

トレンドを考慮した投資信託商品の大局的な要因分析

桐畑 誠† 馬 強†

† 京都大学工学部情報学科 〒 606-8501 京都市左京区吉田本町

E-mail: †kirihiata@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、投資を行う人が増えてきている。投資信託商品は少額から投資が可能で、運用は専門家が行うため、初心者でも始めやすい。しかし、投資信託商品は似たものも多く、それぞれの商品に影響を及ぼす要因を理解して適切な商品を選択することが難しい。そのため、投資信託商品の要因分析手法が多数提案されているが、対象商品の要因を局所的に分析しているものが多く、予測に活用出来ない。そこで、本研究では、時系列のトレンドを動的に検知する TP-TBSM 法、およびそれを用いて検出された変化点を状態空間モデルに組み込んだトレンドシフトモデルを提案する。トレンドシフトモデルは、要因を局面やトレンドを横断して大局的に分析するため、従来手法では困難とされている予測が可能となる。

キーワード 理解支援 要因分析 状態空間モデル トレンド検知

1. はじめに

近年、NISA(少額投資非課税)制度の効果もあって、ますます多くの人が投資を行うようになってきている。金融庁による NISA 制度の効果検証結果 [1] によると、NISA 口座の開設者の約 3 割は投資未経験者であり、NISA の導入によって投資未経験者への投資の裾野拡大の効果があったと見られるとともに、その効果は若い世代ほど大きかった。平成 28 年 6 月時点での NISA 口座は 1030 万、累計買付金額は 8.4 兆円であり、制度の開始以来、順調に推移している。この NISA 口座内の金融商品残高の約 70% を投資信託が占めており、投資信託に注目が集まっている状況である。

投資信託は投資家一人一人は少額からでも投資することが可能で、様々な株式や再建に分散投資することが可能であるため、初心者でも始めやすい金融商品になっている。しかし、投資信託協会の調査 [2] によると、投資未経験者も含めた中での投資信託の現在保有率は 16% に留まり、投資経験者の中でも投資信託を購入したことがある人はおよそ 3 分の 2 程度であり、まだまだ多くの人が投資信託の購入に踏み切れていない。投資信託で不満を感じる点(複数回答可)のアンケート結果では、20.6% の人が「専門知識がないと商品を選びにくい」ことを挙げている。すなわち、投資信託商品の価格に影響を与えている要因やその影響の大きさなどの理解が難しいことが投資に踏み切れていない原因の一つだと考えられる。

そこで、投資を支援するため、市場の変動要因に関する様々な研究が行われている。栗納らや大西らの研究 [3] [4] [5] では、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した。さらに、上昇トレンドや下降トレンドなどの局面で分析を行う事で、投資信託商品の局所的な分析の精度を向上させている。しかし、投資信託商品では、将来の予測などの大局的な分析も重要である。

そこで、我々はトレンドの変化点を動的に検出し、その変化

点を考慮したトレンドシフトモデルを用いて各要因の影響度を分析する手法を提案する。トレンドを検知するため、従来の TBSM(Trend Based Segmantation Method) を拡張し、パラメータへの依存を緩和した TP-TBSM(Two Phase TBSM) 法を提案する。TP-TBSM 法は、Evaluation と呼ぶトレンドにおける誤差を評価し、分割するトレンドを選択するフェーズと Segmentation と呼ぶトレンドをさらに分割するフェーズを交互に繰り返していくことで、トレンドの変化点を検出するのに必要なパラメータ数を軽減する。さらに、検出された変化点を傾きシフト干渉変数として状態空間モデルに導入したトレンドシフトモデルを提案する。トレンドシフトモデルは、従来の状態空間モデルにおけるトレンド項よりも急激な変化を表現出来る。トレンドシフトモデルでは、特定の局面における要因の動きだけではなく、トレンドを横断して要因を大局的に分析しているため、基準価額の予測にも応用可能である。

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 節で我々が提案するトレンド検知手法と状態空間モデルについて説明する。4 節では、11 個の投資信託商品を対象に提案手法の実験及び評価を行っている。5 節で結論および今後の課題を述べる。

2. 関連研究

2.1 テキストデータを用いた金融商品・市場分析

近年、数値データのみでは汲み取れない情報を得るために、テキストデータを分析の対象としている研究が数多く見られ、予測の分野や市場理解において成果を発揮している [5] [6]。

和泉らは、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、回帰分析を用いた分析手法を用いることで、国債市場の価格変動が大きい時期において高い予測力が得られたことを示している [6]。栗納らは、投資信託商品における月次報告書の文章構造を利用した要因抽出を行い、投資信託商品の理解支援手法を提案している [5]。

これらの研究からテキストデータを取り入れる事でより詳細

な市場分析が可能になることがわかる。本研究では、既存手法を用いて投資信託商品の月次報告書から抽出・分類した要因を分析に用いる。

2.2 時系列データを用いた金融商品・市場分析

金融商品や市場というのは時系列データで表されるものが多いため、様々な時系列分析手法が金融商品・市場分析の研究に用いられている。その中でも状態空間モデルは様々なモデルを取り入れることで、目的に合わせたモデルを柔軟に構築することが出来るので、金融商品・市場分析において応用が数多くなされている [3] [4] [7] [8]。

青柳ら [7] は、セールスプロモーションの売上に対する動的效果をテレビ広告の蓄積や参照価格で階層化し、広告ストックや店舗レベル参照価格がセールスプロモーションを通じて、間接的に売上に影響を与えていることを3階層の多変量状態空間モデルを用いる事で示した。山口らは飲食店の売り上げを状態空間モデルによってモデル化し、実際の店舗の売り上げデータを用いて分析を行っている [8]。山口らのモデルでは、飲食店の売り上げは、トレンド成分、週周期成分、雨効果、イベント効果によって説明されるとし、各効果について分析した結果、雨の日は売り上げが上がることなどの知見を得ている。大西らは、運用報告書やニュースなどのテキストデータから投資信託商品に影響を与えているマクロ要因とミクロ要因を抽出・分類し、基準価額の数的データと組み合わせた状態空間モデルを提案している [3] [4]。商品の基準価額の変動要因を定量的に分析し、さらにトレンドを考慮する事で、その精度を高めている。

