

変化点検出と日本株式市場の季節性アノマリー

Burst Detection in time series data of Nikkei 225 and the Calendar Anomaly

山崎高弘^{1*} 岡田克彦²
Takahiro YAMASAKI¹ Katsuhiko OKADA²

¹ 大阪産業大学工学部

¹ Faculty of Engineering, Osaka Sangyo University

² 関西学院大学大学院経営戦略研究科

² Institute of Business and Accounting, Kwansei Gakuin University

Abstract: We detect turning points of the non-stationary time series data of Nikkei 225 index for the period between 1993 and 2010 using 'change finder'. We also calculate the market sentiment using news data prior to the turning point. Our findings are in two-fold. Firstly, the 'change finder' signals the bullish turning points following the rise of the optimistic sentiment and vice versa. Secondly, bullish change occurs significantly more in the first half of the year than the latter half. Our findings are consistent with the view that the reported 'Dekansho-bushi' effect in the Japanese Stock Market is driven by the market psychology.

1 はじめに

株式市場のアノマリーの中でも、古くから指摘されており、その存在が実務界に広く認識されているのが株式市場の季節性である。欧州の金融関係紙はたいてい5月になると“Sell in May”の格言について語る。即ち、保有株は5月中に売ってしまえというのだ。この相場の格言には続きがある。“...but buy back in St. Legers day”と続く。St. Legers dayとは英国で毎年秋口に行われる競馬レースのことである。つまり5月から秋口までの期間は株式市場のパフォーマンスがよくないため投資を避けておいて、秋から再び投資するのが良い方法だというのである。Boumanらによると、彼らが調査した37カ国中36カ国で類似の季節性が見られたとして、そういう効果を「ハロウィーン効果」と呼んでいる[1]。彼らが調査した37カ国の中に日本市場も含まれているが、Sakakibaraらが詳細に調査した結果、日本株式市場の季節性は1月から6月という特殊性を持つことを示し、その効果を「デカンショ節効果」と呼んだ[2]。彼らは、11月から4月まで株式を保有する“Sell in May”戦略と日本市場における1月から6月に株式を買うデカンショ節戦略を比較検討し、期末のwindow dressingや株価指数の銘柄入れ替え等様々な日本市場特有のバイアスをコントロールしても、日

本市場においては後者のパフォーマンスが有意に高いことを報告している。

株式変動の季節性は一時的な市場構造の歪みや、ある特定の国に特有な制度がもたらしているものとはいえない。英国では1694年から“Sell in May”効果（ハロウィーン効果と同義）が存在していることが報告されているし、日本市場のデカンショ節効果も1950年から長期の株価変動において観察されることが明らかになっている。このような季節性アノマリーがどうして発生するのかという原因究明については、はっきりとしたコンセンサスは得られておらず、投資家心理がそうしたサイクルをもたらしているという説が有力である。とりわけ、Kamastraらは、世界の株式市場の季節パターンと緯度との興味深い関係を指摘している[3]。冬場に日照時間が短くなる国ほど顕著な季節性がみられることから、根本原因は投資家のSAD(季節性鬱障害)にあるとの説を打ち出している。

一方、岡田らは、25年にわたる大規模な新聞記事データの中から、1月から6月の期間には楽観的予想記事が多く、7月から12月の期間には慎重な予想記事が多いことを示し、同じ事実に基づく見通しでも、年前半は楽観的に、後半は慎重な予想する傾向があることを発見し、市場全体のムードの揺れが「デカンショ節効果」の原因であると指摘している[4]。ただ、この検証方法では、投資家心理と株価収益率の因果関係の特定が困難である。株価収益率が上がることで、投資家心理が

*連絡先：大阪産業大学工学部電子情報通信工学科
〒574-8530 大阪府大東市中垣内3-1-1
E-mail: yamasaki@eic.osaka-sandai.ac.jp

表 1: 解析対象となった記事数の時系列推移

年度	全記事 (a)	株式関連 キーワードを 含む記事 (b)	株式関連 記事比率 (b/a)	予測文章を 含む記事 (c)	予測記事 比率 (c/b)
1993-95	885,927	26,779	3.02%	14,452	53.97%
1996-98	940,406	29,588	3.15%	14,247	48.15%
1999-01	886,838	34,055	3.84%	13,377	39.28%
2002-04	853,443	29,741	3.48%	13,852	46.58%
2005-07	904,380	31,913	3.53%	14,577	45.68%
2008-10	801,698	13,906	1.73%	6,483	46.62%
合計	5,272,692	165,982	3.15%	76,988	46.38%

