

吉野貴晶 のクオンツ トピックス : NO12

モメンタム&BPR&ROEを入力データとした機械学習モデルの構築と検証

AI/機械学習によるモメンタム効果の改良

- 連載形式でAI（人工知能）と投資手法の関係性を紹介。
- 引き続き、クオンツ手法の一つであるモメンタムが今回のテーマ。

最近、AI（人工知能、以下AI）に関連するニュースが増えています。投資の分野でも研究開発が盛んに行われており、実際に投資手法として利用可能な段階まで進展しています。本レポートでは、AIと投資手法の関係性をご紹介したいと思います。

前回レポートから引き続き、「クオンツとAI/機械学習の融合」をテーマにします。

投資の世界には従来からクオンツという職種が存在します。具体的には、統計解析等の数学的手法を用いて、マーケットの分析や投資戦略の構築をするのが主な仕事になります。今まで、以下のレポートでモメンタム効果への機械学習の応用可能性を検証してきました。今まではシンプルなインプットのみ、具体的にはモメンタムファクター1系列のみを扱っていましたが、今回は入力データ（以下特徴量）を増やします。

今回のレポートでは、特徴量の選定、予測分位毎のリターンシミュレーション、機械学習モデルにおける特徴量寄与度とその解釈まで掲載します。

(ご参考) モメンタム関連の過去レポート

No9：クオンツとAI/機械学習の融合（モメンタム/リバーサル概論）

クオンツ領域において有名な投資手法であるモメンタム/リバーサル効果の概論を紹介しました。

No10：クオンツとAI/機械学習の融合（AI/機械学習によるモメンタム強化）

モメンタム/リバーサル効果の強化（エンハンス）案として、最もシンプルな機械学習モデルである決定木モデルを取り上げ、中期モメンタムを入力データ（1系列のみ）としてパフォーマンスのシミュレーションを行いました。その際、入力データの加工方法による結果の差異も併せて検証しました。

No11：画像認識（CNN）を用いたモメンタム効果のパス情報の取り込み

現在の機械学習の主流である画像認識技術（本レポートではCNN、畳込みニューラルネットワーク）を取り扱いました。データとして、数値情報である個別銘柄のリターン値から画像を作成していきますが、その際に幾つかのデータ前処理を挟むため、その手順も併せて紹介しています。最終的には、予測分位毎のリターンシミュレーションまで掲載しました。

ドメイン知識による入力データ（特徴量）の決定

1. 特徴量の選定

シンプルな問題として扱うために、今までは中期モメンタムリターンの1系列のみを機械学習モデルへの入力データとしていました。このような1系列のみで将来のリターンの説明力が高いモデルが出来れば良いのですが、実際のマーケットは複雑であり、様々な要因がお互いに相互作用しながらリターンを形成していきます。レポートのテーマでもあるモメンタム効果を主軸に扱う上で、そのモメンタム効果の特徴づけたりエンハンスする可能性のある入力データを増やした方が、モデルの効用も改善する可能性があります。

さて、追加の入力データは何にすれば良いでしょうか？

機械学習においては、大量のデータを入れてモデルに学習させれば良い結果がでるはず、と考える向きも見られます。しかし、現実世界では、全く意味の無い、または著しく効果が少ないデータも少なからず存在します。有名な言葉に、「Garbage In Garbage Out」という言い回しがあります。これは、無意味（ごみ）なデータを入力すると、得られる結果は無意味（ごみ）なものになる、という事を表現しています。

では、特徴量はどうやって選択すれば良いでしょうか？現実的には、機械学習を適用しようとする領域の専門家の知見（ドメイン知識）に頼る方が、より良い結果に到達しやすい、と言えるかと思います。今回は、マーケットに携わる先人の研究や、経験則に基づきたいと思います。

結論から言いますと、今回は以下の3系統を特徴量とします。なお、いずれも100分位値化します。

入力データ（3系列）

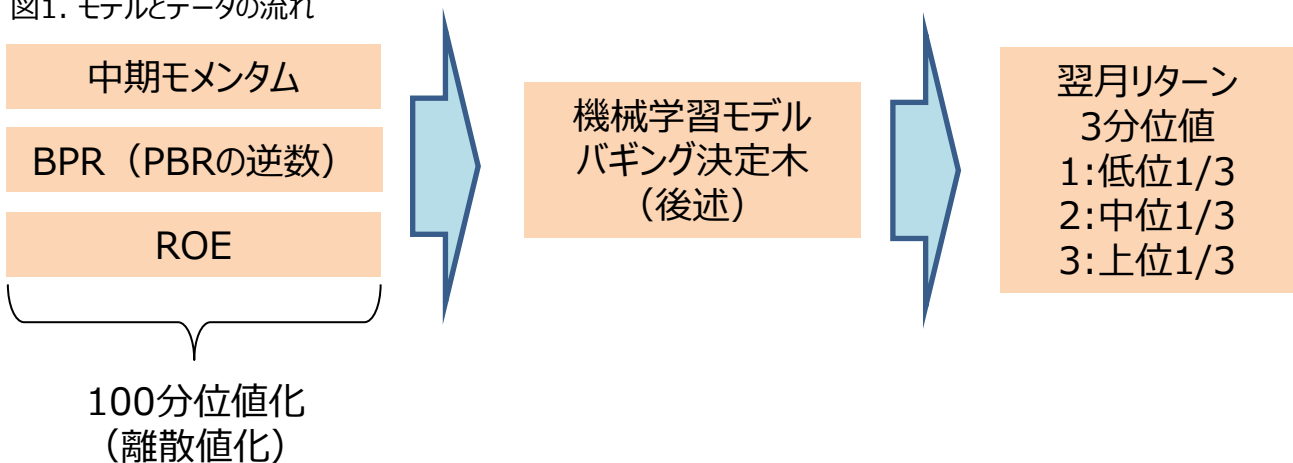
- ・ 中期モメンタム
- ・ BPR（PBRの逆数）
- ・ ROE

出力データは、月次でのリバランスを想定し、以下の1系統とします。

予測データ（1系列）

- ・ 翌月リターン（3分位値）

図1. モデルとデータの流れ



何を入力データ（特徴量）にするか

2. 各特徴量とモメンタム効果の関係性

今回は中期モメンタムリターンに加えて、BPR（PBRの逆数）とROEをインプットとします。BPRとROEですが、経験則的に以下の事象が起こることが知られています。

