

状態空間モデルの拡張による投資信託商品の要因分析

大西 恒彰[†] 馬 強[†]

[†] 京都大学情報学研究科社会情報学専攻

〒 606-8501 京都市左京区吉田本町

E-mail: [†]onishi@db.soc.i.kyoto-u.ac.jp, ^{††}qiang@i.kyoto-u.ac.jp

あらまし NISA 制度が導入され、貯蓄から投資への誘導が進んでいる。中長期的な商品である投資信託は NISA に適しているが、似たような商品が多く、それぞれの商品に影響を及ぼす要因を理解するのが難しい。そこで、本稿では、投資信託の理解支援のため、月次報告書と基準価額の変動を用い、投資信託商品の要因分析手法を提案する。我々は、投資信託商品の基準価額に及ぼす要因を「マクロ要因」と「ミクロ要因」に分類して、その影響度合いを状態空間モデルを用いて、時系列的に分析する手法を提案する。このモデルを用いて、従来のリターンリスクといった数値的な比較では得られない知見、例えば、基準価額の上下の原因を知ることができ、投資信託商品のより一層の理解を支援できる。

キーワード 理解支援 要因分析 状態空間モデル

1. はじめに

近年、NISA(少額投資非課税制度)といった証券優遇制度が普及し、ますます多くの人が投資を行うようになった。特に、投資信託商品は、専門知識がなくても投資が出来るため、金融商品として大変人気がある。しかし、投資信託協会の投資信託に関するアンケート調査報告書 [1] では、投資信託の現在保有率は 16%に留まり、非保有者の多くは投資信託商品の理解度に不安を感じ、投資に踏み切ることができていないことが報告されている。さらに、投資信託商品に伴う月次報告書は、運営会社から公開される信頼性のある価値が高い情報であるが、投資者がそれを詳しく読む割合と全く読まない割合はそれぞれ 8%と 52%となっており、適切な情報利用はできていない。

我々は、利用者の投資情報に対する要求を明らかにするため、クラウドソーシングを用いて、投資経験が異なる 182 名のワーカーに対して調査を行った。調査から彼らが基準価額の変動要因の複雑性に難しさを感じていること、経済全体に影響を与えるマクロ要因だけでなく、個別企業のニュースのようなミクロ要因も理解したいことが分かった。

市場の変動要因理解支援に関する様々な研究が行われている。既存研究の多くは、マーケット全体に影響を与えるマクロ的な要因を重視している。しかし、調査結果で重要性が明らかになったミクロ要因に関する研究がまだ少ない。また、既存のテキストマイニングによる要因分析の手法の多くは、要因の影響度の定量分析を行っていない。さらに、影響の時系列変化もあまり考慮されていない。そのため、同じ要因からもたらされた、基準価額への影響の度合いは、投資信託商品ごとに異なることや、投資信託商品の構成銘柄の入れ替えによって変化することを明らかにできず、投資信託商品の要因理解支援には不十分である。

そこで、我々は、運用報告書やニュースなどのテキストデータと基準価額などの数量データを併用して、投資信託商品の要

因を定量的に分析する手法を提案する。

本研究では、テキストデータとして、投資信託商品の運用報告書である月次報告書と日経 QUICK ニュース社から提供されたニュース記事を用いる。我々は、これらのテキストデータから要因を抽出し、日経シソーラスを用いて分類する。本研究では、日次データと関連づけることが可能である「マクロ変数要因」、経済指標や国内外の政策のようにある時点から影響を与える「マクロ干渉要因」、業績などの個別企業に関わる「ミクロ要因」に分類する。

そして、これらの要因を説明変数とし、状態空間モデルを構築して、要因の影響度を推定する。

本研究における我々の主な貢献は以下の通りである。

- 状態空間モデルを構築するための要因の選択・分類手法を提案する。テキストデータから基準価額に影響を与えた要因を抽出し、日経シソーラスを用いて、要因を「マクロ変数要因」「マクロ干渉要因」「ミクロ要因」に分類することで、基準価額変動要因を理解するための状態空間モデルを構築手法を提案する (4.4.2 節)。また、変数選択では多重共線性を考慮して、モデル構築に適切な必要な要因を選択する (4.4.5 節)。

- 要因の時系列変化を考慮した状態空間モデルを構築するために、動的計画法を用いた基準価格のトレンド分割を提案する。状態空間モデルをトレンド局面ごとに構築することで、その局面で影響を与えている要因のみで分析を行い、要因の影響度をより正確に推定し、要因の影響の時系列変化を明らかにすることが可能となる (4.4.4 節)。

- 2016 年 1 月から 10 月の期間の国内の株式を対象とした投資信託商品を用い、実験およびケーススタディを行った。実験では、要因の抽出および分類を高い精度で行えることを確認した。また、トレンドごとに状態空間モデルを構築した際の状態空間モデルの誤差が、トレンドを考慮せず構築した状態空間モデルよりも優れていることを確認した。また、ケーススタディを通して、提案手法の有用性を明らかにした (5 節)。

本稿の構成は次の通りである。2 節で関連研究について整理し、3 節では、投資信託商品の選択に関して意識調査について述べる。4 節では、実際の月次報告書と日経クイックニュースを用い、要因分析のための状態空間モデルの構築手法について説明する。5 節では、実験結果を用いて提案手法の検証と議論を行う。6 節では、今後の課題および結論を述べる。

2. 関連研究

市場に影響を及ぼす多様な情報の全てに目を通し、意思決定を行うのは極めて困難である。そのため情報技術を用いた市場分析は一定の成果が報告されている [2] [3] [4] [7] [6] [8]。特に近年は、数値のみでくみ取れない情報を得るために、分析対象としてテキスト情報を扱う研究が多くみられるようになった。とりわけ株式市場に関し、数多くの研究事例が報告されており、予測の分野、市場理解において成果を発揮していることが確認されている。

