

地方銀行の貸出明細データ等とマクロ経済指標を用いた信用リスクに関する予兆分析の試行 vol.2

(要旨)

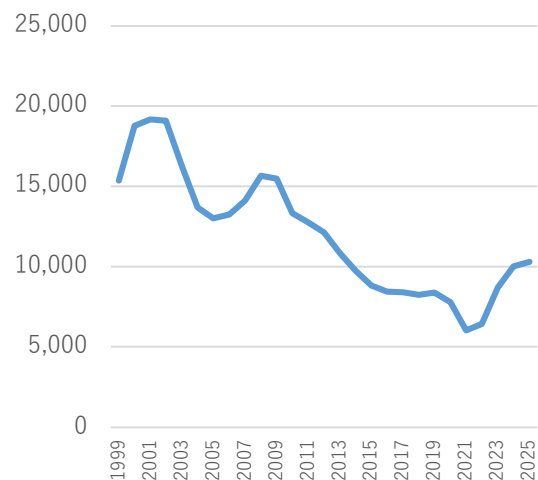
2025年6月に公表した「地方銀行の貸出明細データ等とマクロ経済指標を用いた信用リスクに関する予兆分析の試行」では、地方銀行の経営状況に影響する国内産業の変化の予兆捕捉を目的として、地方銀行の貸出明細データ等とマクロ経済指標を用い、機械学習によって信用リスクの動向（債務者区分の下方遷移）を予測するモデルを構築した。本稿では、分析対象とする業種を拡張するとともに、予測に寄与する指標の抽出方法の改善を試行した。その結果、予測に寄与する指標の解釈性の向上と、寄与する指標の業種間の違いについて示唆を得た。

1. はじめに

近年、我が国の経済情勢は緩やかに回復しつつある一方で¹、各国の通商政策や中東情勢などの地政学的なリスクによる不確実性の高まりや、人件費・原材料費上昇による物価高等を背景として、減少傾向にあった倒産件数は2022年頃から上昇に転じている（図表1）。このように、企業を取り巻く外部環境は大きく変化しており、これらの環境変化がもたらす影響の持続性や波及効果を見通すことは容易ではない。

金融庁では、2025年6月に公表したFSA Analytical Notes²（以下、「FAN」）「地方銀行の貸出明細データ等とマクロ経済指標を用いた信用リスクに関する予兆分析の試行」において、将来的な信用リスクの増加等、金融機関の経営に影響を

図表1 倒産件数推移



（出所）株式会社東京商工リサーチ

¹ 内閣府、「月例経済報告」（令和8年4月）

² 金融庁、FSA Analytical Notes (2025.6) vol.3 「地方銀行の貸出明細データ等とマクロ経済指標を用いた信用リスクに関する予兆分析の試行」

与える変化を早期に検知することを目的として、地方銀行³の貸出明細等と国内外のマクロ経済指標を用いて債務者区分の下方遷移を予測する機械学習モデルを構築⁴し、国内産業の変化の兆し（予兆）を捉えることを試みた。

前稿では製造業を対象としたが、本稿では、新たに建設業を対象に加え、両業種の変化の予兆を捉えることを目指した。また、機械学習モデルの解釈性を向上させることを目的として、事前にマクロ経済指標を対象に相関クラスタリングによる指標選択を実施したうえでモデルを構築した。以下、上記の業種の拡張と解釈性の向上を目的とした取組を加えた結果について解説する⁵。

II. 機械学習モデル構築の流れ

本章では、使用データ、業種および規模の選定、構築手法、指標選定について説明する。

1. 使用データ

本稿では前稿と同様、全国の地方銀行の法人向け貸出明細データ等及びマクロ経済指標データを用いている。両データを基準日単位で結合し、1つのデータセットとして扱う。貸出明細データ等は、全国地方銀行協会の信用リスク情報統合サービス（CRITS）のデータベースから抽出・加工されたデータ及び共同データプラットフォーム（以下、「共同 DP」）における貸出明細データの2種である。前者は2004年度から2023年度までのデータであり、後者は共同 DP が段階的に運用開始された2023年度以降のデータである。

2. 業種および規模の選定

マクロ環境の変化が企業に与える影響は業種等によって異なると考えられることから、機械学習に必要なデータ数が確保できる範囲で、分析対象とする企業群を限定することで、変数選択や結果の解釈が比較的容易になると考えられる。前稿では、モデル構築対象をなるべく均質な集団としつつもサンプル数を確保するため、業種による絞り込みを行いつつも、名目 GDP に占める割合が最大である点と国内銀行からの借入が大きい点を考慮し、製造業を分析対象に選定した。本稿では、製造

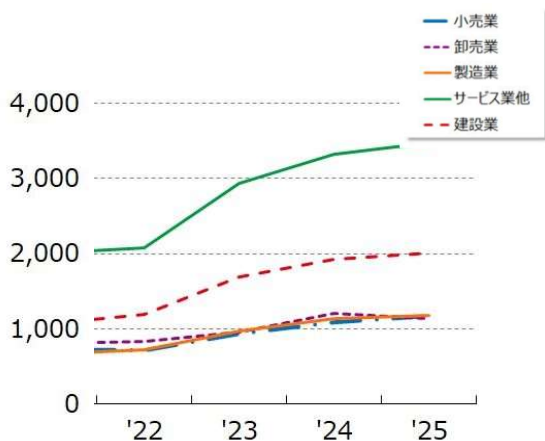
³ 全国地方銀行協会加盟行 62 行（2024 年 9 月末時点）。

⁴ 本稿における機械学習モデルの構築手法、マクロ経済指標の選択にあたっては、2022、2023 年度に金融庁が有限責任あずさ監査法人に調査・研究を委託した「AI や ICT 技術を活用した経営改善支援の効率化に向けた調査・研究」および「AI 技術を活用した経営改善支援の効率化に向けた調査・研究」を参考にしている。

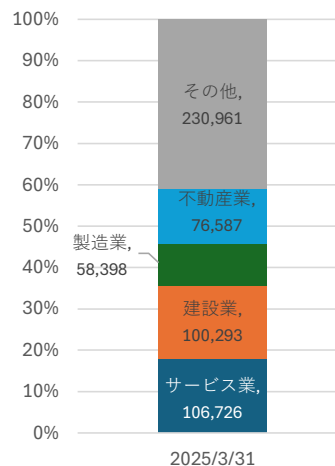
⁵ 本稿に記載の図表のうち、出所の明記のないものについては全て金融庁によるものである。

業以外の業種も分析対象として拡張し、機械学習モデルを構築した際、予測に寄与している指標の比較を行う。拡張する業種は、近年の倒産件数の水準（図表2）が高い点と2025年3月末時点において貸出明細データにおける債務者数（図表3）が多い点を考慮して建設業とする。

図表2 業種別の倒産件数推移



図表3 貸出明細データにおける業種別債務者数



（出所）株式会社東京商工リサーチ、一部抜粋

加えて、企業規模が大企業、中堅企業または中小企業であるかによって、経営環境の変化を背景とした信用リスクの変化の度合に違いがあることが考えられるため、本稿では、地方銀行の主要な貸出先であり、また、サンプル数も確保できる中小企業⁶に分析対象とする企業規模を限定する。

