

インターネット株式掲示板の投稿内容を用いた ファクターモデルの検討

Development of Factor Model based on Stock BBS Posting

阿部有希¹ 梅原英一² 諏訪博彦¹ 太田敏澄¹

Yuki Abe¹, Eiichi Umehara², Hirohiko Suwa¹ and Toshizumi Ohta¹

¹ 電気通信大学

¹ University of Electro-Communications

² 野村総合研究所

² Nomura Research Institute

Abstract: We investigate whether the number and content of postings of the posting of BBS become the explanatory variables of the stocks return. In previous work, the possibility that these factors are related to the stocks return and risk is suggested. We verify whether CAPM approve a portfolio by the number and contents of postings in TSE. We find the possibility CAPM is not approved in a portfolio with the highest number of postings and the fewest number of postings as well as the highest number of bullish postings and the highest number of bear postings. We argue an index of the difference of the return of a portfolio with the highest number of postings and the return of a portfolio with the fewest number of postings as well as the posting contents.

1 はじめに

CGM(Consumer Generated Media)の増加が報告されている。その一つであるインターネット株式掲示板（以下、株式掲示板）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。そのため株式掲示板と株式市場との関連性に注目が集まっており、例えばAntweiler & Frankは、メッセージの投稿が株式指標と関連していること[1]、株式掲示板の投稿量がリターンと関連しており、掲示板活動が標準的なファクターモデルには組み込まれていない要因を反映することを報告している[2]。しかし、アセットプライシングモデルに株式掲示板の投稿をファクターとして加えた研究は、米国市場において投稿数をファクターとしたAntweiler & Frank[2]しか見当たらない。

本研究では日本市場において株式掲示板を利用した新たなファクターモデルを構築し、株式掲示板の投稿数および投稿内容が株式リターンの説明変数になるかを明らかにする。

そのため、2節において先行研究を概観し、それらを踏まえ、3節において仮説を設定する。仮説検証のための分析方法を4節において説明し、その結果と考察を5節において述べる。最後に、6節において結論と今後の展開を述べる。

2 先行研究

ファクターモデルとは、各銘柄・ポートフォリオのリターンをいくつかのファクターを用いて説明するモデルである。これを用いることで、投資家は、株式投資の意思決定支援を受ける事が出来る。代表的なモデルとしてSharp[4]らにより提唱された市場感応度をファクターとした資本資産価格モデル(CAPM)やFama & French[3]の3ファクターモデル等がある。Fama & Frenchは、CAPMのマーケットファクターに、SMBファクターとHMLファクターの2種類のファクターを追加した3ファクターモデルを構築し、その有効性を回帰モデルにより検証している。

丸山ら[5]は、日本の Yahoo! ファイナンス掲示板の投稿数 Top50 (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ている。彼らは、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(Mecab)、ノイズ除去(数字・英字・記号の除去、品詞の制限、否定語の処理)、TF・IDF法)と機械学習(Support Vector Regression(以下 SVR))を用いて、メッセージを「弱気」「中立」「強気」の三種類に分類している。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、投稿内容分析を考慮して考えると、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こと、第三に強気投稿数と弱気投稿数の差により作成した掲示板指標「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。

Antiweiler & Frank[2]は、米国の Yahoo! 掲示板を 1999 年から 2001 年の期間で、上場米国企業の 3500 万以上の投稿を分析し、仮のものであるという注釈のもと、CAPM や Fama & French[3] のモデルに、新しいファクターとして掲示板投稿数ファクター NMQ を加えた所、投稿数は株式リターンに影響を与えている可能性を報告している。

本研究では、これらの先行研究を踏まえ、次節で設定する仮説に基づき、検証を行う。

3 仮説設定

3.1 投稿数に関する仮説

本研究では、Antiweiler & Frank の手法を用いて、日本の Yahoo! 株式掲示板における投稿数ファクター NMQ を作成する。それを CAPM に加えることで投稿数が株式リターンの説明変数になるか検証を行う。投稿数ファクター NMQ は、東証 1 部の企業を投稿数により単純 5 分割し、ポートフォリオを作成する。このポートフォリオが CAMP のマーケットファクターに対してアブノーマルリターンを持つならば、 y 切片が 0 から乖離 (ジェンセンの α) する。また、株式掲示板の投稿がそれ自身ファクターのとして働くならば、投稿数の大きなポートフォリオから小さなポートフォリオになるにつれてジェンセンの α における単調パターンを観測することが出来ることとなる。

