多重共線性などの問題が存在するならばそれを回避するため、散布図とヒストグラム、相関行列を図??にまとめた。図??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、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。

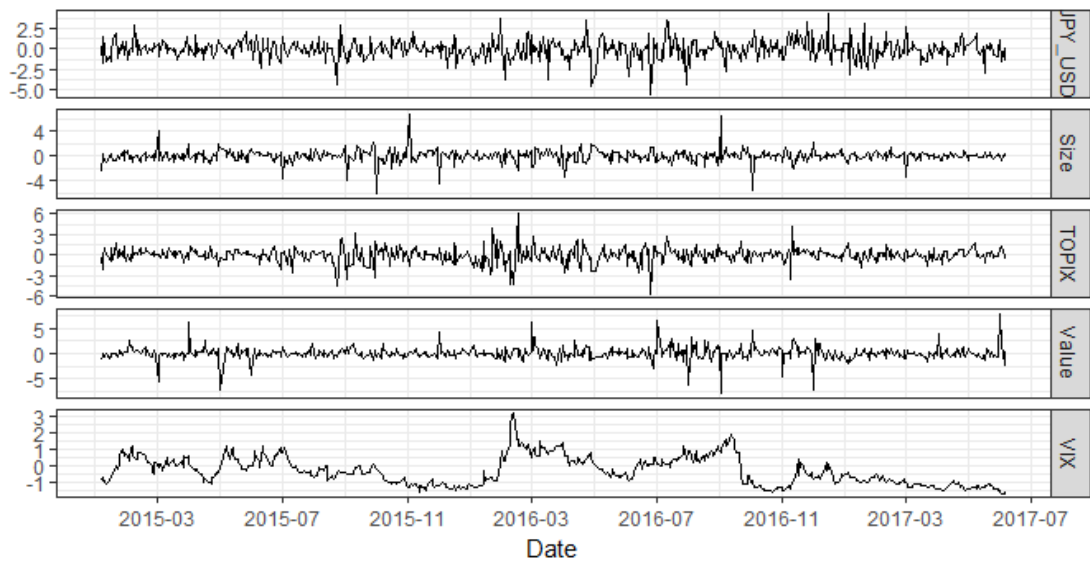


図 3.1: ファクターの時系列プロット

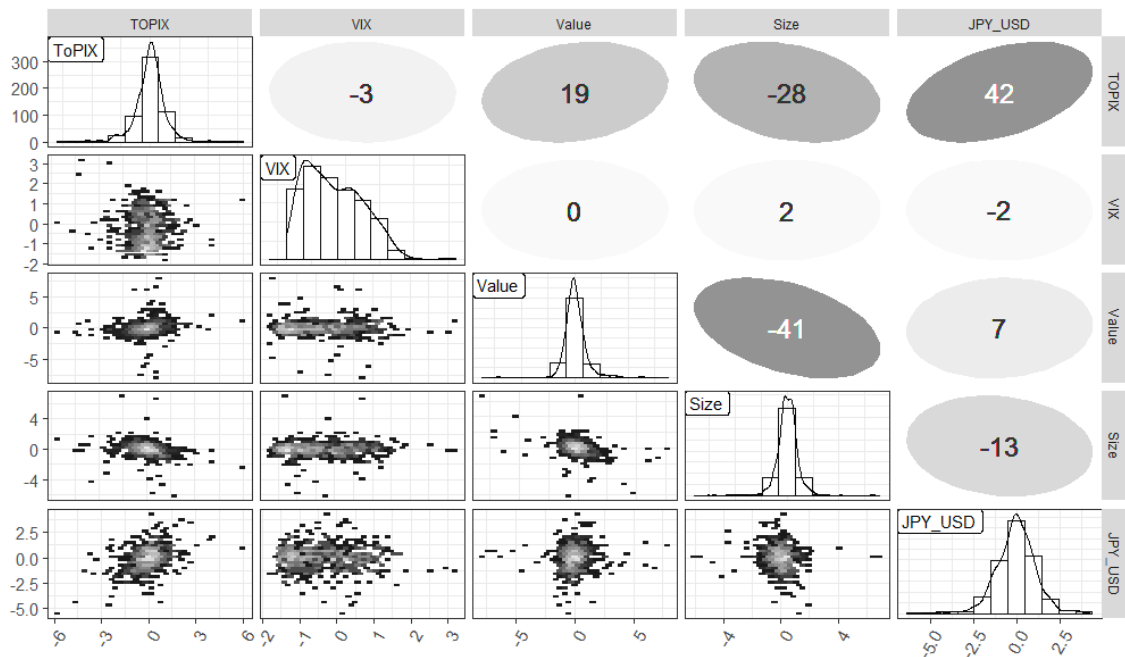


図 3.2: ファクター間の関係

図??のうち TOPIX, Value, Size, JPY_USD はどれも定常のような動きになっていることが分かる。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、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、ボラティリティの大きさをファクターとして扱いたいために原系列をそのまま (差分などを取らずに) 正規化

した。そのため明らかに非定常過程になっており、直近半年ほどは非常に低い値をとっていることが分かる。

次に図??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。このことは TOPIX, Value, Size 同士の散布図が中心付近に集まっており、十字型の広がりを持つことから確認できる。図??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、Value と Size、TOPIX と Size である。相関を持つ説明変数を使用しての回帰を行うといわゆる「多重共線性」の問題が発生し、実際以上に決定係数が大きくなってしまふなどの不都合がある。よってファクター全てを使つてのモデル推定は不適当であると考えられ、次節で述べるように、モデルの推定方法に工夫を施す必要がある。

3.3 ファクターモデルの推定

手に入れやすく扱いやすい日次終値データを対象としていることに加え、時系列データの性質により、推定するパラメータ数に対し十分な標本が得られないという問題がある。また市場の性質が不変であるとは考えにくく、ゆるやか(もしくは何らかの偶発的なショックにより急激に)変化していくものと思われる。そのためバックテストも見据え、3 カ月分のデータを用いてモデルを推定し、その先 1 カ月を対象にバックテストを行うこととした。

マルチ・ファクターモデルを推定する際、Fama-French の 3 ファクターモデルや Carhart の 4 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小さく、ここでは線形モデルのみを考えた。

