Large Volatility Matrix Estimation for High-Frequency Financial Data

Yazhen Wang

University of Wisconsin-Madison

Workshop on Self-normalized Asymptotic Theory in Probability, Statistics and Econometrics

Institute for Mathematical Sciences, National University of Singapore

May 19-23, 2014

Joint work with Donggyu Kim, Yi Liu, Minjing Tao and Harry Zhou

Outline

1. High-Frequency Data
2. Large Volatility Matrix Estimation
3. Convergence Rate and Positiveness
4. Numerical Studies

High-Frequency Finance

High-Frequency Data: Intradaily observations on asset prices such as tick by tick stock price data and minute by minute exchange rate data.

Data Characteristics: High-frequency data have complex structure with microstructure noise.

One-Dim Model: Observed data: $Y_{t_i},\ i = 1, \cdots, n$ and X_t = true log-price of a stock

$$\mathbf{Y_{t_i} = X_{t_i} + \varepsilon_{t_i}}, \qquad i = 1, \cdots, n$$

ε_{t_i}: microstructure noise and independent of X_t.

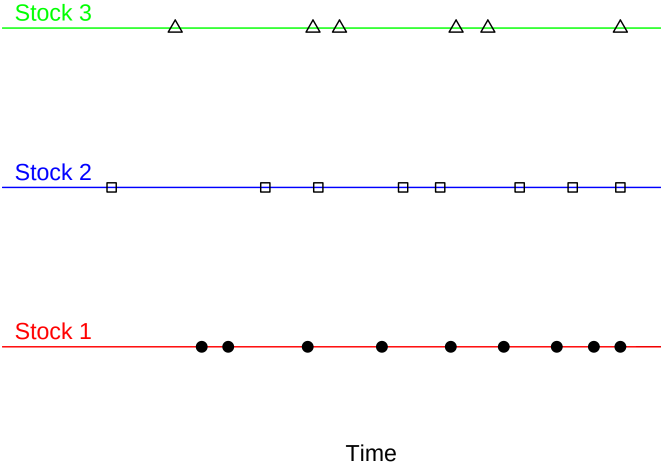
Very High Dim: Large Volatility Matrix

High Dim Model: For the i-th asset, observation times t_{ij}, $i = 1, \cdots, p$, $j = 1, \cdots, n_i$ and observed log price $Y_i(t_{i,j})$,

$$Y_i(t_{i,j}) = X_i(t_{i,j}) + \varepsilon_i(t_{i,j}),$$

$X_i(t)$: true log price of asset i, and microstructure noise $\varepsilon_i(\cdot)$: i.i.d. with zero mean, and independent of $X_i(t)$.

Nonsynchronization: stocks' transactions occur at distinct times and the prices of different stocks are recorded at mismatched time points.



Stock 3
Stock 2
Stock 1
Time

Price Model

$X_t = (X_{1t}, \cdots, X_{pt})^\dagger$: log price of p assets

$$dX_t = \mu_t\, dt + \sigma_t\, dW_t, \qquad t \in [0, 1],$$

where W_t: p-dimensional BM, and σ_t: $p \times p$ matrix.

Integrated volatility matrix:

$$\Gamma = \int_0^1 \gamma(t)\, dt, \qquad \gamma(t) = \sigma_t \sigma_t^\dagger$$

Goal: Estimate Γ based on data $Y_i(t_{ij})$.

Methodology:

1. Form volatility matrix estimators
2. Regularize the matrix estimators

Realized co-volatility based on previous tick

$\tau = \{\tau_r = r/m, r = 1, \cdots, m\}$: pre-determined sampling frequency.

For asset i, select previous-tick times:

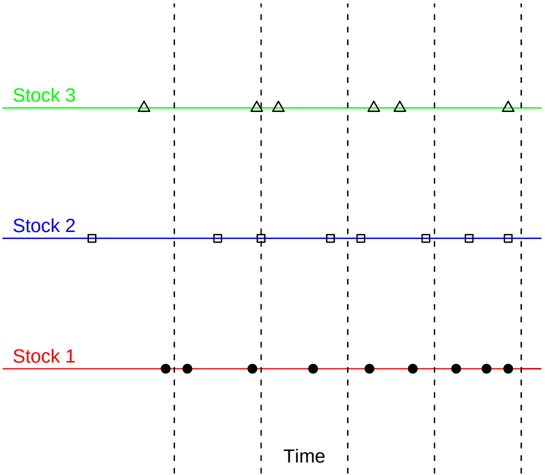
$$\tau_{i,r} = \max\{t_{i_s,j} \le \tau_r, j = 1, \cdots, n_{i_s}\}, \qquad r = 1, \cdots, m$$