トレンドに着目した研究は他にも多くある [9] [10] [11]。鈴木らは時系列の天井と底と呼ばれるトレンドの変化点のみを扱う事で非線形予測法における長期的な予測の精度を向上させている [9]。通常天井と底の予測は長期予測であり、1ステップ先予測を繰り返し行った後に、予測された時系列データの中から天井と底を選ぶが、鈴木らは天井と底のデータのみを用いて1ステップ先予測を行うとその値は次の天井と底になると考え、ショートカット予測法という手法を提案し、これは天井と底の予測において良い成果を出している。

Chang らは Intelligent Piecewise Linear Representation (IPLR) と呼ばれる手法を用いて、株価などの時系列データを買いや売りというトレーディングシグナルに変換してニューラルネットワークを用いて利益を最大にする最適な閾値を学習させた [10]。新たなデータの入力をこの閾値を用いてトレーディングシグナルに変換し、利益を求めたところルールベースの取引などよりも良い利益を得ている。Jheng-Long らは Trend Based Segmentation Method (TBSM) と呼ばれる手法をサポートベクター回帰と共に用いることにより売買タイミングの予測を行い、他の手法に比べて高い精度が得られたことを示している [11]。TBSM では、線形関数を用いて時系列データを近似し、線形関数から距離が一番大きい点をトレンドの変化点として上昇、下降、停滞の3つのトレンドに分割する。

これらの先行研究は、金融市場理解においてトレンドを考慮することや状態空間モデルが有用であることを示している。しかし、既存のトレンド検知手法では、個々の時系列データに対

して適当な閾値を求める必要があり、価格帯が様々である投資信託商品では画一的にトレンドを考えることができない。そこで、TBSM 法 [11] における閾値の数を軽減させることで、画一的にトレンドを求める手法を提案する。また、大西らはトレンドの局面ごとに分析が完結しており、将来の予測やに用いることができない。将来の基準価額の予測というのは投資の支援において大きな影響を持つと考えられるので、本研究では、予測にも対応した状態空間モデルを提案する。

3. 提案手法

本研究では、投資信託商品の基準価額を定量的に分析するため、トレンド検知手法である TBSM [11] を拡張した TP-TBSM と、それに基づいて検出された変化点を状態空間モデルに組み込んだトレンドシフトモデルを提案する。

3.1 TP-TBSM

TBSM 法では、あらかじめ設定する3つの閾値が検知されるトレンドを大きく左右するため、時系列データに合わせて適切なパラメータを決めることが困難である。そこで、本研究では、このパラメータへの依存を緩和した手法を TP-TBSM を提案する。我々はトレンド誤差という概念を導入し、トレンド誤差が大きいトレンドをさらに分割することで、パラメータへの依存を緩和する。

トレンド誤差とは、トレンドの始点と終点を結んだトレンド直線（線形関数で表現出来る）と各データ点の距離について足し合わせ、そのトレンドの長さで割ったものであり、トレンド直線から離れる程度を測る尺度である。トレンド誤差は以下のように定義される：ただし、 t_s 、 t_e はそれぞれトレンドの開始時点と終了時点であり、 $y(t)$ は時系列データで、 $f(t)$ はトレンドの始点と終点を直線で結んだ線形関数である。

$$Error(t_s, t_e) = \frac{\sum_{t=t_s}^{t_e} |f(t) - y(t)|}{t_e - t_s} \quad (1)$$

本研究ではこの誤差が小さければ良いトレンドとする。また、トレンド直線と点の距離として、同じ時点での値の差である式 (1) の $|f(t) - y(t)|$ を今後も用いる。

我々は各トレンドにおいてこのトレンド誤差が平均値より大きければ再分割の対象とし、トレンドの長さが閾値（あらかじめ設定した長さの最小値）を下回る直前まで分割を繰り返す。提案手法は Algorithm1 に示すように、Evaluation と Segmentation の二つのフェーズを交互に繰り返すことでトレンドを検知する。Evaluation のアルゴリズムを Algorithm2、Segmentation のアルゴリズムを Algorithm3 に示す。この2つのフェーズについて説明した後、TP-TBSM のアルゴリズムについて説明する。

Evaluation フェーズはトレンド誤差を用いて、分割するトレンドを選択するフェーズである。

(1) 式 (1) を用いて、各トレンドにおけるトレンド誤差を計算する。(行番号 2-4)

(2) 再分割するトレンドを決める閾値 δ_e をトレンド誤差の平均値とする。(行番号 5)

(3) 閾値 δ_e よりトレンド誤差が大きいトレンドを E に追

加することで、分割するトレンドを選択する。(行番号 6-10)

(4) 選択したトレンド集合 E を返す (行番号 11).

Segmentation フェーズはトレンドをさらに分割するフェーズである。

(1) トレンドを表す直線との距離が最大となる点を求め、その点を変化点の候補とする (行番号 1). ここで、区間 $[start + \delta_x, end - \delta_x]$ を考えているのは、分割したトレンドが一定の長さ以上であることを保証するためである。

(2) 変化点が閾値を超えているかどうかで分割するかどうかを決める (行番号 2). もし閾値を超えていなければ、分割せずにトレンドを返す (行番号 20).

(3) 停滞トレンドを考える (行番号 3-8). 変化点 s_k 周辺において値の変動が小さければ停滞トレンドとする. 変化点 s_k の周辺を区間 $[k - \delta_x, k + \delta_x]$ とし、この範囲において、 s_k との値の差が $\frac{\delta_d}{2}$ より小さい時点を順番にリスト Hold に追加していく.

