

市場急変時における銘柄間情報伝播の変化に関する分析

Analyzing the change of the information flow among stock index and individual stocks at the time of the market sudden change

鈴木 裕士¹ 和泉 潔^{1,2} 吉村 忍¹

Hiroshi Suzuki¹, Kiyoshi Izumi^{1,2}, and Shinobu Yoshimura¹

¹ 東京大学大学院工学系研究科

¹School of Engineering, The University of Tokyo

² 科学技術振興機構 CREST

CREST, JST

Abstract: In this paper, we analyze how the relationship among Nikkei average futures and individual stocks changes when a big event occurs, intending to give investors useful information for the risk management. We showed that the strong relationships had appeared after the Great East Japan Earthquake and we could detect them by using order books properly.

1. 諸言

1.1 背景

近年、株式売買システムの高度化に伴い高頻度取引が増えてきており、複数の銘柄に大量の注文の出し入れが行われるようになった。そのような中、市場が急落する前には銘柄間で価格が連動しているという研究結果が発表されるなど、投資におけるリスク管理のために銘柄間の関係性を理解することが非常に重要になっている[1]。

1.2 既存研究

銘柄間の関係性を分析する研究では、価格データを用いて銘柄間の相関係数を用いる手法や全域木を用いる手法など、様々提案されている[2,3]。Kullmanら[4]は各銘柄の価格データを用い、時間差をつけて相関係数を計算することで、どの銘柄がどの銘柄に価格の影響を与えているのかの分析を試みた。ところが相関係数がピークをとる時間差の時でも値自体が小さく、それ以上の解釈を加えることが困難であった。

Schreiber[5]により導入された移動エントロピー (transfer entropy) は、時系列データの連続性を考慮した計算方法であり、データ間の情報伝播の方向性もわかる手法である。また、確率をもとにした計算方法であるため、値の根拠を分析することで状態推移の確率の変化を分析できるという点でも優れている。Marschinskiら[6]はこの移動エントロピーを用い、ダ

ウ平均株価と DAX 指数の間の情報の流れを分析し、ダウ平均株価から DAX 指数へ価格伝播があることを示した。Kwon ら[7]は対象とするインデックスの種類を 25 に増やし、情報の流れをネットワークとして可視化した。さらに Kwon ら[8]は、個別銘柄を業界ごとにグループ化したデータを用いることでインデックスから個別銘柄への情報伝播があることを示した。また、Shi ら[9]はアジア通貨危機の前後で中国とアメリカの市場間の情報伝播の仕方が変化したことを示した。ところがこれらの研究はすべてインデックスどうしを対象にしていたり、長期間の情報伝播を対象にしていたりしており、日々の投資家の活動に役立つものとは言えない。

1.3 研究の目的

本研究では、日経平均先物を含む個別銘柄間の関係性、すなわち、ある銘柄の板の状態が他の銘柄の板の状態にどのように影響を与えるかを解明することで、投資家のリスク管理に有益な情報を提供することを目的とする。そのために、価格のみのデータよりも多くの情報が含まれている板情報を用い、既存研究よりも時間間隔を短くし、ある大きな出来事が生じた後に銘柄間の関係性がどのように変化するかを分析する。

2. 移動エントロピー

移動エントロピーは、離散的な 2 つの定常過程間の情報の流れを測定するために Schreiber[4]により

導入された。ここで、2つの離散的な定常過程 I と J を考える。過程 I から k 個のサンプル、過程 J から l 個のサンプルを用いて計算する場合、移動エントロピーは以下ようになる。

$$T_{J \rightarrow I} = \sum p(i_{t+1}, i_t^{(k)}, j_t^{(l)}) \log \frac{p(i_{t+1} | i_t^{(k)}, j_t^{(l)})}{p(i_{t+1} | i_t^{(k)})} \quad (1)$$

i_t と j_t はそれぞれ過程 I と J の時間 t における離散状態を表す。 $i_t^{(k)}$ は過程 I の時間 t における状態から k 個分遡った状態までを含めたベクトルであり、 $i_t^{(k)} = (i_t, i_{t-1}, \dots, i_{t-k+1})$ と表される。 $j_t^{(l)}$ も同様に、 $j_t^{(l)} = (j_t, j_{t-1}, \dots, j_{t-l+1})$ と表される。同時確率分布 $p(i_{t+1}, i_t^{(k)}, j_t^{(l)})$ は i_{t+1} 、 $i_t^{(k)}$ 、 $j_t^{(l)}$ がそれぞれ特定の値をとるときの確率である。条件付き確率分布 $p(i_{t+1} | i_t^{(k)}, j_t^{(l)})$ と $p(i_{t+1} | i_t^{(k)})$ はそれぞれ、 $i_t^{(k)}$ と $j_t^{(l)}$ 、あるいは $i_t^{(k)}$ の値が既知であるときに i_{t+1} が特定の値をとる確率である。