また丸山ら[5]より、投稿数がボラティリティと関係している事が示唆されている。投稿数が多いほど、ボラティリティは高くなるならば、投稿数の違いにより CAMP の市場感応度 (β) が変化する。そして、これらが確認されたファクターを加えたファクターモデルは CAPM のマーケットファクターと同等に働くと考えられる。これらに基づき以下の仮説を設定する。

仮説 1: 投稿数が大きなポートフォリオはジェンセンの α の値が 0 から正に乖離し、投稿数が小さなポートフォリオは 0 から負に乖離する。

仮説 2: 投稿数が大きなポートフォリオは β の値が 1 から正に乖離し、投稿数が小さなポートフォリオは 1 から負に乖離する。

仮説 3: 投稿数ファクター NMQ は CAPM に加えると、全てのポートフォリオにおいてジェンセンの α は、0 から乖離しない。

3.2 投稿内容に関する仮説

丸山ら[5]により、投稿内容と株式リターンの関連性が示唆されている。そのため強気指数により投稿内容ファクター BMB を作成し、CAPM にそれを加えることで投稿内容が株式リターンの説明変数になるか検証を行う。そこで 3.1 節と同様に、以下の仮説を設定する。

仮説 4: 強気指数が大きなポートフォリオはジェンセンの α の値が 0 から正に乖離し、強気指数が小さなポートフォリオは 0 から負に乖離する。

仮説 5: 強気指数が大きなポートフォリオは β の値が 1 から正に乖離し、強気指数が小さなポートフォリオは 1 から負に乖離する。

仮説 6: 投稿内容ファクター BMB は CAPM に加えると、全てのポートフォリオにおいてジェンセンの α は、0 から乖離しない。

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

本研究の分析対象企業は、Yahoo!株式掲示板における東証一部上場銘柄のうち取得可能であった1360社である。分析対象期間は、2003/01/01から2008/12/31までの6年間である。

図1は、2005年1月の掲示板のランクに対する投稿数（対数スケール）をプロットしたものである。1番目の企業は最大の投稿数を持ち、2番目の企業は2番目に大きな投稿数を持つ。この結果から、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とが分かる。中央値は180社目である。このことから掲示板投稿と株式市場の関係は投稿数の多い企業と少ない企業では分けて考えられる事が見受けられる。この結果、ポートフォリオの作成を投稿数の多い銘柄、少ない銘柄で単純に5分割しても、投稿数という特徴を捉えたポートフォリオを作ることが出来る。

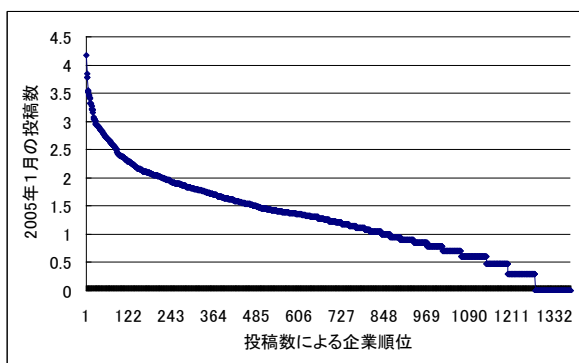


図1 企業順位による投稿数のグラフ

4.2 株式指標

分析対象の株式指標は「超過株価リターン(AR(t))」により表す。株式指標は東洋経済の「株価CD-EOM2009」を使う。資本移動や額面変更があった場合、株価は調整係数で調整し、連続性を保障する。

月次ベースの株価リターン $R(t)$ は、 $R(t) = \text{終値}(t) / \text{終値}(t-1) - 1$ とする。超過リターン $AR(t)$ は、 $R(t)$ より東証株価指数(TOPIX)の月次リターンを差し引く事により算出する。後述のポートフォリオのリターンは、各ポートフォリオに分類されている銘柄のリターンの平均として定義する。

4.3 投稿数ポートフォリオ

4.3.1 投稿数ポートフォリオの作成

仮説1・仮説2・仮説3を検証するため、投稿数によるポートフォリオの作成を行う。まず、投稿の全くない企業をポートフォリオXとして分類した。その後各月でその月の投稿数に従い、残りの対象銘柄を5分割する。ポートフォリオAはその月でもっとも投稿数の多い企業群で、ポートフォリオEは最も投稿の少ない企業群である。ポートフォリオA～Eに分類された企業は月平均約193社、ポートフォリオXに分類された企業は月平均約422社である。

