

RMT 公式を用いた主成分抽出法による 日本及び米国株価の年次トレンドの比較

木戸 丈剛^{†1} 楊 欣^{†1}
田中 美栄子^{†1} 高石 哲弥^{†2}

RMT(ランダム行列理論)を利用して株式市場のトレンド抽出を行う手法を TOPIX500 銘柄の株価に適用し, 2007, 2008, 2009 年の日中データから年次トレンドの変遷を追った. また, 日本経済に多大な影響があるとされているアメリカの市場との比較を行った.

Comparison of yearly trend of Japanese and American stock prices by means of principal component extraction using RMT formula

TAKEMASA KIDO,^{†1} YANG XIN,^{†1}
MIEKO TANAKA-YAMAWAKI^{†1} and TAKAISHI TETUYA^{†2}

In this paper, we apply a method of extracting trends of the stock markets using RMT (random matrix theory), on intraday price data of TOPIX500 in the years of 2007, 2008, and 2009. Also, we compare our result with that of American market, which is closely related to Japanese economy.

^{†1} 鳥取大学大学院工学研究科エレクトロニクス専攻

Tottori University, Graduate School of Engineering, Department of Information and Electronics

^{†2} 広島経済大学

Hiroshima University of Economics

1. はじめに

ランダム行列理論 (Random Matrix Theory:RMT と略) より導かれる公式を株価時系列の相関行列の固有値分布と比較することで特異な成分を分離する方法¹⁾ が注目を集めている. 我々はこれを, ランダム行列理論を応用した主成分分析手法 (RMT oriented Principal Component Analysis:RMT-PCA と略) と名付け, 日次データを使い, 長期的なデータに対して行った²⁾. 本研究ではこの方法を短期的なトレンドの抽出に適用した. 具体的には 2007 年から 2009 年の TOPIX500 の 500 銘柄の 3 年分の日中データを用い, 1 年毎のデータを分析し, その変遷を追った.

2. RMT-PCA 手法に基づく特異な固有値の分離

RMT-PCA 手法の概略は以下のとおりである.

各銘柄に対して 1 時間毎の株価の対数差を正規化して作成した N (銘柄数) \times L (取引日数-1) の時系列行列 G から $C = GG^T$ とする $N \times N$ の自己相関行列 C を作成する. このとき, 自己相関行列 C の各数値は行や列の会社同士の相関を表す値になり自己相関である対角成分は 1 となる.

相関行列 C の固有値を求め, ランダム時系列の固有値分布に対する結果と比較することにより特異な固有値を取り出す. このとき行列 C がランダム時系列よりつくられた場合の固有値の出現範囲と出現頻度はランダム行列理論 (RMT) によって次式で求められることがわかっている. $N < L$ の条件で固有値分布は式 (1) で固有値の上限 (λ_+) と下限 (λ_-) は $Q = L/N$ の関数として式 (2) がわかっている.

$$P_m(\lambda) = \frac{Q}{2\pi} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda} \quad (1)$$

$$\lambda_{\pm} = 1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}}, \lambda_- < \lambda < \lambda_+ \quad (2)$$

式 (1) を図に表すと図 1 の様な形になる.

図 1 は $Q=3$ のときの具体例で, 横軸に固有値, 縦軸にその固有値の出現頻度を示しており, 疑似乱数と理論式 (1) はほぼ重なっている. これを実際の株価データと比較すると図 2 の様になり, 理論式 (破線) の中に収まらない固有値を主成分として分離することができる. 実データの最大固有値が大きく理論式と比較しにくいいため, 4 以上の固有値分布は図 2 の小窓に示す.

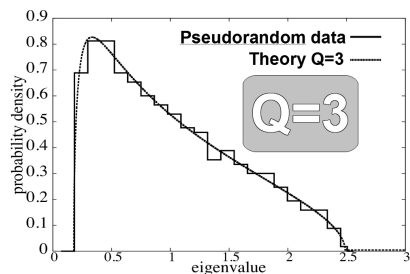


図 1 RMT 公式とランダムデータの固有値分布の比較:Q=3 の場合

Fig.1 Eigenvalue distribution computed from the random data, compered to the RMT formula:Q=3

