

機械学習を用いた J-REIT 市場のパフォーマンス要因分析

2020年6月3日

株式会社三井住友トラスト基礎研究所

REIT 投資顧問部 研究員 小西勝也

- ・ 本稿では、J-REIT 市場のパフォーマンスに影響を与える要因を定量的に分析した。分析手法としては、「株式市場」、「債券市場」、「不動産市場」等の要因別に分解し、線形回帰分析を行う方法が従来から広く採用されているが、金融資産価格は通常、非線形に変動することが知られており、線形性を前提とした回帰モデルでは十分な分析精度が得られないことも多い。
- ・ そこで、本稿では、線形回帰分析に加えて、要因等との関係性を非線形構造¹で捉え、かつ一般的に高い分析精度を得ることが可能な機械学習モデルによる分析を行った。機械学習モデルの一般的な懸念としては、分析結果の判断根拠が不明瞭になる点²があげられることが多いが、本稿では判断根拠の把握が容易なことで知られる決定木²ベースの機械学習モデルを使用した。
- ・ 線形回帰モデルと比較した結果、機械学習モデルを用いることで、より多くの要因を対象としながら、より高い精度が得られることを確認できた。更に、今回採用した機械学習モデルでは、結果に至った要因も各特徴量の寄与度として把握できるなど一定の説明力を維持することも可能であることを確認できた。
- ・ 分析精度の高い機械学習モデルの結果から、J-REIT 市場のパフォーマンス要因は、長期的には市場環境に応じて相当程度変化していること、J-REIT 市場のパフォーマンスは金融危機時には株式との連動性が高まり、2019年には債券代替の性格を強めていたことが、定量的に確認された。足元の J-REIT 市場のパフォーマンス要因を定量的に正確に把握する意義は大きく、本手法は当面の市場動向を見通す上でも有益だと考えている。

1. はじめに

J-REIT 市場のパフォーマンス要因を定量的に分析する手法として、「株式市場」、「債券市場」、「不動産市場」等の要因別に分解し、線形回帰分析を行う方法が従来から広く採用されている。しかし、一般的に金融資産価格は非線形性をもって変動することが知られており、J-REIT もその例外ではない。そのため、線形性を前提とした回帰モデルを用いた分析では十分な推定精度が得られないことも多い。一方で、近年は機械学習モデルと呼ばれる先進的手法がその高い分析精度から注目を集めており、これら手法では分析対象とそれに影響する要因や、各要因間での非線形構造に対応した分析も可能である。そこで、本稿ではこの機械学習モデルを用いた推定結果と、従来の線形回帰モデルの結果の比較を行い、機械学習モデルが分析精度向上に寄与するか検証を行った上で、J-REIT 市場のパフォーマンス要因について、各要因の長期でみた寄与度の時系列変化を分析し、市場環境に応じたパフォーマンス要因の変化を考察した。

¹ 一次式(直線)では表現することのできない関係。相関・逆相関等の単調な変化に限らず、複合的な動きにも対応が可能。

² 木構造(樹形図)を用いる予測モデル。与えられたデータを用いて条件分岐を行っていくことで、目的とする値を導く。

2. 検証・分析方法

2.1 使用データ

目的変数には東証 REIT 指数(配当込)の日次収益率を使用する。本稿では目的変数に影響する要因を、線形回帰では説明変数、機械学習では特徴量と表現している。今回、特徴量(線形回帰では説明変数)の数は3系列と11系列(詳細は後述)の2通りでの分析を行い、特徴量の増加による推定精度への影響も同時に検証した。特徴量(図表1参照)は、先行研究(参考文献[1][2][3])でも多く使用される3系列(図表1赤背景色)に加え、J-REIT 市場パフォーマンスに影響を与えられとされる他の8系列の特徴量を追加選択した。

本稿で用いる機械学習モデルでは、線形回帰モデルと比べ、採用する特徴量の選択上の制約や、特徴量の増加により推定精度が低下する等の懸念が少ないことも長所の一つと考えられる。

なお、分析には機械学習モデルに入力するサンプル数を十分に確保するため日次データを用いるが、不動産関連データは月次のデータであるため図表1の9~11は、月末にのみ値が変化するデータとして扱う。また、特徴量には変化率の算出や、標準化や外れ値の処理等の前処理を行っている。

