

ニュースの新規性と市場の反応

水野貴之^{A,B,C,D}, 大西立顕^{D,E}, 渡辺努^{D,F}

^A 国立情報学研究所情報社会相関研究系

^B 総合研究大学院大学複合科学研究科

^C 独立行政法人科学技術振興機構 さきがけ

^D キヤノングローバル戦略研究所

^E 東京大学大学院情報理工学系研究科

^F 東京大学大学院経済学研究科

Abstract— 2003年から2011年までにThompson Reutersが配信した2400万記事を越えるニュースとニューヨーク証券取引所の取引データを用いて、ニュースの新規性と株式市場のアクティビティへのニュースのインパクトを調べる。コサイン類似度を用いて各ニュースの他のニュースとの記事中の単語の類似性を調べ、類似性の低いニュースを新規性の高いニュースと定義する。新規性の高いニュースはボラティリティや取引数、出来高を増幅させる。一方、新規性の低いニュースはそれらをほとんど上昇させない。つまり、ニュースと市場の反応の関係を調べるうえで、「情報の折込済み」は大変重要な要因である。

1 はじめに

経済物理学において、金融市場、特に株価変動の研究は古くから行われてきた。マイルストーンになった研究は、1995年にMantegnaとStanleyによってNatureに発表された株価変動の分布の研究である[1]。彼らは、株価の変動が様々な時間スケールにおいて、正規分布に比べて非常に大きな変動が頻繁に発生していることを示した。この研究を契機にして、大きな変動が頻発するダイナミクスやメカニズムを解明する研究がスタートした。

大きな変動が生じる要因は、大きく分けて2つある。

1つは、市場参加者が過去の価格変動に対して過敏に反応する内生的ショックと、もう1つは、予期せぬ外的な要因により反応する外生的ショックである。1995年以降、多くの詳細な金融市場の取引データが公開され、まずは、それらのデータを用いて内生的ショックが調べられた。現在、内生的ショックのダイナミクスは、ランダム乗算過程に従うと実証、及び、理論面から考えられている[2,3]。近年、取引データだけではなく、金融市場に流れる網羅的なニュースのビックデータも蓄積、公開され始めた。このニュースデータを用いて、金融市場に、いつ何が起きたのかという外生的ショックを網羅的に調査することができるようになった[4-6]。

外生的ショックの研究には「市場への折込済み」という解決しなければいけない問題がある。市場への折込済みとは、株価は将来に得られる利潤の割引現在価値であるために、同じような外生的ショックが発生したとしても、そのショックが事前に予測されていた場合には、予測ができた時点でショックの影響が株価に現れてしまい、ショックが起きた時点では株価は反応しないという現象である。つまり、ある外生的ショックがいつ予測可能になったのかを明らかにしなければならない。

本発表では、300を越える報道機関の網羅的なニュースデータを用いて、ニュースの新規性と株価の反応の関係について明らかにする。

2 データ

本研究では、株式とニュースのデータを用いる。株価には、ニューヨーク証券取引所に上場されているAIG, Wal-Mart, IBM, P&Gの約定価格を用いる。ニュースには、機関投資家の主要な取引端末であるReuters 3000 Xtraで流れる報道機関約300社のニュースを用いる。分析期間は、2003年1月1日から2011年6月30日までの約8年半である。

本ニュースデータは、速報性を要する見出しであるALERT、通常の見出しであるHEADLINE、ALERTやHEADLINEに対応するニュース記事であるSTORYによって構成されている。本研究ではALERTとHEADLINEを分析に用いる。

3 前処理

3.1 株の季節調整

株式市場は季節や時間帯によりアクティビティが異なる。外生的ショックによる市場の反応を明確に読み取るために、季節や時間帯によるアクティビティの変動を取り除く。このような操作を、季節調整と呼ぶ。出来高、取引数、ボラティリティに対して季節調整をするために、 d 日の時刻 t (分刻み)の出来高 $Vol'(d, t)$ 、取引数 $N'(d, t)$ 、ボラティリティ $V'(d, t)$ に対して次式の変換をおこなう

$$Vol(d, t) \equiv 390 \cdot \frac{Vol'(d, t)}{\sum_a Vol'(d, t)} / \sum_t \frac{Vol'(d, t)}{\sum_a Vol'(d, t)} \quad (1)$$

$$N(d, t) \equiv 390 \cdot \frac{N'(d, t)}{\sum_a N'(d, t)} / \sum_t \frac{N'(d, t)}{\sum_a N'(d, t)} \quad (2)$$

$$V(d, t) \equiv 390 \cdot \frac{V'(d, t)}{\sum_a V'(d, t)} / \sum_t \frac{V'(d, t)}{\sum_a V'(d, t)} \quad (3)$$

ここで、390はニューヨーク証券取引所における1日の取引時間390分間を表す。この変換により時間帯や季節性の周期が取り除かれた出来高 $Vol(d, t)$ 、取引数 $N(d, t)$ 、ボラティリティ $V(d, t)$ を作ることができる。

3.2 各銘柄に関連するニュースの特定

AIG, Wal-Mart, IBM, P&G に関連するニュースをニュースデータから抽出する. IBM を例に取り, 抽出方法を述べる.

一部のニュースには, 記者が関連する銘柄を付随情報として記載している. 私は, IBM という付随情報のある全てのニュースから単語の出現頻度を表す tf 値を求める. 単語 w の tf 値を $tf(w)$ と書く. tf 値が大きい単語ほど, IBM という付随情報のあるニュースの中で頻繁に出現する. 次に, IBM という付随情報を問わず, 全てのニュースから単語の逆文書頻度を表す idf 値を求める. 単語 w の idf 値を $idf(w)$ と書く. idf 値は様々な文書で広く使われる単語ほど小さな値を示す. $tf \cdot idf(w) = tf(w) \times idf(w)$ に注目する. 高い $tf \cdot idf$ 値を持つ単語は IBM という付随情報のあるニュースにしか高頻度で出現しない単語を表す. $tf \cdot idf$ 値の上位 3 単語を持つニュースを, IBM に関連するニュースであると定義する.