好転する側面は少なからず存在し、ムードと株価の相互作用が考えられるからである。そこで本稿では、日経 225 株価指数の時系列データに変化点検出の手法を援用し、変化点の前における投資家心理の変化に着目した。その結果、日経 225 株価指数は年の前半により多くのポジティブな変化点を示し、後半にはネガティブな変化点を示している。更に、ポジティブな変化点の前には、楽観見通しを示す記事率が上昇し、ネガティブな変化点の前には慎重な見通しを示す記事率が上昇することが明らかとなった。

2 投資家心理の代理変数

2.1 データ

本稿では、文献 [4] と同様、投資家心理を表す指標として、新聞記事に含まれる楽観見通し表現と慎重な見通し表現の割合を用いる。新聞記事としては、日本経済新聞朝刊、夕刊、日経産業新聞、日経流通新聞 (MJ)、日経金融新聞 (日経ヴェリタス) の 1993 年から 2010 年までの全記事に含まれるテキスト情報を対象としている。ただし、記事の中には株式市場とは全く無関係のものも多いため、付されているキーワードで株式関係記事を識別した。更に、その中で将来の市場見通しについて述べた文章 (予測文章) だけを対象としている。表 1 に示すのは、対象となった記事数の年度別集計を 3 年毎にまとめたものである。記事データは日経メディアマーケティング社から購入した。

日経四紙全てを対象としているため 18 年間の全記事は 520 万余記事となる。そのうち、約 3% の 165,982 記事は株式関連キーワードが付されていたものである。これらの記事には、当日や過去の株式動向についての事実情報を記述したものと、将来に対する予想を記述したものが混在している。そこで、予測表現の含まれている記事に焦点をあてる。筆者らが「ありそうだ」「とどまるだろう」「・・となる見通し」「もみ合いそう」

「安心感が出てきた」などの予測表現 380 語を特定し、それらの語句を含む記事を予測記事と判断した。株式関連記事のおよそ半分弱が予測記事であり、18 年間で 76,988 記事を抽出した。

2.2 新聞記事の評価方法

予想記事の評価には、サポートベクターマシン (Support Vector Machine : SVM) を用いる。SVM は自然言語処理において良く用いられる線形二値分類器である。基本的にクラスが二つの問題を対象としており、高い分類性能を持っている。M 個の m 次元教師データ $\mathbf{x}_i (i = 1, \dots, M)$ が二つのクラスのいずれかに属し、そのクラスを $y_i = 1$ あるいは $y_i = -1$ で表現する。 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M$ は教師データの特徴ベクトル、 y_1, y_2, \dots, y_M は教師データのラベル (正のクラスの時 +1, 負のクラスの時 -1) を表している。これらのデータが線形分離可能であれば、次式の決定関数を決めることができる。

$$f(\mathbf{x}) = \mathbf{w}^T \cdot \phi(\mathbf{x}) + b \quad (1)$$

ここで、 $\phi(\mathbf{x})$ は入力空間の写像関数、 \mathbf{w} は m 次元係数ベクトル、 b はバイアス項で、 $f(\mathbf{x}) \geq 0$ のときは正のクラス、 $f(\mathbf{x}) < 0$ のときは負のクラスとなる。

多クラスでの分類が必要な場合は、SVM を組み合わせて多クラス識別を行う。k クラスの分類問題を解決する手法として、one versus rest 法がある。あるクラスと残り $k - 1$ クラスを分離する k 個の決定関数が、

$$f_i(\mathbf{x}) = \mathbf{w}_i^T \cdot \phi(\mathbf{x}) + b_i \quad (i = 1, \dots, k) \quad (2)$$

で与えられるとする。未知文書の特徴ベクトル \mathbf{x} は

$$\arg \max_{i=1, \dots, k} f_i(\mathbf{x})$$

のクラスに分類する。概念的には、分割の組み合わせの中で最も近いところに分類する方法である。

また、SVM のアルゴリズムを使い、多クラス分類を入出力関係を近似する関数として生成する回帰問題とする手法もある。サポートベクター回帰 (Support Vector Regression : SVR) では、実際の値から一定の距離内の誤差を無視するように学習する。そのときの損失関数は、

$$\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^M |y_i - f(\mathbf{x}_i)|_\epsilon$$