図表1 使用した特徴量とデータソース

	要因分類	データ	データソース		要因分類	データ	データソース
1	株式市場	TOPIX(配当込)	Bloomberg	7	債券市場	日本10年国債利回り_前日	Bloomberg
2		TOPIX(配当込)_前日	Bloomberg	8		米国10年国債利回り_前日	Bloomberg
3		S&P500_前日	Bloomberg	9	不動産市場	オフィス・都心5区募集賃料単価	三鬼商事
4	債券市場	NOMURA_BPI	Bloomberg	10		オフィス・都心5区空室率	三鬼商事
5		NOMURA_BPI_前日	Bloomberg	11		オフィス・都心5区貸室床面積	三鬼商事
6		日本10年国債利回り	Bloomberg				

注)末尾に「_前日」とつく系列は、時点tの目的変数に対し、t-1時点のデータを用いて回帰することを表す。

(出所) 三井住友トラスト基礎研究所

2.2 使用したモデル

1. 線形回帰モデル

線形回帰モデルは、統計学に基づく一般的な回帰モデル。Python の scikit-learn ライブラリを用いて実装。単純化のため正則化³等は行っていない。

2. ランダムフォレスト(Random Forest 以下、RF)

機械学習モデル使用の際の懸念として、どのような情報を基に判断してその結果に至ったのかが不明瞭となる点あげられることが多いが、本稿では分析結果の判断根拠が比較的明瞭なことで知られる決定木ベースの機械学習モデルを用いることで、その懸念に対処している。

RF は、決定木を用いたアンサンブル学習の一つであり、精度の低いモデルである弱学習器を複数組み合わせることで精度の高い一つの学習モデルを構築する手法である。技術的には、独立する多数の弱学習器を並列に構築し、その結果で多数決や平均をとるバギングという手法を用いる。一般的に、精度面では後述するブースティングを用いるモデルに劣ることが多いが、より少ない計算時間で学習できる等の特徴がある。本分析では Python の scikit-learn ライブラリを用いてモデルを実装。ハイパーパラメータは簡単なグリッドサーチを行い、設定している。

3. 勾配ブースティング決定木(Gradient Boosting Decision Tree 以下、GBDT)

GBDT は、RF と同じ決定木ベースのアンサンブル学習を行う機械学習モデルである。しかし、RF とは異なり弱学習器の構築にはバギングではなく、ブースティングという手法を用いる。ブースティングでは目的変数と、それまでに構築された弱学習器により算出した予測値との差(最初の弱学習器構築の際は目的変数そのもの)に対して学習を行う新たな弱学習器を逐次的に構築していき、その各弱学習器の結果の最終的な和をとることで予測値が計算される。また近年は、その高い分析精度が評価され

³ 過学習を防ぎ、汎化性能(未知のデータに対する推定精度)を高めるため、モデルの複雑さに対して罰則を与える手法

Kaggle等のデータ分析コンペティションでも頻繁に用いられる学習モデルである。本分析ではPythonのxgboostライブラリを用いてモデルを実装。ハイパーパラメータは簡単なグリッドサーチを行い、設定している。

2.3 検証・分析手順

1. ベンチマークとなる一般的な線形回帰モデル(説明変数3つ)の性能を線形回帰モデルの精度指標として頻繁に用いられる決定係数により評価(図表2)。
2. 線形回帰モデル及び機械学習モデル(RFとGBDT)を用いて推定を行い、各モデル間での精度比較を行う。評価指標には、回帰モデル(非線形モデルを含む)の性能評価によく使用される平均平方二乗誤差(以下、RMSE)を用いる。各モデルの学習期間は、①東証REIT指数算出開始以降の2003年4月1日～2019年12月31日の全期間、②2017年1月1日～2019年12月31日の直近3年間、③全期間の中で学習期間3年をとりながら開始月を1ヵ月ずつ後へずらす3年ローリング、の3パターンで行う。また、学習期間終了日の翌日から1ヵ月間をテスト期間とし、学習モデルの汎化性能の検証も行う。
3. 精度比較の結果、最も精度の高い学習モデル(GBDT特徴量11系列)を用いて、推定結果に対する各特徴量の寄与度を求める。本機械学習モデルでは、モデル構築過程における各特徴量の推定誤差低減への貢献度により決まる相対的重要度を用いることで、その推定結果への各要因の寄与度を求めることが可能である。まず、全期間でみた寄与度を確認し、次に3年ローリングでの算出により、学習期間3年で算定した寄与度の推移(時系列変化)を確認する。

