

電子掲示板ネットワークにおける期待形成と投資収益率

Network analysis on stock BBS: Extract excess return from BBS postings

末木将史

Masachika SUEKI

明星大学経済学部(非)

Department of Economics, Meisei University

Abstract: This study attempts to uncover underlying information in stock message board (hereafter BBS) by using network analysis. It is said that many of postings on BBS could be noise. Therefore the overall sentiment of BBS often carries little useful content for future stock investment return. Moreover, under efficient market hypothesis (EMH), it is unlikely that investors (hereafter poster(s)) disseminate valuable information without compensation. However, some empirical research in the United States shows that there are a few posters in the community who post valuable information on BBS. The problem is how to extract such informed posters systematically from the BBS community. In this study, I utilize network analysis to solve this problem.

Results of the empirical study on Yahoo! Stock BBS show that neither number of posting nor degree centrality could extract informed poster. However clique could extract informed poster. Return of informed posters is both statistically and economically significant even after risk adjusted. This indicates that the network analysis approach is valuable in screening out noise in BBS posting.

1.はじめに

直接金融を通じたファイナンスの進展が望まれる今日において、個人の証券市場への参入は政策的課題でもある。その意味で、近年のインターネットの発達による金融取引や情報収集を容易にする各種システムの登場は望ましい方向と言える。さらに投資に関する情報交換を行うことを本来の目的とした電子掲示板(以下、掲示板と略す)も登場した。もともと個人投資家は孤立した存在であり、機関投資家のように社内や同業者間で情報交換することにより自らの判断の適切さを推測する手立てがなかったが、もし掲示板による情報交換や情報生産が行われるならば、いわゆる集合知につながり、ひいては投資効率に寄与する可能性がある。

ところが掲示板上のほとんどの投稿は証券投資に関する情報の観点からみれば、ノイズであると思われる。中には風説の流布や仕手的な行為と疑われる事案も発生している。しかし、米国における実証研究や調査によれば(Gu, et al.(2006), Das, et al. (2004)など)、情報価値がある投稿が含まれているとの報告がある。ただし、その事実だけで掲示板が有益であるとみなすのは早計である。有益な情報が含まれているとしても、それらを膨大な投稿群の中から体系的

に探し出し、その情報を利用して経済的に有意な超過投資収益率が得られるかどうかは別だからである。

本研究は掲示板の投稿の情報価値を検証することが目的であるが、研究の中心に置くのは投稿の中に含まれるノイズ投稿を体系的に除去する方法である。本研究では掲示板コミュニティのネットワーク構造に注目し、社会ネットワーク分析の手法を援用してノイズ排除を試みる。掲示板上のコミュニティがネットワークを形成しているとなれば、情報交換や情報生産の関係性があるであろうし、その前提の下で社会ネットワーク分析の手法を用いても非合理ではない。用いる手法は、具体的には中心性およびクリークである。掲示板コミュニティで情報の交換が行われていると仮定すれば、中心性が高い、あるいはクリークを構成している投稿者には、他の投稿者が返信メッセージを送るのに値する何らかの合理的根拠があると推測される。

分析の結果、中心性尺度を用いた検証ではセンチメントと整合した有意なものが得られなかったが、クリークを用いた検証では、センチメントの一部で有意な結果が頑健に得られた。この結果は、リスク調整済み投資収益率を用いて検証しても同様だった。

2.先行研究

個人投資家の予想に関する研究として Bagnoli, Beneish, and Watts(1999)がある。この研究で用いられたデータは、掲示板における投稿に付されたセンチメント(投資意思)ではなく、米国の *earning whispers.com* 社が個人投資家から収集したウィスパーである¹。内容はアナリスト予想とウィスパーとの統計量、およびそれぞれを元にした投資戦略を行った場合の投資収益率の比較である。分析の結果、ウィスパーはアナリストのコンセンサス予想よりも誤差が小さく、ウィスパーに基づいた投資収益率もアナリスト予想に基づくものよりも超過収益率が大きかった。この研究の限りでは個人投資家の情報生産にも価値があることが示唆される。Zaima and Harjoto(2006)も SSE 社によるウィスパーを対象とした分析を行っている。ウィスパーとアナリスト予想が食い違う場合において、市場反応はウィスパーの方が大きく、実際にウィスパーを元にした投資戦略を採用すると、リスク調整済み収益率はアナリスト予想を用いた場合よりも有意に大きいことも確認している。また、リスク調整済み収益率は経済的にも有効であった。

一方、ウィスパーではなくセンチメントを取り上げた研究としては Wysocki(1998)がある。この研究は Yahoo! Finance(米国)のセンチメントを対象としたもので、分析内容は、投稿件数の変化と、企業特性および市場活動との相関を分析したファクト・ファインディングである。分析の結果、投稿件数は過去のリターンや会計利益、市場時価総額、株価収益率、取引高等について有意に正であったが、超過収益率は経済的には意味があるほど大きくはなかった。