(4) 停滞トレンドを用いて 3 つに分割するか、変化点のみを用いて 2 つに分割するかを判断する (行番号 9). リスト Hold

Algorithm 1 TP-TBSM のアルゴリズム

記号:

s_i : 時系列データ S の時点 i の値

(start, end): トレンドの開始時点を start, 終了時点を end とした時のトレンド

$line(s_i, s_j)$: 点 s_i , と点 s_j を結んだトレンド直線

T_k : 昇順に並べられた集合

δ_d : Segmentation における分割の判断をする閾値

Input: S : 時系列データ, δ_x : トレンドの長さの閾値

Output: T_k : トレンドの変化点の集合, 添字が何回更新されたかを表す

```

1:  $T_0 = \{1, n\}$  //  $S$  の最初の時点 1 と最後の時点  $n$  を変化点として入れる
2:  $E = \{(1, n)\}$  // 分割するトレンドを入れる
3:  $k = 0$ 
4: repeat
5:   if  $k > 0$  then
6:      $E = \text{Evaluation}(T_k, S)$ 
7:   end if
8:    $k = k + 1$ 
9:    $T = T_{k-1}$ 
10:  for  $(start, end) \in E$  do
11:    if  $end - start < 2\delta_x$  then
12:      トレンドの長さが短く分割できないので次のトレンドに移る
13:    else
14:      区間  $[start + \delta_x, end - \delta_x]$  におけるトレンド直線  $line(s_{start}, s_{end})$  との距離が最大になる点を求め、その値を  $dm$  とする
15:       $\delta_d = dm$ 
16:       $T = T \cup \text{Segmentation}(S, \delta_x, \delta_d, start, end)$ 
17:    end if
18:  end for
19:   $T_k = T$ 
20: until  $T_k = T_{k-1}$ 
21: return  $T_k$ 

```

Algorithm 2 Evaluatin のアルゴリズム

記号:

t_i : 昇順に並べられている集合 T の i 番目の要素

p : トレンドの個数 (T に含まれる変化点の個数から 1 引いた数)

e_list : 長さ p のリスト, 計算したトレンド誤差を入れる

$Error$: トレンド誤差を計算する関数

Input: S : 時系列データ, T : トレンドの変化点の集合

Output: E : 分割するトレンドの集合

```

1:  $E = \emptyset$ 
2: for  $i = 1 : p$  do
3:    $e\_list[i] = Error(t_i, t_{i+1})$ 
4: end for
5:  $\delta_e = \text{average}(e\_list)$ 
6: for  $i = 1 : p$  do
7:   if  $e\_list[i] > \delta_e$  then
8:      $E = E \cup (t_i, t_{i+1})$ 
9:   end if
10: end for
11: return  $E$ 

```

Algorithm 3 Segmentation のアルゴリズム

記号:

$Hold$: 停滞トレンドを考えるための変化点リスト

p : Hold に含まれている時点の個数

$line(s_i, s_j)$: 点 s_i , と点 s_j を結んだトレンド直線

Input: S : 時系列データ, δ_x : 閾値, δ_d : 閾値, $(start, end)$: トレンド

Output: T : 変化点の集合

```

1: 区間  $[start + \delta_x, end - \delta_d]$  において  $line(s_{start}, s_{end})$  との距離が最大となる時点を求め、その時点を  $k$ , 対応する距離を  $dm$  とする
2: if  $(dm \geq \delta_d)$  then
3:    $p = 0$ 
4:   for  $i = (k - \delta_x) : (k + \delta_x)$  do
5:     if  $|s_i - s_k| < \frac{\delta_d}{2}$  then
6:        $Then\ Hold[p] = i, p = p + 1$ 
7:     end if
8:   end for
9:   if  $(Hold[p] - Hold[1] > \delta_x)$  and  $(p > \frac{Hold[p] - Hold[1]}{2})$  then
10:     $t_a = \text{Segmentation}(S, \delta_x, \delta_d, start, Hold[1])$ 
11:     $t_b = \{Hold[1], Hold[k]\}$ 
12:     $t_c = \text{Segmentation}(S, \delta_x, \delta_d, Hold[k], end)$ 
13:    return  $\{t_a, t_b, t_c\}$ 
14:   else
15:     $t_a = \text{Segmentation}(S, \delta_x, \delta_d, start, k)$ 
16:     $t_c = \text{Segmentation}(S, \delta_x, \delta_d, k, end)$ 
17:    return  $\{t_a, t_c\}$ 
18:   end if
19: end if
20: return  $\{start, end\}$ 

```

の最初に入っている時点を停滞トレンドの始点とし、最後に入っている時点を終点とする. 停滞トレンドの長さが閾値 δ_d より長く、さらに停滞トレンドの半分以上が Hold に入っている場合のみ停滞トレンドを用いて三つに分割する.

(5) 停滞トレンドを用いて分割する場合 (行番号 10-13).

停滞トレンドを用いてトレンドを3つに分けた後、停滞トレンド以外のトレンドについては、同じ閾値を用いてさらに Segmentation を行う。分割が終了すれば、変化点をまとめて返す。

(6) 変化点のみを用いて分割する場合(行番号 15-17). 変化点を用いてトレンドを2つに分けた後、同じ閾値を用いてさらに Segmentation を行う。分割が終了すれば、変化点をまとめて返す。

TP-TBSM のアルゴリズムの説明を Algorithm1 を用いて行う。

(1) 時系列の最初と最後を結んだ直線をトレンドとし、さらに分割する集合 E にこのトレンドを入れる(行番号 1,2).