- ・ 低BPR（高PBR）銘柄はモメンタム効果が示現しやすい
- ・ 高ROE銘柄はモメンタム効果が示現しやすい

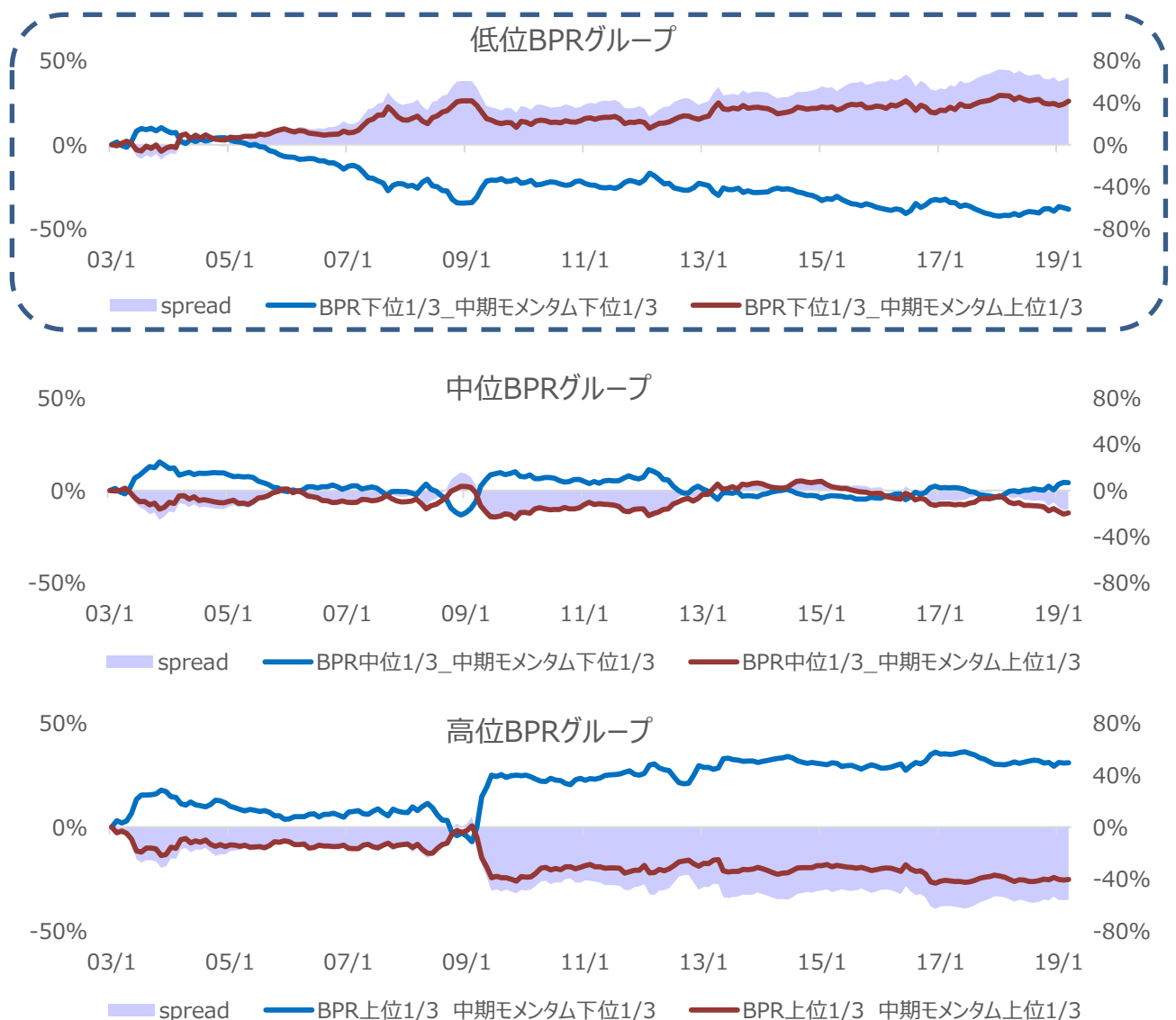
実際に確認してみます。

（検証条件）

- ・ 対象は東証1部（TOPIX銘柄）とし、その中で必要データが取得できる銘柄群とします。
- ・ 3×3分位での分析とします。先にBPRまたはROEで3分位化後、さらに各分位を中期モメンタムリターン値で3分位に分割します。
- ・ spreadはスプレッドリターンであり、上位1/3予想と下位1/3予想との累積リターン差

