

# High Frequency Data and Realized Volatility

Taro Kanatani

Graduate School of Economics, Kyoto University

E-mail: [taro@e02.mbox.media.kyoto-u.ac.jp](mailto:taro@e02.mbox.media.kyoto-u.ac.jp)

January 2005

# Table of Contents

1. Introduction
2. Weighed realized volatility (Chapter 3)
3. Monte Carlo study
4. Summary

# Introduction

- Estimation of volatility (Conditional variance of asset return).  
⇒ VaR, Option pricing, etc.
- High-frequency data.  
e.g. Hourly data, 30 minutes data, ..., 5 minutes data, Transaction data (Raw data)

DGP:

$$\underbrace{dp(t)}_{n \times 1} = \underbrace{\mu(t)}_{n \times 1} dt + \underbrace{\Sigma(t)}_{n \times n} \underbrace{dz(t)}_{n \times 1}, \quad 0 \leq t \leq T$$

Observed points:

$$0 = t_0^i < t_1^i < \dots < t_k^i < \dots < t_{N_i-1}^i < t_{N_i}^i = T$$

Volatility matrix:

$$\Omega(t) \equiv \Sigma(t) \Sigma(t)'$$

Estimation of integrated volatility  $\int_0^T \Omega(t) dt$

⇒ Quadratic variation:

$$\lim_{N_i \rightarrow \infty} \underbrace{\sum_{k=1}^{N_i} (\Delta p_i(t_k^i))^2}_{\text{Realized volatility}} = \int_0^T \omega_{ii}(t) dt$$

## Weighted realized volatility

$$\hat{\omega}_{ij} = \sum_{k=1}^{N_i} \sum_{l=1}^{N_j} w_{kl} \Delta p_i(t_k^i) \Delta p_j(t_l^j) = \underbrace{\Delta p_i'}_{1 \times N_i} \underbrace{\widehat{W}}_{N_i \times N_j} \underbrace{\Delta p_j}_{N_j \times 1}$$

where  $\Delta p_i = (\Delta p_i(t_1^i), \dots, \Delta p_i(t_{N_i}^i))'$ .

Examples:

1. Realized volatility calculated by using interpolated data
2. Realized volatility calculated by using raw data
3. Fourier series estimator (Malliavin and Mancino, 2002)

## Interpolation and realized volatility

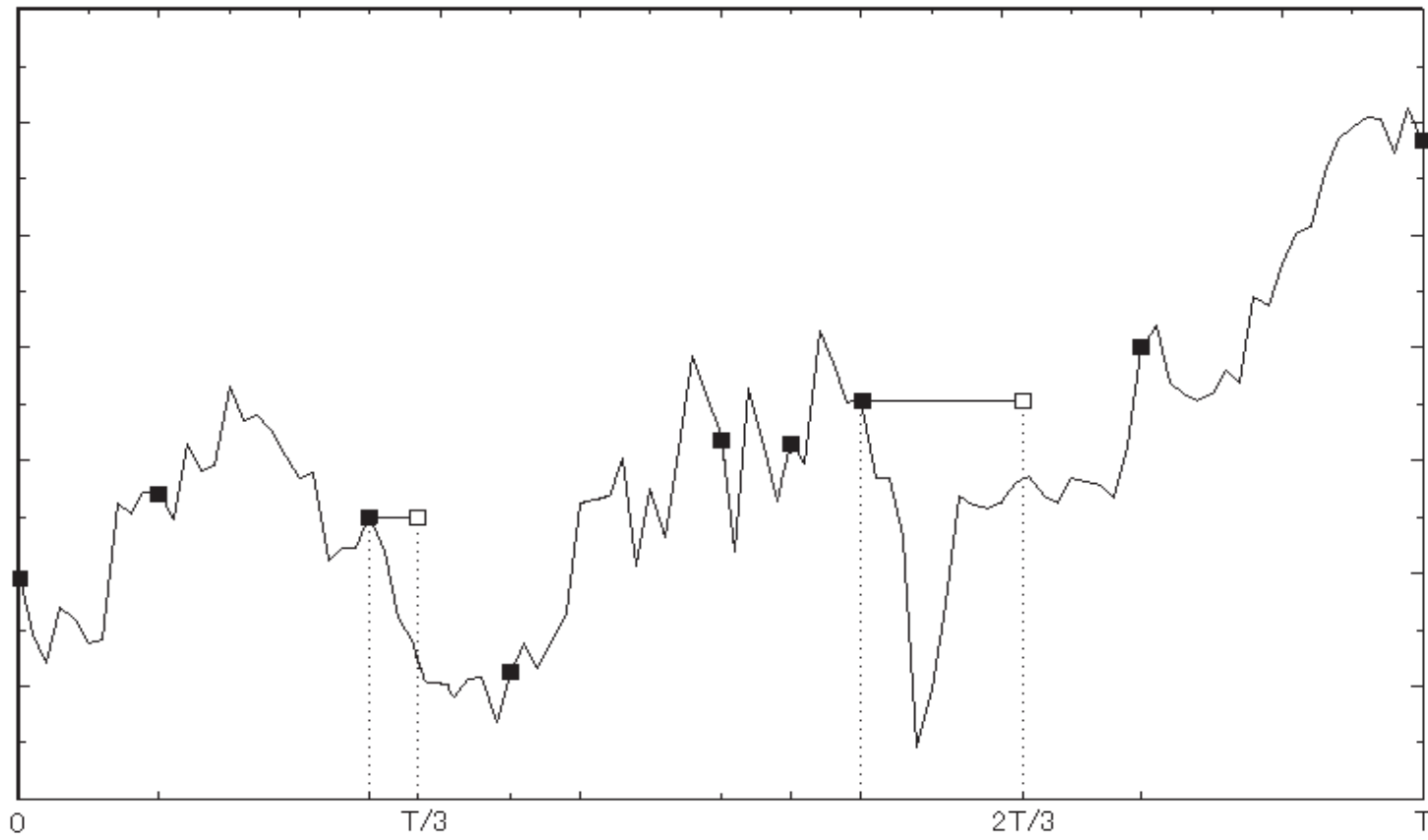
■  $\overbrace{\{p_i(t_k^i)\}_{k=0}^{N_i}}^{\text{Unevenly spaced data}} \rightarrow \boxed{\text{Interpolation}} \rightarrow \overbrace{\{q_i(mT/M)\}_{m=0}^M}^{\text{Evenly spaced data}}$