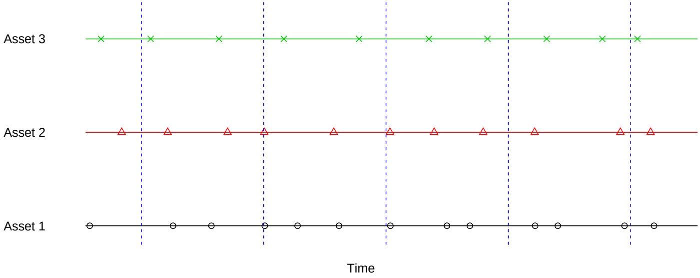
Realized co-volatility $\hat{\Gamma}_{i_1 i_2}(\tau)$ between assets i_1 and i_2:

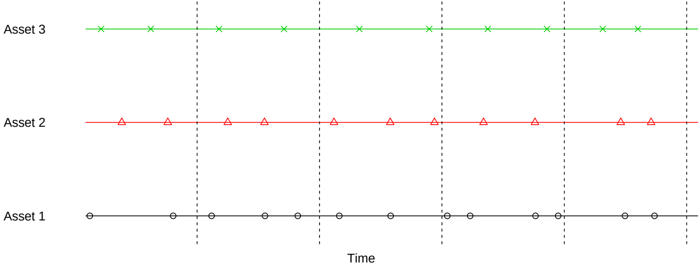
$$\hat{\Gamma}_{i_1,i_2}(\tau) = \sum_{r=1}^{m} [Y_{i_1}(\tau_{i_1,r}) - Y_{i_1}(\tau_{i_1,r-1})]\,[Y_{i_2}(\tau_{i_2,r}) - Y_{i_2}(\tau_{i_2,r-1})].$$

Realized co-volatility matrix: $\hat{\mathbf{\Gamma}}(\tau) = (\hat{\Gamma}_{i_1 i_2}(\tau))$

Data synchronization: Previous tick




Asset 3
Asset 2
Asset 1
Time


Asset 3
Asset 2
Asset 1
Time

Two-scale

Realized volatility matrix estimator

$$\tau^k = \tau + (k-1)/n, \qquad k = 1, \cdots, K = [n/m]$$

$$\widehat{\boldsymbol{\Gamma}}^K = \frac{1}{K}\sum_{k=1}^{K} \hat{\boldsymbol{\Gamma}}(\tau^k), \;\; n = \sum_{i=1}^{p} n_i/p$$

where $\widehat{\Gamma}_{ii}$ are adjusted by subtracting them from estimated noise variance components: $\frac{2\,m}{n_i}\sum_{\ell=1}^{n_i}[Y_i(t_{i,\ell}) - Y_i(t_{i,\ell-1})]^2$

Asymptotics:

$$\text{Entrywise} \qquad \widehat{\boldsymbol{\Gamma}}^K - \boldsymbol{\Gamma} = O_P(n^{-1/6}), \qquad \text{if } K \sim n^{2/3}$$

Multi-scale

Realized volatility matrix estimator

$$\hat{\mathbf{\Gamma}} = \sum_{m=1}^{M} a_m \hat{\mathbf{\Gamma}}^{K_m} + \zeta(\hat{\mathbf{\Gamma}}^{K_1} - \hat{\mathbf{\Gamma}}^{K_M}), \quad (1)$$

where $K_m = m + N$,

$$a_m = \frac{12(m+N)(m - M/2 - 1/2)}{M(M^2 - 1)}, \quad \zeta = \frac{(M+N)(N+1)}{(n+1)(M-1)} \quad (2)$$

Asymptotics:

Entrywise $\quad \widehat{\mathbf{\Gamma}} - \mathbf{\Gamma} = O_P(n^{-1/4}), \quad$ if $M, N \sim n^{1/2}$