図2はBPRと中期モメンタムの3×3分析の結果になります。先の説明通り、低BPR（高PBR）銘柄はモメンタム効果が示現しやすい、と言えるかと思えます。

図2. BPRと中期モメンタムの3×3分析（累積リターン）

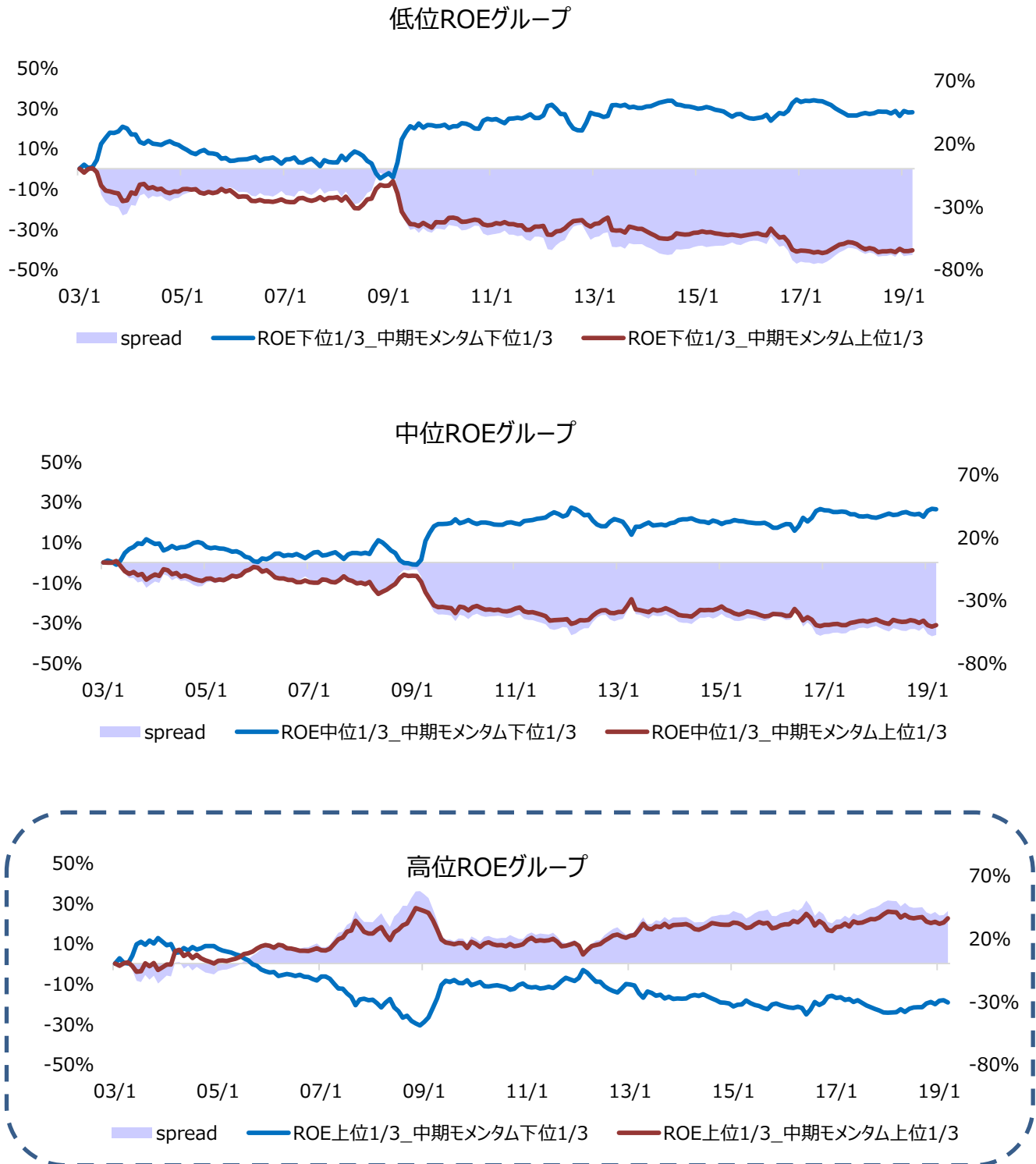


●当資料は、市場環境に関する情報の提供を目的として、ニッセイアセットマネジメントが作成したものであり、特定の有価証券等の勧誘を目的とするものではありません。●当資料は、信頼できると考えられる情報に基づいて作成しておりますが、情報の正確性、完全性を保証するものではありません。●当資料のグラフ・数値等はあくまでも過去の実績であり、将来の投資収益を示唆あるいは保証するものではありません。また税金・手数料等を考慮しておりませんので、実質的な投資成果を示すものではありません。●当資料のいかなる内容も将来の市場環境の変動等を保証するものではありません。

何を入力データ（特徴量）にするか

図3はROEと中期モメンタムの3×3分析の結果になります。こちら、先の説明通り、高ROE銘柄ではモメンタム効果が示現しやすい、と言えるかと思います。このように、モメンタム効果をエンハンス出来る可能性があるこの2系統（BPR、ROE）を、今回は中期モメンタムと共に入力データに採用したいと思います。

図3. ROEと中期モメンタムの3×3分析（累積リターン）



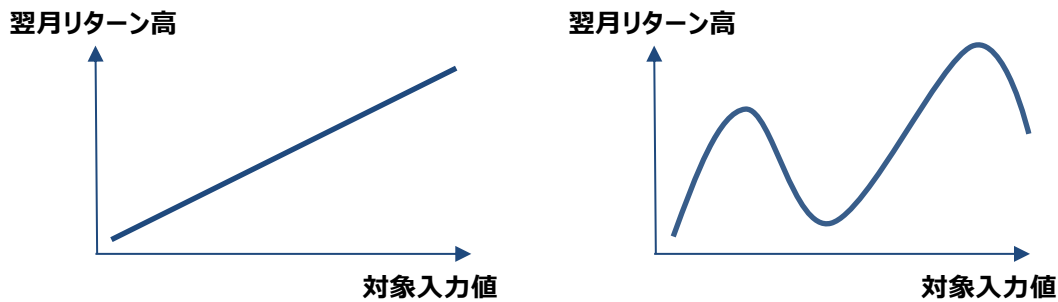
●当資料は、市場環境に関する情報の提供を目的として、ニッセイアセットマネジメントが作成したものであり、特定の有価証券等の勧誘を目的とするものではありません。●当資料は、信頼できると考えられる情報に基づいて作成しておりますが、情報の正確性、完全性を保証するものではありません。●当資料のグラフ・数値等はあくまでも過去の実績であり、将来の投資収益を示唆あるいは保証するものではありません。また税金・手数料等を考慮しておりませんので、実質的な投資成果を示すものではありません。●当資料のいかなる内容も将来の市場環境の変動等を保証するものではありません。

モデルの決定

3. 機械学習を使うべきか？

直線的な関係性、または3×3分析で説明できる関係性のみであるならば、この分析結果をそのまま市場に適用するモデルを作った方が説明性から考えても良いと思います。一方で、線形では無く、ある程度複雑な関係性を相互作用で持っている場合は、非線形型の機械学習モデルに分があります。今回は、このような相互作用の関係性がある程度入り組んでいると仮定して、機械学習モデルを適用します。

