

板情報を用いた金融市場の相転移検出

Detection of the phase transition on financial markets using stock boards

西岡寛兼 鳥海不二夫 石井健一郎

Hirokazu NISHIOKA Fujio TORIUMI Kenichiro ISHII

名古屋大学大学院情報科学研究科

Graduate School of Information Science, Nagoya University

The financial markets are fluctuating consistently. Therefore, it is difficult to analyze the financial market based on the same theory, without depending on the state of the market. So we use the concept called the phase transition in the field of econophysics. To estimate the points when the phase transition occurred in a real market is effective for market analysis. Thus, in this paper, we propose a method to detect the phase transition. Our proposed method focuses on the stock board instead of the price data. We performed an experiment to evaluate the effectiveness of our method. As a result, our method succeeded in detecting the phase transition. In addition, our method detected the phase transition earlier than the price data showed it.

1. 序論

従来の市場理論では、あらゆる時点における市場の動きを、同一の理論に基づいて説明することを目指してきた。これに対し、経済物理学の分野では、相転移と呼ばれる概念を用いる [Mantegna 00]。この概念は、市場に、例えば「バブル時の相」、「不況時の相」といった、市場の内部状態を表す相の存在を仮定し、相が転移することによって市場の動きが変化するというものである。この概念を用いると、例えば 2008 年の金融危機化における市場は特殊な相を持っているために異常な値動きを見せた、という説明が可能になる。このように、相転移という概念は、より現実に即した考え方であるといえる。

相転移という概念に基づいて市場を分析するためには、いつ相転移が起こったのかを知ることが必要である。これにより、相転移の前後で市場にどのような変化が起こったかという、次の段階の分析に移ることが可能になる。また、市場分析以外にも、リアルタイムに相転移を検出することにより、様々な応用が可能であると考えられる。例えば、投資家の投資方針決定に利用することが可能である。また、バブルの検出や価格暴落の予知などにも利用可能であり、投機的な取引の規制など金融市場の健全化に有用であると期待される。そこで本研究では、定量的・客観的に相転移の発生を検出する手法を確立する。

これまで多くの研究では、市場を分析する際に価格データを観測し、利用してきた。それに対し本研究では板情報を利用する。板情報とは、全てのトレーダーの注文状況を表にまとめたものである。金融市場では、図 1 に示したように、まずトレーダーが出した注文が板情報にまとめられる。そして、板情報に基づいて約定が決定され、その結果として、価格が決定される。相転移検出に用いるデータを観測する際には、より多くの情報量を含む対象を観測することが望ましいが、図 1 に示したように、価格データは情報量が減少している。一方、個々のトレーダーの行動や心理を観測することでより詳細な分析が可能となるが、これは非現実的である。そこで本研究では、価格データより多くの情報を含み、かつ観測・分析が現実的である、板情報を用いる手法を提案する。

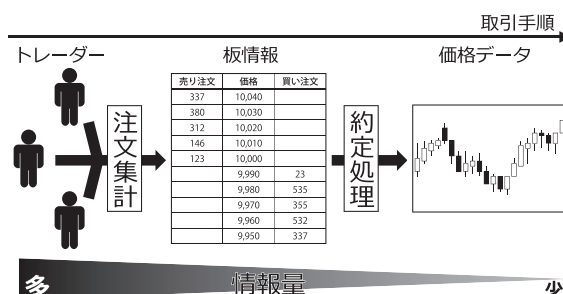


図 1 取引の流れと情報量の減少

2. 相転移検出手法

2.1 概要

本節では、本研究で提案する相転移検出手法について述べる。提案手法では、以下の 2 段階の操作によって相転移を検出する。

1. 市場の類似度計算
2. 類似度時系列から相転移推定

隣接する 2 期間について類似度が小さければ、2 期間の間で市場の動きが変化しており、その点で相転移が発生したとみなすことができる。これを利用して、隣接する 2 期間における類似度を期間をずらしながら求め、その時系列を観察することにより、相転移が発生した時点を推定することが可能である。

2.2 板情報の形式

提案手法では板を観測して得た板情報を分析対象とする。そこで、予め板情報の形式について述べておく。

板に表示される値は、注文が出される度に变化する。よって、板は時間軸を持つリアルタイムな情報である。これに対し、板のある時点における状態を観測したものを、本論文では板情報スナップショットと呼ぶ。