$T_{J \rightarrow I}$ は過程 J の変遷がもう一方の過程 I の遷移確率にどれだけ影響を与えるかを計算する。同時確率分布と条件付き確率分布の i と j を入れ替えることで逆向きの影響を計算することができる。また、移動エントロピーの計算方法は i_t と j_t に関して非対称になっており、これにより2つの時系列データの情報伝播の方向性がわかる。

移動エントロピーは過程 J から過程 I への情報の流れから定量化できる。また、過去の過程 I と過程 J から得られる過程 I の未来の状態に対する情報量と、過去の過程 I のみから得られる過程 I の未来の状態に対する情報量の差をとることで計算することができる。したがって、移動エントロピーは以下のように書き直すことができる。

$$T_{J \rightarrow I} = h_I(k) - h_{IJ}(k, l) \quad (2)$$

ここで、

$$h_I(k) = - \sum p(i_{t+1}, i_t^{(k)}) \times \log p(i_{t+1} | i_t^{(k)}) \quad (3)$$

$$h_{IJ}(k, l) = - \sum p(i_{t+1}, i_t^{(k)}, j_t^{(l)}) \log p(i_{t+1} | i_t^{(k)}, j_t^{(l)}) \quad (4)$$

である。

3. 数値実験

東日本大震災が個別銘柄間の関係性にどのような影響を与えたかを分析する。

3.1 使用データ

日経平均先物と表1に示す25銘柄を対象に分析する。日経平均先物は日本経済の先行きを占う指標と考えることができる。銘柄は、素材、製薬、自動車、銀行、不動産、電力の業界からと、値嵩株から選んでいる。期間は2011年2月28日から2011年3月27日で、市場が開いている平日の9:00~11:00と12:30~15:00の板情報を用いる。

表1: 分析対象銘柄

銘柄コード	銘柄名
3401	帝人
3402	東レ
3405	クラレ
3407	旭化成
4502	武田薬品工業
4523	エーザイ
4568	第一三共
6954	ファナック
7201	日産自動車
7203	トヨタ自動車
7211	三菱自動車工業
7261	マツダ
7267	本田技研工業
8306	三菱UFJFG
8316	三井住友FG
8411	みずほFG
8801	三井不動産
8802	三菱地所
8830	住友不動産
9501	東京電力
9502	中部電力
9503	関西電力
9983	ファーストリテイリング
9984	ソフトバンク

3.2 板の特徴量

リターン(priceReturn)

価格の対数を取り、その差で表される。また、リターンの絶対値(absRetrun)も特徴量として用いる。

約定回数(exeNum)

指定時間内に取引が成立した回数。

出来高(exeVolume)

指定時間内に売買が行われた株数。

平均出来高(exeMean)

指定時間内において、取引 1 回につき売買が行われた平均的な株数。

累積変動価格(accumulated)

指定時間内に変動した価格の累積和。例えば、ある時間内において株価が 100 円、99 円、100 円、101 円のように変化したとき、累積変動価格は 3 円となる。

Best Ask/Bid Volume

Best Ask Volume は最良売り気配の注文数、Best Bid Volume は最良買い気配の注文数。

Order Imbalance

Best Ask Volume と Best Bid Volume の差をとることにより求められる量。

$BD^a(k)/BD^b(k)$

それぞれ ask と bid における最良気配値から第 k 気配値までの累積注文量。 k をパラメーターとする。

Order Imbalance $BD(k)$

$BD^a(k)$ と $BD^b(k)$ の差をとることにより求められる量。

Spread

Spread は最良売り気配値と最良買い気配値との差と、最良売り気配値と最良買い気配値の平均値との比をとることによって求められる。これを式で表すと式(5)のようになる。

$$Spread = \frac{Best\ Ask - Best\ Bid}{(Best\ Ask + Best\ Bid)/2} \quad (5)$$

Slope

Slope は最良気配からの注文数の累積分布の形状を数値化したものであり、式(6)のように表される[10]。

$$Slope = \frac{SE + DE}{2} \quad (6)$$

SE と DE はそれぞれ ask 側と bid 側の Slope を表しており、式(7)と式(8)で与えられる。

$$SE = \frac{1}{N_A} \left\{ \frac{v_1^A}{|p_1^A/p_0 - 1|} + \sum_{\tau=1}^{N_A-1} \frac{v_{\tau+1}^A/v_{\tau}^A - 1}{p_{\tau+1}^A/p_{\tau}^A - 1} \right\} \quad (7)$$

$$DE = \frac{1}{N_B} \left\{ \frac{v_1^B}{|p_1^B/p_0 - 1|} + \sum_{\tau=1}^{N_B-1} \frac{v_{\tau+1}^B/v_{\tau}^B - 1}{p_{\tau+1}^B/p_{\tau}^B - 1} \right\} \quad (8)$$

