

クオントピックス：NO33 ランダム行列理論を用いた市場構造分析

共分散行列の固有値分布の特徴を調べてみよう

- 共分散行列を固有値分解することの意味について
- 市場構造分析から見た実務活用の可能性について

0. 予備知識、キーワード

基本的な金融工学、線形代数、ランダム行列理論の知識を仮定します。

キーワード：ランダム行列理論、固有値分布、共分散行列、シミュレーション

1. イントロダクション

ファイナンスにおける共分散行列は、ポートフォリオ最適化、リスク推定、回帰問題、次元削減等、あらゆる分野で登場します。では、線形関数の枠組みで共分散行列を見た場合、この行列の固有値とは何を意味するのでしょうか？今回は、共分散行列の固有値の特徴、その応用例としての市場構造の分析について紹介します。

2. ランダム行列とWishart分布

まず、N個の銘柄それぞれのT月分の時系列リターンを用意します。つまり、(t, i)成分がt月目の銘柄iのリターンを表しています。

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1N} \\ \vdots & \ddots & \vdots \\ x_{T1} & \cdots & x_{TN} \end{pmatrix} \quad (1)$$

銘柄リターンは確率的に揺らぐため、(1)をランダム行列と呼びます。ここから計算される標本共分散行列(2)もランダム行列です。ただし、 X' は X の転置行列を表します。

$$S = \frac{1}{T} X' X \quad (2)$$

このことを直感的に説明すると、推定された共分散行列 S はサンプルを取り直す度に変わるため、母共分散行列 Σ と異なり、確率変数とみなせるからです。つまり、このような場合は「行列の確率分布」が必要となります。ランダム行列理論において、標本共分散行列 S が従う確率分布をWishart分布と呼びます。

$$S \sim W_N \left(T, \frac{1}{T} \Sigma \right) \quad (3)$$

(3)は、 Σ を母共分散に持つN次元ベクトルをT回サンプリングして得られた S が従う分布を表しています。また、理論上は逆行列を持つための正則性条件 $T \geq N$ を仮定します。

一方、金融データでは上記の仮定が崩れることが多々あります。例えばTOPIX銘柄を分析の対象にした場合、必要な時系列データは2,000ほどになり、日次ベースで8年分が必要です。月次ベースでは用意することは不可能です。この仮定が崩れることで有効自由度が不足し、推定のノイズが発生します。次節以降では、解決策の1つとして共分散行列 S のノイズ・シグナルの判別を行い、さらにノイズ除去の手法を行うことで実際のポートフォリオ戦略にどう活用できるかについて紹介します。

固有値分布とMarčenko–Pastur 定理

3. 共分散行列の固有値分布と、Marčenko–Pastur 定理

本節では、ノイズとシグナルの判別を行うために、ランダム行列理論における以下の極限定理を使います。

【Marčenko–Pastur 定理】

$X \in \mathbb{R}^{T \times N}$ を、各成分が平均0、分散有限で独立同分布のランダム行列とする。このとき、極限を以下の形で取ると、標本共分散行列 $S = \frac{1}{T} X'X$ の固有値分布 $p(\lambda)$ は以下の分布に確率1で収束(概収束)する。

$$p(\lambda) \xrightarrow{\text{a.s.}} MP(q, \sigma^2), \quad T, N \rightarrow \infty \left(q = \frac{N}{T} < \infty \right) \quad (4)$$

確率密度関数で表すと(5)のようになります。ただし、 $\lambda_{\pm} = \sigma^2(1 \pm \sqrt{q})^2$ です。

$$f(\lambda) = \begin{cases} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{2\pi q \sigma^2 \lambda}, & \lambda \in [\lambda_-, \lambda_+] \\ 0, & \lambda \notin [\lambda_-, \lambda_+] \end{cases} \quad (5)$$

この定理で重要なことは、ノイズだけで共分散行列を構成した場合、固有値が区間 $[\lambda_-, \lambda_+]$ から外れることは確率的には100%起こらない。つまり 外側に出た固有値はノイズでは説明できず、シグナル(共通因子・市場構造等)が存在する証拠になるということです。また、ノイズとシグナルの判別においては、 N と T の大小関係の制約が無い点も重要です。そのため、高次元・小標本な金融データに対しても適用可能です。