線形でのマルチ・ファクターモデルを推定する際にも注意しなければならないことがある。むやみに説明変数(ここでは各ファクターを指す)を増やすことにより生じる多重共線性の問題や多重性の問題、みせかけの回帰の問題である。そこで、説明変数を全て使用しての線形回帰ではなく、何らかの工夫を施すことによりむやみにモデルを複雑にしない操作が必要になる。選択肢としては以下の手法が考えられる。

- AIC や BIC などの情報量基準を用いたモデル選択
- リッジ回帰や lasso 回帰などの、正則化項を用いたモデル推定

最終的には lasso 回帰によるモデル推定を行うこととした。詳細は割愛するが、AIC などの情報量基準は用いる説明変数をあらかじめ決め、選択された変数に基づいてモデルの良し悪しを示す値を計算することになる。つまり今回の場合 5 つのファクターを用いているので、一期間の一銘柄に対し $2^5 = 32$ 通りのモデル推定を行わなければならない。全ての銘柄に対し全ての期間でモデル推定を行うと約 160 万回の推定が必要になる。これはいわゆる「組み合わせ爆発」の問題であり、計算時間の観点から現実的ではない。これを回避する手法の一つに変数減少法がある。始めは全ての変数を用いモデル推定を行い、そこから一つずつ不適切な変数をふるい落としていく手法である。しかし、変数減少法を用いても局所的極小や過大な計算時間の不安はまぬがれない。

情報量基準の立場に対し、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．

3.3.1 lasso 回帰について

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「いらぬ変数」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． $t = 1, \dots, T$ 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．

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

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．

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に示した．具体的には，2.6 節の式 (??) 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．