■ Realized (cross) volatility

$$\hat{\omega}_{ij} = \sum_{m=1}^M \left\{ q_i \left( \frac{mT}{M} \right) - q_i \left( \frac{(m-1)T}{M} \right) \right\} \left\{ q_j \left( \frac{mT}{M} \right) - q_j \left( \frac{(m-1)T}{M} \right) \right\}$$

$\Rightarrow$  Interpolation bias

# Previous-tick interpolation

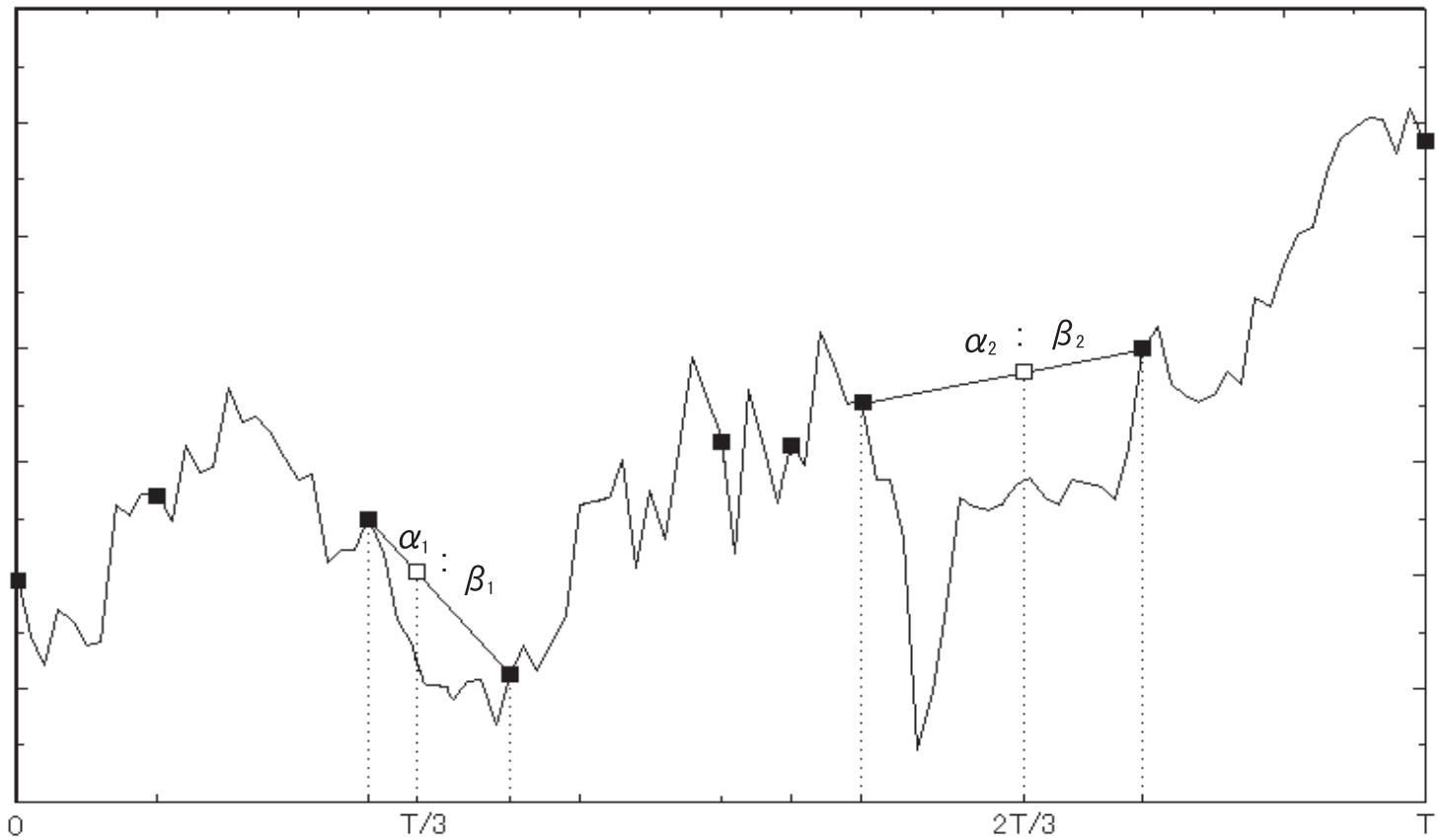


