

日中足株価を用いた株式掲示板と株式市場の関係：SoftBank の事例 The Relation of Intraday stock price and Internet BBS: The Case of SoftBank

梅原 英一[†] 諏訪博彦[‡]
Eiichi Umehara Hirohiko Suwa

1. はじめに

近年の東証レッドアローシステムの稼働などで、高頻度データである日中株価（ティックデータ）を利用した分析が必要になってくると考えられる。日中株価を利用した分析では、経済物理学における株価分析などがあげられる。しかしながら、経済物理学（増川他(2011)）による株価の説明は、株価自体を説明変数としている研究が多い。つまり株価リターン（純資産時価総額比率の高ー低）の分布変化を、過去の株価の動きで説明するアプローチである。一方で、ファイナンスの分野では、マルチファクターモデルに見られるように、株価リターンを別の独立な変数で説明するモデルが提示されている。例えば、Fama and French (1996)は、CAPM に SMB ファクター（時価総額ファクター：大型株－小型株）と HML ファクター（純資産時価総額比率の高ー低）の 2 種類のファクターを追加した 3 ファクターモデルを提案している。しかしながら、日中足株価（ティックデータ）を分析する場合、これに対応する高頻度の説明データは、一般的な企業財務データなどには存在しない。なぜなら、これらのデータは企業決算に合わせて公表されるので、四半期に一度のデータとなる。日中足株価を説明する変数としては不十分である。一方、近年、ソーシャルメディアが発展している。その中にインターネット株式掲示板（以下、株式掲示板）などがある。これは、他の投資家の生の声を直接知ることができるツールである。株式掲示板と株式市場との関連性に焦点をあてた研究としては、Antweiler and Frank(2004)がある。日本における株式掲示板と株式市場との関連性に焦点をあてた研究としては、丸山ら(2008)、諏訪ら(2012)の研究がある。これらの研究によると、株式掲示板の投稿数や投稿内容は、株式市場のボラティリティや出来高を説明すること、手数料等を考慮した場合に経済的な利得を得ることは難しいが、オーバーナイトの投稿内容が翌日のリターンと関係する可能性があることなどの示唆が得られている。そこで本研究では、日中足株価と投稿メッセージの関係をソフトバンク社に関して分析することを目的とする。

本論文の構成は以下である。2 節で先行研究をレビューし、3 節で分析方法について述べる。4 節で分析データについて述べ、5 節で自然言語処理と機械学習を用いた投稿の分類方法と、分類データに基づく強気指数の算出方法について述べる。6 節で分析結果を、7 節は結論である。

2. 先行研究

Wyscocki(1999)は、1998 年 1 月から 8 月の間に Yahoo! 掲示板に投稿された最も投稿数の多い 50 銘柄の投稿数と企業特性や株式市場の活動との関連を時系列で調べた。その結果、前日夜の投稿数が翌日リターンと統計的に有意であ

るが、手数料を考慮に入れた場合、経済的利益を得ることは難しいと述べている。Tumarkin and Whitelaw(2001)は、インターネット関連株 73 社の RagingBull.com の 1999 年 4 月 17 日から 2000 年 2 月 18 日までの 181,133 件のメッセージに関して、投稿数や投稿内容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で検証した。その結果、掲示板のメッセージはリターンを予測しないと述べている。Jones(2006)は、S&P100 企業に対し Yahoo! Finance の投稿前後での株式リターンが変化するかどうか調査した。その結果、投稿後の日次出来高で有意な増加が見られ、日次リターンが投稿後に有意に低下し、日次リターンの変動性が有意に増加することを発見した。この結果は、市場や産業レベルの出来事をコントロールした後も成立したと述べている。