なお、国内銀行の中小企業向け貸出のうち、国内製造業向けは31.9兆円⁷、建設業向けは11.8兆円⁶となっている（いずれも2025年3月末時点）。このうち、本稿のモデル構築に使用したデータでは、製造業は12.6兆円（39.7%）、建設業は5.9兆円（49.5%）が把握可能である。

3. モデル構築手法および指標選定

本分析では、数多くの機械学習分野の調査・研究で利用されており、高い予測性能を発揮することで知られている eXtreme Gradient Boosting (XGBoost) を用いる。

指標は前稿と同じ指標を用いる²。すなわち、マクロ経済指標データは、国内外の約60個の指標を採用しており、金融・経済等に関する代表的なマクロ指標に加え、発電実績や公共機関からの受注額などが含まれ、幅広い業種に関連する指標を選定している。

⁶ 本分析の中小企業の定義は、全国地方銀行協会の信用リスク情報統合サービス（CRITS）のデータベースから抽出・加工されたデータの定義に基づく。

⁷ 日本銀行、「貸出先別貸出金」。それぞれ国内銀行における国内店の銀行勘定のうち製造業向け、あるいは建設業向けを対象とした残高。なお、企業規模は中小企業に限定している。

4. 予測対象

本稿では、機械学習モデルにおける予測対象を、前稿と同様、各金融機関が貸出先の財務状況等を勘案し信用リスクに応じて付与する区分である債務者区分⁸とし、基準日時点で債務者区分が正常先である債務者が1年後の同日時点で要注意先以下へ下方遷移（正常先からの下方遷移）するか否かを予測する。一般に、債務者区分の正常先からの下方遷移は、マクロ経済状況の影響を受けたものを含め、債務者の信用悪化を示しており、その結果、与信費用の増加を通じて金融機関の経営状況にも影響するものと考えられる。したがって、正常先からの下方遷移を予測対象とすることは、地方銀行の経営状況に影響を及ぼすような産業の変調を捉えるという本分析の目的に合致するものと考えられる。

5. 学習におけるデータ分割

本分析においては、機械学習モデルの構築に際して過学習による影響を抑制するため、データを分割⁹して学習させるとともに、予測精度の検証を行う。データの分割方法は、データの時系列性を考慮して、ローリングウィンドウ法を用いる。具体的には図表4に示すように7年分のデータを学習に使用し、学習データの最終年から1年空けた時点のデータを予測精度の確認に用いる。これを開始時点をずらしながら10グループ用意し（学習用データ）、それぞれのグループに対して学習及び検証を行う。予測精度については10回の平均値を採用する。なお、2023年度以降のデータについては学習には用いず、最終的な精度の確認と結果の確認に使用する（確認用データ）。

⁸ 本分析で用いたデータでは、債務者区分を「正常先、要注意先、要管理先、破綻懸念先、実質破綻先、破綻先」の6区分で表記しており、本分析における債務者区分の名称もこれに倣っている。

⁹ データの分割に際しては、次章の図表6、7で示す通り、債務者区分が下方遷移する割合の実績は、年度によっては10%以下と低位であり、そのまま学習した場合、データの偏りの影響を受けることが想定されることから、債務者区分が下方遷移しない先に対しアンダーサンプリングを実施している。

図表4 学習時におけるデータ分割

分割No.	時点																		
	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	
1	学習								検証										
2		学習								検証									
3			学習								検証								
⋮	⋮																		
9								学習								検証			
10									学習										検証

6. 説明変数及び目的変数と前処理

説明変数には、本章第3節で示した変数に前処理を施したものをを用いている。これは各時点における原数値以外に各時点の前年度からの変化による影響も機械学習モデルに取り込むことで、予測精度を向上することを目的としている。具体的には、全ての変数について、原数値そのまま、1年前同日の原数値、1年前との差分（原数値－1年前同日原数値）、1年前との比（原数値／1年前同日原数値）の4パターン（計356変数）を投入している。他方、目的変数は債務者区分下方遷移であり、下方遷移しなかった場合は「0」、下方遷移した場合は「1」としている。

なお、本分析の目的は将来予測であるため、説明変数と目的変数の時点は一致させていない。すなわち、説明変数は基準日とその1年前同日のデータである一方、目的変数は基準日から1年後同日の下方遷移の有無であり、基準日時点に観測可能なデータのみを用いた将来予測を実施している¹⁰。

7. 機械学習モデルの解釈性を向上させるための取組

変数の前処理について前節に記載したが、全ての説明変数を使用した機械学習モデルを構築した場合、説明変数内に強い相関関係を持つ変数が存在することで、予測精度の低下や機械学習モデルの解釈性の低下が考えられる。このため本稿では全ての説明変数で学習を実施した場合と、相関クラスタリングによる事前の変数選択を行った状態で学習を実施した場合の2パターンで機械学習モデルを構築した。

相関クラスタリングとは、相関係数に基づいて対象間の類似性を評価し、その関係性に基づいてクラスターを構成する方法である。相関係数を距離に変換するために、各変数 x_i 、 x_j に対し、 $d(x_i, x_j) =$

¹⁰ 実際には、公的統計や決算情報など、指標によって基準日からは数カ月後に観測可能となるものが多いが、予測対象は1年先であるため、数カ月遅れの情報を加味しても未来情報の先取りは生じない。

$1 - |r(x_i, x_j)|$ と指標 $d(x_i, x_j)$ を定義したうえで、基準となる相関係数の閾値以下の類似した変数同士を同一クラスターに分類する¹¹。分類手法は様々あるが、最もメジャーである Ward 法¹²は分散最小化と整合するユークリッド距離の構造を要求することから、今回は距離にかかる制約が少なく、解釈がよりシンプルな平均連結法を用いた。

なお、相関係数の閾値の設定によって、当然クラスターの分類結果が異なり、閾値を1に近づけるほど含まれる変数が少ないクラスターになる一方、0に近づけるほど多くの変数を含む1つの巨大なクラスターになる傾向がある。本分析に用いるにあたり、適切な個数の変数が含まれるクラスターになるように閾値を設定することが望まれる。今回は参考値として「シルエット平均」¹³によりクラスター内外の距離について確認することで、分類結果の妥当性の目安について指標化しつつ、得られた結果が前稿で得られた結果や一般に企業業績に影響するとされる指標の効果が観測されるか等の検証結果も踏まえ、閾値を設定した。

こうして作成したクラスターについて、各クラスターから代表値を1つずつ選択することで、それ以外の変数を除外した変数グループを用いて機械学習モデルを作成の上、特徴量の寄与度についての確認を試みた。代表値の選択方法は複数考えられるが、地銀の債務者への影響という観点で関連がよりあると思われるものを個別に選択した。実際のクラスター結果と代表値（抽出変数）について、図表5としてまとめる。なお、相関係数の閾値は0.63として階層クラスタリングを実施している。

¹¹ 欠損が多い変数のペアで相関係数を計算すると、関連の薄いもの同士が同一クラスターとなる可能性がある。（極端な例では、ある2つの変数について、分析期間中のある2時点でのみ両変数の値がデータセットに含まれている場合、相関係数は1または-1となる。）結果の安定性のため、今回は6つ以上の時点で両変数の値がデータセットに含まれていない変数の組み合わせについては、機械的に別々のクラスターへ分類する処理としている。

¹² Joe H. Ward, Jr. "Hierarchical Grouping to Optimize an Objective Function", *Journal of the American Statistical Association*, Vol.58, No.301(Mar.,1963), p.236-244.

