

株式市場の主成分追跡 ～RMT-PCAによる四半期間の主要業種抽出～

山本 敦史^{1,a)} 木戸 丈剛¹ 田中 美栄子^{1,b)} 高石 哲也²

概要: ランダム行列理論利用の主成分分析手法 (RMT-PCA) を東証 tick 価格データから抽出した 30 分毎の取引価格データに適用することにより, 株式市場の 4 半期 (3 ヶ月) 毎の主要産業セクタを抽出し, その変動を追跡した. 2007 年から 2009 年にかけての 3 年間の変動を 4 半期毎に解析した結果は, 1 年毎の結果と比較して, 主要産業セクタの変遷を明確に表すと共に, サブプライムローン問題やリーマンショックの影響を如実に見ることができる. 更に主要業種変遷は 2007 年に有力であった鉄鋼・非鉄, および銀行から, 2008 年後半を境に電機・精密, および電力・ガスに移行するが, これを新たに「関連株分類」に直すことで 2007 年の市況関連株から 2008 年以降の外需関連株およびディフェンシブ株への移行と捉えなおすことができ, 2007 年から 2008 年にかけての原油価格の急高騰の影響や 2008 年 11 月に起きたトヨタショック等の影響をより明確に反映した結果が得られることがわかった.

キーワード: RMT-PCA, 主要業種, 株式市場

Tracking the Principal Components of Stock Market ～Extracting the Major Industrial Sectors per Quarter Year Based on RMT-PCA～

ATSUSHI YAMAMOTO^{1,a)} TAKEMASA KIDO¹ MIEKO TANAKA-YAMAWAKI^{1,b)} TETSUYA TAKAISHI²

Abstract: We applied the method of RMT-based PCA on the 30 minutes price time series extracted from the tick-wise traded prices of Tokyo Stock Market and track the quarterly changes of major industrial sectors from 2007 to 2009. Compared to the previous result per year, the result obtained in this analysis shows a change of the major industrial sectors much more clearly so that we can directly read out the effect of the sub-prime loan problem and the bankruptcy of Lehman Brothers in October 2008. Moreover, the change of major sectors from 2007 to 2009 can be described in a clear-cut manner by means of the new classification called "relation-based classification". For example, the change from ferrite and banks to the electro-machinery and utility be viewed as the change from the consumption-based industry to defensive/foreign trading based industry.

Keywords: RMT-PCA, Major sectors, Stock market

1. はじめに

ランダム行列の固有値分布がデータ長とデータ数の比のみで表される簡単な関数となることを利用し, 多量な時系列データの変動から相関の強い変動をするデータとランダムな変動をするデータに分離することができる. これは, 多量の時系列データが測定されている事象ならば

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

a) s082065@ike.tottori-u.ac.jp

b) mieko@ike.tottori-u.ac.jp

適用可能なため、金融工学や地震学等、生態学など様々な分野で研究されている [1][2][3][4][5][6]。我々はこの手法に RMT-PCA (Random Matrix Theory oriented Principal Component Analysis) と名付け、株式データへの適用を試みている [7][8][9]。これにより、大量の株価の数値データから株価が連動する銘柄を抽出できれば銘柄選択の指標になるのではないかと考えた。我々の研究室では TOPIX500 の構成銘柄である 500 銘柄の日中データを用い 1 年毎の主要業種の分析を行っている [7][8][9]。しかし、近年の投資の IT 化により短期間の投資を行う人が増えており、短期間の動向を知る需要が高まっている。そこで本稿では、同データを 3 ヶ月毎に分割し主要業種の抽出を行なった。

2. 相関行列の固有値分布

ランダムデータに対しては相関行列の固有値分布の理論式は

$$Q = \frac{L}{N} < 1, N \rightarrow \infty, L \rightarrow \infty \quad (1)$$

$$\lambda_{max} = \left(1 + \frac{1}{\sqrt{Q}}\right)^2, \lambda_{min} = \left(1 - \frac{1}{\sqrt{Q}}\right)^2 \quad (2)$$

$$P_{rmt}(\lambda) = \frac{Q}{2\pi\lambda} \sqrt{(\lambda_{max} - \lambda)(\lambda - \lambda_{min})} \quad (3)$$

