

インターネット株式掲示板の投稿内容分析に基づく

ファクターモデルの構築
Factor Model of Stock BBS Postings諏訪博彦^{*1}
Hirohiko Suwa梅原英一^{*2}
Eiichi Umehara太田敏澄^{*1}
Toshizumi Ohta^{*1} 電気通信大学
University of Electro-Communications^{*2} 野村総合研究所
Nomura Research Institute

We aim to develop a new factor of stock BBS postings that is different from our BMB factor. In our previous study, the contents of stock BBS postings are classified into two categories; i.e. the bullish postings and the bearish postings, and our BMB factor is based on these categories. The results of recent study suggest that the contents of stock BBS postings may be represented by employing more than one index. To develop a new factor, we use a morphological analysis and a PCA to analyze the contents of stock BBS postings. As results, we develop a new factor that based on principal component score and it represented the return of stock, and in that they are not correlated with our BMB factor.

1. はじめに

本研究の目的は、株式掲示板におけるアノミーを発見すること、株式掲示板の投稿内容より株式リターンを説明するファクターを作成することの2つである。

投資家が投資判断を下す際、数値化された情報(会計データ、株価)だけでなく、ヘッドラインニュースのような非数値的な情報も利用している。特に株式掲示板は他の投資家の生の声を多く蓄積しており、投資判断に有効な判断材料である可能性がある。

これまでの金融市場の研究において、小型株効果やバリュー株効果など、様々なアノミー(CAPMでは説明できないリスクプレミアム)が発見されている[小林 09]。しかし、インターネット株式掲示板(以下、株式掲示板)におけるアノミーはまだ少ない。アノミーの発見はアブノーマルリターンの獲得に繋がる可能性があることから、本研究では日本の株式市場において株式掲示板におけるアノミーの発見を試みる。

また、これまで株式掲示板を利用したファクターモデルを作成が行われている。Antiweiler & Frank は米国市場において、CAPMおよびFama & Frenchの3ファクターモデルに対して、掲示板の投稿数を元とした投稿数ファクター-NMQを加え検証している[Antiweiler 02]。

阿部らは、日本市場においてCAPMに対して投稿数ファクター-NMQ、投稿内容ファクター-BMBを作成し、分析を行っている[阿部 10]。その結果、投稿数は株式リターンを説明しないが、投稿内容は株式リターンを説明するとしている。しかし、現状ではBMBで利用されている強気指数だけでは投稿内容の全てを捉えておらず、残差が存在していることも明らかになっている。また、阿部らの研究では投稿内容を「強気」という一軸しか分類していない。関連研究によるとテキストデータはいくつかの軸で分類出来ることが確認されており、阿部らの研究も同様に投稿内容を「強気」以外の複数の軸で分類できる可能性がある。このことから、株式リターンを説明する株式掲示板に基づくファクターを作成を目指す。

連絡先: 諏訪博彦, 電気通信大学, 〒182-8585 東京都調布市調布ヶ丘 1-5-1, 042-443-5727, h-suwa@is.uec.ac.jp

2. 関連研究

株式掲示板等と株式指標等の関係を論じた数多くの研究がなされている[Wyscoki 99] [Tumarkin 01] [Jones 06] [Antiweiler 02] [丸山 08] [阿部 09] [阿部 10] [和泉 10]。

Wyscoki 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ている[Wyscoki 99]。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 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ている[Tumarkin 01]。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 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調査している[Jones 06]。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でも成立したと述べている。

Antiweiler & Frank は、米国のYahoo!掲示板を1999年から2001年の期間で上企業の3500万以上のメッセージを分析している[Antiweiler 02]。Fama & Frenchの3ファクターモデルに対して投稿数ファクター-NMQを加えたモデルを検証したところ、投稿数は株式リターンに影響を与えている可能性があることを報告している。