Antweiler and Frank(2004)は、ダウ・ジョーンズ工業株指数と同インターネット指数組み入れ 45 社の 150 万件以上の Yahoo! と Raging Bull のメッセージの内容を、Naïve Bayesian 法による機械学習を用いて分析した。その結果、掲示板はリターンを予測しない、強気と弱気が拮抗するとそれが取引を誘発する結果、掲示板が出来高と当日のボラティリティを説明することを示した。また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。Das and Chen(2007)は、5 種類の自然言語処理アルゴリズムによる多数決投票で、分類で生じる誤差が改善することを示した。またモルガンスタンレーハイテク指数の技術セクター株 24 社を対象とした 2001 年 7 月－8 月の 145,110 メッセージに適用した分析結果は、Antweiler and Frank(2004)と同様の結果を得ている。

丸山ら(2008)は、東証 1 部上場企業について、日本の Yahoo! ファイナンス掲示板の投稿数上位 50 社（2005-2006 年）の企業を対象に掲示板指標と株式指標の関係を実証的に検証した。彼らは、掲示板より取得した 110 万件以上の投稿データを、自然言語処理と機械学習（Support Vector Regression）を用いて、「弱気」「中立」「強気」の 3 種類に分類し、強気投稿数と弱気投稿数により作成した強気指数を開発した。その結果、第一に掲示板の投稿数だけでは株式リターンを予測する事は難しいが、弱気投稿数が多いと翌日の株式リターンがマイナスとなる傾向があること、第二に投稿数がボラティリティ、出来高と関連している事、第三に強気指数が株式リターンやボラティリティ、出来高と関連している可能性があることを報告している。

Tetlock(2007)はウォールストリートジャーナルの日次コラムの内容の悲観の程度を自然言語処理で測定した。メディア悲観の程度が高いと、一時的には株価は下落するが 1 週間以内に元の水準に戻ることを発見した。また、悲観の程度の絶対値が高いと出来高が大きくなることも発見した。さらに、メディアの悲観の程度に基づく売買シミュレーシ

[†] 東京都市大学 Tokyo City University

[‡] 電気通信大学 The University of Electro-Communications

ョンを行い、僅かではあるが超過収益が得られることを確認した。

Gilbert and Karahalios(2009)は、ブログから推定した Sentiment が将来の株価に関する市場データにはまだ出てない新しい情報であることを示した。彼らは LiveJournal に投稿された 2000 万以上のデータから不安・心配・恐れを推定した。その結果、不安表現の増加が、S&P500 指数が下落することを予測することを示した。Zhang et al.(2011)は Twitter の投稿を解析することにより、株式市場指標を予測しようと試みた。彼らは日次での集団的希望と恐れを測定し、株式市場指標との相関関係を分析した。その結果、感情的な Tweet の割合がダウ、ナスダックと S&P500 との相関が有意に負であったが、VIX 指数は有意な正の相関であったことを発見した。

3. 仮説

本研究では、丸山他(2009)らの先行研究の結果から仮説を設定する。我々は、以下の 7 つの仮説を設定し、日中足株価データに関して成立するか否かを検証する。

仮説 1：投稿数は株価リターンの一致・遅行指標である。遅行指標としては、当日のリターン、および超過リターンは、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。

仮説 2：弱気投稿数は先行指標として投稿数は、わずかながら株価下落を予測する。

丸山(他)はオーバーナイトの弱気投稿数は翌日の株価下落と相関することを示した。

仮説 3：投稿数は出来高の先行指標・一致指標・遅行指標である。

仮説 4：投稿数はボラティリティの遅行指標、一致指標および先行指標である。

日次で見た場合には、投稿数は出来高およびボラティリティと相関があることを観測されている。

仮説 5：強気指数は、株価リターンの先行指標、一致指標および先行指標である。

丸山他(2009)は、日次株価データでは強気指数が株価リターンと相関のあることを発見した。特に、場引け後の強気比率は翌日のリターンと 1% 有意で相関していることを発見した。

仮説 6：強気指数は、出来高の先行指標、一致指標および先行指標である。

仮説 7：強気指数はボラティリティの先行指標、一致指標および先行指標である。

彼らは、投稿数と同様に強気指数は出来高およびボラティリティと正の相関を観測した。

4. 分析データ

4.1 分析期間

2008 年 12 月の Yahoo! 株式掲示板¹のメッセージを用いた。分析対象は、東証一部上場銘柄である SoftBank 社 (9984) である。この期間のソフトバンクの株価動向は、12 月 1 日寄値が 1305 円、12 月 30 日引値が 1603 円で 22.8% 値上が