で与えられる。いま N 行 L 列のランダム行列があるとすると、 L と N の比率 $Q = L/N$ より最大固有値： λ_{max} と最小固有値： λ_{min} を求めることができる。つまり、固有値分布は Q のみに依存する。実データはランダムではないので式 (3) に一致せず、実際に求めた相関行列の固有値はこの式の範囲から飛び出してしまふ。飛び出した固有値に相当する自由度が、ランダム成分ではない主成分ということになる。

3. 主要業種の抽出方法

3.1 株価データの編集

今回は 2007 年 1 月 4 日 ~ 2009 年 12 月 31 日までの期間の 2009 年の TOPIX500 構成銘柄 (平成 21 年 10 月 30 日現在) から各銘柄 30 分毎の株価をデータから抽出する。ランダム行列理論を用いるには同時刻にすべての銘柄の価格データが必要となる。本研究では過去研究 [7][8][9] 同様、取引が行われなかった時刻を現状維持と考え、一つ前の時系列の株価と同じ値を与えることで補正を行っている。

3.2 データの正規化

まず、各銘柄 30 分毎の株価の変動を調べる。銘柄により株価に大きな差が存在するためここでは対数差を算出し変動の割合に着目する。さらに

$$g_{i,j} = \frac{X_{i,j} - \langle X_i \rangle}{\sqrt{\langle X_i^2 \rangle - \langle X_i \rangle^2}} \quad (4)$$

より求めた対数差の平均値が 0、分散が 1 となるように正規化する。

3.3 相関行列の算出

正規化した値より N 行 L 列の行列 G を作成する。

$$G = \begin{bmatrix} g_{1,1} & \cdots & g_{1,L} \\ \vdots & \ddots & \vdots \\ g_{N,1} & \cdots & g_{N,L} \end{bmatrix} \quad (5)$$

N は銘柄数、 L は時系列長を示す。作成した行列の各銘柄との同時刻の内積を取り相関行列を算出する。

$$C = \frac{1}{L} G^T G \quad (6)$$

相関行列の要素は各銘柄毎の相関係数となる。相関係数は各銘柄による時系列の変動の類似度であり、 $1 \sim -1$ の値で表される。そして相互相関行列 C の固有値を求め、ランダム行列理論の理論式 (3) による固有値分布と比較する。

3.4 固有値の比較による主成分の分離

図 1 は擬似乱数で作成した $N = 486, L = 642$ の行列 (図 2 の株データと同条件) を正規化し、上記と同様の計算を行い平均 0、分散 1 とした相関行列の固有値分布と理論式 (3) により求められる固有値分布を比較した図である。

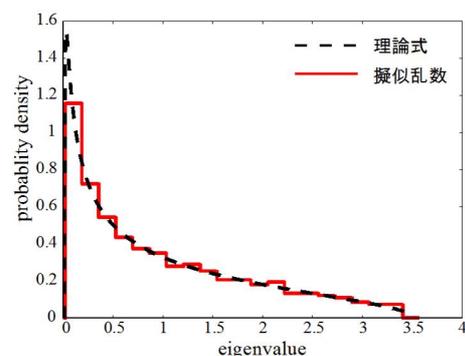


図 1 擬似乱数の固有値分布と理論式の比較

Fig. 1 Comparison of the eigenvalue distribution of pseudo-random and the theoretical formula

(1) 式の条件には $N \rightarrow \infty, L \rightarrow \infty$ とあるが図 1 に使用したデータは N, L ともに有限である。しかし図 1 を見ると、擬似乱数の固有値分布が理論式の固有値分布と類似している様子を見て取れる。これより、 $N = 500, Q = 1.3$ 程度ではデータが有限であっても理論式と同等の結果を得られることがわかる。次に、2009 年 1 ~ 3 月の株価データより作成した相互相関行列の固有値分布と比較したものを