本研究では、板に表示される値が変化する度に板情報スナップショットを観測する。これによって得られた板情報スナップショットの列を、板情報として扱う。

連絡先: 名古屋大学大学院情報科学研究科

〒 464-8603 愛知県名古屋市千種区不老町

E-mail: nishioka@kishii.ss.is.nagoya-u.ac.jp

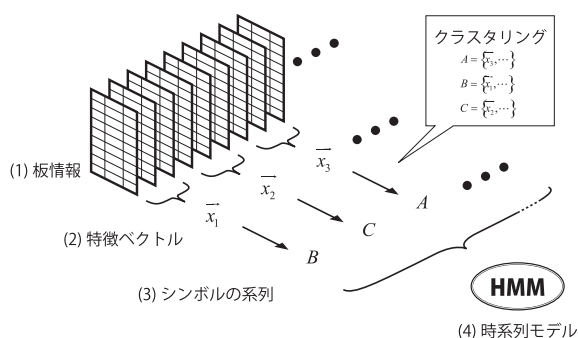


図 2 類似度計算手法の流れ

2.3 類似度計算手法

我々はこれまでに、板情報を用いて市場の類似度を計算する手法を提案している [西岡 09]。本研究では、この手法を応用する。

類似度計算手法の流れを図 2 に示す。初めに板情報を、後述する定められた規則に基づいて分割する。そして、分割された板情報の各組から特徴抽出を行い、特徴ベクトルを生成する。次に、生成された特徴ベクトルに対してクラスタリングを行い、類似した特徴ベクトルを集めたクラスタに分類する。各クラスタにシンボルを付与することにより、シンボル系列が生成される。この系列を基に、HMM (Hidden Markov Models) [Rabiner 89] を作成する。HMM を用いることで、その HMM を作成するために使用した学習系列と、任意の入力系列との類似度を計算することが可能となる。

2.4 特徴抽出

本節では、板情報から特徴ベクトル列を生成する手順について述べる。

まず板情報を分割する。板情報を分割して得られた板情報スナップショットのまとまりをユニットと呼び、ユニット 1 つにつき 1 つの特徴ベクトルを生成する。板情報は、価格変動の発生直前と発生直後で分割する。なお、本論文において価格変動とは、売り注文価格と買い注文価格の境界が移動することを言う。

次に、ユニットから特徴抽出を行い、特徴ベクトルを生成する。特徴抽出の流れを図 3 に示す。

まず前処理として、出現する全ての売り注文数を負の数に変換する。次に、注文数変化量を求める。i 番目の板情報に表示されている価格 p に対する注文数を $Q(i, p)$ としたとき、注文数変化量 $D(i, p)$ を、

$$D(i, p) = Q(i, p) - Q(i - 1, p) \quad (1)$$

とする。ユニット内に出現する全ての注文数 $Q(i, p)$ に対して求められた注文数変化量 $D(i, p)$ を、そのユニット内で出現する注文数変化量とする。本研究では注文が出される度に板情報スナップショットを観測しているため、注文数変化量は、出された注文のサイズを表している。また、予め行った前処理により、売り注文の増加は負数、買い注文の増加は正数の注文数変化量となり、両者を区別することが可能である。

最後に、注文数変化量から、特徴ベクトルを生成する。特徴ベクトルの各要素には、「どのような規模の売り注文（買い注文）が何回出されたか」という情報を持たせる。そのために、まず、注文数変化量を、売りか買いかと、注文の規模に基づいて分類する。売りか買いかについては、前述の通り、注文数変

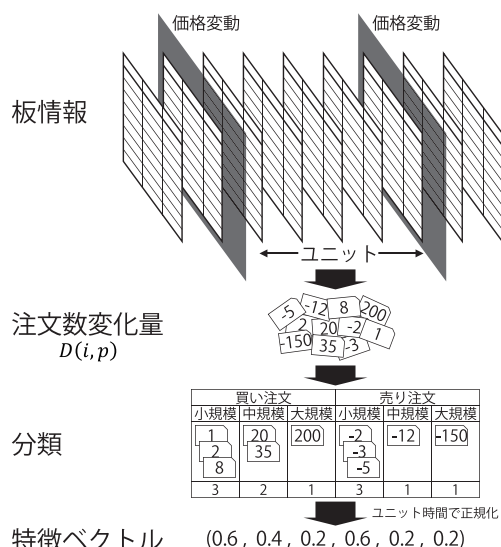


図 3 特徴抽出

化量の符号が正ならば買い注文、負ならば売り注文として分類する。注文の規模については、注文数変化量の絶対値に境界を定めることで分類する。なお、規模の境界の決定方法については後述する。次に、各分類に属する注文数変化量が、ユニット内に何回出現したかを数える。そして、この出現回数を、このユニットの時間で割ることにより正規化し、注文の分類ごとに並べてベクトルとしたものを特徴ベクトルとする。

