平成26年度 博士論文

ランダム行列理論と複素主成分分析:
株式市場における相関構造への応用

新潟大学大学院自然科学研究科 博士後期課程
数理物質科学専攻 数理科学コース

F12J001B

新井 優太

Abstruct

近年、科学技術の発達に伴い膨大なデータの解析が可能となった。その一環として、多変量時系列データ中に潜む有意な相関構造を抽出する研究が盛んに行われている。その際によく用いられる手法の１つとして主成分分析（これによって得られる主成分ベクトルを経験的直交関数とも呼ぶ）がある。

主成分分析では、どこまでを統計的に有意な相関構造とみなすのかが非常に重要な問題である。そこで、統計的に有意な相関構造と統計ノイズとを区別するための理論的な裏付けのある判定基準としてランダム行列理論（Random Matrix Theory; RMT）が注目された。RMT では、ランダム変数の相関行列に対する固有値や固有ベクトル成分の分布について、理論的な表式が $Q \equiv T/N$（データ数:N，時系列長:T）を与えることで一意に定まる。特筆すべき事に、この RMT の導く固有値、固有ベクトル成分の分布がランダム変数の詳細によらず普遍的な性質を持つため、RMT から外れた固有値を有意な相関行列の現れと見なすことができる。それゆえ、RMT に基づく主成分分析を行うことで、実データの持つ本質的で統計的に有意な相関構造の抽出が可能となる。例えば、経済物理学の分野では株価に対して RMT に基づく主成分分析を行った研究が 1990 年に初めて報告されており、多くの興味深い相関構造が明らかになってきた。

しかし、従来の多変量時系列についての主成分分析では、相関行列を求める際に同時刻の変量間で相関をとっているため、タイムラグのある変量間での相関構造を抽出することは難しい。例えば、本研究で対象とする株価時系列の場合、２つの株式銘柄同士に正の相関があっても、それらの株価変動が同時に起きるとは限らない。むしろある一定時間の遅延が存在する動的（異時刻）相関構造が一般的には考えられる。動的相関構造を抽出することは、データ中に潜む相関構造を理解する上で必要なことは言うまでもない。そこで、まずは動的相関構造を抽出するために、ヒルベルト変換を用いた複素主成分分析による動的相関構造の抽出を行った。

次に、主成分分析（複素主成分も含む）では株式銘柄間の詳細な相関構造を理解することには秀でている一方で、株式市場に潜む相関構造を俯瞰的に捉えることは難しい。そこで、本研究では“位相情報を保持した株価相関ネットワーク”の構築・解析を行うことで、株式市場に潜む静的・動的相関構造を俯瞰的に捉える。このようなネットワークモデルは、これまで考案されておらず本研究が初のものとなる。

最後に、株式市場に潜む静的相関構造に基いて構築した株価相関ネットワークの研究から、株式市場にはフラストレーション状態にある多極構造の存在が示唆されている。このような状態は、これまで行われてきた多くの研究のように、株式銘柄間の２体相関を考えているだけでは決して現れることはない。より高次の相関を扱うことによって、初めて現れる構造である。本研究では、高次相関解析の第一歩として３体相関に注目して解析を行う。この高次相関を足がかりに、グループ相関とマーケットモードの関係性について明らかにされることを期待する。

目次

1 **序章** 4
- 1.1 先行研究 4
- 1.2 本論文の構成 8

2 **複素主成分分析** 10
- 2.1 導入 10
- 2.2 解析データ 10
- 2.3 複素主成分分析を行うための準備 10
- 2.4 RMT の拡張 12
- 2.5 拡張した RMT と CPCA を用いた動的相関構造の抽出 15
- 2.6 同期現象 21
- 2.7 リーマン・ショックの影響 23

3 **複素相関ネットワーク** 26
- 3.1 時系列解析からネットワーク解析へ 26
- 3.2 隣接行列 26
- 3.3 無向グラフ 27
- 3.4 位相でリンクに向きをつけた有向ネットワークの構築・解析-有効ネットワーク- . . 31
- 3.5 まとめ 34

4 **株式市場における多体相関** 38
- 4.1 株式市場における多体相関 38
- 4.2 2 体相関と 3 体相関 38
- 4.3 2 項分布を用いた 3 体相関の見積もり 39
- 4.4 まとめ 42

5 **まとめと今後の課題** 43

付録 A ランダム行列理論 45
- A.1 ガウス型直交アンサンブル 45
- A.2 Porter-Thomas 分布 47
- A.3 Wishart 型ランダム行列 50
- A.4 IPR と QOP の導出 56

付録 B 主成分分析 59
- B.1 主成分の導入 59

B.2 主成分の導出 60
B.3 理論的な説明の補足 61
B.4 主成分分析の結果 62
B.5 回転ランダムシャッフルを用いた帰無仮説 63

付録 C 株価相関ネットワーク 66
C.1 株価相関ネットワーク 66
C.2 コミュニティ抽出結果 66

付録 D ヒルベルト変換 68
D.1 連続ヒルベルト変換 68
D.2 フーリエ変換からヒルベルト変換を得る 70
D.3 離散ヒルベルト変換 71
D.4 東証 1 部上場企業に対する解析 72

付録 E 3 体相関 76
E.1 扱うデータ 76
E.2 結果 76

1　序章

1.1　先行研究

近年、科学技術の発達に伴い膨大なデータの解析が可能となり、分野の垣根を越えて学際的な研究が盛んに行われるようになった。その様な分野の１つとして「経済物理学」が挙げられる。経済物理学とはその名の通り、経済現象を統計物理学の考えや手法を活用して理解しようとする分野である。特に、株式市場は多くの銘柄間の相互作用や外生的・内生的ショックなど様々な影響を受ける複雑系の１つと考えられる。そのため、非常に興味深い研究対象の１つであるといえる。また、株式市場では利用可能なデータが豊富 (例えば、株価の日次時系列などは yahoo finance や Mathematica によって誰でも入手可能) であるため、実証的な研究を行うには格好の対象である。

株式市場を対象とした研究の１つとして、株式市場に潜む相関構造の抽出がある。その際によく用いられる手法の１つとして、主成分分析 (Principal Component Analysis; PCA, これによって得られる主成分ベクトルを経験的直交関数とも呼ぶ) がある。解析に用いられる株価時系列データは有限であるため、データには必ず統計ノイズが含まれる。そのため、主成分分析ではどこまでを統計的に有意な相関構造とみなすのかが非常に重要な問題となるが、これまではその判定基準として、半経験的なものが用いられてきた。例えば

1. 相関行列の固有値が 1 以上
2. 固有値の累積和が 80% になるまで

などである。半経験的という意味は、これらがまったく経験的に決められているわけではなく、ある程度の理論的裏付けがあるからである。判定基準 1 は、相関行列の固有値はその主成分がもつ情報量を意味しており、平均 0, 分散 1 に標準化された時系列の場合、その平均値は 1 となる。そのため、固有値が 1 以上というのは、平均以上の情報量を持った主成分である事を意味している。また、判定基準 2 は、元データが持つ情報をどの程度復元するかに相当する。主成分分析では、切り捨てた主成分の寄与の分だけ情報を失うため、80% 程度元データの情報を保持していれば、統計ノイズのみを除去し、有意な相関構造を抽出しているだろうという考えである。しかし、これらの判定基準は人によってまちまちであるため、同じデータでも解析する人によって得られる主成分の数が異なるという問題点がある。

そこで注目されたのがランダム行列理論 (Random Matrix Theory; RMT) である。RMT では，ランダム変数の相関行列に対する固有値や固有ベクトル成分の分布について、理論的な表式が $Q \equiv T/N$(データ数:N, 時系列長:T) を与える事で一意に定まる。特筆すべき事に、この RMT の導く固有値、固有ベクトル成分の分布がランダム変数の詳細に依らず普遍的な性質を持つため、RMT から外れた固有値を有意な相関構造の現れと見なす事ができる。それゆえ、RMT に基づく主成分分析を行うことで、実データの持つ本質的で統計的に有意な相関構造の抽出が可能となる。例えば、経済物理学の分野では株価に対して RMT に基づく主成分分析を行った研究が 1990 年に

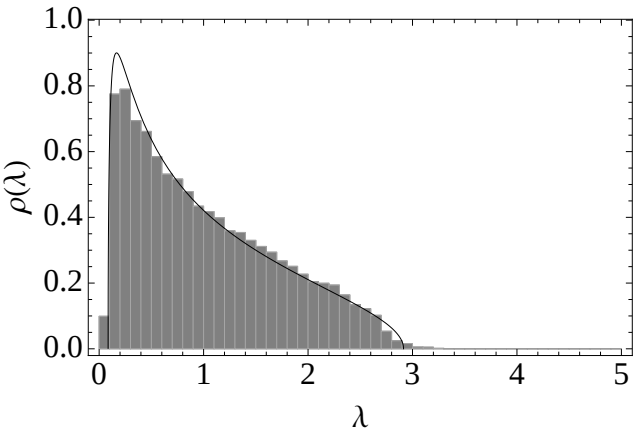

図 1.1　ランダム相関行列の固有値分布における有限サイズ効果

初めて報告されている。そこでは、S&P500 構成銘柄に対する相関行列の固有値分布とランダム行列理論とを比較をすることによって、市場に内在する相関構造について、マーケットモード (重心運動に対応) を含めていくつかの統計的に有意味な主成分がある事が実証されている [1, 2]。そしてこれらの論文を皮切りにして、多くの物理学者が株式市場における相関構造の解析に関する研究を行い、成果を上げてきた [3, 4, 5, 6, 7, 8]。

しかし、主成分分析の帰無仮説として RMT を使うことについても、いくつか問題点がある。それは、RMT から導かれる理論的な表式が (1) 行列サイズ無限大であり、(2) 行列要素が互いに独立同分布な確率変数に従う、という条件のもとで成り立っている事に起因する。

例えば前述したように、実際に解析に用いるデータのサイズは有限であるから、必ずそのことに起因する擬似的な相関が存在するはずである。図 1.1 は $N = 20, T = 40$ という小さなサイズのランダム時系列から成る相関行列を求め、それを対角化することによって得られた固有値の密度分布を表している。ただし、1000 サンプル生成した結果の平均を用いている。この図からわかるように、データサイズが小さい場合には RMT の固有値分布から外れた固有値が現れることが確認できる。そのため、サイズの小さなデータの帰無仮説として RMT を用いる場合には、この有限サイズ効果を評価する必要がある。そこでサイズの小さなランダム時系列を人工的に生成し、RMT より導出される分布からの固有値のズレを評価した。そしてサイズの小さなデータにも適用可能なよう、新たな主成分の判定基準を提案した [9, 10]。

また (2) については、時系列は一般に自己相関を持っている。時系列データを解析する上で、多くの場合には自己相関も非常に重要であるが、多変量の時系列データを扱う場合には自己相関よりもデータ間の相互相関に興味がある。しかし、RMT を帰無仮説とした場合では、自己相関によって生じる擬似的な相関と有意な相互相関を区別することができない。例えば、乱数の差分からなる時系列

$$x(t) = \epsilon(t+1) - \epsilon(t) \tag{1.1}$$

を考える。ただし、$\epsilon(t)$ は平均 0, 分散 1 に従う正規乱数とする。このようにして生成した時系列

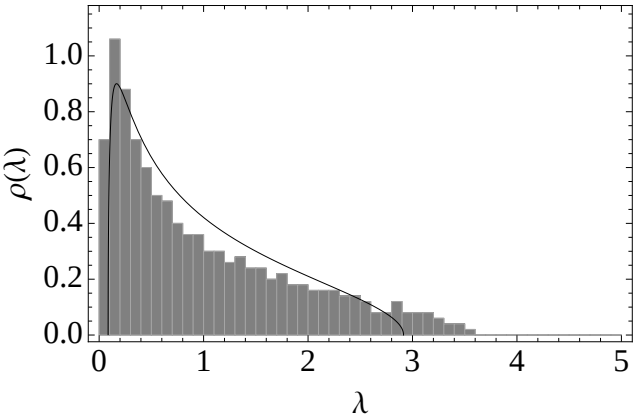

図 1.2　乱数の差分からなる時系列の相関行列と RMT から導かれる固有値分布の比較 ($N = 500, T = 1000$, サンプル数 1)

は、タイムラグを τ として自己相関係数 $\phi(\tau)$ を計算すると

$$\phi(\tau) = \begin{cases} -\frac{1}{2} & (\tau = 1) \\ 0 & (\tau \neq 1) \end{cases} \tag{1.2}$$

となる事がわかる。したがって、時系列 $x(t)$ は時系列間の相互相関は無く、自己相関のみをもつ多変量時系列となる。そして、この時系列を用いてランダム相関行列の固有値分布を求めると図 1.2 のようになる。図 1.2 からわかるように、相互相関がなくても自己相関の影響で RMT の理論曲線から外れた固有値が現れる。そのため、RMT を帰無仮説とした場合、得られた相関構造が自己相関によるものなのか、相互相関によるものなのか区別をすることが出来ない。そこで、私は AR(1) 過程

$$y(t+1) = \rho y(t) + \epsilon(t+1) \quad (0 \leq \rho < 1) \tag{1.3}$$

を用いることで、相互相関はないが自己相関のみを持つ多変量時系列データを人工的に生成し、自己相関が固有値分布に及ぼす影響を数値的に解析した [11, 12]。また、強い自己相関を持ったデータの１つとして知られる海表面気圧アノマリー月次データは、非常に特徴的な自己相関係数の振る舞いを持つことが知られている。具体的には、そのままの時系列に関して自己相関を計算すると AR(1) 過程的な振る舞いをするが、差分をとると乱数の差分の様な振る舞いをするのである (図 1.3)。そのため、その自己相関の振る舞いを再現する様な確率過程をモデル化できれば、海表面気圧アノマリー月次データに主成分分析を行う際のより良い帰無仮説へと応用できる。そこで、ノイズ付き AR(1) 過程という確率過程のモデルを考案し、海表面気圧アノマリー月次データの自己相関係数の振る舞いをモデル化した。そして、その自己相関効果の影響を取り入れた主成分判定基準の提案も行った [13]。また、自己相関と相互相関を区別するための手法として、他にも回転ランダムシャッフル (Rotational Random Shuffling; RRS)[14] も提案されている。以上の研究成果によって、従来のようにただ RMT から導かれる固有値分布を主成分の判定基準として用いるよりも、より高い信頼度を持って主成分の抽出が可能となった。本研究の主な対象として用いる株式市

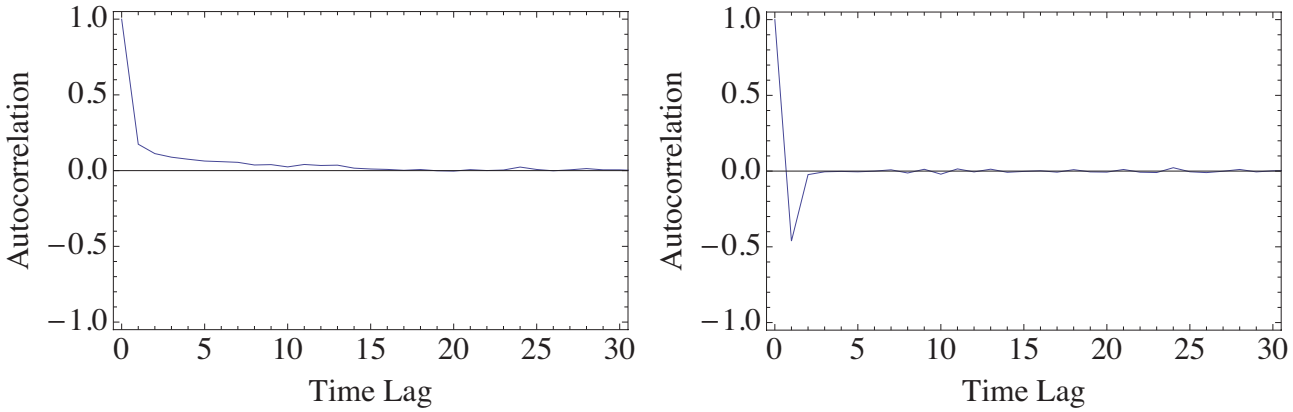

図 1.3　左図:海表面気圧月次アノマリーデータの自己相関，右図:海表面気圧月次アノマリーデータの差分に対する自己相関

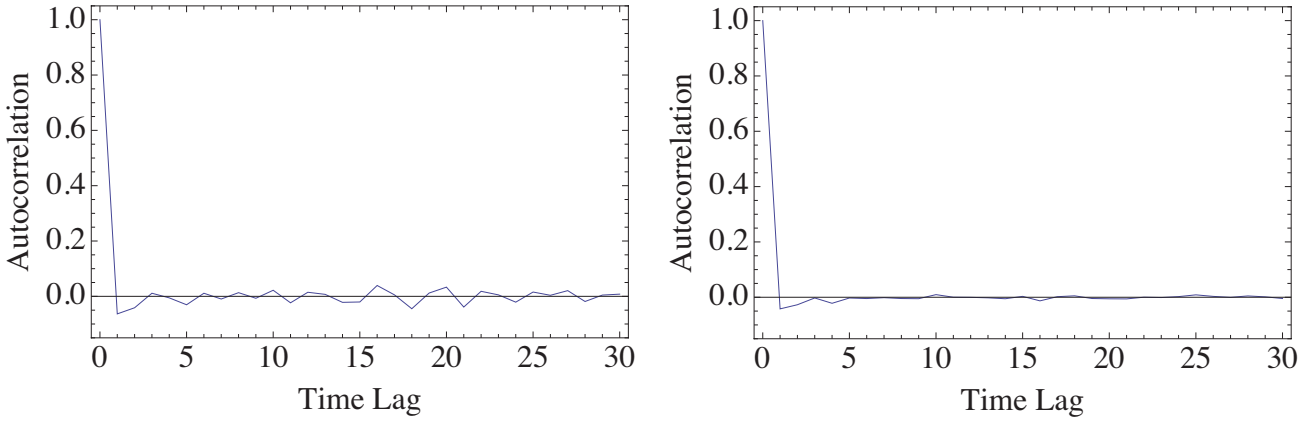

図 1.4　左図:東証 1 部上場企業 1996 年 ∼2006 年の自己相関、右図:S&P500 における 2008 年 ∼2011 年の自己相関

場の場合、対数収益率の差分における自己相関の振る舞いは図 1.4 のようになる。この図からわかるように、日次データについての株価変動時系列にはほとんど自己相関がない。しかし念の為に、本研究で主成分分析を行う場合には RMT と RRS を用いた帰無仮説によって、統計ノイズの除去も行う。これらの手法を用いて、S&P500 構成銘柄 (2008 年 ∼2011 年の全営業日で取引が行われた 483 銘柄) について解析を行った結果を付録 B.4 にまとめてある。

次に、主成分分析は多変量データ中に潜む相関関係の詳細を調べることには適している反面で、相関関係を俯瞰的に捉えることには向いていない。例えば、銘柄 A と銘柄 B の相関関係についての情報を得ることはできるが、全ての有意な相関構造を考慮した上で、市場の相関構造全体を俯瞰的に捉えることは難しい。「木を見て森を見ず」の状態である。そこで、相関構造を粗視化するために、しばしばネットワーク解析の手法が用いられる。この場合、株式銘柄をノード、銘柄間の相関係数をリンクと見なすことでネットワークを構築する。先行研究としては、デンドログラムやミニマムスパニングツリーを用いたものが挙げられる [15]。しかし我々は、RMT と主成分分析を組

み合わせる事で抽出した有意な相関構造から相関行列を再構築し、それを隣接行列と見なすことでネットワークを構築した (以下では、株価相関ネットワークと呼ぶ)。しかしここで注意しなければならないのは、相関係数が正負の値を取るため、リンクの重みとして正負の値をとることである。強い正の相関関係にある銘柄集団に興味があるのはもちろんであるが、負の相関関係にある銘柄集団にも興味がある。そこでこの問題を解決するために、フラストレーションと呼ばれる評価関数を用いた手法に注目した [16]。この評価関数が最小となるネットワークの分割を見つけ出す事で、正の相関関係にある銘柄集団 (コミュニティ) を抽出することが可能となる。この手法を株価相関ネットワークに適用した結果、日米の株式市場双方で 3 つないしは 4 つのコミュニティが抽出された [17, 18, 19, 20]。さらに、得られたコミュニティ間は負の相関関係にあることもわかった。このことから、株式市場にはフラストレーション状態にある多極構造の存在が示唆されている。主成分分析で扱ったのと同じ、S&P500 構成銘柄に対するネットワーク解析の結果を付録 C.1 にまとめてある。

1.2 本論文の構成

上述したように、先行研究として株式市場に潜む有意な相関構造の抽出 (PCA+RMT) と相関構造の粗視化 (ネットワーク解析) を行ってきた。しかしこれらの計算結果は、静的な相関構造、つまり同時刻相関のみを扱って議論している。例えば、株価 A と株価 B があったときに株価 A と B が同時に上がる (下がる) というのが同時刻相関であるが、株価 A が上がった後に株価 B が上がる (下がる) というのが異時刻相関である。一般的には後者のように、時間差のある相関構造の方が自然だと考えられる。そのため、株式市場に潜む相関構造をより理解するためには、同時刻相関のみならず異時刻相関についても抽出・解析する必要がある。

次に、株式市場に潜む相関構造を粗視化した結果得られたフラストレーション状態にある多極構造についてであるが、この構造は非常に不自然である。なぜならばこの構造は、3 つの異なるコミュニティから株式銘柄を 1 つずつ選んで取り出した場合、それら 3 つの銘柄は全て互いに反相関関係にあることを意味しているからである。そのような相関関係は、一見矛盾しているように思える。しかし、その構造の存在が複雑な市場の振る舞いを引き起こしているのではないかと考えている。その根拠としては、フラストレーション構造が現れる物理系の代表にスピングラスが挙げられるが、そこでは多種多様な磁性が発現する。そのため、このようなフラストレーション構造を明らかにすることは、株式市場に潜む相関構造を明らかにする上で非常に重要なのではないかと考えている。しかしながら、フラストレーション構造は静的に安定な状態となりえないため、こちらへの応用の視点からも動的相関の理解が不可欠である。

また、フラストレーション構造というのは、これまでの研究で扱ってきたように 2 銘柄間の相関 (2 体相関) を考えていては決して現れず、3 銘柄以上の高次相関を考えることで初めて現れる。そのため、フラストレーション構造解明に向けて動的相関構造とは別のアプローチとして、株式市場における高次相関の解析を行う。特に本研究では、高次相関解析のための第 1 歩として、3 体相関についての解析も行う。このような高次相関の解析は、フラストレーション構造解明のためのみな

らず、株式市場における相関構造解析の中心的役割を担っていくのではないかと期待している。

以上のように本研究の目的は、これまでの研究を踏まえた上で、実データの解析に基いて市場における動的な性質を明らかにし、フラストレーション構造の詳細に迫ることである。そのために、第 2 章ではヒルベルト変換を用いて時系列を複素数化することによって、市場に潜む動的相関構造の抽出を行う。これによって、これまでの主成分分析では抽出することが困難であったタイムラグのある相関構造の抽出が可能となり、新たな相関構造の発見が期待できる。本研究とは異なるアプローチではあるが、動的相関構造抽出の先行研究として [21] などもある。次に第 3 章では、得られた動的相関構造に基いて株価相関ネットワークを構築・解析を行う。特に本研究の新奇性は、構築されたネットワークにおける各リンクの重みが複素数であり、その大きさがリンクの強さ、位相がノード間のリード・ラグ関係を表す。株価の変動が同期関係にあるリンクに着目し、それらの銘柄集団をコミュニティとして抽出する。また、第 4 章では、フラストレーション構造への別のアプローチとして高次相関についての解析を行う。ここでは簡単のために、株価の変動幅は無視して上がるか下がるかの 2 値化し、 2 体相関効果と 3 体相関効果を 2 項分布を用いて統計的に評価した結果を報告する。最後に第 5 章で、本論文に関してのまとめと今後の課題について述べる。

2　複素主成分分析

2.1　導入

これまでの株価変動における相互相関構造の抽出に関する研究では、主に同時刻相関が扱われてきた。しかし、一般的に相互相関を扱う場合には、同時刻相関だけてなく異時刻相関も存在することが考えられる。そこでこの章では、株式市場に潜む異時刻（動的）相関構造の抽出を行う。

従来の主成分分析では、相関を求める際に同時刻の変量間で相関を計算するため、タイムラグのある相関を抽出する事が困難であった。そのため、従来の主成分分析を拡張した複素主成分分析 (Complex PCA; CPCA) とこれに応じて拡張した RMT を組み合わせることで、株式市場に潜む動的相関構造の抽出を行う。CPCA[22, 23] はもともと気象データを解析するために開発された手法であり、本研究ではそれを株式市場へ応用する。株価に対する CPCA の先行研究として、東証１部上場企業株価へ応用したものがある [24, 25, 26]。この結果の詳細については付録 D.4 にまとめておく。

2.2　解析データ

S&P500 インデックスに採用されている株式銘柄のうち、2008 年 ∼2011 年までの全営業日 (1009 日 $= T + 1$) で取引が行われた 483 銘柄を扱う。含まれる業種と銘柄の数を表 1 にまとめる。

2.3　複素主成分分析を行うための準備

従来の主成分分析では難しかった異時刻相関構造の抽出を、複素主成分分析を用いて行っていく。異なる銘柄間の相関を調べるために、銘柄 $\alpha(= 1, \cdots, N)$ の時刻 $t(= 1, \ldots, T)$ における株価を $s_\alpha(t)$ として、対数収益率 $r_\alpha(t)$

$$r_\alpha(t) = \ln s_\alpha(t+1) - \ln s_\alpha(t) \tag{2.1}$$

をまず求める。そして、$r_\alpha(t)$ のヒルベルト変換 $H[r_\alpha(t)]$ を用いて

$$\xi_\alpha(t) = r_\alpha(t) + iH[r_\alpha(t)] \tag{2.2}$$

と時系列の複素数化を行う。ここで i は虚数単位を表している。ヒルベルト変換の詳細に関しては付録付録 D を参照されたい。

ここで、異なった株価はそれぞれ平均や分散が大きく異るから、このまま $\xi_\alpha(t)$ を解析に用いるとそれらの影響を大きく受けてしまう。そこで、すべての銘柄について標準化

$$\Xi_\alpha(t) = \frac{\xi_\alpha(t) - \langle \xi_\alpha \rangle_t}{\sigma_\alpha} \tag{2.3}$$

IG id	GICS Industry Group	# Stocks
1	Energy	39
2	Materials	29
3	Capital Goods	40
4	Commercial & Professional Services	11
5	Transportation	9
6	Automobiles & Components	4
7	Consumer Durables & Apparel	14
8	Consumer Services	14
9	Media	15
10	Retailing	32
11	Food & Staples Retailing	9
12	Food Beverage & Tobacco	21
13	Household & Personal Products	6
14	Health Care Equipment & Services	31
15	Pharmaceuticals, Biotechnology & LifeSciences	19
16	Banks	15
17	Diversifted Financials	27
18	Insurance	21
19	Real Estate	16
20	Software & Services	28
21	Technology Hardware & Equipment	26
22	Semiconductors & Semiconductor Equipment	16
23	Telecommunication Services	8
24	Utilities	33

表 1　解析データに含まれる業種の一覧と、その業種に含まれる銘柄の数

を行う。式 (2.3) において、$\langle \cdots \rangle_t$ は時間平均、$\sigma_\alpha \equiv \sqrt{\langle (\xi_\alpha - \langle \xi_\alpha \rangle_t)^2 \rangle_t}$ は ξ_α の標準偏差を表している。ここで、$\Xi_\alpha(t)$ の振幅 $A_\alpha(t)$ と位相 $\phi_\alpha(t)$ は極座標表示を用いて

$$\Xi_\alpha(t) = A_\alpha(t) \exp(i\phi_\alpha(t)) \tag{2.4}$$

と定義される。そして標準化した複素時系列 $\Xi_\alpha(t)$ を用いて、複素相関行列 $\tilde{\mathbf{C}}$ は

$$[\tilde{\mathbf{C}}]_{\alpha\beta} = \frac{1}{T} \sum_{t=1}^{T} \Xi_\alpha(t) \Xi_\beta^*(t) \tag{2.5}$$

として計算できる。ただし、Ξ_α^* は Ξ_α の複素共役を表している。$\tilde{\mathbf{C}}$ はエルミート行列であり半正

定値行列なので、固有値はゼロ以上の実数になる。また、$\tilde{\mathbf{C}}$ は

$$\tilde{\mathbf{C}} = \frac{1}{2N}\sum_{\ell=1}^{N}\tilde{\lambda}_\ell \tilde{\mathbf{V}}_\ell \tilde{\mathbf{V}}_\ell^\dagger,\ \ \tilde{\mathbf{V}}_\ell^\dagger \cdot \tilde{\mathbf{V}}_\ell = 2N \tag{2.6}$$

とスペクトル分解できる。ここで、$\tilde{\lambda}_\ell$ と $\tilde{\mathbf{V}}_\ell$ はそれぞれ複素相関行列 $\tilde{\mathbf{C}}$ の ℓ 番目に大きな固有値と対応する固有ベクトルを表している。また、一般的には $\tilde{\lambda}_1 \geq \tilde{\lambda}_2 \geq \cdots \geq \tilde{\lambda}_N$ であるが、株価などの実データを解析する際には重複した固有値は殆ど現れない。

2.4 RMT の拡張

従来の PCA の帰無仮説として RMT を用いてきたように、CPCA の帰無仮説としても RMT を採用する。そのため次に、CPCA へ RMT を応用可能とするための拡張について説明する。PCA から CPCA への変更点としては主に

1. 時系列が複素数になった点
2. 複素数化された時系列の虚部が実部のヒルベルト変換に対応している点

の 2 つである。付録付録 A では、従来の PCA で帰無仮説として用いられている RMT の固有値分布と固有ベクトル成分分布の導出を、ガウス型直交アンサンブルについて行った。一方で、時系列が複素数の場合はガウス型ユニタリーアンサンブルとして知られている。それによると、時系列データが複素数であっても、行列要素の実部と虚部がそれぞれ独立であれば、固有値分布は式 (A.33) と一致し、固有ベクトル成分分布は 2 次元のガウス分布になることが知られている。そのため、ここで考えるべき問題は “虚部が実部のヒルベルト変換” であることの影響がどのように現れるかである。

上述した疑問を解決するために、まず複素数化した時系列データ $\Xi_\alpha(t)$ のフーリエ変換

$$\Xi_\alpha(\omega) = \sum_t \Xi_\alpha(t)\mathrm{e}^{i\omega t} \tag{2.7}$$

を考える。式 (2.7) において、フーリエ変換は直交変換であるから、$\Xi_\alpha(t)$ がガウス分布に従うのであれば、フーリエ係数 $\Xi_\alpha(\omega)$ もガウス分布に従う。その上、$\Xi_\alpha(t)$ は解析信号と呼ばれるものに対応していて、負の周波数領域でフーリエ係数 $\Xi_\alpha(\omega)$ はゼロになることが知られている。これらの事を考慮すると、ヒルベルト変換を用いて複素数化したランダム行列の相関行列は

$$\tilde{\mathbf{C}}_{\mathrm{rand}} \propto \frac{1}{T}[0|\Xi(\omega)][0|\Xi(\omega)]^\dagger \tag{2.8}$$

$$= \frac{1}{T}[\Xi(\omega)][\Xi(\omega)]^\dagger \tag{2.9}$$

と書き表すことができる。式 (2.8) において、[|] は $N \times T$ 行列を表していて、| の左と右はそれぞれ負の周波数領域と正の周波数領域を表している。そして、負の周波数領域はフーリエ係数がゼロであるから、式 (2.8) から式 (2.9) への変形で時系列の長さが半分になり、$[\Xi(\omega)]$ のサイズは

