

# インターネット株式掲示板の投稿内容と株式市場の関連性

## Is Content of Stock BBS related to the Stock Index?

丸山健<sup>1</sup> 梅原英一<sup>2</sup> 諏訪博彦<sup>3</sup> 太田敏澄<sup>3</sup>

Ken Maruyama<sup>1</sup>, Eiichi Umehara<sup>2</sup>, Hirohiko Suwa<sup>3</sup> and Toshizumi Ohta<sup>3</sup>

<sup>1</sup>ネクストソリューションズ株式会社

<sup>1</sup> NEXTSOLUTION CO., LTD

<sup>2</sup>野村総合研究所

<sup>2</sup> Nomura Research Institute

<sup>3</sup>電気通信大学

<sup>3</sup> University of Electro-Communications

**Abstract:** We are to analyze relations between stock index and stock BBS. Previous studies in US find that the stock BBS posting can predict market volatility and trading volume. We will develop hypotheses based on the results of these analyses, and apply statistical analysis to the data of companies that are observed the large amount of message posted in Yahoo! stock topics in 2005-2006. Based on the messages, we will analyze the contents of posting using natural language processing and machine learning. In concerning future stock return, the number of posting is significantly correlated with the volatility and the trading volume, and that significant correlations are observed between the amount of bullish opinion and the stock return.

## 1 はじめに

利用者が発信するメディア CGM (Consumer Generated Media) と株式市場の関係に関する報告が増加している。例えば、アスキー [2006] では、blog と企業株式の関係についての特集がなされた。また、『日本経済新聞』 [2006] によると、「新興市場の株価が下げ続ける背景の中には、買い時を知らせるアナリストが足りないという問題があり、株価の下がった理由を探したら正体不明のネット掲示板の書き込みだった」との報道もある。Van Bommel [2003] は、個人情報拡散モデルを用いて、株式の噂 (うわさ) が拡散し、その結果追従者がその噂に従い取引をし、さらに噂を拡散させることにより、株価がオーバーシュートするモデルを提示している。これらは、インターネット上での投稿が株価と関連している可能性を示唆している。

そこで、本稿では、日本におけるインターネットの株式掲示板のメッセージと株式指標の関係について分析を行う。掲示板に投稿されたメッセージの分析に基づき、そこから得られる掲示板指標 (投稿数や投稿内容) と株式指標 (リターン、出来高、ボラ

ティリティ) の関係性を明らかにすることを目的とする。

本稿では、第2節で先行研究をレビューする。第3節では先行研究を踏まえて仮説を設定する。第4節、第5節、第6節では分析方法、分析結果および結果の考察を行う。第7節は結論である。

## 2 先行研究

米国において、インターネット株式掲示板から得られる掲示板指標と株式指標の関係に関する先行研究がある。Antweiler and Frank [2004] は、ダウ・ジョーンズ工業株指数とダウ・ジョーンズインターネット指数組み入れの45社について150万以上のYahoo! と Raging Bull に投稿されたメッセージの投稿数および内容を、自然言語処理 (NaiveBayes 法) により分析し、それと株式指標の関係を実証的に検証している。結論として、第1に、掲示板は株式リターンを予測しない。第2に、強気と弱気が拮抗する意見相違が取引を誘発し、また、メッセージ投稿は出来高を予測する。第3に、メッセージ投稿は、取引日当日のボラティリティを予測することを指摘している。

投資家間の意見の相違と出来高およびボラティリティ（株価変化の絶対値）の分析に関しては、Harris and Raviv [1993] による投機的取引の数理モデルがある。このモデルでは、投機的取引は、公開情報の解釈の差としての意見の不一致から生じるとされる。分析の結果、株価変化の絶対値と出来高は正の相関がある、投資家による最終損益の予測変化の絶対値と出来高の間には正の相関が存在することを指摘している。また、もし投機家が情報を過大評価（過小評価）するなら、その後の株価変化はマイナス（プラス）の自己相関を示すこと、出来高は正の自己相関を示すことを指摘している。さらに、出来高は、オーバーナイトや週末・休日など、市場がクローズした後の寄り付きで、平均的に大きくなることを指摘している。

Wysocki [1999] は、Yahoo! 掲示板に投稿されたメッセージの投稿数によるクロスセクションと時系列の決定要因を調べ、メッセージの投稿数の変化が企業特性や株式市場活動に関連するかどうか検証している。その結果、累積投稿数は空売り比率、市場評価、機関投資家持ち株比率、取引量、パフォーマンス、アナリストのフォロー等、さまざまな各種企業との関連があることを発見している。投稿数と株価リターンとの関係については、前日夜の投稿数が翌日のリターンを統計的に示すが、手数料を考慮に入れた場合その情報を利用して経済的な利益を得ることは難しいと述べている。