注文の規模に基づいた分類を行う指針を決定するため、板情報から観測された注文について、規模の分布を調査した。2006年の日経 225 先物板情報から観測された全ての注文について、その規模の分布の様子を図 4 に示す。これを見ると、両対数グラフでほぼ直線上にプロットされており、注文の規模はベキ分布に従うと考えられる。

また、図 4 に円で示したように高い頻度を示す点が孤立して出現した。これらの点は、5, 10, 20, 50, 100, 200 といった、人間にとってキリの良いサイズの注文であった。

これらの現象は、日経 225 先物以外の銘柄についても確認された。しかし、孤立して出現する高い頻度を示す点については、銘柄によって異なる注文サイズにおいて現れた。これは取引の最小単位が銘柄によって異なることなどに起因すると考えられる。

以上の結果から、本研究では、分析対象となる銘柄に対する注文の規模の分布を予め調査し、孤立して出現する高い頻度を示す注文サイズを求めた。そして、この値を注文の規模を分類するための境界として用いた。

2.5 類似度の算出

本節では、特徴抽出によって得られた特徴ベクトル列を用いて、市場の類似度を算出する手順について述べる。

まず、特徴ベクトルの各要素を標準偏差で割ることによって正規化する。そして、正規化した特徴ベクトルのクラスタリングを行う。クラスタリングは、k-means 法 [MacQueen 67] を使用する。そして、分類された各クラスタにシンボルを付与することにより、ベクトル量子化を行う。この結果、特徴ベクトルを時系列順に並べた列から、シンボルの系列が生成される。

次に、シンボル系列を学習系列として HMM パラメータを推定する。HMM とは、時系列パターンを確率的に生成するモデルである。パラメータ推定には EM アルゴリズム [Levinson 83]

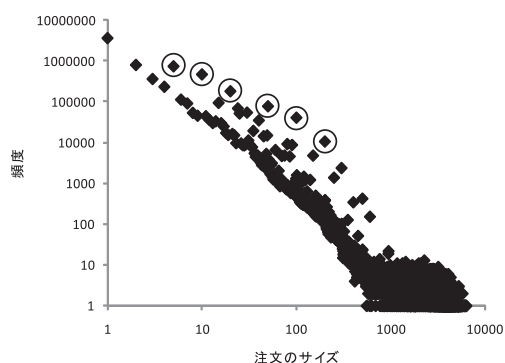


図 4 注文の規模の分布

を用いる。

HMM を作成すると、任意の入力系列が、この HMM によって出力される確率を求めることができる。この出力確率は、入力系列と学習系列との類似性と考えることができる。本研究では、出力確率を入力系列の長さで正規化したものを類似度とする。

ここで、市場から得られるデータと類似度との関係について述べる。板情報は、市場で取引が行われている時間帯のみ観測される。従って、1 日当たり前場と後場、2 系列の板情報が観測される。故に、得られるシンボル系列も 1 日当たり 2 本である。また、HMM のパラメータ推定は、複数の系列を同時に学習系列とすることができる。従って、1 か月分のシンボル系列（取引日数 × 2 本）から、その月を表すモデルを作成することが可能である。一方、モデルに対する類似度を計算する際の入力系列は、1 本でなければならない。従って、計算できることのできる類似度は、例えば「ある日の前場の動きと、モデルを作成した月との類似度」のように、1 本の系列と複数系列とを比較して得られる類似度となる。

3. 提案手法の有効性評価実験

3.1 目的

提案手法を用いて市場の相転移検出シミュレーションを行う。シミュレーションは、現実の市場から観測された板情報を用いて行い、本手法によって市場の相転移検出が可能であるか検証し、提案手法の有効性を確認する。

また、既存の価格データに基づいた市場モデルを作成し、これを用いた分析による結果と本手法の結果を比較する。これにより、価格データではなく板情報を分析対象とすることの優位性を確認する。

3.2 GARCH モデルを用いた類似度算出

GARCH(1, 1) モデル [Bollerslev 86] を用いた類似度算出を行い、提案手法と同様に相転移検出を試みる。GARCH モデルは価格に基づくモデルであり、これと提案手法を比較することで、板情報を用いる優位性を確認する。