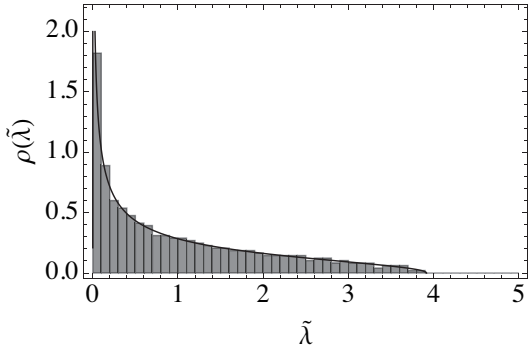

図 2.1　複素ランダム相関行列の固有値分布

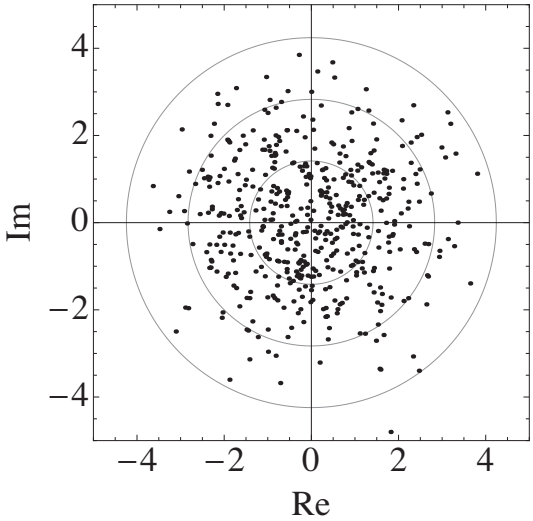

図 2.2　複素ランダム相関行列の固有ベクトル成分分布

$N \times T/2$ となっている。そのため、複素ランダム相関行列 $\tilde{\mathbf{C}}_{\mathrm{rand}}$ の固有値分布は式 (A.33) において $Q \to Q/2$ としたものに等しくなる。

次に、固有ベクトルの成分分布について述べる。すでに述べたように、複素ランダム相関行列 $\tilde{\mathbf{C}}_{\mathrm{rand}}$ は $N \times N$ エルミート行列であるから、ユニタリー行列によって対角化する事ができる。そのため、ガウス型直交アンサンブルでの議論をそのまま適用できるので、固有ベクトルの成分分布は２次元のガウス分布

$$\rho(u, v) = \frac{1}{2\pi} \exp\left(-\frac{u^2 + v^2}{2}\right) \tag{2.10}$$

になる。

例として、今回解析に用いるのと同じ $N = 483, T = 1008$ のランダムデータを用意し、固有値と固有ベクトル成分分布を計算する。図 2.1, 図 2.2 は複素ランダム相関行列 $\tilde{\mathbf{C}}_{\mathrm{rand}}$ の数値シミュレーションと理論曲線を比較したものである。図 2.1 は固有値分布を表していて、黒い実線が理論曲線でヒストグラムが数値計算結果をプロットしたものである。この図から、数値シミュレーションの結果と理論曲線がよく一致しているとわかる。RMT は行列サイズ無限大の極限で導出されているため、実データを扱う場合には有限サイズ効果の影響を考慮する必要があるが、この数値計算の結果から本研究で扱うデータサイズの場合は、有限サイズ効果の影響は無視できるといえる。一方、図 2.2 は最大固有値に対応する固有ベクトルの成分分布を複素平面上にプロットしている。１つの点が１つの成分に対応していて、実線で描かれた円は理論曲線の $\sigma, 2\sigma, 3\sigma$ を表している。ただし σ は、２次元ガウス分布の標準偏差を意味している。ここで、数値シミュレーションの結果が２次元ガウス分布とみなせるかどうかの定量的な評価を行う必要がある。そのためにまず、Inverse

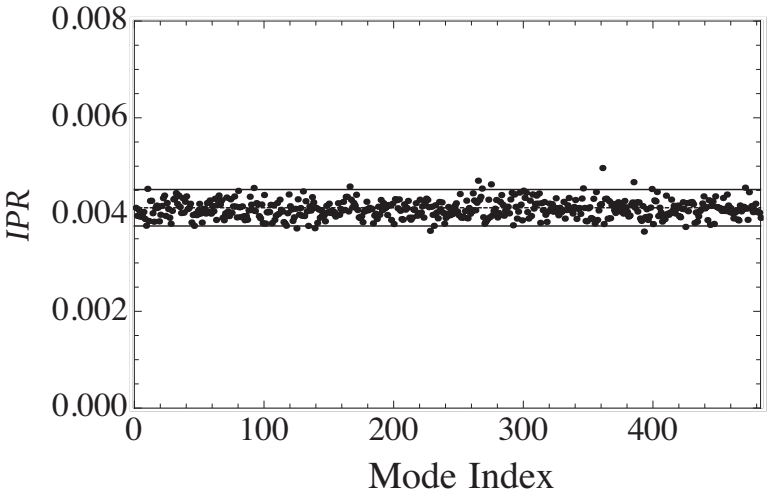

図 2.3　ランダムデータに対する IPR の数値シミュレーション結果

Participation Ratio (IPR) を導入する。ℓ 番目の固有モードに対応する IPR は

$$IPR_\ell = \sum_{\alpha=1}^{N} |\tilde{\mathbf{V}}_\ell^\alpha|^4 / \left(\sum_{\alpha=1}^{N} |\tilde{\mathbf{V}}_\ell^\alpha|^2 \right)^2 \tag{2.11}$$

として計算できる。IPR を計算することで、固有ベクトル成分分布について動径方向のガウス分布からのズレを評価できる。もしも $\tilde{\mathbf{V}}_\ell^\alpha$ の実部と虚部ががそれぞれ独立にガウス分布に従っていれば、行列サイズ $N \to \infty$ の極限で

$$E_{\mathrm{IPR}} = E[\ IPR_{\mathrm{RMT}}\ ] = \frac{2}{N}, \sigma_{\mathrm{IPR}} = \sqrt{VAR[\ IPR_{\mathrm{RMT}}\ ]} = \frac{2}{N^{3/2}} \tag{2.12}$$

となる。詳しい導出については付録 A.4 にまとめておく。式 (2.12) において、$E[\ \ ]$ と $VAR[\ \ ]$ はそれぞれ平均と分散を表している。図 2.3 はランダムデータに対する数値シミュレーションの結果と式 (2.12) の結果を比較したものである。図の横軸はモードインデックス、縦軸は IPR の値を表している。また、図中の破線は式 (2.12) の平均値、実線は平均値 $\pm 2\sigma$ を表している。この図から、ほとんどの点は実線の範囲内に収まっていることがわかる。次に、四極子オーダーパラメータ Q^2 を導入する。本研究では、ℓ 番目に大きな固有値に対応する固有ベクトルの四極子オーダーパラメータ Q_ℓ^2 を

$$Q_\ell^2 = Q_{\ell,xx}^2 + Q_{\ell,xy}^2 + Q_{\ell,yx}^2 + Q_{\ell,yy}^2, \tag{2.13}$$

$$Q_{\ell,xx} = \sum_{\alpha=1}^{N} \cos\left(\mathrm{Arg}\tilde{\mathbf{V}}_\ell^\alpha\right) \cos\left(\mathrm{Arg}\tilde{\mathbf{V}}_\ell^\alpha\right), \tag{2.14}$$

$$Q_{\ell,xy(yx)} = \sum_{\alpha=1}^{N} \cos\left(\mathrm{Arg}\tilde{\mathbf{V}}_\ell^\alpha\right) \sin\left(\mathrm{Arg}\tilde{\mathbf{V}}_\ell^\alpha\right), \tag{2.15}$$

$$Q_{\ell,yy} = \sum_{\alpha=1}^{N} \sin\left(\mathrm{Arg}\tilde{\mathbf{V}}_\ell^\alpha\right) \sin\left(\mathrm{Arg}\tilde{\mathbf{V}}_\ell^\alpha\right), \tag{2.16}$$

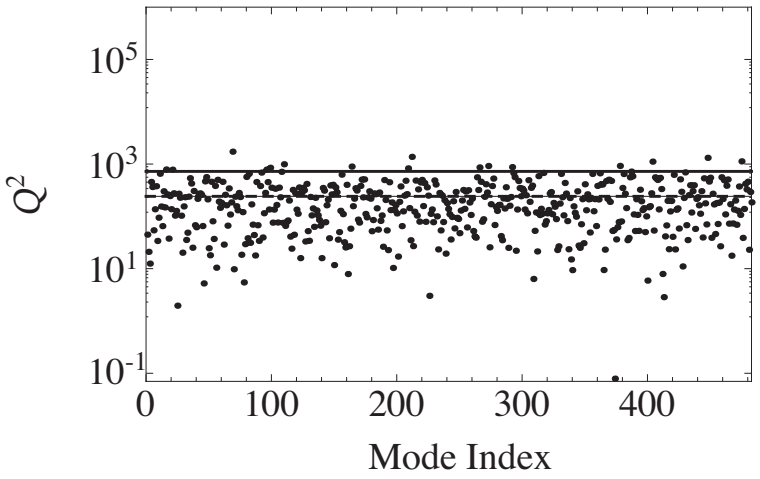

図 2.4　四極子オーダーパラメータの数値シミュレーション結果

として定義する。式 (2.16) において、$\mathrm{Arg}\tilde{\mathbf{V}}_\ell^\alpha$ は $\tilde{\mathbf{V}}_\ell$ の α 番目の成分の位相を表している。Q_ℓ^2 は回転不変であるから、固有ベクトル成分分布の異方性を評価することができる。固有ベクトル成分の分布が 2 次元のガウス分布に従う場合、四極子オーダーパラメータの平均値と標準偏差は

$$E_{Q^2} = E[\ Q^2_{\mathrm{RMT}}\ ] = \frac{N}{2}, \sigma_{Q^2} = \sqrt{VAR[\ Q^2_{\mathrm{RMT}}\ ]} = \frac{N}{2}. \tag{2.17}$$

となる。こちらも詳細は付録 A.4 にまとめておく。図 2.4 はランダムデータに対して計算した四極子オーダーパラメータと式 (2.17) を比較したものである。横軸がモードインデックス、縦軸は Q^2 の値をプロットしてある。また、図中の点線は式 (2.17) の平均値を、実線は平均値 $+2\sigma$ を表している。この図から、ランダムデータに対する四極子オーダーパラメータの結果はほとんど実線の範囲内に収まっていることがわかる。

2.5　拡張した RMT と CPCA を用いた動的相関構造の抽出

この節では、拡張した RMT と CPCA を用いることで S&P500 インデックスを構築する株式銘柄間に潜む動的相関構造の抽出を行う。図 2.5 は複素相関行列に対する固有値分布を表している。ヒストグラムが実データの固有値分布、実線が RMT の理論曲線、破線は回転ランダムシャッフルから得られた固有値分布である。RMT から得られる最大固有値よりも大きな固有値は全部で 7 つある。しかし、ランダム回転シャッフルの最大固有値 $\langle\tilde{\lambda}_{\mathrm{max}}\rangle + 2\sigma = 5.18 + 0.22$ を判定基準とすると、有意とみなせる固有値は 2 つ減って 5 つになる。ただし、$\langle\tilde{\lambda}_{\mathrm{max}}\rangle$ はランダム回転シャッフルしたデータの最大固有値の平均値、σ はその標準偏差を表している。ここで、サンプルの数は 1000 で計算を行った。

自己相関の影響を考慮すると、上述した 5 つの固有値が有意な相互相関構造の現れとみなせるが、この節では RMT で抽出された上位 7 つの固有値について注目して解析を行っていく。図 2.6 は上位 7 つの固有値に対応する固有ベクトル成分の分布を示したものである。固有ベクトルは位相

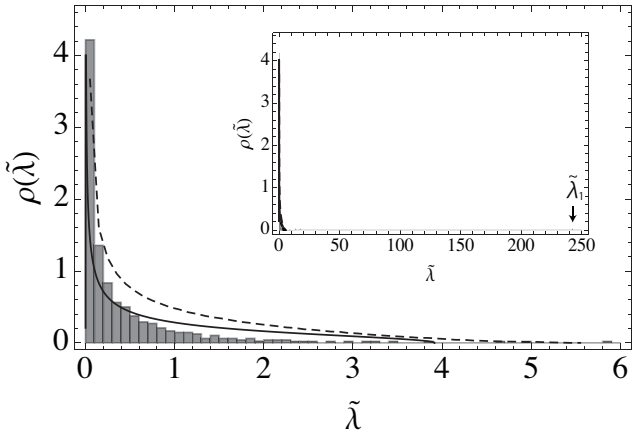

図 2.5　複素相関行列に対する固有値分布

の任意性があるため、分布の主軸が x 軸と重なるように、さらに重心が x 軸の正の方へ近づくように位相を決めてある。図 2.6(a) はほぼすべての成分が一点に集まっていることから、最大固有値に対応する固有ベクトルはマーケットモードだと言える。一方で、モード 2 からモード 4 とモード 6 では固有ベクトルの成分が細長く分布している。位相が π 異なる銘柄同士は互いに反相関関係にあることから、これらのモードでは同時刻相関が支配的であることがわかる。そして、モード 5 とモード 7 では複素平面上に広く分布している。このような分布には、静的相関だけでは説明できない動的な相関構造が潜んでいると考えられる。

図 2.7 は実データの解析から得られた固有ベクトル成分分布の IPR を計算した結果である。横軸がモードのインデックス、縦軸が IPR の値を表している。また、図中の破線と実線は図 2.3 と同様に、２次元のガウス分布の場合の平均と平均 $\pm 2\sigma$ を表している。この図からわかるように、大部分は実線の範囲内に収まっているが、モードインデックスが 1 ～ 100 までと 400 ～ 483 の領域では大きく外れていることがわかる。大きな固有値に対応するモードに注目してみると、モード 1～ モード 6 では実線の範囲内から大きく外れている事が見て取れる。このことから、これらのモードについては、動径方向に偏りのある分布だと言える。

次に、図 2.8 は各固有ベクトル成分分布の四極子オーダーパラメータを図示したものである。横軸がモードのインデックス、縦軸が四極子オーダーパラメータの値を表している。また、図中の破線と実線は図 2.4 と同様、固有ベクトルが２次元ガウス分布に従う場合の平均値と平均値 $+2\sigma$ を表している。多くのモードでは、四極子オーダーパラメータの値が実線の範囲内に収まっている。しかし、モード 1～ モード 6 くらいまでは大きく外れていることがわかる。したがって、角度方向に関しても、モード 1～ モード 6 では固有ベクトル成分の分布に偏りがあることがわかる。

次に、CPCA で抽出された相関構造と PCA で抽出された相関構造の間の関係性に興味がある。そこで、CPCA で得られた固有ベクトルを主軸 (x 軸) とそれに直交する軸 (y 軸) へ射影することを考える。そしてここではそれらを $\mathbf{V}'_\ell$ と $\mathbf{V}''_\ell$ と表すことにする。この表記を用いて、$\mathbf{V}'_\ell(\mathbf{V}''_\ell)$ と

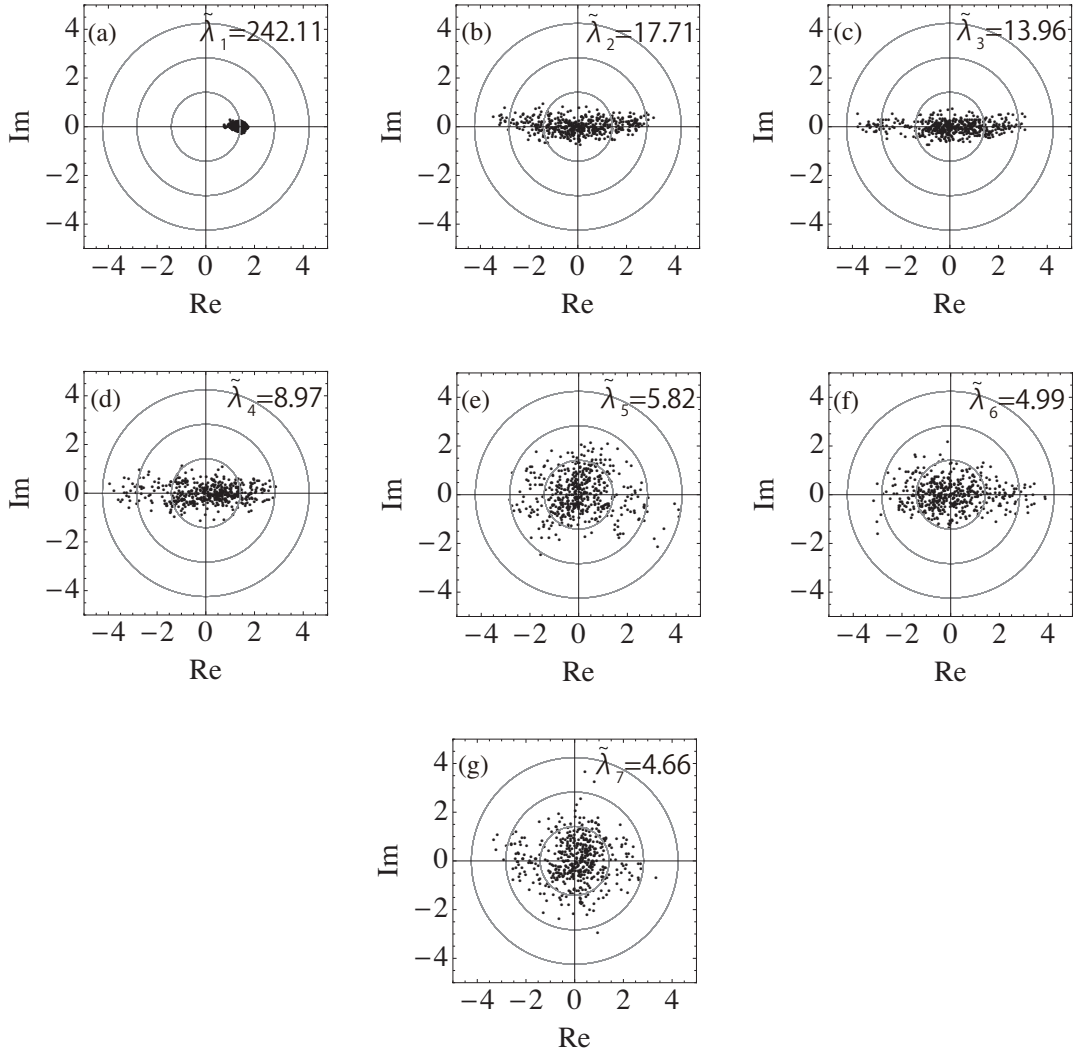

図 2.6　上位 7 つの固有値に対応する固有ベクトル成分の分布

$\mathbf{V}_\ell$ の間の相関を見積もるために、$\mathbf{V}'_\ell(\mathbf{V}''_\ell)$ を

$$\mathbf{V}'_\ell = \psi'_{\ell,1}\mathbf{V}_1 + \psi'_{\ell,2}\mathbf{V}_2 + \cdots + \psi'_{\ell,N}\mathbf{V}_N \tag{2.18}$$

$$\mathbf{V}''_\ell = \psi''_{\ell,1}\mathbf{V}_1 + \psi''_{\ell,2}\mathbf{V}_2 + \cdots + \psi''_{\ell,N}\mathbf{V}_N \tag{2.19}$$

と展開する。そして、展開係数 $\psi'_{\ell,m}(\psi"_{\ell,m})$ の 2 乗を計算する。 表 2 と表 3 は展開係数 $\psi'_{\ell,m}(\psi''_{\ell,m})$ の 2 乗の値をまとめたものである。表 2 より、モード 1 からモード 6 では $\mathbf{V}'_\ell$ と $\mathbf{V}_\ell$ がよく一致している事がわかる。一方で、$\mathbf{V}''_5$ と $\mathbf{V}_7$ も非常に高い相関関係にあることが表 3 よりわかる。このことから、CPCA では動的相関構造に加えて、静的な相関構造も抽出出来ていることがわかる。図 2.9(a) から (f) は $\mathbf{V}'_1$ から $\mathbf{V}'_6$ に対応していて、(g) は $\mathbf{V}''_5$ に対応している。これを PCA で抽出した固有ベクトルとくらべてみると、ほとんど一致していることがわかる。その上、これらの事実は CPCA で抽出できたモード 5 が PCA で抽出したモード 5 とモード 7 との共同現象として生

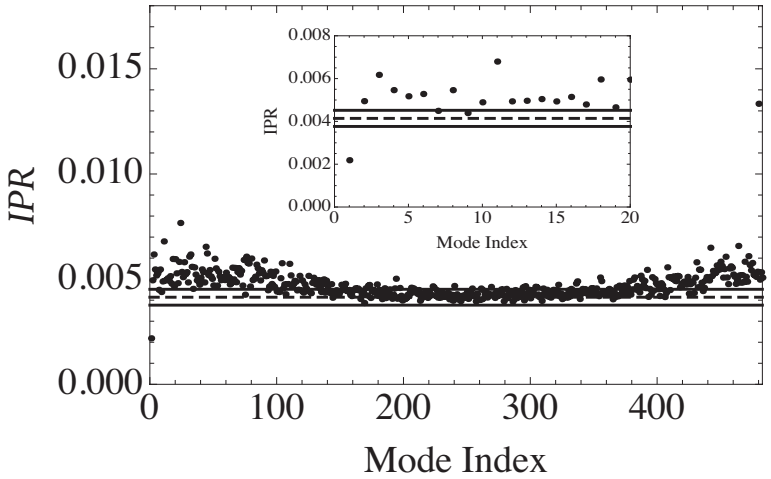

図 2.7　各固有モードの IPR

表 2　主軸上へ射影したベクトルの展開係数 ($\psi'_{\ell,m}$) の 2 乗。ℓ と m はそれぞれ CPCA と PCA のモードインデックスに対応している。

mode number	mode1	mode2	mode3	mode4	mode5	mode6	mode7	mode8	mode9	mode10
mode1	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
mode2	0.000	0.998	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
mode3	0.000	0.001	0.998	0.000	0.000	0.000	0.000	0.000	0.000	0.000
mode4	0.000	0.000	0.000	0.997	0.000	0.000	0.000	0.000	0.000	0.000
mode5	0.000	0.001	0.000	0.003	0.942	0.000	0.005	0.005	0.001	0.000
mode6	0.000	0.000	0.000	0.000	0.014	0.967	0.005	0.001	0.000	0.000
mode7	0.000	0.000	0.001	0.001	0.196	0.026	0.461	0.211	0.007	0.000
mode8	0.000	0.000	0.001	0.001	0.007	0.034	0.037	0.288	0.493	0.004
mode9	0.000	0.000	0.001	0.000	0.000	0.011	0.035	0.264	0.047	0.022
mode10	0.000	0.000	0.002	0.000	0.018	0.035	0.012	0.051	0.266	0.081

じていることを示唆している。

固有ベクトルに関する議論の最後に、$\tilde{\mathbf{V}}_5$ に強く寄与する業種を幾つか抽出したものを図 2.10 に示す。この図から、(a)Utilities と (c)Health Care Equipment & Services また (a) と (d)Pharmaceuticals は反相関関係にあることがわかる。そしてこの反相関関係は PCA によって得られた結果と一致している。そして、(a) と (b)Retailing の間には位相差 $\pi/2$ で、業種 (a) の方が業種 (b) よりも位相が先行しているような動的相関構造の存在が確認できる。

以上のように、CPCA と RMT を組み合わせることで銘柄間のリード・ラグの情報を位相差という形で入手出来た。すると当然、位相と実時間との関係性に興味が湧いてくる。例えば、Utilities と Retailing の間に $\pi/2$ の位相差があるのは分かったが、その位相差は実時間とどのように対応付けられるのであろうか。実は、本研究手法では相関行列を求める際に位相に関して平均化しているため、位相差と実時間とを正確に対応付ける事はできない。しかし、モードシグナルのパワースペクトルを調べることで、ある程度の目安をつけることはできる。

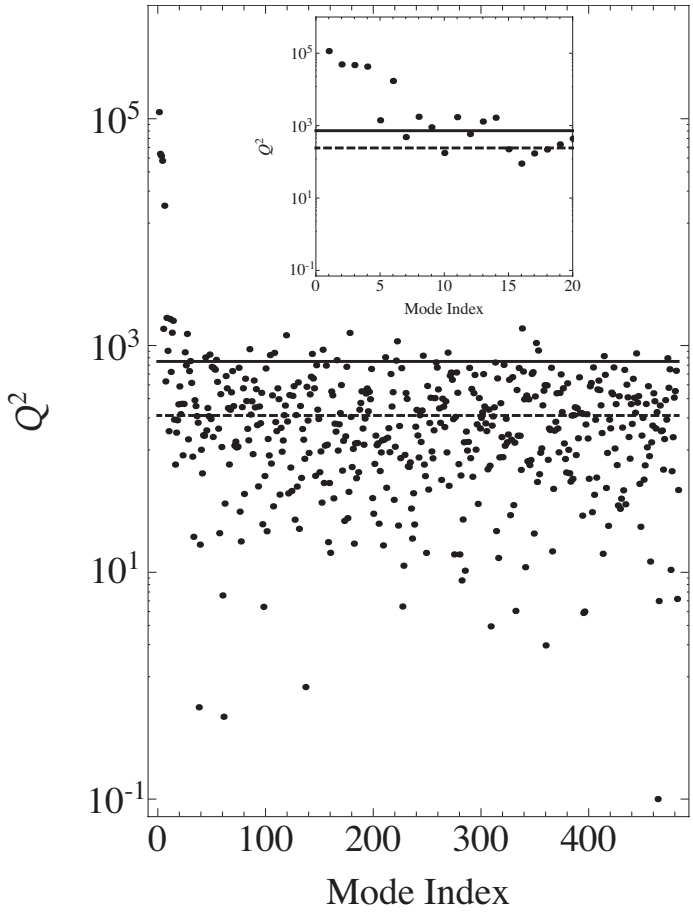

図 2.8　各固有モードにおける四極子オーダーパラメータ

モードシグナルとは複素数化した時系列を、$\tilde{\mathbf{V}}_\ell$ を基底として展開した時の係数であり、

$$\Xi(t) = \sum_{\ell=1}^{N} a_\ell(t)\tilde{\mathbf{V}}_\ell \tag{2.20}$$

において、$a_\ell(t)$ として得られる。このモードシグナル $a_\ell(t)$ は ℓ 番目の主成分として得られた相関構造が、どのような時間的振る舞いをするのかを表す。そのため、モードシグナルをフーリエ変換することで各モードを特徴付ける周期を得られれば、そこから実時間との対応を付けられる。今、$\Xi_\alpha(t)$ と $\tilde{\mathbf{V}}_\ell^\alpha$ が複素数であるから、モードシグナル $a_\ell(t)$ も当然複素数である。したがって、絶対値と位相についての時間変化を調べることができる。例として、モード 1 のモードシグナルについて、絶対値と位相の時間変化の図を示しておく (図 2.11)。

モードシグナルのパワースペクトルを考える際に、実時系列固有のゆらぎを取り除くため

$$\Theta_\ell(t) = \frac{a_\ell(t)}{\sum_{j=1}^{N} |\Xi_j(t)|^2} \tag{2.21}$$

と規格化したモードシグナルを用いる。図 2.12 は上位 7 固有値に対応するモードシグナルのパ

表 3　主軸上へ射影したベクトルの展開係数 ($\psi''_{\ell,m}$) の 2 乗。ℓ と m は表 3 と同様に、それぞれ CPCA と PCA のモードインデックスに対応している。

mode number	mode1	mode2	mode3	mode4	mode5	mode6	mode7	mode8	mode9	mode10
mode1	0.000	0.064	0.123	0.014	0.289	0.084	0.003	0.102	0.041	0.003
mode2	0.005	0.000	0.018	0.260	0.004	0.008	0.008	0.005	0.000	0.032
mode3	0.015	0.022	0.000	0.002	0.120	0.027	0.048	0.000	0.005	0.059
mode4	0.001	0.148	0.001	0.000	0.019	0.047	0.043	0.000	0.051	0.000
mode5	0.002	0.000	0.008	0.003	0.003	0.011	0.786	0.025	0.008	0.003
mode6	0.001	0.001	0.002	0.023	0.006	0.005	0.121	0.227	0.048	0.093
mode7	0.000	0.000	0.000	0.002	0.251	0.001	0.386	0.115	0.000	0.047
mode8	0.000	0.000	0.000	0.001	0.000	0.001	0.089	0.287	0.166	0.025
mode9	0.000	0.002	0.000	0.000	0.021	0.004	0.031	0.182	0.321	0.001
mode10	0.001	0.000	0.001	0.000	0.000	0.012	0.019	0.016	0.000	0.254

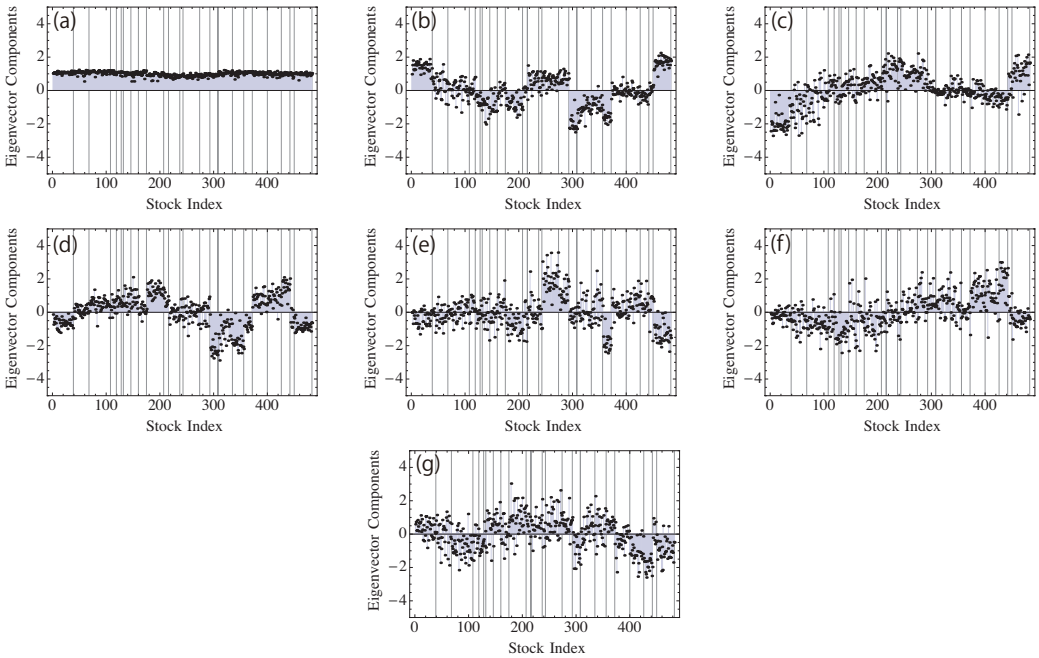

図 2.9　CPCA で得られた固有ベクトルの射影

ワースペクトルを表している。例えば、モード 1 では周期 300 日くらいのところに 1 つの大きなピークがある。そのため、モード 1 における 1 周期 (2π) が 300 日に対応していると考えれば、$\pi/2$ は 75 日に対応する。しかし、モード 1 はマーケットモードに対応し、銘柄間のリード・ラグはほとんどないため実時間との対応にあまり興味が無い。一方で、モード 5 では複数の特徴的なピークが現れているため、代表的な周期を見つけることが難しく、実時間との対応も困難である。

