

わが国における地方債スプレッド 推定モデル構築に対する機械学習の適用可能性

丹 波 靖 博*
原 口 健 太 郎**

1. はじめに

わが国の地方債市場は、株式市場・債券市場などの主要な金融市場の中でも規模的・機能的な重要性を有し、現代社会の発展に欠かせないものとなっている。具体的に見ると、2020年度の公募地方債市場（流通市場）の売買金額は約12兆円に上り世界的にも有数の規模であるとともに、近年、金融機関が国債よりも高い収益性を確保可能な地方債への投資を拡大していることから、投資対象としての重要性も高まっている。さらに、地方債は地方公共団体の財源確保の主要な手段であり、地方財政運営上の役割も極めて大きい。

地方債市場の健全な発展のためには、市場の透明性や効率性を高めることが不可欠である。そのためには、十分な情報のもと投資家がリスク評価を行ったうえで、各自のリスク選好に従い投資できる環境構築が必要となる。

信用リスク評価を適切に行うにあたってはスプレッドの推定が重要な課題である。しかしながら、日本の地方債市場を対象としたスプレッド推定の議論はこれまでほとんど行われてこなかった。その理由として、第一に、日本においては地方債が事実上国債と同リスクであるとする「暗黙の政府保証論」が存在し、地方公共団体の信用リスク差の存在自体が疑問視されてきたことから、地

* 経済学部教授 y-tanba@seinan-gu.ac.jp

** 商学部准教授 haraguch@seinan-gu.ac.jp

方債市場自体がファイナンス領域においてほとんど分析対象とされてこなかったこと、第二に、現在の市場環境ではスプレッドが非常に小さく、OLS 分析等の伝統的な手法を用いた場合、精度確保の点でモデル化が困難なことが挙げられる。これらのいわば「地方債市場の特殊性」により、日本の地方債市場は社債市場や米国地方債市場と比して異質なものとなっており、信用リスク分析・スプレッド推定に関する研究蓄積は著しく少ない状況にある。

地方分権が推進し、地方公共団体固有のリスクが世界的にも注目される状況下で、上記研究が進展していないことは学術上のみならず実務上も重要な問題である。とりわけ、高精度のスプレッド推定モデル構築のための研究蓄積の欠乏は、将来の地方債市場の発展やそれを支える研究の進展にあたって大きな弊害となる。

そこで、本稿では、モデル構築にあたって必要となる前段の知識の整理を行うことで、上記問題解決の糸口を探ることを目的とする。具体的には、柔軟で説明力の高いモデル構築が可能とされる機械学習の試行を行い、地方債のスプレッド推定への適用可能性を検討するとともに、可能であるとすればその精度はどの程度であると想定されるのか、議論を行う。さらに、分析に必要となる技術的な知見や課題についても概観する。これらの知見や課題整理は、今後、高精度かつ安定的な信用リスク分析モデルの構築を通じて、ひいては「暗黙の政府保障論」という特殊性が存在するわが国の地方債市場における市場参加者の投資行動の明確化に寄与するとともに、将来的な市場の透明性や効率性の向上の礎となるものである。

2. 先行研究レビューと本研究の目的

2-1. 先行研究レビュー

地方債市場に関連する学術分野は会計学、財政学、ファイナンスの3つの領域からの研究が行われている。第1に、会計学においては複式簿記・発生主義に基づく公会計の財務諸表等の会計情報と地方債の格付けや地方債金利情報との関連性に係る研究などが進展している。第2に、財政学分野では財政情報等

が地方債市場に対して与える影響などが研究されている。第3に、ファイナンス分野では債券市場の信用リスク評価に係る研究がされている。本論文では、地方債のスプレッドを対象にスプレッド推定モデル構築による分析を進めるが、本稿の主題に関連の深いものを対象にレビューを行う。

地方債の対国債スプレッドの研究は主に米国で進展している。米国における地方公共団体の財務健全性とリスクプレミアムの関係性については、Goldstein, and Woglom (1995), 財務状況以外にも流動性リスクについて考察を行った研究として、Wang, Wu, and Zhang (2008), Schwert (2017) などの研究が見られる。債券投資を行う場合、メインプレイヤーである金融機関は、通常格付けを基本として投資対象を選別していることが多い。地方債の格付けやスプレッドを対象として、財務指標などにより信用リスクを分析したCapeci (1991), 説明変数に格付けを用い、地方債のスプレッドとの関係の分析を行ったLiu and Thakor (1984) などの研究が挙げられる。その他の関連研究として、地方債の流動性と価格ボラティリティに正の関係を示したDowning and Zhang (2004), 地方債のトレーディングコストの推定により価格の透明性の欠如を指摘したHarris and Piwowar (2006), 財政状況が悪いにもかかわらず、発行されている地方債スプレッドは同格の地方公共団体と同水準である地方公共団体も存在するという矛盾を指摘したNovy-Marx and Rauh (2011a, 2011b) の研究も見られる。