¹³ Peter J. Rousseeuw. "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis", *Journal of Computational and Applied Mathematics*, Vol20 (1987), p.53-65.

図表5 相関クラスタリングによるクラスター分割結果（相関係数の閾値=0.63）

NO.	抽出変数	捨象変数
1	為替（ドル／ユーロ）	WTI原油価格
2	機械受注額（原系列・非製造業）	
3	鉱工業出荷指数（2020=100）	居住用_床面積_m2_全国
4	発電実績_全国	
5	製造工業生産能力指数（2020=100）_前年同月差	
6	公共機関からの受注額_前年同月比	
7	総貸出平残（外銀・円貸出）_前年同月比	
8	労働力人口（万人）_前年同月比	
9	為替（円／ドル）	名目実効為替レート指数、失業率（米国）
10	鉱工業在庫指数（2020=100）	
11	公共機関からの受注額	商業販売額_小売業計、地価_全用途_全国、国内企業物価指数（総平均）、労働力人口（万人）、製造工業生産能力指数（2020=100）、将来人口推計_全国、TOPIX、DJI
12	貸出約定平均金利（新規・短期）	旅客輸送量、貨物輸送量、貸出約定平均金利（新規・長期）、貸出約定平均金利（ストック・短期）、貸出約定平均金利（ストック・長期）、他3
13	住民基本台帳人口_全国_前年同月比	
14	商業販売額_商業計_前年同月比	商業販売額_卸売業計_前年同月比
15	実宿泊者数_全国	客室稼働率_全国、訪日外客数_総数_人
16	D.I./業況/中小企業/製造業/実績	D.I./業況/中小企業/非製造業/実績、D.I./業況/中堅企業/製造業/実績、D.I./業況/中堅企業/非製造業/実績、D.I./業況/大企業/製造業/実績、D.I./業況/大企業/非製造業/実績、他8
17	地価_全用途_全国_前年同月比	
18	住民基本台帳人口_全国	商業販売額_商業計、商業販売額_卸売業計、総貸出平残（外銀・円貸出）、国債利回り（米国・5年）、国債利回り（米国・10年）
19	国債利回り（5年）_前年同月差	国債利回り（3年）_前年同月差
20	国債利回り（3年）	無担保コールO/Nレート、国債利回り（1年）、国債利回り（5年）、エネルギー消費量_全国、エネルギー消費量_全国_前年同月比
21	国債利回り（1年）_前年同月差	居住用_床面積_m2_全国_前年同月比、居住用_工事費予定額_万円_全国、居住用_工事費予定額_万円_全国_前年同月比
22	無担保コールO/Nレート_前年同月差	貸出約定平均金利（新規・短期）_前年同月差、貸出約定平均金利（新規・長期）_前年同月差、貸出約定平均金利（新規・総合）_前年同月差、貸出約定平均金利（ストック・短期）_前年同月差、他1
23	為替（円／ユーロ）	
24	為替（円／ドル）_前年同月比	名目実効為替レート指数_前年同月差
25	機械受注額（原系列・非製造業）_前年同月比	
26	為替（ドル／ユーロ）_前年同月比	
27	WTI原油価格_前年同月比	
28	為替（円／ユーロ）_前年同月比	
29	貨物輸送量_前年同月比	
30	商業販売額_小売業計_前年同月比	
31	景気動向指数DI（先行指数）_前年同月差	景気動向指数DI（先行指数）、景気動向指数DI（一致指数）、景気動向指数DI（一致指数）_前年同月差
32	鉱工業在庫指数（2020=100）_前年同月差	
33	実質GDP成長率（前年度比）	完全失業率（%）_前年同月差、有効求人倍率_前年同月差、失業率（米国）_前年同月差
34	D.I./業況/中小企業/製造業/実績_前年同月差	D.I./業況/中小企業/非製造業/実績_前年同月差、D.I./業況/中堅企業/製造業/実績_前年同月差、D.I./業況/中堅企業/非製造業/実績_前年同月差、D.I./業況/大企業/製造業/実績_前年同月差、他6
35	倒産負債額_全国（百万円）_前年同月比	
36	中小企業売上見通LDI_前年同月差	景気動向指数CI（先行指数）_前年同月差、TOPIX_前年同月比、DJI_前年同月比、消費者態度指数_前年同月差
37	第3次産業総合_前年同月差	実質GDP成長率（米国・前期比）
38	中小企業売上見通LDI	実質GDP成長率（ユーロ圏・前期比）、景気動向指数CI（先行指数）、消費者態度指数
39	倒産件数_全国_前年同月比	
40	訪日外客数_総数_人_前年同月比	
41	実宿泊者数_全国_前年同月比	客室稼働率_全国_前年同月差、旅客輸送量_前年同月比、実質GDP成長率（アジア・前期比）
42	国債利回り（米国・10年）_前年同月差	国債利回り（米国・5年）_前年同月差
43	国内企業物価指数（総平均）_前年同月差	
44	発電実績_全国_前年同月比	

8. 構築した機械学習モデルの解釈手法

本稿では、説明変数の予測への寄与の確認手法について述べる。SHapley Additive exPlanations (SHAP) と特徴量重要度 (Feature Importance) による確認を行う。SHAP とは、個々のサンプルにおいて、各変数が予測値にどの程度寄与しているかを示す値 (SHAP 値) を算出、確認する手法で

あり、その絶対値が予測値への寄与の大きさを、符号が寄与の方向をそれぞれ意味する¹⁴。Feature Importance とは、機械学習モデルの学習において各変数が予測精度の改善に寄与する度合を示す指標であり、その値が大きい説明変数ほど相対的に予測精度の向上に寄与している¹⁵。

III. 結果

本章では、構築した機械学習モデルの予測精度と出力結果を確認するとともに、機械学習モデルの解釈を行うことで、予測に寄与する変数を確認する。

1. 債務者区分下方遷移状況の確認

図表 6 は、製造業の基準日時点に債務者区分が正常先であったもののうち、下方遷移した件数と割合を示している¹⁶。下方遷移率は、2009 年度のリーマン・ショック期に大幅に上昇し、その後減少傾向を辿った後、2020 年度の新型コロナウイルス感染症拡大期に上昇し、その後減少している。