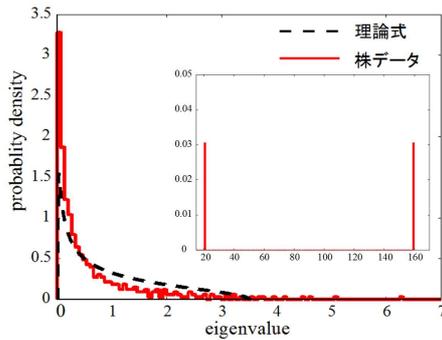


図 2 株データの固有値分布と理論式の比較

Fig. 2 Comparison of the eigenvalue distribution of stock data and the theoretical formula

図 2 に示す。

2009 年 1~3 月のデータでは理論式の最大固有値は 3.5 である。図 2 を見ると、株価データの固有値分布は理論式の固有値分布に収まっておらず 3.5 以上の固有値が現れた。時系列間の変動に相関が無い場合、相関行列より得た固有値分布は、式 (2) によって算出される固有値分布に類似する。反対に相関がある場合、ランダム行列の最大固有値 (式 (3), λ_{max}) より値の大きな固有値が発生する。この固有値をデータの本質的な成分として分離し主成分分析を行う。

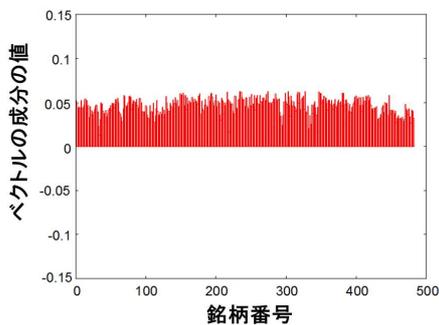


図 3 第一固有ベクトル成分

Fig. 3 Components of the first eigenvector

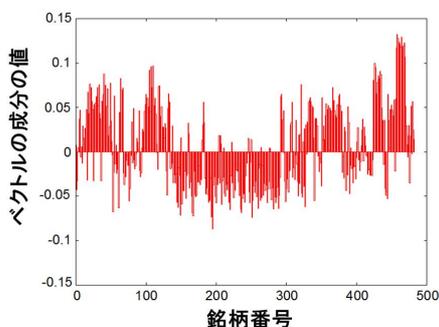


図 4 第二固有ベクトル成分

Fig. 4 Components of the second eigenvector

3.5 固有ベクトルの業種内運動

固有値を大きい順に並べ、 $\lambda_1, \lambda_2, \dots, \lambda_N$ とし、それぞれに対応する固有ベクトルを u_1, u_2, \dots, u_N と呼ぶこととする。固有値が大きいほどその固有値が相関の強い成分を持つことを示すため、まず第 1 固有値 λ_1 に着目する。第 1 固有値の固有ベクトル成分 u_1 を図 3 に示す。図 3 を見ると全ての銘柄のベクトルの成分が正値で平均的であり、銘柄間の差異が見られない。ここで例として 2009 年 1~3 月を挙げると、この期間の λ_1 の値は 157 を取り、全ての固有値の和 ($=N$) は 33 % である。これは固有値 476 個の内、1 つの固有値が約 3 割を占めているということであり、非常に大きい値であることが分かる。文献 [10] には u_1 の固有状態は TOPIX 指数のような市場全体の変動を示すと解釈できることが記述されている。 u_1 から有効な情報が得られないため次点に大きい固有値である第 2 固有値の固有ベクトル成分 u_2 に着目し、これを図 4 に示す。図 4 を見るとベクトルの成分は正負に突出しており銘柄の偏りを顕著に示している。東証株価は業種によって証券コードが割り振られているため、図 4 を見ると同業種と強い相関を示していることが分かる。さらにこの突出した要素は同業種が同じ符号に固まる傾向を持つため、 u_2 成分が正に大きい 10 銘柄、負に大きい 10 銘柄をそれぞれ相関の高い銘柄として抽出し、これらの銘柄の属する業種を表 1 の分類表に従い同定した結果を次の 4 章に示す。図にするにあたり、17 個の業種それぞれに銘柄コードを参考とした番号を付与した。