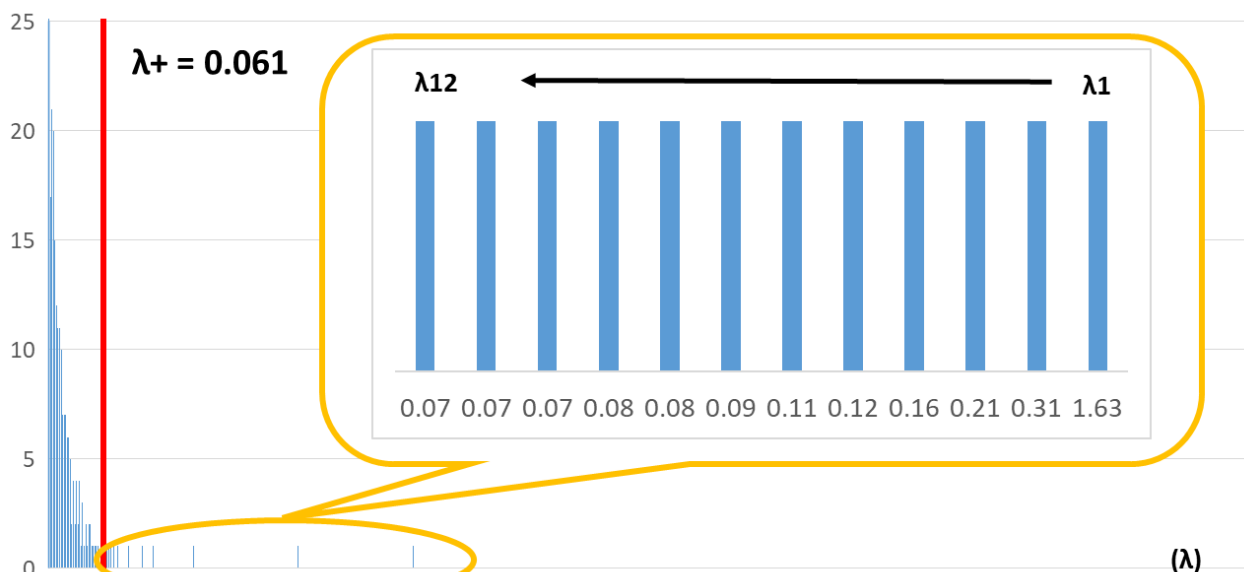
4. 共分散行列のノイズ・シグナル成分の特定

実際に、まず過去25年分のTOPIX構成銘柄の月次リターンデータを用いて、全期間の標本共分散行列 S を計算し、これを固有値分解します。

$$S = P \cdot \text{diag}(\lambda_1, \dots, \lambda_N) \cdot P^{-1} \quad (6)$$

固有値は、 $\lambda_1 > \dots > \lambda_N$ と大きい順に並べ替えていると仮定します。固有値のヒストグラムが(4)の分布に従うことが分かります(図表1)。