ここで、 N_A と N_B はそれぞれ ask と bid の気配の数を表す。つまり、第何気配までを計算に含めるかを表す。 N_A と N_B は同じ値をとり、これをパラメーターとする。 τ はティックレベルであり、 $\tau = 1$ なら最良気配を、 $\tau = 2$ なら第 2 気配を表す。 p_{τ} は第 τ 気配値を表し、 p_0 は最良売り気配値と最良買い気配値の平均を表す。 v_{τ} は第 τ 気配までの注文数の累積和の自然対数をとったものである。また、右肩の A と B はそれぞれ ask と bid を表している。

Dispersion

Dispersion は板の疎密を数値化したものであり、式(9)のように表される[11]。

$$Dispersion = \frac{1}{2} \left(\frac{\sum_{\tau=1}^n w_{\tau}^A Dst_{\tau}^A}{\sum_{\tau=1}^n w_{\tau}^A} + \frac{\sum_{\tau=1}^n w_{\tau}^B Dst_{\tau}^B}{\sum_{\tau=1}^n w_{\tau}^B} \right) \quad (9)$$

ここで、 τ はティックレベルを表しており、 w_{τ} は第 τ 気配における注文数を表している。 n パラメーターとする。 Dst_{τ} は第 τ 気配値と第 $(\tau + 1)$ 気配値との差の絶対値を表している。また、右肩の A と B はそれぞれ ask と bid を表している。

Cost-to-trade

Cost-to-trade は大量の成行注文を出したときにどれだけ公正な価格から離れて約定してしまうかを数値化したものであり、式(10)で表される[11]。

$$Cost - to - trade(t) = \frac{\sum_{\tau=1}^T I_{\tau}^A (p_{\tau}^A - midquote) + \sum_{\tau=1}^T I_{\tau}^B (midquote - p_{\tau}^B)}{T \times midquote} \quad (10)$$

ここで、 p_{τ} は第 τ 気配値を表しており、右肩の A と B はそれぞれ ask と bid を表している。 $midquote$ は最良売り気配値と最良買い気配値の平均を表している。 t はパラメーター、 T は仮想の成行注文数を表しており、平均的な 1 日あたりの出来高の t 分だけが T にあたるとする。 I_{τ} は T だけ成行注文を出した場合に第 τ 気配でどれだけ約定するかを表しており、ask 側と bid 側でそれぞれ式(11)、式(12)のように表される。

$$I_{\tau}^A = \begin{cases} w_j^A & \text{if } T > \sum_{j=1}^{\tau} w_j^A \\ \left(T - \sum_{j=1}^{\tau-1} w_j^A \right) & \text{if } T > \sum_{j=1}^{\tau} w_j^A \text{ and } T < \sum_{j=1}^{\tau} w_j^A \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

$$I_{\tau}^B = \begin{cases} w_j^B & \text{if } T > \sum_{j=1}^{\tau} w_j^B \\ \left(T - \sum_{j=1}^{\tau-1} w_j^B \right) & \text{if } T > \sum_{j=1}^{\tau-1} w_j^B \text{ and } T < \sum_{j=1}^{\tau} w_j^B \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

ここで、 w_{τ}^B は第 τ 気配における注文量を表している。Cost-to-tradeの値が大きいほど、大量の成行注文を出した場合に公正な価格から離れた価格で約定してしまう量が多くなり、取引のコストが大きくなるといえる。

BLM

BLMは成行注文を出した場合に価格がどれだけ変化するかを元の価格で正規化したものであり、式(13)で表される[12]。

$$BLM(s) = \frac{PI^A(s) - PI^B(s)}{midquote} \quad (13)$$

ここで、1回の約定における平均的な出来高の s 倍の成行注文を想定する。 $PI^A(s)$ は仮想的な買いの成行注文を出した場合に到達する価格であり、同様に、 $PI^B(s)$ は仮想的な売りの成行注文を出した場合に到達する価格を表している。 $midquote$ は最良売り気配値と最良買い気配値の平均値である。