りした。しかし市場全体ではリーマンショック後であり、対象期間における株式相場は東証株価指数 (TOPIX) の 12 月 1 日始値が 832.62、12 月 30 日引値は 859.24 で +3.2% の安値圏のボックス相場であった。

4.2 株式データ

株式データは、東京証券取引所が提供している歩み値データレベル II を使用した。このデータには、約定時分、価格、出来高が約定単位で収録されている。超過リターンを算出するために東証株価指数 (TOPIX) の歩み値データを使用した。我々は、日中の立会時間を 30 分ごとに区切りタイムウインドウを設定した。このタイムウインドウは、前場 4 ウインドウ (9:01-9:30、9:31-10:00、10:01-10:30、10:31-11:00) および後場 5 ウインドウ (12:31-13:00、13:01-13:30、13:31-14:00、14:01-14:30、14:31-15:00) である。なお、前場および後場の Open Price(始値)および出来高は板寄せで値付けされるので、通常のザラバ取引とは取引所における取引手法が異なる。そのために、この時点の出来高はザラバよりも多くなることが知られている。この異常値を避けるために、我々は前場、後場の取引開始 1 分間のデータは無視した。故にタイムウインドウには 9:00 および 12:30 の 1 分間のデータは含まれない。また、寄値の時刻が 9 時より遅れることもある。例えば、寄り時刻が 9:08 の場合である。この場合、タイムウインドウは取引開始 1 分間のデータは無視して、9:09-9:30 と設定した。

タイムウインドウを 30 分とした理由は、Yahoo! 掲示板の投稿数がある程度取得できる時間間隔として設定した。これより短いタイムウインドウ (たとえば 15 分) では、投稿量が全くない場合はあった。

リターンは 30 分間の株価の変化で計測した。例えば、9:01-9:30 のタイムウインドウのリターンは、9:30 の最後の株価の自然対数から 9:00 の最後の株価の自然対数を引いたものである。超過リターンを計測するために市場インデックスとして東証株価指数を使用した。東証株価指数は、前場では 9:01-11:00、後場は 13:31-15:00 まで 15 秒間隔で公表されている。我々は、TOPIX の 9 時 1 分 0 秒のデータを 9 時 0 分の最終価格、11 時 0 分 45 秒のデータを 11:00 の最終株価と対応させた。

出来高は各タイムウインドウ内の取引ごとの出来高の集計である。ボラティリティは各タイムウインドウの 1 分間リターンの標準偏差を用いた。Skew および Kurt も 1 分間リターンで計測した値である。

4.3 Yahoo! 株式掲示板の概況

自然言語処理の学習データとしては 2003-2008 年の東証一部銘柄の取得できた総投稿は 3,891,158 件を用いた。この期間で 1 番投稿が多い銘柄はソフトバンクであり、月間平均投稿数は 9364 投稿であり、2 位の SBI ホールディングス社の 2864 投稿の約 3 倍の投稿数があった。投稿は一部銘柄に片寄っており、非常にアクティブな掲示板とそうではない掲示板がある。このために、我々は分析対象としてソフトバンクを選択した。なお、分析対象期間 (2008 年 12 月 1 日 9:00-2008 年 12 月 30 日 11:00) のソフトバンクの総投稿数は 6,108 件であった。

掲示板指標は各タイムウインドウごとに算出した。また、タイムウインドウ以外の投稿メッセージは無視した。前場

¹ Yahoo! 株式掲示板, http://messages.yahoo.co.jp/yahoo/Business_Finance/Investments/Stocks/index.html.

