

地域銀行の信用リスク管理態勢の実態把握に向けた分析 (保全状況に関する分析・債務者区分の遷移予測モデル)

(要旨)

本稿では、共同データプラットフォームで収集された銀行の貸出明細データから、保全状況に関する分析及び債務者区分の遷移予測モデルに関する検証を実施した。保全状況は貸出先企業の規模・特性、資金使途等によって異なるため、その適切性を論じるものではないが、特に複数銀行が貸出を行っている債務者（共通貸出先）や圏外向け（越境貸出）の保全率が低くなる傾向が確認された。債務者区分の遷移予測モデルの検証では、債務者区分が要注意先以上から破綻懸念先以下にランクダウンする際の予測モデルは他の遷移パターンの予測モデルと比較して、財務情報のみで精度の高い予測が可能と示唆された。

1. はじめに

金融庁では金融機関の信用リスク管理態勢に係る実態把握を定量データにより深める観点から、共同データプラットフォームで取得した貸出明細データを使用した分析を行い、FSA Analytical Notes (2025.1) vol.2「共通貸出先に対する債務者区分の付与状況に関する分析」を公表した。本稿ではその取り組みを深化させるため、地域銀行¹の貸出明細データを用いて、保全状況に関する分析及び債務者区分の遷移予測モデルの検証に取り組んだ。前者に関しては、単独共通別（単独貸出先か共通貸出先か）²や越境別³等の観点から、保全状況の実態把握を行った。後者に関しては、債務者区分の決定に財務情報が影響している程度は債務者区分ごとに異なる可能性を確認する観点から、機械学習の手法を用いて財務情報のみで複数の債務者区分遷移の予測モデルを構築し、精度比較を行った。

¹ 本稿における業態区分について、「地方銀行」は埼玉りそな銀行及び全国地方銀行協会加盟行、「第二地方銀行」は第二地方銀行協会加盟行、「地域銀行」は地方銀行及び第二地方銀行を指す。

² 「単独貸出先」とは一つの銀行が貸出を行っている債務者を指し、「共通貸出先」とは複数銀行が貸出を行っている債務者を指す。当該分類は脚注1の地域銀行及び大手行（みずほ銀行（みずほ信託銀行を含む）、三菱UFJ銀行、三井住友銀行、三井住友信託銀行、りそな銀行、あおぞら銀行、SBI新生銀行）を対象として判断しており、信用金庫などの業態は考慮されていない。一部地域では貸出規模の大きい信用金庫等も存在するものの、全体としては、本稿の分析上は影響は限定的と考えられる。

³ 本稿における越境別について、都道府県単位の本店所在地ベースで判断している。なお、各行によって営業基盤とする区域は異なっており、本店所在地外の都道府県への貸出をもって越境貸出とはいえない場合もある点には留意が必要である。

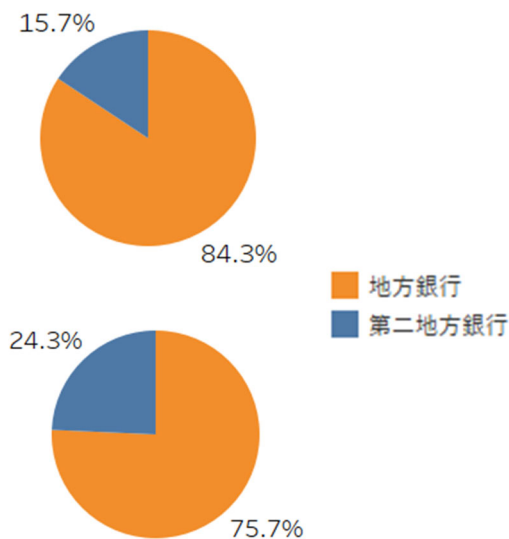
II. 保全状況に関する分析

本章では、保全状況の現状把握とその傾向分析を行った。本章においては、貸出明細データの2024年3月基準日時点の地方公共団体を除く法人貸出先のうち、保全に関する情報が把握可能な先を検証対象とした⁴。当該対象は地域銀行の地方公共団体を除く法人貸出残高全体の約5割を占めている。なお、本分析は、保全状況が貸出先企業の規模・特性や銀行の与信方針等によって異なることを前提とした実態把握であり、保全状況の適切性等を論じることを意図したものではない。

1. 基本データ

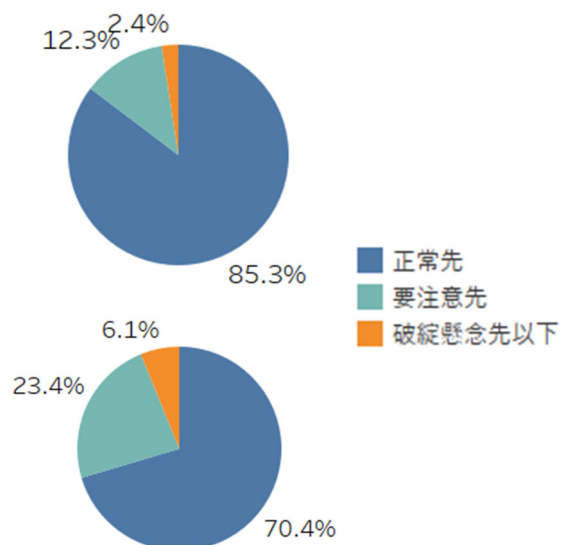
図表1は業態別割合、図表2は債務者区分別割合を示したものである。業態別残高割合では地方銀行が84.3%、第二地方銀行が15.7%を占めている。債務者区分別残高割合では正常先が85.3%、要注意先が12.3%、破綻懸念先以下が2.4%と要注意先以上が大部分を占めている⁵。

図表1 業態別割合
【上図：残高割合、下図：先数割合、表：実数】



地方銀行	貸出残高（兆円）	82.0
	先数	456,725
第二地方銀行	貸出残高（兆円）	15.2
	先数	146,518

図表2 債務者区分別割合
【上図：残高割合、下図：先数割合、表：実数】



正常先	貸出残高（兆円）	82.9
	先数	424,778
要注意先	貸出残高（兆円）	11.9
	先数	141,385
破綻懸念先以下	貸出残高（兆円）	2.4
	先数	37,080