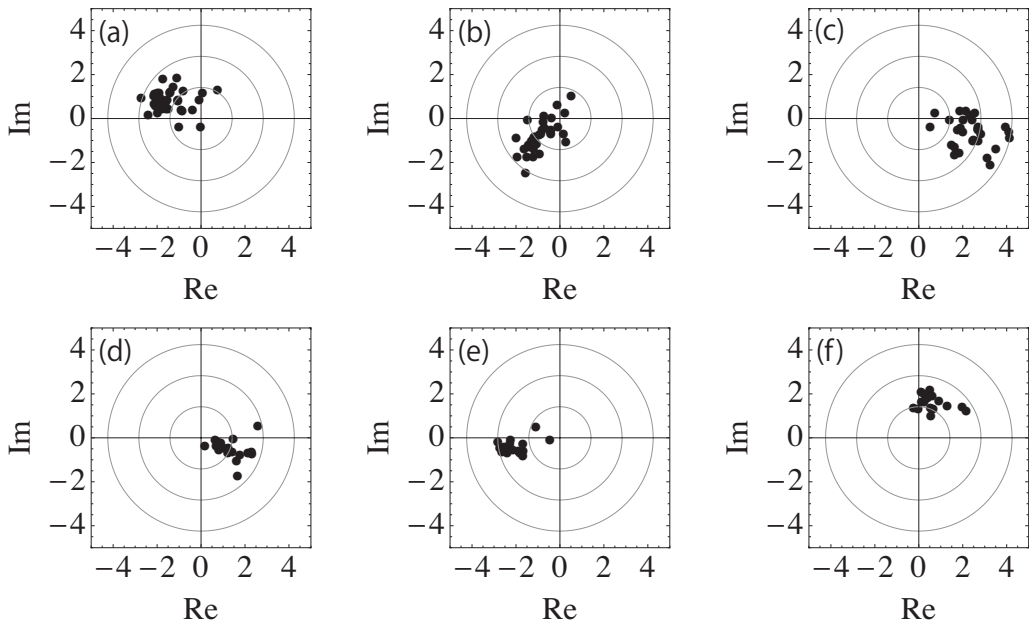

図 2.10　(a)∼(f) はそれぞれ Utilities, Retailing, Health Care Equipment & Services, Pharmaceuticals, Real Estate, Semi-conductors & Semiconductor Equipment に対応している

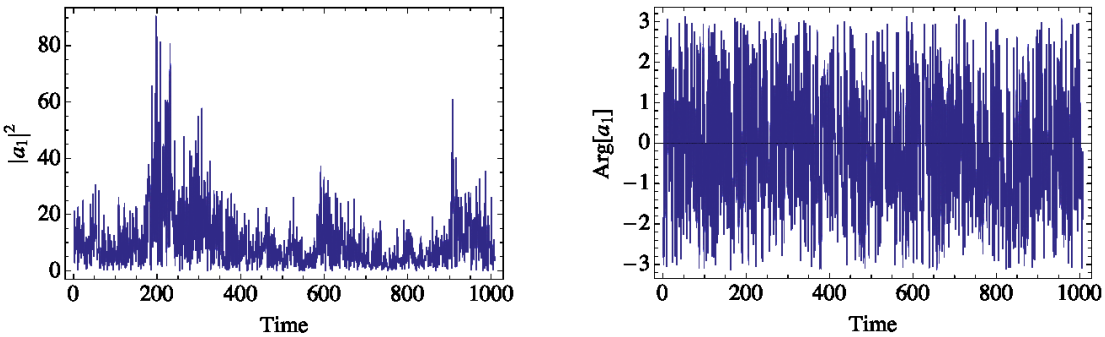

図 2.11　モード 1 におけるモードシグナルの絶対値 (左図) と位相 (右図)

2.6　同期現象

次に、モード間の同期現象について考えてみる。前節でモードシグナルには大きさと位相の情報がある事を述べたが、位相の情報を用いることでモード間の同期・非同期について考えることができる。一般的に、モード間の相関は

$$\langle a_m(t) a_n(t)^* \rangle = \delta_{n,m} \tag{2.22}$$

より、異なるモードシグナル同士は無相関となる。しかし、これはあくまで時間平均がゼロになるだけであって、２つのモード間が全く無関係とはいえない。そこで、モード間のモードシグナルの

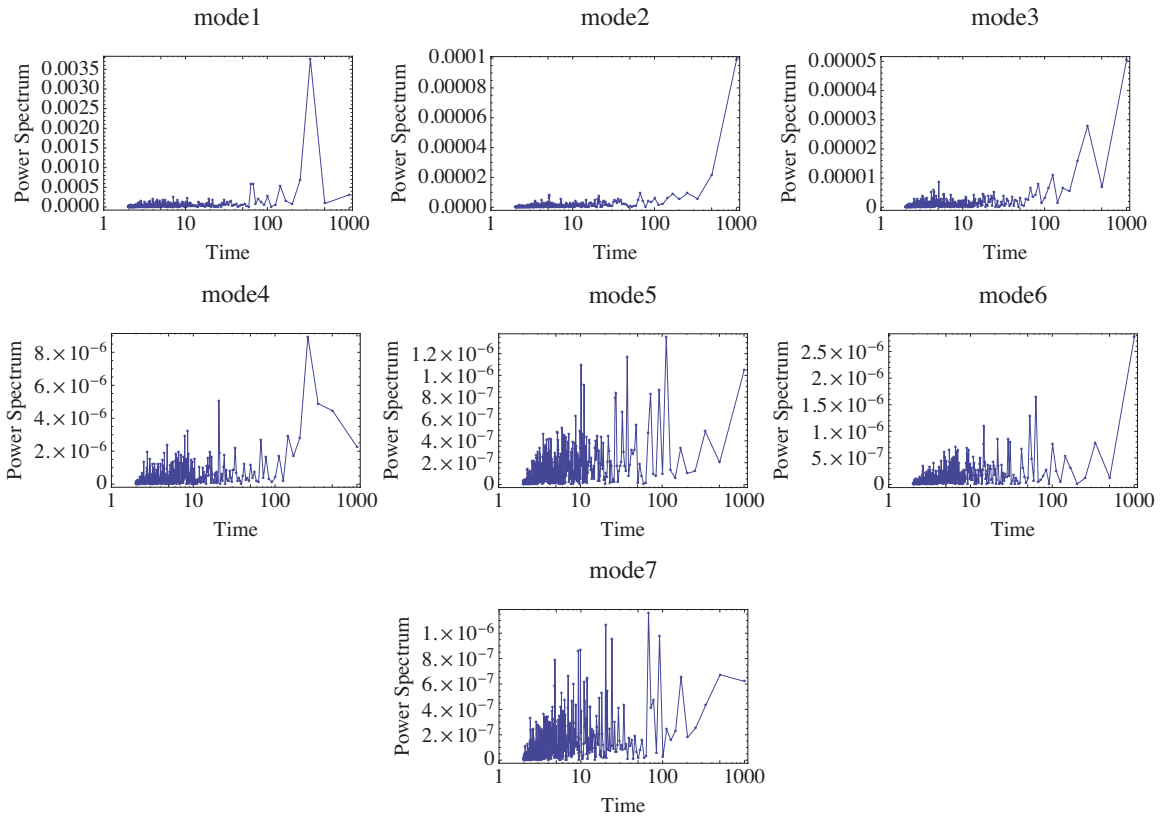

図 2.12　上位 7 固有値に対応するモードシグナルのパワースペクトル

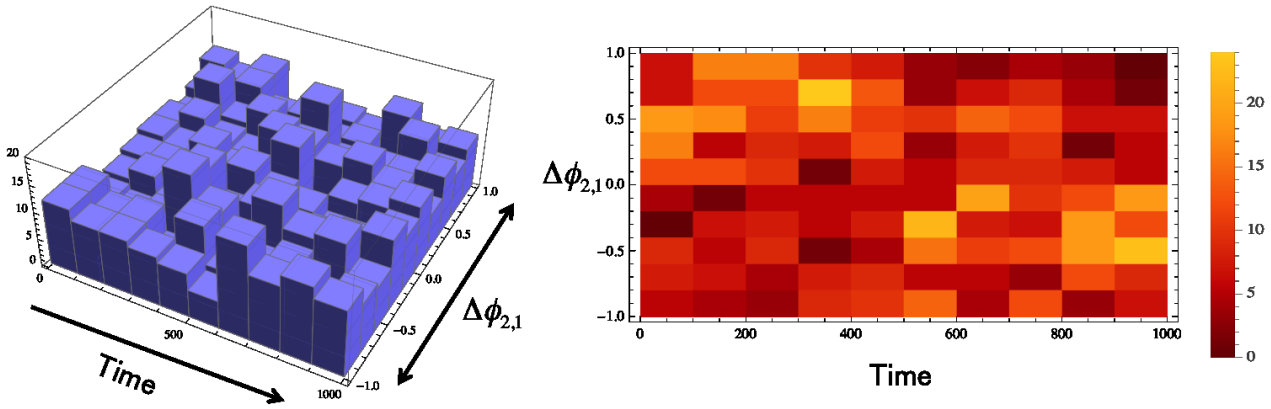

図 2.13　左図:位相差のヒストグラム，右図:ヒストグラムの密度プロット

位相差に注目することで、位相差に偏りがないか調べる。今、モード間の位相差を

$$\Delta\phi_{\ell,m}(t) \equiv \mathrm{Arg}[a_\ell(t)/a_m(t)] \tag{2.23}$$

と定義する。図 2.13 はモード 1 とモード 2 の位相差を示したものであるが、位相差に偏りがあるように見える。もしもこの偏りが統計的に有意なものであれば、モード 1 とモード 2 はある時間に同期しているといえる。そこで、この分布の偏りが統計的に有意であるかどうかを判定するため

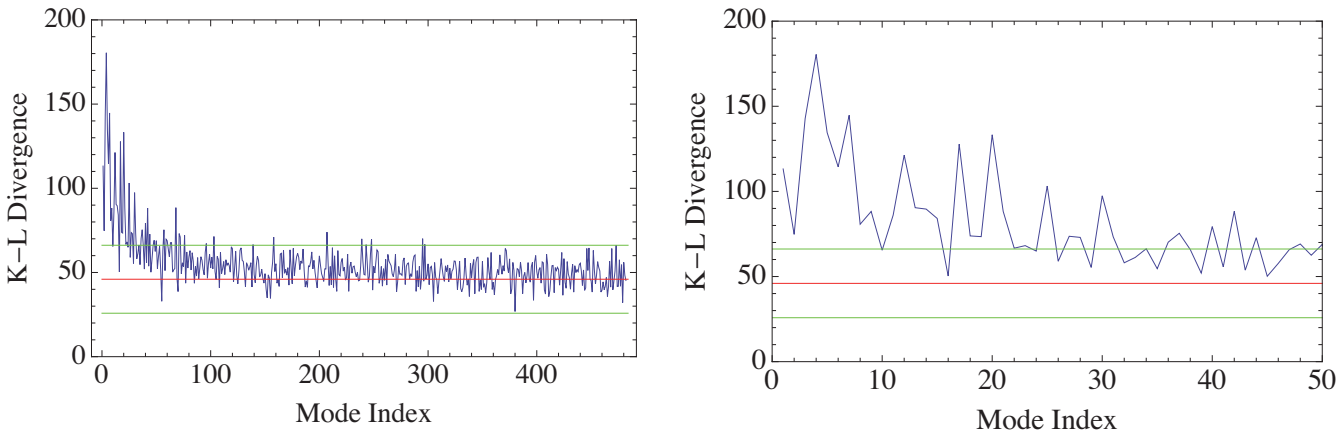

図 2.14　モード 1 との K-L Divergence。左図:全モード，右図:モード 50 までを拡大

に、Kullback-Librer Divergence を導入する。

分布 $P(\Delta\phi, t)$ と $Q(\Delta\phi, t)$ の K-L Divergence は

$$D_{\mathrm{KL}}(P, Q) = \sum P(\Delta\phi, t) \log \frac{P(\Delta\phi, t)}{Q(\Delta\phi, t)} \tag{2.24}$$

として定義されている。今、P を数値計算から得られた分布、Q を真の分布として、分布間の距離を計算する。もしも位相差に偏りがない場合は一様分布になるはずであるから、Q として一様分布を採用する。そのようにして K-L Divergence を計算した結果を、図 2.14 に示す。図中の赤線は、一様分布を数値的に生成した結果得られた K-L Divergence の平均値、緑線は数値計算の結果得られた標準偏差を σ として平均値 $\pm 3\sigma$ を表している。多くのモードでは、K-L Divergence が緑線の範囲内に収まっているが、モード 100 程度までは外れた値が見られる。そこで、モード 2 からモード 50 までを拡大した結果が図 2.14 の右図である。この結果から、今回 RMT で有意とみなされたモード 2 からモード 7 では K-L Divergence が一様分布から大きく外れている事がわかる。したがって、モード 1 となんらかの同期現象があるのではないかと考えられる。

このような同期現象はフラストレーション構造 (詳しくは第 3 章で説明) との関連性からも興味がある。フラストレーション構造は静的には不安定であるため、常に存在しているとは考えにくい。そのため、マーケットモードが弱くなる時に、フラストレーション構造が現れるなど、なんらかの関係があるのではないかと考えている。

2.7　リーマン・ショックの影響

最後に、本研究で用いたデータにはリーマン・ショックが含まれている。そのため、得られた全ての相関構造がリーマン・ショックに影響されたものなのではないかという懸念がある。そこで、モードシグナルの絶対値の 2 乗の累積和である

$$I_\ell(\tau) = \sum_{t=1}^{\tau} |a_\ell(t)|^2 \tag{2.25}$$

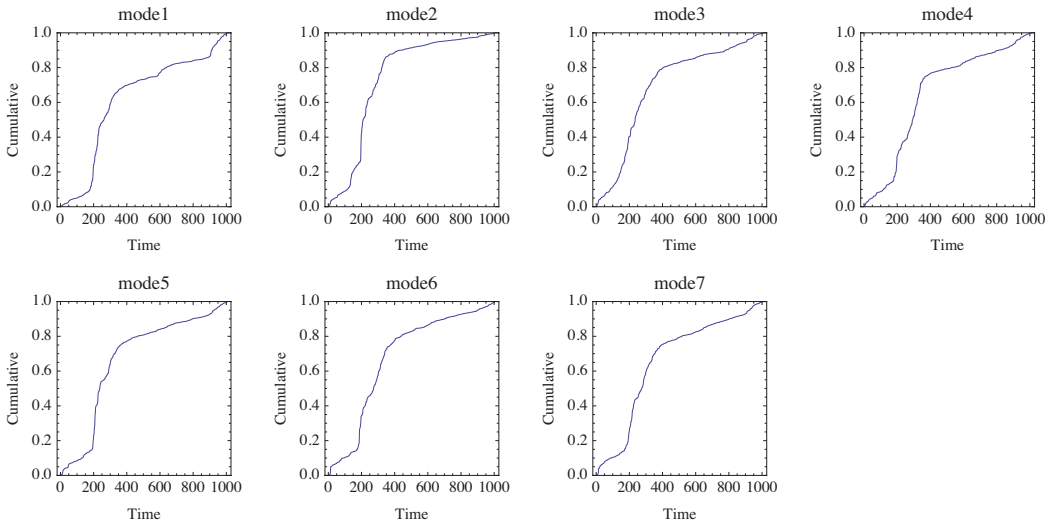

図 2.15　モードシグナルの絶対値の 2 乗の累積和

を考える。モードシグナルの値が大きくなる時点で、対応する相関構造が強く現れる。そのため、全ての相関構造がリーマン・ショックの影響によるものであれば、その様子が $I_\ell(\tau)$ のグラフから読み取れるはずである。図 2.15 を見るとわかるように、リーマン・ショック付近で大きく累積和の値が変化しているように見える。つまり、これらはすべてリーマン・ショックの影響なのであろうか。しかし、モードシグナルは元の時系列のゆらぎの影響を受けるため、その影響をなくすために元の時系列の大きさで割って規格化した式 (2.21) を用いるべきである。規格化した累積和を

$$I'_\ell(\tau) = \sum_{t=1}^{\tau} \Theta_\ell(t) \tag{2.26}$$

とすると 図 2.16 を得る。この図から、どの固有モードもリーマン・ショック時だけでなく常に存在している事がわかる。そのため、得られた相関構造全てがリーマン・ショックの影響によって出現しているわけではないことがわかる。

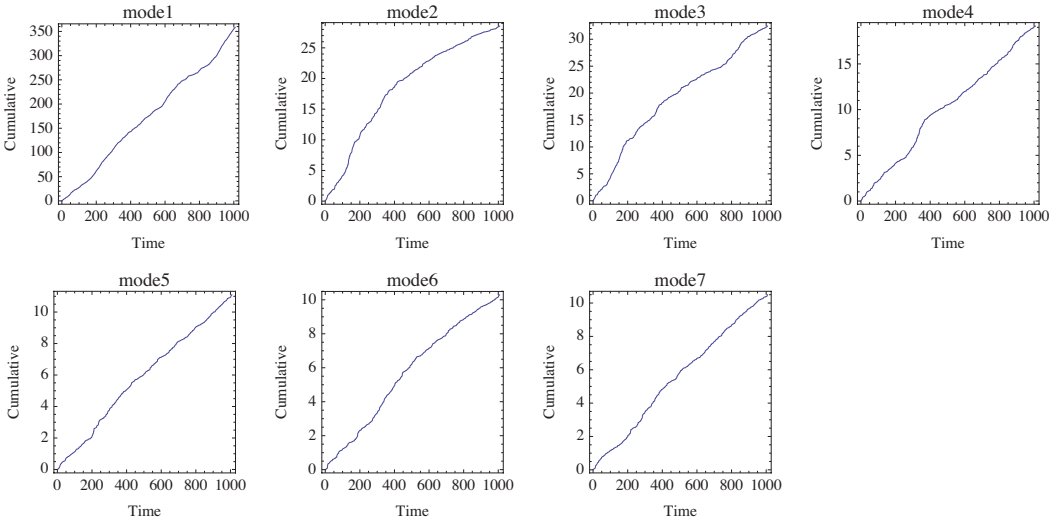

図 2.16　規格化したモードシグナルの絶対値の 2 乗の累積和

3　複素相関ネットワーク

3.1　時系列解析からネットワーク解析へ

主成分分析を用いた株式市場に潜む相関構造の抽出は、銘柄間の相関構造を詳細に把握できる反面で、市場全体の相関構造を俯瞰的に捉えることは難しい。そこで、銘柄間の相関関係をネットワークとみなした解析がこれまで行われてきた。先行研究として我々は、株価相関ネットワークに対してコミュニティ抽出を行うことで株式銘柄間の相関関係を解明する研究を継続して行ってきた [17, 18, 19, 20]。RMT によって、統計ノイズを除去した株価相関行列を隣接行列としてネットワークを構築する。このようにして構築されたネットワークは株式銘柄がノード、相関係数がリンクに対応しており，リンクは正負の重みを持つ。コミュニティ抽出を行うと、ネットワークは数個のコミュニティに分れ，それらが互いに反相関関係にあるような “フラストレーション状態にある多極構造” をしていることがわかった。ここで、コミュニティとは正の重みを持つリンクで結ばれるノード集団を意味する。

しかしこれまで行われてきた株価相関ネットワークの解析では、静的 (同時刻) 相関しか考慮していない。第 2 章では、抽出した株式市場に潜む動的相関を株式市場へ複素主成分分析を用いることで動的相関構造を位相差という形で抽出を行った。本章では、得られた株価間の動的相関構造を反映させた位相ネットワークを構築し、解析を行った結果について報告する。特に、ここでの本研究の新奇性は、このようにして構築したネットワークではリンクの重みが複素数になるため、大きさがリンクの強さと、位相がノード間のリード・ラグ関係を表している点にある。そのため、株価変動時系列間のダイナミクスを反映したネットワーク構造になっていると期待できる。

3.2　隣接行列

この節ではまず、ネットワークを構築するための基礎となる隣接行列の構築方法について説明する。式 (2.6) にある複素相関行列のスペクトル分解を

$$
\begin{aligned}
\tilde{\mathbf{C}} &= \frac{1}{2N}\sum_{\ell=1}^{N}\tilde{\lambda}_\ell\tilde{\mathbf{V}}_\ell\tilde{\mathbf{V}}_\ell^\dagger \\
&= \frac{1}{2N}\tilde{\lambda}_1\tilde{\mathbf{V}}_1\tilde{\mathbf{V}}_1^\dagger + \frac{1}{2N}\sum_{\tilde{\lambda}_+<\tilde{\lambda}}\tilde{\lambda}_\ell\tilde{\mathbf{V}}_\ell\tilde{\mathbf{V}}_\ell^\dagger + \frac{1}{2N}\sum_{\tilde{\lambda}<\tilde{\lambda}_+}\tilde{\lambda}_\ell\tilde{\mathbf{V}}_\ell\tilde{\mathbf{V}}_\ell^\dagger \qquad (3.1)\\
&= \tilde{\mathbf{C}}_{\mathrm{market}} + \tilde{\mathbf{C}}_{\mathrm{group}} + \tilde{\mathbf{C}}_{\mathrm{rand}} \qquad (3.2)
\end{aligned}
$$

と表すことにする。式 (3.2) において $\tilde{\mathbf{C}}_{\mathrm{market}}, \tilde{\mathbf{C}}_{\mathrm{group}}, \tilde{\mathbf{C}}_{\mathrm{rand}}$ はそれぞれ相関行列の、マーケットモード部分、有意相関部分、ランダム部分を表しており、$\tilde{\mathbf{C}}_{\mathrm{market}}, \tilde{\mathbf{C}}_{\mathrm{group}}$ が RMT から統計的に有意とみなされた相関構造部分である。本研究では、マーケットモードを除いた有意な相関構造部分に興味があるので、$\tilde{\mathbf{C}}_{\mathrm{group}}$ を隣接行列とみなしてネットワークを構築する。以下では、$\tilde{\mathbf{C}}_{\mathrm{group}}$ を改めて $\mathbf{A}$ と表す。

$\mathbf{A}$ は複素相関行列のうちの有意な相関構造部分であるから、$|[\mathbf{A}]_{\alpha\beta}|$ と $\Delta\phi_{\alpha\beta} = \mathrm{Arg}[\mathbf{A}]_{\alpha\beta}$ は銘柄 α と銘柄 β の相関係数の絶対値と位相差 (リード・ラグ関係) をそれぞれ意味している。ここで、$\Delta\phi_{\alpha\beta}$ は π で割って規格化し $-1 \leq \Delta\phi_{\alpha\beta} \leq 1$ の値を取るものとする。隣接行列 $\mathbf{A}$ をそのまま用いてネットワークを構築した場合、完全グラフ (自分以外の全ノードと繋がったネットワーク) となってしまう。そこで、位相差に閾値を設けることで

$$\Delta\phi_{\alpha\beta} \leq |\theta_{\mathrm{th}}| \tag{3.3}$$

を満たす銘柄間、つまり株価変動が同期した銘柄間にリンクを張り、その重みとして $|[\mathbf{A}]_{\alpha\beta}|$ を用いたネットワークを構築する。このようにして構築したネットワークに対してコミュニティ抽出を行うと、価格変動が同期しており、相関が強い株式銘柄の集団がコミュニティとして抽出できる。その際に、良いコミュニティ分割かどうかを表す指標としてモジュラリティを用いる [27]。本来であれば、全ての分割パターンを調べあげて最もモジュラリティが高い値になる分割を採用するべきである。しかし、ノード数とリンク数が多くなればなるほど、全分割を試すには膨大な時間が必要となる。そこで本研究では、Fast-Unfolding 法 [28] という確率的な手法を用いることでモジュラリティが高くなるコミュニティ分割を求める。

3.3 無向グラフ

3.3.1 閾値 θ_{th} とコミュニティ数の変化

表 4　各 θ_{th} におけるモジュラリティとコミュニティに所属するノードの数

θ_{th}	0.1	0.075	0.05	0.025	0.01
modurarity	0.60	0.62	0.63	0.65	0.65
com.1	165	161	168	179	179
com.2	146	149	142	146	145
com.3	98	98	101	93	93
com.4	74	75	72	65	65
com.5	0	0	0	0	1(独立)

表 4 は各 θ の時に得られた各コミュニティに，いくつの銘柄が所属しているかをまとめたものである。各コミュニティに所属する銘柄の数を比較すると，θ_{th} を変化させても構造自体は大きく変化していないように見える。あえてあげるとすれば，$\theta_{\mathrm{th}} : 0.05 \to 0.025$ で若干の変化があるようにも思える。

3.3.2 コミュニティの内訳

次に、各コミュニティに所属する銘柄を詳しく調べる。具体的には、業種に注目し、各コミュニティに所属する業種の内訳を調べた。図 3.1 は各閾値 θ_{th} においてどの業種が、どこのコミュニティに、どの程度所属しているのか、を表している。この図から明らかなように、θ_{th} を変化させ

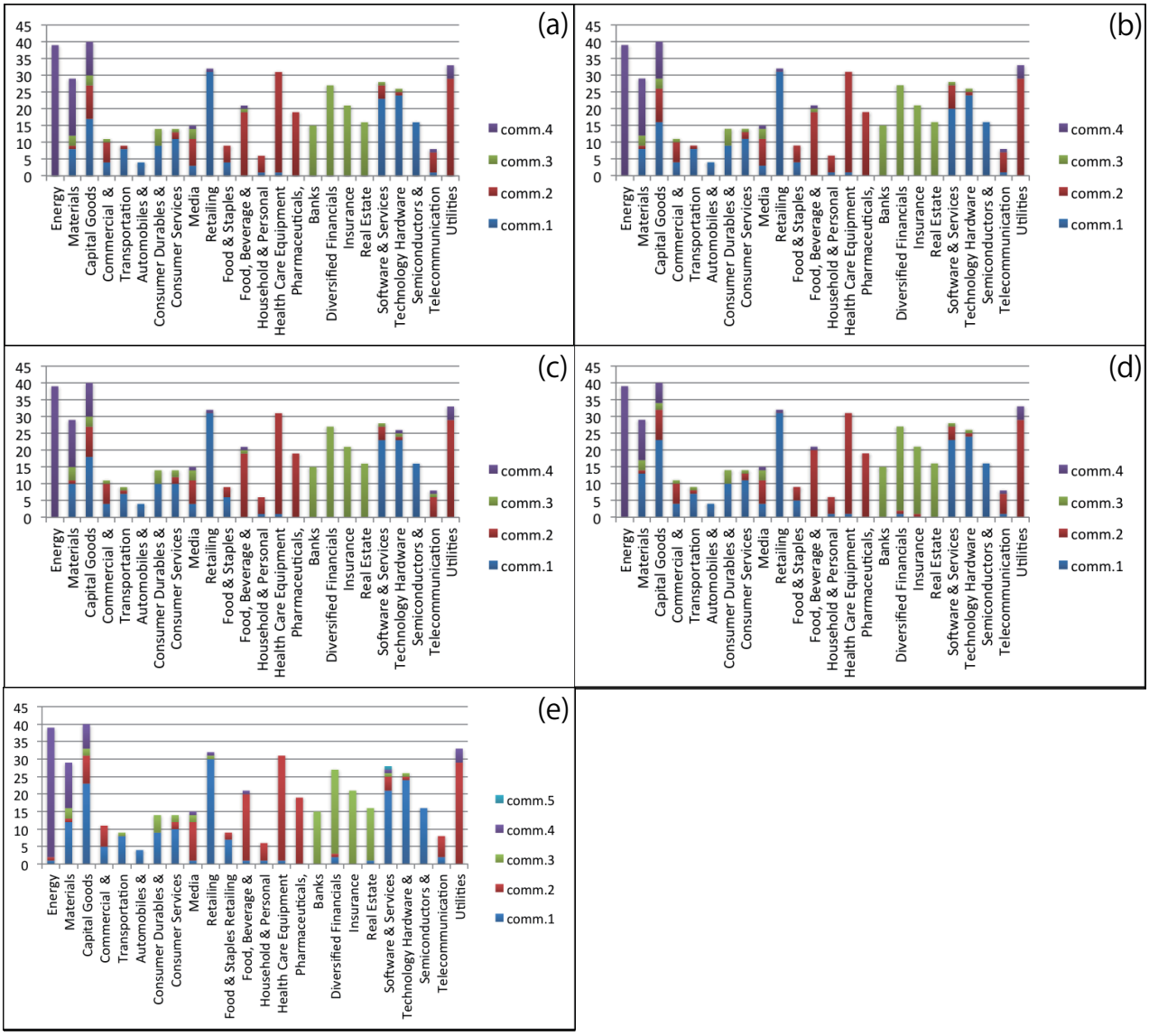

図 3.1　各 θ_{th} でのコミュニティに所属する業種の内訳。(a)∼(e) はそれぞれ $\theta_{\mathrm{th}} = 0.1, 0.075, 0.05, 0.025, 0.01$ に対応している。

てもほとんどコミュニティの構造に変化がないことがわかる。そこで、これを定量的に確かめるため、ジャッカード指数 (Jaccard Index) を用いる。

ジャッカード指数をひと言で説明すると，２つのコミュニティを比較した時にそれらがどの程度類似しているのかを測る指標である。コミュニティ m に所属するノードの集合を c_m とコミュニティ n に所属するノードの集合を c_n とすると，ジャッカード指数 J_{mn} は

$$J_{mn} = \frac{c_m \cap c_n}{c_m \cup c_n} \tag{3.4}$$

として定義される。式 (3.4) からわかるように、$0 \leq J_{mn} \leq 1$ である。$J_{mn} = 1$ となるのはコミュニティ m とコミュニティ n が一致する場合であるが、$J_{mn} \simeq 0$ となるのはコミュニティ m とコ

ミュニティ n が異なる場合だけではない。なぜなら、コミュニティ m とコミュニティ n のどちらかのサイズが極端に大きい (あるいは小さく)、$c_m \subset c_n$ となる場合もジャッカード指数の値は小さくなるからである。そこで、2つのコミュニティが異なるのか？それとも含まれてしまうのか？これを区別するために、本研究では規格化したジャッカード指数 $\hat{J}_{mn}$ を用いる。コミュニティ m の方がコミュニティ n よりも大きい場合、規格化したジャッカード指数 $\hat{J}_{mn}$ は

$$\hat{J}_{mn} \equiv \frac{c_m \cap c_n}{c_m \cup c_n} \Big/ \frac{c_n}{c_m} \tag{3.5}$$

で定義する。これによって、コミュニティ n がコミュニティ m に完全に含まれてしまう場合も $\hat{J}_{mn} = 1$ となる。図 3.2 はジャッカード指数の定義を模式的に表したものである。

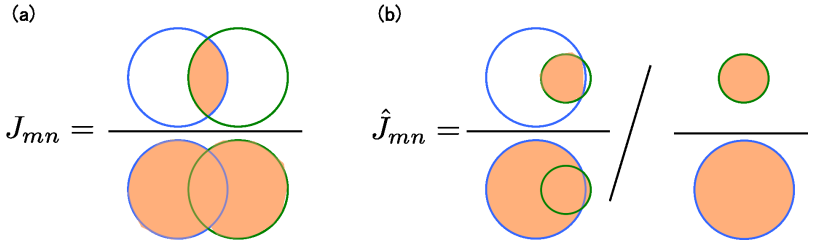

図 3.2　(a) 規格化しないジャッカード指数，(b) 規格化したジャッカード指数。規格化した場合、図のように片方の集合が小さくてもジャッカード指数は高い値になる。

	com.1	com.2	com.3	com.4
com.1	1.00	0.01	0.00	0.01
com.2	0.00	1.00	0.00	0.00
com.3	0.00	0.00	1.00	0.00
com.4	0.00	0.00	0.00	1.00

表 5　$\theta_{\mathrm{th}} = 0.1$ のコミュニティと $\theta_{\mathrm{th}} = 0.075$ のコミュニティのジャッカード指数

	com.1	com.2	com.3	com.4
com.1	0.94	0.00	0.02	0.01
com.2	0.03	1.00	0.00	0.00
com.3	0.01	0.00	0.98	0.00
com.4	0.04	0.00	0.00	0.97