Data Synchronization: Refresh time

Asset i has trading time t_j^i, $j = 1, \cdots, n^i = N_1^i$

$N_t^i = \#$ of t_j^i, $j = 1, \cdots, n^i$, that $\leq t$

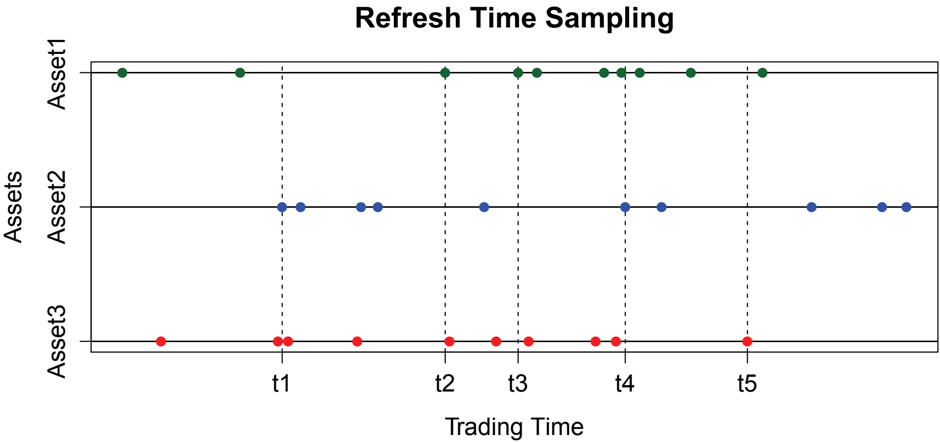
Define 1st refresh time: $\tau_1 = \max(t_1^1, ..., t_1^p)$, subsequent refresh times

$$\tau_{j+1} = \max(t^1_{N^1_{\tau_j}+1}, ..., t^p_{N^p_{\tau_j}+1}).$$

Intuition: τ_1 is the time all their posted prices have been updated (i.e it has taken for all assets to trade); τ_2 is the first time when all the prices are again updated.

At each refreshed time τ_j, one new price and $p - 1$ stale prices

Let $m =$ the resulting Refresh Time sample size


Refresh Time Sampling
Assets
Asset1
Asset2
Asset3
t1
t2
t3
t4
t5
Trading Time

Realized kernel based on refresh time

Synchronized data: at refresh time τ_j, $j = 1, \cdots, m$, define

$$Y_i(\tau_j) = \text{ observation of asset } i \text{ at time point } t^i_{N^i_{\tau_j}}$$

$$\boldsymbol{Y}_j(\cdot) = (Y_1(\tau_j), \cdots, Y_p(\tau_j))', \;\; \mathbf{y}_j = \boldsymbol{Y}(\tau_j) - \boldsymbol{Y}(\tau_{j-1}), \;\; j = 1, \cdots, m$$

$$\gamma_h = \sum_{j=|h|+1}^{m} \mathbf{y}_j \mathbf{y}'_{j-h}, \;\; h \geq 0 \qquad \gamma_h = \sum_{j=|h|+1}^{m} \mathbf{y}_{j-1} \mathbf{y}'_j, \;\; h < 0$$

Realized Kernel estimator $H(\boldsymbol{Y}) = \sum_{h=-m}^{m} H\left(\frac{h}{K+1}\right) \gamma_h$ for kernel $H(\cdot)$

Asymptotics: $H(\boldsymbol{Y})$ can be semi-positive but with entrywise convergence rate $n^{-1/5}$

Or it may achieve $n^{-1/4}$ convergence rate but not semi-positive.

Data Synchronization: Generalized sampling time

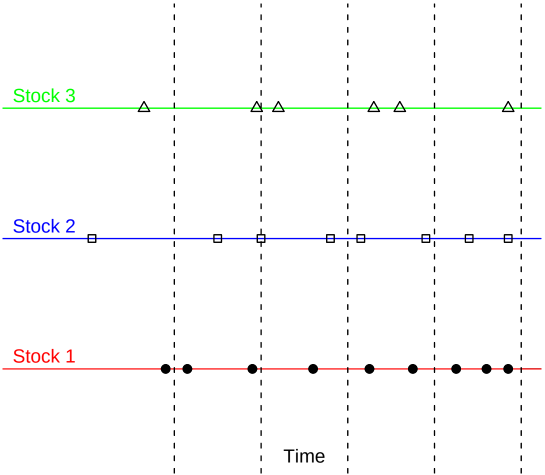
A sequence of time points $\{\tau_0, \tau_1, \tau_2, ..., \tau_m\}$ is said to be the Generalized Sampling Time for a collection of p assets, if

1. $0 = \tau_0 < \tau_1 < \cdots < \tau_m = T$.
2. Each asset has at least one observation between consecutive τ_i's.
3. Time intervals, $\{\Delta_j = \tau_j - \tau_{j-1}, 1 \leqslant j \leqslant m\}$, satisfy $\sup_i \Delta_i \xrightarrow{P} 0$.