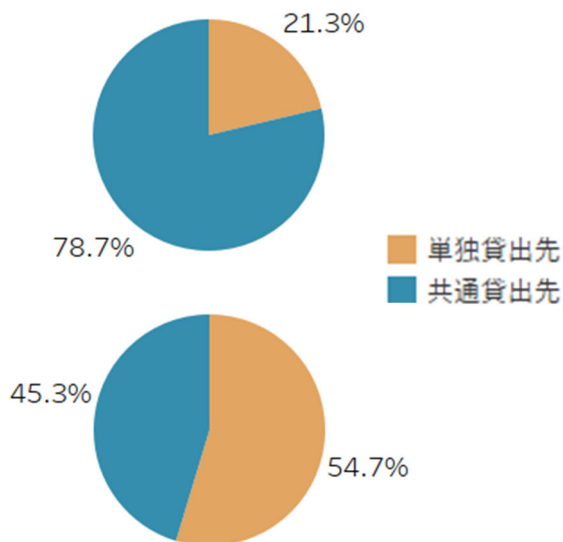
⁴ 共同データプラットフォームにおいて、本格的なデータ収集が開始されるのが2025年3月期からであり、本稿で使用しているデータは、本格的なデータ収集開始に向けたデータ精度向上等の取り組み途中のものである。そのため、本稿の集計では共同データプラットフォーム対象金融機関全てが集計対象とはなっていない。

⁵ 「要注意先」とは、共同データプラットフォームにて報告されている債務者区分「その他要注意先」及び「要管理先」を合わせたものである。また、「破綻懸念先以下」とは、「破綻懸念先」、「実質破綻先」及び「破綻先」を合わせたものである。

図表3は単独共通別割合、図表4は越境別割合を示したものである。どちらも、残高割合だと共通貸出先、圏外向けの割合が大きいものの、件数割合だと単独貸出先、圏内向けの割合が大きくなる。これは、共通貸出先における圏外向けの1先当たりの貸出残高（平均値）が大きいことが要因⁶と考えられる。

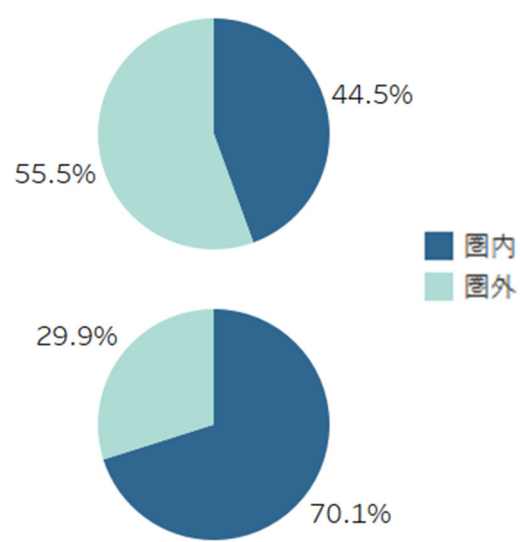
図表3 単独共通別割合

【上図：残高割合、下図：先数割合、表：実数】



図表4 越境別割合

【上図：残高割合、下図：先数割合、表：実数】



2. 保全率の状況

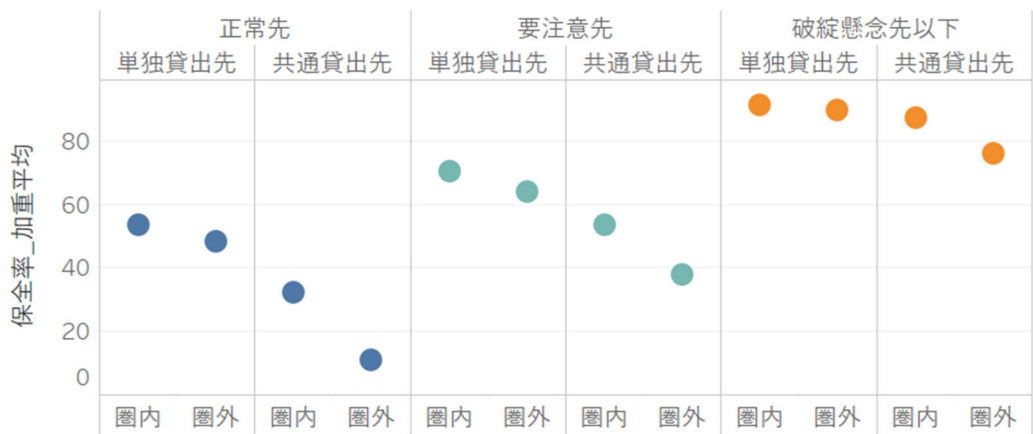
本節では、保全率⁷を用いて、様々な切り口から保全状況を確認した。図表5は、債務者区分別、単独共通別、越境別の保全率の状況を示したものである。これによると、破綻懸念先以下は総じて高い保全率であるものの、正常先及び要注意先においては、単独貸出先・圏内向け、単独貸出先・圏外向け、共通貸出先・圏内向け、共通貸出先・圏外向けの順に保全率が低くなる傾向が確認された。特に、共通貸出先においては、単独貸出先と比較して越境別による保全率の差が大きい。このような傾

⁶ 1件当たりの貸出残高（平均値）について、共通貸出先の圏外向けが約4.1億に対して、単独貸出先の圏内向けが約1.8億、単独貸出先の圏外向けが約0.6億、共通貸出先の圏内向けが約0.8億である。

⁷ 本稿における「保全率」は、データの制約上、「(担保額+保証等額+個別貸倒引当金)/貸出残高*100」で算出している。そのため、一般的な保全率の算出の際分子に含まれている、一般貸倒引当金等が考慮されていない点に留意が必要。

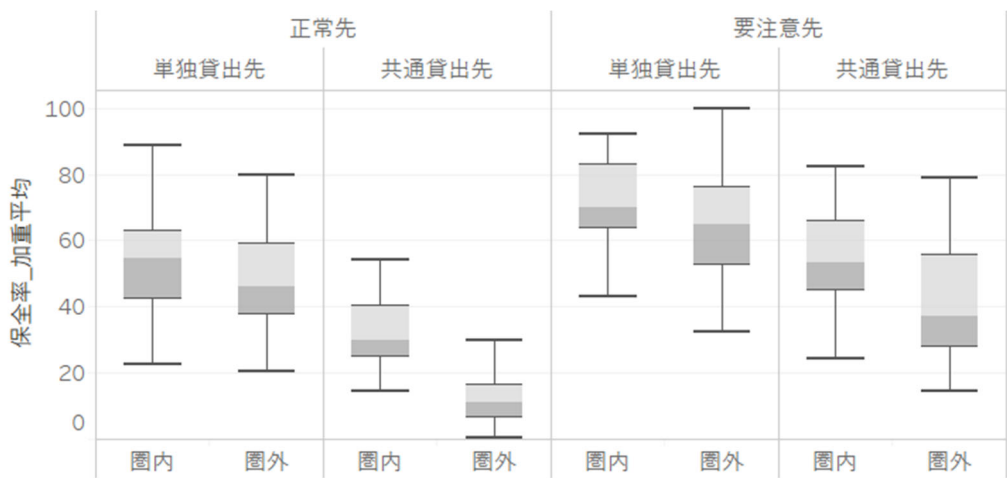
向がみられる理由としては、債務者側からは複数借入となるため単独貸出先よりも共通貸出先の方が保全に係る交渉が困難となることや、圏内よりも圏外の方が希薄なリレーションや新規取引獲得等により保全に係る交渉が困難となっていることが考えられる。なお、破綻懸念先以下の保全率はその属性においても比較的高いため、以降では基本的に正常先及び要注意先における保全率に着目することとする。