わが国の研究では、地方債の対国債スプレッドの研究は主に財政学の分野で行われてきた。例えば、田中 (2012) や中里 (2008), Hattori and Miyake (2015) は、日本の総務省が「地方公共団体の財政の健全化に関する法律」(健全化法) に基づき公表する実質公債費比率等の財政指標(既存指標)と地方債スプレッドとの関連性関係について分析を行っている。既存指標に加え、三宅 (2017) では売買参考統計値最高値と最低値の差を流動性指標として説明変数に取り入れ、地方債のスプレッド推定に関する分析を行っている。さらに、既存指標・流動性指標のみならずマクロ指標である有効求人倍率も説明変数としてモデル構築を行った石川 (2007) の研究も見られる。一方、わが国においては、国が暗黙のうちに地方公共団体の債務を保証しており、地方債と国債が

事実上同リスクと見なされる「暗黙の政府保証論」の存在が言及されている（井潟・三宅，2007）。石田・中里（2019）は、わが国における地方公共団体の複数の既存指標が対国債スプレッドに対して有意な影響を与えることや、既存指標の悪化は他の地方公共団体のスプレッドに負の影響を与えるというスピルオーバーリスクも明らかにした。

しかし、我が国においては地方債の信用リスクの先行研究の事例は限定的であるのが実情であり、特にファンナンスの観点から信用リスク分析に焦点をあてた研究はほとんど見られない。さらに上記の分析は、すべてOLS等のいわゆる伝統的な統計的手法に基づくものである。上述のような先行研究に対し、丹波・原口（2021）では日本の地方債の特殊性や課題を対象とし、ファンナンスの観点からスプレッド推定モデル構築の課題に取り組み、機械学習により高い精度のモデルが構築を行った。さらに、実績スプレッドが相対的に高い地方公共団体について、信用リスク要因の比較を行い信用リスク要因について考察を行った。

2-2. 本研究の目的

上述したとおり、日本の地方債市場の特殊性として、暗黙の政府保証論の存在や信用リスクモデル構築の困難性などが挙げられる。スプレッド推定モデルに関して、スプレッドが低水準である現在の地方債市場の分析を進めるためには高精度のスプレッド推定モデルの構築が不可欠となるが、従来の重回帰モデルでは精度上の限界が生じるため、より高精度の手法を検討する必要がある。近年、コンピュータの計算速度の向上により、コンピュータで膨大な計算を行うことで柔軟かつ高精度の説明力を確保できるAI（人工知能）モデルが進化している。機械学習モデルという呼称は、そのうちのいくつかの計算手法を指す。機械学習モデルには多くの種類が存在するため、適用にあたっては様々な手法や前提条件について検討を行い、最適なものを選定する作業が重要となる。

本稿の目的は、地方債市場における信用リスクを計量化するスプレッド推定モデルの構築のための機械学習モデルの概要や具体的適用方法に関し、技術的な知見や課題の整理を行うことで、機械学習手法の地方債研究への適用可能性

を検討することである。具体的には、丹波・原口（2021）で提示した機械学習モデルの適用方法に焦点を当て、機械学習の計算ロジックや実際に課題となった点について解説を行う。さらに、機械学習モデルの分析方法の詳細な把握を通じて、機械学習モデルの特性や課題についても明確化する。

ここで、本稿の議論が実務性の重要性も有することに関して特に強調したい。安定性が確保された説明力の高い信用リスクモデルを構築することができれば、実務においても地方公共団体の信用リスクの計測が可能となる。つまり、構築された信用リスクモデルにおいて、各地方公共団体の発行した地方債の信用リスクの計量が可能となるため、暗黙の政府保証論に対するより詳細な検証が可能となるのである。暗黙の政府保証論とは、先述のとおり、わが国の地方債は、デフォルトした場合の損失補填は法律上明文化されていないものの、制度背景や総務省の見解に基づき元本が実質的に保証されていることから、そのリスクは国債と事実上同等であると見なす考え方である。ここで、日本政府は地方債のデフォルトリスクを公式に否定している以上、もしも地方公共団体間の信用リスクの差を意識している投資家が存在した場合には、投資家の認識と政府見解との認識との間に乖離が生じていることに注意が必要である。当該乖離は、暗黙の政府保証論を信じる投資家と信じない投資家との併存、つまり情報の非対称性に基づく市場参加者の意思決定の差異をもたらし、ひいては市場の効率性・透明性を損なうおそれがある。このため、信用リスクモデルの構築により地方債のリスクを包括的に明らかにし、投資家が、暗黙の政府保証論が存在する中で、地方債のリスクにはクレジットリスクがそもそも含まれるのか、含まれるとしたらどの程度含まれているのかを正確に認識することは、市場参加者の投資行動の適正化という観点から重要な意義を有するのである。

