

株価の局面変化に対する要因抽出

EXPLORATION OF FACTORS AFFECTING STOCK REGIME CHANGE

瀬之口 潤輔

Junsuke Senoguchi

¹三菱 UFJ モルガン・スタンレー証券

¹Mitubishi UFJ Morgan Stanley

Abstract: Financial crises are typically caused by a chain of credit contractions, which in turns could be caused by the rapid worsening of indexes that indicate people's psychology, such as bank stock prices. The purpose of this analysis is to identify trigger points where bank stocks rise or fall by extracting what common points existed in financial economic indicators immediately before significant fluctuations of bank stocks occurred in the past. To conduct discriminant analysis, we used the traditional statistical method as well as ensemble learning. We also used “bank stock performance” as well as “bank stock regime change” as objective variables. This attempt showed that the money multiplier and 10-year yield of government bonds are important ones that could have an influence on bank stock regime change.

Keywords : Ensemble Learning, Bank Stock, Regime, J48, Random Forest.

1. 序論

1990年代初頭に我が国で発生したバブル崩壊や2008年に全世界的な影響を与えたリーマンショックなどの金融危機は、株式や不動産など資産価格の下落による資本市場の混乱にとどまらず、連鎖的な信用収縮による実体経済の急激な悪化を引き起こした。金融危機は人々の生活に多大な影響を及ぼしたため、その発生するメカニズムの解明や予知に関する研究は重要であり、先行研究も多い。

代表的な先行研究の一つである Kaminsky (1998) は、連鎖的な信用収縮は、株価のように多くの人々の心理を表す指標が急速に悪化することがきっかけになることを示した。我が国を含む多くの国で信用創造の重要な役割を果たしているのは銀行であるため、銀行株の動き、特にレジームの急激な悪化などは、金融危機の重要なきっかけの一つと考えられる。銀行株のレジームが急激に悪化する前の兆候を特定することができれば、金融危機を予知・回避する対策も講じられる。よって本研究では、過去に銀行株のレジームが急激に変化したとき、その直前にどの金融経済指標にどのような共通点があったかを抽出することにより、銀行株のレジームが変化するトリガーを特定することを目的とした。

2. 先行研究

株価には、世間にあるあらゆる情報が織り込まれ

ているため恒常的な利益を上げることができないという、いわゆる効率的市場仮説の Strong 仮説が存在する一方で、株価の予想は可能とする研究も存在する。株価予想に関する研究は、配当割引モデルから派生するもの、経済ファンダメンタルズ分析から派生するもの、時系列モデルから派生するものなどに大別できる。また新たな研究分野として、ノンパラメトリック手法を用いた株価予想がある。以下にそれぞれの概要と課題をまとめた。

2.1. 配当割引モデルによる株価予想

1980年初めまでの株価予想に関する研究は、Shiller (1979)に代表されるよう、株価は将来の配当を現在価値に割り引いた合計で算出されるという、いわゆる配当割引モデルの考え方が一般的であった。しかし Leroy and Porter (1981)は、株価は変動が大きく、配当割引モデルでは説明できないことを示した。

これに対して、一定割引の前提を緩和することにより、Cambell and Schiller (1988)は、株価の動きを、将来の配当と将来の株価リターンに分解した。この結果、株価変動の多くの部分が将来株価リターンの期待の修正によるものであり、将来の配当成長ではないことが示された。

一方で配当割引モデルの場合、将来の期待配当流列と割引率を求める必要があるが、これらの値を正確に推定することは困難である。

2.2. 経済指標による株価予想

企業財務ではなく、経済ファンダメンタルズを株価変動の要因として、株価予想を行った研究も多い。代表的なものは Chen et al.,(1986)により、裁定価格理論を用いて、アメリカの経済ファンダメンタルズと株価の関係を示したものであろう。金利、インフレ率、為替レート、国債利回り、鉱工業生産が、株価に重要な影響を与えていることを示した。

また Bulmash and Trivoli (1991)は、マネーサプライ、政府債務残高、失業率、短期金利などは株価に正の相関があるが、短期政府証券の利回りは負の相関があることを示した。Abdullah and Hayworth (1983)は、マネーの増加率やインフレ率は株価に正の相関があるが、市場金利は負の相関があることを示した。