Dewally(2000)は *Deja-News* のセンチメントを対象に、投稿のセンチメントに基づいて投資を行った場合の日次超過収益率の分析を行った。検証方法はイベント・スタディーであり、イベント日は投稿された日である。分析の結果、投稿前に有意な超過収益率を記録した銘柄に対して投稿がなされる傾向が観察された。超過収益率に関する検証結果は、センチメントに基づく投資戦略の有効性を支持しないものであった。Tumarkin, and Whitelaw (2001)はイベント日を投稿量の急激な増加を記録した日と設定し、*Raging Bull* におけるセンチメントの変更をサンプルとしたイベント・スタディーを行った。イベント日前までの累積超過収益率を観察すると、センチメン

トは前日の超過収益率と有意な関係があったことから、センチメントは前日の価格変動を後追いつていることが推測される。逆に、投稿量に対しては前日の取引高、投稿量、センチメントから予測することができた。以上から、センチメントに基づいて投資を行っても超過収益率を得ることはできず、センチメントには情報価値はないと結論づけている。

Gu, Konana, Liu, Rajagopalan, and Ghosh(2006)は Yahoo! Finance の投稿者の中には優れた情報トレーダーが含まれていると想定してセンチメントを分析している。問題は、どのようにして投稿集合からノイズではないセンチメントを付する投稿者を抽出するかである。この研究では全ての投稿者のセンチメントを一律に用いるのではなく、過去の予想センチメントの正確性を元にウェイト付けて投稿者を抽出している。そしてこのセンチメントに基づいた投資戦略を実行した場合の超過収益率の有意検定を行った。超過収益率の算出モデルは、CAPM および Fama - French factor model など3種を用いたロングとショートを組み合わせた裁定取引ポートフォリオである。検証の結果、ウェイト付けしたポートフォリオによる超過収益率は、3つのモデルとも有意であり、かつ経済的にも意味がある大きさだった。

掲示板システムにセットされているセンチメントではなく、投稿本文を Naïve Bayes モデルおよび Support Vector モデルなど自然言語解析によりセンチメントに変換したものを取り上げた分析に Antweiler, and Frank(2001)がある。この研究は Yahoo! Finance と *Raging Bull* の投稿本文を元に強気/弱気度合い指標や投稿者間での意思合意度合い指標などセンチメントと同等の指標を作成し、これら各種指標に基づく買い持ち戦略の超過収益率を計算した。超過収益率はわずかに有意であったものの経済的には意味がある大きさではなかった。

Das and Chen(2001)も Yahoo! 掲示板における投稿本文を対象に Vector Distance Classifier, Discriminant -Based Classifier など4種類の自然言語解析手法を用いて投資意思を抽出した。分析内容は、投稿者の投資意思(売り/様子見/買い)と価格変動の相関分析である。投稿およびその投資意思が価格変動に対して1分間隔ベースでみてリードしているのか、ラグを伴っているのかを投稿をイベントとしたイベント・スタディーにより判断している。分析の結果、投稿は価格変動に対してラグを伴っていることが明らかにされた。つまり価格変動が生じてから投稿が行われており、投稿に情報効果があるとは言えないという結論である。

比較的最近の研究でもセンチメントの情報効果は明確に確認されていない。Das, Martinez- Jerez, and

¹ センチメントの詳しい意味は3節で述べる。ウィスパーとは個人投資家に対する予想EPS等のアンケートを元に計算したコンセンサス予想を意味する。

Tufano(2004)はMotley Foolに最低25件の投稿があった銘柄の中から、投稿頻度とメディア等によるニュース頻度の2要因について特徴的な4社を対象に、センチメントを独立変数に、超過収益率を従属変数にした重回帰を行った。センチメントはニュースにより影響を受けている可能性があるため、コントロール変数(プレス・リリースやメディア・ニュースの発表件数)も説明変数に加えた。その結果、センチメントと超過収益率の関係は安定的に有意ではなく、センチメントが情報価値を有していると判断するのは微妙であると結論している。

投資掲示板に関するネットワーク分析を行った研究は少ないが、その中にDas and Sisk(2003)がある。この研究はYahoo!、Motley Fool、Raging Bull、Silicon Investorを対象としたものである。内容にはセンチメントと投資収益率の関係についてではなく、投稿者ネットワークと銘柄の特性(価格変動や同分散など)を取り上げている。分析は、まず初めに投稿者が重複する銘柄間でネットワーク分析することにより銘柄間のコミュニティーを抽出し、次いでコミュニティー内の価格変動の共分散や収益率を解明するものである。コミュニティーの抽出方法は、投稿者が重複する銘柄毎に固有ベクトル中心性を計算し、この値が高いものを選択することにより行った。分析の結果、多数の銘柄により構成される大規模コミュニティーのリターンは、小規模のものよりも有意に大きく、共分散は有意に小さかった。また、高い固有ベクトル中心性スコアを有する株式は、そうでない株式よりも高い平均共分散を有する傾向があった。

3.前提と基本仮説

投稿の情報内容がノイズであるのか、あるいは有意な情報を含んでいるのかをネットワーク分析により判断するためには、さまざまな前提を置くことが必要である。ここでは、投稿に対する観察に基づいた前提を作成し、次にその上でネットワーク分析に適合した基本仮説を設定する。