図4. 対象入力値と翌月リターンの関係性



※厳密には多項式で表現可能だが、便宜的に非線形の例として表示

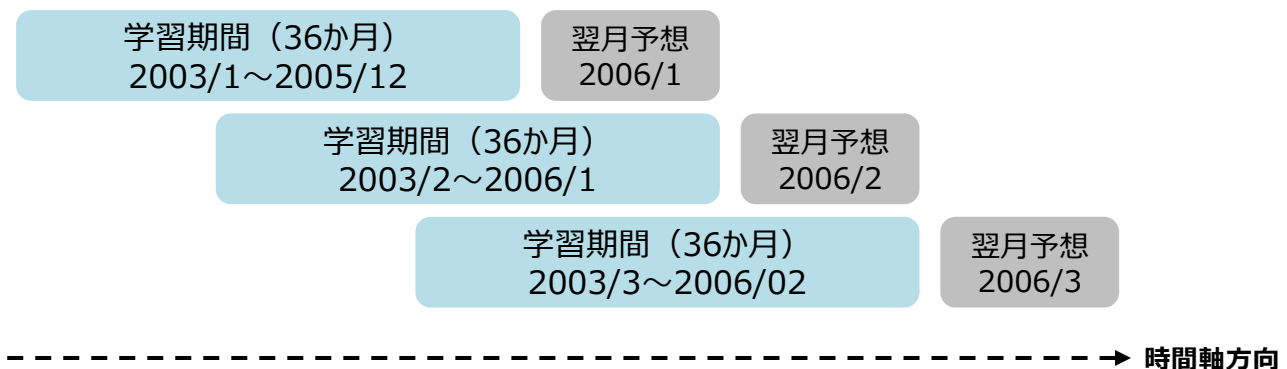
4. バギング決定木の導入

今回は、前々回に引き続き、バギング決定木を導入します。

5. 学習期間とWalk forward Test

より実際のモデル検証に近づけるために、Walk forward Test 方式を採用します。これは、毎月モデルを再学習して、翌月の予測値を作ることを時系列方向に繰り返していきます。毎月最新のモデルになるという意味で、実務で使われやすい手法です。また、重要な変数として、過去どれくらいの期間のデータを学習に使うのか？という点があります。例えば、手に入る過去データは常に全て学習期間に使うという手法もあります。これは感覚的に分かりやすいかと思います。一方、過去Xヶ月分のみを学習期間に使うという方法もあります。これは、ある程度の期間でマーケットの状態が変わっていくとの前提に立っています。まとめると図5の関係性です。今回は、過去36か月を学習期間に使い、Walk forward Test 方式を導入します。なお、この36か月というのは、筆者の経験に基づくものであり、必ずしも正しくは無いかもしれませんが、以降はこの仮定で進めます。