3. 機械学習の適用

3-1. 機械学習の概要

機械学習とは、分析を行う際にモデル式を用いることなく、コンピュータの計算機能を利用してデータから反復的に学習し、潜むパターンや規則性を見つ

け出す手法である。それ故、一般的にはサンプル数が増加するにつれて規則性を見つけやすくなるため、適応的に説明力を改善することが可能であり、伝統的な統計分析手法に比べ、多くのデータを必要とすることが多い。機械学習には、入力データと分析の目的により、教師あり学習、教師なし学習、教科学習の3つの主要な種類に分類される。また、特徴を表すデータを特徴量、または説明変数と呼び、答えであるデータを目的変数、またはラベルと呼ぶ。

教師あり学習は、分析対象である問題の答えである目的変数（教師データ）と特徴を表すデータ（特徴量）をモデルに与え、教師データを対象として特徴量による推定を行い、機械学習のモデルの推定精度を向上するために学習を行う。本稿では、教師あり学習を用いて、地方債のスプレッドを目的変数とした推定モデルを構築する。つまり、地方債のスプレッド水準という答え（目的変数）が与えられており、地方債の特徴を表す別の変数（特徴量）でこのスプレッドの水準を推測するという教師あり学習によりモデル構築を行う。地方債スプレッドのように連続した数値を推定する他にも、属性などのカテゴリーを予測する問題（分類問題）も存在する。分類問題の例としては、デフォルトするかしないかの2値分類や、格付けのような2個以上の属性に分類する多値分類が見られる。

教師あり学習は特徴量と目的変数である正解をセットとしてモデルを学習させる方法である一方、教師なし学習は答えである目的変数である正解を与えないモデルである。教師なし学習では、教師あり学習と同様、特徴を表すデータを使用するが、目的変数を推定するのではなく、データの中に特徴の近い部分集合を見つけたりすることで入力データの構造を理解することなどを目的とする。教師あり学習と比べると、正解となる答えがないため、得られた結果の解釈が困難な場合や、分析者の主観的な解釈が必要かつ重要となる場合もある。教師なし学習の例として、本稿で関連のあるテーマとしては、投資対象となる地方債を選別するために、地方債を同じ属性の地方債にグループ分けするという例が考えられる。このように、複数の似ているデータの集まりに分類する手法はクラスタリングと呼ばれる。3つ以上の変数からなるデータ構造を人間が理解するのは難しい場合が多いが、クラスタリングにより似た属性により分類

することでデータの構造を解釈し、様々な応用が可能となる可能性がある。

3つ目の手法である強化学習とは、設定された環境の中で行動するエージェントが得られる報酬を定義し、エージェントが報酬を最大化するように行動し学習していく手法である。強化学習では定義された報酬を最大化する学習をコンピュータの計算を繰り返し行うことにより、より適切な行動を学習させていく。

3-2. 機械学習の実行環境

本稿では、地方債のスプレッド推定モデルを構築するにあたり、初期環境設定やプログラミング作業の効率化を重視して環境構築を行う。具体的な環境は以下のとおりである。プログラミング言語はPYTHONを使用し、開発環境はGoogle Colaboratoryを用いた。Google Colaboratoryは、開発環境の設定が容易であり、複数の人が同一の作業環境で試行錯誤を繰り返すことができることや、アウトプットを容易に可視化しながら作業を進められること、プログラム内容を容易に共有できる利点などがある。Google Colaboratoryには、計算機能を実行する環境を2つ選択することが可能であり、1つ目の「ホストされているランタイムに接続」はGoogleのサーバーを利用する方法で、無料版と有料版の利用が可能である。また、無料版と有料版とも計算途中に接続が切れるトラブルが多く発生するというデメリットには注意が必要である。2つ目の「ローカルランタイムに接続」は、自己所有のコンピュータを利用して計算を行う方法である。そのため、使用するコンピュータのスペックにより計算速度が決まる。この方法のメリットは、接続が切れるトラブルの頻度が、上述のホスト接続に比べて顕著に低いことであり、本稿ではモデル構築にあたってローカルランタイム接続を採用する。

