

---

# ランダム行列理論を用いた共分散行列のノイズ除去

---

森本 孝之

関西学院大学 理工学部 数理科学科

J S T 新技術説明会  
2009年10月22日 (木)

於 サイエンスプラザ地下1階 JSTホール

# はじめに

---

- [1] 近年、高頻度の金融データが入手可能となり
- 実現ボラティリティ (Realized Volatility, RV) : 日内の高頻度収益率の二乗和
  - 実現共ボラティリティ (Realized Co-volatility, RC) : 2 変量間の交差積の和
- を用い、金融資産のボラティリティを推定できるようになった。
- [2] しかし、これら RV, RC には、リスクの推定に大きなバイアス (ミクロ構造ノイズ) が加わる。しかも、このノイズは高頻度データになるほど、大きくなる。
- [3] そこで本研究では、RV および RC に含まれるノイズを、ランダム行列を用いて取り除く手法について考察。

# ランダム行列とは？

---

- ランダム行列とは，確率変数を要素にもつ行列。
- (例)  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p \in \mathbb{R}^N$  がそれぞれ独立に平均  $\boldsymbol{\mu}$ ，共分散行列  $\boldsymbol{\Sigma}$  の  $N$ 次元多変量正規分布に従うとき

$$\sum_{i=1}^p (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$$

は自由度  $(p - 1)$ ，共分散行列  $\boldsymbol{\Sigma}$  の  $N$ 次元 Wishart 分布に従う。

- $N = 1$  の場合：自由度  $(p - 1)$  の  $\chi^2$  分布。
- つまり，Wishart 分布は  $\chi^2$  分布の多次元への拡張に相当。

# ランダム行列の歴史

---

- (例)  $N$ 次元 Wishart 分布は？
  - $N = 2$  : Fisher (1915), 一般の  $N$  : Wishart (1928).
- (例)  $\{a_{ij}\} \sim \text{i.i.d.} N(0, 1/N)$  を要素に持つ  $N \times N$  の実対称ランダム行列  $A = (a_{ij})$  の固有値分布は？
  - Wigner (1955, 1957).

- [1] R. A. Fisher, "Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population," *Biometrika*, vol. 10, no. 4, pp. 507-521, May, 1915.
- [2] J. Wishart, "The generalised product moment distribution in samples from a normal multivariate population," *Biometrika*, vol. 20A, nos. 1/2, pp. 32-52, Jul., 1928.
- [3] E. P. Wigner, "Characteristic vectors of bordered matrices with infinite dimensions," *Annals of Mathematics*, vol. 62, no. 2, pp. 548-564, Mar., 1955.
- [4] E. P. Wigner, "On the distribution of the roots of certain symmetric matrices," *Annals of Mathematics*, vol. 67, no. 2, pp. 325-327, Mar., 1957.

# Wishart 行列の漸近固有値分布

---

- $\Sigma = I_N$  として, ランダム行列  $X \in \mathbb{R}^{N \times p}$  の各成分が独立に  $\mathcal{N}_1(0, 1^2)$  に従うとき

$$XX^T \sim \mathcal{W}_N(p, I_N)$$

- $XX^T \sim \mathcal{W}_N(p, I_N)$  となるランダム行列  $X \in \mathbb{R}^{N \times p}$  のサイズを

$$\alpha = p/N$$

を一定に保ったまま  $N$  を無限大に拡大していくと,  $XX^T$  の固有値の分布がある関数に収束.

## 固有値の経験分布

---

- $XX^T \sim \mathcal{W}_N(p, I_N)$  から標本された固有値  $\{\lambda_1, \dots, \lambda_N\}$  を

$$u_i = \lambda_i/p, \quad i = 1, \dots, N.$$

- $u$  の経験分布 ( $\delta(u)$  は Dirac 測度) :