図表 5 保全率（加重平均）



図表 6 のとおり、銀行別保全率の分布を確認したところ、銀行によるばらつきがみられた。詳細を確認すると図表 5 で示された傾向と一致しない銀行も存在した。このような傾向の違いがみられる理由として、貸出ポートフォリオ、地域性や保全に関する考え方等による違いが影響していると考えられる。

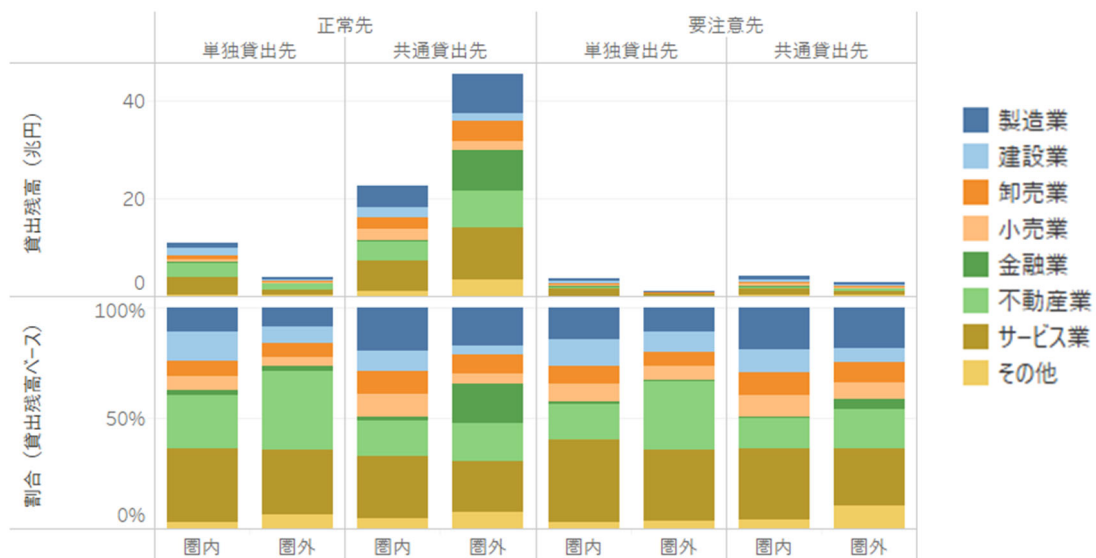
図表 6 銀行別保全率（加重平均）の分布⁸



⁸ 箱の上端、中心、下端、ヒゲの上端、下端はそれぞれ第3四分位、中央値、第1四分位、最大値、最小値を示す。

図表7は、業種別貸出残高・残高割合、図表8は、業種別保全率を示している。両図表より、正常先・共通貸出先・圏外向けにおいて、貸出残高の大きい業種の中では、金融業向けの貸出残高が大きく、保全率が低いことが分かる。この理由として、金融業向けは東京所在の大手金融機関向け貸出残高が大きいことが影響している。また、不動産業向けの保全率は比較的高位であるが、これは不動産業向けは他業種と比較して担保が確保しやすいことが要因として考えられる。

図表7 業種別貸出残高・残高割合



図表8 業種別保全率

	正常先				要注意先			
	単独貸出先		共通貸出先		単独貸出先		共通貸出先	
	圏内	圏外	圏内	圏外	圏内	圏外	圏内	圏外
製造業	44.8	41.6	22.4	6.5	68.9	64.2	43.7	30.7
建設業	58.3	50.7	38.4	17.5	81.4	73.6	65.5	48.6
卸売業	52.4	40.8	30.1	8.5	76.0	71.1	54.8	38.3
小売業	55.8	48.6	25.5	10.9	81.3	73.0	53.8	32.7
金融業	8.1	8.5	9.5	2.0	17.5	58.8	28.9	26.1
不動産業	69.4	62.3	54.8	22.3	73.2	67.7	66.3	57.0
サービス業	48.3	43.3	31.9	14.0	65.8	58.9	54.2	40.5

続いて、図表9は、共通貸出先における債務者区分一致⁹、メイン別¹⁰に各債務者区分における保全率を示したものである。債務者区分一致先のメイン行と非メイン行の保全率を見ると、どの債務者

⁹ 債務者区分を正常先、要注意先、破綻懸念先以下に分類したうえで、全ての銀行が同じ債務者区分を付与している場合を「債務者区分一致」とする。

¹⁰ データの制約上、圏内向けかつ貸出先における貸出割合が最も大きい銀行を「メイン行」、それ以外を「非メイン行」としている。そのため、全て圏外向けの場合や貸出割合が最も大きい銀行が圏外向けである場合には、全て「非メイン行」に分類している。また、「非メイン行」であっても、実質的に「メイン行」に近い「準メイン行」である場合も想定される点には留意が必要である。

区分においても非メイン行の方が保全率が低い。

図表 9 債務者区分一致、メイン別の保全率

			メイン行 国内	非メイン行 国内 国外	
債務者区分一致 正常先	保全率_加重平均		33.7	28.5	17.7
	貸出残高（兆円）		12.5	6.1	20.9
要注意先	保全率_加重平均		61.6	57.0	50.7
	貸出残高（兆円）		1.0	0.3	0.6
破綻懸念先以下	保全率_加重平均		88.1	85.6	76.5
	貸出残高（兆円）		0.3	0.1	0.1

図表 10 は、共通貸出先における債務者区分不一致¹¹のメイン行債務者区分別の保全率を確認したものである¹²。これによると、メイン行の保全率と比較して非メイン行・圏外向けは、非メイン行が破綻懸念先以下を付与している場合を除いて、特に低くなっている可能性がある¹³。なお、債務者区分の組み合わせは多岐に渡るため、解釈に留意する必要がある。

図表 10 債務者区分不一致のメイン行債務者区分別の保全率

			正常先 国内 国外		要注意先 国内 国外		破綻懸念先以下 国内 国外	
メイン行 正常先	メイン行	保全率_加重平均	42.0					
		貸出残高（億円）	20,917					
	非メイン行	保全率_加重平均	31.6	2.4	50.7	26.2	88.5	66.8
		貸出残高（億円）	3,347	124,682	2,498	4,517	30	31
メイン行 要注意先	メイン行	保全率_加重平均			54.6			
		貸出残高（億円）			19,093			
	非メイン行	保全率_加重平均	43.6	24.0	40.7	23.5	81.0	70.3
		貸出残高（億円）	2,138	3,378	1,377	2,548	413	445
メイン行 破綻懸念先以下	メイン行	保全率_加重平均					87.3	
		貸出残高（億円）					4,215	
	非メイン行	保全率_加重平均	66.9	23.8	49.0	30.2	89.7	78.3
		貸出残高（億円）	56	63	335	457	311	510