分析に使用した主要なライブラリとパッケージとして、PYCARET, SCIKIT-LEARN, SHAPの3つが挙げられる。PYCARETはPYTHONのオープンソースの機械学習ライブラリであり、仮説から考察までのサイクルタイムを短縮することを目的としたものである。SCIKIT-LEARNは、多くの機械学習アルゴリズムが実装されたライブラリで、どのアルゴリズムでも同じような書き方で

利用することができるというメリットがある。また、教師あり学習や教師なし学習のアルゴリズムが実装されており、評価用のツールや便利関数、およびサンプルのステータスセットなどが入ったツールキットとなっている。SCIKIT-LEARNはGPUをサポートしておらず、GPUへの対応は少なくとも近い未来には行われない予定であるため、計算の高速化を目的としてコンピュータ環境の整備の際には留意する必要がある。SHAP (SHapley Additive exPlanations)は、予測値に対して各特徴変数とその予想にどのような影響を与えたかを、色付きの図により可視化できる手法である。ある特徴量の値の増減が目的変数に与える影響を、出力図による色分けと数値評価により可視化することが可能となっている。機械学習モデルは、高精度で説明力の高いモデル構築が可能である一方、因果関係などのモデルの解釈が困難であることが大きな欠点である。現時点において、SHAPはモデル解釈の一部を補完できるツールとして有効であると考えられる。機械学習の分析において、因果関係などのモデルの解釈を行うことは、分析の内容をブラックボックス化させないという観点で、実務においても非常に重要な点である。SHAPは、協力ゲーム理論のシャープレイ値 (Shapley Value) を機械学習に応用したもので、シャープレイ値は、協力ゲーム理論において複数プレイヤーの協力によって得られた利得を、各プレイヤーに公正に分配するための評価手法の一つである。しかし、ある特徴量の値の増減が目的変数に与える影響を算出するためには、データの組み合わせを考える必要があるため、計算量が膨大になるというデメリットが存在する。SHAPでは、計算量の軽減のため、算出方法の工夫によりシャープレイ値を機械学習で近似的に計算できるようにしている。SHAPの計算にはいくつかの方法があるが、特徴量の平均的な予測値を周辺化による期待値とし、別の特徴量が追加されたときの予測値の変動量を、その特徴量の限界貢献度として評価していく方法などがある。本稿においては、計算リソースの不足によりSHAP値の計算に時間がかかるため、変数の数を減らして試行を繰り返す必要があった。一般的な性能の家庭用デスクトップコンピュータを用いて計算を実行した結果、約5.7万データの計算に約4～5時間、約14万データの計算に約1日、約34万データの計算に約3日の時間を要した。各種の技術文献によると、SHAP値の算出は

GPUの使用で高速化できるとの記述ある一方で、高速化できなかったとの別の記述もあり、今後の研究において計算の高速化について更に調査する必要があると考えている。

4. スプレッド推定モデルの構築とモデル内容

4-1. スプレッド推定モデルの構築

日本においては地方債のデフォルト実績がなく、デフォルトサンプルを利用したデフォルトモデルを構築できないため、本稿では信用リスクの定量化としてスプレッド推定モデルによる分析を行う。具体的には、上記の計算環境を用いて、複数の機械学習モデルを使ってスプレッド推定モデルを構築し、最も説明力の高い機械学習モデルを選定する。

各機械学習モデルの精度を検討するため、本稿では、試験的に2019年4月15日から2019年5月31日までの約5万7千銘柄の地方債金利データを用いて分析を行う。金利データはロンドン証券取引所グループイールドブック社が提供するデータベース・ソフトウェア「イールドブック」から取得し、各銘柄の利回り（各地方公共団体の地方債利回り）と東京都債の利回りの差（対東京都債スプレッド）を被説明変数に採用する。説明変数には、各銘柄の償還残存年限（YRSTOMAT）、クーポン（COUP）、健全化法に基づき総務省が公表する実質公債費率（jissuitsu）及び将来負担比率（syourai）、流動性リスクの代替変数として、同一団体のスプレッドの分析期間中の標準偏差（sd）、地方債発行高（ISSAMT）を採用する。

各機械学習モデルの説明力を比較した結果が表1である。表1では、決定係数（R2）によりモデルの説明力を比較しており、R2で評価した場合、説明力の高いモデルから順番に機械学習モデルのリストが表示されている。本結果によると、rf（Random Forest Regressor）のR2が最も高く0.9758であることが確認される。リストの1行が1つの機械学習モデルを表している。本稿においては、各機械学習モデルの詳細についての説明を割愛するが、これらの機械学習モデルリストの中には、決定木タイプのモデルや回帰タイプのモデルなど、さ