3.1.基本前提

1.センチメントについて：投稿内容の情報価値の検証にはセンチメントを利用する。ここでのセンチメントとは、当該銘柄に対する投資意思を5段階(とても買いたい/買いたい/様子見/売りたい/とても売りたい)で評価したものと定義する。将来の価格が上昇すると投稿者が予想している場合には「とても買いたい/買いたい」が、下落すると予想する場合には「売りたい/とても売りたい」が表明されるだろう。買いたいセンチメントであれば、ロングするとし、売りセ

ンチメントならばショートするとする。

2.情報価値について：情報価値がある投稿とは、センチメントのとおり投資収益率が事後的に得られたものと定義する。センチメントが買いたい(とても買いたい+買いたい)ならばロング、売りたい(とても売りたい+売りたい)ならばショートとして投資収益率を計算する。投稿者が実際に投資を行ったかどうかは、この分析の範囲外とする。センチメントの予想とは異なる投資収益率であった投稿や、センチメント表明がない投稿は情報価値がないとする。

3.投資期間について：投稿者がセンチメントのとおり投資行為を行なったと仮定しても、果たしてどの程度の期間、ロングまたはショートするのかが明確ではない。そこで本研究では投資期間を1週間(5営業日)と1ヶ月(20営業日)の2種類とする。それぞれの期間中の対前日収益率を幾何平均したものを検証に用いる収益率とした²。これは投稿者が、仮に同期間内ならばいつでも好きな時点でロングまたはショートを終了したとしても柔軟に対応できる上、イベント・スタディーに基づく分析をする必要がなくなるからである³。

3.2.基本仮説

1.ネットワークの発生に関して：掲示板には、ある投稿に対して別の投稿者からコメントを付けるための返信機能がある⁴。この機能を活用した返信行動により、投稿者はそれぞれが孤立した存在ではなく、投稿者間にリンク(任意の投稿とそれに対する返信により結ばれる関係)が張られ、投稿者ネットワークが誕生する。ある投稿者が返信行動をする/しないを意思決定するには相応の理由があるはずであるが、証券投資に関する掲示板であることの趣旨を考えれ

² 通常の幾何平均であり、算式は

$$r_t = \sqrt[n]{(1+r_1) \times (1+r_2)^2 \times \dots \times (1+r_n)^n} - 1$$
である。ここで、 $r_n(n=1,2,\dots,10,\dots,20)$ はn日間のスポット・レートである。

³ この他にも、本研究では投稿動機や投稿内容を考慮しない。投稿内容の信頼性や情報価値などは投稿者の投稿動機に大きく影響を受けると考えられるので、本来ならば分析すべき対象である。しかし、本研究はノイズも含めた総投稿集合の中から情報価値のある投稿を抽出することなので、投稿動機に基づいてサンプルを区別する必要がない。以上により、投稿動機および投稿内容と、投稿の情報価値との関係は独立であるとする。

⁴ 投稿者の同定にはハンドル・ネームを用いる。同一人物が複数のハンドル・ネームを使用することは可能だが、名寄せすることは不可能だし、名寄せしなくても本研究の分析には不都合はないので異なるハンドル・ネームならば異なる投稿者と考える。

ば、投稿内容に情報価値があるからと仮定する⁵。

基本仮説1: 情報価値があると思われる投稿には、別の投稿者から返信があると期待される。

2. サブ・ネットワークの発生: 掲示板内で有益な情報を提供しあう小さな投稿者ネットワークの存在が確認されている⁶。このサブ・ネットワークが成立もしくは維持される根底には相応の信頼関係があると考えるのが合理的であろう。もちろん、このようなグループ内でもノイズの発生は否定できないが、グループ外よりは掲示板の趣旨に沿ったコミュニケーションが行われていると仮定する。

基本仮説2: 情報価値があると評価される投稿者は、特定のサブ・ネットワークに属していると期待される。

次節で上記の前提および基本仮説を元にした検証仮説を提案する。

4. データと統計仮説

4.1. データ

Yahoo! Finance の株式掲示板の投稿をサンプルとする。期間は2008年1月4日から6月30日までの121営業日をプールした投稿を用いる。対象企業は、TOPIX Core100の中の電機セクターに含まれる52社である。超過収益率を計測するためのベンチマークにはTOPIX総合指数を用いる。超過収益率の算出は、日次終値ベースで計算した投資収益率から同リスク調整済み投資収益率を引くことにより求める。リスク調整済み投資収益率は、マーケット・モデルにより計算したベータを用いる。ベータの計算期間は2007年7月の第1営業日から同年12月の最終取引日までである。超過収益率 x_i の計算式は以下の通り。

$$x_i = r_i - \bar{r} = r_i - (\alpha_i + \beta_i r_i)$$

⁵ 投稿者間で非難の応酬が行なわれ、その流れでの返信が存在することも事実である。その場合、返信を受けることが情報価値の代理変数とは言えない。しかし、本研究では投稿内容を考慮しないため、非難の応酬の結果として返信数が増加するのか、何らかの有益な情報交換が行われた結果として返信数が増加するのか、判別できないので、一律に返信数を情報価値の増加関数とみなすことにする。

⁶ Das, Martinez-Jerez & Tufano(2004)は投稿者へのインタビュー調査により、投稿動機のひとつとして優れた分析や正確な予測などを行うことによる投稿仲間内における評判の獲得があると述べている。投稿者コミュニティの中には、少人数ながらも有益と彼ら/彼女らが考える情報を提供しあうネットワークが存在することも明らかにしている。