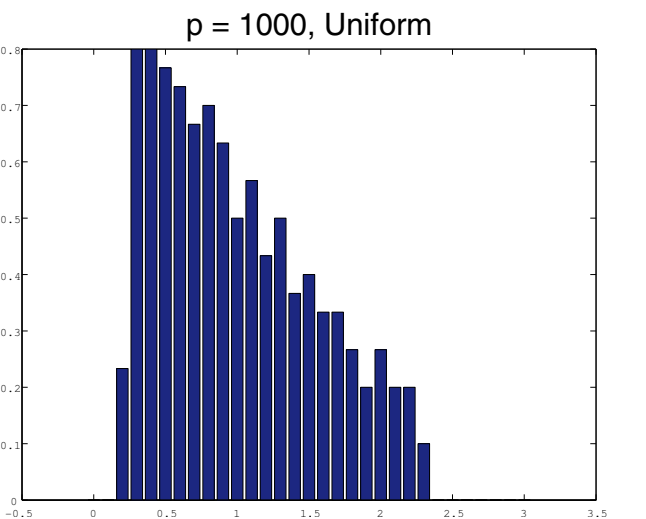
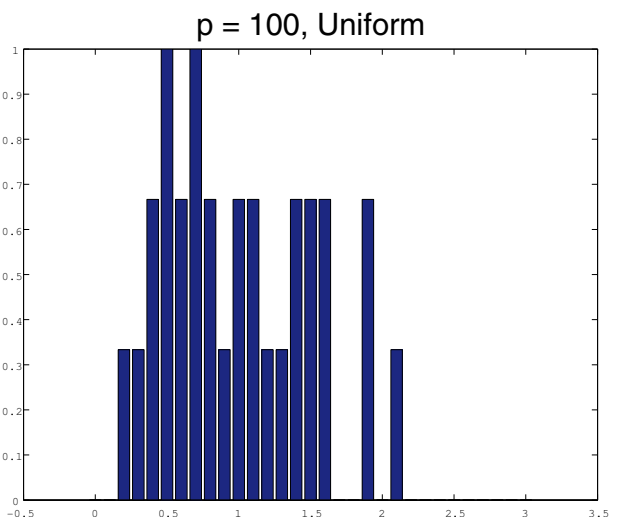
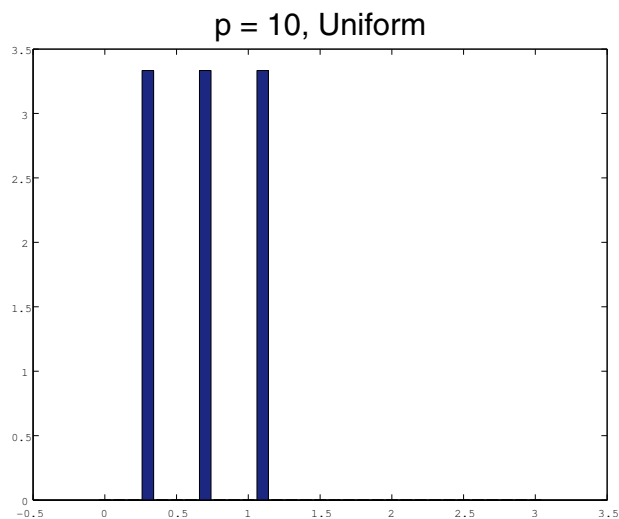
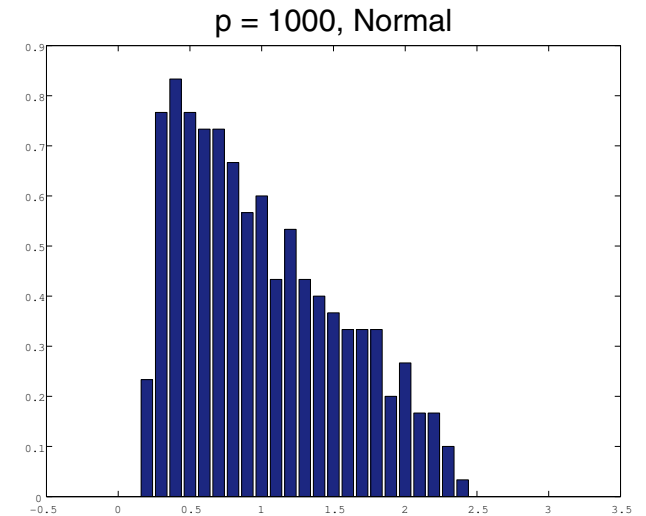
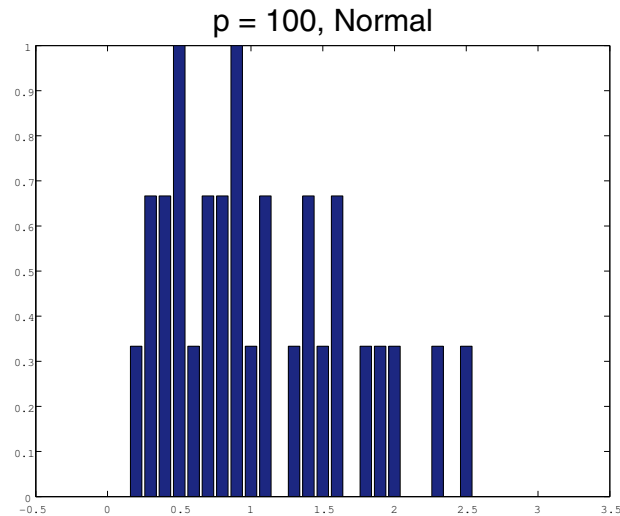
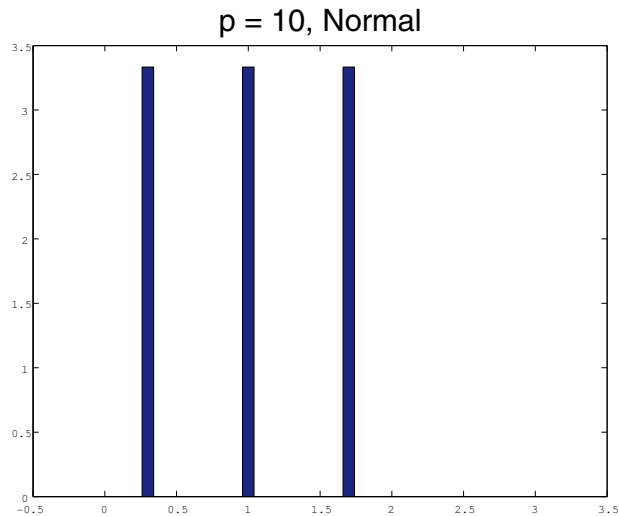
$$\delta P = \frac{1}{N} \{\delta(u_1) + \dots + \delta(u_N)\}$$

- $\alpha = p/N$ ,  $p, N \rightarrow \infty$  で  $\delta P$  は,  $p(u)du$  に概収束.