丸山らは、日本のYahoo!ファイナンス掲示板の投稿数Top50を2005年から2006年の期間でメッセージを分析している[丸山 08]。掲示板より取得した投稿データを、自然言語処理(形態素解析、ノイズ除去)と機械学習により、メッセージを「弱気投稿」「強気投稿」「中立」の3種類に分類している。結論として、第1に投稿数と株式リターンの関連の可能性は低いが、投稿内容を考慮した場合、弱気投稿数が多いと翌日の株式リターンが負になる傾向があること。第2に投稿数がボラティリティ、出

来高と関連していること。第 3 に強気比率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。

阿部らは、日本の Yahoo! 株式掲示板の東証一部上場銘柄 1390 社を、2003 年から 2007 年の期間において分析している [阿部 09]。CAPM に対して投稿数ファクター-NMQ、投稿内容ファクター-BMB を加えた式を検証した結果、投稿内容ファクター-BMB は投稿数ファクター-NMQ よりも優れていることを示唆し、リターンを予測する可能性があることを報告している。

さらに阿部らは、既存研究 [阿部 09] において不足していた株価相場の下降局面も加え再検証を行っている [阿部 10]。また、Fama & French の 3 ファクターモデルに投稿数ファクター-NMQ、投稿内容ファクター-BMB を加えたモデルの検証を行った。結論として、月次において投稿内容ファクターモデルはリターンを説明する可能性が高いことを報告している。

和泉らは、日本銀行の金融経済月報に対して形態素解析と主成分分析を用いることで、実際の金融市場 (株式、国債、円ドルレート) をどの程度説明しているのかについて検証を行っている [和泉 10]。主成分分析の結果、30 個の主成分で累積寄与率が 60% となった。主成分の中には、金利に関する単語で構成されている主成分、企業活動に関する単語で構成された主成分、市場の動きに関する単語から構成された主成分があった。主成分得点を説明変数として回帰分析を行った結果、第 1 に国債 2 年・5 年物に関しては約 85% の確率で、日経平均株価に関しては約 77% の確立でトレンドの予測が可能であること、第 2 にトレンド予想の期間は最大でも三ヶ月程度であることを報告している。

これらの関連研究で確認出来たことは以下の 2 点である。第 1 に、株式掲示板の投稿内容を分類しファクターとすることで、株式リターンを予測することができる。第 2 に、主成分分析を利用することで、「企業活動」「経済のファンダメンタルズ」「市場の動き」に関する単語で構成されたものなど、複数の主成分が抽出できる。

これまでの結果より、投稿内容を利用することで株式リターンを予測すること、主成分分析を利用することでテキスト情報から複数の主成分を抽出出来ることが確認できた。このことから、主成分分析を利用して投稿内容から「強気以外の軸を抽出しファクターとすることで、株式リターンの予測を行うことが可能であると考えられる。

そこで、本稿では投稿内容から作成した主成分がリターンを説明できる可能性があるのか検証を行う。また、作成した主成分が「強気」とは異なったものであることの確認も行う。

3. 分析方法

本節では、株式掲示板においてアノマリーが存在するのか確認を行う。また、主成分得点を利用したファクターを作成し、その検証を行う。

3.1 分析対象

本研究は分析対象掲示板として Yahoo! 掲示板を利用している。対象銘柄は東証一部上場銘柄 1501 社、対象期間は 2003 年 1 月 1 日から 2008 年 12 月 31 日の 6 年間である。

3.2 分析の流れ

以下に分析の流れを示す。

① 投稿の取得

Yahoo! 株式掲示板からの投稿の取得は、プログラムによって自動的に行われる。なお、この方法は、既存研究 [丸山 08] [阿部 09] [阿部 10] と同様である。

② 形態素解析

形態素解析により各投稿に対して分詞処理を行うことで、投稿にどのような単語が何個含まれているのかを調べる。本研究では形態素解析プログラムに MeCab を用いる。

③ TF・IDF 値の算出

各投稿内に出現する単語の TF・IDF 値を算出し、単語の特徴量を求める。単語のフィルタリングのために出現頻度が 1000 回以上の単語のみを使用する。また、同一の単語が複数の投稿で出現する場合は、各投稿毎に求めた TF・IDF 値を合計することで全投稿における単語の特徴量とする。また、本研究においては、TF・IDF 値の合計が大きい上位 700 単語を主成分分析に使用する。