一方でこれらの研究では、マクロ経済指標と株価の強い相関を示しているが、株価の予想精度に関しては概して低い。

2.3. 時系列モデルによる株価予想

VAR など時系列モデルによる我が国の株価予想に関しては、植田 (2007) などが研究を行っている。また観測できない状態変数を用いた状態空間モデルを利用して株価予想を行った研究では、大鋸 (2009) などが代表的である。これらの研究では、学習データに基づく株価予想の精度は高いものの、テストデータを用いると精度は著しく低下する。モデルによる株価予想に関しては、テストデータで高い精度を示すことが最大の課題である一方で、これらの研究ではこの課題に取り組んでいない。

2.4. ノンパラメトリック法による株価予想

株価リターンを、外生変数により予想する場合、外生変数がある一定の値を超えると株価リターンが大きく変動することが考えられる。このような場合、分布を仮定する分析方法では予想精度は落ちると考えられる。このため、ノンパラメトリック手法により株価予想を行う研究も行われている。

左 (2011) は、株価収益率をクラスタリング手法により離散値に変換し、ベイジアンネットワークを用いて、過去の株価収益率から将来の株価を予想した。この結果、正規分布を前提とする一般的な統計手法や時系列分析に比較して、ノンパラメトリック手法の方が高い精度で株価を予測できることを示した。

また Ma (2013)は、ニューラルネットワークを用いて株価の予想を行った。株価予測のような複雑な関係を扱うのにニューラルネットワークは適している

ことを指摘した。また淵井 (2011) は SVM で株価予想を行い、テストデータを用いた場合でも株価予想に関して高い精度を示した。

これらの研究では、テストデータを用いた場合でも株価予想が当たっているが、研究で用いられているテスト期間以外のデータを用いて株価予測を行うと、ほとんど当たらない。またモデルの作成過程は、多くの隠されたノードや決定木に依存しているため、外生変数の寄与度や貢献度が不明である。

2.5. 先行研究のまとめ

以上をまとめると、配当割引モデルや経済ファンダメンタルズモデルでは、過去の株価リターンの動きを、どのような外生変数がどのように寄与していたかといった点から説明することはできるが、株価予想はできない。

ノンパラメトリック手法は、配当割引モデルや経済ファンダメンタルズモデルに比較すると、株価の予測精度は高いと考えられるが、計算過程が隠されているため、株価予想の仕組みを解釈することが困難である。

このような課題が存在することを認識したうえで、本研究では、主にノンパラメトリック手法を用いて、多くの変数の中から株価の先行指標として重要なものを定量的に抽出し、さらにこれまではブラックボックスとなっていた外生変数と目的変数の関係を可視化する試みを行った。

3. 分析に用いるデータと分析手法

本研究では、過去に銀行株が上がる（下がる）直前に、どの金融経済指標にどのような共通の要因があったかを抽出し、それらの要因を用いて将来の銀行株のパフォーマンスを予測する。

必要なデータとしては、銀行株の先行指標になり得る多くの金融経済指標（説明変数）と、銀行株のパフォーマンス（目的変数）である。

3.1. 説明変数

説明変数として、日本銀行の調査などで銀行の業績や長期金利の要因とされた指標、または内閣府で景気に先行性および同時性があるとされた指標など、合計 94 種類の金融経済指標を作成した。これらの指標は、長期金利や為替レートといった市場関連指標、日本銀行ワーキングペーパー「金融活動指標の解説」における「候補となった金融関連指標」、同「金融マクロ計量モデルの概要」における「金融セクター変数」、同「長期金利の変動要因」における「説明変数」