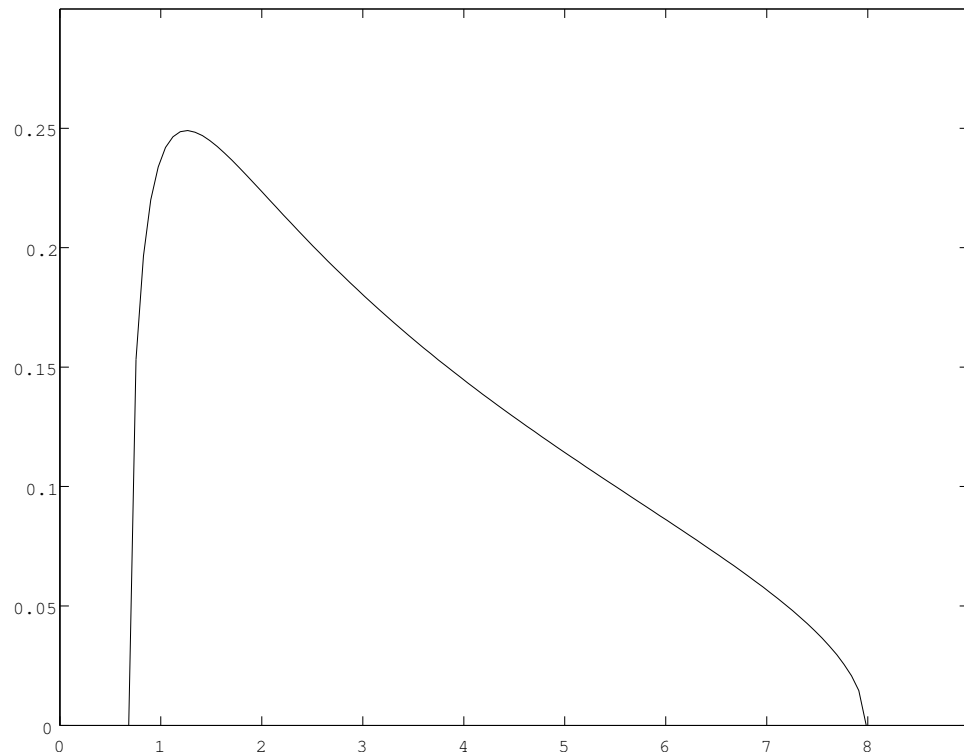
$$p(u) = \begin{cases} \frac{1}{2\pi\alpha} \frac{\sqrt{(u-u_{\min})(u_{\max}-u)}}{u} & \text{if } u_{\min} < u < u_{\max} \\ 0 & \text{otherwise} \end{cases}$$