GARCH モデルは時系列データを表す確率モデルで、金融資産時系列データの分散の不均一性を捉えるモデルとして優れているといわれている。GARCH(p, q) モデルは、時系列 x_t を以下の式で表現する。

$$x_t = x_{t-1} + \epsilon_t \quad (2)$$

$$\epsilon_t \sim N(0, h_t) \quad (3)$$

$$h_t = \alpha_0 + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^p \beta_i h_{t-i} \quad (4)$$

ただし、 $N(0, h)$ は平均 0、分散 h の正規分布を表し、 α_i, β_i はモデルパラメータである。

GARCH モデルは、式 (3) が示すように、分散不定の正規分布に従った価格変動を表す確率モデルであり、最尤推定を用いてモデルパラメータを推定することができる。また、モデルに入力系列を与えると、モデルにおける入力系列の尤度を計算することができる。この尤度を類似度として扱う。

3.3 シミュレーション設定

本シミュレーションは、現実の市場で観測された板情報^{*1}、および価格データを用いて行う。銘柄は新日本製鐵で、2006 年 3 月から 12 月の 10 か月間を対象とする。

提案手法を用いた実験では、1 か月分のデータ学習データとしてモデルを作成し、学習データの次の月の各系列について、モデルとの類似度を計算した。なお、特徴抽出の際に注文の規模を分類する境界として、10000, 100000, 1000000 を用いた。

GARCH モデルを用いた実験では、2 種類のシミュレーションを行った。1 つは、月単位で類似度を算出するシミュレーションで、もう 1 つは半日単位で類似度を算出するシミュレーションである。

1 つ目の月単位の類似度系列を観察するシミュレーションでは、日足終値データを用いた。1 か月分の日足終値系列を学習データとしてモデルを作成し、次の月の日足終値系列について、モデルとの類似度を計算した。

2 つ目の半日単位の類似度系列を観察するシミュレーションは、提案手法と同様に、1 か月分のデータからモデルを作成し、半日分のデータとの類似度を算出することで、提案手法と比較を行うためのものである。このシミュレーションでは、取引時間において 5 分おきに観測した価格データを使用した。これにより、半日あたり前場ならば 24、後場ならば 30 の長さを持つ系列となり、各系列に対する類似度を計算するために十分なデータ量となる。1 か月分の 5 分おき価格データを学習データとしてモデルを作成し、次の月の各系列について、モデルとの類似度を計算した。

3.4 結果と考察

まず、月単位の類似度を計算した結果について、提案手法を用いて計算したものを図 5、GARCH モデルを用いて計算したものを図 6 に示す。なお、提案手法を用いた場合、各系列について類似度が算出されるので、1 か月分の全系列について算出された類似度の平均を示している。

これを見ると、提案手法を用いた場合と GARCH モデルを用いた場合の両方で、11 月まではほぼ同じ類似度を維持しているが 12 月に低い類似度が出現している。2006 年 12 月 21 日に、新日本製鐵によるウジミナス社の持分法適用会社化が発表されており、2006 年 12 月に新日本製鐵株に関する市場は、相転移を起こしているといえる。

以上の結果、提案手法と GARCH モデルを用いた手法のいずれも相転移を検出することに成功しており、両手法とも相転移を検出するために有効であることが示唆された。

次に、提案手法を用いて各系列について計算した類似度を図 7、GARCH モデルを用いて計算したセッション単位の類似度を図 8 に示す。

いずれの手法を用いた場合も、12 月における類似度の値が下がっていることが確認できる。ただし、GARCH モデルを

*1 「NEEDS (日経の総合経済データバンク)」を使用した。

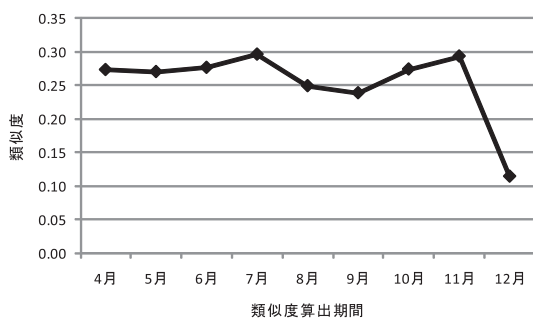


図 5 提案手法により算出した月単位類似度

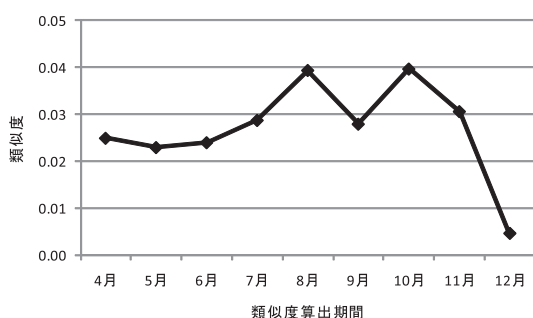


図 6 GARCH モデルにより算出した月単位類似度

用いた手法で算出した類似度は徐々に下がっているのに対し、提案手法で算出した類似度は 11 月 29 日を境に、低い値を示している。つまり、GARCH モデルを用るよりも、板情報を利用した提案手法の方が早い時期に市場の変化を捉えることができたといえる。