となり、学習のときには損失関数を最小化するように回帰関数 $f(\mathbf{x})$ を決定する。回帰関数からは ϵ の距離の誤差を無視して学習を行う。損失関数の帯域を狭く設定するか、広く設定するかは判定対象によって異なる。

本稿での具体的な予想記事判定手法を以下に示す。まず、「今週のマーケット」という毎週日曜日に日本経済新聞に掲載される展望記事を3年分収集し、筆者らで分担して判定した。この特定記事を選択したのは、定期的に週初めに書かれ、記事の目的がその週の株式市場の見通しを記述することだからである。3年分の記事を読み、株式市場が上昇しそうだという楽観的な展望記事であるのか、あるいは上値が重い、下値を試すといった慎重な展望記事であるのか、またあるいはそのどちらでもなく、中立的な展望記事であるかを分類することで「教師データ」を作成した。

次に、この教師データを用いて予測文章を含む 76,988 記事を判定する。手順は、事前に用意した 270 語の特徴語（「期待」「上昇」「警戒」「懸念」「堅調」「意欲」「改善」「重い」「後退」「過熱」「神経質」「不透明」「不安」「安心」「追い風」「慎重」「好調」「鈍化」「転じる」「積極」「先高」「様子見」等の 270 語）を用い、すべての記事データを 270 次元のベクトルとして一律に扱う。次に、教師データとして用意した「慎重な見通し」記事には評価値「1」を与え、「楽観見通し」記事には評価値「2」を与えて、SVM によって機械学習を実行する。これにより「慎重な見通し」記事クラスと「楽観見通し」記事クラスを分離する評価関数を決定する。この評価関数を使って回帰計算を実行し、分類対象となっている 76,988 の予測記事の評価値をそれぞれ求める。予測記事の評価値が 1 に近ければその記事は「慎重」なトーンで記述されていると判断し、評価値が 2 に近ければその記事は「楽観」的のトーンを持つ記事だと判断する。その中間であれば、「中立見通し」記事であるとみなす。本稿では回帰計算された評価値の年間平均値を基準として、評価値が基準から一定以上小さい記事を「慎重な見通し」とし、基準から一定以上大きい記事を「楽観見通し」とした。

こうして、日々の新聞報道から、その日の楽観度合い、慎重度合いの代理変数を楽観見通し記事と慎重な見通し記事の割合を計算することで作成した。本稿ではこの代理変数を市場センチメント指数と呼称する。

3 変化点検出

データマイニング技術を利用して膨大なデータから有用な知識を取り出す試みは幅広く行われている。系列データに対する異常検知はその一つであり、通常と異なる挙動を示す箇所を発見することが目的である。統計的異常検知においては、過去のデータからデータ発生分布の確率モデルを学習し、そのモデルをもとに、データの異常の度合いあるいはモデルの変化のスコアリングを行うことになる。異常検知は、データの外れ値を発見する外れ値検出、時系列上の急激な変化を発見する変化点検出、異常行動パターンを発見する異常行動検出などに分類される [5]。

観測する時系列データの変化を発見する変化点検出は、様々な時系列データに適用することが可能である。例えば、ネットワークセキュリティ分野では通信ログなどのデータを観察しておき、アクセス頻度が急激に増えるといった状況の変化を検出することで、攻撃の早期発見が可能となる。また、コンピュータ上のシステムログを監視しておき、通常と異なる挙動を発見することで、障害検出や故障診断が可能となる。本稿で対象とするのは株価指数という時系列データである。このデータから急激に株価指数の振る舞いに変化する変化点を見つけ出すことで、購買のトレンドが変換する時点を検出する。

変化点検出の基本的な考え方は時系列にモデルを当てはめ、変化点候補となる時点の前後で異なるモデルとした方が、一つのモデルで説明するよりも誤差が少ないかどうかを検定する手法である。検定結果が有意であれば、その候補点を変化点とする方式である。しかしながら、全ての候補点で統計的検定を行う必要があり、計算時間の面で問題がある。山西らはこの問題を解決するため、時系列データに対してリアルタイムに変化点の度合いを計算する Change Finder を提案している [5]。

Change Finder は自己回帰モデル (Auto Regression model : AR モデル) を使用したオンライン 2 段階学習に基づくアルゴリズムによって時系列データの各時点における変化点スコアを求めている。AR モデルは、時系列データにおいて定常状態を仮定し、時系列データの過去の変化から定常状態を予測するモデルである。Change Finder では AR モデルをオンライン忘却型学習アルゴリズム (Sequentially Discounting AR model learning algorithm : SDAR アルゴリズム) を用いて学習している。SDAR アルゴリズムは忘却型のため過去の統計量の影響を減らすことができ、非定常的なモデルの学習も可能としている。また、オンライン型であるため、過去のパラメータと直前の時系列データのみで予測が可能となり、計算量の削減ができリアルタイム的な予測が可能となる。

