

移動エントロピーを用いた 銘柄間影響度ネットワークによる投資指標の分析

Analysis of the investment indexes constructed from influence networks
among stocks by using Transfer Entropy

小村 和輝¹ 鳥海 不二夫¹ 大橋 弘忠¹

Kazuki Komura¹, Fujio Toriumi¹, and Hirotsada Ohashi¹

¹ 東京大学工学系研究科

¹ School of Engineering, The University of Tokyo

Abstract: Recent research has explored the proper method to analyse the relationships in financial markets for risk management. In this paper, we applied and expanded the method of transfer entropy to construct the influence network which represents the information propagation between stocks. First, we demonstrate that this method reveals meaningful hidden relations of cause and effect between stocks, by showing that it is more useful than other methods, such as normal Transfer Entropy network, correlation coefficient network and partial correlation coefficient network. Second, we construct the influence network and mention about its qualitative features. Finally, we examine the indexes which is considering the influence network and useful for individual investors.

1. はじめに

近年、証券市場における個人投資家の比重が増大し、その重要性が高まっている。しかし、個人投資家と機関投資家には未だ大きな情報格差がある[1]。したがって個人投資家に対して投資判断の支援をおこなう技術の必要性が増しており、特に突発的な出来事が及ぼす自身の保有銘柄に対する影響の把握が重要となる。そのような影響度を定量化した銘柄間のネットワークを適切に構築することができれば、ある銘柄の価格やボラティリティの変化が個人投資家の保有する他の銘柄にどう波及するかを事前に知ること役立つ可能性がある。

本研究では、そのような銘柄間の影響度ネットワークを構築する上で、移動エントロピーを拡張した手法が有用であることを示す。さらには個別銘柄間の情報の流れを可視化し、定性的には予測し難い銘柄間の関係性を示し投資指標等に還元することで、個人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。

2. 既存研究

銘柄間の関係性を表すネットワーク構築には、これまで様々な手法が用いられてきた。Plerou らはランダム行列理論を用い、NY 証券取引所の銘柄間についての相互相関行列を分析した[2]。Eom らは最小全域木を用いて構成した銘柄間ネットワークのトポロジー性質から、その関係性を分析した[3]。Kullman らは各銘柄の価格データ

を用い、時間差相関係数を計算することで、銘柄間の影響度分析を試みた[4]。しかしこれらの手法は因果関係を考慮していないことや定常的な性質に着目していることから、突発的な出来事に対する分析には不向きであり、投資家にとっては有用でない。

移動エントロピーを用いると、上記の手法とは異なり、有向性の定義されていない時系列データから因果関係を動的に分析できる。また、確率をもとにした計算方法であるため、状態推移の確率変化を見ることも可能となり、離散値を用いるため外れ値の影響を受けなくなる。

Schreiber は時系列相互相関と比較し、移動エントロピーが因果関係を見る上で優れていることを示した[5]。Greg らは、Twitter の単純な時系列データからノード間の情報伝播を分析し、移動エントロピーの有用性を示した[6]。Marschinski らは本手法を用い、ダウ平均株価からDAX 指数への価格情報の流れがあることを示した[7]。Kwon らは対象とするインデックスの種類を25に増やし、情報伝播のネットワークを構築しその性質を分析するとともに、個別銘柄への情報伝播があることを示した[8] [9]。また、鈴木らは、相互相関係数、偏相関係数、移動エントロピーの3つの指標を用い、移動エントロピーを用いた手法が最も有用であることを示した[10]。

ところがこれらの研究はすべて個別銘柄ではなくインデックスや銘柄グループ間の情報伝播を対象としており、且つ定常的な分析にとどまっているため、個別銘柄への突発的な変化に関心を持つ投資家に役立つものではない

3. 本研究で用いる手法

3.1 拡張移動エントロピー

移動エントロピーは、離散変数で表された 2 つの定常過程間の情報伝播を定量化するために Schreiber[7]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, J からそれぞれ k 個、 l 個のサンプルを用いて計算する場合、過程 J から過程 I の向きへの移動エントロピー $T_{J \rightarrow I}$ は以下ようになる。

$$T_{J \rightarrow I} = \sum p(i_n, i_{n-1}^{(n-k)}, j_{n-1}^{(n-l)}) \log \frac{p(i_n | i_{n-1}^{(n-k)}, j_{n-1}^{(n-l)})}{p(i_n | i_{n-1}^{(n-k)})} \quad (1)$$