などを参考にした。また内閣府が作成した「景気動向指数」を構成する指標のうち、「先行系列」と「一致系列」に属するものも用いた。さらには貸出金や

預金など銀行の財務データに加えて、ROE や自己資本比率など法人企業の財務データ、マネーサプライや GDP などのマクロデータも用いた（表 1）。

表 1：銀行株の先行指標の候補

v1	LIBOR-OIS(日本)	月末	v48	M2/名目GDP	月末
v2	LIBOR-OIS(アメリカ)	月末	v49	M2/ベースマネー	月末
v3	JGB10年利回り	月末	v50	M2/外貨準備	前月差
v4	LIBOR-OIS(日本)	前月差	v51	M2/名目GDP	前月差
v5	LIBOR-OIS(アメリカ)	前月差	v52	M2/ベースマネー	前月差
v6	JGB10年利回り	前月差	v53	CPI(生鮮食品を除く総合)	前月比
v7	貸出金増加率	前月比	v54	CPI(食料及びエネルギーを除く)	前月比
v8	預金増加率	前月比	v55	名目実効為替レート	前月比
v9	預貸率	月末	v56	実質実効為替レート	前月比
v10	(現金+準備預金)/総資産	月末	v57	機械受注	前月比
v11	市場調達資金/総資産	月末	v58	工作機械受注	前月比
v12	流動資産比率	月末	v59	貿易収支	前月比
v13	ROA	月中	v60	貿易収支(輸出)	前月比
v14	貸出金利回り	月中	v61	名目GDP成長率	前期比
v15	預金利回り	月中	v62	実質GDP成長率	前期比
v16	実質貸出金利回り	月中	v63	倒産件数	前月比
v17	実質預金利回り	月中	v64	負債総額	前月比
v18	預貸率	前月差	v65	日米ベースマネー比率	月末
v19	(現金+準備預金)/総資産	前月差	v66	日米マネーサプライ比率	月末
v20	市場調達資金/総資産	前月差	v67	日米ベースマネー比率	前月差
v21	流動資産比率	前月差	v68	日米マネーサプライ比率	前月差
v22	ROA	前月差	v69	最終需要財在庫率指数(逆サイクル)	前月比
v23	貸出金利回り	前月差	v70	鉱工業生産財在庫率指数(逆サイクル)	前月比
v24	預金利回り	前月差	v71	新規求人数(除学卒)	前月比
v25	実質貸出金利回り	前月差	v72	実質機械受注(船舶・電力を除く民需)	前月比
v26	実質預金利回り	前月差	v73	新設住宅着工床面積	前月比
v27	資金繰り	四半期末	v74	消費者態度指数	前月差
v28	貸出態度	四半期末	v75	日経商品指数(42種)	前月比
v29	借入金利水準	四半期末	v76	長短金利差	前月差
v30	インタレスト・カバレッジ・レシオ	四半期末	v77	長期国債(10年)新発債流通利回	前月差
v31	借入返済年数	四半期末	v78	TIBOR(3か月)	前月差
v32	自己資本比率	四半期末	v79	東証株価指数	前月比
v33	当座比率	四半期末	v80	投資環境指数(製造業)	前月差
v34	固定比率	四半期末	v81	総資本営業利益率(製造業)	前月差
v35	ROE(経常利益/純資産)	四半期末	v82	長期国債(10年、流通利回り)	前月差
v36	資金繰り	前期差	v83	中小企業売上げ見通しD.I.	前月差
v37	貸出態度	前期差	v84	生産指数(鉱工業)	前月比
v38	借入金利水準	前期差	v85	鉱工業生産財財荷指数	前月比
v39	インタレスト・カバレッジ・レシオ	前期差	v86	大口電力使用量	前月比
v40	借入返済年数	前期差	v87	耐久消費財財荷指数	前月比
v41	自己資本比率	前期差	v88	所定外労働時間指数(調査産業計)	前月比
v42	当座比率	前期差	v89	投資財出荷指数(除輸送機械)	前月比
v43	固定比率	前期差	v90	商業販売額(小売業)	前年同期比
v44	ROE(経常利益/純資産)	前期差	v91	商業販売額(卸売業)	前年同期比
v45	M2増加率	前月比	v92	営業利益(全産業)	前月比
v46	ベースマネー増加率	前月比	v93	中小企業出荷指数(製造業)	前月比
v47	M2/外貨準備	月末	v94	有効求人倍率(除学卒)	前月差

3.2 目的変数

本研究の目的変数は銀行株のパフォーマンスであるが、これについて以下 2 つの変数を作成した。

3.2.1. 銀行株相対パフォーマンス

第一の目的変数は、毎月 11 日を開始日として、180 日間の銀行株の TOPIX 対比変化率が正の値であれば「上昇群」、負の値であれば「下落群」とした（銀行株相対パフォーマンス）。

開始日を毎月 11 日としたのは、本研究で使用する金融経済指標の多くは、毎月初旬までに発表されるからである。

銀行株相対パフォーマンスの測定期間を 180 日としたのは、短期的な変動の影響を回避し、長期レジームが明確になる期間である必要があること、また実務上、機関投資家の運用成績が半年間のパフォーマンスとなっていることが多いことなどである。