図5. Walk forward Test

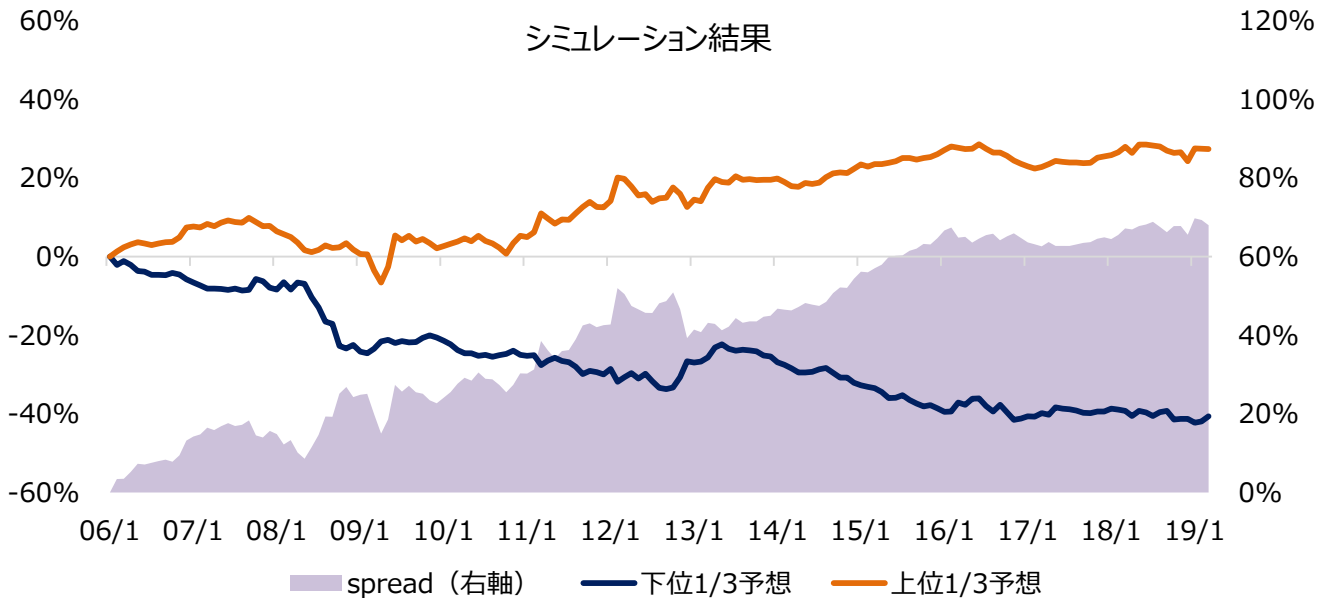


リターンシミュレーションと寄与度推移の確認

6. 各分位毎リターン

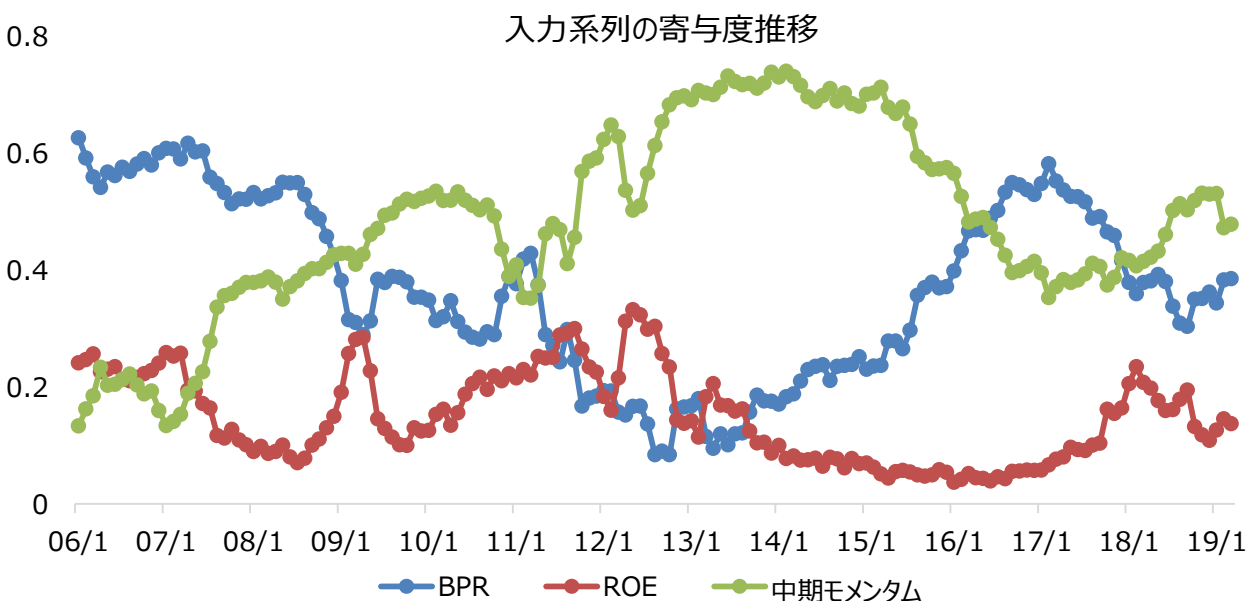
バギング決定木、Walk Forward Testベースでの予測結果を見てみましょう。予想値として吐き出される相対上位1/3、下位1/3のラベルのグループにおいて、等ウェイトポートの月次リターン（全体平均リターン控除後）を累積したものが図6です。ワニ口に開いており、かつSpreadリターン（上位1/3予想と下位1/3予想ポートのリターン差）は右肩上がりです。このことから、予測モデルとして一定の効果がありそうです。

図6.バギング決定木、Walk Forward Testベースでの予測結果（累積リターン）



また、決定木モデル系のモデルでは、各入力系列が予測値の決定にどれくらい寄与したかを確認できます。以下が時系列に見た寄与度推移です（解釈は後述）。

図7.入力系列寄与度の時系列推移



●当資料は、市場環境に関する情報の提供を目的として、ニッセイアセットマネジメントが作成したものであり、特定の有価証券等の勧誘を目的とするものではありません。●当資料は、信頼できると考えられる情報に基づいて作成しておりますが、情報の正確性、完全性を保証するものではありません。●当資料のグラフ・数値等はあくまでも過去の実績であり、将来の投資収益を示唆あるいは保証するものではありません。また税金・手数料等を考慮しておりませんので、実質的な投資成果を示すものではありません。●当資料のいかなる内容も将来の市場環境の変動等を保証するものではありません。

機械学習モデルによるエンハスト効果の検証

7.機械学習を実施した意味があったのか？

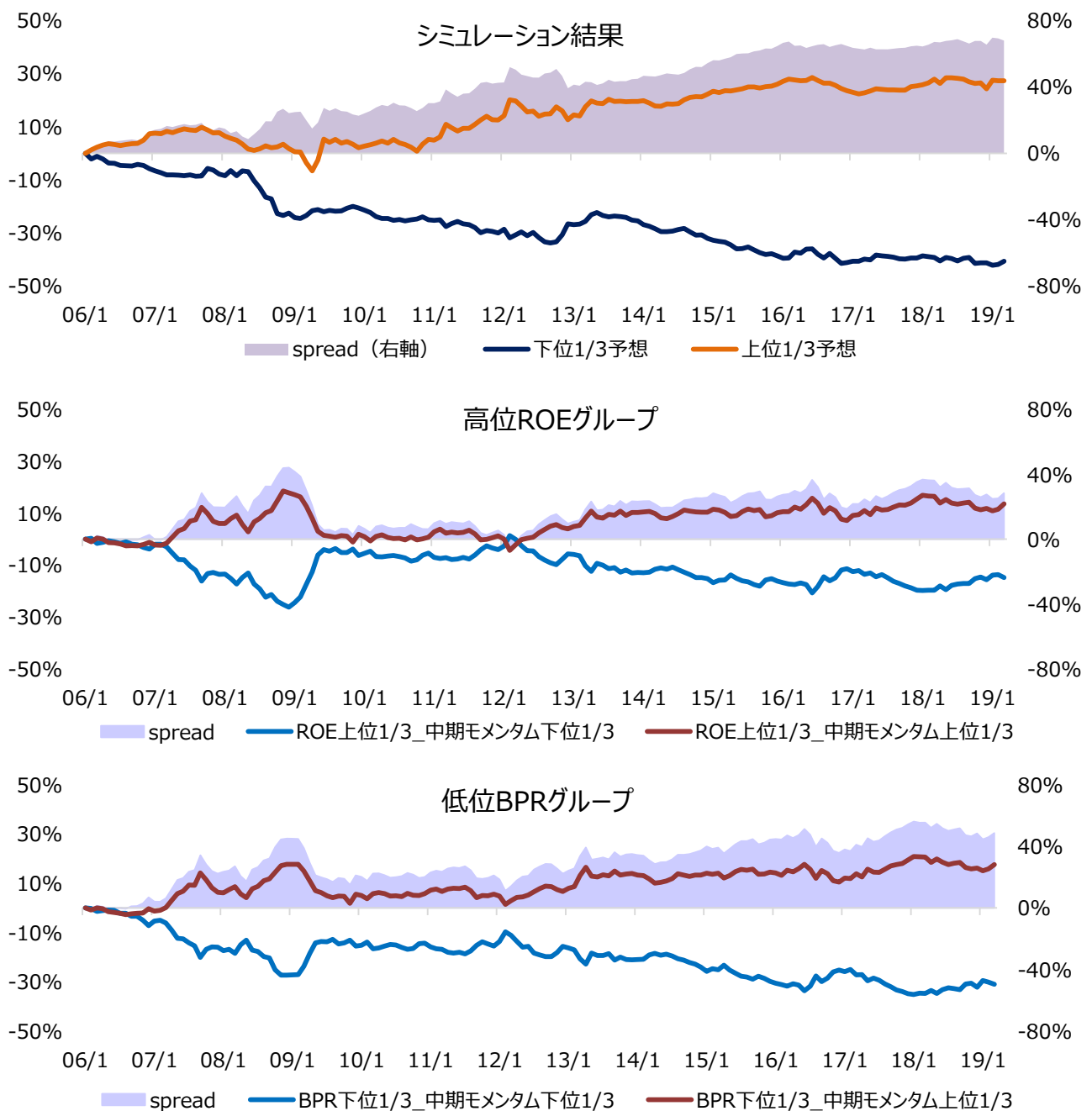
予測モデルとして一定の効果がありそうなことはわかりましたが、もう一つ問題があります。それは、果たして機械学習を実施した意味があったのか？ということです。出てきた結果が比較対象となる結果に劣っているようなら、機械学習モデルは不要です。