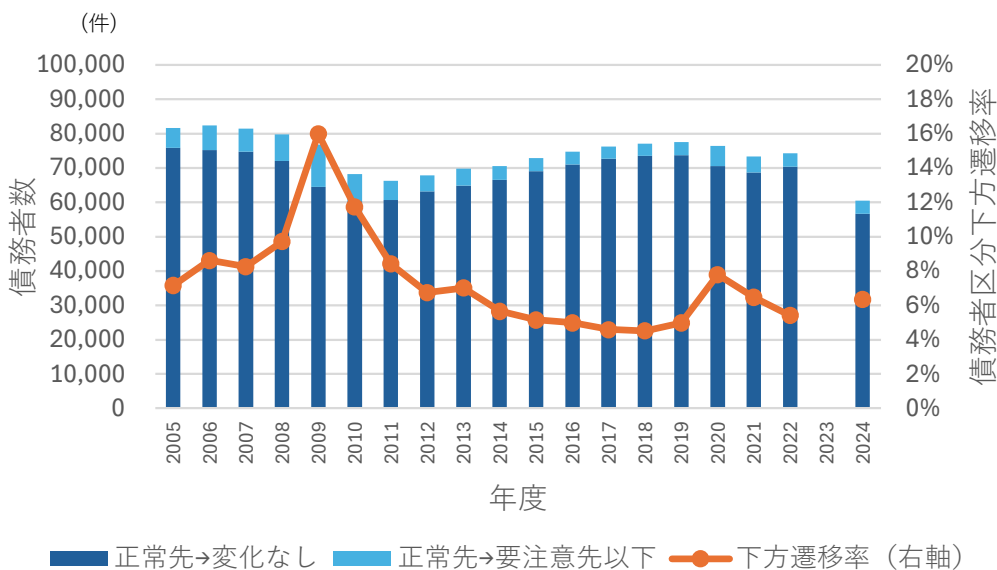
図表 7 は、建設業において同様に債務者区分の下方遷移の状況を示している。2009 年度のリーマン・ショック期にかけて下方遷移率が高止まりした後、減少し、2013 年度ごろから横ばいで推移している。図表 6 で示した製造業に比べて動きは緩やかとなっている。

¹⁴ 詳細は、Lundberg, S. and Lee, S. “A Unified Approach to Interpreting Model Predictions.” Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017, p.4768-4777.、日本銀行ワーキングペーパーシリーズ (2023.3) 「格付け分類モデルにおける機械学習の応用：機械学習の説明可能性を高める手法」、FSA Analytical Notes (2025.6) 「信用保証制度の利用状況に関する実態把握」等を参照。

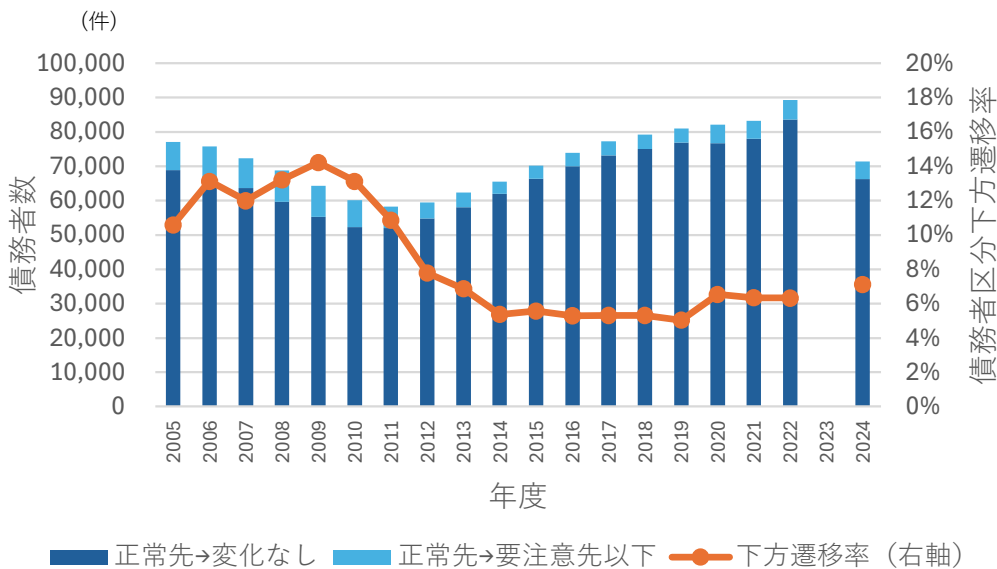
¹⁵ 詳細は、Freidman, J. H. “Greedy function approximation: A gradient boosting machine”, The Annals of Statistics, Vol. 29, No. 5, 2001, p.1189-1232.等を参照。

¹⁶ 2023 年度はデータの切り替えに起因して債務者区分の下方遷移が算出できないため図表上空白となっている。図表 7 も同様。

図表6 債務者区分下方遷移状況（製造業）



図表7 債務者区分下方遷移状況（建設業）



2. 機械学習モデルの予測精度

精度確認の指標は前稿と同様 AUC¹⁷を用いる。機械学習モデルの精度確認には正解率¹⁸、再現率¹⁹

¹⁷ AUC は機械学習モデルの出力結果の序列性について示す指標で 0～1 の間の値をとる。AUC=0.5 である場合に偶然と変わらない精度を示し、数値が高いほど精度が高いことを示す。

¹⁸ 正解率：機械学習モデルが出力したすべての予測のうち、正しく予測した割合。

¹⁹ 再現率：実際に予測ターゲットであった先のうち、モデルが予測ターゲットであると予測できた割合。

等も用いられるが、これらの指標は予測結果の正解／不正解の判定に際して閾値の設定が必要であるのに対し、AUC は閾値の設定を必要とせず、恣意性を排除しやすいことを踏まえて採用している。

構築したモデルの精度について、図表 8 構築した機械学習モデルにおける予測精度を業種別、モデル別（使用変数別）及び分割データ別に示す²⁰。建設業を見ると、相関クラスタリングを実施し、かつ、確認用データを用いた場合と、全変数を用いたモデルに確認用データを用いた場合、その他の場合に比べて若干低いものの、比較的高い精度と考えられる。

図表 8 構築した機械学習モデルにおける予測精度

(製造業)			(建設業)		
データ	機械学習モデル種類（使用変数）		データ	機械学習モデル種類（使用変数）	
	全変数	相関 クラスタリング 実施		全変数	相関 クラスタリング 実施
学習用	0.802	0.803	学習用	0.771	0.771
確認用	0.758	0.757	確認用	0.696	0.696

3. 機械学習モデルの解釈

ここで、説明変数の予測への寄与についてモデル別に確認する。確認には前章第 8 節に示した SHAP および Feature Importance を用いる。図表 9 から図表 12 に全ての変数を使用した機械学習モデルにおける SHAP の結果を示す。図表 9 より製造業においては「当期支払利息_割引料」が最も予測に寄与しているほか、「当期営業利益」、「当期利益」が続くなど、貸出あるいは債務者財務に関する指標が上位に位置していることがわかる。一方、図表 10 より建設業においては「当期純資産」が最も予測に寄与しているほか、「当期支払利息_割引料」、「当期営業利益」が続いている。製造業と建設業を比べると図表全体で見た場合、多少の上下はあるものの、おおむね類似した変数が上位に位置している。なお、図表 9、10 においては、個別債務者に対する SHAP 値を絶対値を取ったうえで平均化したものを示しているため、予測値に対し正負どちらの方向で寄与しているかについては示されていない点に注意が必要である。

