吉野貴晶 のクオンツ　トピックス：*NO17*

NGBoostを用いた市場マクロ変数によるモメンタム効果の将来予測

機械学習モデルで市場マクロ変数からモメンタムを予測できるか

- クオンツ領域の投資手法を紹介
- NGBoostで翌月のリターンと標準偏差を同時に推定

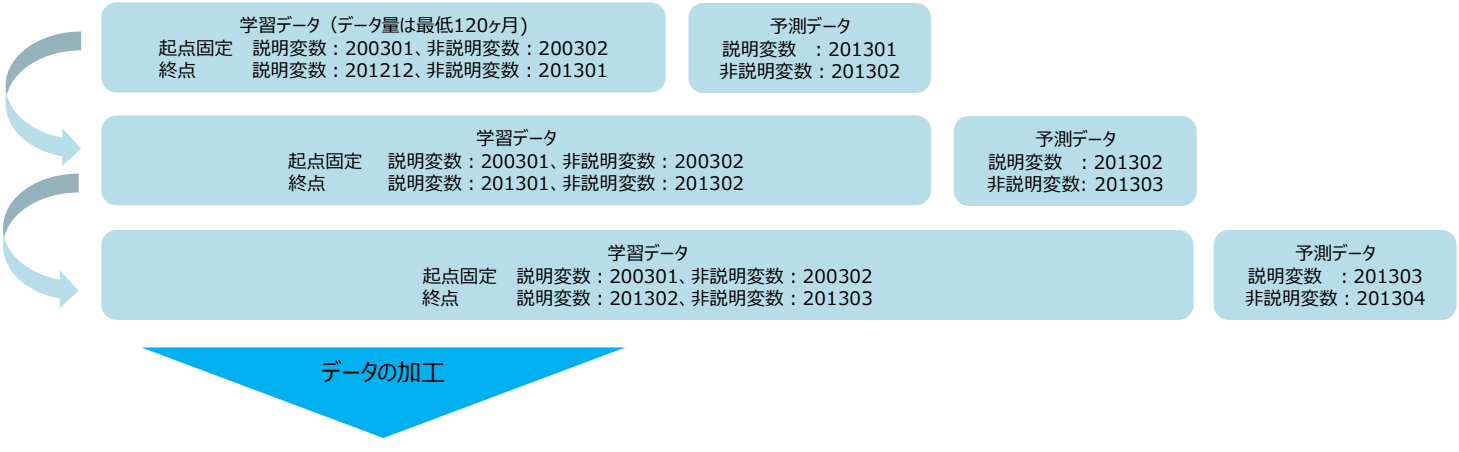
前回のレポートでは、市場から取得されるマクロ関連変数（前回レポートで市場マクロ変数と定義）が日本株のファクタースプレッドリターンにどのように影響を与えるかを確認しました。この際、市場マクロ変数とファクタースプレッドリターンは1対1の関係性でした。今回のレポートでは、入力変数を前回調査した市場マクロ変数全てとし、これらの市場マクロ変数が翌月のモメンタムファクタースプレッドリターンを予測できるか、をテーマにします。予測モデルの枠組みには、NGBoost（Natural Gradient Boosting）を採用します。分布を指定することで、平均値と標準偏差の同時推定が可能なモデルになります。

1.データハンドリング

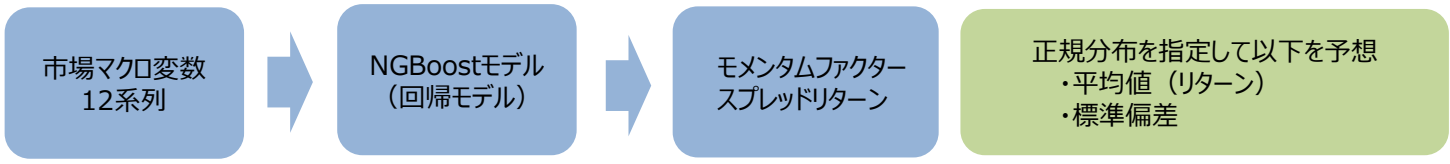
最初に、本レポートにおけるデータハンドリングの枠組みをご説明します。

・説明変数は月次の市場マクロ変数（12系列、定義は前回レポート参照）
・非説明変数は月次モメンタムファクタースプレッドリターン（モメンタムファクターは過去レポート（No9）参照）
・月次でモデルを作成し直し、学習用のデータ期間は1ヶ月ずつ増やす
　→ 月次粒度だとデータセット数がそもそも少ないので、なるべく増やしていくため