(図表1) 固有値のヒストグラムとMarčenko-Pastur分布の比較

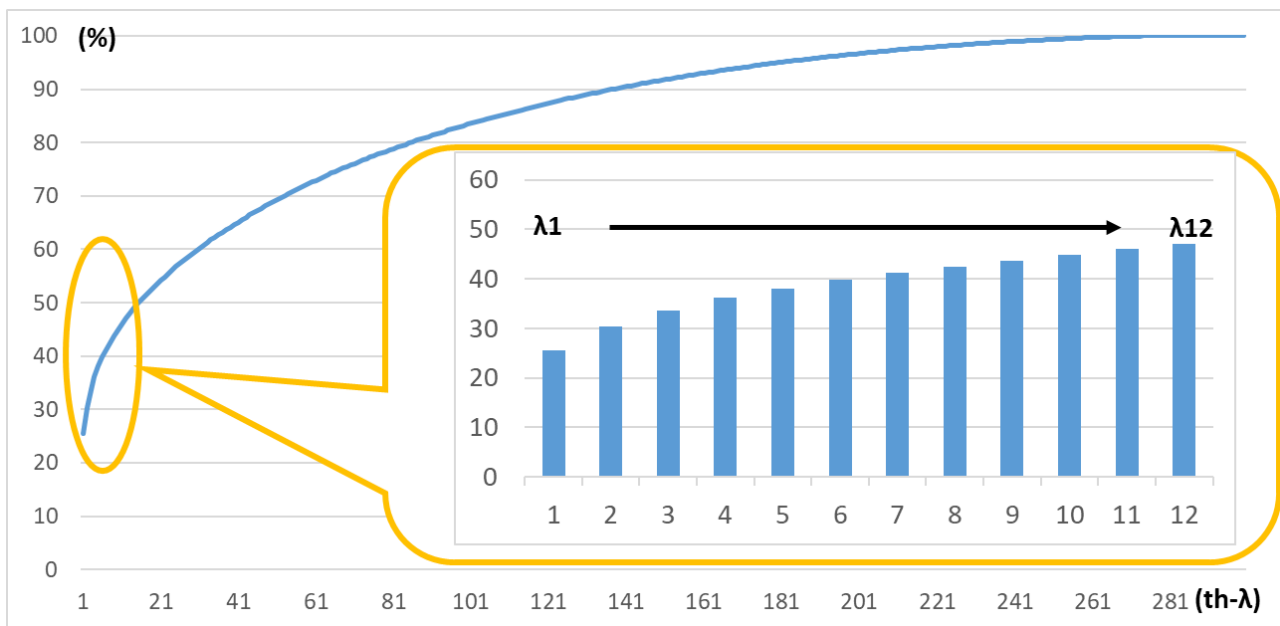


ノイズ・シグナルの判別

Marčenko-Pastur分布の確率密度関数が非ゼロとなる右端（赤線）を閾値として、それより右側にある12個の固有値がシグナルに該当します。一般的には、第1固有値（ $\lambda_1 = 1.63$ ）が市場因子、つまり景気、金利、マクロ要因の影響で市場全体の変動因子に相当します。その他（ $\lambda_2 = 0.31, \dots, \lambda_{12} = 0.07$ ）は「製造業か非製造業か」、「輸出企業か内需企業か」、「バリューかグロースか」といった、市場とは別の集団行動を示す潜在因子となります。

実際、第1固有値から累積させた寄与度が(図表2)であり、第1固有値だけで全体の26%、シグナルとなる第12固有値までで全体の47%を説明できることが分かります（その他はノイズとして扱う）。

(図表2)固有値の累積寄与率



同じ固有値分解を行う主成分分析（PCA）と比較しても、「落とした後の次元数を裁量で決める」主成分分析と、「使うべきでないノイズ部分分かる」Marčenko-Pastur定理では明確な違いがあります。最後に、ノイズとシグナルに関して処理前後のバックテスト結果を紹介します。

5. 最小分散ポートフォリオでの比較

本節では、共分散行列の推定の差が分かりやすい最小分散ポートフォリオの構築を行い、バックテスト結果を比較します。便宜上、投資対象は各月のTOPIXの時価総額上位100銘柄（TOPIX100）とし、毎月リバランスを実施、取引コストは一律0.1%と仮定します。まず、ローリング共分散行列に対して以下の前処理をします。