$$u_{\min} = (\sqrt{\alpha} - 1)^2, \quad u_{\max} = (\sqrt{\alpha} + 1)^2$$

# $XX^T$ の固有値分布 ( $p = 10, 100, 1000$ )



# Marčenko-Pastur 則



- $A$  の固有値分布  $\rho_A(\lambda)$  は

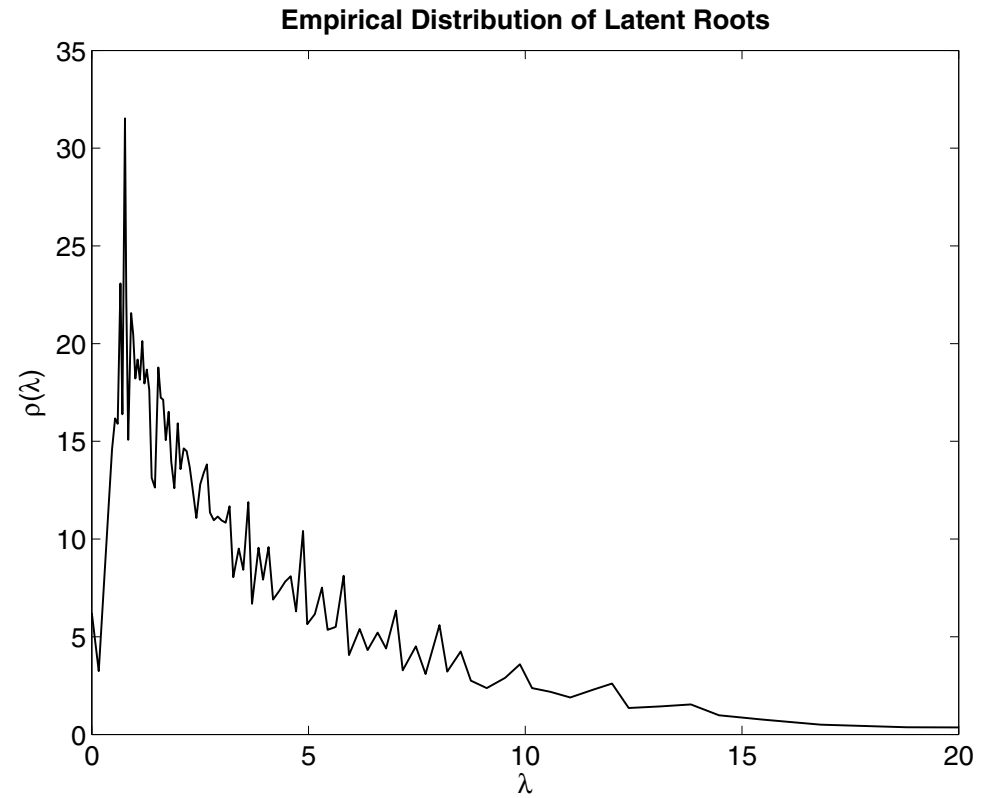
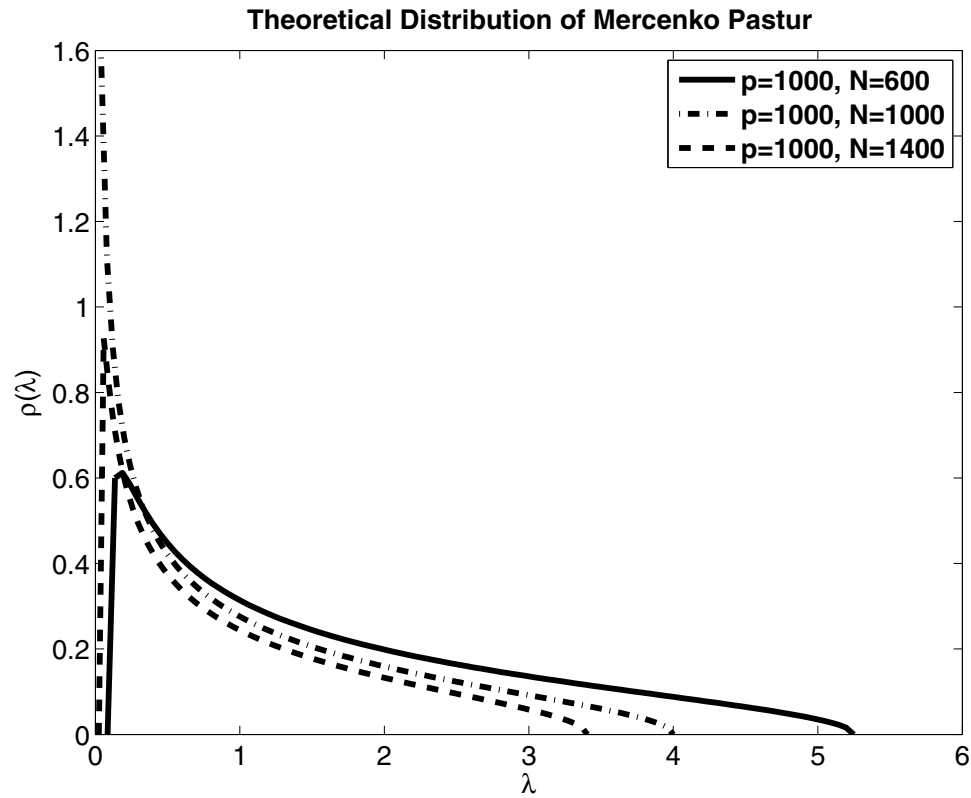
$$\lim_{N \rightarrow \infty} \rho_A(\lambda) = \begin{cases} \frac{1}{2\pi\lambda} \sqrt{(\lambda - \lambda_{\min})(\lambda_{\max} - \lambda)} & \lambda_{\min} \leq \lambda \leq \lambda_{\max} \\ 1 - \alpha & \lambda = 0 \text{ and } \alpha < 1 \\ 0 & \text{otherwise} \end{cases}$$

- 漸近固有値分布は、Marčenko-Pastur 則と呼ばれる以下の式で与えられる。

$$\lambda_{\min} = (1 - \sqrt{\alpha})^2, \lambda_{\max} = (1 + \sqrt{\alpha})^2$$



# 固有値の理論分布と経験分布



# 既存手法による本質的な部分の抽出

---

- そこで、ある閾値  $\theta$  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。
- 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。

## 提案手法による本質的な部分の抽出

---

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、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、先のスライドで説明したように、 $V$ 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。

# 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, 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.
- Tracy-Widom 分布の裾確率を示す下記の表より, 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.

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

$\beta \setminus (1 - \alpha)$	0.995	0.975	0.95	0.05	0.025	0.005
1	-4.1505	-3.5166	-3.1808	0.9793	1.4538	2.4224
2	-3.9139	-3.4428	-3.1945	-0.2325	0.0915	0.7462
4	-4.0531	-3.6608	-3.4556	-1.0904	-0.8405	-0.3400