ただし、 i_n と j_n はそれぞれ過程 I と J の時間における離散状態を表す、 $i_n^{(n-k)}$ は過程 I の時間 n 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、

$$i_n^{(n-k)} = (i_n, i_{n-1}, \dots, i_{n-k+1}) \text{ と表される。} j_n^{(l)} \text{ も同様に、}$$

$$j_n^{(n-l)} = (j_n, j_{n-1}, \dots, j_{n-l+1}) \text{ と表される。}$$

$T_{J \rightarrow I}$ は過程 J の情報が、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あらわす。つまりは j_n の情報が与えられた際に、過程 I の状態遷移の不確かさをどれだけ減少させるかを定量化している。よって、以下の式に書き直すことができる。

$$T_{J \rightarrow I} = H\{I_n | I_{n-1}^{(n-k)}\} - H\{I_n | I_{n-1}^{(n-k)}, J_{n-1}^{(n-l)}\} \quad (2)$$

$$H(A|B) = -\sum_{A, B} P(A, B) \log P(A|B) \quad (3)$$

また、この式から分かるように、移動エントロピーは 2 つの離散変数 i_n と j_n に関して非対称になっており、これにより 2 変数間の情報伝播の方向性がわかる。

しかし、この手法では 2 変数間の状態遷移の符号の関係を区別できず、また遡る状態数に応じて状態の組み合わせが著しく増大してしまう。したがって、今回用いた

手法では、 $i_n^{(n-k)} = (i_n, i_{n-k+1})$ 、 $j_n^{(n-l)} = (j_n, j_{n-l+1})$ とし、確率計算においては、状態遷移が同一の状態間で起こる場合 $(i_{n+1}, j_n) = (+1, +1), (-1, -1)$ と、異なる状態間 $(i_{n+1}, j_n) = (+1, -1), (-1, +1)$ で起こる場合をそれぞれ区別した。前者を正の移動エントロピー、後者を負の移動エントロピーとし、前者から後者を差し引いた値を拡張移動エントロピーまたは拡張 TE とする。

3.2 使用データ

本研究では、比較的市場が安定していたと考えられる 2011 年 1 月 1 日から 2011 年 2 月 1 日で、東京証券取引所の取引時間である平日の 9:00~11:00 と 12:30~15:00 の 1 秒ごとの価格情報から得られるリターン情報を用いる。リターン幅は 5 秒で離散化し、リターンが正ならば 1、

負ならば -1、0 ならば 0 とする。1 つの拡張移動エントロピーの計算に用いるデータ数は一日の総取引時間を秒数で表した数である 16200 となる。今回は期間内の拡張 TE をそれぞれ平均したものを定常的な拡張 TE として用いる。また、今回使用する銘柄は TOPIX CORE30 に含まれる銘柄及び日経平均先物のみとする。これは、流動性が小さく注文情報が疎な他の銘柄では、データ量の少なから十分正確に確率が計算できないためである[8]。

4. テストデータによる妥当性検証

検証内容に応じたテストデータをそれぞれ用意し、本手法の妥当性を確認した。

まず遅延の検証を行った。日経平均先物、ソフトバンク、トヨタ、みずほ FG、セブン&アイのデータを基に、それぞれ 5 秒後、10 秒後、30 秒後、60 秒後、180 秒後に同一の値動きをする 5 つのテストデータを作成した。つまり、常に一定秒数遅れて元データからの影響を受けているデータをテストデータとして使用した。図 1 は遡る変数の数を 1 から 300 まで増加させ、元データからテストデータへの向きの影響を表す拡張 TE が情報伝播の遅延によってどう変化するかを示す。この結果から、5 つのテストデータ全てにおいて遅延秒数と同じ伝播時間に情報伝播のピークが観測できる。したがって、拡張 TE を用いると情報伝播の遅延の様子を捉えられるということがわかった。