Kearney らによるとテキスト分析に用いられるテキスト情報は、メディアが配信するニュースや記事、企業や政府が配信する情報、ネットにおける投稿と 3 つに分類することができる [2]。それぞれ専門性、スピード、信頼性、情報の量で異なる特徴を持っている。本研究で我々が用いる月次報告書は企業や政府が配信する情報に該当し、専門性、信頼性が高い情報である。

酒井らが業績を決算短信文章から抽出する手法を提案している。彼らはポジティブ・ネガティブ表現を「により」や「から」といった手がかり表現により収集することで、要因抽出の精度を向上させている [3]。また、坂路らは機械学習手法を用い、原因と結果を含む文を抽出している [4]。上述の研究は、金融関連テキストにおいてテキストデータの因果関係を適切に把握する際に極めて有用であり、投資信託商品の月次報告書からの因果関係抽出においても効果があると考えられるが、月次報告書は要因のみの記述である場合が多く、結果事象である基準価額の変動に関する記述が少ない。そのため、因果ペアを想定した彼らの抽出がうまくいかない場合がある。そこで本研究では情報抽出によく利用されている CRF(Conditional Random fields) [5] 手法を用いて要因抽出を行う。

ニュースの分析では、Tetlock らが、WSJ のコラムが株価に対し、有意な影響を持つかどうかを検証している。彼らによると、ニュースは短期的にニュースのポジティブ・ネガティブの方向と同じ影響を受けることが確認できたが、その影響が時間とともに失われることを確認した [7]。一方、沖本らが同様の検証モデルを日経 QUICK ニュースを用いて分析したところ、長期的な影響の減衰は見られないという異なる結論を得ている [6]。彼らは扱うニュース媒体の違いにより、異なる結論が得られたと報告しているが、彼らのモデルは関係性があると考えられるマクロ変数を考慮しておらず、モデル構築に問題がある可能性がある。また彼らはニュースをポジティブニュースおよびネガティブニュースの量をもとにニュース指数を設計し、検証をおこなっている。そのため、個別のニュースごとに検証はできておらず、どのニュースが影響しているか考察できていない。

和泉らはテキストに含まれる豊かな情報が、市場構造をより

適切に表現できると報告している。彼らは日銀の金融経済月報において CRP と呼ぶ共起解析、主成分解析、回帰分析を用いた特徴抽出で長期的な市場分析において数値情報だけの予想よりも高い予測力が得られたことを示した [8]。

これらの先行研究は、テキストデータを活用することが、金融市場理解において有用であることを示している。我々は、金融市場全体に対して要因が有意に影響を与えたかの市場理解の議論ではなく、個別の投資信託商品それぞれにおいて有用な要因を提示することに着目している。

3. 利用者調査

3.1 投資信託商品の意識調査

我々は、クラウドソーシングを提供するランサーズ^(注1)にて、投資信託商品の理解に関する調査を行った。

この調査では、投資信託商品の基準価額変動に影響を与える要因に関する意識について調べた。投資に興味がない人 29 人、投資を始めようとしている人 61 人、投資初心者 15 人、投資経験者 46 人、投資を引退した人 31 人の合計 182 人を対象に調査を行った。

調査では、月次報告書とともに日経クイックニュースを見てもらい、いくつかの商品の理解をした後、以下の設問に答えてもらった。

- (問一) 商品選択の際、理解が難しい点
- (問二) 投資信託商品の価格変動の理解でどの要因まで確認したいか

第一の設問では、投資信託商品において難しさを感じる点を調査した。その結果を図 1 に示す。興味がない人は、投資信託商品において経済用語や専門用語の難しさを大きく感じていること、それ以外の層は、基準価額に影響を与える要因の複雑さに難しさを感じていることが分かった。また、30 人と非常に多くの人々が、情報収集自体に難しさを感じている。投資に関する情報収集する際は、どのタイミングで情報を見ればいいのか、情報の多さ、誰を信頼していいかなど、非常に難しい情報の取捨選択を行う必要があり、投資を難しくさせている大きな一因になっていると考えられる。

第二の設問では、投資信託商品の価格変動の理解でどの要因

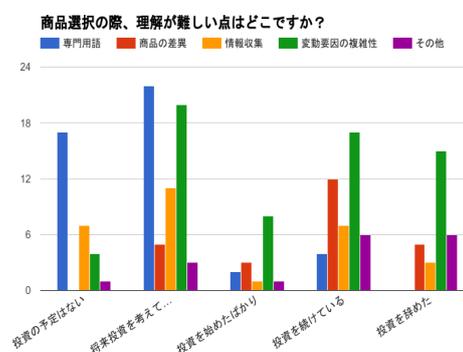


図 1 投資信託商品において難しさを感じる点

(注1) : <http://www.lancers.jp/>

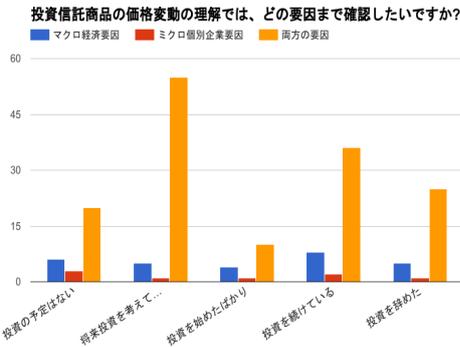


図2 投資信託商品の価格変動の理解でどの要因まで確認したいか

表1 評価値と言及回数の相関係数

| | ひふみ投信 | DIAM 新興市場投信 | J-Open 投信 |
|------|-------|-------------|-----------|
| 相関係数 | 0.961 | 0.856 | 0.828 |

まで確認したいかを調査した。その結果を図2に示す。投資経験を問わず、投資信託の基準価額の変動をマクロ、ミクロの両面の情報から理解したいと考えていたことが分かった。しかし、現状の投資信託における情報提示では、マクロ経済環境を重視した報告がなされているため、ミクロ要因においてもポイントを絞った情報提示を行う必要があると考えられる。