# ノイズ除去の統計的仮説検定

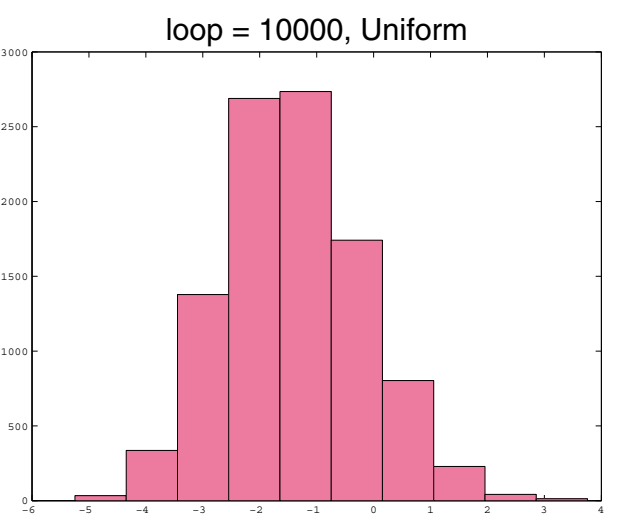
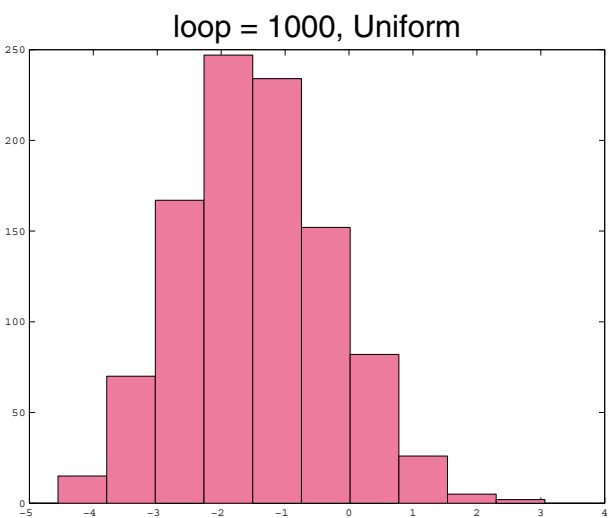
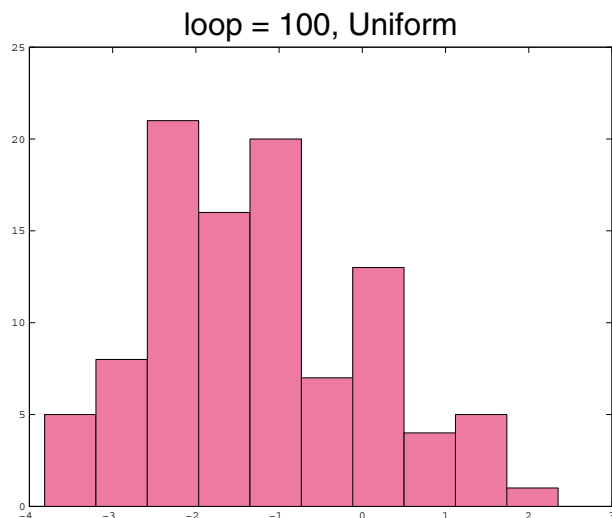
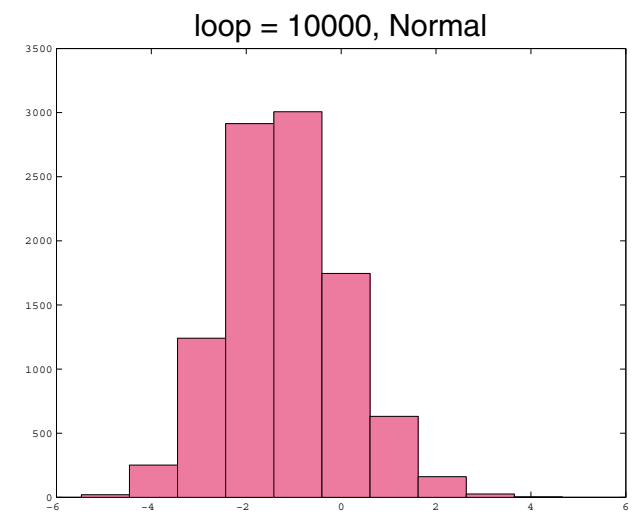
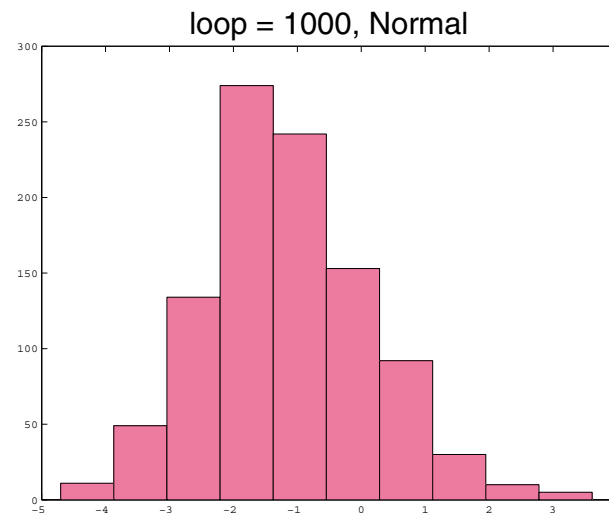
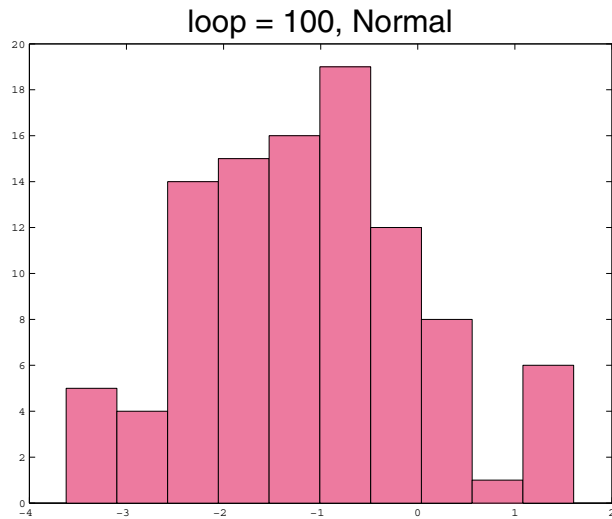
---