次に、方向性の検証を行った。先と同様に、日経平均先物のデータを基に 30 秒後に同一の値動きをするテス

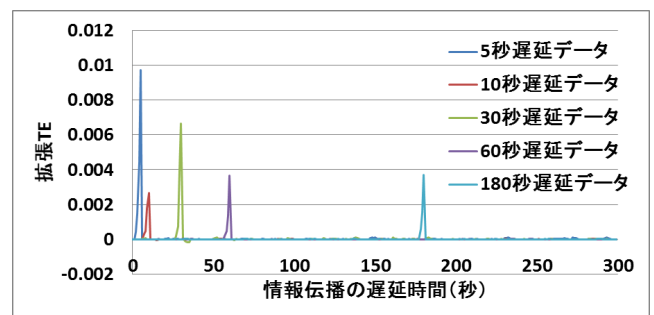


図 1: 情報伝播の遅延の検証

トデータを作成した。また、元データからテストデータへの影響を順向き、テストデータから元データへの影響を逆向きとした。図 2 は拡張 TE、図 3 は 2 銘柄の各リターン情報を用いた通常の移動エントロピーでの結果を表し、情報伝播の遅延時間による変化を観察した。また時系列データ間の関係性を表す一般的手法である、交差相関及び自己相関を取り除いた交差偏相関の結果を図 4 に示す。これらによると、全ての手法において順向きで伝播時間 30 秒の際に情報伝播のピークが見られ、逆向きには見られない。したがって、どの手法においても情報伝播の方向性を捉えられていることがわかった。