この点に関して、図表 11、12 は個々のサンプルにおける SHAP 値の分布を蜂群図で示している。

²⁰ 前章第 5 節で記載した通り、ローリングウィンドウ法でデータを分割しているため、本稿の機械学習モデル構築においては各業種、使用変数パターンの機械学習モデルに対して 10 回算出される。ここではその平均値を示している。

この図は色で変数のとる値の大小を、横軸で個々のサンプルにおける SHAP 値を、各変数の縦方向の広がり度で SHAP 値の分布を示している。図表 11 より製造業においては、「当期支払利息_割引料」が当該変数の値が高い（赤色）場合に、予測値を上げる（債務者区分が下方遷移しやすい）方向へ寄与していることがわかる。逆に、「当期営業利益」は、当期変数の値が高い（赤色）場合に予測値を下げる（債務者区分が下方遷移しにくい）方向へ寄与していることがわかる。図表 12 より建設業においても、個々の変数の振る舞いは図表 11 に示した製造業と同様の傾向となっていることが分かる。

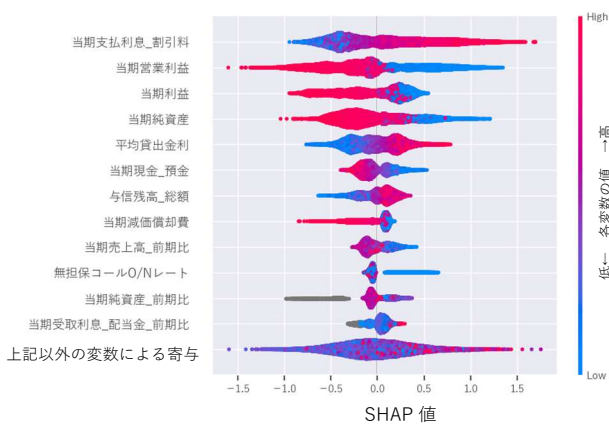
図表 9 予測への寄与が大きい変数
(SHAP 値の絶対平均、製造業)



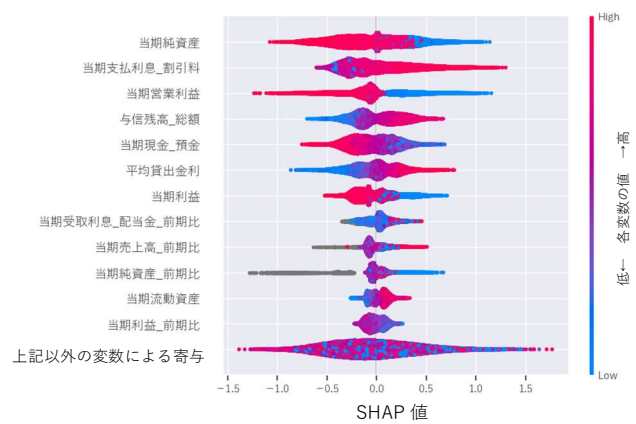
図表 10 予測への寄与が大きい変数
(SHAP 値の絶対平均、建設業)



図表 11 個々のサンプルに対する SHAP 値の分布
(製造業)



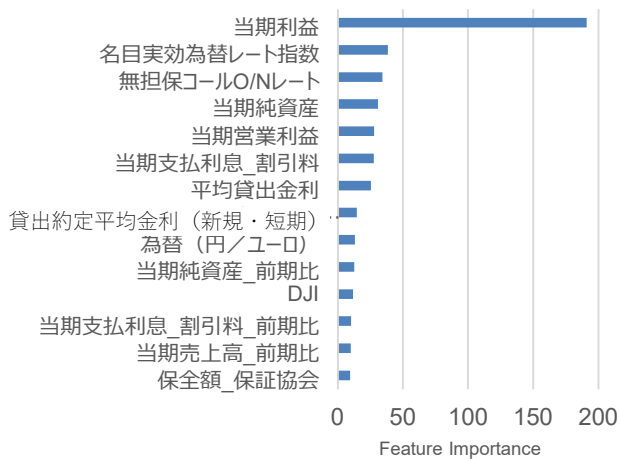
図表 12 個々のサンプルに対する SHAP 値の分布
(建設業)



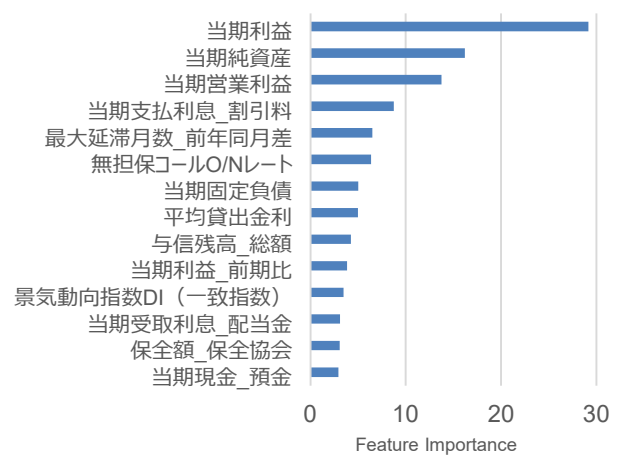
続いて、図表 13、14 に全ての変数を使用した機械学習モデルにおける Feature Importance の結果を示す。いずれの業種においても「当期利益」が共通して最も寄与しているほか、「当期純資産」等の貸出、債務者財務に関する指標が上位に位置する傾向となっている。マクロ経済指標については両業種に共通して「無担保コール O/N レート」が上位に位置しているほか、製造業においては「名目実効為替レート」が上位に位置するなど、建設業に比べて多くのマクロ経済指標が上位に位置する傾向にある。

なお、Feature Importance の結果については機械学習モデルの構築時における決定木の分割プロセスによる影響を受けるほか、説明変数間に相関関係がある場合に結果が不安定になることが指摘されており²¹、幅をもった解釈が必要である。

図表 13 予測への寄与が大きい変数
(Feature Importance、製造業)



図表 14 予測への寄与が大きい変数
(Feature Importance、建設業)



次に、図表 15 から図表 18 に相関クラスタリングを実施した場合の機械学習モデルにおける SHAP の結果を示す。

²¹ Lundberg, S. and Lee, S. "A Unified Approach to Interpreting Model Predictions." Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017, p.4768-4777.