ここでは次の2つの仮説検定が考えられる。

- [1] Type I: 本質的部分であるにもかかわらず、ノイズであると判断する可能性を検定。つまり
- $H_0$ : 本質的部分にノイズが含まれる。
  - $H_1$ : 本質的部分にノイズが含まれない。
- [2] Type II: ノイズであるにもかかわらず、本質的部分であると判断する可能性を検定。つまり
- $H_0$ : ノイズに本質的部分が含まれる。
  - $H_1$ : ノイズに本質的部分が含まれない。

よって Type I は Tracy-Widom 分布に対する上側検定、Type II は下側検定となる。Type I の有意水準を  $1 - \alpha$  とすると、Type II は  $\alpha$  となる。

# Tracy-Widom 分布 ( $p = 100, 1000, 10000$ )



# 高頻度金融データの解析

---

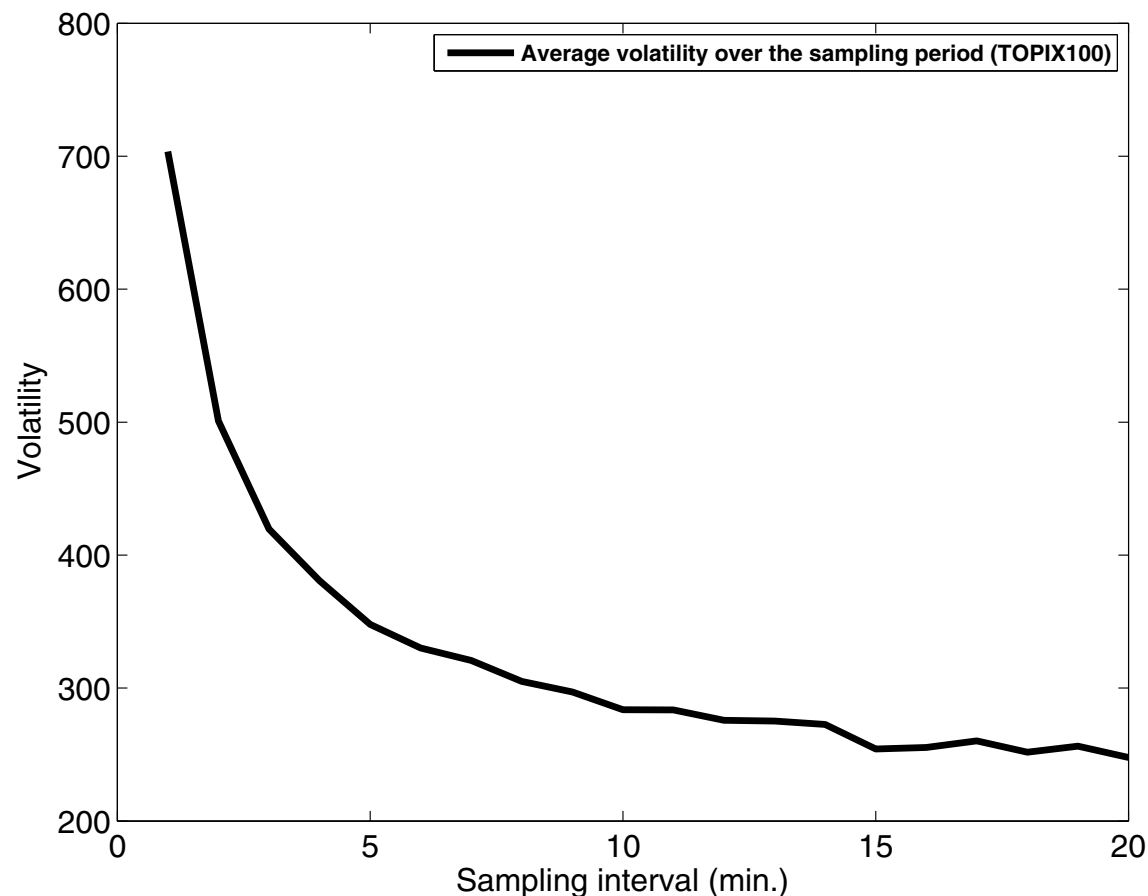
## 実現ボラティリティ (Realized Volatility, RV)

- 近年, [redacted] 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.
  - 積分ボラティリティ (Integrated Volatility, IV) の一致推定量.
  - 高頻度収益率の日内の二乗和.

## 実現共分散 (Realized Covariance, RC)

- また, [redacted] 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.
  - 2変量間の高頻度収益率データの交差積の和.

# ミクロ構造ノイズ



- 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存在.
- ミクロ構造ノイズ: RVによるボラティリティの推定に大きなバイアス.
- 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).