目的変数を、株価の上昇率や下落率といった連続変数ではなく、上述のような離散変数としたのは、目的変数を 2 値の離散変数にすると、SVM やニュー

ラルネットワークのような多くの判別分析手法を用いることができるからである。

銀行株の先行指標の候補を 1991 年 4 月から 2012 年 3 月まで作成したため、銀行株相対パフォーマンスは 1991 年 5 月から 2012 年 4 月までを算出した。合計 252 月（12 ヶ月×21 年）のうち、「上昇群」が 87 月（全体の 34.5%）となり、また「下落群」が 165 月（全体の 65.5%）となった（図 1）。

3.2.2. 銀行株のレジーム

第二の目的変数として、銀行株相対パフォーマンスを隠れマルコフモデルにより 2 値の離散変数に変換したもの（銀行株のレジーム）を作成した。銀行株のレジームが上昇傾向であれば「上昇群」、下落傾向であれば「下落群」とした。

目的変数として、銀行株相対パフォーマンスのほ

かに銀行株のレジームを用いたのは、銀行株相対パフォーマンスの背後には、時間経過と共にマルコフ過程で遷移する 2 つの定常状態（銀行株のレジーム）が存在し、銀行株相対パフォーマンスは各レジームにおける確率分布に従って観測され、また金融経済指標が影響を与えているのは株価パフォーマンスそのものではなく、その背後にあるレジームであるという仮説を立てたからである。

隠れマルコフモデルでは、「上昇群」や「下落群」は観測できない潜在変数とされ、その確率が算出されるが、本研究では確率が 0.5 よりも高い場合はその確率が示す群に分類し、「上昇群」と「下落群」の 2 値の離散地に変換した。

銀行株のレジームは、「上昇群」が 102 月（全体の 40.5%）となり、また「下落群」が 150 月（全体の 59.5%）となった（図 1）。

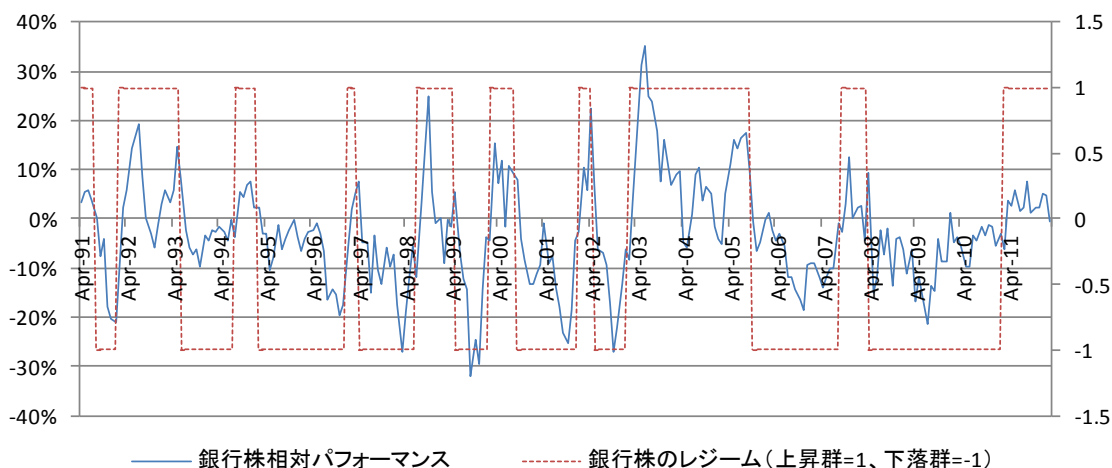


図 1：銀行株相対パフォーマンスと銀行株のレジーム

3.3 分析手法

このようにして作成した説明変数と目的変数を用いて、金融経済指標のうち上昇群、下落群に対応するサンプル集合の間で大きく水準が異なるものを抽出した。そのような指標は、それぞれの群を分類する強い先行指標になり得るからである。

抽出には、ロジスティック回帰分析、SVM、J48、ランダムフォレストの 4 つの手法を用いた。ソフトウェアとして WEKA3 を使用し、それぞれの手法は WEKA3 に内蔵されている分類器 Logistic、SMO、J48、RandomForest を用いた。