まず、冒頭の入力データとしてBPRとROEに決めた理由である、以下の組み合わせでの3分位累積リターンを見てみます。

- ・低BPR（高PBR）銘柄はモメンタム効果が示現しやすい
- ・高ROE銘柄はモメンタム効果が示現しやすい

図を比較すると、今回のシミュレーション結果の方が分位リターン、スプレッドリターン、リターンの傾向（常に右肩上がり）が好まれる）ともに勝っており、機械学習モデルの成果が出ているといえます。

図8. シミュレーション結果と高位ROEグループ、低位BPRグループのモメンタム効果との比較



●当資料は、市場環境に関する情報の提供を目的として、ニッセイアセットマネジメントが作成したものであり、特定の有価証券等の勧誘を目的とするものではありません。●当資料は、信頼できると考えられる情報に基づいて作成しておりますが、情報の正確性、完全性を保証するものではありません。●当資料のグラフ・数値等はあくまでも過去の実績であり、将来の投資収益を示唆あるいは保証するものではありません。また税金・手数料等を考慮しておりませんので、実質的な投資成果を示すものではありません。●当資料のいかなる内容も将来の市場環境の変動等を保証するものではありません。

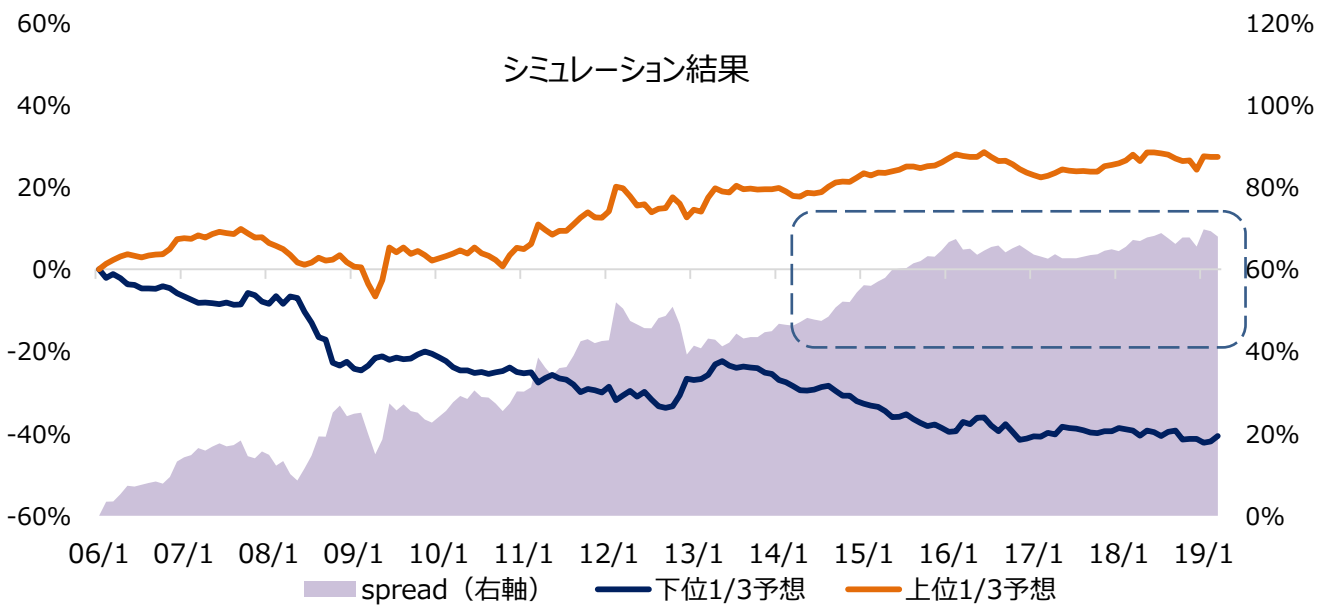
機械学習モデルによるエンハスト効果の検証

8. 入力系列単体（今回はBPRと比較）には勝っているか？

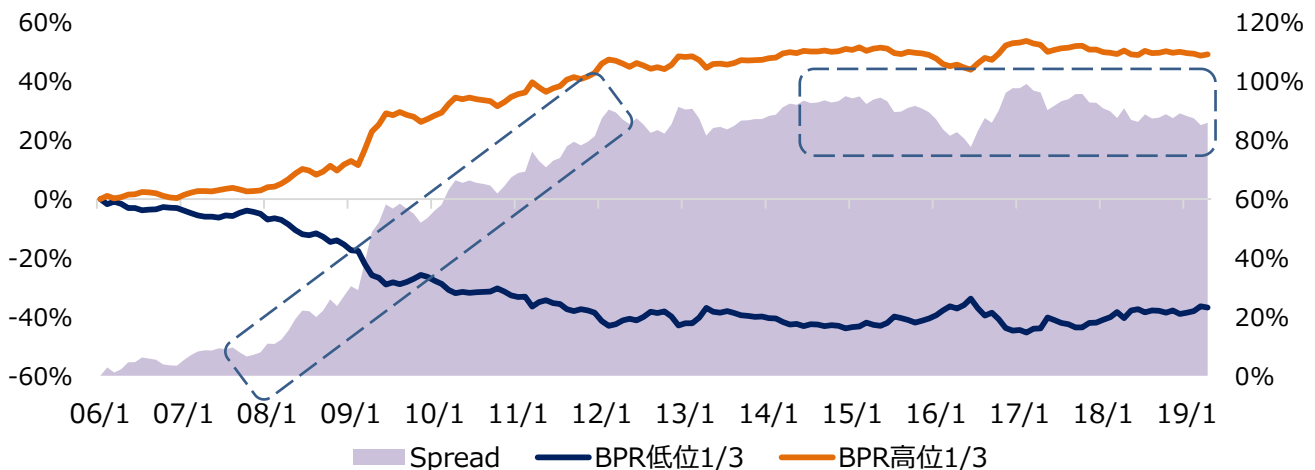
さて、今度は元の入力系列のみで見た3分位累積リターンを見てみたいと思います。中期モメンタム、ROE、BPRの中で、同じ期間に累積リターンが良いBPRを例に取ってみます（本レポートでは中期モメンタムとROEのリターン図は割愛）。図9と図10を比較すると、BPRによる累積スプレッドリターンはBPRの単純3分位の方が勝っています。しかし、この数字だけで単純にBPRの3分位値でポジションを取ればよい、と考えるのは早計です。先にBPRにおける、破線で囲んだ部分を見てみたいと思います。BPRの方は、2008年から2012年ごろに累積リターンのほぼすべてを稼ぎ切り、その後は一進一退の横ばいです。また、2017年以降は緩い右下がりとなり、昨今のパフォーマンスは不調です。対して今回のシミュレーション結果は、期間を通じて全体的に右肩上がり、2017年以降も右下がりにはなっていません。

このように、今回のシミュレーション結果は、入力系列単体（今回はBPRを取り上げています）や、組み合わせである高ROEグループや低BPRグループのモメンタム効果とも違う特徴を持っている、と言えます。