④ 主成分分析

株式リターンを説明する株式掲示板に基づくファクターを作成するために、出現単語を説明変数として主成分分析を行う。

本研究においては、TF・IDF 値の算出によって求めた 700 単語を説明変数として主成分分析を行った。各主成分において、主成分負荷量の絶対値が大きい単語を調べることで、その主成分を構成する単語がどのような傾向を持っているのかを判断する。

⑤ 月次平均主成分得点と月次リターン・強気指数との相関

月次で主成分得点を平均することで、月次平均主成分得点を求めた。

$$\text{月次平均主成分得点} = \sum_{j=1}^n x_{k,j} / n \dots (1)$$

k: 対象とする銘柄

n: 対象する月の総投稿数

x: 投稿 j の主成分得点

月次平均主成分得点と強気指数の相関を求め、主成分が強気指数と同一の意味を持つか否かを明らかにする。また、月次平均主成分得点と月次リターンとの相関を求めることで、主成分が株式リターンを説明できるか否かを確認する。この結果、月次リターンと相関があり強気指数との相関がない主成分に関しては、新たなファクターになり得る可能性がある。

株式リターンと相関がある場合、月次平均主成分得点を利用することで株式リターンの動きを説明することが出来る可能性がある。また、強気指数と相関がない場合、投稿内容ファクター-BMB で利用されている強気指数と相関が無いことから、月次平均主成分得点は従来の投稿内容ファクター-BMB とは異なるファクターになり得る可能性がある。

なお、強気指数は全投稿における強気/弱気投稿の多少を示しており、以下の式で定義される。この定義は、丸山ら [丸山 08] と同様である。

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

⑥ ポートフォリオの作成

月次リターンと相関がみられ、かつ強気指数との相関が見られない主成分を対象に、月次平均主成分得点を利用したポートフォリオ (PF) を作成する。作成方法は、一ヶ月単位で各銘柄の月次平均主成分得点が高い順番に単純 5 等分したものとす。また、その月に月次平均主成分得点がない銘柄は x とする。

この方法は、Antiweiler & Frank の手法に準じている [Antiweiler 02]。

⑦CAPMにおけるアブノーマルリターンの確認

月次平均主成分得点とリターンの間には相関があると仮定すると、CAPM においてはアブノーマルリターンを確認できると考えられる。この場合、月次平均主成分得点が高いポートフォリオの場合には 0 からプラスへの乖離し、低いポートフォリオについては 0 からマイナスの乖離すると考えられる。

CAPM は以下の式で定義されており、回帰分析を行うことでアブノーマルリターン(α)の有無を確認する。この方法は、既存研究[Antiweiler 02] [阿部 09] [阿部 10]と同様である。

$$r_i - r_f = \beta_i(r_M - r_f) + \alpha \dots (3)$$

$r_i - r_f$: ポートフォリオの超過リターン

$r_M - r_f$: マーケットポートフォリオの超過リターン

α : アブノーマルリターン

* r_M は対象銘柄 1501 社の平均リターンとする

⑧主成分得点を利用したファクターの検証

主成分得点を利用したファクター(PCAP)をマルチファクターモデルとして展開し、アブノーマルリターンが出現しないことを確認する。

PCAPファクターは以下の式で定義される。

PCAP=(最も月次主成分得点が高い PF のリターン)

-(最も月次主成分得点が高い PF のリターン) … (4)

p: ファクター作成に使用した主成分

また、マルチファクターモデルは以下の式で定義し、回帰分析によって PCAPファクターの有効性を確認する。PCAPファクターが有効であるならば、アブノーマルリターン(α)は 0 から乖離しないこととなる。

$$r_i - r_f = \beta_i(r_M - r_f) + PCAP_p + \alpha \dots (5)$$

$r_i - r_f$: ポートフォリオの超過リターン

$r_M - r_f$: マーケットポートフォリオの超過リターン

PCAP_p: 主成分得点を利用したファクター

α : アブノーマルリターン

* r_M は対象銘柄 1501 社の平均リターンとする

3.3 分析結果

2003 年 1 月 1 日から 2008 年 12 月 31 日までの投稿を取得した結果、2,950,002 件の投稿が得られた。投稿を分析した結果、3,223,578 種の単語を抽出した。このうち出現頻度が 1000 回以上の単語は 4,885 種であった。4,885 種の単語について TF・IDF 値を求め特徴量を算出した。このうち合計値の大きい上位 700 単語を主成分分析の使用単語とした。