図1.データハンドリング



・説明変数：学習データ期間内で離散化（5分位,離散化に関しては過去レポート（No10）参照）
・非説明変数：小数点以下2桁になるように丸める（モデルの不必要な複雑さを避ける）
・学習データのうち、30%をパラメーター選択の評価用データとして利用
・NGBoostにおける１回の決定木作成には、50%のデータを利用
　→上記の評価用データの分割と併せると、（1-30%)*50%=35%のデータが利用される
・乱数による影響を考慮するため、毎時点でNGBoostを100回試行して平均値を採用
（NGBoostのBoosting回数とは別。例：Boosting回数が100回のモデルを100個作成して平均値を採用）



ニッセイ アセットマネジメント

NGBoostで平均値と標準偏差を予測する

2.NGBoostによる分布推定

今回は、機械学習手法の一つであるNGBoost（Natural Gradient Boosting）を利用します。一般的によく使われる回帰型の機械学習のような点推定（一つの予測値を出す）ではなく、「分布」を予測します。単純な例で言うと、「明日の最高気温は何度ですか？」と言う問いに対して、「25度です」と返すのが点推定です。一方で、分布を返す場合、「平均値は25度、95%の確率で22度から28度に収まると予想します」、となります。NGBoostは後者の予測を試みるモデルです。今回は、予め分布を正規分布と指定した上で、その分布を特徴付ける平均値と標準偏差を算出します。

3.予測結果の確認(平均値)

さて、結果を確認してみたいと思います。回帰モデルを利用しているので平均値予測は数値になります。投資する上で重要なこととして、将来上昇するのか下落するのか、があります。この点を投資戦略につなげられないか検討します。「予測が0%以上の場合は買い、０％未満の場合は売り」として翌月の実績リターンを足し上げた結果を確認してみましょう（図2）。比較として、常に「買い」とした場合のグラフを載せていますが、上記ルールにおいてモデル予測値の累積リターンが上回っています。差が出ている部分を見ると、破線囲みの領域で差が出ており、これはモメンタムファクタースプレッドリターンがマイナスになる領域です。どうやらモデル予測値の優位性は特にマイナス方向の予想で現れているようです。

図2.モデル予測サインによる累積リターンとベース（常に買い）累積リターンとの比較

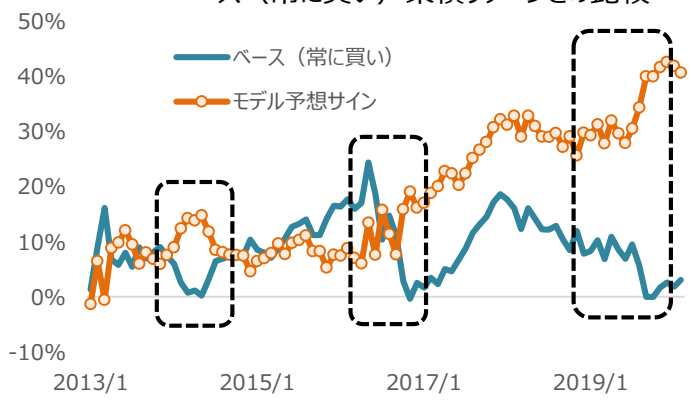


図3.上昇時と下落時の正答率比較

符号正答月数	47
全予測月数	86
符号正答率	54.7%

上昇正答月数	25
上昇月数	47
上昇正答率	53.2%
全て上昇予想とした場合の正答率	54.7%

下落正答月数	22
下落月数	39
下落正答率	56.4%
全て下落予想とした場合の正答率	45.3%

4.予測結果の確認(標準偏差)

NGBoostは標準偏差も予測値として算出しますので、こちらを確認してみます。NGBoostにおける初期設定で正規分布を指定しています。予測値を中心とした95%予想の範囲内に、実際の観測値が収まるかを確認してみました（図4）。標準偏差の予測値がそもそも小さい値にはならなかったので、レンジがかなり広いですが、概ね実際の観測値はレンジ内に収まっています。また、金融実務で過去実績ベースの標準偏差（ヒストリカル標準偏差）が利用されることがあります。比較してみると、モデル予測値は毎月市場マクロ変数の影響を受けて標準偏差の予測値が上下していることが分かります（図5）。

