

テキストモダリティ分析に基づく投資家心理の検証

Investor Sentiment in News and the Seasonal Anomaly

山崎 高弘 *¹ 岡田 克彦 *² 月岡 靖智 *³
Takahiro Yamasaki Katsuhiko Okada Yasutomu Tsukioka

*¹大阪産業大学 工学部 *²関西学院大学 経営戦略研究科
Osaka Sangyo University, Faculty of Engineering Kwansai Gakuin University, Institute of Business and Accounting

*³関西学院大学 商学部
Kwansai Gakuin University, School of Business Administration

The well-known stock market adage “sell in May and go away” arose from long-term stock market seasonality in major financial markets around the globe. Kamastra, Kramer and Levy(2003) present evidence that Seasonal Affective Disorder causes this seasonality, as this condition has a profound effect in people’s mood and makes people increasingly risk averse as daylight diminishes with the onset of winter. In this paper, we present evidence that change in market mood is reflected in the prospect statement in the news text. We used modality analysis to identify the prospect statement in the large textual data.

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. Leegers day”と続く。St. Leegers 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 and Jacobsen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[Bouman 02]。また日本市場においても、Sakakibara, Yamasaki and Okadaが詳細に調査した結果、日本株市場の季節性は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[Sakakibara 13]。

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言い難い。英国では1694年から“Sell in May”効果（ハロウィーン効果）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年からの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。しかしながら、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いない。ハイリスク・ハイリターン原則から言えば、前半と後半にリスクの差異があるがゆえにリターンに差があると考えがちだが、前半にリスクが高いという証拠は見当たらない。また、季節性は、業種サイクル、小型株サイクルとも無関係である。更に、前半に株式に対するポジティブなニュースが多く、逆に後半にネガティブなニュースが多いというわけでもない[Bouman 02]。合理的に推測される様々な原因究明が行き詰まる中、株式の季節性にはなんらかの投資家心理が影響を与えていると推測する研究者が多い。その中でKamastra, Kramer and Leviは、

連絡先: 山崎高弘, 大阪産業大学工学部電子情報通信工学科,
大東市中垣内 3-1-1, yamasaki@eic.osaka-sandai.ac.jp

SAD（季節性鬱障害）という日照時間が短くなるに従って、人間心理に鬱的な傾向が現れることが株価リターンに影響を与えているのではないかと考えている[Kamastra 03]。

本稿では、人間心理の季節による変動を、大規模なマスメディアの記事を解析することで定量的に捉え、投資家心理がこうした季節性を生み出すことを検証する。ただ、マスメディアの記事には、過去の事実を述べたものが多く混在するため、将来について書いている記事を特定する必要がある。このため、本稿ではテキストモダリティ解析を用い、将来見通しにかかる記事をZundaを用いて特定した。それらの記事のセンチメントを算出し、各年、各月ごとに集計する。先行研究では、日本において前半にはお正月、桜の開花、新年度、新学期、ゴールデンウィークなどいわゆる「明るい」行事が多く、投資家心理の好転に結びついているのではないかと推測を述べているが[Sakakibara 13]、センチメントの集計値が前半に高く、後半に低くなっているかを確認する。

2. テキストのモダリティ

一般的に、記述されたテキストには、個々の事象（イベント）だけではなく書き手が表明している態度や真偽判断、価値判断などの情報が含まれている。例えば、以下のテキストからは、次に示したような書き手の態度を読み取ることが可能である。

- (1) 昨日は株価が上昇した。
- (2) 今後、株価が上昇するだろう。
- (3) 株価が上昇すれば、景況感の改善が期待される。

(1)は株価の上昇というイベントが成立していることを示しているのに対し、(2)ではイベントが高い確率で起きるであろうと推量している。(3)ではもしこのイベントが起きたらという条件を仮説的に述べている。一方で、(1)は株価の上昇が実際に起きたという事実を持っているが、(2)(3)では、まだこの事実は起きていないとも判断できる。

このように、記述内容に含まれる態度のことをモダリティと呼び、言語学分野で研究が行われている。テキストのモダリティを分析することで、読み手が情報をどのように受け取っ

