ランダム行列による高頻度金融データの解析

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

1. はじめに

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．

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．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，ランダム行列理論を用いることを提案する．実証研究に関しては，2006 年の TOPIX 100 に選出されている東京証券取引所上場銘柄 100 系列を用いた．

2. 本質的なボラティリティの抽出

本節では，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．ここでは，ランダム行列を用いて得る 2 種類の閾値 θ を提案し（手法 1 と手法 2），さらに，本質的な相関関係をランダム行列により抽出後ボラティリティに換算する手法（手法 3）を提案する．

3. 結果と考察

実証分析での対象とする株価データは，TOPIX100 に含まれる個別銘柄と株価指数 TOPIX の計 $N = 101$ 銘柄を，期間 2006 年 1 月 6 日から 2006 年 12 月 28 日までの計 245 日の分次終値データを用いた．各手法でのミクロ構造ノイズ除去性能を評価するため，各取引日における日内収益率を，標本間隔 Δ を 1 分から 20 分まで 1 分刻みで変えて計算した．

ここでは，横軸は Δ，縦軸はボラティリティをとるいわゆる Volatility Signature Plot を描くことにより，各手法の評価を行った．具体的には，各取引日で計算した $V, V_{+(m)}, (m = 1, 2, 3)$ について $\mathrm{tr}(V), \mathrm{tr}(V_{+(m)})$ を計算し，その 245 日分の平均値を示すことにより評価した．ここで $\mathrm{tr}(\cdot)$ は対角要素の和であり，各銘柄の日内収益率変動について，その大きさの総和を評価している．結果としては，観測周期が短くなるほど大きなバイアスがボラティリティにかかるというミクロ構造ノイズが，提案手法 1 によって少し除去され，提案手法 2 および提案手法 3 によって大幅に除去されたことが分かった．つまり，手法 1 と異なり，手法 2 および手法 3 は銘柄間での分散の違いを反映するため，より効果的にノイズを除去したことが窺える．

次に，横軸は Δ，縦軸は共ボラティリティの総和をとるいわゆる Co-volatility Signature Plot を描くことにより，各手法の評価を行った．具体的には，共ボラティリティの指標として，各取引日における V の非対角要素 $V_{ij} (i > j)$ 全ての和を求め，その 245 日分の平均値と，$V_{+(m)}, (m = 1, 2, 3)$ についても同様に求め評価を行った．結果としては，Raw データで観測周期が短くなるほど共ボラティリティの総和が減っている現象は，所謂 Epps 効果であると思われ，手法 2 でノイズ除去した $V_{+(2)}$ については，この傾向がより顕著となっている．これは，手法 1 及び 3 では分散が等方的なランダム行列を用い，手法 2 では等方的でないランダム行列を用いたことに，少なくとも一部起因すると考えられる．