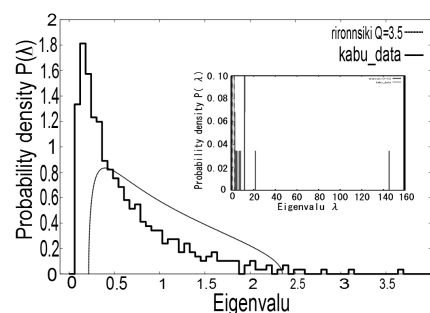


図 2 理論式 (1) と実データの固有値分布の比較

Fig.2 Eigenvalue distribution computed from the price data, compered to the RMT formula

図 2 より小窓には明らかに大きい実データの実線が複数存在している様子が見て取れる。RMT-PCA ではこの逸脱した固有値を特異な固有値とし、主成分とする。この手法は固有値の累積加算が固有値全体の 8 割までを主成分とするような既存の主成分抽出手法に比べ、分析すべき主成分の数を大幅に減らすことができる。この様子を概念的に図 3 に示す。例として 2007 年データの場合、累積寄与率が 80% に達するまでの主成分数は 170 あるのに対し、RMT-PCA では主成分数を 16 にまで絞れた。

3. 固有ベクトルの分析

RMT-PCA により主成分とされた固有値の固有ベクトルに対して分析を行う。第 1 主成分の固有ベクトルの各成分の値を図 4 に示す。第 2 主成分の固有ベクトルの各成分の値を図 5 に示す。第 1 主成分の固有ベクトルには図 4 の様に偏りが表れないが、第 2 主成分以降は図 5 の様に各成分に偏りが生じる。そこで、第 2 主成分以下の固有ベクトルで偏りの大きい成分について分析する。本稿では固有ベクトルの成分のなかで顕著に大きなものから ± の符号毎に 20 成分に注目する。相関行列 C の中で、注目する 20 成分同士の要素をみると、他の要素に比べ相関の値の高いもの同士の集まりであった。相関関係にある銘柄を抽出した結果、特定の業種に関わる銘柄が集まれば、その業種の偏りがその期間のトレンドといえることができる。

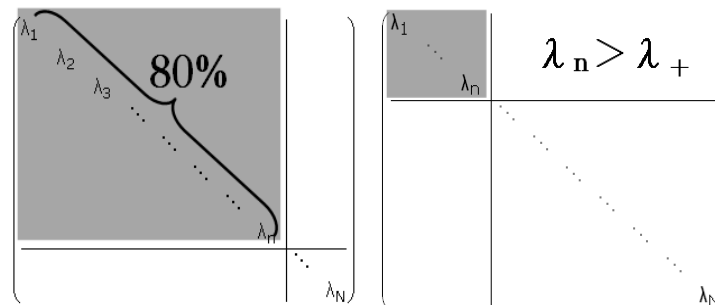


図 3 従来手法 (左) と RMT-PCA (右) の主成分数の比較

Fig.3 Comparison of the conventional method(left) and RMT-PCA(right)

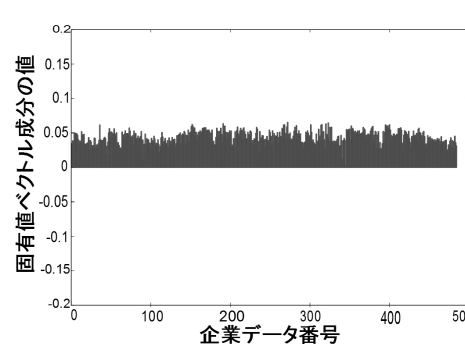


図 4 第 1 主成分の固有値ベクトル
Fig.4 First principal eigenvector

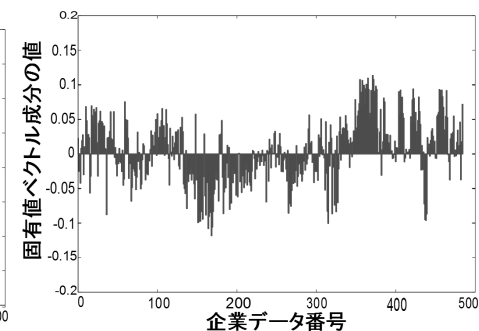


図 5 第 2 主成分以降の固有値ベクトル
Fig.5 Other than the first principal component eigenvector

4. 実データについての分析

東証データの 1997 年-2003 年の日時データを 1 データを対象に行った場合³⁾、特定の業種に偏ることがわかっている。しかし、6 年間もの長期に渡るトレンドでは実際の取引に使用は難しい。そこで日中データを使い 1 年毎のトレンドの追うことにした。

