2009－MPS－73　(21)
2009／3／5

ランダム行列によるノイズ除去の統計的仮説検定とその共ボラティリティへの適用

名古屋大学 大学院工学研究科　橘 完太
一橋大学 大学院経済学研究科　森本 孝之

日内高頻度収益率の交差積和として計算される実現共ボラティリティを用いてボラティリティを推定する際に，ミクロ構造ノイズと呼ばれるバイアスが生じることが知られている．ミクロ構造ノイズを除去する手法として，交差積和行列を固有値分解し，ランダム行列の最大固有値よりも小さな固有値に対応する成分をノイズとみなす手法が既存である．ランダム行列の最大固有値は漸近的に Tracy-Widom 分布に従うが，既存手法では，ランダム行列の最大固有値の漸近性を考慮せず，その収束値のみを用いる．そのため，本質的なボラティリティを誤ってノイズとみなす危険性が定量的に評価されない．そこで，本稿では，ランダム行列最大固有値の Tracy-Widom 分布への漸近性に基づいた，共ボラティリティのノイズに関する統計的仮説検定を提案する．

キーワード：共ボラティリティ，ミクロ構造ノイズ，ランダム行列.

A note on a statistical hypothesis testing for removing noise by the random matrix theory, and its application to co-volatility matrices

Kanta Tachibana[1] and Takayuki Morimoto[2]
[1] *Graduate School of Engineering, Nagoya University*
[2] *Graduate School of Economics, Hitotsubashi University*

It is well known that the bias called market microstructure noise will arise, when estimating realized co-volatility matrix which is calculated as a sum of cross products of intraday high-frequency returns. An existing conventional technique for removing such a market microstructure noise is to perform eigenvalue decomposition of the sum of cross products matrix and to identify the elements corresponding to the decomposed values which are smaller than the maximum eigenvalue of the random matrix as noises. Although the maximum eigenvalue of a random matrix follows asymptotically Tracy-Widom distribution, the existing technique does not take this asymptotic nature into consideration, but only the convergence value is used for it. Therefore, it cannot evaluate quantitatively such a risk that regards accidentally essential volatility as a noise. In this paper, we propose a statistical hypothesis test for removing noise in co-volatility matrix based on the nature in which the maximum eigenvalue of a random matrix follows Tracy-Widom distribution asymptotically.

Keywords: Co-volatility, market microstructure noise, random matrix.

1 はじめに

近年，積分ボラティリティ（Integrated Volatility, IV）の一致推定量であり，高頻度収益率の日内の二乗和として計算される，実現ボラティリティ（Realized Volatility, RV）を用い，より正確にボラティリティを推定，予測できるようになった [1]，[2]．また，このRVの単純な拡張として2変量間の高頻度収益率データの交差積の和として計算される実現共分散（Realized Covariance, RC）もポートフォリオといったファイナンスへの応用として重要である．しかし，これらRV，RCには，高頻度データを用いることに起因するいくつかの問題点が知られている．まず，ボラティリティの推定値に大きなバイアスを与えるミクロ構造ノイズと呼ばれる存在がある．これは，観測されたデータが潜在データとノイズ項の和で表されると仮定することによりモデル化され，横軸に観測周期，縦軸にボラティリティの値をとるボラティリティ・シグニチャー・プロット（Volatility Signature Plot, VSP）を描くことにより，その影響を見ることができる．また，Epps 効果と呼ばれる観測周波数を高くすると2資産間の共分散が小さくなるという現象が知られている．この Epps 効果に関しては，林・吉田推定量により回避できることが知られているが [4]，東証などの最低取引時間間隔が1分以上の場合，この推定量は定義上うまく計算できないため，本研究では Epps 効果に関しては基本的に関与しない．

まず，本研究の理論的背景を示す．ここでは，日次ボラティリティを推定する上で高頻度データを利用する根拠となるものである．前提となることとして，時点 i における対数株価を p_i とする連続時間における拡散過程

$$dp_i = \mu_i dt + \sigma_i dw_i$$

を考える．ここで μ_i を瞬時的ドリフト項，σ_i を瞬時的拡散項，w_i を標準ブラウン運動とする．推定したいボラティリティとは ζ 日での日次の σ，つまり積分ボラティリティ $\int_{\zeta-1}^{\zeta} \sigma_s^2 ds$ である．ただ，これは連続時間モデルなので，現実の離散的なデータにはそのまま適用できない．そこで，Δ を各日内における微少な時間幅とし，ζ 日における τ 番目の日内収益率 $r_{\zeta,\tau} = p_{\zeta,\tau\Delta} - p_{\zeta,(\tau-1)\Delta}$ を考える．そうすると，$\Delta \to 0$ において，実現ボラティリティは

$$RV_\zeta := \sum_\tau r_{\zeta,\tau}^2 \to \int_{\zeta-1}^{\zeta} \sigma_s^2 ds$$