The synchronized data sets are generated by selecting an arbitrary observation $Y_i(\check{t}_j)$ of the ith asset in time interval $(\tau_{j-1}, \tau_j]$. Therefore, the synchronized data are $\{Y_i(\tau_j), 1 \leqslant i \leqslant p, 1 \leqslant j \leqslant m\}$ such that $Y_i(\tau_j) = Y_i(\check{t}_j)$.

It allows to choose the sampling times by requiring each asset to lead in turn.

Data synchronization: Generalized sampling time



Estimators based on generalized sampling time

Previous-tick and refresh time schemes may be treated as special cases of generalized sampling time.

With synchronized data $Y_i(\tau_j)$ we may define MSRV estimator, Realized kernel estimator, and quasi-MLE to achieve entrywise convergence rate $n^{-1/4}$.

Pre-averaging estimator

Define the difference of two local averages at j/n,

$$\bar{Y}_{ij} = \frac{2}{K}\left[\sum_{\ell=K/2}^{K-1} Y_i(t_{i,j+\ell}) - \sum_{j=0}^{K/2-1} Y_i(t_{i,j+\ell})\right],$$

and realized volatility based on all $\bar{Y}_{ij}$

$$ARV = \frac{1}{K}\sum_{i=0}^{n-K+1} (\bar{Y}_{1j}, \cdots, \bar{Y}_{pj})'(\bar{Y}_{1j}, \cdots, \bar{Y}_{pj}).$$

Asymptotics: ARV can be semi-positive but with entrywise convergence rate $n^{-1/5}$.

Or adjust the biases of ARV for the noise variances to obtain PARV with $n^{-1/4}$ convergence rate but not semi-positive.

Matrix Size

Dimension Reduction For moderate to large p, $(p^2 + p)/2$ entries in Γ: too many parameters and too much random fluctuation.

Issue: Usual dimension reduction techniques are **not applicable** to non-synchronized data.

Numerical Illustration

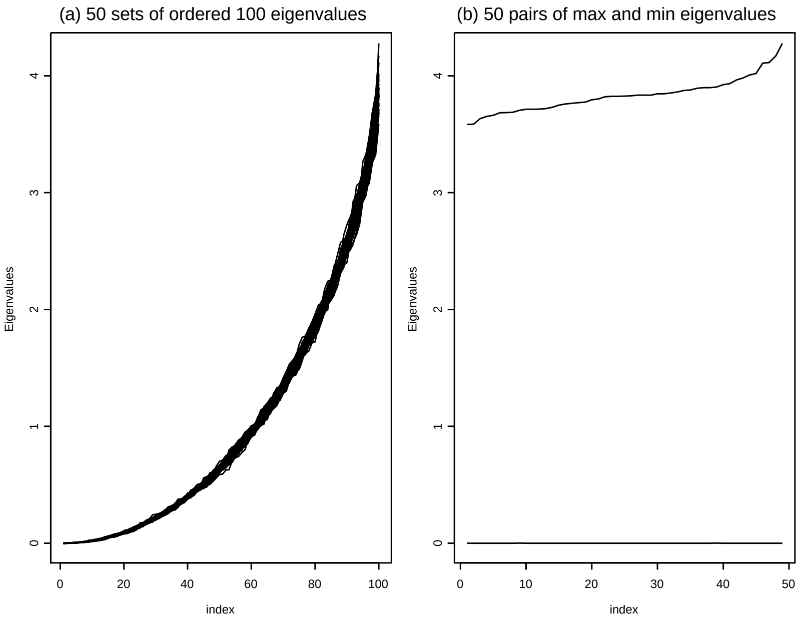
$X(t) = (W_1(t), \cdots, W_p(t))$: vector of p independent Brownian motions.

Observations = $X(k/n)$, $k = 0, 1, \cdots, n$.

$$\Gamma = I_p, \qquad \widehat{\Gamma} = \left(\widehat{\Gamma}_{ij}\right), \qquad \widehat{\Gamma}_{ij} = \frac{1}{n}\sum_{k=1}^{N} Z_{ik} Z_{jk}$$

$$Z_{ik} = \sqrt{n}[W_i(k/n) - W_i((k-1)/n)] \sim N(0,1)$$

Take $n = 100$ and $p = 100$. We compute the eigenvalues of $\widehat{\Gamma}$ in a simulation with 50 replications.