表 1 各機械学習モデルの説明力比較

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE	TT (Sec)
rf	Random Forest Regressor	0.0009	0.0000	0.0013	0.9758	0.0012	0.8496	3.582
et	Extra Trees Regressor	0.0010	0.0000	0.0013	0.9749	0.0013	0.8539	2.121
knn	K Neighbors Regressor	0.0011	0.0000	0.0014	0.9704	0.0014	0.9134	0.063
dt	Decision Tree Regressor	0.0011	0.0000	0.0016	0.9658	0.0015	0.9894	0.060
lightgbm	Light Gradient Boosting Machine	0.0014	0.0000	0.0018	0.9528	0.0017	1.0609	0.161
gbr	Gradient Boosting Regressor	0.0018	0.0000	0.0026	0.9082	0.0024	1.3178	0.918
ada	AdaBoost Regressor	0.0038	0.0000	0.0047	0.6916	0.0043	4.5366	0.481
lr	Linear Regression	0.0045	0.0000	0.0068	0.3565	0.0060	4.1345	0.281
lar	Least Angle Regression	0.0045	0.0000	0.0068	0.3565	0.0060	4.1345	0.014
br	Bayesian Ridge	0.0045	0.0000	0.0068	0.3564	0.0060	4.1285	0.014
ridge	Ridge Regression	0.0045	0.0000	0.0068	0.3553	0.0060	4.0738	0.013
huber	Huber Regressor	0.0043	0.0000	0.0069	0.3238	0.0061	3.4207	0.141
omp	Orthogonal Matching Pursuit	0.0051	0.0001	0.0076	0.1795	0.0068	5.1931	0.015
lasso	Lasso Regression	0.0063	0.0001	0.0084	-0.0003	0.0079	6.8225	0.014
en	Elastic Net	0.0063	0.0001	0.0084	-0.0003	0.0079	6.8225	0.013
llar	Lasso Least Angle Regression	0.0063	0.0001	0.0084	-0.0003	0.0079	6.8225	0.014
par	Passive Aggressive Regressor	0.0088	0.0001	0.0119	-0.9864	0.0118	1.0000	0.017

さまざまなタイプの機械学習モデルが含まれている。

モデルの安定性を確認するため、本データを用いてサンプルをランダムに10グループに復元抽出し、rf (Random Forest Regressor) を適用した結果 (表2) では、決定係数の平均値 (Mean) は表1の結果と同じ (0.9758) 水準であることが確認される。

4-2. ランダムフォレストモデルの概要

次に、上記の分析で最もR2が高かったrf (Random Forest Regressor) について説明を行う。Random Forestは決定木分析の手法の1つである。決定木分析とは、被説明変数に影響する説明変数 (特徴量) により樹木状のモデルを作成する分析方法で、予測、判別、分類などを目的として使われる機械学習の手法のひとつである。Random Forestは、回帰問題と分類問題のどちらにも用いることができる。分類問題を解く例としては、図1のように、決定木により学習

表2 rfモデルの復元抽出における説明力

	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	0.0009	0.0	0.0013	0.9787	0.0012	0.6817
1	0.0009	0.0	0.0013	0.9779	0.0012	1.0499
2	0.0010	0.0	0.0014	0.9745	0.0013	0.8018
3	0.0009	0.0	0.0013	0.9768	0.0012	0.8419
4	0.0010	0.0	0.0014	0.9740	0.0013	0.7403
5	0.0010	0.0	0.0013	0.9754	0.0012	0.9007
6	0.0009	0.0	0.0013	0.9766	0.0013	0.6862
7	0.0009	0.0	0.0013	0.9764	0.0012	0.8958
8	0.0010	0.0	0.0014	0.9720	0.0013	1.0237
9	0.0009	0.0	0.0013	0.9755	0.0012	0.8744
Mean	0.0009	0.0	0.0013	0.9758	0.0012	0.8496
SD	0.0000	0.0	0.0000	0.0019	0.0000	0.1204

図1 rfモデルにおける単体決定木の例



データを条件分岐によってグループを分割していくことを考えるとイメージしやすい。ある1つのサンプルを考えた場合、決定木における所属グループが決まるが、Random Forest Regressor の場合、条件分岐によって分割されたグループの複数木における平均値を取ることで推測値を算出する。Random Forest は決定木モデルを複数利用することで、決定木単体よりも予測精度の向上を図る事ができる。このような手法をアンサンブル学習とよぶ。アンサンブル学習は、個々に別々の学習器として学習させたものを多数決で融合させる事により、予測能力を向上させる学習方法である。決定木単体の性能は必ずしも高くないが、複数の決定木を用いることで、精度の高いモデルを作成できる。このように多数決で結果を決定する方法は、1人に相談した結果だけから物事を判断するのではなく、さまざまな人に聞いて総合的に物事を判断し、より良い結果を得る方法に似ている。機械学習においても同じように複数のモデルを作り多数決をとることで、より適切な結果を得ることが期待できる。

