

機械学習モデルの XAI と統計的手法を用いた モデル構築結果とデータ特性の解釈

丹	波	靖	博*
原	口	健太	郎**
新	原	俊	樹***
大	石	桂	一****

1. はじめに

近年のコンピュータ技術の目覚ましい進化、計算速度の向上、データの重要性に対する認識の深まり、そしてデータ取得技術の飛躍的な進歩は、社会に大きな変革をもたらしている。こうした技術革新は、人工知能（Artificial Intelligence: AI）モデルの利用拡大にも大きく貢献している。近年、AI モデルはコンピュータによる膨大な計算能力を活用することで、従来の手法では実現できなかった柔軟性と高精度を兼ね備えた説明力を提供することが可能となった。AI 技術を用いた計算手法は複数存在するが、近年特に注目を集めているのが、AI 技術の1つである機械学習モデルである。ここで、機械学習モデルとは、データから自動的にパターンや規則を学習することで、複雑な現象を表現する能力を持つモデル群を指す。機械学習モデルには多種多様な種類が存在し、それぞれ異なる特性や適用範囲を持っているため、目的や前提条件を慎重に考慮し、最適なモデルを選定することが重要となる。

また、多くの機械学習モデルは、その内部構造が複雑であり、モデル式を明示的に表現することが困難な場合も少なくない。特に、モデルの構造が非常に

* 西南学院大学経済学部教授

** 西南学院大学商学部准教授

*** 西南学院大学情報処理センター助教 **** 九州大学経済学研究院教授

複雑な場合には、目的変数と特徴量の関係を理解することは容易ではない。分析の目的によっては、モデルの予測精度が最優先される場合もあれば、モデルの説明可能性がより重要視される場合もある。例えば、金融リスク管理など、モデルの解釈が重要な意思決定に直接関わる場合には、モデルの振る舞いについて明確な説明が要求される。しかし、高度に複雑な機械学習モデルは、その仕組みがブラックボックス化してしまうことが頻繁に起こる。これは、モデルがどのように予測を行っているのか、どの特徴量が重要な役割を果たしているのかといった情報が容易に理解できない状態を意味する。

こうした背景から、予測精度と説明可能性を高いレベルで両立させるために、説明可能性を高めるための研究が近年大きな注目を集めている。この研究分野は、「説明可能な AI (eXplainable AI: XAI)」と呼ばれ、機械学習モデルの説明可能性を向上させるための様々な手法が開発されている。しかし、XAI の手法を用いた場合でも直感的な解釈と異なる結果が得られたり、一般的に考えられるモデルの挙動とは異なる結果が得られたりする場合もある。その理由はケースにより様々であるが、例えば機械学習の結果や XAI の結果は正しいものの、そもそものデータが異質なものであったり、特殊なデータであるケース、または使用している機械学習モデルのチューニングがうまくいっていなかったりするケースなどもあり、原因の特定が困難であることも少なくない。特にデータに特殊な構造がある場合には、統計的手法やデータ観察なども同時に併用しながらデータの特性を理解し、機械学習モデルの結果を解釈することも重要となると考える。

本研究では以下の内容で議論を進める。2 章では、機械学習モデルを用いた分析の利用背景と本研究の目的について述べる。3 章では、先行研究や主要な関連資料について紹介する。4 章では、機械学習モデルの基本的な知識と XAI の主要な手法について述べる。5 章では、日本の地方債データを使って機械学習モデルを構築し、構築された機械モデルについて XAI の手法を適用してモデルの解釈を行う。6 章では統計的手法を用いてデータ観察を行い、データの特性と XAI の結果との整合性について考察する。7 章では、最後のまとめを述べる。

2. 本研究の目的

データを利用して分析モデルを構築する際、導出されたモデルから目的変数とその他変数との関係性を理解したり、目的変数に何らかの影響を与えようとしたりするには、変数間の関係性や因果関係を把握することが重要となる。しかしながら、説明可能性の必要性は、分析の目的によって大きく変動する。