3. 分析結果

3.1 モデルの精度比較

図表2記載の線形回帰モデルの当てはまり(精度)を表す決定係数の値は、いずれの学習期間でも0.3を下回っており、分析対象に対する説明力が低いことがわかる。なお、RFやGBDTといった非線形モデルの性能評価には、前述した評価指標がより適していると考え、各モデルの精度比較はRMSEを用いて行った。

モデルの精度比較の結果は図表3の通りである。RMSEは小さいほど推定精度が高いことを示す。3パターンの学習期間及びそのテスト期間のRMSE値を見ると、多くの期間で「GBDT > RF > 線形回帰モデル」の順に推定精度が高くなっていることが確認できる。また、特徴量3系列と特徴量11系列の場合では、概ね後者の推定精度の方が高く、特に機械学習モデルにおいて特徴量増加が推定精度向上に寄与していることがわかる。なお、3年ローリングの場合のテスト期間では線形回帰の結果が機械学習モデルを上回っており、学習モデルに入力していない未学習のデータ(未知のデータ)に対する推定精度では、機械学習モデルが安定して最も高い精度を示す結果とはなっていない。

図表2 線形回帰モデルの推定精度(決定係数)

	全期間	直近3年	3年ローリング(平均)
決定係数(R2)	0.26	0.09	0.27

(出所) 三井住友トラスト基礎研究所が作成

図表3 評価指標(RMSE)に基づく各モデルの精度比較

	全期間		直近3年		3年ローリング(平均)	
	学習期間	テスト期間	学習期間	テスト期間	学習期間	テスト期間
線形回帰_3	1.15%	0.54%	0.57%	0.46%	1.10%	1.02%
線形回帰_11	1.12%	0.53%	0.56%	0.46%	1.08%	1.02%
RF_3	1.10%	0.57%	0.55%	0.43%	1.12%	1.08%
RF_11	1.11%	0.48%	0.52%	0.44%	1.07%	1.06%
GBDT_3	0.87%	0.48%	0.37%	0.40%	0.79%	1.05%
GBDT_11	0.76%	0.39%	0.14%	0.39%	0.59%	1.05%

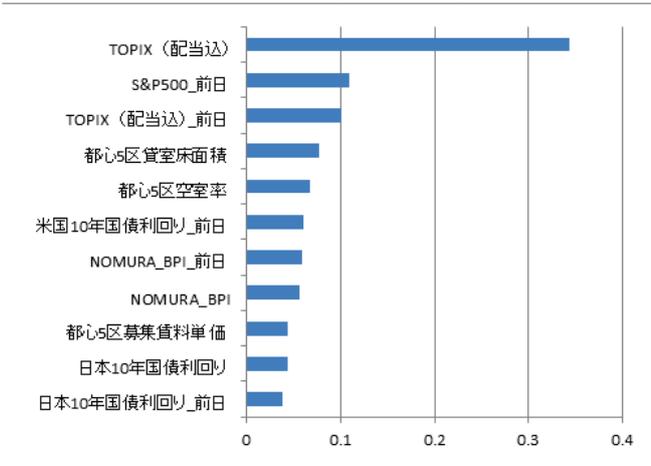
注)モデル名の末尾につく「_3」や「_11」は各モデルに入力する特徴量の系列数を表す。

(出所)三井住友トラスト基礎研究所が作成

3.2 特徴量の相対重要度を用いた変動要因分析

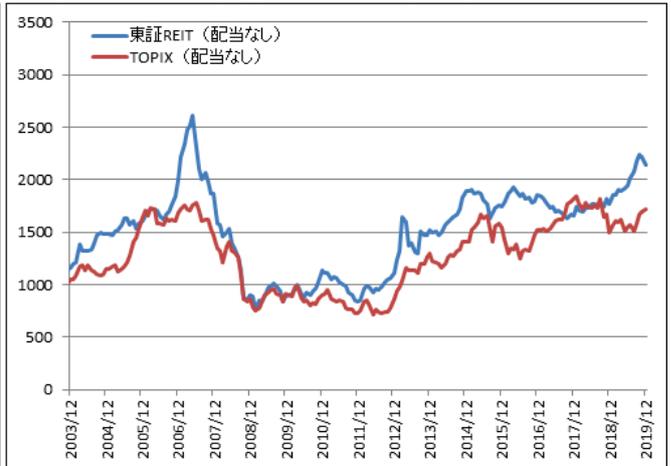
GBDT (特徴量 11 系列) を用い、全期間で算出した各特徴量の相対重要度は図表 4 の通りである。TOPIX (配当込) の重要度が最も高く、株式市場の影響を強く受けていることがわかる。長期ではインカムリターンによる格差が大きくなるため、資産価格変動の連動性、影響を確認するための図表 5 のグラフには配当なし指数を採用している。なお、2014 年半ば以降、TOPIX と REIT 指数は順相関から逆相関へと関係性が変化しているが、このような変化に応じた分析が行えることも機械学習モデルを用いるメリットの一つだと考えられる。また、S&P500_前日や TOPIX (配当込)_前日、都心 5 区貸室床面積、都心 5 区空室率等の重要度も相対的に高い結果となった。