表 6　$\theta_{\mathrm{th}} = 0.075$ のコミュニティと $\theta_{\mathrm{th}} = 0.05$ のコミュニティのジャッカード指数

	com.1	com.2	com.3	com.4
com.1	0.99	0.00	0.00	0.00
com.2	0.00	1.00	0.00	0.00
com.3	0.03	0.02	1.00	0.00
com.4	0.07	0.00	0.00	1.00

表 7　$\theta_{\mathrm{th}} = 0.05$ のコミュニティと $\theta_{\mathrm{th}} = 0.025$ のコミュニティのジャッカード指数

	com.1	com.2	com.3	com.4
com.1	0.87	0.02	0.03	0.03
com.2	0.02	0.90	0.02	0.00
com.3	0.04	0.01	0.86	0.00
com.4	0.02	0.01	0.00	0.91

表 8　$\theta_{\mathrm{th}} = 0.025$ のコミュニティと $\theta_{\mathrm{th}} = 0.01$ のコミュニティのジャッカード指数

表 5～ 表 8 を見るとわかるように、今回用いた θ_{th} の範囲内では得られたコミュニティとしてはほぼ同一である。そのため、次節からは $\theta_{\mathrm{th}} = 0.1$ で得られたコミュニティを用いてコミュニティ内・間の位相差について議論する。

3.3.3 コミュニティ内・間の位相差

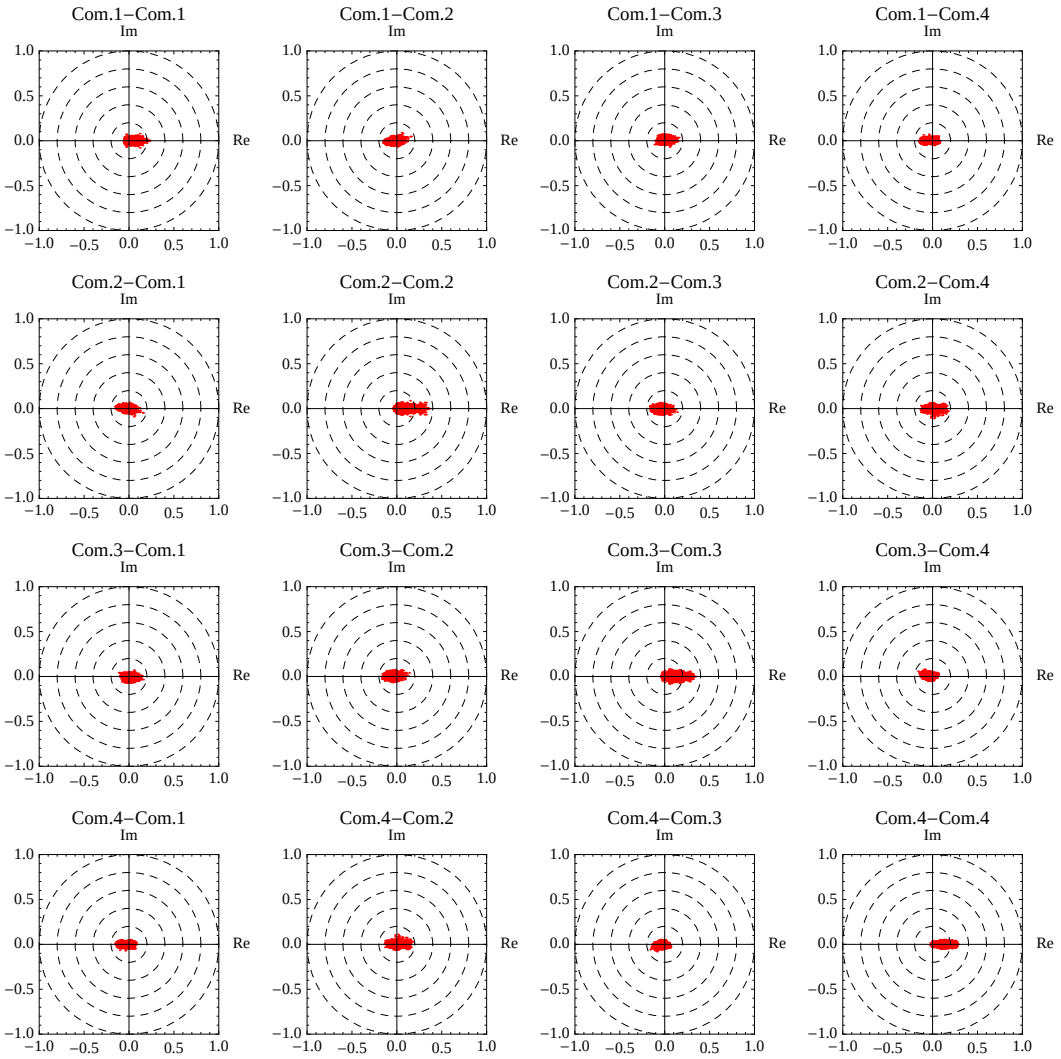

図 3.3　隣接行列要素をコミュニティ内・間で複素平面上にプロット

図 3.3 はコミュニティ内・間の隣接行列要素を複素平面上にプロットしたものである。ネットワークを構築した際には，同期した銘柄間のみにリンクを張ったが，図 3.3 ではコミュニティ内・間において全ての要素間についてプロットしている。この図からわかるように、同一コミュニティに所属するノード間の隣接行列要素は、正の実軸上に散らばって分布しているように見える。コミュニティは同期関係にある銘柄の集団であるから，正の実軸上 (位相差 0) に分布することは当然と言える。一方で，異なるコミュニティに所属するノード間の隣接行列要素は，どちらかと言えば負の実軸まわりに分布しているように見える。このことから，コミュニティ内のノード間における位相差はほぼゼロで同期しており，コミュニティ間では反相関関係にある位相差 $\pm\pi$ をもっていると予想できる。これを確認するために，コミュニティ内・間の銘柄について位相差をヒストグラム

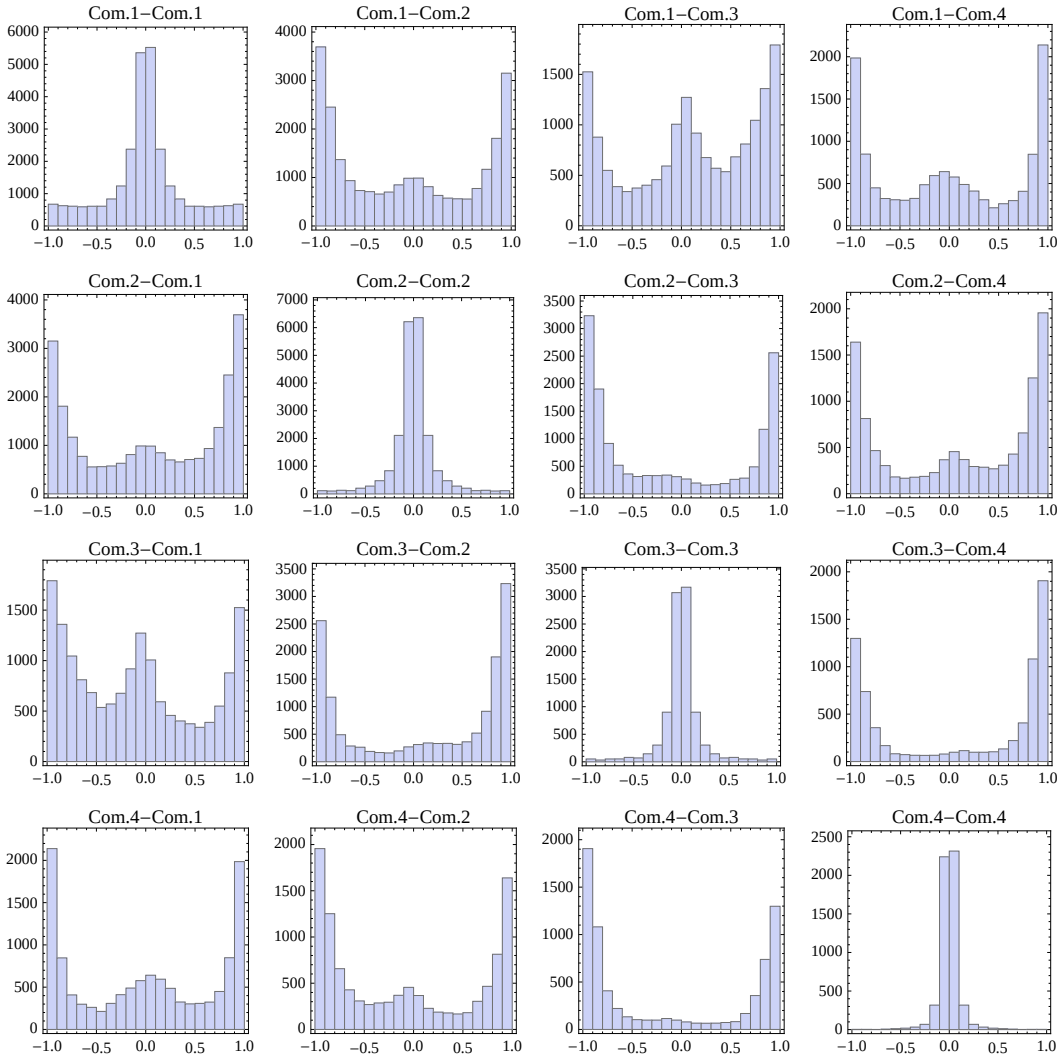

図 3.4　隣接行列要素の位相差をコミュニティ内・間でヒストグラム化

で表したのが図 3.4 である。この図から，先行研究 [17, 18, 19, 20] でも報告されている “フラストレーション状態にある多極構造” の存在が確認できる。そしてこのような構造が，複雑な市場の振る舞いを引き起こす要因の 1 つになっているのではないかと考えている。例えばフラストレーション状態は，物理学におけるスピン系によく現れる。そこでは，多種多様な磁性が発現する。

3.4　位相でリンクに向きをつけた有向ネットワークの構築・解析-有効ネットワーク-

上述したネットワークの構築方法では、銘柄間のリード・ラグ情報を無視した無向ネットワークとなっているため、位相情報を十分に活用できているとはいえない。そこで次に、銘柄間のリード・ラグ情報を用いてリンクに向きをつけた、有向ネットワークの構築・解析を行う。向きの付け方としては、銘柄 α よりも銘柄 β の方が位相が進んでいる場合、銘柄 α から銘柄 β の向きへリン

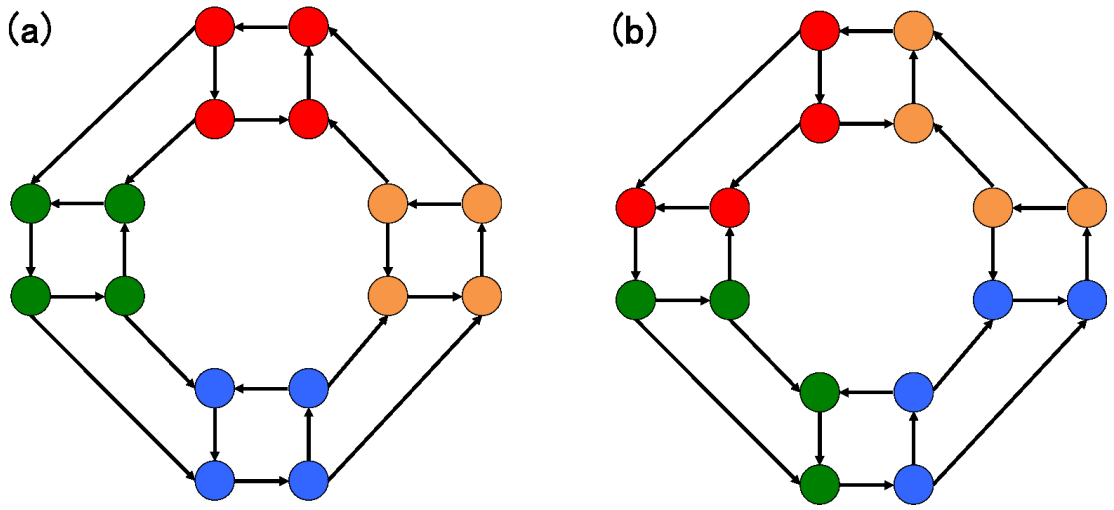

図 3.5　有向ネットワークのコミュニティ分割の例。MapEquation 法では (a) の方が良いコミュニティ分割とされる。

クを張る。このようにして構築したネットワークは、前節で構築したネットワークとトポロジーとしては同一である。しかし、リンクに向きがあるため、ネットワーク上での流れを考慮した解析が可能となる。また、前節でコミュニティ抽出に使用した Fast-Unfolding 法は有向ネットワークに対応していない。そのため今度は MapEquation 法 [29] を用いることでコミュニティの抽出を行う。

図 3.5 は有効ネットワークのコミュニティ分割の例である。同じ色のノードが同じコミュニティに所属していることを表している。図 3.5 においてリンクの向きを無くし、全てのリンクを等価とみなした場合、どちらのコミュニティもモジュラリティの値は等しくなる。しかし、MapEquation 法では (a) の方が良いコミュニティ分割だとされる。つまり、ノード間にループができているような集団がコミュニティして抽出される。

3.4.1　閾値 θ_{th} とコミュニティ数の変化

表 9 に θ_{th} の値と得られるコミュニティ数の変化をまとめた。MapEquation 法では、θ_{th} の値を変化させると小さなコミュニティが増える事がわかる。この理由としては、θ_{th} を小さくするとリンク数が減り、結果としてループ構造が減るためだと考えられる。

3.4.2　コミュニティの内訳

各 θ_{th} において得られた各コミュニティに所属する業種を図 3.6 にまとめた。各コミュニティにおいて、所属する業種は異なるものの、基本的には業種単位のまとまりが見える。

表 10 は $\theta_{\mathrm{th}} = 0.1 \sim 0.025$ 間のコミュニティの類似度をジャッカード指数で比較したものである。無向グラフの時は、θ_{th} を変化させてもあまりコミュニティ構造には変化が見られなかった。しかし、有向グラフの場合は θ_{th} によってコミュニティ構造が異なる。

表 9　各 θ_{th} における MapEquation 法によるコミュニティ抽出結果

θ_{th}	0.1	0.075	0.05	0.025	0.01
com.1	151	155	149	155	143
com.2	128	150	106	98	70
com.3	103	107	85	88	61
com.4	101	71	71	57	38
com.5	0	0	70	33	35
com.6	0	0	2	30	26
com.7	0	0	0	11	15
com.8	0	0	0	11	10
com.9	0	0	0	0	10
com.10	0	0	0	0	8
com.11	0	0	0	0	7
com.12	0	0	0	0	7
com.13	0	0	0	0	7
com.14	0	0	0	0	7
com.15	0	0	0	0	7
com.16	0	0	0	0	7
com.17	0	0	0	0	7
com.18	0	0	0	0	6
com.19	0	0	0	0	3
com.20	0	0	0	0	2
com.21	0	0	0	0	2
com.22	0	0	0	0	1
com.23	0	0	0	0	1
com.24	0	0	0	0	1
com.25	0	0	0	0	1
com.26	0	0	0	0	1(独立)

では、無向グラフとの対応はどうであろうか？無向グラフの $\theta_{\mathrm{th}} = 0.1$ と有向グラフの $\theta_{\mathrm{th}} = 0.1, 0.075$ をそれぞれ比較してみる。表 11 からわかるように、無向ネットワークからえられるコミュニティ構造は有向ネットワークにおける $\theta_{\mathrm{th}} = 0.075$ で得られるコミュニティ構造とよく似ている。つまり、無向ネットワークの $\theta_{\mathrm{th}} = 0.1$ で抽出されたコミュニティは、ループ構造を多く含んでいることがわかる。

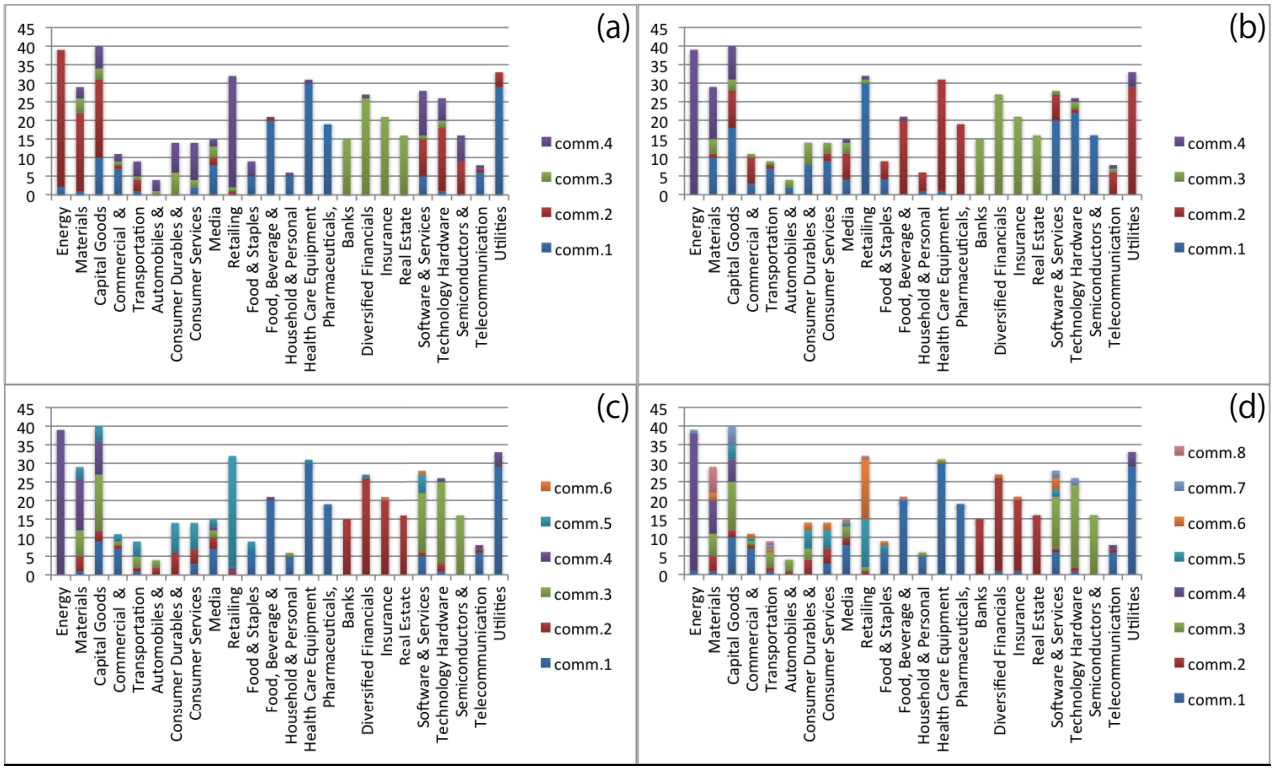

図 3.6 MqpEquation によるコミュニティ分割結果。図 (a)∼(d) は $\theta_{\mathrm{th}} = 0.1, 0.075, 0.05, 0.025$ に対応している。

3.4.3 コミュニティ内・間の位相差

$\theta_{\mathrm{th}} = 0.1, 0.075$ でのコミュニティ内・間の位相差ヒストグラムを示す (図 3.7,3.8)。これらの図からも、有効ネットワークを構築して得られたコミュニティ構造は、フラストレーション状態にある多極構造を構成している事がわかる。そして、$\theta_{\mathrm{th}} = 0.1$ に対する無向ネットワークと $\theta_{\mathrm{th}} = 0.075$ に対する有効ネットワークで得られたコミュニティ構造は似ていることから、図 3.4 と図 3.8 は似た分布を表している

3.5 まとめ

第 2 章で抽出した株式市場に潜む動的相関構造を基に、隣接行列を作成しネットワークを作成した。ただし、最大固有値に対応するモードはマーケットモードと呼ばれ、市場全体の動きを表しているため除外した。

この隣接行列では、各要素が複素数であるため大きさと位相に関する情報を持っている。そこで、ある閾値 θ_{th} を設け、その範囲内にある位相差の銘柄間にリンクを張り、重みを隣接行列の絶対値で与えたネットワークを構築した。その際、θ_{th} の値によってネットワークの構造が異なると予想できるので、θ_{th} の変化によるネットワーク構造の変化を調べた。特に、同期関係にある株式

	com.1	com.2	com.3	com.4
com.1	0.00	0.97	0.00	0.02
com.2	0.32	0.00	0.00	0.96
com.3	0.00	0.00	1.00	0.00
com.4	0.91	0.01	0.02	0.00

	com.1	com.2	com.3	com.4	com.5	com.6
com.1	0.01	0.01	0.98	0.00	0.96	0.00
com.2	0.98	0.00	0.01	0.00	0.01	0.50
com.3	0.00	0.98	0.00	0.00	0.01	050
com.4	0.00	0.00	0.00	1.00	0.00	0.00

	com.1	com.2	com.3	com.4	com.5	com.6	com.7	com.8
com.1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
com.2	0.01	1.00	0.03	0.00	0.05	0.00	0.00	0.00
com.3	0.01	0.00	0.83	0.00	0.02	0.00	0.42	0.08
com.4	0.01	0.00	0.00	1.00	0.00	0.00	0.42	0.70
com.5	0.02	0.00	0.06	0.00	0.87	0.91	0.08	0.16
com.6	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00

表 10　上から $\theta_{\mathrm{th}} = 0, 1 - 0.075$,$\theta_{\mathrm{th}} = 0.075 - 0.05$,$\theta_{\mathrm{th}} = 0.05 - 0.025$ 間のジャッカード指数

	com.1	com.2	com.3	com.4
com.1	0.01	0.30	0.04	0.98
com.2	1.00	0.00	0.00	0.00
com.3	0.01	0.00	0.96	0.01
com.4	0.02	0.96	0.00	0.00

	com.1	com.2	com.3	com.4
com.1	0.93	0.01	0.06	0.02
com.2	0.00	0.99	0.00	0.00
com.3	0.00	0.01	0.98	0.00
com.4	0.05	0.00	0.00	0.95

表 11　左:無向グラフの $\theta_{\mathrm{th}} = 0.1$ と有向グラフの $\theta_{\mathrm{th}} = 0.1$, 右:無向グラフの $\theta_{\mathrm{th}} = 0.1$ と有向グラフの $\theta_{\mathrm{th}} = 0.075$

銘柄をコミュニティとして抽出することで、その変化に注目した。その結果、どの θ_{th} においても基本的には 4 つのコミュニティが得られた。このことから、各 θ_{th} で得られたコミュニティ構造は全て同一のものであると考えられるが、これを定量的に評価するためにジャッカード指数を用いて、コミュニティ構造の比較を行った。すると、やはりコミュニティ構造に大きな変化は見られなかった。そして、$\theta_{\mathrm{th}} = 0.1$ で得られたコミュニティの内部構造を調べた結果、各コミュニティは業種単位で構成されていることがわかった。

次に、$\theta_{\mathrm{th}} = 0.1$ で得られた各コミュニティ内・間の位相差について調べた。その結果、コミュ

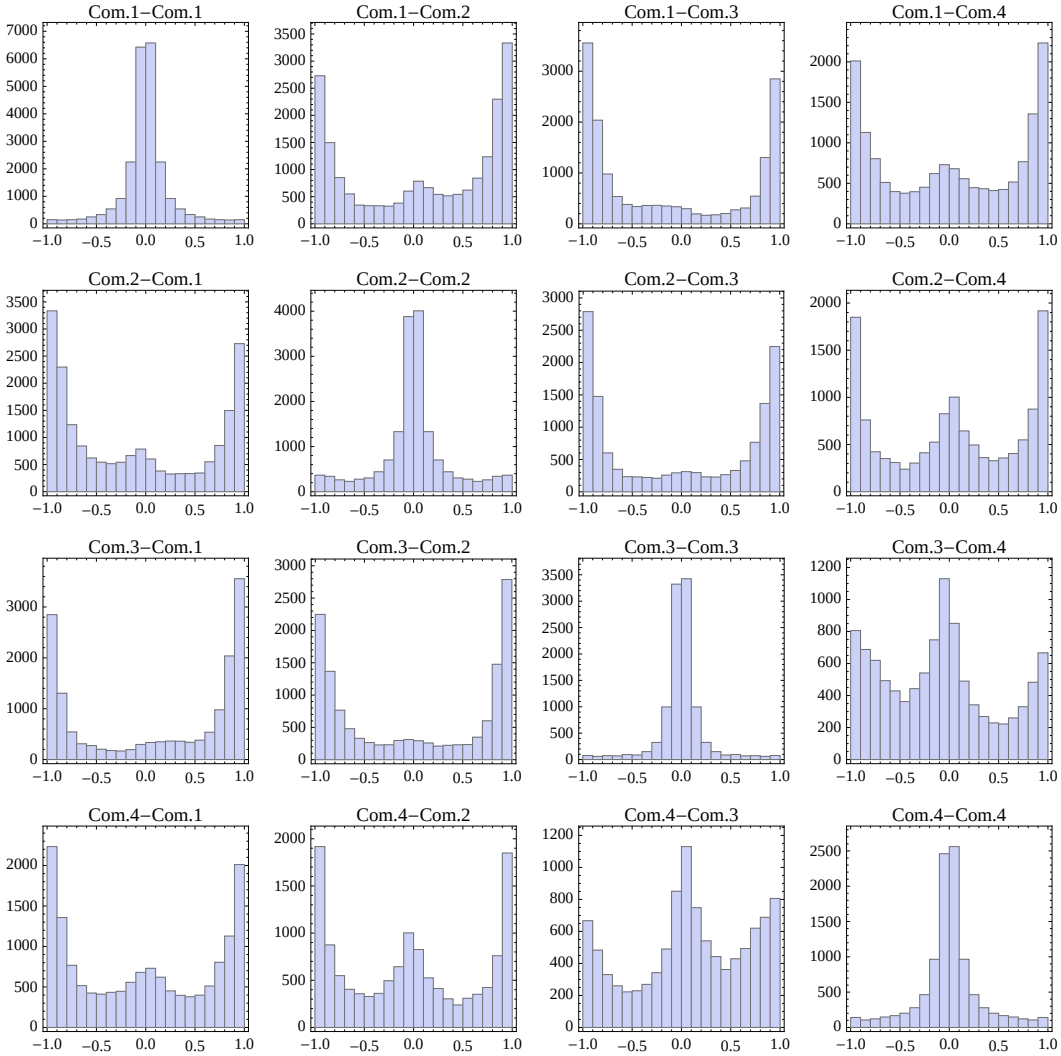

図 3.7　有向ネットワークの $\theta_{\mathrm{th}} = 0.1$ でのコミュニティ内・間の位相差

ニティ内は同位相、コミュニティ間では逆位相の関係にある事が分かった。このような結果は同時刻相関のみを用いて行われた先行研究からも得られている。

ここまでは、無向ネットワークとしての解析であるが、有効ネットワークとしての解析も行った。この場合、同期関係にあり、ループ構造になっているノード集団がコミュニティとして抽出されている。そして、有効ネットワークと無向ネットワークで似たコミュニティ構造が得られていることから、本研究で扱った株価相関ネットワークで同期した銘柄集団はループ状の構造であることを意味している。また、株価相関ネットワークではコミュニティ間が基本的には反相関関係であったが、一般的にはコミュニティ間の流れ (上流・下流) についての解析も可能だと考えられる。例えば、最も簡単な場合としては、コミュニティ間の位相差ヒストグラムから、コミュニティ間の平均的な位相差を見積もることで、ネットワーク上のダイナミクスを解析することができる。

フラストレーション状態は静的には起こりえない、非常に不安定な状態である。そのため、その

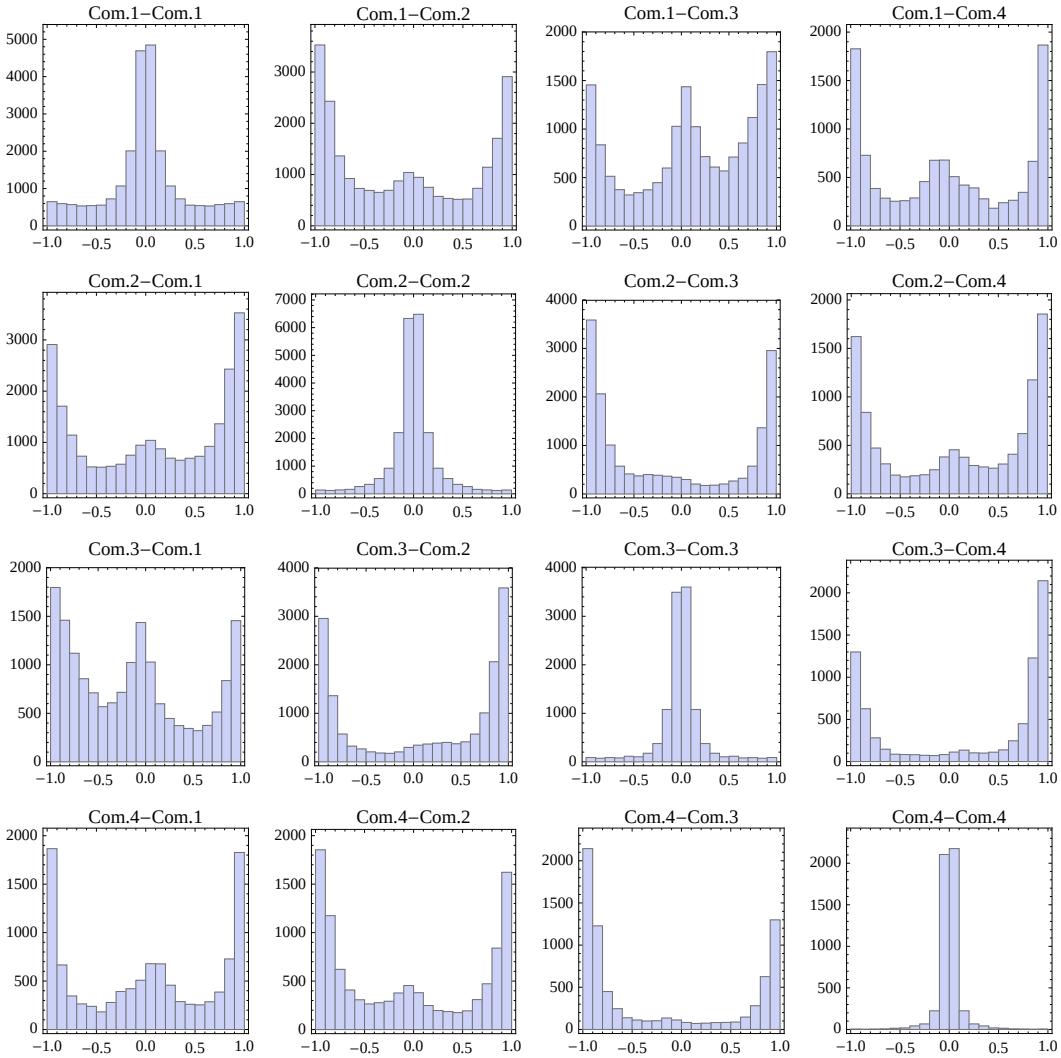

図 3.8　有向ネットワークの $\theta_{\mathrm{th}} = 0.075$ でのコミュニティ内・間の位相差

構造をより詳しく調べるためには動的な相関構造からの視点が必要になるはずである。今後詳細な解析を行っていくことで、フラストレーション状態にある多極構造の解明に本研究手法が有用である事を期待する。具体的には、現在行っている手法では、位相差を計算する際に時間的に平均した位相情報を用いている。しかし、本研究の結果として得られたように、株価相関ネットワークにおいて時間的に平均した位相情報では、銘柄間の位相差を的確に捉えられていない可能性が考えられる。これを解決するためには、位相差を平均せずに、各時点における位相情報を用いるなど、より詳細な解析を行う必要があると考えている。