3.2 マクロ要因の評価

投資を将来的に考えている大学院生3名に、2015年の月次報告書に現れるマクロ要因を基準価額への影響度の観点から評価してもらった。被験者には、5段階評価を用いて、それぞれの投資信託商品に対し、要因の評価をしてもらう。

評価値と言及回数の相関係数を分析してみたらとところ表1のように高い相関係数が見られた。これより、報告書内で言及回数の多い要因を基準価額に影響を与える要因だと考える傾向にあることが確認できる。これより、投資になじみのない人は、影響の大きい要因を誤って判断する可能性があると考えられる。また同じマクロ要因を上位の影響要因として考えており、マクロ要因による投資信託の選別や差別化は難しいことを確認した。

これらの調査結果から、基準価額に影響を及ぼす要因分析の重要性と有用性が明らかになった。

4. 要因分析

4.1 概要

基準価額の変動理由を分析することで、従来のリターン・リスクのみによる商品選択とは異なる、影響要因による商品の理解・選択ができるようになって考えている。

本研究で提案する要因の定量分析手法は以下のステップに分けられる(図3)。

Step1 (要因の抽出) 基準価額に影響を与える要因を月次報告書および日経クイックニュースから抽出する。まず、月次報告書の市況概況からマクロ要因を抽出する。さらに、商品を構成する銘柄に関するニュースを日経クイックニュースよりミクロ要因として抽出する。抽出の際には、その要因がどの時点の要因かを特定するために時間表現やタイムスタンプを同時に取得している。

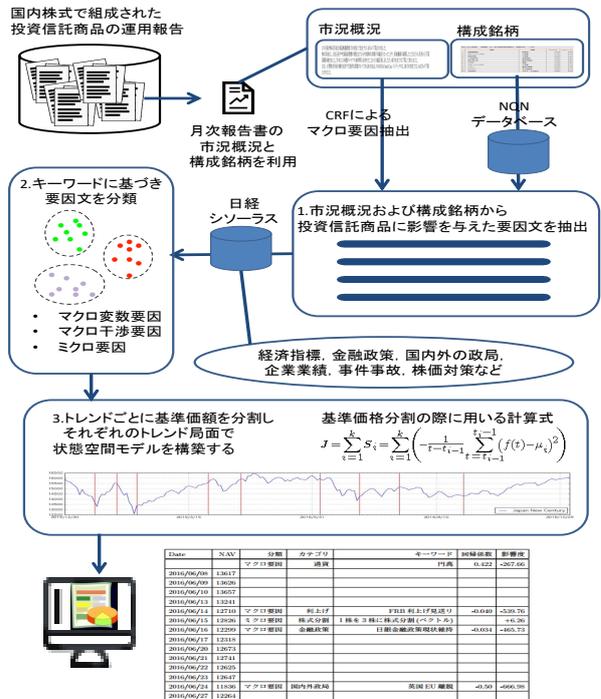


図3 要因分析の処理概要図

Step2 (要因の分類) 抽出された要因文を日経シソーラスを用い分類する。分類後はカテゴリごとにマクロ変数要因、マクロ干渉要因、ミクロ要因として扱う。

Step3 (状態空間モデルの構築) 要因の影響度を定量的に推定するための状態空間モデルを構築する。また、推定される回帰係数から基準価額に与える影響度を求める方法について述べる。状態空間モデルは、基準価額のトレンド局面ごとに構築するものとし、そのトレンド期間で影響のある要因だけでモデルを構築する。変数選択では、多重共線性を持たないように注意し、高い相関を持つ変数であれば、変数選択を行っている。

4.2 要因の抽出

基準価額が高くなったり安くなったりするのは、様々な要因が複雑に影響しあった結果である。我々は月次報告書を利用し、影響を与えた要因の抽出を行う。月次報告書にはファンドの運用状況を確認するために必要な情報、基準価額の推移、基準価額の変動する要因などが記載されており、月に一度と情報の更新頻度が高く、タイムリーな情報を得ることができることから極めて有用な情報源である。

4.2.1 月次報告書からの要因抽出

市況概況から投資信託商品に影響を与えた要因を抽出するために、条件付き確率場 CRF [5] を用いる。CRF は情報抽出でよく使われる手法であり、これを利用し、要因箇所と時間表現の抽出を行う。その前処理として、運用報告書の市況概況の文章に対し、日本語形態素解析器である Mecab [9] を用い、単語分割を行い、それぞれの品詞および品詞細分類を得る。

CRF の素性には品詞細分類、文字種「空白、数字、英字、漢字、ひらがな、カタカナ、その他」、および、要因箇所を表すラベルとして F 、時間表現を表すラベルとして T 、その他は

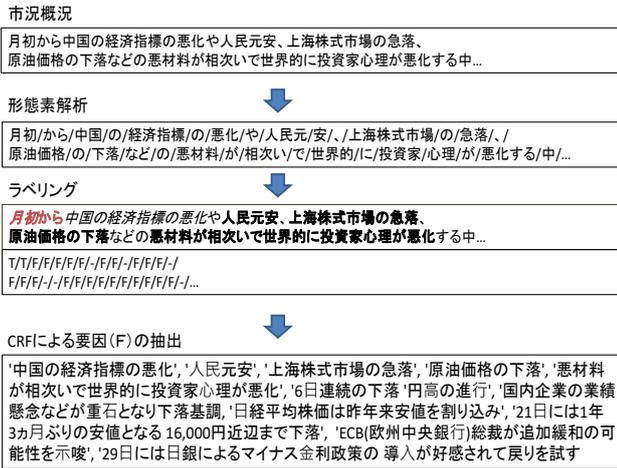


図4 CRFによる要因抽出の概要図

Oを用いた。

要因(F)の抽出の具体例が図4である。ラベルを推定する際は、前後2文字の単語、品詞細分類、文字種と前2文字のラベルタグを利用する。要因抽出および時間抽出の際は図4のようにFの開始文字から終了文字までを取得する。