例えば、顔画像データを用いた感情認識モデルの場合、モデルの説明可能性は必ずしも重要ではなく、喜び、悲しみ、怒りなどの感情をどれだけ正確に識別できたかという正答率が重視される。一方、金融分野の分析においては、得られた結果の説明可能性が重要となるケースが多い。また、保険会社における保険料算出モデルにおいて計量化モデルを用いる場合、個人の年齢、健康状態、職業、生活習慣などの属性情報からリスクが評価され、将来的な保険金支払いの可能性が予測される。そのため、説明変数と目的変数の関係性の解釈がより重要となり、現在のところ機械学習モデルを利用する場合には、モデル利用による判断の精緻性や業務の効率性に重きが置かれる傾向がある。

将来的にはモデルの精度や効率化だけでなく、公平性、説明責任、透明性、信頼性などがより求められる可能性が高いと考えられる。つまり、AI モデルの判断によって、例えば顧客に不利益が生じることを回避し顧客サービス上の説明責任を果たすため、または現在の経験を将来の戦略に活かすため、なぜそのような結果になったのかというモデルの説明可能性を重視しなければならない状況が想定される。モデルが導き出した結果の理由を知ることは、その問題やデータに関する理解を深め、仮にモデルが誤った判断を下した場合においても、その誤りの原因を探ることに役立つ。このように説明可能性が求められつつある社会的背景から、実務においては予測精度が劣っていたとしても、説明可能性の高い伝統的な統計モデルが依然として利用されることが多いという現状がある。

このような背景から、機械学習モデルの説明可能性のデメリットを克服しつつ、より精緻な機械学習モデルを活用する重要性は今後増していくと考えられる。このような機械学習の説明可能性の課題の解決のため、本研究では機械学

習モデルの特徴と説明可能性についての観点をまとめ、急速に研究が進みつつある XAI の手法について主要な手法を紹介する。さらに、日本の地方債金利データを対象として機械学習モデルの構築を行い、その有用性を検証する。地方債金利を分析対象とするのは、先行研究において、地方債市場において機械学習による金利推定が有効に機能する可能性が示されており、XAI の試行に適しているからである（丹波・原口，2023）。さらに XAI を適用し、その結果について統計的手法やデータ観察を用いてより深い解釈を行う。本稿の目的は、これらの分析プロセスを通じて、解釈がブラックボックスとなりがちな機械学習モデルの性質や振る舞いをより深く解釈し、ファイナンス・会計学研究における機械学習の一層の適用可能性を探ることである。

3. 先行研究と関連資料

近年、機械学習モデルの説明可能性は、学術界においても注目を浴びている。機械学習におけるモデルの説明可能性に関する研究は近年急速に進展しており、多くの研究成果が発表されつつある。ここでは、その中でも特に重要な研究や関連資料をいくつか紹介する。

Miller (2017) は、研究者が独自の価値判断により分析モデルを使用する傾向にあることに問題提起している。そして、XAI の分野は、哲学、認知心理学、社会心理学などの既存の広範囲な研究分野の判断基準に基づいて構築されるべきで、関連する哲学、心理学分野を横断するとともに重要な発見を抽出し、XAI の研究にそれらの知見をどのように統合できるかについて議論を行っている。Molnar (2020) では、機械学習におけるモデルの説明可能性について議論を行っている。決定木や線形回帰などの単純なモデルに加えて、特徴量の重要度 (feature importance)、ALE (accumulated local effects)、LIME (Local interpretable model-agnostic explanations)、SHAP (SHapley Additive exPlanations) などの各種 XAI 手法を紹介している。また、それぞれの XAI 手法がどのように機能するのか、そして各手法の長所と短所などについて詳細な解説を行っている。Biecek and Burzykowski (2022) は、構築モデルの説明可能性について、

分類問題や回帰問題に適用可能な手法を紹介している。予測モデリングにおける本質的な課題は、データや計算能力、アルゴリズム、あるいは柔軟なモデルの欠如ではなく、モデルの探索、モデルの説明、モデルの検査のためのツールの不足であることを指摘している。また、予測モデルの改善方法とモデルを監視する手法について議論を展開している。