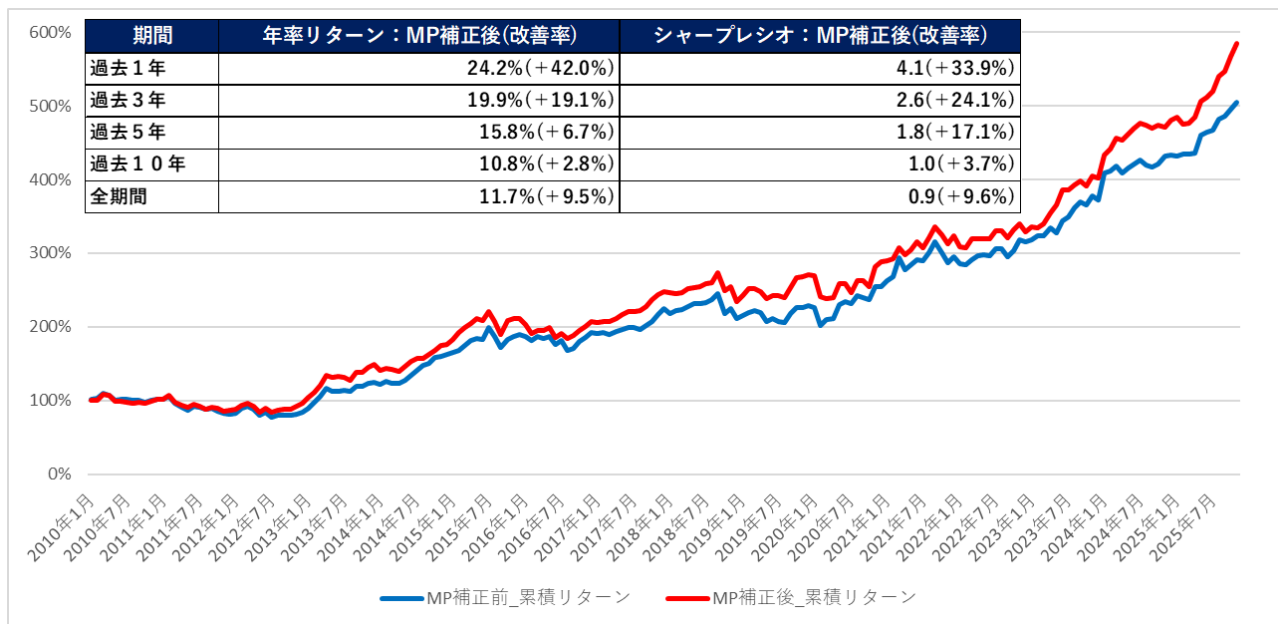
$$\tilde{\lambda}_i = \begin{cases} \lambda_i, & \lambda_i > \lambda_+ \\ \lambda_+, & \lambda_i \leq \lambda_+ \end{cases}, \quad \tilde{S} = P \cdot \text{diag}(\tilde{\lambda}_1, \dots, \tilde{\lambda}_N) \cdot P^{-1} \quad (7)$$

(7)は固有値のノイズ部分を λ_+ に引き上げているノイズ除去であり、固有値クリッピングと呼びます。固有値クリッピング後に共分散行列 \tilde{S} を再構築し、(6)と(7)の共分散行列をそれぞれMP補正前後の共分散行列とします。