一方、Tumarkin and Whitelaw [2001] は、Raging Bull に投稿された投稿数や表明された書き手の意見と、リターンや出来高との因果関係をイベントスタディと多変量自己回帰分析により検証をしている。検証の結果、掲示板メッセージがリターン出来高を予測する因果関係は認められていない。そして、この結果は市場の効率性を示すものであると結論付けている。

山下ほか [2005] は、日本の Yahoo! 掲示板における投稿の特徴を分析し、ブロードバンドの普及と市場の好転が 2003 年春以降の投稿数の増加の原因になっていること、同時期以降の市場における売買代

金や証券会社の約定数と投稿数の相関が極めて強いこと、人々の耳目を集めるニュースが各銘柄の掲示板の投稿数に強く影響を与えていることを報告している。

このように米国においては、株式掲示板の投稿数や投稿内容と、株式指標の関係に関する先行研究が行われている。しかし、日本における研究は、山下ほか [2005] の投稿数と株価変動の関係について検証した研究以外に見当たらない。また、彼らは数銘柄の投稿量と株価変動のみに注目しており、内容まで分析していない。そこでわれわれは、先行研究に基づき仮説を設定し、日本における株式掲示板と株式指標の関係について検証を行う。

### 3 仮説設定

本節では、先行研究に基づき、株式指標と関係のある掲示板指標として、掲示板の投稿数と投稿内容の二つに注目し仮説を設定する。表 1 は、仮説をまとめたものである。なお、株価に影響を与える要因として、企業のファンダメンタルズ、アナリストレポート、マスコミ報道等他にも様々なものがあるが、本稿の分析対象は掲示板の投稿と企業の株式指標に限定する。

#### 3.1 投稿数と株式指標の関係

Wysocki [1999] は投稿数が翌日の超過リターンを予測することができると述べているが、Antweiler and Frank [2004] は、投稿数から株価リターンを予測することはできないと結論づけている。投稿数と出来高の関係について、Wysocki [1999] は、前日夜の投稿数は翌日の出来高を説明できることを実証している。また、投稿数とボラティリティの関係について、Antweiler and Frank [2004] は、投稿数がボラティリティと正の相関をもつと述べている。これらの先行研究に基づき、投稿数と株式指標の関係について、以下の三つの仮説を設ける。

仮説 1：投稿数は株価リターンと相関がない

インターネット掲示板に投稿する人々は、その投稿内容から、明らかに機関投資家ではなく、個人で

表1 本稿の仮説 (○:関係がある ×:関係がない)

	株価リターン	出来高	ボラティリティ
投稿数	×(仮説1) Antweiler and Frank[2004] Tumarkin and Whitelaw[2001]	○(仮説2) Antweiler and Frank[2004] Wysocki[1999]	○(仮説3) Antweiler and Frank[2004]
投稿内容	強気比率	○(仮説4) Antweiler and Frank[2004]	○(仮説5) Antweiler and Frank[2004]
	合意インデックス	×(仮説7)	○(仮説8) Antweiler and Frank[2004] Harris and Raviv[1993]

株式に関心を持つ人々であると考えられる。内容も様々で、「買いたい」、「強く買いたい」だけでなく、「売りたい」、「強く売りたい」や「様子見」を主張している投稿もある。つまり、投稿数自体は、強気（株価上昇）・弱気（株価下降）に対して中立の指標である。そこで、投稿数は株価リターンに対して相関はないと考えられる。

#### 仮説 2：投稿数は出来高と相関がある

もしインターネット掲示板に投稿する人々が、個人のオンライントレード等を利用する投資家であった場合には、投稿内容を見て投資判断を行い、その結果、取引行動を引き起こす可能性が考えられる。掲示板の投稿が活発になり、投稿量が多くなれば、その結果、個人投資家の投資行動に影響を与え、出来高が多くなる可能性が考えられる。つまり、投稿数は、インターネットを利用する個人投資家のその銘柄に対する関心度の代理変数として考えられる。

#### 仮説 3：投稿数はボラティリティと相関がある

仮説 2 で述べたとおり、掲示板への投稿活動が個人投資家の投資行動に影響を与える可能性がある。この結果、売買が活発になり、株価の変動を増幅させる可能性が考えられる。このため、投稿数が増えると、株価のボラティリティが高くなる可能性が考えられる。

### 3.2 投稿内容と株式指標の関係

投稿内容を示す掲示板指標として、強気意見と弱気意見のどちらが多いか（以降、強気比率と呼ぶ）と強気意見と弱気意見の数がどちらか一方に片寄っているか拮抗しているか（以降、合意インデックスと呼ぶ）の二つの指標に注目する。強気比率とは、一定期間における強気投稿数と弱気投稿数の差である。合意インデックスとは、一定期間において投稿が、強気か弱気かどちらか一方に偏っているのか、拮抗しているのかを表す比率である。