図表4 全期間でみた各特徴量の相対重要度



(出所)三井住友トラスト基礎研究所が作成

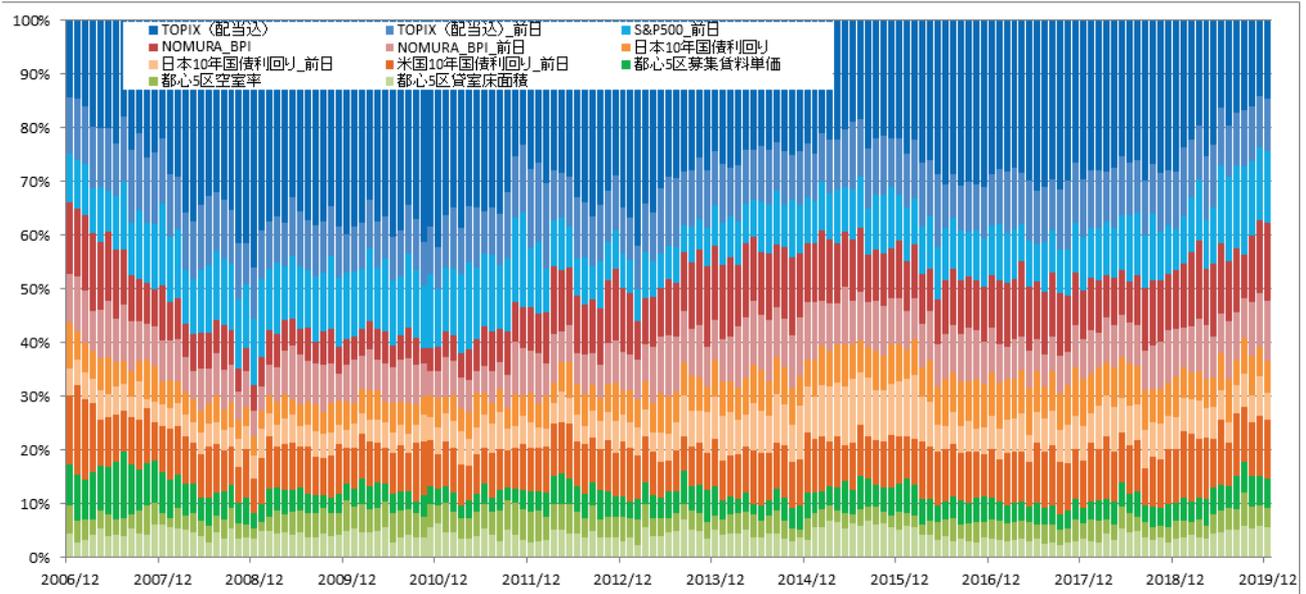
図表5 東証REIT指数とTOPIXの時系列推移



(出所)Bloombergデータをもとに三井住友トラスト基礎研究所作成

各特徴量の寄与度(相対重要度)の時系列推移を示したのが図表 6 である。目的変数である東証 REIT 指数が大きく変動した時期に寄与度も大きく変化していることがわかる。例えば、2007~2008 年の金融危機前後では、信用不安によるリスク回避行動が J-REIT 市場を含む資本市場に大きな影響を与えたことで株式市場要因が大きく上昇。2010 年末に包括的金融緩和策として、日銀による ETF 及び J-REIT の買入を国債などに加えて可能としたことで、債券市場要因が上昇した。2013 年には追加金融緩和期待から、株式市場と連動する形で投資口価格が上昇し、株式市場要因が一時的に上昇したが、その後は再び債券市場要因が上昇。2019 年の好調期も、低金利環境の継続と、安定的な分配金への注目により、債券代替としての J-REIT の魅力度が上昇したことで、株式市場要因が低下し、債券市場要因が上昇している。

