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: October 9, 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

*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/` (The latest version is available here.)

JEL Classification: C14; C32; C63

1 Introduction

1.1 Data generating process and observations

We consider q-dimensional vector of logarithmic asset price $y^*(t)$ for $t \geq 0$. We assume that y^* is a continuous stochastic volatility semimartingle ($\mathcal{SVSM}^c$) with zero drift.[1]

$$y^*(t) = \int_0^t \Theta(u) dw(u),$$

where Θ has elements that are all cadlag and w is a vector standard Brownian motion. We set the drift vector as 0, for the purpose of simplification.[2] We define the instanteneous or spot covariance as

$$\Sigma(t) \equiv \Theta(t)\Theta(t)',$$

that is to say, cross volatility between kth and lth asset is denoted as the (k, l)th element of Σ:

$$\Sigma_{kl}(t) = \sum_{q'=1}^{q} \Omega_{kq'}(t) \Omega_{lq'}(t).$$

Each kth asset price is observed at irregular time points

$$0 = t_0^k < t_1^k < \cdots < t_j^k < \cdots$$

[1] See Barndorff-Nielsen and Shephard (2004) for the $\mathcal{SVSM}^c$.

[2] This 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.

We just impose the assumption on the observation points that the time intervals are small: $\lim_{N_i \to \infty} \sup_{j \geq 1} \left(t_j^i - t_{j-1}^i \right) = 0$.

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 co-volatility bias

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*. The previous-tick interpolation at t' is defined by the following formula.

$$x_k^*(t') = y_k^* \left(\max \left\{ t_j^k : t_j^k \leq t' \right\} \right) \tag{2.1}$$

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

Let $\hbar$ be a fixed interval of time of length. For example, we typically refer to $\hbar$ as representing a day. Then $i\hbar$ denotes the end point of the ith day or the start point of the $(i+1)$th day. We focus on the case where we construct $M+1$ evenly spaced data during each ith day. We define the mth

$\hbar/M$ return for the ith day of kth asset as

$$x_i^k(m) = x_k^*\left((i-1)\hbar + \frac{m\hbar}{M}\right) - x_k^*\left((i-1)\hbar + \frac{(m-1)\hbar}{M}\right).$$

The integrated covariance matrix $\int_{(i-1)^{\sim}}^{i^{\sim}} \Sigma(t)dt$ is measured by the realized covariation matrix

$$\widehat{\Sigma}^i(M) = \sum_{m=1}^{M} x_i(m)x_i(m)', \tag{2.2}$$

that is to say, for each element, the integrated cross volatility $\int_{(i-1)^{\sim}}^{i^{\sim}} \Sigma_{kl}(t)dt$ is measured by

$$\widehat{\Sigma}_{kl}^i(M) = \sum_{m=1}^{M} x_i^k(m)x_i^l(m). \tag{2.3}$$

The bias of $\widehat{\Sigma}_{kl}^i(M)$ is

$$\int_{I_i} \Sigma_{kl}(t)\, dt \tag{2.4}$$

where

$${}_it_m^- = \min\left\{\max\left\{t_{j'}^k : t_{j'}^k \le (i-1)\hbar + \frac{m\hbar}{M}\right\}, \max\left\{t_{j''}^l : t_{j''}^l \le (i-1)\hbar + \frac{m\hbar}{M}\right\}\right\},$$

$${}_it_m^+ = \max\left\{\max\left\{t_{j'}^k : t_{j'}^k \le (i-1)\hbar + \frac{m\hbar}{M}\right\}, \max\left\{t_{j''}^l : t_{j''}^l \le (i-1)\hbar + \frac{m\hbar}{M}\right\}\right\},$$

$$I_i = \bigcup_{m=1}^{M} \left({}_it_m^-, {}_it_m^+\right]$$

Notice that in the case of univariate volatility ($k = l$), for ${}_it_m^- = {}_it_m^+$, the realized volatility through previous tick interpolation is an unbiased estimator.