表 1 17 業種による分類
 Table 1 Classification by 17 industries

「17: 建設・素材」	「80: 商社・卸売」
「20: 食品」	「81: 小売」
「30: 素材・化学」	「83: 銀行」
「45: 医薬品」	「85: 金融」
「50: エネルギー資源」	「88: 不動産」
「54: 鉄鋼・非鉄」	「90: 運輸・物流」
「60: 機械」	「94: 情報通信・サービスその他」
「65: 電機・精密」	「95: 電力・ガス」
「70: 自動車・輸送機」	

4. 主要業種の分析

4.1 TOPIX17 による業種分類

図 5, 図 6 に TOPIX17 による各期間の主要業種の抽出結果を示す。抽出した銘柄を業種に分類すると正と負の 2 つの成分の内、1 方には資源や機械等の業種、1 方には金融やガス等の業種が固まることになった。これを資源、機械と金融、ガスで別々に時系列順に並べると、それぞれに時系列による主要な業種の移り変わりを見ることができた。

まず、先行研究 [7][8][9] で行われていた 1 年間の主要業

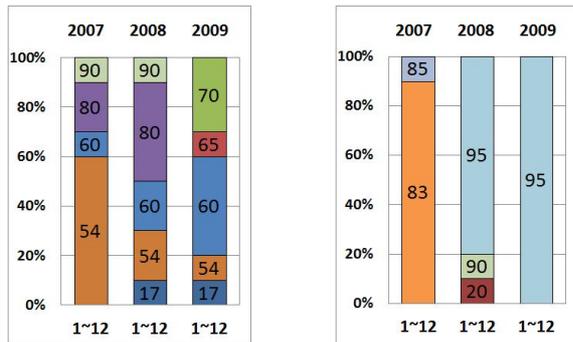


図 5 1年間毎の主要業種の抽出結果,(左)機械等,(右)金融等
Fig. 5 Major sectors extracted from intra-day(1 hour) prices from 2007 to 2009(per year), (left)Machines etc, (right)Financial etc.

種の抽出結果を図5に示す。図5(左)を見ると2007年は「54:鉄鋼・非鉄」の割合が突出しており主要業種であったことが分かる。また図5(右)を見ると,2007年は「83:銀行」,2008年,2009年では「95:電力・ガス」がそれぞれ主成分を占めており,相関の高い変動があったことが分かる。

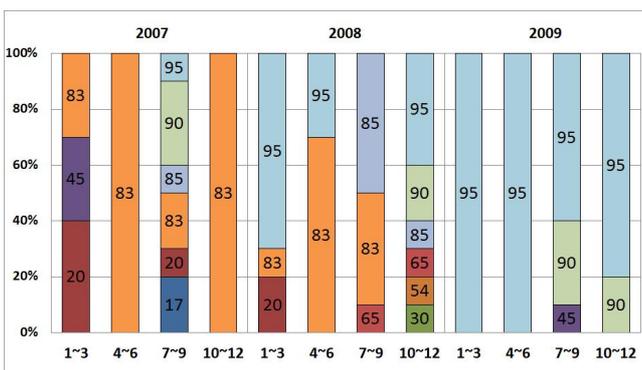
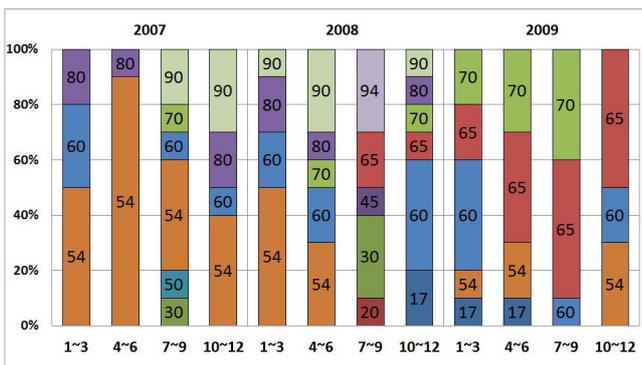


図 6 3ヶ月毎の主要業種の抽出結果,(上)機械等,(下)金融等
Fig. 6 Major sectors extracted from intra-day(30 minutes) prices from 2007 to 2009(per three months), (top)Machines etc, (bottom)Financial etc.