4　株式市場における多体相関

4.1　株式市場における多体相関

第 3 章で述べたように、株式市場にはフラストレーション状態にある多極構造が存在する。存在は確認できているが、未だその発生原因を含めなにもわかっていない。しかし、フラストレーション状態というのは、これまで多くの研究が行われてきたような 2 銘柄間の相関（2 体相関）を考えているだけでは決して現れず、3 銘柄以上の多体相関を扱うことによって初めて現れる現象である。そこで本章では、多体相関を扱うための第一歩として株式銘柄の 3 体相関について解析を行い、その結果を報告する。

4.2　2 体相関と 3 体相関

これまでと同様に、対数収益率を取り標準化した時系列 $r_\alpha(t)$ を用いて解析を行う。2 体相関と 3 体相関はそれぞれ

$$\langle r_\alpha r_\beta \rangle = \frac{1}{T}\sum_{t=1}^{T} r_\alpha(t) r_\beta(t) \tag{4.1}$$

$$\langle r_\alpha r_\beta r_\delta \rangle = \frac{1}{T}\sum_{t=1}^{T} r_\alpha(t) r_\beta(t) r_\delta(t) \tag{4.2}$$

として求められる。今回解析に用いているデータに対して 2 体相関と 3 体相関をそれぞれ計算し、ヒストグラムで表すと図 4.1 のようになる。図をみるとわかるように、2 体相関ではほぼ全てのペアが正の相関を取るにも関わらず、3 体相関では半数以上の組み合わせで負の値をとっている。このことは一見、不自然な様に見える。

そこで本研究では、最もシンプルな場合として 2 体相関と 3 体相関の出現パターンについて調べる。出現パターンを全て列挙すると 図 4.2 のようになる。このうち、2 体相関が正となるのは青線で囲った組み合わせ、3 体相関が負になるのは赤線で囲った組み合わせになっている。つまり、

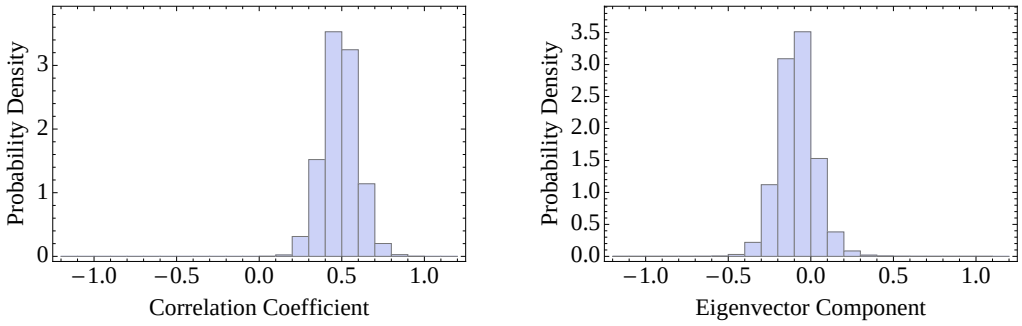

図 4.1　左図: 2 体相関係数のヒストグラム、右図:3 体相関係数のヒストグラム

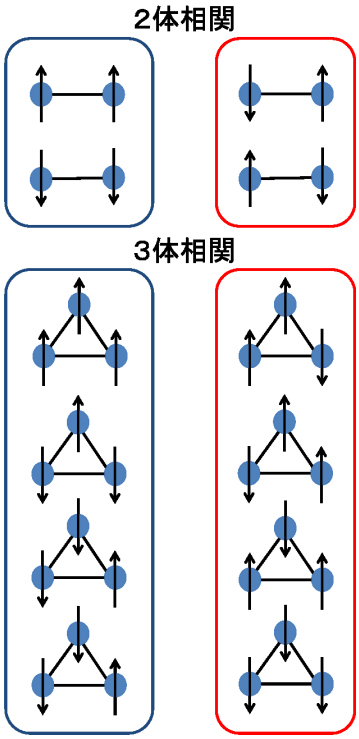

図 4.2　２体相関と３体相関の出現パターン

$\downarrow\downarrow$ のペアが統計的に有意に多く出現するのであれば、$\downarrow\downarrow\downarrow$ の組み合わせも出やすいということが考えられる。そうであれば、３体相関を２体相関で説明できる可能性がある。この場合、本質的な３体相関は存在せず、株価の価格変動の相関を扱う場合、２体相関のみに注目すれば良いことになる。本研究では価格の変動幅は無視して、価格が上がるか下がるかの２値化し、２体相関効果と３体相関効果を２項分布を用いて統計的に評価した。

4.3　２項分布を用いた３体相関の見積もり

解析手順としては

1. $r_\alpha(t)$ の全要素 $T \times N = 483 \times 1008 = 486864$ のうち、$\uparrow$ と $\downarrow$ が出現する回数をカウント
2. 価格変動が正 ($\uparrow$) の場合と負 ($\downarrow$) の場合が等確率で現れるとして、$\uparrow$ と $\downarrow$ の出る確率を評価
3. $\uparrow$ が出る回数と $\downarrow$ が出る回数に統計的有意な偏りがあれば、1 の結果に基いて $\uparrow$ と $\downarrow$ が出る確率 $P_{\uparrow(\downarrow)}$ を見積もる
4. $r_\alpha(t)r_\beta(t)$ の全組み合わせ $\frac{TN(N-1)}{2}$ のうち $\uparrow\uparrow, \uparrow\downarrow$ $(\downarrow\uparrow), \downarrow\downarrow$ の数をそれぞれカウントし、それを $P_{\uparrow(\downarrow)}$ で見積もれるか確認
5. 見積もれない、つまり $\uparrow\uparrow, \downarrow\downarrow$ などの出方に偏りがあれば、それを２体相関の効果として確率に取り入れ、３体相関 $\uparrow\uparrow\uparrow$ などの出方を見積もれるか確認
6. それでも見積もれない場合、有意な３体相関の効果があると考えられる

2 体相関と 3 体相関において、各パターンの出現回数を計算した結果

$$\uparrow: 247027 \tag{4.3}$$
$$\downarrow: 239837 \tag{4.4}$$

$$\uparrow\uparrow: 40023256 \tag{4.5}$$
$$\uparrow\downarrow: 39020502 \tag{4.6}$$
$$\downarrow\downarrow: 38290466 \tag{4.7}$$

$$\uparrow\uparrow\uparrow: 4799443926 \tag{4.8}$$
$$\uparrow\uparrow\downarrow: 4852854358 \tag{4.9}$$
$$\uparrow\downarrow\downarrow: 4531576373 \tag{4.10}$$
$$\downarrow\downarrow\downarrow: 4628712591 \tag{4.11}$$

であった。

まず、$P_\uparrow = P_\downarrow = 0.5$ として、$\uparrow, \downarrow$ の出現回数を見積もる。

$$E[\downarrow] = NTP_\downarrow = 243432 \tag{4.12}$$
$$V[\downarrow] = NTP- \downarrow P \uparrow = 121716 \tag{4.13}$$
$$\sigma[\downarrow] = \sqrt{V[\downarrow]} \simeq 349 \tag{4.14}$$
$$[E[\downarrow] - 2\sigma[\downarrow], E[\downarrow] + 2\sigma[\downarrow]] = [242123, 243517] \tag{4.15}$$

であるから実際にカウントされた $\downarrow$ の回数は 95% 信頼区間に入らない。そのため、$\downarrow$ の出方には偏りがあるとみなし

$$P'_\uparrow = \frac{\#\uparrow}{\#\uparrow + \#\downarrow} = 0.507 \tag{4.16}$$
$$P'_\downarrow = \frac{\#\downarrow}{\#\uparrow + \#\downarrow} = 0.493 \tag{4.17}$$

として確率を再定義する。そして、これらの確率を用いて $\downarrow\downarrow$ の組み合わせが現れる回数を見積もる。すると

$$E[\downarrow\downarrow] = \frac{N(N-1)T}{2} P'_\downarrow P'_\downarrow \simeq 2.847 \times 10^7 \tag{4.18}$$
$$V[\downarrow\downarrow] = \frac{N(N-1)T}{2} P'_\downarrow P'_\downarrow \left(1 -' P''_\downarrow P'_\downarrow\right) \simeq 2.156 \times 10^7 \tag{4.19}$$
$$\sigma[\downarrow\downarrow] = \sqrt{V[\downarrow\downarrow]} \simeq 4644 \tag{4.20}$$
$$[E[\downarrow\downarrow] - 2\sigma[\downarrow\downarrow], E[\downarrow\downarrow] + 2\sigma[\downarrow\downarrow] = \left[2.846 \times 10^7, 2.848 \times 10^7\right] \tag{4.21}$$

である。実際にカウントされた数はこの区間内には含まれないため、2 体相関による相関効果があると考える。そして、その効果によって生じる項を r として相関効果を見積もる。

$$\frac{N(N-1)T}{2} P'_\downarrow P_\downarrow (1+r) = 38290466 \tag{4.22}$$

を r について解くと、$r \simeq 0.345$ が得られる。次に、この相関効果を取り入れて $\downarrow\downarrow\downarrow$ が出る回数を見積もる。擬 Kirkwood 近似を用いて $\downarrow\downarrow\downarrow$ が出る確率を

$$P_{\downarrow\downarrow\downarrow} = P'_{\downarrow}P'_{\downarrow}P'_{\downarrow}(1 + 3r + 3r^2) \tag{4.23}$$

として

$$E[\downarrow\downarrow\downarrow] = \frac{N(N-1)(N-2)T}{6}P_{\downarrow\downarrow\downarrow} \simeq 5.377 \times 10^9 \tag{4.24}$$

$$V[\downarrow\downarrow\downarrow] = \frac{N(N-1)(N-2)T}{6}P_{\downarrow\downarrow\downarrow}\left(1 - P_{\downarrow\downarrow\downarrow}\right) \simeq 3.840 \times 10^9 \tag{4.25}$$

$$\sigma[\downarrow\downarrow\downarrow] = \sqrt{V[\downarrow\downarrow\downarrow]} \simeq 61969 \tag{4.26}$$

$$[E[\downarrow\downarrow\downarrow] - 2\sigma[\downarrow\downarrow\downarrow], E[\downarrow\downarrow\downarrow] + 2\sigma[\downarrow\downarrow\downarrow] = \left[5.3769 \times 10^7, 5.3771 \times 10^7\right] \tag{4.27}$$

なので、実際にカウントされた回数よりも多く、これも 95% 信頼区間に入らない。$\downarrow\downarrow$ が出やすかったのに比べて、$\downarrow\downarrow\downarrow$ は出にくい、つまり抑制されていることがわかる。そこで、この抑制されている効果を $\delta 1$ として

$$P_{\downarrow\downarrow\downarrow} = P_{\downarrow}P_{\downarrow}P_{\downarrow}(1 + 3r + 3r^2 + \delta 1) \tag{4.28}$$

のように取り入れて、実際に得られた値から $\delta 1$ の値を見積もると

$$\delta 1 \simeq -0.333 \tag{4.29}$$

となり、30% 程度の抑制効果がある結果となった。

$\downarrow\downarrow\downarrow$ の時と同様に、それ以外の出現パターンについても計算する。

$$P_{\uparrow\uparrow} = P'_{\downarrow}P'_{\downarrow}\left(1 + s\right) \tag{4.30}$$

$$P_{\uparrow\downarrow} = P'_{\uparrow}P'_{\downarrow}\left(1 + t\right) \tag{4.31}$$

$$P_{\uparrow\uparrow\uparrow} = P'_{\uparrow}P'_{\uparrow}P'_{\uparrow}\left(1 + 3s + 3s^2 + \delta 2\right) \tag{4.32}$$

$$P_{\uparrow\uparrow\downarrow} = P'_{\uparrow}P'_{\uparrow}P'_{\downarrow}\left(1 + 2t + s + 2ts + t^2 + \delta 3\right) \tag{4.33}$$

$$P_{\uparrow\downarrow\downarrow} = P'_{\uparrow}P'_{\downarrow}P'_{\downarrow}\left(1 + 2t + r + 2tr + t^2 + \delta 4\right) \tag{4.34}$$

とおいて、それぞれの項を見積もると

$$s \simeq 0.325 \tag{4.35}$$

$$t \simeq -0.335 \tag{4.36}$$

$$\delta 2 \simeq -0.339 \tag{4.37}$$

$$\delta 3 \simeq 0.128 \tag{4.38}$$

$$\delta 4 \simeq 0.096 \tag{4.39}$$

が求まる。以上の結果から、同じ価格変動を抑制するような 3 体相関の効果が存在するのではないかと期待できる。

4.4 まとめ

本章では、株式市場における多体相関の第一歩として３体相関の解析を行った。特に、最も基本的な性質である株価の変動幅は考えず、上がるか下がるかの２値化した場合について調べた。

その結果、３つの銘柄を任意に選んだ場合、それら全てが同じ変動をするのを阻止するような３体相関効果の存在が確認できた。本章の解析においては、第 2 章と第 3 章と異なり、マーケットモードを取り除いていない。そのため、得られた３体相関効果はこのマーケットモードに起因しているのではないかと考えられる。なぜならば、複数の株価が似た価格変動を促進するような効果があった場合、株価が簡単に高騰 (暴落) してしまうため市場は不安定になってしまう。

今後は、今回得られた３体相関効果が常に存在するのか、それとも逆に作用するような場合もあるのかなど、３体相関効果の安定性についての解析も行う必要があると考えている。

5　まとめと今後の課題

本研究では、株式市場における動的相関構造を調べるため

- 複素主成分分析とランダム行列理論を組み合わせた、株式市場における動的相関構造の抽出 (第 2 章)
- 動的相関構造を反映させた株価相関ネットワークの構築・解析 (第 3 章)

を行った。また、ネットワーク解析の結果得られたフラストレーション構造について迫るため

- より高次の相関を解析するための第一歩として、３体相関効果の基礎的な性質 (第 4 章)

についても調べた。

まず、第 2 章では株価時系列データ間のタイムラグのある相関構造 (動的相関構造) の抽出を行った。従来の主成分分析では、同時刻の変量間で相関をとっているため、たとえば $\sin t$ と $\cos t$ のように位相が $\pi/2$ ズレた、タイムラグのある相関の抽出が困難であった。そこで、そのようなタイムラグのある動的相関構造の抽出を行うために、ヒルベルト変換に注目した。実データから得られる時系列にヒルベルト変換を施すと、位相を $\pi/2$ シフトした時系列を得られる。これを実データの虚部に付け加え、時系列を複素数化することで主成分分析を行う (複素主成分分析)。ここで、主成分分析ではどこまでを有意な相関構造と見なすかが非常に重要な問題であるが、経済物理学の分野ではしばしばランダム行列理論から得られるランダム相関行列の固有値や固有ベクトル成分の分布が帰無仮説として用いられる。しかし、その分布は各行列要素が独立同分布に従うという条件の下で導出されるが、本研究で扱う複素時系列では実部と虚部が独立でない。そのため、本研究へ適用可能なようにランダム相関行列の固有値と固有ベクトル成分の分布について拡張を行った上で、株式市場に潜む動的相関構造の抽出を行った。

S&P500 インデックスに採用されている株式銘柄のうち、2008 年 ∼ 2011 年の全営業日で取引が行われた 483 銘柄に対して複素主成分分析を行ったところ、全部で 7 つの有意な固有モードが得られた。最大固有値に対応するモード 1 は、全ての銘柄からの寄与が等しく同位相の関係にあるため、市場全体の動きを反映したマーケットモードに対応している。そして 2∼4 番目に大きな固有値に対応するモード 2 は、相関 (位相差 0)・反相関 (位相差 $\pm\pi$) が支配的なモードであり、従来の主成分分析でも抽出される構造であることが分かった。一方、モード 5 では固有ベクトル成分が複素平面上に広く分布しているため、動的相関構造が存在している。その上、従来の主成分分析のモード 5 とモード 7 が直角位相で結合して現れていることも分かった。

今後の課題として、現在は複素数へ拡張した解析を行っているが、４元数へ拡張した主成分分析も行っていきたいと考えている。実際、４元数に拡張した主成分分析の手法自体は提案されており [30, 31, 32, 33]、ランダム行列理論に関しても行列要素が４元数であるようなモデルが存在している。そのため、現在の時系列に $\pm\pi/4$ 位相がズレた時系列も組み込んで４元数へ拡張することで、より高い解像度で位相情報を得る事ができるのではないかと期待している。

次に、得られた動的相関構造から相関行列を再構築し、それを隣接行列とみなすことで株価相関ネットワークの解析を行った (第 3 章)。このようにして構築したネットワークは、隣接行列の各要素が複素数であることから、リンクの重みとして複素数の値を持つ。そのため、リンクの強さ (相関係数の絶対値) とリード・ラグ (相関係数の位相) という 2 つの情報を持っている。そして、株価の変動が同期関係にあるリンクに着目し、それらの銘柄集団をコミュニティとして抽出する。その結果、4 つのコミュニティが得られ、コミュニティ内は同位相 (正の相関関係)、コミュニティ間は逆位相 (負の相関関係) であるようなフラストレーション状態にある多極構造の存在が動的にも確認できた。今後の課題として、今回得られた結果ではコミュニティ間が逆位相であったため、コミュニティ間の流れに関する議論を行うことができなかった。しかし、他分野のデータへ本研究手法を応用し、コミュニティ間の位相差に偏りがあれば、コミュニティの上流・下流といった流れの議論を行うことも可能となる。

またフラストレーション構造は、3 つの異なるコミュニティから株式銘柄を 1 つずつ取り出してきた場合、それら 3 つの銘柄は互いに反相関関係にある事を意味している。このような構造は一見不自然であり、静的には不安定である。しかしこのような構造の存在が、株式市場において複雑な振る舞いを引き起こす要因の 1 つになっているのではないかと考えている。たとえば、フラストレーション構造は物理学におけるスピン系でしばしば現れ、そこでは様々な磁性が発現している。今回の解析では、時刻について平均した位相差を用いていたが、各時点の銘柄間の位相差に注目してより詳細な解析を行うことで、フラストレーション構造の解明に迫ることができるのではないかと考えている。

最後に、フラストレーション構造は 3 体以上の高次相関を扱うことで初めて現れる構造である。フラストレーション構造に迫るための、動的相関構造とは別のアプローチとして高次相関についての解析も必要だと考えられる。そのための第 1 歩として、株価の変動幅は無視して上がるか下がるかの 2 値化し、2 体相関効果と 3 体相関効果を 2 項分布を用いて統計的に解析した。その結果、(1) 任意の 2 つの銘柄を選んだ場合、それらが同じ変動するのを促進するように 2 体相関効果が働き、(2) 任意の 3 つの銘柄を選んだ場合、それらが同じ変動するのを抑制するように 3 体相関効果が働くことが分かった。銘柄間の高次相関効果が同じ変動を促進するように働いてしまうと、株価が発散してしまうため、高次相関が抑制効果を持つことは自然のように考えられる。このことから、銘柄間の高次相関を調べることで市場の安定性に関する議論ができるのではないかと考えている。今回の解析では、株価時系列の全期間を用いて高次相関の効果を見積もったが、今後は短いデータ期間に区切って高次相関を調べるなどして、株式市場における高次相関についてより詳細な解析を行っていきたいと考えている。

付録 A　ランダム行列理論

A.1　ガウス型直交アンサンブル

ここでは [34] に沿って、ランダム行列の基本的なアンサンブルの 1 つである、ガウス型直交アンサンブルの説明を行う。ランダム行列に関する詳細については、[35] などを参照されたい。

まずは、最も単純なケースである、2×2 の実行列について考える。ランダム行列とはその名の通りに、行列要素がある確率変数からなる行列である。ここでは、2×2 実行列を $\mathbf{R}$ として、実数の要素 R_{jl} がある確率変数に従っているものとする。

このとき、ランダム行列 $\mathbf{R}$ の対角成分の和について

$$\mathrm{Tr}(\mathbf{R}\mathbf{R}^{\mathrm{T}}) = R_{11}^2 + R_{12}^2 + R_{21}^2 + R_{22}^2 \tag{A.1}$$

が成り立つ。ここで、$\mathbf{R}^{\mathrm{T}}$ は $\mathbf{R}$ の転置行列を表す。

一方、実数 x が平均 0, 分散 σ でガウス分布しているとき、その確率分布関数 $P(x)$ は、

$$P(x)dx = \frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-x^2/(2\sigma^2)}dx \tag{A.2}$$

である。したがって、行列要素 R_{jl} が独立に (平均 0, 分散 1) でガウス分布しているとき、確率分布関数 $P(\mathbf{R})$ は、

$$\begin{aligned} P(\mathbf{R})(d\mathbf{R}) &\propto \mathrm{e}^{-(1/2)\mathrm{Tr}(\mathbf{R}^{\mathrm{T}}\mathbf{R})}(d\mathbf{R}) \\ &= \exp\left(-\frac{1}{2}R_{11}^2 - \frac{1}{2}R_{12}^2 - \frac{1}{2}R_{21}^2 - \frac{1}{2}R_{22}^2\right) dR_{11}dR_{12}dR_{21}dR_{22} \end{aligned} \tag{A.3}$$

と書ける。

さらに、行列が実対称であるという制約を $\mathbf{R}$ に課すと、2×2 実対称行列 $\mathbf{S}$ の要素 S_{jl} がそれぞれ互いに独立でガウス分布しているとき、

$$\begin{aligned} P(\mathbf{S})(d\mathbf{S}) &\propto \mathrm{e}^{-(1/2)\mathrm{Tr}\mathbf{S}^2}(d\mathbf{S}) \\ &= \exp\left(-\frac{1}{2}S_{11}^2 - S_{12}^2 - \frac{1}{2}S_{22}^2\right) dS_{11}dS_{12}dS_{22} \end{aligned} \tag{A.4}$$

になる。ここで、$\mathbf{S}$ は対称行列であるから、$S_{12} = S_{21}$ なので非対角成分の分散は 1/2 とした。また、式 (A.3) では積分するための測度が 4 つ $(dA_{11}dA_{12}dA_{21}dA_{22})$ であったが、式 (A.4) では対称性のために測度が 3 つ $(dS_{11}dS_{12}dS_{22})$ に減っている。このように、行列要素がガウス分布に従っていて、対称であるようなランダム行列をガウス型直交アンサンブルという。

ランダム行列を扱う際に、興味が有るのはその固有値や固有ベクトル成分の分布である。行列サイズが 2×2 程度のランダム行列であれば、それらの分布を簡単に計算する事ができる。しかし一般に、そのような小さなサイズの場合の分布には興味がなく、もっと大きなサイズの場合にどのような分布となるのか知りたい。

そこで、行列サイズを $2\times 2 \to N\times N$ へ拡張したガウス型直交アンサンブルについて考える。するとこの場合の確率分布関数は

$$
\begin{aligned}
P(\mathbf{S})(d\mathbf{S}) &\propto \mathrm{e}^{-(1/2)\mathrm{Tr}\mathbf{S}^2}(d\mathbf{S}) \\
&= \prod_{j=1}^{N} \mathrm{e}^{-(1/2)S_{jj}^2} dS_{jj} \prod_{j<l}^{N} \mathrm{e}^{-S_{jl}^2} dS_{jl}
\end{aligned}
\tag{A.5}
$$

となる。いま、$\mathbf{S}$ の固有値と固有ベクトル成分がどのような分布に従うのか知りたいので、式 (A.5) において、$\mathbf{S}$ から固有値 λ と固有ベクトル $\mathbf{O}$ へ変数変換したい。この場合の問題は、測度

$$
(d\mathbf{S}) = \prod_{j=1}^{N} dS_{jj} \prod_{j<l}^{N} dS_{jl} \tag{A.6}
$$

が、変数変換を行ったときにどう変わるかである。そしてこのような場合、行列のヤコビアンを計算すれば良いことが知られている。

まず、$N\times N$ 実対称行列 $\mathbf{S}$ は、$N\times N$ 直交行列 $\mathbf{O}$ によって、

$$
\mathbf{S} = \mathbf{O}\boldsymbol{\Lambda}\mathbf{O}^{\mathrm{T}} \tag{A.7}
$$

のように対角化できる。ここで、Λ は、$\mathbf{S}$ の実固有値 $\lambda_1,\ldots,\lambda_N$ を対角成分にもつ対角行列とする。式 (A.7) の両辺を微分すると

$$
d\mathbf{S} = d\mathbf{O}\boldsymbol{\Lambda}\mathbf{O}^{\mathrm{T}} + \mathbf{O}\boldsymbol{\Lambda}d\mathbf{O}^{\mathrm{T}} + \mathbf{O}d\boldsymbol{\Lambda}\mathbf{O}^{\mathrm{T}} \tag{A.8}
$$

であり、両辺に左から $\mathbf{O}^{\mathrm{T}}$ を掛け、右から $\mathbf{O}$ を掛けると

$$
\mathbf{O}^{\mathrm{T}}d\mathbf{S}\mathbf{O} = \mathbf{O}^{\mathrm{T}}d\mathbf{O}\boldsymbol{\Lambda} + \boldsymbol{\Lambda}d\mathbf{O}^{\mathrm{T}}\mathbf{O} + d\boldsymbol{\Lambda} \tag{A.9}
$$

を得る。

そして、測度について $(\mathbf{O}^{\mathrm{T}}d\mathbf{S}\mathbf{O}) = (d\mathbf{S})$ という性質を利用して、式 (A.9) の右辺の (j,l) 成分 $(j\le l)$ の外積を計算すると、測度 $(d\mathbf{S})$ を固有値 λ と固有ベクトル $\mathbf{O}$ で書き表す事ができる。固有ベクトル $\mathbf{O}$ は直交行列であるから、$\mathbf{O}^{\mathrm{T}}\mathbf{O} = \mathbf{I}$ より、

$$
\mathbf{O}^{\mathrm{T}}d\mathbf{O} = -d\mathbf{O}^{\mathrm{T}}\mathbf{O} \tag{A.10}
$$

を得る。またこれから、

$$
\left[\mathbf{O}^{\mathrm{T}}d\mathbf{O}\boldsymbol{\Lambda} + \boldsymbol{\Lambda}d\mathbf{O}^{\mathrm{T}}\mathbf{O} + d\boldsymbol{\Lambda}\right]_{jl} = \begin{cases} (\lambda_l - \lambda_j)\left[\mathbf{O}^{\mathrm{T}}d\mathbf{O}\right]_{jl}, & j<l, \\ d\lambda_j, & j=l, \end{cases} \tag{A.11}
$$

である事を用いると、結局、測度 $(d\mathbf{S})$ は

$$
(d\mathbf{S}) = \prod_{j<l}^{N} |\lambda_j - \lambda_l| \prod_{j=1}^{N} d\lambda_j \prod_{j<l}^{N} \left[\mathbf{O}^{\mathrm{T}}d\mathbf{O}\right]_{jl} \tag{A.12}
$$

となる。

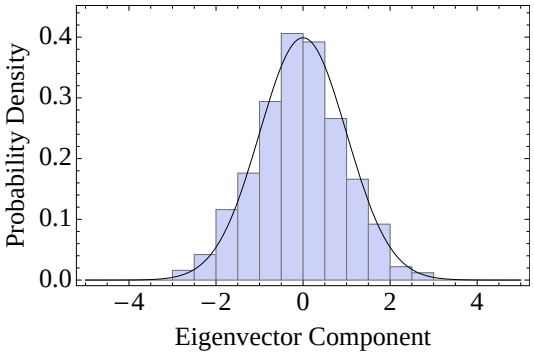

図 A1　Porter-Thomas 分布と数値計算の比較

A.2　Porter-Thomas 分布

次に，ランダム行列理論から得られる固有ベクトル成分分布の導出を行う。図 A1 は Porter-Thomas 分布の理論曲線と数値計算結果の比較である。両者はよく一致していることが分かる。

もしも，行列要素がランダムである場合固有ベクトル成分分布は Porter-Thomas 分布に従う。そのため，実データからなる相関行列に対して固有方程式を解いて得られた固有ベクトル成分分布を Porter-Thomas 分布と比較することでも有意な相関構造を抽出することができる。

A.2.1　Porter-Thomas 分布の導出

ここでも、[34] に沿って Porter-Thomas 分布の導出を行う。付録 A.1 から、ガウス型直交アンサンブルの固有ベクトルの分布関数は

$$P(\mathbf{O})\prod_{j<l}^{N}\left[\mathbf{O}^{\mathrm{T}}\mathrm{d}\mathbf{O}\right]_{jl}\propto\prod_{j<l}^{N}\left[\mathbf{O}^{\mathrm{T}}\mathrm{d}\mathbf{O}\right]_{jl} \tag{A.13}$$

と書けることがわかった。これは固有ベクトル $\mathbf{O}$ を用いて、$\left[\mathbf{O}^{\mathrm{T}}\mathbf{O}\right]_{jl}$ を独立変数としている。この節では、固有ベクトルの成分分布がどのような分布に従うのか明らかにしたい。そのためには、変数変換によって dO_{jl} を独立変数とする形に変えたい。そこで、直交行列による拘束条件

$$\left[O^{\mathrm{T}}O\right]_{jl}=\delta_{jl} \tag{A.14}$$

を用いる。これをデルタ関数に組み込むと、

$$\int\delta(\left[\mathbf{O}^{\mathrm{T}}\mathbf{O}\right]_{jl}-\delta_{jl})\mathrm{d}\left[\mathbf{O}^{\mathrm{T}}\mathbf{O}\right]_{jl}=1 \tag{A.15}$$

と書ける。すると、固有ベクトルの分布関数を改めて

$$
\begin{aligned}
&P(\mathbf{O})\prod_{j<l}^{N}\left[\mathbf{O}^{\mathrm{T}}\mathrm{d}\mathbf{O}\right]_{jl} \\
&\propto \prod_{j\leq l}^{N}\delta(\left[\mathbf{O}^{\mathrm{T}}\mathbf{O}\right]_{jl}-\delta_{jl})\prod_{j<l}^{N}\left[\mathbf{O}^{\mathrm{T}}\mathrm{d}\mathbf{O}\right]_{jl}\prod_{j\leq l}^{N}\mathrm{d}\left[\mathbf{O}^{\mathrm{T}}\mathbf{O}\right]_{jl}
\end{aligned}
\tag{A.16}
$$

と書き換えられる。ところが、積分変数 $\left[\mathbf{O}^{\mathrm{T}}\mathrm{d}\mathbf{O}\right]_{jl}$ と $\mathrm{d}\left[\mathbf{O}^{\mathrm{T}}\mathbf{O}\right]_{jl}$ の外積を考えると

$$
\begin{aligned}
&\left[\mathbf{O}^{\mathrm{T}}\mathrm{d}\mathbf{O}\right]_{jl}\wedge\mathrm{d}\left[\mathbf{O}^{\mathrm{T}}\mathbf{O}\right]_{jl} \\
&=\left[\mathbf{O}^{\mathrm{T}}\mathrm{d}\mathbf{O}\right]_{jl}\wedge\left[\mathrm{d}\mathbf{O}^{\mathrm{T}}\mathbf{O}+\mathbf{O}^{\mathrm{T}}\mathrm{d}\mathbf{O}\right]_{jl} \\
&=\left[\mathbf{O}^{\mathrm{T}}\mathrm{d}\mathbf{O}\right]_{jl}\wedge\left[\mathrm{d}\mathbf{O}^{\mathrm{T}}\mathbf{O}\right]_{jl} \\
&=\left[\mathbf{O}^{\mathrm{T}}\mathrm{d}\mathbf{O}\right]_{jl}\wedge\left[\mathbf{O}^{\mathrm{T}}d\mathbf{O}\right]_{lj}
\end{aligned}
\tag{A.17}
$$