■ Observed point; □ Interpolated point

Weight matrix for the previous-tick interpolation ( $i = j$ ):

$$W = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

# Linear interpolation



■ Observed point; □ Interpolated point

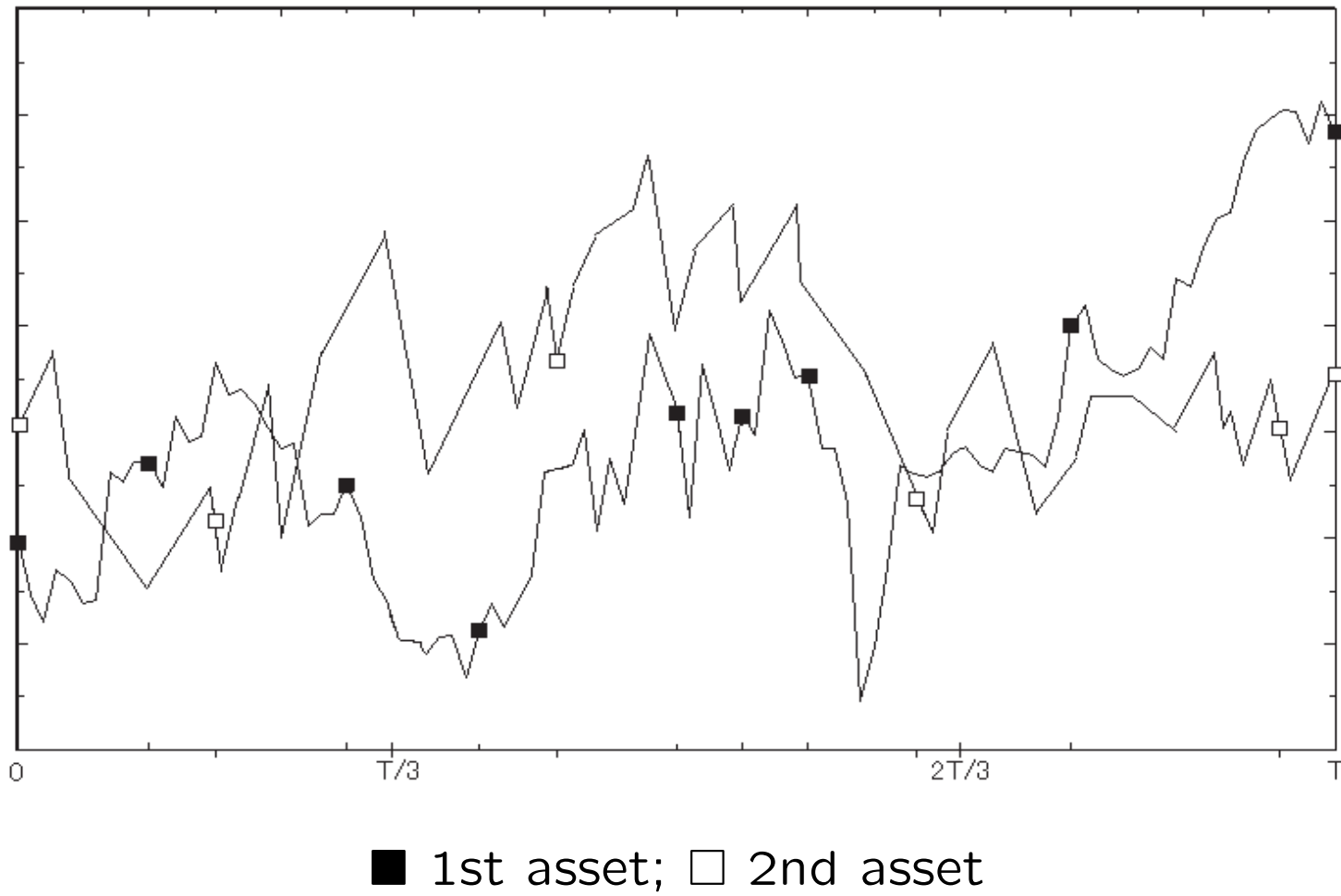
Weight matrix for the linear interpolation ( $i = j$ ):