4.2.2 ニュースからの要因抽出

月次報告書には組み入れ銘柄が十銘柄程度記載されており、その時点での構成比率が記載されている。市況概況および運用報告では、マクロ的な要因に焦点があたっており、個別銘柄の特徴的事象が記述されないことが多い。そこで、当該期間の日経クイックニュースよりこれらの銘柄に対応するマイクロ要因を抽出する。

我々は、各ニュースの見出しに企業名および銘柄略称が含まれているものをマイクロ要因として取得した。例えば、1月にトヨタ自動車の株を保有している投資信託商品の場合、1月1日から1月30日のニュースの見出しを確認し、トヨタ自動車を含むニュースを収集する。

4.3 要因の分類

この節では、抽出した要因を分類する手法を述べる。経済法令研究会によると株式市場の変動要因は、株式市場全体に関する要因と個別企業要因の2つに分けられる[10]。株式市場全体に関する要因には世界の株式市場、金利動向、為替などがあり、個別企業要因には、企業業績や株主構成の変化、資金調達などがある。本研究では、前者をマクロ要因、後者をマイクロ要因として扱う(表2)

表2 株式市場変動要因

| 株式市場全体に関する要因(マクロ要因) | 個別企業に関する要因(マイクロ要因) |
|---------------------|--------------------|
| 世界の株式市場 | 企業業績 |
| 金利・債券の動向 | 組織管理・人事管理 |
| 為替の動向 | 買収・提携・合併 |
| 商品価格 | 株価対策 |
| 経済指標 | 新商品・サービス |
| 国内外の政局 | 配当政策 |
| 金融政策 | 事件・事故 |

4.3.1 知識ベースの分類

我々は日経シソーラスを用いて、要因の分類を行う。日経シソーラスは日本経済新聞社が1982年から作成している新聞記事検索のための用語集であり、記事の出現頻度を考慮して選定された1万3千以上の用語はカテゴリ別に収録されている。本研究では抽出された要因を日経シソーラスに対応させることで分類を行う。

表2の各要因と関連性の高い日経シソーラスのカテゴリを選択し、カテゴリ別に収録されたキーワードを対応させた。分類では、要因文に日経シソーラスのカテゴリに含まれるキーワードが含まれていれば、要因文をそのカテゴリと対応させる。

例えば、「経済指標」というカテゴリには日経シソーラスのカテゴリ「調査・統計」の93キーワードが対応する。「米国雇用統計が予想以上に悪く」といった要因文の場合、雇用統計というキーワードを含むため、経済指標というに分類される。

また、「日経平均」と「日経平均株価」といった省略語による表記ゆれの問題はWeblio^(注2)を用い解決する。もし、抽出した要因文がどのカテゴリとも対応しなかった場合、カテゴリなしとして扱い、その後の分析で用いないものとしている。

さらに、時系列データを持つマクロ要因は、分析する際に日次データを用いることで、要因を正しく表現できると考え、要因を以下の3つに分類している。

・マクロ変数要因

「世界の株式市場」・「金利・債券」・「通貨・為替」・「商品価格」に属する要因であり、時系列データとして日次データと関連づけることが可能である。時系列データは日本銀行^(注3)やInvesting.com^(注4)のデータを用いた。

・マクロ干渉要因

「経済指標」・「国内外の政局」・「金融政策」に属する要因であり、時系列データに干渉し、影響を与える要因である。

・マイクロ要因

「企業業績」・「組織管理・人事管理」・「買収・提携・合併」・「株価対策」・「新商品・サービス」・「レーティング」・「事件・事故」に属する要因であり、個別企業要因である。

4.3.2 発生日の推定

この節では、要因の発生日を推定手法について述べる。要因の発生日を推定することで、影響を与えた期間を特定することができ、影響度の分析が正確になる。

まず、マクロ変数要因が時間表現と共に抽出されているならば、対応する期間を発生日とする。もし、同一の文章内で時間表現が抽出されなかった場合、前の文章内に出ている時間表現を利用する。それでも見つからない場合は、月を通して発生していたマクロ変数要因だと考える。

次に、マクロ干渉要因は、時間表現を用い、日経クイックニュースにてキーワードを探索し、発生日を決定する。例えば、10月の月次報告書にて「雇用統計」というキーワードにより

(注2) : <http://thesaurus.weblio.jp/>

(注3) : <http://www.boj.or.jp/>

(注4) : <http://jp.investing.com/commodities/crude-oil-historical-data>

経済指標に分類されている「前半、雇用統計の悪化～」のような要因文が抽出されているなら、10月の1日から15日において、最も「雇用統計」というキーワードが言及されている日を探す。すると10月7日にもっとも言及回数が多いことから、10月7日に起きた要因として考えることができる。なお、同一文章内で時間表現が抽出されていない場合は、マクロ変数要因と同様の処理を行う。

最後に、個別企業ニュースのミクロ要因の場合は、抽出時にニュースのタイムスタンプを得ることができるため、それを利用し、要因の発生日としている。なお発生時間が市場の後場終了後である場合は翌日に発生したものと処理をする。

4.4 要因分析のための状態空間モデル

4.4.1 概要

基準価額のような時系列データは、観測値が前の時点から影響を受けており、独立なデータではないことから、独立を仮定する通常の回帰モデルではうまくデータ分析を行うことができない。このような複雑な時系列の関係性を表現するために我々は状態空間モデルを利用する。

状態空間モデルは、様々な応用問題に利用されてきており、高い柔軟性を持つ分析手法である。本研究では、この状態空間モデルを使い、基準価額に対する各要因の影響度を推定することを考える。状態空間モデルは残差が等分散性であることを仮定しているため、基準価額やドル円や石油のような変数は推定時の残差が等分散になるように、対数値を用いる。