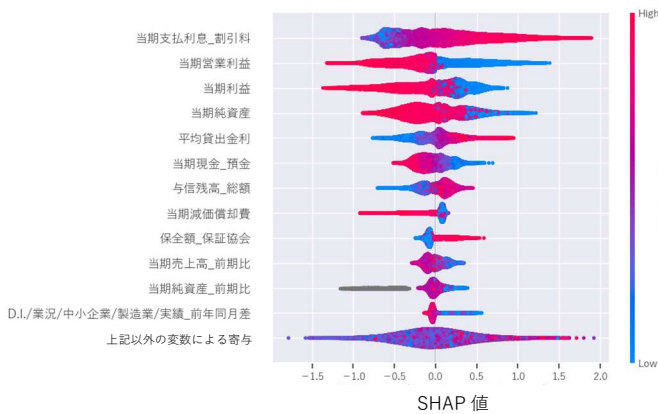
図表 15 予測への寄与が大きい変数
(SHAP 値の絶対平均、製造業)



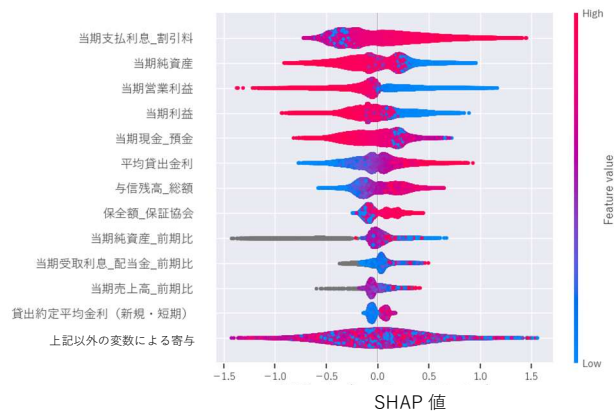
図表 16 予測への寄与が大きい変数
(SHAP 値の絶対平均、建設業)



図表 17 個々のサンプルに対する SHAP 値の分布
(製造業)



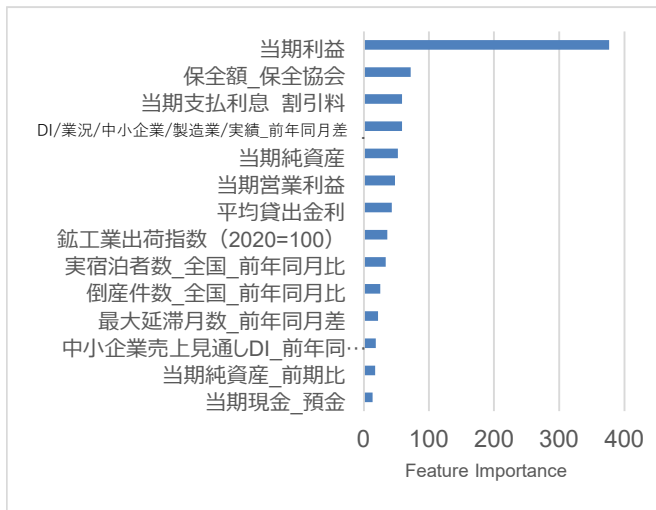
図表 18 個々のサンプルに対する SHAP 値の分布
(建設業)



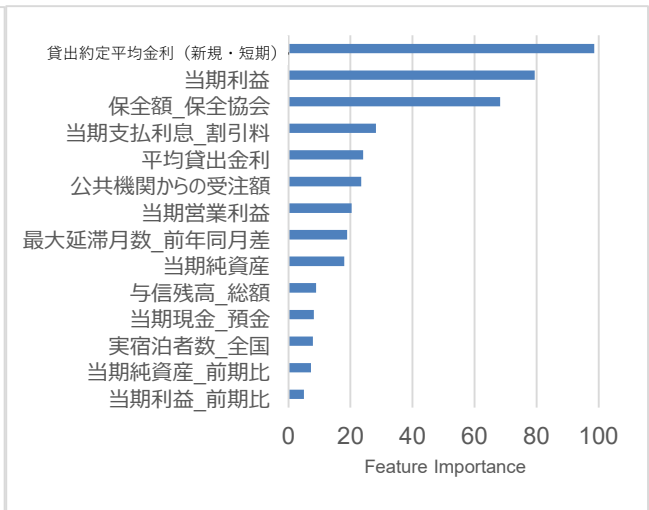
図表 15 から図表 18 の結果を踏まえると、相関クラスタリング実施前と全体的な傾向は変わらないが、製造業においては「D.I./業況/中小企業/製造業/実績 (前年同月差)」の寄与、建設業においては「貸出約定平均金利」の寄与が確認された。相関係数の高かった変数間で分散していた SHAP 値が集約されたものと考えられる。

続いて、Feature Importance の結果を図表 19、20 に示す。

図表 19 予測への寄与が大きい変数
(Feature Importance、製造業)



図表 20 予測への寄与が大きい変数
(Feature Importance、建設業)



図表 15 から図表 18 の SHAP 値のものと同様の傾向が見られ、製造業においては「D.I./業況/中小企業/製造業/実績 (前年同月差)」に加えて「鉱工業出荷指数」の重要度、建設業においては「貸出約定平均金利」に加え、「公共機関からの受注額」の重要度が一定程度確認された。いずれもその業態に関連する指標であると考えられるが、これらの変数はクラスターから代表値を1つ選択したものであり、同クラスターに分類された別指標が関連している可能性もあるため、幅をもった解釈が必要である。

なお、マクロ経済指標について個社の財務指標と比較すると、Feature Importance と比較して SHAP 値のスコアは低くなる傾向がみられる。Feature Importance の高いマクロ経済指標は、その業種全体の景気局面を判別する前提条件として重要だが、個別先のランクダウンを直接決めているのは、受注動向や資金繰りなどの企業固有の要因の方が大きいことを示しているものと思われる。

4. 出力結果²²

機械学習モデルによって算出された個々の債務者に対する債務者区分下方遷移の予測確率について、その分布の時系列推移を示したものが図表 21、22 である²³。

図表 21 は製造業における結果を示している。債務者区分下方遷移の予測確率の平均値や中央値 (50 パーセンタイル値) は、図表 6 で示した債務者区分の下方遷移状況と同様、2008、2009 年度