一方、日本においても近年 XAI について認知が浸透しつつあり、いくつかの文献が得られる。吉田・田嶋・今井（2020）は、テーブルデータを用いて、XGBOOST で学習したモデルに対して SHAP 値を算出し、SHAP 値が概ね正確に特徴量のモデル学習への貢献を評価できることを実証している。大坪（2021）は、AI の説明責任に関する重要性と現在の課題についての解説を行っている。具体的には、実務において誰にどのような説明が求められ、XAI 技術を用いて何が可能なのかについて議論している。技術面では、LIME や SHAP などの XAI の技術について紹介し、説明相手、性能要件、データ特性、AI アルゴリズムなどに応じた適切な手法の選択の必要性について説明を行っている。森下（2021）は、機械学習モデルの説明可能性を向上させるための様々な手法を取り上げている。機械学習の説明可能性と線形回帰モデルの説明可能性を比較しながら解説を行い、複雑なブラックボックスモデルの挙動を理解するための XAI の技術である、Permutation Feature Importance（PFI）、Partial Dependence（PD）、Individual Conditional Expectation（ICE）、SHapley Additive exPlanations（SHAP）などの手法を紹介し、特徴の重要性、特徴と予測値の関係、インスタンス固有の挙動、予測値の理由について解説を行っている。

4. 機械学習モデルの特徴と説明可能性

機械学習モデルはコンピュータの計算能力を駆使し、データから反復的に学習することで、複雑なパターンや規則性を抽出し、高精度な予測を実現する。特に、サンプル数が増加するにつれてその能力は向上するため、伝統的な統計分析手法と比較して、大量のデータが必要となる。本研究では、機械学習の中でも、入力データと出力データの対応関係から学習を行う教師あり学習に焦点

を絞る。教師あり学習の機械学習モデルには、回帰モデルと分類モデルが存在するが、本研究では分類モデルの1つであるランダムフォレストレグレッサーモデルを適用して分析を行う。ランダムフォレストレグレッサーモデルは、決定木分析を拡張した手法であり、目的変数に影響を与える特徴量に基づいて樹木状のモデルを構築する。複数の決定木を組み合わせることで、単体の決定木よりも高い予測精度を実現できる点が特徴である。

決定木は、分岐条件ルールを積み重ねてツリー構造を構築することで、入力データを分類する。このため、各インスタンスがどの条件を満たし、最終的な予測結果に至ったのかを追跡することができる。分析過程の追跡が可能のため、決定木は比較的説明可能性が高いモデルと言える。しかし、条件ルールが複雑に絡み合った決定木の場合には、解釈は困難になる場合もある。ランダムフォレストレグレッサーモデルのように複数の決定木を組み合わせるモデルでは、比較的複雑な構造となるため、説明可能性も決定木に比べると低下する。

機械学習モデルは、線形回帰などの伝統的な手法と比較して、非線形な関係性を表現できるため、より高い説明力を実現できる場合が多い。しかし、機械学習モデルは、高い予測精度を実現する一方で、その複雑な構造ゆえに解釈が困難なモデルの「ブラックボックス化」と呼ばれる問題が生じる。モデルの説明可能性を以下の4つの観点から分解することで、ブラックボックス化の課題をある程度解決し、より多角的な解釈が可能となる。①特徴量の重要度の定量化：各特徴量が予測にどの程度影響を与えているかを定量的に評価する。②特徴量と予測値の平均的な関係性の把握：各特徴量と予測値との平均的な関係性を理解する。③インスタンスごとの特徴量と予測値との関係性の把握：個々のデータにおける特徴量と予測値の関係性を詳細に分析する。④予測の理由の説明：モデルが特定の予測を行った理由を明確に説明する。それぞれの観点に対応する代表的な XAI 手法として、Permutation Feature Importance (PFI), Partial Dependence (PD), Individual Conditional Expectation (ICE), SHapley Additive exPlanations (SHAP) が挙げられる。

これらの手法は、特定の機械学習モデルに依存しないため、様々なモデルに適用できるという利点を持つ。本研究では、これら XAI 技術の中で、SHAP

を用いて分析を行う。SHAP は、協力ゲーム理論に基づき、各特徴量が予測にどれだけ貢献したかを評価することで、予測の根拠を解釈する手法である。特定のインスタンスに対して、各特徴量の貢献度を可視化することで、予測結果の理由を説明することができる。SHAP は XAI の中でも強力なツールであるが、計算負荷が高いこと、理論面の難易度が他の手法に比べると高いというデメリットも存在する。このため SHAP では、できるだけ計算量を軽減するため、シャープレイ値の計算負荷を小さくしながら近似的に計算できるよう工夫されている。SHAP は、Lundberg et al. (2020) が Nature Machine Intelligence に発表して以降、自然科学のみならず社会科学の分野においても XAI の分野において有効なツールの 1 つとして利用され始めている。

