

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 13 日現在

機関番号：34504

研究種目：基盤研究(B) (一般)

研究期間：2013～2015

課題番号：25285127

研究課題名(和文)大規模テキストデータを活用した投資家心理と株価変動の定量的解明

研究課題名(英文)Behavioral anatomy of asset price fluctuations based on large textual data.

研究代表者

岡田 克彦 (OKADA, Katsuhiko)

関西学院大学・経営戦略研究科・教授

研究者番号：90411793

交付決定額(研究期間全体)：(直接経費) 13,600,000円

研究成果の概要(和文)：本研究では、ニュースデータ、ブログデータ、SNSなどのソーシャルデータを大規模に収集し、それらがどの銘柄について語っているものなのかを自然言語処理の諸技術を用いてデータベース化した。この結果、全上場企業約3600社それぞれについて、ニュースおよびソーシャルメディアの情報を紐付けていることになる。次に、市場に流れるコメントやニュースについて、ポジティブな文脈で語られているのか、あるいはネガティブなのかについて、評価表現辞書を作成することでスコアリングした。こうして作成した指標をセンチメント指数として時系列で捉え、金融市場における様々なアナノミー現象と、センチメント指数との関連性を明らかにした。

研究成果の概要(英文)：The purpose of this research is to identify the role of investor psychology in the stock market. For this purpose, we have collected a massive amount of news data as well as the social data. (ex. 5.7 million news paper articles. Five years of data in the stock market messages boards. Past one year data of the number of visitors of company web-site.) These large textual data are then organized in a way that we can identify to what company each news text belongs. Simultaneously, we have created a dictionary in order to determine whether the text is positive or negative. Based on these data base, we have developed a time varying "Sentiment index" for each one of the listed stocks.

We found stock price reactions to a fundamental shock such as earnings announcement surprise / securities analysts' rating revisions are subdued when the sentiment index is in line with the shock, whereas it becomes large when the sentiment is opposite of the surprise.

研究分野：行動ファイナンス、人工知能、

キーワード：ビッグデータ 人工知能 テキストマイニング 投資家心理 ハーディング 行動ファイナンス アノマリー

様式 C - 19、F - 19、Z - 19 (共通)

1. 研究開始当初の背景

これまでのファイナンス研究のパラダイムは、すべての公開情報は株式価格に反映されていると考える、効率的市場仮説であった。効率的市場仮説に従えば、高いリターンは高いリスクを取った結果達成可能なものであり、市場では、ハイリスク・ハイリターン、ローリスク・ローリターンの関係が成立している。これまでのファイナンス理論研究の多くは、こうしたリスクとリターンの関係を定式化するものである。並行して実施された多くの実証研究によって、リターンとリスクには定式化された通りの関係があることが明らかにされている一方、効率的市場仮説とは相容れない市場価格形成も多く報告されている。

こうした中、株式価格はリスクファクターで説明しきれぬのか、否かという議論が2000年代に入って活発となってきた。この議論についてはいまだに決着しておらず、2013年のノーベル経済学賞は、効率的市場を主張するシカゴ大学の Eugene Fama 氏と、市場は投資家心理によって均衡価格から乖離すると主張するイェール大学の Robert Shiller 教授の両者に与えられた。

効率的市場仮説に対する考え方として市場の合理性を疑うのが、行動ファイナンス領域の研究者である。投資家は人間であるがゆえに、リスク評価についても、必ずしもすべて合理的に決定されているのではなく、感情やムード、雰囲気によって価格形成されているという仮説を支持する。極論すれば、市場で取引される企業価値は、投資家の心理状態によって左右されることになり、こうした考え方は、合理的経済人（ホモ・エコノミカス）に依拠したこれまでのファイナンス理論の研究者にとっては、パラダイムシフトを意味する。

市場価格が定式化された理論モデル通りに形成されていないという証拠は多くある

ものの、収益の予測可能性 (return predictability) を示すことは、容易ではない。ここに、Bad model の可能性が指摘される余地がある。すなわち、リスクとリターンの関係性についての定式化の精度に問題があるのであり、ハイリスク・ハイリターンの関係は厳然と存在するというのである。ファイナンス研究の世界にパラダイムシフトが起こっているのではなく、リスクを記述するモデルができていないだけだという主張である。

本研究では、伝統的な投資家心理の代理変数である出来高や売買代金ではなく、大規模テキストデータを自然言語処理の技術を援用することによって解析し、新たな代理変数を作成し、それが収益の予測可能性 (return predictability) を持つかどうかを研究する。