4.4.2 モデル定式化

マクロ変数要因、マクロ干渉要因、ミクロ要因を用いた基準価額への影響度の推定のための状態空間モデルの構築を行う。

$$y_t = \mu_t + \beta_{k,t}x_{k,t} + \lambda_{m,t}w_{m,t} + \sum_{l=1}^n s_{l,t} + \varepsilon_t \quad (1)$$

$$\varepsilon_t \sim NID(0, \sigma_\varepsilon^2) \quad (2)$$

$$\mu_{t+1} = \mu_t + \xi_t \quad \xi_t \sim NID(0, \sigma_\xi^2) \quad (3)$$

$$s_{l,t} = y_{t-1} \times c_{l,t} \times d_{l,t} \quad (4)$$

$$\beta_{t+1} = \beta_t \quad \lambda_{t+1} = \lambda_t \quad (5)$$

ただし、 y_t は t 時点における対数基準価額、 $x_{k,t}$ は t 時点におけるマクロ変数 k の対数値、 $w_{m,t}$ はある時点まで 0 をとり、要因発生時より 1 となるマクロ干渉 m である。また $s_{l,t}$ はある時点 t における、ある銘柄 l のニュースによる騰落率 d である $d_{l,t}$ とその銘柄の構成比率 $c_{l,t}$ を前日の対数基準価額値に乘算して求められるミクロ要因の影響度である。

未知パラメータの推定は最尤推定で求める。この場合、推定で求めるのは、観測値攪乱項 σ_ε^2 、レベル攪乱項 σ_ξ^2 、回帰係数 β 、 λ である。

4.4.3 要因の影響度推定

この節では、推定されたパラメータの解釈し、影響度を推定するための計算式を述べる。まず、マクロ変数要因に関して述べる。説明変数のあるモデルはすべての状態攪乱項をゼロにする式 (6) で表される。このとき、回帰係数 β_1 は弾力性として解釈することができる。

$$y_t = \mu_1 + \beta_1 x_t + \varepsilon_t \quad \varepsilon_t \sim NID(0, \sigma_\varepsilon^2) \quad (6)$$

一般的に弾力性は y のパーセント変化を x のパーセント変化で割ったものであり、式 (7) のように書くことができる。

$$S^* = \frac{x}{y} \frac{\partial y}{\partial x} \quad (7)$$

目的変数 y と説明変数 x が対数表示だとすると、本来の y と x の関係は以下のように書き直せる。

$$\log \hat{y} = \hat{\mu}_1 + \hat{\beta}_1 \log x \quad (8)$$

$$e^{\log \hat{y}} = e^{\hat{\mu}_1 + \hat{\beta}_1 \log x} \quad (9)$$

$$\hat{y} = e^{\hat{\mu}_1} + e^{\hat{\beta}_1 \log x} = e^a x^b = cx^b \quad (10)$$

このとき、

$$S^* = \frac{x}{cx} \cdot \frac{\partial cx^b}{\partial x} = \frac{x}{cx} \cdot \frac{cbx^b}{x} = b = \hat{\beta}_1 \quad (11)$$

となり、 $y = cx^b$ という曲線は固定弾力性を持つ。この場合、説明変数が 1% 下落したとき $\hat{\beta}_1\%$ 目的変数が変動することを示す。

次に、干渉要因による影響度について述べる。干渉変数のあるレベルモデルは、レベルと干渉変数の両方が確定的であるなら、式 (12) のように書ける。

$$y_t = \mu_1 + \lambda_1 w_t + \varepsilon_t \quad \varepsilon_t \sim NID(0, \sigma_\varepsilon^2) \quad (12)$$

干渉変数 w_t は対数ではないので、回帰ウエイト λ_1 の値は弾力性として捉えることはできない。そこで干渉変数の結果としての目的変数のパーセント変化は以下のように考える。 \hat{y}_{pre} を干渉以前の $\mu_1 + \lambda_1 w_t = \mu_1$ 、また \hat{y}_{post} を干渉後の $\mu_1 + \lambda_1 w_t = \mu_1 + \lambda_1$ とする。いま y_t を対数表示されたものとするならば、干渉による目的変数のパーセント変化は次のようになる。

$$100 \left(\frac{e^{\hat{y}_{post}} - e^{\hat{y}_{pre}}}{e^{\hat{y}_{pre}}} \right) = 100 (e^{\lambda_1} - 1) = z \quad (13)$$

となり、マクロの干渉により、目的変数に $z\%$ の変化があったと解釈することができる。

4.4.4 基準価額のトレンド分割

我々は、基準価格のトレンド局面ごとに状態空間モデルを構築することで、変動に影響を与えた説明変数を適切に選択してモデルを構築できると考えている。基準価額系列のトレンド分割を行う際は、動的計画法 [11] を適用する。本研究では、以下のように基準価額の系列をトレンドに分割する。

それぞれのトレンドごとに、近似関数としてトレンドの境界点を結んだ直線を用意し、基準価額との二乗誤差平均 S_i を得る。そして、スコアの合計値 $U = \sum_{i=1}^k S_i$ を最小化することで分割を求める (式 (14))。ここでは動的計画法を適用するため誤差の二乗誤差にマイナスを付加し、最大化の問題に置き換えている。

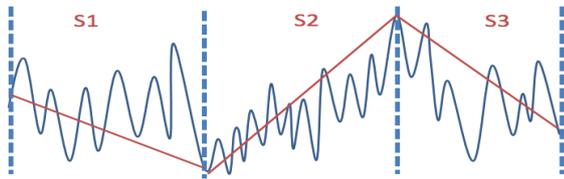


図5 基準価額のトレンド分割

$$U = \sum_{i=1}^k S_i = \sum_{i=1}^k \left(-\frac{1}{t-t_{i-1}} \sum_{t=t_{i-1}}^{t_i-1} (f(t) - g(t))^2 \right) \quad (14)$$

ただし、 $f(t)$ は営業日 t 日での基準価額、 k は分割数。基準価額との二乗誤差平均は S_i とする。 $g(t)$ は境界位置にある点同士を結んできた直線である。図5は分割の具体例である。