(2) トレンドに対して Evaluation を行う(行番号 5-7). トレンドの中で誤差が大きいトレンドを選択して、集合 E に入れる。k=0 に関しては、すでに E を考えてあるので Evaluation をスキップする。

(3) 集合 E に含まれているトレンドに対して長さの判定を行う(行番号 10-12). トレンドの長さが $2\delta_x$ より短ければ、どのような分割をしてもトレンドの長さが δ_x を下回るの、これ以上分割しない。

(4) 分割が可能であれば、分割するための閾値を決め、Segmentation を行い、得られた変化点を用いて集合 T を更新する(行番号 11-13). この時、区間 $[start + \delta_x, end - \delta_x]$ におけるトレンドを表す直線との距離が最大になる時点を求め、その値を閾値 δ_d にすることで、トレンドを必ず1回は分割すると共に分割する回数を抑えている。

(5) 2-4 を Segmentation フェーズの結果が1つ前の結果と一致するまで繰り返す。(行番号 18). すなわち、さらに分割すべきと判断されたトレンドが全て分割できなければ、アルゴリズムを終了するということである。

図1は実際にトレンドの検出を行った例である。図1の(a)では最初に始点と終点を結んだ直線から一番距離が遠い点が変化点として選ばれている。この時停滞トレンドが存在する

かどうかを緑色の四角形で判断する。今回は四角形に含まれているデータ点が少ないので、停滞トレンドは存在しない。図1(a)で、evaluation を行うと2番目のトレンドで選択される。Segmentation を行うと、図1(b)のように、黒い直線で変化点を選ばれ、分割を行った後、さらにその右側のトレンドで直線との距離が閾値を超えているので、このトレンドも分割する。これらを繰り返してトレンドを検知する。

3.2 トレンドシフトモデル

3.2.1 状態空間モデル

大西らが用いた状態空間モデルについて説明する[3][4]. 大西らは月次報告書と日経クイックニュースから抽出・分類した要因を説明変数とした状態空間モデルを構築した。その際用いた要因は、円レートなどの時系列データであるマクロ変数要因、増税の延期などのイベントを表すマクロ干渉要因、個別の企業の業績などのマイクロ要因である。式として表すと以下の通りである。

$$y_t = \mu_t + \sum_k \beta_{k,t} x_{k,t} + \sum_m \lambda_{m,t} w_{m,t} + \sum_l s_{l,t} + \epsilon_t \quad (2)$$

$$\epsilon_t \sim NID(0, \sigma_\epsilon^2) \quad (3)$$

$$\mu_{t+1} = \mu_t + \xi_t \quad (4)$$

$$\xi_t \sim NID(0, \sigma_\xi^2) \quad (5)$$

$$s_{l,t} = y_{t-1} \times c_{l,t} \times d_{l,t} \quad (6)$$

$$\beta_{t+1} = \beta_t \lambda_{t+1} = \lambda_t \quad (7)$$

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要因 k の対数值、 $w_{m,t}$ はある時点まで 0 をとり、要因発生時より 1 となるマクロ干渉要因 m である。また $s_{l,t}$ はマイクロ要因の影響度である。これはある銘柄 l のニュースによる騰落率 d である $d_{l,t}$ とその銘柄の攻勢比率 $c_{l,t}$ を前日の対数基準価額に乘算して求めている。未知パラメータである σ_ϵ^2 , σ_ξ^2 , β , λ は最尤推定で求める。回帰係数 β , λ が各要因の影響度を定量的に表したものになる。 μ_t は一つ前の状態と状態攪乱項を足し合わせていくランダムウォークになっており、このランダムウォークに説明変数の影響を足し合わせることで基準価額を表現する。

3.2.2 トレンドシフトモデル

一般的に株価などの時系列データは平均などが時間と共に変動する非定常時系列であるので、トレンドを処理する必要がある。大西らはトレンドの変化点でデータを区切り、その中で上記の状態空間モデルを構築することでトレンドを処理しているが、トレンド局面で分析が完結してしまっているの、将来の予測に用いることが出来ない。そこで、本研究では検出した変化点を傾きシフト干渉変数として組み込んだ状態空間モデルを提案する。以下このモデルをトレンドシフトモデルと呼ぶこととする。

i 番目の変化点の時刻を τ とすると、傾きシフト干渉変数は

$$z_{i,t} = \begin{cases} 0 & t \leq \tau \\ t - \tau & t > \tau \end{cases} \quad (8)$$

と定義出来る[12]. $z_{i,t}$ は時刻 τ を超えた時点から時刻が 1 つ

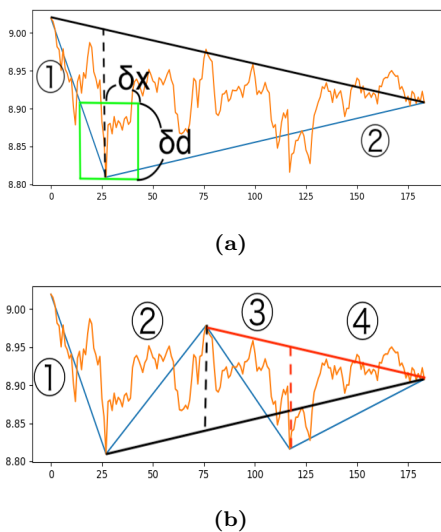


図 1: TP-TBSM の例

進む度に1大きくなる変数である。この変数の回帰係数を求めることで変化点での傾きの変化を求めることができる。

傾きシフト干渉変数を取り入れたトレンドシフトモデルの式は以下ようになる。マイクロ要因に関しては説明を簡単にするため、今回は考えないこととした。

$$y_t = \mu_t + \sum_i \alpha_{i,t} z_{i,t} + \sum_k \beta_{k,t} x_{k,t} + \sum_m \lambda_{m,t} w_{m,t} + \epsilon_t \quad (9)$$

$$\epsilon_t \sim NID(0, \sigma_\epsilon^2) \quad (10)$$

$$\mu_{t+1} = \mu_t + \xi_t \quad (11)$$

$$\xi_t \sim NID(0, \sigma_\xi^2) \quad (12)$$

$$\alpha_{i,t+1} = \alpha_{i,t}, \beta_{k,t+1} = \beta_{k,t}, \lambda_{m,t+1} = \lambda_{m,t} \quad (13)$$