分析手順⑤によって、寄与率の高い上位 10 主成分について月次平均主成分得点を求めた。各銘柄の月次平均主成分得点と月次リターン・強気指数との相関係数、および無相関の検定の結果を表 1 に示す。

この結果、主成分 1、主成分 2、主成分 8 には月次平均主成分得点と月次リターンとの間に相関が確認できた。また、主成分 4、主成分 5、主成分 6、主成分 10 に関しては強気指数と相関が見られた。

表 1: 月次平均主成分得点と月次リターン・強気指数との相関

	月次リターン	強気指数
主成分 1	0.0223**	0.0025
主成分 2	0.0175**	-0.0036
主成分 3	0.0065	0.0075
主成分 4	-0.0032	-0.0362**
主成分 5	0.0061	-0.0374**
主成分 6	-0.0035	-0.0315**
主成分 7	0.0126*	0.0208**
主成分 8	0.0194**	-0.0009
主成分 9	-0.0097	0.0025
主成分 10	-0.0020	-0.0223**

(*5%有意、**1%有意)

表 2: 主成分 1 における回帰分析の結果

	α	β	P 値 (α)	P 値 (β)	決定係数
PF1	0.0032	1.013	0.001	0.469	0.979
PF2	0.0037	1.187	0.009	0.000	0.966
PF3	-0.0006	1.079	0.585	0.001	0.970
PF4	-0.0024	0.894	0.021	0.000	0.967
PF5	-0.0057	0.851	0.000	0.000	0.958
PFx	0.0004	0.922	0.751	0.003	0.949

表 3: 主成分 4 における回帰分析の結果

	α	β	P 値 (α)	P 値 (β)	決定係数
PF1	-0.0003	1.007	0.820	0.761	0.968
PF2	0.0010	1.065	0.460	0.011	0.964
PF3	-0.0004	1.040	0.695	0.051	0.974
PF4	-0.0012	1.009	0.192	0.626	0.978
PF5	-0.0009	0.902	0.353	0.000	0.971
PFx	0.0004	0.922	0.751	0.003	0.949

月次リターンと相関があり、かつ強気指数との相関がみられた主成分を確認出来たため、それぞれの平均主成分得点を用いて主成分ポートフォリオを作成した。作成したポートフォリオを利用し、CAPM におけるアブノーマルリターンの確認を回帰分析によって検証した。主成分 1 を利用したポートフォリオを利用した場合における、回帰分析の結果を表 2 に示す。

この結果より、PF1、PF2、PF4、PF5 においてアブノーマルリターン(α)の存在を確認した。また、主成分得点の高いポートフォリオほど 0 からプラスに乖離し、低いほどマイナスに乖離するという傾向が見られることが確認できた。また、リターンと相関が見られなかった主成分 4 にポートフォリオを作成し、同様に検証を行った(表 3)。その結果、相関がない主成分 4 を利用したポートフォリオを使用した場合には CAPM においてアブノーマルリターンは現れなかった。

表 4: マルチファクターモデルにおける回帰分析の結果 (PCA₁)

	α	β	PCA ₁	P 値(α)	P 値(β)	P 値(PCA ₁)	決定係数
PF1	-0.0002	0.9499	0.3878	0.817	0.000	0.000	0.985
PF2	0.0032	1.1775	0.0571	0.078	0.000	0.651	0.966
PF3	0.0009	1.1076	-0.1775	0.533	0.000	0.097	0.971
PF4	-0.0005	0.9283	-0.2144	0.694	0.000	0.020	0.969
PF5	-0.0002	0.9499	-0.6122	0.817	0.000	0.000	0.979
PFx	-0.0012	0.8929	0.1803	0.483	0.000	0.134	0.951

