O-23

ランダム行列理論に基づいた新たな相関分析手法*

New method of correlation analysis based on random matrix theory

相馬　亘[1]
Wataru Souma[1]

Abstract: The purpose of the principal component analysis is to understand the phenomena in a small number of factors, by extracting the principal components of the correlation matrix between the variables. In many cases, the principal components are extracted by an ad hoc manner. The parallel analysis is well known and frequently applied to extract principal component of correlation matrix in many phenomena. However, we must repeat the Monte Carlo calculations in the parallel analysis. Thus the parallel analysis lacks the usefulness in the case of analyzing a large correlation matrix of different size. In order to overcome the disadvantages of the parallel analysis, we establish a method using random matrix theory. By applying the method to the analysis of stock price correlations, we show the advantage of the method.

１．はじめに

Pearson と Hotelling によって独立に提案された主成分分析の目的は，多くの変量が存在する場合に，それら変量の相関構造から主成分を抜き出し，2 次統計量である共分散行列の大部分を，より少数の因子で理解することにある．主成分分析は多くの分野で用いられていて，直感的に分かりやすく有力な手法である．しかし古くから指摘されている問題の 1 つは, 主成分の抽出基準である. この基準として, 固有値が 1 よりも大きいものを主成分と見なす Kaiser-Guttman 基準や，累積寄与率に対して経験的に閾値を設ける方法などが提案されてきた．これらは分かりやすく便利な方法であり，多くの統計解析ソフトウェアに組み込まれているが，厳密な理論的根拠に欠く．また，実際の研究現場では，このような基準を考慮するよりも，相関行列の固有値のランクサイズ・プロットとして表されるスクリー・グラフを見ながら，研究者の経験と直感を頼りとして，アドホックに主成分を抽出することが多い.

これに対して，Horn によって提案された parallel analysis [1] は，研究の対象とする相関行列の固有値分布と，相関行列と同じ大きさのランダム行列の固有値分布を比較し，これら 2 つの分布がずれる部分に存在する固有値を主成分と見なすものである．ここでランダム行列とは，ランダムな確率変数を成分とする行列であり，原子核物理学や数学でのリーマン予想と関係して，数学や物理学において古くから研究されてきた．Parallel analysis では，ランダム性を無意の証と考えることにより，主成分の抽出に対して論理的な根拠を与える．しかし，ランダム行列の固有値分布を知るには，モンテカルロ計算を繰り返す必要がある．そのため，大規模な相関行列を解析する場合や，大きさの異なる相関行列を頻繁に解析する場合には有用性を欠く．したがって，ランダム行列の固有値が従う分布関数を解析的に求められれば，parallel analysis の欠点を克服でき，主成分分析の発展に大きく貢献できる．

1967 年に Marcenko と Pastur は，ランダム行列の固有値が従う分布関数を解析的に導いた．しかし，この結果がすぐに parallel analysis に応用されることは無く，1999 年になってようやく，Laloux 等 [2] と Plerou 等 [3] によって別々に，株価相関の研究に対して応用された．具体的にこれらの研究では，ニューヨーク証券取引所の S&P500 銘柄（日本の日経平均のようなもの）が作る相関行列の固有値分布とランダム行列理論から導かれた理論式を比較し，これら 2 つの分布がずれる部分に主成分があり，分布が重なる部分の成分はノイズであることを明らかにした．これらの論文を契機に，海外の多くの研究者が，株価相関の研究にランダム行列理論を応用した論文を発表している．また，応用範囲は，気象，情報通信，脳波，経済，音声認識等に広がりつつある．このような状況の中で本研究では，株価の変動に対してランダム行列理論を応用した結果を報告する．そして，これからの展望について議論する.

＊　本研究は，平成 23 年度日本大学理工学部基礎科学研究助成金（研究助成 B）の助成を受けています.