最後に、貸出契約開始時期別の観点から確認を行った。図表 11 のとおり、パターン A のように貸出債権が一つであれば該当債権における取引開始日、パターン B のように貸出債権が複数あった場合は該当債権の中で最も古い取引開始日を、それぞれ貸出契約開始時期（図表 11 であれば、どちら

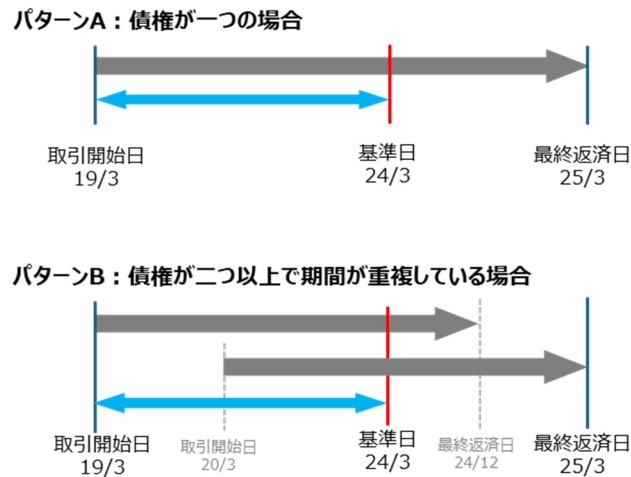
¹¹ 債務者区分を正常先、要注意先、破綻懸念先以下に分類したうえで、1 行でも異なる債務者区分を付与している場合を「債務者区分不一致」とする。

¹² メイン行における債務者区分別の保全率を示しているため、本稿定義上のメイン行が存在しない先は集計対象外となっている。

¹³ 仮に、越境の有無に関わらず、貸出割合が最も大きい銀行とそれ以外の銀行の保全率を比較した場合においても、同様に破綻懸念先以下を除いて、それ以外の銀行の方の保全率が低くなる。