CAPM においてアブノーマルリターンが出現する主成分ポートフォリオの存在を確認出来たため、式(2)の定義に沿って新たなファクターを作成し、回帰分析による検証を行った。主成分 1 を利用したポートフォリオにおける回帰分析の結果を表 4 に示す。すべてのポートフォリオにおいて、 α が 0 から有意に乖離していないことが示されている。この結果より、主成分 1 を利用したファクターをマルチファクターモデルとして展開することで、アブノーマルリターンが出現しないことが確認できる。

4. 考察

株式リターンを説明する株式掲示板に基づくファクターを作成し検証した結果、新たに作成した PCA₁ファクターはリターンを説明しており、株式リターンを説明する株式掲示板に基づくファクターである可能性がある。また、このファクターは従来の BMB と異なるものであると考えられる。以下にその理由を述べる。

第 1 に、PCA₁ファクターを利用したマルチファクターモデルにおいて、アブノーマルリターンが出現していないことである。CAPM において、月次リターンと相関のある月次平均主成分得点を利用したポートフォリオにはアブノーマルリターンが出現していた。しかし PCA₁ファクターを作成しマルチファクターモデルとして展開した場合には、アブノーマルリターンは出現しなかった。この結果より、PCA₁ファクターはリターンを説明していると考えられる。

第 2 に、月次平均主成分得点と強気指数との間に相関がないことである。PCA₁ファクターは、BMB で利用されている強気指数と相関を持たない月次平均主成分得点を利用して作成している。このことから、PCA₁ファクターは従来の BMB とは相関を持たず、BMB とは異なったファクターであると考えられる。

5. 結論

本稿では株式掲示板に対する 2,950,002 件の投稿について形態素解析と主成分分析を行い、株式リターンとの関係を調べた。その結果、月次リターンと相関ある主成分によるポートフォリオを作成することで、CAPM にアブノーマルリターンが出現することを確認した。また、相関がない主成分によるポートフォリオの場合には、CAPM にアブノーマルリターンが出現しなかった。このことから、株式掲示板におけるアノマリーが存在している可能性がある。

また、株式掲示板の投稿内容より作成した PCA₁ファクターをマルチファクターモデルとして展開した結果、アブノーマルリターンは出現しなかった。この結果より、PCA₁ファクターはリター

ンを説明する可能性があり、株式リターンを説明する株式掲示板に基づくファクターであると考えられる。

参考文献

- [Antweiler 02] Antweiler, W. and M.Z. Frank, "Internet Stock Message Boards and Stock Returns", University of British Columbia Working Paper, 2002.
- [Tumarikin 01] Tumarikin, R. and Whitelaw, R.F., "News or Noise? Internet Postings and Stock Prices," Financial Analysts Journal, Vol.57, pp.41-51, 2001.
- [Jones 06] JONES, A., L., "Have internet message boards changed market behavior?," The journal of policy, regulation and strategy for telecommunications, Vol.8, No.5, pp.67-76, 2006.
- [Wysocoki 99] Wysocoki, 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.
- [阿部 09] 阿部有希, 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数および投稿内容を用いたファクターモデルの構築, 2009 年日本社会情報学会 (JSIS&JASI) 合同研究大会 研究発表論文集, pp.130-135, 2009.
- [阿部 10] 阿部有希, 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数および投稿内容を用いたファクターモデルの検討", 第 16 回社会情報システム学シンポジウム学術講演論文集, pp.27-32, 2010.
- [小林 09] 小林孝雄, 芹田敏夫, "新・証券投資論 1 理論編", 本経済新聞出版社, 2009.
- [丸山 08] 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指標の関係", 証券アナリストジャーナル, pp.110-127 2008.
- [和泉 10] 和泉潔, 後藤卓, 松井藤五郎, "テキスト情報による金融市場変動の要因分析", 人工知能学会論文誌, 25 巻, pp.383-387, 2010.