大西らの式との違いは、 $\sum_i \alpha_{i,t} z_{i,t}$ の部分である。 z_i が傾きシフト干渉変数で、 $\alpha_{i,t}$ がその回帰係数である。したがって、この $\alpha_{i,t}$ が変化点における傾きの変化量であり、これを足し合わせた値が線形トレンドの傾きになっている。 μ_t は大西らと同様にランダムウォークになっており、変化する線形のトレンドとランダムウォークを組み合わせることでトレンドを処理している。

4. 評価実験

4.1 目的

本論文で提案しているトレンド検出手法とトレンドシフトモデルの評価実験について述べる。まず、提案手法のトレンドシフトモデルと基本的な状態空間モデルを比較することで、トレンドシフトモデルの有用性を評価する。さらに提案しているトレンド検出手法と既存手法で検出したトレンドの変化点を用いてトレンドシフトモデルを構築し、その分析結果を比較し、提案手法によるトレンド検出手法の有用性も評価する。

4.2 評価方法

推定されたモデルの評価を二つの尺度を用いて行う。

4.2.1 欠測値の補完精度

状態空間モデルは予測に用いられることが少なく、要因の分析に用いられることが多い。そのため、どれだけデータを再現できているかに焦点が当てられることが多い。しかし、投資信託商品においては、予測における精度も重要である。そこで、本研究では未観測データに対する平均予測誤差をモデルの評価に用いる。しかし、モデルには回帰成分が含まれており、完全に未観測なデータに対しては予測を行うことが出来ないの、説明変数のデータに関しては観測されたデータを用いることとなるため、純粋な予測ではなく、観測値の補完である側面が強い。

4.2.2 赤池情報量基準 (AIC)

欠測値の補完の精度だけではなく、赤池情報量基準 (AIC) もモデルの評価に用いる。最大化された対数尤度を L とし、未知パラメータの個数を r 、散漫な初期状態の個数を q 、時点の数を n とすると時系列における AIC は以下のように表される [12]。

$$AIC = \frac{-2L + 2(q+r)}{n} \quad (14)$$

散漫な初期状態とは、推定によって初期状態を求めなければならない状態のことである。したがって、AIC は最大対数尤度に対して推定しなければいけないパラメータの数だけペナルティを課したものになっている。時系列の尤度は一期先予測誤差に基づいているため、この AIC 小さい値になるモデルというのは単純でかつ一期先予測の精度が良いモデルということになる。

4.3 比較手法

ローカルレベルモデルとローカル線形モデルと比較して、トレンドシフトモデルを評価する。また、TP-TBSM 手法を、動的計画法に基づくトレンド検出手法 [3] と比較して評価を行う。

4.3.1 ローカルレベルモデル

ローカルレベルモデルは状態空間モデルにおける基本的なモデルであり、大西らの手法でも用いられていたモデルである。レベルと呼ばれる状態の時間的な変動がランダムウォークであると仮定したモデルである。大西らの式からマイクロ要因の項を取り除いたものを用いる。このモデルをローカルモデルと呼ぶ。

4.3.2 ローカル線形モデル

ローカル線形トレンドモデルはローカルレベルモデルにレベル $\mu_{1,t}$ の平均的な傾きを表す状態であるスロープ $\mu_{2,t}$ を加えたモデルである。大西らの式において μ_t の状態方程式のみが異なるので、その部分を式で表すと以下の通りである。

$$\mu_{1,t+1} = \mu_{1,t} + \mu_{2,t} + \xi_{1,t}, \xi_{1,t} \sim NID(0, \sigma_{\xi,1}^2) \quad (15)$$

$$\mu_{2,t+1} = \mu_{2,t} + \xi_{2,t}, \xi_{2,t} \sim NID(0, \sigma_{\xi,2}^2) \quad (16)$$

未知パラメータの推定をすると状態攪乱項の分散 $\sigma_{\xi,1}^2, \sigma_{\xi,2}^2$ のどちらかは 0 に推定されることが多い。そこで今回は $\sigma_{\xi,1}^2$ を 0 に固定したモデルと $\sigma_{\xi,2}^2$ を 0 に固定したモデルを別々に考える。スロープの状態攪乱項 $\xi_{2,t}$ を 0 に固定するとスロープの値は一定になり、ローカルレベルモデルに線形トレンド項を加えたモデルになる。これを線形モデルと呼ぶこととする。レベルの状態攪乱項 $\xi_{1,t}$ を 0 に固定するとレベルは少しずつ変化するスロープの値を足し合わせていくモデルになるので、ローカルレベルモデルよりレベル項が滑らかに変動する。このモデルをトレンドモデルと呼ぶこととする。

4.3.3 動的計画法に基づくトレンド検知

動的計画法 (DP) によってトレンドを検知する手法は大西らが用いていた手法である [3] [4]。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、基準価額との二乗平均誤差を求め、それらを合計したスコアが最小になるように分割点を動的に計算する。この時、動的計画法はあらかじめ分割する個数を決めておく必要がある。

4.4 実験の実施概要

2016年1月4日から2016年10月31日までの基準価額のデータを用いて実験を行う。このうち最後の20日分のデータは欠測値の補完に用いるデータとし、残りのデータでモデルの推定およびトレンドの検出を行う。この20日は土日など基準価額や円レートなどのデータが無い日を除いた約1か月分のデータになる。TP-TBSMによるトレンドの検知に用いた閾値 δ_x も20日とした。

要因として用いるマクロ変数要因とマクロ干渉要因は大西ら

表 1: 実験に用いた投資信託商品

商品	略称	マクロ変数
投資満々	toshimanman	CNY
グローイング・カバーズ	growing	FTSE
DIAM 新興市場日本株ファンド	diam	CNY
日本株厳選ファンド円コース 4	gensen	DJI
日本株アルファカレット (毎月分配型)	alpha	DJI
ストックインデックスファンド 225	stock	JPY
ニッポンの翼	tsubasa	FTSE
J-フロンティア	jfrontier	FTSE
J オープン (店頭小型株)	jopen	FTSE
日本新世紀	shinseiki	FTSE
ひふみ投信	hihumi	FTSE