Antweiler and Frank[2004]は、強気比率 (bullishness) と株価リターンの関係について、同時間帯では統計的に有意であると主張している。強気比率と出来高及びボラティリティの関係について、Antweiler and Frank[2004]は、有意な関係を確認している。Harris and Raviv[1993]は、市場情報に対するトレーダの解釈の違いが取引を誘発させるモデルを提示している。すなわち、強気の投稿と弱気の投稿が拮抗している場合、出来高及びボラティリティが大きくなることを示している。また、Antweiler and Frank[2004]は、合意インデックスが出来高と負の関係にあることを報告している。なお、先行研究が散見されないので、本稿では合意インデックスと株価リターンに関しては、相関がないという仮説を設定する。これらの先

行研究に基づき、投稿内容と株式指標に関する仮説として、以下の六つを設定する。

#### 仮説 4：強気比率は、株価リターンと相関がある

投稿活動が、個人投資家の投資行動に影響を与えると考えると、投稿内容で、強気意見が多ければ買い行動を引き起こす可能性が、弱気意見が多ければ売りを引き起こす可能性が考えられる。そこで仮説 4 では、強気比率は、株価リターンに正の影響を与えると考える。

#### 仮説 5：強気比率は、出来高と相関がある

強気の見解が弱気の見解より多ければ、その銘柄に対する意見全体は強気の傾向になる。それが個人投資家の投資行動に影響を与え、出来高が増える可能性が考えられる。

#### 仮説 6：強気比率はボラティリティと相関がある

強気の見解が弱気の見解より多ければ、その銘柄に対する意見全体は強気の傾向になる。その結果、それが個人投資家の投資行動に影響を与え、株価が変動し、ボラティリティが増加する可能性が考えられる。

#### 仮説 7：合意インデックスは株価リターンと相関がない

掲示板の投稿内容が強気または弱気のどちらかに偏ったとしても、相場が上昇または下降する可能性はあるが、偏りの程度と株価リターンの相関はないと考えられる。

#### 仮説 8：合意インデックスは出来高と相関がある

Harris and Raviv[1993]では、投機家による公開情報の解釈に関する意見の相違が取引を誘発し、出来高の増加につながるというモデルであった。つまり彼らの米国市場を想定したモデルでは、強気と弱気の見解が拮抗すると、出来高が増加するというを示している。

#### 仮説 9：合意インデックスはボラティリティと相関がある

仮説 8 と同様に Harris and Raviv[1993]は、公開情報の解釈に関する意見の相違により出来高が増加し、その結果、ボラティリティ（株価変化の絶対値）が増加するというモデルを提示している。

## 4 分析方法

本節では、前節にて示した仮説を検証するために、「Yahoo!株式掲示板」を対象に分析を行う。

### 4.1 分析対象

本稿の分析対象企業及び掲示板は、Yahoo!株式掲示板における東証一部上場銘柄のうち投稿数上位 50 社としている。分析対象期間は 2005/01/01 から

2006/12/31 の間である。2005 年度は株価上昇局面の投稿データが多く含まれているが、2006 年度は 2005 年度に比較してボックス相場の投稿データが多く含まれていると考えられる。なお、分析期間では下降局面のデータは少ないと考えられる。下降局面の分析は今後の課題とする。

また、本稿では掲示板への投稿活動と株価の関係について分析することが目的であるため、分析対象を投稿量の多い銘柄 50 社に限定して分析する。投稿量の少ない銘柄に対する分析は、今後の検討課題とする。

## 4.2 株式指標

分析対象の株式指標は「株価リターン( $R(t)$ )」、「超過株価リターン( $AR(t)$ )」、「出来高( $Vol$ )」、「ボラティリティ( $Vlt$ )」、「株価リターン絶対値( $|R(t)|$ )」の計 5 指標である。各銘柄の株式指標は「株価 CD-ROM2006 年版」より算出を行う。株価は修正係数をかけ、出来高は修正係数で割っている。

日次ベースの株価リターンは、 $R(t)=終値(t)/終値(t-1)-1$  とする。超過株価リターン  $AR(t)$  は、 $R(t)$  より東証株価指数(TOPIX)の日次リターンを差し引くことにより算出する。本稿ではボラティリティ( $Vlt(t)$ )の定義を、当日を含めた前後二日間のリターンの標準偏差  $VLT(t)=標準偏差(R(t-2), R(t-1), R(t), R(t+1), R(t+2))$  とする。株価リターンの絶対値 $|R(t)|$ は一日あたりの株価リターンの変動と掲示板の関係を知るために分析に利用する。

## 4.3 掲示板投稿データの収集及び強気・弱気分類

掲示板の投稿を自然言語処理と機械学習の手法に基づき、メッセージを「弱気」「中立」「強気」の三種類に分類する。メッセージ分類は、各投稿メッセージから抽出した特徴ベクトルに対して Support Vector Regression (以下 SVR と呼ぶ)を適用することにより行う。投稿の分類を行うために投稿の表題と本文(以下投稿メッセージ)から特徴ベクトルの抽出を行う。