決定木の作成においては、学習データを分割する際に損失を定義して、どのようにデータを分割すれば良いかを学習していく。決定木では、データの乱雑さを表す不純度が小さくなるようにデータを分割していく。分類されたグループに同一のラベル（区分データ）が多く存在する時に不純度は小さくなり、分類されたグループに異なるラベルが多く存在すると不純度が大きくなる。不純度を表す具体的な指標としては様々なものが存在するが、ジニ係数が代表的なものである。具体的な手順としては、全ての特徴量と分割候補について不純度を計算し、分割により不純度が最も減る分割を選択し、さらに次のグループに分割していくことを繰り返す。Random Forest では、予測結果に対する重要度を特徴量ごとに計算し、不要な特徴量を排除していくことができる。重要度の高い特徴量は不純度を大きく減らすことが期待でき、重要度の小さな特徴量は不純度を減らすことができない特徴量を表す。

しかし、同一のデータを使って同一の学習方法で決定木を作成すると、出力結果も同じになってしまうため、決定木の1つ1つが多様性を持つことが重要となる。Random Forest では、それぞれの決定木に与えるデータを、ブートストラップ法と特徴量のランダム抽出により、分割結果ができるだけ多様性を持

つように学習させる。ブートストラップ法は、学習データから内容の異なる複数の学習データを生成する手法である。ブートストラップ法では、1つの学習データから何度もランダムに復元抽出を行うことによりデータを増やしつつ、データに多様性をもたらすことが可能となる。一方、特徴量のランダム抽出では、特徴量のうち1部の特徴量だけをランダムに用いて学習させる。Random Forest では、ブートストラップ法によりサンプルをランダムに復元抽出すると同時に、決定木のノード分割時に特徴量を都度選択して学習を行う。

4-3. ランダムフォレストによる各変数の説明力

機械学習で分析する際の大きな課題の1つに、モデルの構造を把握しにくいという点がある。つまり、なぜそのような結果や推定になっているのかという理解が難しいという課題である。金融機関などの実務においては、中身が見えないモデルや根拠が分からないモデルから得られた結果の活用は困難であろう。例えば、本稿での地方債のスプレッド推定の例を取って考えよう。金融機関はスプレッド推定モデルにより月次で保有する地方債銘柄の推定スプレッドを算出し、信用リスクの管理を行っている場合がある。例えば、ある地方債銘柄の推定スプレッドが急上昇した場合、金融機関はなぜその銘柄のスプレッドが上昇したのかを把握し、今後のポジションについてのスタンスを検討するであろう。しかし、モデルの内容が把握できない場合は、スプレッド上昇と要因の因果関係が明確ではないため、今後のポジションについての意思決定を行うことが困難となる。つまり、機械学習モデルで評価を行う場合、いかにモデルの内容を把握し、因果関係を解釈できるかという点が非常に重要となる。

この点、今後の研究展開や実務への適用にあたって、SHAP 値の活用が有用である可能性がある。SHAP 値とは、機械学習モデルを用いて被説明変数を推定するにあたり、各変数がその推定にどのような影響を与えたかを算出するもので、モデルの内容の把握可能性を大きく向上させる可能性があり、近時、その有用性が議論されている（吉田ほか、2020；丹波・原口、2021）。当該技術は地方債のスプレッド推定にあたっても活用可能性があることから、その検討は将来の課題としたい。

5. まとめと今後の展望

本稿では、地方債市場における信用リスクを計量化するスプレッド推定モデルの構築のため、機械学習モデルを用いる際のモデルの概要や具体的適用方法をまとめた。具体的には、地方債スプレッド推定モデルを例に取り、機械学習モデルによりモデル構築する際の構築方法の詳細と各分析内容の解説を行った。様々な機械学習手法の推定モデルを比較した結果、最も精度が高かったモデルはランダムフォレストモデルで、その決定係数（ R^2 ）は97%超と非常に高い値を示した。一方、機械学習では、なぜそのような結果となっているかなどのモデルの詳細な理解が難しいという課題がある。特に金融機関における実務では、モデル解釈の困難性のため、機械学習モデルの利用が制限される場面も多いと考えられる。そこで本稿では、SHAP 値の将来の活用可能性についても言及した。