5. 機械学習モデルと XAI によるモデル分析

5-1. 分析データと特徴量

本研究においては分析データに JSPrice を利用している。JSPrice は、日本経済新聞社、日経金融工学研究所、野村証券、野村総合研究所の 4 社が共同開発した時価会計における債券時価評価のための価格データベースである。公募・非公募の国内債を中心に地方債約 6000 銘柄など、合計で約 1 万 2000 銘柄を収録している。地方債のほとんどの取引を取り扱う事実上のプライスメーカーである野村証券が、野村証券店頭を含めたマーケットにおける取引実績、マーケット参加者の投資スタンス、市場動向の変化を反映させ、ロジックに基づく「合理的に算定された価額」（セカンダリーマーケットの理論値）を開示している点が特徴である。本研究では、2015 年度の公募債市場セカンダリーマーケットの日次データを用いて、約 80 自治体の全償還残存年限の地方債銘柄を対象に分析を行う。用いたデータ数は 622,339 件である。モデルに利用する目的変数と説明変数の一覧は表 1、各変数の基本統計量は表 2 の通りである。

ここで、対東京都債スプレッドとは、各銘柄のイールドと、同日・同残存年限の東京都債のイールドとの差のことを指し、原口・丹波（2022）と同一の方法で算出した。sdm は各地方債銘柄データが含まれる月の日次スプレッドデー

表 1 変数一覧

変 数	定 義
TokyoSpreadBP	各銘柄の対東京都債スプレッド（目的変数）
YRSTOMAT	各銘柄の残存年限
jissitsu	各自治体の実質公債比率
syourai	各自治体の将来負担比率
COUP	各銘柄のクーポン金利
sdm	スプレッドの標準偏差

表 2 変数の基本統計量

	TokyoSpreadBP	YRSTOMAT	jissitsu	syourai	COUP	sdm
count	622,339	622,339	622,339	622,339	622,339	622,339
mean	4.83	6.87	12.87	189.94	1.10	2.47
std	10.20	5.97	4.83	77.30	0.69	4.43
min	-64.72	0.01	0.60	0.40	0.00	0.01
25%	0.26	2.50	11.00	152.50	0.47	0.18
50%	1.42	4.92	13.70	198.70	1.03	0.36
75%	4.01	9.23	15.50	232.70	1.69	1.32
max	257.96	29.99	21.30	341.10	2.90	48.71

タの各自治体ごとの標準偏差であり、価格変動の激しさを示し、流動性の代替変数として取り扱う。

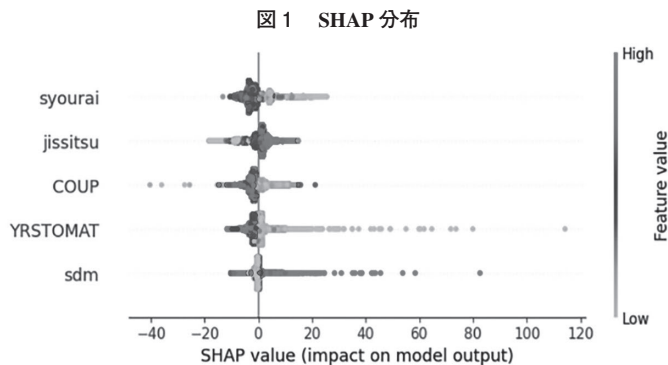
5-2. ランダムフォレストレグレッサーモデルの適用

本研究では、公募債市場におけるスプレッド推定モデルを用いた分析に関して、機械学習モデルの1つであるランダムフォレストレグレッサーモデル（rf）を用いた。学習データに対して、ランダムフォレストレグレッサーモデルを用いた結果、その説明力は約96.23%であった。また、テストデータを用いた検証においても、決定係数は約94.53%と高水準を維持している。学習データとテストデータのデータ数はそれぞれ497,871件、124,468件である。

5-3. SHAP の適用

本節では、XAI の手法の 1 つである SHAP を適用することで、ランダムフォレストレグレッサーモデルの特徴について分析を行う。SHAP を用いることで、ある特徴量の値の増減が目的変数に与える影響や特徴量の貢献度の分布を個別のデータについて色分けによる出力図で可視化することができる。図 1 は、横軸が目的変数への寄与度の大きさ、縦軸が特徴量の貢献度の順位（上が高貢献度の変数）を示し、赤*は特徴量が正の値、青が負の値を表している。例えば、jissitsu は概ね jissitsu の値が小さくなる（青）ほど目的変数への寄与度が小さく（左側）なり、jissitsu の値が大きくなる（赤）ほど目的変数への寄与度が大きく（右側）なる。つまり、jissitsu は目的変数と正の相関があることがわかる。