表 2: 欠測値の補完の誤差 (トレンドシフトモデルの評価実験)

商品 (略称)	ローカル	線形	トレンド	シフト
toshimanman	0.0121257	0.0123519	0.0360227	0.00760425
growing	0.01692213	0.01192982	0.02064795	0.00807945
diam	0.0263603	0.0187761	0.01338155	0.0096469
gensen	0.0091394	0.01260265	0.0132246	0.0095815
alpha	0.02357305	0.0224242	0.01857475	0.01863415
stock	0.01534265	0.0147249	0.0112846	0.0217712
tsubasa	0.019291	0.0176646	0.0425186	0.01261265
jfrontier	0.01504885	0.02040415	0.027809	0.00924125
jopen	0.0211324	0.01933145	0.03037835	0.01348215
shinseiki	0.01992795	0.01959765	0.0345798	0.012273
hihumi	0.01532515	0.01736685	0.0213364	0.00893415

が抽出・分類したものをを用いる。マクロ変数要因として用いたのは、人民元とドルの為替レートを表す CNY, 日本円とドルの為替レートを表す JPY, 原油価格である OIL, ロンドン証券取引所における株価指数である FTSE, アメリカ合衆国の代表的な株価指数である DJI の 5 つである。多重共線性などを考慮して、ローカルレベルモデルを用いて一番 AIC が小さくなる変数の組み合わせを各商品ごとに選択して用いた。表 1 に各商品の略称と選択されたマクロ変数を示す。

マクロ干渉要因に関しては抽出されたもののうち、欠測値の補完の期間に変化点があるものを除いた 26 個を用いた。

状態空間モデルは標準化された予測誤差が独立で正規性であることを仮定しているため、推定したモデルにおいて、予測誤差に系列相関が見られたり、正規分布に従わない場合にはモデルの構築が不十分である可能性がある。系列相関が見られるかどうかの検定として Ljung-Box 検定があり、正規性を調べる検定として Shapiro-Wilk 検定がある。そこで、13 個の商品に対して各モデルでの分析を行い、Ljung-box 検定と Shapiro-Wilk 検定を有意水準 5% で満たしていた 11 個の商品のみを比較に用いた。

4.5 結果・考察

4.5.1 トレンドシフトモデルの評価

ローカルモデル、線形モデル、トレンドモデル、トレンドシフトモデル (TP-TBSM) を比較する。表 2 を見ると欠測値の補完の平均誤差においては、提案手法であるトレンドシフトモデルが 11 個の商品のうち 8 個の商品で一番小さい値になっており、一番良い性能を示している。これはトレンドシフトモデルが要因の影響係数をより正確に推定出来たことを示している。

また、ローカルモデルや線形モデルの予測誤差はほとんどの商品で小さい値を保っている。これはデータ全体で見たトレンドは傾きが小さく気基準価額の変動はほとんどがランダムウォークのレベルに依存し、前の状態から値があまり変動していないためだと考えられる。言い換えれば、これらのモデルはトレンドの影響を十分に考慮出来ていないと思われる。

トレンドモデルに関しては、予測誤差が小さくなる時と大きくなる時の違いが大きい。これはトレンドを表しているスロープ項の値が予測期間の本当に直前だけに影響されており、上昇トレンドであっても直前に少し下降していればその後

表 3: AIC (トレンドシフトモデルの評価実験)

商品 (略称)	ローカル	線形	トレンド	シフト
toshimanman	-4.052401	-3.969067	-3.764244	-3.814294
growing	-4.212589	-4.127697	-3.918999	-3.968967
diam	-3.723839	-3.644144	-3.416982	-3.509882
gensen	-3.730298	-3.64922	-3.416481	-3.584658
alpha	-3.365685	-3.284338	-3.0366	-3.217069
stock	-4.281146	-4.194311	-4.109473	-3.82778
tsubasa	-4.076699	-3.991273	-3.76433	-3.830447
jfrontier	-3.960386	-3.876909	-3.623052	-3.839674
jopen	-4.133614	-4.047846	-3.821848	-3.887454
shinseiki	-4.133536	-4.047639	-3.820015	-3.744925
hihumi	-4.193313	-4.1072	-3.888612	-3.868926

ずつと下降していると推定してしまうからであると考えられる。

表 3 に示されているように AIC に関しては全ての商品においてローカルモデルが一番低い値になり、線形モデルが 2 番目に低い値になっている。AIC では単純なモデルで性能が良いものが良いと評価される傾向があり、この二つは単純なランダムウォークを用いているためであると考えられる。AIC の評価によってパラメータを増やすことによる過適合を防ぐことができるが、状態空間モデルにはランダムウォークするレベル項がデータを追従しすぎてしまい、変動の原因のほとんどランダムウォークとして判断されるという過適合の問題があると考えている。詳細については 4.5.3 節で述べるが、これによってこの二つのモデルは AIC が他の二つより小さくなっているという考え方が出来る。なぜなら、トレンドモデルとトレンドシフトモデルはデータの変動をトレンドによるものとランダムウォークの影響によるものを足し合わせてモデリングしているからである。

トレンドモデルとトレンドシフトモデルの AIC を比べると 11 商品中 8 商品でトレンドシフトモデルの方が良い値になっており、トレンドモデルに比べるとトレンドシフトモデルは良いモデルだと言える。一方、トレンドモデルの方が AIC が低い値になっている商品 stock ではトレンドモデルの方が欠測値の補完の平均誤差が小さくなっており、AIC の値が小さくなるようなトレンドシフトモデルを構築することが良いと分かる。