次に、符号の関係性について検証を行った。日経平均先物のデータを基に 30 秒後に逆の値動きをするテスト

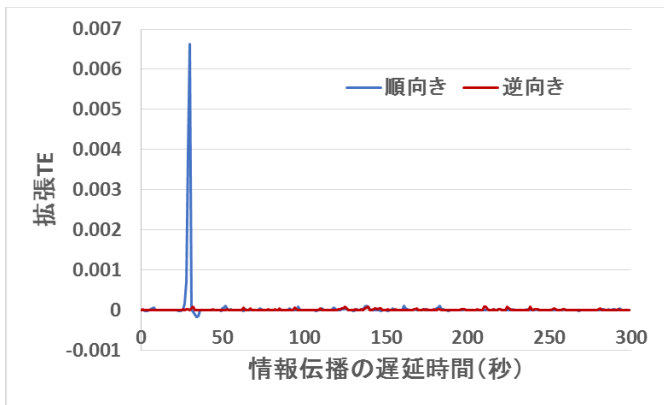


図 2: 拡張 TE での方向性の検証

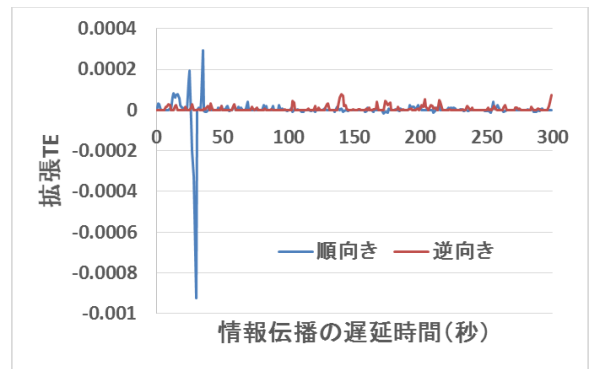


図 5: 拡張 TE での符号の関係性の検証

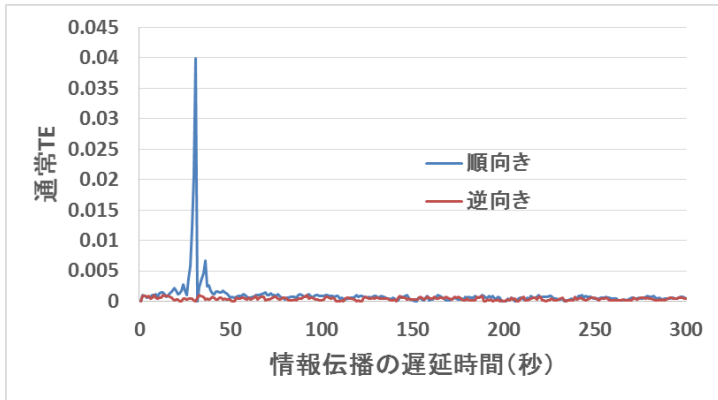


図 3: 通常 TE での方向性の検証

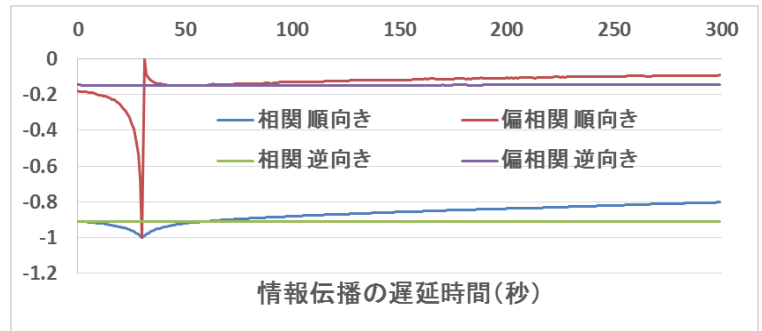


図 6: 相関・偏相関での符号の関係性の検証

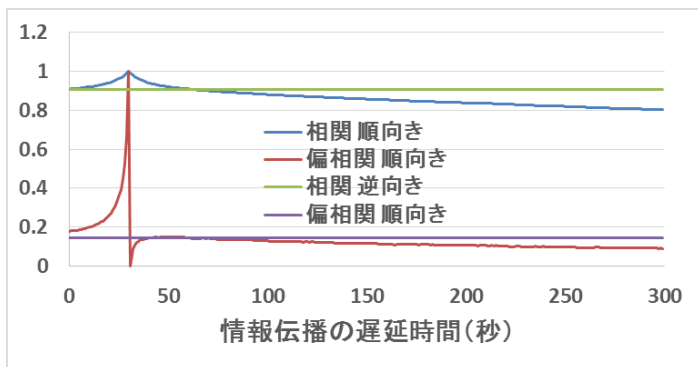


図 4: 相関・偏相関での方向性の検証

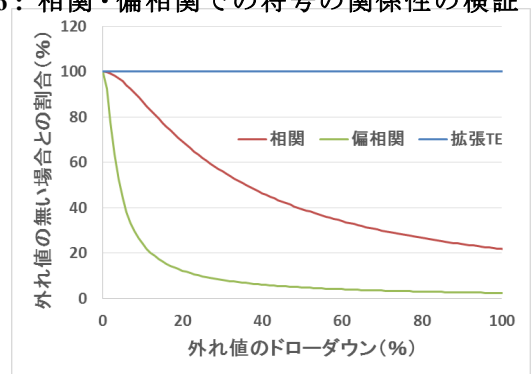


図 7: 外れ値のドローダウンによる影響の検証

データを作成した。つまり、常に一定秒数遅れて元データから負の向きの影響を受けているデータをテストデータとして使用した。図 5, 図 6 はそれぞれ拡張 TE, 相関及び偏相関での結果を示す。これらから、拡張 TE と相関及び偏相関では元データとテストデータ間の符号の関係性を捉えられている。しかし、通常 TE の値の変化は正の向きの場合である図 3 と同じであった。したがって、符号の関係性は分からないといえる。

次に、外れ値による影響について検証を行った。先と同様に、日経平均先物のデータを基に 30 秒後に同一の値動きをするテストデータを作成したが、このとき、元データの 16200 個の価格データのうちの 1 つを外れ値とし下落率(ドローダウン)を 0% から 100% まで変化させた。その際の相関、偏相関、拡張 TE それぞれのピーク値の変化を図 7 に示す。これにより、拡張 TE と比較し他の二つの手法は外れ値の影響を受け、偏相関は特に著しく値が変化することがわかった。

5. 影響度ネットワークの構築

本データをもとに、拡張 TE を計算し有向重み付きネットワークを構築した。その際、遡る変数の数を 1 から 300 まで変化させ、リターン情報の伝播期間が 1 秒から 300 秒までの拡張 TE を各銘柄間で計算し、そのうち最大となるものをエッジの重みとして採用した。また、0, 1, -1 それぞれを等確率で発生させたランダムな離散変数との拡張 TE をそれぞれ計算し、エッジを張る際の閾値とした。その上で、全エッジのうち重みが上位 5% ののみを残し構築したネットワークを図 8 に示す。ただし、エッジの太さは拡張 TE の絶対値の大きさ、矢印は影響の向き、青は正の向きの関係性、赤は負の向きの関係性をそれぞれ表し、ノードは次数でソートして円形に配置した。