一方、図 1 では、COUP（クーポン金利）、YRSTOMAT（残存年限）の右側は青であり、COUP と YRSTOMAT は、それぞれ TokyoSpreadBP（スプレッド）とは逆相関の関係となっている。これら COUP、YRSTOMAT の変数は、目的変数である TokyoSpreadBP とは正の関係となるのが一般的なので、一般的な関係性とは異なった結果が得られている。本研究ではこの問題の原因について詳しく分析を行う。まず、原因の 1 つとして、YRSTOMAT と COUP の高相関が考えられる。説明変数が高相関である場合には、モデル構築において説明変数と目的変数の関係が逆転する場合がある。表 3 の変数の相関係数を見る



*カラー版の論文は「西南学院大学機関レポジトリ 041. 経済学論集 The economic review」を参照のこと。

表3 変数の相関係数

	TokyoSpreadBP	YRSTOMAT	jissitsu	syourai	COUP	sdm
TokyoSpreadBP	1.000	-0.277	-0.023	-0.195	-0.325	0.215
YRSTOMAT	-0.277	1.000	-0.086	-0.028	0.431	-0.369
jissitsu	-0.023	-0.086	1.000	0.819	-0.084	0.035
syourai	-0.195	-0.028	0.819	1.000	0.005	0.056
COUP	-0.325	0.431	-0.084	0.005	1.000	-0.112
sdm	0.215	-0.369	0.035	0.056	-0.112	1.000

と COUP, YRSTOMAT の相関は0.431であり、用いている変数の中では比較的正の高い値を示している。しかし、ここでは分析結果の詳細は割愛するが、YRSTOMAT, COUP どちらか一方の変数を除外し、同様にランダムフォレストレグレサラーモデルを構築しても、各変数の TokyoSpreadBP との逆相関の状況は変わらなかった。

6. 統計的手法とデータ観察による検証

5章で確認された COUP と YRSTOMAT の TokyoSpreadBP との逆相関の原因について、さらに詳しく調べるため、統計的手法を用いてデータ観察を行い、原因を調査することを試みる。

6-1. 各変数の分布図

初めに、ランダムフォレストレグレサラーモデルにおいて、逆相関の関係が見られた COUP と YRSTOMAT の両変数について、各変数における分布図で異常な点が見られないか調査する。TokyoSpreadBP, YRSTOMAT, COUP それぞれの変数の分布は図2, 図3, 図4の通りである。各変数の分布図を見ると TokyoSpreadBP は凸性が強い比較的左右対称の釣鐘状の分布形状であるが、右の裾が長い分布となっており、スプレッド値が高い銘柄が存在していることがわかる。YRSTOMAT の分布図を見ると、地方債の残存年限は10年以下の地方債データが多く、特に比較的短期の5年以内の銘柄が多いことがわかり、長い