$$W = \begin{pmatrix} 1 & 1 & \alpha_1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & \alpha_1 & 0 & 0 & 0 & 0 & 0 \\ \alpha_1 & \alpha_1 & \alpha_1^2 + \beta_1^2 & \beta_1 & \beta_1 & \beta_1 & \beta_1 \alpha_2 & 0 \\ 0 & 0 & \beta_1 & 1 & 1 & 1 & \alpha_2 & 0 \\ 0 & 0 & \beta_1 & 1 & 1 & 1 & \alpha_2 & 0 \\ 0 & 0 & \beta_1 & 1 & 1 & 1 & \alpha_2 & 0 \\ 0 & 0 & \beta_1 \alpha_2 & \alpha_2 & \alpha_2 & \alpha_2 & \alpha_2^2 + \beta_2^2 & \beta_2 \\ 0 & 0 & 0 & 0 & 0 & 0 & \beta_2 & 1 \end{pmatrix}$$

Raw data realized volatility:  $\sum_{k=1}^{N_i} \Delta p_i(t_k^i)^2 = \Delta p_i' \Delta p_i$

$$W = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

# Nonsynchronous observations



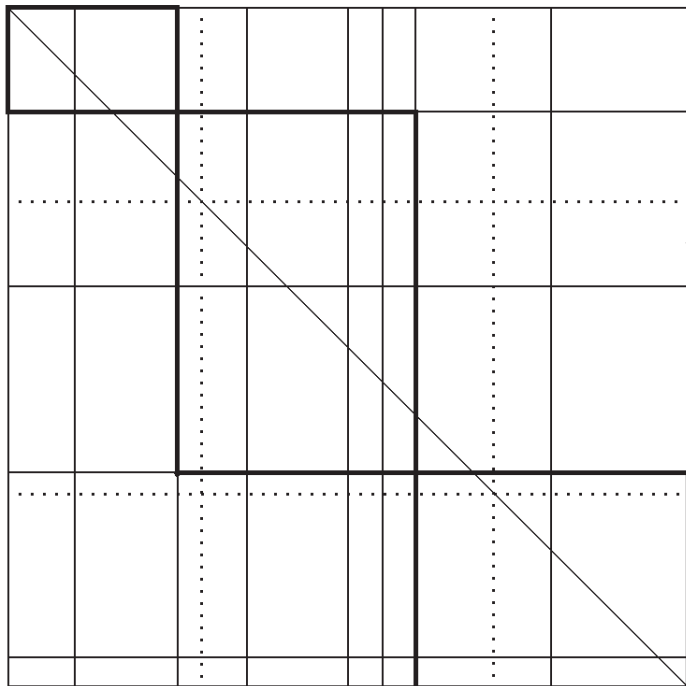
Raw data realized cross  
volatility


$$\hat{\omega}_{ij}^R = \sum_{k=1}^{N_i} \sum_{l=1}^{N_j} \Delta p_i(t_k^i) \Delta p_j(t_l^j) I(A)$$

where  $A = \{(t_k^i, t_{k-1}^i) \cap (t_l^j, t_{l-1}^j) \neq \emptyset\}$ .