これによると、TOPIX CORE30 の各銘柄は正の関係性を持つものが多く、赤いエッジが無いことから上位 5%

の大きさの影響については全て正の向きに伝播していることがわかる。しかし、2銘柄または3銘柄の間でループを形成していたのは野村-三井住友FG間のみであり、自らの影響が他の銘柄を介して自らに伝播し値動きがさらに大きくなるということは起こりにくく、安定的な構造となっていることがわかる。

また、同じ次数の高い銘柄であっても、コマツは影響を受ける関係が多く、みずほFGは影響を与える関係が多い。これは、コマツが比較的時価総額の小さい銘柄であることや、みずほFGの業種の性質によるものと考えられる。

また、特に影響度の大きい三菱商事から三菱UFJ、三菱UFJからコマツ、三菱商事からコマツ、という3つのエッジから、三菱商事から三菱UFJを仲介としてコマツに情報伝播が occurring 可能性があると考えられる。同様の構造はいくつか見られるが、エッジの重みが大きい先の3銘柄の組み合わせで特に顕著に現れている。

6. 情報伝播を考慮した投資指標分析

6.1 遅延の影響の重ね合わせ

本研究の主眼は個人投資家への情報提供にあり、具体的な投資指標への還元方法を検討する必要がある。そこ

で拡張TEによる投資指標を用いて仮想的に投資を行うことで、その有用性を検討する。今回は特に、情報伝播の遅延を考慮した場合について検証する。

具体的な検証方法としては、1つの銘柄の情報を用いて1つの銘柄に投資する場合を考え、元銘柄からの拡張TEと元銘柄の離散値（リターンが正ならば1、負ならば-1、0ならば0）を掛け、それらをリターン幅ごとに足し合わせた値が正の場合は買いポジション、負の場合は売りポジションをとることとする。さらに元銘柄の値動きの大きさ、拡張TEの大きさで重み付けを行う。これにより、300秒前までの元銘柄の情報を全て重ね合わせた投資指標に基づいた仮想的な投資を評価できる。それぞれ1000回の取引を行い、その結果を全エッジについて平均し、元銘柄の情報を用いない単純な買いポジションの場合との差分をとった結果を図9に示す。ただし、勝率は全取引数のうちのリターンが正の取引数の割合を表し、sharpe ratioはリターンを標準偏差で割ることで、リスク調整済みリターンを表す。

これによると、勝率・リターン共に向上しており、情報伝播の遅延を重ね合わせた投資指標が銘柄間の値動きを捉える上で役立つことが分かった。またsharpe ratioの値から、本指標を用いることで有益なリスクを選定できていることも分かった。