それぞれの手法で、学習データによる誤判別率と、10-fold 交差検証法による誤判別率を求め、手法ごとに比較した。

またある期間までの学習データによりモデルを作成し、それ以降の説明変数をモデルに外挿したときの誤判別率も測定し、モデルの予測精度を比較した。

4. 判別精度の比較

4.1 銀行株相対パフォーマンスの分析

銀行株相対パフォーマンスを目的変数として、前述の 4 つの手法で誤判別率を比較したところ、学習データを用いた場合でも、10-fold 交差検証法を用いた場合でも、ランダムフォレストの精度が最も高かった（表 2）。

まずロジスティック回帰分析では、学習データを用いた場合に誤判別率が 6.7% となり、判別精度が比

較的高いことが示された。一方で 10-fold 交差検証法を行うと、誤判別率は 36.9%まで悪化した。ロジスティック回帰分析は、目的変数の判別にすべての説明変数が寄与するため、過学習を起ししやすい特徴がある。本研究の結果には、この特徴が反映されたものと考えられる。ロジスティック回帰分析による判別分析により、誤差分散を正規分布と仮定するパラメトリック法では、銀行株の増減を予想することが難しいことが示された。

SVM では、学習データを用いた場合に誤判別率が 23.0%となり、4 手法のうち最も悪い結果となった。一方で 10-fold 交差検証法を行うと、SVM の誤判別率は 32.5%となり、学習データを使ったときに比較した判別精度の悪化は 4 手法の中で最も小さかった。しかしいずれの判別精度も、4 手法の比較の中で比較的低かった。SVM は、分離超平面に隣接したサンプルからのマージンを最大化するように求められる。このため汎化能力が高い一方で、すべてのサンプルを用いるわけでは無いため、学習データによる判別精度は劣ることがある。本研究の結果は、このような SVM の特徴を示したものと考えられる。SVM による高次元への写像では、銀行株の増減を分離することが難しいことが示された。

J48 では、学習データを用いた場合に誤判別率が 11.9%となり、判別精度が比較的高いことが示された。また 10-fold 交差検証法を行うと、誤判別率は 24.6%となり、汎用性が高いことが示された。一般に J48 は、最終ノードに含まれるサンプル数の下限を小さくすると、学習データの判別精度は高まる一方で、木のサイズが大きくなるため、汎用性は劣化する。本研究の目的は、銀行株の変動要因を特定することであるため、分類木のノード数は小さいことが望ましい。よって 10-fold 交差検証法による誤判別率が大きく悪化しない範囲で、最終ノードのサンプル数の下限を引き上げた。まず最終ノードのサンプル数を 2 とした場合、分岐の数は 19 となり、10-fold 交差検証法による誤判別率は 23.0%であった。これを 10 とした場合、分岐の数は 10 に減少した一方で、誤判別率も 24.6%となり、汎用性の劣化は限定的であった。

またランダムフォレストでは、学習データを用いた場合に誤判別率が 0.8%となり、10-fold 交差検証法

でも 23.4%と、いずれの判別精度も、4 手法のうち最も高かった。J48 による 10-fold 交差検証法でも判別精度が高かったことから、銀行株の増減には、複数の要因の非線形な関係が影響を与えていると考えられる。

表 2 に、それぞれの分析手法ごとの、誤判別率（誤った判別をした予測件数の、全予測件数に対する割合）、適合率（「下落群」と予測した件数のうち、実際に正しかった件数の割合）、再現率（実際の「下落群」の件数のうち、正しく予測された件数の割合）、F 値（適合率と再現率の総合的な評価）を示した。

4.2 銀行株のレジームに関する分析

銀行株のレジームを目的変数として、前述の 4 つの手法で誤判別率を比較したところ、学習データを用いた場合はロジスティック回帰分析とランダムフォレストが最も精度が高く、また 10-fold 交差検証法を用いた場合は、J48 の精度が最も高かった（表 3）。

10-fold 交差検証法を用いた SVM による分析を除いて、すべての場合において、銀行株のレジームを目的変数とするほうが、銀行株相対パフォーマンスを目的変数とするより、誤判別率が低かった。このことから、金融経済指標が影響を与えているのは株価パフォーマンスそのものではなく、その背後にあるレジームであると推察できる。