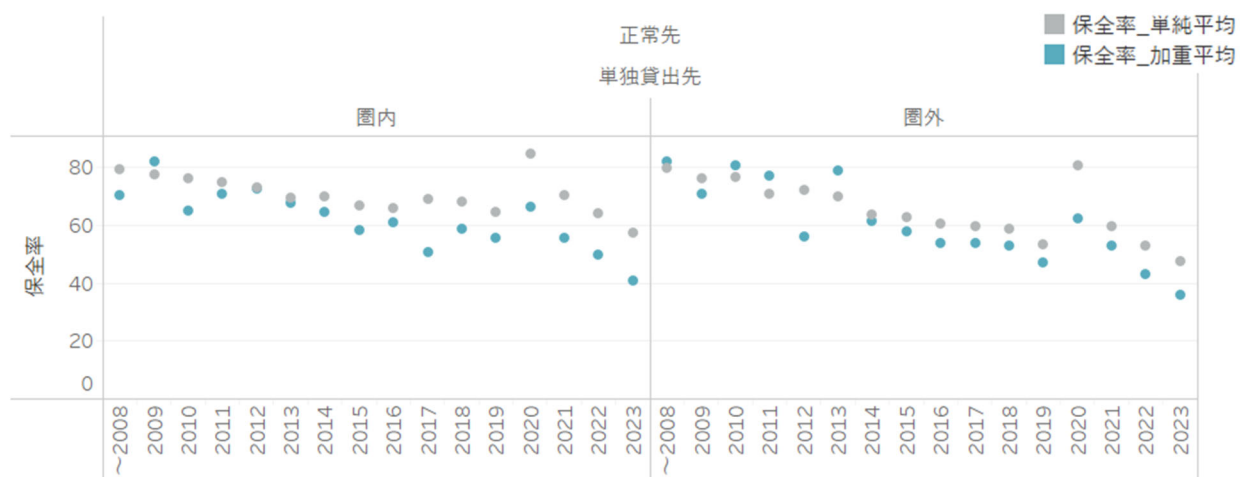
も「19/3」)として使用した。

図表 11 貸出契約開始時期について

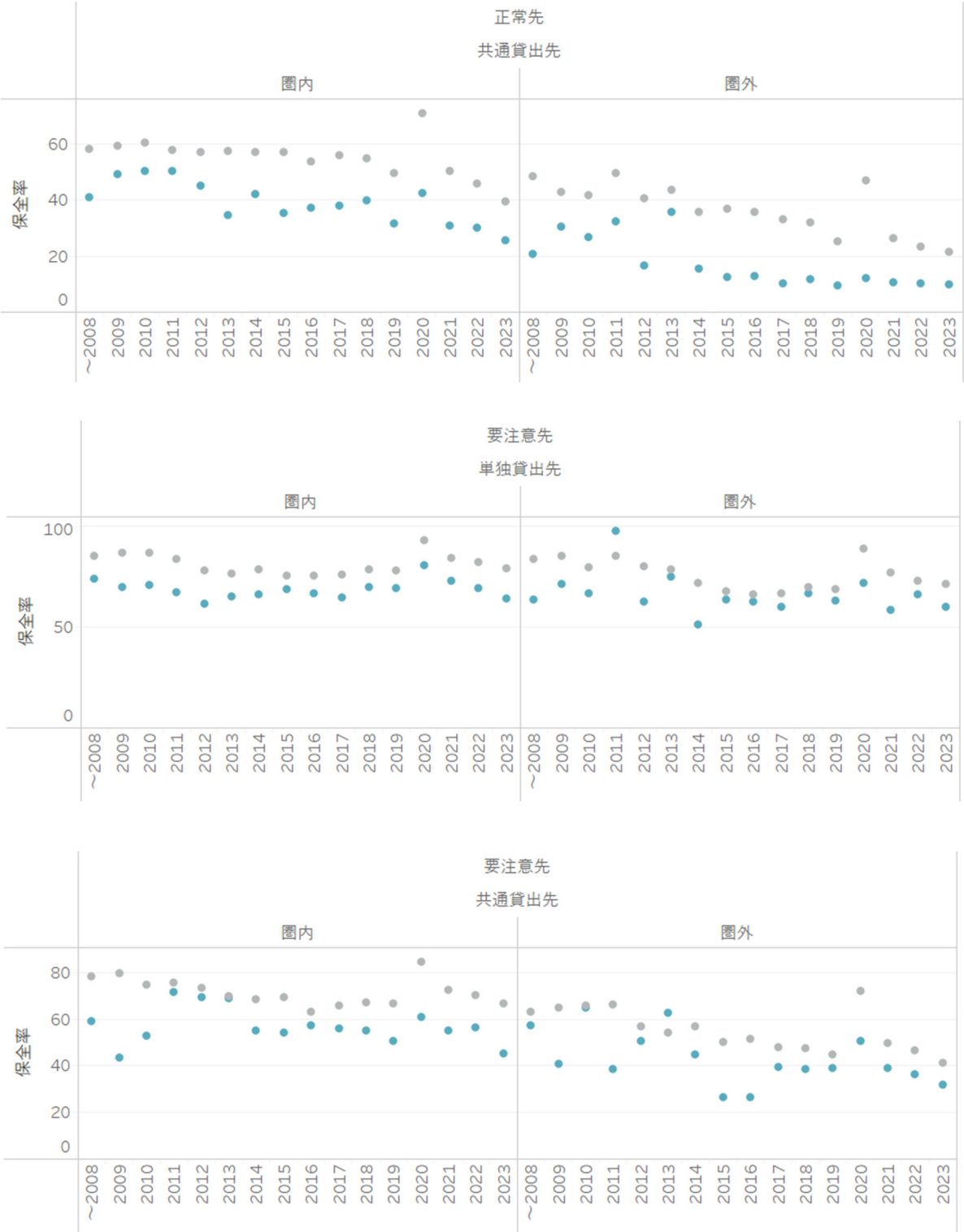


図表 12 は、貸出契約開始時期（年度）別保全率を示したものである。これによると、貸出契約開始時期によって保全率が異なっている様子が確認でき、特に正常先においては貸出契約開始時期が近年であるほど保全率が低いことが分かる。他方で、要注意先は正常先ほどその傾向は確認されないことから、信用力の高い先に対する担保・保証に依存しない融資が促進されている可能性が示唆される。なお、貸出契約開始時期が 20 年度（横軸「2020」）に該当する先の保全率が比較的高い理由は、当該期間において実施されていた、民間金融機関における実質無利子・無担保融資によって、信用保証協会付き融資が多く実行されていたためと考えられる。

図表 12 貸出契約開始時期（年度）別保全率¹⁴



¹⁴ 「～2008」は2008年度以前を指す。



3. 検証

これまでの図示から、保全率は債務者区分以外にも、単独共通別、越境別、貸出契約時期別及び銀行別でそれぞれ傾向があることを確認した。本節では、保全率に影響を与えうる企業財務情報や企業規模等の影響を排除しつつ、特に単独共通別及び越境別に着目し、保全率との関係について重回帰分析を用いて検証した。

本節で用いた検証式は以下のとおりである。地方銀行から借入を受けている債務者区分が要注意先以上の先¹⁵(企業*i*)の保全率を目的変数(y_i)とし、共通貸出先の有無(*Shared borrowers Dummy_i*)、圏外向け貸出の有無(*Cross border Dummy_i*)及びその交差項(*Shared borrowers Dummy_i* * *Cross border Dummy_i*)を説明変数に設定した。また、保全率に影響を与えうる、企業財務情報¹⁶、規模、業種、貸出先所在地及び貸出契約開始時期をコントロール変数(*Controls_i*)とした(変数の詳細な定義：図表 13)。

$$y_i = \beta_0 + \beta_1 \text{Shared borrowers Dummy}_i + \beta_2 \text{Cross border Dummy}_i + \beta_3 \text{Shared borrowers Dummy}_i * \text{Cross border Dummy}_i + \text{Controls}_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2)$$

図表 13 変数一覧

Object variable	y_i	保全率 (%)
Explanatory variables	1.Shared borrowers Dummy	共通貸出先であれば「1」、それ以外「0」
	2.Cross border Dummy	圏外向けであれば「1」、それ以外「0」
	3.Shared borrowers Dummy*Cross border Dummy	1と2の交差項
Controls	Lending start year Dummy	貸出開始年度ダミー
	Order of loan Dummy	貸出順が1番の先を「1」、それ以外「0」
	ROA	営業利益÷総資産
	Interest payable	支払利息割引料÷(短期借入金+長期借入金)
	Debt ratio	(短期借入金+長期借入金)÷総資産
	Cash and deposit ratio	現金預金÷総資産
	Size	企業規模(資本金の常用対数)
	Industry Dummy	業種ダミー(製造業、建設業、卸売業、小売業、金融業、不動産業、サービス業)
	Prefectures Dummy	貸出先所在地ダミー(都道府県ベース)

検証結果は図表 14 のとおりである。今回用いた説明変数については、全て有意にゼロであることが棄却され、保全率と負の相関があることが確認された。つまり、保全率に影響を与えうる財務情報

¹⁵ 加えて、本検証式は、貸出形態がシンジケートローンに分類されるものを除いた債務者の財務情報が取得可能な地方銀行の貸出先を、対象としている。

¹⁶ 前節にて債務者区分も保全率に影響を与えていることを示しているが、債務者区分は金融機関による信用リスク管理の考え方等の影響を受ける指標のため、本節ではより客観的な企業財務情報等をコントロール変数として採用した。

等が同様¹⁷であっても、共通貸出先、圏外向けは、それぞれ単独貸出先、圏内向けと比較して保全率が低くなり、特に共通貸出先かつ圏外向けは追加的に保全率が低くなることが示された。

図表 14 検証結果

	Coefficient	Std.Error	
<i>Shared borrowers Dummy</i>	-12.066	0.185	***
<i>Cross border Dummy</i>	-7.457	0.248	***
<i>Shared borrowers Dummy*Cross border Dummy</i>	-7.778	0.302	***
n	318,828		
Adjusted-R ²	0.249		

***, ** and * indicate significance at the 0.1%, 1%, 5% levels

4. 小括

本章では、様々な切り口から保全率の状況を確認し、特に単独共通別や越境別の保全率の傾向を統計的に検証した。その結果、共通貸出先や越境貸出先は、単独貸出先や圏内貸出先と比べ保全率が低い傾向が見られ、加えて統計的な検証によって、債務者区分が要注意先以上の先においては財務状況等の要因を考慮しても有意に差があることを確認した。ただし、越境貸出について、各行によって営業基盤とする区域は異なっており、本店所在地外の都道府県への貸出をもって越境貸出とはいえない銀行も含まれている。また、メイン別についても、メイン行は主に設備資金目的の貸出を行うなど、対応する資金使途が異なることが保全率に影響を与えている可能性もある。加えて、保全率は貸出先の財務、リレーション、地域経済状況及び銀行における保全の考え方等を踏まえて調整されるものと考えられる。そのため、繰り返しになるが、本検証は現状の保全の適切性等について論じるものではない点に留意が必要である。

¹⁷ なお、企業財務指標及び企業規模の代わりに債務者区分（正常先・要注意先のダミー変数）をコントロール変数として使用した場合も同様な結果になる。

Ⅲ. 債務者区分の遷移予測モデル

これまで FSA Analytical Notes では、信用リスクの動向を定量的に把握すること等を目的に、債務者区分に着目した信用リスク推計モデルを複数の手法で試みてきた¹⁸。他方で、債務者区分は、財務情報以外の定性情報等も勘案し付与されるものであり、財務情報のみで予測できる精度は債務者区分によって異なる可能性がある。そこで、本章では今後の予測モデル等の検討に活かすため、債務者区分の遷移パターンごとに財務情報で構築される予測モデルの精度を確認した¹⁹。