### 4.3.1 形態素解析・ノイズ除去処理

インターネット掲示板からプログラムを用いて自動的に収集した投稿メッセージを、形態素解析プログラムにより文章を形態素ごとに切り分けて、単語を抽出する。さらに、形態素解析の結果得られた単語群から、特徴ベクトルとして適さない単語を除去するために、以下のノイズ除去処理を行う。

- (1) 数字, 英字, 記号, 及び日本語以外の単語の除去

- (2) 不要語 (助詞, 助動詞, 接続詞, 連体詞, 副詞, 数, 代名詞, 感動詞, 固有名詞) の除去
- (3) 否定語の反映

以上のことより、特徴ベクトルを算出するために必要な単語群を抽出している。

### 4.3.2 分類に有用な単語の抽出と特徴ベクトルの算出

次に、各投稿の特徴ベクトルを算出する。このために、まず、各投稿に含まれるノイズ除去処理後の単語の中から、分類に有用と考えられる単語を単語辞書より抽出する。単語辞書は、高村ら[2006]の単語感情極性対応表にある単語と、2005 年ソニー yahoo 掲示板の投稿メッセージから作成した TF・IDF 値を全て合計し、上位 5000 件の中から評価表現に影響すると考えられる単語を、手動で選択したものを合わせたものを用いた。総単語数は、6989 単語である。

各投稿の特徴ベクトルは、6989 次元のベクトルであり、その個々の値は単語の重要度である。重要度とは、投稿における単語の出現頻度の関数であり、TF・IDF 法により算出を行う。なお、この段階で特徴ベクトルを抽出できない投稿についてはノイズとして分類対象そのものから除外する。投稿ごとに算出された 6989 次元ベクトルを、SVR による分類のための特徴ベクトルとする。

### 4.3.3 SVR による分類

本稿では、掲示板の投稿を SVR を用いて、「強気」「中立」「弱気」の三種類に分類する。SVR は、Support Vector Machine (Cristianini[2000]参照)を回帰分析に適用したものであり、回帰式に基づく理論値を出力する。岡野原・辻井[2005]は、SVR を用いて、Amazon.co.jp のレビューの内容とカスタマーレビューのおすすめ度を学習データとし、レビュー内容からおすすめ度が推定できることを示している。本稿では、「投稿者の気持ち」とそれに対応する投稿内容を学習データとして、SVR を用いて投稿内容から「強気」「中立」「弱気」を分類する。

Yahoo!株式掲示板には、「投稿者の気持ち」を公開する機能がある。しかしながら、すべての投稿に「投稿者の気持ち」が公表されているわけではない。「投稿者の気持ち」が公開されたものは全体の 21%にすぎない。そこで、本稿では、SVR を用いて、「投稿者の気持ち」が公開されていないメッセージを含む全投稿の「強気」「中立」「弱気」を判別する。

SVR で内容を分類する際の学習データは、「公表しない」を除いた投稿メッセージとし、入力値は投稿の特徴ベクトル、出力値は「強く買いたい」を 1、「買いたい」を 0.5、「様子見」を 0、「売りたい」を -0.5、「強く売りたい」を -1 として学習を行う。企業ごとに学習を行った SVR を用いて、分析対象期間の分類可能な全ての投稿に SVR の理論値を付与

表2 50社(2005年1月1日-2006年12月31日)のSVRによる分類結果

		分類結果			合計	精度
		弱気 ( $x \leq -0.5$ )	中立 ( $-0.5 < x < 0.5$ )	強気 ( $0.5 \leq x$ )		
投稿者の 気持ち	強く買いたい	591 (0.3%)	9,502 (4.1%)	82,282 (35.5%)	92,375 (39.8%)	84.3%
	買いたい	187 (0.1%)	8,671 (3.7%)	19,122 (8.2%)	27,980 (12.1%)	
	様子見	4,181 (1.8%)	23,311 (10.0%)	37,721 (16.3%)	65,213 (28.1%)	37.9%
	売りたい	1,003 (0.4%)	4,932 (2.1%)	1,836 (0.8%)	7,771 (3.3%)	
	強く売りたい	16,633 (7.2%)	14,456 (6.2%)	7,649 (3.3%)	38,738 (16.7%)	
合計		22,595 (9.7%)	60,872 (26.2%)	148,610 (64.0%)	232,077 (100.0%)	
精度		78.1%	38.3%	68.2%	61.3%	