3.3 実験方法

使用したデータを1週間ごとの区間に分割し、各区間を1つの時系列データとして扱う。各区間をさらにある時間間隔で分割し、この時間間隔で板の各特徴量を計算する。移動エントロピーの計算ではデータを離散化する必要があるため、各特徴量のそれぞれの時系列データにおける平均と標準偏差を用いて1と0と-1に3分割する。データの数が均等になるようにパラメーターを設定し、データを3分割する。この離散化したデータを用い、式(1)で $k = l = 1$ と

し、各区間で移動エントロピーを計算する。

板の時間間隔としては1~10秒の1秒ごと、15秒、20秒、25秒、30秒、45秒、60秒の16種類を用いる。他の銘柄に影響を与える板の特徴量は3.2で述べたものすべてを用い、他の銘柄から影響を受ける板の特徴量はリターンの絶対値と出来高を用いる。

各区間において、時間間隔と他に影響を与える板の特徴量を変化させ、それぞれにおいて、すべての銘柄間の移動エントロピーを計算し、その平均を求めることで、銘柄間の関係性をうまくとらえられる板の特徴量の分析を行う。

3.4 実験結果

実験結果を図1から図5に示す。横軸が他の銘柄の板に影響を与える特徴量を表しており、縦軸がすべての銘柄ペアの移動エントロピーの平均を表している。他の銘柄から影響を受ける特徴量としては出来高を用いており、時間間隔は図1から順に1秒、10秒、15秒、30秒、60秒の結果を示している。また、影響を与える特徴量は一部を省略して示している。

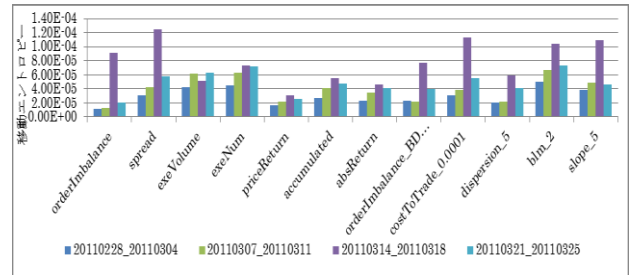


図1: 時間間隔 1 秒

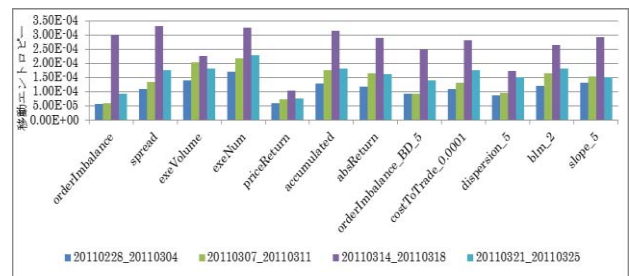


図2: 時間間隔 10 秒

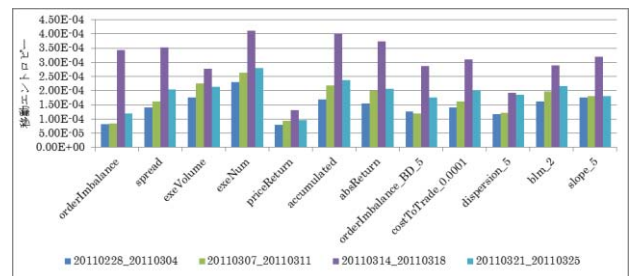


図3: 時間間隔 15 秒

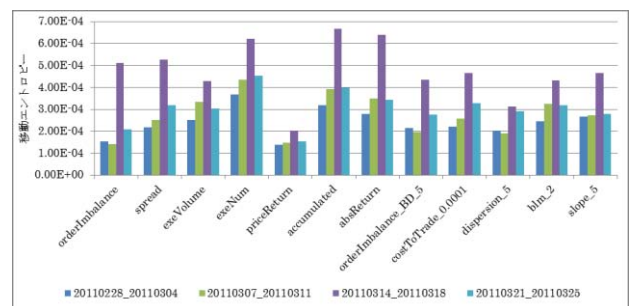


図4: 時間間隔 30 秒

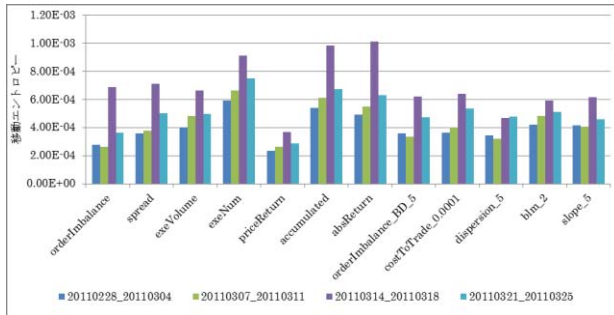


図 5: 時間間隔 60 秒

上記の結果から、いずれの時間間隔、特徴量においても、震災直後の 20110314_20110318 の区間で最も移動エントロピーの平均値が大きくなっている。すなわち、震災の影響により銘柄間の関係性が密になっていることがわかる。