4.3.2 投稿数ポートフォリオのリターン

ポートフォリオ別の月別超過リターンの推移を図2に示す。ポートフォリオAの平均超過リターンは最も高く、ポートフォリオEの超過リターンが最も低い。投稿の多い企業ほど超過リターンが高い結果になっていることは、投稿数が超過リターンと関連している可能性を示唆している。また、ポートフォリオAとTOPIXの軌跡に注目すると、ポートフォリオAは、ほぼ全期間においてリターンが最も高いポートフォリオであった。ただし、TOPIXが急激に低下した時、ポートフォリオAのリターンは急激に低下した。これはポートフォリオAがCAPMの意味で非常にリスクな高 β である事を示唆している。このことから、投稿数が株式リスクと関連している可能性が考えられる。

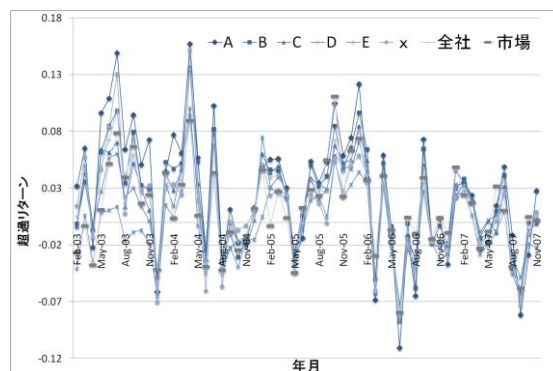


図2 投稿数ポートフォリオ別月別リターンの推移

4.4 投稿内容ポートフォリオ

4.4.1 強気・弱気分類

丸山ら[5]のプログラムを用いて対象期間の全銘柄の投稿メッセージを強気・中立・弱気に

分類した。その結果、強気に分類される投稿が多いことが確認されている。

強気指数とは、月毎における強気/弱気意見の多寡を示す掲示板指標であり、①式で定義する。

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

図3に、強気指数の高低による企業順位毎の強気指数のグラフを示す。1番目の企業は最大の強気指数を持ち、2番目の企業は2番目に大きな強気指数の値を持つ。結果として投稿数と同様に、強気指数が非常に高い値の掲示板と、低い掲示板がある。この事から掲示板投稿と株式市場の関係は強気指数の高い企業と小さい企業では分けて考えられる事が見受けられる。

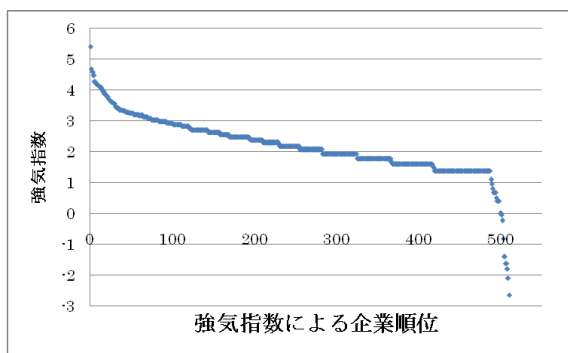


図3 企業順位における強気指数

4.4.2 投稿内容ポートフォリオの作成

仮説4、仮説5、仮説6を検証するため、強気指数によるポートフォリオの作成を行う。丸山ら[5]は強気指数を定義する際、銘柄ごとに8時間の時間の区切りで強気投稿数と弱気投稿数の和が合計3未満のものを分析から除外したが、本研究では1月単位で強気・弱気投稿数の和が3未満のものは強気指数を判断できないと考えポートフォリオXとして分類する。その後各月で強気指数に従い、残りの対象銘柄を5分割する。ポートフォリオ1はその月でもっとも強気指数の高い企業群で、ポートフォリオ5は最も強気指数の低い企業群である。

ポートフォリオ1~5に分類された企業は月平均約129社、ポートフォリオXに分類された企業は月平均約745社である。強気指数の最も高いポートフォリオ1は強気指数が平均3.41である。逆に最も投稿の少ないポートフォリオ5は平均1.03である。