であるから、固有ベクトルの確率分布関数は

$$
P(\mathbf{O})\prod_{j=1}^{N}\prod_{l=1}^{N}\left[\mathbf{O}^{\mathrm{T}}\mathrm{d}\mathbf{O}\right]_{jl}\propto\left(\prod_{j\leq l}^{N}\delta\left(\left[\mathbf{O}^{\mathrm{T}}\mathbf{O}\right]_{jl}-\delta_{jl}\right)\right)\prod_{j=1}^{N}\prod_{l=1}^{N}\left[\mathbf{O}^{\mathrm{T}}\mathrm{d}\mathbf{O}\right]_{jl}
\tag{A.18}
$$

と表すことができる。ここで、独立変数を $\prod_{j,l=1}^{N}\left[\mathbf{O}^{\mathrm{T}}\mathrm{d}\mathbf{O}\right]_{jl}$ から $\prod_{j,l=1}^{N}\mathrm{d}\mathbf{O}_{jl}$ へ変換することを考えると、$\mathbf{O}$ が直交行列であるから、ヤコビアンの絶対値は 1 になる。したがって、固有ベクトルの分布関数は、

$$
P(\mathbf{O})\prod_{j=1}^{N}\prod_{l=1}^{N}\left[\mathbf{O}^{\mathrm{T}}\mathrm{d}\mathbf{O}\right]_{jl}\propto\left(\prod_{j\leq l}^{N}\delta\left(\left[\mathbf{O}^{\mathrm{T}}\mathbf{O}\right]_{jl}-\delta_{jl}\right)\right)\prod_{j=1}^{N}\prod_{l=1}^{N}\mathrm{d}\mathbf{O}_{jl}
\tag{A.19}
$$

となる。

ランダム行列の固有ベクトルは、$\mathbf{O}$ の各列によって与えられる。そこで次に、$\mathbf{O}$ の第 1 列によって与えられる 1 つの固有ベクトルの分布関数について考える。そのためには、全固有ベクトルの分布関数から第 2 列以降の変数 $\mathbf{O}_{jl}$ $(l=2,3\ldots,N)$ を積分して消す必要がある。その際に、第 N 列から順に列番号の小さい方へ積分していくとき、各列に関係した因子を積分した結果、残りの列に依存するような因子が存在してはならない。そのような因子が存在しないことを証明するためにまず、ある n 列についての積分

$$
I_n=\int\prod_{j=1}^{n}\delta\left(\left[\mathbf{O}^{\mathrm{T}}\mathbf{O}\right]_{jn}-\delta_{jn}\right)\prod_{m=1}^{N}\mathrm{d}\mathbf{O}_{mn}
\tag{A.20}
$$

が、残りの列に依存しないことを示す。まず、

$$
\delta(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{e}^{-ikx}\mathrm{d}k
\tag{A.21}
$$

より、

$$I_n = \int \prod_{j=1}^{n} \left\{ \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{e}^{-ik_j\left(\left[\mathbf{O}^{\mathrm{T}}\mathbf{O}\right]_{jn} - \delta_{jn}\right)} \mathrm{d}k_j. \right\} \prod_{m=1}^{N} \mathrm{d}\mathbf{O}_{mn} \tag{A.22}$$

ここで、$j = n$ の時、

$$\mathrm{e}^{-ik_n\left(\left[\mathbf{O}^{\mathrm{T}}\mathbf{O}\right]_{nn} - \delta_{nn}\right)} = 1 \tag{A.23}$$

であるから、かわりに $j = n$ のとき

$$\mathrm{e}^{ik_n\left(\left[\mathbf{O}^{\mathrm{T}}\mathbf{O}\right]_{nn} - \delta_{nn}\right)} = \exp\left[ik_n \sum_{m=1}^{N} O_{mn}^2 - ik_n\right] \tag{A.24}$$

を用いる。すると式 (A.22) は

$$I_n = \frac{1}{(2\pi)^n} \int \prod_{j=1}^{n} \mathrm{d}k_j \mathrm{e}^{-ik_n} \prod_{m=1}^{N} \int \mathrm{d}O_{mn} \exp\left[-i \sum_{j=1}^{n-1} k_j O_{mj} O_{mn} + ik_n O_{mn}^2\right] \tag{A.25}$$

と書ける。$k_n \to k_n^+ = k_n + i\epsilon \quad (\epsilon > 0)$ として、後で $\epsilon \to 0$ の極限をとることにして、O_{mn} に対するガウス積分を計算すると

$$I_n = \lim_{\epsilon \to 0} \frac{1}{(2\pi)^n} \int \prod_{j=1}^{n} \mathrm{d}k_j \left(\frac{\pi}{\epsilon - ik_n}\right)^{N/2} \mathrm{e}^{-ik_n^+} \exp\left[-\frac{1}{4(\epsilon - ik_n)} \sum_{j=1}^{n-1} \sum_{l=1}^{n-1} k_j k_l \sum_{m=1}^{N} O_{mj} O_{ml}\right] \tag{A.26}$$

となる。ここで、まだ積分されていない列 $(j, l = 1, 2, \cdots, n-1)$ の正規直交性

$$\sum_{m=1}^{N} O_{mj} O_{ml} = \delta_{jl}, \tag{A.27}$$

を使うと、これらの列への依存性は消えて、

$$I_n = \lim_{\epsilon \to 0} \frac{1}{2} \pi^{(N-n-1)/2} \int_{-\infty}^{\infty} (\epsilon - is)^{(n-N-1)/2} \mathrm{e}^{-is} \mathrm{d}s \tag{A.28}$$

となる。後は最後に残った積分を計算して、$\epsilon \to 0$ の極限を考えると

$$I_n = \frac{\pi^{(N+1-n)/2}}{\Gamma((N+1-n)/2)} \tag{A.29}$$

となり、残りの列に依存しない因子のみが残る事が分かる。したがって、第 1 列 $x_j = O_{j1}$ の分布関数は

$$P_N(x_1, \ldots, x_N) \prod_{j=1}^{N} \mathrm{d}x_j = (1/I_1)\delta\left(\sum_{m=1}^{N} x_m^2 - 1\right) \prod_{j=1}^{N} \mathrm{d}x_j \tag{A.30}$$

となる。規格化定数は、全部の変数について積分したときに 1 になるように定めた。

この第 1 列の分布関数については、k 個の変数だけ残して残りの変数について積分することもできる。つまり、

$$
\begin{aligned}
P_k(x_1,\ldots,x_k) &\equiv \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\mathrm{d}x_{k+1}\cdots\mathrm{d}x_N P_N(x_1,\ldots,x_N) \\
&= \pi^{-k/2}\frac{\Gamma(N/2)}{\Gamma((N-k)/2)}\left(1-\sum_{m=1}^{k}x_m^2\right)^{(N-k-2)/2}
\end{aligned}
\tag{A.31}
$$

である。とくに、$k=1$ の場合、$\xi=\sqrt{N}x_1$ とおき、Stirling の公式を使って、行列サイズ無限大の極限 $(N\to\infty)$ をとると、ガウス分布

$$
P(\xi)\mathrm{d}\xi = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\xi^2/2}\mathrm{d}\xi \tag{A.32}
$$

が導かれる。これが解析的に導かれた Porter-Thomas 分布であり、その形状はガウス分布と一致することがわかった。

A.3 Wishart 型ランダム行列

付録 A.2 では、実対称なランダム行列の固有ベクトル成分分布の導出を行った。この章では, [36] に沿ってランダム行列からなる相関行列の固有値分布の導出を行う。

本章で扱うランダム相関行列は、Wishart 型のランダム行列と呼ばれ、多変量の共分散行列によって構成されるのが特徴である。まずは、Wishar 型ランダム行列の固有値分布について説明してから、導出を行う。

A.3.1 Wishart 型ランダム行列の固有値分布

図 A2 は、各行列要素が平均 0, 分散 1 のガウス分布に従う 1000×2000 のランダム行列を人工的に生成し、ランダム相関行列を対角化した結果、得られた固有値の分布を示している。この図か

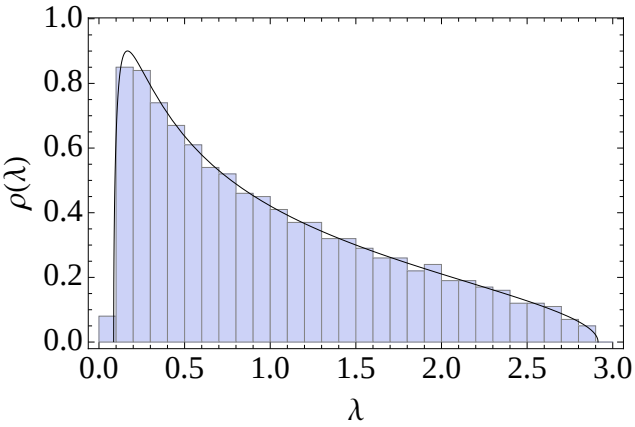

図 A2　固有値の確率密度と理論曲線

ら，行列要素はランダムであるにもかかわらず，固有値 λ の分布は何らかの規則に従っていることがわかる。

図 A2 に描かれた曲線は，$Q \equiv T/N$ を固定したまま N, T が無限大の極限で

$$\rho(\lambda) = \frac{Q}{2\pi\sigma^2}\frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda} \tag{A.33}$$

として得られる。ここで，σ^2 は確率変数の分散で，$\mathbf{R}$ の各成分が平均 0, 分散 1 のガウス分布に従う場合は $\sigma^2 = 1$ である。式 (A.33) において、分布の下限と上限である $\lambda_\pm$ は

$$\lambda_\pm = \sigma^2\left(1 \pm \sqrt{\frac{1}{Q}}\right) \tag{A.34}$$

となっていて、固有値 λ は $[\lambda_-, \lambda_+]$ の範囲にのみ存在する事がわかる。

A.3.2 ランダム相関行列に対する固有値分布の導出

いま、行列 $\mathbf{R} = [R_{ij}]$ が大きさ $N \times T$ の矩形行列で、各行列要素が独立同分布に従うものとする。そして、この $\mathbf{R}$ を用いてランダム相関行列を

$$\mathbf{C}_{\mathrm{rand}} = \mathbf{R}\mathbf{R}^{\mathrm{T}} \tag{A.35}$$

で定義する。この場合、$\mathbf{C}_{\mathrm{rand}}$ は大きさが $N \times N$ の正方行列となる。そして、この行列の固有値を λ_i $(i = 1, 2, \ldots, N)$ とすると、固有値の確率密度 $\rho(\lambda)$ は

$$\rho(\lambda) = \frac{1}{N}\sum_{i=1}^{N}\delta(\lambda - \lambda_i) \tag{A.36}$$

で与えられる。また、行列 $\mathbf{C}_{\mathrm{rand}}$ のレゾルベント $\mathbf{G}(\lambda)$ の成分は

$$G_{ij}(\lambda) = (\lambda\mathbf{1} - \mathbf{C}_{\mathrm{rand}})^{-1}_{ij} \tag{A.37}$$

で与えられ、$\mathbf{G}(\lambda)$ の対角成分の和は

$$\mathrm{Tr}\mathbf{G}(\lambda) = \sum_{i=1}^{N} G_{ii}(\lambda) = \sum_{i=1}^{N}\frac{1}{\lambda - \lambda_i} \tag{A.38}$$

となる。

ここで、デルタ関数を用いると式 (A.36) は

$$\begin{aligned}\rho(\lambda) &= \frac{1}{N}\sum_{i=1}^{N}\lim_{\epsilon\to 0}\frac{1}{\pi}\mathrm{Im}\frac{1}{\lambda - \lambda_i - i\epsilon} \\ &= \lim_{\epsilon\to 0}\frac{1}{N\pi}\mathrm{Im}\mathrm{Tr}\mathbf{G}(\lambda - i\epsilon)\end{aligned} \tag{A.39}$$

と変形できる。このことから、$\mathrm{Tr}\mathbf{G}(\lambda - i\epsilon)$ が計算できれば、固有値分布の形状を知ることができる。では、どのように $\mathrm{Tr}\mathbf{G}(\lambda - i\epsilon)$ を計算すればいいだろうか。式 (A.38) をさらに変形すると

$$\begin{aligned}\mathrm{Tr}\mathbf{G}(\lambda) &= \sum_{i=1}^{N} \frac{1}{\lambda - \lambda_i} = \frac{\partial}{\partial \lambda} \log \prod_i (\lambda - \lambda_i) \\ &= \frac{\partial}{\partial \lambda} \log \det(\lambda \mathbf{1} - \mathbf{C}_{\mathrm{rand}}) \equiv \frac{\partial}{\partial \lambda} Z(\lambda)\end{aligned} \tag{A.40}$$

と書き表すことができる。よって、固有値分布を求めるためには結局、$Z(\lambda) \equiv \log \det(\lambda \mathbf{1} - \mathbf{C}_{\mathrm{rand}})$ が計算できれば良いことになる。

いま、式 (A.40) において行列式をガウス積分を使って書き直すと

$$Z(\lambda) = \log \det(\lambda \mathbf{1} - \mathbf{C}_{\mathrm{rand}}) = -2 \log I(\lambda) \tag{A.41}$$

$$I((\lambda) \equiv \int_{-\infty}^{\infty} \exp\left(-\frac{\lambda}{2} \sum_{i=1}^{N} x_i^2 - \frac{1}{2} \sum_{i,j=1}^{N} \sum_{k=1}^{T} x_i x_j r_{ik} r_{jk} \right) \prod_{i=1}^{N} \left(\frac{dx_i}{\sqrt{2\pi}} \right) \tag{A.42}$$

と表すことができ、これを計算すれば良いことがわかる。ここで、$\mathrm{Tr}\mathbf{G}(\lambda)$ は $N \to \infty$ の極限で自己平均性を示すと考えられる。そのため、$\mathbf{C}_{\mathrm{rand}}$ のアンサンブル平均を計算すればよい。これは、レプリカ法を用いて式 (A.41) を

$$\langle Z(\lambda) \rangle = \langle -2 \log I(\lambda) \rangle = -2 \log \langle I(\lambda) \rangle \tag{A.43}$$

と計算できる。

いま、ランダム行列 $\mathbf{R}$ の $N \times T$ 個の要素 R_{ik} は、平均 0, 分散 σ^2/T のガウス分布に従っているとすると、式 (A.42) と式 (A.43) において、平均の計算に関係するのは

$$\left\langle \exp\left(\frac{1}{2} \sum_{i,j=1}^{N} \sum_{k=1}^{N} x_i x_j R_{ik} R_{jk} \right) \right\rangle \tag{A.44}$$

の部分である。この部分についての計算を行うと

$$\begin{aligned}&\left\langle \exp\left(\frac{1}{2} \sum_{i,j=1}^{N} \sum_{k=1}^{N} x_i x_j R_{ik} R_{jk} \right) \right\rangle \\ &= C \int_{-\infty}^{\infty} dR_{11} dR_{21} \cdots dR_{NT} \exp\left(-\alpha \sum_{m=1}^{N} \sum_{l=1}^{T} R_{ml}^2 \right) \exp\left(\frac{1}{2} \sum_{i,j=1}^{N} \sum_{k=1}^{T} x_i x_j R_{ik} R_{jk} \right) \\ &= C \int_{-\infty}^{\infty} dR_{11} dR_{21} \cdots dR_{NT} \exp\left(-\alpha \sum_{m=1}^{N} \sum_{l=1}^{T} R_{ml}^2 \right) \prod_{k=1}^{T} \exp\left(\frac{1}{2} \sum_{i,j=1}^{N} x_i x_j R_{ik} R_{jk} \right).\end{aligned} \tag{A.45}$$

ただし、C は積分定数を意味している。ここで、R_{ij} は全て独立同分布なので、式 (A.45) の積分

は $\mathbf{R}$ の k 列について計算して、それを T 乗しても同じなので

$$
\begin{aligned}
\text{式}\,(A.45) &= C\left[\int_{-\infty}^{\infty} dR_{1k}dR_{2k}\cdots dR_{Nk}\exp\left(-\alpha\sum_{m=1}^{N}h_{mk}^2\right)\exp\left(\frac{1}{2}\sum_{i,j=1}^{N}x_ix_jR_{ik}R_{jk}\right)\right]^T \\
&= C\left[\int_{-\infty}^{\infty} dR_{1k}dR_{2k}\cdots dR_{Nk}\exp\left(-\alpha\sum_{i,j=1}^{N}\delta_{ij}R_{ik}R_{jk}\right)\exp\left(\frac{1}{2}\sum_{i,j=1}^{N}x_ix_jR_{ik}R_{jk}\right)\right]^T \\
&= C\left[\int_{-\infty}^{\infty}\prod_{m=1}^{N}dR_{mk}\exp\left\{-\sum_{i,j=1}^{N}\left(\alpha\delta_{ij}-\frac{1}{2}x_ix_j\right)R_{ik}R_{jk}\right\}\right]^T \qquad \text{(A.46)}
\end{aligned}
$$

と書ける。ここで $R_{ik}=\phi_i$ とおくと

$$
\begin{aligned}
\text{式}\,(A.46) &= C\left[\int_{-\infty}^{\infty}\prod_{m=1}^{N}d\phi_m\exp\left\{-\sum_{i,j=1}^{N}\left(\alpha\delta_{ij}-\frac{1}{2}x_ix_j\right)\phi_i\phi_j\right\}\right]^T \\
&= C\times\pi^{N/2}\times\left[\det\left(\alpha\delta_{ij}-\frac{1}{2}x_ix_j\right)\right]^{-T/2} \qquad \text{(A.47)}
\end{aligned}
$$

となる。ここで、

$$
\left[\det(\alpha\delta_{ij}-\frac{1}{2}x_ix_j)\right] \qquad \text{(A.48)}
$$

は

$$
\begin{aligned}
\left[\alpha\delta_{ij}-\frac{1}{2}x_ix_j\right] &= \begin{pmatrix} \alpha-\frac{1}{2}x_1^2 & -\frac{1}{2}x_1x_2 & \cdots & -\frac{1}{2}x_1x_N \\ -\frac{1}{2}x_2x_1 & \alpha-\frac{1}{2}x_2^2 & \cdots & -\frac{1}{2}x_2x_N \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{2}x_Nx_1 & -\frac{1}{2}x_Nx_2 & \cdots & \alpha-\frac{1}{2}x_N^2 \end{pmatrix} \\
&= \begin{pmatrix} \alpha & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \alpha \end{pmatrix} - \begin{pmatrix} \frac{1}{2}x_1^2 & \dots & \frac{1}{2}x_1x_N \\ \vdots & \ddots & \vdots \\ \frac{1}{2}x_Nx_1 & \dots & \frac{1}{2}x_N^2 \end{pmatrix} \qquad \text{(A.49)}
\end{aligned}
$$

より、式 (A.49) の第 1 項目と 2 項目をそれぞれ $\alpha, \mathbf{M}$ とおくと

$$
\mathbf{M} = \frac{1}{2}\begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}\begin{bmatrix} x_1 & \cdots & x_N \end{bmatrix} \qquad \text{(A.50)}
$$

と書けるため、$\mathbf{M}$ はランク 1 である事がわかる。したがって求める固有値は

$$
\begin{aligned}
\mathbf{M}\begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} &= \frac{1}{2}\begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}\begin{bmatrix} x_1 & \cdots & x_N \end{bmatrix}\begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} \\
&= \frac{1}{2}\sum_{i=1}^{N}x_i^2\begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} \qquad \text{(A.51)}
\end{aligned}
$$

なので、固有値は

$$\frac{1}{2}\sum_{i=1}^{N} x_i^2 \tag{A.52}$$

だから

$$\begin{aligned}\left[\det(\alpha\delta_{ij} - \frac{1}{2}x_i x_j)\right] &= \alpha^{N-1}\left(\alpha - \frac{1}{2}\sum_{i=1}^{N} x_i^2\right) \\ &= \alpha^N\left(1 - \frac{1}{2\alpha}\sum_{i=1}^{N} x_i^2\right)\end{aligned} \tag{A.53}$$

となる。ここで、いま R_{ij} は平均が 0 で分散が σ^2/T のガウス分布であるから

$$\alpha = \frac{T}{2\sigma^2} \tag{A.54}$$

なので

$$\begin{aligned}\left\langle \exp\left(\frac{1}{2}\sum_{i,j=1}^{N}\sum_{k=1}^{N} x_i x_j R_{ik} R_{jk}\right)\right\rangle &= C \times \left(\alpha^N\right)^{-T/2}\left(1 - \frac{\sigma^2}{T}\sum_{i=1}^{N} x_i^2\right)^{-T/2} \\ &= \left(1 - \frac{\sigma^2}{T}\sum_{i=1}^{N} x_i^2\right)^{-T/2}\end{aligned} \tag{A.55}$$

である。ここで

$$q \equiv \frac{\sigma^2}{T}\sum_{i=1}^{N} x_i^2 \tag{A.56}$$

を定義し、デルタ関数を

$$\delta\left(q - \frac{\sigma^2}{T}\sum_{i=1}^{N} x_i^2\right) \equiv \int \frac{1}{2\pi}\exp\left[i\zeta\left(q - \frac{\sigma^2}{T}\sum_{i=1}^{N} x_i^2\right)\right] d\zeta \tag{A.57}$$

で定義する。これを平均をとった後の式 (A.42) の両辺にかけて、q について積分すると

$$\begin{aligned}\langle\, 式\ (A.42)\rangle &= \frac{1}{2\pi}\int\int\int \exp\left(-\frac{\lambda}{2}\sum_{i=1}^{N} x_i^2\right)\exp\left[-\frac{T}{2}\log(1-q)\right] \\ &\quad \times\exp\left[-i\zeta\left(q - \frac{\sigma^2}{T}\sum_{i=1}^{N} x_i^2\right)\right]\prod_{i=1}^{N}\left(\frac{dx_i}{\sqrt{2\pi}}\right) dq d\zeta \\ &= \frac{1}{2\pi}\int\int\int \exp\left(-\frac{1}{2}\left(\lambda - \frac{2i\zeta}{T}\sigma^2\right)\sum_{i=1}^{N} x_i^2\right) \\ &\quad \times\exp\left[-\frac{T}{2}\log(1-q) - i\zeta q\right]\prod_{i=1}^{N}\left(\frac{dx_i}{\sqrt{2\pi}}\right) dq d\zeta\end{aligned} \tag{A.58}$$

となる。これを x_i について積分すると

$$
\begin{aligned}
\text{式}\,(A.58) &= \frac{1}{2\pi}\int\int\left(\lambda-\frac{2i\zeta}{T}\sigma^2\right)^{-N/2}\exp\left[-\frac{T}{2}\log(1-q)-i\zeta q\right]dqd\zeta \\
&= \frac{1}{2\pi}\int\int\exp\left[-\frac{N}{2}\left\{\log\left(\lambda-\frac{2i\zeta}{T}\sigma^2\right)+Q\log(1-q)+Qq\frac{2i\zeta}{T}\right\}\right]dqd\zeta
\end{aligned}
\tag{A.59}
$$

となる (ただし $Q\equiv T/N$)。ここで $z\equiv\frac{2i\zeta}{T}$ とおくと $\zeta:-\infty\to\infty$ より $z:-i\infty\to i\infty$ であり、$d\zeta=\frac{T}{2i}dz$ であるから結局、式 (A.59) は

$$
\langle I(\lambda)\rangle=\frac{T}{4i\pi}\int_{-i\infty}^{i\infty}\int_{-\infty}^{\infty}\exp\left[-\frac{N}{2}\left\{\log\left(\lambda-\frac{2i\zeta}{T}\sigma^2\right)+Q\log(1-q)+Qq\frac{2i\zeta}{T}\right\}\right]dqdz. \tag{A.60}
$$

よって

$$
Z(\lambda)=-2\log\left[\frac{T}{4i\pi}\int_{-i\infty}^{i\infty}\int_{-\infty}^{\infty}\exp\left[-\frac{N}{2}\left\{\log\left(\lambda-\frac{2i\zeta}{T}\sigma^2\right)+Q\log(1-q)+Qq\frac{2i\zeta}{T}\right\}\right]dqdz\right] \tag{A.61}
$$

となる。

いま、Q を固定したまま $T\to\infty$, $N\to\infty$ の極限をとると、式 (A.61) では鞍点 (Saddle Point) が支配的になる。ここで鞍点の条件は

$$
Qq=\frac{\sigma^2}{\lambda-\sigma^2 z} \tag{A.62}
$$

$$
z=\frac{1}{1-q} \tag{A.63}
$$

である。式 (A.62)、式 (A.63) を $q(\lambda)$ について解くと

$$
q(\lambda)=\frac{\sigma^2(1-Q)+Q\lambda\pm\sqrt{\left[\sigma^2(1-Q)+Q\lambda\right]^2-4\sigma^2Q\lambda}}{2Q\lambda} \tag{A.64}
$$

となる。式 (A.61) を積分して式 (A.40) を用いると

$$
G(\lambda)=\frac{N}{\lambda-\sigma^2 z(\lambda)}=\frac{NQq(\lambda)}{\sigma^2} \tag{A.65}
$$

となる。さらに式 (A.39) を用いると

$$
\begin{aligned}
\rho(\lambda) &= \frac{\sqrt{4\sigma^2Q\lambda-\left[\sigma^2(1-Q)+Q\lambda\right]^2}}{2\pi\lambda\sigma^2} \\
&= \frac{Q}{2\pi\sigma^2}\frac{\sqrt{(\lambda_+-\lambda)(\lambda-\lambda_-)}}{\lambda}
\end{aligned}
$$

となり式 (A.33) が求まった。

A.4 IPR と QOP の導出

ここでは、固有ベクトル成分分布が２次元のガウス分布に従う場合の IPR と QOP の平均値、分散の導出についてまとめておく。IPR と QOP を用いることで、２次元のガウス分布から動径方向と角度方向にそれぞれどれだけズレているかを評価する事が出来る。

A.4.1 IPR の平均値と分散

固有ベクトル成分は複素数であるから、$v_\alpha = v'_\alpha + iv''_\alpha \ \ (\alpha = 1, 2, \cdots, N)$ として IPR は

$$IPR = \frac{\sum_\alpha |v_\alpha|^4}{(\sum_\alpha |v_\alpha|^2)} = \frac{A}{B} \tag{A.66}$$

$$= \frac{\langle A\rangle + \Delta A}{\langle B\rangle + \Delta B} \tag{A.67}$$

$$= \frac{\langle A\rangle \left(1 + \frac{\Delta A}{\langle A\rangle}\right)}{\langle B\rangle \left(1 + \frac{\Delta B}{\langle B\rangle}\right)} \tag{A.68}$$

$$\simeq \frac{\langle A\rangle}{\langle B\rangle}\left(1 + \frac{\Delta A}{\langle A\rangle}\right)\left(1 - \frac{\Delta B}{\langle B\rangle}\right) \tag{A.69}$$

$$\simeq \frac{\langle A\rangle}{\langle B\rangle}\left(1 + \frac{\Delta A}{\langle A\rangle} - \frac{\Delta B}{\langle B\rangle}\right) \tag{A.70}$$

$$= \frac{\langle A\rangle}{\langle B\rangle} + \frac{\Delta A}{\langle B\rangle} - \frac{\langle A\rangle}{\langle B\rangle^2}\Delta B \tag{A.71}$$

と書ける。これから、平均値 $E[IPR]$ と分散 $V[IPR]$ は

$$E[IPR] = \langle IPR\rangle = \frac{\langle A\rangle}{\langle B\rangle} \tag{A.72}$$

$$V[IPR] = E[(\Delta IPR)^2] = \langle IPR\rangle^2 \left(\frac{\langle \Delta A\rangle^2}{\langle A\rangle^2} + \frac{\langle \Delta B\rangle^2}{\langle B\rangle^2} - 2\frac{\langle \Delta A \Delta B\rangle}{\langle A\rangle\langle B\rangle}\right) \tag{A.73}$$

となる。ただし ΔIPR は $\Delta IPR = IPR - \langle IPR\rangle = \frac{\langle A\rangle}{\langle B\rangle}\left(\frac{\Delta A}{\langle A\rangle} - \frac{\Delta B}{\langle B\rangle}\right)$ である。

$$\langle A\rangle = \langle \sum_\alpha |v_\alpha|^4\rangle = 8N \tag{A.74}$$

$$\langle B\rangle = \langle \left(\sum_\alpha |v_\alpha|^2\right)\rangle = 8N + N(N-1)\langle |v_\alpha|^2\rangle^2 \tag{A.75}$$

$$\langle v^2\rangle = 2 \tag{A.76}$$

より

$$\langle B\rangle = 4N^2 + 4N \tag{A.77}$$

であるから

$$E[IPR] = \frac{8N}{4N^2 + 4N} \tag{A.78}$$

となる。これは $N \to \infty$ の極限で

$$E[IPR] \simeq \frac{2}{N} \tag{A.79}$$

次に

$$\langle (\Delta A)^2 \rangle = \langle (A - \langle A \rangle)^2 \rangle = 320N \tag{A.80}$$
$$\langle B^2 \rangle = 16N^4 = 96N^3 + 176N^2 + 96N \tag{A.81}$$
$$\langle (\Delta B)^2 \rangle = 64N^3 + 160N^2 + 96N \tag{A.82}$$
$$\langle \Delta A \Delta B \rangle = 128N^2 + 192N \tag{A.83}$$

であるから、これらを式 (A.73) に代入して $N \to \infty$ の極限を取ると

$$V[IPR] = \left(\frac{2}{N}\right)^2 \times \frac{1}{N} \tag{A.84}$$

となる。

A.4.2 QOP の平均と分散

本研究で扱う四極子オーダーパラメーター (QOP) は

$$Q^2 = Q_{xx}^2 + 2Q_{xy}^2 + Q_{yy}^2 \tag{A.85}$$
$$\tag{A.86}$$

で定義される。本節では QOP の平均値と分散

$$E[Q^2] = \langle Q^2 \rangle \tag{A.87}$$
$$V[Q^2] = E[\left(Q^2 - \langle Q^2 \rangle\right)^2] \tag{A.88}$$
$$= \langle \left(Q^2\right)^2 \rangle = \langle Q^2 \rangle^2 \tag{A.89}$$

を導出する。

$$Q_{xx} = \sum_\alpha \left(\cos^2\theta_\alpha - \frac{1}{2}\right) \tag{A.90}$$
$$Q_{xy} = \sum_\alpha \cos\theta_\alpha \sin\theta_\alpha \tag{A.91}$$
$$Q_{yy} = \sum_\alpha \left(\sin^2\theta_\alpha - \frac{1}{2}\right) \tag{A.92}$$
$$\tag{A.93}$$

であるから

$$\langle Q_{xx} \rangle = 0 \tag{A.94}$$
$$\langle Q_{xy} \rangle = 0 \tag{A.95}$$
$$\langle Q_{yy} \rangle = 0 \tag{A.96}$$
$$\langle Q_{xx}^2 \rangle = \frac{1}{8}N \tag{A.97}$$
$$\langle Q_{xy}^2 \rangle = \frac{1}{8}N \tag{A.98}$$
$$\langle Q_{yy}^2 \rangle = \frac{1}{8}N \tag{A.99}$$

より

$$E[Q^2] = \frac{1}{2}N \tag{A.100}$$

次に

$$\langle Q_{xx}^4 \rangle = \frac{3}{128}N + \frac{3}{64}(N^2 - N) \tag{A.101}$$
$$\langle Q_{xy}^4 \rangle = \frac{3}{128}N + \frac{3}{64}(N^2 - N) \tag{A.102}$$
$$\langle Q_{yy}^4 \rangle = \frac{3}{128}N + \frac{3}{64}(N^2 - N) \tag{A.103}$$
$$\langle Q_{xx}^2 Q_{xy}^2 \rangle = \frac{1}{128}N + \frac{1}{64}(N^2 - N) \tag{A.104}$$
$$\langle Q_{xy}^2 Q_{yy}^2 \rangle = \frac{1}{128}N + \frac{1}{64}(N^2 - N) \tag{A.105}$$
$$\langle Q_{xx}^2 Q_{yy}^2 \rangle = \frac{3}{128}N + \frac{1}{64}(N^2 - N) \tag{A.106}$$
$$\tag{A.107}$$