表 1 各年のデータ数

Table 1 Number of data per year

期間	N(銘柄数)	L(m時系列長)	Q(L/N)
2007年	485	1707	3.52
2008年	483	1707	3.53
2009年	483	1690	3.50

表 2 業種分類表

Table 2 Industry Classification TOPIX500

13:水産・農業	15:鉱業	16:鉱業(石油・ガス)
17:建設	20:食品	30:繊維・紙
40:科学・薬品	50:資源・素材	60:機械・電気
70:自転車・輸送機	80:金融・商業	90:運輸・通信・放送・ソフトウェア

本研究では2007年-2009年の最近の日中データについて1年毎に分析を行い、各年のトレンドの変遷を追うことにする。各年のデータについて表1に示す。RMT-PCAは同時刻相関を利用するため、今回は日中データを1時間毎に全銘柄に対して揃えた結果1700前後のデータが集まった。

図6に2007年、図7に2008年、図8に2009年の主成分の固有ベクトルの業種分類を行った結果を示す。

2007年-2009年の3年間を比較したが、各年度ともに80, 90に分類される企業が集まった。しかしその内訳を詳しく分析した結果、90に集まる企業は電力会社で変わらないのに対し、80番は同じ80番でも地方銀行、信託銀行、貸金業者それぞれが固まった。2007年データでは地方銀行、信託銀行、貸金業者それぞれが、2008年には地方銀行、信託銀行、2009年には地方銀行が2回と信託銀行その内訳が変化している。この3データに対し金融が目立つ原因はこの時期にリーマンショックが起こったためではないかと考えられる。また、東証の業種コードによる振り分けが新規上場銘柄は業種に関係なく20-40番台に振り分けられるため、業種が偏っていても目立たない場合があるが、大まかなトレンドの抽出には成功している。

5. 日本とアメリカのトレンド比較

日中データを利用することで短期的なトレンドも追うことができることが4章でわかった。では日本に多大な影響があるとされるアメリカと同期間で比較した場合はどうなるのか

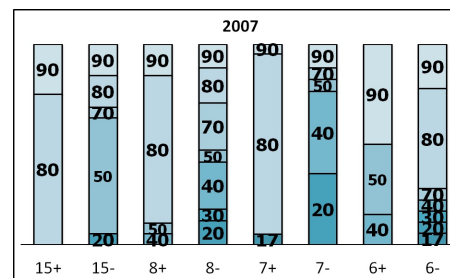


図 6 2007年のトレンド
Fig. 6 Trends of 2007 years

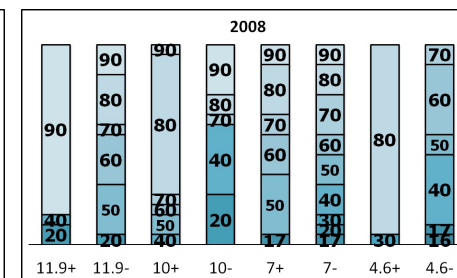


図 7 2008年のトレンド
Fig. 7 Trends of 2008 years

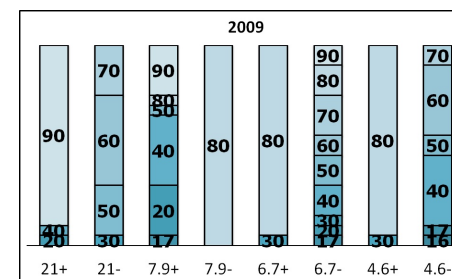


図 8 2009年のトレンド
Fig. 8 Trends of 2009 years

を検討する。TOPIX500と比較するS&P500の1日の終値データが手元にあったので、日本の株も1日の終値を1データとした2007年-2008年、2008年-2009年の2期間の日本株のトレンドとアメリカ株のトレンドについて比較を行った。

2007-2008年のデータを元にした業種分類をTOPIX500は図9にS&P500は図10に図示する。

2008-2009年のデータを元にした業種分類をTOPIX500は図11にS&P500は図10に図示する。

このとき4データともN=約500, L=約500, Q=約1である。

アメリカのデータの方が業種の偏りがあるように見えるが、日本とアメリカ共に金融が目立つ。日本の90は電力会社と石油・ガスの会社も入っている、これに対しアメリカはJが電力会社、Aが石油・ガス関連の会社でありほぼ似通ったトレンドを示している。また、