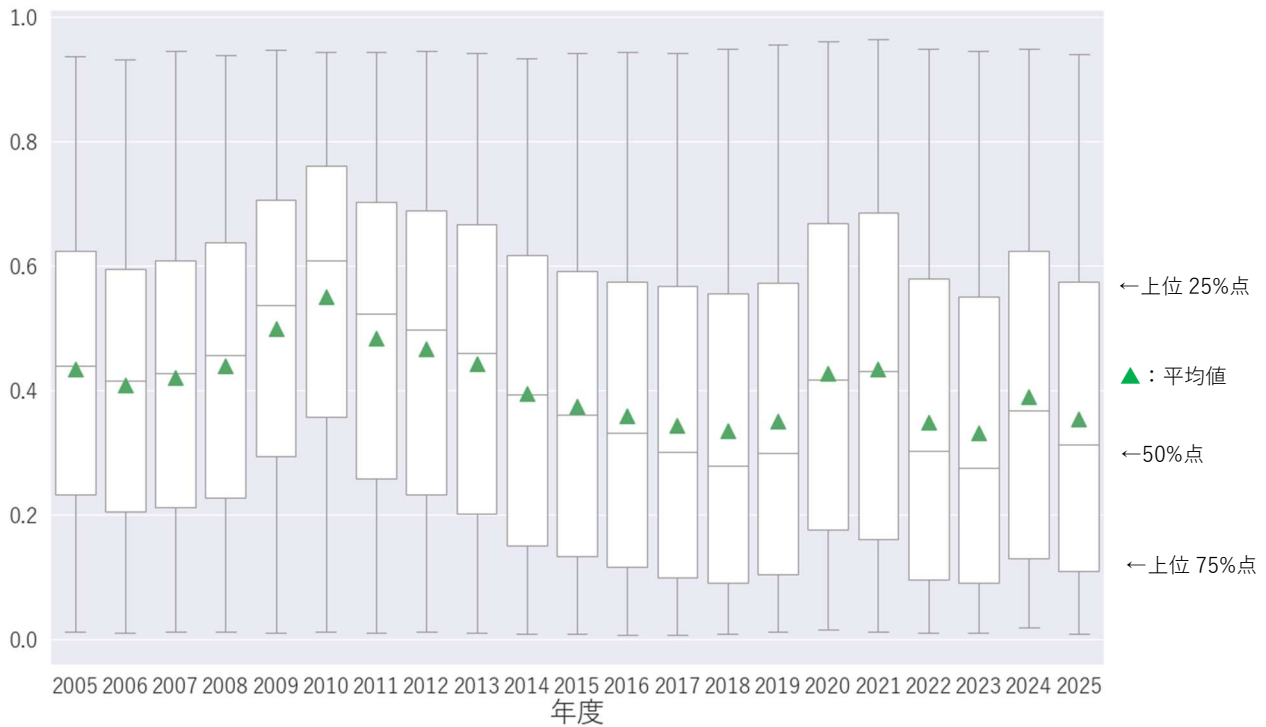
²² 本節においては全サンプルに対して機械学習モデルで債務者区分の下方遷移確率を付与、集約した結果を示している。

²³ 本図表では関連クラスタリング実施前のデータにおける結果を示している。なお、関連クラスタリング実施後のデータを用いた場合も、傾向に大きな変化は見られなかった。

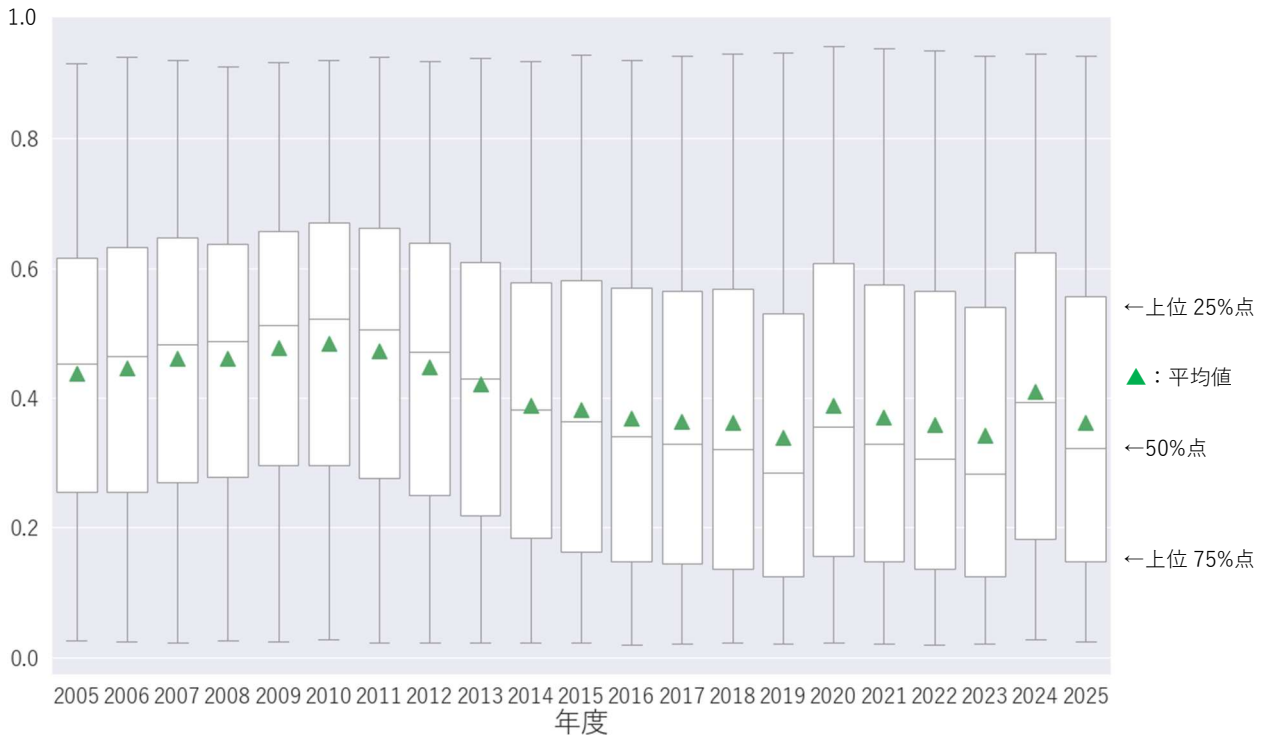
のリーマン・ショック期で大幅に上昇しているほか、2020年度の新型コロナウイルス感染症拡大期においても小幅に上昇が見られ、前稿での分析結果と同様の傾向となっており、学習方法の変更等に関わらず整合的な結果が得られた。

図表 22 は建設業における結果を示している。債務者区分下方遷移の予測確率の平均値や中央値（50%点）は、図表 7 で示した債務者区分の下方遷移状況と同様、2009、2010年度のリーマン・ショック期にかけて高止まりした後、2013年度ごろから横ばいで推移している。図表 21 で示した製造業に比べて全体として動きは緩やかとなっている。

図表 21 機械学習モデル算出の債務者区分下方遷移の予測確率の分布推移（製造業）



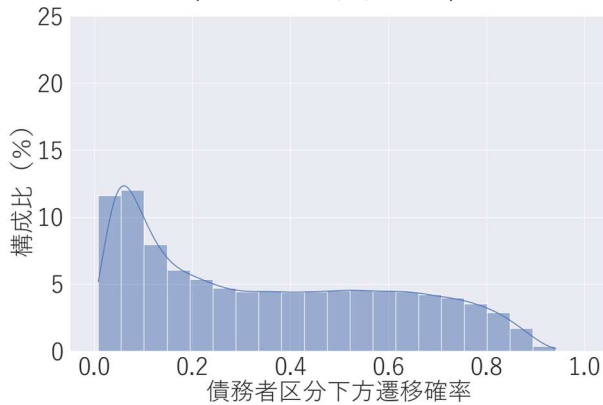
図表 22 機械学習モデル算出の債務者区分下方遷移の予測確率の分布推移（建設業）