次に,今回の結果である3カ月毎の株価による主要業種の抽出結果を図6に示す。図6(上)を見ると,2007年の1月~3月から2008年の4月~6月まで「54:鉄鋼・非鉄」

が常に主要業種を占めており,さらに割合の増減の流れを見ることができる。また2008年の7月~9月,10月~12月を境に,「65:電機・精密」が徐々に増えている事もわかる。図6(下)でも2008年の10月~12月を境に,それまで「83:銀行」が主成分を占めることが多いが,「95:電力・ガス」に移り変わっている様子を見ることができる。なお,図6の2007年の7月~9月,2008年の10月~12月は共に複数の業種が集まっており,主要と言える業種が見られない。これは2007年8月にサブプライムローン問題を発端とした株価急落,2008年10月にはリーマンショックが起こり,多くの銘柄が一斉に下落したため主成分に複数の業種が集まったと考えることができる。このような変動は1年毎の分析では見ることができなかった。3ヶ月毎の分析を行うことで1年毎では見られなかった主要業種の細かい移り変わりを見ることができるようになった。

4.2 関連株による分類

ここまではTOPIX17による業種の分類を行ってきた。この節では,以下の表2の7つの関連株による分類を行い,上記で行った17業種の分類を元に関連株としてまとめ,実際の出来事と対比することで結果の評価を行う。7つの関連株に分けるにあたり,17業種に分類する時と同様に業務内容が複数の関連株に属している銘柄もある。ここでは[11]を参考に関連株による分類を行った。各棒グラフに3ヶ月に区切った業種の比率とその頭文字を示した。

表 2 関連株の分類
Table 2 Relation-based classification

分類名	内訳-TOPIX17による業種分類
外需関連株	「65:電機・精密」,「70:自動車・輸送機」
内需関連株	「17:建設・素材」,「88:不動産」
景気循環株	「30:素材・化学」,「60:機械」
ディフェンシブ株	「20:食品」,「45:医薬品」, 「95:電力・ガス」
消費関連株	「81:小売」,「90:運輸・物流」, 「94:情報通信・サービスその他」
金利敏感株	「83:銀行」,「85:金融」
市況関連株	「50:エネルギー資源」, 「54:鉄鋼・非鉄」,「80:商社・卸売」

図7に3ヶ月毎の主要業種の関連株による分類結果を示す。本研究で扱った時期は2008年11月のリーマンショックの影響が大きかった為に2009年の1月~3月以降,ディフェンシブ株が図3に示す業種関連株に大きく表れる。これは,2008年11月のリーマンショック後に市場が混乱する一方,その影響が少なく変動が安定していた「電力」株が主要関連株に現れたと考えられる。また図9では,2008年の4月6月を境に主要関連株が市況関連株から外需関連株へ徐々に移行していく様子を見ることができる。市況関