となることが知られている．したがって，一日のデータが十分大きい，つまり，観測周期が十分短ければ，RV が IV の一致推定量となる．しかし，観測周期が長ければ，RV が IV の一致推定量とならない．よって，概ね20分以上間隔の RV では長すぎてボラティリティの推定値としては全く意味が無いことになる．したがって，RV を IV の推定量として用いるためには，観測周期を1分-20分程度とするべきである．

RV と同様に，ζ 日の τ 番目の時刻における2銘柄 i, j の収益率 $r_{\zeta,\tau,i}$ および $r_{\zeta,\tau,j}$ を用いて，実現共ボラティリティは

$$CV_{\zeta,ij} := \sum_\tau r_{\zeta,\tau,i} r_{\zeta,\tau,j}$$

と計算される．ここで，N 銘柄について ζ 日の τ 番目の時刻における銘柄 i の収益率 $r_{\zeta,\tau,i}$ を第 τ 行第 i 列に配した $p \times N$ の収益率行列 R_ζ を考えると，実現ボラティリティおよび全銘柄間の実現共ボラティリティをまとめた $N \times N$ の行列が $V_\zeta = R_\zeta^t R_\zeta$ と計算される．ここで右肩の t は行列の転置を表す．次節以降では，簡潔のため，$R_\zeta, V_\zeta, r_{\zeta,\tau,i}$ に付けられた，日を表す添え字 ζ を省略することもある．

続いて，ミクロ構造ノイズをどう考えるかに関して，ノイズ自体は観測不可であるため，現在のところ明確なモデルはない．が，一般的には，効率的な対数価格過程を p^* とすると，ミクロ構造ノイズ過程 u を含んで観測される価格過程 p は恒等式 $p = p^* + u$ で定義される．こうしたノイズを含んだままボラティリティを推定してしまうと上述の RV の推定値にバイアスが生じてしまうため，それを取り除くことが必要となる．そこで，こうしたバイアスを取り除く試

みとして，いろいろな推定量（Bias Correcting Estimators）が考えられてきた [3]．本研究では，このようなノイズの影響を取り除く試みの一つとして，ランダム行列最大固有値の Tracy-Widom 分布への漸近性に基づいた，共ボラティリティのノイズに関する統計的仮説検定を提案する．

2 最大固有値の漸近分布

X を $n \times p$ のランダム行列とし，その共分散行列を $X'X$ とする．この行列の次元の比率がある定数となる場合，単位共分散を持つ Wishart 行列 $X'X$ の最大固有値漸近分布は，1 次の Tracy-Widom 分布に従う [6]．また n と p が 10 程度の大きさであっても，この漸近的性質が失われない [6]．この事実は，次のように定式化できる．

定理： $\mathcal{W}$ を白色 Wishart 行列，γ を任意の定数，l_1 をその最大固有値とし，$n/p \to \gamma \geq 1$ とした場合，

$$\frac{l_1 - \mu_{np}}{\sigma_{np}} \xrightarrow{dist} \mathcal{W} \sim F_1$$

となる．ここで位置定数と尺度定数は，

$$\mu_{np} = (\sqrt{n-1} + \sqrt{q})^2,$$
$$\sigma_{np} = \sqrt{\mu_{np}} \left((n-1)^{-\frac{1}{2}} + p^{-\frac{1}{2}} \right)^{1/3}$$

であり，F_1 は 1 次の Tracy-Widom 法則の分布関数を表す．

この Tracy-Widom 分布は，Painlevé II 型の微分方程式の解として現れる．

この漸近分布関数 F_1 は，F_β の分布族の特殊な場合である．$\beta = 1, 2, 4$ に対して，関数 F_β は，それぞれガウス型直交アンサンブル (GOE)，ガウス型ユニタリ・アンサンブル (GUE)，及びガウス型斜行アンサンブル (GSE) の最大固有値に対する漸近分布として現れる [7]，[8]，[9]．これらの結果は，GOE ($\beta = 1$)，GUE ($\beta = 2$)，あるいは GSE ($\beta = 4$) いずれのランダム行列 $\mathcal{A}$ の最大固有値 $l_{\max}(\mathcal{A})$ に対しても，その分布関数 $F_{N,\beta}(s) \stackrel{def}{=} \mathbb{P}(l_{\max}(A) < s)$, $\beta = 1, 2, 4$ は，次の漸近法則を満たす：

$$F_\beta(s) = \lim_{N \to \infty} F_{N,\beta}(2\sigma\sqrt{N} + \sigma N^{-1/6} s)$$

ここで F_β は，明示的に

$$F_1(s) = \exp\left(-\frac{1}{2} \int^{\infty} sq(x)dx \right) [F_2(s)]^{\frac{1}{2}},$$
$$F_2(s) = \exp\left(-\int_s^{\infty} (x-s))q^2(x)dx \right),$$
$$F_4(2^{-\frac{2}{3}} s) = \cosh\left(-\frac{1}{2} \int_s^{\infty} q(x)dx \right) [F_2(s)]^{\frac{1}{2}}$$