と後場の間 (11:01-12:30)、およびオーバーナイト(15:01-翌日 9:00)のデータは対象外とした。先行研究では、オーバーナイトの投稿メッセージが翌日の株価と相関があることが観測されている。日中足データとオーバーナイト投稿および前場・後場間の投稿の関係に関する分析は今後の研究課題とする。

5. 強気指数の算出

5.1 掲示板投稿データ収集および強気・弱気分類

我々は、Yahoo!掲示板から投稿データを収集し、自然言語処理と機械学習の手法を用いて、投稿データを「弱気」「中立」「強気」の三種類に分類する。分類方法として、丸山ら(2008)が提案した手法を用いる。彼らの手法は、1. 形態素解析・ノイズ除去処理、2. 特徴ベクトルの算出、3. SVR による分類の 3 ステップで構成されている。

5.1.1 形態素解析・ノイズ除去処理

インターネット掲示板から収集した投稿メッセージは、英語等と違い日本語の文章が単語間に分かれていない。このため、形態素解析プログラム (MeCab²) 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の単語群を抽出する。

5.1.2 分類に有用な単語の抽出と特徴ベクトルの算出

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抽出する。SONY の 2005 年の投稿データを形態素解析を行い 22,107 単語を抽出し、目視で意味のある単語を 8,984 単語から、TF・IDF 値上位 5000 単語を抽出した。これに、単語感情極性対応表 (高木他, 2006) の 55,125 単語を加え、SONY の 2005 年投稿メッセージで TF・IDF>10.5 となる単語を抽出した。その結果、総単語数 6989 単語の特徴ベクトルを作成した。特徴ベクトルの個々の値は単語の重要度である。これは(1)式に示すとおり TF・IDF 法に基づく投稿 t 中の単語 d の重要度 $w(t,d)$ の算出を行った。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した。各投稿の特徴ベクトルは、6,989 次元のベクトルである。(1)式により算出する。

$$w(t,d) = \ln(tf_{t,d} + 1) \cdot \ln(N/df_d) \quad (1)$$

N : すべての投稿数

$Td_{t,d}$: 投稿 t に単語 d が出現する頻度

Df_d : 単語 d が出現する投稿の数

単語ごとに算出された重要度 w を用いて投稿の特徴ベクトル f_t を以下の定義より算出する。

$$f_t = (w(t,1), w(t,2), \dots, w(t,l))$$

l : ベクトル総単語数 (6989 ベクトル)

投稿ごとに算出された 6,989 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。

表 1 各気持ちの投稿数と極性値の平均・分散

気持ち	N	平均	分散
強く買いたい・買いたい	417,301	0.744	0.317
様子見	235,671	0.434	0.543
売りたい・強く売りたい	171,475	-0.297	0.612

表 2 SVMの学習データ分類精度

		分類結果			合計	精度
		弱気	中立	強気		
投稿者の気持ち	強く売りたい	8.2%	6.9%	2.0%	17.1%	41.3%
	売りたい	0.4%	2.8%	0.5%	3.7%	
	中立	1.6%	11.7%	15.3%	28.6%	
	買いたい	0.1%	4.7%	9.1%	13.9%	
	強く買いたい	0.2%	3.9%	32.6%	36.7%	
合計		10.5%	30.1%	59.4%	100.0%	
精度		81.6%	38.9%	70.1%	61.9%	

5.1.3 SVR による分類

掲示板の投稿を Support Vector Regression (以下 SVR と記す)を用いて、「強気」「中立」「弱気」の三種類に分類する。SVR の分類プログラムは LibSVM³ を用いる。

4.3 節で述べた取得データ (3,891,158 件) 中で、Yahoo! 掲示板に付加されている機能により「投稿者の気持ち」が判別できる投稿は 824,598 件であった。これを学習データとして用いる。「投稿者の気持ち」は、投稿者が投稿毎に「強く買いたい」「買いたい」「様子見」「売りたい」「強く売りたい」「公表しない」の 6 個の選択の中から任意のものの選び、それを公開する Yahoo! 掲示板の機能である。SVR 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。つまり投稿の特徴ベクトルが強気を示す場合は正、弱気を示す場合は負、様子見の場合は 0 への回帰が期待される学習である。