4.4.5 説明変数の選択

説明変数に極めて高い相関を持った変数の組があると、回帰係数の値が極めて高い値をとったり、回帰係数の正負が逆転するなどといった多重共線性の問題が起り得る。

本研究では、一般的に統計学において強い相関とよばれる0.7以上の相関を持つ変数の組がある場合のみ変数の取捨選択を行うことにする。強い相関がある場合は、その相関関係にある変数のいずれかの変数を捨て、分析を行う。変数の選択はAICの基準を用いて行う。

本研究で用いる状態空間モデルの評価のための赤池情報量基準(AIC)は以下ようになる。

$$AIC = \frac{1}{n} (-2n \log L_d + 2(q + w)) \quad (15)$$

ここで n は時系列の観測値数、 $\log L_d$ は状態空間モデルで最大化する尤度関数の値、 q は状態の初期値の総数、 w は推定される攪乱項分散の総数を示す。一般にAICが小さくなればなるほど、モデルの当てはめがよくできていることを示す。

5. 実験

5.1 概要

2016年1月から10月の月次報告書および日経クイックニュースを用い、提案手法の確認を行った。実験では、状態空間モデルを構築するために必要な要因の抽出と分類が正確に行っているか、また、提案した状態空間モデルの妥当性について検証を行っている。最後にケーススタディとして要因の影響度を定量化している。

5.2 データセット

5.2.1 ニュースデータ

本研究で使用したニュースデータは、2016年1月1日から10月31日までにQUICKで配信された日経ニュース53440件である。これらのニュースは、市況に対するマーケットコメントや企業に業績等の材料で構成され、信頼性の高さ、情報伝達の速さから証券市場に参加する多くの投資家に読まれるテキストデータである。

5.2.2 月次報告書

本研究で扱う投資信託商品の月次報告書は2016年1月から10月までの10ヶ月分である。これらの月次報告書より市況概況

と構成銘柄を抽出する。すべて国内株式に投資する投資信託商品であり、「Japan New Century」「DIAM Emerging Markets」「JPX Premium」「J-Open」「J-Frontier」「Index Fund 225」の6商品を用いた。

5.3 要因抽出・分類

CRFを用い、月次報告書の市況概況からマクロ要因を抽出する。CRFの学習は抽出する同一投資信託商品の月次報告書を用い、予測する当月以外の残り9ヶ月分を用いた。予想するラベルは要因(F)および時間表現(T)、その他(O)である。抽出はラベリングの精度を用い、評価する。

表3がラベリングの結果である。月次報告書の文章が平易で短い「Japan New Century」、「J-Open」は精度の良い抽出ができていた。しかし、月次報告書の文章が長く、記述内容が具体的な地名など詳細に書かれている「DIAM Emerging Markets」の抽出の精度は悪くなった。特に「英国のEU離脱の国民投票で、離脱派が勝利する」のように要因の間に句点が入るなどして、正しくラベリングできなかった場合がある。

これはCRFの学習データが少ないために適切な抽出が行えていないことから発生している失敗だと考えられる。そのため、扱う期間を拡大し、CRFの学習データに使える文章量を増やすと改善すると考えられる。

また、分類は以下の式で評価する。

$$Precision = \frac{\text{抽出した要因で正確に分類された要因数}}{\text{抽出した要因数}} \quad (16)$$

表4が分類の結果である。マクロ変数要因では、「ドル円」「石油」が全ての投資信託商品で共通して言及され、言及回数も非常に多かった。また、新興市場を対象とする「DIAM Emerging Markets」や「J-Frontier」では、市場の需給や心理面に分類されるような要因文が多く、中型・大型を対象とする「JPX Premium」や「Index Fund 225」では、他の商品に比べ、中国経済に分類されるような要因文が抽出された。

マクロ干渉要因は、「米国の経済指標」や「日銀の金融緩和」がどの投資信託商品でも言及されており、小型、中型、大型のどれを対象とした投資信託商品であれ、言及されるマクロ干渉要因は共通であることが多かった。

マイクロ要因は、「Index Fund 225」のような大型銘柄で構成

表3 CRFによるラベリングの精度

| 商品名 | Precision(F) | Precision(T) |
|-----------------------|--------------|--------------|
| Japan New Century | 0.937 | 0.944 |
| DIAM Emerging Markets | 0.887 | 0.841 |
| JPX Premium | 0.881 | 0.861 |
| J-Open | 0.937 | 0.944 |
| J-Frontier | 0.965 | 1 |
| Index Fund 225 | 0.944 | 0.963 |

表4 分類の精度

| 商品名 | Precision | Number of factors |
|-----------------------|-----------|-------------------|
| Japan New Century | 0.929 | 70 |
| DIAM Emerging Markets | 0.862 | 152 |
| JPX Premium | 0.881 | 152 |
| J-Open | 0.917 | 72 |
| J-Frontier | 0.89 | 100 |
| Index Fund 225 | 0.907 | 216 |
| ALL | 0.897 | 762 |

表 5 平均二乗誤差によるモデルの比較

| | Monthly Model | Trend Model | without Segmentation |
|-----------------------|---------------|---------------|----------------------|
| Japan New Century | 0.0028 | 0.0026 | 0.0033 |
| DIAM Emerging Markets | 0.0035 | 0.0032 | 0.0037 |
| JPX Premium | 0.0038 | 0.0033 | 0.0044 |
| J-Open | 0.0028 | 0.0026 | 0.0033 |
| J-Frontier | 0.0045 | 0.0035 | 0.0082 |
| Index Fund | 0.0028 | 0.0018 | 0.0033 |

された商品では、個別企業のニュースを多く取得できた、一方、「J-Frontier」のような新興市場を対象とする投資信託商品では、抽出数が少なかった。これは、日経クイックニュースが大型の個別株のニュースを多く取り扱っていることも原因であると考えられる。

分類がうまくいかない例として、「出尽くし感」「利益確定」「回避する動き」等のキーワードの不足、また「反発する場面」「一進一退の展開」「警戒感が緩和」などの要因文のように主語が欠落して抽出されている場合があった。これらの問題を解決することでより分類の精度は改善すると考えられる。