4.4.3 投稿内容ポートフォリオのリターン

ポートフォリオ別の月別超過リターンの推

移を図4に示す。ポートフォリオ1の超過リターンが最も高く、ポートフォリオ5の超過リターンが最も低い。強気指数の高い企業ほど超過リターンが高い結果になっていることは、強気指数が超過リターンと関連している可能性を示唆している。また、ポートフォリオ1とTOPIXの軌跡に注目すると、ポートフォリオ1は、投稿数のポートフォリオAと同様に、ほぼ全期間においてリターンが最も高いポートフォリオであった。ただし、TOPIXが急激に低下した時、ポートフォリオ1のリターンは急激に低下した。これはポートフォリオ1がCAPMの意味で非常にリスクな高βである事を示唆している。このことから、投稿内容が株式リスクと関連している可能性がある。

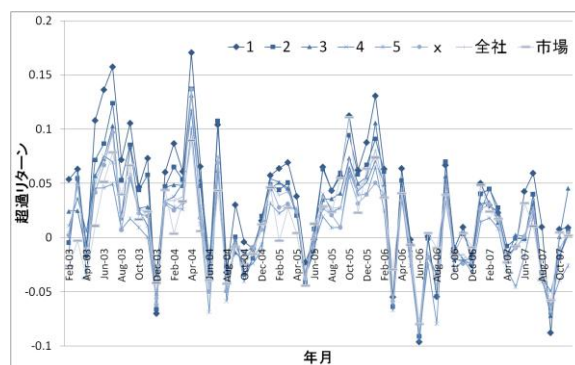


図4 投稿内容ポートフォリオ別月別超過リターン

4.5 ポートフォリオのアブノーマルリターンの算出

TOPIX をベンチマークとした CAPM を、4.4 節、4.5 節で示した投稿数および投稿内容ポートフォリオに対して適用する事で、各ポートフォリオ別のジェンセンの α の導出を行う。ここで CAPM を利用した回帰モデルは、②式で定義する。

$$r_i - r_f = \beta_i(r_M - r_f) + \alpha_i \dots \textcircled{2}$$

r_M は市場のリターンであり、ここでは TOPIX のリターンを使用する。 r_i は投稿数ポートフォリオ i のリターン、 r_f は無リスクリターンであり、今回は日本銀行発表の無担保コールレートをを用いる。また β_i はポートフォリオ i の市場感応度、 α_i はジェンセンの α と呼ばれ、CAPM のアブノーマルリターンを表している。 α の検定は、 α が 0 から有意に乖離するという帰無仮説のもと行う。同様にして β の検定は 1 から有意に乖離するという帰無仮説のもと行う。

4.6 掲示板ファクターモデルの構築

4.6 節の回帰モデルでアブノーマルリターンが観察された場合、掲示板の投稿数および投稿内容（強気指数）が株式リターンの説明変数になる可能性が示唆される。Fama & French[3]の手法を用い、掲示板ファクターNMQとBMBを作成する。NMQ(Noise minus Quiet)は投稿数ファクターであり、③式で定義する。

$$NMQ = (\text{投稿数の最も多いポートフォリオのリターン}) - (\text{投稿数の最も少ないポートフォリオのリターン}) \quad \dots \textcircled{3}$$

BMB(Bull minus Bear)は本研究でオリジナルに作成する投稿内容ファクターであり、④式で定義する。

$$BMB = (\text{最も強気なポートフォリオのリターン}) - (\text{最も弱気なポートフォリオのリターン}) \quad \dots \textcircled{4}$$

NMQおよびBMBをCAPMに加えた掲示板ファクターモデルとしてそれぞれ⑤、⑥式で定義する。

$$r_i - r_f = \beta_i(r_M - r_f) + n_i NMQ \quad \dots \textcircled{5}$$

$$r_i - r_f = \beta_i(r_M - r_f) + b_i BMB \quad \dots \textcircled{6}$$

ここで、⑤、⑥式によりポートフォリオのリターンが全て表現されるなら、⑤、⑥式を検証するための回帰モデル⑦、⑧式の α は0から乖離しないはずである。

$$r_i - r_f = \beta_i(r_M - r_f) + n_i NMQ + \alpha_i \quad \dots \textcircled{7}$$

$$r_i - r_f = \beta_i(r_M - r_f) + n_i NMQ + \alpha_i \quad \dots \textcircled{8}$$

5 分析結果と考察

5.1 投稿数に関する結果と考察

投稿数ポートフォリオのリターンをCAPMに基づき分析し、ジェンセンの α と β を求めた。その結果を表1と図5、図6に表す。ここで全社ポートフォリオとは、分析対象銘柄全てを含んだポートフォリオである。