# 結果と考察

---

- 実証分析には、次の銘柄と期間の分次終値データを用いる。

銘柄数： 2007年9月30日時点でNikkei225に含まれる個別銘柄225系列と株価指数TOPIX1系列の計  $N = 226$ 。

期間： 2007年1月4日—2007年12月28日（計245日）。

- 各手法でのノイズ除去性能を評価するため、標本間隔 $\Delta$ を1分から10分まで1分刻みで日内収益率を計算。
- 1取引日の取引時間が4.5時間。

$\Delta = 1$  [min] :  $p = 270, \alpha = 2.67$

# 最小分散ポートフォリオ

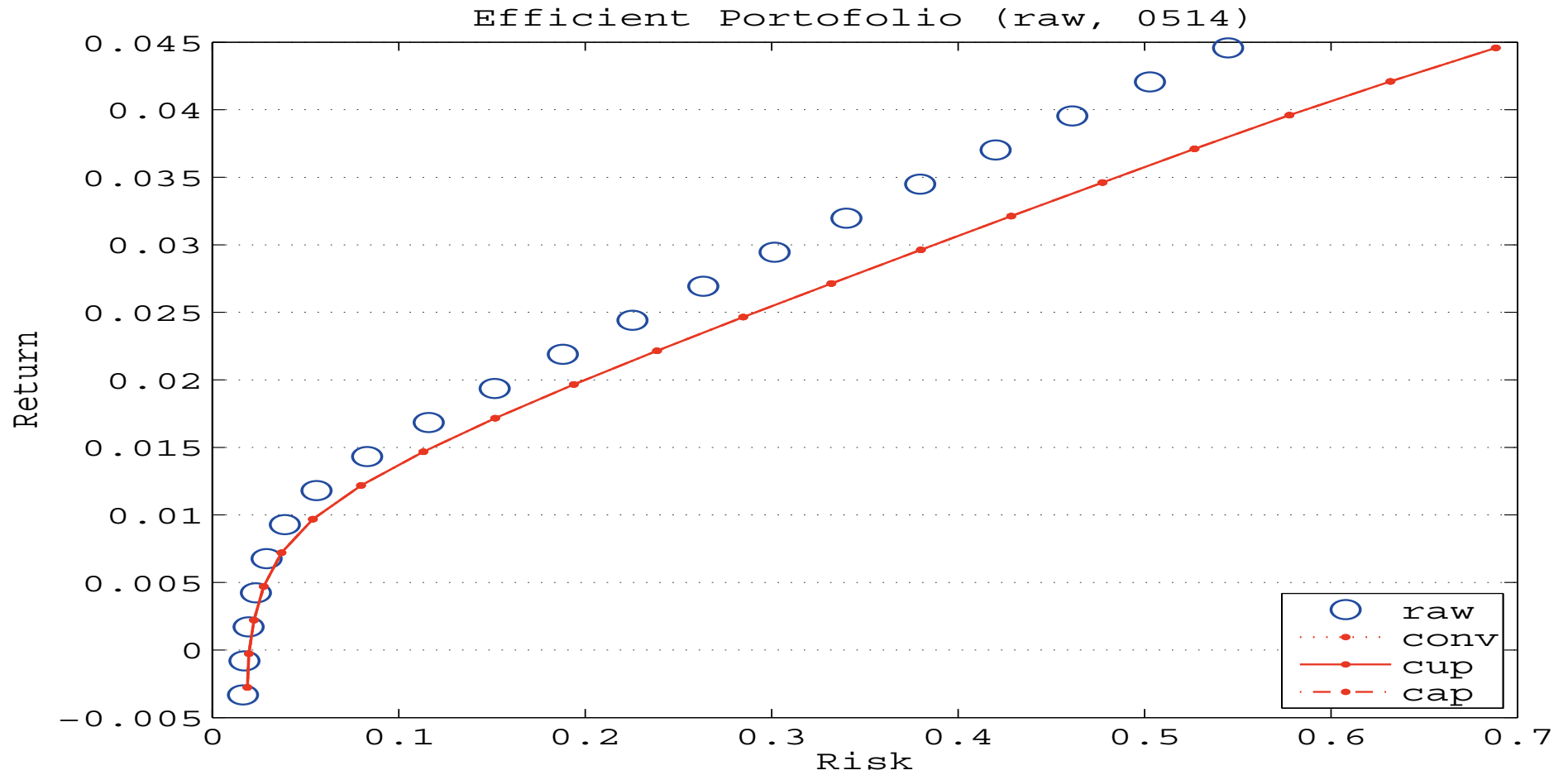
---

- 安全資産が存在しない場合の最小分散（リスク）。

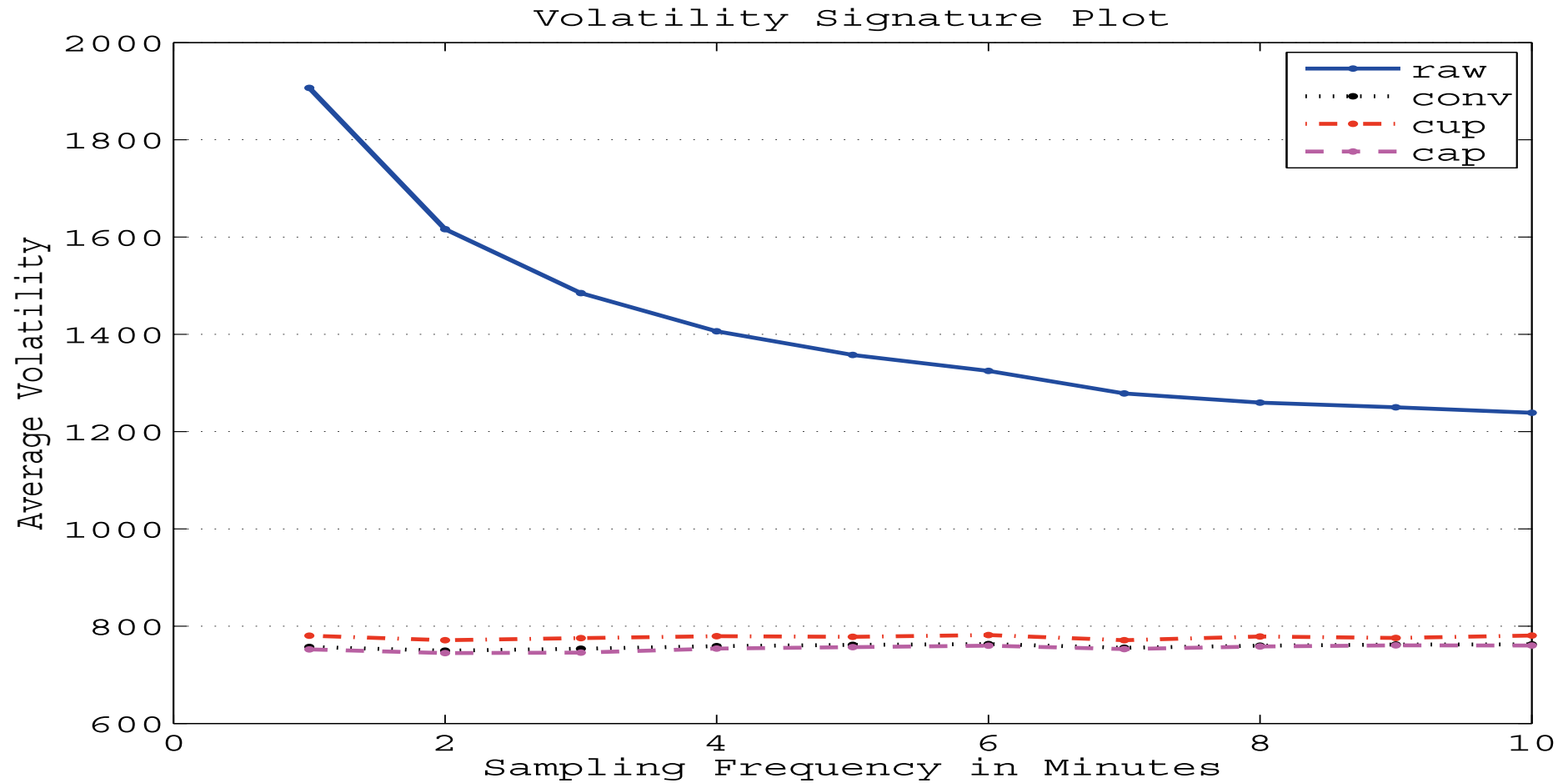
	<i>raw</i>	<i>conv</i>	<i>typeI</i>	<i>typeII</i>
01min.	3.5648	3.3221	3.3123	3.3214
02min.	2.4060	2.2242	2.2130	2.2243
03min.	2.4260	2.3802	2.3807	2.3910
04min.	2.6374	2.3718	2.3553	2.3725
05min.	1.5812	1.4415	1.4195	1.4431
06min.	2.6137	2.4255	2.3724	2.4402
07min.	2.1421	2.3312	2.3266	2.3300
08min.	2.6004	2.2699	2.3505	2.2701
09min.	1.5338	2.1933	2.1939	2.1935
10min.	2.0842	1.7209	1.7280	1.7209

- ノイズの影響が顕著な 05min.以下では，*typeI*が良好な結果。

# 有効フロンティアの比較



# 各標本間隔ごとのボラティリティの平均値



## ボラティリティの平均値

	<i>raw</i>	<i>conv</i>	<i>typeI</i>	<i>typeII</i>