3 An unbiased realized covolatility

We define an unbiased estimator by

$$\begin{aligned}\widetilde{\Sigma}^i_{kl}(M) &= \sum_{(m',m'')\in B} x^k_i(m')x^l_i(m'') \\ &= \widehat{\Sigma}^i_{kl}(M) + \sum_{(m',m'')\in C} x^k_i(m')x^l_i(m'')\end{aligned} \tag{3.1}$$

where

$$(m)^k_i = \min\left\{m' \geq m : x^*_k\left((i-1)\hbar + \frac{m'\hbar}{M}\right) \neq x^*_k\left((i-1)\hbar + \frac{(m'-1)\hbar}{M}\right)\right\}, \tag{3.2}$$

$$B_i = \{((m)^k_i, (m)^l_i)\}^M_{m=1} \cup \{((m)_i, (m-1)_j)\}^M_{m=2} \cup \{((m-1)_i, (m)_j)\}^M_{m=2}, \tag{3.3}$$

$$C_i = \{(m', m'') \in B : m' \neq m''\}. \tag{3.4}$$

The additional part of (3.2) corrects the bias of $\widehat{\Sigma}^i_{kl}(M)$, however increases the variance of the estimator. In order to see trade-off between bias and variance, we use mean intergrated squared errors (MISEs) of the two estimators. We define MISEs of $\widehat{\Sigma}^i_{kl}(M)$ and $\widetilde{\Sigma}^i_{kl}(M)$ on $[0, n\hbar]$ as

$$\begin{aligned}\widehat{MISE}_n &= \sum_{i=1}^n \left(E(\widehat{\Sigma}^i_{kl}(M)) - \int_{(i-1)^{\sim}}^{i^{\sim}} \Sigma_{kl}(t)dt\right)^2, \\ \widetilde{MISE}_n &= \sum_{i=1}^n \left(E(\widetilde{\Sigma}^i_{kl}(M)) - \int_{(i-1)^{\sim}}^{i^{\sim}} \Sigma_{kl}(t)dt\right)^2\end{aligned}$$

respectively. Then the following theorem can work for the comparision of the MISEs.

Theorem 1 *Defining*

$$\delta_n = \widetilde{MISE}_n - \widehat{MISE}_n, \tag{3.5}$$

$$L_n = \left\{ \sum_{i=1}^{n} \sum_{(m',m'') \in C_i} x_i^k(m') x_i^l(m'') \right\}^2 - \frac{1}{2} \sum_{i=1}^{n} \sum_{(m',m'') \in C_i} \left\{ x_i^k(m') x_i^l(m'') \right\}^2, \tag{3.6}$$

$$U_n = \left\{ \sum_{i=1}^{n} \sum_{(m',m'') \in C_i} x_i^k(m') x_i^l(m'') \right\}^2 - \sum_{i=1}^{n} \sum_{(m',m'') \in C_i} \left\{ x_i^k(m') x_i^l(m'') \right\}^2, \tag{3.7}$$

then

$$P(L_n \leq \delta_n \leq U_n) \rightarrow 1 \tag{3.8}$$

as $n \rightarrow \infty$

Proof. See Appendix A ■

This theorem allows us to judge which estimator is better from actual data as follows:

$$\begin{cases} \text{use } \widehat{\Sigma}_{kl}^i(M) & \text{if } L_n > 0 \\ \text{use } \widetilde{\Sigma}_{kl}^i(M) & \text{if } U_n < 0 \\ \text{undecided} & \text{otheriwise} \end{cases}$$

4 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,

$$\begin{pmatrix} dp_1(t) \\ dp_2(t) \end{pmatrix} = \begin{pmatrix} \sigma_{11}(t) & \sigma_{12}(t) \\ \sigma_{21}(t) & \sigma_{22}(t) \end{pmatrix} \begin{pmatrix} dW_1(t) \\ dW_2(t) \end{pmatrix}, \ 0 \leq t \leq T$$

$$d\sigma_{ij}(t) = \kappa_{ij}(\theta_{ij} - \sigma_{ij}(t))\, dt + \gamma_{ij} dW_{ij}(t), \ i, j = 1, 2.$$

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]

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

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

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 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.