2. 研究の目的

本研究では、大規模テキストデータ（ニュース、掲示板、ブログ等）を解析し、投資家心理が資産価格に影響を与えているかどうかを検証する。

従来のアセットプライシングモデルは、リスクファクターが証券のリターンを決定すると考えている。市場がある程度効率的であれば、すべてのニュースがもたらす重要なファンダメンタル情報は既に株価に反映されていると考えられる。したがって、ニュースの内容を吟味しても、その後のリターンを予測することは不可能である。また、Twitter や SNS、ブログ、などのソーシャルメディアに投稿されるテキスト情報については、ファンダメンタルズに関する重要な情報を含んでいないので、情報価値としては低く、ただのノイズだと考えられる。ノイズであればそれぞれが相殺され、資産価

格に対してシステマチックな影響は与えない。

しかしながら、現実の株式市場では、定式化された Asset pricing model では説明できないアノマリー現象が数多く報告されており、これらすべてをリスクの定式化の不十分さのみ帰するには無理がある。本研究の目的は、本来ファンダメンタルズ情報ではないノイズ情報の中から、市場のムードや雰囲気(市場センチメント)を抽出し、それが資産価格形成にどのような影響があるかを明らかにすることである。

3. 研究の方法

3-1 データベースの構築

まず、大規模テキストデータベースを作成した。マスメディアの記事の分類を、記事の中に含まれる特徴語を用いて行う。具体的には「業績関連」「株価水準関連」「対象企業の環境関連」「企業イベント関連」「組織関連」の5分野である。それぞれのカテゴリーの中で、どのような表現が増えることが株価に影響を与えるのか、どのような表現が用いられれば株価への影響度が高いのかを探る準備作業を行う。

同様にソーシャルメディアの記事の分類、Twitter、個人投資家によるブログ、株式投資関連投稿記事を先の分類基準にしたがって作成する。ソーシャルメディアにおける投資家心理(センチメント)を推定するためには、ソーシャル独特の言い回しに注意する必要があり、ニュースとは別の辞書を作成した。

3-2 センチメント指数の算出

日本の株式市場には、約 3600 社の上場銘柄が存在するが、そのすべてについて市場ムードや雰囲気を定量的に表すものとして、当該銘柄のセンチメントを計算した。そし

て、これら合算したものをセンチメント指数として算出している。センチメント指数を時系列で表示することで、株価そのものの動きに加えて、市場センチメントの変化する度合いを時系列でとらえることができる。

3-3 センチメント指数とマーケットアノマリーの関連性研究

マスメディアとソーシャルメディアの大規模テキストデータから、全上場銘柄についてのセンチメントを作成することが可能なインフラが整備できたので、これをもってこれまでパズルと考えられてきた様々なマーケットアノマリーについて、投資家心理との関連性を検証する。

4. 研究成果

本研究の研究成果を2つの側面に分けて報告する。一つは、研究インフラの整備という側面である。本研究プロジェクトで整備されたインフラは、今後のビッグデータと金融市場の研究を行う上で、基盤となるものであり、今後継続的に研究していく礎となる。もう一つは、大規模テキストデータを活用したことにより明らかにされた、投資家心理と資産価格のメカニズムである。

(1) 研究インフラの構築

大規模ニュース記事データベース及び全上場企業に紐付いた形でのソーシャルデータベースの構築：金融情報ベンダーが提供する過去10年分の日本語ニュース(570万記事、3000万文章)及び、過去5年分の掲示板などのソーシャルデータ、過去1年分の企業ホーム

ページへのアクセスデータ等をデータベース化した。

評価表現辞書の構築・整備：570 万のニュースからセンチメントの形成に影響を与えると考えられる極性付き評価表現辞書を、評価表現候補の中から選択し構築した。

大規模データ処理技術の開発：定型化されたデータを高速に処理するために、テキストデータ処理システムを改良した。これまで開発してきた

KGMOD(<http://kgmod.jp/>)を改良し、一般の PC でも数億行ギガ単位のデータも効率的に扱うことができる体制を整え、およびの部分で作業効率を大幅に改善した。

(2) 投資家心理と資産価格の関連性研究

を活かし、証券アナリストの正の評価変動に際して、投資家センチメントが低い銘柄ほど大きな正の反応を示すことや、投資家が見過ごしている企業ほど業績サプライズに対して長期的に正の反応を示すことなど、投資家センチメントの状態によって株価の反応が大きく異なることを明らかにした。これは、投資家センチメントのありようと、投資家ハーディングが深く関係している証拠として捉えることができ、大規模テキストデータを扱う今後の研究の方向性に有用な示唆を与えるものである。

①-③を活かし、マーケットの雰囲気やムード(センチメント)を指標化したものを用いて、時系列での市場センチメントの変化と株価収益率の変化を可視化した。また、の成果にもセンチメント指数を用いている。この研究成果については、他の研究者の利用に資