4 ニュースの新規性

bag-of-words で, ニュースごとに各単語の出現数をベクトルの各成分に持つ多次元ベクトルを構成する. 2 つの異なるニュースの類似度を, この 2 つのニュースのベクトルのコサインを計算することによって測る. k 番目のニュースと $k+i$ 番目のニュースの類似度を $c(k, k+i)$ とする. 全ての k に対して類似度 $c(k, k+i)$ を計算し平均したものを $c(i)$ とする. Fig.1 は Wal-Mart に関連するニュースについて, i 離れたニュースとの平均類似度 $c(i)$ を示している. 時期の近いニュースほど類似度が高いことが分かる. 関数 $c(i)$ はべき関数による近似がよく当てはまり, Wal-Mart の場合,

$$c(i) = 0.224i^{-0.837} + 0.08 \quad (4)$$

である. 残差項の 0.08 は Wal-Mart などの単語がニュースに必ず入るためである.

k 番目のニュースの新規性を, 過去 20 ニュースとの累積類似度の逆数によって定義する.

$$Nov(k) \equiv \left(\frac{1}{20} \sum_{i=1}^{20} c(k-i, k) \right)^{-1} \quad (5)$$

過去のニュースとの類似度が低ければ, 新規性 $Nov(k)$ は大きな値を持つ.

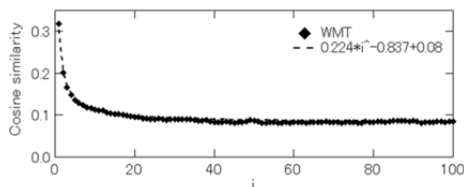


Fig.1 Wal-Mart に関連するニュースの類似度 $c(i)$

5 新規ニュースによる市場の反応

新規性の高いニュースによる市場の反応を調査する. 新規性 Nov が平均以上のニュースについて古いものから順に番号を付ける. k 番目のニュースの時刻 τ_k から T 分離れた時刻の出来高 $Vol(\tau_k + T)$, 取引数 $N(\tau_k + T)$, ボラティリティ $V(\tau_k + T)$ に注目する. 新規性 Nov が平均以上の Wal-Mart に関連した全てのニュースで平均した出

来高 $(Vol(T))$, 取引数 $(N(T))$, ボラティリティ $(V(T))$ を Fig.2 で示す. 同様に, 新規性 Nov が平均以下の Wal-Mart に関連した全てのニュースを用いた結果を Fig.3 で示す.

新規性が高いニュースでは, ニュース発表の時刻 $T = 0$ の前には, 出来高, 取引数, ボラティリティが変化せず, 発表の直後にそれぞれが 1.5 倍に上昇することが分かる. 一方, 新規性が低いニュースでは, ニュース発表の時刻 $T = 0$ の前から出来高, 取引数, ボラティリティは上昇を始めており, ニュースの直後での直前からの上昇は僅かである.

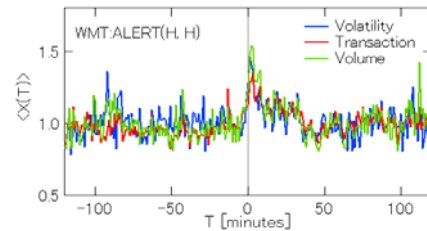


Fig.2 新規性が高いニュースによる市場の反応

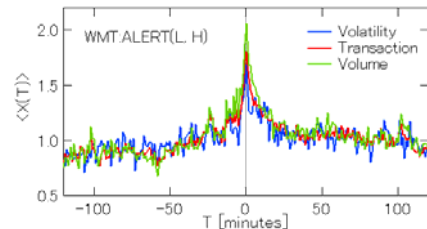


Fig.3 新規性が低いニュースによる市場の反応

6 まとめと展望

網羅的なニュースのデータを用いて, ニュースの新規性を測定し, 新規性の高いニュースでは株式市場は強く反応することを示した. 網羅的に株式銘柄を調査することは今後の課題である. また, ニュースの重要度を測る指標として, 新規性以外に, 話題性と信頼性という指標を新たに導入し, 市場が強く反応するニュースの特徴について調べることも, 今後おこなっていく.

参考文献

- 1) R. N. Mantegna and H. E. Stanley, "Scaling Behaviour in the Dynamics of an Economic Index," Nature 376, 46-49 (1995).
- 2) M. Takayasu, T. Mizuno, H. Takayasu, "Potentials force observed in market dynamics," Physica A 370, 91-97 (2006).
- 3) A. I. Saichev, Y. Malevergne, D. Sornette, "Theory of Zipf's Law and Beyond," Springer Berlin Heidelberg (2009).
- 4) A. Joulin, A. Lefevre, D. Grunberg, J. P. Bouchaud, "Stock price jumps: News and volume play a minor role," arXiv:0803.1769 (2008).

- 5) T. Mizuno, K. Takei, T. Ohnishi, T. Watanabe, "Temporal and Cross Correlations in Business News," *Progress of Theoretical Physics Supplement* 194, 181-192 (2012).
- 6) R. Hisano, D. Sornette, T. Mizuno, T. Ohnishi, T. Watanabe, "High quality topic extraction from business news explains abnormal financial market volatility," *PLoS ONE* 8(6): e64846. doi:10.1371/journal.pone.0064846 (2013).