表 1: 拡張モダリティとラベル

| 項目 | ラベル |
|-------|--|
| 態度表明者 | wr:筆者, wr:筆者_arb:不特定, (その他) |
| 時制 | 非未来, 未来 |
| 仮想 | 条件, 帰結, 0 |
| 態度 | 叙述, 意志, 欲求, 働きかけ-直接, 働きかけ-間接, 働きかけ-勧誘, 許可, 問いかけ |
| 真偽判断 | 成立, 不成立, 不成立から成立, 成立から不成立, 高確率, 低確率, 低確率から高確率, 高確率から低確率, 0 |
| 価値判断 | ポジティブ, ネガティブ, 0 |

たかの手がかりとすることが可能となる。日本語のモダリティに関しては、松吉らが態度表明者、時制、仮想、態度、真偽判断、価値判断、焦点などについてテキスト情報のアノテーションを行っている [松吉 10]。また、表 1 のように各項目についてラベルを与え、拡張モダリティと呼んでいる。例として、次の新聞記事をこの拡張モダリティに従って解析する。

- 東証では日経平均株価の上昇幅が 2 日連続で 200 円を上回った。米雇用統計が予想ほど悪くなかったことが市場のムードを和らげ、膠着状態から上昇相場に移行するだろう。

1 文目のイベントには、「非未来」「叙述」「成立」などのラベルが付き、2 文目のイベントには「未来」「働きかけ-間接」「ポジティブ」などのラベルが付く。詳細は後述するが本稿では 2 文目を対象として、新聞記事の評価を行うことにする。投資家の心理状態を新聞記事中に将来の予測内容をもとに検証を行うためである。

3. 投資家心理の抽出

3.1 対象とする新聞記事

本稿では、投資家心理を表す指標として、経済新聞記事に含まれる楽観見通し記事と慎重な見通し記事の割合を用いる。まず、1986 年から 2010 年の 25 年間について日本経済新聞朝刊・夕刊、日経産業新聞、日経流通新聞、日経金融新聞（日経ヴェリタス）の日経四紙に掲載された全新聞記事約 720 万記事を取得し、その中から、株式市場や市況などマーケットに関する記事約 22 万件を抜粋した。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。本稿の目的は、投資家の楽観（慎重な）見通しに季節性があるかどうかを検証することであるから、ここから将来の市場見通しについて述べている記事（予測記事）だけを抽出することにする。

予測記事の抽出のため、テキストのモダリティの概念を使用する。モダリティの解析には、日本語拡張モダリティ解析器 Zunda*1 を用いた。Zunda は文中のイベント（動詞や形容詞など）に対して、拡張モダリティに従い、その真偽判断（イベントが起こったかどうか）、仮想性（仮定の話かどうか）などを解析するプログラムである。新聞記事の文章を Zunda によって解析し、個々のイベントに対して付与されたラベルを利用し、以下に当てはまるテキストのみを抽出した。

*1 <https://code.google.com/p/zunda/>

- 時制が「未来」
- 態度が「叙述」ではない
- 真偽判断が「成立」ではない

抽出記事が楽観的であるか慎重であるかの判定、すなわち記事のセンチメント分析には、機械学習の手法を用いる。

3.2 新聞記事の評価

予測記事の評価には、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。本稿での具体的な記事センチメント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、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」的トーンを持つ記事だと判断する。その中間であれば、「中立な」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。「慎重」「中立」「楽観」の閾値は平均値から $\pm\alpha$ とした。

| | |
|---|------------|
| 評価値 \leq 平均値 $- \alpha$ | : 慎重な見通し記事 |
| 平均値 $- \alpha <$ 評価値 $<$ 平均値 $+ \alpha$ | : 中立見通し記事 |
| 平均値 $+ \alpha \leq$ 評価値 | : 楽観見通し記事 |