Chage Finder のアルゴリズムを以下に示す。

1. 時系列データの確率モデルとして AR モデルを用意し SDAR アルゴリズムにより入力時系列データの確率密度関数の列 $\{p_t(\mathbf{x}) : t = 1, 2, \dots\}$ を学習する (第 1 段階学習)。
2. 各時点 t のデータ \mathbf{x}_t の外れ値スコアを学習した確率密度関数からの外れ度合いとして計算する。

$$Score(\mathbf{x}_t) = -\log p_{t-1}(\mathbf{x}_t)$$

3. 外れ値スコアの T -平均を計算することでスコアの平滑化を行う。また、時系列として算出される外れ値スコア系列に対して、平均対象スコアをスライドさせることで外れ値の移動平均スコア系列 $\{y_t : t = 1, 2, \dots\}$ を求める。

$$y_t = \frac{1}{T} \sum_{i=t-T+1}^t Score(\mathbf{x}_i)$$

4. 得られた時系列データ $\{y_t : t = 1, 2, \dots\}$ に対して、再び AR モデルを用意して SDAR アルゴリズムにより、平滑化スコアの確率密度関数の列 $\{q_t : t = 1, 2, \dots\}$ を学習する (第 2 段階学習)。
5. 学習した確率密度関数列からの外れ程度を計算し、時刻 t における T' -平均スコアを定める。

$$Score(t) = \frac{1}{T'} \sum_{i=t-T'+1}^t (-\ln q_{i-1}(y_i))$$

本稿では、日経 225 株価指数の時系列に対して Chage Finder を適用することで、単なる株価指数の外れ値ではなく、本質的な変動が生じた時点のみを検出する。変化点スコアが高い値となる時点を株価のトレンド転換点として利用する。

4 変化点検出と市場センチメント

図 1 は標本期間である 1993 年から 2010 年までの変化点検出を行った結果、算出された変化点スコア (Score) と日経 225 株価指数 (NKY) の推移を概観したものである。グラフから、株価指数変動の転換点と変化点のスコアが高い点が一致していることが確認できる。

株価の時系列を対象とした変化点検出の場合、変化点には 2 種類のとらえ方がある。株価が上昇傾向に転換するポジティブ変化点と株価が下落傾向に転換するネガティブ変化点である。本稿ではこの 2 つの変化点を区別するため、変化点スコアの絶対値に株価のトレンド転換を表す正負の符号を付けた TScore を用いる



図 1: 日経 225 株価指数と変化点スコア

ことにする。TScore が正の場合はポジティブ変化点を示し、負の場合はネガティブ変化点を表している。

図 2, 図 3 に示しているのは、市場のセンチメント指数と TScore の関係を、局所的にズームアップしたものである。市場のセンチメント指数は、第 2 節で示した方法で、楽観的な記事と慎重な記事の割合を日々集計した値を、変化点が発見される日から 5 日間過去にさかのぼって求めた 5 日間の平均値である。したがって、グラフで示している市場センチメント指数の推移は変化点スコアの推移とは異なり、1 日前のセンチメントを示す。2 つのグラフがほぼ相関していることから、楽観的なムードが高まると、全体の標本期間を通じてポジティブの変化点が発生し、慎重なムードが高まるとネガティブの変化点が発生することが多いことがわかる。

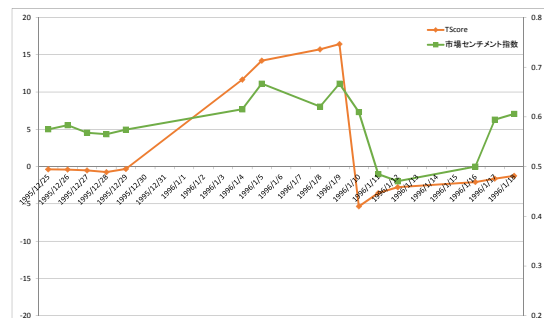


図 2: ポジティブ変化点 (上昇トレンド)



図 3: ネガティブ変化点 (下落トレンド)

