

2ちゃんねると株価

○五島圭一（東京工業大学） 高橋大志（慶應義塾大学）
寺野隆雄（東京工業大学）

The Analysis of Japanese Stock Market through Social Media

*K. Goshima (Tokyo Institute of Technology), H. Takahashi (Keio University) and T. Terano (Tokyo Institute of Technology)

概要— 本稿は、ソーシャルメディアと株価の関連性に注目した。特に、わが国における主要なソーシャルメディアのひとつである2ちゃんねるのスレッド数とマーケット指標との関連性を分析をすることで、ソーシャルメディアが持つマーケット情報の考察を試みた。分析の結果、(1) 場中と閉場後のスレッド数との間に高い相関関係が見られ、マーケットに反応して、書き込み数が増えている可能性があること、(2) 時価総額の大きい株価指数との間に高い相関関係が見られ、大型株に反応している可能性があること、などの結論を見出した。テキスト内容を加味した詳細な分析は今後の課題である。

キーワード: 株価, 2ちゃんねる, ニュース

1 はじめに

投資家は、新聞やテレビ、各企業のプレスリリース、ソーシャルメディアなど、様々なメディアからニュースを入手し、投資先となる企業を選定する。ニュースには数値情報だけでなく、テキスト情報も含まれており、それらを活用することで数値情報だけでは説明することが難しい資産価格の変動やマーケットメカニズムなどの分析や予測ができる可能性がある。そのため、2000年代中頃から、資産価格の分野において、ニュースやソーシャルメディアといったテキストデータを、資産価格評価の分析に用いる試みが模索されている。

例えば、Tetlock (2007) は Wall street Journal column から悲観度を抽出し、ダウ工業平均株価との関連性を見出している⁴⁾。Mitra et al. (2009) では、ニュースを使ったファクターモデルの拡張を提案している³⁾。また、ソーシャルメディアと株価の関連性に言及している研究も存在する。Bollen et al. (2011) は、twitter の投稿内容を利用し、ダウ工業平均株価の変動を87.6%の精度で予測できたとしている²⁾。

このようにテキスト情報とマーケットとの間には一定の関連性があるとされている。一方で、Leinweber (2009) はニュースをその属性によって、ニュース、プレニュース、ソーシャルメディアニュースの3つに分類し、それぞれ各媒体はサービスへの参加者属性や情報伝播の速さ、サービスの仕組みなどが異なっており、それぞれのテキスト情報が内包するマーケット情報に違いがある論じている¹⁾。

そこで本稿では、ソーシャルメディアのひとつである2ちゃんねるを用いて、マーケット指標との関連性を観察することによって、ソーシャルメディアが持つマーケット情報について考察を行った。次章ではデータに触れ、第3章は分析手順・手法、第4章は分析結果をそれぞれ記す。第5章は、まとめである。

2 データ

2.1 マーケットデータ

マーケットデータとして、TOPIX, TOPIX core30, TOPIX large70, TOPIX mid400, TOPIX small それぞれの高値と安値を用いた。また、東証一部の出来高

と売買代金を使用した。

2.2 2ちゃんねるデータ

本稿では、2ちゃんねるデータをCeek.jp¹より入手した。2ちゃんねるは、日本最大規模のソーシャルメディアである。6000以上の掲示板が存在し、そこでは様々なトピックについて議論が交わされている。2ちゃんねるはカテゴリごとの板ごとのスレッドのような構造になっており、スレッドに文章(レス)を投稿する形となっている。

3 分析手順・手法

3.1 2ちゃんねるデータの加工・処理方法

ここでは、2ちゃんねるデータの加工方法について記載する。

はじめに、本稿の分析期間である2007年7月1日から2011年12月31日までのスレッド数を抽出する。そのうち、市況実況1板におけるスレッド数のみを抽出する。市況実況1板は、日本株式市場のみを対象とした実況板であり、日経平均株価やTOPIX指標などの日本株式指標や日本証券市場に上場されている個別銘柄に関して様々な議論や意見表明が毎日行われている。本稿では、市場情報と比較するために、日次スレッド数に加えて、マーケットが開く前の5時から9時までの2ちゃんねるのスレッド数、株式取引時間内の9時から15時までの2ちゃんねるのスレッド数、マーケットが閉まった後の15時から26時までの2ちゃんねるのスレッド数をそれぞれカウントした。