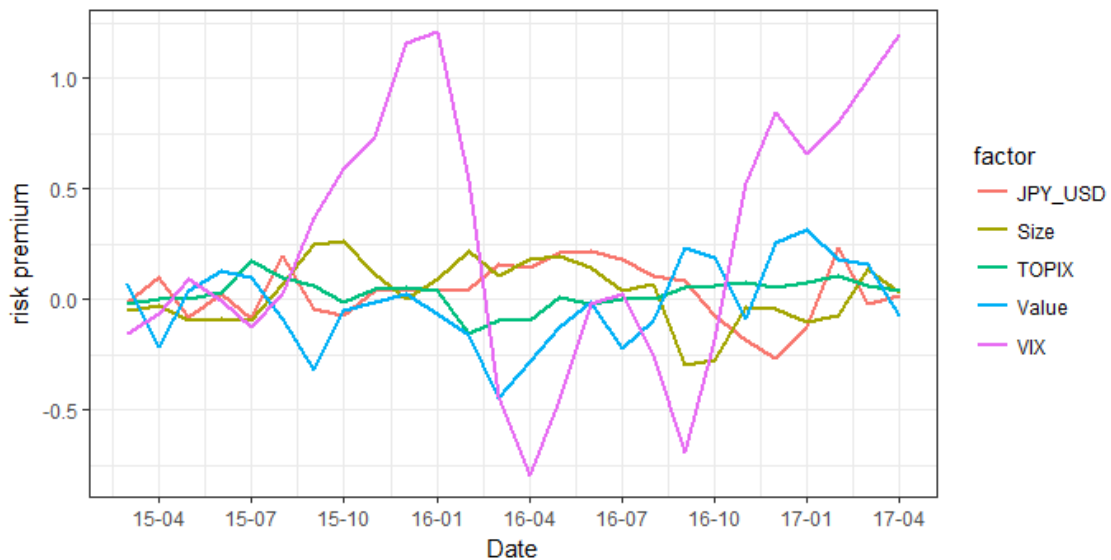


図 3.3: リスク・プレミアムの推移

3.5 様々なポートフォリオ

3.3 節で各銘柄のファクターモデルを推定し、3.4 節にはファクターごとのリスク・プレミアムを算出した。この節では様々なポートフォリオを構築し、バックテストでその性能を確かめると同時に、マルチ・ファクターモデルにおけるリスク・プレミアムが真に価値を持ったものなのかどうかを確かめる。なお、バックテストにおけるリバランスはコンテストを想定し、毎月初に行うこととした。

3.5.1 各ベータの上位 30 銘柄による等加重ポートフォリオ

直近 3ヶ月のデータより推定したマルチ・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、その累計リターンを以下の図??に示した。また、マーケットに対してのアウトパフォームを確認するために、TOPIXの累計リターンを benchmark として共に示した。