1. 検証方法

本章では、図表 15 で示す遷移パターンごとに、財務情報のみを用いた予測モデルを作成し、それぞれの ROC 曲線²⁰及び AUC²¹から精度を比較することで検証を行った。予測精度が高いほど、財務情報のみで十分な予測が可能であり、逆に予測精度が低いほど、財務情報以外の定性情報等が債務者区分の付与に影響を与えていると考えられる。

図表 15 債務者区分の遷移パターン

予測対象	定義	検証対象先数
【Model 1】 ランクダウン（要注意先）	t-1 期に正常先であったものの、t 期に要注意先へランクダウンした先	ランクダウンあり：5,740、 ランクダウンなし：474,959
【Model 2】 ランクダウン（破綻懸念先以下）	t-1 期に要注意先以上であったものの、t 期に破綻懸念先以下へランクダウンした先	ランクダウンあり：659、 ランクダウンなし：563,896
【Model 3】 ランクアップ（正常先）	t-1 期に要注意先であったものの、t 期に正常先へランクアップした先	ランクアップあり：5,159、 ランクアップなし：78,181
【Model 4】 ランクアップ（要注意先以上）	t-1 期に破綻懸念先以下であったものの、t 期に要注意先以上へランクアップした先	ランクアップあり：184、 ランクアップなし：7,140

¹⁸ FSA Analytical Notes (2023.6)「銀行融資の信用リスクに関する分析」、FSA Analytical Notes (2024.7) vol.1「地方銀行における不動産業向け貸出及びその債務者区分の動向に関する分析」

¹⁹ 本章で使用するデータは、地方銀行における貸出先のうち、財務情報が取得可能な先。対象期間は、2023 年 9 月末から 2024 年 6 月末時点。

²⁰ 縦軸に陽性率（真陽性 / (真陽性 + 偽陰性)）、横軸に偽陽性率（偽陽性 / (偽陽性 + 真陰性)）をとり、モデルにおける分類の判定基準を変化させた場合に陽性率と偽陽性率がどのように変化するかを図示したもの。

²¹ ROC 曲線の下部の面積を表し、値が大きいほど予測精度が高いことを意味する。予測精度 100%だと 1、完全ランダムだと 0.5 となる。

予測精度を客観的に評価するため、ランダムフォレスト、XGBoost、ロジスティック回帰、サポートベクターマシン（以下、「SVM」）及び多層パーセプトロン（以下、「MLP」）といった複数の学習モデルを用いて比較を行っている²²。なお、図表 15 の対象先数のとおり対象先数が不均衡であることで生じる過学習²³を避けるため、アンダーサンプリング²⁴によるデータ不均衡を改善する処理を行っている。

図表 16 に本章で使用した特徴量（予測モデルに入力するデータのこと）を示している。特徴量については、財務指標を複数作成したうえで、相関係数の比較や Boruta²⁵といった手法を用いて、選択を行っている。

図表 16 使用した特徴量

特徴量	定義
size	資本金の常用対数
ROE	当期利益/自己資本
ROIC	(営業利益+法人税等充当額)/(株主資本等合計+短期借入金+長期借入金+社債)
net DE ratio	((短期借入金+長期借入金+社債)-現預金)/純資産
net cash ratio	((現預金+有価証券)-(短期借入金+長期借入金+社債))/総資産
capital adequacy ratio	自己資本/総資産
labors share	人件費/付加価値 (付加価値=営業利益+減価償却費+人件費)
DCR	(短期借入金+長期借入金+社債)/(現預金+有価証券+有形固定資産)
sales interest expense ratio	支払利息・割引料/売上高
corporate profit margin	経常利益+営業外費用-法人税等充当額/純資産
common stock ordinary profit ratio	経常利益/資本金
operating cash flow per employee	(当期利益+減価償却費)/期末役員従業員数

2. 検証結果

図表 17 に各予測モデルの ROC 曲線及び AUC を示した。これらを確認すると、Model 2（ランクダウン（破綻懸念先以下））は、Model 1、3 及び 4（ランクダウン（要注意先）及びランクアップ）と比較すると高い予測精度が示された。このことから、特に破綻懸念先以下にランクダウンするような信用リスクが高い事例に対しては、財務情報のみで精度の高い予測モデルの構築が可能である。

²² 各学習モデルの概要は、後述する BOX を参照。

²³ 予測モデルが多数派に偏って学習し、少数派に対する汎用性が低下すること。

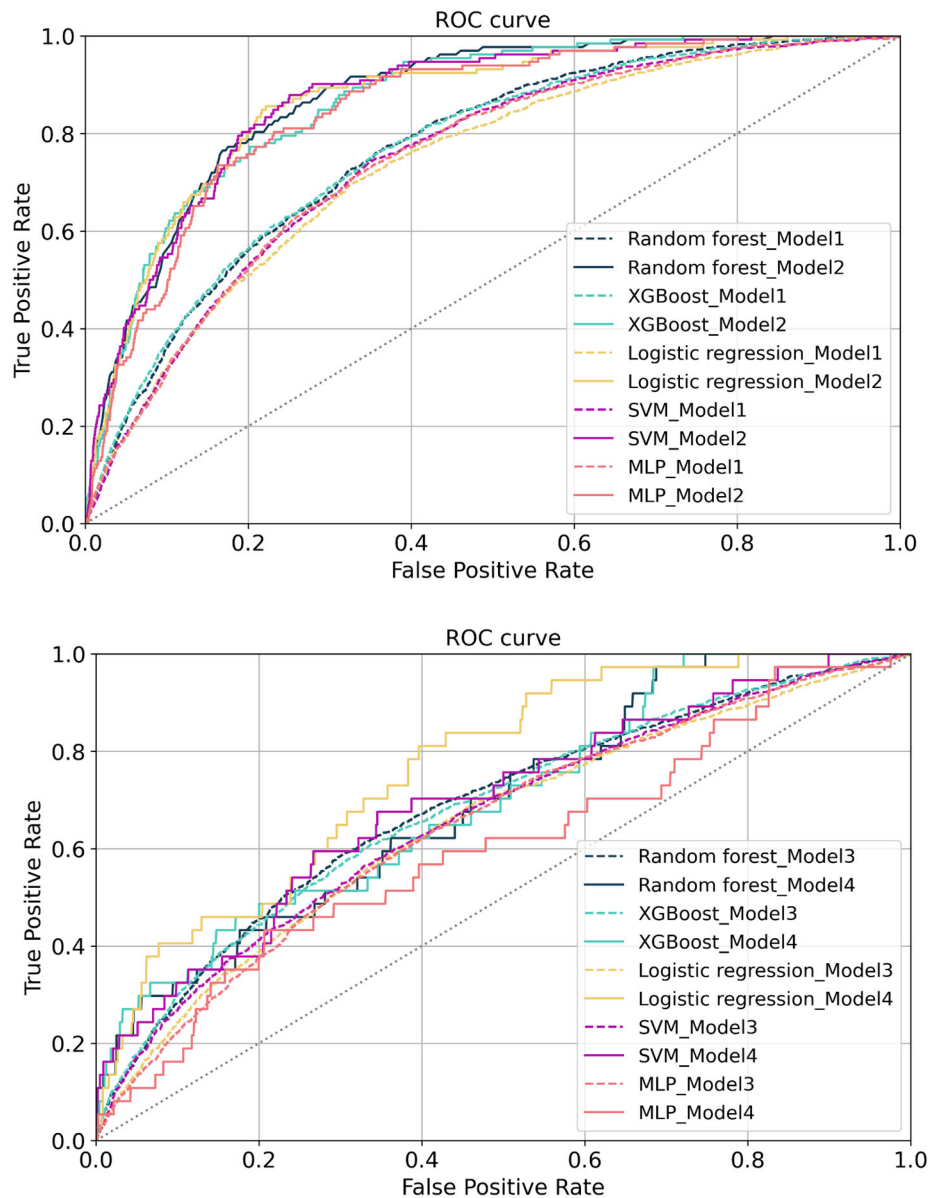
²⁴ 多数派のデータを少数派のデータに合わせて削減することで、データが不均衡な状態を改善する方法。