精度は  の合計である

表3 強気・弱気分類結果

		弱気 ( $x \leq -0.5$ )	中立 ( $-0.5 < x < 0.5$ )	強気 ( $0.5 \leq x$ )	合計
分類 結果	学習データ	22,595 (9.7%)	60,872 (26.2%)	148,610 (64.0%)	232,077 (100.0%)
	気持ちなし	59,564 (6.8%)	318,026 (36.4%)	496,643 (56.8%)	874,233 (100.0%)
	計	82,159 (7.4%)	378,898 (34.2%)	645,253 (58.3%)	1,106,310 (100.0%)

する。その後、SVRの理論値が0.5より大きいものを「強気」投稿、-0.5より小さいものを「弱気」投稿、それ以外を「中立」投稿として分類を行う。SVRの分類プログラムはLibSVMを用いた。

SVRを用いた学習データの分類結果について、「強く買いたい」「買いたい」が強気、「様子見」が中立、「売りたい」「強く売りたい」が弱気に分類されている割合を精度と考えると、精度は61.3%である(表2)。

この内訳を見ると、「強く買いたい」、「買いたい」が強気に分類された割合は、84.3%である。しかし、「様子見」が中立に分類された割合は35.7%であるが、強気に分類された割合は57.8%あった。また、「売りたい」「強く売りたい」が弱気に分類された割合は、37.9%であるが、中立に分類されたものは41.7%、強気に分類されたものは20.4%ある。つまり、本稿で用いたSVRによる分類では、強気にシフトする傾向が出た。分類結果が強気にシフトする傾向に対する補正や、中立の分類精度の向上は、今後の検討課題とする。

以上の学習データをもとに、投稿者の気持ちなしのデータを含めた50社の2005年1月1日から12月31日までの投稿データ1,106,310件を強気・弱気・中立に分類した(表3)。分類の結果、強気が645,253件(58.3%)、弱気が82,159件(7.4%)であった。このデータを以降の強気および弱気の分析に使用する。

#### 4.4 掲示板指標の作成

掲示板指標は「投稿数」、「強気投稿数」、「弱気投稿数」、「強気比率」、「合意インデックス」の計5指

標である。掲示板指標は株式市場の取引時間に合わせ、場前(0:00-9:00)、場中(9:00-15:00)、場後(15:00-24:00)それぞれに対して算出を行う。投稿数を例にとると、総投稿数(t)=場前投稿数(t)+場中投稿数(t)+場後投稿数(t)となる。

##### (1) 投稿数

投稿数は、ある時間間隔に投稿された投稿の数である。

##### (2) 強気投稿数

強気投稿数は、前節のSVR回帰により強気投稿に分類された投稿の数である

##### (3) 弱気投稿数

弱気投稿数は、SVR回帰により弱気投稿に分類された投稿の数である

##### (4) 強気比率

強気比率は、時間間隔における強気投稿数と弱気投稿数の差を示す掲示板指標であり定義は以下である。この指標は、弱気投稿数に比べて強気投稿数が多くなると正の値となり、強気投稿数と弱気投稿数が同数の場合は0、強気に比較して弱気投稿数が多くなると負となる。なお、強気投稿数と弱気投稿数の合計が3未満の場合は、強気比率が判断できないと考え、分析から除外した。この結果、各時間帯のサンプルサイズは、場前7,572件、場中12,351件、場後14,818件となった。

$$\ln\left(\frac{1 + \text{強気投稿数}}{1 + \text{弱気投稿数}}\right)$$

表 4 リターンに関する相関結果

		リターン			超過リターン		
		前日	当日	翌日	前日	当日	翌日
投稿数	場前	-0.017 *		-0.015 *	0.016 *		-0.015 *
	場中			-0.014 *			-0.016 *
	場後		-0.024 **			-0.017 **	
強気投稿数	場前					0.003	
	場中	0.019 **	0.020 **		0.025 **	0.025 **	
	場後						
弱気投稿数	場前	-0.040 **	-0.031 **		-0.040 **	-0.032 **	
	場中	-0.032 **	-0.068 **	-0.016 *	-0.034 **	-0.070 **	-0.021 **
	場後	-0.014 *	-0.058 **	-0.024 **	-0.014 *	-0.059 **	-0.025 **
強気比率	場前	0.075 **	0.063 **		0.082 **	0.065 **	0.024 *
	場中	0.066 **	0.105 **	0.019 *	0.071 **	0.112 **	0.024 **
	場後	0.052 **	0.110 **	0.036 **	0.055 **	0.108 **	0.036 **
合意インデックス	場前	0.052 **	0.039 **		0.054 **	0.035 **	
	場中	0.042 **	0.090 **	0.019 *	0.040 **	0.089 **	0.026 **
	場後	0.050 **	0.096 **	0.035 **	0.049 **	0.087 **	0.033 **