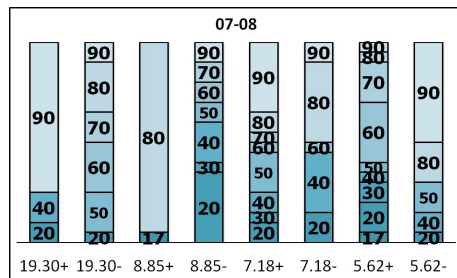


図 9 2007 年から 2008 年の日本株のトレンド
Fig. 9 Trends TOPIX500 from 2007 to 2008

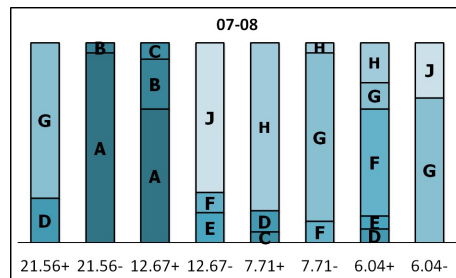


図 10 2007 から 2008 年のアメリカ株のトレンド
Fig. 10 Trends S&P500 from 2007 to 2008

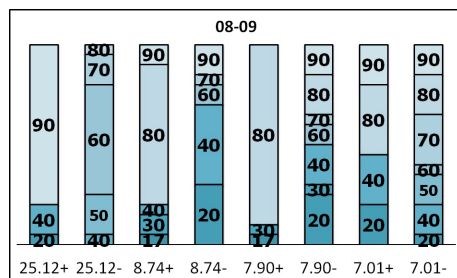


図 11 2008 年から 2009 年の日本株のトレンド
Fig. 11 Trends TOPIX500 from 2008 to 2009

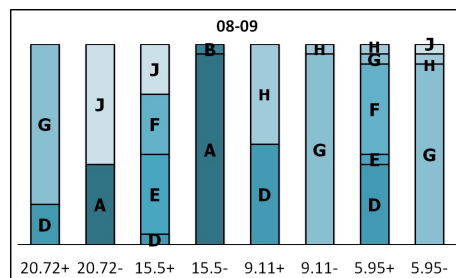


図 12 2008 年から 2009 年のアメリカ株のトレンド
Fig. 12 Trends S&P500 from 2008 to 2009

アメリカの H:情報技術は日本では 60 の電気・機械がメインとなるが 4 章で述べたとおり、新規上場銘柄は業種に関係なく 20-40 番台に振り分けられるため、情報のような歴史の浅い業種は TOPIX のコードだけでは読み取りにくいことがわかった。しかし、少なくとも金融やエネルギーなどの主となるトレンドはアメリカと日本間で似通っていることがわかり、それが RMT-PCA から読み取れることがわかった。

6. まとめ

長期的な場合は比較的容易に入手しやすい日次終値を使うことで、短期間の場合は日中データを使うことで RMT-PCA によりその期間のトレンドを抽出できることがわかった。しかし、抽出した業種を実際の取引の指標として使用する場合 1 年間でもまだ長いため、データを 1 時間毎よりも短くした場合などを試す必要がある。また、図 10・図 12 の様に

表 3 業種分類表

Table 3 Industry Classification TOPIX500 and S&P500

13:水産・農業	15:鉱業	16:鉱業(石油・ガス)
17:建設	20:食品	30:繊維・紙
40:科学・薬品	50:資源・素材	60:機械・電気
70:自転車・輸送機	80:金融・商業	90:運輸・通信・放送・ソフトウェア
A:エネルギー	B:素材	C:資本財
D:サービス	E:生活必需品	F:ヘルスケア
G:金融	H:情報技術	I:電気通信
J:公益事業		

S&P500 と比較した場合、TOPIX500 を元にした図 9・図 11 は、データによる誤差もあるが、コードによる業種分類では直感的に判断しにくい部分が増えるため今のままでは、詳しい分析が何かしらの工夫が必要などの課題が残った。

参考文献

- 1) V.Plerou, P.Gopikrishnan, B.Rosenow, L.A.N.Amaral and H.E.Stanley: Random matrix approach to cross correlations in financial data, *Physical Review E*, p.066126 (2002).
- 2) 木戸文剛: ランダム行列理論との比較による株価日中変動の相関行列解析, FIT2010: 第 9 回情報科学技術フォーラム講演論文集, pp.153-156 (2010).
- 3) 青山秀明, 他 4 名: 経済物理学 第 5 章 (2008).