

A bias correction method for realized covariance calculated using previous-tick interpolation*

Taro Kanatani^{†‡}

Graduate School of Economics, Kyoto University
Yoshida Honmachi, Sakyo-Ku, Kyoto 6068501, JAPAN

Preliminary version: February 27, 2005

Abstract

In this paper we propose an unbiased estimator of cross-volatility (conditional covariance between two asset returns) when we must use evenly spaced data which have already been manipulated by previous-tick interpolation.

Keywords: Integrated cross volatility; Unevenly sampled observations; Previous tick interpolation; Bias correction

JEL Classification: C14; C32; C63

1 Introduction

1.1 Data generating process and observations

We consider n -dimensional logarithmic price $p(t) = (p_1(t), \dots, p_n(t))'$ which follows the stochastic differential equation:

$$dp(t) = \Sigma(t) dz(t), \quad 0 \leq t \leq T$$

*This research was partially supported by the Ministry of Education, Culture, Sports, Science and Technology(MEXT), Grant-in-Aid for 21st Century COE Program “Interfaces for Advanced Economic Analysis”.

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

[‡]URL: <http://www.geocities.jp/kanatanit/>

where $\Sigma(t)$ is an $n \times n$ matrix $[\sigma_{ij}(t)]_{i,j=1,\dots,n}$, and z is an $n \times 1$ vector of independent standard Brownian motions. We set the drift vector as 0, for the purpose of simplification.¹ We define the volatility matrix as

$$\Omega \equiv \Sigma\Sigma',$$

that is to say, cross volatility between i th and j th asset is denoted as the ij element of Ω :

$$\omega_{ij}(t) = \sum_{k=1}^n \sigma_{ik}(t) \sigma_{jk}(t).$$

Each i th asset price is observed at irregular time points $\{t_k^i\}_{k=0}^{N_i}$.² We just impose the assumption on the observation points that the time intervals are small: $\lim_{N_i \rightarrow \infty} \sup_{j \geq 1} (t_j^i - t_{j-1}^i) = 0$. Since we concentrate on the *ex post* cross volatility measuring and do not make any hypothesis on the structure of the underlying probability space Ω , we can construct an auxiliary probability space X where we consider $\Sigma(t)$ as deterministic functions. See Malliavin and Mancino (2002). Throughout this paper, E denotes the expectation on the probability space X .

2 Previous-tick interpolation and realized volatility

The raw data which are unevenly spaced, are converted to evenly spaced data in order to apply to the usual discrete time series analysis. Dacorogna, Gençay, Müller, Olsen, and Pictet (2001) introduces some interpolation methods including *previous tick interpolation*. When constructing $M + 1$ evenly spaced data $\{q(mT/M)\}_{j=0}^M$ from $\{p_i(t_k^i)\}_{k=0}^{N_i}$, previous-tick interpolation is defined by the following formula.

$$q_i\left(\frac{mT}{M}\right) = p_i\left(\max\{t_k^i : t_k^i \leq mT/M\}\right) \quad (2.1)$$

¹This simplification is acceptable not only because it means an efficient market in financial economics, but also because, mathematically, the martingale component swamps the predictable portion over short time intervals.

²For the purpose of simplification, we set $t_0^i = 0$ and $t_{N_i}^i = T$.

where $\max A$ and $\min A$ denote maximum and minimum elements of A , respectively.

Using evenly spaced data of $\{q_i(mT/M)\}_{m=0}^M$ and $\{q_j(mT/M)\}_{m=0}^M$, the integrated cross volatility $\int_0^T \omega_{ij}(t)dt$ is measured by the following estimator,

$$\hat{\omega}_{ij}(M) = \sum_{m=1}^M \Delta q_i \left(\frac{mT}{M} \right) \Delta q_j \left(\frac{mT}{M} \right), \quad (2.2)$$