$$W = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Previous tick interpolation and realized cross volatility



$$\int_{t_2}^{t_1^2} \omega_{21}(t) dt + \int_{t_1^6}^{t_2^3} \omega_{21}(t) dt$$

$$W = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

Fourier series estimator (Malliavin and Mancino, 2002):

$$\hat{\omega}_{ij}^F = \frac{\pi^2}{Q} \sum_{q=1}^Q (a_q(dp_i)a_q(dp_j) + b_q(dp_i)b_q(dp_j))$$

where

$$a_q(dp_i) = \frac{1}{\pi} \int_0^{2\pi} \cos(qt) dp_i(t), \quad b_q(dp_i) = \frac{1}{\pi} \int_0^{2\pi} \sin(qt) dp_i(t).$$

$\Rightarrow$

$$w_{kl} = \begin{cases} 1 & \text{if } t_k^i = t_l^j, \\ \frac{\sin \frac{(Q+1)(t_k^i - t_l^j)}{2} \cos \frac{Q(t_k^i - t_l^j)}{2}}{Q \sin \frac{(t_k^i - t_l^j)}{2}} & \text{otherwise} \end{cases} .$$

MSE-minimizing weight:

$$w_{kl}^A = \frac{\int_0^T \omega_{ij} dt \int_{I(k,l)} \omega_{ij} dt}{v_{kl} \{1 + \sum u_{kl}\}}.$$

where

$$v_{kl} = \left( \int_{I(k,l)} \omega_{ij} dt \right)^2 + \left( \int_{t_{k-1}}^{t_k} \omega_{ii} dt \right) \left( \int_{t_{l-1}}^{t_l} \omega_{jj} dt \right)$$
$$u_{kl} = \frac{\left( \int_{I(k,l)} \omega_{ij} dt \right)^2}{v_{kl}}.$$

- Densely-sampled/less volatile period  $\Rightarrow$  Larger weight
- Coarsely-sampled/volatile period  $\Rightarrow$  Smaller weight

A feasible estimator:

$$\int_{I(k,l)} \omega_{ij} dt \Rightarrow \Delta p_i(t_k^i) \Delta p_j(t_l^j)$$
$$\int_{t_{k-1}}^{t_k} \omega_{ii} dt \Rightarrow \left\{ \Delta p_i(t_k^i) \right\}^2$$
$$\int_0^T \omega_{ij} dt \Rightarrow \sum \Delta p_i(t_k^i) \Delta p_j(t_l^j) I(A)$$