図表6 要因別寄与度の時系列推移



(出所) 三井住友トラスト基礎研究所が作成

4. おわりに

検証の結果、今回採用した機械学習モデルは、本分析の精度向上に有効であった。特に過去データを用いた分析では、従来手法の線形回帰モデルの結果に対して明確な精度改善が認められるなど、過去データのモデル化における十分な可用性が感じられた。また、未知のデータに対する推定精度は、線形回帰モデルを安定して上回る結果とはならなかったものの、その精度差は僅かであり、実務上は特徴量選択や前処理、ハイパーパラメータの調整をより精緻に行うことで線形回帰モデルを安定して上回る精度を出すことも十分可能だと考えている。

更に、今回採用した機械学習モデルでは、結果に至った要因も各特徴量の寄与度として把握できるなど一定の説明力を維持することも可能であることを確認できた。従来手法の線形回帰モデルについても、解釈や説明が容易であり、少ないデータでも学習が可能である等、様々な長所があり、今後も実務で引き続き活用されていくと考えているが、機械学習手法についても、新たな手法が次々に開発されていることから、J-REIT 市場を分析する上でも、その活用は大いに期待できる。

今回の検証において最も精度の高かった、GBDT(特徴量 11 系列)を用いた各特徴量の寄与度の時系列推移を見てみると、J-REIT 市場のパフォーマンス要因については、長期的には市場環境に応じて相当程度変化していることが確認された。本稿の学習期間は、金融危機を含む 2003 年 4 月 1 日～2019 年 12 月 31 日としたが、金融危機等で信用不安によるリスク回避行動が資本市場に影響する局面では、J-REIT 市場は株式市場との連動性を高める特性が確認され、また 2019 年においては、債券代替としての J-REIT の魅力度が上昇したとの一般的な認識が、債券市場要因の寄与度上昇という形で定量的に確認された。このように、足元の J-REIT 市場のパフォーマンス要因を定量的に正確に把握する意義は大きく、本手法は当面の市場動向を見通す上でも有益だと考えている。

【参考文献】

- [1] 三菱 UFJ 信託銀行 受託運用部 2013 「J-REIT のリターン特性について」(三菱 UFJ 信託資産運用情報)
- [2] ニッセイ基礎研究所 原田 2019 「J-REIT のパフォーマンス要因分析 ～J-REIT 市場の上昇・下落要因を調べる～」(ニッセイ基礎研レポート)
- [3] 石井康夫 2018 「J-REIT の投資口価格に影響を与える要因に関する一考察」(大和大学研究紀要 4:13-28)
- [4] 新潟大学経済学部 伊藤教授 2013 「日本の REIT 市場に関する実証分析—株価と金利の影響及び国際間における連動性の検証」(信託研究奨励金論集第 34 号)
- [5] 門脇 阪田 保坂 平松 2019 「Kaggle で勝つデータ分析の技術」(技術評論社)

【お問い合わせ】 REIT 投資顧問部

<https://www.smtri.jp/contact/form-reit/index.php>

1. この書類を含め、当社が提供する資料類は、情報の提供を唯一の目的としたものであり、不動産及び金融商品を含む商品、サービス又は権利の販売その他の取引の申込み、勧誘、あっ旋、媒介等を目的としたものではありません。銘柄等の選択、投資判断の最終決定、又はこの書類のご利用に際しては、お客さまご自身でご判断くださいますようお願いいたします。
2. この書類を含め、当社が提供する資料類は、信頼できると考えられる情報に基づいて作成していますが、当社はその正確性及び完全性に関して責任を負うものではありません。また、本資料は作成時点又は調査時点において入手可能な情報等に基づいて作成されたものであり、ここに示したすべての内容は、作成日における判断を示したものです。また、今後の見通し、予測、推計等は将来を保証するものではありません。本資料の内容は、予告なく変更される場合があります。
3. この資料の権利は当社に帰属しております。当社の事前の了承なく、その目的や方法の如何を問わず、本資料の全部又は一部を複製・転載・改変等してご使用されないようお願いいたします。
4. 当社は不動産鑑定業者ではなく、不動産等について鑑定評価書を作成、交付することはありません。当社は不動産投資顧問業者又は金融商品取引業者として、投資対象商品の価値又は価値の分析に基づく投資判断に関する助言業務を行います。当社は助言業務を遂行する過程で、不動産等について資産価値を算出する場合があります。しかし、この資産価値の算出は、当社の助言業務遂行上の必要に応じて行うものであり、ひとつの金額表示は行わず、複数、幅、分布等により表示いたします。