表1 投稿数ポートフォリオ別 α と β

	A	B	C	D	E	x	全社
投稿数切片	0.01217	0.008205	0.00565	0.002192	-0.01372	0.000588	0.002692
β	1.266749	1.073227	0.937354	0.856014	0.75161	0.990579	1.006201
α 値(切片)	0.0049	0.004865	0.052716	0.403596	1.03E-06	0.863332	0.335209
β 値(β)	0.013835	0.304784	0.388814	0.03288	0.000236	0.913077	0.929705

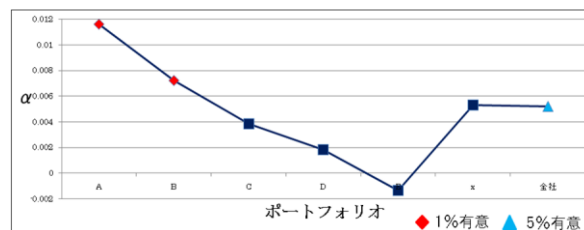


図5 投稿数ポートフォリオによる α

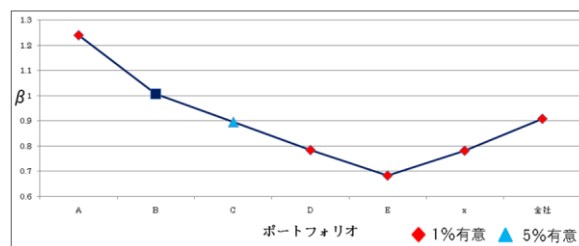


図6 投稿数ポートフォリオによる β

ジェンセンの α は投稿数が大きいポートフォリオAおよびBにおいて0から正に乖離している。この結果は、仮説1を一部支持している。また β について、投稿数が大きいポートフォリオAにおいて1から正に乖離し、投稿数の小さいポートフォリオDおよびEにおいて1から負に乖離している。この結果は、仮説2を支持している。人々はリスクの高い銘柄に投稿している可能性がある。

以上の結果から、投稿数が株式リターンの説明変数になる可能性が示唆されたため、仮説3の検証を行っている。NMQとCAPMのマーケットファクターだけで市場が説明できるのであれば、⑦式の α は観測されないはずである。結果を表2に示す。掲示板投稿数ファクターNMQモデルは、各ポートフォリオに対してアブノーマルリターンが観測されていない。よって仮説3は支持される。これはAntweiler & Frank[2]の結果を支持し、日本において投稿数が株式リターンの説明変数になる可能性を示す結果である。

表2 掲示板投稿数ファクターNMQモデル

NMQモデル	α	β	NMQ	P値(α)	P値(β)	P値(NMQ)	決定係数
ポートフォリオA	-0.00026	0.5414	0.4898	0.9129	3.63E-18	3.25E-16	0.9399
ポートフォリオB	0.002862	0.7640	0.2274	0.2336	3.07E-20	0.000243	0.9020
ポートフォリオC	0.0013	0.8043	0.1464	0.6304	7.07E-17	0.048619	0.8489
ポートフォリオD	0.000745	0.8496	0.0713	0.7760	3E-16	0.373585	0.8199
ポートフォリオE	-0.00026	0.9604	-0.0799	0.9129	3.63E-18	0.331031	0.8110
ポートフォリオx	0.000708	0.6288	0.2760	0.8346	9.07E-09	0.005441	0.7374
ポートフォリオ全社	0.000992	0.7365	0.2386	0.6962	1.5E-16	0.000786	0.8690

5.2 投稿内容に関する結果と考察

5.1 節と同様の手法を用い、投稿内容（強気指数）ポートフォリオの α と β を算出した。結果を表 3 と図 7、図 8 に示す。

表 3 投稿内容ポートフォリオ別 α と β

	1	2	3	4	5	x	全社
投稿内容切片	0.020491	0.010956	0.007534	0.003391	-0.00961	-0.00064	0.002692
β	1.300571	1.167642	0.998574	0.931346	0.896032	0.914312	1.006201
σ 値(切片)	6.5E-06	0.001177	0.015582	0.261649	0.003886	0.82068	0.335209
σ 値(β)	0.005484	0.043023	0.985163	0.367467	0.202659	0.234944	0.929705