$\Rightarrow$  Naive weight

## Monte Carlo study

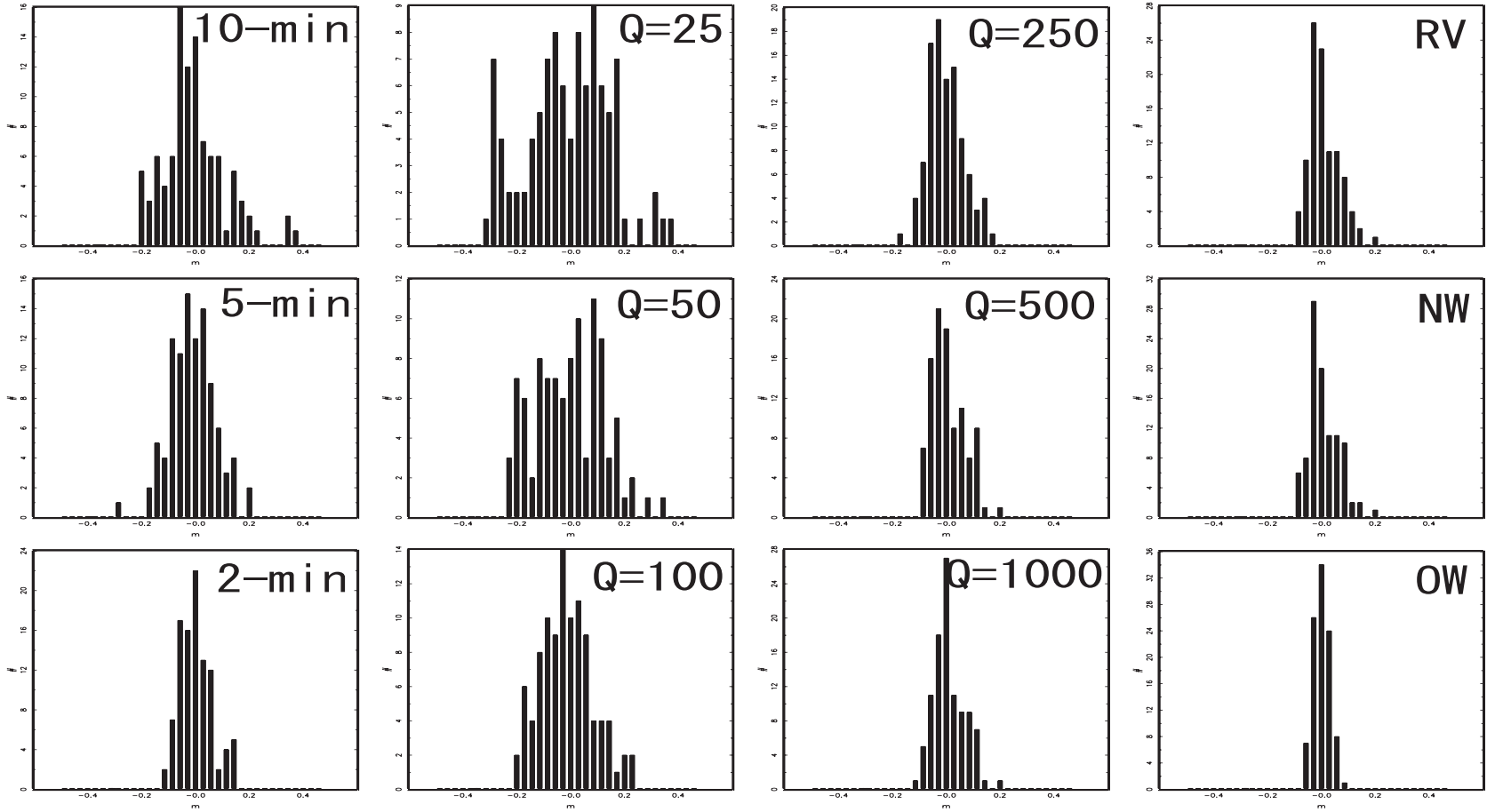
$$\begin{pmatrix} dp_1(t) \\ dp_2(t) \end{pmatrix} = \begin{pmatrix} \sigma_{11}(t) & \sigma_{12}(t) \\ \sigma_{21}(t) & \sigma_{22}(t) \end{pmatrix} \begin{pmatrix} dW_1(t) \\ dW_2(t) \end{pmatrix}, \quad 0 \leq t \leq T$$
$$d\sigma_{ij}(t) = \kappa (\theta - \sigma_{ij}(t)) dt + \gamma dW_{ij}(t), \quad i, j = 1, 2.$$

where  $\kappa = 0.01$ ,  $\theta = 0.01$ , and  $\gamma = 0.001$  and  $T = 60 \times 60 \times 24$  seconds. Time differences are drawn from an exponential distribution with mean 45 seconds for  $p_1$  and 60 seconds for  $p_2$ :