01min.	1.9067	0.7572	0.7803	0.7521
02min.	1.6163	0.7495	0.7709	0.7444
03min.	1.4847	0.7536	0.7753	0.7457
04min.	1.4064	0.7589	0.7793	0.7540
05min.	1.3577	0.7615	0.7780	0.7569
06min.	1.3248	0.7630	0.7818	0.7597
07min.	1.2785	0.7556	0.7711	0.7528
08min.	1.2597	0.7597	0.7788	0.7582
09min.	1.2501	0.7620	0.7757	0.7605
10min.	1.2389	0.7625	0.7806	0.7603
S.D.	210.6982	4.3976	3.8342	5.7993

(各値は  $\times 1/1000$ , S.D. は  $\times 1$ )

- *typeI*のS.D.値を見ると、各標本間隔を通して安定的にボラティリティを推定していることが分かる。

# まとめ

---

- [1] Wigner の半円則, Marčenko-Pastur 則はいずれも, 行列要素が従う分布の詳細に依存しない普遍的な法則であることを説明.
- [2] 実現 (共) ボラティリティを推定する上で, 問題となる観測不可能なミクロ構造ノイズを緩和し, 推定量のバイアスを取り除く試みの一つとして, ランダム行列理論を適用.
- [3] ここで提案した手法が既存手法と比較して全般的に良い結果.
- [4] 今後の課題としては,
  - 時系列構造の導入,
  - 共ボラティリティの予測力比較, など.

## お問い合わせ先

---

関西学院大学 理工学部 数理科学科

専任講師 森本 孝之

TEL 079-565-7118

FAX 079-565-7118

EML morimot@kwansei.ac.jp

URL <http://sci-tech.ksc.kwansei.ac.jp/~morimot/>