ここで r_i は銘柄 i の日次収益率、 r_t はTOPIX総合指数の日次変化率

全投稿者数は5,748人であり、全投稿数は95,393件であった。このうち削除された投稿は72,297件あるので、残った生存投稿は23,096件(全投稿数の約24.2%)である。全投稿数のうち、センチメントがついているものは5,417件、他の投稿者から返信があったものは10,896件、センチメントおよび返信の2つともあったものは2,251件であった⁷。

4.2. 統計仮説

次数中心性⁸

基本仮説1を具体的に計測する尺度として、ネットワーク理論における次数中心性を用いる。次数中心性とは、任意の投稿者から他の投稿者へのリンクの本数である。次数中心性が高いほど、他の投稿者から返信が多いことを示す⁹。次数中心性の定義式は以下の通り。

$$C_{d_i} = \frac{\text{deg}(v_i)}{n-1}$$

ここで $\text{deg}(v_i)$ は投稿 i の次数、 n は総投稿者数
仮説1: 次数中心性が高い投稿者ほどセンチメント通りの投資収益率が得られる。

高い次数中心性の投稿者から順に並べて1/5毎にグループ化する。センチメント毎に投資収益率を t 検定により判断する。

クリーク

基本仮説2整合する投稿者を抽出するための方法として、ネットワーク理論における完全グラフの概念を用いる。完全グラフとは、全てのノードがリンクで相互に結ばれているグラフを指す。これはグラフを構成する全投稿者が直接的に返信をしあうサブ・グループであることを意味する¹⁰。

仮説2: クリークを構成する投稿者群の投資収益率

⁷ 自分の投稿に対する投稿は1,380件を除く。

⁸ 次数中心性は、局所的な各投稿者の次数にのみ注目しているため、ネットワーク全体の構造を反映していないという問題点が指摘されているが、本研究は投稿者ネットワーク全体を考慮しないので支障はない。なお、同様の観点から近接中心性など他の尺度も用いていない。

⁹ 本研究で無向グラフを採用した理由は、有向グラフでは著しくデータが減少したためである。

¹⁰ サブ・ネットワークを抽出する手法として、他に n -クリークや k -クランなどがある。しかしこれらではネットワーク・サイズを小さくすることができなかつたため、制約条件が一層厳しいクリークを用いることにした。

はセンチメント通りである。

クリークの次数は 3、4、5 の 3 段階とする¹¹。センチメント毎に投資収益率を t 検定して判断する。

予備検定と追加検定

準備として独立性検定および投稿シェアに基づく超過収益率の検定を行なう。センチメントをランダムに選んだとしても現実がその通り(正の収益率/不変/負の収益率)となるチャンス・レートは 1/3 である。従って、1/3 以上の確率で期待通りになっていることを確認する必要がある。ただし現実には、センチメントが中立かつ価格変化が不変の組み合わせはデータが少ない。そこで、センチメント 1(とても売りたい)と 2(売りたい)を合計して負の期待投資収益率に、センチメント 4(買いたい)と 5(とても買いたい)を合計して正の期待投資収益率とする。現実の投資収益率もセンチメントの群と整合するように正と負の 2 つだけとする。センチメントと現実投資収益率それぞれ 2 カテゴリーの 2×2 格子の自由度 1 のカイ 2 乗検定を行う。

投稿シェアによる検定は、投稿件数に占めるシェアが多い投稿者から順に並べて 1/5 毎にグループ化し、センチメント毎に投資収益率を t 検定により判断する。

追加検定としてはクリークに属するサンプルと、それ以外のサンプルによる収益率を用いた 2 標本の対照検定を行なう。クリークに属する投稿者の投稿を実験群、センチメントのみの投稿 17,679 件を対照群とする。

5.分析結果

5.1.記述統計

投稿およびネットワークの状況

投稿者 1 人当たりの投稿数は $\mu=4.018$ 件、S.D.=14.50 件、中央値は 1 件、最大は 518 件であった。横軸に対数化投稿数を、縦軸に対数累積確率密度を採ったランクサイズ・プロットを作成すると係数は -1.075 であり、ベキ法則が成立していると推測される(図表 1)¹²。

¹¹ クリークの次数を3から5とした理由は、クリークは2以下ではそもそもクリークの意味をなさないし、6以上のクリークはサンプル集合内に存在しなかったためである。

¹² ベキ法則とは、この場合は投稿数が少ないものから多いものへ並べて描ける分布の累積分布関数がベキ乗分布の形状をしていることを指す。確率変数をx、確率密

ネットワーク密度¹³は $\mu=0.107$ 、S.D.=0.309 である。クラスター係数は 0.365 であり¹⁴、平均パス長は 1.001 である¹⁵。ランダム・グラフと比較してクラスター係数が大きく平均パス長も短いので、スモール・ワールドの可能性はある。