(a) 50 sets of ordered 100 eigenvalues
Eigenvalues
0
1
2
3
4
0
20
40
60
80
100
index
(b) 50 pairs of max and min eigenvalues
Eigenvalues
0
1
2
3
4
0
10
20
30
40
50
index

Regularize Volatility Matrix

Write $\mathbf{\Gamma} = (\Gamma_{ij})$

Sparsity: assume $\mathbf{\Gamma}$ has a sparse representation

$$\sum_{j=1}^{p} |\Gamma_{ij}|^{\delta} \leq M\,\pi(p), \qquad i = 1, \cdots, p, \qquad E[M] \leq C,$$

where $0 \leq \delta < 1$ and $\pi(p) = 1, \log p$, or a small power of p.

Examples: (1) Block diagonal matrix

(2) Matrix with decay elements from diagonal

(3) Matrix with small number of non-zero elements in each row

(4) Random permutations of rows and columns for above matrices

Estimation with regularization

Write $\hat{\mathbf{\Gamma}} = (\hat{\Gamma}_{ij})$ as any of volatility matrix estimators

Thresholding: for sparse $\mathbf{\Gamma}$, regularize $\hat{\mathbf{\Gamma}}$ by hard and soft thresholding rules

$$\text{Hard}: \quad \mathcal{T}_{\varpi}[\hat{\mathbf{\Gamma}}] = \left(\hat{\Gamma}_{ij} 1(|\hat{\Gamma}_{ij}| \geq \varpi, i \neq j)\right) + \left(\hat{\Gamma}_{ij} 1(|\hat{\Gamma}_{ij}| \geq 0, i = j)\right),$$

$$\text{Soft:} \mathcal{T}_{\varpi}[\hat{\mathbf{\Gamma}}] = \left((\hat{\Gamma}_{ij} - sign(\hat{\Gamma}_{ij})\varpi) 1(|\hat{\Gamma}_{ij}| \geq \varpi, i \neq j)\right) + \left(\hat{\Gamma}_{ij} 1(|\hat{\Gamma}_{ij}| \geq 0, i = j)\right),$$

where ϖ is a threshold.

Technical conditions

A1: $\varepsilon_i(\cdot)$, $\mu_i(t)$, and $\sigma_{ii}^{1/2}(t)$ all have bounded 2β moments.

A2: Each asset has at least one observation between consecutive time points of the selected sampling frequency. With $n = (n_1 + \cdots + n_p)/p$,

$$C_1 \leq \min_{1 \leq i \leq p} \frac{n_i}{n} \leq \max_{1 \leq i \leq p} \frac{n_i}{n} \leq C_2, \qquad \max_{1 \leq i \leq p} \max_{1 \leq \ell \leq n_i} |t_{i\ell} - t_{i,\ell-1}| = O(n^{-1}).$$

Asymptotic Theory: Convergence rate

Matrix norm

$$\|\mathbf{\Gamma}\|_2 = \sup\{\|\mathbf{\Gamma}\,\mathbf{x}\|_2, \|\mathbf{x}\|_2 = 1\} = \text{max absolute eigenvalue}$$

Theorem Assume sparsity on $\mathbf{\Gamma}$ and conditions A1-A2. If $\hat{\mathbf{\Gamma}}$ is any of volatility matrix estimators using **(i)** multi-scale based on previous-tick or generalized sampling scheme; **(ii)** realized kernel based on refresh time or generalized sampling scheme; **(iii)** pre-averaging, and $\hat{\mathbf{\Gamma}}$ has entrywise convergence rate $n^{-1/4}$, we have

$$\|\mathcal{T}_{\varpi}[\hat{\mathbf{\Gamma}}] - \mathbf{\Gamma}\|_2 = O_P\left(\pi(p)\,\left[n^{-1/4}p^{2/\beta}\right]^{1-\delta}\right),$$

where $\varpi \sim n^{-1/4}\,p^{2/\beta}$.

Asymptotic Theory: Semi-positiveness

Theorem Assume sparsity on $\mathbf{\Gamma}$ and conditions A1-A2. If $\hat{\mathbf{\Gamma}}$ is one of **(a)** semi-positive pre-averaging estimator; **(b)** semi-positive realized kernel estimator based on refresh time or generalized sampling scheme, and $\hat{\mathbf{\Gamma}}$ has entrywise convergence rate $n^{-1/5}$, then $\mathcal{T}_\varpi[\hat{\mathbf{\Gamma}}]$ is semi-positive and

$$\|\mathcal{T}_\varpi[\hat{\mathbf{\Gamma}}] - \mathbf{\Gamma}\|_2 = O_P\left(\pi(p)\ \left[n^{-1/5}p^{2/\beta}\right]^{1-\delta}\right),$$

where $\varpi \sim n^{-1/5}\, p^{2/\beta}$.