次に、抽出したスレッドのうち100レス以上についているスレッドに限定した。これは、荒らしによるスレッド乱立や市況実況1板の参加者の関心が低いスレッドを排除するためである。これを行うことで、市況実況1板の利用者の関心の高いスレッドタイトルのみを扱うことが可能となる。

3.2 分析手法

分析手法については、前節にて抽出した2ちゃんねるのスレッド数とマーケット指標との関連性を考察する。マーケット指標は2章で述べた株価指標の(高値-

¹<http://open.ceek.jp/>

安値)と東証一部における出来高と売買代金を用いた。分析対象期間は、2007年1月1日から2013年12月31日までとした。具体的には、 i 日における i 日におけるTOPIXの(高値-安値)、出来高、売買代金との相関関係を確認した。

これは、TOPIXの高値と安値との値幅が大きき場況が荒く、また取引額が大きいとき、ソーシャルメディアの反応も大きくなると仮定した。また、どの時間帯との相関が高いか考察するため、開場前、場中、閉場後の3つのスレッド数とのそれぞれとの関連性に注目した。さらに、TOPIX core30、TOPIX large70、TOPIX mid400、TOPIX smallの4つの東証規模別株価指数との関連性を観察することで、ソーシャルメディアは、どの企業規模の銘柄に注目しているかを観察を試みた。

4 分析結果

4.1 各マーケット指標との関連性

Table1は、2007年から2013年における全体と各年ごとの2ちゃんねるのスレッド数と各マーケット指標との相関係数を示したものである。まず、2ちゃんねるのスレッド数と出来高及び売買代金との相関係数はそれぞれ0.22、0.48であり、売買代金のほうがより強い相関関係が見られた。低位株より値嵩株の取引量が多いとき、2ちゃんねるのスレッド数が増えている可能性がある。

また、TOPIX(高値-安値)との相関係数については0.58であり、2ちゃんねるのスレッド数とTOPIX(高値-安値)の間には強い相関関係が見られた。時間帯別のスレッド数との関連性を見ると、場中の9時-15時のスレッド数について、売買代金との相関係数は0.45、TOPIX(高値-安値)との相関係数は0.48であり、他の時間帯に比べ、強い相関関係があることわかった。また、閉場後の15時-26時までのスレッド数について、売買代金との相関係数は0.36、TOPIX(高値-安値)との相関係数は0.47であった。一方で、開場前の5時-9時までのスレッド数について、売買代金との相関係数は0.29、TOPIX(高値-安値)との相関係数は0.37であった。場中や閉場後と比べると低い相関関係であった。

年別にみると、ばらつきはあるが総じて場中と閉場のスレッド数とマーケット指標の間には強い相関関係が見られた。しかしながら、開場前のスレッド数とは、弱い相関関係であり、2ちゃんねるはマーケットに反応して、書き込み数が増えている可能性がある。

4.2 時価総額ごとのインデックスとの関連性

Table2は、2007年から2013年における全体と各年ごとの2ちゃんねるのスレッド数と東証規模別株価指数との相関係数を表している。2ちゃんねるのスレッド数との相関係数について、core30は0.60、large70は0.55、mid400は0.57、smallは0.49であった。企業規模が大きい株価指数ほど相関係数が高い傾向がある。また、東証規模別株価指数についても同様に、場中と閉場後のスレッド数との間に強い相関関係が見られた。

年別にみると、ばらつきはあるが総じて企業規模が大きい株価指数ほど相関係数が高い傾向があり、2ちゃんねるは大型株の動きに反応して、書き込み数が増えている可能性がある。