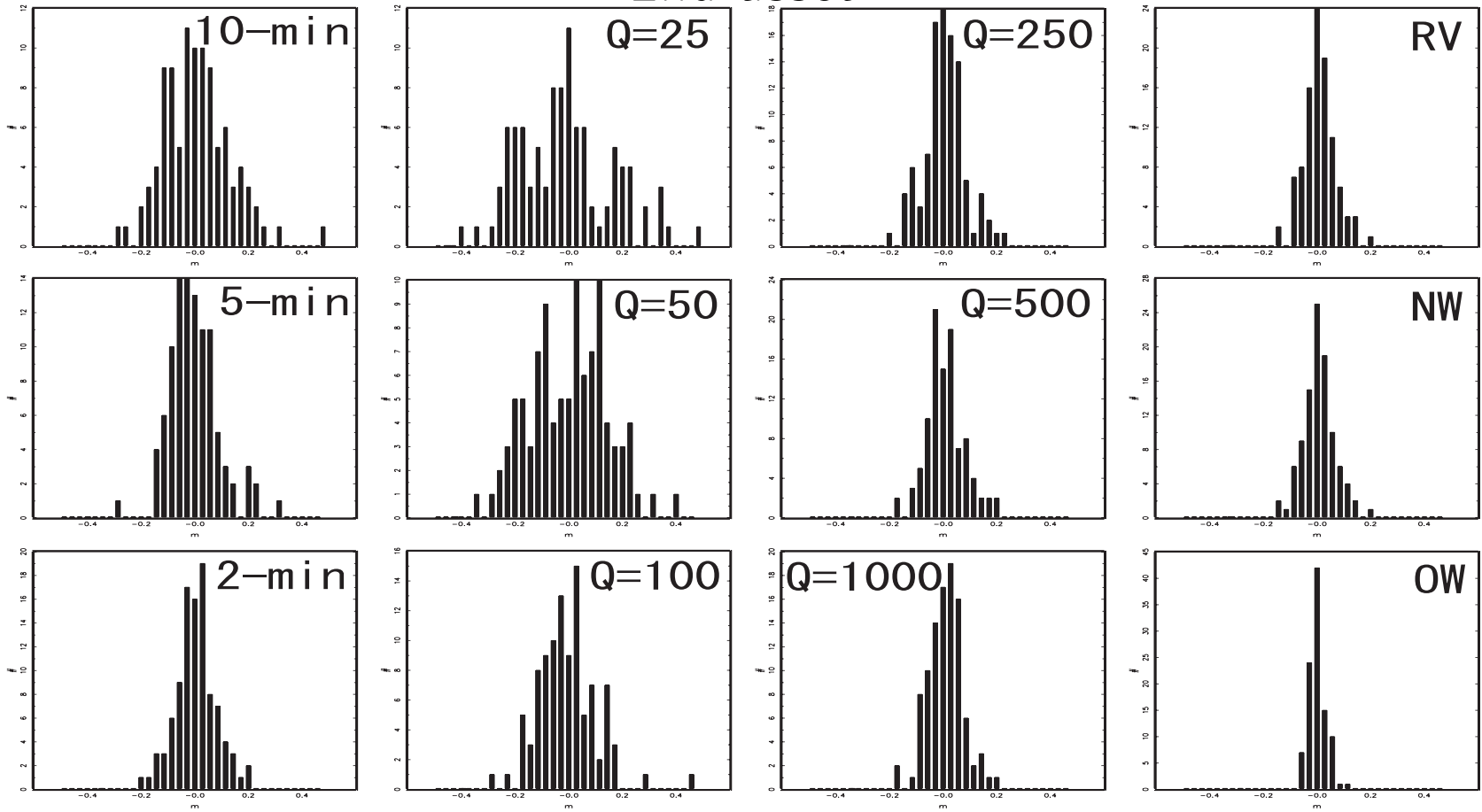
$$F(t_k^i - t_{k-1}^i) = 1 - \exp\{-\lambda_i(t_k^i - t_{k-1}^i)\}, \quad i = 1, 2$$

where  $F(\cdot)$  denotes a cumulative distribution function,  $\lambda_1 = 1/45$  and  $\lambda_2 = 1/60$ .