\*p<.05 \*\*p<.01 なお、有意でない結果は省略している \*p<.05

表 5 出来高に関する相関結果

		出来高		
		前日	当日	翌日
投稿数	場前	0.189 **	0.180 **	0.156 **
	場中	0.226 **	0.251 **	0.206 **
	場後	0.189 **	0.212 **	0.195 **
強気投稿数	場前	0.189 **	0.182 **	0.158 **
	場中	0.224 **	0.248 **	0.208 **
	場後	0.192 **	0.216 **	0.201 **
弱気投稿数	場前	0.077 **	0.075 **	0.067 **
	場中	0.061 **	0.073 **	0.055 **
	場後	0.058 **	0.066 **	0.061 **
強気比率	場前	0.108 **	0.101 **	0.083 **
	場中	0.149 **	0.172 **	0.144 **
	場後	0.129 **	0.147 **	0.137 **
合意インデックス	場前			
	場中	0.042 **	0.049 **	0.046 **
	場後	0.043 **	0.050 **	0.051 **

\*\*p<.01 なお、有意でない結果は省略している

表 6 ボラティリティに関する相関結果

		ボラティティ			絶対値リターン		
		前日	当日	翌日	前日	当日	翌日
投稿数	場前	0.180 **	0.171 **	0.153 **	0.157 **	0.125 **	0.096 **
	場中	0.257 **	0.254 **	0.237 **	0.205 **	0.234 **	0.142 **
	場後	0.217 **	0.218 **	0.209 **	0.151 **	0.206 **	0.148 **
強気投稿数	場前	0.162 **	0.156 **	0.142 **	0.141 **	0.111 **	0.092 **
	場中	0.233 **	0.230 **	0.217 **	0.187 **	0.209 **	0.131 **
	場後	0.198 **	0.199 **	0.192 **	0.140 **	0.186 **	0.135 **
弱気投稿数	場前	0.089 **	0.085 **	0.076 **	0.081 **	0.068 **	0.045 **
	場中	0.116 **	0.117 **	0.108 **	0.088 **	0.115 **	0.067 **
	場後	0.092 **	0.094 **	0.091 **	0.055 **	0.099 **	0.067 **
強気比率	場前	0.092 **	0.094 **	0.089 **	0.074 **	0.063 **	0.057 **
	場中	0.152 **	0.146 **	0.139 **	0.110 **	0.137 **	0.081 **
	場後	0.105 **	0.105 **	0.100 **	0.073 **	0.099 **	0.063 **
合意インデックス	場前						
	場中	0.031 **	0.032 **	0.033 **		0.021 *	
	場後	0.029 **	0.029 **	0.028 **	0.023 **		

\*p<.05 \*\*p<.01 なお、有意でない結果は省略している

### (5) 合意インデックス

合意インデックスは、時間間隔における強気意見と弱気意見の拮抗状態を測るための掲示板指標であり定義は以下である。この指標は強気投稿数と弱気投稿数が同数の場合は0、どちらかに意見が偏ると1に近づく。なお、強気投稿数と弱気投稿数の合計が3未満の場合は、合意インデックスが判断できないと考え、強気指数と同様に除外した。

$$\frac{\text{強気投稿数} - \text{弱気投稿数}}{\text{強気投稿数} + \text{弱気投稿数}}$$

## 5 分析結果

50社を分析対象とし、2005年1月1日から2006年12月31日における取引日のデータを用いて、株式指標と掲示板指標の相関分析を行った。その結果を表4、5及び6に示す。

## 6 分析結果の考察

本節では、3節の仮説検証を行う。分析結果に基づき、掲示板指標が株式指標の先行指標、一致指標、遅行指標のいずれに該当するのか考察を行う。先行指標は、株式指標より以前の掲示板指標（当日場前

の掲示板と当日の株式指標、および当日の掲示板と翌日の株式指標）が有意なものと定義する。一致指標は、当日場中の掲示板指標と当日場中の株式指標が有意となるものと定義する。遅行指標は、株式指標より以降の掲示板指標（当日の掲示板と前日の株式指標、および当日場後の掲示板と当日の株式指標）が有意なものと定義する。

### 6.1 投稿数と株価リターンとの関係(仮説1)

先行指標として投稿数を見ると、場前、場中の投稿数と翌日の株式リターン及び超過リターンとの間には、マイナスの相関で5%有意な関係が見られる。つまり投稿数は僅かだが株価下落を予測している可能性を示す結果が出ている。これは仮説1を棄却する検証結果となっている。なお、遅行指標としては、当日のリターン、及び超過リターンと場後の投稿数は、マイナスで1%有意である。つまり、株価が下落した場合には、投稿者は株価の動きに反応して掲示板に投稿することが多い可能性を示している。

強気投稿数は、場中の強気投稿数が前日と当日のリターンとプラスの相関で1%有意であった。つまり、強気投稿数はリターンに対して遅行指標および一致指標となっている。これは、強気投稿数は、前日や現在の株価の上昇を反映して、投稿されていることが多い可能性を示している。