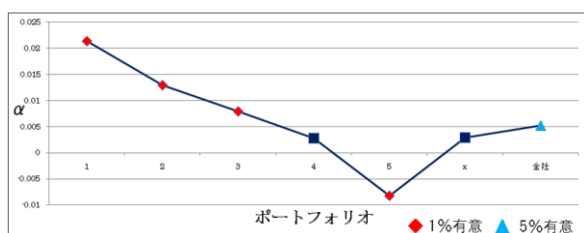


図 7 投稿内容ポートフォリオの α

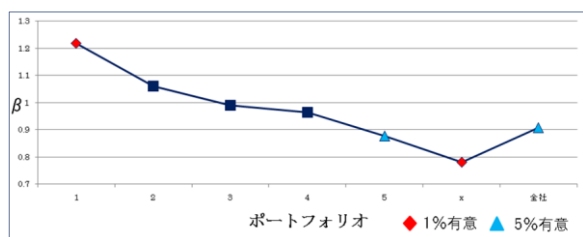


図 8 投稿内容ポートフォリオの β

ジェンセンの α は強気指数が大きいポートフォリオ 1 および 2、3 において 0 から正に乖離し、強気指数の小さいポートフォリオ 5 において 0 から負に乖離している。この結果は、仮説 4 を支持している。また β について、強気指数が大きいポートフォリオ 1 において 1 から正に乖離し、ポートフォリオ 5 において 1 から負に乖離している。この結果は、仮説 5 を支持している。人々はリスクの高い銘柄に強気な投稿をしている可能性がある。

以上の結果から、投稿内容が株式リターンの説明変数になる可能性が示唆されたため、仮説 6 の検証を行っている。BMB と CAPM のマーケットファクターだけで市場が説明できるのであれば、⑧式の α は観測されないはずである。結果を表 4 に示す。掲示板投稿内容ファクター BMB モデルは、ポートフォリオ 1 および 5 に対してアブノーマルリターンが観測されている。よって仮説 6 は棄却される。

表 4 掲示板投稿数ファクター BMB モデル

BMBモデル	α	β	BMB	P値(α)	P値(β)	P値(BMB)	決定係数
ポートフォリオ1	-0.00882	0.6479	0.4017	0.0393	3.4E-22	6.42E-13	0.9161
ポートフォリオ2	-0.00049	0.7849	0.2072	0.9139	1.92E-21	0.00054	0.8683
ポートフォリオ3	-0.00688	0.7621	0.2441	0.0924	8.27E-22	2.92E-05	0.8798
ポートフォリオ4	-0.00509	0.8342	0.1328	0.2399	8.26E-22	0.0293	0.8561
ポートフォリオ5	-0.00882	0.9075	0.0101	0.0393	3.4E-22	0.8747	0.8354
ポートフォリオx	-0.00777	0.7473	0.2156	0.0635	9.09E-17	0.0023	0.8123
ポートフォリオ全社	-0.00588	0.7903	0.2008	0.1314	1.15E-21	0.0007	0.8691

6. 結論と今後の展開

本研究では投稿数および投稿内容が株式リターンの説明変数になる可能性と、それに伴うリスクの代理指標である市場感応度 (β) について検証している。その結果、株式掲示板の投稿数および投稿内容が株式リターンの説明変数になる可能性が示唆されたため、掲示板ファクターとして NMQ と BMB を用いたファクターモデルの作成を行い検証している。結論として、先行研究で提案された投稿数ファクター NMQ モデルが支持されることを確認している。

本研究の問題として、各銘柄の規模が考慮できていないことがあげられる。このため、小型株の影響を受けている可能性が否定できない。今後、Fama & French[3]の 3ファクターモデルへの拡張することで、対応したいと考える。

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] Antweiler, W. and Frank, M. Z., "Internet Stock Message Boards and Stock Returns", University of British Columbia Working Paper, 2002. <http://strategy.sauder.ubc.ca/antweiler/public/returns.pdf> (2009/05/25 現在)
- [3] Fama, E. F. and French, K. R. "Common risk factors in the returns on stocks and bonds," Journal of Financial Economics, 33, 3-56, 1993.
- [4] Sharp, W. F. "Capital asset prices: A theory of market equilibrium under conditions of risk," Journal of Finance, Vol.19 No.3, pp.425-442.,1964.
- [5] 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 pp.110-127, 2008.