するために、www.nysol.jp において公開している。

この公開サイトから、センチメント指数のダウンロードというボタンを押すと、誰でも下図に示す様な時系列のセンチメント指数を CSV でダウンロードできる。これらの値についての算出根拠についてもサイト上に詳細に示している。

date	Pos	Neg	PosNeg	siPos	siNeg	siPosNeg	topix
20120208	5007	-2306	2701	3218.363636	-2614.896364	603.4772727	782.34
20120209	3758	-2826	932	3237.954545	-2652.227273	585.7272727	784.49
20120210	4368	-3190	1178	3288.204545	-2650.795455	637.4090909	779.07
20120213	3003	-1781	1222	3275.681818	-2658.386364	617.2954545	781.68
20120214	3184	-1879	1305	3288.068182	-2656.045455	632.0227273	786.8
20120215	6025	-2252	3773	3383.454545	-2633.568182	749.8863636	802.96
20120216	3813	-3009	804	3388.954545	-2658.818182	730.1363636	800.25
20120217	5159	-1434	3725	3458.886364	-2623.954545	834.9318182	810.45
20120220	4075	-1733	2342	3491.613636	-2559.681818	931.9318182	819.03
20120221	3520	-2427	1093	3532.5	-2515.090909	1017.409091	816.29
20120222	4789	-1716	3073	3562.613636	-2481.068182	1081.545455	825.4
20120223	4084	-2151	1933	3611.295455	-2426.204545	1185.090909	829.35
20120224	4974	-1532	3442	3655.795455	-2398.181818	1257.613636	834.29
20120227	4301	-2678	1623	3667.659091	-2425.340909	1242.318182	835.25
20120228	2666	-4439	-1773	3649.681818	-2461.090909	1188.590909	838.48
20120229	4103	-2015	2088	3689.113636	-2467.590909	1221.522727	835.96
20120301	3012	-2615	397	3706.045455	-2476.409091	1229.636364	831.54
20120302	3862	-1937	1925	3733.25	-2467.977273	1265.272727	837.82
20120305	2714	-2960	154	3757.204545	-2455.159091	1302.045455	832.86
20120306	3097	-3846	-149	3780.227273	-2510.863636	1269.363636	827.35
20120307	2893	-5178	-2285	3758.545455	-2601.636364	1156.909091	822.71
20120308	5099	-1810	3289	3812.772727	-2588.363636	1224.409091	836.16
20120309	4132	-1898	2234	3848.954545	-2550.545455	1298.409091	848.71
20120312	4327	-2172	2155	3871.704545	-2553.409091	1318.295455	845.28
20120313	3966	-2091	1875	3867.295455	-2553.818182	1313.477273	845.33
20120314	4614	-2145	2469	3919.863636	-2545.545455	1374.318182	857.11
20120315	4465	-2311	2154	3937.931818	-2553.636364	1384.295455	863.61
20120316	4319	-1648	2671	3996.545455	-2510.340909	1486.204545	866.73
20120319	4559	-1724	2835	4016.227273	-2504.613636	1511.613636	868.35
20120321	3113	-3447	-334	3997.181818	-2539.931818	1457.25	858.78
20120322	3743	-2067	1676	4000.568182	-2540.181818	1460.386364	862.07

(センチメント指数ダウンロードのイメージ図)

5. 主な発表論文等

[雑誌論文] (計 9 件)

Yuki Kobayashi, Yuya Higashikawa, Naoki Katoh, Adnan Sijoka, “Characterizing Redundant Rigidity and Redundant Global Rigidity of Body-Hinge Graphs,” Information Processing Letters, Vol 116, pp175-178. 2016. (査読有)
Yuya Higashikawa, John Augustine, Siu-Wing Cheng, Mordecai J. Golin, Naoki Katoh, Guanqun Ni, Bing Su and Yin Feng Xu. “Minimax regret 1-sink location problem in dynamic path networks,” Theoretical Computer Science, Vol. 588, pp24-36. 2015. (査読有)

Takanobu Nakahara, Takeaki Uno, and Yukinobu Hamuro, “Prediction Model Using Micro-clustering”, Knowledge-Based and Intelligent Information & Engineering Systems 18th Annual Conference, KES-2014 Gdynia, Poland, September 2014 Proceedings, DOI: 10.1016/j.procs.2014.08.231, Volume 35, pp.1488–1494, 2014. (査読有)

岡田克彦、佐伯政男「注意力の限界と Post-Earnings-Announcement-Drift」証券アナリストジャーナル、Vol.52. pp72-82, 2014. (査読有)