²⁵ 予測対象と関係のない特徴量を追加したうえでランダムフォレストを実施し、変数重要度を本来の特徴量と比較することで、選択すべき特徴量を選定する手法。

Kursa,M.B. and Rudinicki,W.R:"Feature Selection with the Boruta Package", Journal of Statistical Software, Vol.36,Issue11(2010)

一方、債務者区分に関係なくランクアップする際の予測モデルにおいては、財務情報のみでは十分な予測精度は確認できず、その他定性情報が影響を与えている可能性が示された。

図表 17 各予測モデルにおける ROC 曲線及び AUC
【上図：Model 1、Model 2、下図：Model 3、Model 4】



	AUC			
	Model1	Model2	Model3	Model4
Random forest	0.77	0.87	0.68	0.69
XGBoost	0.77	0.87	0.68	0.70
Logistic regression	0.74	0.87	0.65	0.77
SVM	0.75	0.87	0.66	0.69
MLP	0.75	0.85	0.65	0.60

3. 小括

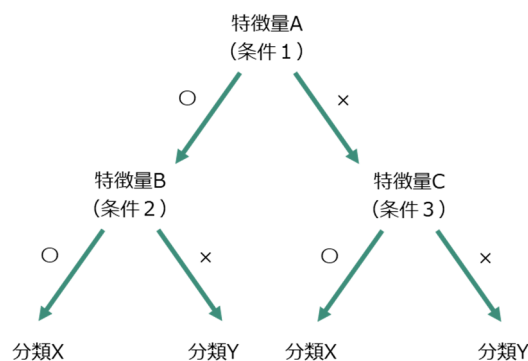
本章では、ROC 曲線及び AUC を用いて、債務者区分の遷移パターンごとに財務情報で構築される予測モデルの精度を比較した。その結果、債務者区分の遷移パターンごとに財務情報のみで予測できる精度が異なることが確認された。特に、債務者区分のランクアップを予測対象とする際には、予測モデル作成の際に財務情報以外の情報を勘案する必要性が示された。なお、本章の検証は、一つの学習モデルによる比較に留まっているため、複数の学習モデルを組み合わせることによって精度が向上する可能性がある点（ランダムフォレストの結果を踏まえた XGBoost の使用等）や学習に用いたデータの期間が限られていることによるサンプリングバイアスの可能性に留意する必要がある。

BOX: 各学習モデルの概要

本 BOX では、本稿で使用した各学習モデルについて、簡単に概要を紹介する。なお、本 BOX での説明は各学習モデルの概要を簡略化したものであり、正確な詳細は注釈で示す各文献等を参照のこと。

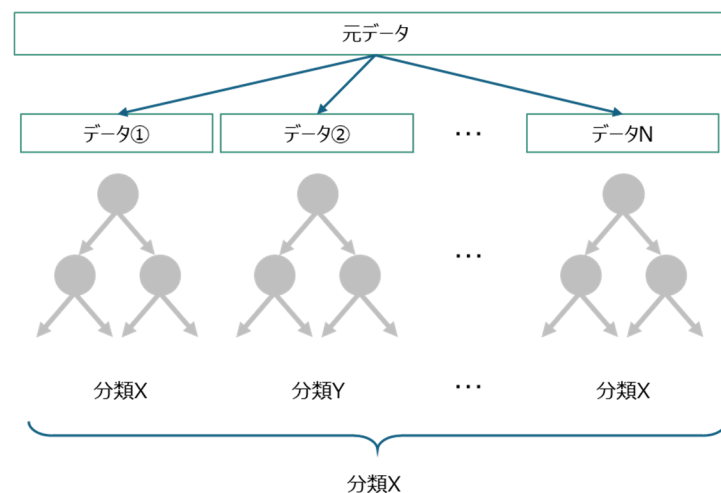
はじめに、ランダムフォレスト及び XGBoost についてである。これらは、決定木と呼ばれるアルゴリズムを基本とする学習モデルである。決定木は、図表 18 のような樹形図を作成し、予測や検証を行う手法のことを指す。

図表 18 決定木



ランダムフォレスト²⁶とは、図表 19 のとおり、アンサンブル学習²⁷のバギング²⁸をベースに、決定木を複数集めたものである。具体的には、元データからランダムにデータをブートストラップサンプリングし、任意の数のデータグループを作成する。次に、各データグループで決定木を作成し、それぞれ予測を行う。最後に、各決定木の予測結果から多数決をとり、最終予測を行う方法である。

図表 19 ランダムフォレスト



XGBoost²⁹とは、図表 20 のとおり、決定木の勾配ブースティングを実装したものである。具体的には、初めに決定木を作成し、その予測モデルの評価を行う。その評価に基づき改良した決定木を再作成し、改めて予測モデルの評価を行う（勾配に基づく探索）。この手順を繰り返し予測モデルを直列に繋げていくことで、高精度な予測モデルを目指すものである（「ブースティング」）。