5 まとめ

2000年代中頃から、資産価格の分野において、ニュースやソーシャルメディアといったテキストデータを、資産価格評価の分析に用いる試みが模索されている。そこで本稿では、ソーシャルメディアのひとつである2ちゃんねるを用いて、マーケット指標との関連性を観察することによって、ソーシャルメディアが持つマーケット情報について考察を行った。特に、わが国における主要なソーシャルメディアのひとつである2ちゃんねるのスレッド数とマーケット指標との関連性を分析をすることで、ソーシャルメディアが持つマーケット情報の考察を試みた。分析の結果、(1)場中と閉場後のスレッド数との間に高い相関関係が見られ、マーケットに反応して、書き込み数が増えている可能性があること、(2)時価総額の大きい株価指数との間に高い相関関係が見られ、大型株に反応している可能性があること、などの結論を見出した。これらの結果から、ソーシャルメディアでは株価の将来予測が難しいことが想定される。テキスト内容を加味した詳細な分析は今後の課題である。

参考文献

- 1) David J. Leinweber.: Nerds on Wall Street, John Wiley & Sons (2009)
- 2) John Bollen, Hunia Mao and Xiaoujun Zeng.: Twitter mood predicts the stock market, *Journal of Computational Science*, 2-1, 1/8 (2011)
- 3) Leela R. Mitra, Gautam Mitra and Dan diBartolomeo.: Equity portfolio risk (volatility) estimation using market information and sentiment, *Quantitative Finance*, 9-8, 887/895 (2009)
- 4) Paul C. Tetlock.: Giving Content to Investor Sentiment: The Role of Media in the Stock Market, *The Journal of Finance*, 62-3, 1139/1168 (2007)

Table 1: スレッド数とマーケット指標との相関係数

相関係数：2007-2013	出来高	売買代金	高値-安値 (TOPIX)
2ちゃんねるのスレッド数	0.22	0.48	0.58
2ちゃんねるのスレッド数：9時-15時	0.29	0.45	0.48
2ちゃんねるのスレッド数：5時-9時	0.12	0.29	0.37
2ちゃんねるのスレッド数：15時-26時	0.13	0.36	0.47
相関係数：2007	出来高	売買代金	高値-安値 (TOPIX)
2ちゃんねるのスレッド数	0.25	0.25	0.26
2ちゃんねるのスレッド数：9時-15時	0.30	0.29	0.30
2ちゃんねるのスレッド数：5時-9時	0.17	0.21	0.19
2ちゃんねるのスレッド数：15時-26時	0.12	0.10	0.11
相関係数：2008	出来高	売買代金	高値-安値 (TOPIX)
2ちゃんねるのスレッド数	0.61	0.28	0.51
2ちゃんねるのスレッド数：9時-15時	0.48	0.37	0.38
2ちゃんねるのスレッド数：5時-9時	0.34	-0.05	0.30
2ちゃんねるのスレッド数：15時-26時	0.44	0.20	0.40
相関係数：2009	出来高	売買代金	高値-安値 (TOPIX)
2ちゃんねるのスレッド数	0.16	0.17	0.17
2ちゃんねるのスレッド数：9時-15時	0.20	0.19	0.18
2ちゃんねるのスレッド数：5時-9時	0.07	0.04	0.02
2ちゃんねるのスレッド数：15時-26時	0.05	0.09	0.16
相関係数：2010	出来高	売買代金	高値-安値 (TOPIX)
2ちゃんねるのスレッド数	0.26	0.24	0.15
2ちゃんねるのスレッド数：9時-15時	0.20	0.18	0.11
2ちゃんねるのスレッド数：5時-9時	0.27	0.21	0.17
2ちゃんねるのスレッド数：15時-26時	0.15	0.14	0.11
相関係数：2011	出来高	売買代金	高値-安値 (TOPIX)
2ちゃんねるのスレッド数	0.57	0.48	0.64
2ちゃんねるのスレッド数：9時-15時	0.43	0.31	0.55
2ちゃんねるのスレッド数：5時-9時	0.17	0.18	0.15
2ちゃんねるのスレッド数：15時-26時	0.51	0.44	0.50
相関係数：2012	出来高	売買代金	高値-安値 (TOPIX)
2ちゃんねるのスレッド数	0.26	0.24	0.26
2ちゃんねるのスレッド数：9時-15時	0.29	0.26	0.33
2ちゃんねるのスレッド数：5時-9時	0.06	0.11	0.08
2ちゃんねるのスレッド数：15時-26時	0.10	0.09	0.07
相関係数：2013	出来高	売買代金	高値-安値 (TOPIX)
2ちゃんねるのスレッド数	0.25	0.22	0.39
2ちゃんねるのスレッド数：9時-15時	0.28	0.29	0.35
2ちゃんねるのスレッド数：5時-9時	0.17	0.11	0.10
2ちゃんねるのスレッド数：15時-26時	0.15	0.09	0.25