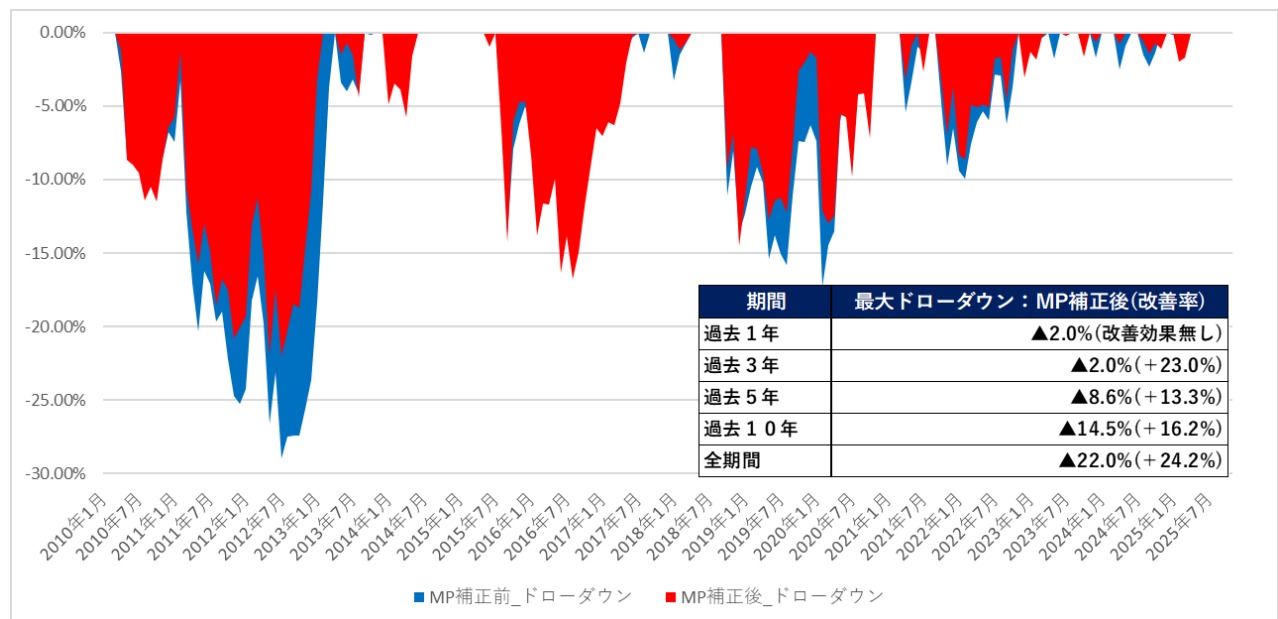
最小分散ポートフォリオによるバックテスト

MP補正前後の共分散行列それぞれを使った最小分散ポートフォリオのバックテスト結果は以下の通りです。

(図表 3) 累積リターン比較(全期間)



(図表 4) ドローダウン比較(全期間)



出所：東京証券取引所のデータを基に、ニッセイアセットマネジメント作成

MP補正効果（MP補正前後の比較）を見ていくと、累積リターンやシャープレシオといったパフォーマンスは安定して上振れ、ドローダウンも抑制されていることが分かります。また、ターンオーバー削減の影響でより安定したウェイトが実現し、過度なリバランスが不要、つまり取引コストの削減効果も見られます。

MP補正効果をまとめたものが(図表 5)です。

まとめ

(図表 5) MP補正の効果 (全期間)

パフォーマンス指標	MP補正前	MP補正後	改善率
年率リターン	10.7%	11.7%	+ 9.5%
シャープレシオ	0.8	0.9	+ 9.6%
リスク指標	MP補正前	MP補正後	改善率
年率ボラティリティ	12.8%	12.8%	+ 0.1%
最大ドローダウン	▲29.0%	▲22.0%	+ 24.2%
取引効率	MP補正前	MP補正後	改善率
平均ターンオーバー	31.0%	10.8%	▲65.0%
勝率	64.4%	65.4%	+ 1.6%

6. まとめ

今回は、ランダム行列理論のMarčenko-Pastur定理を用いた共分散行列のノイズ除去手法について、理論から実証まで一貫した検証を行いました。全体を通した運用実務への示唆は以下の通りです。

第一に、MP補正により、限られたデータからでもノイズとシグナルを統計的に判別し、頑健な共分散行列を構築できることが示されました。これは特に、新興市場や短期間での戦略構築において有用です。

第二に、ターンオーバーが約65%削減されたことで、頻繁なリバランスに伴うスプレッドコストや市場インパクトが大幅に抑制され、特に流動性の低い銘柄を含むポートフォリオでは、この効果は実現リターンに直接反映されます。

本手法は、最小分散戦略に限らず、リスクパリティ、ブラック・リッターマン・モデル、多因子モデル等、共分散行列を用いるあらゆる運用戦略に適用可能です。実装も固有値クリッピングという単純な操作で実現でき、運用現場への応用可能性は高いと言えます。

参考文献

[Kanokogi]マルコス・ロペス・デ・プラド著；鹿子木亨紀訳、アセットマネージャーのためのファイナンス機械学習、金融財政事情研究会、(2020)
[Tao] Terence Tao, Topics in random matrix theory, Graduate Studies in Mathematics(2012)

～執筆者の紹介～

塚本 恵 (写真：右)

ニッセイアセット入社後、ファイナンシャルテクノロジー運用部にて主に機械学習を含む数理的な定量的手法、オルタナティブデータを活用した新たな投資戦略の研究開発を担当。