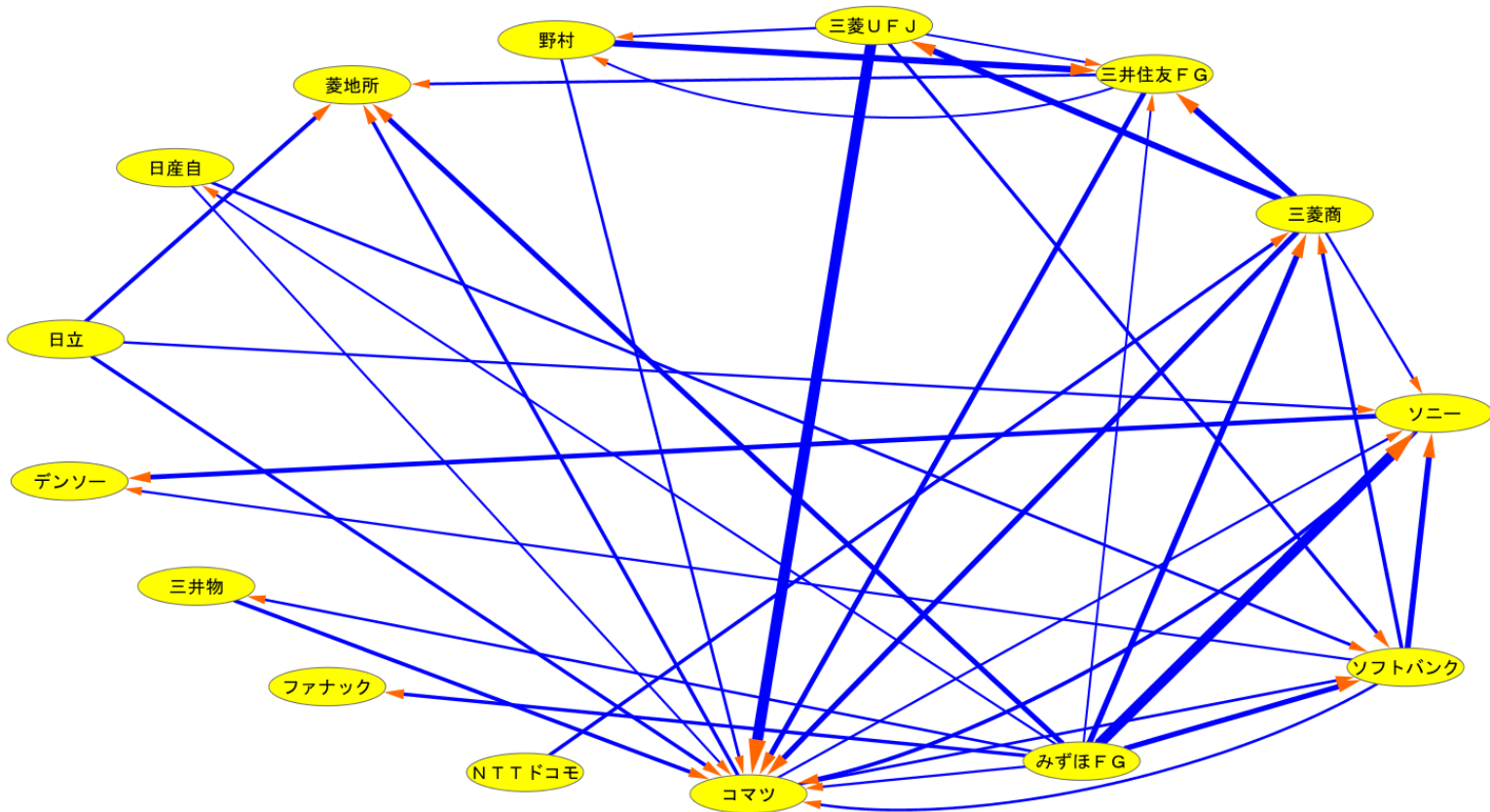


図8：定常的な影響度ネットワーク

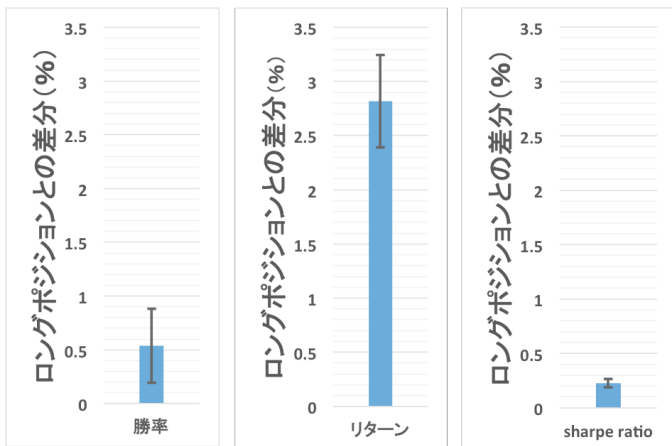


図 9: 情報伝播の遅延を考慮した投資指標の評価

6.2 複数銘柄からの影響の重ね合わせ

先の結果より、さらに複数銘柄の遅延を考慮した投資指標に基づいて仮想的な投資を行い、その有用性を検証する。先と同様に、元銘柄からの拡張 TE と元銘柄の離散値（リターンが正ならば1、負ならば-1、0ならば0）を掛け、それらをリターン幅ごとに足し合わせた値に元銘柄の値動きの大きさ、拡張 TE の大きさと重み付けを行う。これについて複数銘柄の値を重ね合わせ、正の場合は買いポジション、負の場合は売りポジションをとることとする。これにより、300秒前までの複数の元銘柄の情報を重ね合わせた投資指標に基づいた仮想的な投資を評価できる。それぞれ1000回の取引を行い、その結果を全エッジについて平均し、元銘柄の情報を用いない単純な買いポジションの場合との差分をとった結果を図10,11,12,13に示す。横軸は重ね合わせる元銘柄の数を表し、各先銘柄に対して影響度の大きい順に加えていった。