表2に、スコアが大きく変化点として検出した日付、TScore および市場センチメント指数をまとめている。標本期間ではポジティブの方向の変化点が検出された11回のうち9回は市場に楽観的なセンチメントが支配的であった。一方、ネガティブ方向の変化点が検出された20回のうち17回は市場に慎重なセンチメントが支配的であった。また、変化点が検出された日時を確認すると、全部で11回のプラスの変換のうち8回は上半期、20回のマイナスの変換のうち、12回は下半期に起きている。

表2: 変化点スコアと市場センチメント指数

日付	TScore	市場センチメント指数
1993/ 6/22	-9.37	0.48
1993/ 9/16	-6.55	0.52
1994/ 1/27	10.32	0.78
1994/11/25	-6.10	0.63
1995/ 1/25	-102.52	0.40
1995/ 7/13	5.00	0.55
1996/ 1/ 9	16.43	0.67
1996/ 5/16	17.42	0.43
1996/ 7/24	-6.88	0.27
1997/ 1/16	-5.94	0.40
1997/ 5/ 9	5.41	0.88
1997/10/31	-9.70	0.42
1999/ 3/10	14.20	0.44
1999/ 7/26	-5.73	0.45
2000/ 1/11	5.19	0.52
2000/ 4/18	-11.39	0.42
2001/ 9/17	-6.40	0.42
2002/ 2/27	5.35	0.53
2002/ 6/21	-5.70	0.50
2003/ 7/ 3	5.70	0.53
2004/ 5/13	-11.07	0.39
2005/ 4/19	-26.40	0.39
2005/ 8/16	5.59	0.54
2006/11/24	-4.15	0.41
2007/ 1/16	5.61	0.53
2007/ 3/ 5	-15.66	0.57
2007/ 8/20	-6.52	0.38
2008/ 1/23	-6.46	0.39
2008/10/10	-6.43	0.40
2009/ 8/25	-4.95	0.46
2010/ 4/26	-4.13	0.50

5 結論

株価の季節性は運用者の収益に与えるインパクトが強いかかわらず、長期間継続しているアノマリー現象である。ただ、この原因については明らかになっていない。本稿では、Kamastra らの人間心理仮説 [3] を、新聞記事の中の将来見通しが、株式相場の変化点に与える影響を見ることによって検証した。

株式市場は定常過程ではなく、ある日突然変化するものである。明らかな政策変更や、外部環境で変化する場合もあるが、これといったニュースを伴わずに変化する場合も多い。そこで本稿では、非定常過程である株式市場の流れが変化するポイントを、変化点検出の手法で特定した。そして、その変化点があらわれる直前の市場ムードを、「変化点前の日経4紙における将来見通しの楽観度合い」と定義し、市場ムードと変化点の関係を観察した。

その結果、2つのことが明らかになった。第1に、株式市場は、暦年の前半に強気方向への変化点が多く検知され、暦年後半に弱気方向への変化点が多いということである。文献 [2] では25年間の観察期間の中で18年間については暦年前半に高い平均収益率を示すことが指摘されているが、変化点についても同様の傾向が確認された。第2に、変化点の5日前から市場ムードを計測した結果、ムードが良いときには、ポジティブの転換が起こり、逆に市場ムードが悪化しているときには、ネガティブの転換が起こることが明らかになった。これらの結果は、投資家心理の好転が年前半に起こりやすく、それが株式市場のプラス方向への転換に結びついていることを示唆しており、投資家心理が株式市場の季節性の原因であるとする仮説と整合的である。

参考文献

- [1] Bouman, S. and B. Jacobsen: Sell in May and Go Away: Another Puzzle, *American Economic Review*, Vol. 92, pp. 1618–1635 (2002)
- [2] Sakakibara, S, T. Yamasaki and K. Okada: The Calendar Structure of the Japanese Stock Market: The 'Sell in May Effect' Versus the 'Dekansho-Bushi Effect', *International Review of Finance*, Vol. 13, No. 2, pp. 161–185 (2013)
- [3] Kamastra, M., L. Kramer and M. Levy: Winter Blues: A SAD Stock Market Cycle, *American Economic Review*, Vol. 93, No. 1, pp. 324–343 (2003)

- [4] 岡田克彦, 山崎高弘, 榊原茂樹, 山崎尚志: 株価変動の季節性と投資家心理, 証券アナリストジャーナル, Vol. 51, No. 12, pp. 96-105 (2013)
- [5] 山西健司: データマイニングによる異常検知, 共立出版 (2009)