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。

4. 実証結果

4.1 日経四紙にみられる季節性

SVM によって、予測記事を「慎重」「中立」「楽観」に分類し、1986 年から 2010 年までの 25 年間の各種記事の月別出現頻度をまとめたものが表 2 である。25 年間を通じて 139,906 記事が対象となっている。表 2 の最右列には、楽観記事割合（楽観見通し記事数 / (楽観見通し記事数 + 慎重な見通し記事数)）を示しているが、全体期間の 25 年でみれば 3 月を除き上半期に楽観見通し記事が多く、下半期に慎重な見通し記事が多

表 2: 予測記事の分類結果の月別推移

| 月 | 予測記事数 | 楽観見通し 記事数 | 中立見通し 記事数 | 慎重な見通し 記事数 | 楽観記事 割合 |
|-----|---------|--------------|--------------|---------------|------------|
| 1月 | 11,599 | 2,360 | 6,970 | 2,269 | 0.510 |
| 2月 | 10,518 | 2,090 | 6,335 | 2,093 | 0.500 |
| 3月 | 12,211 | 2,400 | 7,337 | 2,474 | 0.492 |
| 4月 | 11,737 | 2,390 | 7,004 | 2,343 | 0.505 |
| 5月 | 10,743 | 2,242 | 6,367 | 2,134 | 0.512 |
| 6月 | 11,555 | 2,490 | 6,851 | 2,214 | 0.529 |
| 上半期 | 68,363 | 13,972 | 40,864 | 13,527 | 0.508 |
| 7月 | 12,050 | 2,519 | 7,172 | 2,359 | 0.516 |
| 8月 | 12,203 | 2,469 | 7,121 | 2,613 | 0.486 |
| 9月 | 11,177 | 2,081 | 6,723 | 2,373 | 0.467 |
| 10月 | 13,042 | 2,443 | 7,709 | 2,890 | 0.458 |
| 11月 | 11,454 | 2,146 | 6,883 | 2,425 | 0.469 |
| 12月 | 11,617 | 2,425 | 6,936 | 2,256 | 0.518 |
| 下半期 | 71,543 | 14,083 | 42,544 | 14,916 | 0.486 |
| 合計 | 139,906 | 28,055 | 83,408 | 28,443 | 0.497 |

いことがわかる。また、上半期、下半期での各記事合計からも同様なことがいえる。従って、暦年の前半に楽観記事が多く、前半における将来予測は楽観に傾き、逆に後半における将来予測は慎重に傾いているといえる。

4.2 予測記事の楽観記事割合と保有リターンの関係性

Sakakibara らによると、日本株式市場においてデカンショ節効果は 1986 年から 2010 年の 25 年間に 18 年間も現れ、7 年間だけが年後半にパフォーマンスが良かったとなっている。表 3 は、楽観記事割合について上半期における値と下半期における値および東証株価指数 (TOPIX) の 1 月からの 6 月までの半年間の保有リターンから 7 月から 12 月までの保有リターンの差を年度ごとに時系列に表示したものを示している。これら 2 つを比較するとデカンショ節効果がみられる年と、上半期に楽観記事割合が上回っている年が連動していることがわかる。連動していないのは唯一 2009 年だけである。2009 年については、リーマン・ショックが起こったのが 2008 年 9 月であり、まだそのショックが冷めやらず、翌年の 2009 年前半にも残っていたと推測できる。25 年間の観察期間の中で、23 年間については株式市場の予想記事が「楽観」に傾いた年は、デカンショ節効果が現れ、「慎重」に傾いた年はデカンショ節効果が見られないのである。これは、株価の季節性が投資家心理の揺れによってもたらされているという仮説と整合的な結果であるといえる。この極端な株価の季節性について報じられても、投資家は年前半に株式を購入し、6 月には売ってしまうとは行動しないようである。株価の季節性が知識として広まったとしても、限定合理性を持った投資家の現実の投資行動は、それ以上に市場のムードに左右されるのかもしれない。

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にもかかわらず、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 [Kamastra 03] 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。