図2 TokyoSpreadBP 分布図

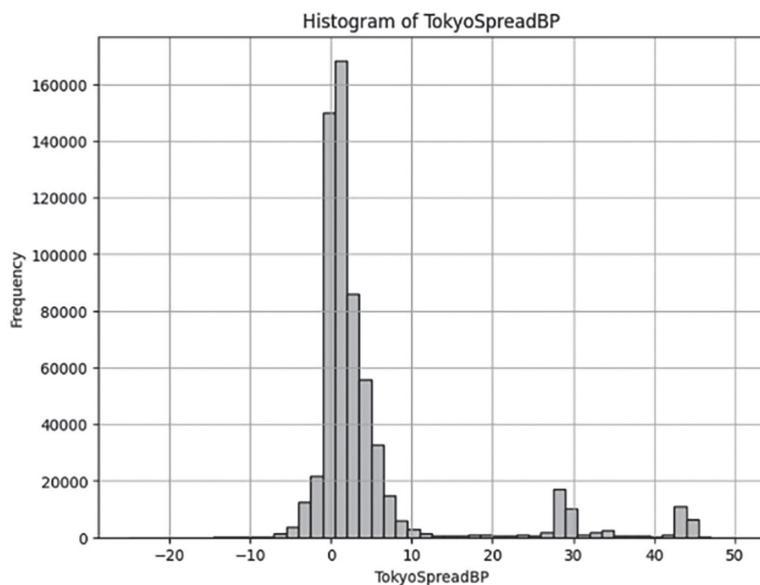


図3 YRSTOMAT 分布図

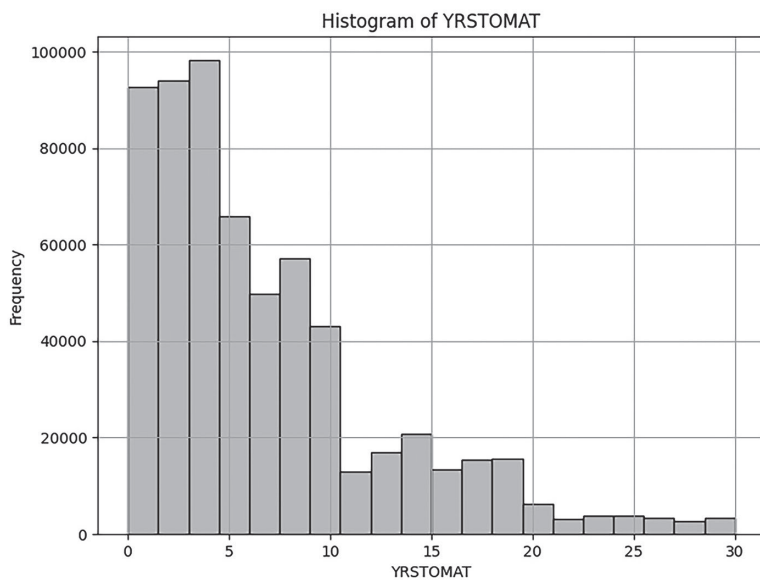
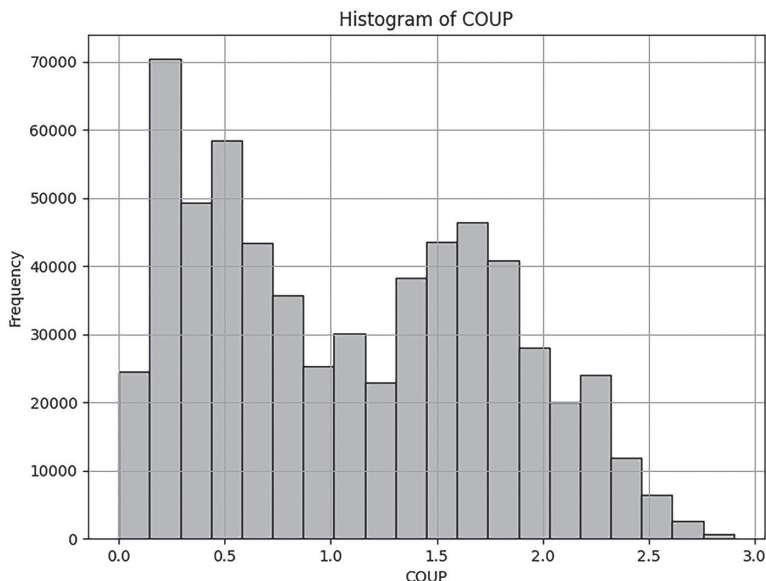