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

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	129.89687	41.754553	130.61587
	(-0.78504928)	(-0.037288398)	(-0.74776088)	
30 min	19.535687	58.910979	19.113176	58.579579
	(-0.86612084)	(-0.53913560)	(-0.32698524)	
10 min	9.5904564	19.267822	8.3008131	19.129370
	(-1.7242417)	(-0.51941316)	(-1.2048285)	
5 min	13.820082	9.6157110	12.080055	9.8853308
	(-3.2669581)	(-0.28829981)	(-2.9786583)	
2 min	49.961383	5.0706777	46.045708	5.0994614
	(-6.9548335)	(-0.29348194)	(-6.6613516)	

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:

$$\frac{1}{R}\sum_{r=1}^{R}\left(\hat{\omega}_{ij}^{r}(M)-\int_{0}^{T}\omega_{ij}^{r}\left(t\right)dt\right)^{2} \text{ and } \frac{1}{R}\sum_{r=1}^{R}\left(\hat{\omega}_{ij}^{r}(M)-\int_{0}^{T}\omega_{ij}^{r}\left(t\right)dt\right),$$

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

$$\frac{1}{R}\sum_{r=1}^{R}\left(\tilde{\omega}_{ij}^{r}(M)-\int_{0}^{T}\omega_{ij}^{r}\left(t\right)dt\right)^{2} \text{ and } \frac{1}{R}\sum_{r=1}^{R}\left(\tilde{\omega}_{ij}^{r}(M)-\int_{0}^{T}\omega_{ij}^{r}\left(t\right)dt\right),$$

We define the estimated bias by

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

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

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

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 (2.4), the shorter the interpolation time intervals is, the more downward biased the previous tick interpolation realized cross volatility $\hat{\omega}_{12}$ is.

5 An application for FX rate

...

6 Concluding remarks

...

A Proof of Theorem 1

Since each product of $\Delta q_i \Delta q_j$ is mutually uncorrelated, the variance of $\tilde{\omega}_{ij}(M)$ is

$$V\left(\tilde{\omega}_{ij}(M)\right) - V\left(\hat{\omega}_{ij}(M)\right) = \sum_{(m',m'')\in C} V\left(\Delta q_i\left(\frac{m'T}{M}\right)\Delta q_j\left(\frac{m''T}{M}\right)\right), \tag{A.1}$$

It is obvious that $V(\hat{\omega}_{ij}(M)) < V(\tilde{\omega}_{ij}(M))$. The variance of $\Delta q_i \Delta q_j$ is

$$\begin{aligned} & V\left(\Delta q_i\left(\frac{m'T}{M}\right)\Delta q_j\left(\frac{m''T}{M}\right)\right) \\ &= \sum_{A(m',m'')}\left(\int_{I(k,l)}\omega_{ij}(t)dt\right)^2 + \int_{t_{k-1}}^{t_k}\omega_{ii}(t)dt\int_{t_{l-1}}^{t_l}\omega_{jj}(t)dt, \end{aligned}$$

where

$$\begin{aligned} I(k,l) &= (t^i_{k-1}, t^i_k) \cap (t^j_{l-1}, t^j_l) \\ A(m',m'') &= \bigcup_{m=1}^{M}((k,l)|k_{m'-1} < k \le k_{m'}, l_{m''-1} < l \le l_{m''}) \\ k_m &= \arg\max_k\{t^i_k : t^i_k \le mT/M\} \\ l_m &= \arg\max_l\{t^j_l : t^j_l \le mT/M\}. \end{aligned}$$

See Kanatani (2004) for the calculation of it. Since

$$E\left(\Delta q_i\left(\frac{m'T}{M}\right)\Delta q_j\left(\frac{m''T}{M}\right)\right)^2$$
$$=\sum_{A(m',m'')} 2\left(\int_{I(k,l)} \omega_{ij}(t)dt\right)^2 + \int_{t_{k-1}}^{t_k} \omega_{ii}(t)dt \int_{t_{l-1}}^{t_l} \omega_{jj}(t)dt,$$

then

$$\frac{1}{2}E\left(\Delta q_i\left(\frac{m'T}{M}\right)\Delta q_j\left(\frac{m''T}{M}\right)\right)^2$$
$$\leq V\left(\Delta q_i\left(\frac{m'T}{M}\right)\Delta q_j\left(\frac{m''T}{M}\right)\right)$$
$$\leq E\left(\Delta q_i\left(\frac{m'T}{M}\right)\Delta q_j\left(\frac{m''T}{M}\right)\right)^2.$$