機械学習モデルのその他の課題としては、モデルの安定性の検証が挙げられると考える。機械学習モデルは説明力が高いゆえ、過度にモデルがデータに柔軟にフィットしてしまう可能性が考えられる。つまり、ある期間に関して高い精度での推定が実現できた場合でも、対象期間を変更した場合にはモデルの構造や因果関係が不安定となる可能性があることから、モデルの安定性の検討が重要である。また、将来的には、コロナウィルスの地方債市場への影響も考慮する必要がある。

一方、現在地方公共団体間のスプレッド水準の差について、機械学習のクラスタリングモデルを用いた分析を進めているが、投資銘柄により地方公共団体間で恒常的な収益性の差が生じているかについて引き続き分析を進める予定である。これらのために、長期データにおける信用リスクを表す重要な財政や市場指標の特定、構築モデルの説明力の検証、地方公共団体間の信用リスクの差の検証をさらに進めていく。

【APPENDIX】

各機械学習モデルの詳細は以下を参照のこと。

ID	Name	Reference
lr	Linear Regression	sklearn.linear_model_base.LinearRegression
lasso	Lasso Regression	sklearn.linear_model_coordinate_descent.Lasso
ridge	Ridge Regression	sklearn.linear_model_ridge.Ridge
en	Elastic Net	sklearn.linear_model_coordinate_descent.Elast...
lar	Least Angle Regression	sklearn.linear_model_least_angle.Lars
llar	Lasso Least Angle Regression	sklearn.linear_model_least_angle.LassoLars
omp	Orthogonal Matching Pursuit	sklearn.linear_model_omp.OrthogonalMatchingPu...
br	Bayesian Ridge	sklearn.linear_model_bayes.BayesianRidge
ard	Automatic Relevance Determination	sklearn.linear_model_bayes.ARDRRegression
par	Passive Aggressive Regressor	sklearn.linear_model_passive_aggressive.Passi...
ransac	Random Sample Consensus	sklearn.linear_model_ransac.RANSACRegressor
tr	TheilSen Regressor	sklearn.linear_model_theil_sen.TheilSenRegressor
huber	Huber Regressor	sklearn.linear_model_huber.HuberRegressor
kr	Kernel Ridge	sklearn.kernel_ridge.KernelRidge
svm	Support Vector Regression	sklearn.svm_classes.SVR
knn	K Neighbors Regressor	sklearn.neighbors_regression.KNeighborsRegressor
dt	Decision Tree Regressor	sklearn.tree_classes.DecisionTreeRegressor
rf	Random Forest Regressor	sklearn.ensemble_forest.RandomForestRegressor
et	Extra Trees Regressor	sklearn.ensemble_forest.ExtraTreesRegressor
ada	AdaBoost Regressor	sklearn.ensemble_weight_boosting.AdaBoostRegr...
gbr	Gradient Boosting Regressor	sklearn.ensemble_gb.GradientBoostingRegressor
mlp	MLP Regressor	sklearn.neural_network_multilayer_perceptron...
lightgbm	Light Gradient Boosting Machine	lightgbm.sklearn.LGBMRegressor

【参考文献】

- 秋庭伸也, 杉山阿聖, 寺田学, 加藤公一 (2019) 『見て試してわかる機械学習アルゴリズムの仕組み機械学習図鑑』 翔泳社。
- 石田三成, 中里 透 (2019) 「地方債の信用リスクとスピルオーバー」『地方債の経済分析』有斐閣。
- 江夏あかね (2018) 「米国における税制改革と地方債市場の行方」『野村資本市場クォーターリー』第21巻第4号, 1~6頁。
- 総務省 (2006) 「新地方公会計制度研究会報告書」総務省公表資料 (平成18年5月)。
- 総務省 (2019a) 「地方公会計の推進に関する研究会報告書」総務省公表資料 (平成31年3月)。