一方、弱気投稿数は、場前の投稿数と翌日のリターンを除き、全て有意なマイナスの相関が示されている。特に、場後の弱気投稿数が先行指標として、翌日リターンと1%有意な結果となっている。これは、当日場後の弱気意見が翌日の株価下落と関連している可能性を示している。本稿の分析期間では、大きな下げ相場はなく、強気意見の方が長い期間であった。そのような状態では、数少ない弱気意見の投稿の方が、個人投資家の投資行動に与える影響が大きい可能性があると考えられる。

## 6.2 投稿数と出来高との関係(仮説 2)

出来高と投稿数、強気投稿数、弱気投稿数の間には 1%有意な正の相関が示されている。この結果は仮説 2 を支持する。分析結果より投稿数、強気投稿数、弱気投稿数は出来高の遅行指標、一致指標及び先行指標であり、掲示板の投稿数は、投資家のその銘柄への関心の高さを示していると考えられる。つまり、投稿数はインターネットを利用する個人投資家の、その銘柄に対する関心度の代理変数として考えられる。

## 6.3 投稿数とボラティリティとの関係(仮説 3)

ボラティリティ及び絶対値リターンと、投稿数、強気投稿数、弱気投稿数との間には、有意な正の相関が示されている。この結果は仮説 3 を支持する。分析結果より投稿数、強気投稿数、弱気投稿数はボラティリティ及び絶対値リターンの遅行指標、一致指標及び先行指標であり、掲示板の投稿数は、Antweiler and Frank[2004]の結果と同様に取引日当日のボラティリティを予測している可能性がある。この結果は、日本の株式市場においても、株式掲示板の投稿活動が、投資家の投資行動に影響を与えている可能性があることを示唆していると考えられる。

## 6.4 強気比率と株価リターンとの関係(仮説 4)

株価リターン及び超過株価リターンと強気比率は、場前強気比率と翌日のリターンを除き、全て有意な正の相関が示されている。これは仮説 4 を支持する結果となっている。特に、場後の強気比率が翌日リターンと 1%有意となっていることから、翌日の株価に反映している可能性を示している。この結果は、前日夕方から夜間の掲示板の投稿内容が、翌日の投資家の売買行動に影響を与えている可能性を示唆していると考えられる。

## 6.5 強気比率と出来高との関係(仮説 5)

出来高と強気比率の間には、正の 1%有意な相関が示されている。この結果は仮説 5 を支持する。分析結果より強気比率は出来高の先行指標、一致指標及び遅行指標であることを示している。分析期間の 2005 年から 2006 年は、上昇相場またはボックス相場であり、下降相場ではなかった。このため、強気比率が高いと、掲示板を見た投資家の注目を集めることとなり、それが投資行動を引き起こし、出来高の増加と関係したと考えられる。

## 6.6 強気比率とボラティリティとの関係(仮説 6)

ボラティリティ及び絶対値リターンと、強気比率

の間には、正の 1%有意な相関が示されている。この結果は仮説 6 を支持している。分析結果より強気比率はボラティリティ及び絶対値リターンの先行指標、一致指標及び遅行指標であることを示している。本稿の分析期間では、上昇相場またはボックス相場であり、下降相場ではなかった。このため、強気比率が高い場合は、それが投資行動を引き起こし、ボラティリティの増加に関係したと考えられる。

## 6.7 合意インデックスと株価リターン・出来高・ボラティリティとの関係(仮説 7, 8, 9)

仮説 7 の検証結果では、株価リターン、超過リターンと合意インデックスの関係は、正の相関があり先行指標、一致指標及び遅行指標である。これは仮説 7 と反する。

仮説 8 の検証では、出来高と合意インデックスの間には、場中、場後で 1%有意な正の相関が示されている。しかしながら、Antweiler and Frank[2004]は、合意インデックスが出来高と負の関係にあることを報告している。つまり、米国の株式市場では、強気/弱気どちらか一方に意見が偏ると、出来高が減少する傾向があることを示している。一方、本稿の結果は、強気または弱気どちらか一方に意見が偏ると、出来高は増加する傾向があることを示している。これは仮説 8 と反する。

仮説 9 の検証では、ボラティリティと合意インデックスの間には、場中及場後とは 1%有意な正の相関が示されている。Antweiler and Frank[2004]による米国の先行研究では、意見がどちらかに偏るのではなく、拮抗するとボラティリティが増加するという結果が出ている。これは、米国では意見の拮抗が投資行動を引き起こすことを示している。つまり仮説 9 は成立していない。

投稿データはその日の場前、場中、場後の値であるが、ボラティリティは前後 5 日間の株価変動(標準偏差)であり、必ずしも期間の対応が取れているとは言えない。そこで、当日の変動に限定するために、当日の株式変動である絶対値リターンとの相関を表 6 右欄に示す。その結果でも、場中と当日絶対値リターンで正の相関があり 5%有意であり、仮説 9 は成立していない。