また J48 による分析では、銀行株のレジームを目的変数とするほうが、銀行株相対パフォーマンスを目的変数とするより、訓練データおよび 10-fold 交差検証法いずれの場合でも誤判別率が大きく改善した。10-fold 交差検証法を用いた場合は、ランダムフォレストよりも誤判別率が低かった。

これは銀行株のレジームを目的変数とした場合、「上昇群」と「下落群」を分類するために、いくつかの少数の金融経済指標が特に重要な役割を示しており、ランダムフォレストのようにランダムに説明変数を選択して分類木を作成する場合、これらの重要な金融経済指標が選択されないことがあるため、これらが選択される J48 に比較すると精度が低下している可能性がある。

表 2：分析手法ごとの判別精度の比較（目的変数：銀行株相対パフォーマンス）

		誤判別率	適合率	再現率	F値
ロジスティック回帰分析	訓練データ	6.7%	93.0%	97.0%	95.0%
	10-fold CV	36.9%	74.3%	66.7%	70.3%
SVM	訓練データ	23.0%	77.7%	90.9%	83.8%
	10-fold CV	32.5%	72.4%	81.2%	76.6%
J48	訓練データ	11.9%	89.9%	92.1%	91.0%
	10-fold CV	24.6%	79.4%	84.2%	81.8%
ランダムフォレスト	訓練データ	0.8%	100.0%	98.8%	99.4%
	10-fold CV	23.4%	81.5%	83.0%	82.3%

表3：分析手法ごとの判別精度の比較（目的変数：銀行株のレジーム）

		誤判別率	適合率	再現率	F値
ロジスティック回帰分析	訓練データ	0.0%	100.0%	100.0%	100.0%
	10-fold CV	27.8%	65.1%	67.6%	66.3%
SVM	訓練データ	22.6%	79.2%	59.8%	68.2%
	10-fold CV	32.9%	62.7%	46.1%	53.1%
J48	訓練データ	1.2%	99.0%	98.0%	98.5%
	10-fold CV	12.7%	86.5%	81.4%	83.8%
ランダムフォレスト	訓練データ	0.0%	100.0%	100.0%	100.0%
	10-fold CV	14.7%	82.2%	81.4%	81.8%

5. 銀行株の変動要因の抽出

5.1 説明変数の絞り込み

前章で示されたように、目的変数として銀行株のレジームを用いた場合、特定の重要な金融経済指標が選択されずに分析が行われたため、ランダムフォレストによる判別精度が J48 によるものより低くなった可能性を指摘した。重要な金融経済指標をまず選択し、それにより分析を行うことにより、この問題は回避されると考えられる。またそもそも本研究で用いた 94 種類もの説明変数による推定モデルは、実用や汎用性の観点でも難点がある。よってこの章では、94 種類の説明変数の中から特に重要なものを絞り込んだうえで、ランダムフォレストと J48 による分析で判別精度の比較を行った。

ランダムフォレストには、ある説明変数が使われなかった場合に、予測精度がどれだけ低下するかを測定し、変数の重要度を評価する手法がある。これを用いて、銀行株のレジームに重要な変数を抽出した。

これによると、銀行株の変動に影響を与える金融経済指標として、M2/ベースマネー (v49)、日米ベースマネー比率 (v65)、国債 10 年利回り (v3)、銀行の現預金 (v10)、法人企業借入返済年数 (v31) が

特に重要な説明変数であることが示された。またそのほかに M2/名目 GDP (v48)、日米マネーサプライ比率 (v66)、預金利回り (v15) が抽出された。8 変数のうち 5 変数がマネー関連となった。

また表 4 に示したとおり、これらの変数を使用して再びランダムフォレストによる判別分析を行ったところ、10-fold 交差検証法を行った場合に判別精度が改善した。これはランダムフォレストでは数多くの変数をランダムに含んだ分類木を生成するため、すべての変数を使用した場合は、必要のない変数が分析に含まれたことが要因と考えられる。

5.2 テストデータの外挿

選択された 8 変数を用いて、ある期間までのデータによりモデルを作成し、それ以降の期間の説明変数をモデルに外挿して得られた結果の誤判別率を図 2 に示した。J48 ではテストデータの外挿により判別精度は大きく低下した。学習データで高い予測精度が示されても、それを予想に使うと精度が大きく落ちるとするのは、これまで資産価格のリターンを予想する多くの研究で見られた結果と同様である。しかしランダムフォレストを使うと、比較的高い精度で将来の株価の上昇または下落を予想できることが示された。