図4.上下2標準偏差バンドと実測値

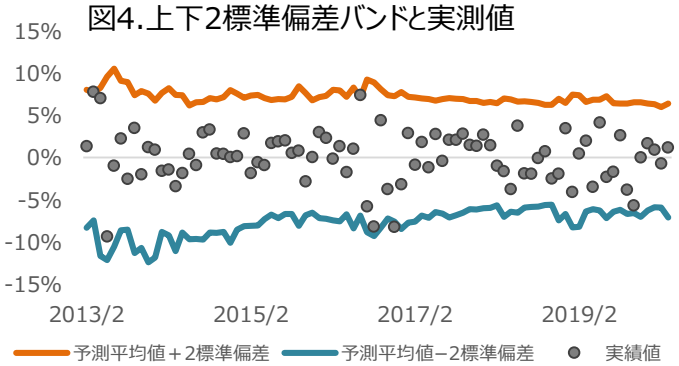
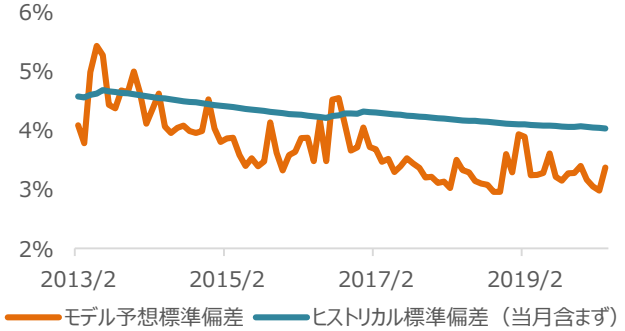


図5.モデル予想標準偏差とヒストリカル標準偏差の比較

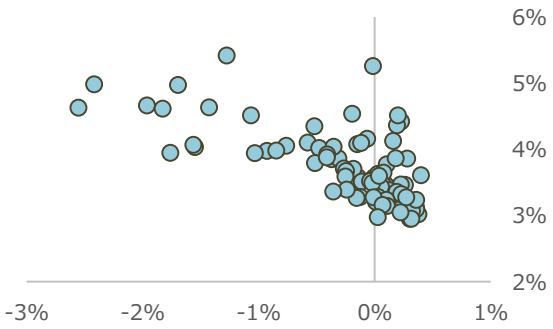


機械学習につきまとうモデルの解釈性（SHAP）

5. 平均値予測と標準偏差予測の関係性

平均予測値と標準偏差の予測値の関係性を散布図で見た結果が図6になります。マイナス方向に大きな値の予測値を示していることが分かります。同時に、大きなマイナスの予測平均値を出す場合に、標準偏差予測もより大きな値になっています。これらの予測に寄与している市場マクロ変数はどれなのでしょうか？

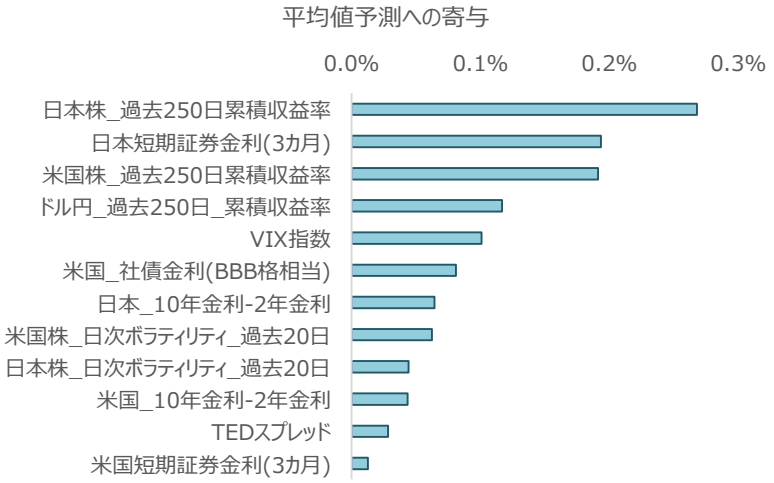
図6. 平均値予測（横軸）と標準偏差予測（縦軸）



6. SHAPによるモデルの解釈性

これまで予測結果を見てきましたが、どの市場マクロ変数が予測に寄与したかは分かりません。機械学習の利用につきまとう、「モデルの解釈性」と言える領域ですが、今回はSHAPと言う手法を使って、各市場マクロ変数がどれくらい影響を与えたかを見てみたいと思います。今回のモデルであるNGBoostにSHAPを適用することで、予測に対する各市場マクロ変数の寄与が簡易的に算出できます。図7が平均値予測への各変数の寄与になります。今回のモデルでは、日本株の過去250日の累積収益率が最も寄与が高くなっています（相関が高い変数同士で寄与が分割されている可能性があり注意が必要です）。

図7.学習期間の入力データに対するSHAP Value絶対値の平均(上位ほど寄与が大きい)



7. 終わりに

次回レポートでは、SHAPによる市場マクロ変数の寄与を個別に見てみたいと思います（筆者都合で変更になる場合があります）。

参考文献
1. NGBoost User Guide
2. Tony Duan, Anand Avati, et. al. *NGBoost: Natural Gradient Boosting for Probabilistic Prediction*. v3. 2020 available on arXiv

～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。ビックデータや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。

髙野幸太（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。