震災直後の 20110314_20110318 の区間に着目する。図 1 と図 2 より、時間間隔が 1 秒から 10 秒の間は spread が最も銘柄間の関係性を強くとらえられていることがわかる。これは、本実験では同業界から複数銘柄ずつ選択しているため、ペアトレードが行われているからであると考えられる。すなわち、ある銘柄で spread が大きくなる時は同様の他の銘柄も spread が大きくなり、収益機会となっているために出来高も増加していると考えられる。また、図 3 より、時間間隔が 15 秒になると約定回数が最も銘柄間の関係性を強くとらえられていることがわかる。図 4 より、時間間隔が 30 秒では累積変動価格が、図 5 より、時間間隔が 60 秒ではリターンの絶対値が最も銘柄間の関係性を強くとらえられるようになる。すなわち、これらの特徴量を指標として収益機会、あるいは株を手放す機会をとらえ、活発な売買につながっていると考えられる。

震災直後の 20110314_20110318 の区間について着目すると、いずれの時間間隔、区間においても、約定回数が最も銘柄間の関係性を強くとらえられている。すなわち、ある銘柄の出来高やリターンよりも、取引頻度が他の銘柄の出来高に影響を与えていることがわかる。

4. まとめ

今回の実験より、震災直後には銘柄間に新たな関係性が生じ、さらにその関係性は、震災直後以外の区間に表れる関係性よりも大きな関係性であることがわかった。また、板を見る時間間隔を変化させると、銘柄間の関係性を強くとらえられる特徴量に変化することがわかり、これは単なる価格データを用いるだけではできない分析であった。この実験を震

災以外の大きな出来事が生じたときにも行うことで、出来事の性質による銘柄間の関係性の変化の仕方を分析する必要がある。

参考文献

- [1] D. Harmon, M. A. M. Aguiar, D. D. Chinellato, D. Braha, I. R. Epstein and Y. Bar-Yam.: Predicting economic market crises using measures of collective panic, arXiv:1102.2620v1, (2011)
- [2] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral and H. E. Stanley.: Universal and Nonuniversal Properties of Cross Correlations in Financial Time Series, Physical Review Letters, 83, 7, pp.1471-1474, (1999)
- [3] C. Eom, G. Oh, W. -S. Jung, H. Jeong and S. Kim.: Topological properties of stock networks based on minimal spanning tree and random matrix theory in financial time series, Physica A, 388, pp. 900-906, (2009)
- [4] L. Kullmann, J. Kertész and K. Kaski.: Time-dependent cross-correlations between different stock returns: A directed network of influence, Physical Review E, 66, 026125, (2002)
- [5] T. Schreiber.: Measuring Information Transfer, Physical Review Letters, 85, 2, pp. 461-464, (2000)
- [6] R. Marschinski and H. Kantz.: Analysing the information flow between financial time series; An improved estimator for transfer entropy, The European Physical Journal B, 30, pp.275-281, (2002)
- [7] O. Kwon and J. -S. Yang.: Information flow between stock indices, Europhysics Letters, 82, 68003, (2008)
- [8] O. Kwon and J. -S. Yang.: Information flow between composite stock index and individual stocks, Physica A, 387, pp. 2851-2856, (2008)
- [9] W. Shi and P. Shang.: Cross-sample entropy statistic as a measure of synchronism and cross-correlation of stock markets, Nonlinear Dynamics, 71, pp. 539-554, (2013)
- [1 0] R. Næs and J. A. Skjeltorp.: Order book characteristics and the volume-volatility relation: Empirical evidence from a limit order market, Journal of Financial Markets, Vol. 9, pp. 408-432, (2006)
- [1 1] W. Kang and W. Y. Yeo.: Liquidity Beyond the Best Quote: A Study of the NYSE Limit Order Book, WFA 2008 Conference Paper, (2008)
- [1 2] P. J. Irvine, G. J. Benston and E. Kandel.: Liquidity Beyond the Inside Spread: Measuring and Using Information in the Limit Order Book, working paper, Emory University, (2000)