図9. シミュレーション結果とBPRの単純3分位とのシミュレーション結果



BPRをそのまま3分位に分けた場合の累積リターン（線形）



●当資料は、市場環境に関する情報の提供を目的として、ニッセイアセットマネジメントが作成したものであり、特定の有価証券等の勧誘を目的とするものではありません。●当資料は、信頼できると考えられる情報に基づいて作成しておりますが、情報の正確性、完全性を保証するものではありません。●当資料のグラフ・数値等はあくまでも過去の実績であり、将来の投資収益を示唆あるいは保証するものではありません。また税金・手数料等を考慮しておりませんので、実質的な投資成果を示すものではありません。●当資料のいかなる内容も将来の市場環境の変動等を保証するものではありません。

予測値の分布を確認する

9. 3次元散布図で見る予測値の分布

シミュレーションという観点では、モデルから予測値が吐き出されればよい、となりますが、予測値の分布はいったいどのようになっているのか確認してみたいと思います。入力が3系列、出力が3レベルですので、3次元散布図に描画することが可能です。

まず、図10は2019年3月値を入力にして2019年4月を予測した場合における、上位1/3予想値の分布図です。奥行きがある空間を創造してください。右奥に伸びているのがBPR、左奥に伸びていくのがROE、そして縦軸が中期モメンタムです。いずれでも100分位値です。図を見ると、モデルによる上位1/3予想の分布が偏っていることがわかります。具体的には、BPRは低め、ROEは高めの領域であり、かつ中期モメンタムが高すぎる領域以外、となります。

そして図11は下位1/3予想値の分布図です。やや散らばっていますが、総じて中期モメンタムの分位値がかなり高い領域に分布しています。

ROEはしばしばクオリティー、低BPRはグロースと呼ばれることがあります。高モメンタムは買われすぎとも解釈できますので、「買われ過ぎていないクオリティーグロース株が翌月好調になるだろう」とモデルが表現していると言えます。このように、入力を3系列にしたことによって、3系列の相互作用をそれぞれ加味した予測値の分布をモデルが学習したことがわかります。

図10. 上位1/3予想の分布図

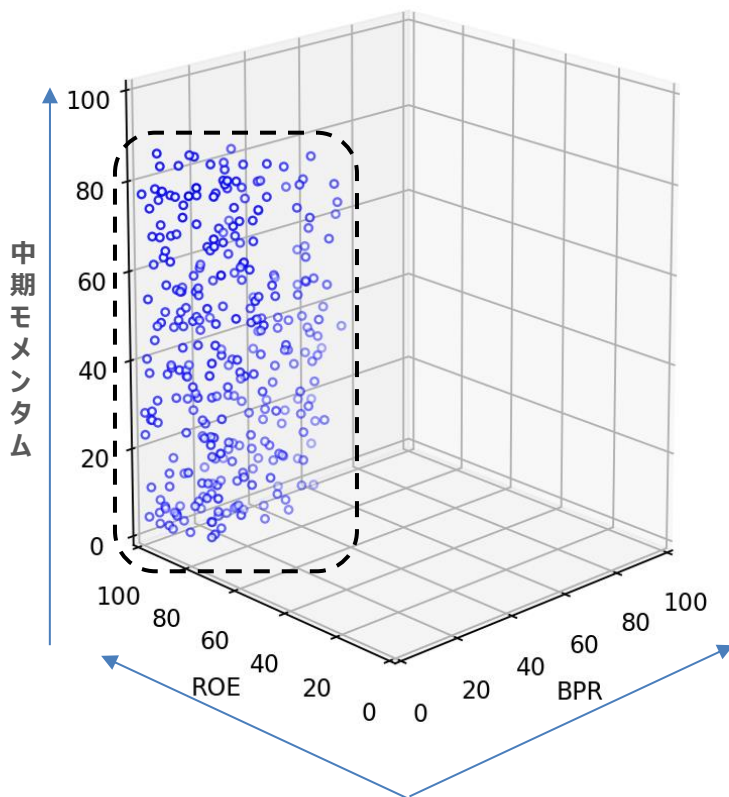
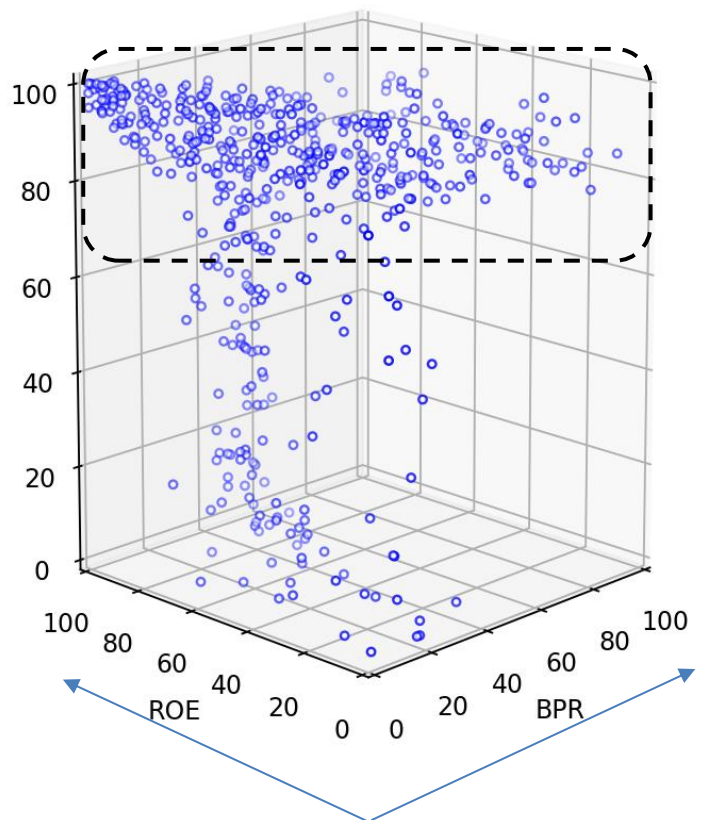