合意インデックスに関する仮説 7, 8, 9 は、Antweiler and Frank[2004]の結果と異なり成立しなかった。この点に関して考察すると、本稿の分析期間である 2006 年から 2007 年では、日本の株式市場は上昇基調であった。つまり、上昇相場による擬似相関である可能性が考えられる。合意インデックスの検証には下降局面を含めた分析が必要であると考えられる。

下降局面を含めた合意インデックスの検証は、今後の検討課題とする。

## 7 結論と今後の展開

本稿では、日本の Yahoo!株式掲示板における投稿数が多い東証一部 50 社を対象に、インターネット株式掲示板と株式指標の関係を、米国の先行研究の結果と比較して、検証を行った。分析結果より、掲示板の投稿数だけからでは、1%有意の水準は得られず、先行指標として将来の株価リターンに関する情報を得ることは難しいと考えられる。しかし、投稿内容を考慮して分析すると、前日場後（15 時–24 時）に投稿された弱気投稿数が多いと、翌日の株価リターンがマイナスとなる傾向が示された。本稿の分析期間は、2005/01/01 - 2006/12/31 で上昇相場またはボックス相場であり、下降相場が含まれていない。下降相場の時期も含めた場合の弱気投稿数と株価リターンの分析は、今後の検討課題である。

投稿されたメッセージの強気比率や偏りは、株式のリターンに反映している可能性がある。また、投稿数および強気比率は、ボラティリティと出来高の先行指標の可能性があると考えられる。

今回の分析においては、本稿の分析対象期間および対象銘柄の範囲では、合意インデックスの効果は検証できなかった。下降相場の期間を含めた分析が必要である。これは今後の検討課題としたい。また、投稿量の多い上位 50 社を分析対象とした。投稿数の少ない企業と株式指標の関係は、本稿では分析は今後の課題である。さらに、本稿では個社毎の分析は行っていないが、個社によって掲示板との関係は異なる可能性がある。今後、個社毎の分析が必要と考える。

本稿では、中立の分類精度が低く、分析から除外した。分析精度の向上には、Das and Chen[2007]の 5 種類の自然言語処理手法を組み合わせる方法などが提案されている。また、本稿の分析では単語の組み合わせである文脈に関しては考慮がなされていない。これらは今後の検討課題である。

## 参考文献

- [1] Antweiler, W. and Frank, M. Z., “Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards,” *Journal of Finance*, Vol. 59, No. 3, pp.1259-1294, 2004.
- [2] Cristianini, N. and Shawe-Talor, J., “An Introduction to Support Vector Machines and other kernel-based learning methods”, Cambridge University Press, 2000. (大北剛訳, サポートベクターマシン入門, 共立出版, 2005.)
- [3] Das, S. R., Chen, M. Y., “Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web”, *Management Science*, Vol.53, No.9, pp. 1375–1388, 2007.
- [4] Harris, M. and Raviv, A., “Differences of Opinion Make a Horse Race,” *Review of Financial Studies*, Vol. 6, pp. 473-506, 1993.
- [5] Tumarkin, R. and Whitelaw, R.F., “News or Noise? Internet Postings and Stock Prices,” *Financial Analysts Journal*, Vol.57, pp.41-51, 2001.
- [6] Van Bommel, J., “Rumors,” *Journal of Finance*, Vol. 58, No. 4, pp.1499-1519, 2003.
- [7] Wyszocki, P. D., “Cheap Talk on the Web: The Determinants of Postings on Stock Message Boards,” *Working paper, University of Michigan*, [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=160170](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=160170), 1999.
- [8] アスキー, BLOG 上場企業ランキング 500, Dec. 2006.
- [9] 岡野原大輔, 辻井潤一, 評価文に対する二極指標の自動付与, 言語処理学会第 11 回年次大会発表論文集, pp. 664-667, 2005.
- [10] 高村大也, 乾孝司, 奥村学, スピンモデルによる単語の感情極性抽出, 情報処理学会論文誌ジャーナル, Vol.47, No.02, pp. 627-637, 2006.
- [11] 東洋経済データバンク, 株価 CD-ROM 2006, 東洋経済新報社, 2006.
- [12] 日本経済新聞, 一目均衡, Nov. 28. 2006.
- [13] 山下一雄, 石上隆達, 佐藤 哲也, インターネット掲示板にみる社会的関心と株価変動の関係, 日本社会情報学会第 20 回全国大会研究発表論文集, pp.237-240, 2005.
- [14] 丸山健, 梅原英一, 諏訪博彦, 太田敏澄, インターネット株式掲示板の投稿内容と株式市場の関係, 証券アナリストジャーナル, 第 46 巻第 11・12 号, pp.110-127, 2008.