表 1 に学習データの気持ち別投稿数および極性値の平均・分散を示す。出力値を投稿の「極性値」と呼ぶ。「強く買いたい」「買いたい」の投稿者の気持ちが公開されている総投稿数は、417,301 件で全学習データの 50.6% である。平均極性値は、0.744 である。「様子見」の投稿者の気持ちが公開されている総投稿数は、235,671 件で全学習データの 28.6% である。平均極性値は、0.434 である。「強く売りたい」「売りたい」の投稿者の気持ちが公開されている総投稿数は、171,475 件で全学習データの 20.8% である。平均極性値は、-0.297 である。

投稿の分類において、メッセージの極性値が 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」投稿として分類した。表 2 に、SVM の学習データ分類精度を示す。「強く売りたい」「売りたい」の気持ちが公開されている投稿が弱気に分類される精度は 81.6%、「強く買いたい」「買いたい」の投稿が強気に分類される精度は 70.1% であった。

³ SVR 使用時のカーネルは線形カーネルを利用。パラメータは C 値を 0.01 に設定した。(Chih-Chung Chang and Chih-Jen Lin, LibSVM, <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>)

² MeCab, <http://mecab.sourceforge.net/>.

この学習データをもとに、気持ちを公表してない投稿を含めた 2008 年 12 月のソフトバンクの投稿を分類した結果を表 3 に示す。投稿総数は 6238 件（12 月 1 日 0 時 10 分～12 月 31 日 23 時 47 分）であった。強気に分類された投稿が 714 件、弱気に分類された投稿が 168 件であり、強気の投稿が多くなっている。

表 3 2012 年 12 月ソフトバンク分類結果

	弱気	中立	強気	合計
売りたい 強く売りたい	90	189	93	372
様子見	14	86	93	193
買いたい 強く買いたい	34	262	417	713
記載なし	30	4,819	111	4,960
合計	168	5,356	714	6,238

5.2 強気指数・強気比率

我々は、Yahoo! 掲示板から投稿データを収集し、自然言語処理と機械強気指数とは、タイムウインドウ毎における強気/弱気意見の多少を示す掲示板指標であり、(2)式で定義する。

$$\text{強気指数}(t) = \ln \left(\frac{1 + \text{強気投稿数}(t)}{1 + \text{弱気投稿数}(t)} \right) \quad (2)$$

この指標は、強気投稿数が弱気投稿数に比べて多くなると正の値となり、同数の場合は 0、強気投稿数が弱気投稿数より少なくなると負の値となる。

強気指数は中立に分類された投稿を無視している。しかしながら、日中のタイムウインドウで見ると、中立に分類された投稿数が多かった。それを反映するために中立を含めた強気比率を(3)式に定義する。

$$\text{強気比率}(t) = \ln \left(\frac{1 + \text{強気投稿数}(t)}{1 + \text{弱気投稿数}(t) + \text{その他投稿数}(t)} \right) \quad (3)$$

6. 分析結果

6.1 投稿数の検定結果

投稿数に関する検定結果を表 4 に示す。30 分タイムウインドウの結果はリターンボラティリティともの投稿数は時点が一致および遅行指標であった。出来高だけは先行指標でも有意であった。この結果は、日足株価での先行研究の結果と異なり、投稿数はリターンもボラティリティも予測してなかった。仮説 1（投稿数は株価リターンの一致・遅行指標である）、および仮説 3（投稿数は出来高の先行指標・一致指標・遅行指標である）は支持された。しかし、仮説 4（投稿数はボラティリティの遅行指標、一致指標および先行指標である）のうち、先行指標に関しては支持されなかった。つまり、分単位の掲示板への投稿活動は、主に相場の状況を反映した結果であると結論づけられる。なお、分布の形状（3 次モーメント（Skew）・4 次モーメント（Kurt））に関しては有意ではなかった。