Asymptotic optimality: Multi-scale RV with sub-Gaussian noise

$\varepsilon_i(\cdot)$ have sub-Gaussian tail, $\mu_i(t)$ and $\sigma_{ii}(t)$ are bounded, data are synchronized.

Theorem For sparse $\mathbf{\Gamma}$, we have for multi-scale RV based on previous-tick scheme,

$$\|\mathcal{T}_{\varpi}[\hat{\mathbf{\Gamma}}] - \mathbf{\Gamma}\|_2 = O_P\left(\pi(p)\ \left[n^{-1/4}\sqrt{\log p}\right]^{1-\delta}\right),$$

where $\varpi \sim n^{-1/4}\sqrt{\log p}$.

Lower bound: For sparse $\mathbf{\Gamma}$, we have that for any estimator $\check{\mathbf{\Gamma}}$

$$\inf_{\check{\mathbf{\Gamma}}} \sup_{\mathbf{\Gamma}} \mathbb{E}\|\check{\mathbf{\Gamma}} - \mathbf{\Gamma}\|_2 \asymp c\pi(p)\ \left[n^{-1/4}\sqrt{\log p}\right]^{1-\delta},$$

where c is a generic constant.

Optimal Estimator: Threshold Multi-scale RV based on previous-tick

Optimality

Take $\boldsymbol{\mu}_t = 0$, $\sigma_t = \sigma$, $\boldsymbol{\Gamma} = \sigma\,\sigma^\dagger$,

$$Y_i(t_j) = \sigma\,\boldsymbol{W}_{t_j} + \varepsilon_i(t_j), \qquad j = 1, \cdots, n$$

$$Y_i(t_j) - Y_i(t_{j-1}) = \sigma\,[\boldsymbol{W}_{t_j} - \boldsymbol{W}_{t_{j-1}}] + \varepsilon_i(t_j) - \varepsilon_i(t_{j-1})$$

Take discrete sine transform

$$\boldsymbol{U}_l = \sigma\,\boldsymbol{V}_l + \kappa \sin\left[\frac{\pi l}{2(n+1)}\right] \boldsymbol{e}_l, \qquad l = 1, \cdots, n$$

where $\boldsymbol{V}_l \sim N(0, \mathbf{I}_p)$, $\boldsymbol{e}_l \sim N(0, \mathbf{I}_p)$

Covariance matrix estimation: Lower Bound

$$\boldsymbol{U}_l \sim N\left(0, \boldsymbol{\Gamma} + \kappa^2 \sin^2\left[\frac{\pi l}{2(n+1)}\right]\mathbf{I}_p\right), \quad l = 1, \cdots, n$$

Theorem For sparse $\boldsymbol{\Gamma}$, we have that for any estimator $\check{\boldsymbol{\Gamma}}$

$$\inf_{\check{\boldsymbol{\Gamma}}} \sup_{\boldsymbol{\Gamma}} \mathbb{E}\|\check{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}\|_2 \asymp c\pi(p) \left[n^{-1/4}\sqrt{\log p}\right]^{1-\delta},$$

where c is a generic constant.

Optimal Estimator: $\mathcal{T}_{\varpi}[\hat{\boldsymbol{\Gamma}}]$.

Some Insights

$$Y_i(t_{i,j}) = X_i(t_{i,j}) + \varepsilon_i(t_{i,j})$$

$$X_i(t_{i,j}) = \int_0^{t_{i,j}} \mu_{i,s}\, ds + \int_0^{\mathbf{t_{i,j}}} \sigma_{\mathbf{i,s}}\, \mathbf{dW_s}$$

(1) Noise = Measurement Error: $n^{-1/4}$

(2) SubGaussian: $\sqrt{\log p}$

(3) Matrix Sparsity: $\pi(p) \& \delta$

Concluding Remarks

• Good matrix estimators may perform poorly when the matrix size is very large. We need to regularize large sample covariance and volatility matrix estimators.

• For sparse volatility matrices, thresholding yields great or even optimal volatility matrix estimators.

http://www.stat.wisc.edu/~yzwang