によって与えられる．ここで $q(s)$ は，Painlevé II 型方程式の唯一の解である．

$$q'' = sq + 2q^3 + \alpha, \ \alpha = 0$$

は，次の境界条件を満たす．

$$q(s) \sim \mathrm{Ai}(s), \ s \to +\infty$$

ここで Ai(s) は，Airy 関数を表す．

また，Tracy-Widom 分布の裾の確率は，数値計算により下記の表 1 のように与えられることが知られている [10]．従って，この値と有意水準 α を用いることにより，共分散行列の最大固有値に対する統計的仮説検定が可能となる．

表 1: Tracy-Widom 分布の値 ($\beta = 1, 2, 4$)

$\beta \backslash 1-\alpha$	0.995	0.975	0.95
1	-4.1505	-3.5166	-3.1808
2	-3.9139	-3.4428	-3.1945
4	-4.0531	-3.6608	-3.4556

3 本質的なボラティリティの抽出

本節では，標準化した収益率行列 R から計算する行列 $V = R^t R$ を本質的な成分と本質的でないノイズ成分とに分ける手法を述べる．V の固有値 $\lambda_1, \ldots, \lambda_N$ と固有値 $\lambda_k, (k = 1, \ldots, N)$ に対応する長さ 1 の固有ベクトル u_k を計算し，

$V_k := \lambda_k u_k u_k^t$ を k 番目の成分とすると，$V = \sum_{k=1}^{N} V_k$ が成り立つので，V が N 個の成分に分解される．これら N 個の成分のうち大きな固有値に対応する成分ほど市場全体の動きに大きな影響を及ぼしており，より必然に起こる，より本質的な成分であるとみなせる．一方，小さな固有値に対応する成分ほど市場全体の動きとは無関係で，より偶然に起こるノイズ成分であるとみなせる．

そこで，ある閾値 θ よりも大きな固有値に対応する成分の和

$$V_+ = \sum_{k|\lambda_k > \theta} V_k$$

を V の本質的な部分とみなすこととする．では，この閾値をいかに設定すればよいだろうか．既存研究では，この閾値を，ランダム行列の最大固有値の漸近性を考慮せず，その収束値のみを用い閾値を決定している [5]．つまり，$\lambda_k > \theta_1$ となるような固有値に対応する成分の和を本質部分，そうでない成分をノイズとみなしている．しかし，こうした確定的でデジタルな方法は，共分散行列が確率変数を要素とするランダム行列であることにより生じる，いわゆる第一種過誤，つまり，本質的なボラティリティを誤ってノイズとみなす危険性が定量的に評価されない．そこで，前節で説明したように，V の最大固有値が Tracy-Widom 分布に従うという性質を用い，本質部分とノイズに関する統計的仮説検定を行う．具体的には，帰無仮説を固有成分がノイズであると設定し，その統計量は標準化した収益率行列 R から計算する V の固有値である．

参考文献

[1] Andersen, T. G., T. Bollerslev, and F. X. Diebold (2007), "Roughing It Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility," *Review of Economics and Statistics*, **89**, 701–720.

[2] Barndorff-Nielsen, O. E., and N. Shephard (2004), "Power and Bipower Variation with Stochastic Volatility and Jumps," *Journal of Financial Econometrics*, **2**, 1-37.

[3] Hansen, P.R. and Lunde, A. (2006). "Realized Variance and Market Microstructure Noise," *Journal of Business and Economic Statistics*, **24**, 127–161.

[4] Hayashi, T. and Yoshida, N. (2005). "On covariance estimation of non-synchronously observed diffusion processes," *Bernoulli*, **11**, 359-379.

[5] Laloux, L., Cizeau, P., Potters, M. and Bouchaud, J. (2000), "Random matrix theory and financial correlations," *International Journal Theoretical Applied Finance*, **3**, 391—397.

[6] Johnstone, I.M. (2001) "On the distribution of the largest eigenvalue in principal component analysis," *Ann. of Stat.*, textbf29, 295–327.

[7] Tracy, C.A., Widom, H. (1993) "Level-spacing distribution and Airy kernel," *Phys. Letts. B*, **305**, 115–118.

[8] Tracy, C.A., Widom, H. (1994) "Level-spacing distribution and Airy kernel," *Comm. Math. Phys.*, **159**, 151–174.

[9] Tracy, C.A., Widom, H. (1996) "On orthogonal and symplectic matrix ensambles," *Comm. Math. Phys.*, **177**, 727–754.

[10] Bejan, A.(2005) "Largest eigenvalues and sample covariance matrices. Tracy-Widom and Painlevé II: computational aspects and realization in S-Plus with applications," Preprint.