Where $\Delta q_i(mT/M) \equiv q_i(mT/M) - q_i((m-1)T/M)$. The bias of $\hat{\omega}_{ij}(M)$ is

$$\sum_{m=1}^M \int_{t_m^-}^{t_m^+} \omega_{ij}(t) dt \quad (2.3)$$

where

$$t_m^- = \min \left\{ \max \{t_k^i : t_k^i \leq mT/M\}, \max \{t_l^j : t_l^j \leq mT/M\} \right\},$$

$$t_m^+ = \max \left\{ \max \{t_k^i : t_k^i \leq mT/M\}, \max \{t_l^j : t_l^j \leq mT/M\} \right\}.$$

Notice that in the case of univariate volatility ($i = j$), for $t_m^- = t_m^+$, the realized volatility through previous tick interpolation is an unbiased estimator. The variance of $\hat{\omega}_{ij}(M)$ is

$$\sum_{A \cap B} \left(\int \omega_{ij}(t) dt \right)^2 + \sum_B \int \omega_{ii}(t) dt \int \omega_{jj}(t) dt,$$

where

$$I(k, l) = (t_{k-1}^i, t_k^i) \cap (t_{l-1}^j, t_l^j)$$

$$A = \{(k, l) | I(k, l) \neq \emptyset\}$$

$$B = \bigcup_{m=1}^M \{(k, l) | k_{m-1} < k \leq k_m, l_{m-1} < l \leq l_m\}$$

$$k_m = \arg \max_k \{t_k^i : t_k^i \leq mT/M\}$$

$$l_m = \arg \max_l \{t_l^j : t_l^j \leq mT/M\}.$$

See Kanatani (2004) for the calculation of it. We define an unbiased estimator by

$$\tilde{\omega}_{ij}(M) = \sum_{m=2}^M \Delta^2 \bar{q}_i \left(\frac{mT}{M} \right) \Delta^2 \bar{q}_j \left(\frac{mT}{M} \right) - \sum_{m=2}^{M-1} \Delta \bar{q}_i \left(\frac{mT}{M} \right) \Delta \bar{q}_j \left(\frac{mT}{M} \right). \quad (2.4)$$

where $\Delta^2 q_i(mT/M) \equiv q_i(mT/M) - q_i((m-2)T/M)$, $\{\bar{q}_i(mT/M)\}_{m=1}^M = \{q_i(mT/M) | \Delta q_i(mT/M) \Delta q_j(mT/M) \neq 0\}$. The variance of $\hat{\omega}_{ij}^P(M)$ is

$$\sum_A \left(\int \omega_{ij}(t) dt \right)^2 + \sum_{B'} \int \omega_{ii}(t) dt \int \omega_{jj}(t) dt,$$

where

$$\begin{aligned} B' &= \bigcup_{m=1}^M ((k, l) | k_{m-1} < k \leq k'_m, l_{m-1} < l \leq l'_m) \\ k'_m &= \min\{k_m : k_m > k_{m-1}\} \\ l'_m &= \min\{l_m : l_m > l_{m-1}\}. \end{aligned}$$

Since $(A \cap B) \subset A$ and $B \subset B'$, it is obvious that $V(\hat{\omega}_{ij}(M)) < V(\tilde{\omega}_{ij}(M))$.

3 Monte Carlo study

We examine the above theory through a Monte Carlo study. Without loss of generality, we set the number of assets as two. We follow the Monte Carlo design of Barucci and Renò (2002) with some modification for multivariate setting: we generate proxy for continuous observations by discretizing following stochastic differential equations with a time step of one second,

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

where $\kappa_{ij} = 0.01$, $\theta_{ij} = 0.01$, and $\gamma_{ij} = 0.001$ for any i, j 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 :³