他からの返信がない投稿者(次数が 0)を除いて計算した次数は、 $\mu=4.188$ 、S.D.=18.449 であり最大は 947(標準化次数は 0.073)であった。横軸に対数化次数中心性を、縦軸に対数累積確率密度を採ったランクサイズ・プロットを作成すると、係数は -0.648 であり、ベキ法則が成立していると判断するのは困難である(図表 2)。クリークの数は次数 3 が 953 個(投稿者数 977 人)、4 が 135 個(投稿者数 204 人)、5 が 10 個(投稿者数 27 人)であった。

個別銘柄とインデックスの投資収益率

計測尺度として、リスク調整済みの 10 日平均収益率(ER10)と同 20 日平均収益率(ER20)を計算する¹⁶。

2008 年 01 月 04 日から 06 月 30 日の間のサンプル企業 52 社の銘柄別の収益率は、ER10 が $\mu=-1.29\%$ 、S.D.=3.82%、ER20 が $\mu=-0.42\%$ 、S.D.=1.04% であった。TOPIX は対前日投資収益率(R1)が $\mu=-0.15\%$ 、S.D.=1.90%、10 日間平均投資収益率(R10)が $\mu=-0.60\%$ 、S.D.=2.90%、20 日平均投資収益率(R20)が $\mu=-0.80\%$ 、S.D.=3.66% であった。いずれにせよ、この間のパフォーマンスはマイナスであり、センチメ

度関数を $p(x)$ とすると、 $\int_0^k p(x) dx \propto x^{-\mu}$ である。つまり、係数 μ が -1.075 のベキ分布が成立している可能性を示している(林他、(2007)、p.40、青山他、(2007)、p.27)。ベキ法則が成立しているということは少数の多くの投稿をした投稿者と、多数の投稿回数の少ない投稿者で構成されていることを意味する。

¹³ グラフ G の密度 $d(G)$ は、グラフ G のリンクの数と、それが完備グラフのときのリンクの数との比率である。 $d(g) = m(G) / m(K_n)$ 、ここで $m(G)$ はグラフ G のリンクの数、 $m(K_n)$ は規模 n のグラフが完全グラフであった場合のリンクの数である。(金光(2003)、P.59)

¹⁴ クラスター係数とは、実際に存在しているクラスターの数 E_i を、存在可能なクラスターの数で割ることにより求める。ここでクラスターとは、3つのノードが相互にリンクで結ばれているものを指す。クラスター係数は、ノード i のリンクを k_i 本あるとすると、個別ノード i のクラスター係数は $C_i = 2E_i / \{k_i(k_i - 1)\}$ であり、これを全ノード数 n について平均したものである。比較のため、同数のノードとリンクのランダム・グラフを作成してクラスタリング係数を算出すると、0.109 であった。

¹⁵ 同数のノードとリンクのランダム・グラフによる平均パス長を算出すると、2.00 であった。

¹⁶ 他に対前日リスク調整済み投資収益率および、リスク未調整の対前日、10日と20日の計4種類も計算した。しかしER10およびER20と大差なかったので報告しない。

ントの超過収益率も負になる可能性が高い。

5.2. 検定結果

独立性検定

センチメントが付されている投稿 3,770 余件を対象に独立性検定を行ったところ、ER10 および ER20 のいずれもが有意でなかった(図表 3)¹⁷。このことから、センチメントには予想能力があるということは難しい。また的中率に関してはいずれの投資収益率尺度でも負のセンチメントはチャンス・レートの 50%を超える一方で正のセンチメントは 50%に達しなかった。期間中の投資収益率は、大半の銘柄において負であったことから、負のセンチメントの的中率が高く、正のセンチメントの的中率が低いということは独立性検定の結果と整合的である。

投稿数シェア(Posting)

投稿者毎の投稿数シェア(Posting)と収益率の関係をセンチメントに応じて分析した結果、特筆すべき安定的な関係は見出せなかった(図表 4A)。センチメント通りの投資収益率が ER10、ER20 とともに 10%水準で有意であったものは 5 つである($0.01 < \text{Posting} \leq 0.03\%$ の S_1、 $0.09\% < \text{Posting} \leq 0.42\%$ の S_1,2 と $\text{Posting} \leq 0.42\%$ 以上の S_1,2)。

40 個のクラス(ER10 および ER20 の Posting 階級 5 つとセンチメント 1,2,4,5 の組み合わせ)のうち 10 個しか有意ではないことから、投稿数に基づいて投稿者を抽出しても、有効な情報をセンチメントから見出すことは難しいと考えられる。

次数中心性(NrmDegree)

次数中心性とセンチメントに応じて分析した結果でも、特筆すべき安定的な関係はなかった(図表 4B)。センチメント通りの投資収益率が ER10、ER20 とともに有意であったものは 3 つである($0.0348 \leq \text{NrD} < 0.842$ の S_2 と $\text{NrD} \leq 0.842$ の S_1,5)。

クリーク(Clique)

クリーク 3、4、5 の結果で共通していることは、S_5 における超過収益率がおおむね有意に正である

¹⁷ センチメントが付されている投稿は、5,417件であるが、この中からS_3(中立)の件数1,638件を除き、さらにごく少数の現実投資収益率が0%であったサンプルを除いた。このため、サンプル数は投資収益率尺度により異なる。ER10は3,775件、ER20は3,778件であった。