表 4 投稿数

	リターン				出来高	超過リターン			
	リターン	Vol	Skew	Kurt		リターン	Vol	Skew	Kurt
一致	0.179 **	0.189 **	0.005	-0.055	0.310 **	0.183 *	0.162 *	0.022	-0.024
先行指標	0.100	0.066	0.060	-0.080	0.194 **	0.087	0.077	0.040	-0.044
遅行指標	0.203 **	0.211 **	0.069	-0.034	0.351 **	0.253 **	0.222 **	0.032	0.085

** 1%有意

* 5%有意

6.2 強気投稿数の検定結果

先行研究における日次株価の分析では、弱気投稿と株式市場の関係において有意な結果が得られていた。そこで日中足株価で同様な結果が得られるか否かを検証した。その結果を表 5 に示す。結果はどれも有意な結果が得られていない。仮説 2（弱気投稿数は先行指標として投稿数は、わずかながら株価下落を予測する）は棄却された。この原因の一つに弱気に分類された投稿メッセージがきわめて少なかったことがあげられる。今後、期間や対象銘柄を拡張して、サンプルサイズを増やした検証を行う必要がある。これは研究課題とする。

表 5 強気投稿数

	リターン				出来高	超過リターン			
	リターン	Vol	Skew	Kurt		リターン	Vol	Skew	Kurt
一致	0.195 **	0.125 **	0.028	-0.036	0.285 **	0.190 **	0.110	0.062	-0.009
先行指標	0.071	0.034	0.039	-0.032	0.233 **	0.058	0.076	0.050	0.025
遅行指標	0.230 **	0.190 **	0.072	-0.047	0.318 **	0.252 **	0.195 **	0.025	0.033

** 1%有意

* 5%有意

6.3 弱気投稿数の検定結果

先行研究における日次株価の分析では、弱気投稿と株式市場の関係において有意な結果が得られていた。そこで日中足株価で同様な結果が得られるか否かを検証した。その結果を表 6 に示す。結果はどれも有意な結果が得られていない。仮説 2（弱気投稿数は先行指標として投稿数は、わずかながら株価下落を予測する）は棄却された。この原因の一つに弱気に分類された投稿メッセージがきわめて少なかったことがあげられる。今後、期間や対象銘柄を拡張して、サンプルサイズを増やした検証を行う必要がある。これは研究課題とする。

表 6 弱気投稿数

	リターン				出来高	超過リターン			
	リターン	Vol	Skew	Kurt		リターン	Vol	Skew	Kurt
一致	-0.057	0.005	-0.018	-0.061	0.037	-0.076	0.010	0.003	-0.081
先行指標	0.084	-0.010	0.042	-0.100	0.009	0.093	-0.019	0.045	-0.137 *
遅行指標	-0.028	0.071	0.050	-0.005	0.141 *	0.047	0.084	0.022	-0.015

** 1%有意

* 5%有意

6.4 強気指数・強気比率の検定結果

強気指数の結果を表7に示す。強気指数に関しては、リターン・超過リターンの一致指標・遅行指標および、出来高の先行・一致・遅行指標であった。この結果、仮説5（強気比率は、株価リターンの先行指標、一致指標および遅行指標である）のうち、先行指標は棄却された。また、仮説6（強気指数は、出来高の先行指標、一致指標および遅行指標である）は支持された。ボラティリティに関しては、超過リターンのボラティリティが先行指標として有意であった。仮説7（強気指数はボラティリティの先行指標、一致指標および遅行指標である）は、超過リターンの先行指標を除き支持されなかった。

なお、中立の分類を含めた強気比率は、すべてにおいて有意な結果は得られなかった（表8）。

表7 強気指数

	リターン				出来高	超過リターン			
	リターン	Vol	Skew	Kurt		リターン	Vol	Skew	Kurt
一致	0.187**	0.119**	0.041	0.022	0.209**	0.184**	0.109	0.045	0.032
先行指標	0.037	0.126	0.028	-0.026	0.199**	0.023	0.159*	0.038	0.110
遅行指標	0.220**	0.113**	-0.018	-0.065	0.163**	0.172**	0.107	-0.023	0.003