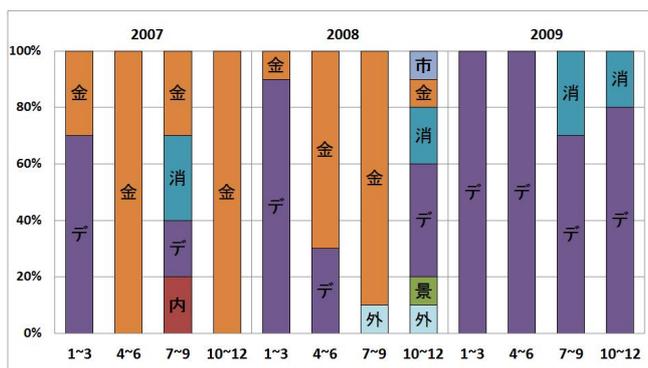
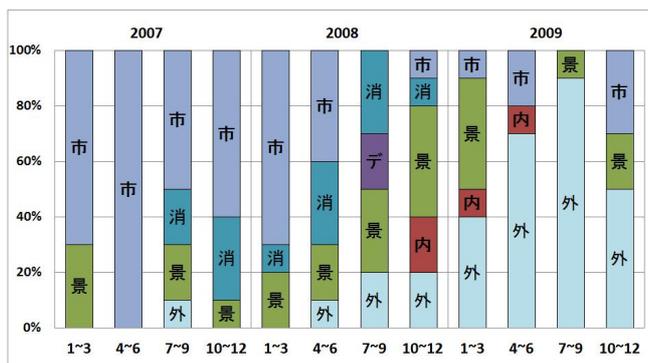


図 7 関連株による分類 (3ヶ月毎),(上)機械等,(下)金融等
Fig. 7 Extracted major sectors of each three months using the relation-based classification, (top)Machines etc, (bottom)Financial etc.

連株は原料の取引価格，外需関連株は国外での業績に影響する銘柄であり，実際に 2007 年から 2008 年にかけて原油価格が急高騰している．2008 年の 11 月にはトヨタショック，その後もクライスラーや GM の破綻など外需に影響する出来事が起きており，図 7 と一致する．このように第 2 固有値の変遷を追うことで株の大まかな動向を知ることができる．

5. まとめ

本研究では 2007 年～2009 年の東証株価データを 3ヶ月ごとに区切り各期間で用いて主要業種の抽出を行った．今回は各期間で 2 番目に大きい固有値の固有ベクトルに着目した．ベクトルの値が正，負に突出したそれぞれ 10 銘柄を主要業種として抽出し，TOPIX17 によって業種分類を行い時系列順に並べると主要業種の変遷を見ることができた．さらに評価として関連株に分類し実際の経済動向との対比を行ったところ，実験結果と実際の経済動向に高い相関を見ることができた．今回ランダム行列理論を用いるにあたり， $Q = L/N > 1$ という条件の下， L を増やすことで短期化が可能になった．また， N を減らして実験を行えば更なる短期化を行うことができると考えられる．

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, Vol. 65, pp. 066–126 (2002).
- [2] M.L.Mehta: Random Matrices, Academic Press, 3rd edition (2004).
- [3] D.Wang, et.al.:Quantifying and modeling long-range cross correlations in multiple time series with applications to word stock indices : Physical Review E, Vol. 83, pp. 046–121 (2011).
- [4] B.Podobnik, et. al.:Time-lag cross-correlations in collective phenomena, EPL, Vol. 90, pp. 68001 (2010).
- [5] R.N.Mantegna, et. al.:Hierarchical structure in financial markets, EUR.Phys.J.B, Vol. 11, pp. 193–197 (1999).
- [6] A.M.Sengupta, P.P.Mitra:Distribution of singular values for some random matrices, Physical Review E, Vol. 60, pp. 33-89 (1999).
- [7] 木戸丈剛，楊欣，田中美栄子，高石哲弥：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，情報処理学会論文誌，数理モデル化と応用，Vol. 4, pp. 104–110, (2011).
- [8] 木戸丈剛，田中美栄子：ランダム行列理論との比較による米国日時変動のトレンド抽出，FIT2010, 第 9 回情報化学技術フォーラム講演論文集（電子情報通信学会・情報処理学会），pp. 157–162 (2010).
- [9] M.Tanaka-Yamawaki, T.Kido, R.Itoi:Trend Extraction of Stock Prices in the American Market by Means of RMT-PCA, Intelligent Decision Technologies, SIST 10, pp. 637–646 (2011).
- [10] 青山秀明，家富宏，池田洋一，相馬亘，藤原義久：経済物理学，第 5 章 (2008).
- [11] NSJ 証券：初めての株式投資，http://www.nsjournal.jp/web_contents/stock_hajimete/3-2.html.