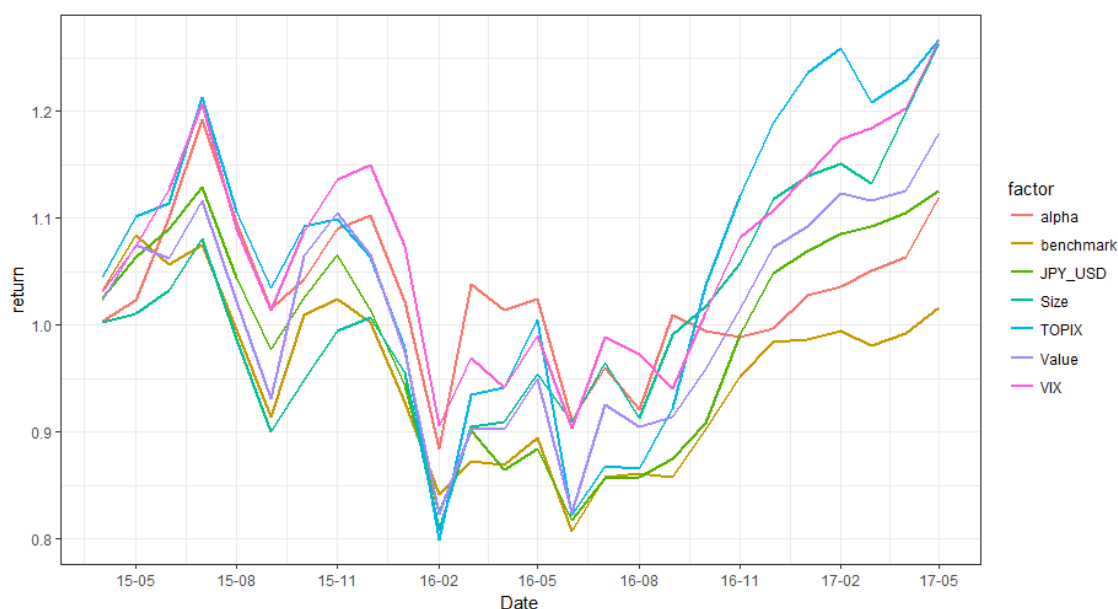


図 3.4: 各ベータ上位 30 銘柄による等加重ポートフォリオ

図??より、どのポートフォリオも似たような動きをしており、マーケット・ファクター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、マーケット・ファクター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

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が、ポートフォリオの構成銘柄は 10~30 銘柄でなくてはならない。銘柄数を考慮しなければ、リスクプレミアムの合計を目的関数とした最適化問題を考えればよいことになる。

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 (??) 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。

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

いま、ファクター F_1, \dots, F_5 のリスク・プレミアムが $\lambda_1, \dots, \lambda_5$ で与えられているとすると、ファクターのリスク・プレミアムを重みとした以下の最適化問題を考えるのが自然である。

$$\begin{aligned}
\text{maximize : } & \beta_{P,1}\lambda_1 + \beta_{P,2}\lambda_2 + \beta_{P,3}\lambda_3 + \beta_{P,4}\lambda_4 + \beta_{P,5}\lambda_5 \\
&= \sum_{k=1}^5 \sum_{i=1}^N w_i \beta_{i,k} F_k \\
\text{subject to : } & \sum_{i=1}^N w_i = 1 \\
& w_i \geq 0 \quad (i = 1, \dots, N)
\end{aligned} \tag{3.3}$$

式 (??) の最適化問題のうち、 $F_k, \beta_{i,k}$ は推定済みであり、 w_i のみを変数である。よってこれは線形計画問題となり、シンプレックス法などのアルゴリズムにより簡単に解くことができる。

次に銘柄数を考慮する。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。

$$\begin{aligned}
\text{maximize : } & \beta_{P,1}\lambda_1 + \beta_{P,2}\lambda_2 + \beta_{P,3}\lambda_3 + \beta_{P,4}\lambda_4 + \beta_{P,5}\lambda_5 \\
&= \sum_{k=1}^5 \sum_{i=1}^N w_i \beta_{i,k} \lambda_k \\
\text{subject to : } & \sum_{i=1}^N w_i = 1 \\
& w_i \geq 0 \quad (i = 1, \dots, N) \\
& 10 \leq \sum_{i=1}^N x_i \leq 30
\end{aligned} \tag{3.4}$$

最後の式が加わったことにより、厳密解を求めることは非常に難しくなった。そこで、特定の最適化問題に依存しないメタ・ヒューリスティクスによる解法が考えられるが、計算量が多く、全ての期間でのバックテストが間に合わなかった。そのため、以下の手法によるポートフォリオ選択を考えた。

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、直近3ヶ月のモメンタムが高い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

第4章 途中經過

4.1 提出したポートフォリオ

企業名	ウェイト (%)	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
G S I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
ソフトクリエイティブホールディングス	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
S O M P O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

4.2 7月中のパフォーマンス

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つた。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つた。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

3.5 節のように、ファクターのリスク・プレミアムに基づいて銘柄ごとのリスク・プレミアムを算出し、上位 30 銘柄によるポートフォリオを構築した。しかしバックテストにおけるそのリターンは思わしくなく、予想とは異なる結果であり、追加の考察が必要であると感じた。

第6章 リバランス

第7章 感想