より

$$V[Q^2] = \langle \left(Q^2\right)^2 \rangle = \langle Q^2 \rangle^2 \tag{A.108}$$
$$= \frac{1}{4}(N^2 - N) \tag{A.109}$$

ここで $N \to \infty$ の極限を考えると

$$V[Q^2] = \frac{1}{4}N^2 \tag{A.110}$$

付録 B　主成分分析

ここでは、主成分分析について簡単な説明を行う。主成分分析は、統計学において古くから使われている多変量解析手法の 1 つである。詳細については、[37] などを参照されたい。

まず、主成分分析について簡単に言葉で説明すると “多くの量的変数がある場合、データに内在する相関構造を考慮し、より低い次元の合成変数（主成分）で代表させることで、データが持つ情報をより解釈しやすくする” というものである。本研究で扱うデータは、基本的には株価時系列データであるため、以下では時系列データを扱うものとして主成分分析が持つ数学的な意味とその取扱について説明していく。

B.1　主成分の導入

時系列長 T の時系列データが N 個あるとする。これを $x_i(t)$ $(i = 1, 2, \cdots, N, t = 1, 2, \cdots, T)$ と表す。データが N 個というのは、株価であれば N 個の銘柄であるし、気象データ（気温、気圧など）であれば N 箇所の観測地点に対応している。一方で時系列長 T は、日次データであれば T 日、月次データであれば T ヶ月というように観測した時点の数を表す。

このような N 個の時系列データが存在する際に、それぞれのデータの間の関係性を知りたいという場合が多々ある。そこで役立つのが主成分分析である。主成分分析では、N 個の観測値の重み付き平均で表される合成変数、式 (B.1) を構成する。

$$\begin{cases} z_1 = w_{11}x_1 + w_{12}x_2 + \cdots + w_{1i}x_i + \cdots + w_{1N}x_N = \sum_{i=1}^{N} w_{1i}x_i \\ z_2 = w_{21}x_1 + w_{22}x_2 + \cdots + w_{2i}x_i + \cdots + w_{2N}x_N = \sum_{i=1}^{N} w_{2i}x_i \\ \vdots \\ z_N = w_{N1}x_1 + w_{N2}x_2 + \cdots + w_{Ni}x_i + \cdots + w_{NN}x_N = \sum_{i=1}^{N} w_{Ni}x_i \end{cases} \tag{B.1}$$

そしてこれを第 1 主成分 (z_1)、第 2 主成分 (z_2)、$\cdots$、第 n$(leqN)$ 主成分 (z_n) に要約する。ここで、重み w_{ij} は定数倍だけ任意性があるので、一意に定めるために

$$w_{\ell 1}^2 + w_{\ell 2}^2 + \cdots + w_{\ell i}^2 + \cdots + w_{\ell N}^2 = \sum_{i=1}^{N} w_{\ell i}^2 = 1 \tag{B.2}$$

と規格化されているものとする。

上述した z_ℓ $(\ell = 1, 2, \cdots, N)$ が求めるべき主成分であるが、これはいくつかの条件を満たしていなければならない。その条件とは

1. 第 1 主成分 z_1 における x_i の係数 $\{w_{1i}\}(i = 1, 2, \cdots, N)$ は、式 $(B.2)$ の条件のもとで z_1 の分散が最大になるように定める。
2. 第 2 主成分 z_2 の係数 $\{w_{2i}\}(i = 1, 2, \cdots, N)$ は、式 $(B.2)$ を満足し、かつ z_2 が z_1 と無相関になるという条件のもとで、z_2 の分散が最大になるよう定める。

3. 以下同様にして、第 ℓ 主成分の係数 $\{w_{\ell i}\}(\ell = 1, 2, \cdots, N)$ は z_ℓ が $z_1, z_2, \cdots, z_{\ell-1}$ と無相関になるという条件のもとで、z_ℓ の分散が最大になるように定める。

である。

B.2 主成分の導出

それでは、上述した主成分を導出するにあたって、z_1 の分散が最大になるとはどういうことであろうか。幾何学的に考えると、長さ T の N 個の時系列というのは、N 次元空間中に散布する T 個の点とみなせる。そしてある軸 $\mathbf{w}_1 = (w_{11}, w_{12}, \cdots, w_{1N})$ 上に T 個の点を射影した時、そのばらつきが最も大きいような $\mathbf{w}_1$ を選べば良い。いま、$w_{11}, w_{12}, \cdots, w_{1N}$ はわかっているものとして、$\mathbf{w}_1$ 上へ射影した点の分散は

$$V[z_1] = \sum_{t=1}^{T} \left\{ \sum_{i=1}^{N} w_{1i}(x_{ti} - \overline{x}_i) \right\}^2 /(T-1) \tag{B.3}$$

$$= \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{i'=1}^{N} w_{1i} w_{1i'} (x_{ti} - \overline{x}_i)(x_{ti'} - \overline{x}_{i'})/(T-1) \tag{B.4}$$

$$= \sum_{i=1}^{T} \sum_{i'=1}^{N} w_{1i} w_{1i'} \sum_{t=1}^{T} (x_{ti} - \overline{x}_i)(x_{ti'} - \overline{x}_{i'})/(T-1) \tag{B.5}$$

$$= \sum_{i} \sum_{i'} w_{1i} w_{1i'} V_{ii'} \tag{B.6}$$

となる。ここで

$$V_{ii'} = \sum_{t=1}^{T} \frac{(x_{ti} - \overline{x}_i)(x_{ti'} - \overline{x}_{i'})}{(T-1)} \tag{B.7}$$

は共分散と呼ばれる。条件 1 によれば、この値が式 $(B.2)$ のもとで最大にならなければいけない。Lagrange の未定乗数法を用いると

$$Q = \sum_{i=1}^{N} \sum_{i'=1}^{N} w_{1i} w_{1i'} V_{ii'} - \lambda \left(\sum_{i=1}^{N} w_{1i}^2 - 1 \right) \to \mathrm{Max} \tag{B.8}$$

を満たすように w_{1i} を決める。そこで、まずこの Q を各 w_{1i} で偏微分してゼロとおくと

$$\frac{\partial Q}{\partial w_{1i}} = 0 \to \sum_{i'} w_{1i'} V_{ii'} - \lambda w_{1i} = 0 \quad (i = 1, 2, \cdots, N) \tag{B.9}$$

を得る。直感的にわかりやすくするために、これを $i = 1, 2, \cdots, N$ について書き下すと

$$|\mathbf{V} - \lambda \mathbf{I}| = \begin{vmatrix} (V_{11} - \lambda) & V_{12} & \cdots & V_{1N} \\ V_{12} & V_{22} - \lambda & \cdots & V_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ V_{1N} & V_{2N} & \cdots & V_{NN} - \lambda \end{vmatrix} \tag{B.10}$$

と表せる。式 $(B.10)$ は対称行列であるから、直交行列で対角化できる。また、半正定値行列でもあるので、対角化によって得られる固有値は非負の実数となる。得られた固有値を大きさの順に並べて

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N \geq 0 \tag{B.11}$$

と書くことにする。

式 $(B.8)$ において Q を微分してゼロより

$$\sum_{i'} w_{\ell i'} V_{ii'} = \lambda_\ell w_{\ell i} \tag{B.12}$$

であるから、これを $V[z_1]$ に代入すると

$$V[z_1] = \sum_i \sum_{i'} w_{\ell i} w_{\ell i'} V_{ii'} = \sum_i w_{\ell i} \left(\lambda_\ell w_{\ell i}\right) \tag{B.13}$$

$$= \lambda_\ell \sum_i w_{\ell i}^2 \tag{B.14}$$

$$= \lambda_\ell. \tag{B.15}$$

つまり、最大にすべき z_1 の分散は λ_ℓ $(\ell = 1, 2, \cdots, N)$ のいずれかに等しくなる。したがって、λ_ℓ のうち最大である λ_1 をとり、λ_1 に対応する固有ベクトルを $\mathbf{w}_1$ とすればよい。第 2 主成分に対しても同様の手順で求めることができる。

B.3 理論的な説明の補足

$m = N$ とし、N 個の主成分を抽出すれば、これは N 次元空間の座標 $(x_1, x_2, \cdots, x_N)$ を直交変換によって、$(z_1, z_2, \cdots, z_N)$ に変えたことになる。これを行列表示すれば

$$\begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_N \end{pmatrix} = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1N} \\ w_{21} & w_{22} & \cdots & w_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ w_{N1} & w_{N2} & \cdots & w_{NN} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} \tag{B.16}$$

つまり

$$\mathbf{z} = \mathbf{W}\mathbf{x} \tag{B.17}$$

$$\sum_{i=1}^{N} w_{ki}^2 = 1 \tag{B.18}$$

$$\sum_{i=1}^{N} w_{ki} w_{k'i} = 0 \tag{B.19}$$

であるから、$\mathbf{W}$ は確かに直交行列である。これにさらに時系列長 T の各時点を対応させる。ここでは簡単のために、x_i および z_k の平均はいずれもゼロとする。すると

$$\mathbf{Z} = \begin{pmatrix} z_{11} & z_{12} & \cdots & z_{1T} \\ \vdots & \vdots & & \vdots \\ z_{N1} & z_{N2} & \cdots & z_{NT} \end{pmatrix}, \mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1T} \\ \vdots & \vdots & & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{NT} \end{pmatrix} \tag{B.20}$$

$$\mathbf{Z} = \mathbf{WX} \tag{B.21}$$

と書けるので，対角行列を Λ とすると

$$\Lambda = \mathbf{ZZ^T}/(T-1) = \mathbf{WXX^TW^T}/(T-1) = \mathbf{WCW^T} \tag{B.22}$$

となる。これは、相関行列 $\mathbf{C}$ の直交変換により、対角行列 $\mathbf{\Lambda}$ を得たことを意味する。これより

$$\mathrm{tr}\Lambda = \sum_{n=1}^{N} \lambda_n = \mathrm{tr}(\mathbf{WCW^T}) \tag{B.23}$$

$$= \mathrm{tr}(\mathbf{CWW^T}) = \mathrm{tr}(\mathbf{C}) \tag{B.24}$$

$$= \sum_{n=1}^{N} C_{nn} \tag{B.25}$$

が成り立つ。

B.4 主成分分析の結果

第 2 章で、動的相関構造と比較を行った静的相関構造の結果についてここに載せておく。図 B1 は実データの固有値分布と RMT の理論曲線を比較している。$Q = 1008/483 \simeq 2.09$ より $\lambda_+ \simeq 2.86$ であり、これより大きな固有値は 9 つある。したがって、RMT を判定基準とみなした

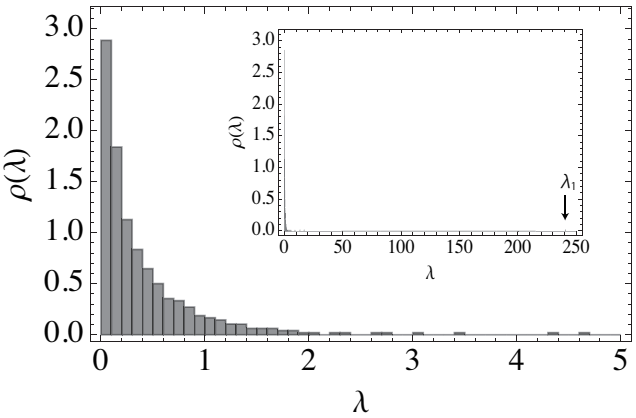

図 B1　相関行列の固有値分布

場合、この 9 つの固有値を有意な相関構造の現れと見なすことができる。また、このことから

$$\mathbf{C} = \lambda_1 \mathbf{V}_1 \mathbf{V}_1^{\mathrm{T}} + \sum_{\ell=2}^{9} \lambda_\ell \mathbf{V}_\ell \mathbf{V}_\ell^{\mathrm{T}} + \sum_{\ell=10}^{N} \lambda_\ell \mathbf{V}_\ell \mathbf{V}_\ell^{\mathrm{T}} \tag{B.26}$$

$$= \mathbf{C}_{\mathrm{market}} + \mathbf{C}_{\mathrm{group}} + \mathbf{C}_{\mathrm{rand}} \tag{B.27}$$

とスペクトル分解することができる。式 (B.27) において、特に 2 項目の $\mathbf{C}_{\mathrm{group}}$ を有意な相関構造によって再構築された相関行列とみなし、付録 C.1 で隣接行列として用いることで、ネットワーク解析を行う。

B.5　回転ランダムシャッフルを用いた帰無仮説

次に、回転ランダムシャッフルを用いた帰無仮説について説明する [14]。回転ランダムシャッフルを模式的に示すと下図のようになる。このようにシャッフルして作った時系列は、自己相関の情

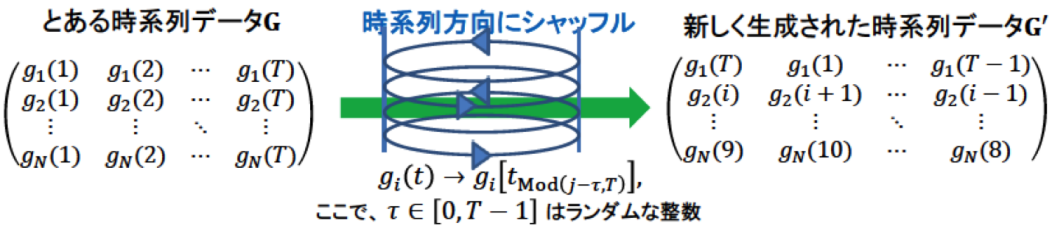

図 B2　回転ランダムシャッフルの模式図

報を保持しているが相互相関の情報は消失している。そのため、回転ランダムシャッフルした時系列を用いて作った相関行列の固有値分布を帰無仮説として用いることで、自己相関の影響を考慮した主成分の抽出を行うことができる。どの行をどれだけシフトするかによって固有値の分布が変化するため、複数回の試行を行い、$\langle \lambda_{\max} \rangle + 2\sigma$ を判定基準として用いる。ただし、$\langle \lambda_{\max} \rangle$ は回転ランダムシャッフルして得られた最大固有値の平均値、σ は最大固有値の標準偏差を意味している。1000 回の試行を行った所、$\langle \lambda_{\max} \rangle + 2\sigma = 3.55 + 0.36$ であり、これを越える固有値は全部で 7 つとなる。よって、これら 7 つの固有値はランダムでも自己相関による擬似相関でもない、真に有意な相関構造だとみなせる。以下では、これら 7 つの相関構造に着目して解析を行う。

ℓ 番目に大きな固有値に対応する固有ベクトルを $\mathbf{V}_\ell$ とし、$\mathbf{V}_1 \sim \mathbf{V}_7$ までを図 B3 にまとめる。図 B3(a) に示したように、このモードではすべての成分が正の値であり、ほぼ一様な値をとっていることがわかる。このことから、市場全体の動きを表すモードに対応しているため、マーケットモードと呼ばれる。一方で、$\mathbf{V}_2 \sim \mathbf{V}_7$ では、業種ごとに特徴的な振る舞いをしている事が見てわかる。

各固有モードにおいて、強く寄与する銘柄の上位 5 つを表としてまとめた。表 B1 では、固有ベクトル成分の値が正と負でそれぞれ絶対値が大きい上位 5 銘柄をリストアップしてある。正方向に寄与する銘柄と負方向に寄与する銘柄ではそれぞれが互いに反相関の関係にある。表 B1 から

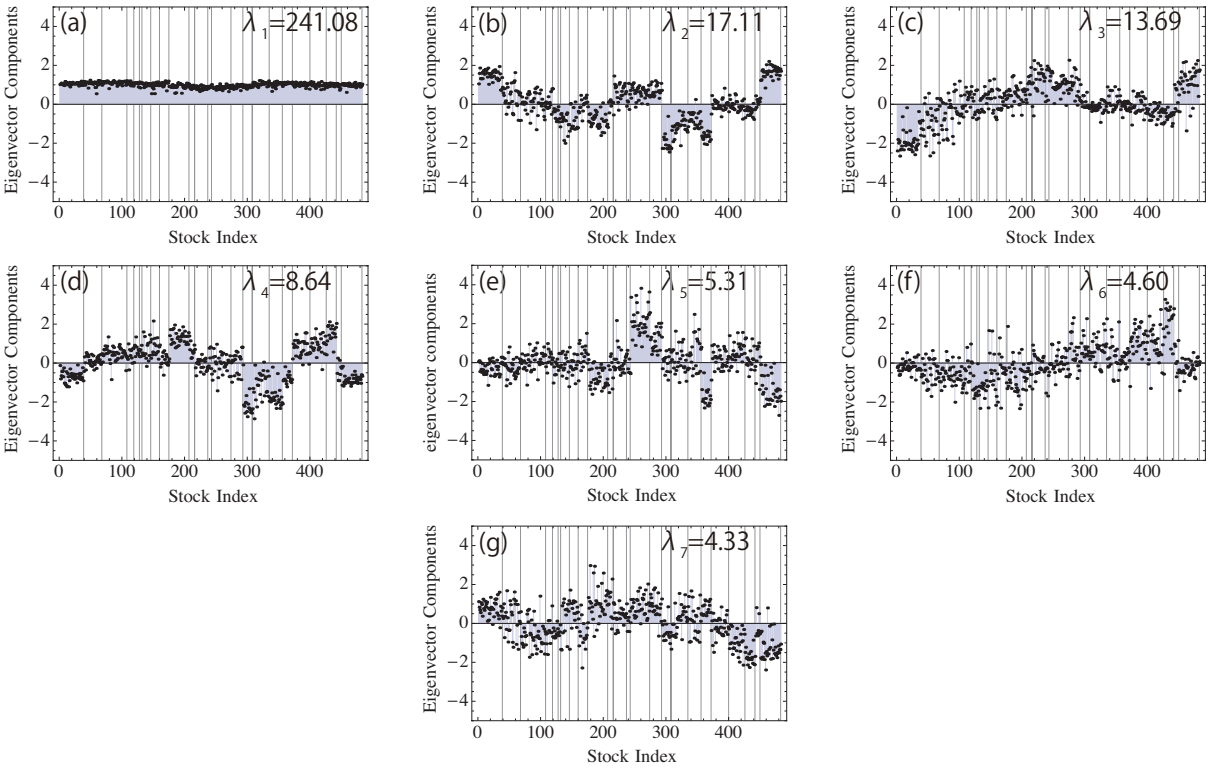

図 B3　最大固有値から 7 番目に大きな固有値に対応する固有ベクトル。横軸が株価のインデックス、縦軸が固有ベクトル成分の値になっている。

- モード 2：Utilities と Banks が強く寄与
- モード 3：Energy, Utilities と Food Bevverage & Tobacco が強く寄与
- モード 4：Finance（Diversified Financials, Banks と Insurance）や Semiconductors & Semiconductor Equipment と Retailing が寄与
- モード 5：Health Care Equipment & Services, Utilities と Real Estate が寄与

ということがわかる。

表 B1　$\mathbf{V}_2$ から $\mathbf{V}_7$ において、強く寄与する銘柄上位 5 つをリストアップした。表において $+(-)$ 記号はそれぞれ正 (負) の方向に強く寄与する事を意味している。それゆえ $+$ 列と $-$ 列は互いに反相関関係にある銘柄と言える。

mode2

+		−	
Ticker symbol	GICS Industry Group	Ticker symbol	GICS Industry Group
ETR	Utilities	STI	Banks
FE	Utilities	PNC	Banks
D	Utilities	WFC	Banks
NEE	Utilities	CMA	Banks
PCG	Utilities	MTB	Banks

mode3

+		−	
Ticker symbol	GICS Industry Group	Ticker symbol	GICS Industry Group
CNX	Energy	ABT	Pharmaceuticals, Biotechnology & Lifescience
FCX	Material	SO	Utilities
BTU	Energy	CPB	Food Beverage & Tobacco
MOS	Material	DUK	Utilities
NFX	Energy	ED	Utilities

mode4

+		−	
Ticker symbol	GICS Industry Group	Ticker symbol	GICS Industry Group
BAC	Diversified Financials	CMG	Consumer Service
KEY	Banks	BRCM	Semiconductors & Semiconductor Equipment
WFC	Banks	XLNX	Semiconductors & Semiconductor Equipment
FITB	Banks	DLTR	Retailing
STI	Banks	ADI	Semiconductors & Semiconductor Equipment

mode5

+		−	
Ticker symbol	GICS Industry Group	Ticker symbol	GICS Industry Group
SO	Health Care Equipment & Services	HUM	Utilities
EQR	Health Care Equipment & Services	WLP	Real Estate
PCG	Health Care Equipment & Services	CI	Utilities
ED	Health Care Equipment & Services	AET	Utilities
PSA	Health Care Equipment & Services	UNH	Real Estate

mode6

+		−	
ticker symbol	GICS Industry Group	Ticker symbol	GICS Industry Groups
JWN	Retailing	ALTR	Semiconductors & Semiconductor Equipment
BEAM	Consumer Durables & Apparel	BRCM	Semiconductors & Semiconductor Equipment
GCI	Media	MCHP	Semiconductors & Semiconductor Equipment
WHR	Consumer Durables & Apparel	ADI	Semiconductors & Semiconductor Equipment
WYN	Consumer Services	XLNX	Semiconductors & Semiconductor Equipment

mode7

+		−	
ticker symbol	GICS Industry Group	Ticker symbol	GICS Industry Groups
ALTR	Semiconductors & Semiconductor Equipment	AZO	Retailing
BRCM	Semiconductors & Semiconductor Equipment	FDO	Retailing
MCHP	Semiconductors & Semiconductor Equipment	DLTR	Retailing
ADI	Semiconductors & Semiconductor Equipment	ROST	Retailing
XLNX	Semiconductors & Semiconductor Equipment	WMT	Food & Staples Retailing

付録 C　株価相関ネットワーク

付録 B.4 の式 (B.27) で得られた $\mathbf{C}_{\mathrm{group}}$ を隣接行列として用いることで、株価相関ネットワークを構築し、相関構造の粗視化を行う。

C.1　株価相関ネットワーク

このようにして構築したネットワークは、ノードが株式銘柄、リンクの重みが相関係数に対応している。株式市場の相関構造をネットワーク的な手法を用いて粗視化する場合、興味が有るのは正の相関で強く結ばれた株式銘柄の集団 (コミュニティ) である。コミュニティを抽出するための有名な指標として、モジュラリティがある [27]。しかし株価相関ネットワークの特徴として、リンクが正と負の重みを持つという事がある。モジュラリティは正の重みを持つネットワークにしか対応しておらず、株価相関ネットワークには適用できない。そこで、社会学の分野で開発されたフラストレーションという指標を用いる [16]。正のリンクと負のリンクを対等に扱う場合、このフラストレーション $F(\{\sigma\})$ は

$$F(\{\sigma\}) = -\sum_{i,j} A_{ij}\delta(\sigma_i, \sigma_j) \tag{C.1}$$

のように書き表すことができる。式 (C.1) において、σ はコミュニティの分割、A_{ij} は隣接行列 (今の場合は $\mathbf{C}_{\mathrm{group}}$)、$\sigma_i$ はノード i が所属するコミュニティのインデックス、$\delta(\sigma_i, \sigma_j)$ はクロネッカーの δ である。つまり、ノード i とノード j の所属するコミュニティが等しい場合に対応する相関係数を足し上げていく事を意味する。そのため、$F(\{\sigma\})$ が最小となる分割を見つけ出すことで、正のリンクで強く結ばれたノード集団を抽出する事ができる。

そこで次に問題になってくるのは、どのようにしてそのようなコミュニティの分割を見つけ出すのか？ということである。もちろん、全ての分割パターンを確かめて、フラストレーションが最小となる分割を見つけ出すことが最も望ましい。しかし、ノードの数が増えれば増えるほど計算時間が膨大となり、実質計算不可能となる。そこで本研究では、アニーリング法 [38] を用いる。

C.2　コミュニティ抽出結果

図 C1 コミュニティ解析の結果得られた業種の所属コミュニティ一覧である。各コミュニティに所属するノードの数は Com.1:166,Com.2:157,Com3:93,Com4:67 である。第 3 章で得られたコミュニティ構造と同様に、業種単位での非常に強い結びつきがある。次に、第 3 章の $\theta_{\mathrm{th}} = 0.1$ で得られたコミュニティとの比較を行うため、規格化したジャッカード指数を計算する。そして、表 C1 はジャッカード指数を計算した結果である。いくつかコミュニティの分割や合併が起こっているため若干異なるものの、大きく異なっていない事がわかる。

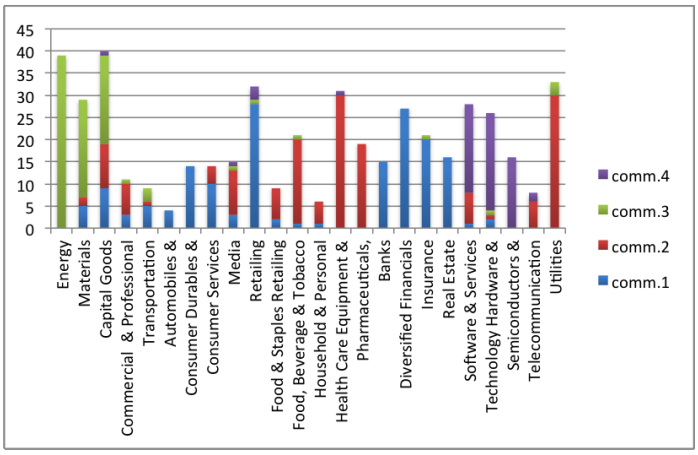

図 C1　コミュニティ解析結果

	com.1	com.2	com.3	com.4
com.1	0.26	0.04	0.15	0.98
com.2	0.00	1.00	0.00	0.00
com.3	0.98	0.00	0.01	0.00
com.4	0.00	0.01	0.95	0.01

表 C1　$\theta_{\mathrm{th}} = 0.1$ のコミュニティと $\mathbf{C}_{\mathrm{group}}$ のコミュニティのジャッカード指数

付録 D　ヒルベルト変換

D.1　連続ヒルベルト変換

信号 $x(t)$ のヒルベルト変換 $\tilde{x}(t)$ は以下の式で与えられる。

$$\tilde{x}(t) = -\frac{1}{\pi}\mathrm{P}\int \frac{x(z)}{t-z}dz \tag{D.1}$$

ここで、積分はコーシーの主値積分を意味する。$\tilde{x}(t)$ のフーリエ変換を考えると

$$\tilde{x}(\omega) = \int_{-\infty}^{\infty} \mathrm{e}^{i\omega t}\tilde{x}(t)dt = \mathrm{P}\int_{-\infty}^{\infty} \mathrm{e}^{i\omega t}\tilde{x}(t)dt \tag{D.2}$$

より

$$\tilde{x}(\omega) = \int_{-\infty}^{\infty} \mathrm{e}^{i\omega t}\left(-\frac{1}{\pi}\right)\mathrm{P}\int_{-\infty}^{\infty} \frac{x(z)}{t-z}dz \tag{D.3}$$

$$= -\frac{1}{\pi}\mathrm{P}\int_{-\infty}^{\infty} dz x(z)\mathrm{e}^{i\omega z}\mathrm{P}\int_{-\infty}^{\infty} dt \mathrm{e}^{i\omega(t-z)}\frac{1}{t-z} \tag{D.4}$$

$$= -\frac{1}{\pi}\mathrm{P}\int_{-\infty}^{\infty} dz x(z)\mathrm{e}^{i\omega z} \times \mathrm{P}\int_{-\infty}^{\infty} dt \mathrm{e}^{i\omega(t)}\frac{1}{t} \tag{D.5}$$

ここで、$x(z)$ は特異点を持っていないと仮定すれば、主値を取る必要はないので

$$\mathrm{P}\int_{-\infty}^{\infty} dz x(z)\mathrm{e}^{i\omega z} = \int_{-\infty}^{\infty} dz x(z)\mathrm{e}^{i\omega z} = x(\omega) \tag{D.6}$$

であるから

$$\tilde{x}(\omega) = -\frac{1}{\pi}x(z) \times \mathrm{P}\int_{-\infty}^{\infty} dt \mathrm{e}^{i\omega(t)}\frac{1}{t} \tag{D.7}$$

ここで

$$\mathrm{P}\int_{-\infty}^{\infty} dt \mathrm{e}^{i\omega(t)}\frac{1}{t} = \begin{cases} \pi i & (\omega > 0) \\ -\pi i & (\omega < 0) \end{cases} \tag{D.8}$$

より

$$\tilde{x}(\omega) = -ix(\omega)\mathrm{sgn}(\omega). \tag{D.9}$$

ここで、

$$\mathrm{sgn}(\omega) = \begin{cases} -1 & \omega < 0 \\ 0 & \omega = 0 \\ 1 & \omega > 0 \end{cases} \tag{D.10}$$

を使うと

$$\tilde{x}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega \mathrm{e}^{-i\omega t} y(\omega) \tag{D.11}$$
$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega \mathrm{e}^{-i\omega t}(-ix(\omega)\mathrm{sgn}(\omega)) \tag{D.12}$$
$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega \mathrm{e}^{-i\omega t}(-i)(x'(\omega)+ix''(\omega))\mathrm{sgn}(\omega) \tag{D.13}$$

ここで

$$\mathrm{e}^{-i\omega t}(-i)(x'(\omega)+ix''(\omega)) \tag{D.14}$$
$$= (\cos\omega t - i\sin\omega t)(-ix'(\omega)+x''(\omega)) \tag{D.15}$$
$$= x''(\omega)\cos\omega t - x'(\omega)\sin\omega t - ix'(\omega)\cos\omega t - ix''(\omega)\sin\omega t \tag{D.16}$$

より

$$\tilde{x}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\,(x''(\omega)\cos\omega t - x'(\omega)\sin\omega t$$
$$-ix'(\omega)\cos\omega t - ix''(\omega)\sin\omega t) \times \mathrm{sgn}(\omega) \tag{D.17}$$

であるが、$x(t)$ が実であれば $x^*(\omega) = x(-\omega)$ であるから (* は複素共役)、

$$x(\omega) = x'(\omega) + ix''(\omega) \tag{D.18}$$

とすれば

$$x^*(\omega) = x'(\omega) - ix''(\omega) \tag{D.19}$$
$$x(-\omega) = x'(-\omega) + ix''(-\omega) \tag{D.20}$$

より

$$x'(\omega) = x'(-\omega) \quad \rightarrow \quad x'(\omega) : \text{偶関数} \tag{D.21}$$
$$x''(\omega) = -x''(-\omega) \quad \rightarrow \quad x''(\omega) : \text{奇関数} \tag{D.22}$$
$$\tag{D.23}$$

である。さらに $\mathrm{sgn}(\omega)$ は奇関数であるから

$$x''(\omega)\cos\omega t\mathrm{sgn}(\omega) : \text{偶関数} \tag{D.24}$$
$$x'(\omega)\sin\omega t\mathrm{sgn}(\omega) : \text{偶関数} \tag{D.25}$$
$$x'(\omega)\cos\omega t\mathrm{sgn}(\omega) : \text{奇関数} \tag{D.26}$$
$$x''(\omega)\sin\omega t\mathrm{sgn}(\omega) : \text{奇関数} \tag{D.27}$$

となる。奇関数は $-\infty \rightarrow \infty$ で積分したら 0 になるので、結局 $x(t)$ のヒルベルト変換は

$$\tilde{x}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega(x''(\omega)\cos\omega t - x'(\omega)\sin\omega t) \times \mathrm{sgn}(\omega) \tag{D.28}$$