- 総務省 (2019b) 「統一的な基準による地方公会計マニュアル」総務省公表資料 (令和元年 8月改訂)。
- 総務省 (2020) 「統一的な基準による財務書類の作成状況等に関する調査 (令和2年3月31日時点)」総務省公表資料 (令和2年8月9日)。
- 加藤公一 (2018) 『機械学習のエッセンス — 実装しながら学ぶ Python, 数学, アルゴリズム』SBクリエイティブ。
- 田中宏樹 (2012) 「地方債をめぐる地方公共団体間信用連関 — 市場公募債パネルデータを用いた実証分析」『証券経済研究』第78号, 69～79頁。
- 丹波靖博, 原口健太郎 (2021) 「機械学習を用いた我が国における地方債の信用リスク分析」『JAFEE2021夏季大会予稿集』65～76頁。
- 中里 透 (2008) 「財政収支と債券市場 — 市場公募地方債を対象とした分析」『日本経済研究』第58号, 1～16頁。
- 原口健太郎 (2017) 「公会計財務諸表分析による地方公共団体の財務健全性の測定」『会計検査研究』第55号, 117～132頁。
- 原口健太郎 (2018) 「地方公共団体における財務健全性の国際比較 — 日本と米国の公会計財務諸表に係る統計分析 —」『会計検査研究』第57号, 13～35頁。
- 原口健太郎 (2019) 「地方公共団体における公会計財務諸表と地方債市場との関連性の発現過程 — 米国各州のデータを用いた時系列分析 —」『会計プロGRESS』第20号, 16～31頁。
- 原口健太郎 (2021) 「米国地方債市場における会計学研究的発展とわが国での展開可能性」『産業経理』近刊。
- 原口健太郎, 丹波靖博 (2021) 「わが国における公会計と地方債市場に関する分野横断的研究の展開可能性」『商学論集 (西南学院大学)』第67巻第3・4合併号, 169～199頁。
- 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』1～3頁。
- Aurélien Géron (2020) 『scikit-learn, Keras, TensorFlow による実践機械学習』下田倫大監修, 長尾高弘訳, オライリー・ジャパン。
- Capeci, J. (1991) Credit Risk, Credit Ratings, and Municipal Bond Yields: A Panel Study, *National Tax Journal*, 44(4), pp.41-56.
- Downing, C. and F. Zhang (2004). Trading Activity and Price Volatility in the Municipal Bond Market, *The Journal of Finance*, 59(2), pp.899-931.
- Duffie, D. and K. J. Singleton (2003) Credit Risk, *Princeton University Press*.
- Goldsterin, M. and G. Woglom. (1991) Market-Based Fiscal Discipline in Monetary Unions: Evidence from the U.S. Municipal Bond Market, *IMF Working Paper*, WP/91/89.
- Hattori, T. and H. Miyake (2015) Empirical Analysis of Yield Determinants in Japan's Municipal Bond Market: Does Credit Risk Premium Exist? *MPRA Paper 67127*, University Library of Munich, Germany.
- Haraguchi, K. and K. Oishi (2019) The International Differences in the Relationship between Public Financial Reports and Bond Markets: A Comparison between Japanese Local Governments and U.S. States, in Proceedings of the 17th Biennial CIGAR Conference in Amsterdam.
- Harris, L. E. and M. S. Piwowar. (2006) Secondary Trading Cost in the Municipal Bond Market, *The Journal of Finance*, 61(3), pp.1361-1397.
- Liu, P. and A. V. Thakor (1984) Interest Yields, Credit Ratings, and Economic Characteristics of State Bonds: An Empirical Analysis, *Journal of Money, Credit and Banking*, 16(3), pp.344-351.

- Novy-Marx, R. and D. R. Joshua (2011a) The crisis in local government pensions in the United States, in Robert Litan and Richard Herring, eds. : *Growing Old : Paying for Retirement and Institutional Money Management after the Financial Crisis* (Brookings Institution, Washington, DC).
- Novy-Marx, R. and D. R. Joshua (2011b) Public pension promises : How big are they and what are they worth? *The Journal of Finance*, 66(4), pp.1211–1249.
- Schwert, M. (2017) Municipal Bond Liquidity and Default Risk, *The Journal of Finance*, 72(4), pp.1683–1722.
- Wang, J. C. Wu, and F. X. Zhang. (2008) Liquidity, Default, Taxes, and Yields on Municipal Bonds, *Journal of Banking & Finance*, 32(6), pp.1133–1149.

【参考 URL】

総務省「地方債 Q&A」

https://www.soumu.go.jp/main_sosiki/c-zaisei/chihosai/chihosai_qanda.html

総務省「統一的な基準による財務書類に関する情報」

https://www.soumu.go.jp/iken/kokaikei/H30_chihou_zaimusyorui.html

総務省「地方公共団体の主要財政指標一覧」

https://www.soumu.go.jp/iken/shihyo_ichiran.html

日本証券業協会

<https://www.jsda.or.jp/>

ロンドン証券取引所 FTSE Russell グループ「The Yield Book Inc」

<https://www.yieldbook.com/m/products/yield-book/index-jpn.shtml>

謝 辞

本研究は科学研究費補助金（JSPS KAKENHI Grant Number JP19K23214, 20K02058, JP21K13412）の交付を受けて行ったものである。また、ロンドン証券取引所・FTSE グループ The Yield Book Inc. には、債券分析ソフトイールドブックを通じて効率的な分析ツールと貴重なデータを提供いただいた。この場を借りて御礼申し上げたい。