ここでは、本ネットワークにある銘柄のうち、先ノードが全銘柄、出次数が入次数を上回るグループ、入次数が出次数を上回るグループの3つをそれぞれ平均した場合の結果を示す。ただし、勝率は全取引数のうちのリターンが正の取引数の割合を表し、sharpe ratio はリターンを標準偏差で割ることで、リスク調整済みリターン、最大ドローダウンは期間中で最大のリターン下落率を表す。

これによると、勝率は出次数の多いグループの方が入次数の多いグループより常にロングポジションとの差分が高くなるにも関わらず、リターンは加える元銘柄数が12までは入次数の多いグループの方が高く、sharpe ratio では常に入次数の多いグループの方が高く、ドローダウンは常にその逆となることが分かった。つまり、入次数の多いグループの方がリスク選定が適切にできており、影響度の情報がリスク軽減に役立っている。これは、入次数の多いグループには影響を与えるよりも受けることが多い銘柄が多く、元銘柄の情報が突発的な出来事の影響をよりよく表しているためであると考えられる。また、入次数の多いグループはどの指標においても元銘柄の数を増やしてもあまりパフォーマンスは変わらないのに対し、出次数の多いグループは銘柄を増やすにつれ、特に

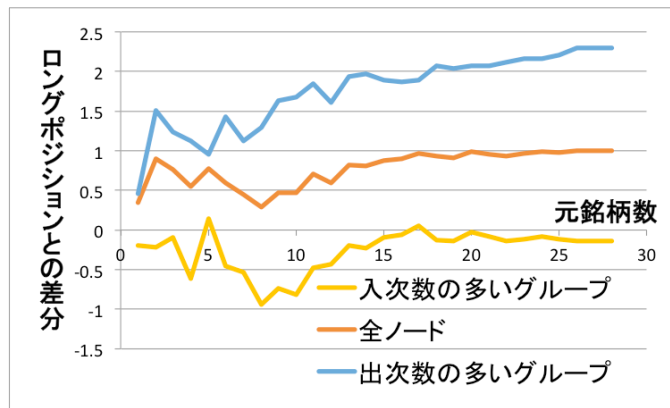


図 10: 複数銘柄からの影響を考慮した投資指標の評価 (勝率)

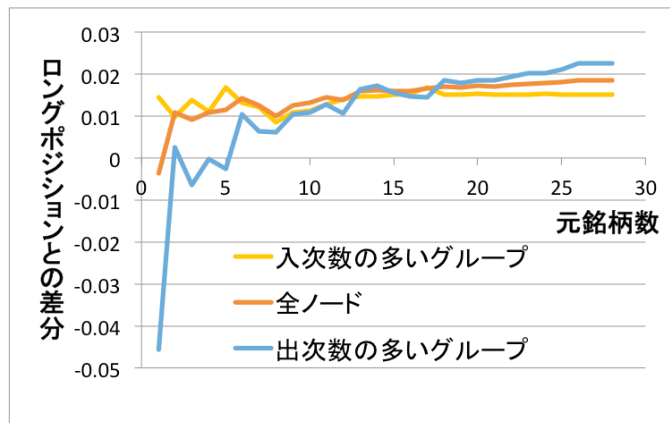


図 11: 複数銘柄からの影響を考慮した投資指標の評価 (リターン)

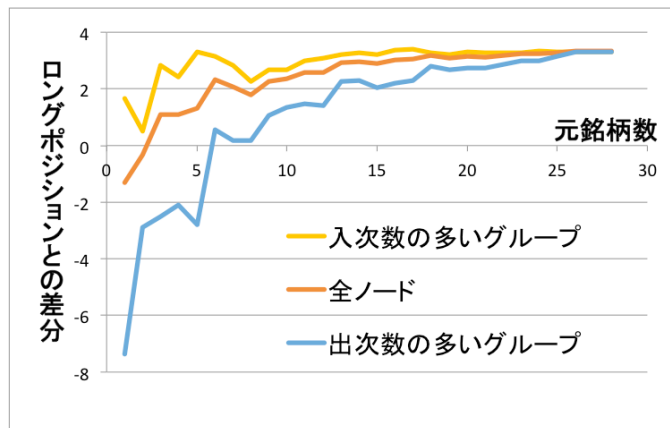
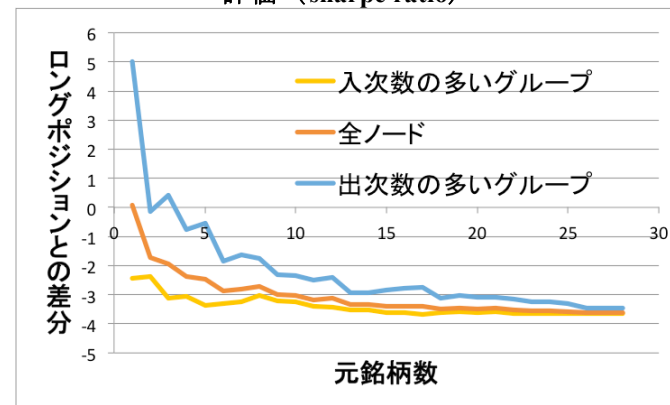