図4 COUP 分布図

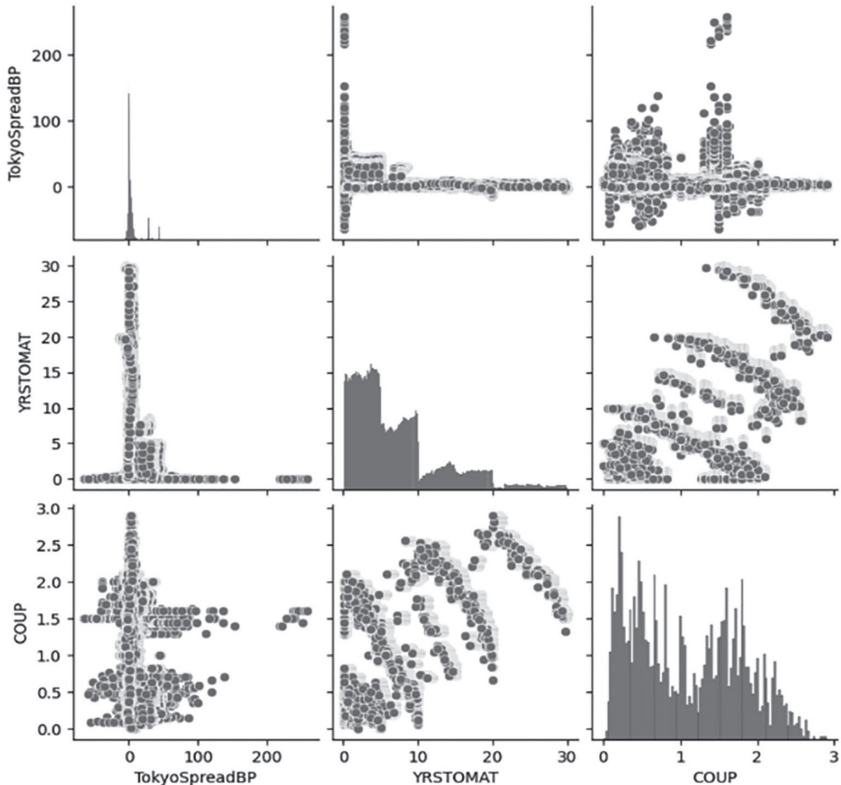


年限になるほどサンプルデータ数が少なくなっていることがわかる。COUP の分布図を見ると、0.2付近のクーポン金利水準が小さい銘柄と1.7付近のクーポン金利水準が中程度の銘柄が相対的に多くなっており、山が2つある分布となっていることがわかる。10年未満の短い年限の地方債銘柄の流通量が多く、クーポン金利が発行した時期により依存するため、分布が複数のピークを持つことは地方債市場の特徴として特段違和感はなく、単一変数の分布状況からは特に大きな異常は見られないと考える。

6-2. 2変数間の分布図

次に、TokyoSpreadBP, YRSTOMAT, COUP の3変数に関する調査を引き続き行うため、各変数の2つの組み合わせによる2次元散布図を分析する。図5のそれぞれの図はこれら3変数のうち、2変数を選択した場合の2次元散布図を表している。図5の下段中央に示されたYRSTOMATとCOUPの散布図に注目する。YRSTOMATとCOUPの相関係数は0.431と正であったが、散布図

図5 2次元散布図



によるとデータの分布は3つのグループに分かれており、各グループ内においては負の相関、つまり年限が長くなるほどクーポンが小さくなる傾向が見られる。これは一般的に考えられる状況とは異なる結果である。この要因として、地方債が発行された時期の金利水準によって関係性が逆転している可能性が考えられる。具体的には、データ分析期間中、日本においては金利低下期間であったため、このような逆転現象が発生している可能性がある。例えば、2007年頃に発行された10年国債のクーポン金利は比較的高く設定されていたが、この国債は2015年には残存期間は2年の債券となっている。一方、2015年に発行された10年国債のクーポン金利は低く残存期間は10年と長くなっている。つま

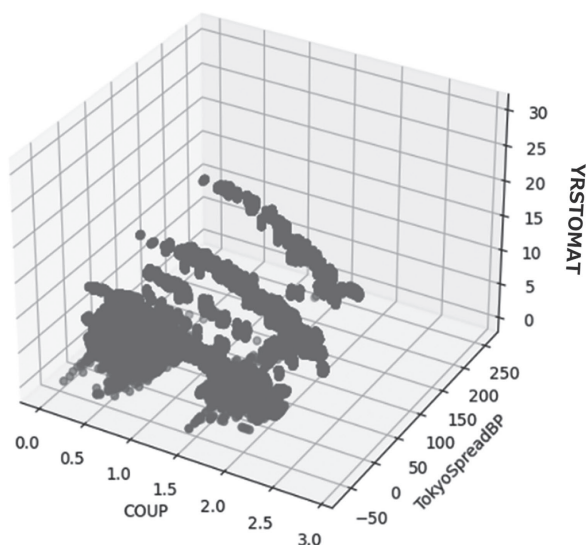
り、年限が長くなるほどクーポン金利が小さくなる傾向は、テクニカルな要因ではなく、日本市場の経済的環境を反映した構造的な問題である可能性が高いと言える。一方、図5の上段中央の YRSTOMAT と TokyoSpreadBP の散布図を確認すると、短い年限のデータで高スプレッドのデータが多いことが確認できる。一方、年限が長くなってもスプレッドの大きさは限定的であることがわかる。これは地方債価格からスプレッドに理論式で変換する際に、年限が短い場合には価格の少しの違いで理論スプレッドが大きく変わる不安定性があり、スプレッドが大きな値となってしまうことが原因ではないかと推測される。次に、図5の上段右側の COUP と TokyoSpreadBP の散布図を確認すると、クーポン金利にはその水準により2つの分布の山があることが確認できる。一方、クーポン金利が高くなってもスプレッドの大きさは限定される。発行された時期によりクーポン金利が変わるため、本データに含まれる地方債銘柄においては、市場金利の推移により、クーポン金利の水準が2グループに分かれていると推測される。COUP、YRSTOMAT どちらの変数に関しても、TokyoSpreadBP との相関係数は結果的にマイナスとなり逆相関の関係になっている。つまり、COUP や YRSTOMAT のスプレッドとの逆相関はテクニカルな要因ではなく、日本市場における構造的な問題による可能性が高い。