表 4：ランダムフォレストと J48 による判別制度の比較

		誤判別率	適合率	再現率	F値
ランダムフォレスト(全94変数)	訓練データ	0.0%	100.0%	100.0%	100.0%
	10-fold CV	14.7%	82.2%	81.4%	81.8%
ランダムフォレスト(選択された8変数)	訓練データ	0.4%	100.0%	99.3%	99.7%
	10-fold CV	13.5%	89.7%	87.3%	88.5%
J48(全94変数)	訓練データ	1.2%	99.0%	98.0%	98.5%
	10-fold CV	12.7%	86.5%	81.4%	83.8%
J48(選択された8変数)	訓練データ	5.2%	97.9%	93.3%	95.6%
	10-fold CV	13.9%	87.6%	89.3%	88.4%

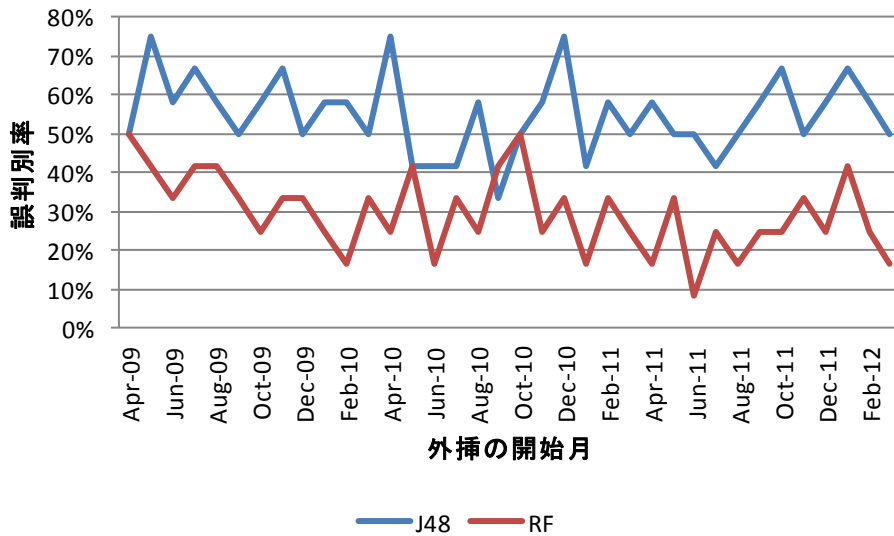


図 2：学習期間後 12 か月の外挿による誤判別率の推移

6. 結論

金融危機や連鎖的な信用収縮は、銀行株のように多くの人々の心理を表す指標が急速に悪化することがきっかけで発生することが多い。銀行株のレジームが急激に悪化する前の兆候を特定することができれば、金融危機を予知・回避する対策も講じられる。よって本研究では、過去に銀行株のレジームが急激に変化したとき、その直前にどの金融経済指標にどのような共通点があったかを抽出することにより、銀行株のレジームが変化するトリガーを特定することを目的とした。

本研究では、説明変数として合計 94 種類の金融経済指標を作成し、また目的変数として「銀行株相対パフォーマンス」および「銀行株のレジーム」を作成し、判別分析および要因抽出を行った。判別分析には、ロジスティック回帰分析、SVM、J48、ランダムフォレストの 4 つの手法を用い、最も適した判別手法を特定した。この結果、ランダムフォレストが最も判別精度が高く、使用する手法として最も適していることが示された。

「銀行株相対パフォーマンス」を目的変数として判別分析を行うより、「銀行株のレジーム」を目的変数として判別分析を行うほうが、10-fold 交差検証法を用いた SVM による分析を除いて、すべての場合において高い判別精度を示した。金融経済指標が影響を与えているのは株価パフォーマンスそのものではなく、その背後にあるレジームであると推察できる。

また「銀行株のレジーム」を目的変数として 10-fold 交差検証法を行った場合は、ランダムフォレストよりも J48 の方が高い判別精度を示した。これは「上昇群」と「下落群」を分類する際に、いくつかの少数の金融経済指標が特に重要な役割を示しており、ランダムフォレストのようにランダムに説明変数を選択して分類木を作成する場合、これらの重要な金融経済指標が選択されないことがあるため、これらが選択される J48 に比較すると精度が低下している可能性がある。