図 12: 複数銘柄からの影響を考慮した投資指標の評価 (sharpe ratio)



- 5 - 図 13: 複数銘柄からの影響を考慮した投資指標の評価 (最大ドローダウン)

元銘柄数が少ない際にパフォーマンスが大きく向上している。これは、前者が比較的様々な銘柄から影響を受け易いため元銘柄数を増やした際に影響度の重ね合わせにおけるノイズが増大してしまうのに対し、後者は比較的少数の銘柄から影響を受けるため、重ね合わせてもノイズが少ないためであると考えられる。また sharpe ratio や ドローダウンといったリスク指標においても、25 を超える程度の多くの銘柄の情報を用いれば入次数の多いグループとほぼ同じ水準となっている。これにより、元銘柄数を一定以上増大させると先銘柄が影響を受け易い銘柄なのか与え易い銘柄なのかに関わらずパフォーマンスはほぼ同様に向上することが分かる。

7. おわりに

本研究では、移動エントロピーの定義を拡張した手法を提案し、株式市場における各個別銘柄のリターンの情報から影響度を定量化する上で、本手法が他の手法に比べ有用であることを示した。また TOPIX CORE 30 に含まれる銘柄間の比較的稳定した市場における影響度ネットワークを構築し、様々な性質を明らかにした。さらに、本手法の情報を還元した投資指標を検討する上で、仮想的な投資を行い、そのパフォーマンスを評価する方法を示し、特に情報伝播の遅延や複数銘柄からの影響を考慮することが有用であることを明らかにした。

今後の課題としては、以下の三点が挙げられる。一つ目に、様々な環境下における影響度ネットワークの分析を行う必要がある。安定した市場の場合と比較し市場急変時には銘柄間の関係性にはどのような変化があるのかや、より流動性の小さな銘柄を加えた際のネットワークがどうなるのか等を明らかにし、その情報の投資指標への還元方法を検討しなければならない。二つ目に、影響度ネットワークで見られる現象をより定量的に示す必要がある。伝播の遅延やネットワーク構造を見ることで、銘柄間の情報伝播の媒介ノードを同定することや、影響力が大きい銘柄を把握すること等を定量的に行うことで、投資指標に応用できる可能性がある。三つ目に、複数銘柄からの影響の重ね合わせを考慮した投資指標をさらに検討する必要がある。今回は単純に影響度の大きい銘柄からの影響の遅延を重ね合わせた投資指標の有用性を示したが、最終的には銘柄選定等も検討し、より確実にリスク管理に役立つ投資指標を提案しなければならない。

8. 参考文献

- [1] 秦, 劼: 情報の非対称性, 能力の非対称性, と個人投資家の変, 大阪大学経済学, 57(4) P. 164-P. 176, (2008)
- [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] Greg Ver Steeg, Aram Galstyan, Information Transfer in Social Media, WWW 2012 – Session: Information Diffusion in Social Networks, (2012)
- [7] 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)
- [8] O, Kwon and J, -S, Yang. : Information flow between stock indices, Europhysics Letters, 82, 68003, (2008)
- [9] O, Kwon and J, -S, Yang. : Information flow between composite stock index and individual stocks, Physica A, 387, pp, 2851-2856, (2008)
- [10] 鈴木智也, 池口徹, 堀尾喜彦, 日経株式市場に隠された複雑ネットワーク構造の推定, THE INSTITUTE OF ELECTRONICS, INFORMATION AND COMMUNICATION, IEICE Technical Report., NLP2005-122, (2006)