Takahiro Azuma, Katsuhiko Okada and Yukinobu Hamuro, “Is No News Good News? : The Streaming News Effect on Investor Behavior surrounding Analyst Stock Revision Announcement,” International Review of Finance, Vol. 14. Pp29-51. 2014 (査読有)

Takanobu Nakahara, Yukinobu Hamuro, “Detecting topics from Twitter posts during TV program viewing”, MoDAT in conjunction with IEEE ICDM 2013 in Dallas, Texas. DOI: 10.1109/ICDMW.2013.48. 2013. (査読有)

Shigeki Sakakibara, Takashi Yamasaki and Katsuhiko Okada, “The Calendar Structure of the Japanese Stock Market;, The ‘Sell in May Effect’ versus the ‘Dekansho-bushi Effect’, International Review of Finance, Vol. 13, pp161-185. 2013. (査読有)

岡田克彦、山崎高広、榊原茂樹、山崎尚志、「株価変動の季節性と投資家心理」証券アナリストジャーナル、Vol.51. pp96-103. 2013. (査読有)

前川浩基、中原孝信、岡田克彦、羽室行信「大規模ニュース記事からの極性

付き評価表現の抽出と株価収益率の予測」オペレーションズ・リサーチ、Vol.58. pp281-288. 2013. (査読有)

〔学会発表〕(計 11 件)

Katsuhiko Okada, Hidenori Takahashi, “A page-view based investor attention and IPO pricing,” Asian Finance Conference in Bangkok, Thailand, Jul. 16. 2016.

Katsuhiko Okada, Masao Saeki, “Inattentive blindness and Post-Earnings Announcement Drift”, World Banking and Finance Conference in Hanoi, Vietnam. Dec. 18. 2015.

岡田克彦・羽室行信 「Period Mining Model」宇野 CREST×SOLEIL DATA DOJO 第一回シンポジウム グランフロント会議場(大阪府・大阪市) 2015.12.12.

羽室行信「Period Mining で探る株価の季節性」行動経済学会、近畿大学(大阪府・東大阪市) 2015. 11 28.

岡田克彦・羽室行信、「グラフ密度による相場大底の検知」人工知能学会、金融情報学・部会、ブルームバーグ東京支店(東京都、千代田区) 2015.10.13.

Masakazu Nakamoto, Stephane Cheung, Yukinobu Hamuro, NYSOL: Command Line Tools for Processing a Large Scale CSV Data, ISM HPCCON Oct. 10. 2015, Tokyo,

Katsuhiko Okada, Yukinobu Hamuro, “Signaling the Rock Bottom,” International Federation of Technical Analysis in Tokyo. Oct. 3. 2015.

Yuya Higashikawa, Remy Belmonte, Naoki Katoh, Yoshio Okamoto, “Polynomial-time approximability of the k-Sink Location problem, Operations

Research, University of Vienna, Vienna, Austria, 2015.9.2.

中原孝信, 宇野毅明, 岡田克彦, 羽室行信, 「強化学習を用いたブーム検知型トレーディングシステムの構築」, 第11回 人工知能学会 金融情報学研究会 (SIG-FIN), 東京大学(東京都文京区), 2013.10.12.

中原孝信, 宇野毅明, 岡田克彦, 羽室行信, 「強化学習を用いた相場のブーム検知と株取引への適用」, 2013年度人工知能学会(第27回), 富山 国際会議場(富山県・富山市), 2013.6.4.

前川浩基, 中原孝信, 岡田克彦, 羽室行信, 「大規模ニュース記事を用いた株価収益率の予測」, 日本オペレーションズ・リサーチ学会, 2013年春季研究発表会, pp.108-109. 東京大学(東京都文京区) 2013.3.5.

[図書] (計 3 件)

Takahiro Azuma, Katsuhiko Okada, Yukinobu Hamuro, Shigeki Sakakibara, Takashi Yamasaki, Nobuyuki Isagawa, Kenya Fujiwara, “Behavioral Interactions, Markets and Economic Dynamics,” Springer, 2015, 663.

Katsuhiko Okada (Editor, Parrisa Haghirián), “Routledge Handbook of Japanese Business and Management,” Routledge 2015, 404.

岡田克彦, 『ビッグデータで株価を読む』中央経済社、2014年、288.

[その他]

ホームページ等

www.nysol.jp

6 . 研究組織

(1)研究代表者

岡田 克彦 (OKADA, katsuhiko)
関西学院大学・経営戦略研究科・教授
研究者番号：90411793

(2)研究分担者

羽室 行信 (HAMURO, yukinobu)
関西学院大学・経営戦略研究科・准教授
研究者番号：90268235

加藤 直樹 (KATOH, naoki)
関西学院大学・理工学部・教授
研究者番号：40145826