5.4 モデルの妥当性

5.4.1 トレンドを考慮したモデル構築の妥当性

ここでは、以下の3つの手法で状態空間モデルの構築の検討を行う。

- 月ごとの月時報告書から抽出した要因を用い、月ごとに状態空間モデルを構築する (Monthly Model)。
- それぞれのトレンド区間で抽出した要因を用い、トレンドごとに状態空間モデルを構築する。分割数は10月分のデータを用いたため、10とした (Trend Model)。
- 分割を行わず、状態空間モデルを構築する (Without Segmentation)。

それぞれのモデル構築において、実際の基準価額とモデルによる推定値の平均二乗誤差は表5のようになった。トレンドごとに状態空間モデルを構築する方が、トレンドを考慮せずに状態空間モデルを構築するより精度の高い推定を行えている。

また、図6を見てもわかる通り、どの投資信託商品も共通して、騰落率が大きく変動する区間は誤差が大きくなっているが、トレンドを考慮して状態空間モデルを構築したモデルが一番精度が良いことがわかる。

分析において、投資信託の基準価額にあまり影響を与えていないマクロ干渉変数が市況概況に記述され、状態空間モデルに組み込まれると推定がうまくいかないことがあったため、モデ

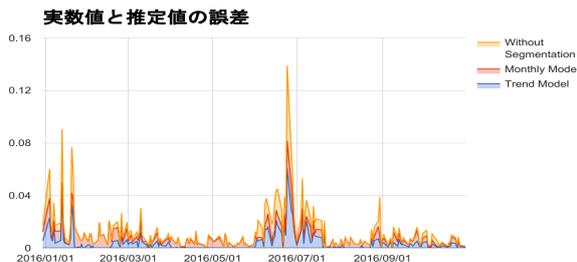


図 6 各モデルの実数値と推定値の誤差

表 6 Ljung-Box Test による残差分析

| | Japan New Century | DIAM Emerging Markets | JPX Premium | J-Open | J-Frontier | Index Fund 225 |
|----|-------------------|-----------------------|-------------|---------|------------|----------------|
| 1 | 0.06515 | 0.09105 | 0.07414 | 0.06029 | 0.9075 | 0.8248 |
| 2 | 0.4493 | 0.3536 | 0.4136 | 0.4544 | 0.1825 | 0.1624 |
| 3 | 0.3308 | 0.1789 | 0.6317 | 0.3148 | 0.7932 | 0.8925 |
| 4 | 0.2117 | 0.07708 | 0.5911 | 0.2209 | 0.8581 | 0.7772 |
| 5 | 0.1752 | 0.00756 | 0.688 | 0.1782 | 0.185 | 0.6183 |
| 6 | 0.9997 | 0.86 | 0.509 | 0.8135 | 0.321 | 0.2494 |
| 7 | 0.1508 | 0.8005 | 0.3914 | 0.6318 | 0.9957 | 0.4982 |
| 8 | 0.8566 | 0.9735 | 0.161 | 0.8741 | 0.9039 | 0.9964 |
| 9 | 0.04484 | 0.292 | 0.6397 | 0.9277 | 0.2461 | 0.802 |
| 10 | 0.1603 | 0.4431 | 0.8026 | 0.1458 | 0.6732 | 0.8964 |

ル構築において改善が必要だと考えられる。

5.4.2 要因変数選択の妥当性

時系列分析する際には、残差分析を行い系列相関が見られるかどうかを確認する必要がある。系列相関とは、時系列データにおいて誤差項が互いに相関している状態のことを指す。残差に系列相関が見られる場合はモデル設定、特に説明変数が不十分である可能性がある。これを確認するための代表的な検定がLjung-box 検定 [12] である。

表6は、トレンドごとに状態空間モデルを構築しLjung-Box 検定による残差分析を用いた結果である。検定ではラグを5とし、0.05以下だと系列相関があると解釈すると、ほぼすべての区間で、系列相関が見られなかった。これより分析に用いた要因が妥当であることが確認できる。

5.4.3 重回帰分析との比較

図7は、トレンドごとに重回帰モデルおよび状態空間モデルで推定を行った結果である。重回帰モデルは、本来の基準価額の推移を必要以上に滑らかにして表現しており、状態空間モデルの方が明らかに推定精度が高いことがわかる。また重回帰モデルの場合、ラグ変数がないため、見せかけの回帰の問題により、回帰係数が必要以上に大きく評価してしまう問題もある。

5.5 ケーススタディ

この節では、実際に投資信託商品に影響を与えた要因の影響度のある局面を題材に計測する。ここでは、「J-Frontier」の第



図 7 重回帰モデルと状態空間モデルの比較

表 7 平均二乗誤差による比較

| | Multiple Regression Model | Trend Model |
|-----------------------|---------------------------|---------------|
| Japan New Century | 0.0144 | 0.0026 |
| DIAM Emerging Markets | 0.0095 | 0.0032 |
| JPX Premium | 0.0221 | 0.0033 |
| J-Open | 0.0124 | 0.0026 |
| J-Frontier | 0.0145 | 0.0035 |
| Index Fund | 0.088 | 0.0018 |

7局面を用いた分析結果を紹介する。この局面では、マクロ変数要因として「円レート」、マクロ干渉要因として「米国の利上げ見送り」「日銀の金融緩和現状維持」「英国のEU離脱」、ミクロ要因として、「ベクトルの株式分割」が得られ、局面内で基準価額が-1353円ほど下落する下落局面である。

ここでは、回帰係数より、この局面でそれぞれの要因が与えた影響度を求める。この局面では、円レートが円高方向に106.98から102に-4.65%ほど推移している、ここで、式(11)により、円レートが1%推移すると目的変数の基準価額が局面の始日の13617円から0.422%ほど推移することになるので、 $13617円 * -4.65 * 0.422\% = -267円$ ほど影響を与えたことがわかる、また、この区間での基準価額下落の $267/1353 = 19.7\%$ を説明している。