ことである(図表 5)。ER10 と ER20 で計測する超過収益率は、クリーク 3 が 0.51%と 0.79%(ただし ER10 の 0.51%は有意ではない)、クリーク 4 は 1.13% と 1.94%、クリーク 5 は 3.14%と 4.06%(図表 4C)である。平均超過収益率の日数が 10 日から 20 日に増加するに従って、またクリークの次数が 3 から 5 に増加するに従って超過収益率も増加する傾向がある。

S_1 および S_2 の符合はおおむね負であるが、ER10 と ER20 の両方とも安定的に有意であるのは、クリーク 5 の S_2 だけである。しかし、次数が増加するに従って収益率がセンチメント通りに減少する傾向がある。

追加検定

クリークを構成する投稿者による投稿をサンプルにしたもの(実験群)と、孤立した投稿者によるもの(対照群)のセンチメント毎の投資収益率を比較した(図表 6)。クリーク 3、4、5 の結果で、S_5 における両群の差が有意であることが共通している。クリークの次数が増加するに従って有意水準が上がっている。クリーク 5 は S_2 も両群の差が有意である。

6. 結果の解釈

6.1. 次数中心性

センチメントどおりの超過収益率を示したのは 40 個のクラスのうちで 6 つしかなかったことから、仮説 1 を支持することはできないだろう。次数中心性を尺度として判断する限り、コミュニケーションを多くの投稿者との間で行っているからと言って優れた投資成果をもたらすとは言えない。ファイナンス論の示す通り、無償で公開されている掲示板投稿には超過収益率をもたらすような私的情報は含まれていないと言えるだろう。

6.2. クリーク

クリークの分析結果から言いうことは、S_5 だけは高度に有意であり、センチメント通りにリスク調整済みの超過収益率を稼得していることである。また、クリークの次数が増加するに従って収益率も増えている。クリークに含まれる投稿者と孤立した投稿者の対照検定を行っても、S_5 の有意な超過収益率は支持された。投資収益は 1.81~3.48%であり、経済的にも投資効果があると言えるだろう。S_1,2 も有意なものは存在するが、安定的とは言えない。

以上から考えると、全般的にはクリークを構成する投稿者の投資収益率がセンチメント通りとする仮説 2 は棄却されるだろう。だが、S_5 だけは安定的

に有意であり、また対照検定の結果もクリークによる抽出の有効性を支持していると言えよう。

7.おわりに

本研究では、掲示板の投稿に含まれるノイズを社会ネットワーク分析の手法を用いて除去し、センチメントの情報価値を検証した。投稿件数に基づく投稿者の抽出や次数中心性に基づく抽出では、明確にセンチメント通りの超過収益率を確認することができず、投稿の情報価値を見出すことはできなかった。しかし、ノイズの中に埋もれてはいるものの、ごく少数の投稿者の間では密な情報ネットワークが構築されており、そのネットワーク内では私的情報がやり取りされている可能性もある。ネットワーク理論に基づくスクリーニング手法により、一部のセンチメント(S_5, とても買いたい)はおおむね安定的かつ有意にセンチメント通りの超過収益率であり、経済的にも有効であった。ただし、クリークを構成する投稿者に関しても全てのセンチメントにおいて有意な結果が得られたわけではない。

当初、掲示板の投稿はほとんどがノイズであろうと推測していたが、投稿群から一部のセンチメントに関しては情報を抽出することができた。しかし、ネットワーク分析の一部を用いて分析しただけであることから、さらに詳しい研究が求められるだろう。

謝辞

この研究を行なうにあたり、中央大学商学部の有賀裕二教授から有益な指示とコメントをいただきました。記して感謝申し上げます。

参考文献

- [1] Antweiler, W. and M. Frank “Is All Talk Just Noise? The Information Content of Internet Stock Message Boards”, working paper, (2001)
- [2] Bagnoli, M., M. Beneish, & S. Watts “Whisper Forecasts of Quarterly Earnings per Share” Journal of Accounting and Economics, Vol. 28, No.1, (1999)
- [3] Gu, B, P. Konana, A. Liu, B. Rajagopalan and J. Ghosh (2006), “Predictive Value of Stock Message Board Sentiments”, working paper
- [4] Das, S. and M. Chen (2001), “Yahoo! For Amazon: Opinion Extraction from Small Talk on the Web”, working paper
- [5] _____ and J. Sisk (2003), “Financial Communities”, working paper
- [6] _____, A. Martinez-Jerez & P. Tufano(2004),

“e-Information: A Clinical Study of Investor Discussion and Sentiment”, working paper

[7] Dewally, M. (2000), “Internet Investment Advice: Investing with a Rock of Salt”, working paper

[8] Knoke, D. and S. Yang (2008) Social Network Analysis second edition, Sage Publications

[9] Scott, J. (2000), Social Network Analysis a handbook second edition, Sage Publications

[1 0] Tumarkin, R. and R. Whitelaw (2001), “News or Noise? Internet Postings and Stock Prices”, Financial Analysts Journal, May/ June

[1 1] Wysocki, P. D. (1998), “Cheap Talk on the Web: The Determinants of Postings on Stock Message Boards”, working paper