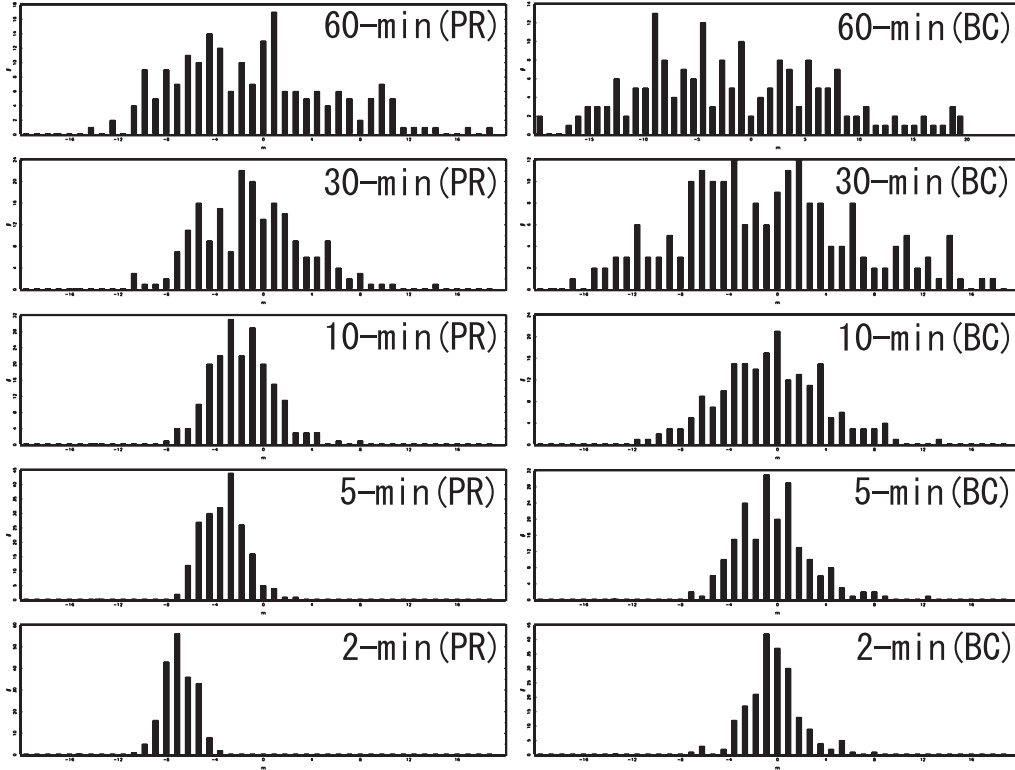
$$F(t_k^i - t_{k-1}^i) = 1 - \exp\{-\lambda_i (t_k^i - t_{k-1}^i)\}, \quad i = 1, 2$$

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

³Of course, our method allows the duration to be correlated or autocorrelated. See Engle and Russell (1998) for an autocorrelated duration model.

We compared the performances of realized volatility $\hat{\omega}_{ij}(M)$ and $\tilde{\omega}_{ij}(M)$. In calculations of the realized volatility of $\hat{\omega}_{ij}(M)$ and $\tilde{\omega}_{ij}(M)$, we set $M = 24, 48, 144, 288$, and 720 , corresponding to so-called daily realized volatility based on 60-min, 30-min, 10-min, 5-min and 2-min returns. We performed 300 replications.

Figure 1: Distribution of errors



Note: 60-min(PR): $\hat{\omega}_{12}(24)$; 30-min(PR): $\hat{\omega}_{12}(48)$; 10-min(PR): $\hat{\omega}_{12}(144)$; 5-min(PR): $\hat{\omega}_{12}(288)$; 2-min(PR): $\hat{\omega}_{12}(720)$; 60-min(BC): $\tilde{\omega}_{12}(24)$; 30-min(BC): $\tilde{\omega}_{12}(48)$; 10-min(BC): $\tilde{\omega}_{12}(144)$; 5-min(BC): $\tilde{\omega}_{12}(288)$; 2-min(BC): $\tilde{\omega}_{12}(720)$; The distribution is computed with 300 ‘daily’ replications.

Figure 1 shows the distribution of errors of $\hat{\omega}_{ij}(M)$ and $\tilde{\omega}_{ij}(M)$:

$$\hat{\omega}_{12}(M) - \int_0^T \omega_{12}(t)dt, \quad \text{and,} \quad \tilde{\omega}_{12}(M) - \int_0^T \omega_{12}(t)dt,$$

respectively.