4.5.2 TP-TBSM に関する評価

トレンドシフトモデル (DP) とトレンドシフトモデル (TP-TBSM) の値について比較する。商品 stock と商品 jopen に関

表 4: 欠測値の補完誤差 (TP-TBSM の評価実験)

商品 (略称)	シフト (DP)	シフト (TP-TBSM)
toshimanman	0.015405445	0.00760425
growing	0.018084755	0.00807945
diam	0.013712115	0.0096469
gensen	0.01047508	0.0095815
alpha	0.017375915	0.01863415
stock	ERROR	0.0217712
tsubasa	0.007685725	0.01261265
jfrontier	0.0199715	0.00924125
jopen	ERROR	0.01348215
shinseiki	0.012518285	0.012273
hihumi	0.009192925	0.00893415

表 5: AIC (TP-TBSM の評価実験)

商品 (略称)	シフト (DP)	シフト (TP-TBSM)
toshimanman	-3.580444	-3.814294
growing	-3.686648	-3.968967
diam	-3.268868	-3.509882
gensen	-3.245881	-3.584658
alpha	-2.816559	-3.217069
stock	ERROR	-3.82778
tsubasa	-3.495332	-3.830447
jfrontier	-3.387825	-3.839674
jopen	ERROR	-3.887454
shinseiki	-3.586449	-3.744925
hihumi	-3.681659	-3.868926

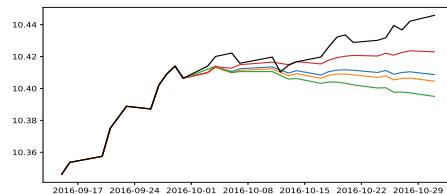


図 2: hihumi の欠測値の補完

(青:ローカル, 黄:線形, 緑:トレンド, 赤:シフト (TP-TBSM))

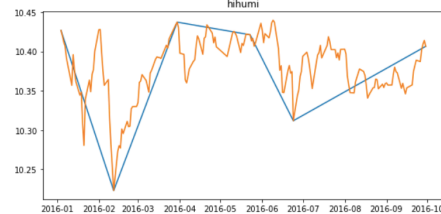


図 3: hihumi のトレンドの変化点

表 6: シフト干渉変数の回帰係数

変化点	回帰係数	時点での傾き
0	-8.382777e-03	-8.382777e-03
1	1.329071e-02	4.907933e-3
2	-4.279738e-03	6.28195e-4
3	-1.872531e-03	-1.244336e-3
4	1.978781e-03	7.34445e-4

してはトレンドシフトモデル (DP) の推定が上手くいかなかったため、ERROR と表記してある。

表 5 を見ると AIC においては全ての商品で TP-TBSM を用いたモデルの方が良い結果を出している。DP はトレンドの数が全て 9 個であるのに対して、TP-TBSM で分けられたトレンドの数は多いもので 7 個、少ないものは 2 個となっており、それぞれの時系列に合わせて動的に変動している。トレンドの変化点が多くなると推定しなければいけないパラメータの数が増えてしまい、結果として AIC が大きくなったと考えられる。

トレンドの数が多くなったということは欠測値の補完を行う期間の直前のトレンドを上手く推定出来ている可能性が高いが、表 4 を見ると欠測値の補完誤差においても TP-TBSM の方が良い結果を得られている。DP 法はトレンドの個数をあらかじめ決めており、トレンドの変化を上手く捉えられなかったと考えられる。TP-TBSM では柔軟にトレンドの個数を決められることが AIC と欠測値の補完における誤差を共に小さくすることに繋がったと考えられる。

4.5.3 ケーススタディ

この節では、商品 hihumi に対して、実際に推定した状態空間モデルを詳細に見ていくことでトレンドシフトモデルがどのような効果をもたらしているかを確かめる。

商品 hihumi のマクロ変数要因としては FTSE が選択された。欠測値の補完は図 2 のようになっており、TP-TBSM を用いたトレンドシフトモデルが一番平均誤差が小さくなっている。この図からもトレンドシフトモデルがトレンドを上手く推定出来ていることがわかる。ローカルモデルや線形モデルは予測が

始まった時点からあまり変動していない。また、トレンドモデルは予測が始まる直前に少しだけ下降した影響を受けてしまい、予測値がどんどん下降して、実測値を離れていってしまっている。

TP-TBSM によるトレンド検知は図 3 のようになった。見た通りでは良くトレンドを上手く検知出来ているように思える。

この時の傾きシフト干渉変数の回帰係数は表 6 のようになっており、この回帰係数を順番に足し合わせた値がその時点の傾きになっている。各時点での傾きの符号や大きさを 3 での線形トレンドと比較すると概ね一致しており、きちんと傾きを推定出来ていることがわかる。

AIC による評価において問題としていたレベル項の影響の大きさについて考察する。各モデルでのレベル項を図 4 のように示した。ローカルモデルや線形モデルのレベル項は図 3 から観測値とほぼ同じ形をしているが分かる。レベル項はランダムウォークによる変動であるので、この変動が大きいと基準価額の変動がほとんど偶然によるものだという説明になってしまい、要因の影響度を正確に測れない。トレンドモデルではこのレベル項が滑らかに変動しているので、ローカルモデルや線形モデルに比べて基準価額の変動とは似ていない。したがって、レベル項の影響が小さくなり、偶然による変動として説明される部分が減る。トレンドシフトモデルでは誤差が大きくなっているところ以外では変動が抑えられており、全体での変動の幅も他のモデルより半分近く小さくなっている。したがって、トレンドシフトモデルでは偶然による変動として推定される部分が少