１：日大理工・教員・一般

２．データとその解析

本稿では，1997 年 1 月 6 日〜2003 年 3 月 31 日までの期間に，東証 1 部に上場していた株式で, すべての日で取引が成立した株式の日次データを用いる．そのため，時系列の長さは最大で 1534 日で銘柄数は 658 銘柄である. これらの株価データの終値を使って日々の対数収益率を計算することによって, 658×658 の相関行列 **C** が得られる．

相関行列に限らず，対象とする行列の性質を明らかにするには，固有値やその分布の解析は重要である．そこで，いま考えている相関行列 **C** に対して固有値λを計算し，固有値λの確率密度 $\rho(\lambda)$ を図 1 に示す．この図では, 比較的小さな固有値の部分（$0<\lambda<10$）を拡大して表示している．図中の小さな図は，分布全体の様子である．

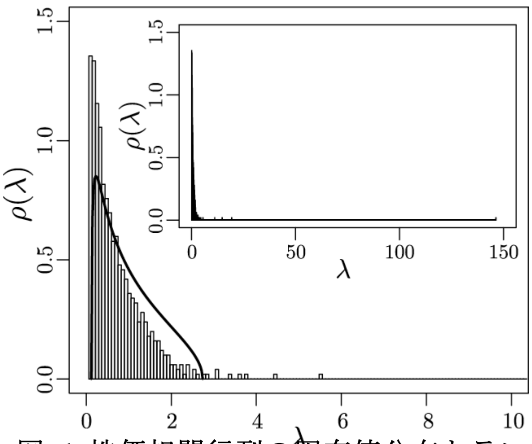

図 1 株価相関行列の固有値分布とランダム行列理論の理論曲線

図 1 に描かれた実線は, ランダム行列の固有値が従う確率密度関数で, Marcenko と Pastur によって解析的に求められたものである．この確率密度関数には，固有値が存在する領域に上限と下限があるという特徴がある．

株価の相関行列から求められた固有値の分布と，ランダム行列理論の理論曲線を比較すると，ズレがあることがわかる．このズレには 2 つのタイプがある. 一つは, 分布形の違いであり，もう一つは, 固有値が分布する範囲の違いである．ここでは，後者のズレに着目し，理論曲線が導きだす上限よりも大きな所に分布している固有値と固有ベクトルに, 有意な相関構造が潜んでいるという立場をとる．

実際に，それらの固有値と固有ベクトルを詳しく見ていくと，最大固有値は，TOPIX や日経平均などの，市場を代表する指数と相関が大きいことがわかった．具体的には，規格化された TOPIX のリターン（横軸）と，第 1 固有値の固有ベクトルの重みで作ったポートフォリオのリターン（縦軸）の相関を調べると，図 2 に示す散布図が得られる．この相関は 0.86 で，強相関であることから，第 1 固有値は市場全体の様子を代弁するものであることがわかる．また，第 2 固有値以下の固有ベクトルの成分を調べることによって，それぞれの固有ベクトルが，ある特定の業種を抽出していることがわかった．このようなことから，ランダム行列を帰無仮説とした主成分分析手法は有効であると期待される．

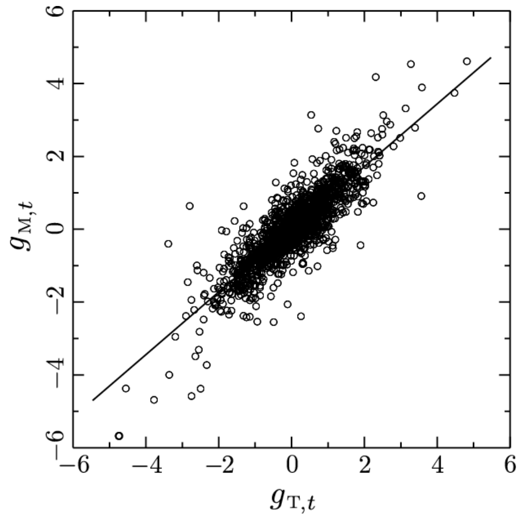

図 2　TOPIX との相関

３．結論と展望

本稿では，株価の相関構造を調べることによって，ランダム行列理論に基づいた新たな相関分析手法の優位性を示した．だが，この手法の優位性を示すには，さらに多くの現象の解析に応用して確かめる必要がある．たとえば，本稿では銘柄間の相関に着目したが，同じデータを用いて時間相関についても解析でき，どのような有意な相関構造が得られるのか興深い．

４．参考文献

[1] J. L. Horn: “A rational and test for the number of factors in factor analysis ”, Psychometrika, Vol. 30, pp. 179 -185, 1965.

[2] L. Laloux, et al.: “Noise dressing of financial correlation matrices”, Physical Review Letters, Vol. 83, pp. 1467-1470, 1999.

[3] V. Plerou, et al.: “Universal and nonuniversal properties of cross correlations in financial time series”, Physical Review Letters, Vol. 83, pp. 1471-1474, 1999.