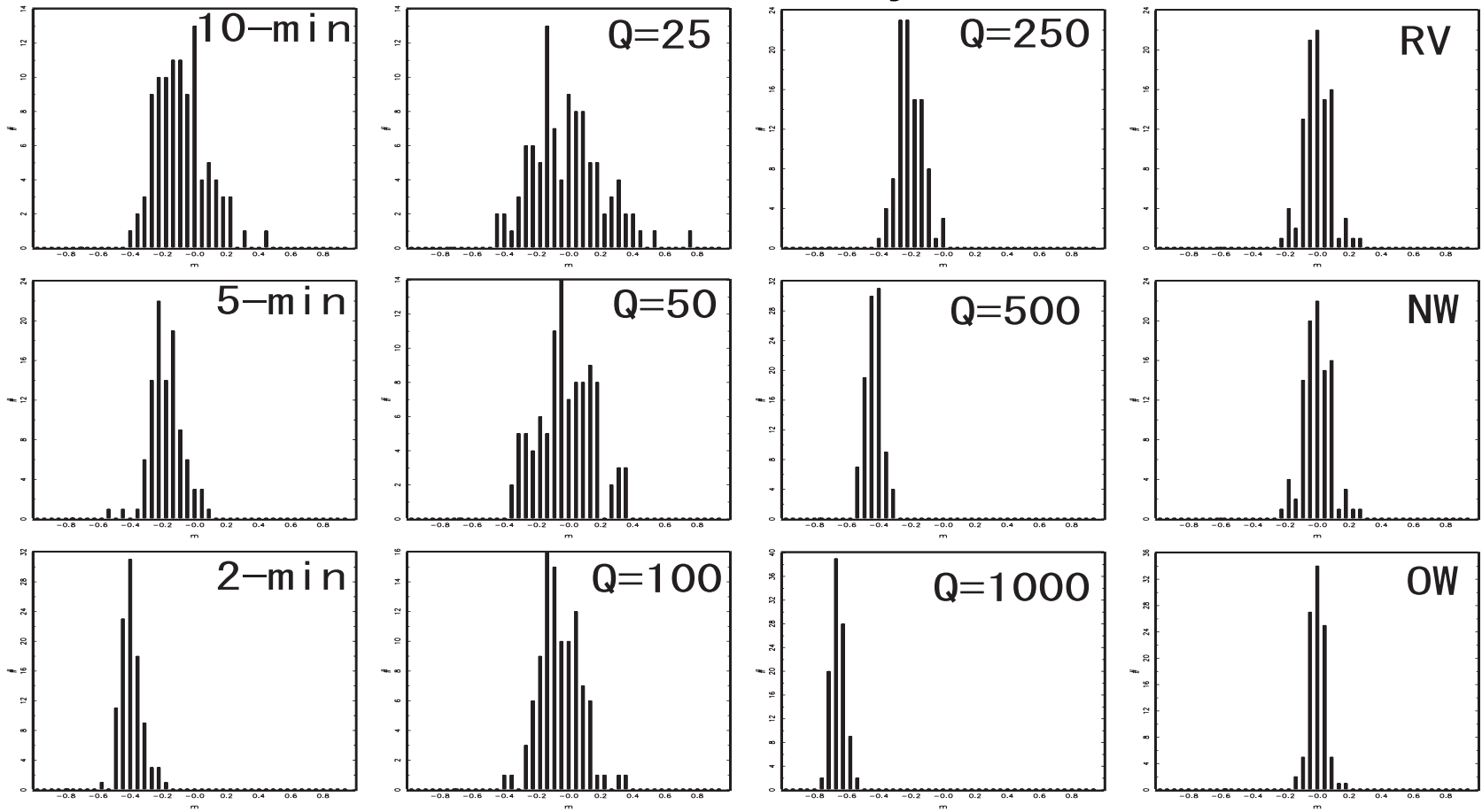
1st asset



2nd asset



cross volatility



# Summary

Contributions:

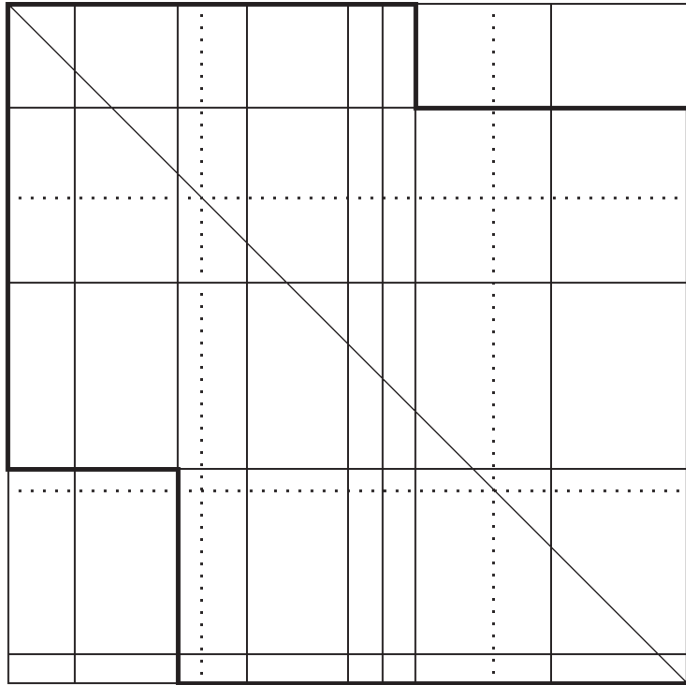
- A framework for comparing several methods
- An estimator of cross volatility

Remaining works:

- To improve feasible estimator
- Correction of interpolation bias

Thank you for your patience.

Bias corrected estimator for previous tick interpolation



Bias corrected estimator:

$$\tilde{\omega}_{ij}^P(M) = \sum_{m=2}^M \Delta^2 q_i \left( \frac{mT}{M} \right) - \sum_{m=2}^{M-1} \Delta q_i \left( \frac{mT}{M} \right)$$

$$W = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$