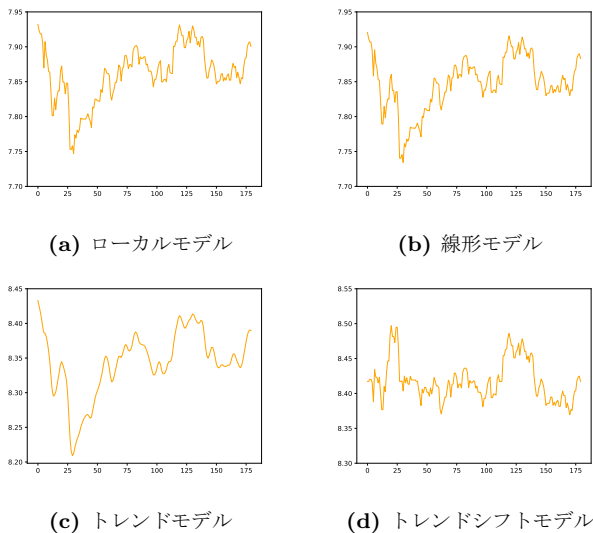


図 4: モデルによるレベルの違い

表 7: 回帰係数の違い

	FTSE の回帰係数	ECB 総裁の回帰係数
レベル	0.2862249844	0.0526380370
線形	0.2875501609	0.0528175387
トレンド	0.228812161	0.062011960
トレンドシフト	0.2305579	0.06222468

なくなっている。

マクロ変数やマクロ干渉変数の回帰係数について考察する。マクロ干渉変数においてどのモデルでも一番影響が大きかったのは「ECB 総裁が追加緩和の可能性を示唆した」というものである。ECB とは欧州中央銀行のことで、マクロ変数として選択された FTSE とも関連が深いものになっている。各モデルによる回帰係数の値を比べると表 7 の通りになっており、ローカルモデルと線形モデルでほとんど同じ値になり、トレンドとトレンドシフトモデルでほとんど同じ値になった。

線形モデルでは傾きは常に一定で、トレンドモデルでは傾きは毎時点変化しているので、今回提案したトレンドシフトモデルはこの二つのモデルの中間に位置している。TP-TBSM によって傾きがどれくらいの頻度で変化するのが良いのかというのを決めることができたため、予測誤差や AIC において良い性能を示せたと考えられる。

5. おわりに

本研究では、投資信託商品の基準価額に影響を及ぼす要因を定量的に分析するため、トレンド検知手法とその変化点を状態空間モデルに組み込んだトレンドシフトモデルを提案している。トレンド検知手法では、トレンドにおける誤差を用いてトレンドに評価と分割を繰り返すことで、従来手法と比べてパラメータへの依存を軽減し、トレンドをより柔軟に検知できる。また、検出した変化点を状態空間モデルに組み込んだトレンドシフトモデルを用いて評価を行ったところ、トレンドシフトモデルは従来のトレンドを考慮した状態空間モデルに比べて、欠測値の

補完精度が高いことが分かった。

今後の課題としては大きく分けて以下の点が挙げられる。

- (1) TP-TBSM の複数時系列への適用
- (2) 複数商品を用いた大局的な分析

課題 1 に関して、複数の時系列を考慮することで、複数の時系列に共通したトレンドを検知することが出来るはずである。共通したトレンドを用いて要因分析を行うことで、より正確な要因の影響度を測ることが出来ると考えている。

課題 2 に関して、投資の支援のためには商品の選択を支援する必要がある。投資信託商品は様々な企業に分散投資しているため、違いが分かりにくく、商品の選択が難しい。類似している複数商品では、マクロ要因の影響度は共通しているだろうという仮定から、複数商品に対して共通した影響度を考えた要因分析を行う事で、類似商品の差が何の影響で起きているのかを分析することができると考えている。

6. 謝 辞

本研究の一部は科研費 (16K12532) による。

文 献

- [1] 金融庁, NISA 制度の効果検証結果, <http://www.fsa.go.jp/policy/nisa/20161021-1.html>, (2016).
- [2] 投資信託協会, 投資信託に関するアンケート調査報告書, <https://www.toushin.or.jp/statistics/report/research>, (2016).
- [3] 大西恒彰, 馬強, 状態空間モデルの拡張による投資信託商品の要因分析, DEIM Forum 2017 E6-2, (2017).
- [4] Nobuaki Onishi, Qiang Ma, Factor analysis of investment trust products by using monthly reports and news articles, ICDIM, pp. 32–37, (2017).
- [5] Awano, Y., Ma, Q. and Yoshikawa, M.: Causal Analysis for Supporting Users' Understanding of Investment Trusts, iiWAS, pp. 524–528, (2014).
- [6] 和泉潔, 後藤卓, 松井藤五郎ほか, 経済テキスト情報を用いた長期的な市場動向推定, 日本オペレーションズ・リサーチ学会和文論文誌, Vol. 52, No. 12, pp. 3309–3315, (2011).
- [7] 青柳 憲治, 佐藤 忠彦, 3 階層多変量状態空間モデリングによる動的市場反応形成メカニズムの解明, 日本オペレーションズ・リサーチ学会和文論文誌, Vol. 58, pp. 70–100, (2015).
- [8] 山口類, 土屋映子, 樋口知之, 状態空間モデルを用いた飲食店売上の要因分解, オペレーションズ・リサーチ, Vol. 49, No. 5, pp. 52–60, (2004).
- [9] 鈴木智也, 太田真喜ほか, 時系列データの天底予測のための非線形予測法, 情報処理学会論文誌数理モデル化と応用 (TOM), Vol. 2, No. 1, pp. 123–132, (2009).
- [10] Chang, P.-C., Fan, C.-Y. and Liu, C.-H.: Integrating a piecewise linear representation method and a neural network model for stock trading points prediction, IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), Vol. 39, No. 1, pp. 80–92, (2009).
- [11] Jheng-Long Wu, Pei-Chann Chang, A Trend-Based Segmentation Method and the Support Vector Regression for Financial Time Series Forecasting, Mathematical Problems in Engineering, Volume 2012, (2012).
- [12] 野村俊一, カルマンフィルタ R を使った時系列予測と状態空間モデル, 共立出版, 東京都, (2016).