** 1%有意

* 5%有意

表8 強気比率

	リターン				出来高	超過リターン			
	リターン	Vol	Skew	Kurt		リターン	Vol	Skew	Kurt
一致	0.085	-0.044	0.034	-0.028	0.104	0.080	-0.022	0.062	-0.073
先行指標	0.027	0.014	0.088	-0.040	0.098	0.016	0.055	0.049	0.112
遅行指標	0.117	0.093	0.000	-0.128	0.109	0.094	0.079	0.016	-0.104

** 1%有意

* 5%有意

7. 結論

本研究では、Yahoo!掲示板の投稿活動とソフトバンクの株式リターン、出来高、ボラティリティ、Skew、Kurtとの関係を調査した。その結果、投稿数・強気投稿数・強気指数はリターンの一致指標・遅行指標、投稿数はボラティリティの一致指標・遅行指標、投稿数・強気投稿数・強気指数は出来高の一致指標・遅行指標・先行指標という関係が観測された。

日中足株価で見た場合、先行指標としては、投稿活動は出来高以外は有意な結果ではなかった。これは2008年12月の一か月だけのソフトバンク1社の結果であった。2008年12月はリーマンショック後で、相場環境も株価があまり動かないボックス相場であった。今後、分析期間や対象銘柄を増やした検証が必要であると考えている。

また、分析をYahoo!掲示板で行ったが、近年はTwitterなどに非常に多くの投稿がある。分析対象を増やすためには、Twitterのメッセージで検証することも有効である可能性がある。これも今後の研究課題である。

なお、統計的には有意ではないが、弱気投稿数および強気比率（＝強気投稿/総投稿）は、超過リターンの4次モ

ーメント（Kurt）の先行指標と相関が高い。つまり、株価のロングテールと関係があるかもしれない。投稿活動とリターンの分布（正規分布からのかい離）に関するテーマも今後の研究課題である。

謝辞

本研究の一部は、科学技術研究費助成事業（基盤C（25380481））の成果の一部である。

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 (2004).
- [2] Das, S. R., Chen, M. Y., "Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web", *Management Science*, Vol.53, No.9 (2007).
- [3] Fama, E. F. and French, K. R., "Multifactor Explanations of Asset Pricing Anomalies," *Journal of Finance*, Vol. LI, No. 1 (1996).
- [4] Gilbert, E., Karahalios, K., "Widespread Worry and the Stock Market", *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media* (2009).
- [5] JONES, A., L., "Have internet message boards changed market behavior?," *The journal of policy, regulation and strategy for telecommunications*, Vol.8, No.5 (2006).
- [6] Tetlock, C. P., "Giving Content to Investor Sentiment: The Role of Media in the Stock Market", *Journal of Finance*, Vol. 62, No. 3 (2007).
- [7] Tumarkin, R. and Whitelaw, R.F., "News or Noise? Internet Postings and Stock Prices," *Financial Analysts Journal*, Vol.57 (2001).
- [8] 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 (1999).
- [9] Zhang, X., Fuehres, H., Peter A. Gloor, A., P., "Predicting Stock Market Indicators Through Twitter "I hope it is not as bad as I fear"", *Procedia - Social and Behavioral Sciences* (2011).
- [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 (2006).
- [11] 増川純一, 水野貴之, 村井淨信, 尹熙元, "株価の経済物理学", 培風館 (2011).
- [12] 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*, Vol.46 No.11・12 (2008).
- [13] 諏訪博彦, 梅原英一, 太田敏澄, "ファクターモデルによるインターネット株式掲示板の投稿と株式リターンの分析", *情報処理学会論文誌*, 53巻1号 (2012).
- [14] 諏訪博彦, 梅原英一, 太田敏澄, "インターネット株式掲示板の投稿内容分析に基づくファクターモデル構築の可能性", *人工知能学会論文誌*, Vol.27, No.6 (2012).