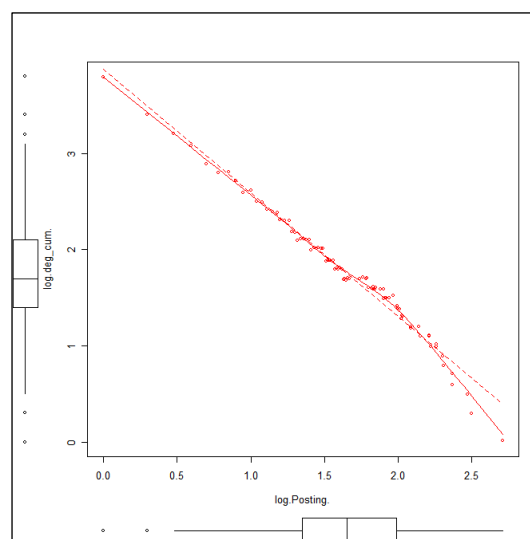
[1 2] Zaima, J. and M. Harjoto (2006), “Conflict in Whispers and Analyst Forecasts: Which One should be Your Guide?” working paper

[1 3] 青山秀明、家富洋、池田裕一、相馬亘、藤原義久(2007)『パレート・ファームズ』, 日本経済評論社

[1 4] 金光淳(2003)『社会ネットワーク分析の基礎』, 勁草書房

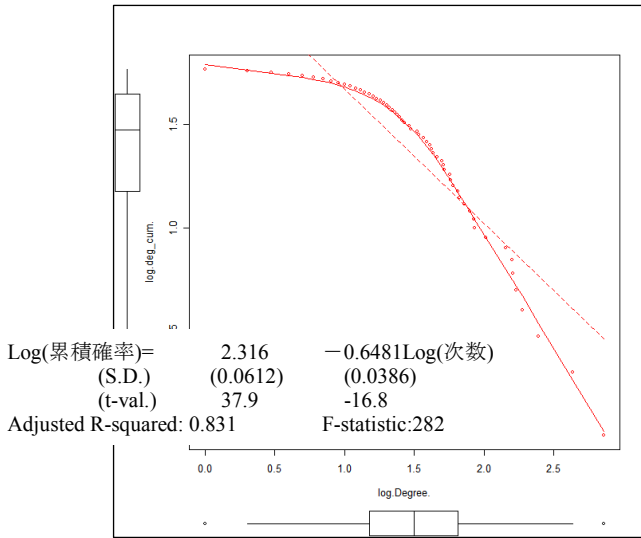
[1 5] 林幸雄編、大久保潤、藤原義久、上林憲行、小野直亮、湯田聰夫、相馬亘、佐藤一憲著(2007)『ネットワーク科学の工具箱』, 近代科学社

図表 1 : 投稿数のランクサイズ・プロット



Log(累積確率)=	3.621	-1.0746Log(投稿数)
(S.D.)	(0.00136)	(0.00283)
(t-val.)	2658.2	-379.6
Adjusted R-squared:	0.962	F-statistic: 1.441e+05

図表 2 : 次数中心性のランクサイズ・プロット



図表 3 : (予備検定)センチメントの独立性検定

尺度	現実の収益率	センチメントのサンプル数(括弧内は的中率)		独立性検定	
		dn (S1&S2)	up (S4&S5)	Chi-sq	p-value
ER10	dn	792 (57.8%)	1339	1.618	0.203
	up	578	1066 (44.3%)		
ER20	dn	807 (58.9%)	136	1.471	0.225
	up	564	1039 (43.2%)		

注 1 : それぞれの投資収益率尺度のグループ内で N=23,096, d.f=1
 注 2 : 的中率の計算は、dn の場合、センチメントが dn かつ現実投資収益率が dn の組み合わせのサンプル数を分子に、センチメントを dn とした総サンプルを分母にして行った。

図表 4 : 収益率の有意検定(投稿件数シェア、次数中心性、クリーク)