6-3. 3変数間の分布図

図6は TokyoSpreadBP、YRSTOMAT、COUP 3変数の3次元散布図を表している。YRSTOMAT と COUP の2次元の散布図で見られた3つのデータグループは、図6においても確認できる。図6のデータ分布を見ると、それぞれのグループではスプレッドの水準が異なっており、スプレッドが高いほど年限が長くなる関係にある。これはこれら2変数の一般的な関係性と一致する結果である。

これらの結果から、TokyoSpreadBP、YRSTOMAT、COUP それぞれの関係は単純な正の関係ではなく、日本市場の状況を反映した結果であり、このことは COUP、YRSTOMAT と TokyoSpreadBP が逆相関を示す SHAP の結果の原因となっていると推測できる。

図6 3次元散布図



7. ま と め

本研究では、機械学習の特徴を概観し、機械学習モデルの説明可能性の必要性を指摘したうえで、近年急速に研究が進みつつある「説明可能な AI (XAI)」の主要な手法の一部を紹介した。さらに、日本の地方債データを使って機械学習モデルの構築を行い、XAI の手法の 1 つである SHAP を用いてモデルの解釈を行うとともに、説明可能性をより向上させるため、統計的手法やデータ観察と XAI との組み合わせに基づく分析を試行した。これらの手法は、これまで一般に解釈が困難とされてきた機械学習モデルの性質や振る舞いの解明に貢献するもので、ファイナンス研究や会計学研究における機械学習活用の基盤となる知見を提供するものである。

参考文献

- 大坪直樹他 (2021) 『XAI (説明可能な AI) そのとき人工知能はどう考えたのか?』 リックテレコム。
- 丹波靖博, 原口健太郎 (2023) 「機械学習を用いたわが国における地方債のスプレッド推定モデルの構築」『ジャフィー・ジャーナル』 第21号, pp.29-42。
- 原口健太郎, 丹波靖博 (2022) 「わが国の統一的な基準に基づく公会計財務諸表が有する増分情報と地方債スプレッドとの関連性分析」『会計プロGRESS』 第23号, pp.85-102。
- 森下光之介 (2021) 『機械学習を解釈する技術～予測力と説明力を両立する実践テクニック』 技術評論社。
- 吉田秀穂, 田嶋優樹, 今井優作 (2020) 「決定木モデルの解釈における SHAP 値の有用性の検証」『人工知能学会第34回全国大会論文集』 pp.1-3。
- Biecek, P. and T. Burzykowski (2022) *Explanatory Model Analysis: Explore, Explain, and Examine Predictive Models*, Chapman and Hall/CRC.
- Lundberg, S. M. G. Erion, H. Chen, A. DeGrave, J. M Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S. Lee (2020) From Local Explanations to Global Understanding with explainable AI for Trees, *Nature Machine Learning*, 2, pp.56-67.
- Miller, T. (2017) Explanation in artificial intelligence: Insights from the social sciences. *arXiv Preprint*, arXiv: 1706.07269.
- Molnar, C. (2020) *Interpretable Machine Learning*, Lulu.com.

謝 辞

本研究は JPS 科研費 JP23K01463, JP24K16490の助成を受けている。