Table 2: スレッド数と東証規模別株価指数との相関係数

相関係数：2007-2013	高値-安値 (core30)	高値-安値 (large70)	高値-安値 (mid400)	高値-安値 (small)
2ちゃんねるのスレッド数	0.60	0.55	0.57	0.49
2ちゃんねるのスレッド数：9時-15時	0.48	0.47	0.48	0.44
2ちゃんねるのスレッド数：5時-9時	0.40	0.35	0.37	0.31
2ちゃんねるのスレッド数：15時-26時	0.50	0.44	0.47	0.38
相関係数：2007	高値-安値 (core30)	高値-安値 (large70)	高値-安値 (mid400)	高値-安値 (small)
2ちゃんねるのスレッド数	0.25	0.25	0.25	0.28
2ちゃんねるのスレッド数：9時-15時	0.27	0.31	0.29	0.29
2ちゃんねるのスレッド数：5時-9時	0.19	0.16	0.19	0.28
2ちゃんねるのスレッド数：15時-26時	0.10	0.10	0.11	0.13
相関係数：2008	高値-安値 (core30)	高値-安値 (large70)	高値-安値 (mid400)	高値-安値 (small)
2ちゃんねるのスレッド数	0.51	0.50	0.50	0.44
2ちゃんねるのスレッド数：9時-15時	0.37	0.37	0.37	0.30
2ちゃんねるのスレッド数：5時-9時	0.29	0.32	0.30	0.30
2ちゃんねるのスレッド数：15時-26時	0.39	0.41	0.39	0.34
相関係数：2009	高値-安値 (core30)	高値-安値 (large70)	高値-安値 (mid400)	高値-安値 (small)
2ちゃんねるのスレッド数	0.17	0.15	0.20	0.12
2ちゃんねるのスレッド数：9時-15時	0.15	0.19	0.20	0.20
2ちゃんねるのスレッド数：5時-9時	0.03	0.03	0.02	-0.01
2ちゃんねるのスレッド数：15時-26時	0.17	0.13	0.18	0.12
相関係数：2010	高値-安値 (core30)	高値-安値 (large70)	高値-安値 (mid400)	高値-安値 (small)
2ちゃんねるのスレッド数	0.16	0.12	0.19	0.07
2ちゃんねるのスレッド数：9時-15時	0.13	0.08	0.15	0.05
2ちゃんねるのスレッド数：5時-9時	0.17	0.16	0.19	0.12
2ちゃんねるのスレッド数：15時-26時	0.12	0.09	0.16	0.04
相関係数：2011	高値-安値 (core30)	高値-安値 (large70)	高値-安値 (mid400)	高値-安値 (small)
2ちゃんねるのスレッド数	0.62	0.64	0.64	0.66
2ちゃんねるのスレッド数：9時-15時	0.53	0.55	0.55	0.55
2ちゃんねるのスレッド数：5時-9時	0.15	0.15	0.15	0.15
2ちゃんねるのスレッド数：15時-26時	0.49	0.49	0.51	0.51
相関係数：2012	高値-安値 (core30)	高値-安値 (large70)	高値-安値 (mid400)	高値-安値 (small)
2ちゃんねるのスレッド数	0.23	0.27	0.24	0.34
2ちゃんねるのスレッド数：9時-15時	0.28	0.34	0.32	0.39
2ちゃんねるのスレッド数：5時-9時	0.10	0.09	0.08	0.08
2ちゃんねるのスレッド数：15時-26時	0.06	0.09	0.04	0.14
相関係数：2013	高値-安値 (core30)	高値-安値 (large70)	高値-安値 (mid400)	高値-安値 (small)
2ちゃんねるのスレッド数	0.37	0.39	0.40	0.33
2ちゃんねるのスレッド数：9時-15時	0.34	0.34	0.35	0.31
2ちゃんねるのスレッド数：5時-9時	0.07	0.10	0.13	0.09
2ちゃんねるのスレッド数：15時-26時	0.24	0.25	0.26	0.21