A) 投稿件数(Posting) N=23,096											C) クリーク(Clique) N=1,3190							
収益率尺度	stmt	Pst<=0.01%		0.01<Pst<=0.03%		0.03<Pst<=0.09%		0.09<Pst<=0.42%		Pst>0.42%		stmt	Clique3		Clique4		Clique5	
		return	p-val	return	p-val	return	p-val	return	p-val	return	p-val		return	p-val	return	p-val	return	p-val
ER10	1	0.00%	1.00	-1.85%	0.00	0.12%	0.71	-1.43%	0.00	-1.86%	0.00	1	-0.72%	0.03	-1.04%	0.02	-1.00%	0.50
	2	0.07%	0.94	-0.39%	0.63	0.86%	0.39	-1.48%	0.09	-1.78%	0.02	2	-0.06%	0.93	-1.10%	0.15	-2.18%	0.02
	4	-0.91%	0.01	0.32%	0.43	-0.20%	0.62	-0.50%	0.37	-0.57%	0.11	4	-0.15%	0.66	0.05%	0.90	-0.45%	0.48
	5	-0.23%	0.38	-0.32%	0.24	0.18%	0.59	-0.24%	0.39	0.60%	0.16	5	0.51%	0.11	1.13%	0.04	3.14%	0.00
	5	0.46%	0.47	-1.93%	0.00	0.80%	0.07	-2.22%	0.00	-1.81%	0.00	1	-0.24%	0.58	-0.73%	0.15	-2.50%	0.13
ER20	2	0.07%	0.95	0.11%	0.92	1.89%	0.19	-3.62%	0.00	-2.65%	0.00	2	0.10%	0.91	-1.29%	0.14	-2.94%	0.00
	4	-1.01%	0.03	0.23%	0.67	-0.31%	0.59	-0.95%	0.18	-2.00%	0.00	4	-0.66%	0.14	-0.37%	0.51	-1.49%	0.06
	5	-0.42%	0.19	-0.27%	0.49	-0.04%	0.94	-0.20%	0.58	0.57%	0.26	5	0.79%	0.06	1.94%	0.01	4.06%	0.00
	5	0.71%	0.71	2.66%	0.01	0.65%	0.33	-0.29%	0.71	-1.61%	0.00	1	-0.24%	0.58	-0.93%	0.15	0.696%	0.14
	2	-1.33%	0.39	5.33%	0.03	-2.00%	0.22	-1.72%	0.09	0.70%	0.67	2	-1.29%	-0.23%	-1.056%	0.31		
ER20	4	-1.05%	0.30	0.32%	0.76	0.14%	0.88	-0.74%	0.27	0.08%	0.90	4	-0.37%	-0.67%	0.304%	0.62		
	5	-1.39%	0.25	0.37%	0.65	-0.34%	0.51	2.03%	0.10	0.71%	0.08	5	1.94%	-0.20%	2.135%	0.00	***	
	5	0.71%	0.71	2.66%	0.01	0.65%	0.33	-0.29%	0.71	-1.61%	0.00	1	-0.73%	-0.93%	0.203%	0.71		
	2	-1.33%	0.39	5.33%	0.03	-2.00%	0.22	-1.72%	0.09	0.70%	0.67	2	-1.29%	-0.23%	-1.056%	0.31		
	4	-1.05%	0.30	0.32%	0.76	0.14%	0.88	-0.74%	0.27	0.08%	0.90	4	-0.37%	-0.67%	0.304%	0.62		
ER20	5	-1.39%	0.25	0.37%	0.65	-0.34%	0.51	2.03%	0.10	0.71%	0.08	5	1.94%	-0.20%	2.135%	0.00	***	
	5	0.71%	0.71	2.66%	0.01	0.65%	0.33	-0.29%	0.71	-1.61%	0.00	1	-0.73%	-0.93%	0.203%	0.71		
	2	-1.33%	0.39	5.33%	0.03	-2.00%	0.22	-1.72%	0.09	0.70%	0.67	2	-1.29%	-0.23%	-1.056%	0.31		
	4	-1.05%	0.30	0.32%	0.76	0.14%	0.88	-0.74%	0.27	0.08%	0.90	4	-0.37%	-0.67%	0.304%	0.62		
	5	-1.39%	0.25	0.37%	0.65	-0.34%	0.51	2.03%	0.10	0.71%	0.08	5	1.94%	-0.20%	2.135%	0.00	***	

図表 6 : (追加検定)クリークに属するサンプルとそれ以外の 2 標本対照

clique	Return metrics	stmt	Expremental group	Control group	Difference	p-val
3	ER10D	1	-0.72%	-1.03%	0.311%	0.40
		2	-0.06%	-0.24%	0.184%	0.81
		4	-0.15%	-0.38%	0.232%	0.55
		5	0.51%	-0.10%	0.618%	0.07 *
		5	0.51%	-0.10%	0.618%	0.07 *
3	ER20D	1	-0.24%	-0.93%	0.696%	0.14
		2	0.10%	-0.23%	0.336%	0.74
		4	-0.66%	-0.67%	0.017%	0.97
		5	0.79%	-0.20%	0.987%	0.03 **
		5	0.79%	-0.20%	0.987%	0.03 **
4	ER10D	1	-1.04%	-1.03%	-0.009%	0.98
		2	-1.10%	-0.24%	-0.852%	0.32
		4	0.05%	-0.38%	0.433%	0.35
		5	1.13%	-0.10%	1.235%	0.03 **
		5	1.13%	-0.10%	1.235%	0.03 **
4	ER20D	1	-0.73%	-0.93%	0.203%	0.71
		2	-1.29%	-0.23%	-1.056%	0.31
		4	-0.37%	-0.67%	0.304%	0.62
		5	1.94%	-0.20%	2.135%	0.00 ***
		5	1.94%	-0.20%	2.135%	0.00 ***
5	ER10D	1	-1.00%	-1.03%	0.033%	0.98
		2	-2.18%	-0.24%	-1.932%	0.05 ***
		4	-0.45%	-0.38%	-0.072%	0.91
		5	3.14%	-0.10%	3.244%	0.00 ***
		5	3.14%	-0.10%	3.244%	0.00 ***
5	ER20D	1	-2.50%	-0.93%	-1.568%	0.16
		2	-2.94%	-0.23%	-2.707%	0.01 **
		4	-1.49%	-0.67%	-0.816%	0.32
		5	4.06%	-0.20%	4.252%	0.00 ***
		5	4.06%	-0.20%	4.252%	0.00 ***

図表 5 : クリーク 3,4,5 サンプルにおけるセンチメント 5 の収益率