この問題は、まず重要な金融経済指標を選択し、それにより分析を行うことにより、回避されると考えられる。そもそも本研究で用いた 94 種類もの説明

変数による推定モデルは、実用や汎用性の観点でも難点がある。よって 94 種類の説明変数の中から特に重要なものをランダムフォレストで抽出したうえで、ランダムフォレストと J48 による分析で判別精度の比較を行った。結果として、判別精度は逆転し、ランダムフォレストの方が J48 より高い判別精度を示した。

また本研究では、ノンパラメトリック手法に特有の問題である計算過程のブラックボックス化に対処するために、ランダムフォレストで抽出された説明変数を用いて J48 により分類木を作成し、説明変数がどのように「銀行株のレジーム」に影響を与えているかを可視化する試みも行った。

この結果、銀行株の変動に影響を与える金融経済指標として、M2/ベースマネー (v49)、国債 10 年利回り (v3)、銀行の現預金 (v10) 法人企業借入返済年数 (v31) などが重要であることが示された。特に、企業の過剰債務の水準が高いことは、銀行株にとって一般的に下落要因となるが、ベースマネーの供給が拡大し貨幣乗数が低下しているときは、企業の過剰債務の水準が高いことは、今後の不良債権処理が加速する期待となり、銀行株にとって上昇要因になることが示された。

また、ある期間までの学習データでモデルを作り、それ以降の説明変数をモデルに外挿すると、ランダムフォレストでは誤判別率が比較的低いことが示された。

今後の課題は、目的変数は「上昇群」と「下落群」の 2 値の離散値だけでなく、それぞれの状態の確率を用いて連続変数とし、分類モデルではなく回帰モデルにすることにより、上昇および下落の程度まで予想することであろう。また月次データではなく、日次データなどを用いてデータ数を多くし、統計的に頑健性を高めることも、重要な課題であろう。

参考文献

Kaminsky, G. L. and Reinhart, C. M. “Financial Crises in Asia and Latin America: Then and Now”, *The American Economic Review*, Vol. 88, No. 2, 1998. pp. 444-448.

Shiller, R. “The Volatility of Long-Term Interest Rates and Expectations Model of the Term Structure”, *Journal of Political Economy*, Vol. 87, 1979. pp. 1190-1219.

LeRoy, S. F. and Porter, R.D “The Present-Value Relation: Tests Based on Implied Variance Bounds”, *Econometrica*, Vol. 49, No. 3, 1981. pp. 555-574.

Campbell, J. Y. and Shiller, R. “Stock Prices, Earnings, and Expected Dividends”, *Journal of Finance*, Vol. 43,

No. 3, 1988. pp. 661-676.

Chen, N. F., Roll, R. and Ross, S. A. “Economic Forces and Stock Market”, *Journal of Business*, Vol. 59, No. 3, 1986. pp. 383-404.

Bulmash, S. B. and Trivoli, G.W. “Time-lagged Interactions between Stock Prices and Selected Economic Variables”, *Journal of Portfolio Management*, Vol. 17, No. 4, 1991. pp. 61-67.

Abdullah, D. A. and Hayworth, S. C. “Macroeconomics of Stock Price Fluctuations”, *Journal of Business and Economics*, Vol. 32, No. 1, 1983. pp. 49-63.

Ma, N., Zhai, Y., Li, W.F., Wang, S.S. and Hjou, L. *Journal of Applied Sciences*, Vol. 13, No. 22, 2013. pp.5384-5390.

植田英三郎「時系列解析による株価予測」『大阪府立大学経済研究』第 53 巻、3 号、2007 年 221 ページ
大鋸崇・大屋幸輔(2009)「株式市場におけるブル相場、ベア相場の日次データを用いた分析—ベイジアンアプローチ」『ジャフィー・ジャーナル (金融工学と市場計量分析)』, pp.112-150.

左毅、原田昌郎、北栄輔「ベイジアンネットワークを用いた株価予測の精度改善」『情報処理学会論文誌:数理モデル化と応用』, 4 巻、4 号、2011 年、92-103 ページ

渕井 亮、鍾 寧、電子情報通信学会技術研究報告. KBSE, 知能ソフトウェア工学 110(386), 43-48, 2011-01-17