以上の結果から、相転移の影響は徐々に価格データに現れるが、板情報を分析対象とした提案手法を用いることで、より早く相転移を検出可能であるということが示唆された。

4. 結論

本研究では、金融市場の相転移を検出する手法を提案した。提案手法は、まず市場の類似度を定量的に算出する方法を確立し、これを利用して、類似度の時間変化から相転移の発生時期を推定するものである。また、類似度算出は板情報を用いることによって、市場の内部状態を反映することを目指した。

提案手法の有効性を評価する実験を行った。実験では、提案手法を現実の取引データに適用し、相転移を検出することができるか否かを確認した。その結果、提案手法は、相転移検出に有効であることが示唆された。また、価格データを用いた既存の手法と比較したところ、板情報を用いた本手法は、相転移の影響が価格データに現れる前に相転移を検出しており、価格データを用いる場合と比べて優位性があることが認められた。

今後の課題として、まず、特徴抽出法を改良が挙げられる。本論文では、注文の規模、および買い注文が売り注文かという点のみに注目して特徴抽出を行った。しかし、板情報からは、例えば注文のキャンセルや、成行注文であるか指値注文であるかなど、より多くの情報を抽出可能である。これらの情報も組み合わせることで、特徴抽出法を改良することができると考えられる。

また、本論文では、1 銘柄 10 か月分のデータを用いて提案

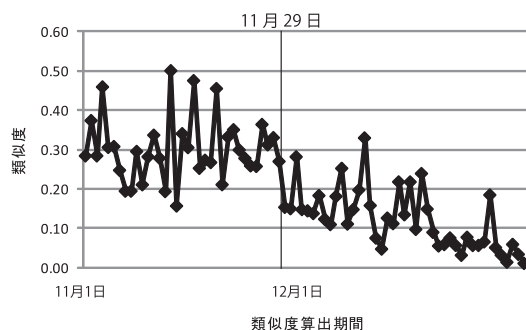


図 7 提案手法により算出したセッション単位類似度

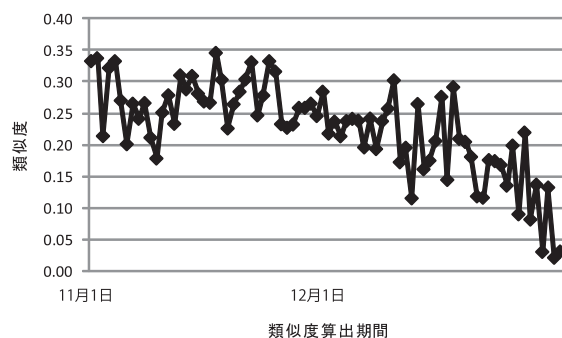


図 8 GARCH モデルにより算出したセッション単位類似度

手法の検証を行ったが、より多くの銘柄、長い期間のデータを用いて更に検証を追加する必要があると考えられる。

参考文献

- [Bollerslev 86] Bollerslev, T.: Generalized autoregressive conditional heteroskedasticity, *Journal of econometrics*, Vol. 31, No. 3, pp. 307-327 (1986)
- [Levinson 83] Levinson, S., Rabiner, L., and Sondhi, M.: An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition, *The Bell System technical journal*, Vol. 62, No. 4, pp. 1035-1074 (1983)
- [MacQueen 67] MacQueen, J., et al.: Some methods for classification and analysis of multivariate observations, in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, Vol. 1967, pp. 281-297 (1967)
- [Mantegna 00] Mantegna, R. and Stanley, H.: *An introduction to econophysics: correlations and complexity in finance*, Cambridge Univ Pr (2000)
- [Rabiner 89] Rabiner, L.: A tutorial on hidden Markov models and selected applications in speech recognition, *Proceedings of the IEEE*, Vol. 77, No. 2, pp. 257-286 (1989)
- [西岡 09] 西岡寛兼, 鳥海不二夫, 石井健一郎: 板情報を用いた市場変化の分析, 人工知能学会研究会資料 SIG-FIN-003, pp. 58-63 (2009)