また、マクロ干渉要因がこの局面に与えた影響度も分析することができる。例えば、式(13)より、「英国のEU離脱」は-4.89%ほどの影響があると考えられ、 $13617円 * -4.89\% = -666円$ の影響を与えていることが分かる。これは基準価額下落の $666/1353 = 49.2\%$ を説明している。

また、ミクロ要因として、この期間では、「ベクトルの株式分割」がある。当日のベクトル株は始値から2.9%ほど推移し、またJ-Frontierは6月にベクトル株を構成比率の1.7%ほど保有している。これより $12710円 * 2.9\% * 1.7\% = 6.26円$ ほどプラスに影響したとしている。

要因の影響度の目安程度ではあるが、各要因の影響の強さを検証する際に、テキスト以上の情報をユーザに与えることができると考えられる。この影響度をユーザに提供することにより、投資信託商品の基準価額の推移を理解することを支援できる。

6. 終わりに

本研究では、投資信託の理解支援のため、月次報告書と日経クイックニュースを用い、基準価格の変動を理解するための、投資信託商品の要因分析手法を提案している。我々は、投資信託商品の基準価額に及ぼす要因を「マクロ要変数要因」と「マクロ干渉要因」「ミクロ要因」に分類して、その影響度合いを状態空間モデルを用いて、時系列的に分析する手法を提案している。

また、2016年1月から10月の月次報告書および日経クイックニュースを用い、提案手法の確認を行った。まず要因の抽出および分類が高い精度で行えたことを確認している。次に、状

態空間モデルをトレンドごとに構築するモデルが、分割を行わず状態空間モデルを構築する場合よりも良い推定が行えていることを確認している。最後にケーススタディでは実際にそれぞれの要因の影響度を確認している。

我々は、国内株式を主に扱う投資信託商品を扱ったが、債権を主に扱う投資信託商品や不動産などを扱う投資信託商品も存在するため、それらの投資信託商品も検証を行うべきであると考えている。また、国内に限らず、海外の株式、債権、不動産を扱う投資信託商品も存在する。特に海外の投資信託商品は、様々な国の影響を受け、それぞれの影響度合いを考慮した新たなモデルを構築する必要があると考えられる。

今後は、状態空間モデルにより明らかにした要因の影響度をユーザに提示するアプリケーションの構築を考えている。

6.1 謝辞

本研究で用いたNQNニュースは(株)QUICKより提供されたニュースである。本研究の一部は、科研費(課題番号25700033)による。

文 献

- [1] 投資信託に関するアンケート調査報告書-2014年(平成26年), <https://www.toushin.or.jp/statistics/report/research2014/>
- [2] Kearney, Colm and Liu, Sha, Textual Sentiment in Finance: A Survey of Methods and Models, International Review of Financial Analysis, 2014.
- [3] Sakai, H. and Masuyama, S., "Assigning Polarity to Causal Information in Financial Articles on Business Performance of Companies", IEICE Trans. Information and Systems, vol.E92-D no. 12, pp. 2341-2350, 2009.
- [4] Sakaji, H., Sekine, S, and Masuyama, S., "Extracting Causal Knowledge Using Clue Phrases and Syntactic Patterns", in 7th International Conference on Practical Aspects of Knowledge Management, pp. 111-122, 2008
- [5] J Lafferty, A McCallum, and F Pereira, Conditional random fields: Probabilistic models for segmenting and labeling sequence data, In Proc. ICML '01 Proceedings of the Eighteenth International Conference on Machine Learning, 2001
- [6] 沖本 竜義, 平澤 英司, ニュース指標による株式市場の予測可能性, 証券アナリストジャーナル 52(4), 67-75, 2014
- [7] Tetlock and Paul C., Giving Content to Investor Sentiment: The Role of Media in the Stock Market, Journal of Finance 62, pp. 1139-1168, 2007.
- [8] 和泉 潔, 松井 宏樹, 松尾 豊, 人工市場とテキストマイニングの融合による市場分析, 人工知能学会誌, Vol. 22, No. 4, pp. 397404.2007
- [9] Taku Kudo, Kaoru Yamamoto, and Yuji Matsumoto, Applying Conditional Random Fields to Japanese Morphological Analysis, Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, pp.230-237, 2004.
- [10] 経済法令研究会, 実践証券業務 実践証券流通市場, pp. 58-59
- [11] R.Bellman, On the Approximation of Curves by Line Segments Using Dynamic Programming, 1961
- [12] G. M. Ljung, G. E. P. Box. On a Measure of a Lack of Fit in Time Series Models, Biometrika 65, pp. 297-303, 1978

表8 「J-Frontier」の第7局面の要因分析

| Date | NAV | 分類 | カテゴリ | キーワード | 回帰係数 | 影響度 |
|------------|-------|-------|-------|------------------|--------|---------|
| | | マクロ要因 | 通貨 | 円高 | 0.422 | -267.66 |
| 2016/06/08 | 13617 | | | | | |
| 2016/06/09 | 13626 | | | | | |
| 2016/06/10 | 13657 | | | | | |
| 2016/06/13 | 13241 | | | | | |
| 2016/06/14 | 12710 | マクロ要因 | 利上げ | FRB 利上げ見送り | -0.040 | -539.76 |
| 2016/06/15 | 12826 | ミクロ要因 | 株式分割 | 1株を3株に株式分割(ベクトル) | | +6.26 |
| 2016/06/16 | 12299 | マクロ要因 | 金融政策 | 日銀金融政策現状維持 | -0.034 | -465.73 |
| 2016/06/17 | 12318 | | | | | |
| 2016/06/20 | 12673 | | | | | |
| 2016/06/21 | 12741 | | | | | |
| 2016/06/22 | 12625 | | | | | |
| 2016/06/23 | 12647 | | | | | |
| 2016/06/24 | 11836 | マクロ要因 | 国内外政局 | 英国 EU 離脱 | -0.50 | -666.98 |
| 2016/06/27 | 12644 | | | | | |