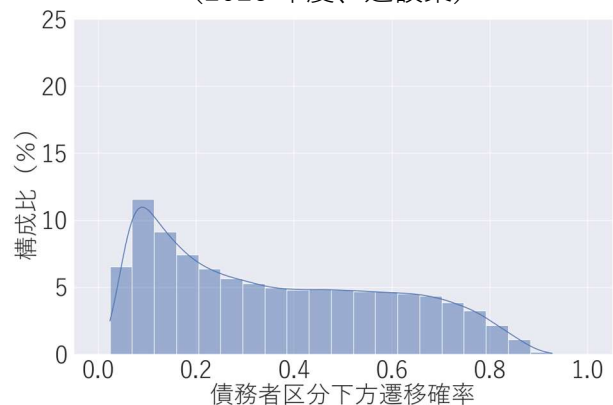
このうち、直近年度（2025年度）における債務者区分下方遷移確率、ヒストグラムおよび密度分布を描いたものを図表 23、24 に示す。両者とも 2025 年度では低位側に分布が集中しており、高位

側に向かって裾野が広がっている。これは経済的に安定している時期の形状²⁴であるが、今後の債務者の財務状況およびマクロ経済指標の変化のほか、使用データの切り替えに伴う影響も考えられることから、今後の挙動に注視が必要と考えられる

図表 23 債務者区分下方遷移確率分布
(2025 年度、製造業)



図表 24 債務者区分下方遷移確率分布
(2025 年度、建設業)

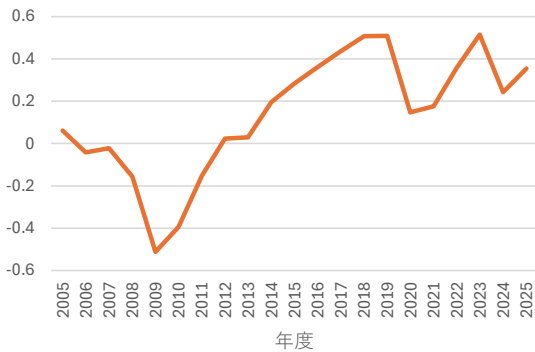


債務者区分下方遷移確率の分布の形状を表す値のうち、歪度²⁵の推移について図表 25、26 において確認を行う。前稿において、通常は債務者区分下方遷移確率が低いところに極めて多く分布している（歪度が大きい）一方、リーマン・ショック期や新型コロナウイルス感染症拡大期のような不安定期に債務者区分下方遷移確率が大きく算出された企業が比較的多くなり、分布の観点からは平準化される（歪度が小さい）ことを指摘した。図表 25 の製造業を対象に構築したモデルの結果における歪度の推移については、前稿で示したものと同様に 2009 年度のリーマン・ショック期で水準が大幅に低下しているほか、2020 年度の新型コロナウイルス感染症拡大期においても水準の低下が見られる結果となっている。また、直近の動きについては 2023 年度以降上昇と下降を繰り返しており、こうした挙動の背景については検証が必要と思われる。他方、図表 26 の建設業を対象に構築したモデルの結果における歪度の推移については、2010 年度のリーマン・ショック期にかけて低位で推移した後、上昇、および高止まりしている傾向となっている。直近の動きは製造業と同様、上昇と下降を繰り返しており、こうした挙動の背景についても注視していくことが必要と思われる。

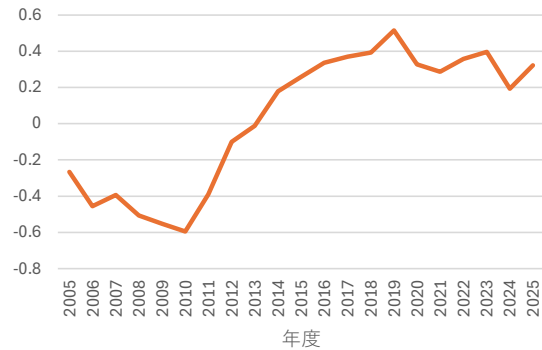
²⁴ 経済的な環境と分布の形状については注釈 2 の文献を参照いただきたい。

²⁵ 歪度：分布の非対称性を示す指標。今回のような度数分布の場合、中心に度数が集中していれば 0、中心より左側に度数が集中していれば正、中心より右側に度数が集中していれば負の値をとる。

図表 25 債務者区分下方遷移確率の分布における歪度の推移（製造業）



図表 26 債務者区分下方遷移確率の分布における歪度の推移（建設業）



4. 本分析の活用と課題

前節までの分析結果より、金融機関の経営に影響を与えるような国内産業の予兆を捉えるにあたり、モデルが出力する分布の形状を確認しつつ、特にその数値の動きが明確で解釈が容易である平均値、中央値、歪度の3指標を中心にモニタリングすることで、足元の状況変化を捕捉できる可能性があるとの示唆が得られた。指標の活用方法としては、前年からの各指標の大小比較や傾向確認、経済状況が不安定であったリーマン・ショック期や新型コロナウイルス感染症拡大期等の特定時期との水準比較を行うことで、今後の地方銀行の経営状況に影響を与えるような予兆を示すものと考えられる。

また、業種ごとに予測に寄与する変数が異なることから、各業種の信用リスクの動向をモニタリングするにあたり、注視すべき変数の探索に活用しうる手法を示すことができたものと考えられる。

一方で課題もある。1つ目が、変数の下方遷移への重み付けについてである。現状、債務者区分の下方遷移については、下方遷移先の区分に関わらず一律に扱っている。この点については下方遷移先の区分によって金融機関に与える影響が異なることから、下方遷移先の区分ごとに重み付けを行うことも考えられる。2つ目が構築した機械学習モデルの解釈に関する点である。今回のクラスタリングに基づき行った予測に寄与する変数分析の結果は、今回構築された機械学習モデルに対する結果であり、別の手法を用いて構築した機械学習モデルに対して同様の解釈手法を用いた場合、異なる傾向の結果が得られる可能性がある。この点に関しては、厳密には複数の手法で機械学習モデルを構築し、それぞれに解釈手法を適用し、共通して抽出される変数を調査することが考えられる。その他、前稿にも記載した予測時点とデータの反映時点のタイムラグや使用データの連続性に関する課題についても引き続き対応していきたい。

IV. まとめ

本分析では、地方銀行の貸出明細データ等及びマクロ経済指標データを用いて金融機関の経営状況に影響を与えるような国内産業の変化の予兆把握を試みた。その結果、前稿では債務者区分の下方遷移を一定の精度で予測可能な機械学習モデルを構築したが、そのモデルをベースとして、対象業種を拡張し、債務者区分の下方遷移の予測に寄与する変数について異なる業種間で比較した。結果、製造業においては「DI（中小企業／製造業）」や「鉱工業出荷指数」、建設業においては「貸出約定平均金利」や「公共機関からの受注額」の寄与が推察されるなどし、足元の予兆把握において、一定の示唆を得られることが確認できた。

金融庁は、金融行政のさらなる進化に向けて、データ分析力の向上やデータ整備等を通じ、データ活用の高度化に取り組んでいる。FSA Analytical Notes（FAN）は、その取組みの一環として金融庁が実施した主なデータ分析をまとめて公表するものである。これまでに公表された FAN は金融庁ウェブサイト（<https://www.fsa.go.jp/common/about/kaikaku/fsaanalyticalnotes/index.html>）で閲覧可能。