表 3: 楽観記事割合と東証株価指数 (TOPIX) の年度推移

| 年 | 楽観記事割合 | | TOPIX | |
|------|--------|-------|--------|--------|
| | 上半期 | 下半期 | 上半期 | 下半期 |
| 1986 | 0.516 | 0.500 | 29.5% | 14.5% |
| 1987 | 0.569 | 0.464 | 31.0% | -15.4% |
| 1988 | 0.521 | 0.469 | 26.5% | 8.0% |
| 1989 | 0.473 | 0.502 | 3.9% | 17.6% |
| 1990 | 0.499 | 0.491 | -18.0% | -26.0% |
| 1991 | 0.544 | 0.449 | 4.9% | -5.7% |
| 1992 | 0.485 | 0.495 | -27.0% | 5.0% |
| 1993 | 0.514 | 0.482 | 20.0% | -8.9% |
| 1994 | 0.536 | 0.450 | 16.3% | -6.8% |
| 1995 | 0.427 | 0.531 | -27.0% | 31.0% |
| 1996 | 0.528 | 0.463 | 8.5% | -14.0% |
| 1997 | 0.567 | 0.421 | 5.6% | -24.4% |
| 1998 | 0.520 | 0.486 | 4.7% | -11.7% |
| 1999 | 0.512 | 0.509 | 30.3% | 21.6% |
| 2000 | 0.538 | 0.500 | -7.6% | -19.3% |
| 2001 | 0.509 | 0.497 | 1.3% | -20.0% |
| 2002 | 0.549 | 0.439 | -0.7% | -17.7% |
| 2003 | 0.455 | 0.546 | 7.1% | 15.5% |
| 2004 | 0.540 | 0.456 | 14.0% | -3.4% |
| 2005 | 0.433 | 0.554 | 2.4% | 40.1% |
| 2006 | 0.475 | 0.524 | -3.8% | 5.9% |
| 2007 | 0.512 | 0.455 | 5.6% | -16.9% |
| 2008 | 0.555 | 0.437 | -10.0% | -34.9% |
| 2009 | 0.472 | 0.535 | 8.2% | -2.4% |
| 2010 | 0.522 | 0.466 | -7.3% | 6.8% |

一流新聞の市場関係記事は、市場の専門家や知識の豊富な記者が記していることから、客観性、中立性が保たれていると考えられる。しかしながら、新聞記事はその性質上、既に明らかになっている事実が多く含まれている。将来に向けての投資

家心理を検証するためには、既知の事実が記載されている箇所を取り除き、将来の見通しに関する箇所のみを特定する必要があった。本稿では、言語学の分野におけるモダリティの概念を組み込むことでその問題を解決した。

テキストモダリティ分析に基いて、25年間にわたる大量の記事データの中の予想表現を解析した結果、上半期の1月から6月の期間には相場の先行きについて楽観的な見通し表現が多く、下半期の7月から12月には先行き慎重な見通し表現が増加することが明らかになった。この傾向は25年間の観察期間のうち18年間に現れ、概ね年前半に将来を楽観するムードがあることを示唆した。更に、年前半に楽観の見通し記事が多い18年間において、その年の株価パフォーマンスを比較してみると、2010年を除いて株価パフォーマンスがより高いリターンを示している。一方、年前半に慎重な見通しが多いのは25年間のうち7年間であるが、株価パフォーマンスは2009年を除いてすべて年前半に低くなっている。すなわち、新聞記事の予想表現が年前半に楽観に傾く年は、デカンショ節効果が現れ、そうでない年は現れないという傾向がはっきり見られたのである。これは、株価変動の季節性が人間心理の季節的変動に依拠するという仮説と整合的な結果であることがわかった。

参考文献

- [Bouman 02] Bouman, S. and B. Jacobsen: Sell in May and Go Away: Another Puzzle, *American Economic Review*, Vol.92, pp.1618–1635, 2002.
- [Sakakibara 13] 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.
- [Kamastra 03] 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.
- [松吉 10] 松吉俊, 江口萌, 佐尾ちとせ, 村上浩司, 乾健太郎, 松本裕治: テキスト情報分析のための判断情報アノテーション, 電子情報通信学会論文誌, Vol.J93-D, No.6, pp.705–713, 2010.