図11. 下位1/3予想の分布図



結果に解釈を与える

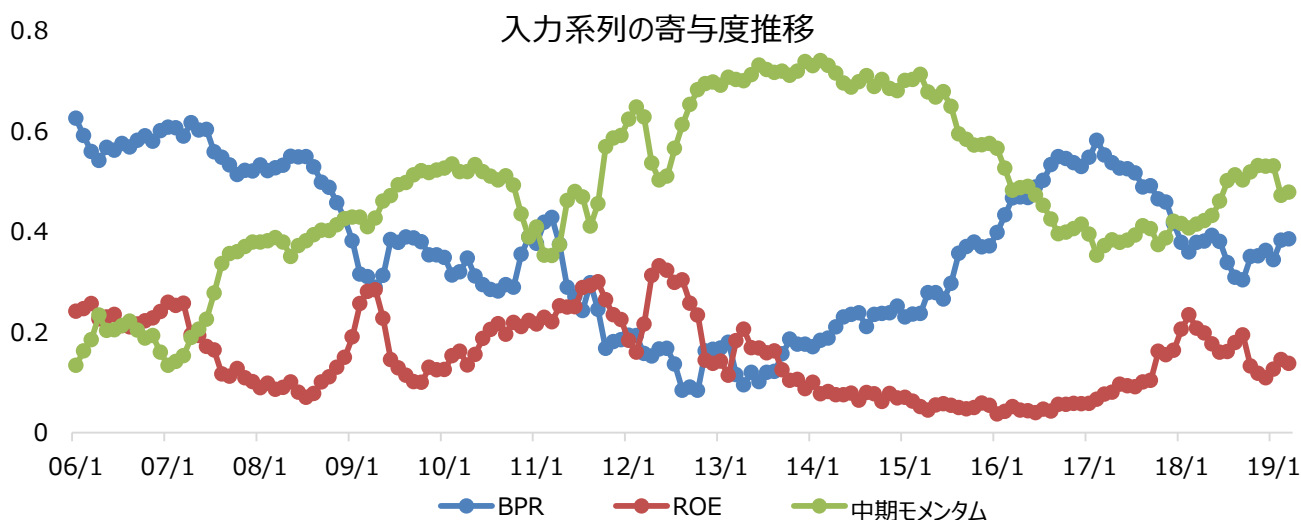
10. 特徴量寄与度の推移

時系列に見た寄与度推移に立ち戻りたいと思います。この寄与度推移を解釈してみます。

中期モメンタムに着目すると、大きく上下していますが、比較的最近に目をやると、2018年1月以降は寄与度1位になっています。中期モメンタムを形成する要素はいろいろと考えられますが、価格動向の重要性が高い可能性が示唆されます。また、よく日本株で注目されるBPRについては、リーマンショック前（寄与度50%以上）に対して、最近の寄与度（40%弱）は低くなっています。バリュー（BPR高）でもグロース（BPR低）でも将来の予測値に寄与するのであれば寄与度は上がるので、単純にBPRによる予測力が下がっていると言えます。

一方、期間を通じてROEの寄与度は上下動こそあれ、低めとなっています。しかし、この結果を受けて、ROEは重要ではない、と切り捨ててしまうのは早計です。先ほどの3次元分布図を見ると、高ROEの領域に上位1/3予想値の分布が存在していたかと思います。ROEによる切り分けもある程度意味を成している可能性もあるからです。

図12. (再掲) 入力系列寄与度の時系列推移



11. 今後の展開

次回以降も、引き続きAIと投資手法の関係性をご紹介したいと思います。具体的に決まっていますが、モメンタム効果を継続するか、ファクターモデルに拡張するか、自然言語処理に戻るか、ボラティリティの予測に挑戦するか、等の範囲を考えています。

(筆者の都合でテーマが変わる場合があります。)

～執筆者の紹介～

吉野貴晶 (写真: 右)

「日経ヴェリタス」アナリストランキングのクオンツ部門で16年連続で1位を獲得。ビッグデータやAIを使った運用モデルの開発から、身の回りの意外なデータを使った経済や株価予測まで、幅広く計量手法を駆使した分析や予測を行う。



高野幸太 (写真: 左)

ニッセイアセット入社後、ファンドのリスク管理、マクロリサーチ及びアセットアロケーション業務に従事。17年4月に投資工学開発室に異動後は、主に計量的手法やAIを応用した新たな投資戦略の開発を担当する。