Table 1: Sample MSE from 300 ‘daily’ replications

	Sample MSE		Estimated MSE	
	$\hat{\omega}_{12}(M)$	$\tilde{\omega}_{12}(M)$	$\hat{\omega}_{12}(M)$	$\tilde{\omega}_{12}(M)$
60 min	41.303275 (-0.78504928)	129.89687 (-0.037288398)	41.754553 (-0.74776088)	130.61587
30 min	19.535687 (-0.86612084)	58.910979 (-0.53913560)	19.113176 (-0.32698524)	58.579579
10 min	9.5904564 (-1.7242417)	19.267822 (-0.51941316)	8.3008131 (-1.2048285)	19.129370
5 min	13.820082 (-3.2669581)	9.6157110 (-0.28829981)	12.080055 (-2.9786583)	9.8853308
2 min	49.961383 (-6.9548335)	5.0706777 (-0.29348194)	46.045708 (-6.6613516)	5.0994614

Note: Sample biases are given in parentheses.

Table 1 reports the sample MSE and bias (in parenthesis) of $\hat{\omega}_{12}(M)$ from 300 replications:

$$\frac{1}{R} \sum_{r=1}^R \left(\hat{\omega}_{ij}^r(M) - \int_0^T \omega_{ij}^r(t) dt \right)^2 \text{ and } \frac{1}{R} \sum_{r=1}^R \left(\hat{\omega}_{ij}^r(M) - \int_0^T \omega_{ij}^r(t) dt \right),$$

where r denotes the number of replications and $R = 300$, and those of $\tilde{\omega}_{12}(M)$:

$$\frac{1}{R} \sum_{r=1}^R \left(\tilde{\omega}_{ij}^r(M) - \int_0^T \omega_{ij}^r(t) dt \right)^2 \text{ and } \frac{1}{R} \sum_{r=1}^R \left(\tilde{\omega}_{ij}^r(M) - \int_0^T \omega_{ij}^r(t) dt \right),$$

We define the estimated bias by

$$\frac{1}{R} \sum_{r=1}^R (\hat{\omega}_{12}^r(M) - \tilde{\omega}_{12}^r(M)),$$

Estimated MSEs of $\hat{\omega}_{12}(M)$ and $\tilde{\omega}_{12}(M)$ are defined by

$$\left(\frac{1}{R} \sum_{r=1}^R (\hat{\omega}_{12}^R(M) - \tilde{\omega}_{12}^r(M)) \right)^2 + \frac{1}{R} \sum_{r=1}^R \left(\hat{\omega}_{12}^r(M) - \frac{1}{R} \sum_{r=1}^R \hat{\omega}_{12}^r(M) \right)^2,$$

and

$$\frac{1}{R} \sum_{r=1}^R \left(\tilde{\omega}_{12}^r(M) - \frac{1}{R} \sum_{r=1}^R \tilde{\omega}_{12}^r(M) \right)^2,$$

respectively. Table 1 also reports the estimated MSE and bias (in parenthesis) of $\hat{\omega}_{12}(M)$ and $\tilde{\omega}_{12}(M)$ from 300 replications.

Under our simulation design, the correlation between the 1st and 2nd asset is on average positive: $\omega_{12}(t)$ varies around a positive mean of 0.0002 because

$$\omega_{12}(t) = \sigma_{11}(t)\sigma_{21}(t) + \sigma_{12}(t)\sigma_{22}(t)$$

and each σ_{ij} has the mean of 0.01. As expected from the bias (2.3), the shorter the interpolation time intervals is, the more downward biased the previous tick interpolation realized cross volatility $\hat{\omega}_{12}$ is.

References

- BARUCCI, E., AND R. RENÒ (2002): “On Measuring Volatility of Diffusion Processes with High Frequency Data,” *Economics Letters*, 74, 371–378.
- DACOROGNA, M. M., R. GENÇAY, U. MÜLLER, R. B. OLSEN, AND O. V. PICTET (2001): *An Introduction to High-Frequency Finance*. Academic Press.
- ENGLE, R. F., AND J. R. RUSSELL (1998): “Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data,” *Econometrica*, 66, 1127–1162.
- KANATANI, T. (2004): “Optimally Weighted Realized Volatility,” *CAEA Discussion Paper*, (26).
- MALLIAVIN, P., AND M. E. MANCINO (2002): “Fourier Series Method for Measurement of Multivariate Volatilities,” *Finance and Stochastics*, 6, 49–61.