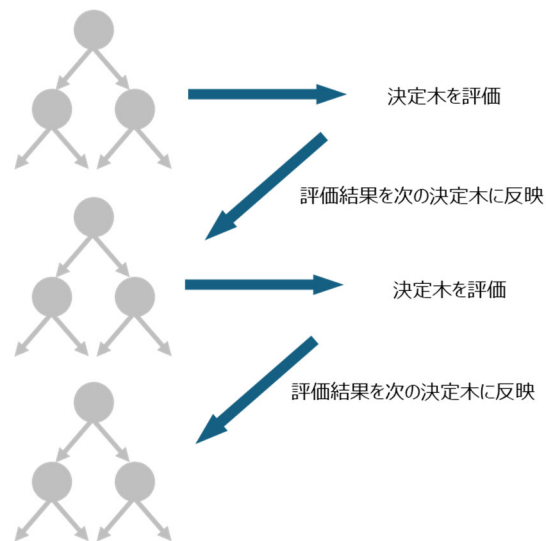
²⁶ L.Breiman: "Random Forests", Machine Learning, 45, 1, p.5-32(2001)

²⁷ 単体だと予測精度が低い機械学習モデルを複数組み合わせ、高精度な予測モデルを作る手法のこと。ランダムフォレストで用いる「バギング」や後述の XGBoost で用いる「ブースティング」もアンサンブル学習の一種である。

²⁸ 母集団から、各決定木で使うデータをランダムに復元抽出するブートストラップサンプリングによって得たデータを用いて、複数の予測モデルを作成し、多数決をとる方法。

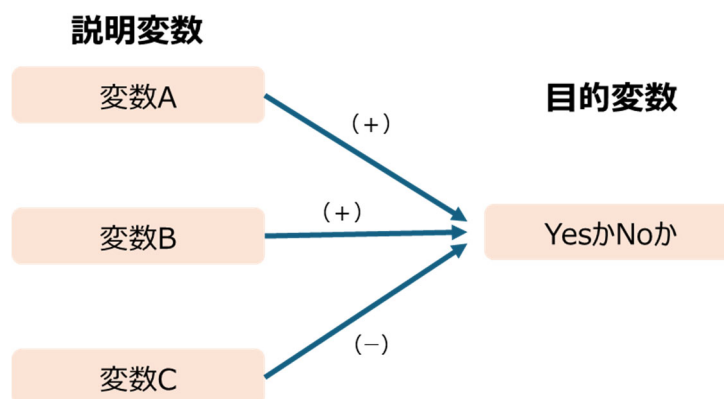
²⁹ T.CHEN and C.GUESTRIIN: "XGboost: A scalable tree boosting system"; Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, p. 785-794(2016)

図表 20 XGBoost



続いて、ロジスティック回帰とは、図表 21 のとおり、複数の要因（説明変数）によって、正誤等の二値（目的変数）を予測する手法である。

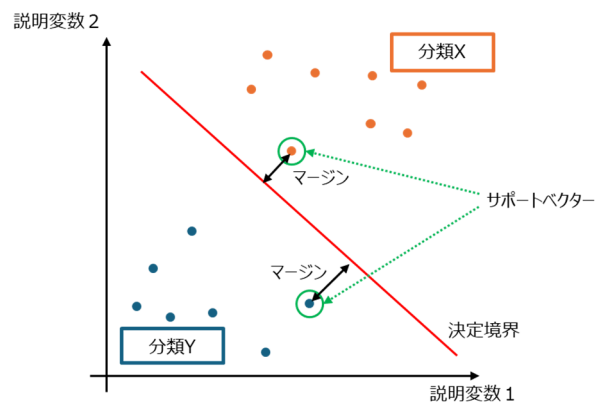
図表 21 ロジスティック回帰



次に、SVM（サポートベクターマシン）³⁰とは、図表 21 のとおり、決定境界までの距離であるマージンを最大化するように学習するアルゴリズムのことである。決定境界とは、分類の境目になる直線や曲線のことである。この決定境界に最も近い対象をサポートベクターといい、サポートベクターと決定強化までの距離であるマージンが最大化するようにモデルを構築する。

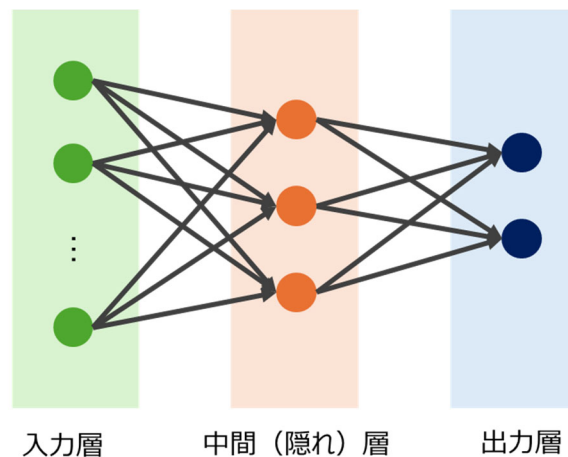
³⁰ Vapnik, V.N.: "Statistical Learning Theory", Wiley (1998)

図表 21 SVM



最後に、MLP（多層パーセプトロン）³¹とは、図表 22 のとおり、ニューラルネットワークの構造を持つ、神経細胞（ニューロン）のネットワークを模倣した数理モデルの一つである。多層パーセプトロンでは最低でも三つの層（入力層、中間（隠れ）層、出力層）から構成され、各層のノードは前の層の全てのノードを接続されており、バックプロパゲーション³²と呼ばれるアルゴリズムを用いて学習を行うものである。具体的には、入力層で外部からの入力データを受け取り、中間層に渡す。中間層にて入力層からの情報を変換して、特徴量を抽出する。出力層にて中間層からの情報を元に最終的な出力を生成するものである。

図表 22 MLP



³¹ 単層パーセプトロンの提唱 F.Rosenblatt: "The perceptron: A preprobabilistic model for information storage and organization in brain"; Psychological Review, Vol.65, No.6, p.386 (1958)。バックプロパゲーションを提唱 D.E.Rumelhart, G.E.Hinton and R.J.Wikkiams: "Learning representations by back-propagating errors"; Nature, Vol.323, No.6088, p.533 (1986)

³² 入力データとパラメータを掛け合わせて予測値を出力し、正解ラベルとの誤差を計算する。その誤差を出力層から入力層に逆に計算し、各層の重みを更新する方法。

V. 総括

本稿では、金融機関の信用リスク管理態勢に係る実態把握を定量データにより深める観点から、保全状況に関する分析及び債務者区分の遷移予測モデルの検証に取り組んだ。前者においては、共通貸出先や越境貸出先は、単独貸出先や圏内貸出先と同様な財務状況にあったとしても、保全率が低い傾向があることを確認した。後者においては、債務者区分が破綻懸念先以下にランクダウンする際は、他の遷移パターンと比較して財務情報のみで精度の高い予測が可能であるものの、特にランクアップする際は、財務情報以外の定性情報等を予測モデルに考慮する必要性が示唆された。

金融庁では、引き続き、貸出明細等の高粒度データ活用を含む多様なデータ分析に関する実績を積み重ね、金融機関との対話への活用やモニタリングの高度化に向けた取り組みを促進していく。