Since

$$\sum \frac{1}{2}E\left(\Delta q_i\left(\frac{m'T}{M}\right)\Delta q_j\left(\frac{m''T}{M}\right)\right)^2$$
$$\leq \sum V\left(\Delta q_i\left(\frac{m'T}{M}\right)\Delta q_j\left(\frac{m''T}{M}\right)\right)$$
$$\leq \sum E\left(\Delta q_i\left(\frac{m'T}{M}\right)\Delta q_j\left(\frac{m''T}{M}\right)\right)^2$$

using

$$\sum_{(m',m'')\in C} \Delta q_i\left(\frac{m'T}{M}\right)^2 \Delta q_j\left(\frac{m''T}{M}\right)^2, \tag{A.2}$$

as an estimate of

$$\sum_{(m',m'')\in C} E\left(\Delta q_i\left(\frac{m'T}{M}\right)\Delta q_j\left(\frac{m''T}{M}\right)\right)^2$$

we can estimate lower and upper bound of $V(\tilde{\omega}_{ij}(M)) - V(\hat{\omega}_{ij}(M))$.

References

Barndorff-Nielsen, O. E., and N. Shephard (2004): "Econometric Analysis of Realized Covariation: High Frequency Based Covariance, Regression, and Correlation in Financial Economics," *Econometrica*, 72, 885–925.

Barucci, E., and R. Renò (2002): "On Measuring Volatility of Diffusion Processes with High Frequency Data," *Economics Letters*, 74, 371–378.

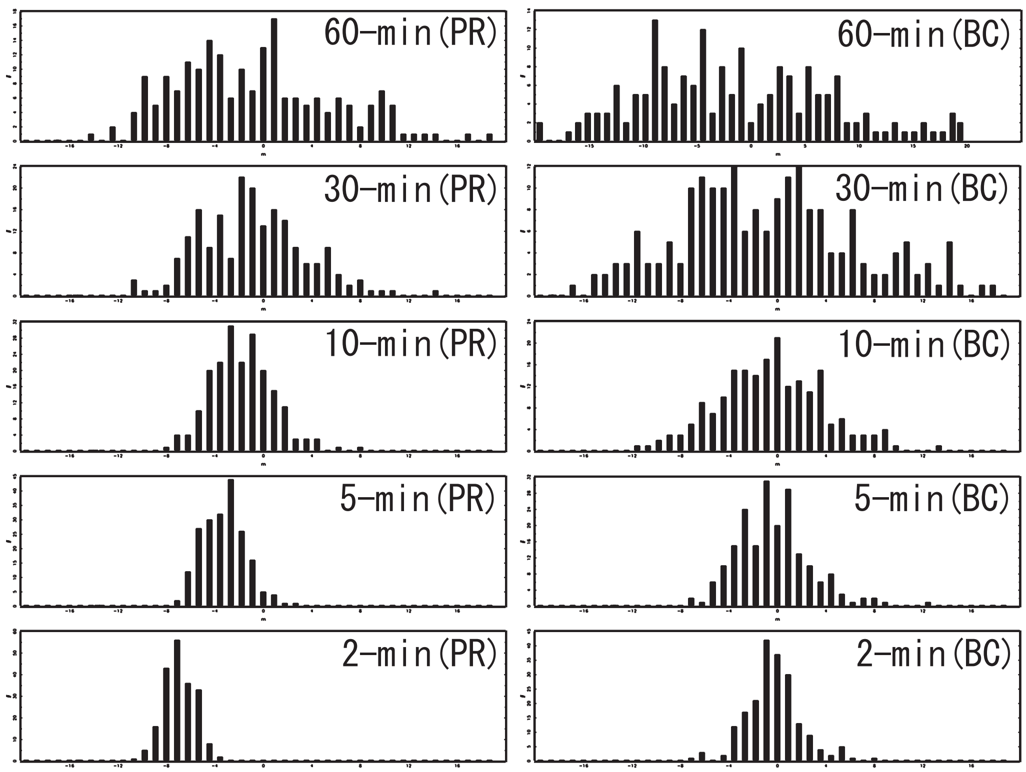
Dacorogna, M. M., R. Gençay, U. Müller, R. B. Olsen, and O. V. Pictet (2001): *An Introduction to High-Frequency Finance*. Academic Press.

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.

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.