で与えられる。

D.2 フーリエ変換からヒルベルト変換を得る

ある連続な時系列データ $x(t)$ があったとき、$x(t)$ を $\pi/2$ シフトしたデータ $\tilde{x}(t)$ を求める方法としてヒルベルト変換を付録 D.1 で説明した。ここでは、$x(t)$ に対してフーリエ変換を行うことで $y(t)$ を求める方法について説明する。

まず、

$$x(\omega) = \int_{-\infty}^{\infty} x(t)\mathrm{e}^{i\omega t}dt \tag{D.29}$$

$$x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} x(\omega)\mathrm{e}^{-i\omega t} \tag{D.30}$$

を考える。$x(t)$ が実であれば $x^*(\omega) = x(-\omega)$ であるから (* は複素共役)、

$$x(\omega) = x'(\omega) + ix''(\omega) \tag{D.31}$$

とすれば

$$x^*(\omega) = x'(\omega) - ix''(\omega) \tag{D.32}$$

$$x(-\omega) = x'(-\omega) + ix''(-\omega) \tag{D.33}$$

より

$$x'(\omega) = x'(-\omega) \quad \rightarrow \quad x'(\omega) : \text{偶関数} \tag{D.34}$$

$$x''(\omega) = -x''(-\omega) \quad \rightarrow \quad x''(\omega) : \text{奇関数} \tag{D.35}$$

$$\tag{D.36}$$

である。これらを用いて、$x(t)$ を $x(\omega)$ のフーリエ変換で表すと

$$x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \left[x'(\omega) + ix''(\omega)\right]\mathrm{e}^{-i\omega t}d\omega \tag{D.37}$$

$$\begin{aligned} &= \frac{1}{2\pi}\int_{-\infty}^{\infty} x'(\omega)\cos\omega t d\omega - \frac{i}{2\pi}\int_{-\infty}^{\infty} x'(\omega)\sin\omega t d\omega \\ &\quad + \frac{i}{2\pi}\int_{-\infty}^{\infty} x''(\omega)\cos\omega t d\omega + \frac{1}{2\pi}\int_{-\infty}^{\infty} x''(\omega)\sin\omega t d\omega \end{aligned} \tag{D.38}$$

であり 2 項目と 3 項目は被積分関数が奇関数なので積分するとゼロになるから、結局

$$\begin{aligned} x(t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\left[x'(\omega)\cos\omega t + x''(\omega)\sin\omega t\right] \\ &= \frac{1}{\pi}\int_{0}^{\infty} d\omega\left[x'(\omega)\cos\omega t + x''(\omega)\sin\omega t\right] \end{aligned} \tag{D.39}$$

と変形できる。そして、$x(t)$ の虚数部として式 (D.39) を $\pi/2$ シフトしたものを選ぶので

$$\tilde{x}(t) = \frac{1}{\pi}\int_{0}^{\infty} d\omega\left[x''(\omega)\cos\omega t - x'(\omega)\sin\omega t\right] \tag{D.40}$$

を得る。ここで、$\omega \to -\omega$ を考えると

$$\begin{aligned}\tilde{x}(t) &= \frac{1}{\pi}\int_{-\infty}^{0} d\omega\,[x''(-\omega)\cos(-\omega t) - x'(-\omega)\sin(-\omega t)] \\ &= \frac{-1}{\pi}\int_{-\infty}^{0} d\omega\,[x''(\omega)\cos\omega t - x'(\omega)\sin\omega t] \end{aligned} \tag{D.41}$$

より $\omega < 0$ では $\omega > 0$ とは符号が逆になる。これより $-\infty \leq \omega \leq \infty$ では

$$\tilde{x}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dw\,[x''(\omega)\cos\omega t - x'(\omega)\sin\omega t] \times \mathrm{sgn}(\omega) \tag{D.42}$$

を得る。これはヒルベルト変換と一致する。

D.3 離散ヒルベルト変換

実際にプログラム上でヒルベルト変換を行う場合には、離散的なヒルベルト変換を行わなければならない。そこで、ここでは離散ヒルベルト変換についての説明を行う。

実時系列データ x_k のフーリエ変換と逆フーリエ変換を

$$x_n = \sum_{k=0}^{T-1} h_k \mathrm{e}^{i\frac{2\pi kn}{T}} \equiv x'_k + ix''_k \tag{D.43}$$

$$x_k = \frac{1}{T}\sum_{n=0}^{T-1} x_n \mathrm{e}^{-i\frac{2\pi kn}{T}} \tag{D.44}$$

と表すことにする。このとき

$$\begin{aligned} x_k &= \frac{1}{T}\sum_{n=0}^{T-1}(x'_n + ix''_n)\left(\cos\left(\frac{2\pi kn}{T}\right) - i\sin\left(\frac{2\pi kn}{T}\right)\right) \\ &= \frac{1}{T}\sum_{n=0}^{T-1} x'_n\cos\left(\frac{2\pi kn}{T}\right) + \frac{1}{T}\sum_{n=0}^{T-1} x''_n\sin\left(\frac{2\pi kn}{T}\right) \\ &\quad -\frac{i}{T}\sum_{n=0}^{T-1} x'_n\sin\left(\frac{2\pi kn}{T}\right) + \frac{i}{T}\sum_{n=0}^{T-1} x''_n\cos\left(\frac{2\pi kn}{T}\right) \end{aligned} \tag{D.45}$$

であるが、式 (D.45) において、第 3 項と第 4 項は足し合わせるとゼロになるので結局

$$x_k = \frac{1}{T}\sum_{n=0}^{T-1} x'_n\cos\left(\frac{2\pi kn}{T}\right) + \frac{1}{T}\sum_{n=0}^{T-1} x''_n\sin\left(\frac{2\pi kn}{T}\right) \tag{D.46}$$

となる。ここで、式 (D.42) を導出したときと同様にして考えると結局、x_k のヒルベルト変換は

$$y_k = \frac{1}{T}\sum_{n=0}^{T-1}\left[-x'_n\sin\left(\frac{2\pi kn}{T}\right) + x''_n\cos\left(\frac{2\pi kn}{T}\right)\right]\mathrm{sgn}\left(n - \frac{N}{2}\right) \tag{D.47}$$

として得られる。ただし

$$\mathrm{sgn}\left(n-\frac{T}{2}\right)=\begin{cases}1 & \left(n<\frac{T}{2}\right)\\ 0 & \left(n=\frac{T}{2}\right)\\ -1 & \left(n>\frac{T}{2}\right)\end{cases} \tag{D.48}$$

である。

ヒルベルト変換に基づく複素主成分分析を行う場合、式 (D.47) を用いて位相が $\pi/2$ シフトしたデータを作成する。

D.4 東証１部上場企業に対する解析

第 2 章では、S&P500 インデックス構成銘柄の 2008 年 ∼ 2011 年に対する結果を報告した。しかし、ここでは東証 1 部上場企業銘柄のうち 1996 年 ∼ 2006 年の全営業日 (2707 日) で取引が行われた 557 銘柄に対して解析した結果を報告する。図 D1 は相関行列 $\tilde{\mathbf{C}}$ の固有値分布を示してお

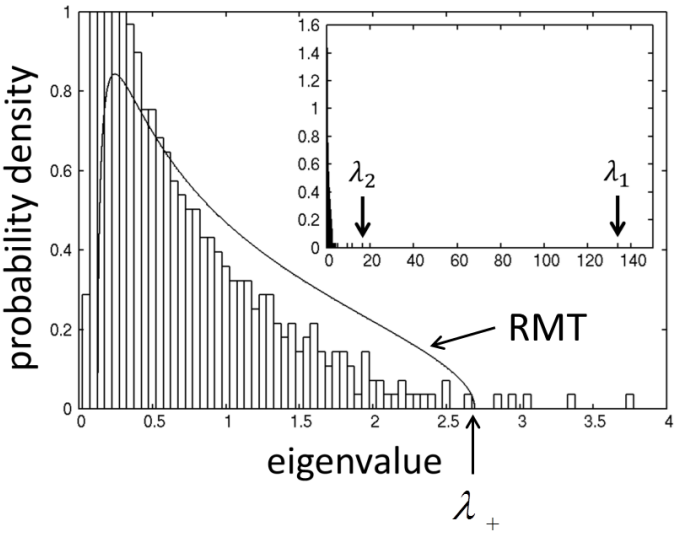

図 D1　固有値分布

り，$\lambda_+ \simeq 2.69$ よりも大きな固有値は全部で 11 個得られた。特に，最大固有値 λ_1 は 133.82 と非常に大きな値であり，これは λ_+ の約 50 倍もある。2 番目に大きな固有値 $\lambda_2 \simeq 16.33$ は λ_1 よりもかなり小さくなるが，λ_+ の約 6 倍近くの大きさがある。有意な相関構造の現れと考えられる固有値 $\lambda_1 \sim \lambda_{11}$ を足し合わせると

$$\sum_{i=1}^{11}\lambda_i \simeq 196.44, \tag{D.49}$$

$$\frac{196.44}{557} \simeq 0.35 \tag{D.50}$$

となり，全固有値の 35% になる。

図 D2 は λ_+ よりも大きな固有値に対応する固有ベクトル成分の分布をすべて表している。モード 1 の固有ベクトルを見ると，固有ベクトル成分のほとんどが一点に集中して分布していることがわかる。これは全ての株式銘柄からの寄与が等しいことを意味しているから，モード 1 はマーケットモードに対応していると言える。他のモードに関しては，モード 2～ モード 7 が帯状の分布，モード 8～ モード 11 が円状の分布をしていることがわかる。帯状の分布をしているモードでは相関・反相関関係が支配的だが，円状の分布をしているモードでは位相差を持った相関関係が存在していると考えられる。そこで、モード 8 の分布を詳しくみてみる。

図 D2 におけるモード 8 の固有ベクトル成分分布をみると，絶対値が大きく $\pi/2$ 近い位相差をもった成分のまとまりが見られる (図 D3)。この図から，電気・ガス業と医薬品業が大きく寄与していてお互いに $\pi/2$ の位相差があることがわかる。ここで抽出された電気・ガス業と医薬品業の株はディフェンシブ株 (景気動向に左右されず、景気が悪化しても余り下落しない銘柄) と呼ばれている。そのため、これらの業種が強く寄与するモードが抽出されることはごく自然なことであるが、実の主成分分析ではこれらの業種が両方とも強く寄与するモードは得られない。

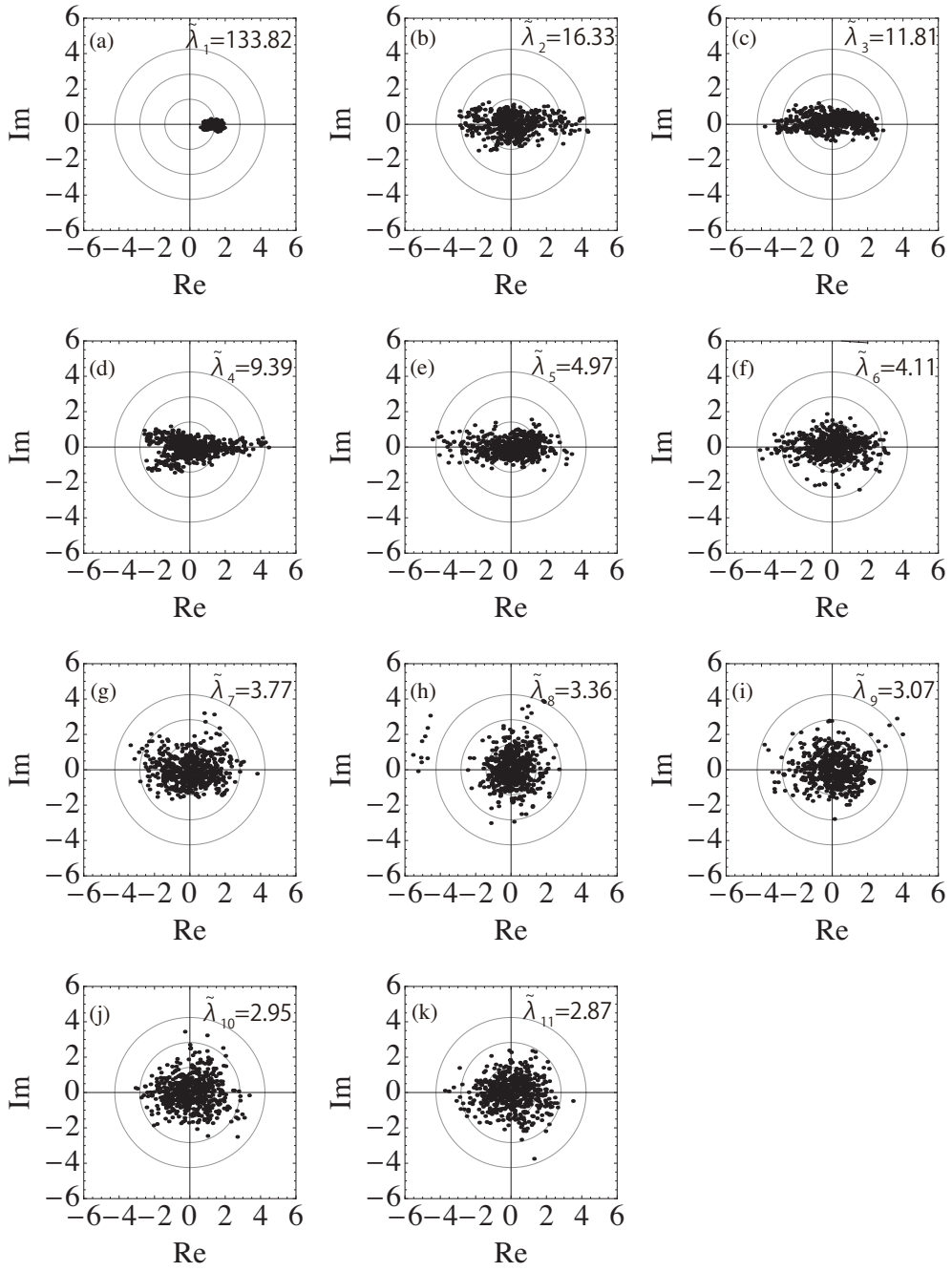

図 D2　固有ベクトル成分分布

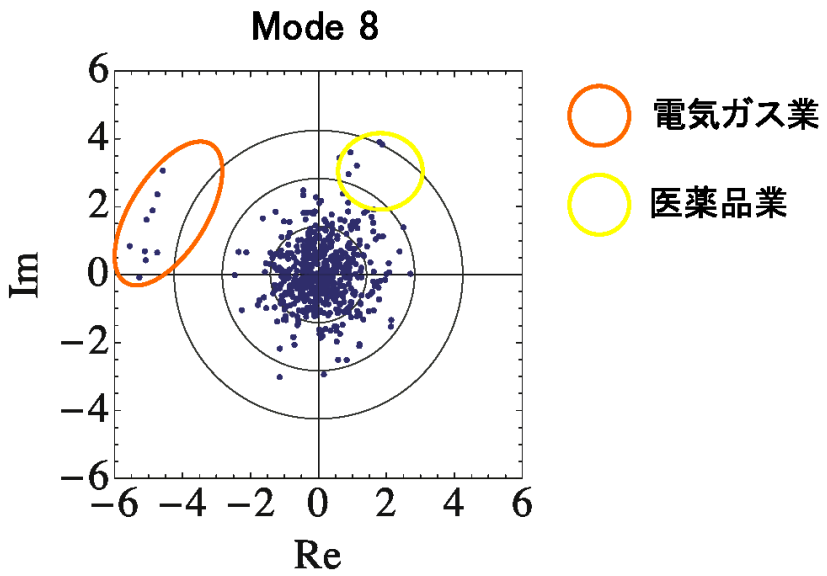

図 D3　モード 8 の固有ベクトル成分分布

付録 E　3 体相関

E.1　扱うデータ

第 4 章では S&P500 インデックスを構成する株式銘柄のうち 2008 年 ∼ 2011 年の全営業日で取引が行われた 483 銘柄を対象とした結果について報告した。ここでは、それ以外の市場及び期間では 3 体相関がどのような振る舞いをするのか調べたので、その結果について報告する。解析に用いたデータは

1. 東証 1 部上場企業のうち 1996 年 ∼ 2006 年の全営業日 (2707 日) で取引が行われた 557 銘柄
2. S&P500 インデックスに採用されている株式銘柄のうち 2001 年 ∼ 2011 年の全営業日 (2766 日) で取引が行われた 407 銘柄
3. Nikkei225 に含まれる銘柄のうち 2001 年 ∼ 2011 年の全営業日 (2699 日) で取引が行われた 180 銘柄

である。それぞれ順にデータ 1∼ データ 3 と名付ける。

E.2　結果

E.2.1　データ 1

$$\uparrow: 740624 \tag{E.1}$$
$$\downarrow: 766618 \tag{E.2}$$

$$\uparrow\uparrow: 119175448 \tag{E.3}$$
$$\uparrow\downarrow: 173436048 \tag{E.4}$$
$$\downarrow\downarrow: 126401780 \tag{E.5}$$

$$\uparrow\uparrow\uparrow: 14137568660 \tag{E.6}$$
$$\downarrow\downarrow\uparrow: 24398835660 \tag{E.7}$$
$$\uparrow\uparrow\downarrow: 23729667660 \tag{E.8}$$
$$\downarrow\downarrow\downarrow: 15251384080 \tag{E.9}$$

が実データから観測される各組み合わせの出現数である。第 4 章と同様に、上述した組み合わせの出現回数を、２項分布を用いて見積もっていく。

$$P_{\downarrow\downarrow} = P'_{\downarrow}P'_{\downarrow}(1+r) \tag{E.10}$$
$$P_{\uparrow\uparrow} = P'_{\uparrow}P'_{\uparrow}(1+s) \tag{E.11}$$
$$P_{\uparrow\downarrow} = P'_{\uparrow}P'_{\downarrow}(1+t) \tag{E.12}$$
$$P'_{\downarrow\downarrow\downarrow} = P'_{\downarrow}P'_{\downarrow}P'_{\downarrow}(1+3r+3r^2+\delta 1) \tag{E.13}$$
$$P'_{\uparrow\uparrow\uparrow} = P'_{\uparrow}P'_{\uparrow}P'_{\uparrow}(1+3s+3s^2+\delta 2) \tag{E.14}$$
$$P'_{\uparrow\uparrow\downarrow} = P'_{\uparrow}P'_{\uparrow}P'_{\downarrow}(1+2t+s+2ts+t^2+\delta 3) \tag{E.15}$$
$$P'_{\uparrow\uparrow\uparrow} = P'_{\downarrow}P'_{\downarrow}P'_{\uparrow}(1+2t+r+2tr+t^2+\delta 4) \tag{E.16}$$

とすると $r, s, t, \delta 1 \sim \delta 4$ は

$$r = 0.166 \tag{E.17}$$
$$s = 0.178 \tag{E.18}$$
$$t = -0.172 \tag{E.19}$$
$$\delta 1 = -0.086 \tag{E.20}$$
$$\delta 2 = -0.092 \tag{E.21}$$
$$\delta 3 = 0.028 \tag{E.22}$$
$$\delta 4 = 0.031 \tag{E.23}$$

である。

E.2.2 データ 2

$$\uparrow: 564530 \tag{E.24}$$
$$\downarrow: 560825 \tag{E.25}$$

$$\uparrow\uparrow: 71893560 \tag{E.26}$$
$$\uparrow\downarrow: 85412060 \tag{E.27}$$
$$\downarrow\downarrow: 71141445 \tag{E.28}$$

$$\uparrow\uparrow\uparrow: 6780236898 \tag{E.29}$$
$$\downarrow\downarrow\uparrow: 8519761044 \tag{E.30}$$
$$\uparrow\uparrow\downarrow: 8776181106 \tag{E.31}$$
$$\downarrow\downarrow\downarrow: 6764174727 \tag{E.32}$$

が実データから観測される各組み合わせの出現数である。以下、データ 1 の場合と同様に上述した

組み合わせの出現回数を、２項分布を用いて見積もっていく。

$$P_{\downarrow\downarrow} = P'_{\downarrow}P'_{\downarrow}(1+r) \tag{E.33}$$
$$P_{\uparrow\uparrow} = P'_{\uparrow}P'_{\uparrow}(1+s) \tag{E.34}$$
$$P_{\uparrow\downarrow} = P'_{\uparrow}P'_{\downarrow}(1+t) \tag{E.35}$$
$$P'_{\downarrow\downarrow\downarrow} = P'_{\downarrow}P'_{\downarrow}P'_{\downarrow}(1+3r+3r^2+\delta 1) \tag{E.36}$$
$$P'_{\uparrow\uparrow\uparrow} = P'_{\uparrow}P'_{\uparrow}P'_{\uparrow}(1+3s+3s^2+\delta 2) \tag{E.37}$$
$$P'_{\uparrow\uparrow\downarrow} = P'_{\uparrow}P'_{\uparrow}P'_{\downarrow}(1+2t+s+2ts+t^2+\delta 3) \tag{E.38}$$
$$P'_{\uparrow\uparrow\uparrow} = P'_{\downarrow}P'_{\downarrow}P'_{\uparrow}(1+2t+r+2tr+t^2+\delta 4) \tag{E.39}$$

とすると $r, s, t, \delta 1 \sim \delta 4$ は

$$r = 0.254 \tag{E.40}$$
$$s = 0.251 \tag{E.41}$$
$$t = -0.252 \tag{E.42}$$
$$\delta 1 = -0.183 \tag{E.43}$$
$$\delta 2 = -0.199 \tag{E.44}$$
$$\delta 3 = 0.073 \tag{E.45}$$
$$\delta 4 = 0.054 \tag{E.46}$$

である。

E.2.3 データ 3

$$\uparrow: 243094 \tag{E.47}$$
$$\downarrow: 242546 \tag{E.48}$$

$$\uparrow\uparrow: 13900298 \tag{E.49}$$
$$\uparrow\downarrow: 15713230 \tag{E.50}$$
$$\downarrow\downarrow: 13851252 \tag{E.51}$$

$$\uparrow\uparrow\uparrow: 591142038 \tag{E.52}$$
$$\downarrow\downarrow\uparrow: 697650540 \tag{E.53}$$
$$\uparrow\uparrow\downarrow: 700826930 \tag{E.54}$$
$$\downarrow\downarrow\downarrow: 589290772 \tag{E.55}$$
$$\tag{E.56}$$

が実データから観測される各組み合わせの出現数である。以下、データ 1、データ 2 の場合と同様

に上述した組み合わせの出現回数を、２項分布を用いて見積もっていく。

$$P_{\downarrow\downarrow} = P'_{\downarrow}P'_{\downarrow}(1+r) \tag{E.57}$$
$$P_{\uparrow\uparrow} = P'_{\uparrow}P'_{\uparrow}(1+s) \tag{E.58}$$
$$P_{\uparrow\downarrow} = P'_{\uparrow}P'_{\downarrow}(1+t) \tag{E.59}$$
$$P'_{\downarrow\downarrow\downarrow} = P'_{\downarrow}P'_{\downarrow}P'_{\downarrow}(1+3r+3r^2+\delta 1) \tag{E.60}$$
$$P'_{\uparrow\uparrow\uparrow} = P'_{\uparrow}P'_{\uparrow}P'_{\uparrow}(1+3s+3s^2+\delta 2) \tag{E.61}$$
$$P'_{\uparrow\uparrow\downarrow} = P'_{\uparrow}P'_{\uparrow}P'_{\downarrow}(1+2t+s+2ts+t^2+\delta 3) \tag{E.62}$$
$$P'_{\uparrow\uparrow\uparrow} = P'_{\downarrow}P'_{\downarrow}P'_{\uparrow}(1+2t+r+2tr+t^2+\delta 4) \tag{E.63}$$

とすると $r, s, t, \delta 1 \sim \delta 4$ は

$$r = 0.278 \tag{E.64}$$
$$s = 0.2761 \tag{E.65}$$
$$t = -0.278 \tag{E.66}$$
$$\delta 1 = -0.230 \tag{E.67}$$
$$\delta 2 = -0.231 \tag{E.68}$$
$$\delta 3 = 0.078 \tag{E.69}$$
$$\delta 4 = 0.076 \tag{E.70}$$

を得る。

以上のように、第 4 章の結果と同様に、異なる市場また異なる期間でも同様に３体相関として、すべての銘柄が同じ向きに価格変動するのを抑制する効果が観測されている。

謝辞

研究の遂行、及び本論文をまとめるにあたり、多くの方々にご協力して頂きました。この場を借りて感謝の意を述べさせて頂きたいと思います。

学生相談員としてご指導頂いた頃から数えると、9 年間にわたり家富洋教授には大変お世話になりました。本研究を進めるにあたり非常に多くのご助言を頂いただけでなく、研究者としての姿勢を学ぶことが出来ました。心より感謝致します。

また、新潟大学物性理論研究室の大野義章教授には、博士後期課程入学時から 1 年半の間、主指導教員としてお世話になりました。同じく物性理論研究室の奥西巧一准教授には、ランダム行列理論について学ぶ上で色々とご助言頂きました。それから、数理科学コースへ転コース後に副指導教員を引き受けてくださった、田中環教授、蛭川潤一准教授、並びに学位論文審査員を引き受けて下さった、山田修司教授、三浦毅教授にもお世話になりました。先生方のおかげで、研究に対する理解を深めることが出来ました。感謝致します。

新潟大学物性理論研究室の方々にも大変お世話になりました。特に、飯野隆史博士には研究に関する議論だけでなくプログラミングや計算機サーバーの管理など、多くの事に関してご指導頂きました。研究室の同期である、金山雅人氏、松浦裕貴氏、丸山大輔氏、吉川丈夫氏、吉田智大氏にも大変お世話になりました。皆さんのおかげで、辛くても挫けずに博士論文を完成させることが出来ただけでなく、非常に充実した研究生活を送ることが出来ました。また卒業された関川直毅氏、林健太氏、永井端女氏にも大変お世話になりました。たまに研究室へ遊びにきていただき、よい息抜きになりました。

また、後輩である吉川悠一氏、小船井博紀氏、伊藤亮人氏には研究に関する議論をさせて頂きました。私自身、自分の考えを整理し、研究に対する理解を深めることが出来ました。

最後に、私の学生生活を支えてくださった友人と祖父母、両親、妹のさゆりに深く心より感謝致します。

参考文献

[1] L. Laloux, P. Cizeau, J.-P. Bouchaud, and M. Potters, Phys. Rev. Lett. **83**, 1467 (1999).

[2] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. Nunes Amaral, and H. E. Stanley, Phys. Rev. Lett. **83**, 1471 (1999).

[3] M. S. Santhanam and Prabir K. Patra, Phys. Rev. E **64**, 016102 (2001).

[4] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, T. Guhr, and H. E. Stanley, Phys. Rev. E **65**, 066126 (2002).

[5] A. Utsugi, K. Ino, and M. Oshikawa, Phys. Rev. E **70**, 026110 (2004).

[6] D.-H. Kim and H. Jeong, Phys. Rev. E **72**, 046133 (2005).

[7] V. Kulkarni and N. Deo, Eur. Phys. J. B. **60**, 101 (2007).

[8] R. K. Pan and S. Sinha, Phys. Rev. E **76**, 046116 (2007).

[9] Y. Arai, K. Okunithi, and H. Iyetomi, in Intelligent Decision Technologies, SIST 10, edited by J. Watada et al. (Springer-Verlag Berlin Heidelberg) pp. 557–565 (2011).

[10] 新井優太、奥西巧一、家富洋，統計数理研究所共同研究リポート経済物理学とその周辺 (7) **259**，統計数理研究所，pp. 8–18 (2011).

[11] 新井優太、家富洋，統計数理研究所共同研究リポート経済物理学とその周辺 (8) **271**，統計数理研究所，pp. 85–89 (2012).

[12] Y. Arai and H. Iyetomi, Prog. Theor. Phys. Suppl. **194**, 84 (2012).

[13] Y. Arai and H. Iyetomi, in Intelligent Decision technologies, Vol. **2**, SIST 16, edited by J. Watada et al. (Springer-Verlag Berlin Heidelberg) pp. 415–422 (2012).

[14] H. Iyetomi, Y. Nakayama, H. Aoyama, Y. Fujiwara, Y. Ikeda, and W. Souma, Phys. Rev. E **83**, 016103 (2011).

[15] G. Bonanno, G. Caldarelli, F. Lillo and R. N. Mantegna, Phys. Rev. E **68**, 046130 (2013).

[16] V. A. Traag and J. Bruggeman, Phys. Rev. E **80**, 036115 (2009).

[17] T. Yoshikawa, T. Iino and H. Iyetomi, Intelligent Decision Technologies, SIST 10, edited by J. Watada et al. (Springer-Verlag, Berlin), pp. 511–518 (2011).

[18] T. Yoshikawa, T. Iino and H. Iyetomi, Prog. Theor. Phys. Suppl. **194**, 55 (2012).

[19] T. Yoshikawa, T. Iino and H. Iyetomi, Intelligent Decision Technologies, Vol. **1**, SIST 15, edited by J. Watada et al. (Springer-Verlag, Berlin), pp. 481–486 (2012).

[20] T. Yoshikawa, Y. Arai and H. Iyetomi, Frontiers in Artificial Intelligence and Applications, Vol. **255**, pp. 104–110 (2013).

[21] Y. Nakayama and H. Iyetomi, Prog. Theor. Phys. Suppl. **179**, 60 (2009).

[22] T. P. Barnett, Monthley Weather Review **111**, 756 (1983).

[23] K. Stein, A. Timmemann and N. Schneider, Phys. Rev. Lett. **107**, 128501(2011).

[24] 新井優太、家富洋，統計数理研究所共同研究リポート経済物理学とその周辺 (9) **292**，統計数

理研究所，pp. 19–24 (2013).
[25] 新井優太、家富洋，統計数理研究所共同研究リポート経済物理学とその周辺 (10) **311**，統計数理研究所，pp. 31–38 (2014).
[26] Y. Arai, T. Yoshikawa and H. Iyetomi, Frontiers in Artificial Intelligence and Applications, Vol. **255**, pp. 111-119 (2013).
[27] M. E. J. Newman, Phys. Rev. E **69**, 066113 (2004).
[28] V. D. Blondel, J.-L. Guillaume, R. Lambiotte and E. Lefebvre, J. Stat. Mech. : Theory and Experiment 2008, P10008 (2008).
[29] M. Rosvall and C.T. Bergstrom, Proceedings of the National Academy of Sciences, Vol. **105**, No. 4, pp. 1118–1123 (2008).
[30] P. Denis, P. Carre, and C. Fernandez-Maloigne, Computer Vision and Image Understanding **107**, 74 (2007).
[31] L. Shi and B. Funt, Computer Vision and Image Understanding **107**, 88 (2007).
[32] D. Assefa, L. Mansinha, K. F. Tiampo, H. Rasmussen, and K. Abdella, Signal Processing **90**, 1825 (2010).
[33] F. nian Lang, J. liu Zhou, S. Cang, H. Yu, and Z. Shang, Expert Systems with Applications **39**, 12046 (2012).
[34] 永尾太郎, 『ランダム行列の基礎』(東京大学出版会, 2005).
[35] M. Mehta, *Random Matrices* (Academic Press, New York, 1995).
[36] 青山秀明, 家富洋, 池田裕一, 相馬亘, 藤原義久, 『経済物理学』(共立出版, 2008).
[37] 奥野忠一, 芳賀敏郎, 久米均, 古澤正, 『多変量解析法〈改訂版〉』 (日科技連出版社, 1999).
[38] S. Kirkpatrick, C. D. Gelatt and M. P. Vecchi, Science New Series, Vol. **220**, 671 (1987).